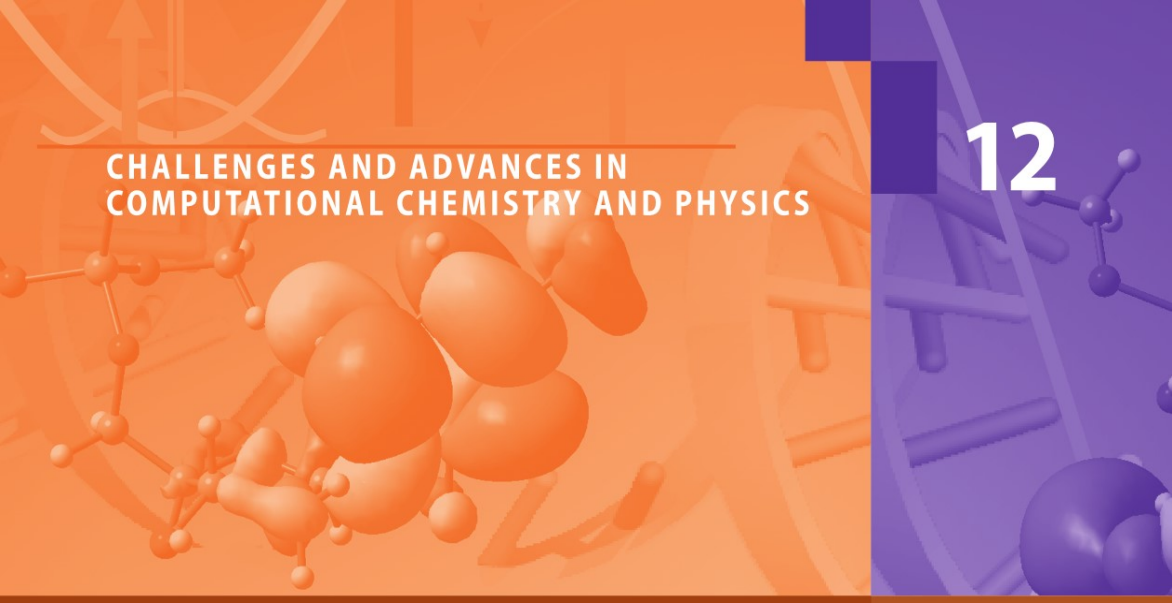


CHALLENGES AND ADVANCES IN COMPUTATIONAL CHEMISTRY AND PHYSICS

12

Series Editor J. Leszczynski

Volume Editors P. Paneth · A. Dybala-Defratyka

Kinetics and Dynamics

From Nano- to Bio-Scale

Springer

Kinetics and Dynamics

CHALLENGES AND ADVANCES IN COMPUTATIONAL CHEMISTRY AND PHYSICS

Volume 12

Series Editor:

JERZY LESZCZYNSKI
Department of Chemistry, Jackson State University, U.S.A.

For further volumes:
http://www.springer.com/series/6918

Kinetics and Dynamics

From Nano- to Bio-Scale

Edited by

Piotr Paneth
Technical University of Lodz, Poland

Agnieszka Dybala-Defratyka
Technical University of Lodz, Poland

Springer

Editors
Prof. Piotr Paneth
Technical University of Lodz
Inst. Applied Radiation
Chemistry
Faculty of Chemistry
Zeromskiego 116
90-924 Lodz
Poland
paneth@p.lodz.pl

Dr. Agnieszka Dybala-Defratyka
Technical University of Lodz
Inst. Applied Radiation
Chemistry
Faculty of Chemistry
Zeromskiego 116
90-924 Lodz
Poland
Agnieszka.Dybala-Defratyka@p.lodz.pl

ISBN 978-90-481-3033-7 e-ISBN 978-90-481-3034-4
DOI 10.1007/978-90-481-3034-4
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2010932677

Cover design: WMXDesign GmbH

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

PREFACE

Advances in computer hardware together with recent enormous progress in linear scaling algorithms and developments in theory make molecular modeling attractive research tool for various fields starting from gas-phase reactivity of simple chemical molecules, through nanochemistry, to enzymology and pharmacology. Studies of kinetics and dynamics frequently involve modeling of the short-lived intermediates and transition states. These studies are conceptually demanding and are still stirring the scientific community. A compilation of reviews from different but related fields collected together in this volume can facilitate the knowledge transfer and interdisciplinary cooperation. Molecular modeling of dynamic processes of systems of increasing complexity is reviewed by experts in computational chemistry, wherever possible in collaboration with experimentalists. The first five chapters are devoted to progress in studies of gas phase reactions as well as catalysis and reactivity of the organic reactions. They also present novel approaches allowing to study dynamics of chemical reactions. In the following chapters dynamics of more complex systems like DNA constituents or polymerization reactions and phenomena such as charge transfer in biological systems are presented. The remaining, about half of the book will addresses kinetics of biosystems, including mechanisms of enzymatic reactions, whole protein motions concluding with tunneling contribution to the reaction dynamics.

The presented volume starts with very thorough chapter devoted to Ca^{2+} reactivity by Manuel Yáñez and his coworkers. Based on the most up-to-date knowledge from both experimental and theoretical studies the most fundamental issues regarding this important and common metal ion behavior are discussed. Ca^{2+} binding energies as well as created basicity scale based on the gas-phase DFT calculations deserve a special attention. It is followed by the contribution from Kyrychenko and Waluk that presents recent developments in computational modeling of hydrogen-bonding induced phenomena and excited-state properties in a series of biologically relevant bifunctional proton donor–acceptor heteroazaaromatic compounds. The authors show how the structure and dynamics of these complexes change gradually going from gas-phase calculations, via solvation simulation to the computations where lipid membrane environment is included. It is a very nice example of using MD simulations to explore hydrogen-bonding modes and the factors controlling their photochemistry and photophysics in the studied molecules of biological importance. The third chapter by Maksić and coworkers constitutes a nice continuation of the previous contribution as it also aims at describing an ultrafast phenomena in photochemistry and photobiology by the use of dynamics simulations. The authors present an overview of computational methods adequate for studying fast photodissociation process, among which a recently developed hybrid nonadiabatic photodynamics QM/MM method seems to be a promising tool for approaching such problems. In the following chapter the

predictive power of quantum mechanical methods in designing catalysts for asymmetric applications as exemplified with an organocatalyzed aldol reaction and a sulfur-ylide promoted aziridination is presented. The next chapter takes the readers into the world of chemical reactions and their representations using empirical force fields. The importance of the dynamics of bond-breaking and bond-forming processes is emphasized and the application of four different approaches is illustrated. Two novel ones, Molecular Mechanics for Proton Transfer (MMPT) and Adiabatic Reactive Molecular Dynamics (ARMD) are tested against several systems including the large biological ones. This chapter closes the first section devoted to the reactivity of small systems and available approaches capable of exploring dynamics and kinetics of reaction they take part in. The proceeding chapter opens up another section focused on more complex systems as it touches the molecular modeling of polymerization reactions. The authors present limitations of the available methods and give an overview of problems which may arise while using various levels of approximation. In the following chapter, Leszczynski and coworkers provide an important review on ab initio based kinetics simulation technique and show its performance on the example of the proton transfer reaction in DNA bases, nucleotides and their complexes with metal ions in the presence of water molecules. DNA theme is further explored in the chapter by Kubaň and Elstner. It is devoted to the charge transfer (CT) in DNA, the computational framework developed for treating the main factors influencing it in this important biological system is presented. The CT process is also the focus of attention in the chapter by Cavasotto and Anisomov. The authors present semiempirical linear scaling quantum-mechanical methods and their application to studies of CT using quantum mechanical molecular dynamics simulations in explicit water. The first chapter of the second half of the presented volume introduces the readers to the world of enzymatic systems by presenting the quantum theory of enzymology with the special emphasis on the role of the transition state structures. In the proceeding account the authors (Ramos and coworkers) review MD simulations in the study of metalloproteins. The most common obstacles and the available methods helping to solve the problems are presented. The next chapter by Liu provides a nice overview of QM/MM methodology and its application on enzymatic catalysis. It is followed by the work of Tuñón and coworkers in which on the example of lactate dehydrogenase the evolution of techniques and methods allowing for simulation of chemical reaction catalyzed by an enzyme is shown. This chapter constitutes also a nice illustration of possible shortcomings of all approaches chosen so far to study this particular enzymatic system. The topic of the role of protein dynamics and its impact on the kinetics and catalysis is further explored in the next chapter by McGeagh and Mulholland. Several examples of reactions catalyzed by enzymes are presented along with the applied QM/MM schemes.

In the subsequent chapter Cao and coworkers review their computational approach for studying transport mechanism in the membrane protein and show how theoretical studies can constitute a complementary tool towards deeper understanding of a biological problem. Another example of how MD methods can support experimental studies is presented in the next chapter by Liedl and

coworkers. The authors illustrate the applicability of different simulation techniques in drug discovery process. The following chapter is an introductory review for contributions focused on tunneling phenomena in enzymes. The contribution by Siebrand and Smedarchina provides a new model – analysis tool allowing to determine, based on the observed kinetic isotope effect and its temperature dependence, whether the tunneling process occurs via adiabatic or non-adiabiatic mechanism. The theoretical background of quantum effects is further discussed in the subsequent chapter by Meana-Pañeda and Fernández-Ramos. Using Variational Transition State Theory the authors propose an alternative approach for taking into account large curvature tunneling mechanism, especially for large systems. Enzymatic H-tunneling reactions and the role of so-called promoting vibrations are reviewed in the last chapter of this volume by Scrutton and coworkers. The effect of the fluctuations coupled to the reaction coordinate in enzyme kinetics is illustrated by several biological H-transfers.

The presented studies do not only provide very useful information regarding the most recent achievements in computational chemistry and show the usefulness and applicability of the employed approaches but they also outline the major limitations of current simulations and draw possible future perspectives. There is a constant need for force fields improvement and development in order to guarantee the reliable models for simulating target molecules. The existing computational tools and theoretical frameworks require an ongoing feedback from the experiment for better understanding of factors governing dynamics and kinetics and the link between them in many systems, especially enzymes. And not only classical methods suffer from certain shortcomings, also quantum approaches necessary to study phenomena which explanation is only available by using higher level theories need more accurate and at the same time reasonable models with respect to the complexity and the cost of computations. Despite of listed limitations the field of molecular modeling is advancing at an easy to notice and acknowledge speed and its impact on the current knowledge status in such areas as photochemistry, material science, enzymology, or biophysics is growing and its role will definitely continue to be crucial in resolving many existing and long-lived puzzles.

Lodz
March 2010

Agnieszka Dybala-Defratyka
Piotr Paneth

CONTENTS

CHAPTER 1

Ca^{2+} REACTIVITY IN THE GAS PHASE. BONDING, CATALYTIC EFFECTS AND COULOMB EXPLOSIONS

INÉS CORRAL[1], CRISTINA TRUJILLO[1], JEAN-YVES SALPIN[2,3], AND MANUEL YÁÑEZ[1]

[1]*Departamento de Química, C-9. Universidad Autónoma de Madrid, Cantoblanco 28049, Madrid, Spain*
e-mail: manuel.yanez@uam.es

[2]*CNRS – UMR 8587, Laboratoire d'Analyse et Modélisation pour la Biologie et l'Environnement – Bâtiment Maupertuis, Boulevard François Mitterrand 91025, Evry, France*

[3]*Université d'Evry Val d'Essonne – Laboratoire d'Analyse et Modélisation pour la Biologie et l'Environnement – Bâtiment Maupertuis, Boulevard François Mitterrand 91025, Evry, France*

Abstract: In this chapter we present an overview of what is known today about the reactivity of Ca^{2+} with different species, both from the theoretical and the experimental points of view. We have paid particular attention to the bonding characteristics of Ca^{2+} complexes and to the electron density re-distribution undergone by the neutral interacting with this doubly charged metal, due to the quite intense electric field around it, which results sometimes in, dramatic polarization effects. Another important question is related with the stability of Ca^{2+} complexes in the gas phase, in clear contrast with the instability of analogous complexes involving other metal dications, such as Cu^{2+} or Pb^{2+}, which only produce singly charged species when interacting with neutral bases. The last part of this chapter will be devoted to review the different combined experimental and theoretical studies of several reactions involving Ca^{2+}. This survey will include a comparison of the reactivity exhibited by Ca^{2+} with respect to the reactivity shown by other divalent metal ions, as well as some preliminary results of the reaction of Ca^{2+} with uracil. Particular attention will be also paid to the catalytic effect of Ca^{2+} in the tautomeric process of these kinds of biochemical compounds.

Keywords: Alkaline-earth dications, Coulomb explosions, Bonding, Reactivity, Mass spectrometry, Theoretical calculations, Catalytic effects

P. Paneth and A. Dybala-Defratyka (eds.), *Kinetics and Dynamics*, Challenges and Advances in Computational Chemistry and Physics 12,
DOI 10.1007/978-90-481-3034-4_1,

1.1. INTRODUCTION

Doubly charged metal ions play an important role in biochemistry, chemical biology and biophysics [1–11]. The formation of polypeptides from amino acids or the thermal denaturation of DNA are enhanced by the presence of divalent metal ions. It has been also well established, that the association of transition metal cations to DNA bases leads to an increase of the melting temperature of the DNA double helix [12]. Among doubly charged metal ions, alkaline-earth metal dications (Mg^{2+}, Ca^{2+}, Ba^{2+}) deserve particular attention, because they are very often found in the physiological media. It has been proved, for example, that when interacting with Watson–Crick base pairs they contribute to increase the base pair dissociation energy [3], and favor proton transfer processes between guanine and cytosine. Ca^{2+} also regulates muscle contraction [13, 14] and stabilizes DNA triplexes [15–18], rybozimes and interprotein complexes. It has also been shown [19, 20] that calcium ions binding increases the protein thermal stability by enhancing the alpha helix content.

Although it seems well established that when interacting with DNA, the association to the phosphate group is largely favored [3, 21, 22], a lot of attention was also paid, essentially from a theoretical point of view, to the complexes Ca^{2+} may form with different DNA components [23–32]. Since DNA–metal cation interactions take place in the biochemical media, where water is ubiquitous, much work has been done, both experimentally and theoretically on the solvation of alkaline-earth dications [33–40]. Also, some theoretical studies dealt with the interaction of hydrated metal dications with nucleotides and other biochemical systems [29, 32, 41–47]. A parallel experimental information is however, practically inexistent. It must be taken into account that, in general, the information on the specific interactions between a metal ion and a base are of a double origin, they come either from theoretical calculations or from mass spectrometry studies. This actually explains the scarcity of experimental studies involving doubly charged metals due to the difficulty to generate stable doubly charged species in the gas phase. On the other hand, a great majority of the theoretical investigations were focused on the elucidation of the structures of the complexes formed and/or on the estimate of the corresponding binding energies, but very little on reactivity.

In this chapter we will present an overview of what is known today about the reactivity of Ca^{2+} with different species, both from the theoretical and the experimental points of view. Probably, one of the most fundamental questions as far as Ca^{2+} reactivity is concerned, is why it yields stable complexes when interacting with different organic bases [48–53], when other doubly charged metal ions such as Cu^{2+} or Pb^{2+} only produce singly charged species by the spontaneous loss of a proton [54–56].

It is quite obvious that the answer to these and other questions concerning the reactivity of Ca^{2+} are directly related with the electron density re-distribution undergone by the base, triggered by the quite intense electric field created by the

doubly charged ion, which results in sometimes, dramatic, polarization effects. Hence, we will pay also some attention to the bonding characteristics of Ca^{2+} complexes, as well as to the calculated binding energies. It has been shown previously in the literature that already some simple bases, such as ammonia and formaldehyde, have a basicity with respect to Ca^{2+} opposite to that exhibited in normal protonation processes [57]. In this chapter we will present, for the first time, a Ca^{2+} basicity scale including more than 60 different bases, whose basic sites are both first- and second-row atoms.

The last part of this chapter will be devoted to review the different combined experimental and theoretical studies of different reactions involving Ca^{2+}. This survey will include a comparison of the reactivity exhibited by Ca^{2+} with respect to the reactivity shown by other divalent metal ions, as well as some preliminary results of the reaction of Ca^{2+} with uracil.

1.2. STRUCTURE AND BONDING OF Ca^{2+} COMPLEXES

As mentioned above, the intrinsic reactivity of Ca^{2+} is dictated by the perturbation that this doubly charged metal ion produces on the electron density of the base with which it interacts. Hence, this is one of the first points to be discussed in this chapter. No matter the nature of the basic site, it has been systematically found that the interactions involving Ca^{2+} are essentially electrostatic in nature [23, 48, 52, 53, 58, 59]. This has been well illustrated for the particular case of the complexes formed by this metal dication with urea [48] and thiourea [50], where different theoretical approaches, namely the atoms in molecules (AIM) theory [60], the electron localization function (ELF) theory [61–63] or the natural bond orbital (NBO) method [64], all lead to the same conclusion [52].

The electron density of urea–Ca^{2+} complex (see Figure 1-1) shows the existence of a bond critical point (BCP) between oxygen and calcium, with a small electron density, typically associated with closed-shell interactions. Furthermore, the energy density [65] evaluated at the BCP is positive, which indicates that kinetic energy contributions dominate over potential energy distributions, as it is typically the case in ionic bonds.

Similarly, the ELF [62, 63, 66] of the urea– and thiourea–Ca^{2+} complexes does not show any disynaptic basin in which Ca and O or S participates, clearly showing the absence of any covalent interaction between both atomic systems. Concomitantly, the NBO approach describes these complexes as the interaction of two separated subunits, namely Ca and the base. Also importantly, the net natural charge on the metal within the complex is very close to that in the isolated doubly charged metal ion (typically ca. +1.98), which indicates the absence of a significant electron exchange between both interacting subunits.

As a consequence of the dominant contribution of electrostatic and polarization terms in the Ca^{2+}-complexes, they resemble closely those formed by alkali metal ions. Hence, in general, alkali metal ions and Ca^{2+} have a propensity to form

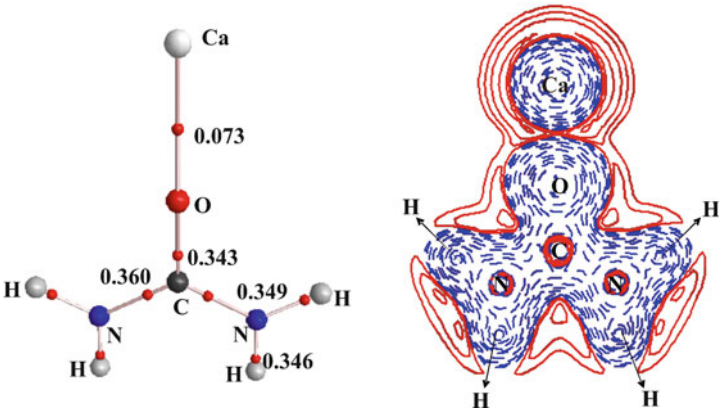


Figure 1-1. Molecular graph and two-dimensional plot of the energy density for urea–Ca^{2+} complex. In the molecular graph the red dots correspond to bond critical points. The electron density, in a.u. is shown. In the energy density *dashed blue lines* and *solid red lines* correspond to negative and positive values, respectively

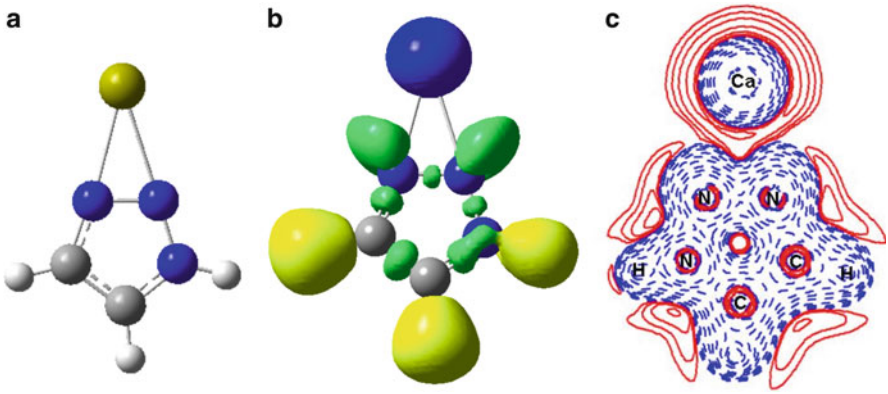


Figure 1-2. Structure (**a**), tridimensional ELF (**b**) and energy density plot for the 1,2,3-triazole-Ca^{2+} complex. In the ELF representation, *yellow lobes* denote disynaptic C–H and N–H basins. *Green lobes* correspond to disynaptic basins involving heavy atoms and *blue lobes* to atomic core monosynaptic basins. For the energy density we have used the same conventions as in Figure 1-1

chelated structures, in which the metal bridges between two basic sites, when the base has two or more basic centers physically close. This is the case for instance, when Ca^{2+} interacts with 1,2,3-triazole (see Figure 1-2a) in which Ca^{2+} bridges between two neighbor N atoms. The same kind of structures were reported for Li^+ complexes, in order to explain the enhanced basicity that some triazoles and tetrazole have with respect to Li^+ [67–69].

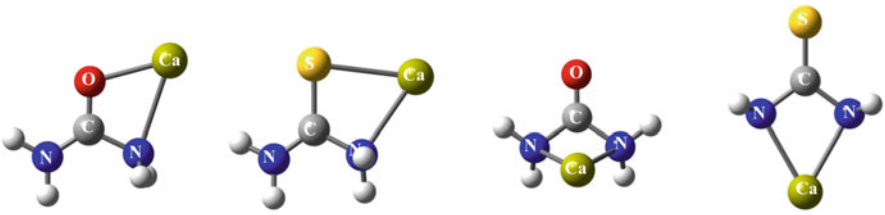
It is important to recall here that electrostatic interactions exclusively cannot account for the formation of these chelated structures, which are extra-stabilized

with respect to the linear arrangements by the enhancement of polarization effects, since, as shown by the ELF and the energy density plots (Figure 1-2b, c), when the metal bridges between both basic sites it is able to polarize both lone-pairs simultaneously. However, as explained by Alcamí et al. [70], the electrostatic interaction with the base, when the metal cation occupies a central position between both basic sites, is not different from the one it experiences in the linear arrangement, simply because the metal ion moves so far away from the nuclei of the basic sites, that it cannot be trapped by electrostatic potential minima, but moves along the isopotential curves connecting them. Hence, the higher polarization effects, when the metal is in the middle, result in an extra-stabilization of the chelated structure. This ability of Ca^{2+} to interact with several basic sites was repeatedly reported in the literature [9, 22, 24, 26, 28–31, 41, 71, 72].

It is worth noting however, that the existence of basic centers close to each other does not yield necessarily to a chelated structure. The interactions with urea and thiourea are again suitable examples [48, 50]. In these two cases, two chelated minima have been located in the potential energy surface (see Figure 1-3). In one of them the metal dication interacts simultaneously with the oxygen (sulfur) atom and with the nitrogen of one of the amino groups. In the other, it bridges between both amino groups. However, these two minima are higher in energy than the global minimum in which the metal ion shows a linear arrangement with respect to the C=O (C=S) bond (Figure 1-1).

This clearly indicates that in spite of the fact that in local minima **2** and **3**, the metal interacts with two basic sites, this requires the pyramidalization of one or two of the amino groups, whose energetic cost is higher that the enhancement in the interaction energy associated with the interaction of the metal with the second basic site. This is actually consistent with the fact that minima **3** is less stable than minima **2**, because in the former it is necessary to change the local conformation of both amino groups. It is also important to note that the energy gap between the global minimum and these chelated structures is much higher for Li^+ than for Ca^{2+} complexes, simply because the polarization of the basic site is much larger when the metal bears a net charge +2.0.

This ability of Ca^{2+} to polarize the lone pairs of two basic sites close to each other, has an important consequence when interacting with uracil and its sulfur and



2 **3**

Figure 1-3. Bidentate local minima 2 and 3 formed by the interaction of Ca^{2+} with urea and thiourea

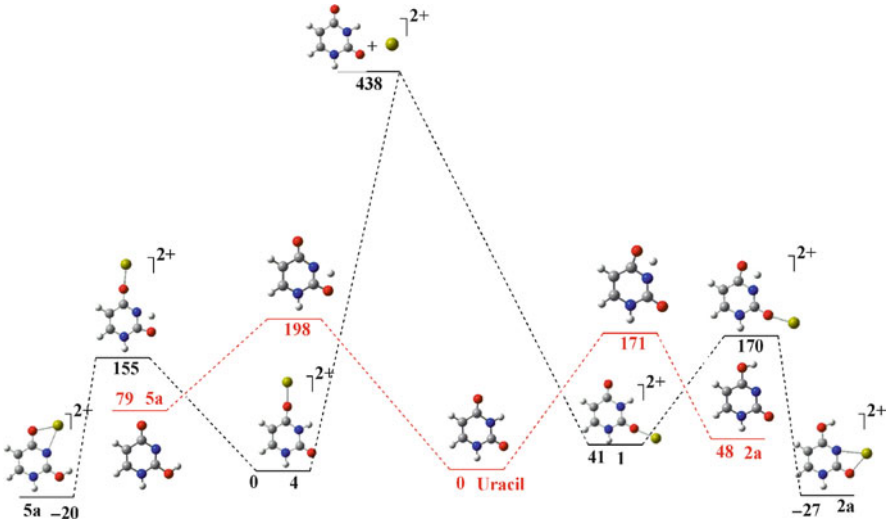


Figure 1-4. Energy profile showing the stabilizing effect of Ca^{2+} on the enolic forms of uracil. The catalytic effect of this doubly charged metal ion on the tautomeric processes connecting the diketo forms (1 and 4) with the enolic ones (2a and 5a) is apparent by comparing the barriers involved in uracil–Ca^{2+} complexes and those of the uracil isolated molecule (*in red*). All values in kJ mol^{-1}

selenium derivatives [53]. It is well established that uracil and its sulfur and selenium derivatives only exist in their keto forms in the gas phase, the corresponding enolic forms being much higher in energy [73–76], and separated from the keto forms by quite high activation barriers. However, this stability order is reversed upon Ca^{2+} association. The possibility of Ca^{2+} to interact simultaneously with two basic sites in the enolic forms leads to binding energies which are not only larger than those found for the keto forms, but bigger than the keto-enol energy gap (see Figure 1-4), so that the enolic complexes become lower in energy than the corresponding keto ones.

No less important is the fact that Ca^{2+} has a clear catalytic effect on these tautomerization process [51, 53, 77], because the polarizations induced by the metal ion contribute to significantly stabilize the transition states associated with the different hydrogen transfer mechanisms connecting the keto forms (**1** and **4**) with the enolic ones (**2a** and **5a**). Also important, as illustrated in Figure 1-4, the activation barriers in the presence of Ca^{2+} are not only smaller, but much lower in absolute value than the binding energy of the metal dication to the base, so, in principle, the enolic forms, which are not accessible for the isolated compound in the gas phase, would be accessible in presence of Ca^{2+}.

Even though, there is not a significant charge transfer between the base and the cation, the strong electric field created by the dication polarizes the electron density of the base causing the reinforcement of some of its bonds and the activation of some others. These bonding perturbations are easily detected by any of the three

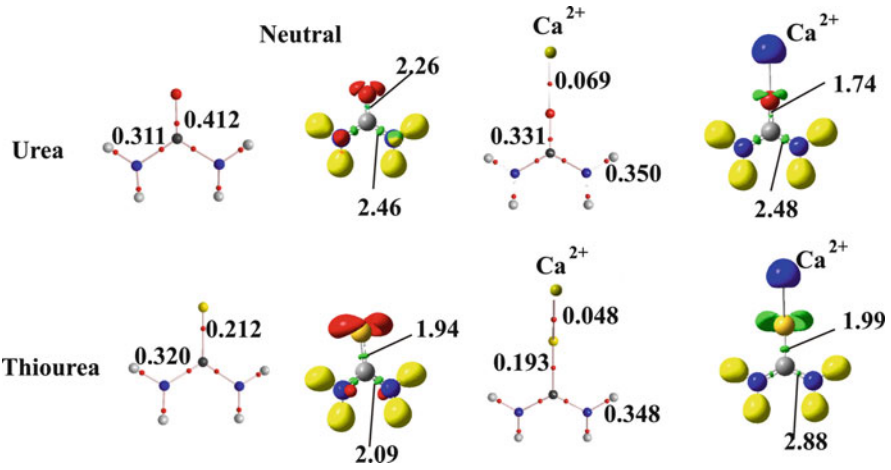


Figure 1-5. Molecular graphs and tree-dimensional plots of the ELF for urea, thiourea and their Ca^{2+} complexes. Conventions as in Figures 1-1 and 1-2. The population of the C=O, C=S and C–N disynaptic basis is shown

methods mentioned above. A comparison of the molecular graphs and the ELF of the Ca^{2+}-complexes and those of the isolated bases permits an easy visualization of these changes. As an illustration we present such a comparison in Figure 1-5 for the particular case of the complexes of urea and thiourea. The first conspicuous fact is that the interaction with the metal ion leads to a significant decrease of the electron density at the C=O (or C=S) BCPs. Why the electron density at the C=O (C=S) BCP decreases even though there is not a significant charge transfer from the base to the cation can be easily understood if one takes into account that when the complex is formed the lone-pairs of the basic center are strongly polarized towards the metal dication. The immediate consequence is an increase in the population of the corresponding monosynaptic basins (see Figure 1-5).

This increase comes from a concomitant depopulation of the C=O (C=S) bonding region. Consequently, the population of the V(C,O) (V(C,S)) disynaptic basin as well as the electron density at the C–O (C–S) BCP decrease, the bond length increases and the stretching frequency of the bond appears red-shifted, by 200 (84) cm^{-1}. In the NBO analysis this is mirrored in a decrease of the C=O bond order, which in isolated urea is 1.65, whereas in the Ca^{2+} complex comes down to 1.15. These polarization effects enhance the electronegativity of the carbon atom and contribute to transmit them to the rest of the molecule. In fact, the decrease of the electron density at the carbon atom favors the conjugation with the amino lone-pairs. This is reflected in the planarity of both groups, which in the isolated base are not planar, and in the reinforcement of the C–N bond, which shortens and whose stretching frequency is blue-shifted.

Polarization effects are also behind the subtle differences found between the structural characteristics when the basic site is an atom of the first- or of the second

or third-row. When the basic site is an atom of the first row, like oxygen in urea, the metal polarizes the two lone pairs, which are not very far apart from each other, and accordingly the urea-Ca^{2+} adduct has C_{2v} symmetry,[48] with the metal ion linearly bound to the carbonyl oxygen. Conversely, for thiourea and selenourea, where the sulfur and the selenium lone pairs are much more diffuse than those of oxygen, the interaction occurs specifically with one of them and not with the two leading to C_s complexes [50, 78], where the metal ion lies in the plane which bisects the NCN angle, the C–S–Ca and the C–Se–Ca bond angles being equal to 109.3° and 113.8°, respectively. Similar bent structures were also found for thiouracil and selenouracil derivatives [53].

Particular attention deserves the effect of polarization contributions on the calculated Ca^{2+} binding energies as well. As expected Ca^{2+} binding energies are about 2.5 times larger than Na^+ binding energies, but they are also about 1.7 times larger than Li^+ binding energies, in spite of the fact that Ca^{2+} is a much bigger ion than Li^+. In fact, this trend is a consequence of the larger charge of Ca^{2+} which counterbalances the opposite size effect. However, a quite surprising result predicted in the theoretical assessment carried out by Corral et al. [57], regarding the appropriate theoretical model to describe Ca^{2+} complexes, was that Ca^{2+} binding energies do not followed the same trends as the corresponding proton affinities. It has been found, for instance, that the Ca^{2+} binding energy to formaldehyde is larger than that to ammonia, even though, the intrinsic basicity of the latter, as measured by its gas-phase proton affinity is much larger than that of the former. When this basicity inversion was analyzed in terms of electrostatic, polarization and covalent contributions, as well as geometry distortion, it was found that although the neutral–ion interaction energy for alkali and alkaline-earth metal cations is dominated by electrostatic contributions, in many cases the correct basicity trends are only attained once polarization effects are also included in the model [57]. This is actually the case as far as Ca^{2+} binding to formaldehyde and ammonia are concerned. In some other cases, the correct basicity order is only attained when accounting for the geometry distortions caused by the strong polarization of the base. This was found, for instance, in the bonding of Ca^{2+} to HCN and methylamine. Hence, although the neutral–ion interaction energy for Ca^{2+} is dominated by electrostatic contributions, in many cases the correct basicity trends are only attained once polarization effects and/or distortion energy effects are also included in the model.

In view of the complete lack of experimental information on Ca^{2+} binding energies, we have considered it of interest to build up a Ca^{2+} basicity scale in the gas phase, based on B3LYP/cc-pWCVTZ calculations. This basicity scale includes a total of 81 different bases, whose basic sites are C, N, O, and S atoms. The results obtained for the calculated Ca^{2+} binding enthalpies are summarized in Table 1-1. The values presented include the basis set superposition error [79] corrections, which in average amounts to 8.7 kJ mol^{-1}.

Since as it has been mentioned above, an inversion of the basicity of some compounds was observed when the Ca^{2+} binding enthalpy was compared with the

Table 1-1 B3LYP/cc-pWCVTZ calculated Ca^{2+} binding enthalpies (in kJ mol^{-1}) of different bases

Carbon bases			Carbonyl derivatives		
CH_4		119	R1	R2	
CH_3CH_3		151	H	NMe_2	446
CH_2CH_2		180	H	OMe	378
CHCH		178	H	CF3	269
Nitrogen bases			H	OOH	321
NH_3		269	H	NHMe	430
NH_2Me		294	H	SMe	363
$NHMe_2$		310	H	Adamantyl	404
Me_3N		321	Me	Me	375
Aniline		382[a]	NH_2	NH_2	446
Acetamidine		340	Me	NH_2	429
Guanidine		474	CF3	NH_2	381
Pyridine		364	OH(trans)	NH_2	420
Pyrazol		353	OH(cis)	NH_2	380
Imidazol		397	F	NH_2	350
NO		113	CN	NH_2	336
N_2O		172	F	OH	298
HCN		270	CF_3	OH	304
Oxygen bases			F	F	218
H_2O		238	Cl	Cl	267
CH_3OH		274	Sulfur bases		
HOOH		263	SH_2		193
CH_3OCH_3		294	CH_3SH		244
Phenol		346	HSSH		250
pyrrol		348	CH_3SCH_3		281
CO		121	Thiocarbonyl derivatives		
CO_2		103	R1	R2	
SO_2		170	H	H	232
NO		233	H	Me	287
N_2O		97	H	NH_2	350
NO_2		176	H	F	200
Carbonyl derivatives			H	Cl	227
R1	R2		H	NMe_2	409[a]
H	H	275	H	OMe	326
H	Me	334	H	CF_3	273
H	NH_2	395	CH_3	CH_3	317
H	OH	337	NH_2	NH_2	400
H	F	244	Me	NH_2	484
H	Cl	267	CF_3	NH_2	368
H	Et	349	F	F	196
H	tBut	368	Cl	Cl	270

[a]These values do not include the BSSE correction, due to convergency problems when ghost atoms are used in these two specific cases

corresponding proton affinity (PA) we have considered it of interest to verify if a correlation exists between our calculated Ca^{2+} binding enthalpies and the experimental PAs, for those systems for which they are available [80].

As shown in Figure 1-6 there is a significant dispersion, so only a very rough correlation seems to exist between the two magnitudes. It is also worth noting that the quality of the correlation improves when the systems analysed are grouped in series of homologous compounds. Hence, the correlation between the Ca^{2+} binding enthalpies and the PAs for carbonyl derivatives shows a much smaller dispersion (see Figure 1-7), and the regression coefficient of the linear correlation between the two sets of values is $r^2 = 0.947$.

Sometimes the deviations observed have their origin in the enhanced stability of complexes with several basic centers when the reference acid is Ca^{2+}. For instance,

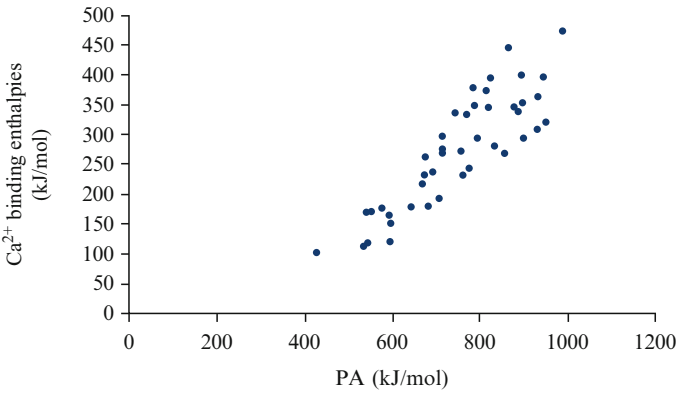


Figure 1-6. Correlation between calculated Ca^{2+} binding enthalpies and experimental proton affinities (PA) for the bases included in Table 1-1

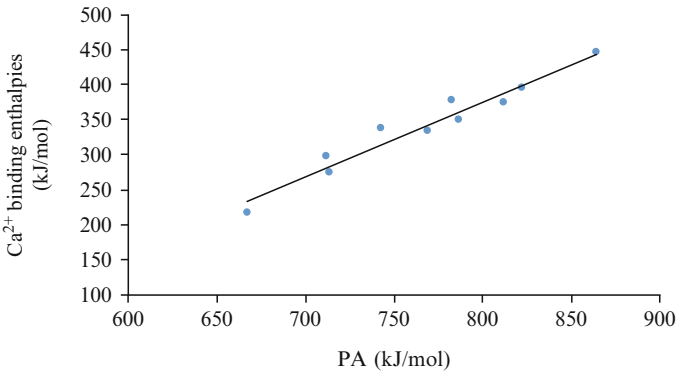


Figure 1-7. Correlation between the calculated Ca^{2+} binding enthalpies and the experimental proton affinities (PA) for the carbonyl derivatives included in Table 1-1

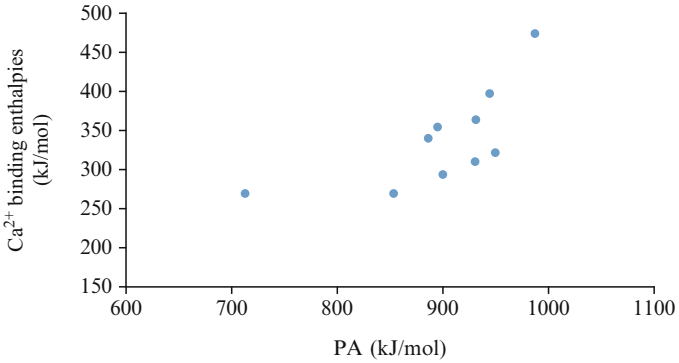


Figure 1-8. Correlation between the calculated Ca^{2+} binding enthalpies and the proton affinities (PA) of the nitrogen containing bases included in Table 1-1

on going from formaldehyde to acetaldehyde, the PA increases by 55 kJ mol^{-1}, and the Ca^{2+} binding enthalpy by 59 kJ mol^{-1}. However, whereas the basicity of formic acid is only 29 kJ mol^{-1} higher than that of formaldehyde, its Ca^{2+} binding enthalpy is 62 kJ mol^{-1} higher due to the ability of the doubly charged metal ion to polarize simultaneously, the carbonyl and the hydroxyl oxygen lone pairs.

The dispersion is also quite significant when considering different nitrogen bases, which include a N(sp), N(sp^2) and N(sp^3) atoms as basic sites (see Figure 1-8). It is interesting to note however the existence of four points perfectly aligned in the lower right corner of the figure which correspond to the series of the methyl amines, including ammonia. In this case, where the nature of the basic center remains the same, and where no chelation effects are possible, the correlation between PAs and Ca^{2+} binding enthalpies is very good, with a regression coefficient, $r^2 = 0.999$.

The number of available experimental Li^+ binding enthalpies for the bases investigated here is small, but, as expected, the correlation between these values and the calculated Ca^{2+} binding enthalpies is better than that found when the PAs are used as a reference (see Figure 1-9). This is not surprising, if one takes into account that in both cases the dominant interactions are of electrostatic origin and that for both reference acids, chelation is equally probable when the base presents two or more basic sites. There are however significant deviations. For instance, whereas the Li^+ binding energy of acetamide [81] was reported to be 16 kJ mol^{-1} smaller than that of formamide [82] for the calculated Ca^{2+} binding enthalpies is the other way around and the estimated value for acetamide is 51 kJ mol^{-1} larger than that for formamide.

1.3. STABILITY OF Ca^{2+} COMPLEXES IN THE GAS PHASE

The stability of diatomic doubly charged species has been fully discussed by Schröder and Schwarz [83]. An AB^{2+} doubly charged species is thermodynamically stable when the ionization energy (IE) of B is larger than that of A^+ (Figure 1-10a).

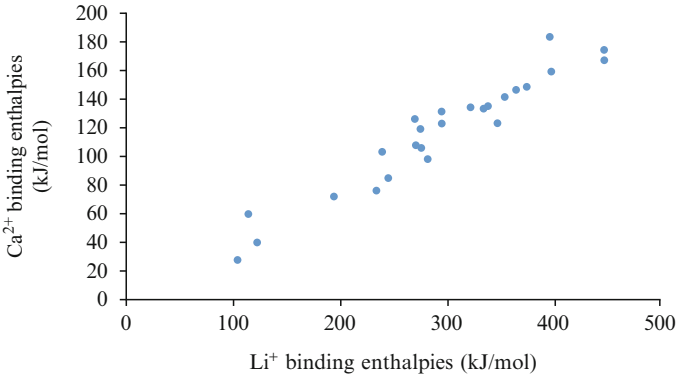


Figure 1-9. Correlation between the calculated Ca^{2+} binding enthalpies and the experimental Li^{+} binding enthalpies for the bases included in Table 1-1 for which the latter were available

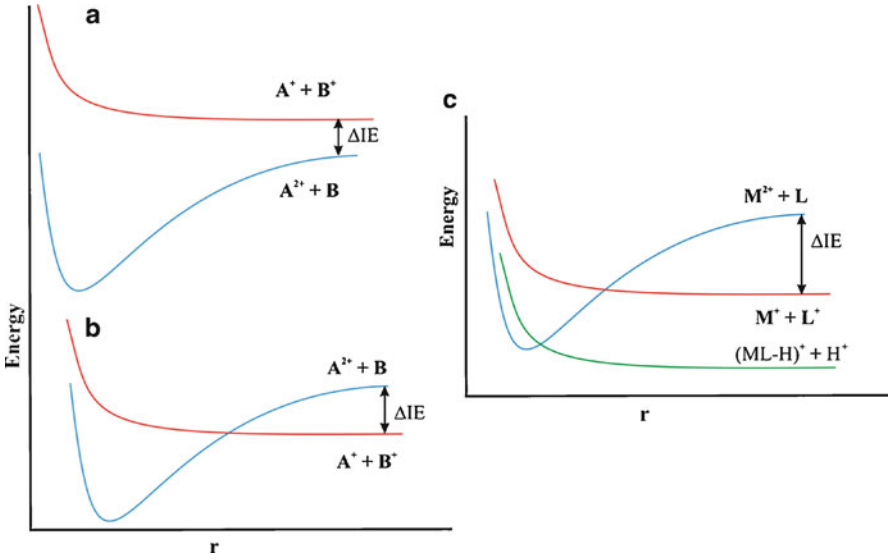


Figure 1-10. (**a**) and (**b**) Potential energy curves corresponding to thermodynamically stable AB^{2+} doubly charged diatomic systems. (**c**) Potential energy surfaces corresponding to a complex between a doubly charged metal ion (M^{2+}) and a ligand (L)

It can be also stable if the IE of B is smaller than that of A^{+} provided that the energy difference between both ionization energies is smaller than the $AB^{2+} \rightarrow A^{2+} + B$ dissociation energy (Figure 1-10b). The situation is however, a little bit more complicated when the doubly charged species is the result of the interaction of a

metal dication with an organic ligand containing hydrogen atoms, because, besides these two possible dissociation pathways, there is a third one, usually more exothermic, which corresponds to a Coulomb explosion in which a proton is lost (see Figure 1-10c). The situation shown in Figure 1-10c can still correspond to a kinetically stable doubly charged species provided that the energy barrier to pass from the covalent (blue) potential energy curve to the coulombic (green) curve dissociating into $(ML–H)^+ + H^+$ is high enough so that several vibronic resonant states can be accommodated below the top of the barrier.

The interactions between uracil and Cu^{2+} nicely illustrate these possibilities. As shown in Figure 1-11a, the $[uracil–Cu]^{2+}$ binding energy is estimated to be 973 kJ mol^{-1}. Although the dissociation limit of Coulomb explosion yielding $uracil^+ + Cu^+$ is 145 kJ mol^{-1} lower in energy, the $[uracil–Cu]^{2+}$ is kinetically stable with respect to this process because the crossing between the covalent (blue) and the coulombic (red) potential energy curves leads to a significantly high barrier for this dissociation to take place. This is not the case for the second Coulomb explosion associated with the deprotonation of the system, which is not only much more exothermic (464 kJ mol^{-1}) than the previous one, but which involves a negligible activation barrier, indicating that $[uracil–Cu]^{2+}$ complexes are completely unstable with respect to the proton loss. This is in nice agreement with the experimental evidence which shows that in the gas-phase uracil–Cu^{2+} reactions only the $[uracilCu–H]^+$ monocations are detected. This is a clear consequence of the oxidative character of Cu^{2+} reflected in the quite large second ionization energy of Cu (20.29 eV). However, the second ionization energy of Ca is much lower (12.03 eV) and therefore no oxidation of the base should be expected. The main consequence is that

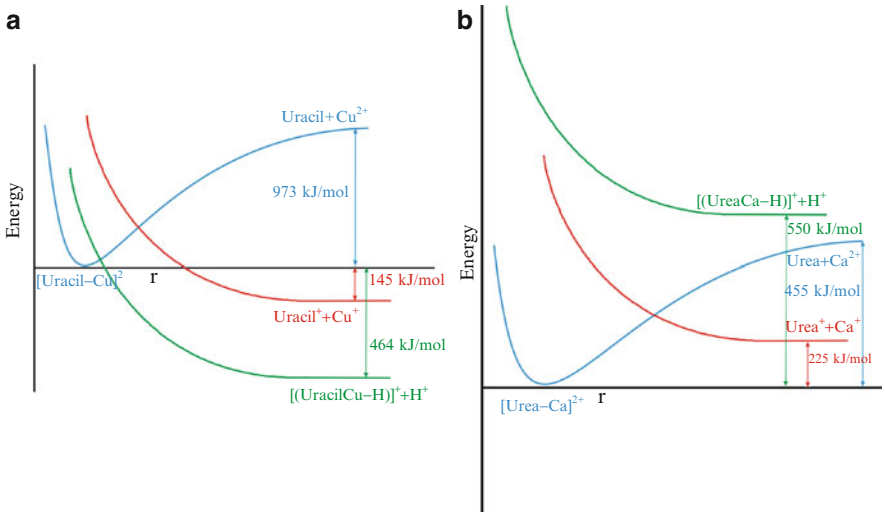


Figure 1-11. Potential energy curves associated with the formation of uracil–Cu^{2+} and urea–Ca^{2+} complexes and their Coulomb explosions yielding Ca^+ and H^+ as products

in general Ca^{2+} complexes are stable in the gas phase. As a matter of fact, as illustrated in Figure 1-11b for the particular case of urea–Ca^{2+} reactions, both Coulomb explosions, leading to the loss of Ca^{+} or H^{+} are endothermic process and therefore the $[urea–Ca]^{2+}$ complexes should be observed in the gas phase, as it is indeed the case and will be discussed in forthcoming sections.

1.4. REACTIVITY OF Ca^{2+} IONS IN THE GAS PHASE

As already mentioned in the introduction, gas-phase studies about metal-ion solvation by various molecules have been of interest for a very long time. One of the motivations is that the information that can be deduced from these studies provides strong insights into the metal-ion chemistry in solution. In addition, such studies also give useful information about the structure and functions of many biomolecules interacting with metal ions. However, until quite recently, most work was limited to singly charged systems because solvated multiply charged metal ions were not accessible experimentally. The difficulty in generating solvated divalent metals experimentally is due to the large difference between the second ionization energy (IE2) of a metal and the first ionization energy of a ligand (IE1). As a matter of fact, IE1 of most common ligands ranges from 8 to 12 eV while IE2 of almost all metals is above 12 eV. Consequently, as indicated in previous sections, addition of a ligand to a bare dication leads to a spontaneous exothermic electron transfer on contact followed by dissociation driven by Coulomb explosion according to reaction (1-1).

$$ML^{2+} \rightarrow ML^{+} + L^{+} \tag{1-1}$$

However, as most divalent M^{2+} metals (and more particularly alkali-earth metals) are stable in bulk solution in water or other solvents, generation of solvated metal polycations in the gas phase could be possible by transferring already existing $M^{2+}(L)_n$ ions from the solution to the gas phase. This is what is achieved by the electrospray ionisation (ESI) process. This method was pioneered by Kebarle's group in the early 1990s [84–87] and is now widely used worldwide to generate multiply charged metal ion complexes with various ligands. An alternative process has been developed by Stace and coworkers. This so-called pick-up technique [88] consists in generating in a first step a complex of needed size containing a neutral metal center. Then in a second step its charged state is raised up by high-energy electron impact ionization.

The fact that solvated divalent metal cations could be readily produced led to significant interest in obtaining thermochemical information about these ions. In this respect, the first solvent considered was water and many experimental and theoretical studies [33, 39, 84, 86, 89–93] aimed at determining the enthalpies, free energies, and entropies for the sequential dehydration reaction,

$$M^{2+}(L)_n \rightarrow M^{2+}(L)_{n-1} + L \tag{1-2}$$

(L=H_2O) for alkali earth metal dications, and n = 5–14 in the gas phase. For all these metal hydrates, dissociation only occurred by loss of one water molecule. The experimental dehydration enthalpies deduced either from equilibrium [33, 84, 86, 90], blackbody infrared radiative dissociation (BIRD) [89, 92, 93] or threshold collision induced dissociation (TCID) [39] measurements are globally in good agreement. For the particular case of Ca^{2+} ions, the measured dehydration enthalpy increases only slightly from n = 14 to n = 7, and then increases substantially for n = 6. These results clearly established the preference of Ca for a six-water first solvation shell. The same findings were reported for strontium and barium. For magnesium, it was found that the $M^{2+}(H_2O)_6$ corresponds to a mixture of two structures, one involving six water molecules in the first solvation shell and a second one which, according to the experimental information [93], most likely corresponds to four water molecules in the inner shell and two water molecules in the second shell.

According to both DFT and MP2 calculations [39], dehydration enthalpies for n < 5 are rather large (for example, $\Delta H^\circ_{1,0}(Ca^{2+})$ = 229 kJ mol^{-1} $\Delta H^\circ_{2,1}(Ca^{2+})$ = 197 kJ mol^{-1}) and therefore cannot be measured by the aforementioned experimental approaches. An alternative mean to study the reactivity $M^{2+}(H_2O)_n$ (n ≤ 4) is to perform low-energy CID experiments in a tandem mass spectrometer. Dissociation reactions can be studied by selecting these complexes with a first mass analyzer, collide them in a collision chamber by a collision gas, and detect the fragmentation products in a second analyzer. For example, CID spectra of $Ca^{2+}(H_2O)_n$ (n = 1–4) recorded with our QqTOF instrument are given in Figure 1-12.

As illustrated in this figure, when the number n of water decreases, new reaction channels may compete with the loss of intact ligand. With n = 4, the MS/MS spectrum is dominated by sequential losses of water. For n = 2 (Figure 1-12b), one can see that the predominant process leads to the formation of calcium hydroxide $CaOH^+$, while the smallest calcium hydrate loses a molecule of water to generate bare Ca^{2+} ions. Interestingly an ion resulting from the association of Ca^{2+} with N_2 (the collision gas used during this set of experiments) is also detected at *m*/*z* 33.99. The fact that neither *m*/*z* 19.99 nor 33.99 are observed on Figure 1-12d supports this interpretation.

A general scheme for the various ligands investigated since the early 1990s is that two main charge reduction processes compete with the loss of intact ligand as the number n of ligands surrounding the metal decreases. These two processes, described by reactions (1-3) and (1-4) imply electron transfer and interligand proton transfer, respectively, followed by charge separation driven by Coulomb explosion.

$$M^{2+}(L)_n \rightarrow M^+(L)_{n-1} + L^+ \tag{1-3}$$

$$M^{2+}(L)_n \rightarrow M(L-H)^+(L)_{n-k-2} + LH^+(L)_k \tag{1-4}$$

Consequently, the calcium hydroxide ion observed in Figure 1-12b is generated via the dissociative proton transfer (with n = 2 and k = 0). Note that $CaOH^+$ is also detected in weak abundance in Figure 1-12a probably through transient formation

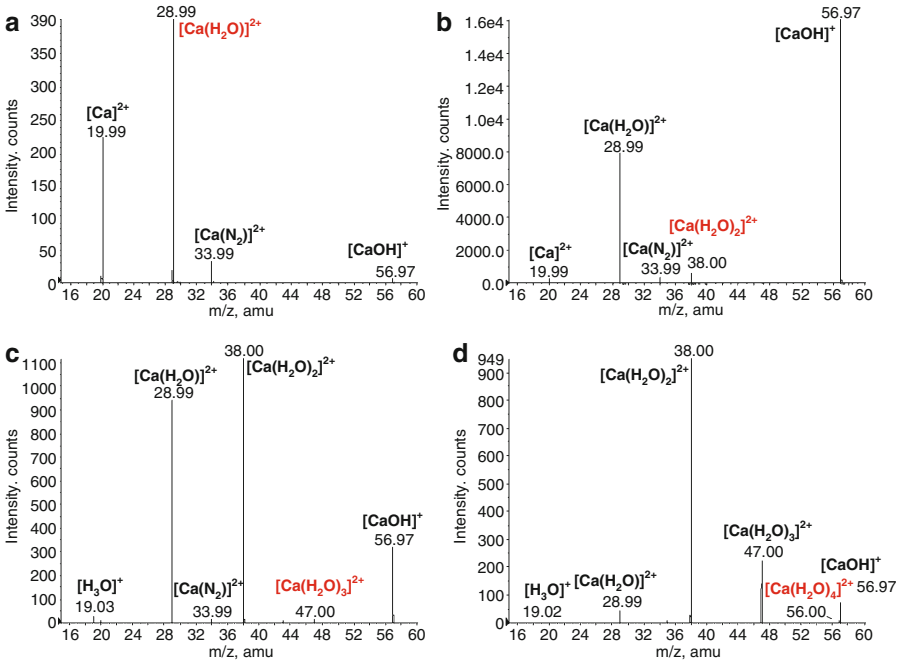


Figure 1-12. Low-energy CID spectra of $Ca^{2+}(H_2O)_n$ ions recorded with a QqTOF instrument at a collision energy of 8 eV (laboratory frame), with (**a**) n = 1, (**b**) n = 2, (**c**) n = 3 and (**d**) n = 4

of $[Ca(H_2O)_2]^{2+}$. Several observations support this interpretation. First, direct formation of $CaOH^+$ from $Ca^{2+}(H_2O)$ is highly endothermic [94]. Spears and Fehsenfeld [95] investigated the addition of H_2O to Ca^{2+} in a flowing afterglow apparatus at 296 K. One H_2O molecule was observed to attach to Ca^{2+}. On the other hand, addition of a second water molecule resulted in spontaneous charge separation by a proton-transfer reaction to form the metal hydroxide $CaOH^+$ and H_3O^+. One can also mention the recent TCID study of Carl et al.[39], who observed that H_3O^+ was detected once the $Ca^{2+}(H_2O)_2$ complex was generated. Note that electron transfer (reaction (1.3)) leading to Ca^+ ions is not observed, in agreement with previous findings [33, 39, 84, 86, 89–93, 95, 96].

Beyer et al. explored in detail the interligand proton transfer reaction involving the $M^{2+}(H_2O)_2$ complexes (M=Be, Mg, Ca, Sr and Ba) [97] and proposed that the reaction proceeds through a salt-bridge mechanism (Figure 1-13), for which the first step corresponds to migration of one molecule from the inner to the outer solvation shell.

No crossing points between the water loss and the charge-transfer curve were located for Ca^{2+}, Sr^{2+}, and Ba^{2+}. For these metal ions, IE2 of the metal is lower than the ionization energy of water so that curve crossing to the charge-transfer exit channel cannot occur. In fact under CID conditions, such electron transfer reaction to water has never been observed experimentally for any alkali-earth metal.

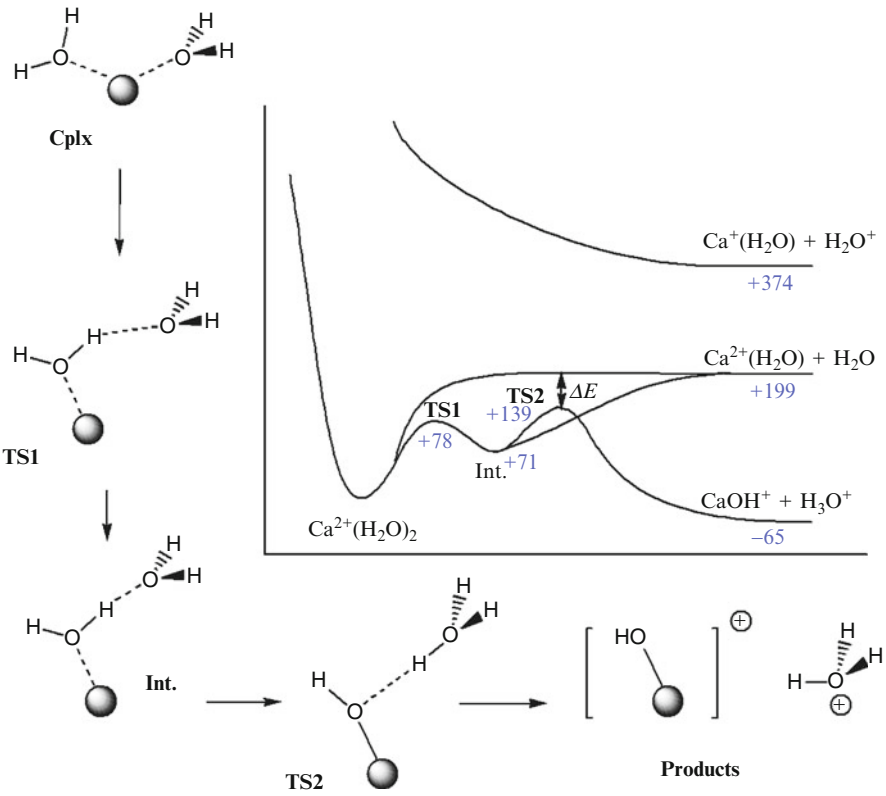


Figure 1-13. "Salt-bridge" mechanism associated with the formation of MOH^+ and H_3O^+ from the $M^{2+}(H_2O)_2$ complex and the corresponding potential energy surface calculated for M=Ca. See reference [97] for details

The competition between loss of water and the proton transfer is found to depend on the relative energy between TS2 and products corresponding to water loss (ΔE, Figure 1-13). For instance, in $Mg^{2+}(H_2O)_2$ TS2 is 132 kJ mol^{-1} lower in energy than water loss. Thus, hydroxide formation is highly energetically favored and is even observed with three and four molecules of water. This energy difference decreases to 60 and 10 kJ mol^{-1} for calcium, and strontium, respectively. For barium, the relative energies are reversed; loss of water is favored by 20 kJ mol^{-1} over the proton-transfer reaction and consistently formation of bare Ba^{2+} ions is the prominent fragmentation observed [98]. Interestingly, the net charge on the metal remains unchanged during this process, suggesting that IE2 of the metal is not necessarily the driving force for this mechanism. Instead, energetics of the process leading to MOH^+ and H_3O^+ is correlated with the radius of the metal ion.

In order to rationalize the reactivity of the doubly charged complexes, two parameters n_{min} and n_{crit} were introduced [98]. n_{min} is defined as the minimum

number of ligand for which $M^{2+}(L)_n$ complexes, under CID conditions, remain stable against dissociative electron transfer (1-3) or proton transfer (1-4) reactions. n_{crit} is defined as the largest value of n at and below which $M^{2+}(L)_n$ decomposes not solely by neutral evaporation but via charge (electron and/or proton) transfer as well. Values of n_{min} and n_{crit} obtained for three solvents (water, acetonitrile, and dimethylsulfoxide) are gathered in Table 1-2.

Note that these values may depend on the mass spectrometer used to record the CID spectra, and more particularly to its sensitivity and dynamic range. Examination of Table 1-2 indicates that n_{min} and n_{crit} depend on the nature of the ligand. For aqua complexes, electron transfer is not competitive and is not encountered for any metal, with any number of ligands. In the case of acetonitrile, electron and interligand proton transfer are often competitive. The latter appears less facile than for protic solvent such as water. Expectedly, the electron transfer becomes more and more favourable as the ionization energy of the ligand decreases. In this respect, pyridine (IE = 9.3 eV) behaves like DMSO with Ca^{2+} and transition metal dications [101, 102]. Examination of Table 1-2 also reveals that the second ionization energy of the metal affects the critical number of ligands below which either electron or proton transfer occurs. If one considers the acetonitrile case, n_{crit} shifts from 0 to 4 as IE2 increases from 10 to 20 eV. The IE2 also determines whether the solvent evaporation can proceed all the way to bare M^{2+} ion.

Table 1-2 Values of n_{min} and n_{crit} determined for different metal ions and three solvents: water (IE1 = 12.6 eV [80]), acetonitrile (IE1 = 12.2 eV [80]) and DMSO (IE1 = 9.1 eV [80]). (e) stands for electron transfer while (p) refers to as the interligand proton transfer

Metal	IE2(eV)	H_2O[a]		CH_3CN[b]		$(CH_3)_2SO$[c]	
		n_{min}	n_{crit}	n_{min}	n_{crit}	n_{min}	n_{crit}
Ba	10.0	0	2(p)	0	0	0	0
Sr	11.0	0	2(p)	0	0,2(p)[e]	0	1(e)
Ca	11.9	0[d]	2(p)	0	0,2(p)[e]	0	1(e)
Mg	15.0	1	4(p)	0	3(e,p)	1	3(p)
Mn	15.6	1	4(p)	0	3(e,p)	1	3(p)
Fe	16.2	1	5(p)	0	4(e,p)	1	2(e)
Cd	16.8	1	4(p)	0	3(e,p)	2	2(e)
Co	17.1	1	5(p)	0	3(e,p)	1	3(e,p)
Zn	18.0	1	6(p)	1	3(e,p)	2	3(p)
Ni	18.2	1	5(p)	1	3(e,p)	2	3(e)
Be	18.2	2	8(p)	1	4(e,p)	2	–
Cu	20.3	1	6(p)	1	3(e,p)	2	3(e)

[a]From [98]
[b]From [99]
[c]From [100]
[d]This work
[e]From [101]

In addition to neutral ligand loss and charge reduction processes, some ligand-specific fragmentation products have been detected. With acetonitrile, the heterolytic cleavage of the C–C bond leading to $[M^{2+}CN^{-}]^{+}$ ion and CH_3^{+} is observed.

$$M^{2+}(CH_3CN)_n \rightarrow M^{+}CN(CH_3CN)_{n-1} + CH_3^{+} \quad (1\text{-}5)$$

This process was observed for Ca^{2+} but in very weak intensity compared to Co and Mn [99, 101, 102]. According to a recent computational study [103], the $Ca^{2+}(CH_3CN)$ complex exhibits a linear geometry, the metal interacting with the nitrogen atom. N≡C and C–C bond are slightly elongated. The electron withdrawing effect of Ca^{2+} is such that the metal dication polarizes the ligand, through a metal–nitrogen interaction, and weakens the C–C bond thus promoting heterolytic bond cleavage. The potential energy surface (PES) associated with the unimolecular reactivity of $M^{2+}(CH_3CN)$ complexes has been described in details by Xiao et al. [103]. The dissociation processes studied included the loss of the neutral ligand, the dissociative electron transfer, and the elimination of neutral and charged methyl groups (CH_3 and CH_3^{+}). Results obtained for calcium showed that elimination of neutral acetonitrile is the most favourable dissociation route. Formation of the methyl cation is the less energy-demanding process. However, the associated activation barrier is higher than for the loss of CH_3CN. The formation of CH_3^{+} and $CaCN^{+}$ is expected to be less pronounced than neutral elimination, in agreement with experimental findings [99, 101]. In addition, loss of the methyl radical is highly unfavourable. Other metals have been studied, and the computed PESs clearly show that the loss of neutral acetonitrile becomes less favourable than the electron transfer as the second ionization energy of the metal increases.

Unlike acetonitrile, prominent elimination of methyl radical is characteristic of the reactivity observed with DMSO [100]. Indeed, $Ca^{2+}(DMSO)_n$ complexes ($n \leq 4$) fragment by homolytic cleavage of the C–S bond, yielding one or more methyl radicals:

$$Ca^{2+}(DMSO)_n \rightarrow Ca^{2+}\left[(DMSO)_n - m \cdot CH_3\right] + m \cdot CH_3 \quad (1\text{-}6)$$

This is a particular behaviour as there is no charge reduction and therefore this cleavage leads to a doubly charged fragment ion. Another major difference with respect to acetonitrile is that the C−S cleavage in $M^{2+}(DMSO)_n$ may also involve a proton transfer to the methyl leaving group that converts it into methane:

$$Ca^{2+}(DMSO)_n \rightarrow Ca^{2+}\left[(DMSO)_n - mCH_4\right] + mCH_4 \quad (1\text{-}7)$$

In summary, this reactivity is markedly different from what is observed with water or acetonitrile, in which ligand cleavage is always accompanied by charge reduction.

Briefly, other solvent molecules have been considered such as methanol [101], pyridine [101, 102] and acetone [104]. In addition to ligand loss and dissociative electron and proton transfer, a particularly interesting reaction observed with the

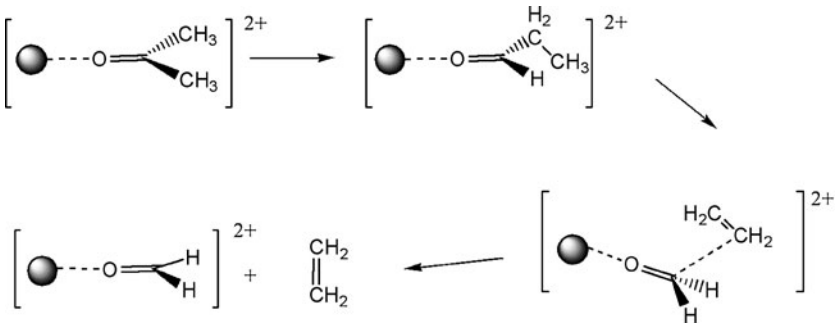


Figure 1-14. Proposed mechanism for the elimination of ethylene from the Ca^{2+}(acetone) complex

latter is elimination of ethylene (reaction (1-8)), observed for alkali-earth and transition metal dications, and particularly intense in the case of calcium.

$$Ca^{2+}(CH_3COCH_3) \rightarrow Ca^{2+}(CH_2O) + C_2H_4 \tag{1-8}$$

The proposed mechanism is given in Figure 1-14.
It first implies simultaneous H and methyl migrations in a single step leading to a Ca^{2+}(propanal) intermediate. A second hydrogen migration onto the carbonyl group results in a single-step formation of an ion–neutral complex, which ultimately leads to the loss of ethylene. Although this limited number of computed steps may appear surprising, the computed PES indicate that this ethylene elimination is the most favourable dissociation pathway, in agreement with experimental findings [104].

Finally, another striking feature observed with pyridine is the capacity of divalent metal ions, and notably Ca^{2+} cation, to destroy the aromaticity of an aromatic ligand [101, 102]. While singly charged metal ions do not normally activate aromatic rings, a heterolytic cleavage leading to the formation of metal cyanide ions has been observed for pyridine besides charge reduction by electron or proton transfer.

$$M^{2+}(\text{pyridine})_n \rightarrow [MCN]^+(\text{pyridine})_{n-1} + C_4H_5{}^+ \tag{1-9}$$

Note that like for DMSO, a homolytic cleavage leading to the elimination of NH_2 or CH_3 radicals were also observed. Although computations were not carried out, the facile formation of $M^{2+}(\text{pyridine})_n$ with $n > 2$ suggests that M^{2+} coordinates not to the π-ring, but in-plane to nitrogen like in acetonitrile.

Apart from solvents, the reactivity of alkali-earth dications with some other organic small molecules has been studied. One may cite the work of Feil et al. [105] about the gas-phase reactions of both Ca^+ and Ca^{2+} with ozone [105], or the study of the complexation process with acetamides [33, 106].

Our group focused its attention to the reactivity of Ca^{2+} ions towards urea and its thio and seleno derivatives by combining mass spectrometry experiments and computational calculations [48, 50, 78, 107]. One reasonable way of approaching the problem of the reactivity of large biochemical compounds such as the

nucleotides is by exploring the behavior of small model systems, such as urea. Under electrospray condition, aqueous mixture of calcium chloride and urea leads to facile formation of $Ca^{2+}(urea)_n$ (n = 1–3) complexes in the gas phase. Their low-energy CID spectra recorded on a QqTOF instrument are presented in Figure 1-15. The minimum amount of collision gas was used in order to limit the extent of multiple collisions. One can see that the reactivity significantly changes as the

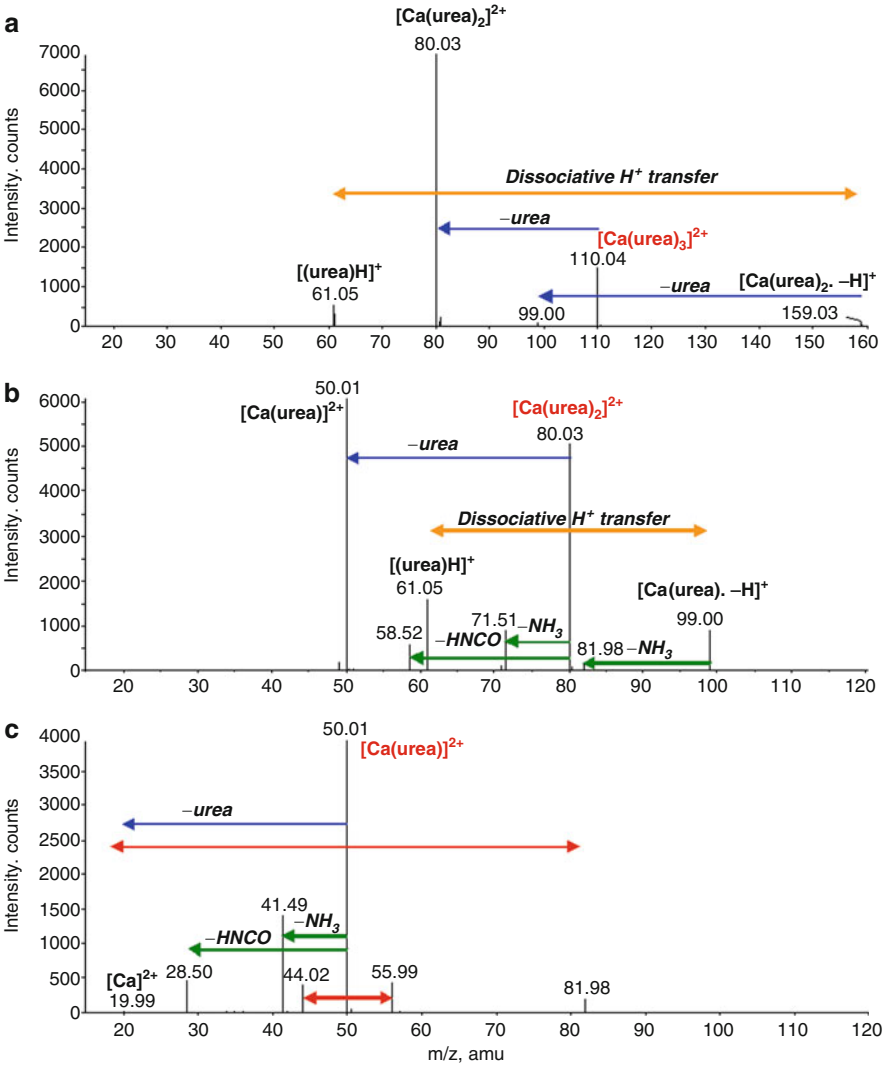
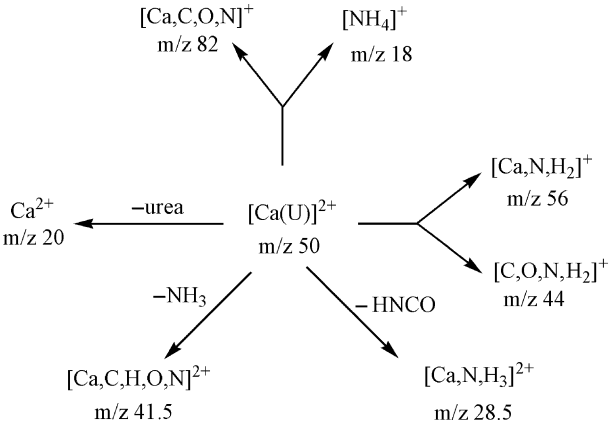


Figure 1-15. Low-energy CID spectra of $Ca^{2+}(urea)_n$ ions recorded with a QqTOF instrument at a collision energy of 10 eV (laboratory frame), with (**a**) n = 3, (**b**) n = 2, (**c**) n = 1

number of urea molecules decreases. With n = 3, the main process corresponds to elimination of urea, leading to $Ca^{2+}(urea)_2$ complex (*m*/*z* 80.00). Charge reduction through proton transfer is already significant for n = 3. So, n_{crit} is at least equal to 3 (the $Ca^{2+}(urea)_4$ ion was detected but its intensity was too low to record its MS/MS spectrum). When n = 2, loss of intact urea and proton transfer still dominate, but new fragmentation patterns appear, corresponding to elimination of neutral molecules such as ammonia and [H,N,C,O].

Finally, the reactivity of the smallest Ca^{2+}/urea complex (Figure 1-15c) is particularly rich. Besides limited elimination of urea and predominant elimination of neutrals, two additional charge separation processes, via appropriate Coulomb explosions, are observed, leading to the formation of *m*/*z* 44, 56, 82, but also *m*/*z* 18 (only detected at higher collision energies). The overall fragmentation pattern is summarized below.



In comparison, the MS/MS spectrum of the $[Ca(urea)–H]^+$ complex is remarkably simpler since only two product ions corresponding to elimination of ammonia and [H,N,C,O] are observed. These two particular processes have already been encountered in reactions of urea with Cu^+ [108, 109] and Ni^+ [110] ions. On the other hand, unlike the corresponding situations for Cu^+ or Ni^+, loss of water, which has been attributed to a partial collision-induced isomerization of M^+(urea) (M = metal) in the cone region prior to mass selection [109, 110], is not observed. Moreover, no ions corresponding to loss of CO were detected.

In order to propose structures for the various dissociation products and to rationalize these experimental findings, we have studied the PES of the $[Ca(\mathbf{U})]^{2+}$ system (Figure 1-16). A complete description of this PES is beyond the scope of the present chapter, but several comments can be made. First, this study [48] suggests that ions arising from the two Coulomb explosions are NH_4^+ (*m*/*z* 18), $CaNH_2^+$ (*m*/*z* 56), H_2NCO^+ (*m*/*z* 44) and $[Ca,N,C,O]^+$ (*m*/*z* 82). For the latter species, $CaNCO^+$ is strongly preferred thermodynamically while formation of $CaOCN^+$ is slightly favored kinetically. One can also see that elimination of 43 Da probably

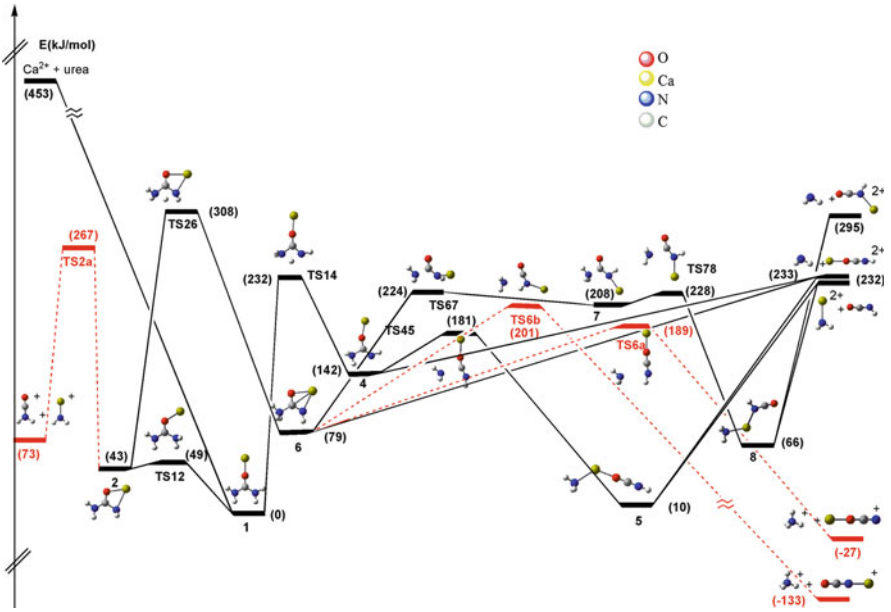


Figure 1-16. Potential energy surface associated with the fragmentation of the Ca^{2+}(urea) complex. Coulomb explosion dissociation pathways are in *red dotted lines*

corresponds to the elimination of isocyanic acid HNCO and implies, like the elimination of NH_3, a 1,3-H shift from one amino group to the other.

A striking difference with the experimental results is that the formation of NH_4^+ and $CaNCO^+$ should be more favorable than the elimination of ammonia. In this respect, the kinetic study [107] carried out recently provided useful insights. In fact this study showed that at 300 K, the preferred product ion (loss of ammonia) from the kinetic viewpoint is not the most thermodynamically stable (formation of NH_4^+). These results are in agreement with the experimental spectrum, thus suggesting that the present system is rather under kinetic control.

Replacing urea by either thiourea or selenourea results in significant differences. As illustrated by Figure 1-17 for selenourea, many more dissociation products are observed.

Partner peaks arising from the Coulomb explosion processes should have in principle the same intensity. However, as it was already found for the urea–Ca^{2+} system, the lightest ions (m/z 18.03, 39.96, 43.02) are systematically less intense than the heaviest ones (m/z 145.86, 123.94, 120.86, respectively). Similar findings have been also previously reported for alcohols [111] and interpreted in terms of different radial ion energies, with the lighter ions generated by the Coulomb explosion gaining most of the radial energy and therefore having a much higher velocity than the relatively high mass ions. This can result in an unstable ion trajectory within the instrument and explains why lighter ions are detected in the

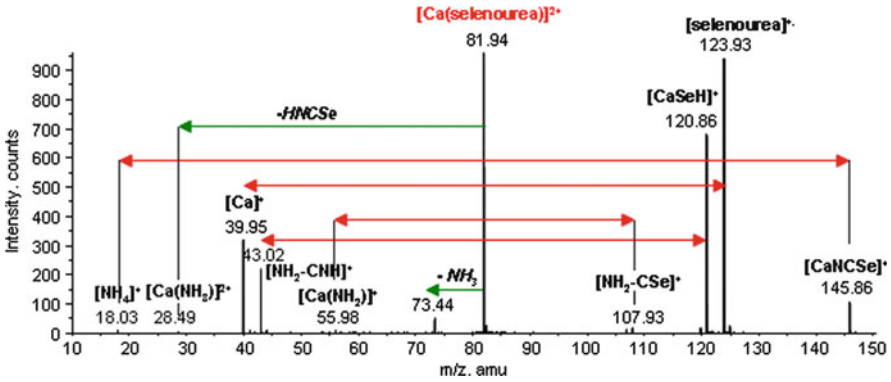


Figure 1-17. Low-energy CID spectrum of Ca^{2+}(selenourea) ions recorded with a QqTOF instrument at a collision energy of 11 eV (laboratory frame)

Table 1-3 Summary of the fragment ions observed on the MS/MS spectra of Ca^{2+}(L) complexes (L = urea, thiourea or selenourea)

Ligand L fragment ions (m/z)[a]	Urea (X = O)	Thiourea (X = S)	Selenourea (X = Se)
NH_4^+ (18)	Low	Low	Low
Ca^{2+} (20)	Sig	Low	Low
Ca^{2+} (NH_3) (28.5)	Sig	Low	Low
Ca^+ (40)	–	Sig	**Str**
$CH_3N_2^+$ (43)	–	Sig	Sig
$Ca(NH_2)^+$ (56)	**Str**	**Str**	Low
Ca^{2+}(H)NCX	**Str** (41.5)	Low (49.5)	Low (73.5)
NH_2CX^+	**Str** (44.0)	**Str** (60.0)	Low (107.8)
$CaXH^+$	–	**Str** (72.9)	**Str** (120.9)
L^+	–	**Str** (76)	**Str** (124)
$CaNCX^+$	Sig (81.9)	Sig (97.9)	Sig (145.9)

[a]Sig: significant; **Str**: strong

MS/MS spectrum with smaller abundance. This effect may be even more pronounced in our QqTOF instrument, the lightest ions being less effectively transferred into the TOF region during the orthogonal injection step, because of their higher velocity.

The main fragments ions obtained with urea and its derivatives are gathered in Table 1-3. A first important difference is the very abundant ion corresponding to calcium thio- and seleno hydroxides. Such a process was not observed with urea. This fragmentation is particularly interesting since a single molecule of thiourea or selenourea is necessary, while we have seen previously (vide supra) that two molecules of water were required to generate the $CaOH^+$ ion.

Furthermore, examination of the potential energy surfaces [50, 78] indicate that formation of $CaSH^+$ and $CaSeH^+$ correspond to the lowest energy paths. This may be explained by the substantial difference between the barriers associated with the hydrogen shifts from the amino groups towards the C = X (X = O, S, Se) group. In both thiourea and selenourea, these hydrogen shifts involve activation barriers that are rather similar to those of the different Coulomb explosions and to those of the hydrogen shifts between the amino groups. Conversely, in urea, these barriers are much higher in energy, and therefore the associated processes very unfavorable.

One of the most important dissimilarities between urea and its S and Se containing analogues arises from the large difference between the corresponding ionization energies, which for urea (10.27 eV [80]) is sizably larger than for thiourea (8.50 eV [80]) and for selenourea (7.80 eV, as estimated at the B3LYP/6-311 + G(3df,2p)/B3LYP/6-311 + G(d,p) level). Accordingly, one can see that the formation of the radical cation of the ligand is particularly favorable as illustrated by the intense peak at *m*/*z* 123.94 (selenourea$^+$), while such a charge transfer is not observed with urea. Furthermore, this is a more and more facile process ongoing form thiourea and to selenourea, as attested by the increasing abundance of these species and Ca^+ ion at a given collision energy. Finally, it is worth noting that these experiments unambiguously demonstrated that the $Ca^{2+}(NH_3)$ is a stable species in the gas phase. Calculations also showed that the Ca^{2+} affinity of ammonia is almost identical to that of HNCO if the metal cation attachment takes place at the oxygen atom in the latter case. On the other hand, the Ca^{2+} affinity of ammonia is greater than that of HNCX (X = S, Se), regardless of the attachment site of the metal to the HNCX moiety. Data also show that HNCX behaves preferentially as a nitrogen base towards Ca^{2+}.

Interaction of metal cations with amino acids can provide useful insights into the corresponding interactions with more complex biologically relevant systems such as peptides or proteins. In this respect, glycine represents the simplest model for such a theoretical and experimental investigation, and we decided to study [49] how its reactivity with Ca^{2+} ions compares with that of urea derivatives [48, 50, 78]. Use of labeled compounds was necessary not only to gain more information about the dissociation patterns, but also to achieve better separation of the Ca^{2+}(glycine) complex (*m*/*z* 57.5) from calcium hydroxide $CaOH^+$ (*m*/*z* 56.97) which is produced in abundance by ESI and may interfere during the selection of the complex.

The MS/MS spectrum (Figure 1-18) appears remarkably simple compare to those obtained with urea derivatives. First, there is practically no loss of neutral fragments. Two charge reduction processes are observed leading to two pairs of product ions, 32 + 85 and 57 + 60. For unlabeled glycine, the first one corresponds to the immonium ion $CH_2{=}NH_2^+$ and $CaCOOH^+$, while the second one is attributed to $CaOH^+$ and $NH_2CH_2CO^+$. This reactivity differs markedly from that observed under FAB (Fast Atom Bombardment) conditions with Ni^+ ions [112], which is characterized by elimination of CO, H_2O and 46 Da corresponding to H_2O/CO. This latter process is commonly observed for aminoacids and notably in the case of the Cu^+/glycine system [113].

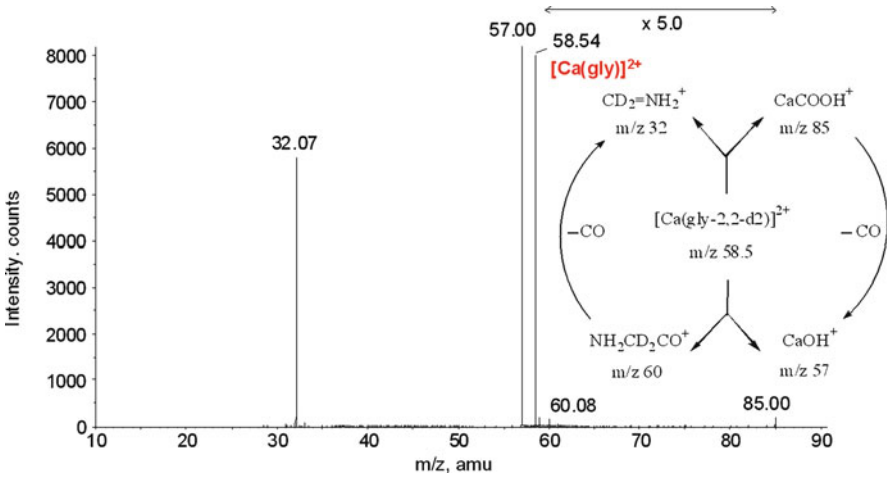


Figure 1-18. Low-energy CID spectrum of Ca^{2+}(glycine-2,2-d_2) complex recorded with a QqTOF instrument at a collision energy of 11 eV (laboratory frame)

The striking feature of Figure 1-18 is that *m*/*z* 32 and *m*/*z* 57 peaks are significantly more intense than their partner peaks at *m*/*z* 85 and *m*/*z* 60, respectively. This is the reverse of our previous observations for urea derivatives. This finding was interpreted by the fact that *m*/*z* 60 and 85 may undergo a secondary fragmentation (loss of CO) leading to ions with *m*/*z* 32 and 57, respectively. This seems very likely since our theoretical study [49] suggests that *m*/*z* 60 and 85 ions certainly correspond to $NH_2CH_2CO^+$ and $[CaCOOH]^+$, respectively.

Amino acids exist as zwitterions in aqueous solution within a wide range of pH. In contrast, all experimental evidences indicate that the ground-state structure of amino acids in the gas phase is the neutral form. The differences between the gas-phase and solution-phase structures of amino acids are due to solvent preferentially stabilizing the zwitterionic form. Monovalent metal ion-chelated zwitterionic glycine complex is not the global minimum. Recent studies about the interaction of glycine with alkali metals [114], Ni^+ [112] or Cu^+ [113, 115] have shown that the most stable form involves metal chelation between the nitrogen atom and the oxygen atom of the carbonyl group of neutral glycine. The situation is radically different for metal dications. Dications strongly stabilize the zwitterionic form of glycine in the gas phase [49]. Indeed, the structures in which the dication interacts with the carboxylate group of the zwitterionic form were found to be the global minimum for all alkali-earth metals but Be [49, 116], and also for Zn(II) [114] (Figure 1-19).

Furthermore, a theoretical study has demonstrated that in presence of Ca^{2+} ions, the intramolecular proton transfer leading to the zwitterionic glycine occurred without any activation barrier [117]. Consequently, the presence of the zwitterionic form of glycine at least as an intermediate, might explain the difference of reactivity

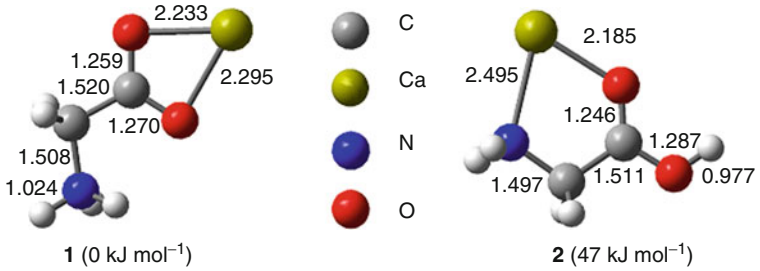


Figure 1-19. The two most stable structures of the Ca^{2+}(glycine) complex. See reference [49] for details

observed between Ca^{2+} and other metal ion such as Ni^{+} or Cu^{+}. Our computational results [49] indicate that the least-energy demanding dissociation pathways leading to the observed fragment ions involve the zwitterionic complex (**1**). In addition, it is worth noting that several experimental evidences support the existence of zwitterionic aminoacids in presence of metal ions. More particularly, IRMPD (InfraRed Multiple Photon Dissociation) spectroscopy has emerged as a powerful method to directly probe gas-phase structures of cationized amino acids. Kapota et al. [118] provided the first experimental evidence that the zwitterion is formed in the case of Proline–Na^{+}. In contrast, chelation of Na^{+} between nitrogen and the carbonyl oxygen of the neutral form seems to be most favorable for Glycine–Na^{+}. Concerning cationization by alkali-earth metal, Dunbar and Oomens [119] reported the IRMPD spectrum of divalent barium complexed with tryptophan and concluded that tryptophan is zwitterionic in this ion. Similarly, it was recently demonstrated that arginine, glutamine, proline, serine, and valine all adopt zwitterionic structures when complexed with divalent barium [120]. The same conclusion could be drawn for Sr^{2+}/arginine system.

Metal cations can both stabilize and destabilize DNA [121]. The interaction of divalent cations with nucleic acids plays an important role in promoting and maintaining their functionalities [121, 122]. The interactions and the coordination chemistry of metal cations with nucleic acid building blocks have been studied quite extensively in the condensed phase. By contrast, data about the intrinsic gas-phase behavior of such interactions are rather scarce. In this context, we began to study the interaction of Ca^{2+} ions with nucleic acid building blocks and the present paragraph will present some preliminary results obtained with uracil. Recently, a mass spectrometric study of the Ca^{2+}/uracil system has been published [123], but surprisingly no doubly charged complexes were detected. However, with our experimental setup production of the $Ca^{2+}(uracil)_n$ complexes (n = 1–3) is straightforward. On the other hand, such doubly charged ions have not been observed with copper [55]. MS/MS spectra of Ca^{2+}(uracil) and $Ca^{2+}(uracil)_2$ are presented in Figure 1-20.

The dissociation pattern of the bisligated complex is classical, with elimination of neutral ligand (*m*/*z* 76.02) and charge reduction by interligand proton transfer giving rise to protonated uracil (*m*/*z* 113.06) and $[Ca(uracil)–H]^{+}$ ion (*m*/*z* 151.01).

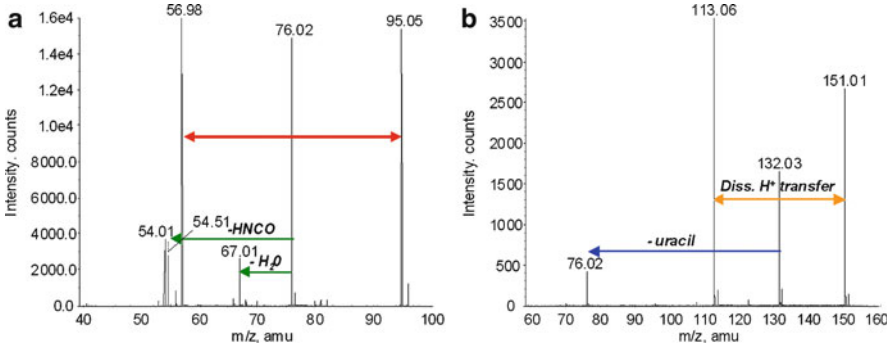


Figure 1-20. Low-energy CID spectra of (**a**) Ca^{2+}(uracil) and (**b**) $Ca^{2+}(uracil)_2$ complexes

The MS/MS spectrum of this latter species is simple, characterized by prominent loss of [H,N,C,O], very weak elimination of water, and finally formation of a $[Ca,N,C,O]^+$ ion. The two first processes are also observed for the doubly charged complex (Figure 1-20a). While elimination of 43 Da is a common feature for metal-cationized uracil [55, 56], loss of water appears characteristic of calcium. Consequently, unlike alkali metals [124], interactions of Ca^{2+} strongly activate the aromatic ring and promote its cleavage upon collision. The prominent fragmentation is a charge reduction process leading to calcium hydroxide (*m*/*z* 56.98) and formation of $[C_4,H_3,N_2,O]^+$ ion. An additional fragment ion is observed at *m*/*z* 54.01 but its identity is still unclear. Theoretical calculations and experiments involving labelled uracil are under progress in order to ascertain the structure of the various product ions.

To conclude this chapter, is it important to remind that fundamental studies about the Ca^{2+} reactivity have not been limited to small molecules. One can also mention studies about interactions with diketones [125] or crown ethers.[126] Furthermore, given the key role played by Ca^{2+} in biology, interactions of Ca^{2+} with peptides and proteins have been studied by mass spectrometry both in positive and negative ion mode. As matter of fact, as metal-binding sites in interiors of proteins are hydrophobic, interactions in the gas phase are an appropriate model for those in hydrophobic regions. Gross and co-workers used FAB to generate complexes of the type $[Ca^{2+} + \text{tripeptide}–3H]^-$ in the gas phase [127, 128]. It was determined that Ca^{2+} interacts with the deprotonated C-terminal COOH group and two deprotonated amide nitrogens. Labelling experiments confirmed that the terminal amino group was not deprotonated. More recently [129], the same group has determined the metal-binding sites in small peptides that are models for the amino-acid sequence in large metal-binding proteins in an effort to get more insight about Ca^{2+} bonding sites of calcium binding proteins. By using ESI and H/D exchange experiments [130], they also demonstrated that calmodulin adopts a tighter, less solvent-accessible structure as the protein binds with Ca^{2+}. Furthermore, it is that tighter structure which binds with proteins, peptides, and other small molecules.

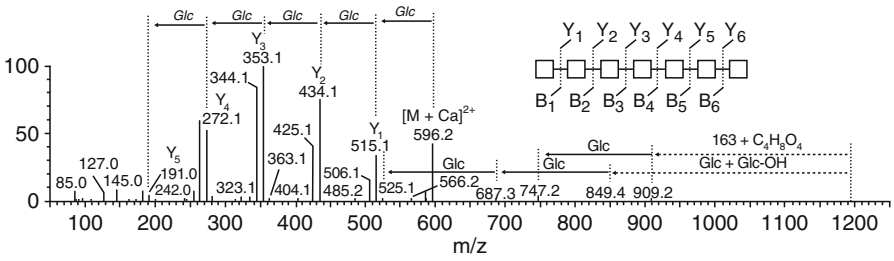


Figure 1-21. MS/MS spectra of the electrospray-generated [Ca+maltoheptaose]$^{2+}$ complex. See reference [133] for details (Reproduced by permission of Elsevier)

On the other hand, addition of Mg^{2+} to calmodulin does not cause any significant change in the tertiary structure of the protein.

Attempts were also made to apply the Ca^{2+} chemistry to the analysis of cyclic peptides (CP) [131]. In the case of cycloprin A, complexation with strontium and calcium ions resulted in the most intense complexes relative to formation of the protonated species. However, MS/MS spectra of [Ca+ CP]$^{2+}$ did not provide any particular additional information compared to the protonated peptides.

Alkaline-earth metal ions have also been applied to polysaccharide analysis. Mass spectrometric characterization of these biopolymers is a challenging task. A complete structural characterization of carbohydrates implies notably exact mass measurements, determination of the sequence, of the position of the glycosidic linkages and location of functional groups. Several reports have shown that Ca^{2+} ions can be particularly helpful for the structural characterization of polysacharides [24, 132–134]. As illustrated above (Figure 1-21), MS/MS spectra of doubly charged complexes obtained with Ca^{2+} ions provide a wealth of information.

Like small molecules, these complexes fragment according to neutral losses and charge reduction processes. The most intense ions presently correspond to glycosidic cleavages (Y and B ions) that involve single bond rupture between the sugar rings and therefore provide sequence information. Cross-ring cleavages of the rings themselves are also observed, these processes giving additional information on the linkage position of one residue to the next. Ca^{2+} cationization also gave useful insights in the analysis of sulfated polysaccharides such as glycosaminoglycans (GAG) [134]. The high fragility of sulfate groups has limited the usefulness of mass spectrometry in the analysis of GAGs. However, it has been shown that pairing the sulfated oligosaccharide anions with Ca^{2+} served to increase sulfate stability such that abundant backbone cleavage fragments are observed, allowing more complete sequence-related data to be obtained.

REFERENCES

1. Weeds A (1982) Nature 296:811–816
2. Sigel H, Sigel A (eds) (1988) Nickel and its role in biology, vol 23. Marcel Dekker, New York
3. Sigel H (1993) Chem Soc Rev 22:255–267

4. Illangasekare M, Sanchez G, Nickles T, Yarus M (1995) Science 267:643–647
5. Sigel H, Sigel A (eds) (1996) Interactions of metal ions with nucleotides: nucleic acids, and their constituents, vol 32. Marcel Dekker, New York
6. Sigel H, Sigel A (eds) (1996) Probing of nucleic acids by metal ion complexes of small molecules, vol 33. Marcel Dekker, New York
7. Herlitze S, Garcia DE, Mackie K, Hille B, Scheuer T, Catterall WA (1996) Nature 380:258–262
8. Lis H, Sharon N (1998) Chem Rev 98:637–674
9. Babu CS, Dudev T, Casareno R, Cowan JA, Lim C (2003) J Am Chem Soc 125:9318–9328
10. Mank M, Griesbeck O (2008) Chem Rev 108:1550–1564
11. Caputo A, Caci E, Ferrera L, Pedemonte N, Barsanti C, Sondo E, Pfeffer U, Ravazzolo R, Zegarra-Moran O, Galietta LJV (2008) Science 322:590–594
12. Jia X, Marzilli LG (1991) Biopolymers 31:23–44
13. Silver PJ (1985) Rev Clim Basic Pharm 5:341–395
14. Ashley CC, Griffiths PJ, Lea TJ, Mulligan IP, Palmer RE, Simnett SJ (1993) Rev Phys Biochem Pharmacol 122:149–258
15. Kang S, Wells RD (1992) J Biol Chem 267:20887–20891
16. Malkov VA, Voloshin ON, Soyfer VN, Frankkamenetskii MD (1993) Nucl Acids Res 21:585–591
17. Potaman VN, Soyfer VN (1994) J Biomol Struct Dyn 11:1035–1040
18. Noonberg SB, Francois JC, Garestier T, Helene C (1995) Nucl Acids Res 23:1956–1963
19. Saboury AA, Atri MS, Sanati MH, Moosavi-Movahedi AA, Haghbeen K (2005) Int J Biol Macromol 36:305–309
20. Serban DN, Serban IL, Petrescu G (2004) Rev Med Chir Soc Med Nat Iasi 108:441–451
21. Martin RB (1985) Acc Chem Res 18:32–38
22. Torres J, Veiga N, Gancheff JS, Dominguez S, Mederos A, Sundberg M, Sanchez A, Castiglioni J, Diaz A, Kremer C (2008) J Mol Struct: THEOCHEM 874:77–88
23. Burda JV, Sponer J, Leszczynski J, Hobza P (1997) J Phys Chem B 101:9670–9677
24. Zheng YJ, Ornstein RL, Leary JA (1997) J Mol Struct: THEOCHEM 389:233–240
25. Sponer J, Sponer JE, Gorb L, Leszczynski J, Lippert B (1999) J Phys Chem A 103:11406–11413
26. Wong CHS, Siu FM, Ma NL, Tsang CW (2001) J Mol Struct: THEOCHEM 536:227–234
27. Russo N, Toscano M, Grand A (2003) J Phys Chem A 107:11533–11538
28. Reddy AS, Sastry GN (2005) J Phys Chem A 109:8893
29. Nicolas I, Castro M (2006) J Phys Chem A 110:4564
30. Allen RN, Shukla MK, Burda JV, Leszczynski J (2006) J Phys Chem A 110:6139
31. Allen RN, Shukla MK, Leszczynski J (2006) Int J Quant Chem 106:2366–2372
32. Remko M, Fitz D, Rode BM (2008) J Phys Chem A 112:7652–7661
33. Peschke M, Blades AT, Kebarle P (2000) J Am Chem Soc 122:10440–10449
34. Trachtman M, Markham GD, Glusker JP, George P, Bock CW (2001) Inorg Chem 40:4230–4241
35. Bako I, Hutter J, Palinkas G (2002) J Chem Phys 117:9838–9843
36. Piquemal JP, Perera L, Cisneros GA, Ren PY, Pedersen LG, Darden TA (2006) J Chem Phys 125:054511
37. Jiao D, King C, Grossfield A, Darden TA, Ren PY (2006) J Phys Chem B 110:18553–18559
38. Tofteberg T, Ohrn A, Karlstrom G (2006) Chem Phys Lett 429:436–439
39. Carl DR, Moision RM, Armentrout PB (2007) Int J Mass Spectrom 265:308–325
40. Rao JS, Dinadayalane TC, Leszczynski J, Sastry GN (2008) J Phys Chem A 112:12944
41. Ai HQ, Bu YX, Li P, Zhang C (2005) New J Chem 29:1540–1548
42. Li HF, Bu YX, Yan SH, Li P, Cukier RI (2006) J Phys Chem B 110:11005–11013
43. Sundaresan N, Pillai CKS, Suresh CH (2006) J Phys Chem A 110:8826–8831
44. Reddy AS, Zipse H, Sastry GN (2007) J Phys Chem B 111:11546–11553

45. Gavryushov S (2007) J Phys Chem B 111:5264–5276
46. Liu HC, Zhang L, Li P, Cukier RI, Bu YX (2007) Chemphyschem 8:304–314
47. Potoff JJ, Issa Z, Manke CW, Jena BP (2008) Cell Biol Int 32:361–366
48. Corral I, Mó O, Yáñez M, Salpin JY, Tortajada J, Radom L (2004) J Phys Chem A 108: 10080–10088
49. Corral I, Mó O, Yáñez M, Salpin JY, Tortajada J, Moran D, Radom L (2006) Chem Eur J 12:6787–6796
50. Trujillo C, Mó O, Yáñez M, Salpin J-Y, Tortajada J (2007) Chemphyschem 8:1330–1337
51. Lamsabhi AM, Mó O, Yáñez M, Boyd RJ (2008) J Chem Theory Comput 4:1002–1011
52. Trujillo C, Lamsabhi AM, Mó O, Yáñez M (2008) Phys Chem Chem Phys 10:3229–3235
53. Trujillo C, Lamsabhi AM, Mó O, Yáñez M, Salpin JY (2008) Org Biomol Chem 6:3695–3702
54. Lamsabhi AM, Alcamí M, Mó O, Yáñez M, Tortajada J (2006) J Phys Chem A 110:1943–1950
55. Lamsabhi AM, Alcamí M, Mó O, Yáñez M, Tortajada J, Salpin J-Y (2007) Chemphyschem 8:181–187
56. Guillaumont S, Tortajada J, Salpin JY, Lamsabhi AM (2005) Int J Mass Spectrom 243:279–293
57. Corral I, Mó O, Yáñez M, Radom L (2005) J Phys Chem A 109:6735–6742
58. Burda JV, Sponer J, Hobza P (1996) J Phys Chem 100:7250–7255
59. Corral I, Mó O, Yáñez M, Scott AP, Radom L (2003) J Phys Chem A 107:10456–10461
60. Bader RFW (1990) Atoms in molecules. A quantum theory. Clarendon Press, Oxford
61. Becke AD, Edgecombe KE (1990) J Chem Phys 92:5397–5403
62. Savin A, Nesper R, Wengert S, Fassler TF (1997) Angew Chem Int Ed 36:1809–1832
63. Silvi B, Savin A (1994) Nature 371:683–686
64. Reed AE, Curtiss LA, Weinhold F (1988) Chem Rev 88:899–926
65. Alcamí M, Mó O, Yáñez M, Cooper IL (1999) J Phys Chem A 103:2793–2800
66. Abboud JLM, Notario R, Yáñez M, Mó O, Flammang R, Jagerovic N, Alkorta I, Elguero J (1999) J Phys Org Chem 12:787–795
67. Alcamí M, Mó O, Yáñez M (1989) J Phys Chem 93:3929–3936
68. Alcamí M, Mó O, Depaz JJG, Yáñez M (1990) Theor Chim Acta 77:1–15
69. Alcamí M, Mó O, Yáñez M, Anvia F, Taft RW (1990) J Phys Chem 94:4796–4804
70. Alcamí M, Mó O, Yáñez M (1996) Modelling intrinsic basicities: the use of the electrostatic potentials and the atoms–in–molecules theory, vol 3. Elsevier, Amsterdam
71. Karamat S, Fabian WMF (2008) J Phys Chem A 112:1823–1831
72. Dybal J, Makrlik E, Vanura P (2009) J Radioanal Nucl Chem 279:553–559
73. Estrin DA, Paglieri L, Corongiu G (1994) J Phys Chem 98:5653–5660
74. Gould IR, Burton NA, Hall RJ, Hillier IH (1995) J Mol Struct: THEOCHEM 331:147–154
75. Lamsabhi M, Alcamí M, Mó O, Bouab W, Esseffar M, Abboud JLM, Yáñez M (2000) J Phys Chem A 104:5122–5130
76. Trujillo C, Mó O, Yáñez M (2007) Org Biomol Chem 5:3092–3099
77. Lamsabhi AM, Mó O, Yáñez M, Guillemin JC, Haldys V, Tortajada J, Salpin JY (2008) J Mass Spectrom 43:317–326
78. Trujillo C, Mó O, Yáñez M, Tortajada J, Salpin JY (2008) J Phys Chem B 112:5479–5486
79. Boys SF, Bernardi F (1970) Mol Phys 19:553–&.
80. Linstrom PJ, MAllard WG (2005) NIST chemistry webbook. Standrard Reference Database Number 69. National Institute of Standards and Technology, Gaithersburg, MD
81. Burk P, Koppel IA, Koppel I, Kurg R, Gal JF, Maria PC, Herreros M, Notario R, Abboud JLM, Anvia F, Taft RW (2000) J Phys Chem A 104:2824–2833
82. Remko M, Rode BM (2000) J Mol Struct: THEOCHEM 505:269–281
83. Schröder D, Schwarz H (1999) J Phys Chem A 103:7385–7394

84. Blades AT, Jayaweera P, Ikonomou MG, Kebarle P (1990) Int J Mass Spectrom Ion Proc 102:251–267
85. Blades AT, Jayaweera P, Ikonomou MG, Kebarle P (1990) Int J Mass Spectrom Ion Proc 101:325–336
86. Blades AT, Jayaweera P, Ikonomou MG, Kebarle P (1990) J Chem Phys 92:5900–5906
87. Jayaweera P, Blades AT, Ikonomou MG, Kebarle P (1990) J Am Chem Soc 112:2452–2454
88. Stace AJ (2002) J Phys Chem A 106:7993
89. Rodriguez-Cruz SE, Jockusch RA, Williams ER (1998) J Am Chem Soc 120:5842–5843
90. Peschke M, Blades AT, Kebarle P (1998) J Phys Chem A 102:9978–9985
91. Pavlov M, Siegbahn PEM, Sandstrom M (1998) J Phys Chem A 102:219
92. Rodriguez-Cruz SE, Jockusch RA, Williams ER (1999) J Am Chem Soc 121:1986–1987
93. Rodriguez-Cruz SE, Jockusch RA, Williams ER (1999) J Am Chem Soc 121:8898–8906
94. Feil S, Koyanagi GK, Bohme DK (2009) Int J Mass Spectrom 280:38–41
95. Spears KG, Fehsenfeld FC (1972) J Chem Phys 56:5698–5705
96. Spears KG, Fehsenfeld GC, McFarland M, Ferguson EE (1972) J Chem Phys 56:2562–2566
97. Beyer M, Williams ER, Bondybey VE (1999) J Am Chem Soc 121:1565–1573
98. Shvartsburg AA, Siu KWM (2001) J Am Chem Soc 123:10071–10075
99. Shvartsburg AA, Wilkes JG, Lay JO, Siu KWM (2001) Chem Phys Lett 350:216
100. Shvartsburg AA, Wilkes JG (2002) J Phys Chem A 106:4543
101. Kohler M, Leary JA (1997) J Am Soc Mass Spectrom 8:1124–1133
102. Shvartsburg AA (2003) Chem Phys Lett 376:6
103. Xiao CY, Walker K, Hagelberg F, El-Nahas AM (2004) Int J Mass Spectrom 233:87–98
104. Wu JH, Liu D, Zhou JG, Hagelberg F, Park SS, Shvartsburg AA (2007) J Phys Chem A 111:4748–4758
105. Feil S, Koyanagi GK, Viggiano AA, Bohme DK (2007) J Phys Chem A 111:13397–13402
106. Shi TJ, Siu KWM, Hopkinson AC (2006) Int J Mass Spectrom 255:251–264
107. Cimas A, Gamez JA, Mó O, Yáñez M, Salpin JY (2008) Chem Phys Lett 456:156–161
108. Luna A, Amekraz B, Morizur JP, Tortajada J, Mó O, Yáñez M (2000) J Phys Chem A 104:3132–3141
109. Schröder D, Weiske T, Schwarz H (2002) Int J Mass Spectrom 219:729–738
110. Rodriguez-Santiago L, Noguera M, Sodupe M, Salpin JY, Tortajada J (2003) J Phys Chem A 107:9865–9874
111. Kohno J, Mafune F, Kondow T (1999) J Phys Chem A 103:1518–1522
112. Rodríguez-Santiago L, Sodupe M, Tortajada J (2001) J Phys Chem A 105:5340–5347
113. Hoppilliard Y, Ohanessian G, Bourcier S (2004) J Phys Chem A 108:9687–9696
114. Hoyau S (2001) Pélicier J P, Rogalewicz F, Hoppilliard Y, Ohanessian. Eur J Mass Spectrom 7:303–311
115. Hoyau S, Ohanessian G (1997) J Am Chem Soc 119:2016–2024
116. Strittmatter EF, Lemoff AS, Williams ER (2000) J Phys Chem A 104:9793–9796
117. Ai HQ, Bu YX, Li P (2003) Int J Quant Chem 94:205–214
118. Kapota C, Lemaire J, Maitre P, Ohanessian G (2004) J Am Chem Soc 126:1836–1842
119. Dunbar RC, Polfer NC, Oomens J (2007) J Am Chem Soc 129:14562
120. Bush MF, Oomens J, Saykally RJ, Williams ER (2008) J Am Chem Soc 130:6463–6471
121. Lippert B (2000) Coord Chem Rev 200:487–516
122. Burrows CJ, Muller JG (1998) Chem Rev 98:1109–1151
123. Franska M (2007) Eur J Mass Spectrom 13:339–346
124. Cerda BA, Wesdemiotis C (1996) J Am Chem Soc 118:11884–11892
125. Hall BJ, Brodbelt JS (1999) J Am Soc Mass Spectrom 10:402–413

126. Shen N, Pope RM, Dearden DV (2000) Int J Mass Spectrom 195–196:639
127. Hu PF, Gross ML (1992) J Am Chem Soc 114:9153–9160
128. Hu PF, Gross ML (1992) J Am Chem Soc 114:9161–9169
129. Nemirovskiy OV, Gross ML (1998) J Am Soc Mass Spectrom 9:1020–1028
130. Nemirovskiy O, Giblin DE, Gross ML (1999) J Am Soc Mass Spectrom 10:711–718
131. Williams SM, Brodbelt JS (2004) J Am Soc Mass Spectrom 15:1039–1054
132. Fura A, Leary JA (1993) Anal Chem 65:2805–2811
133. Harvey DJ (2001) J Am Soc Mass Spectrom 12:926–937
134. Zaia J, Costello CE (2003) Anal Chem 75:2445–2455

CHAPTER 2

FROM THE GAS PHASE TO A LIPID MEMBRANE ENVIRONMENT: DFT AND MD SIMULATIONS OF STRUCTURE AND DYNAMICS OF HYDROGEN-BONDED SOLVATES OF BIFUNCTIONAL HETEROAZAAROMATIC COMPOUNDS

ALEXANDER KYRYCHENKO[1] AND JACEK WALUK[2]
[1]*V.N. Karazin Kharkov National University, Institute of Chemistry, 4 Svobody Sq., 61077, Kharkov, Ukraine,*
[2]*Polish Academy of Science, Institute of Physical Chemistry, Kasprzaka 44, 01-224, Warsaw, Poland,*
e-mail: waluk@ichf.edu.pl

Abstract: We present a review of our recent developments in computational modeling of hydrogen-bonding-induced phenomena in a series of biologically relevant bifunctional proton donor–acceptor heteroazaaromatic compounds. Different types of hydrogen-bonded solvates, in which water or alcohol molecules form a bridge connecting the proton donor (pyrrole NH group) and the acceptor (pyridine or quinoline nitrogen) atoms of bifunctional solutes, are explored by combining density functional theory (DFT) and molecular dynamics (MD) simulation approaches. Structure and dynamics of multiple hydrogen-bonded solute-solvent complexes are studied starting from isolated complexes in the gas phase, elucidating their solvation dynamics in solutions and, finally, in a heterogeneous environment of a lipid bilayer. Our results indicate that the structure, stoichiometry and hydrogen bond strength in such solvates are tuned by local topologies of the hydrogen-bonding sites of a bifunctional proton donor–acceptor molecule. A role of such solvates in hydrogen-bond-dependent photophysics and in controlling excited-state behavior of heteroazaaromatic compounds is discussed.

Keywords: Excited-state proton transfer, Hydrogen-bonded complexes, Hydrogen-bonded network, Density functional theory, TDDFT, Fluorescent probe, Lipid bilayer, Molecular dynamics simulations

P. Paneth and A. Dybala-Defratyka (eds.), *Kinetics and Dynamics*,
Challenges and Advances in Computational Chemistry and Physics 12,
DOI 10.1007/978-90-481-3034-4_2,

2.1. INTRODUCTION

Heteroazaaromatic molecules containing both pyrrole/indole/carbazole and quinoline/pyridine heterocyclic units reveal complex spectroscopy and photophysics in hydroxylic solvents [1]. In such bifunctional compounds one moiety acts as a hydrogen bond donor, whereas the other plays a role of a hydrogen bond acceptor. The electronic excitation of such species can induce remarkable and cooperative acid–base changes, occurring on the opposite parts of an electronically excited chromophore. These changes provide a strong driving force for proton translocation from the NH donor moiety towards the accepting pyridine-type nitrogen atom [2]. If geometrical criteria are favorable, i.e., the donor and acceptor are hydrogen-bonded, such a proton transfer reaction can occur intramolecularly, as shown schematically in Figure 2-1a [3–7]. If it is not the case, the reaction can still occur; however, it now requires assistance of external hydrogen-bonding partners (Figure 2-1b) [8,9]. It has been demonstrated that in the presence of protic solvents, bifunctional hydrogen-bond donor–acceptor compounds are able to form cyclically hydrogen-bonded complexes with the solvent, in which solvent molecules act as a proton wire connecting the donor and acceptor parts of the chromophore. In such cyclically hydrogen-bonded solute–solvent complexes, solvent-assisted excited-state multiple proton transfer becomes possible [10–13].

One of the best known systems in this field is the dimer of 7-azaindole, for which the excited-state double proton transfer reaction was observed for the first time [14]. In such symmetric, doubly hydrogen-bonded system two possible, stepwise versus

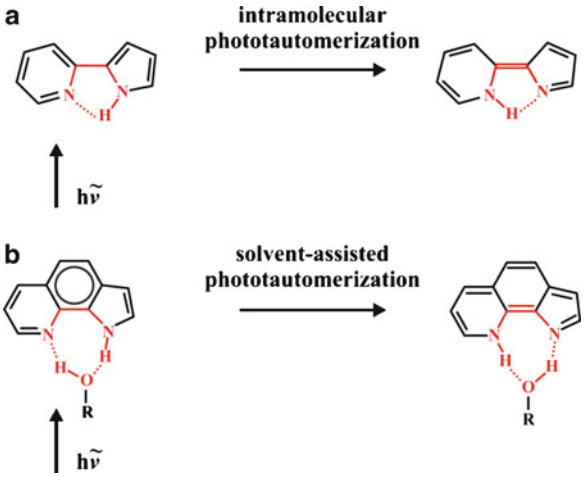


Figure 2-1. (**a**) Direct, intramolecular photoinduced proton-transfer tautomerization in 2-(2′-pyridyl) pyrrole. (**b**) Solvent-assisted proton-transfer reaction in a cyclic, hydrogen-bonded complex of 1*H*-pyrrolo[3,2-*h*]quinoline with a solvent molecule

concerted, proton transfer mechanisms have extensively been discussed [15–27]. Other interesting examples of bifunctional hydrogen-bond donor–acceptor molecules are provided by 6- and 7-hydroxyquinolines, in which migration of a proton occurs from a hydroxyl group towards a quinoline nitrogen atom via a solvent molecule bridge. It has been demonstrated that in 7-hydroxyquinoline proton transfer may proceed via solvent bridges composed of alcohols [28–30] or ammonia [31–37]. Proton translocation in 6-hydroxyquinoline has been shown to occur via a hydrogen-bonded molecular wire composed of two molecules of acetic acid [38, 39]. The excited-state multiple-proton relay along a hydrogen-bonded proton wire is believed to play an important role in the emitting mechanism of green fluorescent protein and many of its mutants [40–44]. Multiple proton migration in proton pump membrane proteins, such as bacteriorhodopsin, drives the proton transport across a biomembrane of a cell and it also proceeds through a proton wire acting against a pH gradient [45–47]. In many respects, these light-driven biological systems are also important for understanding the cooperative phenomena in proton conductivity and reorganization of the hydrogen-bonded networks.

Examples of several bifunctional molecules that have been studied with regard to structure and dynamics of hydrogen-bonded complexes are shown in Figure 2-2. While presenting our methodology and reviewing the results, we will focus mainly on 1H-pyrrolo[3,2-*h*]quinoline (PQ), a compound which has been investigated most extensively.

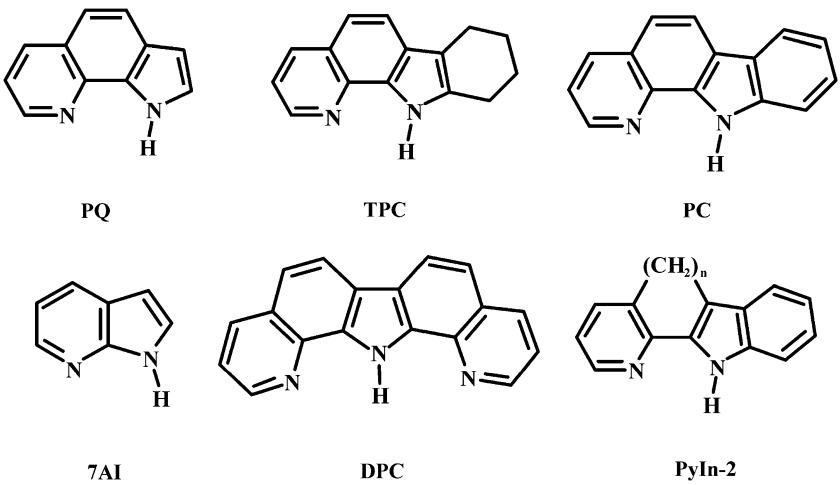


Figure 2-2. Structure of bifunctional hydrogen-bond donor/acceptor compounds: 1H-pyrrolo[3,2-*h*] quinoline (PQ), 7,8,9,10-tetrahydropyrido-[2,3-*a*]carbazole (TPC), pyrido[2,3-*a*]carbazole (PC), 7-azaindole (7AI), dipyrido-[2,3-*a*: 3′,2′-*i*]carbazole (DPC) and 2-(2′-pyridyl)indole, bridged by a chain consisting of two methylene units (n = 2, PyIn-2)

2.2. ELECTRONIC STRUCTURE OF 1H-PYRROLO[3,2-*h*] QUINOLINE

The ground-state (S_0) structure of PQ has been studied by means of the DFT method using the B3LYP hybrid functional and the correlation-consistent polarized valence cc-pVDZ basis set. Although valence excited states are of main interest, diffuse basis functions are required to calculate correctly low-lying Rydberg states, as well as πσ* excited states [48, 49]. Therefore, aug-cc-pVDZ basis functions are added on each of the two N1 and N4 nitrogen atoms and on the pyrrole hydrogen H5 of PQ (Figure 2-3a). Vertical electronic singlet and triplet excitation energies are

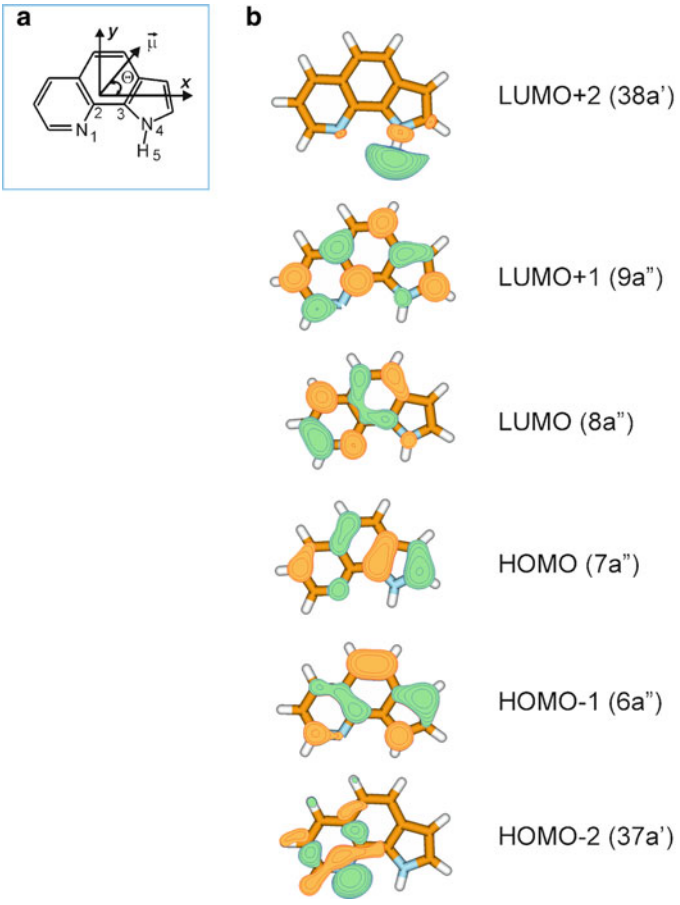


Figure 2-3. (**a**) Structure and atom labeling of PQ. The *arrow* defines a positive angle Θ of the direction of the excited-state transition moments and also, the ground and excited-state permanent dipole moments μ with respect to the long molecular axis *x* [50]. (**b**) Molecular orbitals involved in the low-lying electronic transitions of PQ

calculated for the S_0-geometry of PQ applying the TDDFT methodology and using the same B3LYP functional and the basis set level of theory [50].

Figure 2-3b shows the molecular orbitals participating in the description of $S_0 \rightarrow S_i$ electronic transitions in the low-energy part of the absorption spectrum of PQ. Oscillator strengths, permanent/transition dipole moments and their directions, as well as main orbital configurations are summarized in Table 2-1. The energies of the two lowest electronic transitions, $S_0 \rightarrow S_1$ and $S_0 \rightarrow S_2$, are calculated at 3.89 and

Table 2-1 TD-B3LYP/cc-pVDZ results for isolated PQ: vertical excitation energies (eV), oscillator strengths (*f*), transition dipole moments (M, Debyes), transition moment directions (ω(M), degrees), permanent dipole moments (μ, Debyes), permanent dipole moment directions (Θ(μ), degrees), and main electronic configurations [50] (Reproduced from [50] with permission from *Journal of Physical Chemistry*. A. Copyright © 2006 American Chemical Society)

State	TDDFT	Exp.	*f*	M	ω(M)	μ	Θ(μ)[a]	Main configuration (%)	
Singlet									
$1^1A'$						0.07	66.0		
$2^1A'$ ($\pi\pi^*$)	3.89	3.66[c]	0.0197	1.15	+52.1	4.57	−1.2	H→L	85.8
	(3.52)[b]							H − 1→L + 1	9.1
$3^1A'$ ($\pi\pi^*$)	4.00	3.7–4.0[d]	0.0304	1.42	−10.3	4.25	10.7	H − 1→L	80.8
								H→L + 1	15.0
$1^1A''$ ($n\pi^*$)	4.64	–	0.0020	0.33	[f]	2.19	–	H − 2→L	98.7
$4^1A'$ ($\pi\pi^*$)	4.76	4.6–4.8[d]	0.0795	2.09	+69.4	3.47	1.6	H − 1→L + 1	81.3
								H→L	6.8
$5^1A'$ ($\pi\pi^*$)	4.95	4.6–4.8[d]	0.4949	5.13	−6.4	2.11	−3.4	H→L + 1	68.6
								H − 1→L	13.6
$2^1A''$ ($\pi\sigma^*$)	5.14	–	0.0020	0.30	–	12.38	–	H→L + 2	95.0
$3^1A''$ ($\pi\sigma^*$)	5.25	–	0.0012	0.24	–	13.65	–	H − 1→L + 2	94.8
$4^1A''$ ($n\pi^*$)	5.34	–	0.0001	0.05	[f]	3.07	–	H − 2→L + 1	98.3
$6^1A'$ ($\pi\pi^*$)	5.60	–	0.1953	3.03	13.6	–	–	H − 3→L	63.3
Triplet									
$1^3A'$ ($\pi\pi^*$)	2.85	2.7[e]	–	–	–	1.66	–	H→L	44.7
								H − 1→L	33.0
$2^3A'$ ($\pi\pi^*$)	3.36	–	–	–	–	4.17	–	H→L	40.4
								H − 1→L	52.3
$3^3A'$ ($\pi\pi^*$)	3.64	–	–	–	–	1.75	–	H→L + 1	62.0
								H − 1→L + 1	19.0
$4^3A'$ ($\pi\pi^*$)	3.82	–	–	–	–	1.55	–	H − 1→L + 1	66.0
								H→L + 1	25.7
$1^3A''$ ($n\pi^*$)	4.08	–	–	–	–	2.19	–	H − 2→L	93.8

[a]See Figure 2-3 for angle definition
[b]Zero-point energy corrected value for the energy gap between the $1^1A'$ and $2^1A'$ states
[c]0-0 absorption band in supersonic jet [53]
[d]Absorption band maxima in *n*-hexane and 1-butanol [55]
[e]Phosphorescence band maximum in 1-propanol and ethanol [55]
[f]Transition moment directed along the out-of-plane *z* axis

4.00 eV, respectively. These vertical transitions correspond to $2^1A'$ and $3^1A'$ states. The vertical transition to the lowest excited state ($2^1A'$) has a small oscillator strength and its main electronic configuration (86%) is described by excitation from the highest occupied molecular orbital (HOMO) to the lowest unoccupied molecular orbital (LUMO). The excitation results in a significant intramolecular charge transfer from the pyrrole ring (HOMO) to the quinoline moiety (LUMO) (Figure 2-3b). The transition to the second ($3^1A'$) excited-state is described by the electronic configuration of HOMO-1→LUMO (81%). Similarly to $2^1A'$, this state is also characterized by a significant charge transfer character directed from the pyrrole towards the quinoline part of the chromophore. The S_1 and S_2 states are characterized by permanent dipole moments of 4.57 and 4.25 D, respectively. Comparison with the value of the ground-state dipole moment, 0.07 D, shows that, due to a charge transfer character, the $S_0 \rightarrow S_1$ and $S_0 \rightarrow S_2$ excitations result in a significant increase of the dipole moment of PQ. The $4^1A'$ and $5^1A'$ states are calculated to lie at 4.76 and 4.95 eV, respectively. These states have high oscillator strengths and they appear as a strong absorption band in the electronic spectrum (Figure 1 in reference [55]). The TDDFT calculation also predicts a forbidden electronic transition $1^1A' \rightarrow 1^1A''$ characterized by small oscillator strength of 0.002 and located at 4.64 eV. The orbital analysis indicates that this transition originates mainly from the HOMO-2→LUMO configuration and it may be assigned to an $n\pi^*$ state.

A pair of the $2^1A''$ and $3^1A''$ states is calculated to lie above $4^1A'$ and $5^1A'$. The electronic transitions to these states are described by electronic excitation from HOMO or HOMO-1 to a diffuse, Rydberg-type, LUMO+2 orbital (Figure 2-3b). The latter is of σ^*-type and it is localized on the pyrrole hydrogen atom. The $2^1A''$ and $3^1A''$ states can therefore be assigned as the $\pi\sigma^*$ excited states. The $\pi\sigma^*$ states have a very large excited-state dipole moment of about 12–13 D, as seen in Table 2-1.

The TDDFT-calculated properties of the two lowest excited states reveal evidence for their significant charge-transfer character. Therefore, acid and base properties of PQ in these excited-states are also expected to increase significantly, providing a driving force for the proton transfer reaction. $2^1A'$ and $3^1A'$ states are predicted to lie very close in energy to each other ($\Delta E \sim 0.11$ eV, Table 2-1). Since the TDDFT calculation are carried out for the gas phase, the ordering of these polar $2^1A'$ and $3^1A'$ states may be reversed in polar solutions. Additionally, it has been demonstrated that single-reference methods such as TDDFT tend to overestimate the energy of an excited-state with a multiconfigurational character by 0.2–0.4 eV [51]. The energy difference between $2^1A'$ and $3^1A'$ is in this overestimation range. Therefore, the low-energy part of the spectrum of PQ has been recalculated using the second-order coupled cluster method implemented with the resolution of the identity approximation (RI-CC2) [52]. This approach reproduces the same, $2^1A'$ and $3^1A'$, state ordering.

TDDFT calculations of the $S_0 \rightarrow S_n$ vertical excitation energies have shown a good agreement with the electronic absorption spectra measured in solution, as well as under jet-isolated conditions. The predicted energy of the lowest excited state S_1,

corrected for the zero-point energy terms, is 3.52 eV, which corresponds well with the 0-0 absorption band of 3.66 eV measured for the jet-isolated PQ [53]. The maximum of the strongest absorption band of PQ in *n*-hexane, assigned to the S_2 state, has been measured at about 3.9 eV, in good agreement with the value of 4.0 eV calculated by TDDFT [50].

2.3. STRUCTURE OF GAS-PHASE COMPLEXES

2.3.1. Hydrogen-Bonded Complexes with Water

Hydrogen bonding of PQ with water is considered for 1:1 and 1:2 PQ:water stoichiometries. In the case of 1:1, PQ forms a cyclic, hydrogen-bonded complex with water as shown in Figure 2-4 (center column). Such a complex, possessing two

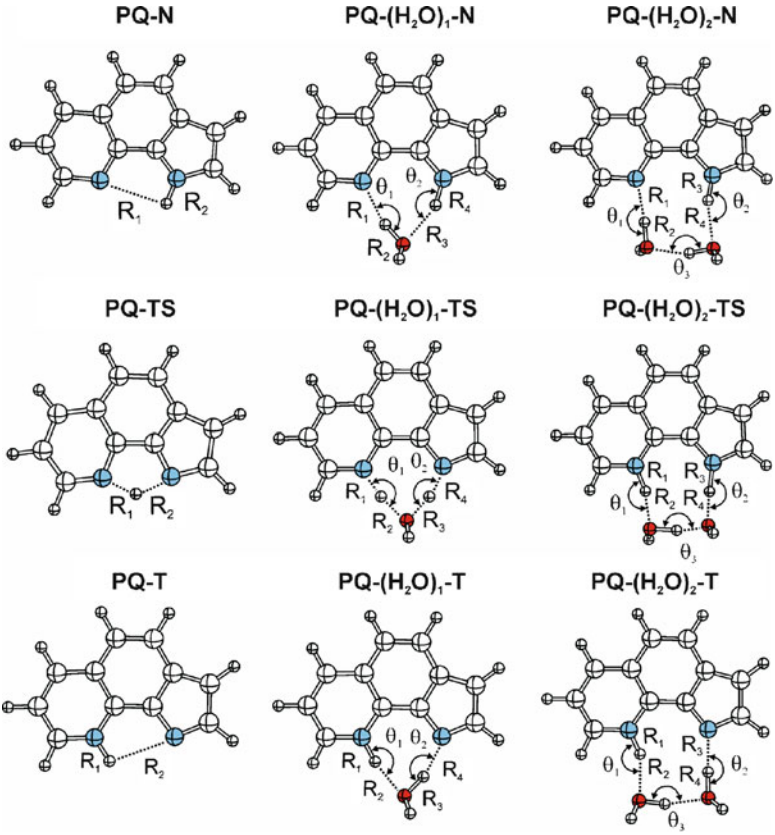


Figure 2-4. Geometry-optimized structures of the stationary points along the proton transfer reaction coordinate in bare PQ (*left*) as well as in PQ–$(H_2O)_1$ (*center*) and PQ–$(H_2O)_2$ (*right*). The proton-transfer relevant structural parameters are summarized in Table 2-2

hydrogen bonds, is calculated as the most energetically stable in the S_0 and S_1 states. The topology of the hydrogen-bond donor (NH) and hydrogen-bond acceptor (quinoline-type nitrogen) favors the formation of two almost linear hydrogen bonds characterized by very similar values of the H-bond lengths and angles, as shown in Table 2-2. The structure of PQ–$(H_2O)_1$ in the S_1-state, optimized using the adiabatic TDDFT methodology [54], reveals that upon electronic excitation the two hydrogen bonds in PQ–$(H_2O)_1$ are strengthened compared to the situation in S_0 (Table 2-2).

In the case of 1:2 PQ:water stoichiometry, different types of hydrogen-bonded solvates are found. The first group of PQ–$(H_2O)_2$ complexes is generally characterized by a structure which is similar to that of the above-described 1:1 cyclically hydrogen-bonded form, PQ–$(H_2O)_1$. In this species, one water molecule forms a cyclic hydrogen-bonded structure with PQ, similarly as in PQ–$(H_2O)_1$, whereas the other water molecule is attached to the first one, acting as either an hydrogen bonding donor or as an acceptor. In terms of the photophysics of PQ, this set of hydrogen-bonded solvates is very similar to that of cyclic 1:1 PQ–$(H_2O)_1$ [53]. However, another type of 1:2 PQ:water complex was found to be energetically more stable. The most energetically favorable structure of PQ–$(H_2O)_2$ is shown in Figure 2-4 (right). In this structure, two water molecules form a cyclic, hydrogen-bonded chain connecting the pyrrole NH hydrogen and the quinoline-type nitrogen atom. The most relevant geometrical parameters of the H-bonded network in PQ–$(H_2O)_2$ are summarized in Table 2-2. The hydrogen bond lengths and angles

Table 2-2 Selected distances (R_i, Å) and angles (Θ_i, degrees) calculated for the stationary points (N – normal, TS – transition state, T – tautomer) along the proton transfer coordinate in PQ–$(H_2O)_n$ in the ground state (in brackets) and in the first excited singlet state [50] (Reproduced from [50] with permission from *Journal of Physical Chemistry* A. Copyright © 2006 American Chemical Society)

R_i and Θ_i	PQ			PQ–H_2O			PQ–$(H_2O)_2$			
	N	TS	T	N	TS	T	N	TS_a	TS_b	T
R_1	2.63	1.34	1.01	1.67	1.26	1.02	1.67	1.14	1.02	1.02
	(2.79)	(1.46)	(1.02)	(1.82)	(1.13)	(1.05)	(1.78)	(–)	(1.09)	(1.03)
R_2	1.01	1.26	2.55	1.02	1.24	1.89	1.02	1.38	1.94	1.95
	(1.01)	(1.42)	(2.54)	(0.99)	(1.42)	(1.72)	(1.00)	(–)	(1.51)	(1.84)
R_3	–	–	–	1.73	1.44	0.99	1.03	1.09	1.15	1.77
				(1.91)	(1.24)	(1.02)	(1.02)	(–)	(1.16)	(1.68)
R_4	–	–	–	1.04	1.12	1.80	1.79	1.51	1.34	1.00
				(1.02)	(1.26)	(1.67)	(1.91)	(–)	(1.16)	(1.01)
Θ_1	–	–	–	165.7	167.3	167.9	175.6	171.0	153.7	154.4
				(165.3)	(165.7)	(168.9)	(173.1)	(–)	(164.8)	(154.6)
Θ_2	–	–	–	156.1	158.7	160.1	162.1	169.1	175.3	177.4
				(158.9)	(160.8)	(160.4)	(163.8)	(–)	(175.6)	(178.9)
Θ_3	–	–	–	–	–	–	162.3	167.3	160.8	163.2
							(161.7)	(–)	(168.5)	(164.1)

in the cyclic H-bond network are found not to be equal, indicating the existence of some sterical strain in the PQ–$(H_2O)_2$ complex. The steric strain is partially reduced by out-of-plane rearrangement of the triple hydrogen-bonded chain, so that the oxygen atom of one of the water molecules is shifted below the molecular plane of PQ.

2.3.2. Excited-State Proton Transfer Through Water Bridges

The TDDFT analysis of the properties of PQ in the lowest excited singlet state reveals the evidence for the important role the charge-transfer type excitation may play in the excited-state behavior of the bifunctional heteroazaaromatic compounds. Our recent works on a number of bifunctional systems indicate that microsolvation involving hydrogen-bonding may trigger the excited-state translocation of proton from a proton-donating imino moiety to a proton-accepting pyridine-type nitrogen atom [1, 8, 9, 12]. Experimental spectroscopic studies have also shown that the phototautomerization is not the only excited state deactivation channel switched on by the formation of a hydrogen-bonded complex with a protic partner. Another channel is a rapid $S_0 \leftarrow S_1$ internal conversion [55]. The experiment suggests that the latter deactivation path is favored for 1:2 species, whereas 1:1 complexes undergo excited state proton transfer. We have examined the energetics and the potential energy profiles leading from the normal (N) form to the tautomer (T) via the transition state (TS) applying the adiabatic TDDFT method. In order to consider the proton transfer reaction path in PQ and PQ–$(H_2O)_n$ solvates with n = 1,2 geometries of PQ-N, PQ-TS and PQ-T stationary points were optimized for the S_0 and S_1 states. As shown in Figure 2-3a, in the normal form PQ-N, before the phototautomerization, the migrating proton H5 is attached to the pyrrole nitrogen atom N4. In the tautomeric form PQ-T, this proton becomes covalently attached to the nitrogen atom N1 of the quinoline-type moiety (Figure 2-4). The energy of the initial and final states of the reaction is summarized in Table 2-3, in which ΔE refers to the difference in total electronic energy between the corresponding stationary points. The energy of the normal form was set to zero. Table 2-3 also shows the enthalpy difference, ΔH_0, which includes the zero point vibrational energy (ΔZPE) correction term to ΔE. In the S_0 state, in the absence of hydrogen-bonding with water, the PQ-N form is characterized by the lowest energy, corresponding to the global minimum on the S_0 potential energy surface. The energy of the ground state tautomer PQ-T is calculated to be higher than PQ-N by 18.6 kcal/mol in terms of ΔH_0. In S_1, however, the reversed energy ordering of PQ-N* and PQ-T* is observed. The TDDFT-optimized structure of PQ-T* is found to be most energetically stable, being lower in energy than PQ-N* by 16.4 kcal/mol. Thus, the TDDFT calculations demonstrate that the reaction N*→T* is strongly energetically favorable. Despite of the favorable energetics, the energy barrier to the intramolecular proton transfer is, however, found to be large. The barriers of 37.8 and 20.9 kcal/mol are calculated for S_0 and S_1 states, respectively. In the

Table 2-3 Proton transfer energetics (*kcal/mol*) for the stationary points calculated along the proton transfer reaction path in PQ–$(H_2O)_n$. (Reproduced from [50] with permission from *Journal of Physical Chemistry* A. Copyright © 2006 American Chemical Society)

Structure	Stationary points	Energy (kcal/mol)			
		S_0-state		S_1-state	
		ΔE	ΔH_0	ΔE	ΔH_0
PQ	N	0.0	0.0	0.0	0.0
	TS	37.8	–	20.9	–
	T	21.0	18.6	−17.2	−16.4
PQ–H_2O	N	0.0	0.0	0.0	0.0
	TS	16.1	–	3.0	–
	T	14.6	12.7	−12.5	−11.8
PQ–$(H_2O)_2$	N	0.0	0.0	0.0	0.0
	TS_a	17.8	–	5.6	–
	TS_b	–	–	1.5	–
	T	12.8	12.6	−12.0	−11.4

transition state PQ-TS*, the geometry of the aromatic backbone of PQ is significantly distorted in comparison with the equilibrium structures PQ-N* and PQ-T* (Figure 2-4 and Table 2-2). The large energy barrier for the excited-state proton transfer reaction, estimated by TDDFT, is consistent with the spectroscopic results obtained in nonpolar and polar aprotic solvents. The fluorescence studies show that, due to the large activation energy, no photoinduced proton transfer is spectroscopically detected for PQ in the aprotic media [55].

The tautomerization energetics for the cyclic hydrogen-bonded complexes PQ–$(H_2O)_n$ with one and two water molecules are presented in Table 2-3. In the case of the complex with one water, the TDDFT calculations show that the tautomeric form PQ–$(H_2O)_1$–T* is lower in energy than PQ–$(H_2O)_1$–N* by 12.5 kcal/mol in the S_1 state (Table 2-3). Thus, the formation of the cyclic hydrogen-bonded complex decreases the difference in the excited-state energy between PQ–$(H_2O)_1$–N* and PQ–$(H_2O)_1$–T*. It is interesting that a more pronounced effect is observed for the energy barrier for the transfer of the proton from the pyrrole group to the quinoline nitrogen, assisted by the water molecule. In the cyclic complex, the steric strains within the hydrogen-bonded network are reduced, and this effect leads to the stabilization of the energy of the transition state structure PQ–$(H_2O)_1$–TS*, so that the barrier height for double proton transfer is reduced from 20.9 to 3.0 kcal/mol. Table 2-2 shows that when the system is passing the stationary point PQ–$(H_2O)_1$–TS* the two protons move asynchronously. The water proton starts shifting to the quinoline-type nitrogen before the pyrrole proton detachment is initiated. The geometry of PQ–$(H_2O)_1$–TS* indicates that this proton is located midway between the water oxygen and the quinoline nitrogen atom, as can be noticed from the distances R_1 and R_2, equal to 1.26 and 1.24 Å, respectively

(Figure 2-4 and Table 2-2). The TDDFT calculations predict, therefore, that the phototautomerization in PQ–$(H_2O)_1$ occurs by a concerted, asynchronous double proton transfer, which is assisted by a water molecule forming a hydrogen-bonded wire connecting the proton-donor and acceptor atoms of PQ.

A more complex proton transfer profile was found for the phototautomerization in PQ–$(H_2O)_2$. Two transition states TS_a* and TS_b* were identified along the excited-state reaction path connecting PQ–$(H_2O)_2$–N* and PQ–$(H_2O)_2$–T*. The order of the TS_a* and TS_b* transition states on the pathway from the initial N* state to the final T* state is mainly determined by the structural similarity of TS_a* and TS_b* to either N* or T*. As can be seen from Table 2-2, the values of the hydrogen bond lengths R_3 and R_4 in TS_a* ($R_3 = 1.09$ Å and $R_4 = 1.51$ Å) are closer to those of the initial state N* ($R_3 = 1.03$ Å and $R_4 = 1.79$ Å) rather than to the final state T* ($R_3 = 1.77$ Å and $R_4 = 1.00$ Å). The structural difference between N* and TS_a* is located, therefore, in different values of R_1 and R_2. Upon going from N* to TS_a*, R_1 and R_2 are rearranged from $R_1 = 1.67$ Å and $R_2 = 1.02$ Å (in N*) to $R_1 = 1.14$ Å and $R_2 = 1.38$ Å (in TS_a*). This structural transformation corresponds to the transfer of one proton from the water molecule to the nitrogen atom of the quinoline ring of PQ. In TS_a*, the other two protons involved in tautomerization are still connected to the corresponding heavy atoms. On the other hand, the structure of the TS_b* state is closer to that of the final state T*, because R_1 and R_2 are similar in both TS_b* ($R_1 = 1.02$ Å and $R_2 = 1.94$ Å) and T* ($R_1 = 1.02$ Å and $R_2 = 1.95$ Å). That means that in order to reach the TS_b* state directly, the system should already have R_1 and R_2 rearranged, but this R_1 and R_2 rearrangement is actually passing through TS_a*. Based on these considerations, the phototautomerization in PQ–$(H_2O)_2$ proceeds as N*→ TS_a*→ TS_b*→T*. The first state TS_a* is characteristic for highly asynchronous concerted proton movement along the hydrogen-bonded network (Table 2-2). In this state, the transfer of one proton from the oxygen atom of the first water to the quinoline nitrogen atom becomes almost complete, whereas the other two transferring protons are still attached to the pyrrole nitrogen and the oxygen atom of the second water, respectively. In the case of the second transition state TS_b*, one proton is moved from the water molecule and becomes attached to the quinoline nitrogen, whereas the pyrrole hydrogen only starts moving towards the water dimer. The ΔE barriers between the initial form, PQ–$(H_2O)_2$–N* and the corresponding transition states, PQ–$(H_2O)_2$–TS_a* and PQ–$(H_2O)_2$–TS_b* are calculated to be 5.6 and 1.5 kcal/mol, respectively. Thus, the first transition state determines the overall rate for the triple proton transfer in PQ–$(H_2O)_2$. The analysis of bond lengths and angles in the hydrogen-bonded network of PQ–$(H_2O)_2$ connecting the hydrogen-bond donor and acceptor atoms indicates that the complete tautomerization is not able to proceed by simultaneous and cooperative transfer of the three protons. This type of multiple proton transfer along a hydrogen-bonded proton wire has been reported for the transfer via hydrogen-bonded water bridges composed of equivalent or similar hydrogen bonds [56–60]. Such concerted reorganization of a hydrogen-bonded network may proceed via cooperative breaking and re-forming of hydrogen bonds possessing very similar geometry. It has also

been reported that the cooperative transfer may be blocked in the case of unfavorable geometry of one of hydrogen-bonds participating in a bridge [61–64]. The significant nonequivalence of the three hydrogen bonds in PQ–$(H_2O)_2$–N* leads to strongly asynchronous movement of the two transferring protons. The triple proton transfer occurs, therefore, in a step-by-step fashion through PQ–$(H_2O)_2$–TS_a* and PQ–$(H_2O)_2$–TS_b*, which requires a significant rearrangement of the positions of heavy oxygen atoms of the water cluster. This asynchronous transfer leads to the situation when, at a certain time, only one hydrogen bond is broken along the hydrogen-bonded bridge. After this single proton transfer step, the hydrogen-bonded network is being rearranged and, after that, the transfer of the second proton requires less activation energy. Therefore, we found that even though tautomerization in PQ–$(H_2O)_2$ is possible, it requires large relaxation and reorganization of the hydrogen-bonded water bridge. This reorganization may, therefore, trigger the excited state depopulation of PQ via internal conversion and thus block the complete tautomerization in this type of microsolvates.

2.3.3. Hydrogen-Bonded Complexes with Methanol

The hydrogen-bonded complexes of 7-azaindole (7AI, Figure 2-2) with protic solvent molecules have previously been investigated at various levels of theory [65]. The ground-state structure of 7AI complexes with the smallest protic solvent, water, have been studied in detail, including such methods as DFT [66, 67] and MP2 [68]. The excited-state structure and properties of 7AI–$(H_2O)_n$ have been evaluated using ab initio approach employing a configuration interactions scheme (CIS), involving single excited configurations [69–72], as well as high level multiconfigurational schemes such as MCSCF [73] and CASSCF [74]. Recently, geometries, electronic structure, excited-state permanent dipole moments, as well as the orientations of transition dipole moments in 7AI–$(H_2O)_n$ clusters have been re-evaluated using robust, second-order approximated coupled cluster model within the resolution-of-the identity approximation (RI-CC2) [75,76]. The RI-CC2 structure calculations combined with laser-induced fluorescence experiments have also been reported for jet-cooled clusters of 7AI with ammonia [77]. Sekiya and coworkers have applied supersonic jet-cooled fluorescence experiments combined with quantum chemistry calculations to study the structures of microsolvated clusters of 7AI with methanol molecules. They reported the ground-state structures of 7AI–$(MeOH)_n$ with $n = 2, 3$ and the excited state proton transfer reaction path calculations carried out at MP2/6-31++G** and CIS/6-31++G** levels, respectively [62,78]. The structure and the reaction path for the proton transfer have also been reported for 7AI/carboxylic acid complexes calculated at Hartree-Fock/DFT [79] and CIS levels [80]. These computational studies have shown that the structure and proton transfer photoreactivity of hydrogen-bonded complexes of 7AI with various protic partners are strongly dependent on the structural topology of solute/solvents hydrogen-bonding centers.

Here we have compared the ground-state structure of methanol complexes of 7AI with those of two other bifunctional compounds, PQ and PyIn-2 (Figure 2-2). In the latter ones, the topology of hydrogen-bonding sites differs from that of 7AI. The structure of hydrogen-bonded complexes with one and two methanol molecules studied at the B3LYP/6-31G(d,p) level is presented in Figure 2-5. Hydrogen-bond lengths and angles are summarized in Table 2-4. All three solutes form two kinds of solute-solvent complexes with methanol: a π-complex and a hydrogen-bonded

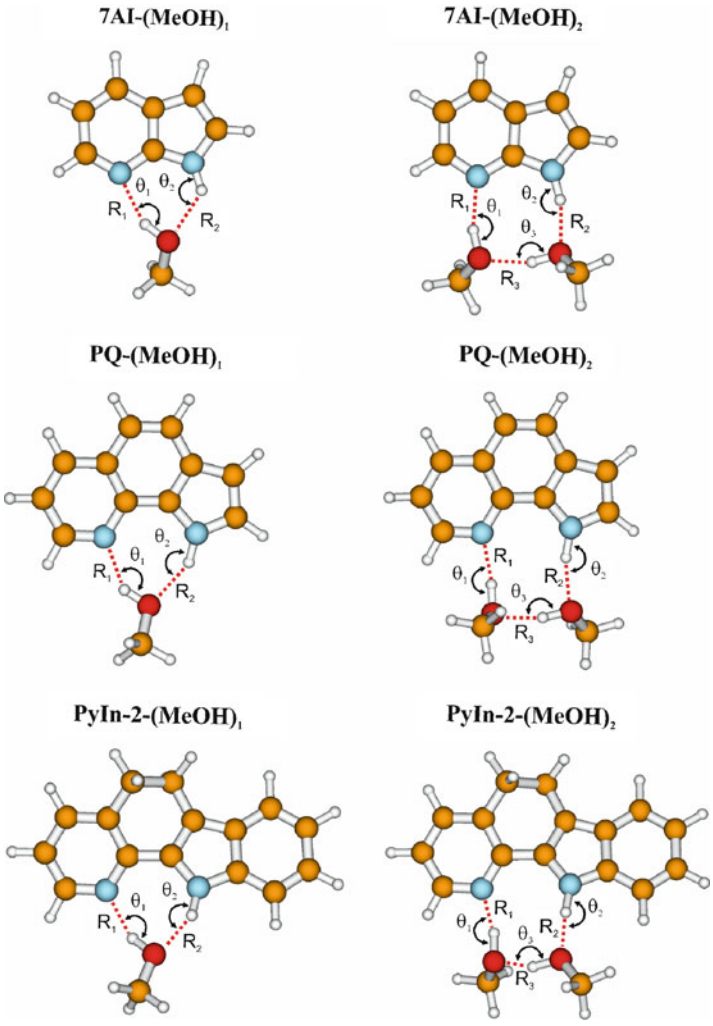


Figure 2-5. B3LYP/6-31G(d,p) optimized geometries of 1:1 and 1:2 cyclic hydrogen-bonded complexes of 7AI, PQ and PyIn-2 with methanol [83–85]. The most important hydrogen-bond structural parameters (R_i and Θ_i) are collected in Table 2-4

Table 2-4 Selected hydrogen bond lengths (R_i, Å) and angles (Θ_i, degrees) for methanol complexes of 7AI–$(MeOH)_n$, PQ–$(MeOH)_n$ and PyIn-2-$(MeOH)_n$ with n = 1, 2 calculated at the B3LYP/6-31G(d,p) level [83]

Hydrogen-bond parameters	7AI		PQ		PyIn-2	
	7AI + 1 MeOH	7AI + 2 MeOH	PQ + 1 MeOH	PQ + 2 MeOH	PyIn-2 + 1 MeOH	PyIn-2 + 2 MeOH
R_1	1.95	1.81	1.84	1.79	1.85	1.79
R_2	2.00	1.79	1.84	1.85	1.87	1.84
R_3	–	1.72	–	1.72	–	1.74
Θ_1	149.7	170.8	163.2	170.4	164.1	172.7
Θ_2	136.2	172.3	157.5	162.5	158.6	166.6
Θ_3	–	165.1	–	164.8	–	159.7
ΔE[a]	−7.8 −8.5[b] −9.2[c]	−18.4 −17.4[b]	−9.9	−15.1	−8.1	−14.8
ΔH[d]	−6.3 ± 1.5[e]	–	−9.2 ± 0.7[f]	–	−8.7 ± 0.2[g]	–

[a]Binding energy in kcal/mol
[b]Calculated at the MP2/6-31++G** level with applying BSSE and ZPE corrections [78]
[c]Calculated at the HF/6-31G* level with no BSSE corrections [81]
[d]Experimental binding enthalpy in kcal/mol
[e]Methanol complex in cyclohexane [81]
[f]Complex with *n*-butanol in *n*-hexane [55]
[g]Methanol complex in *n*-hexane [82]

complex. The most stable structures are found to be hydrogen-bonded species. For all three solutes, the geometry optimization of a hydrogen-bonded complex with a 1:1 stoichiometry has always been converging to a cyclic, NH···OH···N, double hydrogen-bonded structure, as shown in Figure 2-5 (left column). The structural analysis of 1:1 methanol complexes has shown that in PQ–$(MeOH)_1$ and PyIn-2-$(MeOH)_1$ the two hydrogen bonds, characterized by Θ_1 and Θ_2 angles, as indicated in Figure 2-5 and Table 2-4, are more linear than those in 7AI–$(MeOH)_1$. On the basis of simple electrostatic and hybridized valence bond considerations a linear hydrogen-bond is expected to be energetically preferred, so that the strength of hydrogen-bonding association is well correlated with simple hydrogen-bonding geometrical parameters, such as the X–H···Y bond length and angle. Table 2-4 shows that the largest deviations from the idealized hydrogen-bonded geometry within the series of the 1:1 solute-methanol complexes are observed for 7AI–$(MeOH)_1$. The geometrical criteria for the hydrogen bond strength in the 1:1 complexes also correlate well with the energy of their formation, ΔE, calculated at the same B3LYP/6-31G(d,p) level and corrected for zero-point energy (ZPE) and basis set superposition error (BSSE) contributions (Table 2-4). The energy of the formation of the 1:1 complex has been calculated as 7.8, 8.1 and 9.9 kcal/mol for 7AI–$(MeOH)_1$, PyIn-2-$(MeOH)_1$, and PQ–$(MeOH)_1$, respectively. Thus, it has been concluded that PQ forms the strongest hydrogen-bonded complex with methanol, whereas the hydrogen-bonding

in 7AI–$(MeOH)_1$ is the weakest among the three studied compounds. The DFT-computed binding energies are also in good agreement with the experimentally determined enthalpies of hydrogen-bonding association, ΔH (Table 2-4). Chou and co-authors have reported the enthalpy of the formation of the hydrogen-bonded complex between 7AI and one methanol molecule in a cyclohexane solution as -6.3 ± 1.5 *kcal/mol* [81]. For the hydrogen-bonding association between PyIn-2 and methanol in *n*-hexane solution the enthalpy is -8.7 ± 0.2 kcal/mol [82]. Using spectroscopic titration methods, we have estimated the enthalpy for the association reaction between PQ and one *n*-butanol molecule in *n*-hexane solution as -9.2 ± 0.7 kcal/mol [55]. Therefore, both experimental and theoretical findings have provided substantial evidence that the hydrogen-bonding strength decreases in the following sequence: PQ–$(MeOH)_1$ > PyIn-2-$(MeOH)_1$ > 7AI–$(MeOH)_1$.

In the case of 1:2 solute-methanol complexes, a cyclic, triply hydrogen-bonded complex containing two methanol molecules, bound in a closed OH hydrogen-bonded network, has been found as the most stable structure compared to other possible hydrogen-bonded configurations for all three bifunctional solutes, as shown in Figure 2-5. Table 2-4 demonstrates that, in contrast to 7AI–$(MeOH)_1$, in the case of 7AI–$(MeOH)_2$ two hydrogen bond angles Θ_1 and Θ_2 have approached the values which are close to the ideal hydrogen-bond angle of 180°. The hydrogen bond lengths in the triple bound "eight-membered ring" structure of 7AI–$(MeOH)_2$ have become shorter than those in 7AI–$(MeOH)_1$, indicating the increase in the hydrogen bond strength in the former complex. Hydrogen bonds in cyclic "nine-membered ring" structures of PQ–$(MeOH)_2$ and PyIn-2-$(MeOH)_2$ have also rearranged toward the idealized hydrogen bond parameters. Despite of an overall increase in the hydrogen bonding strength among the 1:2 solute-methanol complexes, the opposite tendency has been observed in the binding energy for the 1:2 cyclic, triply hydrogen-bonded complexes than that obtained for the 1:1 hydrogen-bonded solvates. As can be seen in Table 2-4, the B3LYP/6-31G(d,p) binding energies are 15.1, 14.8 and 18.4 kcal/mol for PQ–$(MeOH)_2$, PyIn-2-$(MeOH)_2$ and 7AI–$(MeOH)_2$, respectively. Thus, 7AI–$(MeOH)_2$ has been found to form the most stable complex among the three studied bifunctional solutes. The larger exothermicity of the complex formation of 7AI–$(MeOH)_2$ can be explained by the possibility of the adjustment of the hydrogen bond distances and angles within the triple hydrogen-bonded solvent wire toward an idealized wire configuration, in which each N–H···O and N···H–O bonds would be almost linear.

We have also re-evaluated the ground-state structure and normal vibrational modes of PQ–$(MeOH)_n$ using more precise ab initio methodology based on the resolution of identity (RI) method combined with the MP2 procedure referred to as RI-MP2/TZVP [86, 87]. Applying the TDDFT methodology described in Section 2.3.1 for PQ–$(H_2O)_n$ solvates, we have studied the influence of the $S_1 \leftarrow S_0$ electronic excitation on the structure of hydrogen-bonded complexes in the lowest singlet-excited state at the TD-B3LYP/cc-pVDZ level. Crucial hydrogen bond parameters for PQ–$(MeOH)_n$ with $n = 1, 2$ are summarized for both the S_0 and S_1 states in Table 2-5. Our DFT and RI-MP2 studies on the structure and

Table 2-5 Selected hydrogen bond lengths (R_i, Å) and angles (Θ_i, degrees) for methanol complexes of PQ–$(MeOH)_n$ with n = 1, 2 in the S_0 and S_1 states calculated at the B3LYP/cc-pVDZ level [85]. The hydrogen bond parameters for the S_0 state are also compared with the RI-MP2/TZVP data

Hydrogen-bond parameters	PQ + 1 MeOH			PQ + 2 MeOH		
	MP2	DFT		MP2	DFT	
	S_0	S_0	S_1	S_0	S_0	S_1
R_1	1.84	1.83	1.68	1.79	1.80	1.72
R_2	1.87	1.88	1.71	1.77	1.87	1.73
R_3	–	–	–	1.76	1.74	1.67
Θ_1	162.5	165.3	164.8	174.7	174.5	176.8
Θ_2	157.5	156.8	159.1	173.1	166.0	166.8
Θ_3	–	–	–	154.6	162.4	163.4
ΔE	–	−10.4	–	–	−13.9[a]	–

[a]The binding energy was estimated for the binding interaction between PQ and a hydrogen-bonded methanol dimer, so that ΔE does not include a contribution of methanol-methanol hydrogen-bonding, which was calculated to be −3.9 kcal/mol at the same level of theory

hydrogen-bonding interactions of PQ with one and two methanol molecules have shown that even DFT calculations with medium size basis sets are entirely sufficient to provide a correct picture of hydrogen bond structures of PQ, which is evident from comparison with the structures obtained by the more accurate RI-MP2 method (Tables 2-4 and 2-5). As it was described in Section 2.2, the $S_1 \leftarrow S_0$ electronic transition is accompanied by the intramolecular charge transfer from the pyrrole ring toward the quinoline moiety. Therefore, the electronic excitation should result in an overall strengthening of the solute–solvent hydrogen-bonding in the S_1-state. This hydrogen bond strengthening in S_1 has indeed been observed for both PQ–$(MeOH)_1$ and PQ–$(MeOH)_2$, as can be noticed from Table 2-5.

2.3.4. Cluster Size Effect on Fluorescence Quenching in Hydrogen-Bonded Complexes of PQ with Methanol

It has recently been found that, in the gas phase, photophysical properties of PQ depend strongly on stepwise microsolvation by methanol molecules [84]. In particular, no laser-induced fluorescence (LIF) has been observed for the jet-cooled PQ–$(MeOH)_1$, while strong LIF signals were detected for PQ–$(MeOH)_2$ and its larger size methanol clusters. In PQ–$(MeOH)_2$, the emission was characterized by a structured excitation spectrum up to excess energies of 663 cm^{-1}, above which the LIF spectra were vanishing [85]. The difference in the LIF behavior for PQ–$(MeOH)_n$ clusters with different sizes indicates the existence of a fast relaxation channel in PQ–$(MeOH)_1$ which is not observed in PQ–$(MeOH)_2$. Further

detection and characterization of PQ–methanol clusters by measuring their infrared (IR) spectra in the N–H and O–H stretch regions using femtosecond multiphoton ionization (fsMPI) methodology have allowed to observe a short-lived PQ–methanol complex which was not originally detected using nanosecond laser excitation. Based on IR/fsMPI spectrum and DFT vibrational frequency calculations, the complex possessing a short excited-state lifetime has been assigned to a cyclic, double hydrogen-bonded complex of PQ with one methanol molecule, the structure of which may schematically be outlined as PQ–$(MeOH)_1$ in Figure 2-5 [84]. It has been proposed that the short lifetime of the excited state of PQ–$(MeOH)_1$ is due to a rapid double proton transfer reaction which is followed by nonradiative relaxation. It is interesting that, using the mass-resolved vibronic and the dispersed fluorescence spectra, Sekiya and co-authors have recently observed the photoinduced proton/hydrogen atom relay for supersonically cooled 7AI–$(MeOH)_2$, but not for 7AI–$(MeOH)_1$ [62]. These experimental findings correlate well with our DFT-based conclusions that, in the gas phase, the formation of 7AI–$(MeOH)_2$ is more favorable than of 7AI–$(MeOH)_1$. The former is characterized by the cyclic, triple hydrogen-bonded structure with close-to-ideal hydrogen bond network. In PQ, the two almost perfect hydrogen bonds are formed already in the case of PQ–$(MeOH)_1$, so that the formation of the cyclic, triply hydrogen-bonded structure in PQ–$(MeOH)_2$ introduces some steric perturbations in the hydrogen-bonded OH network. The cluster size effect on the excited-state proton transfer reaction, which is being driven by an electron transfer along the ammonia wire, has also been reported by Leutwyler and coworkers for 7-hydroxyquinoline (7HQ)-ammonia clusters [33,88]. In addition, combining the fluorescence-detected IR spectra in the OH stretching region and DFT-based normal mode analysis, Matsumoto and coworkers have also found that, similarly to ammonia, at least three molecules of methanol are required to form a jet-cooled 7HQ–$(MeOH)_3$ cluster, in which the photoinduced keto-enol tautomerization of 7HQ starts to appear [30].

2.4. HYDROGEN BONDING OF HETEROAZAAROMATICS IN SOLUTION

Bifunctional H-bond donor–acceptor molecules, such as 7AI and PQ, exist in bulk alcohol or water solvents in a wide range of hydrogen-bonded species. Only a selected fraction of the equilibrium population possesses an appropriately hydrogen-bonded structure, able to tautomerize directly and rapidly after electronic excitation [55]. The detailed structure of such a phototautomerization precursor is, however, difficult to determine in bulk solution. To minimize the effect of solvent self-aggregation, the dilution of a small amount of alcohol or water in non-polar hydrocarbon solvents may be considered [55, 89]. Solvation dynamics and hydrogen bonding of 7AI, PQ and PyIn-2 in bulk alcohol or water solutions, as well as in *n*-hexane solution containing only one and two molecules of these

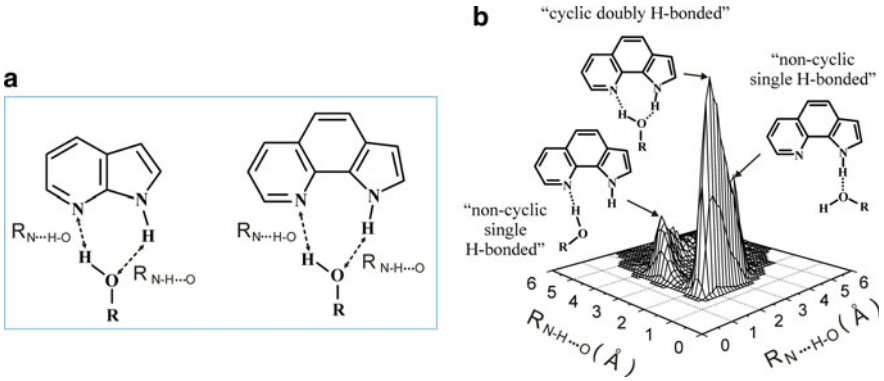


Figure 2-6. (**a**) The definition of two hydrogen-bonding parameters, $R_{N–H\cdots O}$ and $R_{N\cdots H–O}$, in 1:1 cyclic complexes of 7AI and PQ with protic solvents. (**b**) An example of a hydrogen-bonding distance distribution (HBD) displaying a relative frequency of a simultaneous occurrence of a pair of distances simulated for PQ in *n*-hexane in the presence of two methanol molecules. Populations of different kinds of hydrogen-bonded solvates of PQ, involving cyclic and non-cyclic H-bonded species, can be reconstructed from a 3D representation of an HBD distribution. Both $R_{N–H\cdots O}$ and $R_{N\cdots H–O}$ distances are calculated at every step during MD simulations; however, the corresponding H-bond values are included in the HBD distribution only when both H-bond distances are simultaneously found to be smaller than the cutoff of 7 Å

solvents, have been explored by classical molecular dynamics simulations using the GROMOS96 package of programs [83,90].

The bifunctional compounds possess two centers able to form hydrogen bonds with the solvent, acting either as a donor or an acceptor. Hydrogen bond definition and geometrical criteria for 7AI and PQ are shown in Figure 2-6a. The hydrogen bond between the pyridine-type nitrogen atom of the solute and the hydrogen atom of hydroxylic solvents (alcohol or water) is labeled as "N···H–O", while that between the hydrogen atom of the pyrrole NH group and the solvent oxygen atom is referred to as "N–H···O". During MD simulations, the solute-solvent hydrogen bonding is analyzed in terms of hydrogen-bonding distance distribution (HBD distribution), which displays a relative frequency of a simultaneous occurrence of a pair of $R_{N–H\cdots O}$ and $R_{N\cdots H–O}$ distances within hydrogen-bonded complexes, as shown in Figure 2-6b. Such 3D representation of hydrogen-bonding allows one to reconstruct structures and populations of different kinds of H-bonded complexes between the solute and surrounding solvent molecules, accumulated and averaged over a long MD simulation period.

2.4.1. Hydrogen-Bonded Complexes with Methanol and Water

The details of the MD simulations have been presented in reference [83]. Briefly, the simulation was based on a classical MD approach implemented in the GROMOS96 program package [90]. The 43A1 force field was used for modeling

bulk solvent and solvent mixtures, whereas the corresponding force field and charges for the azaaromatics were developed from the gas-phase B3LYP/6-31+G (d,p) structure calculations. The simulations were performed using periodic boundary conditions at the constant number of particles, constant pressure of 1 atm, and constant temperature of 300 K. The simulated systems included 116 and 216 solvent molecules for *n*-hexane, and methanol or water, respectively.

The HBD distributions for 1:1 complexes of 7AI, PQ and PyIn-2 in *n*-hexane solution in the presence of one methanol molecule are shown in Figure 2-7 (top). The distribution is characterized by one peak in which the most probable values of a pair of both hydrogen bond distances span a narrow range from 1.9 Å to about 2.5 Å. The population of hydrogen-bonded species, which may be described by such a kind of hydrogen bonding, corresponds to a complex in which, most of the time, a single methanol molecule is simultaneously H-bonded to both H-bonding donor and acceptor atoms of the solute. The structure and hydrogen bond lengths of such a solvate conform to a cyclic, doubly hydrogen-bonded complex between the solute and one methanol molecule, as shown schematically for a cyclic 1:1 PQ–water complex in Figure 2-6b. The comparison between 7AI, PQ and PyIn-2 shows that in the cases of PQ and PyIn-2 a cyclic, doubly hydrogen-bonded complex has a more "rigid" structure than the analogous cyclic species of 7AI. In the latter case, a fraction of the H-bonded species with one of the distances exceeding the maximum cutoff value of 2.5 Å for H-bonding is observed. Therefore, in such a cyclic solvate of

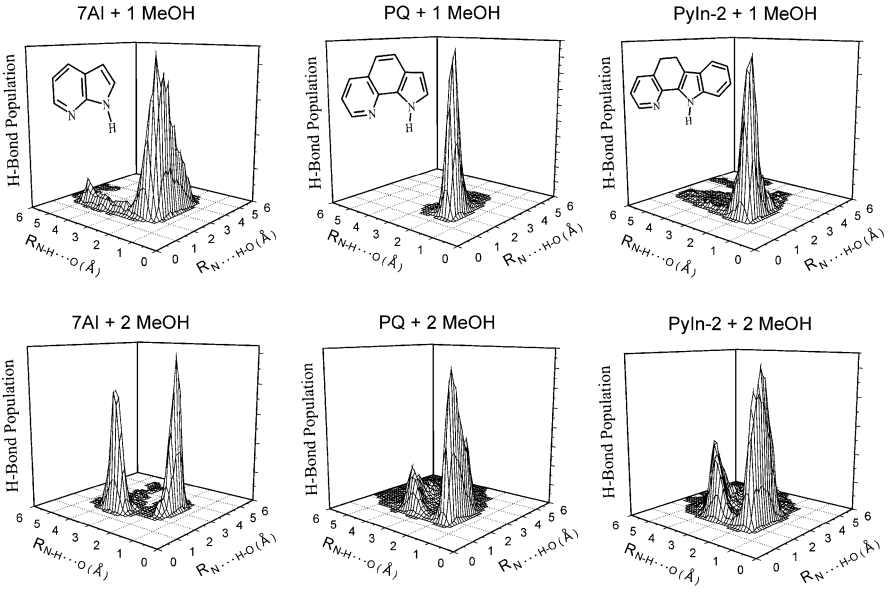


Figure 2-7. The HBD distributions for 7AI, PQ and PyIn-2 simulated for *n*-hexane in the presence of one (*top*) and two (*bottom*) molecules of methanol [83]

7AI, one of the two hydrogen-bonds periodically becomes broken. Applying the two cutoff criteria of 2.5 and 2.2 Å for weak and strong hydrogen bonds, respectively, a relative equilibrium population of the cyclic complexes may be estimated. For 7AI–$(MeOH)_1$, the equilibrium fractions of the cyclic complexes are 37% and 13%, respectively. Thus, only for about 13–37% of the time, 7AI is solvated in a "cyclic manner" and, for the rest of the time, one of the two hydrogen bonds, the N···H–O, is broken more frequently than the N–H···O bond. Such analysis can be used to reconstruct the average structure and hydrogen bonding dynamics for PQ–$(MeOH)_1$ and PyIn-2-$(MeOH)_1$. Populations of the complexes, their average hydrogen bond lengths and relative strength, estimated as the full width at half-maximum (*fwhm*) of the corresponding HBD distribution peak, are summarized in Table 2-6. As can be seen, in the case of PQ–$(MeOH)_1$, the cyclic complex is observed to persist for 69–88% of the simulation time depending on the hydrogen bond cutoff criteria. PyIn-2-$(MeOH)_1$ is found to be an intermediate case between 7AI and PQ in terms of the equilibrium population of the cyclic species, 50–72%, respectively.

Addition of a second methanol molecule to 7AI–$(MeOH)_1$ complex changes the hydrogen bonding pattern in a new formed 7AI–$(MeOH)_2$ solvate dramatically (Figure 2-7, bottom left). In the HBD distribution, the single peak corresponding to the 1:1 cyclic complex has disappeared and two new peaks have appeared. This new solvation structure conforms to the complex in which each methanol molecule is strongly hydrogen-bonded to only one of the two centers of H-bonding of the solute. In addition, the two methanol molecules are hydrogen-bonded to each other, forming a triply hydrogen-bonded wire connecting the pyrrole hydrogen (NH) and pyridine-type nitrogen of 7AI. Thus, a cyclic, doubly hydrogen-bonded complex in 7AI becomes unfavorable when more than one methanol molecule is present in solution. The fraction of the cyclic 1:1 species decreases below 1% (Table 2-6). In this context it is interesting that, as already mentioned, a recent supersonic jet study of 7AI–$(MeOH)_n$ clusters has demonstrated the presence of 1:1, 1:2 and 1:3 hydrogen-bonded complexes, among which the phototautomerization was only observed in the 1:2 species [62]. In the cases of PQ–$(MeOH)_2$ and PyIn-2-$(MeOH)_2$, a different MD picture is observed when a second methanol molecule becomes available. The addition of the second methanol molecule only leads to some broadening of the major peak in the HBD distribution. In the case of PQ–$(MeOH)_2$, about 26–39% of the hydrogen-bonded species still correspond to the 1:1 cyclic, doubly hydrogen-bonded complex. The appearance of a small population of non-cyclic hydrogen-bonded species is seen as two minor peaks in the HBD distribution. In PyIn-2-$(MeOH)_2$, the same pattern of three H-bonded complexes is found, and the cyclic complex is present for about 15–26% of the time. Solute-solvent solvates with higher solvent stoichiometry, up to 1:5, show that the addition of subsequent molecules of methanol to a 1:2 solvate does not cause significant changes in the corresponding HBD distribution. The most dramatic changes are observed when the first two solvent molecules are added.

Considering water as another H-bonding solvent containing multiple H-bonding centers, MD simulations show that the population of different hydrogen-bonded

Table 2-6 Molecular dynamics results: the maxima of main peaks in the HBD distributions and the corresponding *fwhm* values; The equilibrium fractions (in percent) of the 1:1 cyclic doubly hydrogen-bonded complexes (see text for details) (Reproduced from [83] with permission from *Journal of Physical Chemistry A*. Copyright © 2000 American Chemical Society)

	Peak maximum [Å][a]		fwhm [Å]		Equilibrium fraction cutoff [Å]	
	N···H–O	N–H···O	N···H–O	N–H···O	2.5	2.2
PQ + 1 MeOH	2.02	1.76	0.48	0.26	88	69
PQ + 2 MeOH	2.02	1.76	1.00	0.26	39	26
	2.02	3.58	0.44	0.54		
PQ in bulk MeOH	2.02	1.76	1.00	0.26	35	24
	2.15	4.49	0.56	1.22		
PQ + 1 H_2O	1.89	1.89	0.44	0.34	89	70
PQ + 2 H_2O	2.02	1.89	1.00	0.40	42	26
	2.15	3.71	0.52	0.54		
PQ in bulk H_2O	2.02	4.10	0.66	1.30	10	5
	3.19	2.02	1.10	0.40		
PyIn-2 + 1 MeOH	2.02	1.89	0.50	0.42	72	50
PyIn-2 + 2 MeOH	2.15	3.71	0.54	0.52	26	15
	2.54	1.89	1.02	0.30		
PyIn-2 in bulk MeOH	2.15 (a)[b]	1.76	0.58	0.36	15	9
	2.02 (b)	3.45	0.60	0.90		
	3.06 (c)	1.89	0.80	0.34		
PyIn-2 + 1 H_2O	1.89	2.02	0.58	0.36	74	47
PyIn-2 + 2 H_2O	1.89	2.02	0.44	0.40	31	19
	2.02	3.32	0.42	0.70		
PyIn-2 in bulk H_2O	2.15	1.89	0.52	0.32	5	3
	2.02	3.84	0.52	1.10		
	3.32	2.02	0.86	0.56		
7AI + 1 MeOH	2.28	1.89	1.28	0.36	37	13
7AI + 2 MeOH	1.89	3.45	0.34	0.50	1	0.1
	3.19	1.89	0.56	0.24		
7AI in bulk MeOH	1.89	5.14	0.30	0.86	0.9	0.1
	3.84	1.76	0.92	0.26		
7AI + 1 H_2O	2.15	2.15	0.60	0.68	35	11
7AI + 2 H_2O	1.89	3.45	0.26	0.46	0.1	0.01
	3.19	1.89	0.48	0.28		
7AI in bulk H_2O	1.89	4.49	0.30	1.00	0.01	–
	3.97	1.89	0.98	0.34		

[a]Accuracy: ±0.07 Å
[b]Compare with Fig 2-9

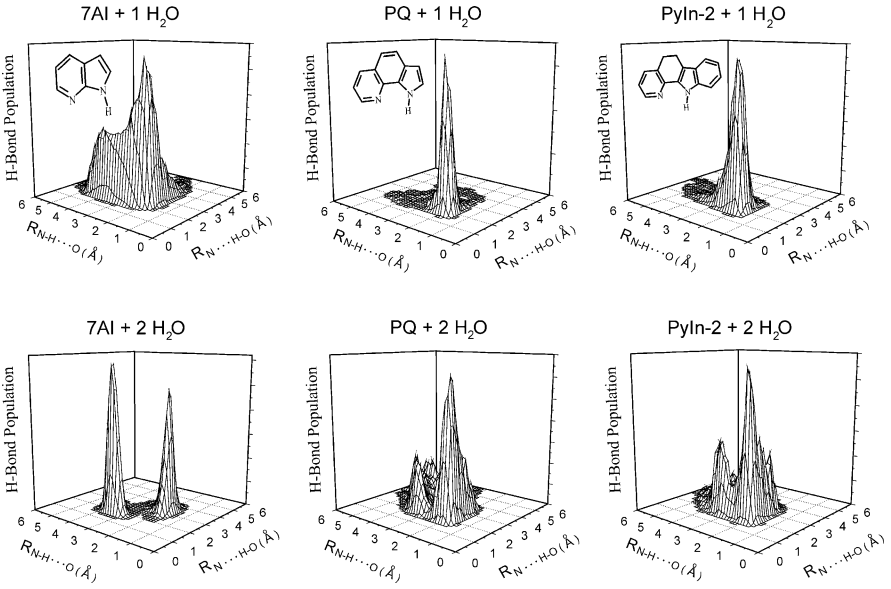


Figure 2-8. The HBD distributions for 7AI, PQ and PyIn-2 simulated for *n*-hexane in the presence of one and two molecules of water [83]

species and their structure are found to be in many aspects very similar to those obtained from the MD simulations in the presence of methanol. In the case of PQ and PyIn-2 with 1:1 and 1:2 solute:water stoichiometries, the cyclic, doubly hydrogen-bonded complex is present most of the time, as seen in Figure 2-8. Significant changes are observed for 7AI–$(H_2O)_1$, in which a considerable broadening of the HBD distribution is observed. The H-bonding intensity along the N–H···O coordinate is broadened and, at the same time, the corresponding distance along the N···H–O bond is shortened. In the case of 7AI–$(H_2O)_2$, similarly to 7AI–$(MeOH)_2$, most of the time the cyclic, triple hydrogen-bonded chain structure is found. Therefore, the MD simulations for the dilute solution with alcohol and water show that the topology of the H-bonding centers in PQ and PyIn-2 strongly favors the formation of the 1:1 cyclic hydrogen-bonded structure, even if the second solvent molecule is present in solution. 7AI also forms the cyclic, weakly hydrogen-bonded complex with protic solvent at 1:1 stoichiometry. However, in contrast to the behavior of PQ and PyIn-2, another, cyclic, 1:2 triply hydrogen-bonded structure is favored when more than one solvent molecule is available.

2.4.2. Hydrogen-Bonding Dynamics in Bulk Solvents

The MD simulations of the hydrogen bonding behavior of 7AI, PQ and PyIn-2 in bulk water and methanol show the increase of the probability of solute-solvent

hydrogen bonding in a long-distance region of HBD distribution, which is consistent with the formation of additional solvation shells around the solute. The corresponding intensity in the HBD graphs was cut off in order to improve the visualization of the hydrogen bonding distribution in the first solvation shell, as shown in Figure 2-9. The hydrogen bonding distribution for 7AI in bulk methanol shows the presence of two strong peaks. The comparison of peak positions and their breadth, seen as *fwhm* values in Table 2-6, with those observed in 7AI–$(MeOH)_1$ and 7AI–$(MeOH)_2$ demonstrates a significant peak broadening and shifts of the maximum of both peaks towards a long-range distance region compared to the case of the complexes in the dilute *n*-hexane solution (Figure 2-7). By tracing the hydrogen-bonding dynamics around 7AI in bulk methanol, it was shown that the two peaks correspond rather to a hydrogen-bonded complex in which 7AI is simultaneously hydrogen-bonded by its both, donor and acceptor atoms, with two molecules of methanol [83]. These methanol molecules are involved in independent hydrogen-bonding interactions with other molecules from the surrounding bulk methanol environment rather than with each other. It was also observed that during MD simulations these methanol molecules are being periodically exchanged by methanol molecules from the bulk. Therefore, the triple 1:2 hydrogen-bonded complex of 7AI with two methanol molecules, observed in the case of 7AI–$(MeOH)_2$, is strongly disfavored in bulk solution. What is most interesting, the equilibrium fraction of the 1:1 cyclic, doubly hydrogen-bonded complex was also found to decrease dramatically, so that its population is estimated to be smaller than 1%, as seen in Table 2-6.

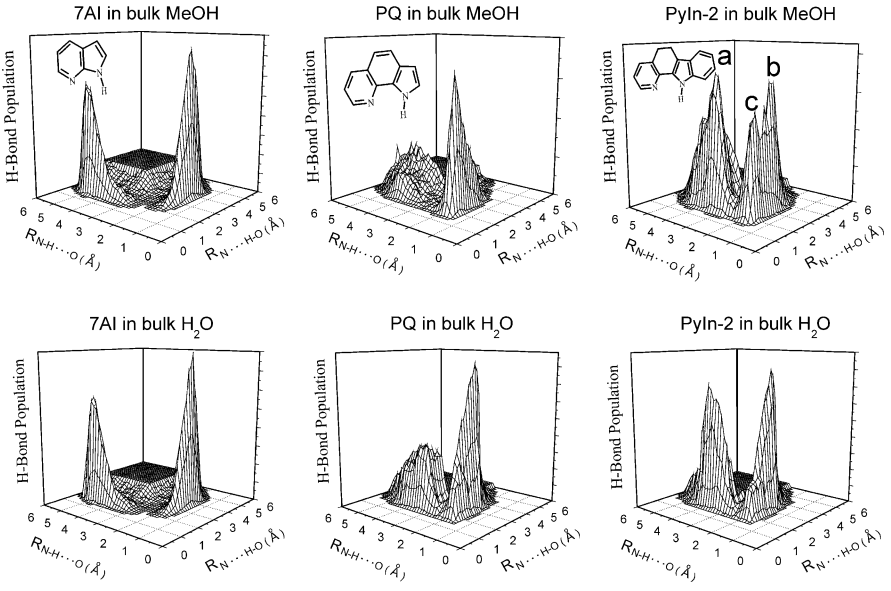


Figure 2-9. The HBD distributions for 7AI, PQ and PyIn-2 simulated for bulk methanol and for water solution [83]

The HBD distributions for PQ and PyIn-2 in bulk methanol and water are presented in Figure 2-9. The analysis of these distributions has shown that, despite of an extensive hydrogen-bonding nature of the bulk solvents, the significant equilibrium population of the 1:1 cyclic, double hydrogen-bonded species of PQ and PyIn-2 is still present (Table 2-6). In the case of PyIn-2, the three peaks (a, b and c in Figure 2-9, top-left) correspond to three differently hydrogen-bonded solvates, among which the peaks *a* and *b* can be assigned to the non-cyclic single hydrogen-bonded complexes, as schematically shown in Figure 2-6b for PQ, whereas the peak *c* is related to the 1:1 cyclic, double hydrogen-bonded species.

2.5. HYDROGEN-BONDING-INDUCED AND EXCITED-STATE PHENOMENA IN BIFUNCTIONAL DONOR–ACCEPTOR MOLECULES

Fluorescence spectra of hydrogen-bond bifunctional compounds such as PQ, DPC and PC are characterized by a single emission band in nonpolar (*n*-hexane) and polar aprotic (acetonitrile) solvents. Fluorescence quantum yields of these compounds have been reported to be quite high, 0.15–0.25 [8, 10, 55]. In alcohol and water solutions, the fluorescence quantum yield becomes very weak. The most interesting feature is the observation of dual fluorescence in the protic solvents [8–12, 55]. In addition to the first "normal" fluorescence band, referred to as F_1, a second, long-wavelength band, F_2, appears. The spectral position of F_1 is similar to that of the emission typically observed in polar aprotic solvents. The long-wavelength band F_2, observed in protic media only, has been attributed to the fluorescent product of double proton tautomerization occurring in a photoexcited cyclically hydrogen-bonded complex of bifunctional azaaromatic compounds with protic solvent molecules [8–10]. The electron density redistribution occurring from the pyrrole to quinoline moieties upon excitation has been proposed to be the driving force for the excited-state tautomerization. PQ reveals an increase of the excited-state pK_a of the nitrogen atom in the acceptor quinoline moiety and a decrease of the corresponding pK_a of the donor NH group: $\Delta pK_a(N) = 9.6$ and $\Delta pK_a(NH) = -6.0$, respectively [55]. Such an acid–base excited-state behavior has been corroborated by the TDDFT calculations, which have shown a significant electron density redistribution from the donor to the acceptor part of PQ (Section 2.2). In addition to steady-state fluorescence studies, subpicosecond time-resolved fluorescence transient measurements carried out for PQ and PC have shown that the F_1 band contains two fast decay components with time constants of $\tau_1 = 0.6$–0.9 ps and $\tau_2 =$ 6–11 ps, and a slower decay component, τ_3, characterized by a time constant between 50 and 150 ps, depending on the properties of the alcohol [11]. The F_2 band has also revealed a fast biexponential rise occurring at the same rate as the fast initial decay of the emission band F_1. It is interesting that no rise time corresponding to the slow decay of F_1 has been detected. Therefore, the time-resolved fluorescence experiments have provided convincing evidence that the F_1

and F_2 bands are emitted by different fluorescent species. The fast time decay components observed for F_1 and in the rise of F_2 have been attributed to the presence of cyclic hydrogen-bonded complexes of PQ with solvent molecules that are able to undergo solvent-assisted proton transfer. The longer decay component τ_3 of F_1 has been assigned to solute-solvent hydrogen-bonded complexes which are not involved the fast phototautomerization. An important finding is that the decay components τ_1 and τ_2 are only weakly dependent on temperature and viscosity [11]. This is in contrast to a strong temperature dependence of the slow component τ_3. It has been concluded that cyclic, hydrogen-bonded complexes of the heteroazaaromatic species, with the structure appropriately prepared for a proton transfer reaction, are formed already in the ground state. Therefore, they do not require significant solvent relaxation around an excited bifunctional chromophore. This conclusion is consistent with the observation that the additional component τ_3 starts to appear in F_1 only when such a solvent relaxation process is activated in liquid solution at higher temperatures. It has also been proposed that since the slow component τ_3 has no corresponding counterpart in the rise of the fluorescence F_2, the decay characterized by τ_3 should correspond to internal conversion from the photoexcited chromophore to its ground state [55]. The photophysics of different hydrogen-bonded species of bifunctional compounds with protic solvents is schematically summarized in Figure 2-10.

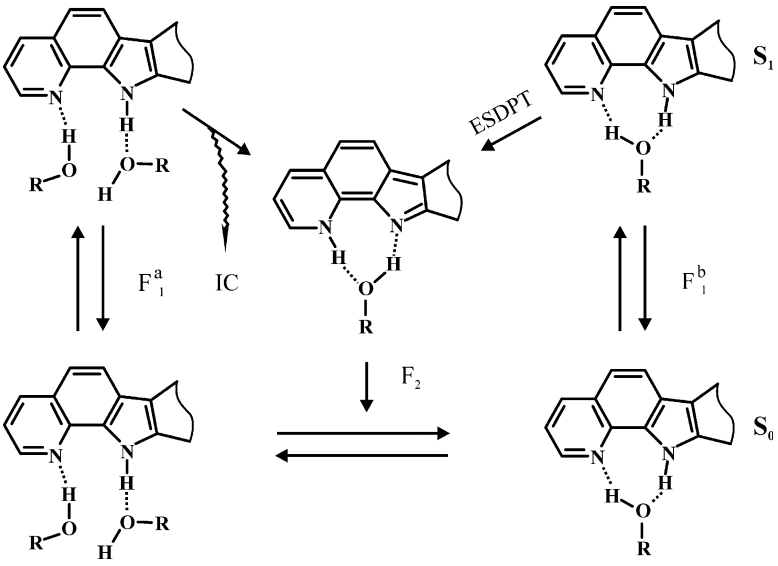


Figure 2-10. General scheme of the ground-state equilibrium between hydrogen-bonded complexes of bifunctional heteroazaaromatic molecules with protic solvents and photophysical processes occurring in hydrogen-bonded solvates [55]

The photophysics of bifunctional compounds composed of indole and pyridine moieties has recently been studied by Sobolewski and Domcke using ab initio electronic structure calculations [91–94]. It has been shown that, upon the electronic excitation, electron- and proton transfer processes are strongly correlated in these systems. It has been proposed that the photochemistry of such bifunctional compounds can be described in terms of the electron-driven-proton-transfer mechanism [92]. A common feature of such mechanism is the role of highly polar charge-transfer states, which induce the proton transfer reaction, leading in many cases to a conical intersection of the S_1 and S_0 state surfaces observed experimentally as ultrafast internal conversion.

2.6. INTERACTION OF HETEROAZAAROMATICS WITH LIPID MEMBRANES

Tryptophan, one of natural amino acids, is known to bind preferentially to the interfacial region of lipid membranes [95–97]. Therefore, membrane-bound tryptophan and other indole derivatives are found to be in easy contact with interfacial water. 7-azatryptophan and its analogs, closely related to 7-azaindole, have originally been designed as fluorescent probes with improved emission characteristics for studying solvation dynamics in biological systems [98]. The diversity of physical and chemical properties of these probes and, also, those that mimic natural amino acids and base-pair analogues, such as PQ derivatives, suggest that, due to their different hydrophobicities, binding modes of these probes with lipid membranes may differ significantly. To address this question, we used classical all-atom MD simulations to study interactions and favorable locations of several heteroazaaromatic compounds, such as 7AI, PQ and PC inside a lipid membrane [99].

A membrane is modeled using a phospholipid bilayer composed of 128 dipalmitoylphosphatidylcholine (DPPC) lipids surrounded by 4310 water molecules. In such a system, a lipid-to-water ratio is about 1:33, so that the lipid bilayer is considered to be in the fully hydrated state [100]. In order to model a bilayer existing in a biologically relevant liquid crystalline phase the simulation is carried out at the temperature of 323 K. Figure 2-11 shows an example of the simulated system containing a hydrated DPPC bilayer with several molecules of PQ partitioned inside the hydrophobic core of the membrane. Although such a snapshot represents a single, instantaneous configuration of the system, averaging over many configurations and for a long period of MD runs allows one to obtain average density profiles for individual components of a bilayer [101], as well as a distribution profile of probe molecules diffusing freely across an MD box during simulations (Figure 2-11, right).

Interactions and binding modes of a series of heteroazaaromatic probes with a lipid membrane are modeled using classical MD simulations at the constant number of particles, constant pressure of 1 atm, and constant temperature, $T = 323$ K (NPT

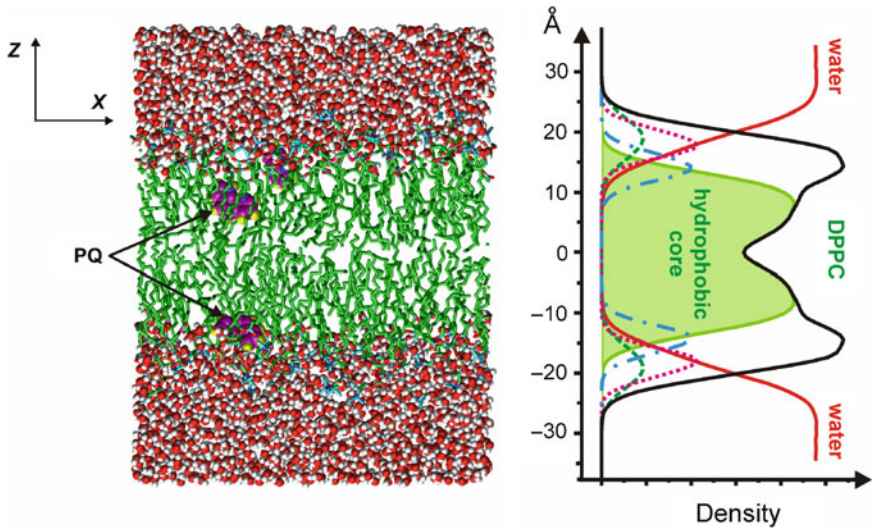


Figure 2-11. (*Left panel*) Snapshot of a hydrated DPPC bilayer after 50 ns of MD simulations. PQ molecules are drawn in vdW representation. (*Right panel*) Mass density profiles of different molecular components of a DPPC bilayer. The mass densities were obtained by averaging MD trajectories for the last 40 ns. *Solid black* and *red* lines show the densities of DPPC and water across the MD box. Other individual components of the bilayer are plotted as follows: acyl chains, showing a hydrophobic core of the bilayer, (*solid light green line*), choline (*dashed green line*), phosphate (*dotted magenta line*) and glycerol (*dash-dotted blue line*) moieties, respectively

ensemble). Statistical averaging is applied to accumulate a distribution profile of a probe between water solution, at biological pH (pH about 7), and a bilayer. The simulations are designed to reproduce the passive, thermally-driven diffusion of the probes in the simulation box. At the beginning of the simulation, eight molecules of each probe (7AI, PQ and PC) are added, in a random fashion, into the aqueous solution at the vicinity of the bilayer interface (four probes per each, top and bottom, leaflet of the bilayer). The initial MD configuration of a PQ/DPPC system containing eight PQ molecules at $t = 0$ ns is shown in Figure 2-12. The MD simulations demonstrate that, during the first 10 ns, the probe molecules tend to diffuse from the aqueous phase into the polar interfacial region of the DPPC bilayer. After 20 ns, all the probe molecules are found to be located preferentially at the bilayer interface, as shown in Figure 2-12 for the PQ/DPPC system at $t = 20$ ns. Further time evolution of membrane interactions of 7AI, PQ and PC with the bilayer shows a similar behavior for all the probes, so that when they reach the headgroup region of the membrane, their diffusion slows down significantly. At the end of the 50 ns simulation period, 7AI was found to be distributed broadly across the bilayer interface, whereas PQ and PC were located deeper inside the hydrocarbon core of the DPPC bilayer. A typical distribution of the PQ molecules at the end of the 50 ns MD run is shown in Figure 2-12.

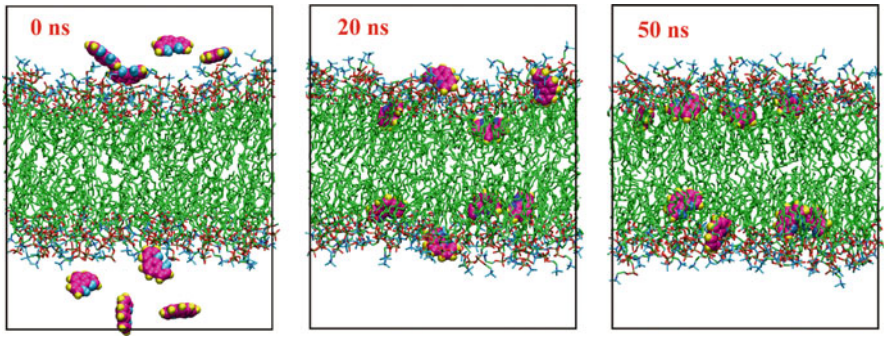


Figure 2-12. Snapshots of the MD boxes taken from the simulations of eight PQ molecules in the DPPC bilayer after $t = 0$, 20 and 50 ns of the NPT simulation. Water molecules are not shown for clarity. The PQ molecules are rendered with enlarged atomic van der Waals radii [99]

The quantitative evaluation of the distribution of 7AI, PQ and PC within the DPPC bilayer is represented by mass density profiles of different molecular components across the bilayer. Figure 2-13 shows the density distributions for choline, phosphate, as well as the glycerol backbone and acyl chains of the phospholipid molecules across the bilayer. The density of 7AI, PQ and PC molecules within the DPPC bilayer is also shown in Figure 2-13a–c. As can be seen, the three probes demonstrate a different localization behavior and different penetration depths in the DPPC bilayer. The density profiles also reveal a different number of the binding sites for these probes. 7AI is found to be distributed throughout the interfacial region to the hydrocarbon acyl chain core of the bilayer. The preferable localization of PC is much deeper inside the hydrocarbon region of the bilayer, below the glycerol moiety of lipids. In the case of PQ, two binding sites, located at two different depths inside the bilayer, can be identified; a major site is localized in the hydrocarbon core and a second, minor population is observed at the membrane interface (Figure 2-13b).

The results of decomposition analysis of the density profiles of the binding sites of the three probes are summarized in Table 2-7. Comparison of the site localization shows that the molecular shape and aromaticity of the probes play an important role in their localization in the lipid membrane. The favorable probe location is found to be determined by the hydrophobic effect. The three probes differ in the number of their aromatic rings, increasing in the following sequence: 7AI–PQ–PC. Therefore, the hydrophobic character of the probes also increases in the same order. The tendency to drive the aromatic hydrocarbon probe out of polar water results in their deep membrane localization inside the bilayer. As can be seen in Table 2-7, the depth of the probe localization in this sequence is gradually shifted from interfacially localized (sites B and C) to deeply inserted into the hydrocarbon acyl chain core of the bilayer (site A) (Table 2-7).

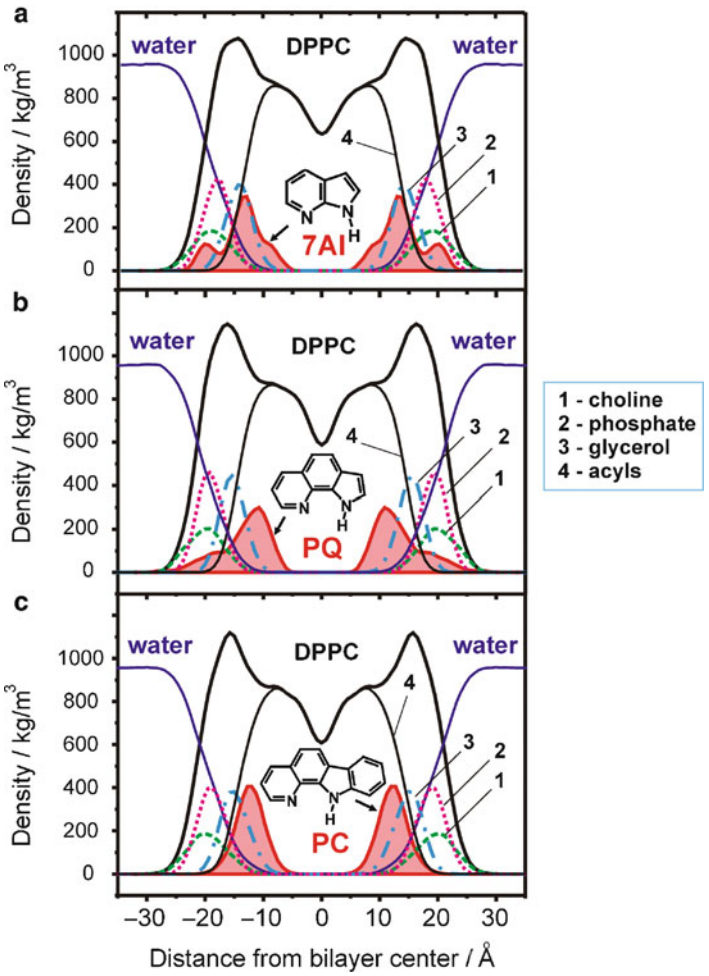


Figure 2-13. Mass density distribution profiles for individual components of the DPPC bilayer and for the total density distribution of eight molecules of 7AI (**a**), PQ (**b**) and PC (**c**). The DPPC components are: 1 – choline group; 2 – phosphate group; 3- glycerol moiety; 4 – acyl chains. All the density profiles are plotted with respect to the center of the z-axis of the MD box [99]

2.6.1. Hydrogen-Bonding at the Membrane Interface

The three probes, 7AI, PQ, and PC, bind to the DPPC membrane in a different fashion. They also penetrate into the hydrocarbon region of the bilayer at different depths. As a result, hydrogen-bonding of these probes with water molecules and with polar headgroups of the bilayer may also differ significantly. Figure 2-14a shows the geometrical criteria for the existence of a hydrogen bond between a probe

Table 2-7 Summary of Gaussian decomposition analysis for probe-membrane binding sites [99] (Reproduced from [99] with permission from *Biophysical Chem*istry. Copyright © 2008 Elsevier)

Binding site	7AI			PQ			PC		
	P[a] (%)	Max[b] (Å)	Fwhm[c] (Å)	P (%)	Max (Å)	Fwhm (Å)	P (%)	Max (Å)	Fwhm (Å)
A	15	8.6	3.0	62	11.0	4.6	100	12.2	4.9
B	65	13.3	3.8	38	17.2	8.3	–	–	–
C	20	19.8	3.3	–	–	–	–	–	–

[a]Population (P) of a binding site
[b]Maximum of a Gaussian band
[c]Full width at half maximum

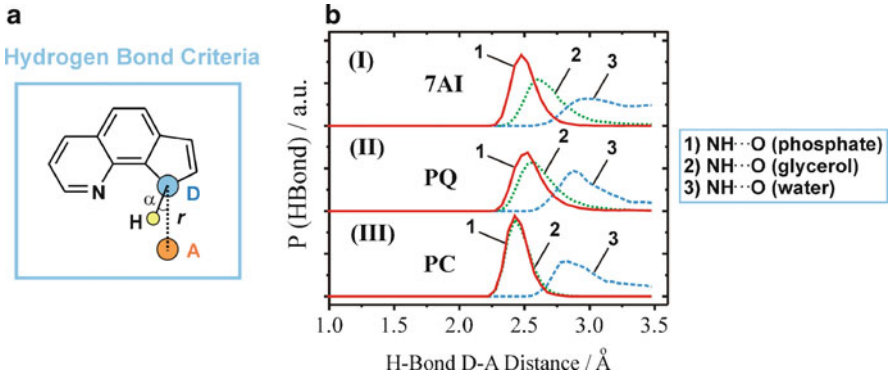


Figure 2-14. (**a**) Geometrical criteria ($r_{HB} \leq 3.5$ Å and $\alpha_{HB} \leq 30°$) for hydrogen-bonding between a donor (D) and an acceptor (A) in PQ. (**b**) Distribution of hydrogen bond donor–acceptor distances for 7AI (I), PQ (II), and PC (III). The distance distribution between the donor (the imino nitrogen atom (NH) of the probe molecule) and the acceptor (oxygen atoms of a lipid phosphate group) is shown by a *solid line* (1); between the donor and oxygen atoms of a lipid glycerol moiety by a *dotted line* (2); between the donor and water oxygen atoms by a *dashed line* (3) [99]

molecule acting as a donor and the oxygen atoms of water or lipids as acceptors. A hydrogen bond between a donor and an acceptor is assumed to exist if the D–A distance (r_{HB}) and the angle (α_{HB}) fulfill the following criteria: $r_{HB} \leq 3.5$ Å and $\alpha_{HB} \leq 30°$. The imino nitrogen atom of a probe was chosen as an H-bond donor and oxygen atoms of water, phosphate and glycerol groups served as H-bond acceptors.

To determine the nature of favorable hydrogen bonds formed between the probe and interfacial waters or the DPPC lipids, the relative frequency of the occurrence of each type of these H-bonds was estimated. The results of the hydrogen-bonding analysis are summarized in Table 2-8. As can be seen, the majority (∼95%) of 7AI population is found to be hydrogen-bonded with either water or lipid molecules of the bilayer, whereas a significant fraction (about 24% and 33%) of PQ and PC, respectively, exists in a hydrogen-bond-free form. 7AI was found to be localized on the water/bilayer interface and, therefore, a large population of probe molecules

Table 2-8 Results of the analysis of hydrogen-bonding between the probe molecules and lipids and water (Reproduced from [99] with permission from *Biophysical Chemistry*. Copyright © 2008 Elsevier)

Hydrogen bond	7AI				PQ				PC			
	F[a] (%)	r_{HB}[b] (Å)	fwhm[c] (Å)	α_{HB}[d] (deg)	F[a] (%)	r_{HB}[b] (Å)	fwhm[c] (Å)	α_{HB}[d] (deg)	F[a] (%)	r_{HB}[b] (Å)	fwhm[c] (Å)	α_{HB}[d] (deg)
NH–water oxygen	44.4	2.98	0.58	11.5	26.8	2.88	0.37	12.0	22.8	2.83	0.42	14.0
NH–phosphate oxygen	39.5	2.49	0.19	5.5	21.8	2.52	0.22	8.5	14.3	2.42	0.17	6.5
NH–glycerol oxygen	10.9	2.60	0.27	6.5	27.3	2.57	0.26	9.0	29.4	2.44	0.16	8.5
no H-Bonds	5.2	–	–	–	24.1	–	–	–	33.5	–	–	–

[a]Relative frequency (F) of the occurrence of hydrogen-bonding
[b]Distribution peak of the H-bond donor–acceptor distance, see Figure 2-14 for details
[c]Full width at half maximum of the H-bond donor–acceptor distance distribution
[d]Average deviation of the H-bond from linearity

exist in permanent contact with interfacial water, as also seen from the corresponding density profiles in Figure 2-13a. Therefore, the relative frequency of the occurrence of hydrogen-bonds between 7AI and water was estimated to be about 44% (Table 2-8). For comparison, in the case of PQ and PC, hydrogen-bonding with water is gradually decreased to ~27% and ~23%, respectively, in agreement with the deeper located major binding sites of these probes inside the hydrophobic, water-free region of the bilayer. The distribution of the hydrogen bonds distances is shown in Figure 2-14b and Table 2-8. The H-bond distance and angle distributions demonstrate that all the three probes, 7AI, PQ and PC, form relatively strong hydrogen bonds with the oxygen atoms of the phosphate group of the lipids. However, this effect may not be decisive in favoring the probe localization in the bilayer, because significant populations of PQ and PC are still found to exist as hydrogen-bond-free, indicating that the favorable binding of these probes into the hydrocarbon core of the bilayer is not completely driven by hydrogen-bonding.

There is a significant overlap between the density distribution profiles of the probe and water molecules, penetrating deeply into the headgroup region of the bilayer, as seen in Figure 2-13a–c. The above-presented hydrogen-bonding analysis described the statistically averaged hydrogen-bonding interactions accumulated over a long MD simulation period of 40 ns. We also evaluated a partial radial distribution function *g(R)* as a complementary tool for the H-bonding analysis. The *g(R)* function was calculated between the imino NH hydrogen atom of the probe, assumed to be an H-bond donor, and the oxygen atoms of the lipid molecules (Figure 2-15). The analysis shows that, for all the three probes, the first peak of *g(R)* is observed between the NH hydrogen atom of the probe and the oxygen atoms of either the phosphate or glycerol groups of the DPPC lipids. The *g(R)* peak for 7AI, PQ and PC are observed at distances of 1.5–1.6 Å, respectively. This distance range corresponds to values of hydrogen-bond lengths evaluated by MD simulations for H-bonded water complexes of these compounds in bulk solution [83, 99].

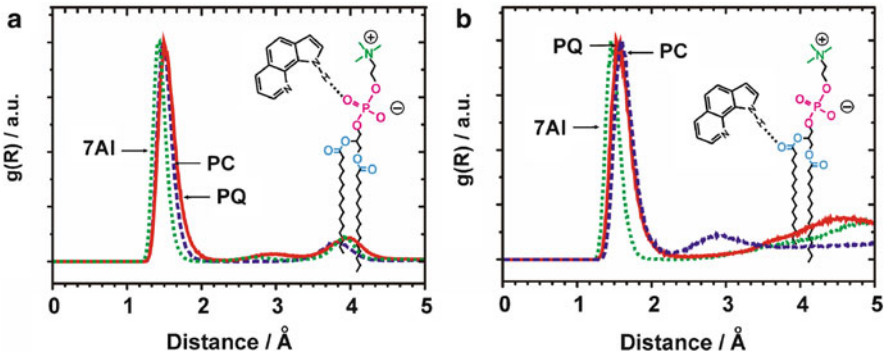


Figure 2-15. Radial pair distribution functions *g(R)* of the N-H hydrogen atom of PQ (*solid line*), PC (*dashed line*) and 7AI (*dotted line*) around the oxygen atoms of the phosphate (**a**) and glycerol (**b**) groups of the DPPC lipids. The corresponding hydrogen-bonding between the probe and lipid molecules is schematically shown for PQ [99]

2.7. PROBING THE ACID–BASE EQUILIBRIUM AT THE MEMBRANE INTERFACE

Excited-state proton transfer (ESPT) has been reported for 2-(2′-pyridyl)benzimidazole (2PBI) in aqueous solutions at acidic pH [102]. It has been observed that the reaction occurs predominantly in the cationic form, $2PBI^+$, in which the second benzimidazole nitrogen atom is protonated, as shown schematically in Figure 2-16a. It has also been proposed that the ESPT in 2PBI is assisted by hydrogen-bonding with water molecules. Therefore, according to refs [103, 104], one of the requirements for ESPT in 2PBI to occur is the formation of a cyclic, hydrogen-bonded complex with protic partner molecules (Figure 2-16b).

The solvent-mediated ESPT in such systems is often characterized by high sensitivity to the properties of a surrounding medium. It has been recently shown that ESPT in hydrogen-bonded complexes of 2PBI is selectively enhanced at a micelle-water interface [103]. One of possible reasons for this phenomenon is that the cationic form $2PBI^+$ could be stabilized by favorable electrostatic interactions with the charged surface of micellar systems [104]. This effect could have important implications in fluorescence sensing of structure and dynamics of the interface of cell membranes. Therefore, the knowledge of interactions and binding modes of 2PBI derivatives with lipid membranes could allow understanding of possible mechanisms of the modulation of an ESPT reaction at a heterogeneous membrane environment.

In order to understand interactions and preferable binding modes of 2-(2′-pyridyl) benzimidazole with a model membrane we applied a molecular dynamics approach

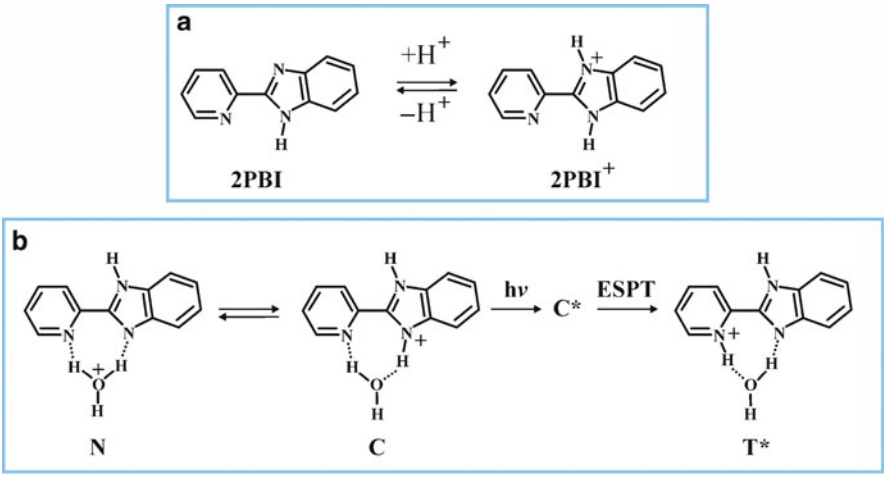


Figure 2-16. (**a**) Scheme of acid–base equilibrium in 2-(2´-pyridyl)benzimidazole. (**b**) Mechanism of the photoinduced proton transfer in a cyclic, hydrogen-bonded complex of a cationic form of 2PBI with water

described in Section 2.6. To model an acid–base equilibrium in solution and at a membrane interface, both the neutral (2PBI) and cationic ($2PBI^+$) forms were simultaneously present in MD simulations. Two neutral and two cationic species were initially placed in the water region near the top and the bottom leaflets of a DPPC bilayer, so that a total number of eight solute molecules were sampled in MD simulations. The initial configuration of the simulated system at $t = 0$ ns is shown in Figure 2-17. Simultaneous MD sampling of the membrane interaction of the two different, neutral and cationic, forms of 2PBI under the same simulation conditions allows us to understand in more detail preferable distribution of each of these forms between the aqueous solution and the lipid bilayer. Different polarities of 2PBI and $2PBI^+$ imply that these two species will partition into different, apolar and polar, regions of a bilayer [105], where concentrations of water molecules are different [106, 107]. Therefore, the formation of hydrogen-bonded complexes of 2PBI and $2PBI^+$ with water may be modulated by localization of these species in different regions of a bilayer.

In the applied MD approach, both 2BPI and $2PBI^+$ species were allowed to diffuse freely between the aqueous solution and the lipid bilayer, so that sampling over a long MD simulation period can provide the overall probability to find each of

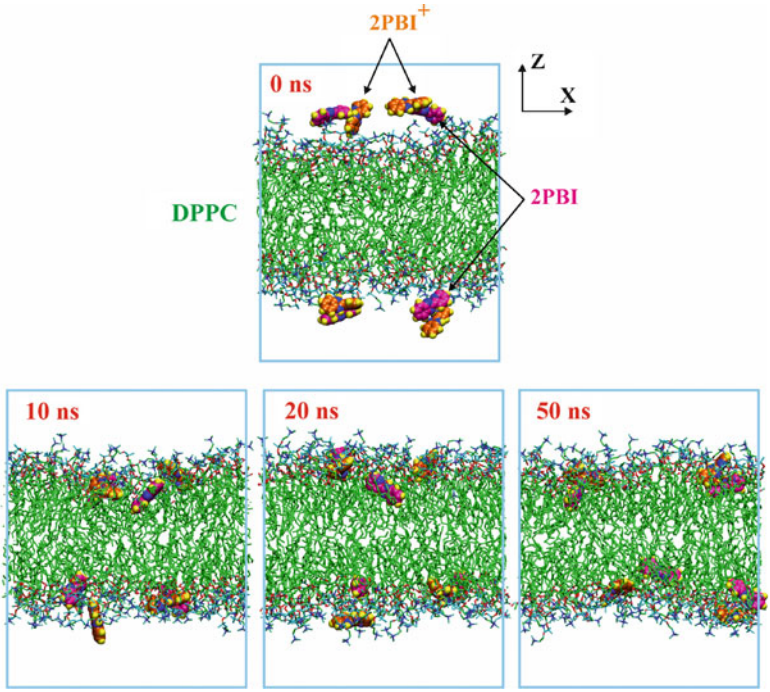


Figure 2-17. MD simulations of the interactions of neutral (2PBI) and cationic ($2PBI^+$) forms of 2-(2´-pyridyl)benzimidazole with a DPPC bilayer. Snapshots are shown for an initial MD configuration ($t = 0$ ns) and after 10, 20 and 50 ns. Water molecules are not shown

these species across bulk water/interface/membrane core regions. The time evolution of the distribution of 2BPI and $2BPI^+$ between bulk solution and the DPPC bilayer is illustrated in Figure 2-17. As can be seen from the first 10 ns of the MD sampling, both the neutral and cationic forms are moving towards the bilayer interface. At longer simulation times of 20–50 ns, further redistribution occurs between the species located at the membrane interface and the solute molecules deeply-inserted into a hydrocarbon core of DPPC.

The localization probability of the probe molecules can be described by their mass density profile acquired across the bilayer [99,101]. The mass densities of neutral 2BPI and cationic $2BPI^+$ were therefore evaluated along the bilayer axis z, as defined in Figure 2-17. Figure 2-18 shows the time evolution of the mass densities which are averaged starting from $t = 0$ ns and ending at time intervals of 2, 10, 20 and 50 ns, respectively. As can be noticed, at these time intervals the major peak of the mass density of the polar cationic form $2BPI^+$ is always found to be located in more polar interfacial regions of the bilayer compared to that of the neutral form 2BPI. At $t = 2$ ns $2BPI^+$ is preferably located at 22.1 Å from the bilayer center (Figure 2-18). A second binding site for $2BPI^+$ is also observed at 17.4 Å. Both these populations of $2BPI^+$ can be assigned to interfacial binding sites. In the same time interval of 2 ns, 2BPI is also characterized by a dual peak distribution with the maxima located at 18.7 and 22.7 Å, respectively. The major population of 2BPI is, however, found to be shifted towards the hydrophobic region of the bilayer. The time evolution of the densities shows that, at longer time intervals of 10–20 ns, the major population of $2BPI^+$ becomes preferably concentrated at the binding site located at 16.5–17.4 Å, whereas the population originally observed at 22.1 Å gradually decreases and its peak position is shifted to 20.7 Å (Figure 2-18). A similar population redistribution occurs between two originally observed binding sites of 2BPI. The population of 2BPI, which is observed at 22.7 Å at $t = 2$ ns, has a very small contribution to the overall population evaluated after 20 ns of the MD sampling. At the same time, the peak of the major population of 2BPI is localized at 15.0–15.4 Å, respectively. At $t = 50$ ns, $2BPI^+$ can be characterized by the mass density distribution with one, well-defined, population located at 17.0 Å. At the same time, 2BPI is preferably located at 13.7 Å, so that it is more deeply buried into the bilayer compared to the polar cationic form $2BPI^+$.

The density distributions of 2BPI and $2BPI^+$, as well as those of the DPPC bilayer and water, averaged over all 50 ns of the MD sampling period are presented in Figure 2-19. As can be seen, the preferable localizations of 2BPI and $2BPI^+$ are driven by different factors. The major population of neutral and, therefore, more hydrophobic 2BPI is located inside the hydrophobic core of the bilayer formed by its acyl chains. In contrast, $2BPI^+$ is found to be strongly localized at the interface. The depth of its distribution into the bilayer correlates and coincides well with the depth of water penetration into the bilayer, as is clearly seen in Figure 2-19. Therefore, polar cationic $2BPI^+$ prefers to be located in the region of the bilayer with high concentration of water molecules.

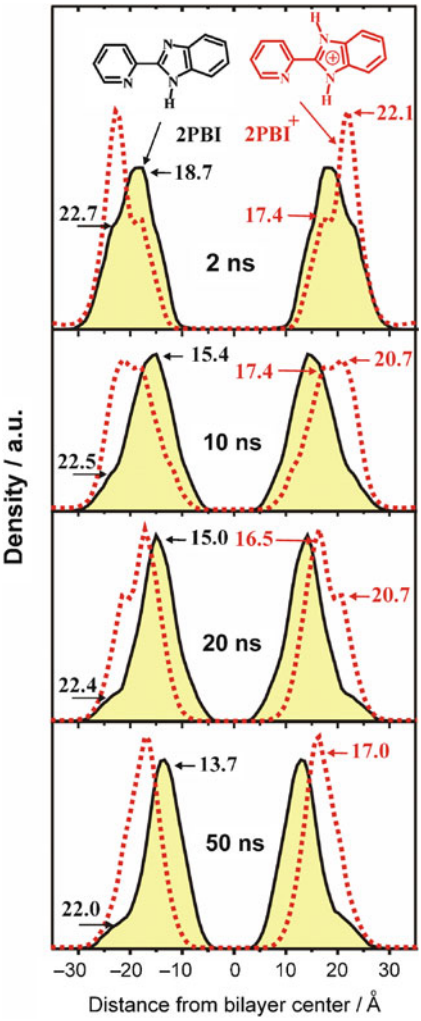


Figure 2-18. Time evolution of the mass density of neutral 2PBI and the cationic form $2PBI^+$ calculated across an MD simulation box along the z axis of the DPPC bilayer (see Figure 2-17 for details). Peak positions of major binding sites are shown in angstroms

Owing to distinct binding localizations and the different hydrogen-bonding-dependent fluorescence behavior of neutral and cationic forms 2BPI and $2BPI^+$ in the lipid bilayer, it is possible to apply this compound for studying the properties of biological water of the bilayer interface. Recently, it has been shown, by measuring time-resolved fluorescence spectra and kinetics, that the excited-state proton transfer dynamics of 7AI in water nanopools of reverse micelles is slowed down compared to bulk water solution [108]. It has been concluded that, because 7AI is preferably located in the bound-water regions of reverse micelles, higher free

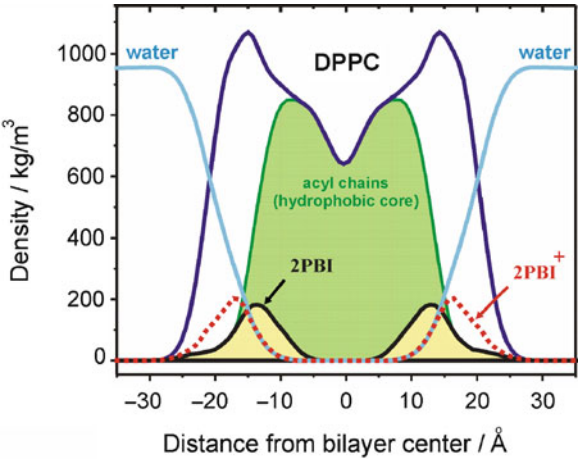


Figure 2-19. Average density distributions of the DPPC bilayer, water and probe molecules (2PBI and $2PBI^+$) along the *z* axis of the simulation box (see Figure 2-17 for details). The density averaging was performed over 50 ns of the MD simulation

energy is required for solvation of reactive, cyclically hydrogen-bonded species of 7AI in confined water environment. The cationic form $2BPI^+$ is able to form hydrogen-bonded solvates with interfacial water molecules in which its excited-state behavior is altered by proton transfer reaction. Therefore, the fluorescence of $2BPI^+$ bound to a lipid membrane may be used as a tool for better understanding of water dynamics in structured biological environments.

2.8. CONCLUSIONS

We have presented an overview of the results of computational and experimental works concerning structures, hydrogen-bonding dynamics and excited-state properties of a series of bifunctional hydrogen-bond donor–acceptor heteroazaaromatic compounds of biological importance. The common thread in the photophysics of these compounds is the important role of cyclic hydrogen-bonded solvates formed with protic partner molecules. The cyclic hydrogen-bonding with external partners, connecting proton donor and acceptor parts of a heteroazaaromatic molecule, enables multiple proton transfer phototautomerism, which is not possible in the case of an isolated, hydrogen-bond-free chromophore. Using ab initio and DFT computational approaches we have provided physical insights and qualitative interpretation of the role of hydrogen-bonding and cluster size effects for the photophysics of jet-isolated PQ clusters [84, 85]. The ultrashort excited state lifetime observed for the cyclic, double hydrogen-bonded complex PQ–$(MeOH)_1$ has been found to be due to the rapid double proton transfer reaction followed by nonradiative relaxation. In the larger size, jet-cooled, clusters of PQ with methanol, the cyclic, however, less

favorable hydrogen bonding blocks the fast proton transfer tautomerization. The structure of hydrogen-bonded solvates and preferable hydrogen-bonding modes of bifunctional compounds in bulk solvents have been further explored by classical molecular dynamics. We have found, using several lines of evidence based on both experimental results and quantum chemical/molecular dynamics simulations, that the structural topology of proton donor and acceptor parts of bifunctional solute molecules composed of a pyrroloquinoline-like moiety favors the formation of the cyclic, 1:1 double hydrogen-bonded solvates even in bulk hydroxylic solvents. Such hydrogen-bonding behavior of PQ and related compounds is quite different from that of another family of bifunctional donor–acceptor molecules based on 7AI and 1-azacarbazole. The latter compounds appear to form hydrogen-bonded complexes with the 1:2 solute-solvent stoichiometry [62, 78]. These results show that a mere shifting of the donor and acceptor groups by just one bond away from each other may lead to dramatic structural effects which, in turn, control the photophysics and photochemistry of hydrogen-bonded complexes.

The possible effect of microsolvation on the excited-state proton transfer and fluorescence quenching of bifunctional compounds in a heterogeneous environment of a lipid bilayer has also been discussed. The MD simulations have shown that the depth of the localization of the heteroazaaromatics probes inside a lipid membrane may be tuned by small structural modification of a parent probe molecule [99]. It appears that the bifunctional chromophores can be used as probes of local polarity, viscosity, and hydrophilicity, as well as indicators of specific hydrogen-bonding topologies [13].

Finally, it should be noted that the MD methodologies similar to those presented above have been used to explain other types of environment-specific phenomena: (i) solvent-induced conformational changes [1, 109]; (ii) rare gas matrix deposition, origin of site-structure and its energetics [110–113].

ACKNOWLEDGEMENTS

JW acknowledges support by the grant N N204 3329 33 from the Polish Ministry of Science and Higher Education. The experiments referred to in this work were based on a long-time collaboration with the groups of Prof. R.P. Thummel (University of Houston) and Prof. B. Brutschy (University of Frankfurt).

REFERENCES

1. Waluk J (2003) Acc Chem Res 36:832–838
2. Chou PT (2001) J Chin Chem Soc 48:651–682
3. Nosenko E, Wiosna-Sałyga G, Kunitski M, Petkova I, Singh A, Buma WJ, Thummel RP, Brutschy B, Waluk J (2008) Angew Chem Int Ed Engl 47:6037–6040
4. Kijak M, Nosenko E, Singh A, Thummel RP, Waluk J (2007) J Am Chem Soc 129:2738–2739
5. Wiosna-Sałyga G, Dobkowski J, Mudadu MS, Sazanovich I, Thummel RP, Waluk J (2006) Chem Phys Lett 423:288–292

6. Kijak M, Zielińska A, Chamchoumis C, Herbich J, Thummel RP, Waluk J (2004) Chem Phys Lett 400:279–285
7. Wiosna G, Petkova I, Mudadu MS, Thummel RP, Waluk J (2004) Chem Phys Lett 400:379–383
8. Herbich J, Dobkowski J, Thummel RP, Henge V, Waluk J (1997) J Phys Chem A 101:5839–5845
9. Kyrychenko A, Herbich J, Izydorzak M, Gil M, Dobkowski J, Wu FY, Thummel RP, Waluk J (1999) Isr J Chem 39:309–318
10. del Valle JC, Dominguez E, Kasha M (1999) J Phys Chem A 103:2467–2475
11. Marks D, Zhang H, Borowicz P, Waluk J, Glasbeek M (2000) J Phys Chem A 104:7167–7175
12. Kijak M, Zielińska A, Thummel RP, Herbich J, Waluk J (2002) Chem Phys Lett 366:329–335
13. Herbich J, Kijak M, Luboradzki R, Gil M, Zielińska A, Hu YZ, Thummel RP, Waluk J (2002) J Photochem Photobiol: A Chem 154:61–68
14. Taylor CL, El-Bayoumi MA, Kasha M (1969) Proc Natl Acad Sci USA 63:253–260
15. Douhal A, Kim SK, Zewail AH (1995) Nature 378:260–263
16. Folmer DE, Poth L, Wisniewski ES, Castleman AW Jr (1998) Chem Phys Lett 278:1–7
17. Takeuchi S, Tahara T (1998) J Phys Chem A 102:7740–7753
18. Takeuchi S, Tahara T (2001) Chem Phys Lett 347:108–114
19. Fiebig T, Chachisvilis M, Manger M, Zewail AH, Douhal A, Garcia-Ochoa I, de La Hoz AA (1999) J Phys Chem A 103:7419–7431
20. Folmer DE, Wisniewski ES, Castleman AW Jr (2000) Chem Phys Lett 318:637–643
21. Catalán J, del Valle JC, Kasha M (1999) Proc Natl Acad Sci USA 96:8338–8343
22. Catalán J, Pérez P, del Valle JC, de Paz JLG, Kasha M (2002) Proc Natl Acad Sci USA 99:5793–5798
23. Takeuchi S, Tahara T (2007) Proc Natl Acad Sci USA 104:5285–5290
24. Kwon OH, Zewail AH (2007) Proc Natl Acad Sci USA 104:8703–8708
25. Catalán J (2008) Proc Natl Acad Sci USA 105:E78
26. Kwon OH, Zewail AH (2008) Proc Natl Acad Sci USA 105:E79
27. Sekiya H, Sakota K (2008) J Photochem Photobiol: C Photochem Rev 9:81–91
28. Fang WH (1998) J Am Chem Soc 316:7568–7576
29. Kohtani S, Tagami A, Nakagaki R (2000) Chem Phys Lett 316:88–93
30. Matsumoto Y, Ebata T, Mikami N (2002) J Phys Chem A 106:5591–5599
31. Meuwly M, Bach A, Leutwyler S (2001) J Am Chem Soc 123:11446–11453
32. Coussan S, Meuwly M, Leutwyler S (2001) J Chem Phys 114:3524–3534
33. Tanner C, Manca C, Leutwyler S (2003) Science 302:1736–1739
34. Bach A, Tanner C, Manca C, Frey HM, Leutwyler S (2003) J Chem Phys 119:5933–5942
35. Manca C, Tanner C, Coussan S, Bach A, Leutwyler S (2004) J Chem Phys 121:2578–2590
36. Tanner C, Manca C, Leutwyler S (2005) J Chem Phys 122:204326/1–204326/11
37. Fernández-Ramos A, Martïnez-Núñez E, Vázquez SA, Ríos MA, Estévez CM, Merchán M, Serrano-Andrés L (2007) J Phys Chem A 111:5907–5912
38. Mehata MS (2007) Chem Phys Lett 436:357–361
39. Mehata MS (2008) J Phys Chem B 112:8383–8386
40. Zimmer M (2002) Chem Rev 102:759–781
41. Vendrell O, Gelabert R, Moreno M, Lluch JM (2008) J Chem Theory Comp 4:1138–1150
42. Pakhomov AA, Martynov VI (2008) Chem Biol 15:755–764
43. Vendrell O, Gelabert R, Moreno M, Lluch JM (2008) J Phys Chem B 112:5500–5511
44. Vendrell O, Gelabert R, Moreno M, Lluch JM (2008) J Phys Chem B 112:13443–13452
45. Kandori H (2000) Biochim Biophys Acta 118:177–191
46. Sato Y, Hata M, Neya S, Hoshino T (2006) J Phys Chem B 110:22084–22812
47. Bondar AN, Baudry J, Suhai S, Fischer F, Smith JC (2008) J Phys Chem B 112:14729–14741

48. Serrano-Andrés L, Roos BO (1996) J Am Chem Soc 118:185–195
49. Sobolewski AL, Domcke W (1999) Chem Phys Lett 315:293–298
50. Kyrychenko A, Waluk J (2006) J Phys Chem A 110:11958–11967
51. Parac M, Grimme S (2002) J Phys Chem A 106:6844–6850
52. Hättig C, Köhn A (2002) J Chem Phys 117:6939–6951
53. Nosenko Y, Kunitski M, Riehn C, Thummel RP, Kyrychenko A, Herbich J, Waluk J, Brutschy B (2008) J Phys Chem A 112:1150–1156
54. Furche C, Ahlrichs R (2002) J Chem Phys 117:7433–7447
55. Kyrychenko A, Herbich J, Izydorzak M, Wu F, Thummel RP, Waluk J (1999) J Am Chem Soc 121:11179–11188
56. Kornyshev AA, Kuznetsov AM, Spohr E, Ulstrup J (2003) J Phys Chem B 107:3351–3366
57. Agmon N (2005) J Phys Chem A 109:13–35
58. Lapid H, Agmon N, Petersen MK, Voth GA (2005) J Chem Phys 122:014506/1–014506/11
59. Mezer A, Friedman R, Noivirt O, Nachliel E, Gutman M (2005) J Phys Chem A 109:11379–11378
60. Friedman R, Fischer S, Nachliel E, Scheiner S, Gutman M (2007) J Phys Chem B 111:6059–6070
61. Wang S, Smith SC (2006) J Phys Chem B 110:5084–5093
62. Sakota K, Komoto Y, Nakagaki M, Ishikawa W, Sekiya H (2007) Chem Phys Lett 435:1–4
63. Tanner C, Thut M, Steinlin A, Manca C, Leutwyler S (2006) J Phys Chem A 110:1758–1766
64. Agmon N (2007) J Phys Chem B 111:7870–7878
65. Mente S, Frankland CJV, Reynolds L, Maroncelli M (1998) Chem Phys Lett 293:515–522
66. Yokoyama H, Watanabe H, Omi T, Ishiuchi SI, Fujii M (2001) J Phys Chem A 105:9366–9374
67. Taketsugu T, Yagi K, Gordon MS (2005) Int J Quant Chem 104:758–772
68. Smedarchina Z, Siebrand W, Fernández-Ramos A, Gorb L, Leszczynski J (2000) J Chem Phys 112:566–573
69. Gordon MS (1996) J Phys Chem 100:3974–3979
70. Fernández-Ramos A, Smedarchina Z, Siebrand W, Zgierski MZ, Rios MA (1999) J Am Chem Soc 121:6280–6289
71. Shukla MK, Mishra PC (1998) Chem Phys 230:187–200
72. Casadesúus R, Moreno M, Lluch JM (2003) Chem Phys 290:319–336
73. Chaban GM, Gordon MS (1999) J Phys Chem A 103:185–189
74. Fernández-Ramos A, Smedarchina Z, Siebrand W, Zgierski MZ (2001) J Chem Phys 114:7518–7526
75. Svartsov YN, Schmitt M (2008) J Chem Phys 128:214310/1–214310/9
76. Vu TBC, Kalkman I, Meerts WL, Svartsov YN, Jacoby C, Schmitt M (2008) J Chem Phys 128:214311/1–214311/10
77. Koizumi Y, Jouvet C, Norihiro T, Ishiuchi SI, Dedonder-Lardeux C, Fujii M (2008) J Chem Phys 129:104311/1–104311/10
78. Sakota K, Kageura Y, Sekiya H (2008) J Chem Phys 129:054303/1–054303/10
79. Hu WP, You RM, Yen SY, Hung FT, Chou PH, Chou PT (2003) Chem Phys Lett 370:139–146
80. Hung FT, Hu WP, Chou PT (2001) J Phys Chem A 105:10475–10482
81. Chou PT, Wei CY, Chang CP, Meng-Shin K (1995) J Phys Chem 99:11994–12000
82. Herbich J, Hung CY, Thummel RP, Waluk J (1996) J Am Chem Soc 118:3508–3518
83. Kyrychenko A, Stepanenko Y, Waluk J (2000) J Phys Chem A 104:9542–9555
84. Nosenko Y, Kunitski M, Thummel RP, Kyrychenko A, Herbich J, Waluk J, Riehn C, Brutschy B (2006) J Am Chem Soc 128:10000–10001
85. Nosenko Y, Kyrychenko A, Thummel RP, Waluk J, Brutschy B, Herbich J (2007) Phys Chem Chem Phys 9:3276–3285
86. Weigend F, Häser M, Patzelt H, Ahlrichs R (1998) Chem Phys Lett 294:143–152

87. Schäfer A, Huber C, Ahlrichs R (1994) J Chem Phys 100:5829–5835
88. Manca C, Tanner C, Leutwyler S (2005) Int Rev Phys Chem 24:457–488
89. Mente S, Maroncelli M (1998) J Phys Chem A 102:3860–3876
90. Scott WRP, Hünenberger PH, Tironi IG, Mark AI, Billeter SR, Fennen J, Torda AE, Huber T, Krüger P, van Gunsteren WF (1999) J Phys Chem A 103:3596–3607
91. Sobolewski AL, Domcke W, Dedonder-Lardeux C, Jouvet C (2002) Phys Chem Chem Phys 4:1093–1100
92. Sobolewski AL, Domcke W (2007) J Phys Chem A 111:11725–11735
93. Rode MF, Sobolewski AL (2008) Chem Phys 347:413–421
94. Lan Z, Frutos LM, Sobolewski AL, Domcke W (2008) Proc Natl Acad Sci USA 105:12707–12712
95. Yau WM, Wimley WC, Gawrisch K, White SH (1998) Biochemistry 37:14713–14718
96. Grossfield A, Woolf TB (2002) Langmuir 18:198–210
97. Gaede HC, Yau WM, Gawrisch K (2005) J Phys Chem B 109:13014–13023
98. Smirnov AV, English DS, Rich RL, Lane J, Teyton L, Schwabacher AW, Luo S, Thornburg RW, Petrich JW (1997) J Phys Chem B 101:2758–2769
99. Kyrychenko A, Waluk J (2008) Biophys Chem 136:128–135
100. Tieleman DP, Berendsen HJC (1996) J Chem Phys 105:4871–4880
101. Tieleman DP, Marrink SJ, Berendsen HJC (1997) Biochim Biophys Acta 1331:235–270
102. Rodríguez-Prieto F, Mosquera M, Novo M (1990) J Phys Chem 94:8536–8542
103. Mukherjee TK, Ahuja P, Koner AL, Datta A (2005) J Phys Chem A 109:12567–12573
104. Mukherjee TK, Panda D, Datta A (2005) J Phys Chem A 109:18895–18901
105. Ulander J, Haymet ADJ (2003) Biophys J 85:3475–3484
106. Milhaud J (2004) Biochim Biophys Acta 1663:19–51
107. Berkowitz ML, Bostick BL, Pandit S (2006) Chem Rev 106:1527–1539
108. Kwon OH, Jang DJ (2005) J Phys Chem B 109:20479–20484
109. Kyrychenko A, Herbich J, Wu F, Thummel RP, Waluk J (2000) J Am Chem Soc 122:2818–2827
110. Kyrychenko A, Waluk J (2003) J Chem Phys 119:7318–7327
111. Kyrychenko A, Gorski A, Waluk J (2004) J Chem Phys 121:12017–12025
112. Kyrychenko A, Waluk J (2005) J Chem Phys 123:064706/1–064706/10
113. Kyrychenko A, Gawinkowski S, Urbańska N, Pietraszkiewicz M, Waluk J (2007) J Chem Phys 127:134501/1–134501/12

CHAPTER 3

FORMAMIDE AS THE MODEL COMPOUND FOR PHOTODISSOCIATION STUDIES OF THE PEPTIDE BOND

MIRJANA ECKERT-MAKSIĆ[1], IVANA ANTOL[1], MARIO VAZDAR[1], MARIO BARBATTI[2], AND HANS LISCHKA[2,3]
[1]*Division of Organic Chemistry and Biochemistry, Rudjer Bošković Institute, 10002, Zagreb, Croatia*
e-mail: mmaksic@emma.irb.hr
[2]*University of Vienna, Institute for Theoretical Chemistry, Waehringerstrasse 17, A1090, Vienna, Austria*
[3]*Academy of Sciences of the Czech Republic, Institute of Organic Chemistry and Biochemistry, Flemingovo nam. 2, CZ-16610, Prague 6, Czech Republic*

Abstract: Dynamics simulations are an essential step in exploring ultrafast phenomena in photochemistry and photobiology. In this chapter we present results of photodynamics studies for some model compounds for the peptide bond using the on-the-fly surface hopping method. The mechanism of photodissociation of formamide, its protonated forms and methyl substituted derivatives in their lowest singlet excited states in the gas phase is discussed in detail. Merits and demerits of using these simple molecules as models in exploring photochemical and photophysical properties of more complex systems, like peptides and proteins, are emphasized. It is found that in all examined model molecules the major deactivation process after excitation to the S_1 state is dissociation of the peptide C–N bond. The same holds for the deactivation path from the S_2 state, with exception of the O- protonated formamide in which C–O dissociation becomes the major deactivation process. Furthermore, it is shown that substitution by the methyl group(s), as well as protonation, strongly influence the lifetime of both excited states. In the last section application of the newly developed hybrid nonadiabatic photodynamics QM/MM approach in calculating photodissociation of formamide in argon matrix is illustrated.

Keywords: Excited state, Peptide bond, Photodissociation, Nonadiabatic dynamics, Formamide, Substituted formamides, Surface hopping, Environmental effects, QM/MM

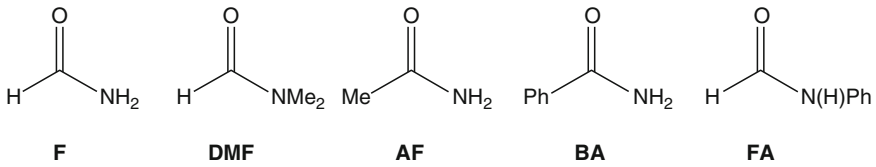
P. Paneth and A. Dybala-Defratyka (eds.), *Kinetics and Dynamics*,
Challenges and Advances in Computational Chemistry and Physics 12,
DOI 10.1007/978-90-481-3034-4_3,

3.1. INTRODUCTION

The peptide functional group (–CO–NH–) is responsible for many important chemical and biochemical processes, from serving as a linkage in proteins to acting as a building block in materials science [1]. Its dual ability to act as both hydrogen bond acceptor (due to the presence of carbonyl oxygen) and donor (due to the presence of the NH group) makes it in particular versatile in molecular assembly and recognition. For example, the H-bonds among peptide units in proteins are the key driving forces for forming organized α-helix and β-sheet secondary structures. The peptide group also plays important role in the pharmacophores of some antibacterial agents, as well as, in designing enzyme inhibitors or synthetic polymers containing peptide linkages [1–3]. Furthermore, interaction of the peptide bond with ultraviolet light is of fundamental importance for understanding the effect of light irradiation on various life forms. It also has significant implications in many other fields such as structural biology [4], laser surgery [5], or photodegradation of biopolymeric materials [6]. Despite a wealth of experimental investigations of photochemistry of polypeptides and proteins, some aspects of the electronic states involved in the initial photoexcitation and subsequent processes are still not well understood. Therefore, thorough exploration of photochemical and photophysical properties of simple model compounds plays an important role in reaching these goals. Among them simple amides which can be considered as the prototypes of peptide bond, like formamide and its alkyl derivatives, have attracted considerable attention both experimentally [7–11] and computationally [12–19]. Most of the older computational studies of these species were limited to investigation of the potential energy surfaces and concomitant conclusions about possible photodeactivation pathways. This approach can be quite useful, but full understanding of the process requires studying dynamics of the molecule.

In photochemical processes usually sufficient initial energy is available from the vertical excitation implying that existing small energy barriers on the potential energy surface can be surmounted and the region of the intersection seam can be reached in less than hundred of femtoseconds to few picoseconds. Because of the relatively short simulation times needed, such cases are specially tailored to an application of surface-hopping on-the-fly dynamics. Most of the dynamics simulations results reviewed in this chapter were obtained by this approach, which will be described in more detail in the following section.

In the present chapter recent results on photodissociation of formamide, its oxygen and nitrogen protonated forms and some of its methyl- and phenyl-substituted derivatives will be reviewed (Scheme 3-1). The emphasis will be on dynamics simulations of their gas phase photolysis in the low excited valence singlet states that have been recently carried out in our group [15–17]. In the last section preliminary results for photodissociation of formamide in the argon cage using recently developed hybrid nonadiabatic photodynamics QM/MM approach [20] will be presented. Since most of practical photochemistry occurs in condensed



Scheme 3-1. Molecules considered in this chapter

media, it is important to understand the role of complex environments in shaping and directing photochemical processes. In that sense, photodissociation of small molecules in rare gas matrices, which can be considered as a prototype of reaction dynamics in condensed media, has been the subject of several recent experimental and theoretical studies [21].

3.2. AN OVERVIEW OF COMPUTATIONAL METHODS FOR STUDYING DYNAMICS OF FAST PHOTODISSOCIATION PROCESSES

Dynamics simulations of photodissociation processes are specially challenging to perform. They demand that the electronic structure method employed to describe the Born–Oppenheimer surfaces is robust enough to describe the dissociation process properly, giving correct dissociation energies, charge distribution between fragments, state crossings and degeneracies. Moreover, when energy barriers are present along the dissociation path, the dynamics method should be efficient enough to deal with relatively long processes. If light fragments are expected to be formed, the method should also be able to take into account tunneling processes in competition with classical dissociation. If the dissociation channel has a minor output for a given excitation energy, the dynamics methods should allow to assess statistically relevant information. The traditional approach to satisfy these requirements is to perform wavepacket propagation by solving the time-dependent Schroedinger equation (TDSE) on a Born–Oppenheimer surface of reduced dimensionality [22]. Although precise information can be obtained in such way, it has a serious drawback: educated guesses about what are the relevant nuclear degrees of freedom should be done and potential energy surfaces must be built for these degrees. These are cumbersome procedures which demand individualized treatment of each molecular system under investigation and intrinsically restrict the research to small molecules.

With development of computational resources, dynamics simulations of photoexcited-state processes for relatively large molecular compounds have become available by using mixed-quantum classical dynamical (MQCD) approaches [23]. These approaches combine, as we shall explain in this section, classical information coming from the Newton's equation of motion with quantum information coming from both, time-dependent and time-independent Schroedinger equations. The methodology is deeply simplified in comparison to the full quantum treatment,

allowing for the development of general procedures to treat any molecule. The problems related to the reduction of dimensionality and to the construction of potential energy surfaces are avoided by the usage of on-the-fly methodologies, where energies, energy gradients, non-adiabatic coupling terms and any other property that may be needed during the dynamics are computed for all degrees of freedom as needed along classical trajectories.

Naturally, there is a price to pay for such simplifications. As in the full quantum treatment, the electronic structure methods should still be able to give a robust description of the multitude of electronic states for bound and dissociated nuclear conformations. Thousands of single point electronic structure calculations are necessary for the propagation of a single trajectory. This is the main time consuming step of the simulations and effectively limits the time propagation to few picoseconds and few hundreds of trajectories. Because of the local approximation implicit in the classical equations and enforced by the on-the-fly approach, non-local phenomena like tunneling cannot be described. Because of the limited number of trajectories, low probability dissociation channels cannot be adequately described too.

If, however, the process to be investigated involves several degrees of freedom and many electronic states occur within few picoseconds, and one is interested in the statistically most relevant channels, MQCD approaches can be used and may be expected to give relevant information. These are exactly the case studies discussed in this chapter.

The MQCD approaches constitute a class of methods for dynamics simulations that, as mentioned above, combine quantum and classical information for describing the nuclear time development. They include the surface hopping [24], the multiple spawning [25], and the Ehrenfest [26] methods. In this section we shall focus on the surface hopping method, which we have used in our investigations. Here only brief overview will be given and for more detailed description the reader is referred to [27, 28].

In the surface hopping method, the nuclear motion is propagated on a single Born–Oppenheimer surface by solving Newton's equations. The instantaneous non-adiabatic transition probabilities to other states are simultaneously computed every time step and a stochastic algorithm decides in which state the molecule will be propagated in the next steps of a single trajectory. This process is repeated a large number of times so as to have an ensemble of trajectories, which should statistically emulate the quantum nuclear wavepacket.

The integration of the Newton's equation is carried out by some standard algorithm such as the Velocity Verlet algorithm [29]. A time step of 0.5 fs or smaller is good enough to describe the fast hydrogen vibrational stretching motions keeping a good level of energy conservation along the dynamics simulations.

The computation of the non-adiabatic transition probabilities is what distinguishes the several surface hopping approaches. The simplest way to estimate them is assuming that the hopping probability is unity as soon as the energy gap between the current state where the molecule is propagated and some other state drops below a pre-defined threshold [30]. Naturally, this approach gives a

significantly approximated non-adiabatic treatment. More advanced algorithms employing the Landau–Zener formula [31] and the change of configuration coefficients in the electronic wave function [32] have been proposed as well. In the early nineties, Tully proposed the fewest switches approach [24], which is used in our simulations. In this approach, a local approximation of the TDSE for the nuclei is propagated simultaneously to the Newton's equation and the adiabatic population obtained as the solution of the TDSE is employed to compute the non-adiabatic transition probabilities. For integration of the TDSE it is necessary to compute non-adiabatic coupling terms between electronic states. This can be done either by spatial derivatives or by time derivatives of the electronic wavefunction [33]. In our simulations, the TDSEs are integrated using the Butcher algorithm [34] with steps of 0.01 fs. Note that electronic structure calculations are performed only for the classical intervals (0.5 fs) and the quantities necessary for the TDSE integration are obtained by interpolation between two classical time steps. Decoherence of the density matrix is imposed by an ad hoc algorithm that corrects the adiabatic populations every step [35].

When a hopping event between two surfaces takes place, it does it through a small but finite energy gap. To ensure total energy conservation, the change in the potential energy due to the hopping is balanced by a corresponding change in the kinetic energy. This can be done either by rescaling the momentum or by adding momentum to some specific direction [36]. In our investigations, momentum is added to the non-adiabatic coupling vector direction. In some instances, it may occur that the stochastic algorithm attempts to make a hopping to an upper surface at a nuclear conformation for which there is not enough energy. In such cases, it is usual to forbid the hopping occurrence. The time uncertainty principle can also be used to search for a neighbor conformation where the hopping would be allowed [37].

Initial conditions for the trajectories can be generated in a number of different ways. In our gas phase investigations, a Wigner distribution within the harmonic approximation is built for the nuclear degrees of freedom in the electronic and vibrational ground state [22]. Each point of this distribution is taken as initial condition for one single trajectory starting in the excited state. This simulates a ground state wavepacket that is instantaneously photoexcited in the Franck–Condon region. In our simulations of formamide in an argon matrix, this approach cannot be used because the normal modes for the whole system are not available. Moreover, care should be taken that the two subsystems have correct vibrational temperatures. This can be achieved by relaxing the whole structure and starting the dynamics always from this particular conformation but with different random velocities. The velocities are generated by sampling Gaussian-distributed random numbers for each degree of freedom [38] and imposing that the norm of the formamide velocity is always equivalent to half of its zero point energy as predicted by the Virial theorem and the norm of the argon matrix velocity is equivalent to the usual $3N_a k_B T/2$, where N_a is the number of atoms in the matrix, k_B is the Boltzmann constant and T is the temperature.

As already mentioned, the investigation of photodissociation processes requires robust electronic structure methods. They should be able to describe the ground and the excited states of the bound and the dissociated molecule preferentially at the same level. As we discuss in the next section, our choice has been to employ the CASSCF method, which is sometimes supplemented by the MRCI method. Besides being a computationally efficient ab initio method with well established procedures to compute energy gradients and non-adiabatic coupling vectors analytically [39–42], it allows the computation of excited states with correct description of multireference conformations and of energy dissociation levels. It is important to stress that this method represents a satisfactory compromise between accuracy and computational efficiency demanded by the on-the-fly dynamics simulations, where a full quantum chemical calculations have to be performed at every time step. However, since the CASSCF method does not include electron dynamic correlation potential energy surfaces computed at this level of theory should always be validated by comparison with the surfaces computed at higher levels of theory. Additionally, the building of computationally manageable active space often leads to a neglecting of σ–π correlation [43], which causes overestimation of the vertical excitation energies of the ππ* ionic states and considerable overshooting in photodynamics starting in these states. This problem was also encountered in the dynamics simulations of the molecules considered in this chapter and its impact on the results of dynamics simulations was thoroughly discussed in our previous studies on formamide and its derivatives [15–17]. Besides that, in the investigation of photodissociation processes, the active space must be well balanced to properly describe the bond breaking. For formamide, for instance, a relatively large active space composed of ten electrons in eight orbitals including two pairs of σ/σ* orbitals should be employed to achieve this requirement (see Section 3.3).

The inclusion of environmental effects on the molecule is achieved by means of a hybrid quantum-mechanical/molecular-mechanical method (QM/MM). The QM treatment is restricted to the fraction of the molecular system affected by the excitation and where the dissociation occurs. The remaining atoms are treated by molecular mechanics. The interaction between the two subsystems can be included either by treating the electrostatic interactions at the MM level (mechanical embedding) or at the QM level by representing the environment as point charges (electrostatic embedding) [44]. Usually, the latter is the most adequate when dealing with electronically excited molecules. Later in this chapter, we shall discuss photodissociation of formamide embedded into an argon matrix. In this case, electrostatic interactions can be neglected and the environment influence is reduced to the van der Waals interactions, which are computed at the MM level.

3.3. COMPUTATIONAL DETAILS

For the characterization of the potential energy surfaces (PES) in our previous photodissociation studies of formamide and its derivatives, which will be reviewed here, state-averaged complete active space (CASSCF) and/or multireference

configuration interaction methods with singles and doubles (MR-CISD) in connection with the 6-31G(d) basis set were used, if not stated otherwise [15–17]. In all considered molecules three states with equal weights (SA-3) were taken in the state average procedure. The active space comprised ten electrons in eight orbitals, which in case of the parent molecule were two σ/σ^* pairs, two nonbonding lone pairs and one π/π^* pair of orbitals. For the MR-CISD calculations, the total configuration state function (CSF) space was constructed by applying single and double excitations from all internal (active and double occupied) orbitals into all virtual orbitals for all reference CSFs. In this case, the generalized interacting space restriction [45] was employed. Inner shell orbitals were always kept frozen. The stationary points and the minima on the crossing seams (MXS), analytic gradients and nonadiabatic coupling vectors were computed based on the methods described in [15–17]. In discussing results from literature sources the method used in calculations will be specified at spot.

Simulations of photodissociation dynamics were carried out using the on-the-fly surface hopping method described in the previous section based on ab initio CASSCF(10,8)/6-31G(d) calculations of the electronic energies and wave functions. For each of the considered states batches of 100 trajectories (120 in case of protonated formamide) were calculated. The initial Cartesian coordinates and momenta were selected from a Wigner quantum harmonic oscillator distribution at the ground electronic and vibrational state [22].

In QM/MM dynamics calculations of formamide in Ar matrix, the mechanical embedding scheme was used, implying that electrostatic interactions were neglected. The formamide was treated fully quantum mechanically using gradients and energies from the COLUMBUS program at the same level of theory as in our gas phase study. The Ar matrix and formamide-argon interaction were treated on the molecular mechanics level using energies and gradients from TINKER program employing a modified OPLSAA force field [46] with ChelpG [47]-charges for formamide calculated at the B3LYP/cc-pVTZP level of theory. Description of creating initial conditions generation for subsequent dynamic calculations will be described in detail in a forthcoming paper [48]. All CASSCF and MR-CISD calculations were performed with COLUMBUS programs [49–51] using analytic MRCI gradients and nonadiabatic coupling vectors as decribed in references [39–42]. The molecular mechanics calculations of energy and gradients were performed by using TINKER program [52]. Dynamic calculations were performed using the NEWTON-X program [53].

3.4. SIMULATIONS OF NON-ADIABATIC PHOTODYNAMICS OF FORMAMIDE

3.4.1. Gas Phase Studies

Photodissociation of formamide has been studied experimentally in the gas phase [8, 11], in solution [9] and in argon and xenon matrices [7, 10], as well as computationally in the gas phase [7, 12, 15]. Photochemical experiments in the

gas phase with 206.2 nm (6.01 eV) radiation revealed that formamide undergoes photodecomposition at 150°C [11]. Products detected were CO, H_2 and NH_3 with quantum yields of 0.8, 0.6 and 0.2, respectively. Their formation was proposed to arise from three major dissociation processes of formamide affording $NH_2 + CO + H$, $HNCHO + H$ and $NH_3 + CO$, respectively, as the primary products. Subsequently, Liu and coworkers calculated energies of the transition structures for possible photodissociation paths at the S_1 and T_1 surfaces using CASPT2/CASSCF(8, 7)/6-31G(d) method [12]. They found that the transition state for the C–N dissociation in both states has the lowest energy relative to the ground state, indicating that this decomposition channel is the most probable one. In another set of experiments photodissociation of formamide at 205 nm was investigated by measuring laser induced fluorescence spectra [8]. It was shown that under these conditions formamide undergoes cleavage of the hydrogen atom in the α- or β-position to the carbonyl group. From the measured channel branching ratio and translational energy releases and CASSCF calculations, it was concluded that dissociation takes place on the triplet surface with considerable recombination barriers along the reaction coordinate.

A more detailed investigation of the potential energy surface for formamide photodissociation using CASSCF and MR-CISD calculations was recently carried out in our group [15]. In this work the reaction paths proposed experimentally [11] and considered by Liu and coworkers [12] were recalculated. Additionally, the minima on the crossing seams (MXSs) for dissociation from the first and the second valence excited singlet states were characterized. The resulting energy scheme for dissociation from both states calculated at the CASSCF level is displayed in Figure 3-1. It is appropriate to mention at this place, that although the calculated vertical excitation energy for the S_1 state agrees within 0.3 eV with the experimental value, the computed energy of the S_2 state is too large by 1.3 eV, due to the missing σ–π correlation mentioned in Section 3.2. A similar trend, as will be shown later, was observed in all other formamide derivatives. Therefore, for each of the considered species energies of all critical points on the potential energy surface were also recalculated at the MR-CISD and MR-CISD+Q levels of theory and in each case the energetic ordering of the TSs and MXSs was qualitatively similar to that obtained with the CASSCF method. Hence, it is reasonable to expect that dynamical calculations starting in the S_2 state for all of the studied species will be affected to a similar extent by the excess of the vertical energy of this state, thus enabling reliable insight into the effect of structural modifications on the photodissociation mechanism of the peptide linkage. An additional test was made by running a set of 50 trajectories starting from the S_2 state of formamide using the MR-CISD(6, 4)/SA-3-CASSCF(10, 8)/6-31G(d) method [54]. It was found that the fraction of trajectories for the major dissociation path, i.e. CN dissociation, remained essentially the same as in the CASSCF calculations, thus providing additional justification for using the CASSCF method in the present case. It should be, however, mentioned that some differences in distribution of other trajectories were observed, the main one being the disappearance of trajectories for CO

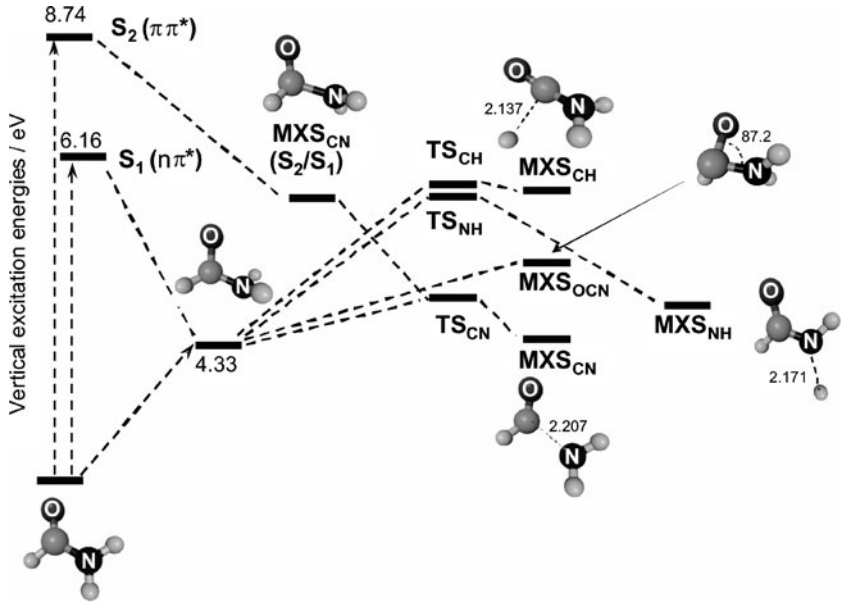


Figure 3-1. Energy scheme for formamide displaying vertical excitation energies, saddle points and MXSs calculated at the CASSCF(10,8)/6-31G(d) level

dissociation. In case of CASSCF calculations, as will be seen later trajectories of this type constituted ca. 13% of the total number of trajectories.

An important piece of information obtained from analyzing the results of dynamics simulations concerns the lifetime of the involved states during photodissociation process, defined as the time necessary for the fraction of trajectories in the excited state to drop to 1/e. For all of the molecules discussed in this chapter the lifetime was calculated by exponential fitting the occupation in the considered excited state with the function $f(t) = f_0 + \exp(-(t - \tau_d)/\tau_e)$ where $f(t)$ is the fraction of trajectories and τ_d and τ_e are fitting parameters which represent initial delay and the decay time constant, respectively. The parameter f_0 is the fraction of trajectories that do not deactivate at all. The overall lifetime is defined as the sum of τ_d and τ_e. The lifetime of the S_1 state of formamide obtained from the average of 100 trajectories was found to be 441 ± 4 fs. It is interesting to note that for the dynamics simulations started in the S_1 state, the average occupation remains practically unchanged up to 43 fs. After this initial stage its population starts to decay exhibiting an exponential profile (Figure 3-2a).

Dynamics from the $S_2(\pi\pi^*)$ state proceeded much faster than that starting in the S_1 state. This state starts to depopulate with a delay of ca. 12 fs and has a lifetime of 66 ± 1 fs (Figure 3-2b). The reason for the very fast $S_2 \rightarrow S_1$ decay is mainly due to the presence of the S_2/S_1 conical intersection in the vicinity of the Franck–Condon region (Figure 3-1). It is also likely that overestimation of the initial vertical

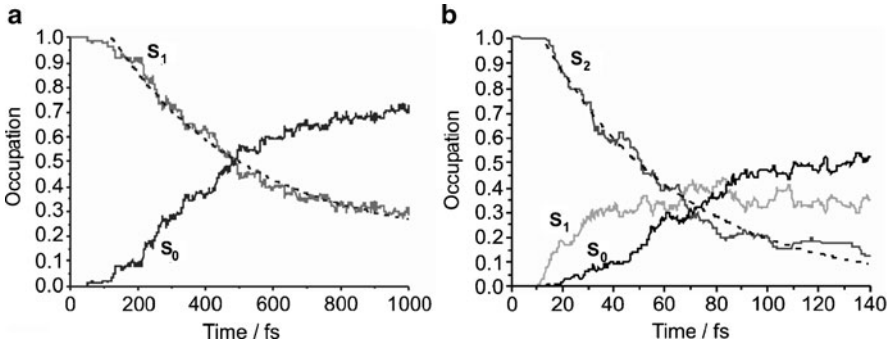
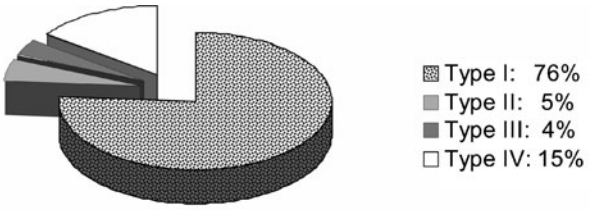


Figure 3-2. Fraction of trajectories for each state as a function of time following photoexcitation of formamide into the S_1 (**a**) and S_2 (**b**) initial states. The *dashed line* shows the exponential fitting curves



Type I: C-N dissociation followed by separation of the HCO and NH_2 fragments or formation of CO and NH_3 in the ground state

Type II: formamide switches to the S_0 either via $\mathbf{MXS_{CN}}(S_1/S_0)$ and recombines back or gets to the ground state via $\mathbf{MXS_{OCN}}(S_1/S_0)$

Type III: stretching of the C-H bond after passing to the S_0

Type IV: no dissociation and no decay to the ground state

Figure 3-3. Photodeactivation paths observed in dynamics simulations of formamide starting from the S_1 state

excitation energy to the ππ* state by the CASSCF method also contributes to this feature.

Analysis of the trajectories calculated in dynamics simulations starting in the S_1 state, showed four types of trajectories (Figure 3-3) representing (I) C–N dissociation, (II) return to the non-dissociated formamide ground state and (III) C–H dissociation. In the last type (IV), formamide remained non-dissociated in the S_1 state till the end of the simulation.

Most of the trajectories (76%) belong to type I, in which hopping to the ground state takes place in a region of the crossing seam along the C–N dissociation

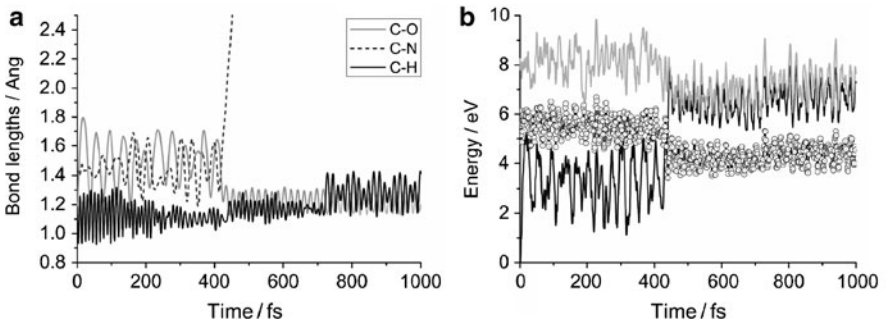


Figure 3-4. Time evolution of C–N, C–O and C–H bond distances (**a**) and potential energies of the S_0, S_1 and S_2 states in a typical trajectory for type I (**b**) deactivation paths following excitation to the S_1 state. The *dots* in the potential energy graph indicate the current state of the system in each time step

coordinate. An example of a typical trajectory is shown in Figure 3-4. In this particular trajectory, the initial excited-state motion starts with C–O stretching which gradually decreases during the evolution to the hopping point and the C–N stretching amplitude increases. After hopping to the ground state the energy gets transferred to the translational motion of the HCO and NH_2 fragments and in ca. 16% of the trajectories, subsequent C–H dissociation from the vibrationally hot HCO radical and transfer of hydrogen atom to NH_2 is observed.

In trajectories of type II, which constitute only 5% of the total number of trajectories, formamide decays to the ground state either via the C–N conical intersection $MXS_{CN}(S_1/S_0)$ and reverts to the starting structure on the S_0 PES or switches to the ground state via a non-dissociated conical intersection MXS_{OCN} (Figure 3-1) and does not dissociate later on. In trajectories of type III (4%), hoppings to the ground state trigger activation of the C–H vibrations and the molecule undergoes gradual lengthening of the C–H bond. Full C–H dissociation was noted in only one trajectory. Finally, in 15% of trajectories formamide does not decay to the ground state (Type IV) and does not dissociate within the simulation time. In this type of trajectories the overall energy of the system is simply distributed among C–O, C–N and C–H stretching modes.

Similar to the S_1 state, the major deactivation channel from the S_2 state is found to be C–N bond cleavage. It is found in 68% of the trajectories (type I). However, in contrast to dissociation from the S_1 state, the overall process is, as mentioned above, much faster. The stretching of the C–N bond starts immediately after photoexcitation and proceeds on the S_2 surface till the decay to the S_1 state, which occurs in an average time of 31 fs. The system remains on the S_1 surface only for a short time and then switches to the S_0 state in an average time of ca. 65 fs (Figure 3-5a). It is important to note that the latter process is much faster than in case of dissociation starting in the S_1 state. The reason for that can be attributed to an impact of the energy transferred from the S_2 state and the momentum that the system acquires along the C–N stretching coordinate during its motion from the

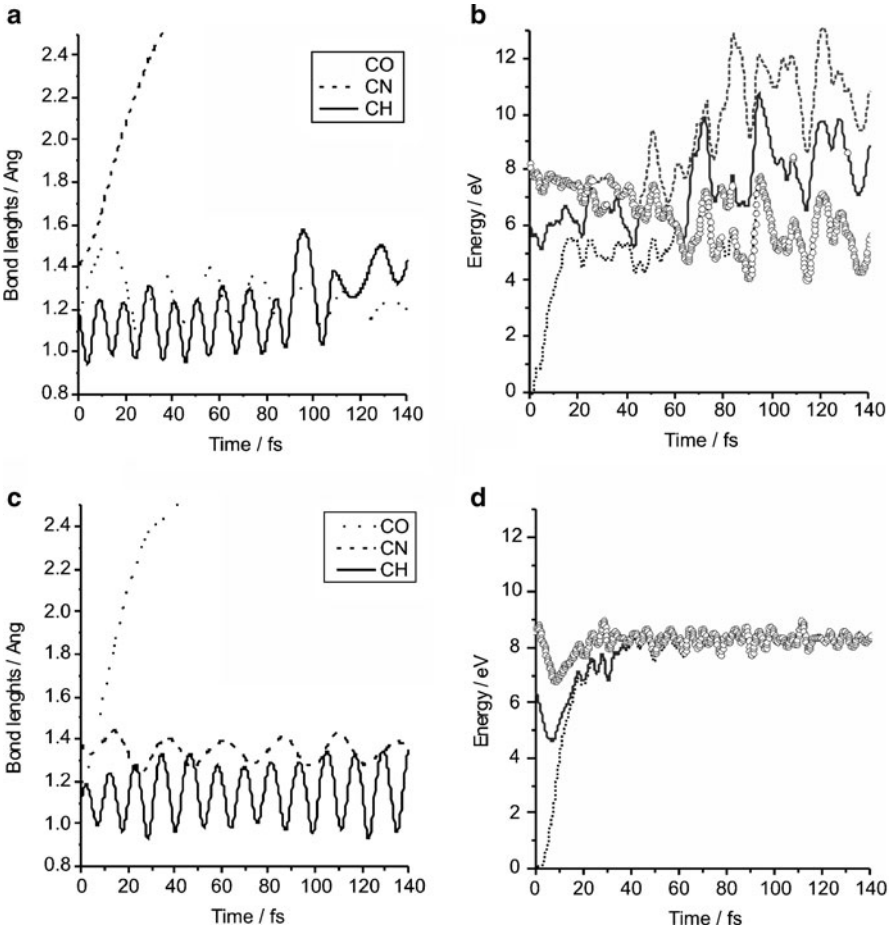


Figure 3-5. Time evolution of C–N, C–O and C–H bond distances and potential energies of the S_0, S_1 and S_2 states in two typical trajectories for the C–N (**a**, **b**) and the C–O (**c**, **d**) deactivation paths following excitation to the S_2 state. The *dots* in the potential energy graphs indicate the current state of the system in each time step

Franck–Condon region to the $MXS_{CN}(S_1/S_0)$. A similar trend has been also observed in the dynamics calculations of the protonated formamide [16] and the methaniminium cation [55].

Another interesting feature, which is not observed in dissociation from the S_1 state concerns activation of the C–O dissociation channel. It is observed in 13% of trajectories, resulting in dissociation of the oxygen atom and a concomitant merging of all three potential energy surfaces along the trajectory (Figure 3-5c and d). On average, in trajectories of this type excited molecule remains in S_2 state for 106 fs. In the remaining trajectories (not shown here) initial motion follows the C–O

stretching mode, but C–O bond cleavage is prevented by recombination to the starting molecule, followed by subsequent C–N bond dissociation. In these trajectories, average times for the first hop from the S_2 to the S_1 state and from the S_1 to the S_0 (73 and 100 fs, respectively) are somewhat longer than in the case of fast C–N dissociation in previously described trajectories.

3.4.2. Photodissociation of Substituted Formamides

Photodissociation of substituted formamides in the gas phase has attracted much less attention than the photodissociation of formamide. The most detailed experimental work available is that published by Butler and coworkers on photodissociation of *N,N*-dimethylformamide [56, 57]. Some earlier gas phase studies [58], as well as photolysis studies of acetamide in aqueous solution [59], in water [60], and in dioxane or hexane solutions [61] have been also reported. Computational results are available for *N,N*-dymethylformamide (DMF) [12], acetamide (AF) [14] and phenyl substituted formamides benzamide (BA) and formanilide (FA) [18, 19]. Since we are primarily interested in the effect of substituents on the photolysis of the C–N bond within the peptide linkage emphasis in presenting results will be put on dynamics of this process. In this regard substitution by the methyl groups is particularly interesting since they, to some extent, mimic the effect of a protein chains.

3.4.2.1. *N,N*-Dimethylformamide

Based on the measuring the photofragment velocity and angular distribution of *N,N*-dimethylformamide (DMF) at 193 nm excitation in the second excited singlet state ($\pi\pi^*$) Butler and coworkers proposed that the cleavage of the central C–N bond in this compound proceeded either along the excited state pathway leading to HCO in the ground state and electronically excited $N(CH_3)_2$ radicals (Eq. (3-1)) or through a decay channel resulting in HCO and $N(CH_3)_2$ (Eq. (3-2)) fragments in the ground state [56, 57]. Besides, the first of these two processes was found to be much slower than the second. They also discussed cleavage of the N–CH_3 bond leading to formation of $HCONCH_3$ + CH_3 (Eq. (3-3)):

$$HCON(CH_3)_2 + 193\ \text{nm} \rightarrow HCO\ (\tilde{X}^2A') + N(CH_3)_2(\tilde{A}^2A_1)(\text{slow}) \quad (3\text{-}1)$$

$$\rightarrow HCO\ (\tilde{X}^2A') + N(CH_3)_2(\tilde{X}^2B_1)\ (\text{fast}) \quad (3\text{-}2)$$

$$\rightarrow HCONCH_3 + CH_3 \quad (3\text{-}3)$$

Subsequently, Fang and coworkers carried out CASSCF calculations for the above proposed photodissociation paths in the first excited singlet ($n_O\pi^*$) and triplet states [12]. Based on the obtained results they concluded that most of the cleavage of the central C–N bond (i.e. the bond between carbon atom of the carbonyl group

and the nitrogen atom) in DMF takes place on the S_1 surface and partially on the T_1 surface. This process will be hereafter denoted as N–CO cleavage in order to differentiate it from the cleavage of the bond between nitrogen atom and the methyl group which will be denoted as N–Me. In addition, they found evidence for formation of $HCONCH_3$ and CH_3 from the T_1 surface. It is also important to mention that they excluded the possibility of forming products from the S_2 state, which according to Butler's work was responsible for formation of the ground state HCO radical and the excited state $N(CH_3)_2$ radical. Namely, they claimed that the molecule upon excitation of 193 nm, which was employed experimentally, could not reach the S_2 state, but instead resulted in molecules in the excited vibrational states of the S_1 electronic state. However, because of the vibronic coupling between the S_2 and the S_1 electronic states, some of the molecules in the highly excited vibrational state of the S_1 electronic state dissociate into ground state HCO and the excited state $N(CH_3)_2$ fragments along the S_2 pathway, whereas the rest of them relax toward the S_1 minimum.

Recently, we carried out dynamics simulations of the processes starting from the S_1 and S_2 states [17] which led to somewhat different conclusions to those obtained by Fang [12]. Our main aim in undertaking this study was to provide a deeper insight into mechanism and dynamics of photodissociation of the N–CO bond in DMF in the S_1 ($n_O\pi^*$) and the S_2 ($\pi\pi^*$) excited singlet states. Given that in the experimental setup, the 193 nm light was used (with excitation energy of 6.42 eV), and the experimental absorption maximum for the $\pi\pi^*$ state of DMF appears at 197.4 nm (6.28 eV), involvement of the S_2 state in considering photodissociation processes is, in our opinion, fully justified. Stationary points (energy minima and saddle points) and minima on the crossing seam obtained in this work on the singlet energy surfaces of the considered singlet excited states are shown in Figure 3-6.

Their energy ordering was found to be similar to that in formamide with one notable exception. Namely, in DMF the global minimum on the crossing seem corresponds to the MXS responsible for dissociation of the N–CH_3 bond ($MXS_{N\text{-}Me}$), while in formamide the lowest MXS corresponds to the N–CO bond dissociation. However, although the $MXS_{N–Me}$ is the lowest energy MXS in DMF it is not expected that it will get activated since the activation of conical intersections is regulated by the corresponding dissociation barrier, which is by 0.69 eV (0.59 ZPVE corrected) higher than the barrier for the cleavage of the central C–N bond [17]. The S_1/S_0 MXSs for the N–CO ($MXS_{N\text{-}CO}$), and the H–CO (MXS_{CH}) dissociation were found to be higher in energies than $MXS_{N–Me}$ by 0.40 and 1.67 eV, respectively. In addition, similarly as in formamide, an additional MXS related to the OCN bending angle (MXS_{OCN}) is located at 5.40 eV. It is important to note that, likewise to formamide (Figure 3-1), all calculated MXSs possess energies below the vertical excitation energy to the S_1 state.

Comparison with the results for formamide discussed in the previous section reveals that replacement of the amino hydrogen atoms by methyl groups leads to pronounced lengthening of the lifetime in the S_1 excited state (by almost 700 fs), while the deactivation process in the S_2 state shortens by almost 50%.

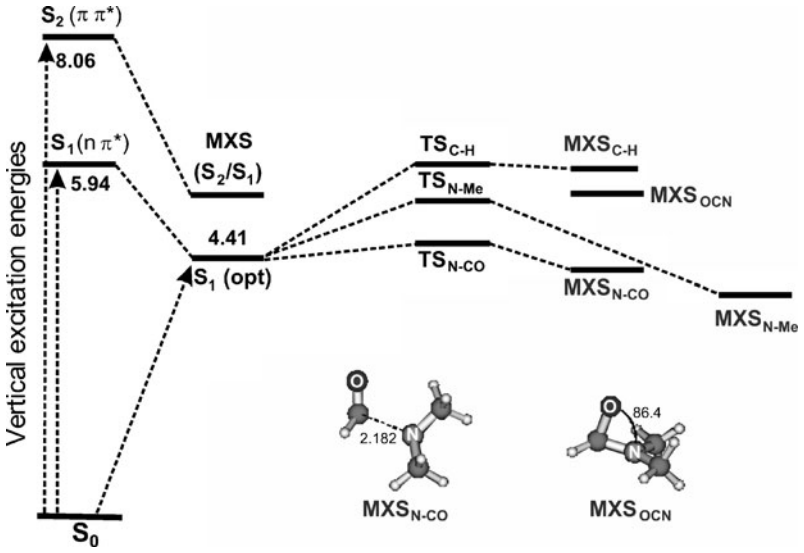


Figure 3-6. Energy scheme for *N,N*-dimethylformamide displaying vertical excitation energies, saddle points and MXSs calculated at the CASSCF(10, 8)/6-31G(d) level. Also included is a schematic presentation of optimized structures of MXSs for dissociation of the central C–N bond ($MXS_{N\text{-}CO}$) and for nondissociative switch (MXS_{OCN}) to the ground state

Results of statistical analysis of the trajectories obtained in dynamics simulations starting in the first excited singlet state ($n_O\pi^*$) are summarized in Figure 3-7.

Their perusal shows that 57% of trajectories of S_1 excited DMF decays to the ground state at the $\mathbf{MXS_{N-CO}}$ region of the crossing seam, related to the N–CO bond stretching. In 41% of trajectories, the excited molecule evolves along the S_1 PES and ends up in the minimum where it stays trapped till the end of the simulation time. It is noteworthy that in 37% of calculated trajectories for dissociation, initial relaxation leaves in the system enough energy for surmounting the excited state energy barrier for N–CO dissociation enabling the molecule to move to the MXS_{N-CO} region, followed by decay to the ground state in an average time of 470 fs. In the ground state, the two already separated fragments move further apart and the energy deposited by excitation is mostly transferred to their kinetic energy. In contrast to that, in 13% of trajectories DMF decays to the ground state via the $\mathbf{MXS_{N-CO}}$ leading to HCO and $N(CH_3)_2$ fragments, which recombine into the starting structure. In remaining 7% of trajectories deactivated to the ground state, the $N(CH_3)_2$ radical abstracts the hydrogen atom from the hot CHO yielding $NH(CH_3)_2$ and CO as the final products. Additionally, in one of two remaining atypical trajectories N–CH_3 dissociation on the S_1 PES is observed and in another one DMF switches to the ground state via a hopping point very close to the strongly pyramidalized $\mathbf{MXS_{OCN}}$ seam (Figure 3-6) and undergoes N–CO dissociation on the S_0 PES. It should be stressed that almost at all hopping points (97% of population

57% Deactivation near the crossing seam related to the N-CO bond stretching.	37% Energy deposited by excitation is transferred to the kinetic energy of the $N(CH_3)_2$ and HCO fragments.
	13% The $N(CH_3)_2$ and HCO fragments recombine to the parent molecule on the S_0 PES
	7% $N(CH_3)_2$ radical abstracts the H atom from HCO fragment
41% Molecules do not dissociate and do not switch to the S_0 state till the end of the simulation time	

Figure 3-7. Statistical analysis of different trajectory types in the excited-state dynamics of *N*, *N*-dymethylformamide starting in the S_1 state

which decays to the S_0) the N–CO bond in DMF is found to be strongly stretched (average value 2.3 Å), indicating that the N–CO dissociation process in the first excited singlet state is the major primary deactivation channel from the S_1 state to the ground state photoproducts.

It is instructive to compare these results with those described earlier for formamide. Both molecules exhibit similar types of trajectories, but their proportion is different. This holds in particular for the percentage of trajectories that remains undissociated in the S_1 state, which is by far larger in DMF than in formamide (41% in DMF vs 15% in F). There is also a difference observed in the behavior of the NR_2 (R=H, CH_3) and HCO fragments after reaching the ground state. Specifically, the fraction of trajectories characterized by recombination into the parent molecule in the ground state is larger in DMF, whereas abstraction of the H atom from HCO by NR_2 radical is more efficient in formamide. It is also interesting to mention that the fraction of trajectories which proceeds through MXS_{OCN} in formamide is larger than in DMF (7% in F vs 1% in DMF). It should be noted that this change is in line with an increase of relative energy of this MXS on passing from formamide to DMF as predicted by static calculations (Figure 3-6).

In simulations started in the second excited state all trajectories deactivate through $MXS(S_2/S_1)$ by very fast N–CO stretching. However, the largest portion of the population overshoots the $MXS_{N–CO}$ and the N–CO dissociation continues in the S_1 state leading to fully dissociated electronically excited HCO and $N(CH_3)_2$ radicals. In other words, full N–CO dissociation is achieved before reaching the ground state (Figure 3-8).

Analysis of the calculated critical points on the ground and the first excited state for HCO and $N(CH_3)_2$ radicals, allowed us also to propose a mechanism for decay of the HCO ($1^2A''$) and $N(CH_3)_2$ (1^2A_1) radicals to the ground state. This analysis showed that the HCO ($1^2A''$) species decay to the ground state via a conical intersection, which is lower in energy than the S_2/S_1 MXS and consequently it is easily accessible. On the other hand, deactivation of $N(CH_3)_2$ (1^2A_1) requires

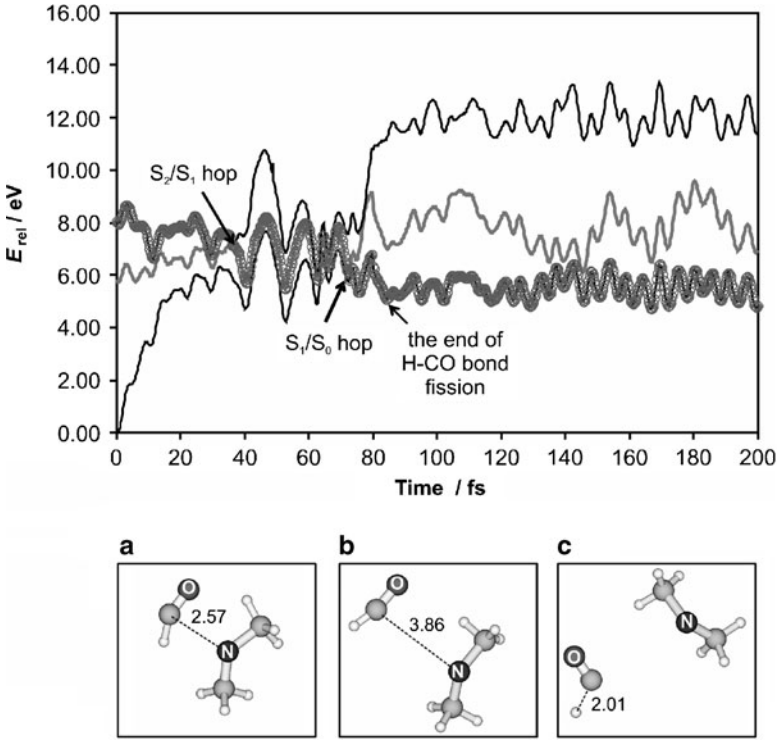


Figure 3-8. Time evolution of the potential energy of the excited states of DMF in a characteristic trajectory representing the N–CO stretching path with S_2 being the initial state. The *dots* in the potential energy graphs indicate the current state of the system. Also shown are the structures of the S_2/S_1 (**a**) and S_1/S_0 (**b**) hopping points and the structure of the end-product (**c**) in the ground state

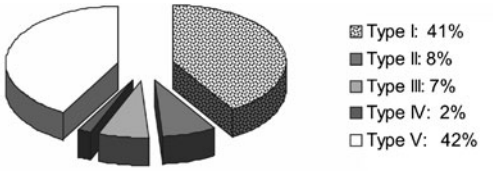
passage through the uphill MXS and its decay to the ground state is less probable, unless there is enough energy in the system to overcome the barrier.

3.4.2.2. Acetamide

The last of small amides for which dynamics simulations of photodissociation was investigated is acetamide. Here we present only preliminary results of this study. The more detailed discussion will be reported in a forthcoming paper [62]. As mentioned earlier photodissociation of acetamide has been studied experimentally in the gas phase [14] and in solution [59–61]. In both cases the C–C and the C–N bond fissions were proposed as the primary photodissociation channels. In addition direct molecular formation of $CH_3CN + H_2O$ and $CH_3NH_2 + CO$ was observed. Recently Fang and coworkers reported a comprehensive computational study of acetamide photodissociation using DFT approach (for the ground and triplet states) and the CASSCF/cc-pVDZ method followed by CASSCF/MR-CISD single point calculations (for the singlet excited states) [14]. On the basis of calculated potential

energies in the ground state and the lowest singlet excited and triplet states and surface crossing points they proposed several dissociation pathways for each of these states. As to the photodissociation processes, it was concluded that the C–C and C–N bond cleavage, as well as dehydration can take place in the T_1 state, whereas from the S_1 state molecules can relax via three nonradiative channels, besides radiation decay. They were identified as internal conversion to the ground state, internal system crossing to the T_1 state and direct dissociation in the S_1 state. They also proposed that the C–N bond dissociation at the S_1 state does not compete with the C–C dissociation due to much larger barrier.

The dynamics simulations of acetamide starting with the excitation to the S_1 state were carried out at 2000 fs time scale, whereas photodissociation from the S_2 was followed within 500 fs. The calculated lifetimes of the system in the S_1 and S_2 state are 1031 $\pm$ 3 fs and 44 $\pm$ 1fs, respectively. Comparison of these lifetimes with those calculated for F and DMF reveals that substitution by the methyl group(s) has a similar effect on the lifetime of each of the considered valence states irrespective of the site of substitution: for the S_1 state the lifetime increases and for the S_2 state it decreases. Initial excited-state motion in all trajectories starting in the S_1 state is dominated by a stretching mode of the C–N bond. Concomitantly, in all S_1/S_0 hopping points, irrespective of the dissociation mode, the C–N bond distance is found to be strongly stretched relative to the minimum energy structure (from 1.88 to 5.23 Å). Most of the activated channels during the process end up in formation of fully separated MeCO and NH_2 radicals, while in 8% of them recombination of these two fragments into the starting molecule at the S_0 is observed (Figure 3-9). Both types of trajectories were also observed in dynamic



Type I:	C-N dissociation followed by separation of the CH_3CO and NH_2 fragments
Type II:	C-N dissociation followed by recombination of the fragments
Type III:	C-N dissociation followed by additional C-CH_3 dissociation on the S_0 PES
Type IV:	atypical C-N dissociation on the S_1 PES
Type V:	no dissociation and no decay to the ground state

Figure 3-9. Statistical analysis of different trajectory types in excited-state dynamics of acetamide following excitation to the S_1 state

simulations of F and DMF, although in different proportions. This holds in particular for the first type of trajectories which in formamide and DMF where represented by 75% and 50%, respectively. This, in turn, indicates that attachment of the methyl group to the carbon atom of the carbonyl group has a more pronounced effect on the efficacy of dissociation of the central C–N bond than attachment to the amino nitrogen atom. Besides these two channels, an additional $C–CH_3$ dissociation process in MeCO fragment is observed in 7% of trajectories. Interestingly, average C–N and $C–CH_3$ bond distances in the hopping points in these trajectories are found to be 2.58 and 1.496 Å, respectively, implying that the primary process in this type of trajectories is also stretching of the C–N bond. Finally, likewise to DMF and O-protonated formamide (see Section 1.4) large fraction of trajectories (42%) do not activate till the end of simulation time.

In dynamics simulations starting from the S_2 state an even more complex dissociation pattern (Figure 3-10) is observed. Briefly, the trajectories can be divided into seven different types. They differ in mode of initial motion of the system, the bond that undergoes dissociation and the electronic state in the last time step of the simulation. Initial motion of the system can be C–N bond stretching that leads to full C–N bond dissociation in all cases or C–O bond stretching followed by delayed C–N dissociation. Second criterion for distinguishing trajectories is the existence or nonexistence of additional $C–CH_3$ dissociation. Most of the trajectories start with very fast elongation of the C–N bond and deactivate to the S_1 state in a very short period of time. In a smaller fraction initial motion is dominated by C–O bond stretching with large amplitude (C–O bond distance exceeds the values above 2 Å) followed by delayed C–N dissociation and hopping to the S_1 state. The time in which the secondary C–N dissociation starts varies from trajectory to trajectory and lies in a range of 60–300 fs. After switching to the S_1 state four different patterns are observed: (i) the molecule remains in the S_1 state till the end of the simulation time; (ii) additional dissociation of the $C–CH_3$ bond in the S_1 state occurs; (iii) hopping to the ground state followed by dissociation of the $C–CH_3$ bond; and (iv) hopping to the ground state without additional $C–CH_3$ dissociation. It should be mentioned that cleavage of the $C–CH_3$ bond on the S_0 PES is by far more abundant.

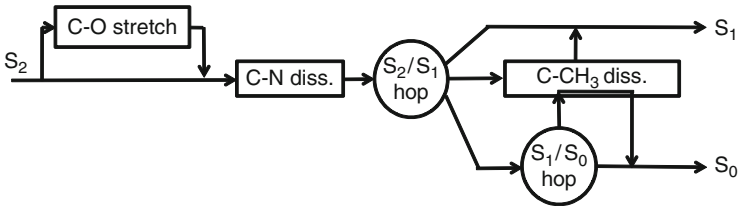


Figure 3-10. Schematic presentation of observed dissociation paths following excitation to the S_2 state of acetamide

3.4.2.3. Phenyl Substituted Formamides

Among phenyl substituted formamides calculations of PES for benzamide (BA) and formanilide (FA) have been carried out with the aim of obtaining insight into their photodissociation pathways [18, 19]. The mechanism for their initial relaxation and subsequent dissociation processes have been proposed on the basis of the calculated potential energy surfaces and their intersections [19]. The most important conclusion from this work is that their dissociation behavior is quite different from that of aliphatic amides. This holds in particular for benzamide in which the C–N bond is found to dissociate via $S_1/T_1/T_2$ three-surface intersection, irrespective of which excited state is initially populated upon irradiation. In case of formanilide this process is less probable due to high energy of the $S_1/T_1/T_2$ intersection region with respect to the S_1 minimum. Therefore, the main channel for deactivation for this molecule is internal conversion to the ground state followed by C–N fission to HCO + PhNH. More recently, Zheng and coworkers [18] reported on resonance Raman and theoretical investigations of the photodissociation dynamics of BA in the S_3 state. Based on differences in geometries in the ground state and the S_2/S_3 intersection point they concluded that BA in the S_3 state undergoes considerable vibrational distribution before getting to the conical intersection point. However, it is fair to say that full understanding of the proposed dissociation mechanism in both cases requires further investigations that would enable determination of the lifetimes of the excited molecules in the involved excited states and identification of the actual conical intersections that participate in the deactivation process. This, as illustrated in previous examples, can only be obtained by dynamics simulations.

3.5. EFFECT OF PROTONATION ON PHOTODISSOCIATION OF FORMAMIDE

Although, photodissociation of protonated formamide has not been studied experimentally as yet, it is of fundamental interest for understanding the behavior of a photoexcited peptide bond exposed to acidic conditions. From a practical point of view, results of these studies are of relevance for understanding mechanisms of photoinduced fragmentation of the peptide bond(s) in charged small polypeptides occurring in 'soft' ionization methods such as MALDI and ESI [63–67].

Formamide possesses two competing basic centers, the amino nitrogen and carbonyl oxygen atoms, with the latter being more basic in the ground, as well as in low valence excited states. The calculated gas phase energy difference between the O- and the N-protonated formamides in the ground state lies in the range of 13–22 kcal mol^{-1} depending on the level of theory (HF, B3LYP, MP, G2, MR-CISD+Q) used in calculations [68]. The basicity of the oxygen site in the ground state calculated by MR-CISD+Q/6-31-G(d) method amounts to 194.6 kcal mol^{-1} [68] as compared to the experimental value [69] of 196.5 kcal mol^{-1}. Upon excitation to the first excited valence state the difference in basicity, calculated at

the same level of theory, diminishes from 16 to 6 kcal mol^{-1} [68]. The effect of protonation on the vertical excitation energies to the low-lying excited valence states is also remarkable. Thus, in the O-protonated form the $n_O\pi^*$ valence state gets shifted to higher energy by 2.63 eV compared with the neutral molecule, whereas the excitation energy to the bright $\pi\pi^*$ state remains almost the same [68]. This in turn leads to interchanged character of the first two valence states with respect to the parent formamide. Consequently, a considerable change in mechanism of photodissociation is expected. This is, among others, reflected in considerable shortening of the time scale of photodissociation of the O-protonated formamide. The calculated lifetime of the S_1 state shortens to 33 fs. The lifetime of the S_2 state is even more reduced and amounts only 11 fs.

Starting from the FC region of the S_1 state in the O-protonated species three photodissociation pathways to MXSs with lower energy than the vertical excitation energy have been located on the PES (Figure 3-11). It is interesting to mention that for dissociation of the C–N bond one more MXS with significantly shorter C–N bond than the one shown on Figure 3-11 was found, but this dissociation channel did not become activated in dynamics simulations. A glance at Figure 3-11 shows that the lowest energy intersection MXS_{CO} corresponds to dissociation of the C–O bond. It is followed by the MXS related to dissociation of the N–H bond (MXS_{NH}), while the highest energy MXS is responsible for dissociation of the C–N bond (MXS_{CN}). All of them, in contrast to the S_1 surface in the neutral molecule, can be reached without energy barrier. However, analysis of the calculated trajectories

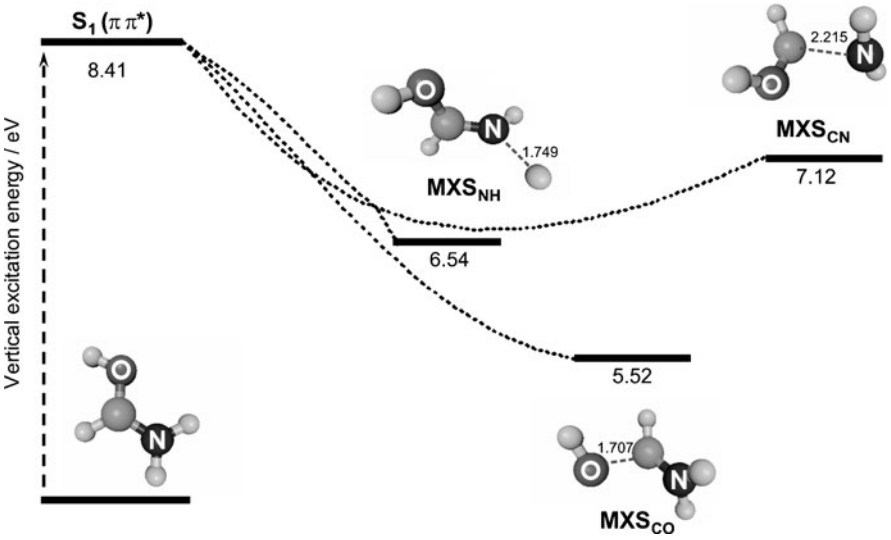


Figure 3-11. Energy profile for possible dissociation pathways of O-protonated formamide from the S_1 state observed in dynamic simulations calculated at the CASSCF level. Also shown is a schematic presentation of the S_1/S_0 MXSs for C–N, C=O and N–H bond dissociation

(Figure 3-12) revealed that only paths to the MXS_{CN} and MXS_{CO} get activated, with the former being prevalent.

Another interesting feature is that the major deactivation path occurs in the region of the highest energy MXS, presumably due to differences in their slopes on the PES after initial motion. The slope of the reaction path to the MXS_{CN} is the steepest in the FC region (Figure 3-11) and directs the course of dynamics. In other words, as the system in the FC region does not have any information about the energies of the accessible MXSs, it selects the energetically most favorable dissociation path and not the one leading to the most stable MXS. A similar effect was also observed in dynamics simulations for aminopyrimidine [70]. These examples clearly underline the importance of carrying out dynamics simulations in discussing the mechanism of non-adiabatic photodissociation processes.

The time evolution and the potential energies for characteristic trajectories representing the C–N (Figure 3-13a, b) and C–O (Figure 3-13c, d) stretching deactivation paths following excitation to the S_1 state of O-protonated formamide are shown in Figure 3-13.

In dynamics simulations starting in the S_2 state, two dissociation processes (C–O and C–N) are activated, but in contrast to the dynamics from the S_1 state, the C–O dissociation becomes the major process (in 83% of trajectories). More detailed analysis of the calculated trajectories shows that during this process S_1 and S_0 potential energy curves become degenerate allowing random redistribution of occupations between these two states. Interestingly, in 7% of trajectories C–O dissociation in the S_2 state proceeds without deactivation, yielding excited OH

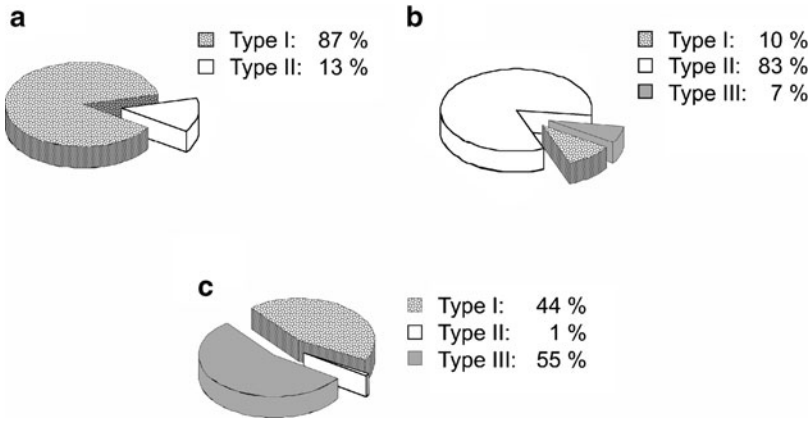


Figure 3-12. Statistical analysis of different trajectory types in excited-state dynamics starting in the S_1 (**a**) and S_2 (**b**) states of O-protonated and in the S_1 (**c**) state of the N-protonated formamide. In (**a**) and (**b**) fraction of trajectories of types I, II correspond to C–N and C–O dissociation, respectively, while in (**b**) trajectories of type III correspond to C–O dissociation without deactivation. In (**c**) trajectories of type I and II represent C–N and N–H dissociation, respectively, while trajectories of type III are undeactivated and undissociated

and $HCNH_2$ radicals. The energy profile of these trajectories is very similar to that shown in Figure 3-13d with the only difference that a switch from S_2 to S_1 does not occur when these two states get close in energy (at ca. 10 fs). With development of the C–O dissociation the energy gap between S_2 and S_1 increases disabling any hopping event.

It is particularly important to note that dynamics on the S_1 PES starting from the S_2/S_1 conical intersection differ considerably from dynamics on the S_1 PES starting in the FC region. The major difference is in the branching ratio between CN and CO dissociation pathways. Specifically, as mentioned earlier in this section, the S_1/S_0 MXS for C–N dissociation is activated in only 10% of trajectories suggesting that the momentum in the S_2/S_1 conical intersection favors the C–O reaction path on the S_1 surface. Consequently, the major photoproduct after excitation to the S_2 state is the OH radical. This is an interesting finding which indicates that dissociation of the protonated peptide bond in the S_2 state might be a generator of potentially toxic OH radical.

Finally, upon protonation on the N-atom, the vertical energy of the first excited state moves to lower energy, and the energy barrier for C–N photodissociation increases. Therefore, the N-protonated formamide decays to S_0 only partially via

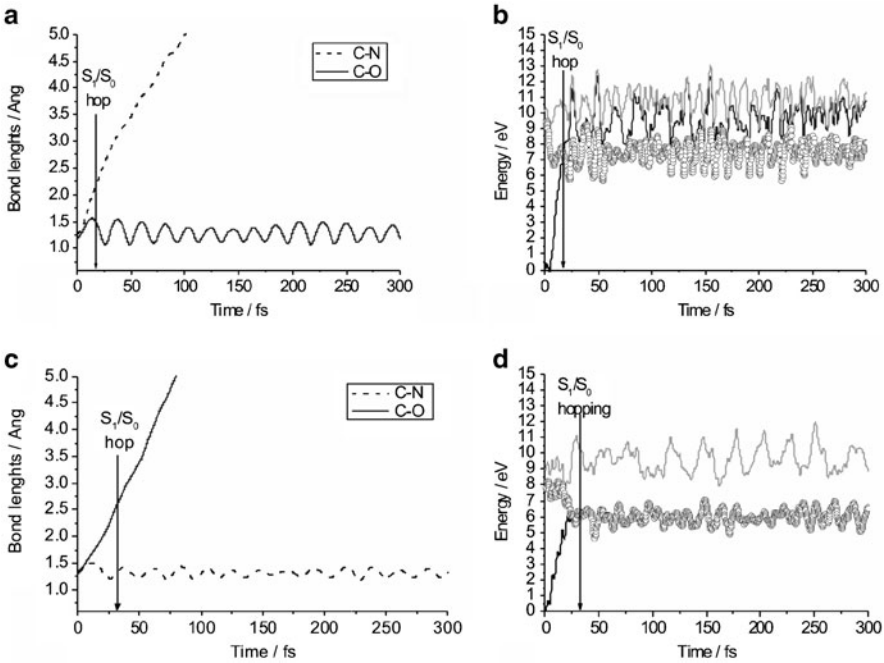


Figure 3-13. Time evolution of C–N and C–O bond distances and potential energies of S_0, S_1 and S_2 states following excitation to the S_1 state in a typical trajectory for type I (**a** and **b**) and type II (**c** and **d**) deactivation paths. The *dots* in the potential energy graph indicate the current state of the system

C–N dissociation pathway, leaving a substantial number of trajectories (55%) undeactivated and undissociated with nuclear motions oscillating within the well of the S_1 minimum.

3.6. EFFECT OF ENVIRONMENT ON PHOTODISSOCIATION OF FORMAMIDE

The photodissociation of small molecules in rare gas matrices, which can be considered as a prototype of reaction dynamics in the condensed phase, has been a subject of several recent experimental and theoretical studies [7, 10, 21, 71–76]. Photodissociation of formamide molecule is of considerable interest in this regard due to availability of experimental data for photodissociation in the gas phase [11] as well as in argon and xenon matrices [7, 10].

Lundell and coworkers studied photodissociation of formamide in the argon and xenon matrices using matrix isolation FTIR spectroscopy [7,10]. They found that excitation to the S_1 state in solid Ar matrix led to formation of a weakly bound NH_2 + CHO radical pair, followed by donation of the hydrogen from the hot HCO radical to the NH_2 radical, forming a NH_3–CO complex. On the other hand, in the Xe matrix, H_2 and HNCO were detected as the major products. The change in composition of products on going from argon to xenon was rationalized by a strong external heavy-atom effect which induced intersystem crossing to the triplet surface, yielding HNCO and H_2.

The structure of the NH_3–CO complex deserves some attention. In principle, formation of two types of complexes are possible: one with the CO attached from the carbon end to the nitrogen atom of the amino group (NH_3–CO) and the other one in which the oxygen atom of the CO group is attached to the nitrogen atom (NH_3–OC). In earlier molecular beam experiments only the former complex was detected [77]. However, Lundell and coworkers observed formation of both complexes in argon matrix with strong prevalence of the NH_3–CO complex. In a subsequent ab initio study they also located both structures. Contradicting the expectations from the experiment the oxygen-attached complex NH_3–OC was found to be slightly more stable. At the highest level of theory used in this work (CISD/cc-pVTZ) the two complexes were found to differ in energy by 0.2 kcal mol^{-1}. Furthermore, analysis of the calculated structures indicated that in both complexes the hydrogen atom is situated between two electronegative atoms in an almost linear arrangement.

Lundell and coworkers also calculated interaction energies between NH_3 and CO for both complexes, which turned to be very weak and of similar magnitude as typical van der Waals interaction energies of rare gas atoms [7]. Consequently, they concluded that formation of the higher energy species NH_3–CO as the primary photoproduct was due to energetically comparable interactions with the cage atoms. This in turn, indicates that interactions with the matrix environment, coupled with the low temperature at which experiments are carried out, play a crucial role in

determining stability of the weak complexes and strongly influence the type of complex structures observed in the matrix cage.

Recently, we carried out dynamics simulations for the photodissociation of formamide in an argon matrix at 18 K. Preliminary results for dissociation starting in the first excited singlet state showed that the 62% trajectories underwent dissociation of the C–N bond, resulting in formation of either a NH_2 + HCO radical pair (in 47% trajectories) or a NH_3–CO complex (in 15% trajectories). It is noteworthy that in the latter type of trajectories we also observed motion of the hot NH_3 and CO components resulting in both of above mentioned bonding possibilities. The cartoon showing dissociation of formamide (Figure 3-14a) in selected trajectories to NH_2 + HCO radical pair (Figure 3-14b), and formation of the NH_3–CO (Figure 3-14c) and NH_3–OC (Figure 3-14d) complexes in the argon matrix is shown in Figure 3-14.

Comparison with previously described results for the gas phase photodissociation reveals that the argon matrix has strong impact on the outcome of the dissociation process. This is illustrated by formation of the NH_3–CHO complex and prevention of separation of the NH_2 and CHO fragments in the NH_2 + HCO radical pair. The influence of the Ar matrix is also evident in elongation of the lifetime of

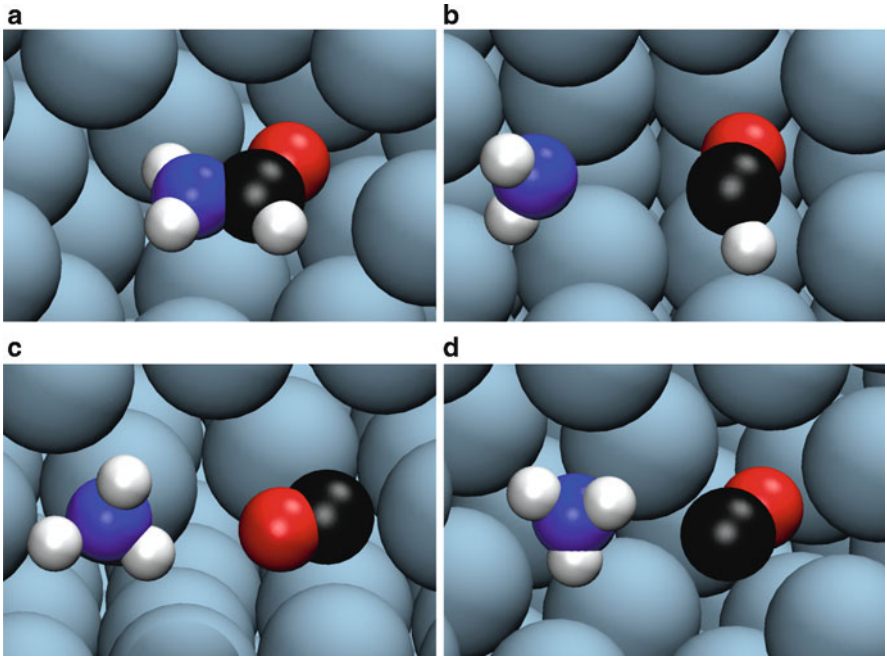


Figure 3-14. Snapshots of selected trajectories for dissociation of formamide in an argon matrix showing the starting molecule (**a**), the NH_2 + HCO radical pair (**b**), the NH_3–CO (**c**) and NH_3–OC (**d**) complexes embedded into argon matrix

the S_1 state by about 130 fs relative to the gas phase. The full account of this work will be reported in a forthcoming publication [48].

Photodissociation of formamide has been also studied in solution. For instance, Volman addressed the primary photochemistry of simple aqueous amides as early as in 1941 [59]. This was followed by microsecond flash photolysis study by Hayon and Nakashima [78]. These groups observed no transient species in the wavelength region from 250 to 700 nm, indicating very little, if any, photochemical activity of the examined amides.

More recently, Keiding and coworkers [9] reported on photolysis of formamide in water using 200 nm femtosecond pulses and following formation and relaxation of products with time-resolved absorption spectroscopy in the range from 193 to 700 nm. They observed that the majority of the excited formamide molecules (ca. 80%) convert the electronic excitation energy to vibrational excitation, which becomes dissipated to the solvent through vibrational relaxation within the first few picoseconds after excitation. A smaller fraction of the excited molecules dissociated to CHO + NH_2, of which 50% escaped recombination within experimental time scale. However, from the experimental results it could not be concluded whether the excited formamide molecules follow the same potential energy surface to dissociation, whereupon some of them retain in the cage and recombine to the ground-state structure, or if the hot formamide ground state is formed in an internal conversion process different from the path leading to dissociation of the molecule. A simple computational investigation of the involved processes has also been carried out using time-dependent wavepacket propagation on the first and second excited singlet valence state of formamide, which provided a qualitative confirmation of the observed spectral modulation of the measured absorption spectrum of formamide in water. On the calculated potential energy curves for the ground and the S_1 and S_2 states Keiding and coworkers [9] observed near crossing between these two states which suggested presence of a conical intersection calling for more advanced computational studies, including in particular dynamics simulations.

3.7. CONCLUSIONS AND FINAL REMARKS

In this chapter an overview of the computational methods for studying dynamics of fast photodissociation processes is presented. This is followed by presenting the recent results of dynamics simulations of possible deactivation processes for formamide and its methyl substituted derivatives, acetamide and *N,N*-dimethylformamide, in the lowest lying valence excited singlet states. Besides, effect of protonation of the oxygen and nitrogen atoms within peptide linkage on the mechanism of photodissociation is considered in some detail. It is shown that in all of the considered molecules the major deactivation process after excitation to the S_1 state is dissociation of the peptide C–N bond. The same holds for deactivation following excitation to the S_2 state, with exception of the O–protonated formamide

in which C–O dissociation prevails. Special attention is paid to the effect of substituents on the lifetime of the systems in the considered states and their impact on activation of possible deactivation paths. In particular, it is shown that substitution by the methyl group(s) strongly elongates the lifetime of the S_1 state, as well as percentage of trajectories that remain undissociated in the S_1 state during dynamics simulations. Relevant recent literature results for photodissociation of phenyl substituted formamides, benzamide and formanilide, are also briefly reviewed for the sake of completeness.

In the last section of the chapter, photodissociation of formamide in argon matrix in low lying valence excited singlet states using hybrid nonadiabatic photodynamics QM/MM method is discussed. It was shown that embedding of the formamide molecule in the argon matrix has strong impact on the outcome of the photodynamical process. This is illustrated by formation of the NH_3–CO complex and prevention of separation of the NH_2 and CHO fragments in the NH_2 + HCO radical pair. This is the first example of exploring photodissociation of an organic molecule using this approach and it nicely illustrates the importance of considering environmental effects on photodissociation behavior of the peptide bond. Therefore, it would be of considerable interest to extend these computational studies to the solvents, where electrostatic effects could play a significant role. Previous experimental studies of photodissociation in water indicated that the solvent influences the relaxation of the reaction products and transition states to such extent that the outcome of the photochemical reaction is largely determined within the first contact between photoexcited solute molecule and the solvent. In other words, the solvent, and in particular water, is so efficient in absorbing the excess energy of the reaction that very little energy remains for further reactions after the first few hundredths of femtoseconds. Formamide is a particularly suitable model compound for such studies due to availability of reliable experimental results for dissociation in water.

ACKNOWLEDGEMENTS

This work has been supported by the Ministry of Science, Education and Sport of Croatia through the project. No. 098-0982933-2920 (Mirjana Eckert-Maksić, Ivana Antol, and Mario Vazdar) and by the Austrian Science Fund within the framework of the Special Research Program F16 (Advanced Light Sources) and Project P18411-N19 (Hans Lischka and Mario Barbatti). H.L. acknowledges support by the grant from the Ministry of Education of the Czech Republic (Center for Biomolecules and Complex Molecular Systems, LC512) and by the Praemium Academiae of the Academy of Sciences of the Czech Republic, awarded to Pavel Hobza in 2007 and by the research project Z40550506 of the Institute of Organic Chemistry and Biochemistry of the Academy of Sciences of the Czech Republic. The support by the COST D37 action, WG0001-06 and WTZ treaty between Austria and Croatia (Project No. HR17/2008) are also acknowledged. The authors

especially acknowledge the technical support and computer time at the Linux PC cluster Schrödinger III of the computer center of the University of Vienna. Zagreb group (Mirjana Eckert-Maksić, Ivana Antol, and Mario Vazdar) also thanks allocation of computation time on the ISABELLA cluster at the University Computing Center in Zagreb.

REFERENCES

1. Greenberg A, Breneman CM, Liebman JF (2002) The amide linkage: structural significance in chemistry, biochemistry and materials science. Wiley, New York
2. Boyd DB (1993) J Med Chem 36:1443–1449
3. Jungheim LN, Boyd DB, Indelicato JM, Pasini CE, Preston DE, Alborn WE Jr (1991) J Med Chem 34:1732–1739
4. Challender RH, Dyer RB, Gilmanshin R, Woodruff WH (1998) Ann Rev Phys Chem 49:173–202
5. Nikogosyan DN, Görner HJ (1998) J Photochem Photobiol B 47:63–67
6. Sionkowska A, Wisniewski M, Skopinska J, Mantovani D (2006) Int J Photoenergy. Article ID 29196
7. Lundell J, Krajewska M, Räsänen M (1998) J Mol Struct 448:221–230
8. Kang TY, Kim HL (2006) Chem Phys Lett 431:24–27
9. Petersen C, Dahl NH, Jensen SK, Poulsen JA, Thogersen J, Keiding SR (2008) J Phys Chem A 112:3339–3344
10. Lundell J, Krajewska M, Räsänen M (1998) J Phys Chem A 102:6643–6650
11. Boden JC, Back RA (1970) Trans Faraday Soc 66:175–182
12. Liu D, Fang W-H, Fu X-Y (1999) Chem Phys Lett 318:291–297
13. Liu D, Fang W, Lin Z, Fu X (2002) J Chem Phys 117:9241–9247
14. Chen X-B, Fang W-H, Fang D-C (2003) J Am Chem Soc 125:9689–9698
15. Antol I, Eckert-Maksić M, Barbatti M, Lischka H (2007) J Chem Phys 127:234303
16. Antol I, Vazdar M, Barbatti M, Eckert-Maksić M (2008) Chem Phys 349:308–318
17. Eckert-Maksić M, Antol I (2009) J Phys Chem A. doi:10.1021/jp9046177
18. Pei K-M, Ma Y, Zheng X (2008) J Chem Phys 128:224310
19. Chen X-B, Fang W-H (2004) J Am Chem Soc 126:8976–8980
20. Ruckenbauer M, Barbatti M, Müller T, Lischka H J Phys Chem A. doi: 10.1021/jp103101t
21. For a recent review see e.g. Bargheer M, Borowski A, Cohen A, Fushitani M, Gerber RB, Gühr M, Hamm P, Ibrahim H, Kiljunen T, Korolkov MV, Kühn O, Manz J, Schmidt B, Schröder M, Schwentner N (2007). In: Kühn O, Wöste L (eds) Analysis and control of ultrafast photoinduced reactions. Springer Series in Chemical Physics, vol 87. Springer, Heidelberg, pp 257–385
22. Schinke R (1995) Photodissociation dynamics: spectroscopy and fragmentation of small polyatomic molecules. Cambridge University Press, Cambridge
23. Tully JC (1998) Faraday Discuss 110:407–419
24. Tully JC (1990) J Chem Phys 93:1061–1071
25. Ben-Nun M, Quenneville J, Martínez TJ (2000) J Phys Chem A 104:5161–5175
26. Li XS, Tully JC, Schlegel HB, Frisch MJ (2005) J Chem Phys 123:084106
27. Barbatti M, Granucci G, Persico M, Ruckenbauer M, Vazdar M, Eckert-Maksić M, Lischka H (2006) J Photochem Photobiol A: Chem 190:228–240
28. Barbatti M, Sellner B, Aquino AJA, Lischka H (2008) In: Shukla MK, Leszczynski J (eds) Radiation induced molecular phenomena in nucleic acids. Springer Science+Business Media, The Netherlands, pp 209–235

29. Swope WC, Andersen HC, Berens P, Wilson KR (1982) J Chem Phys 76:637–649
30. Hayashi S, Taikhorshid E, Schulten K (2009) Biophys J 96:403–416
31. Lasser C, Swart T (2008) J Chem Phys 129:034302
32. Fabiano E, Groenhof G, Thiel W (2008) Chem Phys 351:111–116
33. Pittner J, Lischka H, Barbatti M (2009) Chem Phys 356:147–152
34. Butcher J (1965) J Assoc Comput Mach 12:124–135
35. Granucci G, Persico M (2007) J Chem Phys 126:134114
36. Muller U, Stock G (1997) J Chem Phys 107:6230–6245
37. Jasper AW, Stechmann SN, Truhlar DG (2002) J Chem Phys 116:5424–5431
38. Sellner B, Barbatti M, Lischka H (2009) J Chem Phys 131:024312
39. Shepard R, Lischka H, Szalay PG, Kovar T, Ernzerhof M (1992) J Chem Phys 96:2085–2098
40. Lischka H, Dallos M, Shepard R (2002) Mol Phys 100:1647–1658
41. Lischka H, Dallos M, Szalay PG, Yarkony DR, Shepard R (2004) J Chem Phys 120:7322–7329
42. Dallos M, Lischka H, Shepard R, Yarkony DR, Szalay PG (2004) J Chem Phys 120:7330–7339
43. Angeli C (2009) J Comput Chem 30:1319–1333
44. Lin H, Truhlar DG (2007) Theor Chem Acc 117:185–199
45. Bunge A (1970) J Chem Phys 53:20–28
46. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) J Am Chem Soc 118:11225–11236
47. Breneman CM, Wiberg KB (1990) J Comput Chem 11:361–373
48. Eckert-Maksić M, Vazdar M, Ruckenbauer M, Barbatti M, Müller T, Lischka H (2010) Phys Chem Chem Phys (submitted)
49. Lischka H, Shepard R, Brown FB, Shavitt I (1981) Int J Quant Chem S15:91–100
50. Lischka H, Shepard R, Pitzer RM, Shavitt I, Dallos M, Müller T, Szalay PG, Seth M, Kedziora GS, Yabushita S, Zhang ZY (2001) Phys Chem Chem Phys 3:664–673
51. Lischka H, Shepard R, Shavitt I, Pitzer RM, Dallos M, Müller T, Szalay PG, Brown FB, Ahlrichs R, Boehm HJ, Chang A, Comeau DC, Gdanitz R, Dachsel H, Ehrhardt C, Ernzerhof M, Hoechtl P, Irle S, Kedziora G, Kovar T, Parasuk V, Pepper MJM, Scharf P, Schiffer H, Schindler M, Schueler M, Seth M, Stahlberg EA, Zhao J-G, Yabushita S, Zhang Z, Barbatti M, Matsika S, Schuurmann M, Yarkony DR, Brozell SR, Beck EV, Blaudeau J-P (2006) COLUMBUS: an ab initio electronic structure program, release 5.9.1. www.univie.ac.at/columbus. Accessed date June 11 2010
52. Ren PY, Ponder JW (2003) J Phys Chem B 107:5933–5947
53. Barbatti M, Granucci G, Lischka H, Ruckenbauer M, Persico M (2007) NEWTON-X: a package for Newtonian dynamics close to the crossing seam, version 1.1.0a. www.newtonx.org. Accessed date June 11 2010
54. Vazdar M, Eckert-Maksić M (work in progress)
55. Barbatti M, Aquino AJA, Lischka H (2006) Mol Phys 104:1053–1060
56. Forde NR, Butler LJ, Abrash SA (1999) J Chem Phys 110:8954–8968
57. Forde NR, Myers TL, Butler LJ (1997) Faraday Discuss 108:221–242
58. Spall BC, Steacie EWR (1952) Proc R Soc Lond A239:1–15
59. Volman DH (1941) J Am Chem Soc 63:2000
60. Bosco SR, Cirillo A, Timmons RB (1969) J Am Chem Soc 91:3140–3143
61. Booth GH, Norrish RGW (1952) J Chem Soc 188–198
62. Antol I, Vazdar M, Eckert-Maksić M, Lischka H (2009) (manuscript in preparation)
63. Thompson MS, Cui W, Reilly JP (2007) J Am Soc Mass Spectrom 18:1439–1452
64. Cui W, Thompson MS, Reilly JP (2005) J Am Soc Mass Spectrom 16:1384–1398
65. Jeong HM, Young SS, Hyun JC, Myung SK (2007) Rapid Commun Mass Spectrom 21:359–368
66. Yoon SH, Kim MS (2007) J Am Soc Mass Spectrom 18:1729–1739

67. Grégoire G, Dedonder-Lardeux C, Jouvet C, Desfrancois C, Fayeton JA (2007) Phys Chem Chem Phys 9:78–82
68. Antol I, Eckert-Maksić M, Lischka H (2004) J Phys Chem A 108:10317–10325 and references cited therein
69. Hunter EP, Lias SG (1998) J Phys Chem Ref Data 27:413
70. Barbatti M, Lischka H (2007) J Phys Chem A 111:2852–2858
71. Macas EMS, Khriachtchev L, Petersson M, Fausto R, Räsänen M (2004) Vib Spec 34:73–82
72. Macas EMS, Khriachtchev L, Petersson M, Fausto R, Lundell J, Räsänen M (2005) J Phys Chem A 109:3617–3625
73. Duvernay F, Trivella A, Brget F, Coussan S, Aycard J-P, Chiavassa T (2005) J Phys Chem A 109:11155–11162
74. Vaskonen KJ, Kunttu HM (2003) J Phys Chem A 107:5881–5886
75. Sałdyka M, Mielke Z (2003) Phys Chem Chem Phys 5:4790–4797
76. Chen M, Yang R, Ma R, Zhou M (2008) J Phys Chem A 112:7157–7161
77. Fraser GT, Nelson DD Jr, Peterson KI, Klemperer W (1986) Chem Phys 84:2472–2480
78. Hayon E, Nakashima M (1971) J Phys Chem 75:1910–1914

CHAPTER 4

DESIGN OF CATALYSTS FOR ASYMMETRIC ORGANIC REACTIONS THROUGH DENSITY FUNCTIONAL CALCULATIONS

C.B. SHINISHA, DEEPA JANARDANAN, AND RAGHAVAN B. SUNOJ
Department of Chemistry, Indian Institute of Technology Bombay, Powai, Mumbai 400076, India

Abstract: The current decade being a golden era in the history of organocatalysis, designing new organocatalysts for synthetically valuable reactions is of high importance. A fine blend of theoretical techniques and knowledge gathered from the experimental observations can help one design highly selective organocatalysts. The present chapter summarizes our efforts in designing organocatalysts for two synthetically important reactions; namely, the aldol reaction and sulfur ylide mediated ring formation reactions. In order to identify the crucial elements that affect the stereoselection process, detailed mechanistic studies are performed initially. Thus, factors controlling the vital energy differences between the diastereomeric transition states are identified and rationalized. Later on, insights from these model studies are utilized toward designing the new catalyst framework. In the last stage, the catalytic efficiency with the designed catalysts is evaluated for selected reactions. Conformationally constrained catalysts designed in this manner are predicted to be more effective with improved selectivities in comparison to the experimentally employed analogues.

Keywords: Density functional theory, Organocatalysis, Transition state, Stereoselectivity, Catalyst design, Aldol reaction, Sulfur ylide mediated reactions

4.1. INTRODUCTION

The computational sciences have long been recognized as one of the key components of enabling technologies. The present decade has witnessed an unprecedented growth of computational chemistry as one of the most efficient tools toward understanding chemical reactivity. Equally important is the increasing attention on the power of these methods as a predictive tool. The modern quantum chemical methods can offer reliable levels of accuracy for small molecules and are even

P. Paneth and A. Dybala-Defratyka (eds.), *Kinetics and Dynamics*,
Challenges and Advances in Computational Chemistry and Physics 12,
DOI 10.1007/978-90-481-3034-4_4,

capable of handling larger molecular systems of the order of thousands of atoms, with varying degrees of approximation.

An interesting subset of the predictive applications of computational quantum chemical methods includes studies on chemical reactivity problems. The electronic structure calculations can offer insights on the mechanism and fundamental stereoelectronic factors responsible for stereoselectivity in organic reactions. This approach can indeed enable the development of novel catalysts with improved catalytic efficiency. Theoretical understanding of reactions at the molecular level can serve as an effective complementary knowledge to the vast amount of experimental literature. However, interesting examples are also available wherein highly reactive and unstable species that may elude the experimental observations, have been studied using electronic structure calculations. In addition, new concepts as well as molecules exhibiting non-conventional structures have been identified, initially through computational methods, and subsequently formed the subject of successful experimental verification [1].

The quantum mechanical methods have also been successful in designing catalysts for number of interesting reactions [2]. Recently, Houk and Cheong highlighted the role of quantum chemical approaches toward the prediction of small-molecule catalysts for carbon–carbon bond formation reactions [3]. One of the primary steps toward catalyst design consists of establishing the mechanistic insights of the catalyzed reaction. The mechanistic details on the catalyzed reaction pathway will help propose intuitive modifications on the catalyst. Computational approaches will be able to offer some preliminary information on the potential and effectiveness of the catalyst. The electronic structure features of the transition state exert the critical controlling factor in catalysis. The identification of optimal interactions between charged (electrostatic) or neutral (non-electrostatic effects) moieties in a transition state are therefore highly important. Rational modifications of weak interactions operating in the transition state can help modulate the properties of the catalyst and thus could lead to improved reactivity or/and selectivity of chemical reactions. The aforementioned method has an underlying parallelism with the virtual screening protocols adopted in drug design using computational methods [4, 5]. Herein we present our efforts in designing catalysts for asymmetric applications with the help of two important reactions. The aims of the present chapter are twofold: (a) to understand the intricacies associated with the stereoselection process of two popular chemical reactions (an organocatalyzed aldol reaction and a sulfur-ylide promoted aziridination), and (b) computational design of catalysts that could offer improved enantio- and diastereoselectivities for the above reactions.

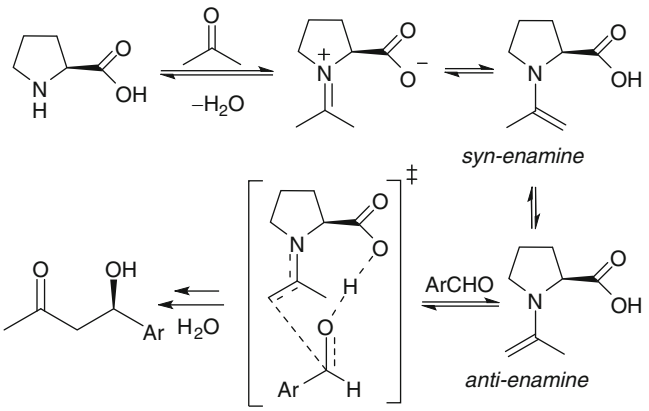
4.1.1. Organocatalytic Reactions and Theoretical Models

Over the last few years, the field of organocatalysis has witnessed tremendous growth as an effective protocol for asymmetric synthesis. In the recent times, the use of small organic molecules as asymmetric catalysts has emerged as an

alternative to traditional metal-based catalysis [6, 7]. The synthesis of several complex chiral molecules with higher levels of selectivity have been achieved through asymmetric organocatalysis [8, 9]. The development of efficient organocatalysts and strategies such as cascade reactions [10], has made possible the synthesis of complex natural products with high selectivity [11]. Proline has evolved as a prototypical example of organocatalyst with capabilities for a diverse range of asymmetric transformations [12]. In the last few years, many proline analogues are found as capable of catalyzing many asymmetric reactions with high selectivity [13]. Both experimental and theoretical studies have been used in evaluating the mechanisms of organocatalytic reactions. The concurrence between the predicted and experimentally determined enantiomeric excess has generally been quite impressive. Some such examples include the studies on proline catalyzed aminoxylation [14], Mannich reaction [15], and α-alkylations [16]. Several notable successes of computational methods in gaining valuable insights on the mechanism and stereoselectivity [17], evidently highlight the likely potential of such methods toward the design of asymmetric catalysts [2, 3, 18].

4.1.1.1. Proline Catalyzed Intermolecular Aldol Reaction

In early 2000s, List et al. reported the first proline catalyzed intermolecular aldol reaction between acetone and *p*-nitrobenzaldehyde [19]. The latest experimental results, in concert with the density functional theory studies, support the proposed mechanism as described in Scheme 4-1 [20, 21]. Initially, acetone on reacting with the catalyst forms an enamine. Depending on the orientation of the enamine double bond with respect to the carboxylic acid group *syn* or *anti* conformers are possible. Through a series of theoretical and experimental studies on proline-catalyzed intermolecular aldol reactions, the stereoselectivity determining step is proposed to involve the nucleophilic addition of the enamine to the electrophilic aldehyde



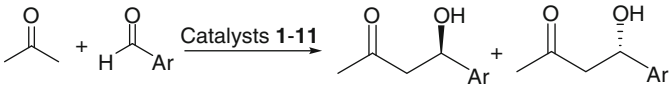
Scheme 4-1. Proline catalyzed intermolecular aldol reaction and Houk transition state model for stereoselective C–C bond formation

and simultaneous proton transfer from the carboxylic acid group to the developing alkoxide ion (Scheme 4-1) [21]. The hydrolysis of the resulting adduct, furnishes the desired β-hydroxy ketone with defined stereochemistry, as the product. The product ratios and the enantiomeric excess calculated using this type of transition state model has been found to be in good agreement with the experimental observation.

4.1.1.2. Modified Proline Analogues as Catalysts

With the successful use of proline as a catalyst for a number of asymmetric reactions, increasing number of proline analogues began to appear in literature as potential organocatalysts. The major modifications on the proline framework include conversion/replacement of carboxylic acid group as well as elaborations at the β and γ positions [22, 23]. All such attempts thus far have revolved around the monocyclic proline analogues, except for a couple of reports on 4,5-methanoprolines [24]. The pyrrolidine ring conformation in proline is believed to be an important factor in chirality transfer [25], which could play a vital role in organizing the stereoselectivity determining transition state. Pyrrolidine ring is also known to be best suited for aldol reactions as compared to other cyclic secondary amines such as pipecolic acid and 2-azetidinecarboxylic acid [26]. Primary amino acids have also been studied as catalysts for aldol reactions, where good enantiomeric excess was observed only when the substrates employed were cyclic ketones with reduced conformational flexibility [27]. Proline can exist in different interconvertible conformations by virtue of its flexible puckered pyrrolidine ring [28]. Therefore, it is logical to anticipate that restricting the conformational freedom of the pyrrolidine ring could have a direct impact on the stereoselectivity of proline-catalyzed reactions. Improved rigidity of the catalyst could impart improved stability and organization of the transition states and help minimize the entropy losses. Moreover, detailed knowledge on the controlling elements such as structure, conformation and energetics of the catalyst, substrate and the key transitions states will be valuable towards the design of new catalysts for asymmetric applications [29] (Scheme 4-2).

We reasoned that introducing geometrical constraints on the catalyst can lead to relatively ordered transition states capable of influencing the stereochemical outcome of the reaction. In this chapter, we summarize how we designed a series of novel bicyclic-bifunctional analogues of proline. These catalysts are evaluated for their ability to promote stereoselective aldol reaction between acetone and *p*-nitrobenzaldehyde, by using the density functional theory methods. One of the major questions at this juncture relates to the ability of the density functional theory



Scheme 4-2. General scheme of model aldol reaction investigated in the present work

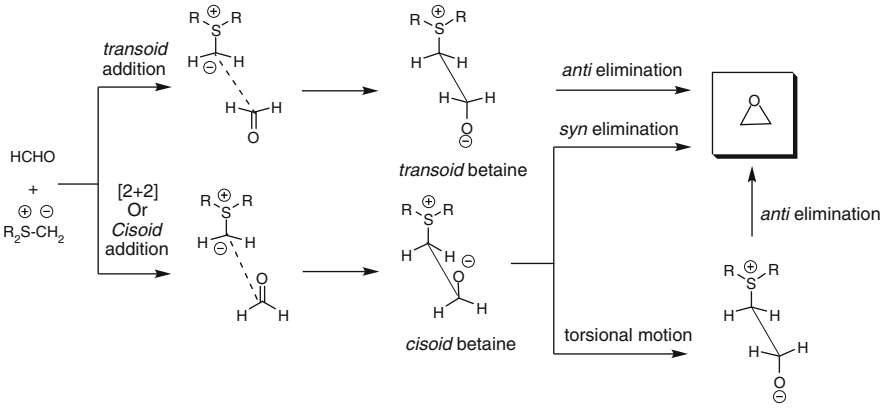
methods in accurately capturing the essential electronic structure features of the transition states that decide the stereochemical outcome of organic reactions. More importantly, how good are the relative energies calculated using the density functional theories for the diastereoselective transition states that ultimately exerts a direct influence on the stereochemical outcome of the reaction. A closer perusal of the available literature reveals several interesting applications of the density functional theory toward obtaining useful insights on reaction mechanisms as well as stereoselectivity [30]. We believe that the predictive potential of theoretical models can be further exploited in accelerating this burgeoning area of chemical research. Our DFT based approach towards designing new organocatalysts is presented as an illustration of this approach [31]. In view of the recent debate on the merits and limitations of predicting new molecules by using computational methods [32], we would like to stress that the proposed catalysts are either synthetically available or structurally very similar to the known compounds.

The design concepts for two important carbon–carbon formation reactions are described in this chapter. The first one relates to the design of a sulfur-ylide catalyst for asymmetric aziridination reaction while the later pertains to the in-silico design of proline analogues for asymmetric aldol reaction.

4.1.2. Sulfur Ylide Promoted Reactions

Sulfur ylide promoted ring formation [33] is an efficient strategy in the construction of three-membered carbo/heterocyclic compounds of synthetic [34] as well as biological [35] importance. This methodology has witnessed revitalized progress in the recent past, owing to the developments of catalytic methods towards the generation of chiral sulfur reagents [36]. Pioneering works from the research groups of Aggarwal [37], Dai [38], Solladie-Cavallo [39], and Metzner [40] deserve special mention in terms of popularizing the asymmetric version of S-ylide mediated reactions. The remarkable recent success with this protocol obviously demands better understanding of the stereoelectronic effects that control the stereochemical outcome of such reactions. Such insights can contribute to the development of new catalysts which might offer improved selectivity.

Despite its synthetic popularity, mechanistic details on sulfur ylide promoted ring formation reactions have only been scarcely found in the literature [41]. A few reports on diastereoselection and mechanism of selected achiral substrates have lately appeared [42], although, most of these reports focus primarily on the successful application of Corey–Chaykovsky epoxidation reaction (Scheme 4-3) [33b]. For instance, a model computational study on the reaction between formaldehyde and sulfonium/phosphonium ylides ascertained that the kinetic preference of Corey–Chaykovsky and Wittig models were controlled by the leaving group ability of $–SH_2$ group in the former and ring opening of the cyclic intermediate in the latter [41a]. In another report, Lindvall and Koskinen examined the concerted, end-on, and torsional pathways for the reaction between dimethyl sulfonium methylide and formaldehyde and proposed the feasibility of an end-on addition between the



Scheme 4-3. The general mechanism of Corey–Chaykovsky reaction

reactants [41b]. However, none of the above studies provide sufficient molecular insights on the selectivity or mechanistic details owing to the absence of the analysis of possible reaction pathways.

Later, in a remarkable work, Aggarwal et al. revisited the mechanistic issues with the inclusion of solvation effects as well as with real-life substrates [42a]. Interestingly, their findings were at variance to the previous reports, and it was shown that the favorable pathway was a quasi [2 + 2] addition of ylide to aldehyde rather than an end-on (*transoid*) addition. It was found that torsional motion from *cisoid* to *transoid* betaine controls the diastereoselection in the sulfur ylide mediated epoxidation reactions [41a]. Further, the authors could satisfactorily reproduce the experimental observation that the formation *anti* and *syn* betaine is reversible.

Although mechanistic rationale on sulfur ylide promoted epoxidation reaction was established, such insights on the generation of cyclopropanes or aziridines through a similar protocol continued to remain as lacunae. It is conspicuous that the difference in the nature of the electrophiles capable of furnishing aziridines or cyclopropanes could offer different mechanistic features than what was earlier noticed for epoxidation reaction. The introduction of a third ring substituent in aziridination can further influence the relative orientations of the substituents around the newly forming C_2–C_3 bond as these carbo/heterocycles usually prefers a *trans* disposition among the adjacent substituents (vide infra).

A far more important, yet unsolved issue relates to chiral induction in S-ylide promoted reactions. It is known that optically pure sulfides can act as the source of chirality in asymmetric version of these reactions. However, the exact mechanism of chirality transfer is not established. Although qualitative proposals based on ylide conformation are available, those models fail to offer a convincing depiction of how the observed enantioselectivities come about. The major limitation pervading such models using achiral ylides arises to the assumption that the initial addition between ylide and imine controls the stereochemical outcome of the reaction.

This very assumption may not always be the rate-limiting step, as described in the later sections of this chapter. Thus, detailed investigations are necessary in order to improve our understanding on the stereoselection process, which can as well help contribute the design of improved catalyst or chiral auxiliary. Keeping these in mind, the present chapter also summarizes our efforts toward understanding the mechanistic and selectivity aspects of sulfur ylide mediated ring formation reactions as well as to utilize these insights thus obtained toward the design of improved molecules as catalysts.

4.2. COMPUTATIONAL METHODS

All calculations were done using Gaussian suite of quantum chemical programs [43]. The geometry of reactants, intermediates, and transition states were first optimized. The characterization of the nature of the stationary points were carried out by evaluating the corresponding Hessian indices at the B3LYP/6-31G* or other level of theory as applicable [44, 45]. Careful verification of the unique imaginary frequencies for the transition states has been carried out to check whether the frequency indeed pertains to the desired reaction coordinate. Further, intrinsic reaction coordinate (IRC) calculations were carried out to authenticate most of the transition states [46]. Cases where enthalpies are calculated were obtained by adding the scaled ZPVE (0.9806) [47] and thermal contributions to the 'bottom-of-the-well' energy values. The single-point energies were then calculated using a more flexible triple zeta quality basis set, namely the 6-311+G** with the continuum solvation model using the SCRF-CPCM method [48], using the United-Atom Kohn-Sham (UAKS) radii. DMSO was used as the continuum solvent dielectric ($\varepsilon = 46.7$) for the intermolecular aldol reaction. For sulfur ylide reactions, the single-point solvent energies were obtained at the B3LYP/6-311+G** or B3LYP/6-311G** levels using the SCRF-PCM method in acetonitrile continuum. All the SCRF calculations were performed with the default options as implemented in Gaussian03. Full geometry optimizations within the continuum solvation model might lead to changes in geometries and energetics. Unfortunately, such calculations are computationally expensive on relatively larger molecules considered here at the levels of theory employed herein. It is to be further noted that the focus of this chapter is on the relative energies of the diastereomeric transition states than on the absolute activation parameters. One can therefore anticipate that the computed values are sufficiently reliable to be able to draw meaningful conclusions. Further, enantiomeric excess is calculated by assuming that the Boltzmann distribution of transition states leading to each enantiomer give the final ratio between the enantiomers. The %ee can be computed using the following equation (4-1).

$$\%ee = \frac{1 - e^{\frac{-\Delta G^{\ddagger}_{R/S}}{RT}}}{1 + e^{\frac{-\Delta G^{\ddagger}_{R/S}}{RT}}} \times 100 \tag{4-1}$$

Here enantiomeric excess is computed by assuming that the final R/S ratio is given, at a certain temperature, by the Boltzmann distribution of the transition states leading to each enantiomer. In the equation given above, $\Delta G^{\ddagger}_{R/S}$ refers to the free energy difference between the pro-*R* and pro-*S* transition states. In our studies, we found that the incorporation of free energy corrections (from the corresponding geometry optimizations and frequency calculations) into the solvent phase single-point energies (free energy of solvation G_s, denoted as *E* in the text) does not alter the trends predicted from the relative energies (*E*) as such.

4.2.1. Terminology

4.2.1.1. Bicyclic Proline Analogues as Catalyst for Intermolecular Aldol Reaction

The enamines derived from acetone by the secondary amine catalyst can either exist as *syn*-enamine or *anti*-enamine depending on how the enamine double bond is positioned with respect to the carboxylic acid group. In the present study, *re* and *si* facial attack on *p*-nitrobenzaldehye by *anti* and *syn* enamines are considered. The transition state corresponding to the attack of *anti*-enamine on the *re* and *si*-face of the aldehyde is respectively denoted as a-*re* and a-*si* and that of *syn*-enamine is referred as s-*re* and s-*si*.

4.2.1.2. Design of Catalyst for Sulfur Ylide Mediated Reaction

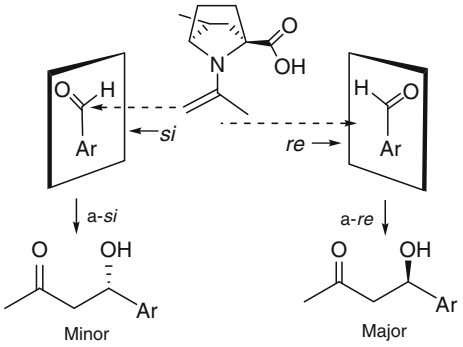
For achiral model studies, total of five ylides were considered (Scheme 4-5, **12**–**16**). For the C–C bond-forming step, both *cisoid* and *transoid* addition transition states represented as TS-**n***c* and TS-**n***t* for ylide **n** (**n** = **12**–**16**) are examined. The intermediate betaines are termed as I-**n***c* for *cisoid* and I-**n***t* for *transoid* orientations (Scheme 4-6). Similarly, the transition state for *cisoid* to *transoid* rotation is designated as TS-**n**R and that for elimination as TS-**n**E. In the alternative (*re*,*re*) approach, the ylides, intermediates as well as transition states are designated using **13**′ through **16**′.

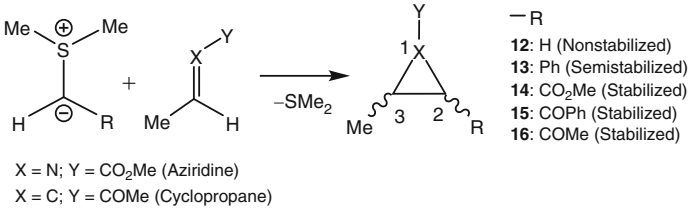
4.3. RESULTS AND DISCUSSION

4.3.1. Intermolecular Aldol Reaction

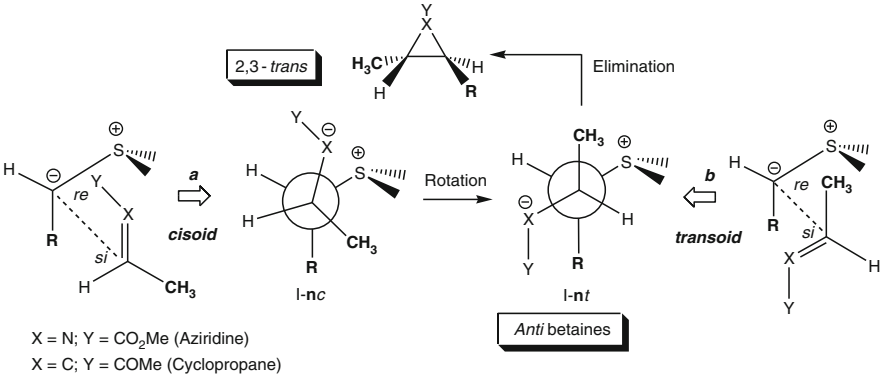
The stereoselectivity determining C–C bond formation step in the aldol reaction between acetone and *p*-nitrobenzaldehyde has been modeled with all the proposed catalysts. This step involves the addition of enamine formed between acetone and the catalyst (secondary amine) to the electrophilic aldehyde. The stereoselectivity of the overall reaction depends on the preferred stereochemical modes of approach between the enamine and the aldehyde. Key stereochemical modes of addition involving *anti*-enamine are depicted in Scheme 4-4. Possible transition states are

Scheme 4-4. Different possible stereochemical modes of additions for anti-enamine (derived from catalyst **2** and acetone) with electrophile (ArCHO, Ar → $PhNO_2$)





Scheme 4-5. Model aziridination and cyclopropanation reactions



Scheme 4-6. Various steps in the reaction between S-ylide and acceptors leading to the formation of carbo/heterocycles

located and the enantiomeric excess is then calculated from the relative energies between the TSs leading to possible enantiomers.

To begin with, we have modeled the stereoselectivity determining step for the proline catalyzed aldol reaction between acetone and *p*-nitrobenzaldehyde. The transition states corresponding to the different modes of addition of enamine

formed between acetone and proline to *p*-nitrobenzaldehyde are located. The computed value of enantiomeric excess 75% is found to be in excellent agreement with the experimentally observed value 76% for the proline catalyzed intermolecular aldol reaction with the same set of substrates [25]. Such an impressive agreement between the computed and the experimentally observed enantioselectivity for the proline-catalyzed aldol reaction provides considerable support to the Houk transition state model for proline-catalyzed asymmetric aldol reaction that operates through an enamine mechanism. Inspired by the successful application of the DFT method with the parent proline, we have considered several proline analogues as summarized in the following sections. The designed bicyclic catalysts are grouped into smaller sets depending on the position of the amino group.

4.3.1.1. Bicyclic Systems Where –NH Is Out-of-Plane in the Envelope Conformer of Proline

The catalytic ability of bifunctional amino acids, when tailored onto a rigid bicyclic framework, in aldol reaction is studied in detail. In the first set of catalysts, an envelope conformer of proline with an out-of-plane –NH is considered in a [2.2.1] or [2.1.1] bicyclic framework (Figure 4-1). The synthesis and other applications of these proposed azabicyclic compounds are known [49]. Some of these azabicyclics have been reported to be useful as peptidomimetics since they are capable of inducing conformational changes in peptides, which could be useful in studying receptor recognition [50].

The pyrrolidine envelope conformer with an out-of-plane –NH group, when inscribed on a bicyclic framework as in **1** and **2**, the inherent chirality will apparently be lost due to a plane of symmetry. While restricted conformers, such as that in a [2.2.1] bicyclic system, could have significant nitrogen inversion barriers ($\Delta G_{Expt} = 13.8$ kcal/mol), the molecule will be achiral under most practical conditions [51]. Such high barrier probably arises due to the repulsion between nitrogen lone pairs and bonding electrons of the two carbon–carbon bridges and the lack of flexibility for the CNC bond angle. To impart inherent chirality to these catalysts and also as a preliminary step towards investigating the effect of substituents, we have introduced methyl groups at suitable positions on the bicyclic system as shown (**1** and **2**). The azanorbornyl systems bearing nitrogen at the seventh position are known to be highly pyramidal around the nitrogen atom [52]. The orientation of the carboxylic acid group at the bridge-head position is

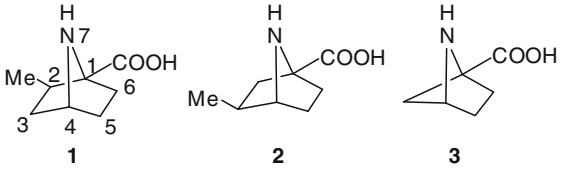


Figure 4-1. Catalysts **1**, **2** and **3**

therefore expected to be restricted due to the intramolecular hydrogen bonding with the amino nitrogen. Another logical extension at this juncture is to restrict the amino group in an out-of-plane position with the help of a methylene bridge, as in a [2.1.1] bicyclic system. This approach results in catalyst **3**.

We have considered two important enamine conformations of the catalysts. The *syn*-enamines formed from catalysts **1** through **3** are found to be more stable than the corresponding *anti*-enamines. More importantly, the TSs for the addition of enamine to aldehyde are found to be energetically more favorable for the *syn*-enamines as compared to corresponding *anti*-enamines. As a representative example, optimized TS geometries for four possible addition modes of **3-**enamine to *p*-nitrobenzaldehyde are provided in Figure 4-2.

When the enamine adds to the aldehyde, the developing alkoxide ion tends to abstract the proton from the carboxylic acid group (Scheme 4-1). The analysis of the TSs revealed that the geometry for such proton transfer is optimal with the TSs involving *syn*-enamine. Similar geometric features are reported previously for proline-catalyzed aldol reactions. Further, the geometric distortion suffered by the ensuing iminium ion in the case of *syn*-enamines is found to be only minimal as compared to an ideal planar geometry around the nitrogen atom, exhibiting a closer resemblance with the product geometry. These factors would undoubtedly contribute toward imparting improved stabilization for the *syn*-enamine TSs. The planarity of the putative iminium moiety for each TS is analyzed in detail using the C_{ring}–N–C–$C_{iminium}$ dihedral angles, described as θ_1–θ_4 (inset, Figure 4-2). Interesting correlations between the activation barriers and these dihedral angles emerge when different TSs are compared. Larger deviations from planarity are generally found for additions involving higher activation barriers. In [2.1.1] bicyclic catalyst **3**, both *syn* and *anti* enamines geometries deviate from planarity, with a more prominent deviation in *anti*-enamines. Interestingly, the differences in the activation barriers between *syn* and *anti*-enamine additions are much more pronounced in [2.1.1] bicyclic catalysts (**3**) than that with [2.2.1] bicyclic systems (**1** and **2**). It is noticed that the rigidity of [2.1.1] bicyclic framework leads to a relatively less favorable geometry for the proton transfer from the carboxylic acid group to the developing alkoxide in the *anti*-enamine TSs. Such proton transfer is facilitated at the expense of larger geometric distortion around the developing iminium nitrogen. Another contributing factor helping to achieve additional stabilization for s-*re* TS presumably originates from the C–H$\cdots\pi$ stabilizing interaction noticed between the CH_3-hydrogens of the enamine/iminium with the aryl group of the aldehyde (Figure 4-2a). Other stereoisomeric TSs involved in this example, i.e., s-*si* as well as a-*re*/a-*si* TSs lack such interactions. The relative activation barriers would eventually depend on the presence or absence of all these stabilizing interactions. The relative activation enthalpies calculated with respect to the lowest energy TSs and the corresponding enantiomeric excess for all the proposed catalysts in this series are summarized in Table 4-1.

The nature and position of the substituents on the catalyst could be employed as an effective method in fine-tuning the stereoselectivity. This is evident from the

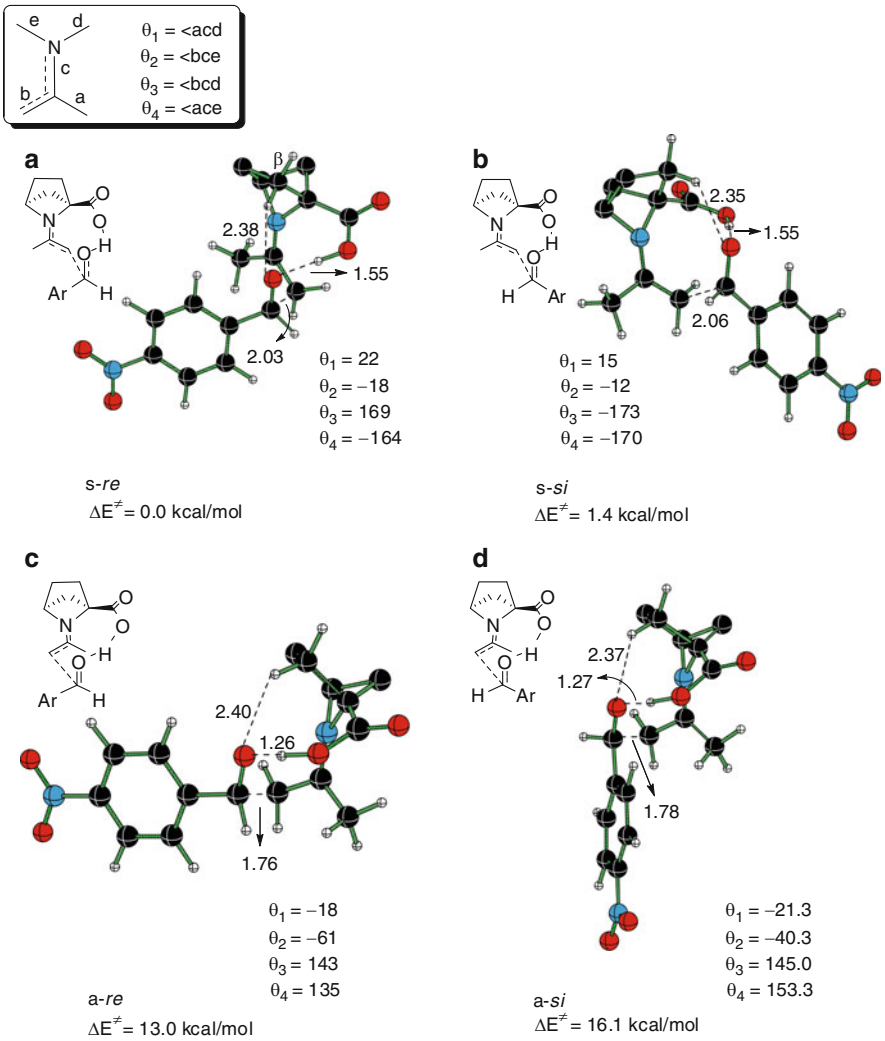
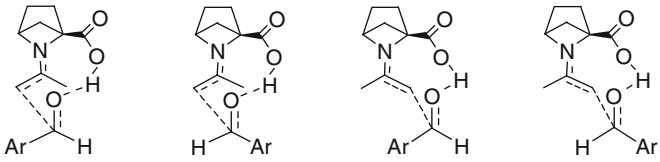


Figure 4-2. The B3LYP/6-31G* optimized geometries of the transition states for the stereoselectivity determining C–C bond formation between the enamine derived from catalyst **3** and acetone to p-nitrobenzaldehyde (Shinisha and Sunoj [31] http://dx.doi.org/ 10.1039/b701688c. Reproduced with permission from the Royal Society of Chemistry)

modest improvement in the enantiomeric excess noticed for catalyst **1** as compared to **2**. When the methyl substituent is placed closer to the enamine, as in **1**, the stereoselectivity is found to be better. Further elaborations using larger substituents at this position could be valuable towards improving the overall selectivity. Such modulations on the catalysts might be a useful strategy even for sterically unbiased electrophilic substrates.

Table 4-1 Computed relative activation barriers ($\Delta E^{\neq}$) obtained at the $CPCM_{(DMSO)}$/B3LYP/6-311+G**//B3LYP/6-31G* level for the addition of enamines to p-nitrobenzaldehyde and the corresponding enantiomeric excess for catalysts **1–7**[a]



Catalysts	Relative activation energies $\Delta E^{\neq}$ in kcal/mol[b]				Enantiomeric excess (in %)
	Modes of approach				
	a-*re*	a-*si*	s-*re*	s-*si*	
1	2.1 (2.2)	5.0 (5.0)	0.0 (0.0)	1.6 (1.6)	87 (87)
2	2.4 (2.8)	5.4 (5.7)	0.0 (0.0)	1.5 (1.5)	85 (85)
3	13.0 (14.6)	16.1 (15.9)	0.0 (0.0)	1.4 (1.6)	82 (87)
4	3.8 (4.0)	6.7 (10.8)	0.0 (0.0)	1.8 (1.9)	91 (92)
5	13.3 (14.3)	32.5 (2.9)	0.0 (0.0)	1.9 (1.7)	92 (89)
6	20.6 (20.5)	7.9 (8.1)	0.0 (0.0)	1.8 (1.8)	90 (91)
7	16.1 (16.2)	19.3 (19.3)	0.0 (0.0)	1.5 (1.5)	84 (85)

[a]Schematic representation of TSs corresponding to the attack of anti/syn enamine on re/si face of aldehyde (for catalyst **3**) is provided.

[b]Gas phase relative activation barriers $\Delta H_{298K}^{\neq}$ including scaled zero-point energies obtained at the B3LYP/6-311+G**//B3LYP/6-31G* are given in parentheses (Shinisha and Sunoj [31] http://dx.doi.org/ 10.1039/b701688c. Reproduced by permission of the Royal Society of Chemistry)

4.3.1.2. Structural Modification of N Out-of-Plane Bicyclic Proline Analogues by Fine-Tuning the Weak Interactions

On the basis of the calculated absolute and relative activation barriers, it is noticed that the *syn*-enamines (s-*re*) derived from Catalysts **1–3** tend to exhibit a general preference toward the *re*-face attack on the aldehyde. The TSs belonging to this series enjoy an additional C(β)H···O hydrogen bonding stabilization between the developing alkoxide and a suitably aligned C(β) hydrogen, as depicted in Figure 4-2. The conformational features of the pyrrolidine ring in these bicyclic systems help to modulate such favorable weak interactions that could influence the relative stabilization of the diastereomeric TSs. As a rational design strategy, we envisaged that fine-tuning the acidity of C(β)H might have a direct bearing on the relative energies of the key TSs. Thus, the replacement of the adjacent C(γ) methylene group of the azabicyclic system by a more electronegative heteroatom is considered. The presence of an α-heteroatom will impart enhanced acidity to the C(β) hydrogen and thus will improve its ability towards stabilizing the developing alkoxide in the TS. We have studied a series of α-heteroatom (oxygen and sulfur) substituted azabicyclic

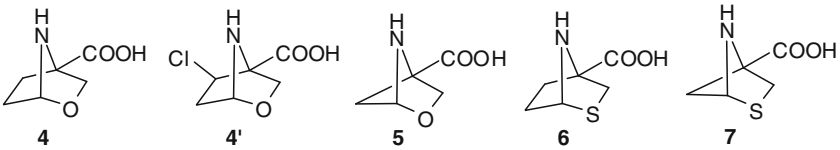


Figure 4-3. Catalysts **4–7**

catalysts (Figure 4-3, **4–7**). The examination of the optimized TS geometries provided in Figure 4-2 indicates that the C(β′) substituent can have a direct influence on the orientation of carboxylic acid group, which in turn, can affect the crucial proton transfer process. To verify the existence of any such effects, the C(β′)H in a representative case (catalyst **4**) is substituted by chlorine atom.

The computed enantiomeric excess with these catalysts is found to be indeed encouraging (Table 4-1). The enantiomeric excess is consistently higher than the corresponding unmodified bicyclic catalysts as well as the parent proline. For example, the predicted enantiomeric excess for catalyst **3** is 82% while that for **5** is as high as 92%. These values are noticeably higher than the experimental as well as DFT predicted *ee*'s for proline-catalyzed direct aldol reaction. The computed enantiomeric excess for **4′** is found to be quite similar in range as that of the corresponding unsubstituted system (**4**). In these cases, computed activation barriers of the C–C bond formation step with *syn*-enamine derived from [2.1.1] catalysts are found to be much lower than that with the [2.2.1] catalysts. Further, the energy differences between TSs for *syn* and *anti* enamine additions to the electrophile are more pronounced in [2.1.1] system. Another key position on the catalyst framework is the C(β) position. In the lowest energy addition TSs for [2.1.1] and [2.2.1] catalysts, the C(β)H$\cdots$O$^{\delta-}$ interaction is found to be stronger in [2.1.1] according to the optimized distances.

4.3.1.3. Bicyclic Systems Where C-β Is Out-of-Plane in the Envelope Conformer of Proline

In the second set, the constraints are applied to have C-β (C-7) of the pyrrolidine ring at the apical position in the [2.2.1] bicyclic system (Figure 4-4, **8**). The synthesis of this compound is reported [53].

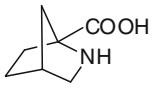
The *syn*-enamine generated from catalyst **8** is found to be marginally more stable than the corresponding *anti*-enamine. Among the four stereochemical modes of addition of enamine to the electrophile, a-*re* TS is identified as the lowest energy TS for the C–C bond formation. The geometrical features provided additional insights on the factors contributing to the vital energy differences between these TSs. A network of stabilizing weak interactions is found to be relatively more in favor of the *anti*-enamine TSs. These include (i) the hydrogen bonding interaction between the developing alkoxide and the partially positive hydrogen of the carbon adjacent to the nitrogen, (ii) intramolecular hydrogen bonding between C(β)H and the developing alkoxide, and (iii) Coulombic interaction between the incipient

iminium nitrogen ($N^{\delta+}$) and the alkoxide. In an effort to achieve optimal proton transfer distance between the carboxylic acid group and the developing alkoxide in *syn*-enamine TSs, change in orientation of the substituents around the incipient C–C bond takes place and results in lowering of intramolecular stabilizing interactions. The orientation of substituents around the incipient C–C bond is relatively more eclipsed in the *syn*-enamine TSs. The highest eclipsing interactions are also found with the s-*si* TS. The computed activation barriers for the addition of *syn*-enamines derived from catalyst **8** are much higher than that for the corresponding *anti*-enamines. Now, among the two lower energy diastereomeric TSs from *anti*-enamines, a-*re* TS is found to be the lowest energy TS, where the phenyl substituent is in a least hindered position. The energy difference between the two lower energy TSs (a-*re* and a-*si*) leading to diastereomeric products is found to be 2.1 kcal/mol. This corresponds to an enantiomeric excess of 95%.

4.3.1.4. Bicyclic Systems Where C-γ Is Out-of-Plane in the Envelope Conformer of Proline

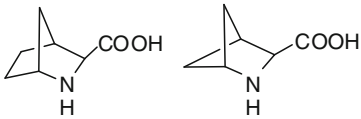
In the third set of catalysts, the most stable conformer of proline is constrained as a part of [2.2.1] or [2.1.1] bicyclic systems (Figure 4-5, **9** and **10**) [54, 55].

These catalysts also exhibited a preference towards *anti*-enamine addition involving a-*re* TS as the lowest energy pathway as with the parent proline. Relatively more staggered arrangement of the substituents around the new C–C bond is noticed with the *anti*-enamine TSs while it is more eclipsed with the *syn*-enamine TSs. Such geometric features evidently lead to higher energy TSs for the *syn*-enamine addition pathway. Between the *anti*-enamine TSs, the *re*-facial attack on the aldehyde is more favored over the corresponding *si*-facial approach, since the aryl substituent on the aldehyde in the former is found to be sterically better positioned. Perhaps the most striking feature emerging from this investigation relates to the correlation between the catalyst structure and stereoselectivity. It is worthwhile to compare the enantiomeric excess calculated for catalysts **9** and **10** with that of proline, from a structure-selectivity point of view. The predicted *%ee* for catalyst **9** is



8

Figure 4-4. Catalyst **8**



9 **10**

Figure 4-5. Catalyst **9** and **10**

conspicuously quite close (74%) to the parent proline (75%). Even though in **9** and **10**, the same conformation of proline can be thought of as constrained in a bicyclic framework, the increased rigidity of the [2.1.1] system is found to be good in improving the enantiomeric excess in **10** (Table 4-2, entry 3).

4.3.1.5. Bicyclic System Where C-β Is Out-of-Plane in the Envelope Conformer of β-Proline

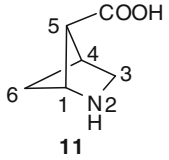
The position, orientation and acidity of the carboxylic acid group are known to be important in contributing to the catalytic ability of proline and its derivatives in aldol reactions. For instance, proline and pyrrolidine-3-carboxylic acid have been reported to yield products of opposite stereochemistry in the Mannich reaction between 3-pentanone and N-PMP-protected α-iminoester [56]. To verify how a bifunctional variant of proline, with a 1,3 relationship between the amino and the carboxylic acid groups perform, as compared to other bicyclic catalysts, we have considered a [2.1.1] bicyclic system of β-proline (Figure 4-6, catalyst **11**) [57].

The computed activation barriers and the enantiomeric excess are given in Table 4-2. Though the carboxylic acid group is not adjacent to the enamine nitrogen, the distance is found to be close enough to facilitate the crucial proton transfer to the developing alkoxide. In the case of catalyst **11**, the *syn*-enamine is found to be more stable than the *anti*-enamine. Further, the TSs resulting from the *syn*-enamines are much more stable than those formed from the *anti*-enamines.

Table 4-2 Computed relative activation barriers ΔE obtained at the $CPCM_{(DMSO)}$/B3LYP/6-311+G**//B3LYP/6-31G* level for the addition of enamines to *p*-nitrobenzaldehyde and the corresponding enantiomeric excess for **8–11**

Catalysts	Relative activation energy $\Delta E^{\dagger}$ in kcal/mol [a]				Enantiomeric excess(% *ee*)
	Modes of approach				
	a-*re*	a-*si*	s-*re*	s-*si*	
8	0.0 (0.0)[a]	2.1 (2.2)	8.2 (8.6)	7.3 (6.9)	95 (95)
9	0.0 (0.0)	1.1 (1.7)	10.5 (2.8)	2.9 (3.6)	75 (88)
10	0.0 (0.0)	1.3 (1.9)	1.2 (2.0)	2.1 (2.8)	80 (91)
11	3.6 (7.6)	4.3 (7.2)	0.1(0.0)	0.0 (0.9)	5 (65.2)

[a]Gas phase activation barriers $\Delta H_{298K}^{\dagger}$ including the scaled zero-point energies obtained at the B3LYP/6-311+G**//B3LYP/6-31G* are given in parentheses (C.B. Shinisha, R.B. Sunoj (2007) *Organic Biomolecular Chemistry* http://dx.doi.org/ 10.1039/b701688c. Reproduced with permission from the Royal Society of Chemistry)



11

Figure 4-6. Catalyst **11** (2-azabicyclo[2.1.1]hexane-6-carboxylic acid)

But the relative energies between the *syn*-enamine TSs are very low. Hence the overall enantioselectivity offered by this catalyst is found to be the lowest among the present series of catalysts investigated. Based on the computed enantiomeric excess, it seems evident that the 1,2-relationship between the secondary amino group and the carboxylic acid group is a highly desirable feature for amino acids to act as a potential asymmetric catalysts for aldol reactions. A recent experimental work on intermolecular aldol reaction with constrained β-proline as catalyst gave lower enantioselectivity than parent proline catalyzed aldol reaction [58]. These results are in accordance with our findings where 1,2-relationship between amino and carboxylic group was found to be more desirable. However, the most encouraging aspect is that this study reports an improved enantioselectivity while using constrained β-proline than when a parent β-proline is employed.

We have also considered substrate level changes (electrophilic aldehyde) for this reaction. Different substitutions on the aromatic aldehyde are therefore studied for their reaction with the enamine derived from a representative catalyst (**2**). Methyl substitutions at the 2,6 positions of the aromatic aldehyde are found to be quite effective in enhancing the enantiomeric excess up to 99%. Further, the role of electronically active substituents on the energetics of addition, when placed at the para-position of electrophilic aldehyde, has also been investigated. The calculated activation barrier as well as the enantiomeric excess exhibited little variation compared to the original *p*-nitrobenzaldehyde.

In the next section, our studies on the mechanistic details of sulfur ylide mediated achiral model reactions, as well as our efforts towards designing a bicyclic S-reagent for asymmetric transformations are elaborated.

4.3.2. Sulfur Ylide Promoted Reactions

4.3.2.1. Mechanism and Diastereoselectivity Studies on Achiral Models

In order to understand the mechanistic features of S-ylide promoted reactions, density functional calculations on achiral model reactions as shown in Scheme 4-5 were performed. Differently substituted ylides [59] were chosen and energetic features for addition to acceptors such as imine (aziridination) [60] and enones (cyclopropanation) [61] were investigated.

According to the generally accepted mechanism [41, 42] a more elaborate picture of the reaction mechanism illustrating various elementary steps involved in the (*re*,*si*) attack of ylide to the acceptor (imine or enone) leading to the formation of 2,3-*trans* product is presented as a representative example (Scheme 4-6). As shown, the reaction proceeds via three distinct and consecutive steps namely, (i) the attack of ylide on the electrophile (or the acceptor), (ii) rotation from the *cisoid* to *transoid* betaine intermediate (around the newly formed C–C bond), and (iii) the *antiperiplanar* elimination of SMe_2 group leading to the ring-closed product. A similar treatment involving a (*re*,*re*) diastereomeric reactant approach will result in the formation of 2,3-*cis* products via the formation of *syn* betaine intermediates. The other two possibilities, resulting from the *si* face of ylide approaching the

electrophile, namely, (*si*,*re*) and (*si*,*si*) will result only in enantiomeric pairs of the above products, and hence are not considered.

Depending on the nature of substituents present on the ylide and acceptor moieties, barriers of any of the three steps mentioned above can be crucial toward deciding the rate and diastereoselectivity of reaction. For instance, the rate and selectivity in the formation of epoxide from dimethylsulfonium benzylide and benzaldehyde is predicted to be controlled by the torsional barriers (for the *cisoid–transoid* betaine interconversion) [42a]. This is an example illustrating the key role of charge stabilizing substituents on the acceptor molecule. Since electrostatic factors is the major stabilizing factor in the case of epoxides, torsional motion from *cisoid* to *transoid* form of betaine intermediate is highly energy demanding. At the same time, this effect is expected to be minimal in the related aziridine or cyclopropane formation where electron withdrawing substituents are present on the acceptor (imine or enone) double bond. In a similar manner, the electron withdrawing ability of ylidic substituent can as well regulate the final outcome of S-ylide promoted reactions (vide infra).

The relative energies of the crucial TSs along the (*re*,*re*) and (*re*,*si*) diastereomeric pathways for selected examples of substituted ylides for aziridination and cyclopropanation reactions are presented (Table 4-3). Representative examples from each semistabilized (**13**/**13′**) and stabilized (**16**/**16′**) class of ylides are chosen for further illustration. The reaction profiles for the corresponding data are provided in Figures. 4-7–4-10.

Table 4-3 The relative energies and activation barriers (in parentheses) of crucial TSs for the formation of aziridination and cyclopropanation shown in Scheme 4-6 [a]

Ylide	TS-**n***c*[b]	TS-**n***t* [c]	TS-**n**R[d]	TS-**n**E[e]
Aziridines[f]				
13	1.8	4.0	–12.1 (9.4)	–13.4 (0.8)
13′	2.0	3.1	–15.4 (6.5)	–10.2 (5.5)
16	8.1	10.4	7.8 (9.6)	9.2 (6.6)
16′	9.9	9.3	5.7 (7.5)	9.9 (8.1)
Cyclopropanes[g]				
13	8.8	10.1	8.8 (15.8)[h]	1.5 (3.7)
13′	10.2	10.4	−1.5 (7.4)	−0.8 (0.7)[h]
16	20.9	24.3	23.2 (9.3)	21.6 (4.6)
16′	20.2	19.9	18.5 (3.5)	17.0 (3.7)

[a]Relative energies (in kcal mol^{-1}) with respect to the separated reactants
[b]*Cisoid* addition
[c]*Transoid* addition
[d]Torsional barrier with respect to the nearest *cisoid* betaine intermediate I-**n***c*
[e]Elimination barrier with respect to *transoid* betaine I-**n***t*
[f]$PCM_{(MeCN)}$/B3LYP/6-311G**//B3LYP/6-31G* energies (Reproduced from [60] with permission from *Chemistry A European Journal*. Copyright ©2007 Wiley-VCH Verlag GmbH & Co. KGaA)
[g]$PCM_{(MeCN)}$/B3LYP/6-311G**// B3LYP/6-31G*energies (Adapted in part with permission from *Journal of Organic Chemistry* 61. Copyright ©2007 American Chemical Society)
[h]$PCM_{(MeCN)}$ /B3LYP/6-311+G**//HF/6-31G* energies

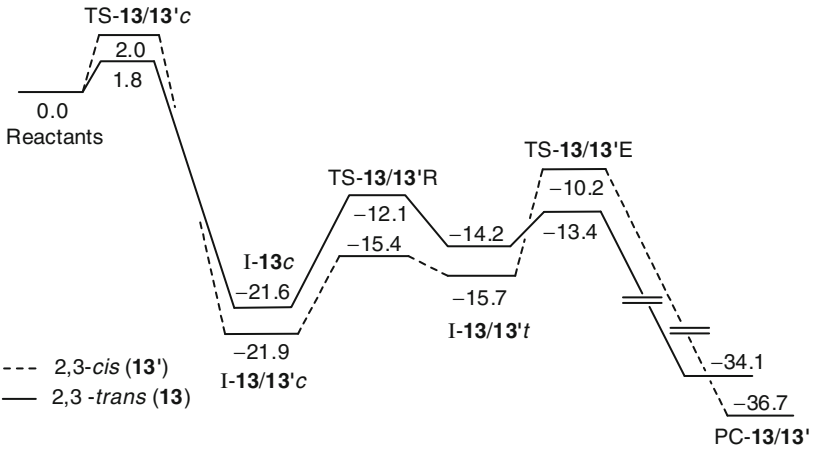
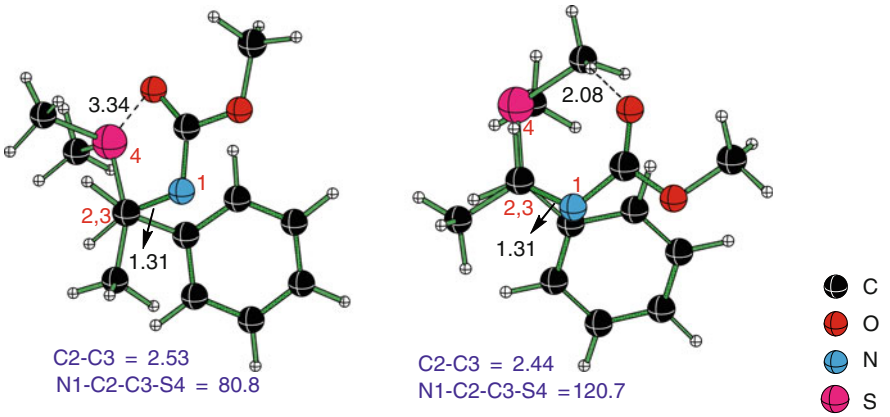


Figure 4-7. Reaction profiles for diastereomeric aziridine formation from semistabilized ylide **2** (R = Ph) and imines. ΔE (in kcal mol^{-1}) computed at the PCM$_{(MeCN)}$/B3LYP/6-311G**//PCM$_{(MeCN)}$/B3LYP/6-31G* level. Energies are reported with respect to the isolated reactants. PC-**13/13′** refer to the diastereomeric product complexes (Reproduced from [60] with permission from *Chemistry A European Journal*. Copyright Wiley-VCH Verlag GmbH & Co. KGaA)



TS-**13***c* (1.8) TS-**13***t* (4.0)

Figure 4-8. The PCM$_{(MeCN)}$/B3LYP/6-31G* optimized transition state geometries (view along C_2–C_3 bond) and activation barriers (in parentheses) of *cisoid* and *transoid* addition TSs for semistabilized ylide **13** [Distances in Å, Angles in °, ΔE in kcal mol^{-1} at the PCM$_{(MeCN)}$/B3LYP/6-311G**//PCM$_{(MeCN)}$/B3LYP/6-31G* level. Energies are reported relative to separated reactants]. (Reproduced from [60] with permission from *Chemistry A European Journal*. Copyright 2007 Wiley-VCH Verlag GmbH & Co. KGaA)

The activation energies presented in the above table clearly show that the addition to imine is favored than to enone. While the initial addition barriers increases according to the inherent stability of the ylide, the ensuing torsional and elimination barriers are not greatly affected. It can as well be noticed that, in

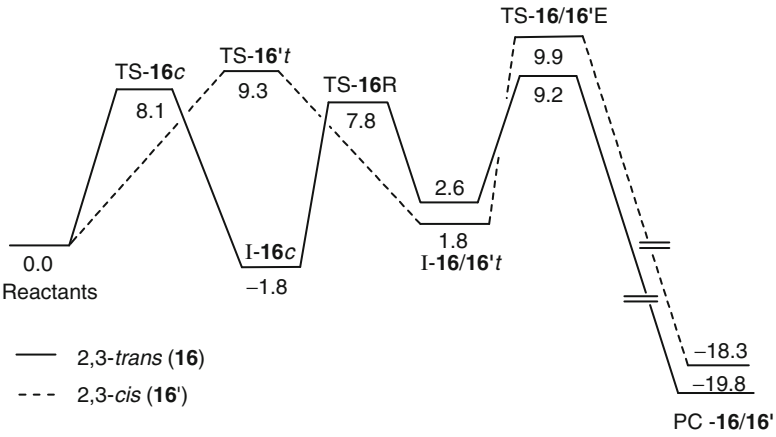


Figure 4-9. Reaction profiles for diastereomeric aziridine formation from stabilized ylide **16** (R = COMe) and imines. ΔE (in kcal mol^{-1}) computed at the $PCM_{(MeCN)}$/B3LYP/6-311G**// $PCM_{(MeCN)}$/B3LYP/6-31G* level. Energies are reported with respect to the isolated reactants. PC-**16**/**16′** refer to the diastereomeric product complexes (Reproduced from [60] with permission from *Chemistry A European Journal*. Copyright 2007 Wiley-VCH Verlag GmbH & Co. KGaA)

general, *cisoid* addition TSs are lower in energy than the corresponding *transoid* congeners. The data additionally indicate that the same ylide can behave differently with the two acceptors, rendering different extent of diastereoselectivity. In the case of aziridination of stabilized ylides, for instance, the highest energy stationary point is the ring-closure TS (TS-**16**E and TS-**16′**E), which controls the rate and diastereoselectivity of the reaction. At the same time, in the case of cyclopropane formation from stabilized ylide, the torsional TSs are higher in energy than the elimination TSs (Table 4-3, also see Figure 4-10). On the other hand, the semistabilized ylide behaves almost similarly with the both kinds of acceptors. Here, the addition TS is the highest energy point on the reaction profile and it plays a key role toward rendering diastereoselection (Figure 4-7).

An inspection of the above reaction profile reveals that the initial addition of semistabilized ylide to imines is irreversible. Thus, the addition step controls the rate and selectivity in this reaction. The subsequent torsional and elimination steps are highly facile as indicated by the lower energies of the corresponding TSs and intermediates. However, smaller energy differences between the diastereomeric addition TSs implies only a very low diastereoselectivity when semistabilized ylide **13** is employed for aziridination reaction.

In order to understand the reasons behind the predicted relative energy distributions, the optimized TS geometries of diastereoselectivity deciding addition TSs are analyzed in detail (Figure 4-8). It is noticed that the *cisoid* addition TS is generally favored because of the proximal arrangement of developing charge centers ($S^{\delta+}$ and $N^{\delta-}$ (or $C^{\delta-}$)) around the newly forming C–C bond. The optimized geometries show that in addition to the electrostatic effects, steric interaction between the different substituents on the C2–C3 bond as well affect the relative energy

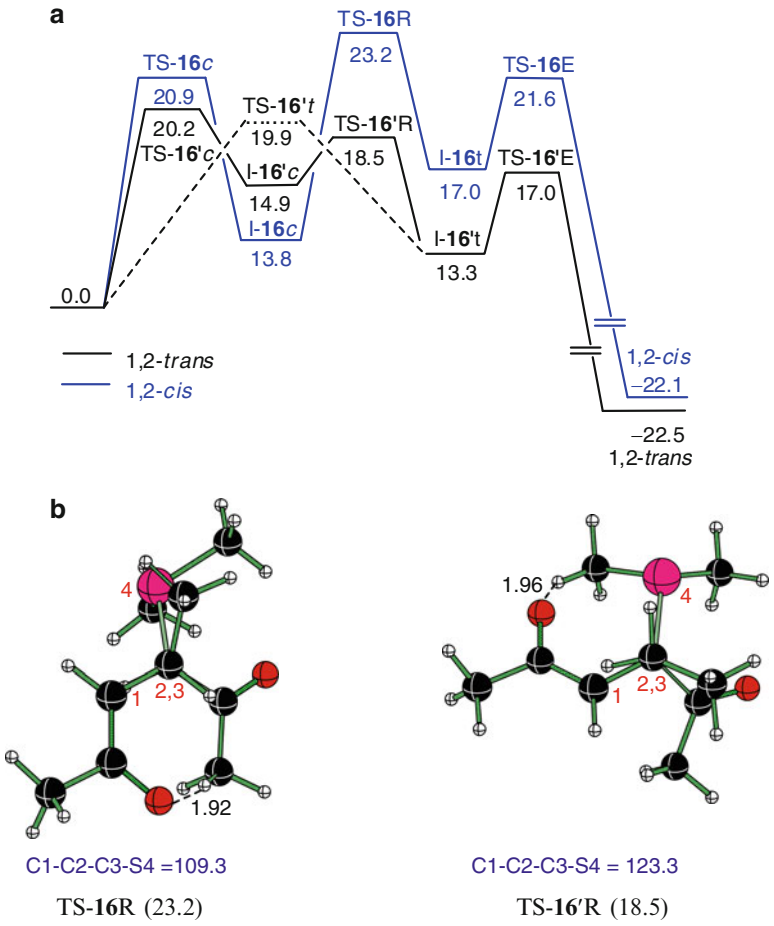


Figure 4-10. (**a**) Reaction profiles for diastereomeric cyclopropane formation from stabilized ylide **16** (R = COMe) and enones. Energies (ΔE in kcal mol^{-1}) at the PCM$_{(MeCN)}$/B3LYP/6-311+G**//B3LYP/6-31G* level are relative to infinitely separated reactants. (**b**) The B3LYP/6-31G* optimized geometries and relative energies (ΔE in kcal mol^{-1}) of diastereomeric torsional TSs Distances in Å and angles in °; See Figure 4-8b for color schemes adopted) (Reprinted with permission from D. Janardanan. R.B. Sunoj, J. Org. Chem. Copyright 2007, American Chemical Society)

differences significantly (notice the unfavorable interaction between CO_2Me/Ph substituent pair in TS-**13***t*, Figure 4-8). However, in the case of semistabilized ylide (in the formation of both aziridines and cyclopropanes) smaller energy differences are predicted between the diastereomeric addition TSs. Therefore, the *trans* diastereoselectivity in product formation is found to be very low with this kind of ylides [59, 60].

On the contrary, the stabilized ylide presents a different scenario. In the case of aziridine formation from stabilized ylide, the addition and torsional steps are found to be reversible and hence the rate and selectivity are controlled by the ring-closure

barriers. Since the energy differences between the diastereomeric elimination TSs are not large, the reaction is predicted to be only moderately selective towards 2,3-*trans* aziridines (Figure 4-9) [60].

With enones, stabilized ylide addition is characterized by a rate-limiting torsional barrier (Table 4-3). The large energy difference of the order of 4.7 kcal mol^{-1} between the diastereomeric torsional TSs leads to almost complete selectivity toward the formation of 1,2-*trans* cyclopropanes (Figure 4-10). An inspection of the optimized TS geometries reveals that steric effects play a major role in deciding the energetic preferences. The unfavorable eclipsing interaction between -SMe_2 and -Me substituents on the C2–C3 bond keeps the energy of TS-**16**R high such that 1,2-*trans* cyclopropane is formed exclusively from TS-**16′**R. Other weak stabilizing interactions such as H-bonding are found to be comparable for both these TSs (Figure 4-10).

From the above discussions, we have attempted to clarify that the reaction mechanism and stereoselectivity of S-ylide promoted aziridination and cyclopropanation reactions are highly dependent on the nature of ylide and the acceptor. While comparable electronic and steric factors lead to smaller energy differences between diastereocontrolling TSs in the case of semistabilized ylide additions, moderate to high *cis*/*trans* selectivities are predicted for the stabilized ylide additions.

4.3.2.2. Designing Catalyst for Asymmeric Aziridination

Insights obtained from the model reactions, as described above, are extended toward designing a chiral S-reagent that can offer high enantio- and diastereoselectivity [62]. Even though asymmetric aziridination employing chiral S-reagents has received increased attention in the recent years, little is known about the chiral induction associated with the reaction. We recognized that a detailed survey on the available addition possibilities is highly essential since the existing rationalization on chirality transfer does not provide enough details except for certain qualitative viewpoint [63]. Further, it is very clear from the available model studies that even predicting the characteristics of a simple achiral model reaction with differently substituted ylide is difficult to generalize. Thus we aim at understanding the factors controlling the enantioselection process associated with asymmetric aziridine formation, along with the particular objective of designing a better catalytic S-reagent for ylide mediated asymmetric transformations. A bicyclic chiral sulfur reagent is chosen [64] and suitably modified so as to obtain good facial selectivity in the addition step. A schematic representation of the chiral model reaction is presented (Scheme 4-7). The structure of chiral S-reagent originally reported by Durst et al. for the asymmetric epoxidation reaction is also provided in the inset (Scheme 4-7).

Since the facial discrimination and selectivity in stereochemically different approaches between reactants is the primary factor responsible for enantioselection, all possible reactant approaches and ylide conformations are considered for the initial addition step (Scheme 4-8). For representative cases, diastereomeric reaction energy profiles are also evaluated.

4.3.2.2.1. Aziridines from Stabilized Chiral S-Ylides

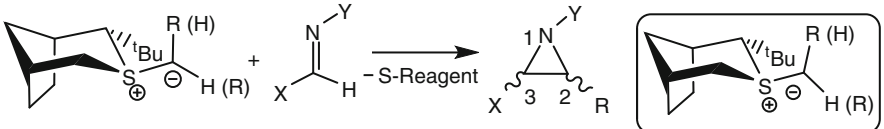
After evaluating all key TSs for the initial addition step, entire reaction energy profiles were constructed for the addition of stabilized ylide (R = COMe, Scheme 4-7) to mesylimine **17** (X = Ph; Y = SO_2Me) and carbamate **19** (X = Ph; Y = CO_2Me) as representative examples (Scheme 4-7). This was guided by the fact that diastereoselectivity in aziridination is controlled by the ring-closure step and not by the initial addition in the case of stabilized ylides (see Section 4.3.2.1). The consequences of the above being the requirement that, in order to form the major product, the betaine resulting from the lowest energy addition TS should also possess the lowest barrier for the ensuing ring-closure step (vide infra).

We found that reaction profiles computed for the chiral version of the reaction display similar features as that noticed with the achiral model studies. The computed diastereomeric reaction profiles for the formation of aziridines from stabilized chiral S-ylide show that in the case of N-methoxycarbonyl imine, the major enantiomer can be correctly predicted only if the relative energies of diastereoselective elimination TSs are known (Figure 4-11). It can be seen that the major product does not result from the lowest energy addition TS (**19d**-A-$c^{\ddagger}$), but from the next higher energy TS **19c**-A-$c^{\ddagger}$, which has a lower energy for the ring-closure step. Thus, energy difference of the order of 2.8 kcal mol^{-1} between **19c**-A-$c^{\ddagger}$ and the corresponding TS resulting in the enantiomeric product (**19b**-A-$c^{\ddagger}$) leads to a high selectivity (98%*ee* and 64%*de*) toward the formation of (2*R*,3*S*)-*trans* aziridine.

In the case of mesyl substituted imine on the other hand, the extent of enantioselection can be computed solely from the relative energies of the addition TSs, since the elimination as well as the addition steps favor the same diastereomeric pathway. In this case as well, high enantio- and diastereoslectivities are predicted.[62]

4.3.2.2.2. Aziridines from Semistabilized Chiral S-Ylides

Available experimental reports [37]b as well as the computational results on the model systems [60] indicate that the addition step is non-reversible in aziridination

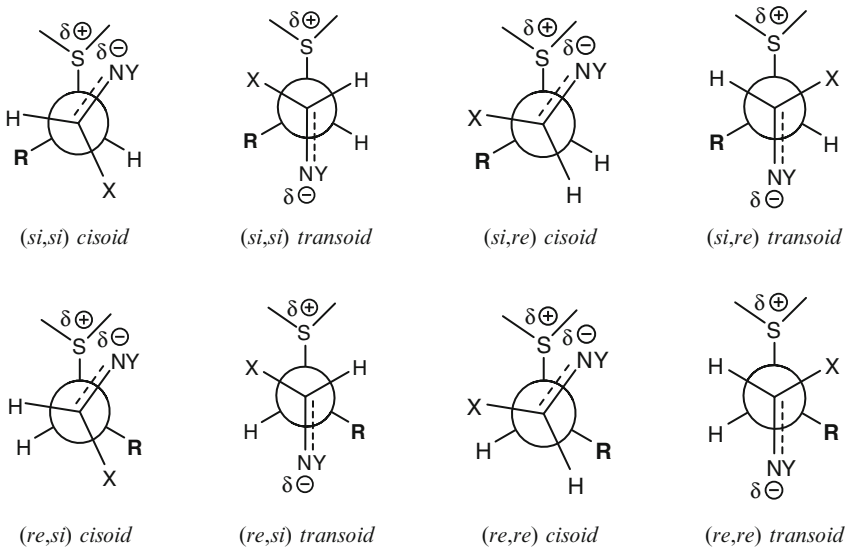


R = COMe, Ph, Me

X = Ph, Me; Y = SO_2Me, CO_2Me

X = Ph, Y = SO_2 Me, R = COMe (**17**)	X = Ph, Y = SO_2Me, R = Ph (**21**)
X = Me, Y = SO_2Me, R = COMe (**18**)	X = Me, Y = SO_2Me, R = Ph (**22**)
X = Ph, Y = CO_2Me, R = COMe (**19**)	X = Ph, Y = CO_2Me, R = Ph (**23**)
X = Me, Y = CO_2Me, R = COMe (**20**)	X = Me, Y = CO_2Me, R = Ph (**24**)

Scheme 4-7. Chiral sulfur ylide promoted aziridination reactions investigated



Scheme 4-8. Important stereochemical modes of approach between ylide and aldimine in the initial addition step, leading to enantiomeric and diastereomeric aziridines (Reprinted with permission from D. Janardanan, R.B. Sunoj, J. Org. Chem. 73, 8163. Copyright 2008, American Chemical Society)

from semistabilized ylides and imines. The enantio- and diastereoselectivities involving semistabilized ylides are therefore decided in the initial addition step. Akin to stabilized ylide, seven stereochemically distinct addition TSs are first identified and the corresponding energies are summarized in Table 4-4.

According to the computed relative energies of TSs presented in Table 4-4, the energy difference between TSs *si*,*si*-$c^{\ddagger}$ and *re*,*re*-$t^{\ddagger}$ leading to enantiomeric aziridines is found to be 3.1 kcal mol^{-1} for mesylbenzaldimine (**21**). Such large energy difference corresponds to an *ee* of the order of 99% in favor of (2*S*,3*R*)-*cis* aziridine. On the other hand, the difference in energy between TSs that are responsible for diastereomeric products is comparatively small (1.3 kcal mol^{-1}) and leads to a moderately good *de* of 80%. In general, enantio- and diastereoselectivities associated with the addition of semistabilized ylide to carbamates (**23** and **24**) is lower than that of mesyl imines (**21** and **22**).

The key geometric features of the optimized TS geometries for the representative addition TSs are provided in Figure 4-12 for the mesyl system. It is seen that the interaction between phenyl substituents at the C2 and C3 positions play a major role in regulating the relative energy order. In the lowest energy TS (*si*,*si*-$c^{\ddagger}$), the phenyl substituents are oriented in a *trans* fashion around the C2–C3 bond (with a dihedral angle of 101° between them) to minimize the unfavorable interactions. It is worthy of mention that all other higher energy TSs have a smaller dihedral angle between the C2–C3 phenyl substituents. The optimized TS geometries are further characterized by weak interactions between the C_{α}–H atoms and the developing

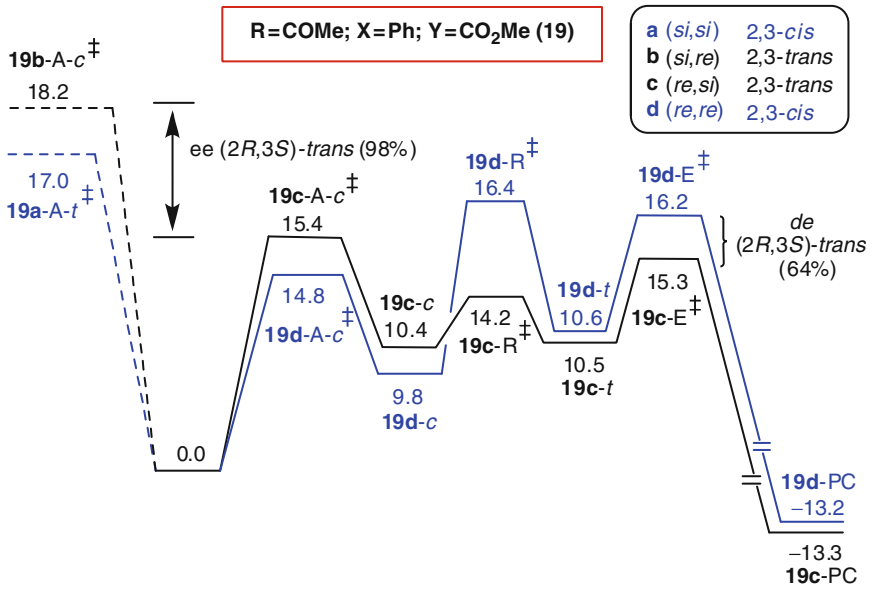


Figure 4-11. The reaction profiles for the lower energy diastereomeric pathways (**c** and **d**) for the addition of stabilized ylide to methyl benzylidenecarbamate. Addition TSs for **a** and **b** modes are also included. The notations **19c-A**-*c*‡, **19c**-*c*, **19c**-*t*, **19c**-R‡, **19c**-E‡ and **19c**-PC respectively denote *cisoid* addition TS, *cisoid* betaine, *transoid* betaine, torsional TS, elimination TS and product complex for pathway '**c**'; Similar notations used for pathway '**d**'; $\Delta E^{\ddagger}$ (kcal mol^{-1}) is computed at the PCM$_{(CH3CN)}$/B3LYP/6-311G**//B3LYP/6-31G* level (Reprinted with permission from [62] D. Janardanan, R.B. Sunoj, J. Org. Chem. 73, 8163 (2008). Copyright 2008, American Chemical Society)

Table 4-4 The computed relative activation energies ($\Delta\Delta E^{\ddagger}$ in kcal mol^{-1}) obtained at the PCM$_{(MeCN)}$/B3LYP/6-311G**//B3LYP/6-31G* level for aziridination reaction between semistabilized ylide (R = Ph) and N-substituted benzaldimines (systems **21**–**24**)[a, b]

Aldimine		(*si*, *si*)		(*si*,*re*)[c]	(*re*, *si*)		(*re*, *re*)		Product Configuration[d]	*ee*, *de* (%)
		c	*t*	*c*	*c*	*t*	*c*	*t*		
21	X = Ph Y = SO_2Me	0.0	3.7	1.3	3.4	3.5	3.3	3.1	(2*S*,3*R*) *cis*	99, 80
22	X = Me Y = SO_2Me	0.0	3.5	1.4	2.1	3.9	3.1	–[c]	(2*S*,3*R*) *cis*	99, 83
23	X = Ph Y = CO_2Me	0.0	2.4	1.1	3.4	3.4	2.5	3.8	(2*S*,3*R*) *cis*	97,73
24	X = Me Y = CO_2Me	0.0	1.7	0.8	1.4	4.0	1.8	3.2	(2*S*,3*R*) *cis*	91,59

[a] $\Delta\Delta E^{\ddagger}$ relative to the most stable addition transition state

[b] '*c*' and '*t*' denote *cisoid* and *transoid* addition modes

[c] *transoid* TS could not be located

[d] The product corresponding to the lowest energy TS (Reprinted with permission from D. Janardanan, R.B. Sunoj, J. Org. Chem. 73, 8163–8174 (2008). Copyright 2008 American Chemical Society)

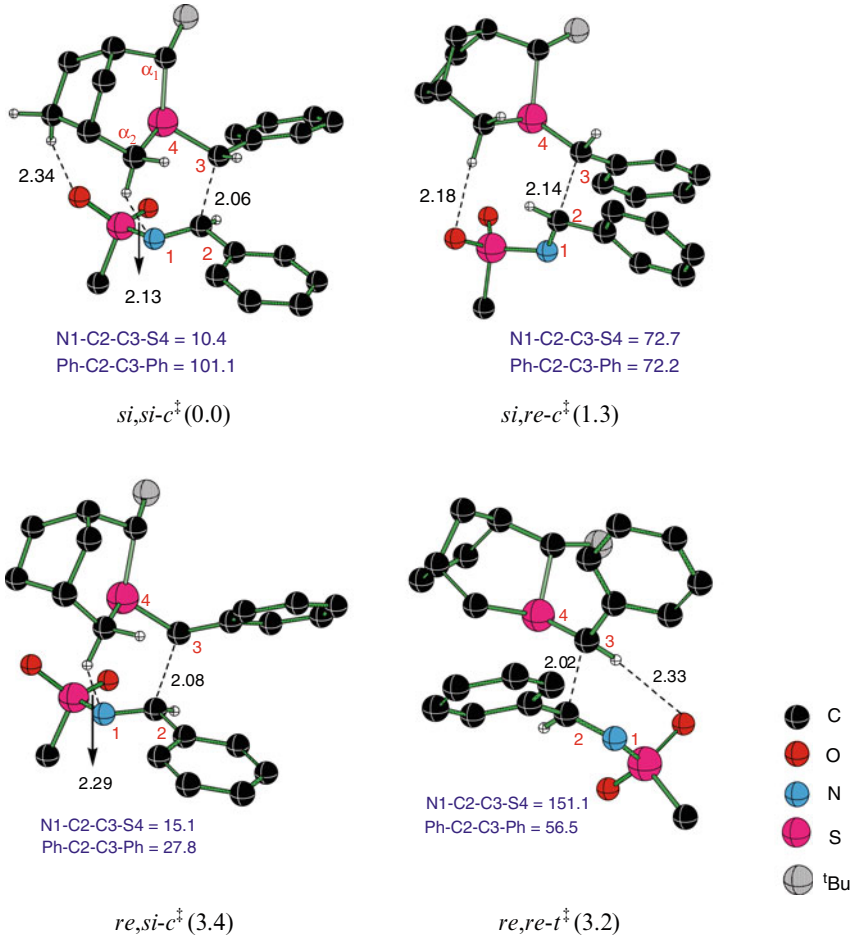


Figure 4-12. The B3LYP/6-31G* optimized geometries and relative energies of lower energy TSs for the addition of semistabilized ylide to substituted imines for system **21** (distances in Å and angles in °). [$\Delta\Delta E^{\ddagger}$ in kcal mol^{-1} obtained at the PCM$_{(MeCN)}$/B3LYP/6-311G**//B3LYP/6-31G* level are provided in parentheses.] (Reprinted with permission from D. Janardanan, R.B. Sunoj, J. Org. Chem. 73, 8163 (2008)62]. Copyright 2008, American Chemical Society)

charges, either $N^{\delta+}$ or $O^{\delta-}$ of the mesyl group. However, these interactions are quite comparable in nature and are not expected to affect the relative energy order significantly.

The above discussion on geometric features convey the important role of steric interactions in deciding the energetic preferences associated with the enantio- and diastereomeric addition TSs for aziridine formation from semistabilized ylides.

Thus, it has been demonstrated that both stabilized and semistabilized ylides derived from a chiral bicyclic sulfur compound can provide excellent enantioselectivities toward 2,3-*cis* aziridines. At this juncture, it is of significance to reckon that many of the available chiral sulfur reagents employed in asymmetric aziridination lead to *trans* aziridines. Our results additionally suggest that the remote alkyl substituents on the bicyclic sulfur compound are not quite necessary to induce high stereoselectivity. The model sulfur reagent devoid of such alkyl groups does not significantly lower the extent of enantio- and diastereoselection. The strategies as described for aziridination could be extended to other ring-formation reactions such as epoxidation and cyclopropanations.

4.4. SUMMARY

Through the present chapter, we have tried to illustrate how the knowledge on the mechanism and factors controlling stereoselectivity can be effectively employed toward the design of new catalysts for asymmetric application. Two key examples have been selected in illustrating this idea. The potential of azabicyclic compounds as new organocatalysts in asymmetric aldol reaction between acetone and *p*-nitrobenzaldehyde are evaluated. The stereoselectivity-determining step, similar to that of the established proline catalyzed aldol reactions, are carefully examined by locating all stereochemically pertinent transition states using the density functional theory methods. The calculations showed that these catalysts could be highly effective for aldol reactions as compared to the popular organocatalysts such as proline. Catalysts **1**–**8,** which are predicted to give enantiomeric excess from 82% to 95% are expected to be superior over proline for which the calculated enantiomeric excess is only about 75%. The reaction is expected to proceed with greater ease as the predicted activation energies are relatively lower. In another example, on the basis of the mechanistic insights on sulfur ylide promoted reactions, a new chiral S-reagent has been predicted as capable of offering an impressively high levels of enantio- and diastereoselectivity toward the formation 2,3-*cis* aziridines from stabilized and semistabilized ylides. The computed results indicate that the rate and stereoselectivity of ylide mediated ring formations depend significantly on the nature of substituents present on the ylide as well as on the acceptor. Importance of weak bonding interactions in regulating the relative energies of key TSs is established. Insights on enantio- and diastereoselection unraveled through this study are expected to facilitate rational design of novel chiral molecules for asymmetric applications.

ACKNOWLEDGEMENTS

Generous computing facilities from IITB computer centre and senior research scholarship from CSIR New Delhi (Shinisha C. B. and Deepa Janardanan) are gratefully acknowledged.

REFERENCES

1. (a) Hoffmann R, Alder RW, Wilcox CF Jr. (1970) J Am Chem Soc 92 (16): 4992–4993 (b) Schleyer P v R. Boldyrev AI (1991) J Chem Soc, Chem Commun 1536–1538 (c) Boldyrev AI, Schleyer P v R (1991) J Am Chem Soc 113 (24):9045–9054
2. Young KJH, Oxgaard J, Ess DH, Meier SK, Stewart T, Goddard WA III, Periana RA (2009) Chem Commun 3270
3. Houk KN, Cheong PH-Y (2008) Nature 455:309–313
4. Grzybowski BA, Ishchenko AV, Shimada J, Shakhnovich EI (2002) Acc Chem Res 35(5):261–269
5. Kitchen DB, Decornez H, Furr JR, Bajorath J (2004) Nat Rev Drug Discov 3(11):313–317
6. Dalko PI, Moisan L (2004) Angew Chem Int Ed 43(39):5138–5175
7. List B, Yang JW (2006) Science 313:1584–1586
8. RMde F, Christmann M (2007) Eur J Org Chem 16:2575–2600
9. Carpenter J, Northrup AB, Chung D, Wiener JJM, Kim S-G, Macmillan DWC (2008) Angew Chem Int Ed 47(19):3568–3572
10. Enders D, Grondal C, Hüttl MRM (2007) Angew Chem Int Ed 46(10):1570–1581
11. Dondoni A, Massi A (2008) Angew Chem Int Ed 47(25):4638–4660
12. List B (2002) Tetrahedron 58(28):5573–5590
13. Tang Z, Jiang F, Yu L-T, Cui X, Gong L-Z, Mi A-Q, Jiang Y-Z, Wu Y-D (2003) J Am Chem Soc 125(18):5262–5263
14. Cheong PH-Y, Houk KN (2004) J Am Chem Soc 126(43):13912–13913
15. Mitsumori S, Zhang H, Cheong PH-Y, Houk KN, Tanaka F, Barbas CF III (2006) J Am Chem Soc 128(4):1040–1041
16. (a) Vignola N, List B (2004) J Am Chem Soc 126(2):450–451 (b) Fu A, List B, Thiel W (2006) J Org Chem 71(1):320–326
17. Balcells D, Maseras F (2007) New J Chem 3:333–343
18. Brown JM, Deeth RJ (2009) Angew Chem Int Ed 48(25):4476–4479
19. List B, Lerner RA, Barbas CF III (2000) J Am Chem Soc 122(10):2395–2396
20. Bahmanyar S, Houk KN, Martin HJ, List B (2003) J Am Chem Soc 125(9):2475–2479
21. Zotova N, Broadbelt LJ, Armstrong A, Blackmond DG (2009) Bioorg Med Chem Lett 19 (14):3934–3937
22. Guillena G, Nájera C, Ramón DJ (2007) Tetrahedron Asymmetr 18(19):2249–2293
23. Hayashi Y, Sumiya T, Takahashi J, Gotoh H, Urushima T, Shoji M (2006) Angew Chem Int Ed 45 (6):958–961
24. Cheong PH-Y, Houk KN, Warrier JS, Hanessian S (2004) Adv Synth Catal 346(9–10):1111–1115
25. Bahmanyar S, Houk KN (2001) J Am Chem Soc 123(51):12911–12912
26. Sakthivel K, Notz W, Bui T, Barbas CF III (2001) J Am Chem Soc 123(22):5260–5267
27. Bassan A, Zou W, Reyes E, Himo F, Córdova A (2005) Angew Chem Int Ed 44(43):7028–7032
28. Czinki E, Császár AG (2003) Chem – Eur J 9(4):1008–1019
29. Wittkopp A, Schreiner PR (2003) Chem – Eur J 9(2):407–414
30. (a) Cheong PH-Y, Houk KN (2004) J Am Chem Soc 126(43):13912–1313 (b) Bahmanyar S, Houk KN (2003) Org Lett 5(8):1249–1251 (c) Fu A, List B, Thiel W (2006) J Org Chem 71(1):320–326
31. Shinisha CB, Sunoj RB (2007) Org Biomol Chem 5:1287–1294
32. Hoffmann R, Schleyer P v R HF III (2008) Angew Chem Int Ed 47(38):7164–7167
33. (a) Trost BM, Melvin LS Jr. (1975) Sulfur ylides –emerging synthetic intermediates. Academic, New York (b) Corey EJ, Chaykovsky M (1965) J Am Chem Soc 87(6):1353–1364 (c) Johnson CR, Schroeck CW (1973) J Am Chem Soc 95(22):7418–7423

34. (a) Hodgson DM, Gibbs AR, Lee GP (1996) Tetrahedron 52(46):14361–14384 (b) Taylor SK (2000) Tetrahedron 56(516):1149–1163
35. (a) Sweeney JB (2002) Chem Soc Rev 31:247–258 (b) Brackmann F, de Meijere A (2007) Chem Rev 107(11):4538–4583 and references therein
36. (a) Li A-H, Dai L-X, Aggarwal VK (1997) Chem Rev 97(6):2341–2372 (b) Blot V, Briere J-F, Davoust M, Miniere S, Reboul V, Metzner P (2005) Relat Elem 180:1171–1182 (c) McGarrigle EM, Myers EL, Illa O, Shaw MA, Riches SL, Aggarwal VK (2007) Chem Rev 107 (12) 5814–5883
37. (a) Aggarwal VK, Alonso E, Fang G, Ferrara M, Hynd G, Porcelloni M (2001) Angew Chem Int Ed 40(8):1433–1436 (b) Aggarwal VK, Charmant JPH, Ciampi C, Hornby JM, O'Brien CJ, Hynd G, Parsons R (2001) J Chem Soc Perkin Trans 1:3159–3166
38. (a) Yang X-F, Zhang M-J, Hou X-L, Dai L-X (2002) J Org Chem 67(23):8097–8103 (b) Deng X-M, Cai P, Ye S, Sun X-L, Liao W-W, Li K, Tang Y, Wu Y-D, Dai L-X (2006) J Am Chem Soc 128 (30):9730–9740
39. (a) Solladié-Cavallo A, Diep-Vohuule A, Sunjic V, Vinkovic V (1996) Tetrahedron: Asymmetry 7(6):1783–1788 (b) Solladié-Cavallo A, Diep-Vohuule A, Isarno T (1998) Angew Chem Int Ed 37(12):1689–1691
40. (a) Zanardi J, Lamazure D, Minière S, Reboul V, Metzner P (2002) J Org Chem 67(25):9083–9086 (b) Davoust M, Brière J-F, Jaffrès P-A, Metzner P (2005) J Org Chem 70(10):4166–4169
41. (a) Volatron F, Eisenstein O (1987) J Am Chem Soc 109(1):1–14 (b) Lindvall MK, Koskinen AMP (1999) J Org Chem 64(13):4596–4606
42. (a) Aggarwal VK, Harvey JN, Richardson J (2002) J Am Chem Soc 124(20):5747–5756 (b) Silva MA, Bellenie BR, Goodman JM (2004) Org Lett 6(15):2559–2562
43. Gaussian 03, Revision C.02, Frisch M J, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr., Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA, Gaussian, Inc., Wallingford CT, 2004.
44. (a) Becke AD (1993) J. Chem. Phys. 98:5648–5652 (b) Becke AD (1988) Phys Rev A 38:3098–3100 (c) Lee C, Yang W, Parr RG (1988) Phys Rev B 37:785–789
45. (a) Ditchfield R, Hehre WJ, Pople JA (1971) J Chem Phys 54(2):720–723 (b) Hehre WJ, Ditchfield R, Pople JA (1972) J Chem Phys 56:2257 (c) Hariharan PC, Pople JA (1973) Theor Chim Acta 28(3):213–222
46. (a) Gonzalez C, Schlegel HB (1989) J Chem Phys 90:2154 (b) Gonzalez C, Schlegel HB (1990) J Phys Chem 94(14):5523–5527
47. Scott AP, Radom L (1996) J Phys Chem 100(41):16502–16513
48. Cossi M, Barone V, Cammi R, Tomasi J (1996) Chem Phys Lett 255(4–6):327–335
49. (a) Gil AM, Bunuel E, Diaz-de-Villegas MD, Cativiela C (2003) Tetrahedron: Asymmetry 14(11):1479–1488 (b) Gil AM, Bunuel E, López P, Cativiela C (2004) Tetrahedron: Asymmetry 15(5):811–819
50. Gante J (1994) Angew Chem Int Ed 33(17):1699–1720
51. Nelsen SF, Ippoliti JT, Frigo TB, Petillo PA (1989) J Am Chem Soc 111(5):1776–1781

52. Alemán C, Jiménez AI, Cativiela C, Pérez JJ, Casanovas J (2005) J Phy Chem B 109 (23):11836–11841
53. Grygorenko OO, Artamonov OS, Palamarchuk GV, Zubatyuk RI, Shishkin OV, Komarov IV (2006) Tetrahedron Asymmetry 17(2):252–258
54. Brandt P, Andersson PG (2000) Synlett 8:1092–1106
55. Jenkins CL, Lin G, Duo J, Rapolu D, Guzei IA, Raines RT, Krow GR (2004) J Org Chem 69 (25):8565–8573
56. Zhang H, Mifsud M, Tanaka F, Barbas CF III (2006) J Am Chem Soc 128(30):9630–9631
57. Krow GR, Huang Q, Lin G, Centafont RA, Thomas AM, Gandla D, DeBrosse C, Carrol PJ (2006) J Org Chem 71(5):2090–2098
58. Armstrong A, Bhonoah Y, White AJP (2009) J Org Chem 74(14):5041–5048
59. Different substituents on the ylidic center were chosen according to their electron withdrawing capacities. Based on literature reports and proton affinity evaluation, the various ylides have been classified into stabilized, semistabilized and non-stabilized categories (see ref. [60] for details).
60. Janardanan D, Sunoj RB (2007) Chem – Eur J 13(17):4805–4815
61. Janardanan D, Sunoj RB (2007) J Org Chem 72(2):331–341
62. Janardanan D, Sunoj RB (2008) J Org Chem 73(21):8163–8174
63. Aggarwal VK, Ferrara M, O'brien CJ, Thomson A, Jones RVH, Fieldhouse R (2001) J Chem Commun Perkin Trans 1(14):1635–1643
64. Breau L, Durst T (1991) Tetrahedron Asymmetr 2(5):367–370

CHAPTER 5

REACTIVE PROCESSES WITH MOLECULAR SIMULATIONS

SABYASHACHI MISHRA AND MARKUS MEUWLY
Department of Chemistry, University of Basel, Klingelbergstr 80, CH 4056, Switzerland,
e-mail: m.meuwly@unibas.ch

Abstract: Describing chemical reactions is one of the most challenging aspects of current computational approaches to chemistry. In this chapter established (EVB, ReaxFF) and novel (MMPT, ARMD) approaches are discussed that allow to study bond forming and bond breaking processes in a variety of chemical and biological environments. Particular emphasis is put on methods that enable investigating the dynamics of chemical reactions. For MMPT and ARMD methods, a number of model studies are discussed in more detail. They include the kinetics of NO rebinding to Myoglobin and the conformational transition in Neuroglobin which explores the full functionality of ARMD. The chapter closes with an outlook of possible generalizations of the methods discussed.

Keywords: Reactive Molecular Dynamics, Proton Transfer, Ligand Binding in Heme Proteins, NO Detoxification

5.1. INTRODUCTION

Chemical reactions involve bond-breaking and bond-forming processes. As such, they are fundamental to chemistry and biology alike. In catalysis, the activation and subsequent breaking of H_2 is of considerable importance. The same applies to activation of C-H bonds. In many cases, mechanistic aspects of the reactions ("which reaction partners interact at which time with each other") are of interest. However, many details are elusive experimentally, because "the reaction" itself is a transient process, the transition state is unstable and thus, the most interesting regions along a reaction path cannot be investigated experimentally in a direct fashion. To shed light on such questions, theoretical and computational work have

P. Paneth and A. Dybala-Defratyka (eds.), *Kinetics and Dynamics*,
Challenges and Advances in Computational Chemistry and Physics 12,
DOI 10.1007/978-90-481-3034-4_5,

become invaluable companions to experimental efforts in understanding particular reaction schemes.

The computational investigation of a chemical or biological system assumes our ability to compute quantities such as the total energy of the system for different arrangements of the atoms. There are two fundamentally different concepts to do that: either by solving the electronic Schrödinger equation (quantum mechanics), or by assuming a suitably defined empirical potential energy function (molecular mechanics). The first approach has been developed over the past 20 years to a degree that allows us to carry out calculations with "chemical accuracy" – which implies accuracies for relative energies to within 1 kcal/mol. A quantum chemical calculation makes no assumption whatsoever about the bonding pattern in the molecule and as such, is ideally suited to answer the question which atoms are bonded to which ones given a particular relative orientation. To obtain realistic reaction profiles it is, however, necessary to carry out calculations at least at the level of density functional theory (DFT) with a suitable basis set. In technical terms this would amount to at least a standard (and validated) basis set of 6–31G quality ideally with diffuse and polarization functions. With such an approach, it is in principle possible to address questions related to the structure and relative energetics of product, educt, and transition state of a given chemical reaction. Also, through elementary statistical mechanics, entropic effects can be included and thus, free energies can be calculated. However, although such computations are by now standard, they can realistically and routinely only be carried out for systems including several tens of heavy atoms. This is due to the N^3 scaling of the secular determinants that need to be diagonalized, where N is the number of primitive Gaussian functions.

The realization that the electronic Schrödinger equation can only be solved for small systems led to the development of important alternative approaches to study chemical reactivity. They date back to London's work on the H + H_2 reaction for which he used a 2 × 2 valence bond treatment [1]. Building on this, refined and extended approaches led to the London-Eyring-Polanyi (LEP) [2], and to the London-Eyring-Polanyi-Sato (LEPS) surfaces [3, 4]. A development that continued the efforts to use valence bond theory to describe multi-state chemical systems, is the diatomics-in-molecules (DIM) theory [5]. Following a slightly different perspective, Pauling profoundly influenced the theoretical description of chemical reactivity through his work on molecular structure and the nature of the chemical bond [6, 7]. Significant relations, such as the one between bond length and bond order later became foundations for empirical descriptions of reactivity [8, 9].

Explicit exclusion of all electronic effects finally leads to empirical fields. They were developed with the emphasis to carry out studies of the structure and dynamics of macromolecules, including peptides and proteins [10–17]. Thus, their primary application area were sampling and characterizing conformations of extensive molecular structures where reorganization of the bonds would not occur. The mathematical form

$$\begin{aligned} V_{\text{bond}} &= \sum K_b (r - r_e)^2 \\ V_{\text{valence}} &= \sum K_\theta (\theta - \theta_e)^2 \\ V_{\text{dihe}} &= \sum K_\phi (1 + cos(n\phi - \delta)) \end{aligned} \tag{5.1}$$

of empirical force fields is not suitable to describe chemical reactions where chemical bonds are broken and formed. Here, the K are the force constants associated with the particular type of interaction, r_e and θ_e are equilibrium values, n is the periodicity of the dihedral and δ is the phase which determines the location of the maximum. The sums are carried out over all respective terms. Nonbonded interactions include electrostatic and van der Waals terms which are

$$\begin{aligned} V_{\text{elstat}} &= \frac{1}{4\pi\varepsilon_0} \sum \frac{q_i q_j}{r_{ij}} \\ V_{\text{vdW}} &= \sum \varepsilon_{ij} \left[\left(\frac{R_{\min,ij}}{r_{ij}} \right)^{12} - \left(\frac{R_{\min,ij}}{r_{ij}} \right)^{6} \right] \end{aligned} \tag{5.2}$$

where the sums include all nonbonded atom pairs. q_i and q_j are the partial charges of the atoms i and j involved and ε_0 is the vacuum dielectric constant. For the van der Waals terms, the potential energy is expressed as a Lennard-Jones potential with well depth $\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}$ and range $R_{\min,ij} = (R_{\min,i} + R_{\min,j})/2$ at the Lennard-Jones minimum. This interaction captures long range dispersion ($\propto -r^{-6}$) and exchange repulsion ($\propto r^{-12}$) where the power of the latter is chosen for convenience. The combination of Eqs. 5.1 and 5.2 constitutes a minimal model for a force field which might be extended by using explicit terms for hydrogen bonds or for metal-containing systems [18].

An important step to investigate reactions by simulation methods has been the introduction of mixed quantum mechanical/classical mechanics methods (QM/MM) [19–21]. In QM/MM the total system is divided into a (small) reaction region for which the energy is calculated quantum mechanically and a (bulk) environment which is treated with a conventional force field. The majority of applications of QM/MM methods to date use semiempirical (such as AM1, PM3 [22], SCC-DFTB [23, 24]) or DFT methods [25] on isolated structures. Only a small number of studies including the dynamics of the nuclei have been reported [26–33]. Typically, the QM part contains of the order of several tens of atoms. It should also be noted that studies of reactive processes in the condensed phase often employ energy evaluations along pre-defined progression coordinates [22, 28], i.e., the system is forced to move along a set of more or less well suited coordinates. One of the main reasons why molecular orbital QM/MM calculations are not yet used routinely in fully quantitative studies is related to the fact that the evaluation of the intermolecular interactions in the QM region is computationally too expensive to allow meaningful configurational sampling which is required for reliably estimating essential quantities such as free energy changes. Alternatives to QM/MM methods have been developed whereby empirical force fields are used to investigate

chemical reactions through combination of their ability to describe large systems with the methods developed in quantum mechanics. They include RMD (Reactive molecular dynamics) [34–37], EVB (Empirical Valence Bond) [38] and its variants AVB (Approximate Valence Bond) [39], and MCMM (multiconfiguration molecular mechanics) [40].

5.2. CONCEPTUAL APPROACHES

In this section the techniques to investigate the energetics and dynamics of chemical reactions based on empirical force fields are discussed. Particular emphasis is put on the methods that allow to follow the rearrangements of atoms along the progression of a chemical reaction. What is not being discussed are nonadiabatic effects and quantum dynamics. For such topics we refer to reviews in the literature.

5.2.1. Molecular Mechanics with Proton Transfer

Proton transfer processes play a central role in maintaining metabolic processes in living organism. From a computational point of view, proton transfer processes pose some unique challenges due to the intrinsic quantum nature of the H atom and due to the fast motion of the proton. Ideally, such processes are best studied employing quantum mechanics except for the fact that the reaction coordinate for the proton transfer process is often sensitive to the environment surrounding of the H-donor and H-acceptor atoms, thus making quantum mechanical study prohibitively expensive. A popular alternative is to use QM/MM strategies where the reactive part is treated by quantum mechanics and the surrounding part by classical force fields. One variant of such a strategy has been developed by our group which combines a quantum chemically determined potential energy surface, suitable for describing proton transfer between a donor and an acceptor, and a force field to account for the remaining degrees of freedom [41–43]. The proton transfer event is then described by molecular dynamics simulations. The total interaction energy for the system with coordinates $\mathbf{Q}$ is given by the following form [41]

$$V(\mathbf{Q}) = V_{\mathrm{PT}}(R, r, \theta) + V_{\mathrm{MM}}(\mathbf{q}) \tag{5.3}$$

where the proton transfer motif, described by V_{PT}, is the QM determined potential energy surface along R (the distance between donor and acceptor atoms), r (the distance between donor and H atom), and θ (the angle between the unit vectors along R and r). The remaining degrees of freedom of the system ($\mathbf{q}$) are described by a classical force field. The resulting potential is called Molecular Mechanics for Proton Transfer (MMPT) [41]. In MD simulations with MMPT, the bonding pattern changes upon proton transfer. The algorithm is designed to add, modify, and remove force-field terms, that include bonded and non-bonded interactions, in a smooth and energy conserving fashion (by using appropriate switching functions,

such as $f_{sw}(R, r) = 0.5(\tanh(2R(r - R/2)) + 1)$, whenever the migrating H attempts to transfer from donor to acceptor [41].

MMPT treats the proton transfer process in full dimensionality while satisfying three important aspects of the problem: speed, accuracy, and versatility. While speed and accuracy are obtained from combining MM and QM techniques, respectively, the versatility of the approach is exploited by using the morphing potential method [44]. To this end, it is important to realize that without loss of generality, a wide range of proton transfer processes can be described by three prototype model systems: (a) symmetric single minimum (SSM, optimized structure of the system has equal sharing of the proton), (b) symmetric double-minimum (SDM, optimized structure of the system has unequal sharing of the proton but is symmetric with respect to the transition state), and (c) asymmetric double minimum (ADM, optimized structure of the system has unequal sharing of the proton and is asymmetric with respect to the transition state) [41]. The potential energy surface of these three model systems are fitted to generic forms and these zeroth order potential energy surfaces (SSM, SDM, or ADM) are morphed into a suitable PES to approximately reproduce important topological features of the target PES by a transformation of the type

$$V_{morph}(R', r', \theta') = \lambda(R, r, \theta)V_{orig}(R, r, \theta) \tag{5.4}$$

where λ can either be a constant or a more complicated function of one or more of the proton transfer coordinates. The morphing of PES approach not only avoids computing a full PES for the proton transfer motif but also reduces the rather laborious task of fitting an entirely new parametrized PES.

5.2.2. Reactive Molecular Dynamics

Reactions involving two potential energy surfaces where quantum effects such as tunneling and coherence are less important due to the adiabatic nature of the transitions, can be effectively studied by force-field based classical molecular dynamics simulations. The functional form of conventional force fields does not allow to break or form chemical bonds and is therefore unsuitable to describe reactive process. To overcome this limitation, our group developed a conceptually simple surface crossing approach to study reactive processes, involving bond-formation and bond-breaking, suitable for high-dimensional systems within classical MD simulations [36, 37, 45]. The algorithm, called adiabatic reactive molecular dynamics (ARMD) is implemented in CHARMM [14].

The algorithm involves two potential energy surfaces defined by two sets of force-field parameters corresponding to reactant (V_R) and product (V_P), differing by a number of force field parameters. For macromolecular systems the number of energy terms by which (V_R) and (V_P) differ is much smaller compared to the total number of energy terms. Thus, by providing only a smaller number of additional parameters compared to a standard MD simulation, it is possible to describe the

difference between the reactant and product with limited computational overhead [45]. Since the PES for the reactants and products differ by several force field terms, (V_R) and (V_P) differ by a constant Δ which corresponds to the asymptotic energy difference between the two states [45].

The dynamics of the system is initiated and propagated in the reactant state and the energy difference $\Delta E = V_R - V_P + \Delta$ is monitored at each time step. When ΔE changes sign, a crossing is detected. Since prior to crossing the system is propagated only on the reactant state, the mixing of surfaces induced by coupling is introduced a posteriori. This is achieved by restoring the configuration at $T_{mix}/2$ time earlier and the trajectory is restarted on an appropriately mixed surface for a T_{mix} period of time such that at the end of the mixing, the system smoothly approaches the product state [45]. The choice of mixing time, typically a of the order of 10 fs, has only a small effect on the resulting trajectories [37, 45]. During crossing, the energies and forces from the newly created or broken bonded terms are added or subtracted and the corresponding nonbonded interactions are updated.

Apart from the parametrization of the potential energy surfaces – which should be "validated" vis-a-vis experimental data or high level ab initio calculations – ARMD has only one free parameter: the asymptotic energy difference Δ. In multi dimensional potential energy surfaces of the reactant and product, Δ is expected to vary with individual degrees of freedom. Since determining the contribution of each degree of freedom to Δ is not possible, Δ is considered as an averaged quantity. For a given problem, Δ is only known approximately either from ab initio calculations or fitting to experimental observables such as reaction rates or product yield. Therefore it is advisable to carry out test simulations to determine the sensitivity of the calculated observables on the value of Δ used.

5.2.3. Empirical Valence Bond

One of the established methods to investigate chemical reactions based on empirical force fields is the empirical valence bond (EVB) method [38, 46]. EVB starts from the realization that valence bond states are suitable to distinguish between ionic and covalent resonance forms of a chemical bond that is close to chemical intuition. Since the environment in which a chemical reaction takes place interacts primarily through electrostatics with the reactive species, empirical force fields can be used to describe the resonant forms of the product and the educt. In a simple bond-breaking reaction between fragments A and B that constitute the molecule AB, three resonance forms are introduced: $\psi_1 = AB$, $\psi_2 = A^-B^+$, and $\psi_3 = A^+B^-$. If A is more electronegative than B, resonance structures Ψ_1 and Ψ_2 are most relevant. For a collection of covalent and ionic states, matrix elements for the EVB Hamiltonian have to be determined. They include diagonal elements for the covalent and ionic states, and off-diagonal elements that couple configurations (bonding patterns) that differ by the presence of one bond. All other off-diagonal matrix elements $H_{ij} = 0$. The justification for this is that such matrix elements are

proportional to the square or higher powers of the overlap between atomic orbitals, but they may also be retained [38, 47]. The covalent diagonal matrix elements H_{ii} correspond essentially to an empirical force field, whereas for the ionic diagonal matrix elements the bonded terms are replaced by electrostatic interactions between the charged fragments and the formation energy of A^-B^+ from AB has also to be added. For the two-fragment system AB the matrix elements are $H_{11} = D_e(1 - \exp[-\beta(r - r_e)])^2$ and $H_{22} = \Delta - \frac{e^2}{r} + V_{nb}$, where Δ is the gas-phase formation energy of A^-B^+ from AB at infinite separation, and V_{nb} is the nonbonded interaction potential such that the minimum of $(\frac{-e^2}{r} + V_{nb})$ is given by the sums of the ionic radii of A^+ and B^-. In the original version of EVB the off-diagonal element $H_{12} = H_{21}$ is determined through the requirement that the eigenvalues of the Hamiltonian E satisfy the relation $H_{12} = \sqrt{(H_{11} - E)(H_{22} - E)}$ and E is the experimentally determined ground-state bond energy. In a later and slightly more general approach, the off-diagonal elements are parametrized functions of the form $H_{ij} = A\exp(-\mu(r - r_0))$ [46]. The definition of the off-diagonal terms has been a source of considerable discussion in the field, in particular the assumption that upon transfer of the reaction from the gas phase to the solution phase these elements do not change significantly. This assumption has been recently tested [46]. Alternative forms which also capture the shape and energetics of the potential energy surface around the transition state use generalized Gaussians [48]. A comparison of different diabatic models has been recently given and provides a notion of the common features and the differences between various approaches [49–51]. A useful comparison of the similarities and differences between the various methods can be found in [52]. Applications of EVB are widespread and include enzymatic reactions (for which it was originally developed [53]), proton transfer processes (see Section 5.3.1), and the autodissociation of water [54].

5.2.4. ReaxFF

Starting from Pauling's realization that bond order and bond length are related [55], a bond energy bond order (BEBO) potential was developed by Johnston and Parr [8]. It was found that, in addition to the nearly linear relationship between bond order and bond length [55], a log-log plot of dissociation energies against bond order is also almost linear. This approach yielded activation energies within ≈2 kcal/mol and rate constants within an order of magnitude for reactants with well known bond energies. One of the essential assumptions underlying this approach is that – at least for hydrogen-atom transfer reactions, the sum of the bond orders n_1 of the breaking and the newly formed bond n_2 is unity, i.e. $n_1 + n_2 = 1$. Or in the words of the authors, "At all stages of the reaction the formation of the second bond must be 'paying for' the breaking of the first bond" [8].

A more general method that is based on the concept of bond order and its relationship to bond length and bond energy is ReaxFF [9]. In this force field van

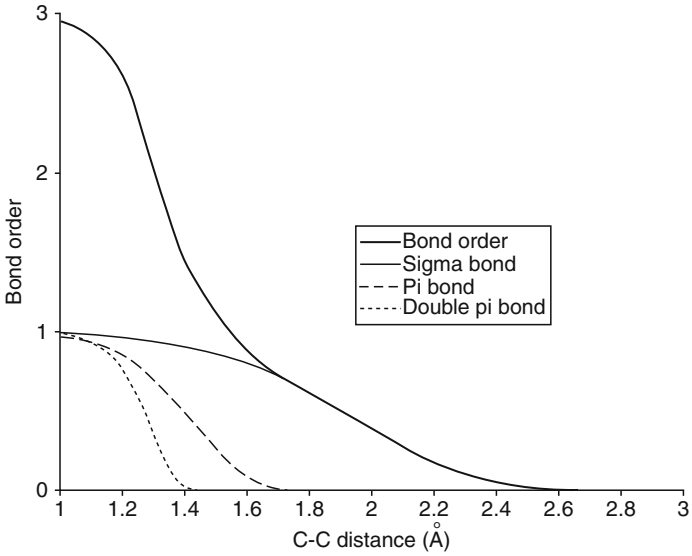


Figure 5-1. The relationship between C–C bond distances and bond orders (The figure is taken from [9])

der Waals and Coulomb terms are included from the beginning and the dissociation and reaction curves are derived from electronic structure calculations. Central to ReaxFF is that the bond order can be calculated from the distance between two atoms. This is illustrated in Figure 5-1 which shows that the bond order is a real number that varies between 0 and 3 for a carbon–carbon bond. This expresses the fact that two carbon atoms can be found to form anything in between "no bond" (bond order = 0) to triple bond. From the bond order the bonded energy term E_{bond} is calculated. To correct for over-coordination, a penalty term E_{over} is added to ReaxFF and for under-coordinated atoms additional favorable energy terms E_{under} reflecting resonance energies between π–electrons are introduced. A last, non-standard term usually not present in conventional force fields is the conjugation energy E_{conj}. With these terms, the total potential energy in ReaxFF can be written as

$$E = E_{\text{bond}} + E_{\text{over}} + E_{\text{under}} + E_{\text{conj}} + E_{\text{val}} + E_{\text{pen}} + E_{\text{dihe}} + E_{\text{VdW}} + E_{\text{coul}} \qquad (5.5)$$

Here, E_{val}, E_{dihe}, E_{VdW} and E_{coul} are the well-known valence-angle, dihedral, Van der Waals, and electrostatic terms, whereas E_{pen} reproduces the stability of systems with two double bonds sharing an atom in a valence angle. All energy terms and their parametrizations are explained in detail in [9].

Illustrative applications of ReaxFF range from the study of shock-induced chemistry in high energy materials [56] to activation and dissociation of H_2 on platinum surfaces [57] and the oxidation of nanoparticles on aluminum surfaces [58]. Using ReaxFF and nonequilibrium MD simulations it was found that

depending on the impact velocities cyclic-$[CH_2N(NO_3)]_3$ decomposes into a variety of small molecules on the picosecond time scale or only into NO_2, both of which are consistent with experiments [56]. Such simulations provide considerable insight into the time dependence of concentration changes of particular species.

5.2.5. Other Approaches

Multiconfiguration Molecular Mechanics (MCMM) More recently, a procedure that is parametrized entirely with respect to information from ab initio calculations has been put forward. It was termed multiconfiguration molecular mechanics (MCMM). In the light of other existing algorithms MCMM is probably best viewed as a particular variant of EVB [49–52]. One of the particular features of MCMM is that it uses Shepard interpolation to represent the off-diagonal matrix element H_{12} (see also section on EVB). A stringent comparison of MCMM with EVB still has not been presented.

Approximate Valence Bond (AVB) An approach that is based on parametrizing EVB entirely from electronic structure calculations is AVB [39]. It has been primarily used for the investigation of proton transfer reactions in enzymes, such as the hydrolytic activity of phospholipase A_2 [59].

Quenching dynamics Instead of explicitly breaking and forming chemical bonds the possibility has been explored to approximately locate the transition state between the reactant and the product states of a system and subsequently use quenching (downhill) dynamics to relax the system. Such an approach was employed to investigate the rebinding of CO in myoglobin [35]. More generally, the approach is reminiscent of using an interpolating Hamiltonian as in the theory of electron transfer [60].

5.3. APPLICATIONS

5.3.1. Proton Transfer Reactions

5.3.1.1. Protonated Water Dimer ($O_2H_5^+$)

The protonated water dimer has received substantial attention from both experiment [61–64] and theory [33, 65–68] alike, discussing its structural and dynamic behavior. The MMPT potential described in the previous section has been used to investigate proton transfer dynamics and to characterize infrared spectra of $O_2H_5^+$ [69].

The relaxed PES of $O_2H_5^+$ was obtained by scanning the collinear conformations of the proton transfer motif over a grid along R, r dimensions employing MP2/6-311++G(d,p) level of electron correlation [69]. The resulting surface belongs to the symmetric-single-minimum (SSM) category described in the previous section was fitted to a generic form involving a combination of Morse potentials [69]. The

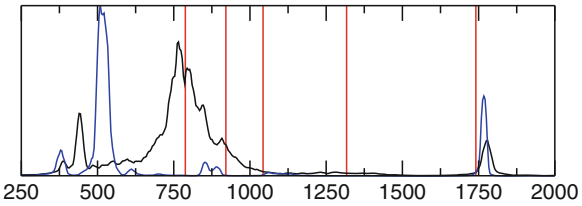


Figure 5-2. Averaged infrared spectra for $(O_2H_5^+)$ (black) and $(O_2DH_4^+)$ (blue) at 50 K together with the experimental line positions (red lines) (The figure is taken from [69])

resulting potentials together with conventional force-field terms for the rest of the system are then used for MD simulations at different temperatures. This MMPT potential allows to investigate long time (several 100 ps) bond breaking and bond formation processes [69].

MD simulations provide time series of coordinates and dipole moments of the system which are then used to obtain vibrational spectra by Fourier transforming the dipole-dipole autocorrelation functions. For calculation of spectra, the 15-dimensional dipole moment surface of Huang et al. [70] was used. The spectral lines can be assigned to particular motion using normal-mode analysis and by comparing the spectra with isotopically substituted molecules [69]. The averaged infrared spectra for $O_2H_5^+$ and $O_2DH_4^+$ at 50 K together with the experimental line positions are shown in Figure 5-2. Most experimental line positions, for example the absorptions at 750, 900, 1330, and 1770 cm^{-1}, correspond to features in the calculated infrared spectra. Analyzing power spectra associated with different degrees of freedom, it was established that motion along O–H* coordinate involved in the O–H*–O asymmetric stretch is coupled to the O–O stretching and the OH*O and HOH* bending coordinates [69].

Vibrational excitations estimated from other calculations are well reproduced by the MD simulations with MMPT thus validating the approach [69]. In particular, the MD simulations by Bowman and co-workers [70] on a high-quality 15-dimensional surface find the O–H*–O stretching vibration at 860 cm^{-1} which agrees well with 830 cm^{-1}, calculated with the MMPT potential.

5.3.1.2. Protonated Ammonia Dimer $(N_2H_7^+)$

Unlike protonated water dimer, the protonated ammonia dimer represents a symmetric double minima (SDM) proton transfer where the transfer of the proton from one ammonia to another involves a barrier. Earlier computational studies on protonated ammonia dimer include semiempirical calculations, EVB methods as well as approximate DFT (SCC-DFTB) [27, 71–73].

Employing the same strategy described for protonated water dimer, the MMPT potential was constructed for protonated ammonia dimer which was then used for MD simulations. The time series of the distance between donor and acceptor atoms and donor and H atoms in an example trajectory is shown in Figure 5-3. Most proton

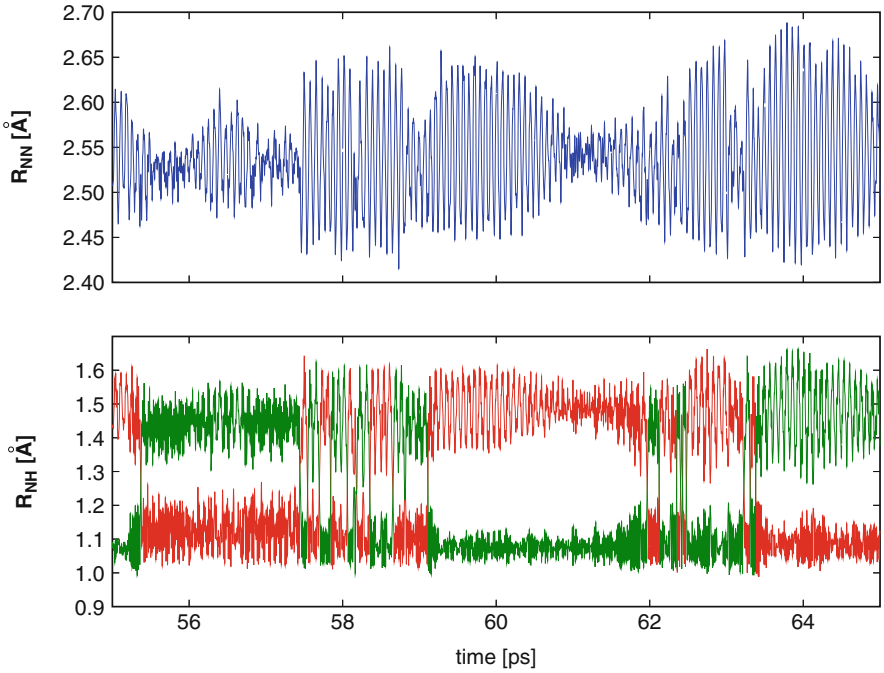


Figure 5-3. Typical time series for donor-acceptor (R_{NN}) and donor/acceptor-H (R_{NH}) distances in $NH_4^+ \cdot NH_3$ from a trajectory at 300 K and with scaling $\lambda E = 1.0$. The top panel shows the distance fluctuations of the donor-acceptor distance. The bottom panel shows distances between the nitrogen atoms and the proton, N1 → N1 shown in green and N2 → N2 shown in red. Residence times range from subpicoseconds to multiple picoseconds (The figure is taken from [41])

transfer events are found if the D-A distance is shorter than 2.45 Å. During 10 ps several such events can be observed with duration between less than 1 ps up to about 3 ps.

The temperature dependence of the proton transfer process, multiple MD simulations at several temperatures between 30 and 300 K were carried out and the averaged temperature-dependent rates were determined. For the SDM potential, no proton transfer was seen below 40 K whereas between 40 and 150 K a nearly linear increase in the rate is found, beyond which the rate remains constant, see Figure 5-4. To further characterize the rate dependence on the barrier separating the reactant and product state, simulations were carried out on morphed PES where the morphing parameter lambda in Eq. (5.4) was chosen as 0.5, 0.75, and 2.0. The onset of proton transfer is seen at lower temperatures for potentials with lower barrier, while the high-temperature limit of the proton transfer rate remains unchanged, suggesting a property common to all surfaces control the maximum possible rate. Proton transfer probabilities of 8.8–10 transfers/ps are found for all barriers investigated. This corresponds to a frequency range associated with the symmetric stretch vibration of the donor and acceptor atoms, which acts as a gating mode [41].

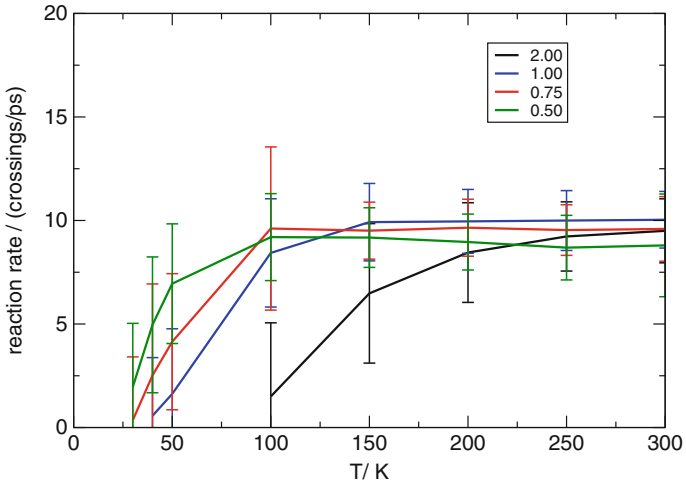


Figure 5-4. Proton transfer rates with typical error bars as a function of T for $NH_4^+ \cdot NH_3$. Rates are given for different energy scaling factors λE (proton-transfer barrier heights): $\lambda E = 2.0$ (black), 1.0 (blue), 0.75 (red), and 0.5 (green). For high temperatures the energy scaling has a smaller effect on the barrier than for low temperatures. This also explains why the rates are similar for $T \approx 300$ K (The figure is taken from [41])

5.3.2. Ligand Binding in Heme Proteins

5.3.2.1. Rebinding Dynamics in MbNO

Myoglobin, besides being an important model system for understanding the relation between structure and function of proteins, has also been of interest due to its ligand binding properties. In particular, the migration pathways and rebinding dynamics of diatomic ligands such as O_2, NO, and CO inside the protein matrix have been studied by both experimental and computational methods. While rebinding of CO is nonexponential at low temperature but becomes exponential at high temperature with a time scale of 100 ns, rebinding of NO remains nonexponential at all temperature with time constants of the order of tens of picoseconds [74, 75].

The rebinding dynamics of MbNO was studied employing the ARMD method. To this end, two force-fields were prepared corresponding to the bound and dissociated states, differing in a number of energy terms. The dissociating Fe-N bond was described by a Morse potential to describe the anharmonic nature of the bond. Multiple trajectory simulations were carried out for $\Delta = 60$, 65, and 70 kcal/mol [45].

To visualize the crossing seam all observed crossing geometries are projected onto a plane containing the Fe–N distance and the angle formed by Fe and the ligand NO. The most probable iron–ligand distance lies around 3 Å [45]. The distribution is rather wide along the bond angle coordinate and the crossing seam is found to be rather insensitive towards the value of Δ, see Figure 5-5.

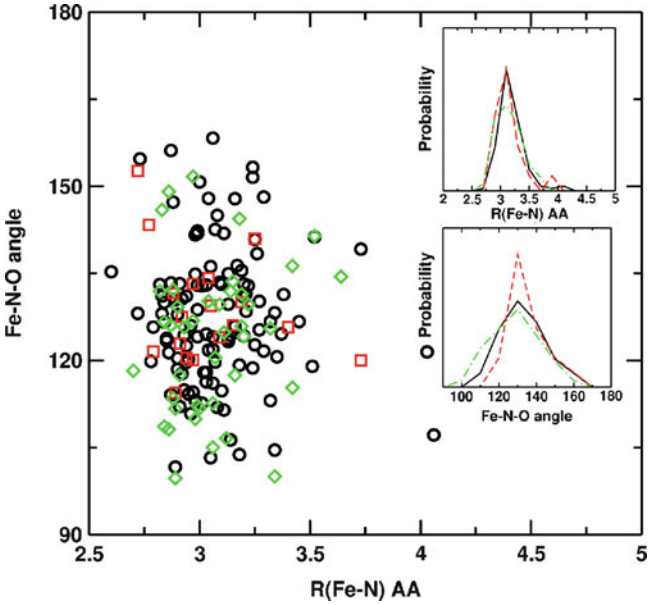


Figure 5-5. The crossing geometries of NO rebinding to Mb with three different values of Δ (red squares $\Delta = 60$, black circles $\Delta = 65$, and green diamonds $\Delta = 70$ kcal/mol) projected onto the distance between iron and the nitrogen of the ligand R_{FeN} and the angle between iron and the NO ligand θ_{FeNO}. The insets show the distributions of crossing points along the R_{FeN} (upper) and θ_{FeNO} (lower) coordinates. For the insets, broken, solid, and dash-dotted lines denote Δ = 60, 65, and 70 kcal/ mol, respectively (The figure is taken from [45])

The time series of fraction of trajectories without showing crossing provides information about the kinetics of rebinding. The choice of Δ is found to have a substantial effect on the time constant associated with the rebinding reaction, although for all values of Δ the rebinding remains nonexponential. For Δ = 65 kcal/mol, the time constants are found to be 3.6 and 373 ps [45] compared to the experimental (from ultrafast IR spectroscopy) value of 5.3 and 133 ps [75, 76]. While the fast rebinding component is well reproduced by ARMD, the agreement for the slower component is poor which arises due to an insufficient sampling of the slow time scale by ARMD.

5.3.2.2. NO Detoxification Reaction in Truncated Hemoglobin (trHbN)

Truncated hemoglobin is a recently discovered heme protein found in plants, bacteria, and lower eukaryots. The trHbN of Mycobacterium tuberculosis has been proposed to play an important role in the survival of the bacteria causing tuberculosis in host cells by converting toxic NO to harmless NO_3^-. Although the vertebrate Hb can perform the same reaction, the trHbN does it at least 20 times

faster, with a second order rate constant of $7.5 \times 10^8\ M^{-1}\ s^{-1}$ [77]. The fast reaction has been attributed to the presence of a continuous tunnel inside protein matrix assisting ligand migration and due to the presence of ligand stabilizing residues in the active site of the protein [77, 78]. However, a clear understanding about the mechanism of the detoxification reaction had remained illusive.

We have employed ARMD method to shed light on the reaction by dividing the overall reaction into following four steps [79]:

$$\mathrm{Fe(II)} - \mathrm{O_2} + \mathrm{NO} \rightarrow \mathrm{Fe(III)[-OONO]} \quad \mathrm{O2-N\ distance}, \tag{I}$$

$$\mathrm{Fe(III)[-OONO]} \rightarrow \mathrm{Fe(IV)} = \mathrm{O} + \mathrm{NO_2} \quad \mathrm{O1-O2\,distance}, \tag{II}$$

$$\mathrm{Fe(IV)} = \mathrm{O} + \mathrm{NO_2} \rightarrow \mathrm{Fe(III)[-ONO_2]} \quad \mathrm{O1-N\ distance}, \tag{III}$$

$$\mathrm{Fe(III)[-ONO_2]} \rightarrow \mathrm{Fe^+(III)} + \mathrm{[NO_3]^-} \quad \mathrm{Fe-O1\,distance}. \tag{IV}$$

The right-hand side column of the above reaction steps indicates the bond broken/formed in the reaction step. In the first step oxy-trHbN reacts with free NO and forms a peroxynitrite intermediate complex which then undergoes homolytic fission followed by the rebinding of free NO_2 to the oxo-ferryl species to form the heme-bound nitrato complex which then undergoes heme-ligand dissociation resulting in free NO_3^- and penta-coordinated heme.

The force field parameters associated with the reactants and products of each of the reaction steps are obtained from ab initio calculations. Each of the reaction step was then studied by running multiple ARMD trajectories with a range of Δ values [80]. For reaction steps I, III, and IV, the ARMD simulations yielded rate constants on the picosecond time scale. The choice of the free parameter Δ had only limited effects on the timescale of the reaction. The crossing seam of the reaction steps were found to be broad in phase space and were found to be largely independent of the choice of Δ, as was the case with MbNO.

However, for step II of the detoxification reaction, ARMD simulations could not find crossing between the reactant and products despite running simulations of the order of nanoseconds [80]. From DFT calculations, it is known that this reaction step involves a barrier of 6.7 kcal/mol [81]. To further investigate the activation energy associated with the reaction step, free-energy profiles for the reaction step II were calculated using umbrella sampling method which yielded a barrier of 12–15 kcal/mol. A barrier of 6–12 kcal/mol corresponds to timescale on the order of micro- to milli-second, assuming an entirely diffusion-limited reaction to be occurring in the picoseconds time scale [82, 83]. Simulating such a slow process is certainly computationally demanding. Moreover, since the overall reaction is quite fast (order of picoseconds), it is unlikely that the reaction goes via step II. This proposition is in line with the lack of experimental detection of free NO_2 radical in several studies which propose an alternative mechanism where peroxynitrite intermediate rearranges to nitrato complex [84, 85]. We carried out ARMD simulations for this rearrangement reaction and found the reaction occurring in the time scale of picoseconds, explaining the fast overall detoxification reaction [80].

One of the reasons attributed to the fast NO detoxification reaction in trHbN compared to vertebrate Hb, is the presence of ligand stabilizing residues (Tyr33 and Gln58) in the active site of the former [77, 78, 86]. The role of these residues in detoxification reaction is investigated using ARMD simulations for mutants of the protein. While the rate constants of step I and IV were found to be little affected by the mutations, step III and rearrangement reactions were slowed by an order of magnitude with respect to their wild-type values. In particular, ARMD simulations find that Tyr33, which is a conserved residue along the family of truncated hemoglobins, plays an important role in facilitating the reaction in trHbN [80].

5.3.2.3. Conformational Equilibrium in Neuroglobin

Neuroglobin is a recently discovered member of the globin protein family expressed in the nerve cells of vertebrates [87]. Unlike Myoglobin, the unligated neuroglobin is hexacoordinated, with active-site histidines acting as axial ligands [88]. However, the X-ray experiments found that the measured electron density in the active site could not be fitted with a single conformation. Using a second conformation with 30% occupancy and a different heme orientation the quality of the fitting was considerably improved. The biological significance for these almost degenerate conformations is unknown, although NMR experiments indicate that binding of CN^- is significantly slower in one of the conformations [89].

To study the conformational equilibrium between two states via ARMD, two separate potential energy surfaces were prepared and Δ was chosen to reproduce the correct occupation of the two states [45]. The barrier for the conformational transition is found to be high for a free dynamics simulation, instead umbrella sampling was chosen as a method to guide the transition. ARMD simulations were carried out for a range of Δ values between -3 to -6.2 kcal/mol and relative population of the two conformations was determined. It is found that $\Delta = -3.3$ kcal/mol provides the experimentally observed population [45]. From the crossings obtained with $\Delta = -3.3$ kcal/mol, a free energy profile was determined, see Figure 5-6. The free energy difference of 1.1 ± 0.25 kcal/mol between the two minima is in quite good agreement with the value derived from experiment (1.4 kcal/mol) [88].

5.4. OUTLOOK

This outlook summarizes some new and expected future developments in investigating reactions using force fields.

Improved and purpose-built force fields Most force fields for biomolecular simulations have been developed with the purpose to be as universal as possible for particular applications such as the simulation of proteins or DNA. As such, these force fields have proven to be reliable and robust under most circumstances. However, improved results can be gained by modifying the underlying parameters and by incorporating more elaborate models to describe the systems of interest.

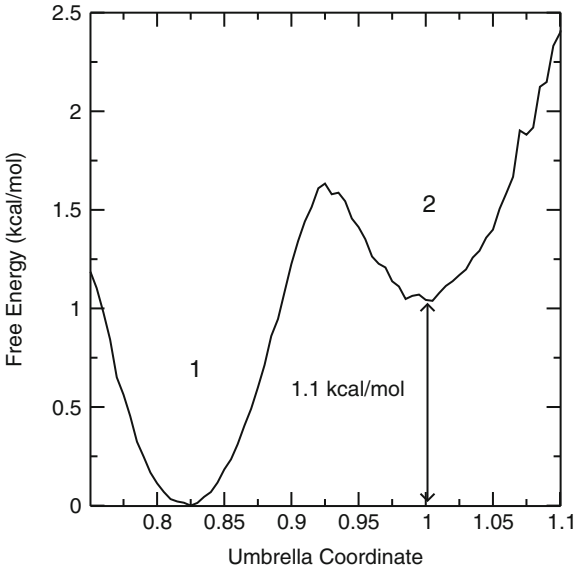


Figure 5-6. Free energy profile along umbrella coordinate in neuroglobin from umbrella sampling calculations with $\Delta = -3.3$ kcal/mol. The higher energy state 2 is destabilized by 1.1 kcal/mol relative to the more stable state 1 (The figure is taken from [45])

Recent improvements in force fields include effects that go beyond point charge models and include multipolar force fields [90–95]. Furthermore, empirical force fields are usually fitted to structural, thermodynamic and/or spectroscopic data – primarily from vibrational spectroscopy in the harmonic approximation. Improved fitting, which uses additional data sensitive to the intermolecular dynamics, such as NMR (H-bonded network), anharmonic dynamics, or relaxation of vibrational excitations, will lead to more reliable force fields. Initial steps in this direction have already been undertaken [95–97].

The definition of reaction coordinates Many schemes discussed in the present chapter require – in one form or another – the definition of a progression coordinate. A suitable progression coordinate varies between a minimal and a maximal value which characterize the educt and the product, respectively, whereas intermediate values correspond to transition states. Commonly used reaction coordinates are often a single internal coordinate or a combination of them. For condensed phase reaction dynamics, finding a suitable reaction coordinate in this way is a formidable challenge. An alternative approach which is coordinate free is to use the energy gap between the reactant and product as the essential reaction coordinate. This is realized in EVB as well as in RMD methods. Improved schemes to define reaction coordinates will greatly assist in removing ambiguity and bias (introduced by resorting to "chemical intuition") from current simulations. One attractive possibility is to combine a coordinate-free approach, with schemes to extract suitable

combinations of internal coordinates for defining reaction coordinates [98]. With such an approach it will then also be possible to re-score the energetic profile along reaction pathways by using more accurate methods.

Extensions to ARMD Currently, adiabatic reactive MD correctly interpolates between initial and final state of an elementary step. Although the potential energy surfaces are individually validated for the product and reactant states, the correct shape of the PES around the transition state is not guaranteed by this procedure. Including geometrical and energetic properties of the transition state will greatly extend the validity of ARMD. In its current implementation, ARMD uses one (conformationally averaged) value Δ for the asymptotic separation of the product and reactant state. However, it is known that the asymptotic energy difference between the two states can also change as a function of the relative orientation of the reaction partners. Thus, using a conformationally dependent Δ will allow to model reactions more realistically.

This work is supported by the Schweizerischer Nationalfonds. Special thanks go to Jonas Danielsson, Sven Lammers, Stephan Lutz, and David Nutt. The major part of this review is based on results obtained in their work. The authors thank Professor Arieh Warshel for useful comments on the manuscript.

REFERENCES

1. London F (1929) Z Elektrochem 35:552–555
2. Eyring H, Polanyi M (1931) Z Phys Chem Abt B 12:279–311
3. Sato S (1955) J Chem Phys 23:592–593
4. Sato S (1955) J Chem Phys 23:2465–2466
5. Ellison FO (1963) J Am Chem Soc 85:3540–3544
6. Pauling L (1960) The nature of the chemical bond. Cornell University Press, Ithaca, NY
7. Pauling L (1932) J Am Chem Soc 54:3570–3582
8. Johnston HS, Parr C (1963) J Am Chem Soc 85:2544–2551
9. van Duin ACT, Dasgupta S, Lorant F, Goddard WA III (2001) J Phys Chem A 105:9396–9409
10. Lifson S, Warshel A (1968) J Chem Phys 49:5116–5129
11. Levitt M, Lifson S (1969) J Mol Biol 46:269–279
12. Hwang MJ, Stockfisch TP, Hagler AT (1994) J Am Chem Soc 116:2515–2525
13. Maple JR, Hwang MJ, Stockfisch TP, Dinur U, Waldman M, Ewig CS, Hagler AT (1994) J Comput Chem 15:162–182
14. Brooks B, Bruccoleri R, Olafson B, States D, Swaminathan S, Karplus M (1983) J Comput Chem 4:18–217
15. Weiner SJ, Kollman PA, Case DA, Singh U, Ghio C, Alagona G, Profeta S Jr, Weiner P (1984) J Am Chem Soc 106:765–784
16. Jorgensen WL, Tirado-Rives J (1988) J Am Chem Soc 110:1657–1666
17. Hermans J, Berendsen HJC, van Gunsteren WF, Postma JPM (1984) Biopolymers 23:1
18. Deeth R, Anastasi A, Diedrich C, Randell K (2009) Coord Chem Rev 253:795–816
19. Warshel A, Levitt M (1976) J Mol Biol 103:227–249
20. Alagona G, Ghio C, Kollman PA (1986) J Mol Biol 191:23–27
21. Bash PA, Field MJ, Karplus M (1987) J Am Chem Soc 109:8092–8094

22. Claeyssens F, Ranaghan KE, Manby FR, Harvey JN, Mulholland AJ (2005) Chem Commun 40:5068–5070
23. Zhou HY, Tajkhorshid E, Frauenheim T, Suhai S, Elstner M (2002) Chem Phys 277:91–103
24. Konig PH, Ghosh N, Hoffmann M, Elstner M, Tajkhorshid E, Frauenheim T, Cui Q (2006) J Phys Chem A 110:548–563
25. Altun A, Guallar V, Friesner RA, Shaik S, Thiel W (2006) J Am Chem Soc 128:3924–3925
26. Mei HS, Tuckerman ME, Sagnella DE, Klein ML (1998) J Phys Chem B 102:10446–10458
27. Meuwly M, Karplus M (2002) J Chem Phys 116:2572–2585
28. Cui Q, Elstner T, Karplus M (2003) J Phys Chem B 106:2721–2740
29. Meuwly M, Müller A, Leutwyler S (2003) Phys Chem Chem Phys 5:2663–2672
30. Devi-Kesavan LS, Gao J (2003) J Am Chem Soc 125:1532–1540
31. Zoete V, Meuwly M (2004) J Chem Phys 120:7085–7094
32. Xu D, Wei Y, Wu J, Dunaway-Mariano D, Guo H, Cui Q, Gao J (2004) J Am Chem Soc 126:13649–13658
33. Sauer J, Doebler J (2005) ChemPhysChem 6:1706–1710
34. Li H, Elber R, Straub JE (1993) J Biol Chem 268:17908–17916
35. Zheng C, Makarov V, Wolynes PG (1996) J Am Chem Soc 118:2818–2824
36. Meuwly M, Becker OM, Stone R, Karplus M (2002) Biophys Chem 98:183–207
37. Nutt D, Meuwly M (2006) Biophys J 90:1191–1201
38. Warshel A, Weiss RM (1980) J Am Chem Soc 102:6218–6226
39. Grochowski P, Lesyng B, Bala P, McCammon JA (1996) Int J Quant Chem 60:1143–1164
40. Kim Y, Corchado JC, Villa J, Xing J, Truhlar DG (2000) J Chem Phys 112:2718–2735
41. Lammers S, Lutz S, Meuwly M (2008) J Comput Chem 29:1048–1063
42. Lammers S (2006) Simulations of proton transfer processes using reactive force fields. Ph.D. thesis, University of Basel
43. Lammers S, Meuwly M (2004) Aust J Chem 57:1223–1228
44. Meuwly M, Huston J (1999) J Chem Phys 110:8338–8347
45. Danielsson J, Meuwly M (2008) J Chem Theory Comput 4:1083–1093
46. Hong G, Rosta E, Warshel A (2006) J Phys Chem B 110:19570–19574
47. Coulson CA, Danielsson U (1954) Ark Fys 8:239–244
48. Chang YT, Miller WH (1990) J Phys Chem 94:5884–5888
49. Valero R, Song L, Gao J, Truhlar DG (2009) J Chem Theory Comput 5:1–22
50. Valero R, Song L, Gao J, Truhlar DG (2009) J Chem Theory Comput 5:2191–2191
51. Kamerlin SCL, Cao J, Rosta E, Warshel A (2009) J Phys Chem B 113:10905–10915
52. Florian J (2002) J Phys Chem A 106:5046–5047
53. Warshel A (1984) Proc Nat Acad Sci USA 81:444–448
54. Strajbl M, Hong G, Warshel A (2002) J Phys Chem B 106:13333–13343
55. Pauling L (1947) J Am Chem Soc 69:542–553
56. Strachan A, van Duin ACT, Chakraborty D, Dasgupta S, Goddard WA III (2003) Phys Rev Lett 91:098301–1
57. Ludwig J, Vlachos DG, van Duin ACT, Goddard WA III (2006) J Phys Chem B 110:4274–4282
58. Vashishta P, Kalia RK, Nakano A (2006) J Phys Chem B 110:3727–3733
59. Bala P, Grochowski P, Nowinski K, Lesyng B, McCammon JA (2000) Biophys J 79:1253–1262
60. Marcus RA (1993) Angew Chem Int Ed 32:1111–1121
61. Asmis KR, Pivonka NL, Santambrogio G, Bruemmer M, Kaposta C, Neumark DM, Woeste L (2003) Science 299:1375–1377
62. Fridgen TD, McMahon GB, MacAleese L, Lemaire J, Maitre P (2004) J Phys Chem A 108:9008–9010

63. Moore DR, Oomens J, van der Meer L, von Helden G, Meijer G, Valle J, Marshall AG, Eyler JR (2004) ChemPhysChem 5:740–743
64. Hammer NI, Diken EG, Roscioli JR, Johnson MA, Myshakin EM, Jordan KD, McCoy AB, Huang X, Bowman JM, Carter S (2005) J Chem Phys 122:244301
65. Dai J, Bacic Z, Huang X, Carter S, Bowman JM (2003) J Chem Phys 119:6571–6580
66. Kaledin M, Kaledin AL, Bowman JM (2006) J Phys Chem A 110:2933–2939
67. Lobaugh J, Voth GA (1996) J Chem Phys 104:2056–2059
68. Cho HM, Singer SJ (2004) J Phys Chem A 108:8691–8702
69. Lammers S, Meuwly M (2007) J Phys Chem A 111:1638–1647
70. Huang S, Braams BJ, Bowman JM (2005) J Chem Phys 122:044308
71. Asada T, Haraguchi H, Kitaura K (2001) J Phys Chem A 105:7423–7428
72. Li GS, Costa MTCM, Millot C, Ruiz-Lopez MF (1999) Chem Phys 240:93–99
73. Wang Y, Gunn JR (1999) Int J Quant Chem 73:357–367
74. Austin RH, Beeson KW, Eisenstein L, Frauenfelder H, Gunsalus IC (1975) Biochemistry 14:5355–5373
75. Kim S, Jin G, Lim M (2004) J Phys Chem B 108:20366–20375
76. Petrich JW, Lambry JC, Kuczera K, Karplus M, Poyart C, Martin JL (1991) Biochemistry 30:3975–3987
77. Ouellet H, Ouellet Y, Richard C, Labarre M, Wittenberg B, Wittenberg J, Guertin M (2002) Proc Nat Acad Sci USA 99:5902–5907
78. Milani M, Pesce A, Ouellet Y, Dewilde S, Friedman J, Ascenzi P, Guertin M, Bolognesi M (2004) J Biol Chem 279:21520–21525
79. Bourassa JL, Ives EL, Marqueling AL, Shimanovich R, Groves JT (2001) J Am Chem Soc 123:5142–5143
80. Mishra S, Meuwly M (2010) Peroxide bond cleavage versus rearrangement reaction in nitric oxide detoxification: a reactive molecular dynamics study of truncated hemoglobin J Am Chem Soc 132:2968
81. Blomberg LM, Blomberg MRA, Siegbahn PEM (2004) J Biol Inorg Chem 9:923–935
82. Nauser T, Koppenol WH (2002) J Phys Chem A 106:4084–4086
83. van Holde KE (2002) Biophys Chem 101:249–254
84. Herold S, Exner M, Nauser T (2001) Biochemistry 40:3385–3395
85. Herold S (1999) FEBS Lett 443:81–84
86. Mishra S, Meuwly M (2009) Biophys J 96:2105–2118
87. Burmester T, Weich B, Reinhardt S, Hankeln T (2000) Nature 407:520
88. Vallone B, Nienhaus K, Brunori M, Nienhaus G (2004) Proteins 56:85
89. Du W, Syvitski R, Dewilde S, Moens L, La Mar G (2003) J Am Chem Soc 125:8080
90. Price SL (1996) J Chem Soc Faraday Trans 92:2997
91. Karamertzanis PG, Price SL (2006) J Chem Theory Comput 2:1184
92. Plattner N, Meuwly M (2008) Biophys J 94:2505–2515
93. Plattner N, Bandi T, Doll J, Freeman DL, Meuwly M (2008) Mol Phys 106:1675
94. Plattner N, Meuwly M (2008) ChemPhysChem 9:1271
95. Devereux M, Meuwly M (2009) J Phys Chem B 113:13061
96. Schmid FFF, Meuwly M (2008) J Chem Theory Comput 4:1949
97. Huang J, Meuwly M (2010) J Chem Theory Comput 6:467
98. Ma A, Dinner AR (2005) J Phys Chem B 109:6769–6779

CHAPTER 6

THEORETICAL STUDIES OF POLYMERISATION REACTIONS

MAREK CYPRYK AND GRZEGORZ KRASIŃSKI
Centre of Molecular and Macromolecular Studies, Polish Academy of Sciences, Sienkiewicza 112, 90-363 Łódź, Poland
e-mail: mcypryk@cbmm.lodz.pl

Abstract: In this review, the applicability of computational chemistry procedures for modelling of macromolecules and large macromolecular systems is shortly reviewed. In particular, the recent achievements in studying the reactions leading to polymers by quantum chemical and molecular mechanical methods are presented.

Keywords: Molecular mechanics, Semiempirical methods, Molecular dynamics, Ab initio, Density functional theory, Hybrid methods, QM/MM, Polymerisation mechanism, Radical polymerisation, Ionic polymerisation, Ring-opening polymerisation

6.1. INTRODUCTION

Polymers are macromolecules built up by linking together a large number of small molecules, called monomers. The reactions by which they combine with each other are termed polymerisations. What makes macromolecules different from small molecules, such as ethylene or styrene, is their size. A typical polymer chain may consist of several thousand atoms which are connected together by the same types of bonding that occur in small molecules. The size difference is crucial to make polymers useful and unique materials. Physical and chemical properties become a function of molecular weight.

As a consequence of size and methods of making, the molecules of synthetic polymer are never identical, because there always occurs the statistical distribution of molecular sizes. Moreover, even the chemical composition of chains is often not identical [1]. Scheme 6-1 shows the difference between homopolymers in which the repeating units are identical, and copolymers, which consist of two or more

P. Paneth and A. Dybala-Defratyka (eds.), *Kinetics and Dynamics*, Challenges and Advances in Computational Chemistry and Physics 12,
DOI 10.1007/978-90-481-3034-4_6,

-AAAAAAAAAAAA-	Homopolymer
-AABABABBAAAB-	Random copolymer
-ABABABABABAB-	Alternate copolymer
-AAAAAABBBBBB-	Block copolymer

Scheme 6-1. Microstructure of linear polymers

different repeating units (mers) distributed in various ways along the chain. For simplicity, only 'linear' macromolecules (chains) are shown although polymer molecules may have various structures such as star-shaped, comb-like, branched, and dendritic [1].

The distribution of mers along a chain statistically differs from one copolymer molecule to another. When modelling of a selected polymer molecule which has an 'average' composition, we assume that statistical differences in the chain microstructure do not affect the molecular properties noticeably.

The end groups also affect the properties of a polymer molecule and this effect is size-dependent. It becomes insignificant when the size of polymer (molecular weight) increases to infinity. The reactivity of the functional chain end also depends on the size of the chain. Because of large size, polymer molecules are represented by a countless number of energetically equivalent conformations. Finally, one of the main difficulties encountered in simulations of polymer molecules is a great variety of intermolecular interactions which essentially affect the material properties. The chains of a given polymer may interact primarily by van der Waals forces, hydrogen bonding, π stacking, or charge transfer interactions [2].

The main goal of theoretical methods is to develop general procedures to compute the structures and predict the properties without synthesizing the actual material. Thus, modelling should help us in designing new materials with the desired properties. There is a number of books and reviews dealing with the problems of polymer modelling in detail [3–9]. For a very helpful compilation of earlier books and reviews on polymer modelling see also bibliography in [2].

6.1.1. Methods for Modelling of Polymers

6.1.1.1. Ab Initio Calculations

Molecular modelling applied to polymers is, in principle, an extension of the concepts applied to small molecules. Potentially the most accurate model for evaluation of physical properties of chemical species is that which uses quantum chemistry methods (for a nice introduction to quantum chemistry methods used in studies of polymerisation reactions see [10]). Properties computed with ab initio methods are conformational energies, polarizabilities and hyperpolarizabilities, spectral properties, dielectric properties, charge distributions, intermolecular interactions, and reaction pathways [11].

A major problem, however, related to the accurate analysis of the potential energy surfaces for chemical reactions follows from the fact that the computational effort required by advanced ab initio techniques scales very unfavourably with the size of the reacting system making them impractical in treatment of very large reacting systems (which are quite common in polymer and biopolymer chemistry). Thus, various levels of approximation must be applied.

The simplification that is necessary to use high level ab initio methods on polymer-related systems is to approximate the propagating polymer chain (which may be hundreds of units long) with a small molecule with the active centre at the chain end. This is usually a reasonable procedure as the substituent effects decrease rapidly with the increasing distance from the reaction centre. For example, the terminal and penultimate units in a propagating polymer chain have often a remarkable effect on the reactivity and selectivity of the active site in polymerisation. However, substituent effects beyond the penultimate position are rarely considered in copolymerisation models [1]. In order to include the most important substituent effects, it is usually sufficient if propagating chains are represented as trimers (chains that are three units long).

6.1.1.2. DFT Methods

Density functional theory (DFT) is widely used in the computational chemistry community as the methods developed on the basis of this theory combine relatively high accuracy with computational efficiency [12]. The main weakness of this theory is such that the exact form of the functional relating the energy to the electron (or spin) density is unknown. The great challenge is to create a functional which describes the real problems the most accurately. The most popular density functional, B3LYP [13, 14], has, however, some serious shortcomings: (i) it performs better for main-group elements than for transition metals; (ii) it systematically underestimates reaction barrier heights; (iii) it is inaccurate for interactions dominated by medium range correlation energy, such as van der Waals attraction.

A variety of new density functionals have recently been designed. Extensive tests and benchmarks published in [15, 16] show both advantages and weaknesses of this class of techniques. DFT is primarily a ground-state oriented method [17]. Hence, DFT cannot normally compete with semiempirical and post-Hartree–Fock methodologies in calculations involving excited states having the same symmetry as the ground state, and its application to degenerate states is also particularly troublesome [18]. Another important limitation of DFT methods concerns its nonvariational nature. In fact, energy values below the true ground-state energy can be obtained, and the use of more complete basis sets does not necessarily lead to an improvement in accuracy.

Truhlar and co-workers have developed the M0X-class (X = 5,6) functionals, for which some fundamental exact constraints such as the uniform-electron-gas limit and the absence of self-correlation energy were enforced. These new functionals overcome some of the above-mentioned difficulties and are reported to be

generally the most accurate so far [19, 20]. The rapid development of the density functional theory continues and the performance of the methods hopefully will be constantly improving.

6.1.1.3. Semiempirical Methods

The application of high level ab initio and DFT calculations to polymerisation processes requires a compromise in which small model systems (oligomers, short-chain polymers) are used to mimic the behaviour of macromolecules. For polymers having symmetrical periodic structure band-structure calculations may be applied, which can reveal some long range effects and properties (photoelectron spectra, semiconducting properties, non-linear optical properties and charge-transfer effects) [8, 9]. In the cases when the effect of polymer chain cannot be neglected, it is necessary to make additional simplifications. This led to the class of approaches known as semiempirical methods, which omit or simplify the most time-consuming fragments of the ab initio calculations [21–24]. In practice, conventional semiempirical SCF-MO calculations can be easily performed on current hardware for molecules containing up to about 1,000 nonhydrogen atoms. For much larger molecules, it is advisable to employ alternative algorithms that attempt to achieve a linear scaling of the computational effort with system size, by exploiting the local character of most relevant interactions [23, 24].

Applications of linear scaling semiempirical methods have focused on large biomolecules, and calculations have been reported for molecules up to about 20,000 atoms [25]. Such calculations are most useful for large systems with long-range charge transfer or long-range charge fluctuations since such effects can only be captured by quantum–chemical approaches that cover the complete system.

6.1.1.4. Molecular Mechanics

Simplifying the model still more, we get to a commonly used level of abstraction which is the atomic ball-and-spring model (molecular mechanics, parameterized by force fields) [24, 26–28]. The major simplification is that the electronic structure of atoms is neglected, thus for example, the carbon atom is treated as a single particle instead of a nucleus with six electrons. This reduces the number of particles in the system by approximately an order of magnitude. The force field is represented in terms of intuitive parameters such as bond stretching, valence angle deformation, rotation about bonds and some nonbonded interactions.

The force field approach has the advantage of great speed. Very large systems consisting of thousands of atoms can be treated alone or surrounded by solvent molecules. Additionally, this approach is extensively used in molecular dynamics, a technique, in which the state of the system is calculated as a function of time, assuming initial distribution of energy in the system [27–29].

However, unlike ab initio basis sets, force field parameterisation is not universal. From one force field to the next, the balance of energy terms may be different. Thus, the parameters are generally non-transferable from one FF to another, or may

be transferable only with great care. Moreover, a set of parameters correspond to only a limited number of elements and bonding types. Generally, force field methods are good in predicting properties for classes of molecules where a lot of information is already available. Thus, for different classes of polymers there are different FF which work reasonably well. For less typical molecules, their use is very limited [2, 24].

6.1.1.5. Coarse Grained Models

A coarse-grained simulation model is the next step of approximation, in which a polymer is represented by a still more simplified structure. The chain consists of small molecular fragments (for example, repeating units) which are treated as stiff beads connected together with the deformable bonds. A priori, two classes of coarse-grained models may be distinguished: those which are designed for a specific polymer and reflect the underlying atomistic details to some extent, and those which retain only the most basic features of a polymer chain (chain connectivity, short-range excluded-volume interactions, etc.). Generic models are suited to explore general and universal properties of polymer systems, which occur particularly in the limit of long chains. A variety of variants of the coarse-grained model are discussed in [30, 31].

Figure 6-1 depicts a progression of the most commonly used models, starting from the one of electronic structure "resolution" and showing how detail is removed to arrive at coarse-grained models.

6.1.1.6. Hybrid Methods

When modelling of the large systems like polymers a question always arises of optimal compromise between accuracy and efficiency of chosen theoretical methods. One possible solution, which has become very popular, is to use different methods for different parts of the system, restricting the use of the more accurate high-level method to a relatively small 'active' region, while the remaining part of the system is treated by an inexpensive method such as molecular mechanics (MM), semiempirical or low-level ab initio methods. There are two approaches, one represented by the combined quantum mechanics and molecular mechanics (QM/MM) techniques [11, 32–34], and the other involving a suite of integrated methods known as ONIOM (our own *n*-layered integrated molecular orbital and molecular mechanics method) [35–37].

There are many processes in large systems where the electronically active part is quite localized, for example, polymerisation reactions involving a well-defined active site or enzymatic catalysis. In such cases, hybrid methods provide a promising solution. These methods offer a versatile approach that can be tailored to the specific systems studied, by a suitable choice of the QM/MM partitioning and of the QM and MM methods employed. A thorough review of QM/MM methods is beyond the scope of this article, and the reader is referred to the extensive literature for this purpose [38–42].

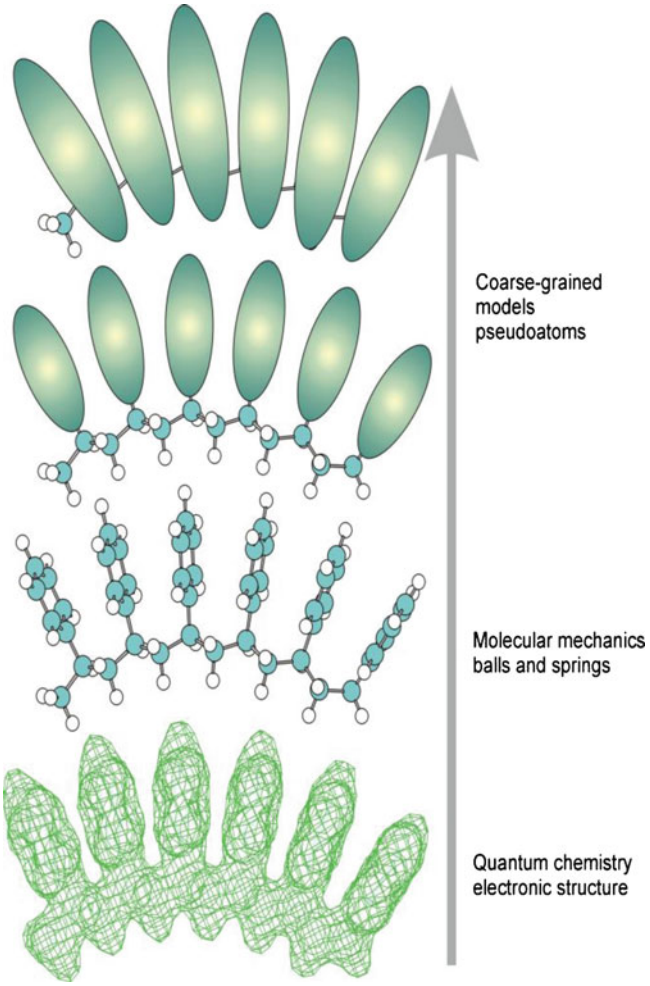


Figure 6-1. Different scales of molecular modelling

Based on these and other methodological developments, QM/MM approaches have become a popular tool for studying large (bio)systems. Both semiempirical, and ab initio and DFT methods can serve as QM component in QM/MM calculations. Semiempirical QM/MM methods are computationally much less demanding than pure QM methods [43]. In practice, semiempirical QM/MM methods are therefore often used for extensive potential energy explorations and sampling in biomolecular simulations.

Despite the obvious advantages, the hybrid methods suffer from several drawbacks that stem from the application of two fundamentally different approaches. Most of them are related to the discontinuity at the boundary between QM and MM

regions and several methodological issues were developed to deal with them. These include the choice of a proper QM/MM coupling scheme [44], the treatment of the QM/MM boundary, the specific parameterisation of QM/MM interaction terms (both for van der Waals and electrostatic contributions), and the treatment of long-range electrostatic interactions [45–51]. It was shown that the size of the QM region affects considerably the computations and have large influence on the computed quantities [52].

6.1.1.7. Molecular Dynamics

Molecular dynamics is a class of molecular modelling methods which follow explicitly the evolution of a system in time, by using the potentials together with Newton's laws of motion. These methods are a great step forward in approaching the real world, but unfortunately, they are computationally extremely expensive [2, 27, 29, 53–55].

A dynamics simulation requires a set of initial coordinates and velocities, and an interaction potential (energy function). For a short time step, the interaction may be considered constant, allowing a set of updated positions and velocities to be estimated, at which point the new interaction can be calculated. By taking a large number of small time steps, the time behaviour of the system can be obtained in the region close to the starting point (usually within a time range of a few picoseconds).

Classical reaction dynamics provides a strategy for calculating the rate coefficients of larger chemical systems. Having used quantum–chemical techniques to calculate the potential energy surface, the motions of the nuclei are studied by solving the laws of Newtonian dynamics. This is often a reasonable approximation, since the atoms (being heavier) are considerably less subject to quantum effects than the electrons.

Classical reaction dynamics of much larger systems can be studied using approximate potential energy surfaces, constructed using empirical or semiempirical procedures. In particular, the method of molecular mechanics (MM) is commonly used to simulate the motion of polymers and proteins. However, the accuracy of MM simulations are limited by the accuracy of the force field.

6.1.2. Large-Scale Molecular Modelling Calculations on Biological Systems

Molecular modelling methods are now routinely used to investigate the structure, dynamics and thermodynamics of inorganic, biological, and polymeric systems. The types of biological activity that have been investigated using molecular modelling include protein folding [24, 27, 56], enzyme catalysis [57–62], protein stability, conformational changes associated with biomolecular function, and molecular recognition of proteins, DNA, and membrane complexes [2, 24, 27, 63].

Quantum chemical calculations have been applied to biological systems almost from the beginning of their applications. Initially one could study only small

models of the actual biochemically relevant molecules. Over recent years, the situation is changing rapidly, and it is becoming possible even for nonexperts to carry out calculations on large biomolecules routinely [63–65].

6.1.3. Molecular Modelling Software for Describing Structures and Energies

Although the optimisation algorithms within most modelling software can be used to study different aspects of the potential surfaces for chemical reactions, they exhibit substantial differences in their capabilities and performance characteristics.

For the good, although partly obsolete, description of these and other molecular modelling programs see [11]. A comprehensive list molecular modelling programs with links can be found at www.redbrick.dcu.ie/∼noel/linux4chemistry.

There are a large number of software packages available for performing computational chemistry calculations. Among the programs available are: ACES II [66], ADF [67], GAMESS [68], GAUSSIAN [69], Jaguar [70], MOLPRO [71], and QCHEM [72]. Other programs, such as POLYRATE [73], have been designed to use the output of quantum-chemistry programs to calculate reaction rates and tunnelling coefficients. Whereas computational chemistry software traditionally operated on large supercomputers, versions for desktop personal computers are becoming increasingly popular. In addition, programs for visualizing the output of computational chemistry calculations are also available. These programs allow one to visualize the geometry and electronic structure of the resulting molecule, animate its vibrational frequencies or follow the dynamic trajectory.

6.2. COMPUTATIONAL QUANTUM CHEMISTRY STUDIES OF POLYMERISATION MECHANISMS

From the point of view of the polymerisation chemistry and mechanism, the reactions leading to macromolecules may be divided into two groups:

1. Step polymerisations where the macromolecules grow stepwise as a result of many subsequent reactions between the functional groups located at chain ends of growing molecules; usually, such reactions are accompanied with evolution of low-molecular by-product (polycondensation, Scheme 6-2a); or, if no by-product is evolved, the process is termed polyaddition.
2. Chain polymerisations, in which an initiator is used to produce a species R* with a reactive centre; polymerisation occurs by the successive additions of monomer molecules with regenerating of the active centre (Scheme 6-2b).

The mechanism of polycondensation is not much different from the mechanism of simple condensation of the corresponding small functional compounds. For example, formation of polyesters from difunctional alcohols and carboxylic diacids is simply the sequence of esterification reactions. Polyamides are formed by

a

step polymerisation (polycondensation):

$$\text{A-M}_1\text{-A} + \text{B-M}_2\text{-B} \rightarrow \text{A-M}_1\text{-M}_2\text{-B} + \text{AB}$$

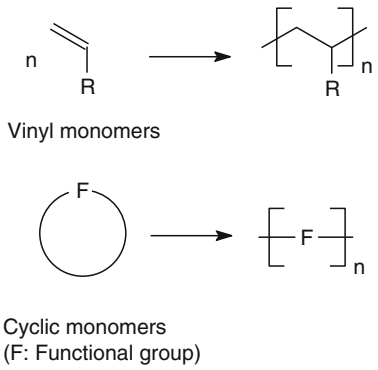
$$\text{A-(M}_1)_k(\text{M}_2)_l\text{-B} + \text{A-(M}_1)_m(\text{M}_2)_n\text{-B} \rightarrow \text{A-(M}_1)_{k+m}(\text{M}_2)_{l+n}\text{-B} + \text{AB}$$

b

chain polymerisation (addition):

$\text{R}^* + \text{M} \rightarrow \text{RM}^*$	initiation
$\text{RM}^* + \text{n M} \rightarrow \text{RM}_n\text{M}^*$	propagation
$\text{RM}_n\text{M}^* + \text{X}^* \rightarrow \text{RM}_n\text{MX}$	termination

Scheme 6-2. Step and chain polymerisations



Scheme 6-3. Polymerisation of a vinyl monomer and ring-opening polymerisation of a heterocyclic monomer

sequential condensation of amine with acid groups, as in the case of simple amines and carboxylic acids. Thus the mechanistic studies of these reactions are usually performed using small models.

The reactive centre in polymerisation (Scheme 6-2b) may be of anionic, cationic or radical nature. Monomer may be olefin or a heterocycle (Scheme 6-3).

The chain growth reactions shown in Scheme 6-2 are usually accompanied by other concurrent reactions (initiation, termination, chain transfer, decomposition, etc.) depending on the nature of active centre (radical, coordination, cationic or anionic) which make the process much more complex [1]. The knowledge of all elementary reactions taking place in the system is necessary in order to gain a control over the process, which in turn makes possible to obtain tailor-made materials. Polymerisation mechanisms are studied experimentally by kinetic methods, which often are not able to provide full required information about the system. Theoretical methods offer an invaluable tool for studying the mechanism and kinetics of polymerisation and thus importantly complement experimental procedures.

6.2.1. Solvent Effects

Typically, polymerisation reactions take place in condensed phase, in bulk or in solution. Unfortunately, the treatment of solvent effects at a high level of theory is not possible with currently available computing power, and hence the neglect of these effects (or their treatment at a low level of theory) is a potential source of error. Solvent may affect the polymerisation process in different ways [1]. For example, if polar interactions are significant in the transition state of the reaction, a high dielectric constant solvent lowers the reaction barrier due to stabilisation of the transition structure. The solvent effect is exceptionally strong for ionic reactions where the specific intermolecular electrostatic interactions dominate. Radical polymerisations are generally much less sensitive to effects of solvation unless some specific interactions with solvent such as hydrogen bonds occur.

The treatment of solvent effects may be effected in two ways. The most computationally efficient methods are continuum models, in which a solute molecule is placed in a cavity surrounded by a continuum of dielectric constant ε [74]. More sophisticated continuum models, such as the conductor-like solvation model (CPCM) [75] and the polarisable continuum model (PCM) [76] include also terms for the nonelectrostatic contributions of the solvent, such as dispersion, repulsion and cavitation. A variety of continuum models can be used for studying the influence of the dielectric constant on polar reactions. However, when specific interactions with the solvent are important, as it is the case for ionic species, it is strongly recommended to include solvent molecules in the calculation explicitly. In practice only a very limited number of individual molecules can be considered, but this system may be combined with a continuum model [77]. Unfortunately, this approach increases significantly the cost of the calculation.

The other problem concerning the study of the kinetics and thermodynamics of association and dissociation processes in solution by continuum models is that they ignore the changes to the entropic partition functions upon solvation [78]. Moreover, the outcome of calculations strongly depends on the initial geometry of the molecule and on definition of the cavity (size and shape) and should be interpreted with caution. Although some practical guidelines for dealing with solvent effects in modelling of macromolecular systems have been suggested [10], the simplified yet accurate models for dealing with solvent effects are urgently needed [79].

6.2.2. Free-Radical Polymerisation

6.2.2.1. General Remarks

Free radical polymerisation is of a great importance in technology. Therefore it has been extensively studied experimentally and theoretically. Free radical polymerisation proceeds via a chain mechanism, which basically consists of four types of reactions: (i) generation of radicals from nonradical species (initiation, usually by homolytic dissociation of a catalyst), (ii) addition of a monomer

(alkene) to radical species (propagation), (iii) atom transfer and atom abstraction reactions (chain transfer and termination by disproportionation), and (iv) radical–radical coupling reactions (termination by recombination) [80]. The knowledge of the rates of these reactions is essential for a good control of the process and, consequently, for a design and control of the product structure. For example, the relative rates of the chain transfer and termination determine the molecular weight of the resulting polymer. In copolymerisation, the relative rates of addition of the different monomers to the various types of active centres are the key parameters governing the composition and sequence distribution of the resulting copolymer.

Computational chemistry has already made a significant contribution to the understanding of conventional radical polymerisation [10, 81–91]. First of all, computational quantum chemistry allows to calculate directly the rate coefficients of individual reactions [10, 87, 92]. It also offers access to detailed mechanistic information such as transition state geometries, electron densities, substituent and chain-length effects on the kinetics and thermodynamics of key reactions [81, 89, 93, 94]. Computational studies provided evidence for the importance of penultimate unit effects [82, 83, 87, 95]. More recently, quantum chemical calculations have been used to provide an insight into the kinetics of controlled radical polymerisation processes. Excellent reviews on this topic are available [10, 84, 87, 92, 96, 97].

Numerous studies have examined the applicability of theoretical methods for investigating the reactions involved in free-radical polymerisation [10]. Geometry optimisations are generally not very sensitive to the level of theory and even the low cost HF/6-31G(d) and B3LYP/6-31G(d) methods provide reasonable geometries and frequencies [98]. UMP2 methods should generally be avoided for these reactions, as they suffer from spin-contamination problems [98, 99].

Barriers and enthalpies of reactions relevant to radical polymerisation are much more sensitive to the level of the theory, and, wherever possible, high level procedures should be used. The composite procedures, which approximate high level calculations through a series of lower level calculations proved themselves to be very useful. The main classes of procedures currently used are the G3 [100], Wn [101], and CBS [102] methods. In particular, the "RAD" variants of G3 provide the best approximations to the most accurate Wn methods [99, 103, 104]. Unfortunately, these methods are still limited to rather small systems. When they cannot be afforded, the R(O)MP2/6-311+G(3df,2p) method was shown to give reasonable values for the barriers and enthalpies of radical reactions. Hybrid DFT procedures are a good and cost-effective description of the system, although they must be applied with caution [104]. Semiempirical methods are used to study large systems for which ab initio calculations are not feasible. They perform fairly well, if the species of interest are similar to those for which the method was parameterized. However, in many situations these methods fail, as for example, in the case of radical addition to C=C bonds [105]. Hence current semiempirical methods are not generally recommended for studying free-radical polymerisation.

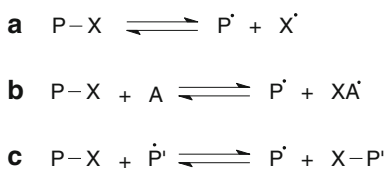
The QM/MM and ONIOM procedures appear to be accurate alternatives to high-level calculations on the full system, provided that the lower level of theory measures the substituent effects accurately [36]. Obviously, the accuracy of hybrid procedures depends upon the level of theory chosen and the way in which the core system is defined. Based on a careful study, the general guidelines for partitioning the core and full systems have been suggested [106]. For larger systems, where G3 (MP2)-RAD calculations are not feasible, the ONIOM procedure G3(MP2)-RAD: RMP2/6-311+G(3df,2p) was shown to perform very well [104].

6.2.2.2. Controlled Radical Polymerisation

Controlled radical polymerisation has recently become the most important and promising perspective for practical synthesis of polymers of well defined structure. The major growth of living (or controlled) free radical polymerisation occurred in the 1990s, commencing around 1994 with the exploitation of nitroxide-mediated polymerisation (NMP or, more general, Stable Free Radical Polymerisation, SFRP, Scheme 6-4a), atom transfer radical polymerisation (ATRP, Scheme 6-4b), degenerative transfer with alkyl iodides, and addition–fragmentation transfer (RAFT, Scheme 6-4c) approaches allowing for the facile production of a multitude of polymer architectures from simple narrow polydispersity chains to more complex stars, combs, brushes, and dendritic structures [80]. Therefore, these processes have been subjects of extensive theoretical studies in which radical addition and propagation, reaction barriers, radical stabilisation energies, polar and steric effects, and kinetic isotope effect were explored [10, 87, 92, 97].

Bond dissociation energies, transitions state structures and energy barriers for elementary reactions in model SFRP-type nitroxide-mediated polymerisations were studied by DFT methods [107–112]. Theoretical study of BDE points to the penultimate effect for the styrene-acrylonitrile copolymerisation system in the presence of 2,2,6,6-tetramethyl-1-piperidinyloxy (TEMPO) radical. The presence of two alongside styrene units in the propagating chain dramatically weaken the C-ON bond in TEMPO-derived alkoxyamines [111].

DFT calculations were carried out to find the bond dissociation energies (BDE) and free energies for a number of R-X systems (X) Cl, Br, I, N_3, S_2CNMe_2) that have been or can potentially be used as initiators for atom transfer radical polymerisation (ATRP). These calculations revealed the effectiveness of the DFT-computed BDEs for predicting polymerisation rates for new monomers in ATRP



Scheme 6-4. Types of controlled radical polymerisation

processes [113]. Comparison of DFT-derived bond dissociation energies of dormant species relevant to degenerative transfer (RAFT) and atom transfer radical polymerisation (ATRP) reveal significant steric and polar effects on the BDE [114]. Penultimate effect in ATRP copolymerisation using propylene, methyl acrylate, and methyl methacrylate monomers was studied [115]. The possibility of the heterolytic outer sphere single electron transfer bond dissociation being responsible for the activation step in single electron transfer mediated living radical polymerisation (SET-LRP) to compete with homolytic bond dissociation responsible for ATRP mechanism has been discussed [116, 117]. Mechanistic implications of thermodynamic and electrochemical properties of alkyl halides as well as effects of ligand and initiator structures for ATRP process were explored in the recent papers [117, 118].

Theoretical calculations have helped to provide an insight into structure-reactivity patterns in RAFT polymerisation [114, 119–122] and have led to the design of a new class of RAFT agents [123, 124]. The theoretical work on cobalt-mediated polymerisation has been reviewed in [125]. With continuing advances in computer power, more applications are anticipated and computational chemistry will take its place along side experimental methods as a practical kinetic tool.

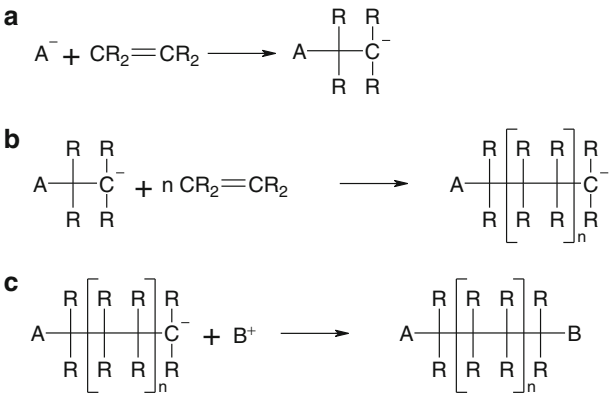
6.2.2.3. Radical Ring-Opening Polymerisation

Although typical ring-opening polymerisation proceeds via ionic routes, some cyclic monomers bearing vinyl or exomethylene groups are polymerised radically through a ring-opening fashion. Free radical ring-opening polymerisation is a less common process than addition polymerisation of olefins but it has been the focus of great interest because functional groups such as esters, amides, and carbonates can be incorporated into the backbone of chain polymers. A range of theoretical methods from semi-empirical and DFT calculations [126] to high-level G3(MP2)-RAD calculations have been used to determine the propensities of various monomers towards radical ring-opening polymerisation [127, 128]. The terminal and penultimate unit models were investigated for copolymerisation systems of 2-methylene 1,3-dioxepane and methyl methacrylate (MMA) [129].

6.2.3. Ionic Polymerisation

6.2.3.1. General Remarks

The ionic chain polymerisation of the carbon–carbon double bond by cationic and anionic initiators constitute very important class of reactions alternative to radical route. Polymerisations initiated by coordination and metal oxide initiators are usually also ionic in nature. These are called coordination polymerisations and are considered separately in the next section. The most important difference between radical and ionic polymerisation mechanism comes from the fact that ionic species in condensed phases are involved in strong intermolecular electrostatic interactions.



Scheme 6-5. General scheme of anionic polymerisation: (**a**) initiation; (**b**) propagation; (**c**) termination

Therefore ionic polymerisations are much more sensitive to the medium and the counter-ion than the free radical process. The ion pairs show also a strong tendency to aggregation. These factors must be taken into account in molecular modelling and they make the modelling of ionic polymerisation processes very difficult.

Ionic polymerisation as any kind of polymerisations, consist of three main steps: initiation, propagation and termination (Scheme 6-5). Two classes of ionic polymerisations are distinguished: cationic where the active species bears the positive charge; and anionic with negative charge.

Both types of ionic polymerisations have some common features but differ in some extent due to the different nature of both processes. First of all, they require different initiators and/or catalysts. Typical initiators of anionic polymerisation are: Brønsted and Lewis bases, in particular, carboanions or radical carboanions. Cationic polymerisation may be initiated by Brønsted and Lewis acids such as carbocations. These elementary reactions may be accompanied by the chain transfer to polymer, transfer or termination to solvent and/or to impurities, reaction with the counter-ion, ion pair aggregation, and others. For more details see, for example, [1]. These side reactions make the polymerisation process extremely difficult to study by kinetic methods. Due to the experimental difficulties, the reaction mechanisms are usually not well understood. As it was already mentioned, the reactivity of the active centres in ionic polymerisation strongly depends on the polarity of solvent.

Computational chemistry methods are employed for a wide range of problems encountered in polymer science such as mechanism and energetics of each elementary reaction and stereoselectivity of polymerisation. Large group of publications concern the modelling of initiation step of the polymerisation itself and its influence on the properties of the obtained polymer. A glance through the literature shows that the most extensively used computational method is DFT B3LYP functional alone with moderate basis sets such as 6-31+G(d). Owing to ionic character of the propagating centre, the diffusion functions must be included in the basis set.

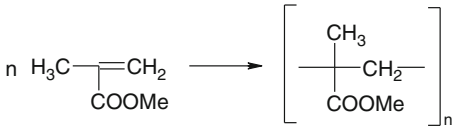
The overall tendency in modelling polymers is to deal with this in a QM/MM manner. The QM part includes the active centre, and the MM region being the polymer chain. Recently, the hybrid methods were successfully combined with the polarizable continuum model for solvation [130, 131].

6.2.3.2. Ionic Polymerisation of Olefins

Stereoselectivity of polymerisation is an important problem often investigated by QM methods. One of the earliest works in this field concerned the stereoselectivity of bonding between monomer and the active centre in the anionic polymerisation of methyl methacrylate (MMA) (Scheme 6-6). Calculations at the SCF-HF/3-21G level pointed to the two key factors affecting the stereoselectivity: intramolecular coordination of the growing centre to the penultimate unit which directs the monomer to the isotactic bonding; and energetic preference of the syndiotactic bonding of the monomer to the non-coordinated or active centre [132].

The effect of σ-ligands on the structure of the lithium ester enolates as initiators of anionic polymerisation of MMA in THF was studied in more detail [133]. The special care was taken to incorporate all possible interactions with solvent (explicit THF molecules and SCRF scheme). Studies on lithium ester enolate of methyl isobutyrate and tetraphenylphosphonium and other phosphorus-containing metal-free counterions were carried at the PB86/6-31G(d) level [134]. The ylide formed after the first polymerisation step is more stable than the ion pair, whereas after addition of one molecule of methyl methacrylate, the ylide of the dimer is destabilized through steric interactions and the polymerisation can proceed much faster. This explains the observation of a pronounced induction period when using the ester enolate of methyl isobutyrate as an initiator. Initiators of MMA anionic polymerisation were the starting point for the studies on reactivity of n-butyllithium towards pyridine derivatives [135]. Study at the B3LYP/6-31+G(d) level covers the initiation reaction between MMA and the active center made by *n*-buthyllitium and pyridine derivatives. Possible structures of highly aggregated reactive ends in polymerisation of olefins with dilithium initiator were illustrated by DFT quantum-chemical calculations and their relative stabilities were analyzed [136].

Budzelaar et al. in the series of works examined the mechanism of olefin polymerisation in the presence of aluminum catalysts. A computational strategy to study olefin polymerisation at Al systems is proposed [137–139]. The nature of the active centres in anionic polymerisation of (meth)acrylates in the presence of aluminum complexes was discussed based on the results of DFT calculations



Scheme 6-6. Polymerisation of MMA

[140, 141]. Theoretical study of the effect of triethylaluminum on the chain-end structure and tacticity of poly(N,N-dimethylacrylamide) with lithium counterion in THF has shown that the polymerisation kinetics is determined by the competition for complexation with Et_3Al between solvent, monomer, and carbonyl groups of predominantly heterotactic polymer chains [142]. Computational chemistry was also able to estimate the effect of LiCl on rates of anionic polymerisation of styrene [143]. In polar solvents the dissociation of LiCl affects the rates either ways. Free cations Li^+ are able to suppress the dissociation of polystyryllithium (PStLi) or, if scavenged by LiCl, promote the dissociation of PStLi. In contrast to this dual effect of LiCl, only enhancement (retardation) of the styrene polymerisation is observed when *n*-BuOLi or *t*-BuOLi is added. DFT methods were used to investigate this interesting phenomenon [144]. The effect of π-complexing agents on the anionic polymerisation of styrene initiated by lithium butoxides was also investigated by DFT methods [145].

A new mechanism of isoprene anionic polymerisation was proposed on the basis of theoretical caclulations. Its central moment is thermal electronic excitation of a living polyisoprene-isoprene complex into the quasi-degenerate electronically excited state, which is of the charge transfer character [146]. Analysis of energies of excited isoprenyl lithium–isoprene complexes shows that the formation of 1,4(4,1)-cis-polyisoprene in an inert media is most preferable. Similar mechanism was examined for anionic polymerisation of butadiene [147].

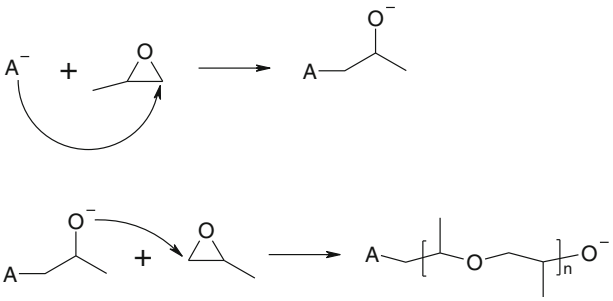
Quantum mechanical B3LYP/6-31+G(d,p) calculations were applied to explain the role of the electrolyte in the course of the anionic electropolymerisation of α-tetrathiophene. The reaction occurs in organic solvents in the presence of $LiClO_4$, while no reaction is observed when LiCl, or LiBr were used. Calculations showed that responsible for this behaviour are the differences in structures of complexes of α-tetrathiophene with the anions added [148].

The performance of DFT methods in predicting ethylene polymerisation and/or oligomerisation activity in selected aluminium and scandium based complexes was examined using both DFT and post-Hartree–Fock CCSD(T) calculations. It was shown that DFT methods underestimates the barriers for elementary reactions compared to CCSD(T) methods [149].

6.2.3.3. Ring Opening Polymerisation (ROP) of Cyclic Monomers

Cationic and anionic polymerisation mechanisms of ethylene and propylene oxides were among the most extensively investigated ROP systems by quantum mechanical methods [150, 151]. Zhang observed that the butoxy anion dominantly attacks, according to S_N2 process, the less hindered carbon of propylene oxide, producing a new secondary alkoxide (Scheme 6-7).

Cendejas examined the influence of various initiators, KOH, C_2H_5OK and potassium salt of ethylene glycol, on the rate of initiation of propylene oxide polymerisation [152]. The polymerisation was investigated at the B3LYP/6-31+G(d) level using model oligomers consisting of up to ten propylene oxide units in the



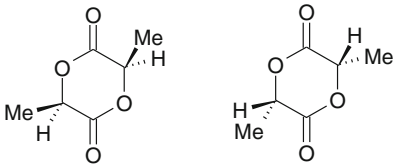
Scheme 6-7. Anionic ring opening polymerisation of propylene oxide

chain in order to obtain the significant energetic correlations as a function of the chain length. The dipotassium salt of ethylene glycol was the most efficient initiator due to the most energetically favoured configurations and regioselectivity.

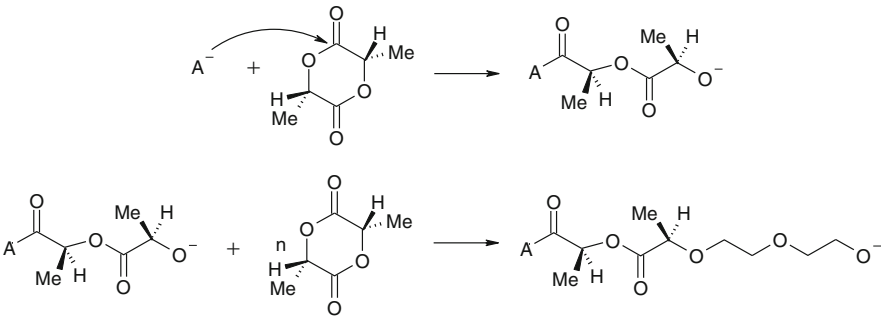
A combination of generalized energy-based fragmentation (GEBF) quantum chemical method and MM was proposed as a new QM/MM scheme to perform calculations on poly(ethylene) oxide and polyethylene in dilute aqueous solution [153]. GEBF belongs to the set of molecular fragmentation methods that are able to perform ab initio calculations for large systems containing thousands of atoms [154]. These methods rely on the locality and transferability of properties of a molecular subsystem what makes them a powerful tool in polymer research. The proposed method was demonstrated to give satisfactory energies, forces, geometries, electrostatic potentials and other properties, when compared with conventional QM/MM method. The advantages of the fragmentation QM/MM method over others are: incorporation of the polarisation of solvent and good description of hydrogen-bonding interactions.

The mechanism of copolymerisation of CO_2 with cyclohexene oxide catalyzed by the Zn(II) organometallic compound (BDI)$ZnOCH_3$, (BDI = N(2,6-*i*-$Pr_2C_6H_3$)C(Me)CHC(Me)N(2,6-*i*-$Pr_2C_6H_3$) chelating β-diimine ligand) has been studied using ONIOM method, combining the density functional method B3LYP/LANL2DZ(d) with the semiempirical PM3 method. The hybrid ONIOM (B3LYP:PM3) method was found to be a good approximation to the much more expensive standard B3LYP method [155].

The other monomer which has drawn a considerable attention of researchers is lactide (Scheme 6-8) which occurs as three, L-, D- and meso, stereoisomers. Stereoseletive polymerisation of these isomers leads to polymers stereochemically different which results in very interesting differences of physical properties. Various polymerisation mechanisms leading to poly(lactic acid), i.e., polycondensation of lactic acid, as well as anionic and cationic polymerisations of lactide, using different initiators – (a) H_2SO_4 for polycondensation, (b) aluminum isopropoxide for coordination-insertion ROP, (c) methyl triflate for cationic ROP, and (d) potassium methoxide for anionic ROP, were studied in detail at the B3LYP/6-31G(d) level



Scheme 6-8. L-lactide and D-lactide



Scheme 6-9. Anionic ring-opening polymerisation of lactide

of theory (6-31G(d,p) for metal atoms) [156]. The mechanism of the anionic ring-opening polymerisation (AROP) of lactide (Scheme 6-9) was studied in order to investigate the difference in reactivity of aluminum trialkoxide and dialkylaluminium monoalkoxide [157].

Ryner et al. studied the mechanism of ring-opening polymerisation of lactones and L-lactide with Sn(IV) initiators, stannous 2-ethylhexanoate, evaluating the activation barrier of the L-LA polymerisation initiated by $Sn(Oct)_2$ [158, 159].

Single-site metal catalysts draw much attention in polymerisation of lactide due to perfect stereocontrol of the reaction. To elucidate the observed experimental results, the B3LYP calculations were applied for the ROP of rac-lactide initiated by single-site (BDI)MgOMe(THF) complexes (BDI = $HC[CMeN\text{-}2,6\text{-}iPr_2C_6H_3]_2$) [160]. Calculations were performed using B3LYP functional and three types of basis sets: 6-311G(3d) for Mg atom to obtain the realistic coordination, 6-31G(d) for BDI moiety and LA, and small STO-3G for bulky substituents governing the stereoselectivity of the overall process. The mechanism of initiation of ROP of L-lactides by the cationic (N-heterocyclic carbine) (NHC) silver complex was investigated by DFT-B3LYP [161]. Model reactions for the 4-dimethylaminopyridine (DMAP)-catalyzed ring-opening polymerisation of lactide and the corresponding lactic O-carboxylic anhydride (lacOCA) have been studied computationally at the B3LYP/6-31G(d) level of theory. The solvent effect of dichloromethane was

regarded at PCM/SCRF approximation [162]. Cerium(IV) alkoxide catalyst for lactide polymerisation was reported [163].

Methanolysis of ε-caprolactone as a model for initiation of ROP of this monomer was studied by B3LYP/6-31G(d) method to evaluate the solvent effect [164]. Specific solvation by explicit methanol molecules organized in cyclic hydrogen-bonded clusters are significant and influences considerably the energy barriers. Opposite conclusion can be derived when the polarizable continuum model is employed for solvation. ROP of caprolactone proceeds according to stepwise addition/elimination mechanism. Similar mechanism was proposed for ROP initiated by yttrium catalyst using B3LYP functional and 6-31G(d) basis set for all atoms except to Y when the LANL2DZ was used [165]. Calculations show that the rate-determining step is nucleophilic attack of the alkoxide on the carbonyl-carbon atom of the monomer. Another functional B3PW91 with the 6-31G(d,p) basis set and Stuttgart–Dresden pseudopotential for Eu was used to explore the mechanism of ROP of ε-caprolactone initiated by hydride $Cp_2Eu(H)$ and boronhydrides $Cp_2Eu(BH_4)$ and $(N_2NN')Eu(BH_4)$ where N_2NN'=(N_2NN'=(2-C_5H_4N)$CH_2(CH_2CH_2NMe)_2$) [166].

The mechanisms for ring-opening polymerisation of cyclic esters initiated by various initiators were investigated at the B3LYP/6-311++G(d,p)//B3LYP/6-31+G(d) level [167, 168]. Mechanism of initiation and propagation of substituted maleimides using Zn^{2+} complexes was studied by ONIOM(B3LYP/LANL2DZ/B3LYP/LANL2MB) method [169]. The use of a chiral catalyst results in chiral polymer which is in accord with experimental observations.

The mechanism of cationic and anionic ring-opening polymerisation of cyclosiloxanes has been explored by ab initio methods [170–172] The effect of counter-ion-monomer interaction on the chain microstructure was analysed by NMR and ab initio calculations [173].

6.2.4. Coordination Polymerisation

Polymerisation of α-olefins is one of the cornerstone reactions in organometallic catalysis. Ziegler–Natta olefin polymerisation is an important chemical process leading to polymers of defined stereochemical structure [1]. Because of its importance it has drawn a great attention of theoretical chemists. This chapter presents only the most recent papers; for earlier literature the readers are recommended to follow the references cited therein. A typical catalyst for the heterogeneous polymerisation reactions is $TiCl_3/AlR_3$. The function of the group I–III metal compound is to modify and activate the transition metal compound for initiation. However, the discovery and development of metallocene initiators in the mid-1980s revolutionised the coordination polymerisation process. The metallocene initiators, unlike the traditional Ziegler–Natta initiators, are the molecules that can be isolated, analyzed, and purified. They are also homogeneous in the reaction systems and have the potential for tailor-making initiators to perform highly stereoselective

polymerisations of specific monomers. Bis(cyclopentadienyl)titanium dichloride and bis(indenyl)zirconium dichloride are examples of simple metallocene initiators.

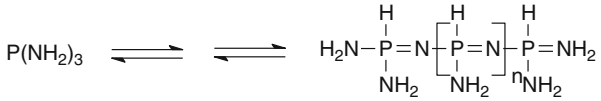
Many works in this area attempt to apply molecular mechanics (or combined quantum mechanics/molecular mechanics methods) to understand the mechanism of polymerisation, and, in particular, the origins of stereoselectivity in polymer production. The other direction is search for new transition metal complexes which may be active as catalysts of the polymerisation. Earlier results in this area have been comprehesively reviewed in [174–176].

For heavy elements like transition metals, it is usually necessary to make use of effective core potentials (ECPs), in which the inner-shell electrons are replaced with a model potential [28, 177]. The advantage of this approach is that the computational effort is significantly reduced, since only the valence electrons are explicitly considered. However, due to complex electronic structure of transition metals it is often impossible to write a single, simple Lewis structure for compounds that contain transition metal. This generally means that it is difficult to find a single electronic configuration that adequately describes what the electrons are doing in such species. Then a single configuration wavefunction is unlikely to be appropriate, even as a starting point for subsequent correlated calculations. In such cases a multiconfigurational (MC) wavefunction must be considered. The most common approach is the complete active space (CAS) self-consistent field (SCF) or fully optimized reaction space (FORS) method. Both these approaches are specific examples of the more general MCSCF method [178].

Because of the cost of MCSCF methods, most authors prefer using the DFT (BP86, PW91, B3LYP) methods which are significantly less computationally demanded. There is a huge amount of theoretical work in this field focusing on the search of the new catalytically active complexes of practically all transition metals starting from titanium, zirconium, vanadium, nickel, cobalt, palladium, platinum, rhodium, hafnium and others. They aim at deeper understanding the mechanistic features of the process, such as the relationship between transition metal, its ligands and structure of the complex as well as the structure of support in the case of heterogeneous systems on tacticity and degree of stereoregularity of polymerisation of various monomers.

A thorough understanding of the relationship between the chemical structure of donor and the isotacticity (and productivity) of polypropylene are essential for molecular design of better donor to yield highly isotactic polypropylene with high productivity. The mechanism of internal (or external) donors, added to supported heterogeneous Ziegler–Natta catalyst has been the subject of extensive studies [179–182] and references cited therein.

The role of titanium oxidation states in olefin polymerisation activity was rationalised based on DFT calculations [183]. Catalytic properties of Ni(II) in polymerisation of polar olefins (acrylates) were recently explored [184]. Stereoregularity, regioselectivity, and reactivity of various metallocene systems were the subject of extensive theoretical studies by Ziegler's group and others [185–191]. The influence of the ligand structures of hafnocene polymerisation catalysts on



Scheme 6-10. Polycondensation of phosphazenes

ethene insertion and chain propagation was systematically investigated by quantum chemical methods [192].

Molecular dynamics methods were applied to study the whole catalytic process. Ziegler et al. presented calculations on the ion-pair recombination based on the MD (Car-Parrinello ab initio molecular dynamics (CP-AIMD)) method. Further, MD methods are well suited to evaluate free energy profiles for a catalytic process, a task that can be expensive and inaccurate with static methods where use is made of the harmonic approximation [193].

A number of studies on the ring-opening metathesis polymerisation (ROMP) of cycloolefins catalysed by ruthenium and rhenium complexes have appeared recently [194–197].

6.2.5. Polycondensation

Very few reports on theoretical studies of polycondensation reactions can be found. Ab initio calculations were carried out to study the mechanism of polycondensation of PH-functionalized phosphazenes. Starting from the parent triaminophosphine-triaminophosphine $P(NH_2)_3$, the relative stabilities of corresponding aminophosphine–phosphazene tautomers for the idealized polycondensation reaction were compared (Scheme 6-10). This idealized reaction mechanism is based on several condensation/elimination steps [198].

Polycondensation of silanols to obtain cage-like structures was systematically studied in a series of works by Kudo and Gordon [199, 200]. The influence of water on lowering the barrier of acid-catalysed hydrolysis/condensation of siloxane bonds was reported [201].

6.3. ENZYMATIC REACTIONS

Biological systems are specific both in structure and in chemical reactivity. Enzymes are highly selective and effective catalysts showing kinetics different from other types of catalytic reactions. On the other hand, there is a vast literature on enzymatic reactions including a great amount of theoretical work. Therefore the enzymatic catalysis is treated here as the separate topic, although very briefly. Nowadays we do not possess the sufficient computer power to deal with the enzymatic reactions on the quantum mechanical level. QM/MM method proved to be a most successful and powerful tool for investigating complex enzymatic reactions [57, 202]. The substrate and active site of the enzyme are the obvious

choice for the QM region, and the remainder being treated on the MM level of accuracy. Since the natural environment of the enzymes, directly affecting their properties, is water solution, the solvent effects must be taken into account. This is usually done by explicit water molecules present in the MM region, or using polarizable continuum model, or more complex methods [203]. From the variety of quantum mechanical approaches, the DFT methods are most often chosen, due to their good performance and smaller computational effort in comparison to other methods of similar accuracy [58]. Especially QM/MM studies on mechanisms involving transition-metal enzymes apply DFT, mostly B3LYP functional, as the QM part [42, 204]. Other high accuracy models like CCSD(T) and CASPT2 are also used [205]. For the MM part various force fields are used. Many FF were derived for peptides, like CHARMM [206], AMBER [207], GROMOS [208], and OPLS-AA [209] and have been successfully implemented in the QM/MM scheme. The QM/MM mechanistic studies on enzymes focus on optimisation of the stationary states, evaluation of activation barriers and the geometry of transition states [210]. By means of QM/MM procedure it is possible to calculate the NMR shielding tensors of the enzymes [211]. For reviews see [39, 57–59, 61, 62, 212] and citation herein.

Large biomolecules with thousands of atoms are an attractive target for semiempirical work. Higher-level ab initio and DFT calculations are very costly for such systems even when the QM/MM approach is used, especially when considering the need for an adequate sampling of conformational space. Therefore, semiempirical QM/MM studies on large biomolecules have become very popular in recent years. Many enzymatic reactions have been examined at semiempirical QM/MM level [23, 39, 213, 214]. Semiempirical QM/MM calculations are fast enough to compute potentials of mean force along the reaction coordinate so that free energy barriers of enzymatic reactions can be obtained. Comparing with the results from an analogous treatment of the reaction in aqueous solution and in the gas-phase allows an analysis of the origin of the rate enhancement in the enzyme. Examples of such semiempirical QM/MM work are the recent investigations on chorismate mutase [215, 216] and on catechol O-methyl transferase [217]. Dynamical effects on enzyme kinetics have been incorporated on the basis of semiempirical QM/MM potential surfaces by a variety of techniques, including mixed quantum-classical molecular dynamics and variational transition state theory (see [39] for a review), and kinetic isotope effects have been evaluated for a number of enzymatic reactions [39, 218, 219]. Some of these QM/MM studies employ specific reaction parameters (mostly AM1-SRP or PM3-SRP) to improve the accuracy of the QM component [39, 219, 220].

Large biomolecules can also be treated by linear scaling semiempirical QM calculations, but at notably higher costs as compared to analogous QM/MM work [43]. Such calculations are thus done less often. Applications of the linear scaling divide-and-conquer approach to biological systems have been reviewed [221]. Topics that have been addressed by linear scaling semiempirical methods include charge transfer interactions in macromolecular systems such as the protein/water interface, the electrostatic potential in the potassium channel, charge fluctuations in

the nucleic acids DNA and RNA in solution, the regioselectivity and RNA-binding affinity of HIV-1 nucleocapsid protein, and the calculation of NMR chemical shifts in biomolecules using a divide-and-conquer implementation of the GIAO-MNDO method. In addition to such large-scale calculations on systems with thousands of atoms, there are conventional semiempirical calculations that address biochemical problems using smaller model. DFT has been the choice for geometry optimisations and single-point energy calculations albeit with a larger basis set. An evaluation of the general performance and limitations of density functionals, compiling hundreds of benchmarks, can be found in ref [16]). With technological advances in computer power, the choice of QCISD(T) becomes now feasible, and energy values do improve with this methodology. A recent work has shown that very high theoretical levels (in this case coupled-cluster) might be needed for the high-accuracy calculation of enzymatic barrier heights [51].

6.4. STRUCTURAL STUDIES

Computational methods are very useful in analysis of conformations and electronic structures of polymer chains. Conjugated organic polymers have been of great interest as an important class of electronic materials because of their high electrical conductivities and good non-linear optical (NLO) properties [222]. Structural parameters, conformations, spectral properties and the energy gaps between valence and conduction band based on the HOMO/LUMO differences have been studied by the semiempirical, ab initio and DFT methods taking advantage of the highly symmetrical structures of these polymers [2, 8].

Due to the size of polymer systems, the choice of theoretical methods has been limited. Previous theoretical studies had been based on semiempirical optimized geometries, and band structure calculations using extended Hückel theory (EHT), the Su–Schrieffer–Heeger model (SSH), or the VEH method [223–225]. Recently, higher level ab initio and DFT calculations become more and more popular showing an improved agreement between experiment and theory.

One of the most explored semiconducting polymer is polyaniline. Band gap in polyaniline has been calculated for oligomers of different forms of polyaniline by use of molecular mechanics (MMFF), semiempirical (AM1), ab initio (HF/6-31G (d)), and density functional theory (B3LYP/6-31G(d)) methods and the performance of the methods was critically discussed [226]. The other promising class of semiconducting materials which have been widely studied by theoretical methods are polythiofenes and thiophene-derived polymers [227–230].

Very interesting class of polymers are polyphosphazenes. The structure of polyphosphazenes in solid state and in solution was studied by crystallographic and theoretical methods. It is known that they tend to crystallize in a planar *cis*, *trans* conformation with minor P–N bond-length alternations. In contrast, recent theoretical studies of several phosphazene oligomers suggest that isolated polyphosphazene molecules preferentially form a twisted helical conformation [231, 232].

6.5. SUMMARY

The authors are aware that this short review is far from covering or at least mentioning all the problems associated with molecular modelling of large polymeric systems. Also the contemporary knowledge in this area is too vast to be comprehensively reviewed. We just selected a few topics to mention, and in particular, the state-of-art in modelling of the reactions leading to polymers. The main limitation in modelling of macromolecular and supramolecular systems is the lack of computing power. However, as the dynamical progress in hardware and the development of theoretical calculation methods continues, the importance and reliability of modelling of large chemical systems will still increase.

REFERENCES

1. Odian G (2004) Principles of polymerization. Wiley, Hoboken, NJ
2. Young D (2001) Computational chemistry. Wiley, New York
3. Bicerano J (ed) (1993) Computational modeling of polymers. Marcel Dekker, New York
4. Bicerano J (ed) (1996) Prediction of polymer properties. Marcel Dekker, New York
5. Gelin BR (1994) Molecular modeling of polymer structures and properties. Hanser Publishers, Munich
6. Galiatsatos V (1995) In: Lipkowitz KB, Boyd DB (eds) Reviews in computational chemistry, Vol 6. VCH, New York
7. Askadskii AA (2003) Computational materials science of polymers. Cambridge University Press, Cambridge
8. André J-M, Delhalle J, Brédas J-L (eds) (1991) Quantum chemistry aided design of organic polymers. Worlds Scientific Publications, Singapore
9. André J-M (2005) In: Dykstra CE, Frenking G, Kim KS et al (eds) Theory and applications of computational chemistry. Elsevier BV, Amsterdam
10. Coote ML (2004) In: Mark HF (ed) Encyclopedia of polymer science and technology, vol 9, 3rd edn. Wiley, New York
11. Karadakov PB (2001) Annu Rep Progr Chem Sect C 97:61–90
12. Parr RG, Yang W (eds) (1989) Density-functional theory of atoms and molecules. Oxford University Press, Oxford
13. Becke AD (1993) J Chem Phys 98:5648–5652
14. Lee C, Yang W, Parr RG (1988) Phys Rev B 37:785–789
15. Riley KE, Op't Holt BT, Merz KM (2007) J Chem Theory Comput 3:407–433
16. Sousa SF, Fernandes PA, Ramos MJ (2007) J Phys Chem A 111:10439–10452
17. Perdew JP, Ruzsinszky A, Constantin LA et al (2009) J Chem Theory Comput 5:902–908
18. Bersuker IB (1997) J Comput Chem 18:260–267
19. Zhao Y, Truhlar DG (2006) J Chem Theory Comput 3:289–300
20. Zhao Y, Truhlar D (2008) Acc Chem Res 41:157–167
21. Stewart JJP (1998) In: PvR S, Allinger NL, Kollman PA et al (eds) Encyclopedia of computational chemistry, 3rd edn. Wiley, Chichester
22. Thiel W (2000) In: Grotendorst J (ed) Theory and applications of computational chemistry, vol 3, 2nd edn. John von Neumann Institute for Computing, Jülich

23. Thiel W (2005) In: Dykstra C, Frenking G, Kim K et al (eds) Theory and applications of computational chemistry. Elsevier, Amsterdam
24. Ramachandran KI, Deepa G, Namboori K (2008) Computational chemistry and molecular modeling: principles and applications. Springer, Berlin
25. Daniels AD, Miliam JM, Scuseria GE (1997) J Chem Phys 107:425–431
26. Machida K (1999) Principles of molecular mechanics. Wiley, New York
27. Leach A (2001) Molecular modelling: principles and applications. Prentice Hall, London
28. Jensen F (2007) Introduction to computational chemistry. Wiley, Chichester
29. Rapaport DC (2004) The art of molecular dynamic simulation. Cambridge University Press, Cambridge
30. Baschnagel J, Wittmer JP, Meyer H (2004) Monte Carlo simulation of polymers: coarse-grained models, NIC Series edn, vol 23. John von Neumann Institute for Computing, Jülich
31. Faller R (2007) In: Lipkowitz KB, Cundari TR (eds) Reviews in computational chemistry, vol 23. Wiley-VCH, Weinheim
32. Gao J (1996) In: Lipkowitz KB, Boyd DB (eds) Reviews in computational chemistry, vol 7. VCH, New York
33. Gao J (1998) In: PvR S, Allinger NL, Kollman PA et al (eds) Encyclopedia of computational chemistry, vol 2. Wiley, Chichester
34. Senn HM, Thiel W (2009) Angew Chem Int Ed 48:1198–1229
35. Froese RDJ, Morokuma K (1998) In: PvR S, Allinger NL, Kollman PA et al (eds) Encyclopedia of computational chemistry, vol 2. Wiley, Chichester
36. Vreven T, Byun KS, Komaromi I et al (2006) J Chem Theory Comput 2:815–826
37. Lundberg M, Morokuma K (2009) Multi-scale quantum models for biocatalysis, vol 7. Springer, Dordrecht
38. Monard G, Merz KM (1999) Acc Chem Res 32:904–911
39. Gao J, Truhlar DG (2003) Annu Rev Phys Chem 53:467–505
40. Gao J (1996) Acc Chem Res 29:298–305
41. Mulholland AJ (2007) Chem Centr J 1:19. doi:10.1186/1752-153X-1-19
42. Lundberg M, Kawatsu T, Vreven T et al (2009) J Chem Theory Comput 5:222–234
43. Titmuss SJ, Cummins PL, Bliznyuk AA et al (2000) Chem Phys Lett 320:169–176
44. Bakowies D, Thiel W (1996) J Phys Chem 100:10580–10594
45. Thompson MA (1996) J Phys Chem 100:14492–14507
46. Gao J (1997) J Comput Chem 18:1061–1071
47. Gao J, Alhambra C (1997) J Chem Phys 107:1212–1217
48. Reuter N, Dejaegere A, Maigret B et al (2000) J Phys Chem A 104:1720–1735
49. Pu J, Gao J, Truhlar DG (2004) J Phys Chem A 108:632–650
50. Heyden A, Lin H, Truhlar DG (2007) J Phys Chem B 111:2231–2241
51. Thiel W (2009) In: Grotendorst J (ed) Multiscale simulation methods in molecular sciences, vol 42. John von Neumann Institute for Computing, Jülich
52. Solt I, Kulhanek P, Simon I et al (2009) J Phys Chem B 113:5728–5735
53. Binder K (1995) Monte Carlo and molecular dynamics simulations in polymer science. Oxford University Press, Oxford
54. Frenkel D, Smit B (2002) Understanding molecular simulation: from algorithms to applications. Academic, San Diego, CA
55. Sun L, Hase WL (2003) In: Lipkovitz KB, Boyd DB (eds) Reviews in computational chemistry, vol 19. Wiley, Hoboken

56. Strunk T, Verma A, Gopal SM, Schug A, Klenin K, Wenzel W (2009) In: Grotendorst J (ed) Multiscale simulation methods in molecular sciences, vol 42. John von Neumann Institute for Computing, Jülich
57. Gao J, Ma S, Major DT et al (2006) Chem Rev 106:3188–3209
58. Siegbahn PEM, Borowski T (2006) Acc Chem Res 39:729–738
59. Senn HM, Thiel W (2007) Curr Opin Chem Biol 11:182–187
60. Van der Kamp MW, Mulholland AJ (2008) Nat Prod Rep 25:1001–1014
61. Ramos MJ, Fernandes PA (2008) Acc Chem Res 41:689–698
62. Senn HM, Thiel W (2009) Angew Chem Int Ed 48:1198–1229
63. Friesner RA, Dunietz BD (2001) Acc Chem Res 34:358
64. Mulholland AJ (2006) In: Hinchliffe A (ed) Chemical modelling: applications and theory, vol 4. The Royal Society of Chemistry, Cambridge
65. Woods CJ, Mulholland AJ (2008) Multiscale modelling of biological systems, vol 5. The Royal Society of Chemistry, Cambridge
66. Stanton JF, Gauss J, Watts JD et al (1992) Int J Quant Chem Symp 26:879–894
67. Te Velde G, Bickelhaupt FM, Baerends EJ et al (2001) J Comput Chem 22:931–967
68. Gordon MS, Schmidt MW (2005) In: Dykstra CE, Frenking G, Kim KS et al (eds) Theory and applications of computational chemistry: the first forty years. Elsevier, Amsterdam
69. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery JA, Jr., Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam JM, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas O, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ. Gaussian 09. [Revision A.02]. 2009. Wallingford, CT, Gaussian, Inc. 2009
70. Jaguar. [Rev. B] (2006) http://www.schrodinger.com, Schrödinger, LLC
71. Hetzer G, Pulay P, Werner HJ (1998) Chem Phys Lett 290:143
72. Shao Y, Fusti-Molnar L, Jung Y et al (2006) Phys Chem Chem Phys 8:3172–3191
73. Fernandez-Ramos A, Ellingson BA, Garrett BC, Truhlar DG (2007) In: Lipkowitz KB, Cundari TR (eds) Reviews in computational chemistry, vol 23. Wiley-VCH, Hoboken
74. Tomasi J, Mennucci B, Cammi R (2005) Chem Rev 105:2999–3094
75. Cossi M, Rega N, Scalmani G et al (2003) J Comput Chem 24:669–681
76. Mennucci B, Tomasi J (1997) J Chem Phys 106:5151–5158
77. Pliego JR, Riveros JM (2001) J Phys Chem A 105:7241–7247
78. Leung BO, Reid DL, Armstrong DA et al (2004) J Phys Chem A 108:2720–2725
79. van Dam HJJ, Guest MF, Sherwood P et al (2006) J Mol Struct:Theochem 771:33–41
80. Matyjaszewski K, Davis TP (eds) (2002) Handbook of radical polymerization. Wiley, New York
81. Heuts JPA, Gilbert RG, Radom L (1995) Macromolecules 28:8771–8781
82. Coote ML, Davis TP, Radom L (1999) Macromolecules 32:5270–5276
83. Coote ML, Davis TP, Radom L (1999) Macromolecules 32:2935–2940
84. Coote ML, Davis TP (1999) Prog Polym Sci 24:1217–1251
85. Filley J, McKinnon JT, Wu DT et al (2002) Macromolecules 35:3731–3738
86. Zhan CG, Dixon DA (2002) J Phys Chem A 106:10311–10325

87. Heuts JPA (2002) In: Matyjaszewski K, Davis TP (eds) Handbook of radical polymerization. Wiley-Interscience, Hoboken
88. Van Cauter K, Hemelsoet K, Van Speybroeck V et al (2005) Int J Quant Chem 102:454–460
89. Heuts JPA, Russell GT (2006) Eur Polym J 42:3–20
90. Van Cauter K, Van Speybroeck V, Vansteenkiste P et al (2006) Chemphyschem 7:131–140
91. Van Cauter K, Van Speybroeck V, Waroquier M (2007) Chemphyschem 8:541–552
92. Coote ML (2009) Macromol Theory Simul 18:388–400
93. Huang DM, Monteiro MJ, Gilbert RG (1998) Macromolecules 31:5175–5187
94. Toh JSS, Huang DM, Lovell PA et al (2001) Polymer 42:1915–1920
95. Coote ML, Davis TP, Radom L (1999) J Mol Struct: THEOCHEM 461–462:91–96
96. Krenske EH, Izgorodina EI, Coote ML (2006) In: Matyjaszewski K (ed) Controlled/living polymerization, vol 944. American Chemical Society, Washington DC
97. Coote ML, Krenske EH, Izgorodina EI (2008) In: Barner-Kowollik C (ed) Handbook of RAFT polymerization. Wiley-VCH, Weinheim
98. Gómez-Balderas R, Coote ML, Henry DJ et al (2004) J Phys Chem A 108:2874–2883
99. Wong MW, Radom L (1998) J Phys Chem A 102:2237–2245
100. Curtiss LA, Raghavachari K (2002) Theor Chem Acc 108:61–70
101. Martin JML, Parthiban S (2001) In: Cioslowski J (ed) Quantum mechanical prediction of thermochemical data. Kluwer, Dordrecht, The Netherlands
102. Montgomery JA Jr, Frisch MJ, Ochterski J et al (2000) J Chem Phys 112:6532–6542
103. Coote ML (2004) J Phys Chem A 108:3865–3872
104. Izgorodina EI, Brittain DRB, Hodgson JL et al (2007) J Phys Chem A 111:10754–10768
105. Wong MW, Radom L (1995) J Phys Chem 99:8582–8588
106. Izgorodina EI, Coote ML (2006) J Phys Chem A 110:2486–2492
107. Marsal P, Roche M, Tordo P et al (1999) J Phys Chem A 103:2899–2905
108. Gigmes D, Gaudel-Siri A, Marque SRA et al (2006) Helv Chim Acta 89:2312–2326
109. Kaim A, Megiel E (2006) J Polym Sci A – Polym Chem 44:914–927
110. Kaim A (2007) J Polym Sci A – Polym Chem 45:232–241
111. Megiel E, Kaim A (2008) J Polym Sci A – Polym Chem 46:1165–1177
112. Zarycz N, Botek E, Champagne B et al (2008) J Phys Chem B 112:10432–10442
113. Gillies MB, Matyjaszewski K, Norrby PO et al (2003) Macromolecules 36:8551–8559
114. Matyjaszewski K, Poli R (2005) Macromolecules 38:8093–8100
115. Lin CY, Coote ML, Petit A et al (2007) Macromolecules 40:5985–5994
116. Guliashvili T, Percec V (2007) J Polym Sci A – Polym Chem 45:1607–1618
117. Lin CY, Coote ML, Gennaro A et al (2008) J Am Chem Soc 130:12762–12774
118. Tang W, Kwak Y, Braunecker W et al (2008) J Am Chem Soc 130:10702–10713
119. Chiefari J, Mayadunne RTA, Moad CL et al (2003) Macromolecules 36:2273–2283
120. Coote ML, Henry DJ (2005) Macromolecules 38:1415–1433
121. Coote ML (2005) J Phys Chem A 109:1230–1239
122. Feldermann A, Coote ML, Stenzel MH et al (2004) J Am Chem Soc 126:15915–15923
123. Coote ML, Henry DJ (2005) Macromolecules 38:5774–5779
124. Theis A, Stenzel MH, Davis TP et al (2005) Austr J Chem 58:437–441
125. Debuigne A, Poli R, Jérôme R et al (2009) Prog Polym Sci 34:211–239
126. Ochiai B, Endo T (2007) J Polym Sci A – Polym Chem 45:2827–2834
127. Hodgson JL, Coote ML (2005) Macromolecules 38:8902–8910
128. Coote ML, Hodgson JL, Krenske EH et al (2007) Austr J Chem 60:744–753
129. Roberts GE, Coote ML, Heuts JPA et al (1999) Macromolecules 32:1332–1340
130. Vreven T, Mennucci B, da Silva CO et al (2001) J Chem Phys 115:62–72

131. Mo SJ, Vreven T, Mennucci B et al (2003) Theor Chem Acc 111:154–161
132. Dybal J, Křiz J (1997) Macromol Theory Simul 6:437–450
133. Yakimansky AV, Muller AHE (2001) J Am Chem Soc 123:4932–4937
134. Weiss H, Steiger S, Jungling S et al (2003) Macromolecules 36:3374–3379
135. Catak S, Aviyente V, Nugay T (2005) J Polym Sci A – Polym Chem 43:455–467
136. Janssens K, Loozen E, Yakimansky A et al (2009) Polymer 50:5368–5373
137. Talarico G, Budzelaar PHM (2000) Organometallics 19:5691–5695
138. Talarico G, Barone V, Budzelaar PHM et al (2001) J Phys Chem A 105:9014–9023
139. Talarico G, Busico V, Budzelaar PHM (2001) Organometallics 20:4721–4726
140. Schmitt B, Schlaad H, Muller AHE et al (1999) Macromolecules 32:8340–8349
141. Schmitt B, Schlaad H, Muller AHE et al (2000) Macromolecules 33:2887–2893
142. Yakimansky AV, Muller AHE (2006) Macromolecules 39:4228–4234
143. Yakimansky AV, Muller AHE, Van Beylen M (2000) Macromolecules 33:5686–5692
144. Yakimansky A, Van Beylen M (2002) Polymer 43:5797–5805
145. Yakimansky A, Wang G, Janssens K et al (2003) Polymer 44:6457–6463
146. Kalninsh KK, Podolsky AF (2005) Int J Quant Chem 104:114–125
147. Kalninsh KK, Podolsky AF (2002) Int J Quant Chem 88:624–633
148. Aleman C, Oliver R, Brillas E et al (2005) Chem Phys 314:1–7
149. Meier RJ, Koglin E (2004) Macromol Theory Simul 13:133–139
150. Zhang Z, Yin H, Fang X (2006) J Zhejiang Univ Sci A 7:325–329
151. Zhang X, Chu W, Chen JJ et al (2009) Acta Phys-Chim Sinica 25:451–456
152. Cendejas G, Flores-Sandoval CA, Huitron N et al (2008) J Mol Struct 879:40–52
153. Li H, Li W, Li S et al (2008) J Phys Chem B 112:7061–7070
154. Yang W (1991) Phys Chem Lett 66:1438–1441
155. Liu Z, Torrent M, Morokuma K (2002) Organometallics 21:1056–1071
156. Khanna A, Sudha YS, Pillai S et al (2008) J Mol Model 14:367–374
157. Equiburu JL, Fernandez-Berridi MJ (1999) Macromolecules 32:8252–8258
158. Ryner M, Stritsberg K, Albertsson A et al (2001) Macromolecules 34:3877–3881
159. Von Schenk H, Ryner M, Albertsson A et al (2002) Macromolecules 35:1556–1562
160. Marshall EL, Gibson VC, Rzepa HS (2005) J Am Chem Soc 127:6048–6051
161. Samantaray MK, Katiyar V, Roy D et al (2006) Eur J Inorg Chem 15:2975–2984
162. Bonduelle C, Martin-Vaca B, Cossio FP et al (2008) Chem Eur J 14:5304–5312
163. Broderick EM, Diaconescu PL (2009) Inorg Chem 48:4701–4706
164. Buis N, French SA, Ruggiero GD et al (2007) J Chem Theory Comput 3:146–155
165. Liu JZ, Ling J, Li X et al (2009) J Mol Cat A-Chem 300:59–64
166. Barros N, Mountford P, Guillaume SM et al (2008) Chem Eur J 14:5507–5518
167. Simon L, Goodman JM (2007) J Org Chem 72:9656–9662
168. Ling J, Shen JG, Hogen-Esch TE (2009) Polymer 50:3575–3581
169. Hori T, Yoshimura K, Ohno H et al (2003) Tetrahedron 59:6301–6309
170. Cypryk M, Apeloig Y (1997) Organometallics 16:5938–5949
171. Kress JD, Leung PC, Tawa GJ et al (1997) J Am Chem Soc 119:1954–1960
172. Kress JD, Leung PC, Tawa GJ, Hay PJ (2000) In: Clarson SJ, Fitzgerald JJ, Owen MJ et al (eds) Silicones and silicone-modified materials, vol 729. ACS, Washington DC
173. Cypryk M, Kaźmierski K, Fortuniak W et al (2000) Macromolecules 33:1536–1545
174. White DP, Douglass W (2001) In: Cundari TR (ed) Computational organometallic chemistry. Marcel Dekker, New York
175. Rappé AK, Skiff WM, Casewit CJ (2000) Chem Rev 100:1435–1456
176. Resconi L, Cavallo L, Fait A et al (2000) Chem Rev 100:1253–1346

177. Cundari TR (ed) (2001) Computational organometallic chemistry. Marcel Dekker, New York
178. Schmidt MW, Gordon MS (1998) Annu Rev Phys Chem 49:233–266
179. Lee JW, Jo WH (2009) J Organomet Chem 694:3076–3083
180. Flisak Z, Ziegler T (2005) Macromolecules 38:9865–9872
181. Cavallo L, Del Piero S, Ducere JM et al (2007) J Phys Chem C 111:4412–4419
182. Busico V, Causa M, Cipullo R et al (2008) J Phys Chem C 112:1081–1089
183. Bhaduri S, Mukhopadhyay S, Kulkarni SA (2006) J Organomet Chem 691:2810–2820
184. Berkefeld A, Drexler M, Möller HM et al (2009) J Am Chem Soc 131:12613–12622
185. Yang SY, Szabo MJ, Michalak A et al (2005) Organometallics 24:1242–1251
186. Wang DQ, Tomasi S, Razavi A et al (2008) Organometallics 27:2861–2867
187. Ning Y, Caporaso L, Correa A et al (2008) Macromolecules 41:6910–6919
188. Wondimagegn T, Wang D, Razavi A et al (2008) Organometallics 27:6434–6439
189. Tomasi S, Razavi A, Ziegler T (2009) Organometallics 28:2609–2618
190. Wondimagegn T, Wang DQ, Razavi A et al (2009) Organometallics 28:1383–1390
191. Sauriol F, Wong E, Leung AMH et al (2009) Angew Chem Int Ed 48:3342–3345
192. Karttunen VA, Linnolahti M, Pakkanen TA et al (2008) Organometallics 27:3390–3398
193. Yang SY, Ziegler T (2006) Organometallics 25:887–900
194. Frech CM, Blacque O, Berke H (2006) Pure Appl Chem 78:1877–1887
195. Frech CM, Blacque O, Schmalle HW et al (2006) Chem Eur J 12:3325–3338
196. Fomine S, Ortega JV, Tlenkopatchev MA (2005) Organometallics 24:5696–5701
197. Fomine S, Tlenkopatchev MA (2007) Organometallics 26:4491–4497
198. Raab M, Schick G, Fondermann R et al (2006) Angew Chem Int Ed 45:3083–3086
199. Kudo T, Gordon MS (1998) J Am Chem Soc 120:11432–11438
200. Kudo T, Gordon MS (2000) J Phys Chem A 104:4058–4063
201. Cypryk M, Apeloig Y (2002) Organometallics 21:2165–2175
202. Bruice TC (2006) Chem Rev 106:3119–3139
203. Orozco M, Luque FJ (2000) Chem Rev 100:4187–4226
204. Lundberg M, Morokuma K (2007) J Phys Chem B 111:9380–9389
205. Siegbahn PEM, Himo F (2009) J Biol Inorg Chem 14:643–651
206. Brooks BR, Brooks CL, Mackerell AD Jr et al (2009) J Comput Chem 30:1545–1614
207. Cornell WD, Cieplak P, Bayly CI et al (1995) J Am Chem Soc 117:5179–5197
208. Christen M, Hünenberger PH, Bakowies D et al (2005) J Comput Chem 26:1719–1751
209. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) J Am Chem Soc 118:11225–11236
210. Acevedo O, Jorgensen WL (2009) Advances in quantum and molecular mechanical (QM/MM) simulations for organic and enzymatic reactions. Acc. Chem. Res. 43:142–151 (2010)
211. Gascon JA, Sproviero EM, Batista VS (2005) J Chem Theory Comput 1:674–685
212. Antoniou D, Basner J, Nunez S et al (2006) Chem Rev 106:3170–3187
213. Mordasini TZ, Thiel W (1998) CHIMIA 52:288–291
214. Case DA, Cheatham TE III, Darden T et al (2005) J Comput Chem 26:1668–1688
215. Guimaraes CRW, Repasky MP, Chandrasekhar J et al (2003) J Am Chem Soc 125:6892–6899
216. Repasky MP, Guimaraes CRW, Chandrasekhar J et al (2003) J Am Chem Soc 125:6663–6672
217. Roca M, Marti S, Andres J et al (2003) J Am Chem Soc 125:7726–7737
218. Marti S, Moliner V, Tunon I et al (2003) Org Biomol Chem 1:483–487
219. Tresadern G, Wang H, Faulder PF et al (2003) Mol Phys 101:2775–2784
220. Bash PA, Ho LL, MacKerrell AD Jr et al (1996) Proc Natl Acad Sci U S A 93:3698–3703
221. Van der Vaart A, Gogonea V, Dixon SL et al (2000) J Comput Chem 21:1494–1504
222. Heeger AJ (2001) Angew Chem Int Ed 40:2591–2611
223. Barta P, Kugler T, Salaneck WR et al (1998) Synth Met 93:83–87

224. Vaschetto ME, Retamal BA (1997) J Phys Chem A 101:6945–6950
225. Vaschetto ME, Retamal BA, Monkman AP et al (1999) J Phys Chem A 103:11096–11103
226. Kwon O, McKee ML (2000) J Phys Chem B 104:1686–1694
227. Tsai FC, Chang CC, Liu CL et al (2005) Macromolecules 38:1958–1966
228. Gama Amazonas J, Guimaraes JR, Cristina Santos Costa S et al (2006) J Mol Struct: THEOCHEM 759:87–91
229. Shen W, Li M, He R et al (2007) Polymer 48:3912–3918
230. Fu Y, Shen W, Li M (2008) Macromol Theory Simul 17:385–392
231. Sun H (1997) J Am Chem Soc 119:3611–3618
232. Fondermann R, Dolg M, Raab M et al (2006) Chem Phys 325:291–298

CHAPTER 7

EVALUATION OF PROTON TRANSFER IN DNA CONSTITUENTS: DEVELOPMENT AND APPLICATION OF AB INITIO BASED REACTION KINETICS

DMYTRO KOSENKOV[1,2], YANA KHOLOD[1,3], LEONID GORB[1,4], AND JERZY LESZCZYNSKI[1,5]

[1]*Interdisciplinary Center for Nanotoxicity, Department of Chemistry and Biochemistry, Jackson State University, Jackson, Mississippi, MI 39217, USA*

[2]*Department of Chemistry, Purdue University, West Lafayette, IN 47906, USA*

[3]*US DoE Ames Laboratory, Iowa State University, Ames, IA 50011, USA*

[4]*Department of Molecular and Quantum Biophysics, Institute of Molecular Biology and Genetics, National Academy of Sciences of Ukraine, Kyiv 03143, Ukraine*

[5]*Army Engineering Research and Development Center, Vicksburg, MS 39180, USA*

e-mail: jerzy@icnanotox.org

Abstract: The kinetics of chemical reactions characterizes the rates of chemical processes, i.e. distribution of all reactants, intermediates and products over time. This information is of vital importance for all areas of chemistry: chemical technology to control organic or inorganic syntheses, chemical construction of nanomaterials, as well as for the investigation of biochemical processes. The chemical kinetics data provide a possibility to investigate the effect of different chemical, physical and environmental factors on the rate of a reaction, final products and by-products distribution, and even the direction of a chemical process.

In the first part of the chapter the general introduction to the kinetics of chemical reactions is given. The classical kinetics of chemical reactions uses the outcome from experimental measurement of reaction rates. However, currently available reliable computational ab initio methods provide an alternative efficient way for estimation of the rate constants even for stepwise and multidirectional reactions. Another benefit of the computational investigations is the possibility to simulate a wide range of processes with duration from picoseconds to hours, days, or even for much longer time scales. Contemporary ab initio methods have been used for estimation and prediction of reaction rates for a number of different chemical reactions. Until recently most of the theoretical studies on kinetic parameters have not been extended beyond the calculations of the rate constants of chemical reactions. In the present review we describe the simulation of the chemical kinetics of proton transfer (tautomerization) in nucleic acid bases and their complexes with metal ions, also in the presence of water molecules. The considered models are based on the ab initio calculated rate constants of chemical reactions. Then, such predicted rate constants are used for further kinetic simulations. Biological consequences of investigated processes are also discussed.

P. Paneth and A. Dybala-Defratyka (eds.), *Kinetics and Dynamics*, Challenges and Advances in Computational Chemistry and Physics 12,
DOI 10.1007/978-90-481-3034-4_7,

Keywords: Chemical reactivity, Gas-phase experiments, Nucleobases, Laser desorption, Tautomerization, Point mutations

7.1. INTRODUCTION

Ab initio quantum chemistry methods, such as the Hartree–Fock (HF) [1, 2] theory, Møller–Plesset Perturbation (MP) Theory [3] and Density Functional Theory (DFT) [4] have been successfully applied for the studies of various static properties of chemical systems [5, 6]. During the last decade a considerable advance in development of methods capable to evaluate time-dependent properties of the molecular systems has been reported. First principles molecular reaction dynamics, e.g. Car-Parrinello [7, 8] and other DFT based methods [9–11] have been developed and applied to the study of diverse types of systems from semiconductors [12] and quantum dots [13] to biological molecules [14, 15]. These molecular dynamic approaches are based on the atomic level description of the system. In contrast to the molecular dynamics methods, the chemical kinetics is phenomenological theory based on reaction rates, which describe macroscopic properties of the system. Application of the basic principles of quantum chemistry along with statistical mechanics is combined in ab initio based chemical kinetics for estimations of the thermally-averaged rate constants. These rate constants determine probability of the system to follow a certain reaction coordinate.

At a glance ab initio based chemical kinetics can be compared with the molecular dynamics methods. There are several advantages of chemical kinetics approach in comparison to molecular dynamics:

1. Ability to simulate fast (femtoseconds) and slow (days or even years) reactions.
2. No constraints implied on the activation barriers of the considered reactions.
3. Effects of the temperature and solvents can be accounted for by using ab initio calculations, experimental estimations, or both sources together.
   The main limitation of the kinetic approach is the requirement of the implicit definition of the reaction coordinate in a simulation. Thus, a priori knowledge of the location of local minima and transitions states is required.

These advantages and limitations of the chemical kinetics simulation arise from the basic idea of the method. In the kinetic approach quantum chemical methods are used initially, for investigation of a potential energy surface (PES) of the chemical system. Local minima corresponding to reactants, products and transition states of the considered reactions are located. Then, analysis of the PES results in determination of the reaction coordinates. This evaluation of the molecular structures can be done with help of available quantum chemical software (e.g. Gaussian [16], GAMESS [17, 18])

Ab initio based reaction kinetics uses formally quite simple computational protocol:

1. The methods of statistical mechanics are applied for the estimation of the Gibbs free energies of the reactants, transition states and products involved in the reaction, along the selected reaction coordinates using information from the PES analysis.
2. Evaluation of free energies using statistical mechanics methods that follows the first step yields reaction rate constants.
3. The obtained reaction rate constants are used to compute the changes of populations of the reactants and products of the reactions.

Since, statistical mechanics is used for evaluation of reaction rates there are no limitations on considered reactions on the basis of the barrier heights. The reactions with any activation barriers can be considered. Such requirement however, creates considerable hurdle for molecular dynamics simulations. This feature also explains why fast (low activation barrier) and slow (high activation barrier) reactions can be easily considered using the kinetic approach. The effects of temperature are accounted for during thermochemical evaluations of free energies and rate constants in the kinetic approach. In addition, the solvent effects are usually included during ab initio quantum chemical investigations of the PES using explicit or implicit solvent models.

7.2. METHODOLOGY

7.2.1. Ab Initio Based Computation of Reaction Rates

In this section a brief overview of the calculation of the rate constants based on the ab initio obtained Gibbs free energies is given. Following [19, 20], in accordance with the standard transition state theory the equilibrium constant ($K_{TS}^{\neq}$) between the reactant R and activated complex TS (transition state) can be written as:

$$K_{TS}^{\neq} = \frac{[TS]}{[R]} \tag{7-1}$$

Then the change of the concentration of the product P can be expressed via the equilibrium constant $K_{TS}^{\neq}$ and frequency (ν) of transformation of the activated complex (TS) to the final product (P):

$$\frac{d[P]}{dt} = [R]K_{TS}^{\neq}\nu \tag{7-2}$$

Introducing the rate constant (k) for the reaction of the interest: $R \rightarrow P$, Eq. (7-2) can be rewritten:

$$\frac{d[P]}{dt} = k[R] \tag{7-2a}$$

where the rate constant k can be written as:

$$k = K_{TS}^{\neq} v \tag{7-2b}$$

The equilibrium constant $K_{TS}^{\neq}$ can be decomposed into two parts: $K_{trans}^{\neq}$ which corresponds to the translational motion along the one dimensional reaction coordinate and $K^{\neq}$, which corresponds to the rest of the degrees of freedom:

$$K_{TS}^{\neq} = K_{trans}^{\neq} K^{\neq} = \frac{(2\pi m k_b T)^{1/2}}{h} v \delta K^{\neq} \tag{7-3}$$

where m is the mass of the activated complex, δ is a range of reaction coordinate at the top of the barrier where the activated complex is located.

Since the product $v\delta$represents an averaged speed $\langle \upsilon \rangle$ at which the activated complex achieves the barrier top using Boltzmann distribution, the product can be expressed as follows:

$$\delta v_c = \langle \upsilon \rangle = \left(\frac{k_b T}{2\pi m} \right)^{1/2} \tag{7-4}$$

Then, after substituting Eq. (7-4) into Eq. (7-3), the rate constant can be written as:

$$k = \frac{k_b T}{h} K^{\neq} \tag{7-5}$$

The equilibrium constant $K^{\neq}$ of the formation of the activated complex can be expressed through the change of the Gibbs free energy between the reactant R and activated complex TS, so called activation energy:

$$\Delta G^{\neq} = -RT \ln(K^{\neq}) \tag{7-6}$$

Finally, substitution (7-6) into (7-5) yields:

$$k(T) = \frac{k_b T}{h} \exp\{-\Delta G^{\neq}/RT\} \tag{7-7}$$

In the derivation above the tunneling effect has been neglected.

One of the possibilities to introduce a tunneling is to use first order Wigner's correction [21]. In this approach the rate constant is represented as:

$$k(T) = \kappa \frac{k_b T}{h} \exp\{-\Delta G^{\neq}/RT\} \tag{7-8}$$

where κ is the transmission coefficient, which accounts for quantum tunneling effect [21]; $\Delta G^{\neq}$ is the activation energy (the difference in Gibbs free energy between the reactant and transition state).

$$\kappa = 1 + \frac{1}{24}\left(\frac{\hbar\left|\omega^{\neq}\right|}{k_b T}\right)^2 \tag{7-9}$$

Another approach that accounts for the tunneling correction during the calculation of the rate constants is an instanton method [22]. This method is based on the consideration of the action integrated along the instanton path. The instanton path is the trajectory, which follows the minimum energy path near the equilibrium configuration, and follows a shorter and higher energy path in the region of the transition state. This trajectory corresponds to a combination of path length and energy that minimizes the corresponding instanton action. In this method the tunneling rate constant is calculated as follows:

$$k(T) = \frac{\Omega_0^i}{2\pi}\exp\{-S_I(T)\} \tag{7-10}$$

where Ω_0^i is the effective tunneling frequency in the metastable equilibrium (tautomeric) configuration, and $S_I(T)$ represents the multidimensional instanton action (derived from the integration along the instanton path) measured in units of $\hbar$.

7.2.2. Numerical Solution of a System of Rate Equations

As it is well known, equations which describe the rate of chemical reactions generate the system of linear ordinary differential equations. It relates concentrations of reactants and products of a chemical reaction and their change over time. The rates of forward and reverse reactions are proportional to the concentrations ($n_j^{(s)}$) of the reactants and products (Eq. (7-11)) and depend on the rate constants ($k_{i,j}$), which can be deduced from ab initio computations using Eq. (7-8) or Eq. (7-10).

$$\frac{dn_i^{(s)}}{dt} = \sum_{\substack{j=1 \\ (i\neq j)}}^{N} n_j^{(s)} k_{i,j} - \sum_{\substack{l=1 \\ (l\neq i)}}^{N} n_i^{(s)} k_{l,i}, \quad i = 1...N \tag{7-11}$$

Various numerical methods are available [23] for the solution of the equations of this type. One of the simplest and efficient approximations [23] is represented by numerical integration of rate equations (Eq. (7-11)) using Runge–Kutta method of fourth order [23].

The numerical scheme based on the linear interpolation of the values of concentrations inside the time interval Δt could be presented by the following set of equations:

$$
\begin{aligned}
A_{i,1}^{(s)} &= f_i\left(t_s, n_i^{(s)}\right) \\
A_{i,2}^{(s)} &= f_i\left(t_s + \frac{\Delta t}{2}, n_i^{(s)} + \frac{\Delta t}{2} A_{i,1}^{(s)}\right) \\
A_{i,3}^{(s)} &= f_i\left(t_s + \frac{\Delta t}{2}, n_i^{(s)} + \frac{\Delta t}{2} A_{i,2}^{(s)}\right) \\
A_{i,4}^{(s)} &= f_i\left(t_s + \Delta t, n_i^{(s)} + \Delta t A_{i,3}^{(s)}\right)
\end{aligned}
\tag{7-12a}
$$

$$
f_i\left(t_s, n_i^{(s)}\right) = \sum_{\substack{j=1 \\ (i \neq j)}}^{N} n_j^{(s)} k_{i,j} - \sum_{\substack{l=1 \\ (l \neq i)}}^{N} n_i^{(s)} k_{l,i}
$$

$$
n_i^{(s+1)} = n_i^{(s)} + \frac{\Delta t}{6}\left(A_{i,1}^{(s)} + 2A_{i,2}^{(s)} + 2A_{i,3}^{(s)} + A_{i,4}^{(s)}\right)
$$

$$
t_{s+1} = t_s + \Delta t \tag{7-12b}
$$

where Δt is the time step of the integration, t_s represents s^{th} step of the integration, $A_{i,1}^{(s)} \ldots A_{i,4}^{(s)}$ are the parameters used for the interpolation (for details see reference [23]).

Finally, to propagate the concentrations ($n_j^{(s)}$) with time using Eq. (7-11), the initial ($t = 0$) concentrations should be specified:

$$
n_i^{(0)} = c_i \tag{7-13}
$$

where c_i is the initial concentration of i^{th} component of the mixture.

This approach based on ab initio calculated rate constants has been implemented recently as a computer code and applied to studies of the tautomerization kinetics in the DNA bases [24].

7.3. APPLICATIONS OF THE REACTION KINETICS MODELS TO THE STUDIES OF PROTON TRANSFER IN DNA CONSTITUENTS

The described above methodology has been recently applied for extensive studies of the process of proton transfer in the DNA constituents. In particular, kinetics of proton transfer in the gas phase and in the presence of water molecules, metal ions and phosphate groups has been investigated [25–29]. The following sections review the application of those kinetic simulations.

7.3.1. Tautomerization of Nucleobases in the Gas Phase

Nucleotides are the building blocks of the nucleic acids. They have been subject to a number of experimental [30–37] and computational studies [30, 31, 38] due to their biological and chemical importance.

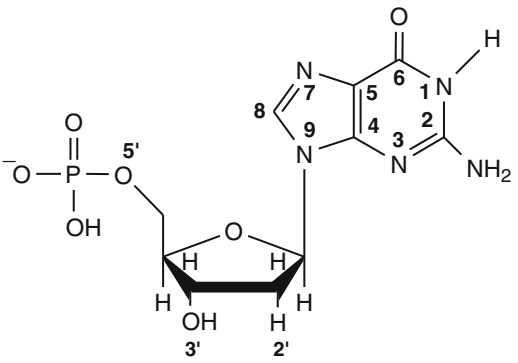


Figure 7-1. 5′-Guanosine-monophospate in anionic form (Reproduced by permission of American Chemical Society [27])

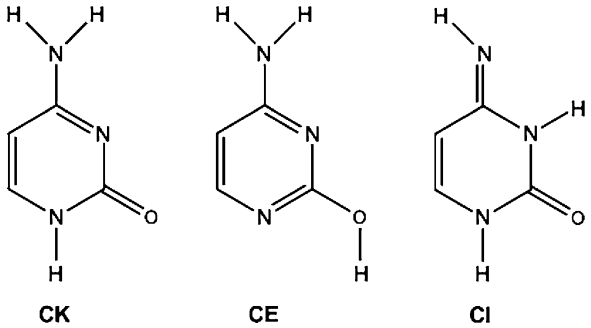


Figure 7-2. Cytosine canonic keto (CK) and rare enol (CE), imino (CI) tautomers (Reproduced by permission of American Chemical Society [24])

Each nucleotide (Figure 7-1) consists of several units: a phosphate group, ribose ring, and nucleic acid base (Figures 7-2 and 7-3). The most abundant nucleic acid bases are cytosine, guanine, adenine, thymine and uracil.

According to generally accepted paradigm, which originates from the famous idea of Watson and Crick, the existence of nucleobases in a canonic structure is one of the most important requirements for storage and transduction of hereditary information [30]. However, Watson and Crick also hypothesized [39] that extremely small concentrations of noncanonic forms of nucleic bases (‘rare’ tautomers), which are different from canonic tautomer by a location of just a single proton, are responsible for occurring of spontaneous point mutations. To study theoretically the probability of those mutations one needs the information on the rate constants and kinetics of tautomerization. Nowadays, such information is available from the high level ab initio calculations of the tautomerization processes in isolated or hydrated nucleobases that are considered as a simplest model to understand the tautomeric properties of nucleic acids [40, 41]. Those ab initio calculations suggest that only cytosine and guanine could contribute to the frequency of spontaneous point mutations

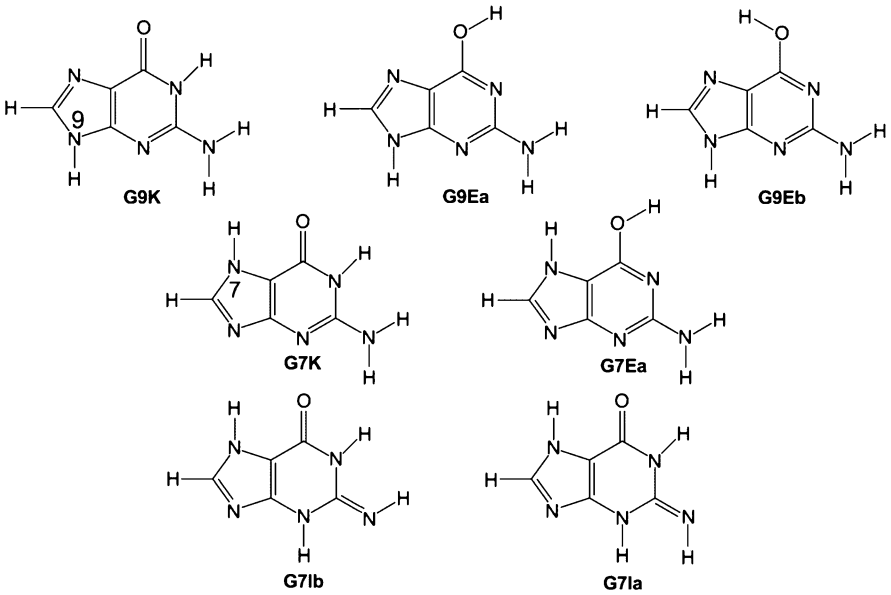


Figure 7-3. Guanine canonic *keto* (G9K), *keto* (G7K), *enol* (G9Ea, G9Eb, G7Ea), and *imino* (G7Ia, G7Ib) tautomers (Reproduced by permission of American Chemical Society [24])

through the mechanism of tautomerization, since only these two molecules have low energy 'rare' tautomeric forms. The structures of those tautomers along with the canonic form are presented in Figures 7-2 and 7-3.

As it was already mentioned, the nucleobases in the canonic *keto* forms are absolutely dominating in the DNA molecule. Tautomers of nucleic acid bases could cause spontaneous point mutations with the frequency of 10^{-8}–10^{-10} (in vivo) and 10^{-6}–10^{-5} (in vitro) [42, 43], relative to concentration of nucleotide per nucleotide synthesized. It means that their concentrations are so small that they hardly could be detected even with modern experimental techniques.

However, in contrast to the biochemical environments, several cytosine and guanine rare tautomers were unambiguously shown to co-exist in experiments performed in the gas phase or in noble gas matrix [44, 45]. Therefore, experimental techniques such as matrix isolation [44, 45], He-nanodroplets [46, 47] and resonance enhance multi photon ionization (REMPI) spectroscopy [40, 48–51] are the main sources of experimental information on physical and chemical properties of canonic and rare tautomers. Unfortunately, each technique requires very different experimental conditions. Thus, some discrepancy between results obtained in these experiments exists. Furthermore, the REMPI spectroscopy is limited in ability to measure concentrations of the observed tautomers.

Theoretical calculations could assist evaluation of the experimental outcome. The ab initio based kinetic simulation of these experiments can be performed in order to reveal detail mechanism of tautomer formation and to estimate relative

populations of the tautomers produced in these experiments. Below the kinetic simulations of tautomerization processes in the gas phase experiments, in presence of water molecules and hydrated sodium cations are reviewed.

7.3.1.1. Kinetic Model for the Gas Phase Experiments

As it was already mentioned, the results of matrix isolation [44], He-nanodroplets [46, 47] and resonance enhance multi photon ionization spectroscopy [40, 48–51] experiments on nucleobases sometimes differ in assigning the structures and concentration of the observed tautomers. Moreover, REMPI spectroscopy is limited in ability to measure concentrations of the observed tautomers. The kinetic simulations [24] have been performed to overcome these restrictions of the experimental techniques and to estimate relative populations of the tautomers observed in the experiments.

Some initial assumptions are necessary before simulations. It was assumed in the performed study [24] that after thermal or laser induced desorption the nucleobases are present only in form of monomers and dimers and all the tautomeric transitions occur in the gas phase. The considered dimers are shown in Figure 7-4 (cytosine) and Figure 7-5 (guanine).

Thus, the desorption of DNA bases has been described as follows:

1. Formation of a dimer D from the nucleobases M_1 and M_2
2. Tautomerization of two monomers via the bimolecular mechanism with formation of the tautomeric dimer D*
3. Dimer decomposition with formation of two tautomers M^*_1 and M^*_2:

$$M_1 + M_2 \overset{K_D}{\rightleftarrows} D \underset{k_D^{-1}}{\overset{k_D}{\rightleftarrows}} D^* \overset{K_D^*}{\rightleftarrows} M_1^* + M_2^* \tag{7-14}$$

The monomolecular mechanism of transformation for each canonic form (M_i) to tautomer (M_i^*) has been considered using a simple scheme:

$$M_i \underset{k_M^{-1}}{\overset{k_M}{\rightleftarrows}} M_i^* \tag{7-15}$$

These equilibrium relations determine relative concentrations of the canonic (D) and rare (D*) dimers:

$$\begin{aligned} K_D &= \frac{D}{M_1 M_2} \\ K_D^* &= \frac{D^*}{M_1^* M_2^*} \end{aligned} \tag{7-16}$$

where K_D and K_D^* are the equilibrium constants for canonic and rare dimers, respectively. Laser desorption represents the evaporation technique widely used in the REMPI experiments [40, 48–51]. The rate constant of solid sample

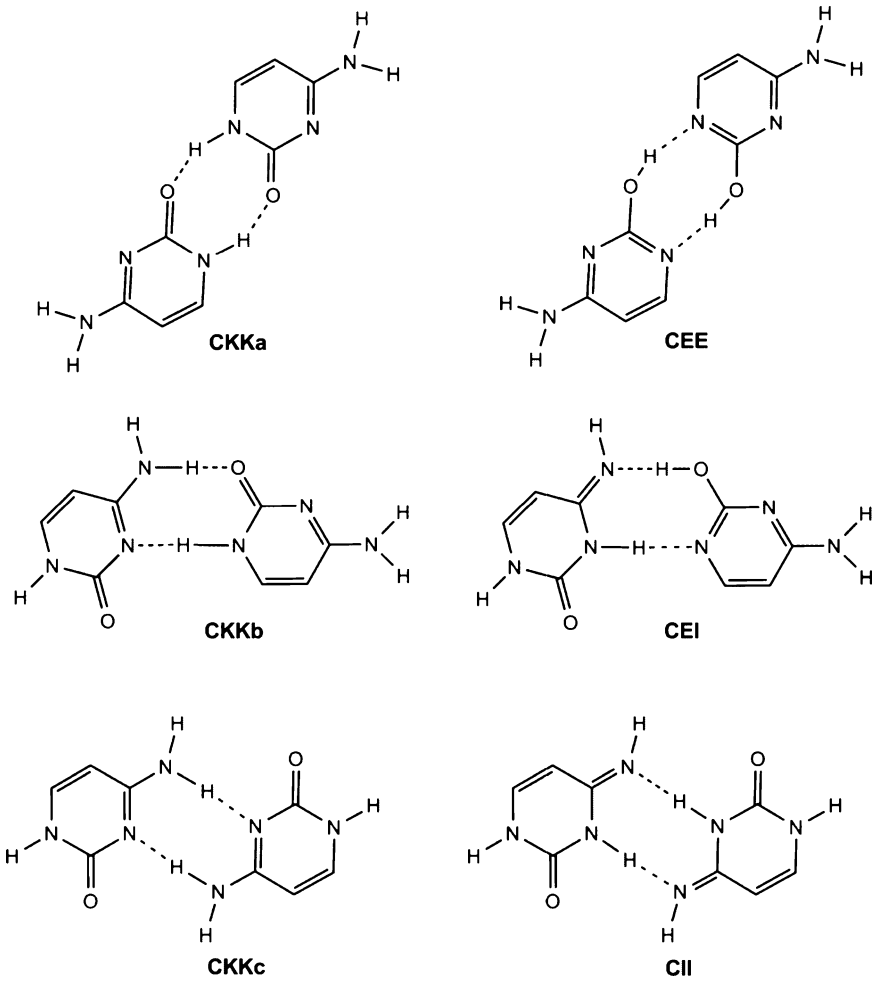


Figure 7-4. Considered dimers of cytosine (Reproduced by permission of American Chemical Society [24])

evaporation (k_{ev}) for all tautomeric forms depends on the laser pulse parameters and intermolecular interactions on the solid sample surface. Thus, the evaporation rate constant k_{ev} is the only empirical parameter. The kinetic equations describing laser induced desorption of the tautomer can be written as:

$$\begin{aligned} &\frac{dL^{gas}}{dt} = k_{ev}L^{solid} \\ &\frac{dL^{solid}}{dt} = -k_{ev}L^{solid} \\ &L \in \{M_i, M_i^*, D, D^*\} \end{aligned} \tag{7-17}$$

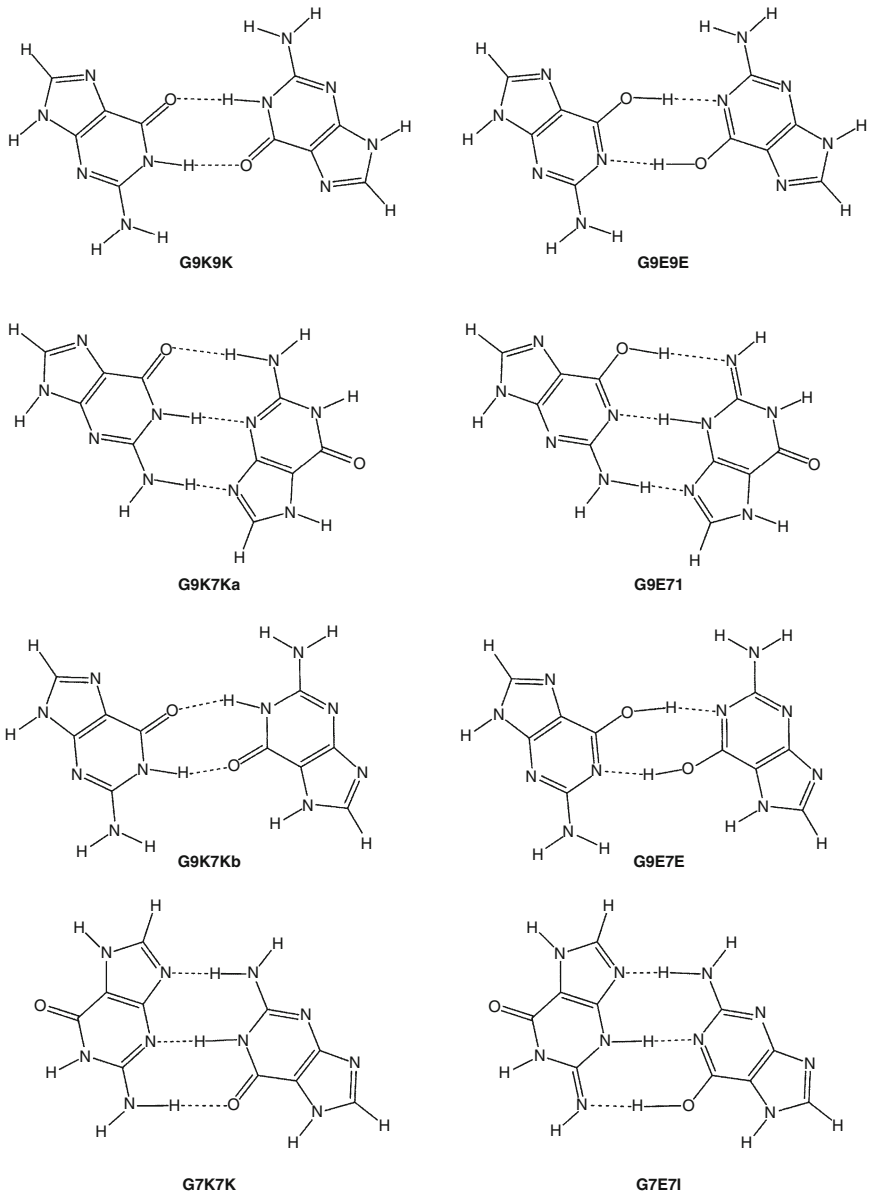


Figure 7-5. Considered dimers of guanine (Reproduced by permission of American Chemical Society [24])

where L^{solid} denotes the population (on the surface) in the solid phase of canonic and rare tautomers in both monomolecular and dimer forms; L^{solid} represents corresponding populations after the evaporation in the gas phase, k_{ev} is the rate constant of evaporation.

It should be also emphasized that no electronic excitations in nucleobases take place during the laser desorption [24]. The rate constants for the rate equations have been estimated from ab initio computations using Eq. (7-8). The tunneling effect in Eq. (7-8) has been neglected due to high temperatures of the evaporation.

7.3.1.2. Ab Initio Based Kinetics of Laser-Induced Desorption

The Gibbs free energies have been calculated using MPWB1K functional [52], which is specifically designed for estimation of the thermochemical and kinetic parameters. The rate constants have been calculated using Eq. (7-8). The activation barriers and rate constants are presented for cytosine in Table 7-1 and for guanine in Table 7-2.

The kinetic curve representing the laser induced desorption of cytosine (Figure 7-6) consists of two stages:

1. Fast tautomerization via the bimolecular mechanism
2. Evaporation of all cytosine forms with the most abundant CK and CE forms

According to the Figure 7-6 the first tautomerization stage lasts approximately 10^{-9} s and is characterized by fast depletion of CK form and fast rise of CE on the solid surface, due to very fast tautomerization through the bimolecular mechanism with rate constant of 10^{11} s^{-1}. The mechanism described by the scheme: $2CK \rightarrow CKKa \rightarrow CEE \rightarrow 2CE$ is the most favorable. At the second stage, after 2×10^{-8} s period, desorption of the tautomers dominates and mixture of tautomers starts to evaporate into the gas phase.

The concentrations of CK and CE forms are the highest and the process of desorption ends within 1×10^{-7} s, based on the rate of desorption. It means that all considered equilibria, except CK↔CE, are strongly shifted towards stabilization of canonic CK form. This is a plausible reason for the absence of the *imino* form in REMPI spectra of cytosine [48, 49].

Table 7-1 Calculated at MPWB1K/aug-cc-pVDZ level kinetic parameters (activation barrier $\Delta G_f^{\neq}$, kcal/mol, forward rate k_f, s^{-1}, reverse rate k_r, s^{-1}) of tautomerization in cytosine at T = 1,000 K under laser desorption conditions (Reproduced by permission of American Chemical Society [24])

	$\Delta G_f^{\neq}$	k_f	k_r
CK ⇄ CE	35.72	3.25×10^5	4.71×10^5
CK ⇄ CI	44.13	4.72×10^3	7.44×10^4
CKKa ⇄ CEE	4.81	1.85×10^{12}	3.69×10^{13}
CKKb ⇄ CEI	13.89	1.92×10^{10}	2.14×10^{12}
CKKc ⇄ CII	17.11	3.80×10^9	3.08×10^{11}

See Figures 7-2 and 7-4 for description of the molecules

Table 7-2 Calculated at MPWB1K/aug-cc-pVDZ level kinetic parameters (activation barrier $\Delta G_f^{\neq}$, kcal/mol, forward rate k_f, s^{-1}, reverse rate k_r, s^{-1}) of tautomerization in guanine at T = 1,000 K under laser desorption conditions (Reproduced by permission of American Chemical Society [24])

	$\Delta G_f^{\neq}$	k_f	k_r
G9K ⇄ G9Ea	34.75	5.30×10^5	4.87×10^5
G9Ea ⇄ G9Eb	9.39	1.85×10^{11}	1.88×10^{11}
G7K ⇄ G7Ea	37.57	1.28×10^5	5.48×10^5
G7K ⇄ G7Ib	43.79	5.60×10^3	1.20×10^5
G7Ib ⇄ G7Ia	19.98	8.95×10^8	9.22×10^8
G9K9K ⇄ G9E9E	11.00	8.22×10^{10}	1.00×10^{13}
G9K7Ka ⇄ G9E7I	16.10	6.31×10^9	1.71×10^{12}
G9K7Kb ⇄ G9E7E	11.88	5.28×10^{10}	1.04×10^{13}
G7K7K ⇄ G7E7I	18.61	1.78×10^9	2.01×10^{12}

See Figures 7-3 and 7-5 for description of the molecules

It was also shown [24] that the distribution of the cytosine tautomers is close to the equilibrium populations at 1,000 K. As follows from Figure 7-7, guanine tautomerization under laser desorption conditions is more complex than the corresponding process for cytosine, due to the larger number of intermediates and products. The principal tautomerization pathway is the following: $2G9K \rightarrow G9K9K \rightarrow G9E9E \rightarrow 2G9Ea$.

It was found that in spite of the high rate of laser desorption (4.17×10^7 s^{-1}) bimolecular tautomerization (10^9–10^{12} s^{-1}) is many orders of magnitude faster than the rate of laser desorption. The presented results of the simulation suggest the domination of the thermodynamically most stable *keto* and *enol* guanine forms in the gas phase mixture [24]. The absence of three major components of the mixture (G9K, G7K, G9Ea) in the spectra [51] should be ascribed to the limitations of the REMPI detection scheme itself.

One of the reasons for the absence of the four most stable forms in the spectrum was believed to be due to the intrinsic ultrafast radiationless $S_1 \rightarrow S_0$ transitions that prevent their observation [51, 53, 54].

Finally, we would like to consider the effect of change of the rate constant of the laser desorption (k_{ev}) on the relative populations of the guanine tautomers. For the considered cases of laser evaporation of cytosine and guanine the constant value of the rate constant of evaporation has been estimated as $k_{ev} = 4.17 \times 10^7$ s^{-1} [24]. The effect of the variation of this rate constant on the relative populations of the tautomers in the gas phase is presented in Figure 7-8. One can notice that the slower evaporation is (longer time of evaporation) the lower are relative concentrations of the most stable *keto* forms (G9K, G7K) and higher concentrations of the *enol* and *imino* tautomers. It can be concluded that varying parameters of the laser pulse and

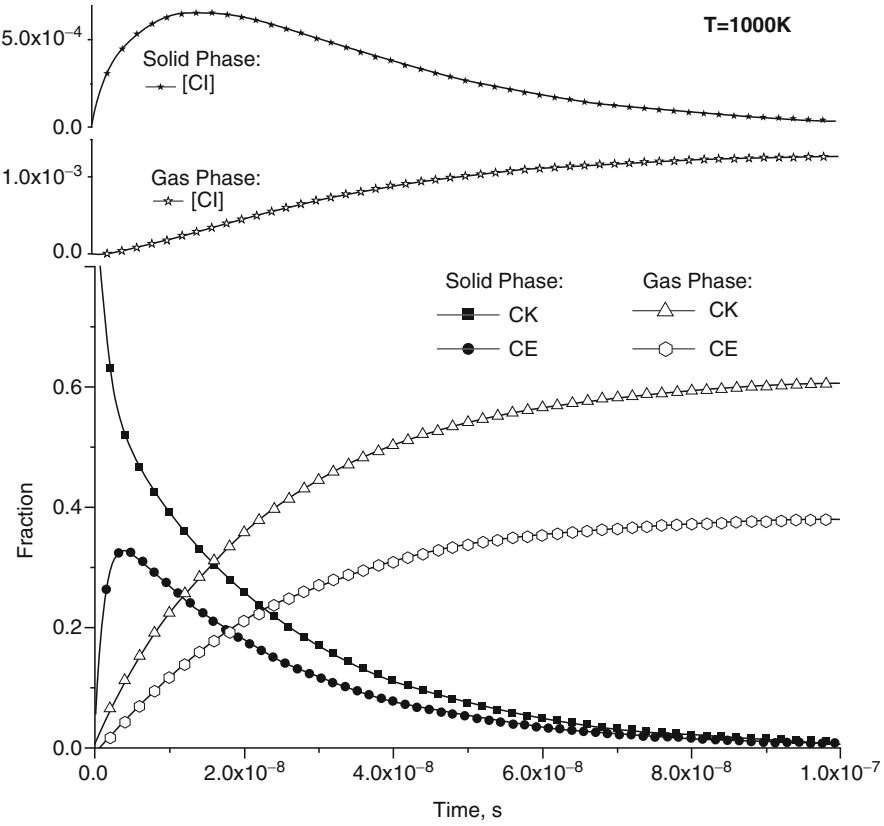


Figure 7-6. Kinetics of cytosine tautomerization under laser desorption conditions at 1,000 K with the rate constant of evaporation (k_{ev}) 4.17×10^7 s^{-1} (Reproduced by permission of American Chemical Society [24])

consequently changing k_{ev} one can select the mode optimal for registration of the low-concentration *imino* tautomers.

Therefore, the present simulations suggest that the laser induced tautomerization of nucleic bases proceeds via the bimolecular mechanism. The simulated data are in close correspondence to the results obtained experimentally, and thus enhance experimental findings by prediction that a laser-induced desorption of cytosine and guanine tautomers should proceed instantly to the equilibrium state at the temperature of desorption.

However, the presented results cannot be applied directly to the explanation of tautomerization mechanisms in living cell, at least because of the fact that tautomerization in cell takes place at room temperature. In addition, the DNA and RNA bases are surrounded by water molecules and ions of metals (mainly Na^+). Very simple models that are able to approximate such situations are presented in the next sections.

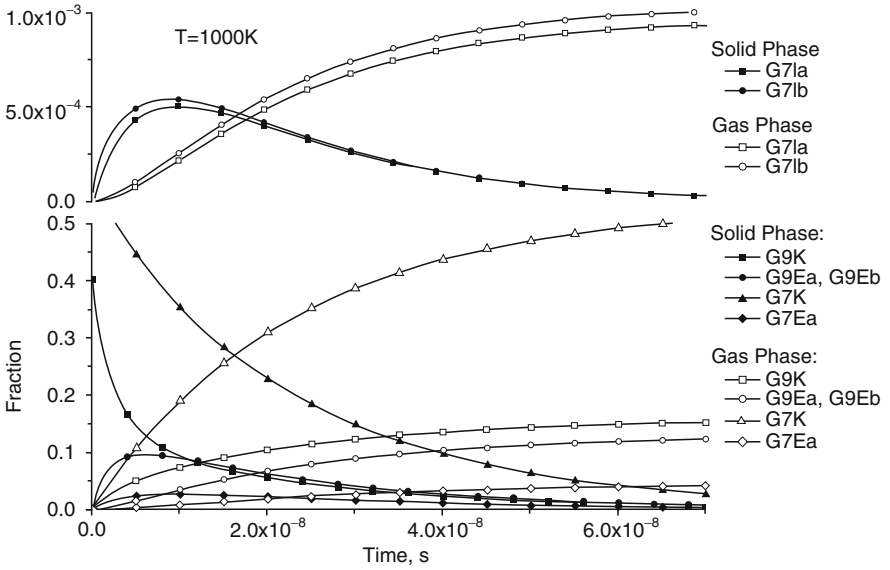


Figure 7-7. Kinetics of guanine tautomerization induced by laser desorption mixture at T = 1,000 K with the rate constant of evaporation (k_{ev}) $4.17 \times 10^7\ s^{-1}$ Note: Only tautomers having major contributions are shown (Reproduced by permission of American Chemical Society [24])

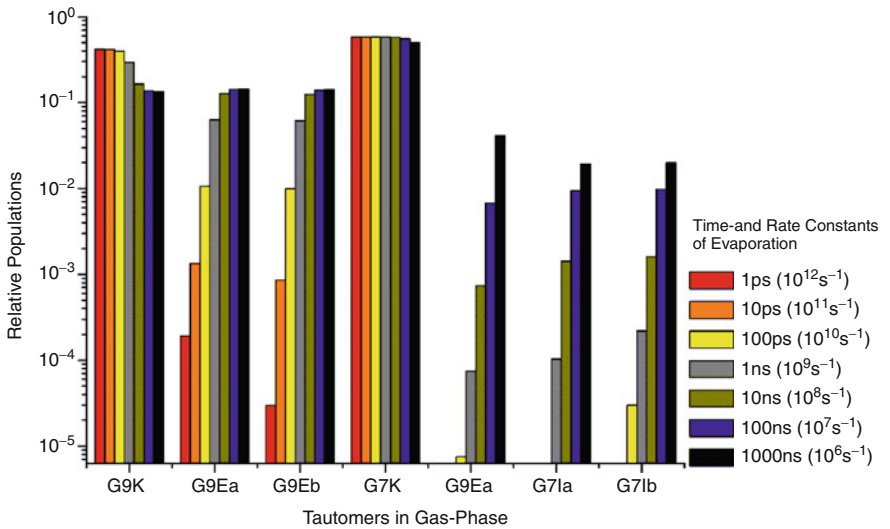
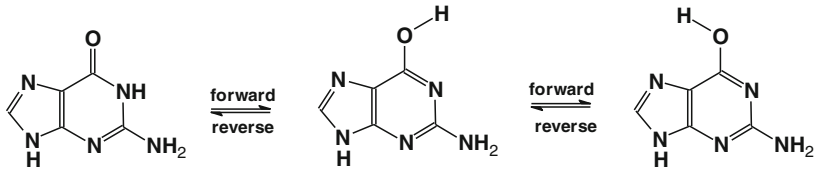


Figure 7-8. The effect of variation of the rate constant of laser desorption: $10^6 < k_{ev} < 10^{12}\ s^{-1}$ on the relative populations of the guanine tautomers

7.3.2. Tautomerization of Isolated and Monohydrated Cytosine and Guanine at Room Temperature

7.3.2.1. Tautomerization of Isolated Species

The modeling of tautomerization kinetics for isolated guanine has been carried out recently [25]. The following equilibria have been considered:



G9K G9Ea G9Eb

(Reproduced by permission of American Chemical Society: reference [25])

The geometry optimizations and vibrational analysis have been performed using the MP2/6-311++G(d, p) level of theory. Based on the obtained data the kinetic parameters have been calculated. In view of the fact that the proton transfer has been studied at room temperature, the values of rate constants have been calculated within the instanton approach [22]. They are collected in Table 7-3. The system of rate equations for the reactions displayed in the scheme above has been solved analytically and predicted kinetic curves are shown in Figure 7-9

According to the results presented in Table 7-3 and the kinetic curve in Figure 7-9, the gas-phase tautomerization processes are very slow. The reactions reach the equilibrium in about 2,500 h (104 days). However, it has been concluded [25] that even such a slow chemical reaction still could be considered as the source of point mutations.

Table 7-3 Calculated at MP2/6-311++G(d, p) level thermodynamic (relative Gibbs free energy ΔG, kcal/mol, equilibrium constant K_{eq}) and kinetic parameters (activation barrier $\Delta G_f^{\neq}$, kcal/mol, forward rate k_f, s^{-1}, reverse rate k_r, s^{-1}) of the guanine tautomerization reactions at T = 298 K (Reproduced by permission of American Chemical Society [25])

	ΔG	K_{eq}	$\Delta G_f^{\neq}$	k_f	k_r
G9K ⇄ G9Ea	0.01	9.36×10^{-1}	33.72	2.2×10^{-7}	2.35×10^{-7}
G9Ea ⇄ G9Eb	0.72	3.83×10^{-2}	8.09	6.9×10^{6}	1.8×10^{8}

See Figure 7-3 for description of the tautomers

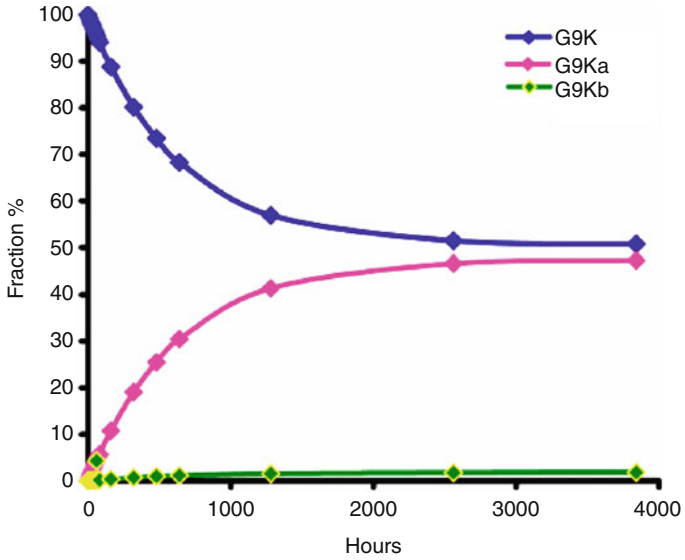
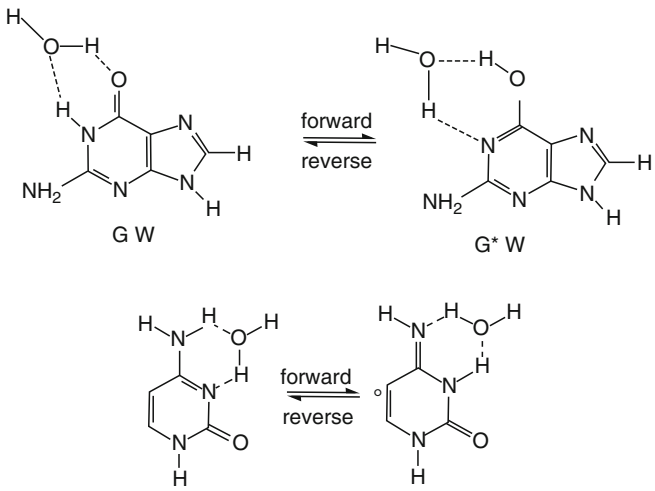


Figure 7-9. Ab initio predicted kinetics of guanine tautomerization (Reproduced by permission of American Chemical Society [25])

7.3.2.2. Tautomerization of Monohydrated Species

To study tautomerization kinetics of monohydrated nucleic acid bases the following tautomeric equilibria have been considered [28]:



(Courtesy of Molecular Diversity Preservation International, Basel, Switzerland. (Reproduced under Creative Commons Attribution: reference [28])

Table 7-4 Calculated thermodynamic (relative Gibbs free energy ΔG, kcal/mol, equilibrium constant K_{eq}) and kinetic parameters (activation barrier $\Delta E_f^{\neq}$, kcal/mol, forward rate k_f, s^{-1}, reverse pate k_r, s^{-1}) of the tautomerization reactions at T = 298 K (Courtesy of Molecular Diversity Preservation International, Basel, Switzerland. Reproduced under Creative Commons Attribution [28])

	ΔG	K_{eq}	$\Delta E_f^{\neq}$	k_f	k_r
C ⇄ C*[a]	0.82	2.5×10^{-1}	44.41	1.0×10^{-10}	4.4×10^{-10}
C·W ⇄ C*·W [a]	2.07	3.0×10^{-2}	20.38	1.8×10^{2}	5.7×10^{3}
G ⇄ G* [a]	2.38	1.8×10^{-2}	36.15	1.5×10^{-7}	8.6×10^{-6}
G·W ⇄ G*·W [a]	3.65	2.1×10^{-3}	14.74	9.5×10^{4}	4.5×10^{7}

Canonic form of cytosine (C), *Imino* tautomer of Cytosine (C*), Canonic G9K Guanine, *Enol* tautomer of Guanine (G*), Water molecule (W)
[a]MP2/6-31G(d) [28]

The calculations have been carried out at the MP2/6-31G(d) level of theory. The rate constants have been calculated within instanton approach [22] using equation (Eq. (7-10)). The thermodynamic and kinetic parameters of the tautomerization process are collected in Table 7-4.

The most profound difference between isolated and monohydrated bases is observed for the height of the tautomerization barriers $\Delta E_f^{\neq}$ (see Table 7-4). The absence of water makes the process of proton transfer very slow. The addition of one water molecule decreases activation barrier by about one half. This result in increases of the rate constant by 11–12 orders of magnitude (Table 7-4).

The data presented in the Table 7-6 have been also used to estimate a possible effect on the frequency of point spontaneous mutations in *E. coli*. It was found that equilibrium tautomerization constants for both hydrated cytosine and guanine are several orders of magnitude higher than experimentally observed frequency of spontaneous point mutations (experimentally measured is in the range of 10^{-5}–10^{-10} [42, 43]). In contrast, the kinetics of tautomerization that includes non-hydrated species fits reasonably well to both – the range of frequencies and time of DNA synthesis in the cells of *E. coli*.

7.3.3. Role of Hydrated Metal Ions for Nucleic Acids Stabilization

In this section we will discuss the kinetics of tautomerization of methyl-cytosine under model biochemical conditions, which are characterized by the presence of hydrated sodium ion.

Among a number of different metal ions which are important for the interaction with DNA bases [30, 55–57] the alkali cations play a vital role. It has been found [58–60] that at low concentration, the alkali cations interact with the negatively charged phosphate groups of DNA and stabilize the structure due to charge

Table 7-5 Calculated at the MP2/cc-pVDZ//B3LYP/cc-pVDZ level thermodynamic (relative Gibbs free energy ΔG, kcal/mol, equilibrium constant K_{eq}) and kinetic parameters (activation barrier $\Delta G_f^{\neq}$, kcal/mol, forward rate k_f, s^{-1}, reverse rate k_r, s^{-1}) of the proton transfer reactions in methyl cytosine complexes at T = 298 K (Reproduced by permission of American Chemical Society [26])

	ΔG	K_{eq}	$\Delta G_f^{\neq}$	k_f	k_r
MeC	2.4	1.7×10^{-2}	40.3	1.2×10^{-13}	7.1×10^{-12}
MeCW	2.1	2.9×10^{-2}	15.6	7.9×10^{5}	2.7×10^{7}
$MeCNa^{+}W$	19.0	1.2×10^{-14}	25.2	3.2×10^{-16}	2.7×10^{-2}
$MeCNa^{+}2W$	16.5	8.4×10^{-13}	23.0	3.8×10^{-17}	4.5×10^{-5}

See Figure 7-5 for description of the complexes

Table 7-6 Relative Gibbs free energies of the base pair dimerization (ΔG_{dim}), transition state ($\Delta G^{\neq}$), rare tautomeric forms ($\Delta G_{C\rightarrow R}$), and dimerization ($\Delta G^{r}_{dim}$) of the rare tautomeric forms (in kcal/mol) (Reproduced by permission of American Chemical Society: reference [29])

Method	ΔG_{dim}		$\Delta G^{\neq}$		$\Delta G_{C\rightarrow R}$		ΔG^{r}_{dim}	
	AT	GC	AT	GC	AT	GC	AT	GC
B3LYP/6-31G(d)	0.3	−11.8	13.8	13.4	15.2	11.1	−11.3	−5.4
MP2/6-31G(d)	0.1	−10.5	15.1	14.6	16.4	9.7	−9.8	−3.9
MP2/cc-pVDZ// MP2/6-31G(d)	0.7	−8.6	11.4	10.8	13.3	6.7	−9.3	−2.8
MP2/cc-pVTZ// MP2/6-31G(d)	−1.4	−12.1	10.2	10.6	12.5	7.6	−11.8	−5.4
MP2/infinite// MP2/6-31G(d)	−3.0	−14.5	9.4	10.2	12.1	7.8	−13.6	−7.2

neutralization [58–60]. At higher concentrations the ions interact with the nucleic acid bases so that the ions can destroy hydrogen bonding in the base pairs which leads to denaturation of the nucleic acid structure [58–60].

The effects of sodium cations on a water-assisted proton of methyl cytosine have been investigated recently [26]. The canonic *keto* form of methyl cytosine (MeC) and its corresponding *imino* tautomeric form (MeC*) have been studied at the MP2/cc-pVDZ//B3LYP/cc-pVDZ level of theory. The rate constants of tautomerization have been calculated using the instanton approach. In order to study effect of Na^{+} ions and water molecules on proton transfer in the complexes of MeC with one (MeCW), two (MeC2W) water molecules, and the corresponding complexes with sodium cation ($MeCNa^{+}W$, MeC $Na^{+}2W$) have been considered. The optimized structures obtained in [26] are shown in Figure 7-10.

The following tautomerization scheme has been assumed in the study:

MeC↔MeC*

MeCW↔MeC*W

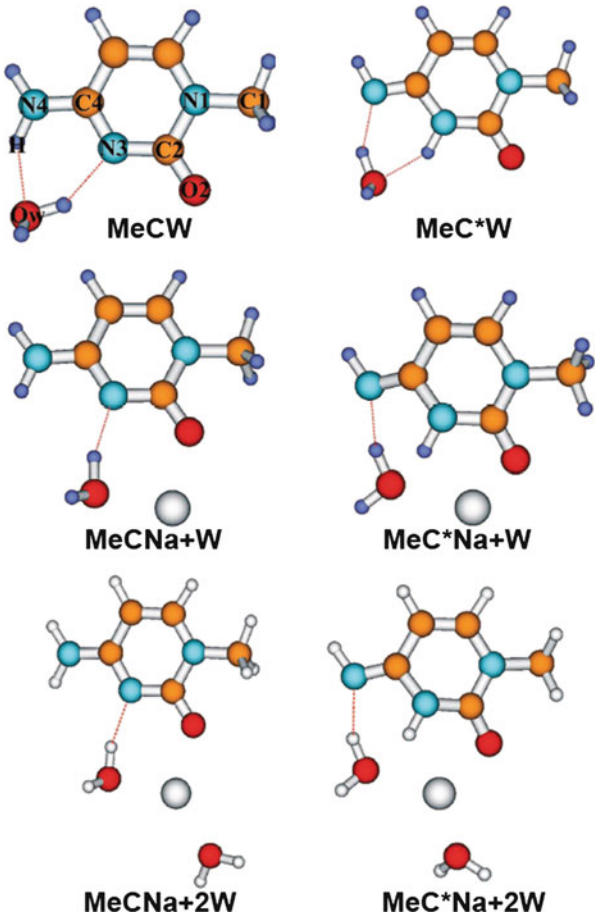


Figure 7-10. Optimized structures of the considered complexes of methyl cytosine (Reproduced by permission of American Chemical Society [26])

$MeCNa^+W \leftrightarrow MeC^*Na^+W$

$MeCNa^+2W \leftrightarrow MeC^*Na^+2W$

The thermodynamic and kinetic parameters characterizing the proton transfer are collected in Table 7-5. We believe that the most important finding of this study is that the interaction of MeC* with Na^+ has dramatic effect on the rate constants of tautomerization, and makes concentration of MeC* coordinated by hydrated Na^+ absolutely negligible.

As follows from Figure 7-11, the kinetics that characterizes the tautomerization could be very different and depends on the type of species involved in the tautomerization. The first type characterizes a very fast chemical reaction represented by the equilibration of MeC*W tautomer. The concentration of MeC*W undergoes the

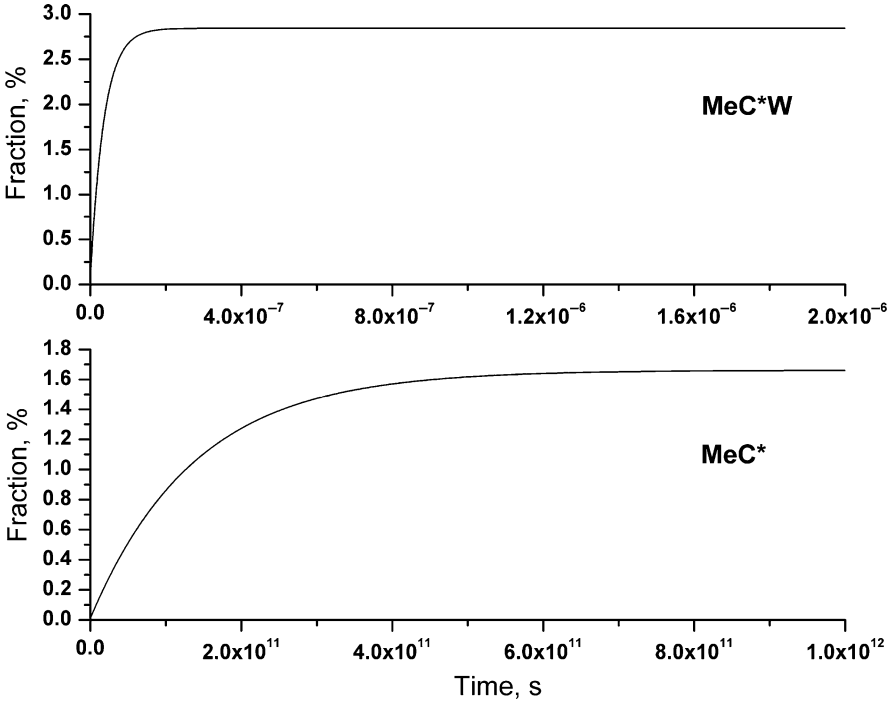


Figure 7-11. Relative concentrations of tautomers (MeC*W and MeC*) with respect to their canonic forms (MeC and MeCW) (Reproduced by permission of American Chemical Society [26])

most significant changes and reaches the equilibrium concentration (2.9×10^{-2}) within about 2×10^{-7} s (see Figure 7-11).

The second type of kinetics involves equilibration of MeC*, MeC*Na^+W, and MeC*Na+2W species. The equilibration time for this kinetics type is 10^{11}–10^{16} s. During this period, MeC*, MeC*Na^+W, and MeC*Na^+2W reach the equilibrium concentrations with the domination of MeC* over MeC*Na^+W and MeC*Na^+2W species. It has been found [26] that the equilibrium concentrations for MeC*Na^+W and MeC*Na^+2W forms are absolutely negligible (1.2×10^{-14} and 8.4×10^{-13}, respectively). Moreover, the rates of tautomerization reactions for MeC*Na^+W and MeC*Na^+2W are 21 orders of magnitude slower than those for MeC*W, and three orders of magnitude slower than those for MeC (Table 7-5).

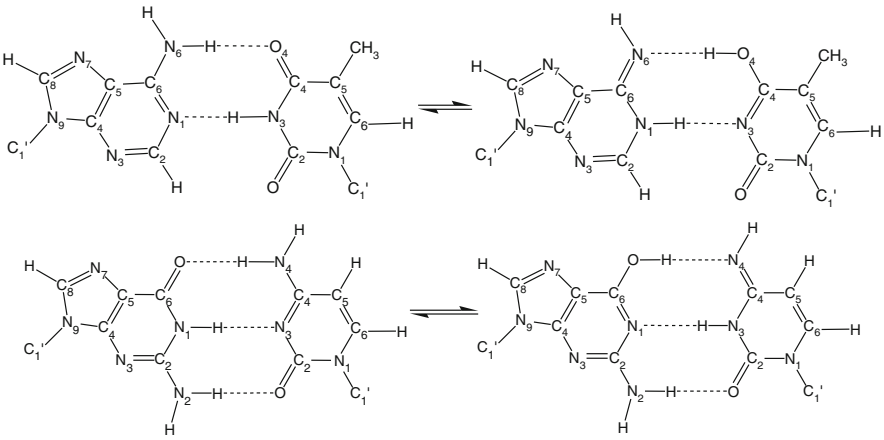
The kinetics analysis of the tautomerization allows one to estimate the approximate rates of the point mutations caused by the tautomerization of the hydrated MeC in the presence of the Na^+ cation. It appears that the interactions of hydrated MeC with Na^+ contribute to decrease of the concentrations of MeC tautomers. The MeCNa^+W and MeCNa^+2W complexes are characterized by equilibrium concentrations of 10^{-13}–10^{-14}, which is much lower than the level corresponding to the point mutations: 10^{-10}–10^{-8} in vivo and 10^{-6}–10^{-5} in vitro [42, 43], and thus

cannot contribute to point mutations in DNA. In contrast to the MeC*Na^+W and MeC*Na^+2W complexes, MeC*W reaches a relative concentration of 2.9×10^{-2} within 10^{-7} s (see Figure 7-11), and thus could contribute to the formation of point mutations. However, the actual rates of mutations are expected to be lower because of the replication fidelity mechanisms in DNA polymerases [61].

We also would like to mention the recent investigations [27, 62] which reported only thermodynamic parameters of the interaction of the DNA sugar–phosphate backbone and hydrated Mg^{2+} It was found [27, 62] that the tautomeric properties of the nucleotides in biologically relevant anti conformations do not depend on the charge of the phosphate group. However, the investigation of the effect of hydrated Mg^{2+} on tautomeric equilibrium in 2′-deoxyguanosine monophosphate (GMP) revealed the suppression of the GMP tautomerization. The results of calculations at the MP2/cc-pVDZ//B3LYP/cc-pVDZ level [27] suggest that the equilibrium constant of tautomerization (*Keq*) for biologically relevant anti conformations of GMP are $(3.7–8.4) \times 10^{-5}$, in comparison to 2.0×10^{-1} for pure guanine. Thus the hydrated magnesium dication shifts tautomeric equilibrium in GMP towards stabilization of the canonic *keto* form of guanine.

7.3.4. Gas Phase Tautomerization in AT and GC Pairs of DNA Bases

All the tautomerization processes considered above result in single intramolecular proton transfer. However, there is another intermolecular way of obtaining the rare forms of DNA bases. It is represented by the following equilibria:



(Reproduced by permission of American Chemical Society: reference [29])

Thermodynamic and kinetic of such tautomerization processes have been recently studied [29]. Due to a complexity of the system the kinetics of these processes has not been investigated in such details as in previous studies devoted

to single nucleobases. Therefore we also will discuss it only briefly, based on the parameters presented in Table 7-6.

The most remarkable result is obtained for the AT → A*T* proton transfer. As follows from the data presented in Table 7-4, the values of $\Delta G^{\neq}$ (activation Gibbs free energy) are smaller than the values of $\Delta G_{C \to R}$ (tautomerization Gibbs free energy) at all considered levels of theory. This causes disappearance of the local minimum corresponding to the A*T* base pair from the surface of the Gibbs free energy. In other words these chemical processes could not be described by equations of traditional chemical kinetics, which always should have local minima on the surface of Gibbs free energy, connected by a transition state. Therefore, only the values of the rate constants that describe the kinetics of the proton transfer in forward and reverse directions of GC nucleic acid pair have been estimated using instanton approach [22]. At the MP2/infinite//MP2/6-31G(d) level the considered rate constants amount to 1.2×10^7 s^{-1} and 1.2×10^{13} s^{-1}, respectively. It means that the equilibrium concentration of the G*C* form of the base pair will be achieved instantaneously, as soon as guanine and cytosine form a canonic structure of the base pair.

7.4. CONCLUSIONS

The ab initio based kinetics simulation technique becomes a powerful tool, which has been used recently for study of the proton transfer processes in DNA bases, nucleotides and their complexes with metal ions and in presence of water molecules. Considered methodology includes there steps: ab initio calculation of the relative Gibbs free energies of local minima and transition states of the reactants and products; estimation of the equilibrium and rate constant of the reactions under study; and finally solution of the rate equations. The kinetic simulation provides the information on relative populations of the reactants and products of chemical reactions and their change in the course of the reactions. This technique has been also applied for the simulation the gas phase experiments and allows one to estimate the effect of the experimental concentrations of the relative populations of the tautomers of the nucleobases. We believe that the proposed approach gains a fine standing among the available methods which facilitates its further applications for accurate kinetic simulation of a wide range of chemical processes.

ACKNOWLEDGEMENTS

Support has been provided by NSF CREST grant HRD-0833178. The authors are grateful to the Mississippi Center for Supercomputer Research for the computational facilities. The authors would thank to Professor Michel Mons, Professor Oleg V. Shishkin, Professor Dmytro M. Hovorun and Dr. Andrea Michalkova for the collaboration and active discussion of the kinetic project.

REFERENCES

1. Hartree DR, Hartree W (1936) Proc R Soc Lond A 157:490–502
2. Hartree DR (1961) Calculation of atomic structure. Wiley, New York
3. Moller C, Plesset MS (1934) Phys Rev 46:0618–0622
4. Kohn W, Becke AD, Parr RG (1996) J Phys Chem 100:12974–12980
5. Sponer J, Leszczynski J, Hobza P (2001) J Mol Struct – THEOCHEM 573:43–53
6. Nowek A, Sims R, Babinec P, Leszczynski J (1998) J Phys Chem A 102:2189–2193
7. Car R, Parrinello M (1985) Phys Rev Lett 55:2471–2474
8. Kuhne TD, Krack M, Mohamed FR, Parrinello M (2007) Phys Rev Lett 98:066401–04
9. Furthmuller J, Hafner J, Kresse G (1994) Phys Rev B 50:15606–15622
10. Kresse G, Furthmuller J (1996) Comp Mater Sci 6:15–50
11. Kresse G, Furthmuller J (1996) Phys Rev B 54:11169–11186
12. Prezhdo OV, Duncan WR, Prezhdo VV (2009) Prog Surf Sci 84:30–68
13. Kilina SV, Kilin DS, Prezhdo OV (2009) ACS Nano 3:93–99
14. Barbatti M, Ruckenbauer M, Szymczak JJ, Aquino AJA, Lischka H (2008) Phys Chem Chem Phys 10:482–494
15. Spiegel K, Rothlisberger U, Carloni P (2004) Phys Chem B 108:2699–2707
16. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery Jr.JA, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill P, MW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA. (2004) Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford CT.
17. Gordon MS, Schmidt MW (2005) In: Dykstra CE, Frenking G, Kim KS, Scuseria GE (eds) Theory and applications of computational chemistry, the first forty years. Amsterdam, Elsevier
18. Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, Jensen JJ, Koseki S, Matsunaga N, Nguyen KA, Su S, Windus TL, Dupuis M, Montgomery JA (1993) J Comp Chem 14:1347–1363
19. McQuarrie DA (2000) Statistical mechanics, 2nd edn. University Science Books, Sausalito, CA
20. McQuarrie DA, Simon JD (1997) Physical chemistry: a molecular approach. University Science Books, Sausalito, CA
21. Henriksen NE, Hansen FY (2008) Theories of molecular reaction dynamics: the microscopic foundation of chemical kinetics, 1st edn. Oxford, Oxford University Press
22. Smedarchina Z, Siebrand W, Fernandez-Ramos A, Gorb L, Leszczynski J (2000) J Chem Phys 112:566–573
23. Press WH, Flannery BP, Teukolsky SA, Vetterling WT (1988) Numerical recipes in C. Cambridge University Press, Cambridge, UK
24. Kosenkov D, Kholod Y, Gorb L, Shishkin O, Hovorun D, Mons M, Leszczynski J (2009) Phys Chem B 113:6140–6150
25. Gorb L, Kaczmarek A, Gorb A, Sadlej AJ, Leszczynski J (2005) J Phys Chem B 109:13770–13776
26. Michalkova A, Kosenkov D, Gorb L, Leszczynski J (2008) Phys Chem B 112:8624–8633

27. Kosenkov D, Gorb L, Shishkin OV, Sponer J, Leszczynski J (2008) J Phys Chem B 112:150–157
28. Podolyan Y, Gorb L, Leszczynski J (2003) Int J Mol Sci 4:410–421
29. Gorb L, Podolyan Y, Dziekonski P, Sokalski WA, Leszczynski J (2004) J Am Chem Soc 126:10119–10129
30. Saenger W (1984) Principles of nucleic acid structure. Springer-Verlag, New York
31. Neidle S (1994) DNA structure and recognition. Oxford University press, Oxford
32. Sato T (1984) Acta Cryst C 40:736–738
33. Rivas L, Sanchez-Cortes S, Garcia-Ramos JV (2002) Phys Chem Chem Phys 4:1943–1948
34. Young DV, Tollin P, Wilson HR (1974) Acta Cryst B 30:2012–2018
35. Reddy BS, Viswamitra MA (1975) Acta Cryst B 31:19–26
36. Sugawara Y, Nakamura A, Iimura Y, Kobayashi K, Urabe H (2002) Phys Chem B 106: 10363–10368
37. Jardetsky O, Roberts GC (1981) NMR in molecular biology. Academic, New York
38. Shishkin OV, Gorb L, Zhikol OA, Leszczynski J (2004) J Biomol Struct Dyn 22:227–243
39. Watson JD, Crick FHC (1953) Nature 171:964–967
40. Mons M, Dimicoli I, Piuzzi F, Tardivel B, Elhanine M (2002) J Phys Chem A 106:5088–5094
41. Piuzzi F, Mons M, Dimicoli I, Tardivel B, Zhao Q (2001) Chem Phys 270:205–214
42. Strazewski P (1988) Nucl Acid Res 16:9377–9398
43. Topal MD, Fresco JR (1976) Nature 263:285–289
44. Szczesniak M, Szczepaniak K, Kwiatkowski JS, KuBulat K, Person WB (1988) J Am Chem Soc 110:8319–8330
45. Sheina GG, Stepanian SG, Radchenko ED, Blagoi YuP (1987) J Mol Struct 158:275–292
46. Choi MY, Dong F, Miller RE (2005) Phil Trans R Soc A 363:393–413
47. Choi MY, Miller RE (2006) J Am Chem Soc 128:7320–7328
48. Nir E, Muller M, Grace LI, de Vries MS (2002) Chem Phys Lett 355:59–64
49. Nir E, Hunig I, Kleinermanns K, de Vries MS (2003) Phys Chem Chem Phys 21:4780–4785
50. Nir E, Hunig I, Kleinermanns K, de Vries MS (2002) Eur Phys J D 20:317–329
51. Mons M, Piuzzi F, Dimicoli I, Gorb L, Leszczynski J (2006) J Phys Chem A 110:10921–10924
52. Zhao Y, Truhlar DG (2004) J Phys Chem A 108:6908–6918
53. Martin RB (1985) Acc Chem Res 18:32–38
54. Pugliesi I, Muller-Dethlefs K (2006) Phys Chem A 110:13045–13057
55. Sigel H (1993) Chem Soc Rev 4:255–267
56. Sigel A, Sigel H (1996) Metal ions in biological systems. Marcel Dekker, New York
57. Sigel A, Sigel H (1996) Probing of nucleic acids by metal ion complexes of small molecules. Marcel Dekker, New York
58. Lippard SJ, Berg JM (1994) Principle of bioinorganic chemistry. University Science Books, Mill Valley, CA
59. Kaim W, Schwedersky B (1994) Bioinorganic chemistry: inorganic elements in the chemistry of life. Wiley, Chichester
60. Loeb LA, Zakour AR (1980) In: Spiro TG (ed) Nucleic acid-metal ion interactions. New York, Wiley
61. Kunkel TA (2004) J Biol Chem 279:16895–16898
62. Gorb L, Shishkin O, Leszczynski J (2005) J Biomol Struct Dyn 22:441–454

CHAPTER 8

SIMULATION OF CHARGE TRANSFER IN DNA

TOMÁŠ KUBAŘ AND MARCUS ELSTNER
Theoretical Chemical Biology, Karlsruhe Institute of Technology, Germany,
e-mail: tomas.kubar@kit.edu; marcus.elstner@kit.edu

Abstract: The phenomenon of charge migration in DNA has attracted considerable interest of experimental as well as computational research in the last decade. It poses a huge challenge for simulation, due to the large system size and the long relevant time scales. Simple modeling frameworks may miss or overapproximate several important factors influencing the charge transfer in DNA, most prominently the dynamical and solvent effects. Therefore, modern approaches make use of multi-scale coarse-grained computational schemes, which have been developed in several labs recently. These techniques combine empirical force fields to capture the DNA dynamics and quantum-chemical methods to describe the actual charge transfer events. The performed simulations show that the dynamics and solvent effects play a major role in DNA charge transfer.

Keywords: DFT, Electronic coupling, Time-dependent Schrödinger equation, Ionization potential, Mechanism of transfer, Molecular dynamics, Multi-scale techniques, QM/MM, Reorganization energy, SCC-DFTB, Solvent effects

8.1. INTRODUCTION

In recent years, DNA has been the subject of intensive research not only due to its primary purpose as a carrier of genetic information, but also because of its ability to support the migration of electric charge over large distance [1, 2]. The long-range charge transfer discovered in double-stranded DNA [3, 4] was also observed in single-molecule conductivity experiments [5–7]. Charge transport was shown to play a fundamental role in the processes of radiatively induced DNA damage as well as in the repair mechanisms [8, 9]. In a more general view, the research of electrochemical properties of DNA dates 50 years back [10], and may be documented by the construction of DNA-based electrochemical biosensors [11, 12]. It is also worth mentioning the suggested application of DNA species in the design of nanoelectronic devices [13, 14].

P. Paneth and A. Dybala-Defratyka (eds.), *Kinetics and Dynamics*,
Challenges and Advances in Computational Chemistry and Physics 12,
DOI 10.1007/978-90-481-3034-4_8,

8.1.1. Basics of Hole Transfer in DNA

The excess charge subject to transfer can be negative (an electron) or positive (missing electron, or a hole). Because of its biological relevance, this work deals with the transfer of hole (i.e. a radical-cation state), and the same is true about many other computational as well as experimental studies.

In vivo, a radical-cation may be introduced by means of a radiation damage of DNA by UV irradiation. Experimental approaches may use various techniques, like the photolysis of an appropriate molecule linked to DNA, or of a modified nucleotide, or the photoexcitation of a linked molecule, leading to charge separation. In spite of the presence of negative charge on the phosphate moiety of the nucleotide, the most easily oxidizable parts of DNA are the nucleobases, as illustrated in a recent study of oxidation of nucleotides in aqueous solution [15]. The hole transfer then proceeds via the nucleobases along the DNA strand.

The relative stability of the radical-cation state on the individual nucleobases, and thus the energy profile of hole transfer in DNA is determined by the nucleobases' ionization potentials (IP). Since the IP of guanine (G) is the smallest and adenine (A) follows with a 0.4 eV larger IP [16, 17], the radical-cation state will be accommodated predominantly by the purine bases, rather than the pyrimidines with an IP more than 1 eV larger than that of G.

If several consecutive Gs are accumulated on one DNA strand, they work as a hole trap with an increased probability of reaction with water, which terminates the transport [18–20]. The hole transfer thus proceeds between individual (groups of) Gs that are separated by one or several AT base pairs. This principle was named G-hopping by Bixon and Jortner [21], and further characterized by Berlin et al. [20]. The hopping along the DNA strand then competes with the reaction with water.

This is sketched in Figure 8-1. Two principally different modes of motion of the hole between (groups of) Gs are possible. These are one-step *superexchange tunelling* across the bridge formed by AT base pair(s) [21], which obviously requires a relatively short bridge for the tunneling probability to reach non-negligible values. On the other hand, *thermally induced hopping* (or A-hopping) proceeds in multiple steps [20, 22], with the hole being transiently present within the bridge. Large amount of research of charge transfer in DNA has concentrated on the distinction between these paradigmatic modes of charge migration.

8.1.2. Experimental Studies

The experimental research of long-range charge transport in DNA dates back to the study performed by Barton et al. in 1993 [23].

8.1.2.1. Rate of Hole Transfer

Following studies by Barton et al. concentrated on the effect of distance and DNA sequence [24], also by using spectroscopy with high temporal resolution [25]. The

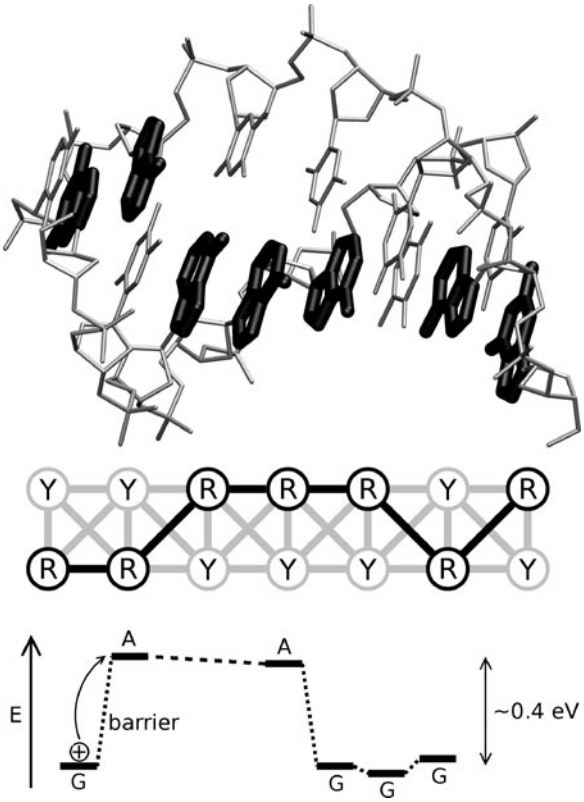


Figure 8-1. *Coarse graining* of the hole transfer in DNA. *Top*: The transfer of hole is assumed to proceed via the purine bases (Gs and As – black), whereas the pyrimidine bases as well as the backbone (grey) do not take part in the transfer. *Center*: The model of DNA for hole transfer. Purine bases (R, black) constitute a chain of sites, over which the hole migrates; pyrimidine bases (Y, grey) stand out of the way of transport and are not considered. *Bottom*: Since the IP of A is 0.4 eV larger than that of G, the hole has to overcome this energy difference if transferring from one G to another. The A-bridge constitutes a barrier for the transfer. However, the magnitude of the energy difference may be affected by the dynamics and environment, as described in Section 8.3

results showed a very shallow distance dependence of CT efficiency (or rate). As for the sequence dependence, it was AT-rich sequences that proved themselves to be particularly disruptive to charge migration. In another work, Barton et al. emphasized the strength of π-stacking between nucleobases to be the determinant of charge transfer efficiency [26].

Giese et al. performed extensive studies of transport of a hole from an initial guanine to a terminating G triplet. They first reported strongly distance-dependent transfer over one to four AT base pairs [27]. A substitution of AT by GC accelerated the transfer largely, which led to the idea of charge transport as a sequence of individual hops between Gs across AT pairs. The efficiency of these single hops

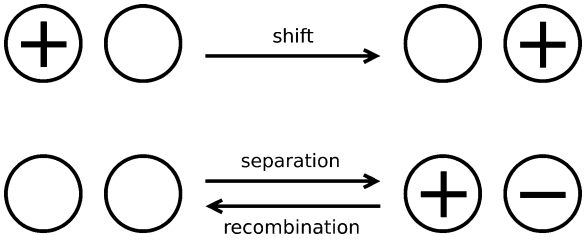


Figure 8.2. The different charge transfer reactions investigated in experiments. Charge shift (e.g. Giese et al.) as well as charge separation and recombination (e.g. Lewis et al.)

would then contribute to the overall distance dependence, which need not be exponential any more. A further investigation involved hole migration across even longer AT bridges and revealed the crossover of distance dependence from exponential to algebraic, at four AT base pairs [28]. This was attributed to a difference in CT mechanism in short and long bridges, namely tunneling and thermally induced hopping.

Charge migration in stilbene-modified DNA hairpins was the subject of study of Lewis et al., using time-resolved spectroscopic measurements [29, 30]. The observed rate of hole arrival at the acceptor molecule, and in particular its dependence on the length of intervening AT-bridge was, however, the same as in other experiments: A strong, exponential dependence observed in short bridges changed to a weak one in long bridges.

It is of interest to note that the charge-transfer reactions investigated in these experiments were different (cf. Figure 8-2). Giese et al. injected a net positive charge onto the DNA double strand and observed its migration, also called *charge shift*. On the other hand, the sequence of CT processes in the experiments by Lewis et al. was initiated by a *charge separation*; here, a hole was injected to a nucleobase while the corresponding electron remained localized on the charge donor molecule. The hole then travelled to the charge acceptor, and the last step of the reaction sequence consisted of a *charge recombination*, which involved the charge donor and acceptor molecules.

8.1.2.2. Mechanism of Charge Migration

Schuster et al. performed experiments with DNA sequences containing isolated GG steps, capped with anthraquinone derivatives, which inject a hole onto the DNA helix upon irradiation [31]. The observed distance dependence of CT efficiency was nearly unaffected by the exact sequence of AT pairs between the GG steps, and suggested the idea of dynamic averaging of structure. Structural distortions seemed to play a role, i.e. the variation of helical structure or hydrogen bonding. This lead to the proposal of *phonon-assisted polaron-like hopping*. The polaron should constitute of a self-trapped structural distortion of DNA, which would be caused by the motion of DNA to migrate.

A model of interaction with the *polaron* was presented by Conwell and Rakhmanova in a computational study [32] based on a Su-Schrieffer-Heeger-type model (SSH) [33]. Here, the result is that the injection of excess charge into DNA would indeed lead to the formation of a polaron, under favorable circumstances. Such a polaron was reported to be rather robust – independent of the parameterization of the theory.

Based on QM calculations of a micro-solvated DNA oligomer, Schuster et al. came up with yet another possible mechanism of charge migration in DNA [34]. The energetics of charge transfer, and in particular the energy barrier of AT bridges, should be largely affected by the thermal motion of solvated counterions. They observed only a small fraction of (ion) configurations to be "charge-transport effective", while most others were not. Then, the time scale of hole migration would be controlled by the rate of appearance of such (ion) configurations. Support for such *ion-gated charge transport* was sought in the experimental as well as computational study of CT in a DNA analog with uncharged backbone, which does not require the presence of counterions.

The early suggestion by Barton et al. that the conformational dynamics associated with the intervening segment of DNA (between the hole donor and acceptor) must play a role [35], formed the basis of the proposed *conformationally gated hopping* [36]. Here, CT in DNA was said to be closely related to the dynamics of the intervening bridge, rather than simply dependent on distance. The hole transfer should take place between groups of four to five base pairs, which form a transient well-stacked domain. Such a domain is constituted by a π-orbital momentarily extended over several nucleobases, over which a hole may be delocalized. Also, they note that there is merely a small subset of DNA conformations active towards CT, of the vast conformational space.

In another work [37], Conwell acknowledged the findings by Giese of the tunneling–hopping crossover, as well as by Barton of the delocalization of hole. She predicted a polaron to be spread over about four nucleobases, and noted the equivalence of polaron transport and conformationally gated transport. The polaron might be transported either by drifting (in homogeneous sequences) or by hopping. Obviously, the polaron transport would be unlikely in vacuo, as there is no polarizable medium around DNA.

8.1.3. Theory and Computation

In principle, a full description of the CT process would be achieved using ab initio quantum-chemical methodology or a DFT framework, as already outlined in several studies [34, 38]. However, given the necessity to take into account the entire DNA molecule and the solvent, the concerned molecular system is way too large to be described by means of such an approach. Thus, appropriate approximations or alternative methods to describe CT have to be sought.

8.1.3.1. Phenomenological Description: Marcus' Theory

The usual means of a classical description of charge transfer between molecules in a medium is the Marcus' theory [39–41], which has the form of a closed expression for the rate of charge transfer k from the charge donor to the acceptor

$$k = \kappa\, \nu\, exp\left[-\frac{(\Delta G_0 + \lambda)^2}{4\lambda k_B T}\right] \tag{8-1}$$

in the adiabatic regime (κ – electronic transmission coefficient, ν – effective frequency for nuclear motion along the reaction coordinate, k_B – Boltzmann constant, T – temperature), or

$$k = \frac{2\pi}{\hbar} |H_{DA}|^2 \frac{1}{\sqrt{4\pi\lambda k_B T}} exp\left[-\frac{(\Delta G_0 + \lambda)^2}{4\lambda k_B T}\right] \tag{8-2}$$

under the assumption of weak electronic coupling ($\hbar$ – Dirac constant).

These equations contain three parameters: ΔG_0 is the free energy change of the charge transfer reaction, which vanishes if the donor and acceptor are chemically equivalent. The donor–acceptor electronic coupling H_{DA} is then one of the decisive factors. It may be thought of as a term involving implicitly the IPs of the charge donor and acceptor, as well as that of the possible bridge separating the donor and acceptor, and further the individual charge-transfer integrals between all these elements of the system. The reorganization energy λ formally corresponds to the necessary change of structure of the entire system before the actual CT can take place and has the meaning of activation energy if the free energy change vanishes. It is often considered to be composed of the internal reorganization energy λ_i, which describes the change of molecular structure of the donor and the acceptor, and the solvent reorganization energy λ_s, which may be attributed to the (change of) orientation polarization of solvent.

Examples of the application of Marcus' theory for CT in DNA are the studies by Beratan et al. [42, 43] and Olofsson and Larsson [44], as well as the "atomistic model of DNA charge transfer" introduced by Cramer et al. [45], who made use of the SSH model [33] and re-parameterized it in order to describe heteroatoms.

In spite of its formal simplicity, the Marcus' expression of rate has proven to take into account the decisive factors affecting the rate of CT. As such, it has been used for considerations about the dependence of CT rate k on the donor-acceptor distance R_{DA}. By various authors, the exponential dependence of CT rate k over short distances

$$k \propto exp[-\beta R_{DA}] \tag{8-3}$$

has been attributed to the exponential distance dependence of the coupling ($|H_{DA}|^2 \propto exp[-\beta_c R_{DA}]$), or to the linear distance dependence of the solvent reorganization energy ($\lambda_s = \beta_s R_{DA}$ and λ_s enters an exponential term), or both ($\beta = \beta_c + \beta_s$).

The Marcus' theory captures the underlying principles of charge transfer in biomolecular systems, including DNA. However, its rather macroscopic character disregards certain details, like the specific molecular character of the bridge connecting the charge donor and acceptor and, not least importantly, the dynamics of the entire system and the features of solvent. Therefore, more elaborate approaches were developed, which take these factors into account.

8.1.3.2. Kinetic Studies

One of the first approaches that considered the distinct structure of DNA as the charge-transfer system were kinetic and quantum – kinetic analyses. Here, the hole is considered to be transferred between the individual nucleobases, before being eventually trapped and annihilated at the end of a DNA strand. Also, it may react with water on every nucleobase, and to each of these elementary processes is ascribed a rate constant.

Such schemes were used to assess the efficiency of hole transfer once the necessary rate constants were determined, mostly on the basis of available experimental data. Alternatively, it was possible to estimate the rate constants of the elementary processes by comparison with experimentally obtained rates. Studies of this type led to the introduction of G-hopping [21] and thermally induced hopping [20, 22], as well as to some more detailed mechanistic concepts [46].

8.1.3.3. Stochastic Approaches

Several recent studies used the classical theory in an advanced way, estimating the rate and probability of the elementary CT steps with Marcus' equation.

Jakobsson and Stafström performed Monte Carlo simulation of CT in DNA [47]. The individual CT steps were described with Marcus' theory, which was extended to involve special hole states de-localized over the entire adenine bridge. The necessary parameters were chosen from previous experiments and calculations. The authors pointed out that this dependence is brought about by both the donor– acceptor coupling H_{DA} and the reorganization energy λ in the Marcus' picture. Also, they reported a quite continuous crossover from superexchange to hopping.

The Monte Carlo concept was also used by Řeha et al. [48], who developed yet another charge-hopping model involving the Marcus' equation for the rate of individual hops. A multi-scale scheme including MD simulation, semiempirical quantum chemistry and a Monte Carlo engine was used. The hole was considered localized on a single nucleobase at a time. They concluded that CT in DNA was driven by local changes of relative configuration of stacked bases with high frequency, supporting the idea of conformationally gated hole transfer of Barton et al., see above.

8.1.4. Subject of This Contribution

The following sections deal with several key features of hole transfer in DNA. First are the effect of solvent on the CT parameters, together with the effect of dynamics of the molecular system. A quantum dynamics study of hole transfer follows, which assesses critically the influence of the individual CT parameters and their features. Further presented is a study of solvent reorganization energy, which constitutes the other factor affecting the charge transfer. Also introduced is a current development – hybrid quantum–classical framework accounting for the polarization of solvent.

8.2. CHARGE-TRANSFER PARAMETERS

As described in Section 8.1.1, the economical way to deal with the complex problem of hole transfer in DNA is a sort of *coarse graining*. Every purine nucleobase is considered an elementary entity or *site*, which may accommodate the excess charge (hole – a radical-cation), and these purine bases form a pathway, through which the hole migrates (see Figure 8-1 again). Such a coarse-grained model then involves two kinds of parameters: site energies and electronic couplings.

8.2.1. Ionization Potentials

The energetics of hole transfer in DNA is determined by the ionization potentials (IP) of nucleobases. The values of these IPs *in vacuo* were determined experimentally long ago [16], which was followed by a study on solvated nucleosides [17]. At the same time, it is relatively straightforward to obtain these *site energies* computationally. Applicable methodologies include a wide range of quantum-chemical approaches – ab initio, DFT and semiempirical. IP may be calculated using the ΔSCF procedure, as the difference of energy of the neutral molecule and its radical-cation. Alternatively, Koopmans' theorem (KTA) can be applied, which approximates the IP by the energy of the highest occupied molecular orbital (HOMO), making the calculation of the radical-cation unnecessary.

In one of the first studies, Voityuk et al. presented IPs of nucleobase triplets in representative configurations using a semiempirical method and KTA, for all combinations of nucleobases [49], They found that the IP of a nucleobase depends on the neighboring nucleobases. Computational studies then concentrated on the effect of varying structure of the particular nucleobase as well as its surroundings in DNA, which will be the topic of Section 8.3.2.

8.2.2. Electronic Couplings

Electronic couplings are the other set of parameters that take part in the Marcus' term H_{DA} as well as in the framework of other methodologies. Unlike IPs, the

calculation of electronic couplings had been initially a tough challenge, but all the applied methodologies converged eventually to consensual results. A number of approaches may be used to calculate the electronic coupling of a pair of nucleobases, as reviewed in [50]. Very efficient is the method based on the calculation of diabatic states: for hole transfer, the donor and acceptor states correspond to the HOMO of one nucleobase and the other. Then, their electronic coupling equals the Hamilton matrix element between the respective HOMOs of both molecules.[1] This is the essence of the fragment-orbital (FO) approach [51–53], which will be described in detail in Section 8.2.3.1.

A first study by Voityuk [54], using the Hartree–Fock method, also described the basic features of the couplings: these were the exponential distance dependence, structural dependence, general asymmetry in the 5'→3' and 3'→5' direction, as well as the comparison of intra- and interstrand couplings. Several in-depth studies of dynamics effects on the electronic couplings will be mentioned in Section 8.3.1.

8.2.3. CT Parameters Within the Fragment-Orbital Approach

In the last years, we have implemented a multi-scale scheme to calculate the site energies and electronic coupling [55]. The coarse graining of the quantum-chemical problem arises from the application of a fragment-orbital (FO) approach [51–53]. Favorable efficiency is achieved by the use of the approximative DFT method SCC-DFTB. The influence of solvent water, counterions, DNA backbone as well as pyrimidine bases is taken into account by means of a quantum mechanics–molecular mechanics (QM/MM) implementation within SCC-DFTB. The dynamics of the entire molecular system is described by classical MD simulations. The CT parameters are then evaluated using the FO-DFTB methodology along the classical MD trajectories. This methodology allows the study of CT processes in DNA and related molecules in a realistic fashion.

8.2.3.1. Fragment-Orbital Approach

In the first step, the electronic structure of every individual nucleobase i is solved (with a Hamiltonian $\hat{H}_i$) and the expansion coefficients of the HOMO ϕ_i are stored. Also, the i-th site energy is approximated as the HOMO energy:

$$\varepsilon_i = \langle \phi_i | \hat{H}_i | \phi_i \rangle \tag{8-4}$$

The obtained coefficients of the HOMOs are then used in the second step. The Hamiltonian $\hat{H}$ for the entire molecular system composed of all nucleobases is set

[1] More precisely, this holds only if the diabatic states, i.e. HOMOs are orthogonal. If this is not the case, a correction should be introduced, as mentioned in Section 8.2.3.1.

up and the electronic couplings T_{ij} are evaluated as the Hamilton matrix elements between the HOMOs:

$$T_{ij} = \left\langle \phi_i | \hat{H} | \phi_j \right\rangle \tag{8-5}$$

In this way, the SCF procedure for the entire system is omitted, which leads to a significant decrease of computational complexity.

The HOMOs are expressed by an expansion into the atomic-orbital basis set η_μ with expansion coefficients c_μ^i:

$$\phi_i = \sum_\mu c_\mu^i \eta_\mu \tag{8-6}$$

Then, the coupling and overlap integrals may be evaluated using the Hamilton and overlap matrices calculated in the atomic-orbital basis set (this is what we mean by "set-up of the Hamiltonian for the entire molecular system"):

$$\begin{aligned} T_{ij} &= \sum_{\mu\nu} c_\mu^i c_\nu^j \langle \eta_\mu | \hat{H} | \eta_\nu \rangle = \sum_{\mu\nu} c_\mu^i c_\nu^j H_{\mu\nu} \\ S_{ij} &= \sum_{\mu\nu} c_\mu^i c_\nu^j \langle \eta_\mu | \eta_\nu \rangle = \sum_{\mu\nu} c_\mu^i c_\nu^j S_{\mu\nu} \end{aligned} \tag{8-7}$$

The matrix of couplings T_{ij} is here obtained for diabatic states (HOMOs), which are generally non-orthogonal. However, a representation in an orthogonal basis set is often preferred. In such a case, we may convert the couplings to those corresponding to a representation in a set of orthogonal states, using a procedure proposed by Loewdin [56].

$$T' = S^{-\frac{1}{2}} \cdot T \cdot S^{-\frac{1}{2}} \tag{8-8}$$

where T' is the matrix of "orthogonalized" couplings.

8.2.3.2. The SCC-DFTB Method

We perform the actual quantum-chemical calculations with the self-consistent-charges density-functional tight-binding method (SCC-DFTB) [57]. It consists of a second-order expansion of total DFT energy around a reference electron density, where the fluctuations of density are approximated by charge monopoles, and an effective damped Coulomb interaction operates between the net atomic charges. The total SCC-DFTB energy then reads

$$E^{SCC-DFTB} = \sum_i \sum_{\mu\nu} c_\mu^i c_\nu^i H_{\mu\nu}^0 + \frac{1}{2} \sum_{\alpha\beta} \gamma_{\alpha\beta} \Delta q_\alpha \Delta q_\beta + \frac{1}{2} \sum_{\alpha\beta} E_{\alpha\beta}^{rep} \tag{8-9}$$

where the index i refers to molecular orbitals, $H_{\mu\nu}^0$ is the Hamilton matrix in the minimal basis set of atomic orbitals, c_μ^i are the expansion coefficients, $\gamma_{\alpha\beta}$ describes the effective damped Coulomb interaction between the atomic charges Δq_α and Δq_β, and $E_{\alpha\beta}^{rep}$ is a parameterized repulsive potential between the atoms α and β.

The elements of the Hamilton matrix as well as the overlap matrix (which is also necessary) are pre-calculated and stored in tables for all combinations of orbitals in the minimal atomic basis set, for the chemical elements of interest and on a sufficient interval of distances. The values needed for every particular molecular geometry (distance and orientation of atoms in space) can then be quickly obtained by a table look-up and a simple interpolation. No further integral evaluation is thus required during a SCC-DFTB run, which greatly enhances the computational efficiency compared to standard DFT approaches.

Application of the variational principle leads to an eigenvalue problem with the following Hamiltonian

$$H_{\mu\in\alpha,\nu\in\beta} = H^0_{\mu\nu} + \frac{1}{2}S_{\mu\nu}\cdot\sum_{\delta}\Delta q_\delta\left(\gamma_{\alpha\delta}+\gamma_{\beta\delta}\right) \tag{8-10}$$

which is self-consistent due to the presence of (Mulliken) atomic charges Δq_δ. Therefore, the problem must be solved iteratively – the coefficients c^i_μ obtained from the diagonalization of the Hamiltonian are used to calculate the Mulliken atomic charges; these are then cast into the Hamiltonian, which is used in the next iteration. This scheme proceeds in a loop until self-consistence is reached.

8.2.3.3. Coupling to the Environment: QM/MM SCC-DFTB

As noted in the beginning of this contribution, the goal that we are aiming at is the description of the entire molecular system, including the molecular environment of the nucleobases. A particularly efficient way to do so is to combine the quantum-chemical description of the nucleobases with a simplified, molecular-mechanics treatment of the environment – QM/MM [58, 59].

QM/MM was implemented in SCC-DFTB [60] and has been widely applied since then [61]. Within this framework, every atom A of the environment is represented by its MM atomic charge Q_A, and the modified Hamilton matrix element then reads

$$H_{\mu\nu} = H^0_{\mu\nu} + \frac{1}{2}S_{\mu\nu}\left(\sum_{\delta}\Delta q_\delta\left(\gamma_{\alpha\delta}+\gamma_{\beta\delta}\right) + \sum_{A} Q_A\left(\frac{1}{r_{A\alpha}}+\frac{1}{r_{A\beta}}\right)\right) \tag{8-11}$$

with $r_{A\alpha}$ being the distance of the MM atom A from the QM atom α. Using this simple formalism, the electric field induced by the DNA backbone, pyrimidine bases, solvent water as well as the counterions is taken into account, and these parts of the molecular system will now polarize the electron density on the purine bases.

The whole procedure to obtain the CT parameters, using the FO approach and the SCC-DFTB method then runs as follows: First, a full SCC-DFTB calculation is performed for every purine nucleobase considered, using a QM/MM implementation, where all other parts of the system are represented by MM atomic charges (including all other purine bases as well). The IPs, which have the meaning of site energies, are obtained as the energies of HOMO orbitals of the nucleobases. Then,

the SCC-DFTB Hamilton and overlap matrices[2] are set up for a QM system consisting of the purine bases together, with all other atoms represented by MM point charges. The elements of coarse-grained Hamiltonian T_{ij} and overlap matrix S_{ij} are evaluated as in Eq. (8-7). The conversion to a representation in an orthogonal basis set is then performed according to Eq. (8-8).

Very good performance of this procedure was established by thorough testing and comparison with the results of ab initio as well as DFT calculations. Importantly, this approach is so efficient that the CT parameters can be evaluated along multi-nanosecond MD trajectories, using a mainstream PC [55].

8.2.4. Summary

Many quantum-chemical methods have been used to evaluate the parameters for CT in DNA, and very good agreement among all of them has been reported. The IPs of nucleobases are obtained in accordance with experiment (G < A < C ≈ T, difference of G and A of about 0.4 eV). The electronic couplings reach values of up to 0.1 eV, with all methods reproducing the dependence on the orientation of nucleobases as well as the difference between various base pair-steps (e.g. GG versus GA).

Our FO-based approach using the QM/MM implementation in SCC-DFTB provides results in agreement with other methodologies as well, and its remarkable efficiency makes it useful even in computationally intensive applications, as will be presented in further sections.

8.3. EFFECT OF DYNAMICS AND ENVIRONMENT ON CT PARAMETERS

The process of hole transfer in DNA occurs neither in a stationary molecule nor (mostly) in vacuo. Therefore, it is essential to take into account the dynamics of DNA as well as the environment, mostly solvent. Within the coarse-grained model of hole transfer, it is the IPs and the electronic couplings that determine the transfer. This section thus deals with the way these CT parameters are influenced by the phenomena in question.

Several authors have previously studied the features of CT parameters, evaluated for ensembles of DNA conformations, most often generated by MD simulations. They focused on various aspects of the influence of both the DNA conformation and the solvent motion. A common finding was that of IPs and the electronic couplings fluctuating with a considerable magnitude, see further on.

[2]In this second phase of an FO calculation, a different DFTB basis set is used. The standard DFTB basis set uses markedly compressed electron density, which leads to underestimated electronic couplings. Therefore, a re-optimized basis set featuring nearly uncompressed density is used in the second phase of FO, in order to provide couplings of good quality. See [55] for details.

8.3.1. Electronic Couplings

Troisi et al. concentrated on the factors affecting the electronic couplings. They evaluated those for nucleobases in a number of DNA crystal structures [51], as well as along classical MD simulation [62]. Observed were couplings in a broad interval, which was explained by their marked dependence on the DNA structure, in particular on the mutual orientation of the nucleobases. Similarly, Voityuk et al. pointed out the extraordinary sensitivity of couplings with respect to the DNA conformation [63]. The *dynamic nature of electronic coupling* was highlighted by Troisi et al. [64] for electron transfer in an example of organic molecules. The coupling was found to fluctuate vigorously, so that just a minority of configurations were likely to contribute to CT.

Hatcher et al. concentrated on the donor–acceptor coupling H_{DA} (in Marcus' sense) and its distribution, in connection with the mechanism of charge transport [65]. They performed nanosecond MD simulations of DNA as well as of a nucleic-acid analog with peptide-like backbone, which exhibits larger structural flexibility. The crucial observation was that only a few MD snapshots exhibited large couplings whereas all others were nearly negligible. There were MD snapshots with vanishing energy barrier, which led to large H_{DA} – "CT-active" conformations on a "CT-silent" background. A conclusion was drawn that the transport was gated by *local conformational fluctuations*, which would mean intramolecular motion.

In another study on general charge transfer, Balabin et al. evaluated electronic couplings in various donor–acceptor systems in water and protein [66]. They reported that beyond a critical distance (of several ångströms), the charge transport was determined by so-called *non-equilibrium conformations* of the system, which are brought about by its dynamics. The different configuration of the medium between donor and acceptor played a major role on shorter distances, and its influence pertained on long distances to some extent.

8.3.2. Ionization Potentials

Concerning the IPs, two major sources for the fluctuations have been described. On the one hand, the intramolecular vibrations of nucleobases lead to fluctuations of IPs (site energies) of about 0.1 eV, as described by various authors [48, 65, 67]. However, in these studies the interaction with the solvent was neglected.

Voityuk et al. [68] included the solvent in the computational scheme using a QM/MM implementation, and calculated the IP of nucleobases for snapshots along classical MD simulation. Observed were much more pronounced fluctuations; the standard deviation of IPs was found to be 0.4 eV, which reaches the IP difference between guanine and adenine. Thus, it may well happen that these transient fluctuations of IP of neighboring G and A cancel the in-vacuo difference of 0.4 eV, for a brief period of time. A possible consequence was suggested in the title of this paper: "Environmental fluctuations facilitate electron-hole transfer from G to A". This

observation was confirmed in later studies by Steinbrecher et al. [69] and by ourselves [55, 70]. It should be compared to the aforementioned results of calculations performed with the neglect of environment (no QM/MM) – then, the standard deviation was 0.1 eV, and the probability of the energy barrier to vanish was much smaller.

Therefore, a proper description of solvent degrees of freedom is of utmost importance, and the following sections present our research on this topic.

8.3.3. Computation of CT Parameters Along MD Trajectories

With all the necessary components of the multi-scale approach available, namely the MD simulation to catch the dynamics of DNA in water solution and the QM/MM implementation to include the influence of molecular environment, one can aim at a realistic description of the charge transfer in DNA.

8.3.3.1. Magnitude of Oscillations of CT Parameters

Several researchers have already studied the effect of solvent environment on the energetics of hole transfer, expressed in terms of IPs/site energies [68, 69]. Having applied the formalism described above [70], we observed large oscillations of these site energies in agreement with previous studies: Instead of a value of around 0.1 eV if the environment is not taken into account ('in vacuo'), the standard deviation of IPs reaches 0.4 eV with a QM/MM treatment of the environment, see Figure 8-3.

Interesting is the comparison of IP of guanine and adenine. The difference of 0.4 eV, which was observed in experiments, remains the same in the solvent (if we

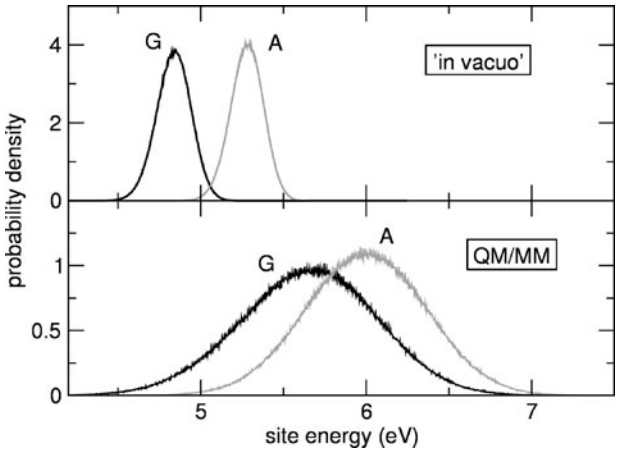


Figure 8-3. The distribution of IP of guanine and adenine in DNA calculated with the neglect of environment (*top*) as well as taking the environment into account with QM/MM (*bottom*). The difference of mean values is the same, but the much larger standard deviations with QM/MM cause a significant overlap of the distributions

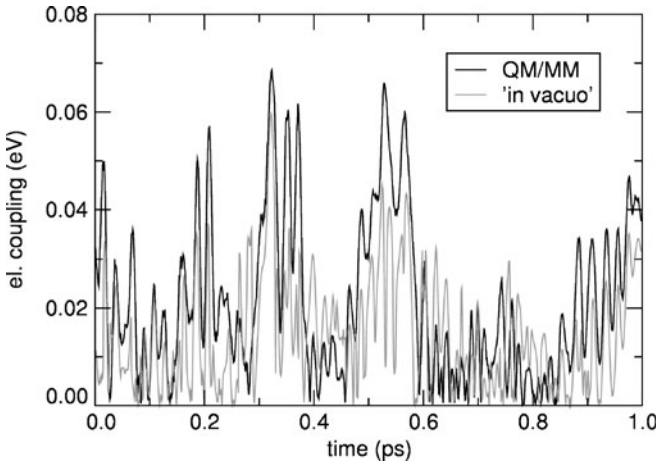


Figure 8-4. The electronic coupling T_{ij} evaluated 'in vacuo' as well as with QM/MM. The time course exhibits partially similar features but the magnitude of individual peaks differs. This does not affect the form of the distribution of the couplings whatsoever

consider the difference of mean values). However, the distributions of IP becomes much broader, due to the motion of the environment and fluctuation of the induced electric field. In turn, this broadening makes the distributions overlap to a large extent, which greatly enhances the possibility for the IP of adenine to reach the (inherently smaller, by 0.4 eV) IP of guanine. In this way, the environment modulates the energy barrier for the hole transfer from a G to an A.

Unlike the case of IPs, the effect of molecular environment on the electronic couplings T_{ij} is much less distinct. Whereas the magnitude of couplings changes when the environment is switched on in the calculation, it does not seem to be possible to formulate a similarly clear conclusion as for what the effect is like, see Figure 8-4. An important point may be that even though the particular time course of the coupling is different in both cases i.e. if the environment is taken into account by means of QM/MM or not, the respective distributions of the coupling are nearly identical (presented in the original work [70]).

8.3.3.2. The Electrostatic Potential Induced by the Environment

The effect of the electric field at a certain nucleobase i induced by the environment may be quantified by evaluating the electrostatic potential (ESP) at that nucleobase summed over the MM point charges Q_A:

$$ESP_i = \frac{1}{4\pi\varepsilon_0} \sum_A \frac{Q_A}{r_{iA}} \tag{8-12}$$

If the calculated IP is plotted versus the ESP obtained in this way, a definite correlation is revealed, see Figure 8-5. Correlation coefficient reaching values

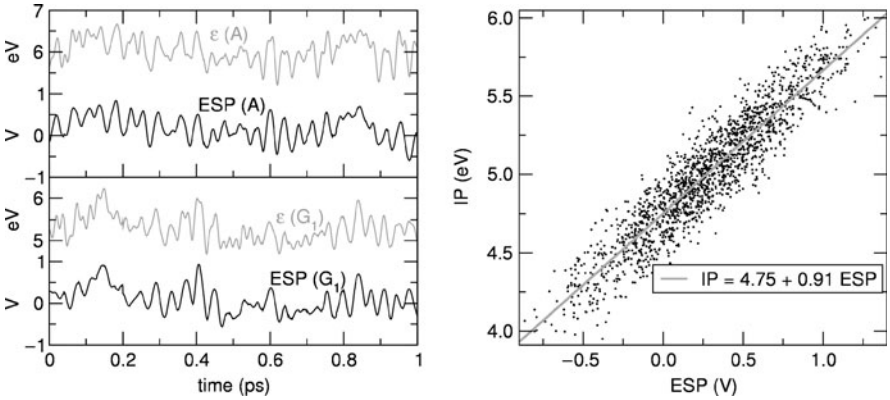


Figure 8-5. IP of nucleobases and the electrostatic potential (ESP) induced by the environment (DNA backbone, all other nucleobases, water and counterions). *Left*: Time course of IP (grey) of two bases and ESP (black) at these molecules. Note the strikingly similar features of both curves, for both G_1 and A. *Right*: The correlation coefficient of 0.92 proves the unambiguous dependence of IP on ESP, with the bias being brought about by the intramolecular vibrations of the nucleobase

larger than 0.9 documents the decisive influence of MM environment on the IP of nucleobases, with the internal dynamics of the nucleobase playing only a minor role, which acts as a bias in the correlation of IP with ESP.

A more detailed analysis of electrostatic potentials may reveal the vibrational modes of the system contributing to the fluctuation of IPs and the characteristic time scales of CT. Because of the additive character of ESP, it is possible to decompose this quantity to the contributions due to the various parts of the system, for instance water, counterions and the DNA backbone. Here, it turned out that the fluctuations of ESP coming from water largely compensate for those of the ESP of counterions, and this expresses the shielding of the counterions' charge by the solvent water. Put together with the large negative ESP induced by the close negatively charged phosphate groups in the DNA backbone, a total value in the range of tenths to units of volt is obtained. This magnitude is truly significant if one considers the effect of such ESP on the IP of nucleobases, yet still quite small if compared to the absolute magnitude of its components.

Another interesting observation is that of the frequency of oscillations of ESP due to the water and counterions. In spite of the considerable mutual compensation described above, the electric field induced by the water molecules fluctuates quite quickly, with a period of 40 fs ($\approx$ 800 cm^{-1}). This reflects most probably the vibration of hydrogen bonds between the water molecules and the respective nucleobase, and oscillations with this frequency have been observed in the IP of nucleobases as well. Contrary to that, the counterions move much more slowly, leading to much slower oscillation of ESP due to the counterions. Thus, the compensation of the contributions to ESP coming from the water and the counterions can never be complete, and non-zero effective fluctuations of ESP always

remain. A similarly imperfect compensation will be observed in case of this combined contribution from the solvent and that from the DNA backbone; resulting is the mentioned total value of tenths to units of volt.

8.3.3.3. Correlation of CT Parameters

It was shown that the molecular environment of DNA is the decisive factor that drives the IP of nucleobases. As the environment of neighboring nucleobases cannot be considered independent,[3] it seems to make sense to look if the IP of such nucleobases are correlated in some way or are free to fluctuate independently. The correlation coefficients were evaluated for IPs of guanines in a poly(G) DNA double strand, calculated along a MD simulation, see Figure 8-6.

Evident is a strong correlation of IP of neighboring bases, with a correlation coefficient of 0.7, and there is even a (weaker) correlation between second neighbors. This observation supports the idea of short runs of three to five purines along the strand, which exhibit correlated site energies. The existence of such regions constitutes an important element of several proposed mechanisms of long-range charge migration in DNA.

Unlike that, there was no such correlation observed for electronic couplings of neighboring base-pair steps. This is probably no surprise as the effect of environment (the source of correlation of IPs) on the couplings could not be characterized clearly (see Section 8.3.3.1).

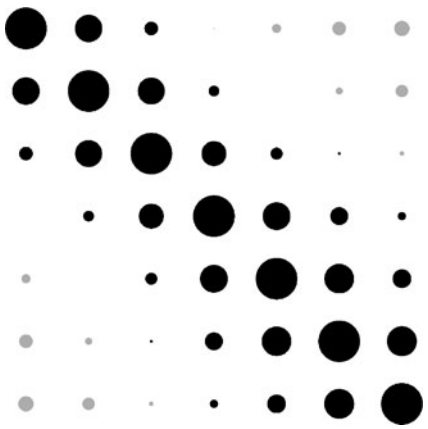


Figure 8-6. The correlation of IPs (site energies) of individual guanines in a poly(G) DNA strand. Black circles – correlation, grey – anti-correlation. The size of every circle codes the magnitude of the respective correlation coefficient. (The large black circles on the main diagonal correspond to correlation of a site with itself, which is unity.)

[3]Note the distance of stacked nucleobases in the A- or B-DNA structure of as little as 3.5 Å.

8.3.4. Summary

All methods used to simulate CT so far agree fairly well in the numerical values of the computed CT parameters. The difference of IP of G and A is estimated to be about 0.4 eV, while the couplings are computed to assume values of up to 0.1 eV. The estimates of the magnitude of oscillations are highly convergent, too. The IPs fluctuate by about 0.1 eV due to the intramolecular vibrations of nucleobases, however, if the interaction with the solvent is included, large fluctuations of 0.4 eV are found by all methods. Therefore, any method aiming at a realistic description of CT should include the interaction with solvent using a QM/MM scheme.

The IPs were shown to be driven by electrostatic interaction with the environment, with the ESP being a good measure for this effect. This allows to study the molecular character of this phenomenon, for instance to investigate the role of individual counterions. Further, a significant correlation of IPs along the strand was observed, which is likely to affect the transfer of hole along the DNA strand.

8.4. QUANTUM DYNAMICS OF A HOLE IN DNA

Having the CT parameters at disposal, it is possible to study the actual development of the excess charge in time – to follow its migration along the DNA strand. A straightforward way to do so is to cast the pre-evaluated CT parameters into the form of a (time-dependent) Hamiltonian, and integrate the time-dependent Schrödinger equation (TDSE) for the hole with this Hamiltonian.

8.4.1. Integration of the Time-Dependent Schrödinger Equation

Grozema et al. [53, 67, 71] and others exploited this idea of a simple coarse-grained way to solve the dynamics of a hole. Here, the wave function of the hole is represented in the basis of HOMOs of the participating nucleobases as $\Psi(t) = \sum_i a_i(t) \cdot \phi_i$, where ϕ_i are the HOMOs and a_i are the corresponding expansion coefficients. Then, the diagonal elements H_{ii} of the coarse-grained Hamiltonian are represented by the IPs of nucleobases, while the electronic couplings T_{ij} represent the off-diagonal terms (H_{ij}). The TDSE may then be solved in the form

$$\frac{\partial \vec{\mathbf{a}}(t)}{\partial t} = -\frac{i}{\hbar} S^{-1} \cdot H \cdot \vec{\mathbf{a}}(t) \tag{8-13}$$

to compute the time evolution of the hole (S – overlap matrix, H – coarse-grained Hamilton matrix, $\mathbf{a}$ – vector of time-dependent expansion coefficients a_i). It should be noted again that the Hamiltonian shall be time-dependent, if the dynamic character of the system is to be considered. To model the annihilation of hole by means of reaction with water, an additional negative-imaginary potential is added to the diagonal elements corresponding to the terminal guanine(s), mimicking a charge sink: $\varepsilon_i' = \varepsilon_i - i\tau$, with an appropriately chosen rate constant τ.

Using this approach, Grozema et al. [67] studied CT in the DNA hairpins as investigated in the experiments by Lewis et al. The CT parameters were computed along an MD trajectory, and the distributions of CT parameters were then used to generate the time-dependent coarse-grained Hamiltonian, resulting in a stochastic model. This indirect procedure was forced by the need to evaluate the CT parameters for an extensive set of MD snapshots. Although the solvent was not included in the calculation of parameters, the significance of the fluctuation of site energies was pronounced.

Such a methodology may be easily improved once a more efficient method to evaluate the CT parameters is available. The favorable performance of the computational framework based on the fragment-orbital approach and the SCC-DFTB method makes it possible to calculate the CT parameters along multi-nanosecond MD trajectories. In this way, the time-dependent Hamiltonian is given by time series of CT parameters evaluated along classical MD trajectories. Importantly, the sets of CT parameters yielded by this computational scheme will involve all the features described in Section 8.3, allowing to abandon stochastic models of CT parameters, and preserving the possible correlations of parameters.

8.4.2. Simulation of Hole Transfer Over Adenine Bridges

In our case study [72], the hole was propagated across adenine bridges in DNA sequences GT_nGGG with $n = 1 \ldots 14$, used in the experiments by Giese et al. (e.g. [28]). The hole was placed onto the first G in time $t = 0$, and negative-imaginary potentials were installed on the guanines of the terminal triplet GGG. Temporal evolution of the hole was then studied by solving the TDSE in the form

$$\frac{\partial \vec{\mathbf{a}}(t)}{\partial t} = -\frac{i}{\hbar} T'(t) \cdot \vec{\mathbf{a}}(t) \tag{8-14}$$

using the fourth-order Runge–Kutta method. T' is the time-dependent coarse-grained Hamilton matrix evaluated along a MD trajectory. It is already expressed in an orthogonal basis set due to Eq. (8-8), thus the overlap matrix S is a unit matrix and can be left out. The part of the hole localized on a particular nucleobase i, or the "occupation" of this site, is obtained as $q_i = |a_i|^2$. The sum of these occupations then gives the total magnitude of the excess charge present in the system, or "survival" of the hole P:

$$P(t) = |\Psi(t)|^2 = \sum_i |a_i(t)|^2 = \sum_i q_i(t) \tag{8-15}$$

Note that the hole starts to disappear from the system once it reaches the terminal GGG due to the action of the negative-imaginary potential, which models the final reaction with water following the trapping of hole on that triplet.

8.4.2.1. Effect of Dynamics

First, we discuss simulations of hole transfer in the DNA oligomer of sequence GTGGG. The results depend on the chosen initial conditions (positions and velocities of all atoms), therefore it is necessary to perform several simulations to assess the variation of results with respect to these initial conditions.

The survival of hole in 20 independent simulations is presented in Figure 8-7 (left). Large variance of the time course of hole survival is observed, yet the survival drops below 0.1 within 20 ps in 19 out of 20 simulations. The strikingly different trajectories show already the dramatic effect of dynamics and solvent on the hole transfer.

A simulation for a "static model" of DNA was performed as well, see Figure 8-7 (right). Here, the time-dependent CT parameters were substituted by their mean values.[4] We can see an exponential falloff of the survival on a much longer time scale compared to the full dynamical simulations. Only now is it possible to appreciate fully the significance of the dynamic approach: It is the variation of CT parameters, brought about by the structural fluctuations of the molecular structure of DNA and environment, and captured by the appropriate model, that has a decisive importance for the control of charge migration.

Note in particular the difference between the individual simulations (Figure 8-7 left): in some of them, not much charge is transferred during several ps, while in others the charge comes through within five ps, and the hole is rapidly disappearing by means of annihilation at the terminal GGG triplet. This seems to be caused by some combinations of CT parameters being more favorable with respect to hole

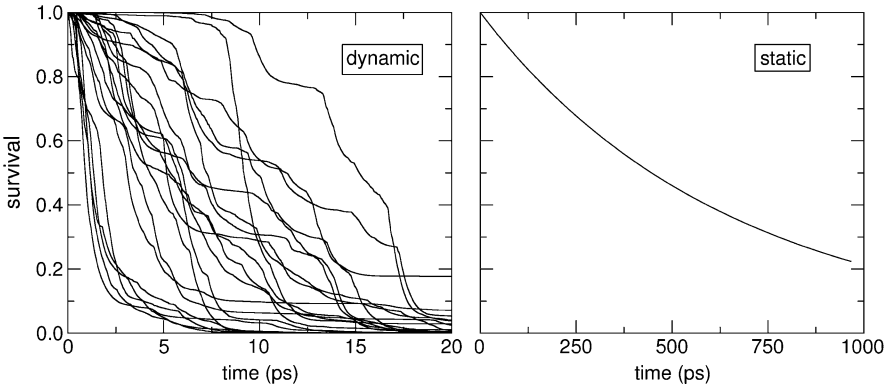


Figure 8-7. Hole transfer in the DNA oligomer GTGGG. Survival of hole in 20 simulations with time-dependent Hamiltonian (*left*) as well as in a static simulation (*right*, note the 50× longer time scale)

[4]Alternatively, it would be possible to evaluate the CT parameters for a static DNA structure – for instance idealized B-DNA, an X-ray structure or an averaged structure from MD simulation.

transfer than others. Our analysis (in [72]) revealed that the energy barrier, i.e. the difference of site energies of G and A must nearly vanish in order to allow for fast CT. In addition, the respective electronic coupling must assume a non-zero value. Only if both conditions are met simultaneously, the hole is transferred rapidly. In principle, this picture of hole transfer corresponds to the conformationally gated model by Barton et al. [36] in a more generalized sense – the term *conformation* would involve not only the DNA, but also the solvent.

8.4.2.2. On the Mechanism and Rate of the Transfer

Obviously, such considerations apply only to a couple of neighboring nucleobases in the DNA double strand, and this would implicitly anticipate a hopping mechanism of CT rather than tunneling. This question, which impacts the mechanism of CT crucially, can be answered by looking at the occupation of the bridge sites (i.e. adenines) by the hole. This should nearly vanish if the hole tunneled through the bridge, contrary to the case of hopping, which would involve non-negligible hole population on the bridge.

The hole occupation of the bridge sites is presented in Figure 8-8, for selected bridge lengths. Already in the shortest bridge composed of a single A, we observe a significantly non-zero occupation. The bridge occupation grows for longer bridges due to the longer traveling times of the hole. This finding indicates clearly a hopping mechanism, as opposed to superexchange tunneling, even for the shortest A-bridges. A similar conclusion has already been drawn earlier. Since simulations using static structures and Marcus' equation (see above) cannot account for the experimentally measured rates (with the theory being off by up to five orders of magnitude), it was concluded that the single-step superexchange is not operative in DNA charge transfer (Tong et al. [73]), and rather a hopping mechanism should be considered.

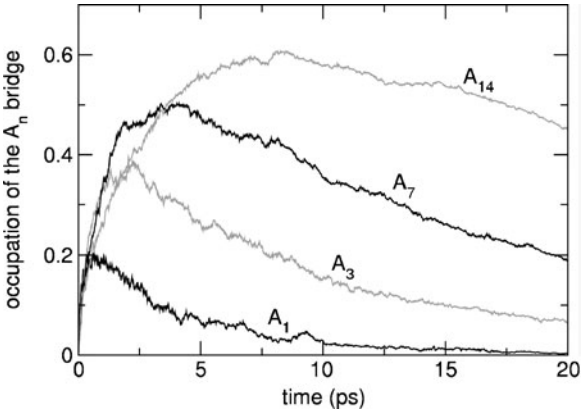


Figure 8-8. Occupation of the adenine bridge by the hole, in simulations of hole transfer in DNA oligomers GT$_n$GGG. Presented are data for four DNA species differing in the length of the bridge. Every curve is averaged over 100 independent simulations of 20 ps each

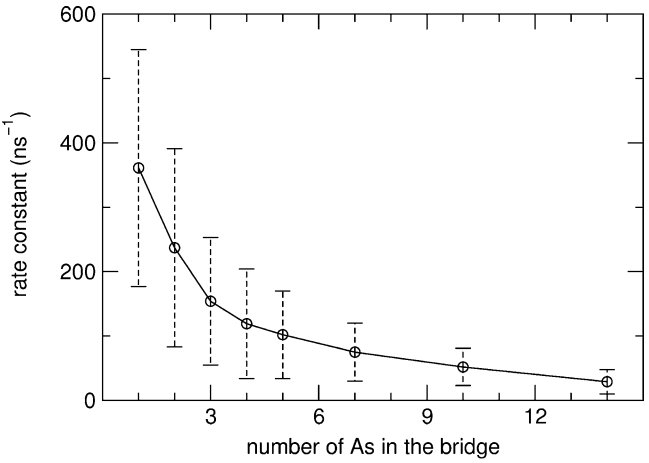


Figure 8-9. Rate of hole transfer from the initial G to the terminal GGG triplet, for the considered DNA oligomers GT_nGGG. Presented are mean values from 100 simulations, error bars denote standard deviations

The rate of hole transfer k is then calculated from the hole survival $P(t)$ at a defined time (e.g. at the end of the simulation, $t = 20$ ps):

$$k = -log[P(t)]/t \tag{8-16}$$

This rate, evaluated for every DNA species as an average from 100 independent simulations, is presented in Figure 8-9.

As expected, the hole transfer is the fastest across the shortest bridge of a single guanine and decreases as one proceeds to longer bridges. However, the distance dependence found in experiment is exponential for up to four AT base pairs in the bridge (Giese et al. [28] as well as Lewis et al. [30] and others), in contrast to our simulations, where the distance dependence for short bridges is much weaker. This indicates that a crucial ingredient is still missing in the simulations, and this will be discussed in the following sections in detail.

8.4.2.3. Effect of the Features of CT Parameters on the Rate of Transfer

As described in Section 8.3.3, the site energies as well as electronic couplings possess interesting features [72], which may manifest themselves in the dynamics of the excess charge and the rate of its transfer.

As for the mere fact that the structural fluctuations of the molecular system make the CT parameters oscillate, an interesting observation has been made: It is exclusively the fluctuations of site energies that accelerate the hole transfer, while the (however immensely) broad distribution of electronic couplings does not play any role from the point of view of the rate of transfer. Thus, if the electronic couplings assume constant values (for instance, their mean values), then the rate of hole

transfer is nearly unaffected. On the other hand, the oscillations of IP of nucleobases (site energies) modulate the energy barrier of the A bridge, which the hole has to overcome. If the site energies are considered constant, the rate of hole transfer converges to zero already for GT_3GGG.

We have also seen that the magnitude of oscillation of site energies is brought about by the molecular environment of the nucleobases, rather than the intramolecular vibrations of these molecules. This is expressed by the much larger oscillations of these if calculated with the QM/MM treatment of the environment (0.4 eV compared to the "in-vacuo" value of 0.1 eV). Consequently, the rate of hole transfer is roughly halved, if the CT parameters are evaluated without the QM/MM inclusion of the environment.

Another interesting observation was that of a significant correlation of site energies for neighboring nucleobases, and it may be of interest to investigate how this affects the actual transfer of charge. In order to answer this question, a set of simulations was performed where the real time series of CT parameters were replaced by sets of values drawn randomly from the corresponding distributions. Slightly slower CT rates were observed with these sets of CT parameters. Remarkably, the difference from the "full" simulations was most profound in the longest bridges. It seems thus evident that the correlations influence mostly the transfer of hole between chemically identical molecules (as is the case of an adenine bridge), with IPs oscillating around the same mean value.

8.4.3. Summary

Most early attempts to simulate CT in DNA used a static picture, with the CT parameters computed for an idealized DNA structure and kept fixed. However, modeling the CT using Marcus' equation with properly chosen values for the electronic couplings and solvent reorganization energy rendered hole transfer to be several orders of magnitude too slow, compared to the experiment [73]. Similarly, our simulation of hole transfer with constant CT parameters showed a too small CT rate (Section 8.4.2.1).

The importance of fluctuations of nucleobases' IP was emphasized in two recent studies [65, 67]. It was shown that these fluctuations bring on a significant population of the A-bridge, which means that the picture of tunneling dominating the CT in short A-bridges is probably incorrect. Note that these studies only included the internal dynamics of nucleobases, neglecting the interaction with solvent. Our simulations discussed in Sections 8.3 and 8.4 show that these effects are significantly enhanced due to the dynamics of solvent.

On the other hand, there was also much speculation regarding the importance of fluctuations of electronic couplings T_{ij}. For instance, Řeha et al. [48] observed the couplings to change along certain normal modes of DNA, however, they did not prove that this actually enhances CT. Our results indeed suggest the opposite: If some individual vibrational modes had the effect of promoting CT, this should be detectable by the dynamical simulations. Our analysis shows that only the

average values of couplings are important, while the exact form of their time dependence does not matter for the CT. Nevertheless, some DNA conformations may be identified as CT-active, while most others suppress the CT. This was analyzed in detail in our recent work [74].

8.5. SOLVENT REORGANIZATION ENERGY AND DE-LOCALIZATION OF THE HOLE

The mere presence as well as the dynamics of the solvent determine the CT parameters to an overwhelming extent, and thus drive the features of transfer of a hole in this way. However, this need not be the only effect of the environment. It must be noted that the study of hole transfer described in Section 8.4.2 (as well as several others) dealt with the dynamics of an eigenstate of the hole rather than with an actual hole. This was caused by the fact that the classical MD simulation did not involve the excess charge, and thus the environment–water, counterions and the DNA backbone–did not "feel" the hole.

Apparently, the presence of a positive charge on a nucleobase[5] will have a strong impact on the molecular environment. While the previous section focused on the effect of solvent on the hole, in the following we will investigate the other part of the interaction – the influence of the hole on the environment.

8.5.1. Polarization of the Environment by the Hole Charge

If a nucleobase in the DNA strand is oxidized, then the nearby water molecules (as well as probably all other parts of the system) change their orientation in order to guarantee the most favorable of possible interactions. This *orientation polarization* will in turn induce electrostatic field on the respective nucleobase, and on the neighboring ones as well.

If the importance of the interaction between the hole and the environment is to be quantified, then it shall be beneficial to evaluate the ESP induced by the environment at the nucleobases, as the ESP directly drives the IPs of these molecules. The result of such a calculation is presented in Figure 8-10. Evidently, the effect of solvent polarization is huge – remember that the decrease of ESP by 1 V would decrease the IP of a nucleobase by 0.9 eV (see Figure 8-5), in the absence of other factors. It is also important to note that large negative values of ESP are experienced not only by the nucleobase that actually carries the positive charge, but also by its neighbors.

The orientation polarization of solvent is changing as the hole is migrating along the DNA strand.[6] The energy cost of the solvent molecules being re-oriented results

[5]Or on a group of nucleobases, if the hole can spread over them.

[6]This phenomenon may be described as a *polaron*, which is accompanying the migrating hole.

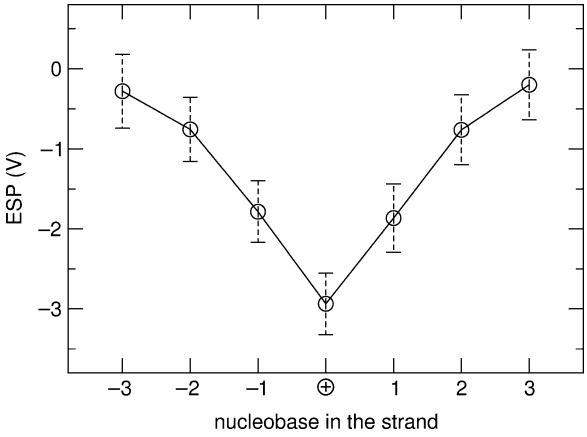


Figure 8-10. Orientation polarization of the environment by the hole. Electrostatic potential induced by the solvent on the guanines of a poly(G) DNA double strand, where the central guanine is oxidized (+). Mean values from MD simulation are presented, error bars denote standard deviation. The nucleobases experience a strong negative ESP due to the re-orientation of the hydration shell

in an additional energy barrier to the transfer, which constitutes the solvent reorganization energy λ_s in the Marcus' equation.

8.5.2. Solvent Reorganization Energy

While it seems to be difficult to extract the magnitude of λ_s from experimental data, the computational efforts aimed in this direction have been quite successful.

Several studies appeared that used dielectric continuum models for the solvent to evaluate λ, however involving a difficult estimate of dielectric constants. In a first work [75], both charge shift and charge separation/recombination reactions were considered. The distance dependence of λ_s and the consequences on the rate of CT were also discussed. LeBard et al. [76] evaluated λ_s for charge separation/recombination reactions, using a special parameterization of continuous solvent, which took into account different dielectric properties of different solvent regions. They further argued that the exponential distance dependence of CT rate may be due to both the donor–acceptor coupling and the reorganization energy. Although the absolute value of resulting λ_s depended on the preset parameter values, the shape of distance dependence was rather invariable. Siriwong et al. used a heterogeneous dielectric model for a charge shift reaction [77]. In addition, the reorganization energy was evaluated for the transfer of a de-localized hole; in this case, smaller values were obtained, with the difference from the case of localized hole of around 0.5 eV.

The studies involving explicit solvent modeled by means of molecular mechanics avoided the problem of ill-determined parameters [78, 79]. MD simulation was applied for hole transfer in DNA as well [80], calculating the reorganization

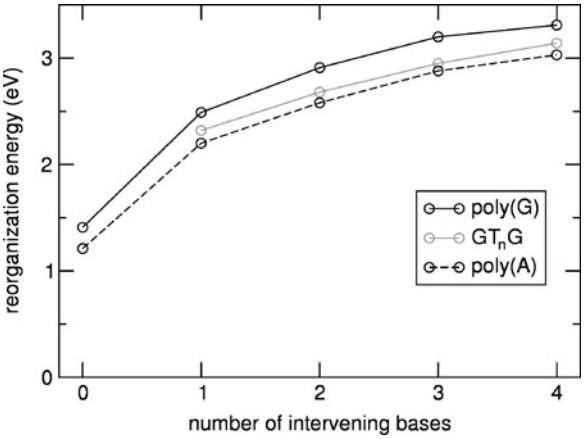


Figure 8-11. Solvent reorganization energy λ_s of hole transfer in DNA. Considered were "regular" sequences poly(G) and poly(A), spanning both A- and B-like conformation; as well as the adenine bridges GT_nG used in the experiments by Giese et al.

energy from the definition – as the difference of total potential energy of the initial state evaluated in an ensemble sampled with the Hamiltonian of the final state, and the potential energy of the initial state.

Obtained were both absolute values of λ_s and its distance dependence, see Figure 8-11. The steep increase on short distances is followed by weak dependence, starting roughly at three intervening nucleobases between the hole donor and acceptor. Also studied was the case of transfer of a hole de-localized over two neighboring nucleobases. Here, the calculated λ_s was 0.9 eV smaller than obtained previously for a localized hole (for every donor–acceptor distance). This result may have consequences for the mechanism of hole transfer.

8.5.3. Spatial Extent of the Hole

Another interesting question, related to the role of solvent, is that of the amount of de-localization of the hole. Whereas the initial contemplations on the mechanism of hole transfer involved a hole confined to a single nucleobase in every instant, more elaborate models (see Section 8.1.2.1) removed this constraint and started to work with a hole spread over several nucleobases. This issue has also been a subject of dedicated computational work.

The important point of hole de-localization was addressed by Kurnikov et al. [81] in a study using a continuum solvation model as well as MD simulation. There is a competition between the quantum effects, which show themselves by spreading the hole over several nucleobases, and the solvation energy, which is more favorable for a hole confined on a single nucleobase. The authors concluded that the

balance between these effects leads to a smaller difference of IP of a single guanine and a GGG triplet (hole trap), and further that the hole will be most likely de-localized over three nucleobases or fewer.

The issue of the solvent effect on the spatial extent of excess hole in DNA was also touched by Mantz et al. [82], who used a special version of DFT corrected for the self-interaction error. The study involved Born–Oppenheimer-type quantum dynamics, i.e. the minimization of energy of the quantum system in every MD step. The reported results prove the utmost close connection between the degree of localization of the hole and the features of the hydration shell, which are manifested by the orientation polarization of water.

Crucial role that the solvent plays in controlling the de/localization of excess charge in DNA is thus given sufficient evidence.

8.5.4. How to Include the Response of Solvent in the Simulation?

Neither the solvent reorganization nor the control of hole de-localization by solvent were captured in the simple simulations of hole transfer presented in Section 8.4.2. For instance, the too large transfer rates as well as the incorrect distance dependence were most likely caused by the missing description of solvent reorganization. Also, the distribution of hole among the individual adenines within the bridge may have been problematic, above all in longer bridges. Therefore, it is of fundamental importance to incorporate these phenomena into the quantum-dynamics scheme.

A step in this direction was made by Steinbrecher et al. [69]. They used MD simulation to sample the DNA configurations, and a QM/MM implementation to capture the effect of solvent. To describe the nucleobases, a one-electron tight-binding Hamiltonian for π-orbitals was employed. The time development of the hole was rendered in Born–Oppenheimer-approximation dynamics, which effectively means that the state of the hole is being energy-minimized in every step of the classical MD. Necessary to mention is here the fact that the environment (solvent and the DNA backbone) is exposed to the effect of excess charge on the nucleobases, which makes the desired polarization effects possible. The authors were able to perform multi-nanosecond simulations of DNA containing an excess hole.

The way being pursued in our laboratory consists in the development of a time-dependent DFT framework to describe the entire system, and in the application of well-defined approximations in order to obtain a computationally efficient scheme. We use the same framework for the coarse-grained Hamiltonian as implemented by Niehaus et al. [83] for the AO-based SCC-DFTB scheme.

The Hamiltonian used in Section 8.4.2 (Eq. 8.14) is essentially a non-self-consistent coarse-grained Hamiltonian evaluated for a neutral reference system (i.e. the complex of DNA bases with no excess charge), and can therefore be likened to the non-self-consistent DFTB Hamiltonian $H^0_{\mu\nu}$ in Eq. (8-9). As in SCC-DFTB, the effect of the excess charge is incorporated by means of a

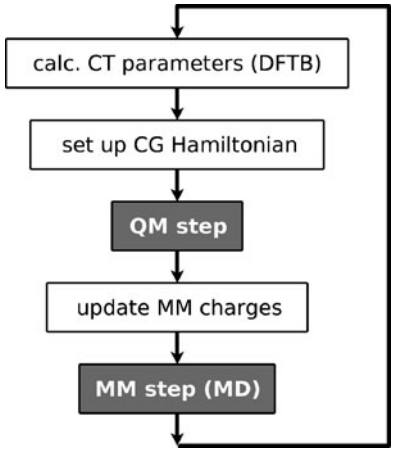


Figure 8-12. The sequence of steps performed to solve the coupled equations of motion

second-order expansion of the DFT total energy. This brings additional terms in the Hamiltonian, which cover the interaction of the hole charge with the environment, as long as a QM/MM formalism is used.

Following Niehaus et al., we have derived a set of coupled equations, which describe the simultaneous propagation of the hole charge and the degrees of freedom of the classical MM system. The hole is propagated by solving the time-dependent DFT equations with the coarse-grained Hamiltonian. Then, in every time step, the hole charge is projected onto the atoms, effectively modifying the atomic MM charges. In this way, the classical MD simulation describes the entire molecular system as well as its interaction with the excess charge (hole).

These sets of equations of motion are then solved simultaneously, see also Figure 8-12: a step of classical MD is followed by an SCC-DFTB calculation of CT parameters, which are cast into the coarse-grained Hamiltonian. Next, a step of TDSE integration is performed; the new wave function of the hole is then used to update the MM atom charges on the nucleobases. With that, the loop is closed and a next iteration begins, again with a step of classical MD. Note that the nucleobases possess time-dependent MM atomic charges, which reflect the current state of the excess charge in the DNA strand.

In principle, this scheme should account for the dynamics of excess charge (hole) in DNA, as ab initio as possible – without the need to adjust any further parameters.

8.6. SUMMARY, CONCLUSIONS AND OUTLOOK

The aim of this contribution has been to introduce the topic of charge transfer in DNA as well as the approaches to describe the actual dynamics of the excess charge. Special attention has been paid to the computational framework being

developed in our laboratory, and some of the results obtained so far were presented. In this concluding section, we will first discuss two phenomena of crucial importance in the hole transfer in DNA (Sections 8.5.1 and 8.5.2). Finally, we will summarize the necessary features of a computational framework to describe the hole transfer.

8.6.1. Fundamental Mechanism of Charge Transfer

Two possible modes of hole transfer were proposed to apply in DNA – the superexchange tunneling and the thermally induced hopping. The former has been considered to dominate on shorter donor–acceptor distances, whereas the latter prevails on longer distances, of roughly more than three nucleobases between the donor and the acceptor. This picture was initially supported by the shape of distance dependence of CT rate, which is exponential on short distances and weak on longer ones.

However, in light of recent studies, there seems to be another possible explanation. The steep, exponential distance dependence may have been brought about by the steep distance dependence of solvent reorganization energy, and not by a tunneling nature of the CT mechanism. A strong argument against the tunneling hypothesis is the considerable hole population on the bridge sites, which was observed in recent simulations of hole transfer across adenine bridges, already in a bridge composed of a single adenine.

Once the issue of fundamental mechanism is resolved, the research interest shall concentrate on the effects fine-tuning this mechanism, most likely with regard to the proposals summarized in Section 8.1.2.2. Since it should be possible to probe the elements of these hypotheses (like the exact role of dynamics and/or the solvent water and the counterions) within simulations based on theoretical models, there is a good chance for computational chemistry to provide valuable information. A topic of interest on an application level is then how the sequence and structure of DNA affect the CT rate.

8.6.2. De/localization of the Hole

Almost all of the many complex mechanisms of hole transfer in DNA that have been proposed so far involve a defined state of the hole, with regard to its spatial extent. The factors affecting the de/localization of hole have been already identified – it is the quantum effects making the hole spread over several nucleobases and, on the other hand, the solvation which acts to confine the hole to smaller volume (down to a single nucleobase). Clearly, the actual state of the hole will be determined by the balance between these effects, and it is almost sure that the dynamics of the system will play a huge role in driving these effects. This includes the dynamics of both DNA and the solvent.

8.6.3. Requirements on a Computational Model

Even though some of the studies that appeared recently disregarded the DNA backbone and the solvent, there seems to be a certain consensus on their key role. The electrostatic field induced by the environment has a huge influence on the IP of nucleobases and thus on the energetics of hole transfer. This phenomenon is likely to be captured by any QM/MM implementation in the form of external point charges. Since the application of such a procedure does not compromise the efficiency of the QM method used to describe the nucleobases, it is possible to say that the environment should be definitely taken into account in the future studies.

Connected to this point is the issue of dynamics of the system. The effect of environment is revealed to the full extent if the energetics of hole transfer is studied in the course of time. Only then is it possible to observe the massive fluctuations of IPs caused by the thermal motion of the DNA backbone, water and counterions. Maybe even more pronounced is the influence of dynamics on the electronic couplings, which exhibit extremely broad distributions. Unless justified by some further considerations (based e.g. on the magnitude of relevant time scales), it seems to be a very crude approximation to substitute the time course of electronic coupling by a single (mean) value.

Also related to the effect of environment on the hole is the other part of this interaction: the influence of the hole on the environment. The solvation shell of a nucleobase undergoes marked changes if the nucleobase is charged. This may be described roughly by the orientation polarization of the solvent, which in turn affects the energetics of hole in the system of nucleobases. Therefore, it is necessary to make the environment feel the charge of the hole, and this practice guarantees a realistic description of the response of environment (solvent).

The right way to obtain knowledge of the microscopic fundamentals of hole transfer in DNA seems to be to use some of the available quantum dynamics approaches (Born–Oppenheimer-like dynamics, non-adiabatic dynamics, surface hopping technique etc.). Given the complexity and computational cost of such methodologies, there is no doubt that the description must be simplified by using reasonable approximations. However, it is indeed desirable to keep the computational model as ab initio as possible.

ACKNOWLEDGEMENTS

The authors are grateful to Ben Woiczikowski for his contribution to this research. The fruitful collaboration with Rafael Gutiérrez and Giovanni Cuniberti is acknowledged. This work was supported by the Deutsche Forschungsgemeinschaft, Project DFG-EL 206/5-2.

REFERENCES

1. Schuster GB (ed) (2004) Long-range charge transfer in DNA I & II, Topics in Current Chemistry, vol 236 and 237. Springer, Heidelberg

2. Wagenknecht HA (ed) (2005) Charge transfer in DNA: from mechanism to application. Wiley, Weinheim
3. Hall DB, Holmlin E, Barton JK (1996) Nature 382:731–735
4. Gasper SM, Schuster GB (1997) J Am Chem Soc 119:12762–12771
5. Porath D, Bezryadin A, De Vries S, Dekker C (2000) Nature 403:635–638
6. Xu B, Zhang P, Li X, Tao N (2004) Nano Lett 4:1105–1108
7. van Zalinge H, Schiffrin DJ, Bates AD, Haiss W, Ulstrup J, Nichols RJ (2006) ChemPhysChem 7:94–98
8. Boon EM, Livingston AL, Chmiel NH, David SS, Barton JK (2003) Proc Natl Acad Sci USA 100:12543–12547
9. Holman MR, Ito T, Rokita SE (2007) J Am Chem Soc 129:6–7
10. Paleček E (1996) Electroanalysis 8:7–14
11. Kelley SO, Barton JK, Jackson NM, Hill MG (1997) Bioconjugate Chem 8:31–37
12. Wong ELS, Gooding JJ (2006) Anal Chem 78:2138–2144
13. Cuniberti G, Fagas G, Richter K (eds) (2005) Introducing molecular electronics, Lecture notes in physics, vol 680. Springer, Berlin
14. Joachim C, Ratner MA (2005) Proc Natl Acad Sci USA 102:8801–8808
15. Slavíček P, Winter B, Faubel M, Bradforth SE, Jungwirth P (2009) J Am Chem Soc 131:6460–6467
16. Orlov VM, Smirnov AN, Varshavsky YM (1976) Tetrahedron Lett 17:4377–4378
17. Steenken S, Jovanovic SV (1997) J Am Chem Soc 119:617–618
18. Saito I, Takayama M, Sugiyama H, Nakatani K (1995) J Am Chem Soc 117:6406–6407
19. Lewis FD, Kalgutkar RS, Wu Y, Liu X, Liu J, Hayes RT, Miller SE, Wasielewski MR (2000) J Am Chem Soc 122:12346–12351
20. Berlin YA, Burin AL, Ratner MA (2001) J Am Chem Soc 123:260–268
21. Jortner J, Bixon M, Langenbacher T, Michel-Beyerle ME (1998) Proc Natl Acad Sci USA 95:12759–12765
22. Bixon M, Jortner J (2001) J Am Chem Soc 123:12556–12567
23. Murphy CJ, Arkin MR, Jenkins Y, Ghatlia ND, Bossmann SH, Turro NJ, Barton JK (1993) Science 262:1025–1029
24. Núñez ME, Hall DB, Barton JK (1999) Chem Biol 6:85–97
25. Wan C, Fiebig T, Schiemann O, Barton JK, Zewail AH (2000) Proc Natl Acad Sci USA 97:14052–14055
26. Kelley SO, Barton JK (1999) Science 283:375–381
27. Meggers E, Michel-Beyerle ME, Giese B (1998) J Am Chem Soc 120:12950–12955
28. Giese B, Amaudrut J, Köhler AK, Spormann M, Wessely S (2001) Nature 412:318–320
29. Lewis FD, Liu JQ, Zuo XB, Hayes RT, Wasielewski MR (2003) J Am Chem Soc 125:4850–4861
30. Lewis FD, Zhu H, Daublain P, Fiebig T, Raytchev M, Wang Q, Shafirovich V (2006) J Am Chem Soc 128:791–800
31. Henderson PT, Jones D, Hampikian G, Kan Y, Schuster GB (1999) Proc Natl Acad Sci USA 96:8353–8358
32. Conwell EM, Rakhmanova SV (2000) Proc Natl Acad Sci USA 97:4556–4560
33. Su WP, Schrieffer JR, Heeger AJ (1980) Phys Rev B 22:2099–2111
34. Barnett RN, Cleveland CL, Joy A, Landman U, Schuster GB (2001) Science 294:567–571
35. Williams TT, Odom DT, Barton JK (2000) J Am Chem Soc 122:9048–9049
36. O'Neill MA, Barton JK (2004) J Am Chem Soc 126:11471–11483
37. Conwell E (2005) Proc Natl Acad Sci USA 102:8795–8799
38. Gervasio FL, Carolini P, Parrinello M (2002) Phys Rev Lett 89:108102

39. Marcus RA (1956) J Chem Phys 24:966–978
40. Marcus RA, Sutin N (1985) Biochim Biophys Acta 811:265–322
41. Newton MD (1991) Chem Rev 91:767–792
42. Priyadarshy S, Risser SM, Beratan DN (1996) J Phys Chem 100:17678–17682
43. Lewis FD, Liu J, Weigel W, Rettig W, Kurnikov IV, Beratan DN (2002) Proc Natl Acad Sci USA 99:12536–12541
44. Olofsson J, Larsson S (2001) J Phys Chem B 105:10398–10406
45. Cramer T, Krapf S, Koslowski T (2004) J Phys Chem B 108:11812–11819
46. Bixon M, Jortner J (2006) Chem Phys 326:252–258
47. Jakobsson M, Stafström S (2008) J Chem Phys 129:125102
48. Řeha D, Barford W, Harris S (2008) Phys Chem Chem Phys 10:5436–5444
49. Voityuk AA, Jortner J, Bixon M, Rösch N (2000) Chem Phys Lett 324:430–434
50. Rösch N, Voityuk AA (2004) Top Curr Chem 237:37–72
51. Troisi A, Orlandi G (2001) Chem Phys Lett 344:509–518
52. Endres RG, Cox DL, Singh RRP (2002) Electronic properties of DNA: structural and chemical influence on the quest for high conductance and charge transfer. arXiv:cond-mat/:0201404
53. Senthilkumar K, Grozema FC, Guerra CF, Bickelhaupt FM, Lewis FD, Berlin YA, Ratner MA, Siebbeles LDA (2005) J Am Chem Soc 127:14894–14903
54. Voityuk AA, Rösch N, Bixon M, Jortner J (2000) J Phys Chem B 104:9740–9745
55. Kubař T, Woiczikowski PB, Cuniberti G, Elstner M (2008) J Phys Chem B 112:7937–7947
56. Loewdin PO (1950) J Chem Phys 18:365–375
57. Elstner M, Porezag D, Jungnickel G, Elsner J, Haugk M, Frauenheim T, Suhai S, Seifert G (1998) Phys Rev B 58:7260–7268
58. Warshel A, Levitt M (1976) J Mol Biol 103:227–249
59. Field MJ, Bash PA, Karplus M (1990) J Comput Chem 11:700–733
60. Cui Q, Elstner M, Kaxiras E, Frauenheim T, Karplus M (2001) J Phys Chem B 105:569–585
61. Elstner M (2006) Theor Chem Acc 116:316–325
62. Troisi A, Orlandi G (2002) J Phys Chem B 106:2093–2101
63. Voityuk AA, Siriwong K, Rösch N (2001) Phys Chem Chem Phys 3:5421–5425
64. Troisi A, Ratner M, Zimmt M (2004) J Am Chem Soc 126:2215–2224
65. Hatcher E, Balaeff A, Keinan S, Venkatramani R, Beratan DN (2008) J Am Chem Soc 130:11752–11761
66. Balabin IA, Beratan DN, Skourtis SS (2008) Phys Rev Lett 101:158102
67. Grozema FC, Tonzani S, Berlin YA, Schatz GC, Siebbeles LDA, Ratner MA (2008) J Am Chem Soc 130:5157–5166
68. Voityuk AA, Siriwong K, Rösch N (2004) Angew Chem Int Ed 43:624–627
69. Steinbrecher T, Koslowski T, Case DA (2008) J Phys Chem B 112:16935–16944
70. Kubař T, Elstner M (2008) J Phys Chem B 112:8788–8798
71. Grozema FC, Berlin YA, Siebbeles LDA (2000) J Am Chem Soc 122:10903–10909
72. Kubař T, Kleinekathöfer U, Elstner M (2009) J Phys Chem B 113:13107–13117
73. Tong GSM, Kurnikov IV, Beratan DN (2002) J Phys Chem B 106:2381–2392
74. Woiczikowski PB, Kubař T, Gutiérrez R, Caetano RA, Cuniberti G, Elstner M (2009) J Chem Phys 130:215104
75. Tavernier HL, Fayer MD (2000) J Phys Chem B 104:11541–11550
76. LeBard DN, Lilichenko M, Matyushov DV, Berlin YA, Ratner MA (2003) J Phys Chem B 107:14509–14520
77. Siriwong K, Voityuk AA, Newton MD, Rösch N (2003) J Phys Chem B 107:2595–2601

78. Ando K (2001) J Chem Phys 115:5228–5237
79. Vladimirov E, Ivanova A, Rösch N (2008) J Chem Phys 129:194515
80. Kubař T, Elstner M (2009) J Phys Chem B 113:5653–5656
81. Kurnikov IV, Tong GSM, Madrid M, Beratan DN (2002) J Phys Chem B 106:7–10
82. Mantz YA, Gervasio FL, Laino T, Parrinello M (2007) Phys Rev Lett 99:058104
83. Niehaus TA, Heringer D, Torralva B, Frauenheim T (2005) Eur Phys J D 35:467–477

CHAPTER 9

QUANTUM-MECHANICAL MOLECULAR DYNAMICS OF CHARGE TRANSFER

VICTOR M. ANISIMOV AND CLAUDIO N. CAVASOTTO
School of Health Information Sciences, University of Texas at Houston, Houston, 7000 Fannin St TX 77030, USA
e-mail: Victor.Anisimov@uth.tmc.edu; Claudio.N.Cavasotto@uth.tmc.edu

Abstract: Computational studies of biological macromolecules are challenging due to large size of biomolecules, their conformational flexibility, and the need in explicit water solvation in order to simulate conditions close to experiment. Under these circumstances studying molecular systems via quantum-mechanical calculations becomes exceedingly difficult. Natural is the attempt to reduce the complex quantum-mechanical picture to a more tractable one by accommodating classical-mechanical principles. However, the simplified models may overlook important physics details of atomic interactions. To avoid such potential pitfalls higher level of theory methods should be available to conduct validation studies. Using semiempirical linear scaling quantum-mechanical LocalSCF method we performed molecular dynamics simulation of ubiquitin in explicit water. The simulation revealed various deviations from the classical mechanics picture. The average charge on amino acids varied depending on their environment. We observed charge transfer channels transmitting electric charge between amino acids in sync with protein motion. We also noticed that the excess charge transferred from protein to water creates a charge cloud around the protein. The observed global dynamic effects of charge transfer represent a new previously unaccounted degree of freedom of biomolecules which requires QM treatment in order to obtain more accurate dynamics of biomolecules at atomic resolution.

Keywords: NDDO method, PM5 Hamiltonian, Ubiquitin, Water droplet, Spherical boundary potential, QM MD, Charge transfer, VFL approximation, LocalSCF, Linear scaling

Abbreviations

AM1	Austin model 1
DFT	Density functional theory
HF/6-31G*	Hartree–Fock method using Pople 6-31G* basis set
LocalSCF	Local self consistent field
MD	Molecular dynamics

P. Paneth and A. Dybala-Defratyka (eds.), *Kinetics and Dynamics*,
Challenges and Advances in Computational Chemistry and Physics 12,
DOI 10.1007/978-90-481-3034-4_9,

NDDO	Neglect of diatomic differential overlap
PM3	Parametric method 3
PM5	Parametric method 5
QM	Quantum mechanics
RM1	Recife model 1
SCF	Self-consistent field
VFL	Variational finite localized molecular orbital approximation

9.1. INTRODUCTION

Despite of their computational cost quantum-mechanical (QM) methods offer many important potential uses in computational biology. The data obtained from QM calculations can be used to validate and assist in improving empirical force fields. QM methods provide a possibility to study problems which are beyond the reach of classical models, particularly the processes accompanied by significant redistribution of electronic density e.g. enzymatic reactions and photochemical processes, to name a few. However the progress of such applications is limited by severe performance bottleneck of QM methods. Going to progressively larger systems the resource requirement of QM calculations quickly becomes prohibitive. The resource requirement scales as power of 4 of the number of basis functions for Hartree–Fock and Density Functional Theory (DFT) methods and as power of 5 for MP2 method. Coupled cluster calculations using singles and doubles scale as power of 6. Therefore depending on the level of theory the conventional QM methods are practically limited to systems containing 100–1,000 atoms, whereas typical biological systems including solvent encounter tens of thousands of atoms minimum.

Performance of QM methods has recently been significantly improved due to the development of linear scaling methods [1, 2]. In the linear scaling regimen the computation time increases linearly of the number of employed basis functions. Now when the scalability problem seems to be solved it might look like the road for QM methods to biological application is eventually going to be wide open. However, making such significant step forward was only to discover even more impenetrable wall erected by structural flexibility of biological systems.

Since the pioneering work of McCammon et al. [3] it has long been recognized that biomolecules are dynamic entities and that their biological structure and function are linked to their thermal motions. This makes insufficient the standard method of attack of quantum chemistry in the form of geometry optimization which was developed for and very efficiently applied to small molecule problems. Applying geometry optimization to large and flexible biomacromolecules is going to give information only about one of the million of energetically possible conformational local minima. Collecting such limited data about few conformations is obviously

insufficient to make computational predictions about macroscopic properties of large biomolecular systems.

This problem is known as configurational sampling and it requires application of molecular dynamics (MD). According to statistical mechanics the complexity of free energy profiling of a flexible molecular system grows exponentially with the number of atoms in the system. After facing this sobering fact it becomes apparent that the further expansion of QM methods toward biological problems should go via development of computationally inexpensive approximate QM methods. Semiempirical QM methods based on neglect of diatomic differential overlap (NDDO) are particularly attractive in this regard. They are about 1,000-fold faster than chemically accurate DFT methods. Correspondingly, same order of the resource saving is expected during MD simulation based on the given level of QM theory. Therefore semiempirical QM methods provide the first outpost on the long journey to bring QM platform to solution of biological problems.

QM/MM is another example of the approximate platform where a most critical part of the system is treated at QM level whereas the least important part of the system is treated by classical force fields. Typical applications of QM/MM include enzymatic reactions [4]. The QM/MM methods are not necessarily limited to semiempirical theory; quite contrarily selecting small QM area makes feasible using a higher level of theory, usually DFT methods. Although useful this approach cannot be a substitute for all-atom QM treatment. For instance, QM/MM cannot be used to validate and further improve force fields or to further advance our knowledge about structure and dynamics of biomacromolecules, the task which requires consistent application of a single level of theory to the entire molecular system.

Resorting to semiempirical formalism does not exhaust the potential of approximate QM methods. Additional performance can be obtained by reformulating the large-scale problem. Such method was proposed by Anikin et al. [5] suggesting to seek the molecular density matrix in the basis of constrained molecular orbitals. The complete account of the theory of the variational finite localized (VFL) molecular orbital approximation and the linear scaling LocalSCF method was given by Anisimov et al. [6].

9.2. THEORETICAL PART

In LocalSCF method and in the underlying VFL approximation the solution to large-scale molecular problem is sought via decomposition of the potential energy of the system, E, as function of density matrix, **P**, in Taylor series and neglecting the second and higher order terms $\delta^{(2)}$

$$E[\mathbf{P}] = E[\mathbf{P}(0)] + \partial E/\partial \mathbf{P} \cdot [\mathbf{P} - \mathbf{P}(0)] + \delta^{(2)} \tag{9-1}$$

Such expansion given by Eq. (9-1) basically implies that if we know an approximate density matrix, **P(0)**, and the perturbation is small we can improve the density matrix toward the final solution. This is effectively utilized in the VFL

approximation to maintain orthogonality of compact molecular orbitals (CMO). Considering the non-orthogonality as a small perturbation to the fully orthogonal case we can additionally expand the inverse overlap matrix $\mathbf{S^{-1}}$ in Taylor series preserving only the linear terms

$$\mathbf{S}^{-1} = \mathbf{I} - (\mathbf{S} - \mathbf{I}) + \cdots \tag{9-2}$$

This leads to the final expression for the total energy, where quadratic term represents a penalty function which helps minimizing the nonorthogonality between CMOs

$$E \approx E[\mathbf{P}(0)] - \partial E/\partial \mathbf{P} \cdot [\mathbf{C} \cdot (\mathbf{S} - \mathbf{I}) \cdot \mathbf{C}^t] + \mathbf{W} \cdot [\mathbf{S} - \mathbf{I}]^2 \tag{9-3}$$

Here, $\mathbf{W}$ is penalty weight, $\mathbf{I}$ is diagonal unit matrix, $\mathbf{C}$ is matrix of linear coefficients of CMOs, and $\mathbf{C}^t$ is its transpose.

Starting from the initial atomic orbital (AO) expansion of CMOs the VFL approximation can variationally determine density matrix corresponding to this expansion. During the self consistent field (SCF) refinement of linear coefficients of CMOs the VFL method holds fixed the AO expansion of each CMO. Therefore VFL approximation is the method for variational determination of density matrix under the constraint of fixed CMO expansion. Correspondingly, besides pointing to the compact size of molecular orbitals the CMO abbreviation also stands for constrained molecular orbitals.

Since the VFL method assumes that the CMO expansion is known before starting the calculations and provides no mechanism to determine the optimal expansion the VFL method alone is insufficient to perform QM calculations on large molecular systems. The complementary task of finding the optimal CMO expansion and the corresponding molecular density matrix is resolved in the linear scaling LocalSCF method, which is based on the VFL approximation. Similar to VFL the LocalSCF method also relies on the perturbation expansion (Eq. (9-1)) but uses it to determine the most energetically favorable direction for expansion of CMO i on trial AO μ based on the value of energy derivative

$$\frac{\partial E[\mathbf{P}(0)]}{\partial C_{i\mu}} \tag{9-4}$$

The trial atomic orbital μ will be added to CMO i if the corresponding energy derivative defined by Eq. (9-4) is the largest among the derivative values corresponding to other alternative AO expansion directions. The success of accurately determining the most energetically favorable expansion depends on the quality of the available density matrix. Having the density matrix $\mathbf{P}(0)$ fully converged assures determining best possible and minimally necessary AO expansion leading to rapidly converging total energy and density matrix via the series of successive expansions.

The localized character of CMOs offers a beneficially restricted choice of possible expansions onto only the nearest atomic centers. Further limiting the CMO expansion

to the most energetically favorable AOs helps keeping CMOs maximally localized while providing the best possible improvement to the density matrix. In this way LocalSCF elegantly demonstrates the flexibility advantages of the MO-based linear scaling formalism versus density functional (DFT) approach in dealing with large-scale molecular problems.

Useful in interpretation of the results of quantum-chemical calculations is the concept of partial atomic charges, Q_A, where capital letter A denotes atom index, index i runs over CMOs, and index a corresponds to AOs centered on atom A.

$$Q_A = \sum_{a \in A} P_{aa} = \sum_a \sum_i C_{ia}^2 \tag{9-5}$$

Throughout this work we will be using Mulliken algorithm as our method for calculation of partial atomic charges (Eq. (9-5)). In semiempirical methods Mulliken charges reduce to Coulson charges due to the NDDO approximation but referring to them as Mulliken charges is still formally correct and preferable for the sake of theoretical consistency. The Mulliken charges are most natural for the semiempirical framework and unlike other charge derivation schemes do not invoke additional parameters. Besides, since semiempirical methods are parametric ones they are free from the basis set superposition error. Consequentially, the troubles ab initio methods have with the Mulliken population analysis do not extend to semiempirical methods.

9.3. THE NOTION OF CHARGE TRANSFER

When considering molecular dynamics of biomolecules it is common to question the need in the QM level of theory when classical force fields are doing seemingly well. This is a rhetorical dilemma of any established field of science. However, it has long been recognized that the fixed-charge approximation staying behind the empirical force fields may not be optimal when considering environments of different polarity [7]. Correspondingly, force field parameters optimized to reproduce properties of a model compound in hydrophobic environment may poorly perform in hydrophilic environment and vice versa. Addressing such problems has long been the focus of polarizable force field development [8–10]. However, despite of the achieved significant progress [11], the description of induced electrostatic effects at classical mechanics level turned out to be an exceedingly difficult problem, and after 2 decades of intense efforts additional development is still necessary [12]. Besides, the classical polarizable model captures only a portion of the large spectrum of induced electrostatic effects while the other part – intermolecular charge transfer, an inherently quantum mechanical phenomenon, has no classical mechanics formulation. This is a particularly important aspect to consider in the design of computational methods, since according to Nadig et al. [13] energetic consequence of charge transfer effects is greater than that of polarization effects.

Historically there was much uncertainty about the significance of charge transfer effects in molecular systems. Small systems were too small to show any sensible amount of charge transfer whereas studies on large systems were technically not feasible. However the progress in computer hardware and development of linear scaling methods provided the opportunity to reconsider the previous assumptions. The subject of charge transfer in ion solvation was extensively studied by theoretical [14–16] and experimental methods [17]. Charge transfer in DNA has received particularly strong attention [18, 19]. Charge transfer in glycine dimer in water solution was confirmed via DFT MD by Peraro et al. [20] Merz and co-workers [13] predicted −2.0 electron units of charge transfer from a small cold shock protein to water by performing single-point PM3 calculations on 100 protein–water snapshots generated by force field MD. Despite of the later finding [21] that single-point calculations and particularly PM3 are not fully reliable for estimation of charge transfer, this work along with other studies [22] fueled the interest to the subject of charge transfer in biological systems.

In another study Komeiji et al. [23] predicted protein-to-water charge transfer of −0.8 electron units for protein ubiquitin by performing HF/6-31G* single-point calculations over five empirical snapshots. The limited number of configurations considered in their study was compensated by the use of ab initio Hartree–Fock level of theory, which is the highest level of theory so far applied to the problem in question. This result most definitely confirmed the presence of charge transfer in biological systems. Yet two major questions remained – first, what are the implications of charge transfer for biological simulations and, second, how to turn the knowledge about charge transfer into a predictive biomolecular simulation?

Considering the first question it is not entirely clear that the omission of charge transfer of −0.8 electron units by classical force fields is something to worry about. For a typical protein–water system consisting of 10,000 atoms the charge transfer would be at the order of 10^{-4} electron units per atom. Such electrostatic perturbation will be entirely lost in the numerical noise of MD simulation. So the significance of charge transfer for biological applications still remains to be demonstrated.

Second, putting temporarily aside the proof and simply assuming the significance of charge transfer effects the derived question would be how to properly revise the computational protocol of biomolecules in order to make the practical use of charge transfer effects? Both above simulations of Merz and Komeiji were performed via single-point QM calculations on snapshots extracted from force field MD trajectories. Unfortunately, such two-step procedure has little or no use for typical biomolecular applications which are usually concerned with sampling a canonical statistical mechanics ensemble in order to relate the computational model to the actual experimental conditions. The QM total energies computed based on the empirical snapshots do not represent any statistical ensemble because of the expected differences between the QM and force field potential energy hypersurfaces. Therefore averaging the results of such QM calculations is not going to produce a value corresponding to any experimental condition. To generate a proper statistical ensemble one has to run a QM MD simulation.

Such QM MD simulation could also bring us closer to the solution of the problem of significance of charge transfer. There is some sort of inconsistency in studying charge transfer effects based on empirical trajectories which are generated under the convention of fixed charges. In computational biology a significance of a phenomenon is necessarily understood via its effect on structure and dynamics of a biomolecule. Putting it forward, if the effect in question does not influence the dynamics trajectory it has likely no significance at all for biomolecular applications. Hence, if charge transfer is important it must be present during the production of the MD trajectory. Since this condition is not satisfied when QM calculations are performed over empirical trajectories, it becomes impossible to determine physical significance of the computed charge transfer. This dilemma can only be resolved by performing QM MD simulation and analyzing the generated data.

9.3.1. QM MD of Ubiquitin in Explicit Water

To clarify the issue of significance of charge transfer in protein–solvent systems we performed a series of QM MD simulations on 12,199 atoms system representing ubiquitin in 30 Å water droplet for the duration of 20 ps [21] using AM1 [24], RM1 [25], PM3 [26] and PM5 [27] Hamiltonians. The size of the water droplet was selected to provide 10 Å water cushion between protein and vacuum. The system boundaries were restrained by application of a spherical boundary potential preventing water molecules from leaving the system. Constant temperature of the system was maintained at 310 K by application of Nose-Hoover thermostat [28]. No restraints were applied to bonds involving hydrogen atoms. Integration time step was set to 1 fs. Numerical integration of dynamics equations was performed by using a constant volume – constant temperature algorithm [29].

Despite of finding of Merz and co-workers that AM1 provides best agreement with correlated ab initio methods on the value of charge transfer [30] this Hamiltonian turned out to be unsuitable for MD simulations in presence of explicit water. AM1 resulted in unrealistically high water density of about 2 g/cm^3 and showed an unphysical tendency to water dissociation at very first picosecond of the simulation. From this we conclude that AM1 is not applicable for QM MD simulations of biomolecules in presence of explicit water. This is kind of a sad finding considering the widespread popularity of AM1 in biomolecular applications [31].

Another Hamiltonian, RM1, representing recent reparameterization of AM1, showed improved results. It produced correct water density and stronger OH bond than that in AM1. Unfortunately the original deficiency of, AM1 related to the underestimated strength of OH bond in water, is still somehow lurking in RM1 leading to water dissociation happening after 5 ps of simulation. Correspondingly, both AM1 and RM1 Hamiltonians were excluded from our further QM MD simulations of ubiquitin in water.

For information, considering the experimental strength of OH bond in water, a spontaneous dissociation of a given water molecule is expected on 11-h time scale

[32] which makes it unobservable in regular MD simulations. Contrarily to AM1 and RM1, the PM3 and PM5 MD simulations were entirely stable and predicted protein-to-water charge transfer of −1.60 and −0.45 electron units, respectively [21].

Having the data of Komeiji as the reference point (−0.8) we can see now that PM3 overestimates charge transfer as previously noticed by Merz [30]. However, the reparameterized variant, PM5, shows quite promising result. The agreement between PM5 MD and HF/6-31G* is nearly quantitative. There are also reasons to believe that both semiempirical and ab initio levels of theory might come to even a closer agreement. Since the HF/6-31G* simulation was performed on a rigid force field geometry one may expect that the HF/6-31G* result would approach the PM5 MD value given the opportunity to relax the structure at the HF/6-31G* level. In addition to that, a smaller value for charge transfer seems more credible considering the multitude of factors increasing the appearance of charge transfer in ab initio calculations. These could be strained geometry, basis set superposition error, insufficient sampling, missing correlation effects and so on.

9.3.2. Charge Transfer Inside Protein

In our PM5 MD simulation of formally neutrally charged ubiquitin in water droplet [21] the instantaneous net protein charge fluctuates around the average value of

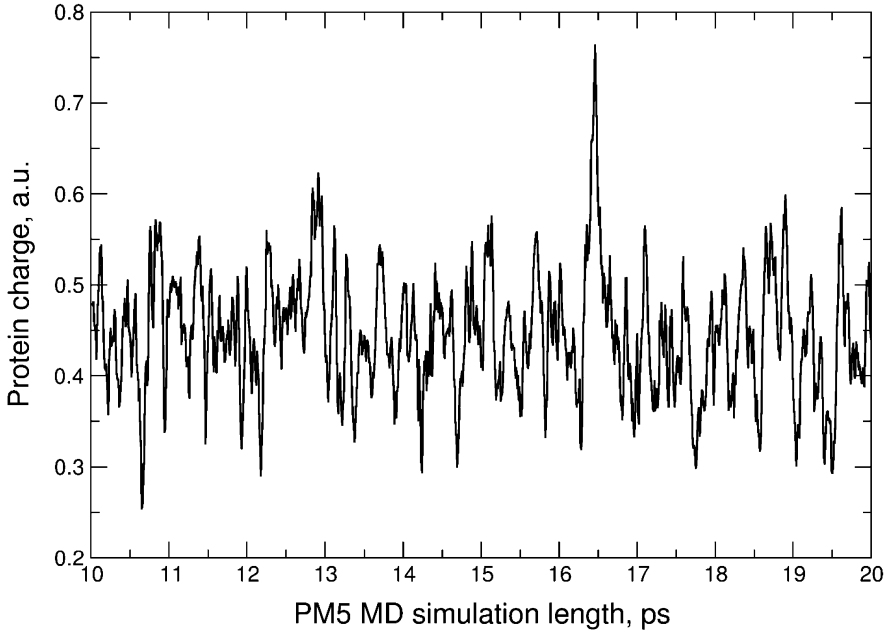


Figure 9-1. Protein net charge fluctuation during MD simulation. Reproduced from Ref. 21 with permission. Copyright Wiley-VCH Verlag GmbH & Co. KGaA

0.45 electron units (Figure 9-1). Standard deviation of the protein charge is 0.07 electron units; minimal and maximal values of the protein charge are 0.25 and 0.76 electron units, respectively.

At this point the obtained results alone do not move us much further in answering the primary question of this study whether charge transfer has or has no any biological significance. Yet a critical step is achieved in making sure that our semiempirical QM MD prediction of protein-to-water charge transfer closely agrees with the reference ab initio value of Komeiji. This assures the predictive capability of our method and supports going into deeper analysis of the accumulated data. The obtained QM MD trajectory provides the necessary means to reconstruct the time evolution of charge transfer taking place inside and outside of the protein and obtaining expected values of physical quantities as the canonical ensemble trajectory average.

9.3.3. Charge Transfer Channel

When considering charge transfer in macromolecules first step coming to mind is to identify the location of charge donors and recipients. According to the convention adopted by polarizable force field the charge fluctuations are confined to single residues [33] and cannot go beyond their boundaries. Any attempt to permit the charge to freely travel between amino acids would be technically so difficult to implement and to optimize the necessary parameters so such opportunity is even not considered in the modern polarizable force fields. Of course, in reality it is hard to know beforehand the location of charge donors and charge recipients. This makes the charge transfer phenomenon beyond the rich of empirical force fields. Contrarily to that, quantum mechanics permits any imaginable charge distribution in the system and offers a ready mathematical apparatus to find the distribution corresponding to the minimum of potential energy of the system.

In the domain of electrostatic phenomena, protein-to-water charge transfer is only a tip of the iceberg. As our QM MD calculations reveal, all parts of the simulation system are actively involved in various charge transfer processes. Tracking these processes is not all easy. Since all electrons in the system are non-distinguishable we can only hope to find patterns in the way some regions of the system depleting the electron density while other areas simultaneously acquiring the excess density. Following this technique we observed a distinctly localized charge flow channel (Figure 9-2) formed by amino acids E34 and K11 participating in a salt-bridge interaction.

The top panel in Figure 9-2 shows net charge fluctuation from -1.11 to -0.61 electron units on glutamic acid residue E34 along the dynamics trajectory. The net charge on its salt-bridge partner lysine K11 varies exactly in anti-phase from $+0.60$ to $+1.12$ electron units. In any moment of time this pair carries nearly perfect zero net charge, which helps us to conclude that these two residues are involved in some sort of a space-confined dynamic charge exchange, operating almost independently from the rest of charge transfer processes happening around. During their thermal

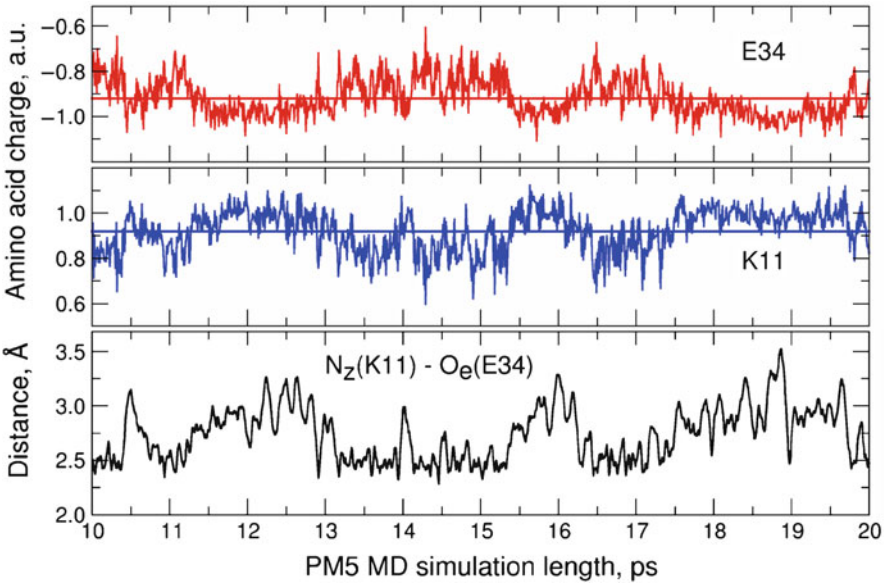


Figure 9-2. Synchronous charging and discharging of E34–K11 residues following the pattern of their motion. *Horizontal lines* across the charge fluctuation profiles denote average charge value. Reproduced from Ref. 21 with permission. Copyright Wiley-VCH Verlag GmbH & Co. KGaA

fluctuation these amino acids pump through their bodies about 0.5 units of electron charge, which is alone equal to the value of protein-to-water charge transfer. Since 0.5 units is a half of the effective charge associated with these titratable amino acids such magnitude of charge fluctuation strongly disagrees with the popular idea of force fields assigning a fixed unit charge to amino acid residues. This also shows limitations of confining charge transfer to the boundaries of individual amino acids.

The non-classical character of this salt bridge is revealed in the way the amino acid charges change. We already mentioned that during dynamics trajectory this pair maintains electric neutrality in the sea of charge fluctuation. Nearly all the electric charge lost by one amino acid goes to its partner. Most remarkably, the charge pump operates synchronously as the distance between amino acids changes (Figure 9-2). When E34 and K11 come closer to each other the electron density travels from E34 to K11 discharging the so called capacitor formed by the amino acid pair. During the discharging process the energy is released to produce a mechanical work. When distance between the donor and acceptor sites increases the electron density flows in opposite direction from K11 to E34 thus charging up the capacitor and accumulating the potential energy. This complex picture of intermolecular charge exchange directly connects the charge transfer effects with protein structural dynamics making them functionally linked.

The effect discussed above is a distinct quantum-mechanical phenomenon, which is overlooked in typical biomolecular simulations. Yet some may argue

that it also can be described by using classical mechanics methods. For instance enclosing the charge transfer channel in a single structural unit within a charge fluctuation polarizable force field scheme may do the work. However, since we do not know beforehand where such charge transfer channels will be located and how to handle the dynamic processes of creating and disrupting the charge transfer channels during biomolecular dynamics the empirical force field treatment of intermolecular charge transfer will not be practically feasible. Although similar problems are successfully handled by reactive force fields [34] they are typically limited to isotropic systems while biomolecules are highly anisotropic. Therefore, QM methods are necessary in order to treat intermolecular charge transfer in biological systems.

The analysis of net charge associated with E34 and K11 residues shows additional deviations from the classical picture. The average charge on E34 and K11 is -0.92 and $+0.92$ units, respectively, which is by 8% smaller from the unit charges of -1 and $+1$ assigned to these amino acids by classical force fields. The charge on E34 and K11 amino acids fluctuates around the average with the standard deviation of 0.08 and 0.09, respectively. Similar in magnitude charge fluctuation was earlier observed by Yang and co-workers [35] in their QM/MM treatment of crambin in water where protein was treated quantum mechanically whereas solvent was considered at force field level.

It is apparent that fixing the residue charges to unit values as it is done in empirical force fields would result in a different interaction energy profile between amino acids and would ultimately lead to a different trajectory. It would also result in predicting different mechanical properties of macromolecules. Since mechanical flexibility of proteins is an inherent component of their biological function the correct handling of charge transfer is necessary in order to further improve the accuracy of computer simulations of biomolecules.

Looking at the smooth work of the charge transfer channel E34–K11 it is tempting to suggest a potential biological function of this unit. The mechanic motion of the protein generates an alternating electric current at the E34–K11 salt bridge which in turn produces an infra-red frequency electromagnetic wave. Such electromagnetic wave can be sensed by other charge transfer channels and through that obstruct or facilitate a specific protein motion. In fact the E34–K11 salt bridge in ubiquitin may itself serve as an electromagnetic sensor by converting the absorbed electromagnetic energy into mechanic motion. Such signaling mechanism may control protein–protein binding recognition. This in turn may be a part of a signal transduction mechanism consisting of the cascade of protein–protein binding events [36]. These are of course only abstract thoughts at this moment. Additional studies are necessary in order to prove the feasibility of such mechanism.

9.3.4. Inequality Among Same-Type Amino Acids

It is common in empirical force fields that same-type amino acids carry same set of parameters – charges and polarizabilities. Therefore despite of saddle differences

in their environment same-type amino acids usually exhibit very similar physical response. Such model is ideally suited to describe bulk properties of materials where average response is what matters. In such models large fluctuations of properties at the level of individual blocks are even unwelcomed due to the accompanied computational noise. Such problem already becomes visible in polarizable force fields versus non-polarizable ones. It takes longer time to obtain same level of convergent properties in polarizable force fields, particularly in weak electrostatic field environments, than in non-polarizable ones [37]. Therefore, because of adopting a simplified electrostatic model empirical force fields are not capable to fully substitute quantum-mechanical treatment when atomic level resolution is necessary [38]. Next we will see what kind of unusual effects one can observe at atomic level.

While performing charge transfer analysis along dynamics trajectory it is reasonable to expect significant changes in charge distribution taking place at electron-rich or electron-deficient sites. To our great surprise, ubiquitin showed strong charge fluctuations happening at the sites of neutral amino acids too. Large charge fluctuation in the range of 0.50 electron units took place at formally neutral amino acid tyrosine Y59 (Figure 9-3). During the course of the simulation its net charge varied from -0.39 to $+0.11$. This charge fluctuation could happen due to hydrogen-bond interaction of Y59 with negatively charged glutamic acid E51. The question is can we predict this outcome without performing QM calculation? As a neutrally charged amino acid the tyrosine is not expected to exhibit such large charge fluctuation. Definitely, typical tyrosine in a protein behaves exactly as expected. What we see in ubiquitin is a unique environment of Y59 which forces it to deviate

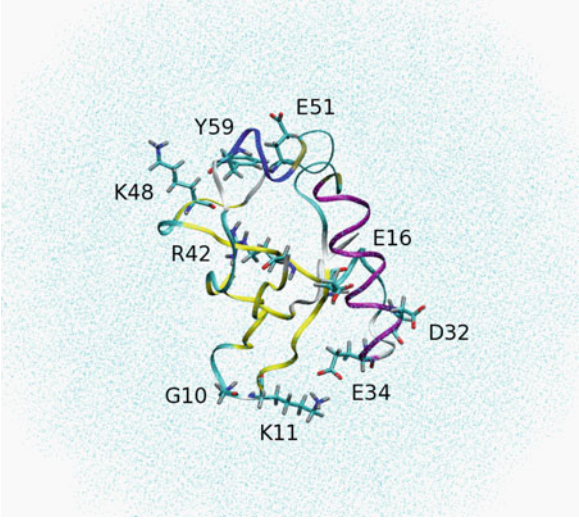


Figure 9-3. Ubiquitin in water droplet. Denoted are amino acids showing the largest charge transfer effects

from the average. The ability to pinpoint such details makes QM calculations irreplaceable when considering the ways of improving the predictive ability of computational methods in biology.

Even more unusual is large charge fluctuation of 0.43 electron units happening on amino acid glycine G10. This simple amino acid assumes extreme net charge in the range from −0.20 to +0.23 electron units during the simulation. G10 does not participate in hydrogen-bond interactions with other amino acids but its neighbor K11 could be responsible for inducing the charge fluctuations on G10. Glycine G10 is fully exposed to solvent and its carbonyl oxygen is in dynamic charge exchange with water serving as the charge transfer partner to G10. Since glycine is not known for high polarization response abilities and because force fields treat all glycine amino acids equally, it is likely that the unique electrostatic properties of G10 in ubiquitin would be overlooked by polarizable force fields. According to our QM MD data, other glycine amino acids in ubiquitin behave nothing but ordinarily. This is another example of quantum-mechanical effects playing a significant role in the place where they are least expected.

Besides of showing deviations from the classical-mechanical picture the analysis of QM data also provides some evidences supporting classical model. For the majority of amino acids their average net charge agrees quite well with the classical assignment of unit charges. Same fact was earlier noticed by Yang and co-workers in their QM/MM simulation of crambin [35]. Excellent agreement between average QM and classical charges can be seen on amino acids G10, E16, and K48 in Table 9-1. Particularly interesting is the case of glycine G10. In the light of the above discussion we know that this amino acids exhibits quantum-mechanical effects in the form of unusually large instantaneous charge fluctuations whereas on average G10 looks like a classical neutrally charged amino acid. This is an illustrative example of a difference between the average response and actual atomic

Table 9-1 Average[a] and extreme net charges on selected amino acids

Residue	qQM_{ave}	qQM_{min}	qQM_{max}	qFF[b]	Δq[c]
N-ter	0.87 (0.04)	0.74	0.98	1	−0.13
G10	0.03 (0.05)	−0.20	0.23	0	0.03
K11	0.92 (0.09)	0.60	1.12	1	−0.08
E16	−0.99 (0.05)	−1.13	−0.84	−1	0.01
D32	−0.91 (0.05)	−1.08	−0.76	−1	0.09
E34	−0.92 (0.08)	−1.11	−0.61	−1	0.08
R42	0.95 (0.05)	0.81	1.07	1	−0.05
K48	1.00 (0.04)	0.85	1.15	1	0.00
E51	−0.84 (0.09)	−1.10	−0.55	−1	0.16
Y59	−0.12 (0.08)	−0.39	0.11	0	−0.12
C-ter	−0.82 (0.04)	−0.92	−0.71	−1	0.18

[a]Data in brackets represent standard deviation
[b]Charge assigned by force field
[c]$\Delta q = qQM_{ave} - qFF$

level resolution processes. Force fields are definitely good in describing the average properties whereas QM is necessary when simulations are performed to obtain atomic level resolution.

Concerning the differences between the quantum- and classical-mechanical models, according to our QM MD data for 76 ubiquitin amino acids, their average charge deviates from the classical values on overall by 3% only. It explains the general success of empirical force fields in computational biology. However, the differences between classical and quantum mechanical pictures increase when looking at the details of individual amino acids. Most noticeable is charge difference on terminal residues C-ter and N-ter, which participate in charge transfer to water. Since these residues are fully exposed to solvent they act almost like free ions.

Even though for the majority of amino acids the classical units charge is a good match, it is not that good for K11, D32, E34, G51, and Y59. These are the examples where average charge of amino acids strongly depends on their environment. Most interesting is to compare average charge on same-type amino acids. The classical expected charge of $+1$ agrees with our QM MD prediction for K48 ($Q_{ave} = 1.00$) but it is not applicable to K11 ($Q_{ave} = 0.92$). The most unusual average charge of -0.84 electron units is predicted for glutamic acid E51, whereas charge on same-type amino acid E16 is quite classical (-0.99; see Table 9-1). From this comparison it becomes apparent that the assignment of a unified unit charge on amino acids in anisotropic protein environment is not fully correct. Hence, this is a factor reducing accuracy of computer simulation of biomolecules. Classical force fields can in principle manipulate with fractional amino acid charges. However in order to derive such charges one would need to run QM MD simulation first. Such computational protocol may indeed be of certain utility.

Besides of the non-equivalent average charge determined for same-type amino acids by our QM MD calculation, we can also see global-scale dynamic charge fluctuations in ubiquitin as displayed in Figure 9-4. According to these data every amino acid including neutral ones experiences large charge fluctuation of 0.25 electron units or more. Supporting the obtained numerical values is our previous observation that PM5 MD is even less prone to exaggerate charge transfer than ab initio HF/6-31G* level. In contrast to empirical polarizable models artificially confining charge fluctuation to individual amino acid fragments, in QM picture the charge can travel from and to any distant part of the system. Since this effect cannot be described by polarizable force fields this makes the current findings even more compelling. By averaging out charge fluctuation and resorting to fixed charges the classical model makes itself computationally more attractive but also less sensitive to perturbations. The charge density freely floating in the system, as in QM picture, is a sensitive probe to the environment changes. Subtle changes in the environment alter protein charge flow and the latter instantaneously transmits this information to the level of macromolecular motion. Therefore integrating global charge transfer effects into computational scheme is the necessary way to improve accuracy of biomolecular simulations.

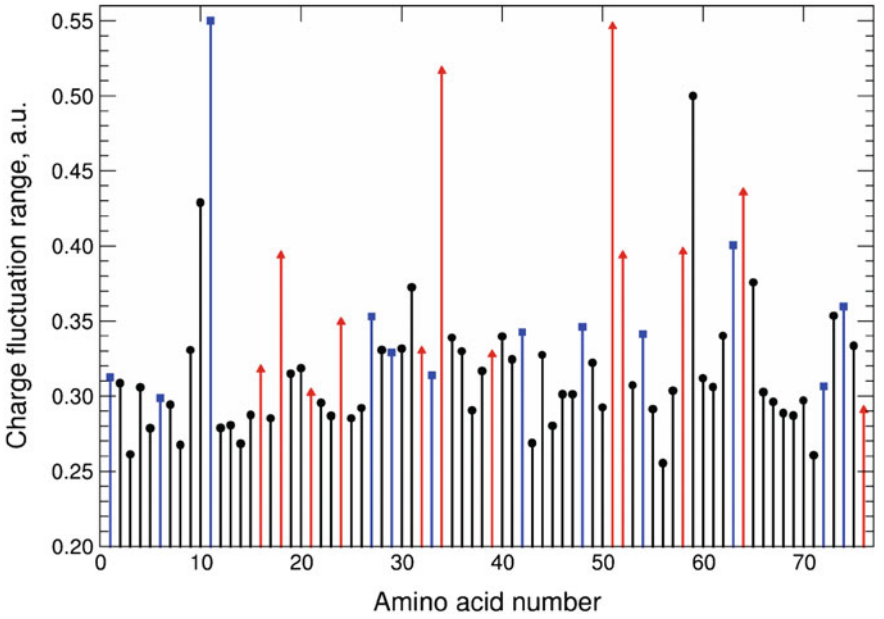


Figure 9-4. Extreme range of amino acid net charge fluctuation. *Rectangles*, *circles*, and *triangles* denote positively charged, neutral, and negatively charged amino acids, respectively

9.3.5. Protein-Solvent Charge Transfer

So far we limited the analysis of charge transfer effects solely to the processes happening inside the protein. However, since the original charge transfer effects pertaining to biological systems were observed on protein–water interactions [13] it is reasonable to expect large charge transfer effects taking place at the protein–water interface. Therefore it is necessary to extend the charge transfer analysis to solvent molecules. Among water molecules the largest charge fluctuation of 0.24 electron units happens in vicinity of protein N-terminus. Residue K48, which is fully exposed to solvent, induces maximal charge fluctuation of 0.21 electron units to the nearest water molecules. Similar charge fluctuation on water molecules happens in vicinity of the other solvent exposed ionized amino acids.

Now we can go back to the original question of significance of protein-to-water charge transfer. It was presumed that the charge transfer value divided by the total number of atoms will be entirely lost. However, as we could see above, large charge fluctuations happen at the level of individual solvent molecules. It helps identifying a flaw in the initial reasoning involving division by the total number of atoms, which is unacceptable. The charge transfer effect would be really negligible when dealing with isotropic systems like neat liquids where the bulk properties dominate over microscopic details. But proteins are highly anisotropic systems and averaging over the number of atoms is not physical. Such averaging compromises the main

purpose of biomolecular simulations in deducing the atomic-level details which requires the precise accounting of actual physics at every single atomic position. We could also see that water molecules are not all equivalent as presumed in empirical force fields. The water molecules close to protein participate in dynamic charge transfer and carry large excess charge. This makes them different from the bulk water molecules.

The charge transfer effects go somewhat further than just invalidating the charge equity between individual solvent molecules at microscopic level. A deviation from the classical picture can even be seen on a mesoscopic scale of large group of solvent molecules. The charge transfer between protein and solvent molecules induces excess charge on solvent molecules close to protein and the charge value depends on the distance between water molecules and protein surface as depicted on Figure 9-5. To obtain these data the charge on water molecule is computed as the sum of partial atomic Mulliken charges on hydrogen and oxygen atoms. The distance is measured between oxygen atom of water molecule and the nearest protein atom, which may be hydrogen; next the computed charge is averaged over other water molecules located at the same distance and over the QM dynamics trajectory.

According to our data the charge on water molecules changes its sign as the function of distance from protein (Figure 9-5). Due to water geometry and our choice of measurement of the protein–water distance the nearest water molecules are

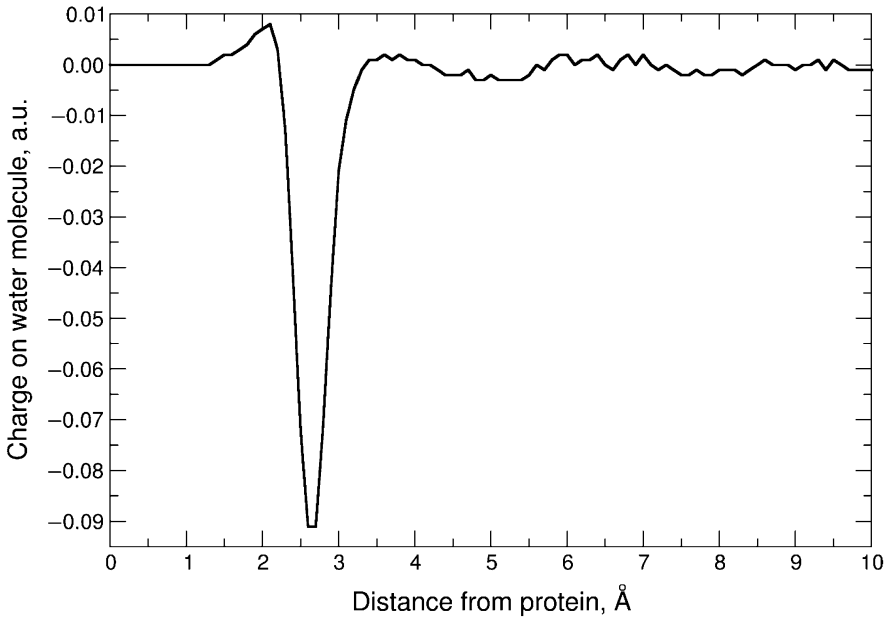


Figure 9-5. Excess charge on individual water molecules as a function of distance to protein, averaged over PM5 MD trajectory

those involved in solvation of positively charged protein sites (protein–$H^+\cdots$OHH) with the charge intensity peak happening at 2.2 Å. Second characteristic peak takes place at 2.6 Å. It presumably describes water molecules solvating electron-donor sites of the protein (protein–$O^-\cdots$HHO). After the distance of 3.5 Å the charge on individual water molecules becomes close to zero.

Figure 9-5 reveals an asymmetry in the ability of water molecules to hold excess charge of particular sign with the apparent preference to acquire excess negative charge. This indicates that greater solute-to-solvent charge transfer might be expected for highly negatively charged biomolecules, e.g. DNA and RNA. No such charge asymmetry is assumed in the design of empirical water models, i.e., the instantaneous charge fluctuations on water molecule averaged due to mean field approximation are expected to be zero. Moreover, since empirical water models do not take into account intermolecular charge transfer their radial charge distribution function will be a flat horizontal line at value zero rather than the shape shown in Figure 9-5.

Having obtained a non-classical water radial charge distribution function, next obvious step is to integrate this function in order to obtain a picture of cumulative excess charge transferred from protein to water as a function of distance. Such integral plot is displayed in Figure 9-6. It basically goes over the main conclusions derived from Figure 9-5. The nearest water shell of thickness of 2.2 Å carries total net charge of +0.04 electron units. The net positive charge quickly becomes

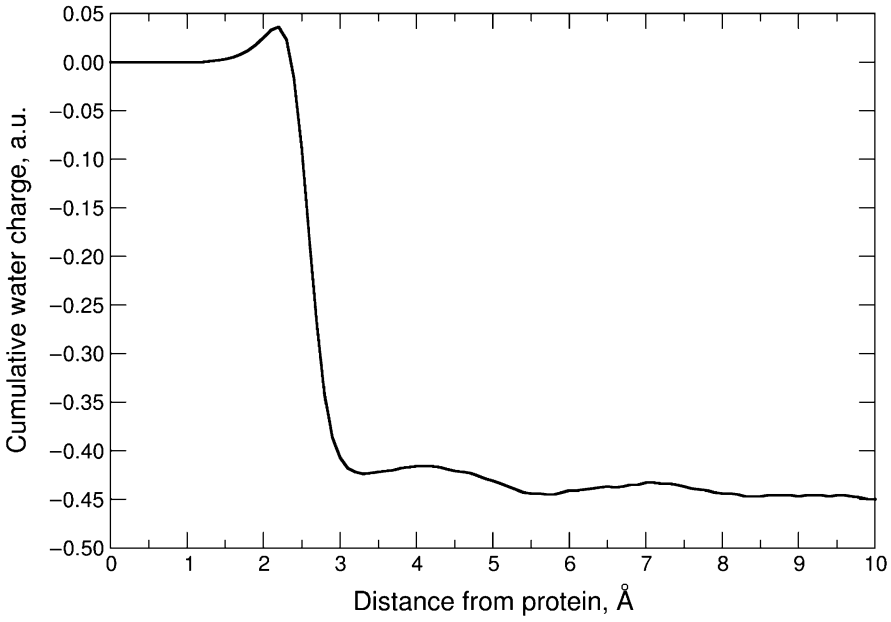


Figure 9-6. Integral plot of cumulative charge on water molecules as a function of distance to protein, averaged over PM5 MD trajectory

outweighed by the excess negative charge and the water shell of thickness of 3.3 Å already encloses 95% of the entire electron charge transferred from the protein to water. The second minimum at 5.7 Å encloses almost 100% of the excess charge with the value reaching plateau at 8 Å.

Somewhat more diffuse excess charge localization as a function of thickness of water shell was suggested by HF/6-31G* computations of ubiquitin–water system by Komeiji et al. [23]. According to their studies the protein-to-water charge transfer value converges at about 6 Å water shell whereas in our case it converged at 3.3 Å. Although the superiority of the ab initio level over semiempirical one is unquestionable it is possible that the ab initio data were biased toward more diffused picture because of geometric strain in the studied structures. Unlike our PM5 MD simulation the HF/6-31G* computations were performed on the structures which do not correspond to the corresponding QM energy minimum. The energy strain present in the molecular structure can artificially increase the appearance of charge transfer. To test this assumption we performed single-point (rigid-structure) PM5 calculations over 1,000 ubiquitin–water snapshots generated by force field MD and got a prediction of −1.6 electron units for protein-to-water charge transfer. It clearly showed that performing QM calculations over not self-consistent geometries leads to overestimated charge transfer predictions.

In cases when QM MD is not available the accuracy in handling charge transfer can be significantly improved by performing geometry relaxation at the corresponding QM level. Performing PM5 geometry optimization on 100 empirical snapshots changed the prediction from −1.6 to −0.6 electron units [21], which is in excellent agreement with our QM MD data and the data of Komeiji. Therefore taking into account unavailability of ab initio MD data on charge transfer for ubiquitin–water system and based on the variety of indirect evidences discussed above there are reasons to believe that our PM5 MD predictions on charge transfer in ubiquitin– water system are the most accurate to date.

9.4. IMPLICATIONS OF CHARGE TRANSFER

In the conclusion, the undertaken QM MD simulation revealed complex physics of dynamic charge transfer in the ubiquitin–water system. We observed a distinctly localized charge transfer channel pumping electron density between interacting amino acids in sync with their thermal motion. Delocalized effects of charge transfer are responsible for global instantaneous charge fluctuations exceeding 0.25 electron units and affecting every amino acid in the system including neutrally charged and weakly polarizable ones. Same-type amino acids which are treated as equal by force fields show different average net charges in QM picture due to their interaction with environment. It means that the assignment of the same unit charge to same-type amino acids is not actually correct. As it follows from our QM MD data water molecules of the solvent are also not equal. Those water molecules located in close proximity to the protein carry instantaneous excess charge of up to

0.2 electron units per water molecule resulting in accumulation of a charge cloud around the protein.

This atomic-level picture is largely overlooked by classical mechanics treatment of electrostatic interactions. Due to significant charge transfer effects taking place at every part of the protein–water system and because of the link between charge transfer and protein dynamics the biomolecular systems can no longer be considered as merely classical mechanical entities unless at the cost of omission of important atomistic details. Although empirical force fields dealing with fixed atomic (or amino acid) charges are computationally attractive, the oversimplified electrostatic model is the major factor limiting accuracy of computer simulation of biological systems. Therefore QM MD represents an important physics-based approach for improvement of accuracy of biomolecular simulations.

ACKNOWLEDGEMENTS

The Authors are grateful to the Texas Advanced Computing Center (TACC) (http://www.tacc.utexas.edu) for providing high-performance computing resources for this project. Support from a R.A. Welch Foundation Chemistry and Biology Collaborative grant from the John S. Dunn Gulf Coast Consortium for Chemical Genomics is greatly acknowledged.

REFERENCES

1. Goedecker S (1999) Rev Mod Phys 71:1085–1123
2. Goedecker S, Scuseria GE (2003) Comp Sci Eng 5:14–21
3. McCammon JA, Gelin BR, Karplus M (1977) Nature 267:585–590
4. Warshel A, Levitt M (1976) J Mol Biol 103:227–249
5. Anikin NA, Anisimov VM, Bugaenko VL et al (2004) J Chem Phys 121:1266–1270
6. Anisimov VM, Bugaenko VL (2009) J Comp Chem 30:784–798
7. Halgren TA, Damm W (2001) Curr Opin Struct Biol 11:236–242
8. Sprik M, Klein ML (1988) J Chem Phys 89:7556–7560
9. Caldwell J, Dang LX, Kollman PA (1990) J Am Chem Soc 112:9144–9147
10. Rick SW, Stuart SJ, Berne BJ (1994) J Chem Phys 101:6141
11. Lopes PEM, Roux B, MacKerell AD (2009) Theor Chem Acc 124:11–28
12. Anisimov VM, Vorobyov IV, Roux B et al (2007) J Chem Theory Comput 3:1927–1946
13. Nadig G, Van Zant LC, Dixon SL et al (1998) J Am Chem Soc 120:5593–5594
14. Thompson WH, Hynes JT (2000) J Am Chem Soc 122:6278–6286
15. Tanaka M, Aida M (2004) J Sol Chem 33:887–901
16. Ikeda T, Hirata M, Kimura T (2005) J Chem Phys 122:024510
17. Cappa CD, Smith JD, Wilson KR et al (2005) J Phys Chem B 109:7046–7052
18. Voityuk AA, Siriwong K, Rösch N (2004) Angew Chem Int Ed 43:624–627
19. Kubar T, Kleinekathofer U, Elstner M (2009) J Phys Chem B 113:13107–13117
20. Peraro MD, Raugei S, Carloni P et al (2005) Chemphyschem 6:1715–1718
21. Anisimov VM, Bugaenko VL, Cavasotto CN (2009) Chemphyschem 10:3194–3196
22. Balabin IA, Beratan DN, Skourtis SS (2008) Phys Rev Lett 101:158102
23. Komeiji Y, Ishida T, Fedorov DG et al (2007) J Comp Chem 28:1750–1762

24. Dewar MJS, Zoebisch EG, Healy EF et al (1985) J Am Chem Soc 107:3902–3909
25. Rocha GB, Freire RO, Simas AM et al (2006) J Comp Chem 27:1101–1111
26. Stewart JJP (1989) Method J Comp Chem 10:209–220
27. Stewart JJP (2002) Mopac 2002. Fujitsu Ltd, Tokyo, Japan
28. Hoover WG (1985) Phys Rev A 31:1695–1697
29. Ferrario M, Fionino A, Ciccotti G (1997) Physica A 240:268–276
30. van der Vaart A, Merz KM (2002) J Chem Phys 116:7380–7388
31. Cavalli A, Carloni P, Recanatini M (2006) Chem Rev 106:3497–3519
32. Geissler PL, Dellago C, Chandler D et al (2001) Science 291:2121–2124
33. Patel S, Mackerell AD Jr, Brooks CL III (2004) J Comp Chem 25:1504–1514
34. van Duin ACT, Dasgupta S, Lorant F et al (2001) J Phys Chem A 105:9396–9409
35. Liu H, Elstner M, Kaxiras E et al (2001) Proteins 44:484–489
36. Tong L, Warren TC, King J et al (1996) J Mol Biol 256:601–610
37. Vorobyov IV, Anisimov VM, MacKerell ADJ (2005) J Phys Chem B 109:18988–18999
38. Giese TJ, York DM (2004) J Chem Phys 120:9903–9906

CHAPTER 10

BEYOND STANDARD QUANTUM CHEMICAL SEMI-CLASSIC APPROACHES: TOWARDS A QUANTUM THEORY OF ENZYME CATALYSIS

ORLANDO TAPIA
Department of Physical Chemistry and Analytical Chemistry, Uppsala University, Ångström Building, Box 259, 75105 Uppsala, Sweden

Abstract: The role transition structures (TSs) and vectors have played in discussing issues associated to enzyme catalysis is examined with focus on RubisCO; computations belong to standard semi-classic Born-Oppenheimer model with one-electron orbitals located at nuclear position coordinates. Here, theory is brought a step beyond starting from exact quantum schemes to get types of semi-classic Hamiltonians allowing for clear definitions of electronuclear separable models. Electronuclear quantum states (QSs) are given by linear superpositions over eigenfunctions of semi-classic Hamiltonians; base state functions are products of space and spin components. For frozen nuclei models amplitudes depend on nuclear configuration multiplying electronic diabatic base functions characterized by nodal pattern distributions. QS' time evolution requires couplings to external fields and does not necessarily conserve total spin. RubisCO's dioxygen and ethene-fragment quantum reactivity are examined and generalized.

Conservation of nodal patterns permits links to exact semi-classic schemes ensuring correct presentation of bond forming processes; possible failures of LCAO-based computations without nodal control are discussed. These issues are examined in relation to catalysis representation. Pauling's idea that TS signals bent out of equilibrium shape of reactant and product is framed in terms of transition QSs. A full quantum catalysis model is introduced.

Keywords: Abstract quantum mechanics, Electromagnetic fields, Quantum catalysts, H + H system, RubisCO, Quantum transition states, Photorespiration, Angular momentum, Diabatic states, Quantum base states: dioxygen, Ethylene, Carbene

P. Paneth and A. Dybala-Defratyka (eds.), *Kinetics and Dynamics*,
Challenges and Advances in Computational Chemistry and Physics 12,
DOI 10.1007/978-90-481-3034-4_10,

10.1. INTRODUCTION

Electronic studies of catalytic mechanism rely on semi-classic quantum chemical models, where electronic and nuclear degrees of freedom are separated and potential energy (hyper)-surfaces (PESs) play paramount roles. Classical physics aspects enter once electronic quantum mechanical elements are averaged out with fixed nuclear geometry arrangements in real space and atomic orbital basis planted on, thereby PESs and concepts such as reaction coordinates and structure in three-dimensional space become well defined in real space. But thus far, more basic descriptions using quantum state concept represented as linear superpositions over electronuclear base states [1–5] occupy a marginal place. Yet, quantum states are those related to chemical change and sensed via spectroscopic measurement wherefrom most experimental data obtain; a host of chemical processes are catalyzed by light, more generally by electromagnetic (EM) fields, including thermal baths. Thus, model development relying on full-fledged quantum mechanical (QM) presentation seems timely. Yet, there is a key question: Why should we change our semi-classic approach including BO-PESs to replace it by a difficult and abstract QM view? Answer: it is not desirable to replace one by the other; a target is to exploit the quantum picture to get a deeper understanding of the chemical change in general and possibly new vistas on catalysis while standard computation schemes conveniently rigged would continue yield numerical results.

So far we have not realized that standard models may produce results contradicting experimental experience in unexpected ways. One example discussed below concerns one of the more fundamental reactions in the Universe: H + H. Thus, H_2 dissociation into two entangled hydrogen atoms can be considered one of the triumph of computational quantum chemistry [6]; and rightly so. A problem is found in the association reaction. While, the PES for H_2 used in the direction opposite to dissociation, i.e. scattering s-wave setup, predicts a no-barrier process, experimentally, the reaction must be catalyzed by a third body to get a signal related to H_2 (cf. Section 10.5.1 for details and references). One may wonder whether many of free radical reactions mechanisms may be faltering due to unsuspected quantum mechanical factors.

To make progress one thing actually needed is to dig out a better understanding of semi-classic models; this is not a problem of level of theory used to calculate electronic wave functions but of getting to know its connections with abstract Quantum Mechanics; see [4, 5]. Furthermore, because the Hamiltonian we use in quantum chemical computations is time independent no quantum process changing the amplitudes of a linear superposition is allowed unless couplings to external fields are let in. These issues are examined in this contribution. At this stage it is worth mentioning that what comes out is a conceptual framework and not computing algorithms; the latter aspect is discussed in connection with semi-classic schemes, relations with configuration interaction (CI) are established.

Chapter 10 is organized as follows. Section 10.2 includes discussions on enzyme catalyzed reactions with ribulose-1,5-bisphosphate carboxylase/oxygenase (RubisCO) as key example; see also [7, 8]. In Section 10.3, the terms required in an exact quantum theory including electromagnetic (EM) fields are examined. Section 10.4 presents semi-classic schemes and beyond; the algorithm representing atomic orbitals on top of nuclei is formalized in a general manner. Section 10.5 continues with a discussion of quantum aspects of catalysis; using H + H case a model quantum treatment of catalysis is presented; quantum transition states are introduced. In Section 10.6 spin-space separation and chemical reactivity analyses are examined. Ethylene and Carbene fragments are specially discussed. Section 10.7 addresses to activation of dioxygen; the mechanisms shed new light on photorespiration step in Rubisco. Section 10.8 shed light on radical reaction processes as one applies the line of thinking presented here. The QM scheme must include electronic and nuclear degrees of freedom, the connection to semi-classic scheme are examined to make clear differences with standard CI methods; yet some similarities survive.

10.2. ENZYME CATALYZED REACTIONS

The quality and efficiency of a biologic catalyst cannot be measured by kinetic parameters only. RubisCO (EC 4.1.1.39) catalyzes incorporation of inorganic carbon dioxide (CO_2) into the organic molecules of life. Its importance is difficult to exaggerate [9]. Yet there is consensus that this enzyme is extremely inefficient in carboxylase activity let alone the fact that it is compromised by numerous side reactions, see Figure10-1; most notable is oxygenation activity occurring at the same point where carbon dioxide fixes. An issue for theoretical chemists and people working in computer simulations is to understand why such is the case and whether partial modulation of one activity instead of another can be made possible. In so doing one may progress along a better description of catalysis with theoretical and computational tools.

The substrate, ribulose-1,5-bisphosphate (RuBP) in water, is barely (if at all) reacting in presence of CO_2 or dioxygen $O_2(^3\Sigma_g^-)$; the need for a catalyst is apparent. Linus Pauling before any computing could be done (1948) suggested that enzymes are molecules complementary in structure to the activated complexes of the reaction they catalyze; in other words, they mold (bent) reactants to get the geometry arrangement of the activated complex. By 1988 experience gathered for liver alcohol dehydrogenase (LADH) where the transition structure calculated for the hydride transfer step [7] was just in agreement with Pauling's idea. Carbonic anhydrase also show to be a case [8]; a host of examples followed reported by many groups in the world and the saddle point of index 1 started to make sense as transition state structure even if calculated with models in vacuum.

Pauling's hypothesis brings to the fore active site and molding mechanisms; reactants are forced, as it were, to adopt geometry configurations far away

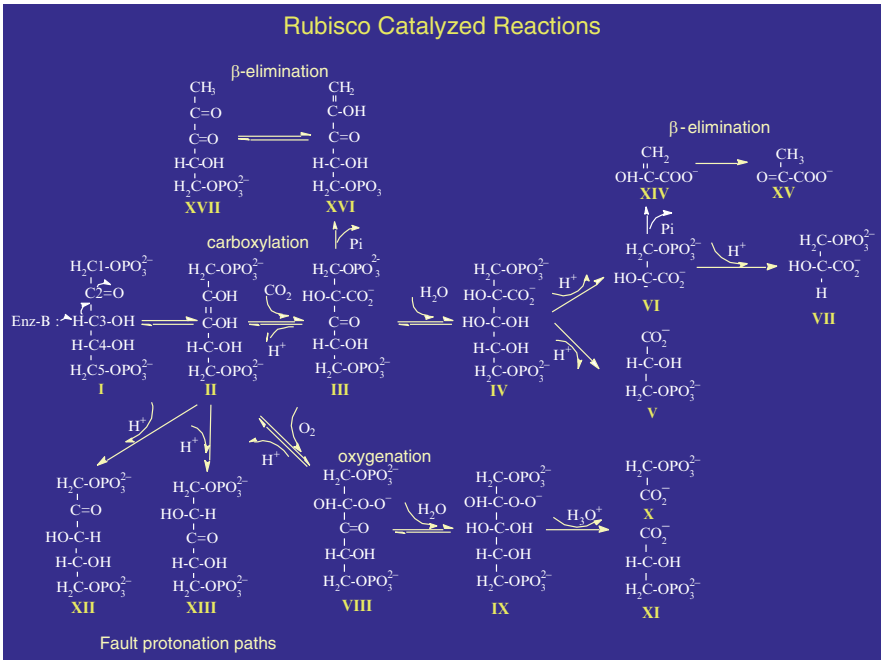


Figure 10-1. Schematic description of RubisCO catalyzed reactions

equilibrium ones. Intrinsic deformation work can be estimated in absence of the enzyme as a first step. Vibration activation permits (large) structure fluctuations, a binding site anchoring the substrate bent into a complementary shape; binding energy provides work source or sink. The reaction direction is modulated by differential binding for reactant and product species. Binding energy can be estimated with computer simulations as well. That much qualitative insight one gets from Pauling's seminal idea. Enzymes whose catalytic paths can be described in this manner belong to Pauling's class. It goes without saying that not all known enzymes belong to this class.

RubisCO presents an ordered catalytic mechanism. Once the enzyme is activated, substrate (RuBP) binds and enolization is the first step followed by either carboxylation or oxygenation steps [9]; in Figure 10-1 we summarize reactions RuBP can undergone that are catalyzed by RubisCO. The enzyme depends on effector molecules to modulate its activity and on ancillary proteins to control its activation state and mediate correct folding and last but not least assembly in the cell; the so-called chaperone protein play key roles [9]; all these subjected to control involving light.

A key activation factor is fixation of one carbon dioxide to carbamylate Lysine 201 (spinach numbering) empowering the residue to contribute to magnesium ion binding and complete the catalytic site.

10.2.1. Transition Structures and Chemical Mechanisms

The enzyme RubisCO acts as a receptor to reactants. Thus, activation and catalytic factors can be separated from chemical events undergone by RuBP under attack from CO_2, dioxygen or $O_2(^3\Sigma_g^-)$ and water. This special situation was used in our quantum chemical studies of saddle point structures of index-1 calculated for a host of mechanisms involving carboxylation, oxygenation, hydration and inhibition pathways [10, 11]; (incidentally, a saddle point results from finding stationary values on an energy hypersurface, showing negative eigenvalues once the Hessian is transformed into a diagonal form, index-1 signals presence of only one negative eigenvalue, the corresponding geometry defines a transition state (TS) structure; the Hessian is a matrix representing the second derivative of energy change with respect to geometry changes, in a mass weighted coordinates model one gets a force constant matrix). Models with 3- and 5-carbon atoms backbone reflected similar behavior. Invariance of TS structures to different levels of theory and models was found. The models in vacuum reflected all chemistry the enzyme was known to catalyze. While the statement indicates possibility for the enzyme to do so, it does not detract import acid-base groups in contributing to catalysis. This fit theoretical exploring where possibilities and not only actual events are sought.

The key TS appears in enolization step. RuBP chemical formula including carbon numbers is: $PO_4^{(-2)}$–CH_2(1)–CO(2)–CHOH(3)–CHOH(4)–CH2(5)–$PO_4^{(-2)}$ carbon center C(2) sustains a carbonyl group while C(3) put up with an alcohol function. The TS correlates with dienol-like form: –COH(2)=COH(3)–; this opens the field for a host of mechanistic possibilities; parentage to ethylene is clear.

There are two spin electronic structures (a singlet and a triplet) that can be seen label by the equivalent of a disrotatory angle signaling displacement of the hydroxyl groups COH(2)–COH(3) with respect to a fixed plane; at $\pi/2$ angle of the OH frames the triplet spin state shows energy lower than spin singlet. The first paper on this system [10] observed that a triplet state would open a pathway to interact with another triplet such as dioxygen. A clue to solve the mystery of oxygenation is found in the first activation step concerning enolization. With relatively simple basis sets carboxylation reaction was explored. It took the labors of many collaborators before a fairly complete set of TSs was actually calculated describing inhibitory paths.

In fact, enolization-TS conceals a key to describe inhibitory pathways: Xylulose, Ribulose, 3-keto-Ribitol and 3-keto-Arabinitol were found at the end of a series of new TSs described in reference [11]; cf. Figure 10-1.

Which role does the enzyme has if all chemistry is written down in the TSs? A first role is evident: TSs are not equilibrium structures so that substrates must be bent out of (equilibrium) shape and hold out by docking at the active site. In the same manner LADH's TS could be docked with the help of the X-ray structure from a slow substrate, for RubisCO we had a high resolution X-ray structure with inhibitor CABP (2′-carboxyarabinitol-1,5-bisphosphate). The docking showed TSs could be overlay to the coordinates of CABP. Pauling's lemma works here!

At variance with standard studies focusing on energy variables we stick to the structural ones using three- and five-carbon models. To our dismay we found that the TS leading to the dienol structure in 5-C model where only the phosphate groups are absent lost to a great extent the local out-of-plane configuration in the –COH (2)–COH(3)– fragment. This factor modulates the singlet-triplet energy gap and a cis-conformer lost access to spin triplet.

Inclusion of carbamylated lysine fragment in Magnesium coordination shell led to an astonishing result: the carbamylated group induced an out-of-plane configuration; so that the new set of TS coordinates could be overlay with CABP in a totally satisfactory manner.

Transition state structure invariance to model system size and computing level was studied with QM/MM and ab initio MO (molecular orbitals) methods for the carboxylation step catalyzed by RubisCO [12]. The TS calculated within the model enzyme and also in vacuum show a high degree of invariance. But not only this, the TS couples carbon dioxide attack to the C(2) center of the dienol form of the substrate to a simultaneous hydroxyl hydrogen transfer from –CHOH(3)–; the product corresponds to a carboxylic acid group and carbonyl function at C(3). Hydration at C(3) leads to a geminal diol (gemdiol) structure; see structure IV in Figure10-1. The break of C(2)–C(3) bond is the step producing complexes in the product pathway via a back hydride transfer to C(2) [13, 14]. A first 3PGA (3-Phospho glycerate) can be released and a stereospecific protonation of the second fragment leads to another 3PGA molecule; TSs were all determined with ab initio schemes of varied quality.

For RubisCO the electronic structural description of catalytic steps was more than satisfactory. Yet, no answer to the invariance towards model and level of theory in the Hartree-Fock scheme was found. The need for a more fundamental description became more and more clear. The mechanical description of processes such as bond breaking and knitting was not sufficient. Computer algorithms veiled quantum mechanical aspects. Digging on more basic theory became necessary.

10.3. EXACT QUANTUM SCHEMES

Quantum mechanics is about quantum states. To link abstract to laboratory levels one introduces an abstract configuration space via an inertial frame (I-frame) whose origin lies somewhere at a laboratory space with respect to another I-frame. The degrees of freedom required to describe the quantum system define the dimension of the configuration space. The space part of a system sustained by n-electrons and m-nuclei requires real coordinate vectors of dimension 3n + 3m: ($\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{x}_{n+1}, \ldots, \mathbf{x}_{n+m}$). The coordinates serve as label to configuration basis vectors constituting a rigged Hilbert space: $|\mathbf{x}_1, \ldots, \mathbf{x}_n, \mathbf{x}_{n+1}, \ldots, \mathbf{x}_{n+m}>$. To simplify notation one may write: $|\mathbf{x}_e > |\mathbf{x}_N>$ for separable EN space or $|\mathbf{x}_e,\mathbf{x}_N>$ to signal the exact space. The symbols inside the kets (bras) in abstract QM do not stand for position vectors, they are just numeric labels that may take a meaning if we use position vector operator; e.g. $\mathbf{x}_{op.e}(k)\ |\mathbf{x}_e > |\mathbf{x}_N> = \mathbf{x}_e(k)\ |\mathbf{x}_e > |\mathbf{x}_N > = \mathbf{x}_{ek}\ |\mathbf{x}_e > |\mathbf{x}_N>$, $k \leq n$; in

other words one picks up the k-th label that operator $\mathbf{x}_{op}(k)$ indicates and transform it into a real value vector. An abstract Hamiltonian operator H_{op} projected in the EN-space reads as a bracket: $<\mathbf{x}_e,\mathbf{x}_N|H_{op}(\mathbf{x}_{op,e},\mathbf{x}_{op,N})|\mathbf{x}_e,\mathbf{x}_N> = H_{op}(\mathbf{x}_e,\mathbf{x}_N)$. This is a configuration space defined operator.

An abstract quantum state $|\Psi>$ projected into the configuration space is a scalar product $<\mathbf{x}_e,\mathbf{x}_N\ |\Psi>$, namely, a complex number. It is customary to take the assignment $<\mathbf{x}_e,\mathbf{x}_N\ |\Psi> \rightarrow \Psi(\mathbf{x}_e,\mathbf{x}_N)$ as a wave function mapping in coordinate space an abstract quantum state space (Hilbert space) [5].

The configuration space is just a mathematical space where a complex function is defined; this is the reason to name it abstract.

For a time independent Hamiltonian, separating time from space coordinates via the constant E leads to the famous Schrödinger equation:

$$H_{op}\left(x_{op,e},x_{op,N}\right)\Psi(x_e,x_N) = E\Psi(x_e,x_N) \tag{10-1}$$

For a self-adjoint H_{op} and standard boundary conditions on the wave function the system show a complete set of eigenfunctions and eigenvalues: $\Phi_{i,j(i)}(\mathbf{x}_e,\mathbf{x}_N)$ and energy eigenvalues $E_{i,j(i)}$; the first label is electronic the second is nuclear label subsidiary to the electronic quantum number, information gleaned from molecular spectra.

Spin space basis states are constructed separately. Eigen functions of total ${S_{op}}^2$ and its third component $S_{op,3}$ for the global electronic system are determined in a manner independent from the basis function for the space part. For the electronic labels, the direct product of the space and spin basis functions must be anti symmetric to label permutations.

Energy eigenvalue differences are associated to spectroscopic responses. An arbitrary quantum state is represented by a linear superposition over the eigenstates for bound states:

$$\Psi(x_e,x_N) = \sum\nolimits_{i,j} C_{ij}(\Psi) \quad \Phi_{ij}(x_e,x_N) \tag{10-2}$$

Implicit spin dependence is implied albeit not shown here. The quantum state corresponds to the set of i-ordered complex numbers $C_{ij}(\Psi)$. For time dependent processes, the basis set is invariant, only the amplitudes change in time: $C_{ij}(\Psi,t)$.

The Hamiltonian with the electronic and nuclear kinetic energy operators are given as $\mathbf{K}_{op,e}$, $\mathbf{K}_{op,N}$ and a potential energy operator V_{op} has the form:

$$H_{op}\left(x_{op,e},x_{op,N}\right) = K_{op,e} + K_{op,N} + V_{op}\left(x_{op,e},x_{op,N}\right) \tag{10-3}$$

For charged systems the four-potential of the electromagnetic (EM) field $A = (A_o, A_1, A_2, A_3) = (A_o, \mathbf{A})$ permits describing the interaction with charge e. The quantity $(-e/c)\mathbf{A}$ corresponds to a momentum in the EM field so that in classical physics the total linear momentum reads as $(\mathbf{p}\text{-}(e/c)\mathbf{A})$. Schematically, the momentum operator reads as $(\mathbf{p}_{op} - (e/c)\mathbf{A}_{op})$ and in quantum electrodynamics one has to provide the quantum state of the EM field; basis vectors $|n_\omega,\mathbf{k}_\omega>$ are label by the amount of energy at disposal (photon number n_ω, frequency ("color") ω or momentum $\mathbf{k}_\omega$; a vacuum base state indicates that no photon available for exchange:

$|n_\omega = 0, \mathbf{k}_\omega>$. Direct products of different colors and photon numbers form a Fock space. The Hamiltonian for the material system in an EM field has the generic form for a particle of mass M and charge e in a Coulomb field $A_o = \phi$:

$$\frac{1}{2M}\left(p_{op} - \left(\frac{e}{c}\right)A_{op}\right)^2 + e\phi_{op} = \frac{1}{2M}\left(-i\hbar\nabla - \left(\frac{e}{c}\right)A_{op}\right)^2 + e\phi_{op} \qquad (10\text{-}4)$$

The Hamiltonian contains $e\phi$ the standard Coulomb term we use in computations of atomic and molecular systems; a coupling term $(e/c)\mathbf{A}_{op}$. $\mathbf{p}_{op}$ is effective in changing the amplitudes of given quantum states in a time dependent framework. The term $(e/c)^2\mathbf{A}_{op}{}^2$ is effective in presence of high intensity (lasers) fields. By enclosing the system in a box of length L, the transverse EM vector operator $\mathbf{A}_{op}(\mathbf{q})$ is given by the sum of creation ($a_{op}{}^+$) and annihilation (a_{op}) operators:

$$A_{op}(q) = \sum\nolimits_j \left[\frac{\hbar}{2\omega_j L^3}\right]\left[a_{op,j}n_j \exp\left(ik_j \bullet q\right) + a_{op,j}{}^+ n_j \exp\left(-ik_j \bullet q\right)\right] \qquad (10\text{-}5)$$

The polarization vector $\mathbf{n}_j$ is orthogonal to the propagation vector $\mathbf{k}_j$; the EM field is transverse with the electric and magnetic field defining a plane orthogonal to propagation vector. Because we control the EM by giving non-zero amplitudes to those states able to exchange energy, these infinite sums reduce to few terms.

The kinetic energy operators $\mathbf{K}_{op,e}$ and $\mathbf{K}_{op,N}$ get their standard forms when M is replaced by the electron mass m_e and nuclear mass m_N, respectively. For a nucleus with atomic number Z_N the nuclear charge is positive: $|e|Z_N$. The sum of kinetic energy and Coulomb interaction define the Coulomb Hamiltonian: $H_{op}(\mathbf{x}_{op,e},\mathbf{x}_{op,N})$.

The inclusion of quantized EM fields, that is the natural thing to do for realistic process descriptions, provides us with mechanism to change the quantum state sustained by specific material systems and ensure energy conservation in a simple manner [5]. Because the number of material constituents is fix the only thing that can change is their quantum state related to change of EM quantum state.

10.4. SEMI-CLASSIC SCHEMES AND BEYOND

We speak of semi-classic approach whenever the nuclear part is separated from the electronic one [5]. The positive charge background (PCB) set up Coulomb sources to the electronic element. This is a model; it is not a representation of material systems in the laboratory. The quality of the model is gauged by its ability to agree with experimental data. The statement functions in only one direction: i.e. agreement with experiments does not prove the model.

10.4.1. Semi-classic Hamiltonian Models

We distinguish five types of Hamiltonian starting from the most general one:

(a)$H_{op}(\mathbf{x}_{op,e},\mathbf{x}_{op,N})$; (b) $H_{op}(\mathbf{x}_{op,e},\mathbf{X})$; (c) $H_{op}(\mathbf{x}_{op,e};\mathbf{X})$; (d) $H_{op}(\mathbf{q};\mathbf{X})$; (e) $H_{op}(\mathbf{q}(\mathbf{X}),\mathbf{X})$ where cases (a)–(c) the electronic subspace is exact and the nuclear space

operator changes from the abstract ($\mathbf{x}_{op,N}$) to the operator projected in real space ($\mathbf{X}$). Case (c) is the frozen nuclei model but exact electronic part. The operator (d) stands for a fix nuclei model (no nuclear kinetic energy operator) and the electronic term is a particle model in real electronic configuration space ($\mathbf{q}$).

For Hamiltonian (e), namely, $H_{op}(\mathbf{q}(\mathbf{X}),\mathbf{X})$ the electronic configuration space is made to depend upon nuclear positions in real space. For those among you that are used to the molecular orbital (MO) algorithm in the LCAO model can see Hamiltonian (e) is just telling that model in a general fashion.

All operators above except (a) are semi-classic in one way or another. Finally, the frozen nuclei model is overcome to read as $H_{op}(\mathbf{q}(\mathbf{X}),\mathbf{X})$ so that:

$$H_{op}(q(X),X) = K_{op,N} + H_{op}(q(X);X) \tag{10-6}$$

Chemical structural concepts are explicitly included within this Hamiltonian operator; note that in Eq. (10-3) classical variables are absent. The operator (10-6) defines a general Born-Oppenheimer (BO) framework [5]. For bounded states, because the Hamiltonian is self-adjoint one can get a denumerable set of basis functions $\Phi_{n(i)}(\mathbf{q};\mathbf{X}^{(i)})$ where $\mathbf{X}^{(i)}$ is a stationary nuclear configuration. For the ground state $\Phi_{n\,=\,0(i)}(\mathbf{q};\mathbf{X}^{(i)})$ set up an attractor for the nuclear configuration that is minimum energy at $\mathbf{X}^{(i)}$. The set of eigenfunctions and eigen energies is complete; the electronic functions can be characterized by the distribution of nodal planes. For two different attractors $\mathbf{X}^{(i)}$ and $\mathbf{X}^{(j)}$ their ground state basis functions read as $\Phi_{n\,=\,0\,(i)}(\mathbf{q};\mathbf{X}^{(i)})$ and $\Phi_{n\,=\,0(j)}(\mathbf{q};\mathbf{X}^{(j)})$ each attractor sustain a complete set of electronic basis functions. The two sets of eigenfunctions are the same but ordered in a different manner; the difference is imposed by the PCBs. A correlation between vibronic levels belonging to these attractors follows. Take $E_{n',\neq 0(i)}(\mathbf{X}^{(i)})$ to be the vibronic energy that (almost) corresponds to the energy an excited electronic state at the j-th attractor that shows the same electronic nodal pattern than the i-th attractor. This means that the nodal pattern of say $\Phi_{n',\neq 0(j)}(\mathbf{q};\mathbf{X}^{(j)})$ can be mapped to a particular attractor sustained by $\Phi_{n\,=\,0(i)}(\mathbf{q};\mathbf{X}^{(i)})$ because they share the same pattern; *mutatis mutandis* $\Phi_{n'',\neq 0(i)}(\mathbf{q};\mathbf{X}^{(i)}) \rightarrow \Phi_{n''\,=\,0(j)}(\mathbf{q};\mathbf{X}^{(j)})$. In one word, diabatic potential energy surfaces (D-PES) $E_{n''\,=\,0(i)}(\mathbf{X})$ and $E_{n'\,=\,0(j)}(\mathbf{X})$ cross each other [5].

The crossing has no special consequences for computing algorithms. The geometry domain where D-PES cross is designated as $\mathbf{X}^{(ij)}$. Note that Eq. (10-6) with the replacements $H_{op}(\mathbf{q}(\mathbf{X}^{(i)}),\mathbf{X})$ and $H_{op}(\mathbf{q}(\mathbf{X}^{(j)}),\mathbf{X})$ yield vibration spectrum for each attractor, $E_{i\,=\,0,k(i\,=\,0)}$ and $E_{j\,=\,0,k'(j\,=\,0)}$; k(j = 0) and k'(i = 0) take on integer values (harmonic oscillator) at each well. Bear in mind that at each DP the electronic spectra contains an energy eigenvalue that is also identifiable by the nodal pattern of its electronic eigenfunction; Thus, sooner or later, the vibration ladders will approach the energy level for the electronic excited state of interest. By hypothesis, at $\mathbf{X}^{(ij)}$ there might be electronuclear levels almost degenerate and belonging to either attractor; these levels would be taken to form quantum states that can bridge the attractor wells. The corresponding abstract basis states are designated as |i,k(i)> and |j,k(j)> with quantum numbers k(i) = 0_i,1_i,... and

$k(j) = 0_j, 1_j, \ldots$ One has to search after those k(i) and k(j) that fulfill the condition just discussed. Note that this is an analysis not a construction algorithm.

A model linear superposition the vibronic levels selected above would stand for: $\Sigma_{k(i)}\ C_{k(i)}(t)\ |i, k(i)> + \Sigma_{k(j)}\ C_{k(j)}(t)\ |j,k(j)>$ permits a first analysis of the chemical process that would bring the electronic state response of the system from attractor $\mathbf{X}^{(i)}$ to $\mathbf{X}^{(j)}$.

Prepare the *i*th attractor in its EN-ground state so that all amplitudes are zero except $C_{k(i=0)=0}(t=0) = 1$. The chemical conversion expresses itself as the state where all amplitudes are zero except $C_{k(j=0)=0}(t \neq 0) = 1$. How to get from one state to the other is discussed now.

It is apparent that a change of quantum state involves amplitudes over intermediate base states $|i,k(i)\neq 0>$ and $|\ j,k(j) \neq 0>$ that cover the energy domain associated to the zone $\mathbf{X}^{(ij)}$.

For the simplest model start with vibration activation and consider Fock base states $|n_\omega = 0>$ and $|n_\omega = 1>$ so that direct product base states $|i,k(i) = 0>|n_\omega = 1 >$ and $|i,k(i) \neq 0>\ |n_\omega = 0>$ may have the same energy (degenerate levels). The operator coupling these two base states leads to the time dependent quantum state:

$$|X^{(ij)},t > = \cos(\omega t)\ |i,k(i) = 0) > |n_\omega = 1 > + \sin(\omega t)\ |i,k(i) \neq 0 > |n_\omega = 0 > \tag{10-7a}$$

The quantum system beats between the states as a function of time when all remaining amplitudes are (almost) zero. This is an entangled state between the EM field and the material system. This is a mechanism to produce an effect of EM energy over the material system that would function even off resonance.

Consider an j-th attractor and assume that there are those base states $|j,k(j) \neq 0>|\ n_{\omega'} = 0>$ and $|j,k(j) = 0>\ |n_{\omega'}=1>$ that added to the previous ones generate a four-state base set. We want to move the system at a neighborhood of the diabatic crossing. Observe that we can find two frequencies ω and ω' such that the base states are almost degenerate: $|j,k(j) = 0>\ |n_{\omega'} = 1>$ and $|i,k(i) = 0>\ |n_\omega = 1>$ are quasi resonant. This condition determines the appropriate k(i) and k(j) in the superposition. Let us write a model TQS (transition quantum state):

$$|TQS;X^{(ij)},t> = C_{i,k} = 0(t)\ |i,k(i) = 0)>|n_\omega = 1> + C_{i,k\neq 0}(t)\ |i,k(i) \neq 0|n_\omega = 0> + C_{j,k} = 0(t)\ |j,k(j) = 0)>|n_{\omega'} = 1 + C_{j,k\neq 0}(t)\ |j,k(j) \neq 0>|n_{\omega'} = 0> + \ldots \tag{10-7b}$$

We want to "move" amplitude from base state $|k(i) = 0) > |n_\omega = 0>$ that is ground state reactant to the one in the product channel $|j,k(j) = 0 > |n_{\omega'} = 0>$ and assume one can do it via processes involving states in the $\mathbf{X}^{(ij)}$ region (or above). In QM the possibility exists to put amplitude at the ground state $|j,k(j) = 0>$ with a photon of right frequency in the EM field, namely, with base states $|n_{\omega'} = 1>$.

Thus, as a matter of principle, a QM description of a chemical change can be implemented with the present approach. With the help of external sources activate

the states in an energy shell covering states found within $\mathbf{X}^{(ij)}$ region, this is necessary but not sufficient condition.

The QM scheme critically depends upon the transition amplitude $<i|T_{op}|f>$ that relates initial |i> to final |f> base states. Here, $T_{op} = (e/c)\mathbf{A}_{op}$. $\mathbf{p}_{op}$ in generic terms may do the job. The linear momentum operator is a sum of electronic $\mathbf{p}_{op,e}$ and nuclear $\mathbf{p}_{op,N}$ terms. Only the electronic term contributes and because the diabatic electronic basis function $\Phi_{n(i)}(\mathbf{q};\mathbf{X}^{(i)})$ does not depend upon nuclear geometry we write them simply as: $\Phi_{n(i)}(\mathbf{q})$; one obtains

$$<\Phi_{n(i)}(q)|p_{op,e}|\Phi_{n(j)}(q)><k(i) \neq 0\ |k(j) \neq 0> \tag{10-8}$$

The second term is the Franck-Condon overlap factor. The first term depends upon the parity of the diabatic electronic states involved. For electronic function with equal parity the integral is zero. The mechanism underlying a chemical process shows similarities to spectroscopic phenomena.

Dipole magnetic transition is another source of coupling opening some reaction channels. See below, Section 10.7.

10.4.2. Invariant Electronic Configuration Space Models

Theoretical models such as the one discussed by Born and Huang [15] are based on the Hamiltonian type $H_{op}(\mathbf{q};\mathbf{X})$. In principle, the electronic configuration space should be defined once and for all from a fix inertial frame.

The generalized electronic diabatic scheme was presented in this framework in our previous publications [2, 3, 16]. Zones $\mathbf{X}^{(ij)}$ can be identified as related to energy funnels as well as quantum transition state domains. It is at this level that transition structures play a key role helping identify crossing-like domains.

From a chemical viewpoint what is missing concerns asymptotic fragment-like states. For bimolecular and termolecular processes the reactants separately prepared are brought into regions equivalent to our $\mathbf{X}^{(ij)}$. But they are not solutions of $H_{op}(\mathbf{q};\mathbf{X})$ [5]. For each fragment there is one I-frame with a given assignment of matter, charge and angular momentum so that the sum total of fragments match the 1-system material system. We need to include fragments spectra (many I-frames) into the one-I-frame spectrum to complete a chemical basis set; this point is critical, one should be able to detect spectral features characterizing fragments (reactants) as such.

Each fragment presents the solutions to its own Schrödinger Eq. (10-1) while the sum of fragment is the sum of fragments Hamiltonians. There are no interactions at the stage of basis set construction.

To put together fragments and 1-system assume that a nodal hyper surface separates them; relate to one ordered configuration coordinate set defined from a unique origin. Because $H_{op}(\mathbf{q};\mathbf{X})$ contains all interactions, expectation values with fragment base sets will show up fragment-fragment interactions (intermolecular forces). All sorts of charge transfer state interactions appear as we

include all possible portioning (partitioning). In practice, only a few are put in computations.

The key qualitative idea is the inclusion of a sort of anti-bonding interaction between centers belonging to different fragments in so far spatial part is concerned. This is also valid for radical atom centers in a given molecule. This is a new idea examined below.

10.4.3. "Mobile" Electronic-Configuration-Space: Nodal Envelope States

How do we explain that solutions to the operator $H_{op}(\mathbf{q}(\mathbf{X}),\mathbf{X})$ where atomic functions (orbitals) follow the displacement of the PCB actually produce fragments and chemical change by numeric computing?

To discuss this issue take the material system formed by two electrons and two protons: $H + H$, H_2, $H^+ + H^{-1}$ or $2e + 2H^+$. Let us focus on the first two members. The algorithm itself $\mathbf{q}(\mathbf{X})$ "dissociates" into disjoint $\mathbf{q}(\mathbf{X}_1)$ and $\mathbf{q}(\boldsymbol{X}_2)$ as the proton distance increases. A laboratory experiment where each H-atom is locally prepared quantum descriptions naturally involve two I-frames, independent of each other. The Hamiltonian looks as a sum of two terms where no interactions are allowed for; $H_{op}(\mathbf{q}_1(\mathbf{X}_1),\mathbf{X}_1)$ and $H_{op}(\mathbf{q}_2(\boldsymbol{X}_2),\boldsymbol{X}_2)$. These operators can be used to generate fragment (asymptotic-like) base states fixed at their respective I-frame; relative velocity can easily be incorporated as classical magnitude between I-frames. In a laboratory we are free to fix I-frame origins at will as well as imparting classical linear momenta.

For our example we have two hydrogen atom systems with two I-frames. Focus the two beams into a head-on (s-wave) collision course. Including the system in a box (trapping potential, for instance) the quantum state is describable as two bosons states in the box. With this trick, the two I-frames fade away; the physical box provides one I-frame system. Note, however, the possibility to probe either a one I-frame state of two bosons that would correspond to a first step in a Bose-Einstein condensate or it may be an aggregate of two I-frame systems.

Because one wants to implement a computing quantum chemical scheme the two bosons are included in a unique I-frame and impose one condition over the base states: they must have nodal envelopes separating the composite bosons. With the semi-classic model we run into trouble because it is a classical Coulomb model only with respect to nuclei. We retain the general idea concerning two fragment systems each defined with an I-frame that, including the system referred to one I-frame, there must be now a repulsive state between them; a nodal envelope must separate the joint system.

The nodal-envelope constraint can be easily implemented within the independent particle model pervasively used in computational quantum chemistry.

Consider the two-atom toy model where one atomic orbital is planted at nuclei positions and move with them. The anti-bonding molecular orbital (MO) shows a node between nuclei and it will be an element to model a nodal-envelope.

The subtleties concerning permutation symmetry are not considered yet to help illustrate the qualitative aspects. The two hydrogen atoms are in a repulsive state, two electrons in anti-bonding MO as it were. The lowest attractive basis state of hydrogen molecule is always there with zero occupation (amplitude).

If we use the bonding MO it will remain bonding because that is the choice we impose: there always be a marginal attraction. The theoretical description of two H-atoms requires enforcement of nodal-envelope (NE) state between them. This NE-state is presented as $|H + H; NE> \rightarrow |H \perp H>$.

The repulsive interaction allowed for in the present model actually would forbid short-range approaching distances.

Let allow now van der Waals interactions; the base set contains direct products of bosonic systems. In laboratory set up, one may expect appearance of weak attractor states that can be responsible for stabilization when acting on top of external trap-potentials thereby leading at very low temperatures to condensation on the lowest box energy eigenvalues. As hydrogen quantum states conserve their bosonic nature it is natural to see appearing of Bose-Einstein condensates of hydrogen atoms ($F = 0, 1$) and not the solutions of the complete Coulomb Hamiltonian. Bear in mind, quantum physics is about quantum states sustained by material systems.

Now the experience just acquired permit to go back to discussions of computing algorithms when one comes to handle composites and component systems even if fermionic composites are present.

Hypothesis: The nodal envelope states characterizes separated composite sub-systems that can be brought in to correlate to one-I-frame full composite system in a semi-classic model.

Therefore, embedding a multi-composite system with the full 1-system requires use of NE-states to physically describe the asymptotic base states. These latter states are included with the base set of the full system. Thus, all chemical possibilities are taken into account [4, 5].

10.5. QUANTUM ASPECTS OF CATALYSIS

10.5.1. Model Quantum Catalyst: H + H and H_2

A model can be constructed. Let the two base states $|Cat_{i\,=\,0}; E_0>$ and $|Cat_{i\,=\,1}; E_1>$ including energy eigenvalue labels belong to a material catalyst; the base states $|H>|H>$ and $|H_2, v = 0>$ stand for the system of interest. Energy and momentum conservation forbid a direct change of amplitudes signaling the chemical reaction.

Select now a catalyst such that $E_1 - E_0$ equals dissociation energy from the vibration state $v = 0$ of the hydrogen molecule. Quantum states can be written as:

$$|\Phi_a,t> = C_1(t)\ |H>|H>|Cat_{i=0};\ E_0> + C_2(t)|H_2,v = 0>|Cat_{i=1};E_1> \qquad (10\text{-}9)$$

The state ($C_1 = 1\ C_2 = 0$) describes two hydrogen atom incident onto the catalyst in its ground state. A discharge couples the states and ($C_1 = 0\ C_2 = 1$) is

a possibility. This one tells us formation of hydrogen molecule in the ground vibration state and the catalyst in its excited electronic state. Complete description by including another base state to get:

$$\begin{aligned}|\Phi,t>_b &= C_1(t)\ |H>|H>|Cat_{i=0};\ E_0>|n_\omega = 0> + C_2(t)|H_2,v = 0>|Cat_{i=1};\ E_1>|n_\omega \\ &= 0> + C_3(t)|H_2,v = 0>|Cat_{i=0};\ E_0>|n_\omega = 1> \\ &= C_1(t)\ |1> + C_2(t)|2> + C_3(t)|3>\end{aligned} \tag{10-10}$$

The three base states belong to the same energy shell. If the catalyst radiates a photon in some direction the new base state would look like base state $|4> = |H_2,v = 0>\ |Cat_{i=0};\ E_0>|\mathbf{k}_\omega \neq 0>$; quantum states would change as:

$$\begin{aligned}&(C_1 = 1\ C_2 = 0\ C_3 = 0\ C_4 = 0) \rightarrow (C_1 = 0\ C_2 = 1\ C_3 = 0\ C_4 = 0) \\ &\rightarrow (C_1 = 0\ C_2 = 0\ C_3 = 1\ C_4 = 0) \rightarrow (C_1 = 0\ C_2 = 0\ C_3 = 0\ C_4 = 1)\end{aligned} \tag{10-11}$$

The last step would be irreversible were the continuum of the photon field directions be coupled to base state |3> leading to a non-zero amplitude over base state |4>. One would observe the catalyst glowing as light propagates to your eyes, namely $|\mathbf{k}_\omega \neq 0>$ [17]. Otherwise, the energy might be dissipated non-radiative processes via couplings to base states of the catalyst [18].

Inverting the process an incident photon with the frequency required in base state |4> $\mathbf{k}_\omega \neq 0$ if properly entangled one would be putting amplitude at base state |3>. Propagation leading to the reverse of Eq. (10-11) would end up into photo-catalytic dissociation of hydrogen molecule system.

Note that for quantum states $|\Phi,t>_b$ the vacuum states of the photon field must be present. This is a very simplified picture; entangled states ought to be considered as well. But, the formulation makes the point clearly.

In summary, two hydrogen atoms in s-wave arrangement (and zero relative kinetic energy do not yield hydrogen molecule as if it were an adiabatic process. Equation (10-10) underlies the type of basis states required to produce a chemically significant change of quantum state. A catalyst, among other things must break the symmetry of the diatomics (in this case). A third (quantum) body takes away energy that must be released thereby acting as (quantum) catalyst. These are among basic factors of catalysis in general.

Biomolecules acting as catalysts are usually massive. They present vibration and libration modes with high density of states; they can be considered a sort of aperiodic solids. Besides, there are plenty of chromophoric residues with spectral features that can interact with reactive subsystems.

The large mass permits absorbing high momentum exchange processes without increment of relative velocity. Analogously, exchange of energy required to pay for bond forming/breaking can easily be ensured by the biocatalyst.

The classical physics description used can be translated into quantum mechanics. The very simple example shown in the preceding section can be incorporated into a quantum description of enzyme catalyzed reactions involving hydrogen. But now we move on to examine more general formalism.

Let us introduce some concepts within the more abstract scheme of linear superpositions standing for quantum states.

10.5.2. Quantum Transition States

QTSs belong to subsets of electronuclear quantum states corresponding to few particular amplitudes $\{C_{n(i)}(\mathbf{X})\}$ for base states $\Phi_i(\mathbf{q})$ involved in coupling different reaction channels:

$$\Psi(q;X) = \sum_{i,n} C_{n(i)}(X)\ \Phi_i(q) \tag{10-12}$$

The generalized electronic diabatic function $\Phi_i(\mathbf{q})$ determines the stationary nuclear configuration $\mathbf{X}^{(i)}$ for the "supermolecule", namely a one I-frame system. Asymptotic base states are included in the list as discussed above to open amplitude transfers signaling dissociation processes. The construction of QTS requires knowledge of product end channels to help searching nuclear configuration domains ($\mathbf{X}^{(i,j)}$) where the diabatic states show overlapping vibronic states (D-PES crossings). Detailed discussions over some model systems were given above. Here, a short analysis is presented of the formalism necessary to set up computing procedures.

A 1-system showing a bonding between two centers as it is stretched increases energy, but they would never break because a basis state is identified by quantum numbers and not by regions of configuration space; but we can check that it can *correlate* with defined fragments; for the fragments antibonding case yield a repulsive diabatic PES that cross the bonding state just above dissociation energy. A common situation will be two-electrons case so that for reactant and product base states show same parity under inversion of the I-frame. Because the coupling operator for its electronic part has parity one, the transition amplitude is zero; when both states have different parity transition amplitude can be different from zero. In this model, the initial bonding quantum state is not changing at crossing although it must be activated to get at a crossing domain.

To get non-zero transition amplitude it is required inclusion of an electronic excited state with angular momentum L differing in one unit to the reactant states. As a matter of fact, one would consider the total angular momentum. Excited states would play the role of QTS in the diabatic basis set that includes both dissociated and bound states [16, 19].

In the diabatic basis set Hamiltonian Eq. (10-3) is diagonal thereby providing a first ingredient for a theoretically grounded computing framework:

$$H_{op}(x_{op,e},X)_{in(i),jn'(j)} = \delta_{ij}\ \delta_{nn'} H_{op}(x_{op,e},X)_{in(i),in'(i)} \tag{10-13}$$

Models of diabatic potential energy functions can be implemented to get chemical reactions simulators [2, 3, 19]. Consider two diabatic base states $\Phi_i(\mathbf{q})$ and $\Phi_j(\mathbf{q})$ with $\mathbf{X}^{(i,j)}$ being the crossing point between the D-PES:

$$\begin{aligned} E_i(X) &= <\Phi_i(q)|H_{op}(q;X)|\ \Phi_i(q)>_q\ \&\ \ E_j(X) \\ &= <\Phi_j(q)|H_{op}(q;X)\ |\Phi_j(q)>_q \end{aligned} \tag{10-14}$$

Take a domain around $\mathbf{X}^{(i,j)}$ so that the energy gaps between vibronic states are minimal or quasi-degenerate.

To set up the basis pick up the corresponding linear superposition with two components. The initial state is taken to be ($C_{n(i)}(\mathbf{X}^{(i,j)}) = 1\ C_{n'(j)}(\mathbf{X}^{(i,j)}) = 0$). The change to get ($C_{n(i)}(\mathbf{X}^{(i,j)}) = 0\ C_{n'(j)}(\mathbf{X}^{(i,j)}) = 1$) would mark a chemical reaction if the system could relax along the vibronic states of the j-D-PES. For this to be true, the transition amplitude $\mathbf{T}_{ij} = <\Phi_i(\mathbf{q})|\mathbf{T}_{op}|\ \Phi_j(\mathbf{q})> = <\Phi_i(\mathbf{q})|\mathbf{p}_{op}.\mathbf{A}(t)\Phi_j(\mathbf{q})>$ must be different from zero. In this case one gets a time dependent state that for the two-state model loks like:

$$\Psi(q;X^{(i,j)},t) = C_{n(i)}(X^{(i,j)},t)\ \Phi i(q) + C_{n'(j)}(X^{(i,j)},t)\ \Phi_j(q) \tag{10-15}$$

Let $t = t_0$ the system state represented by the row vector ($C_{n(i)}(\mathbf{X}^{(i,j)},t = t_0) = 1\ C_{n'(j)}(\mathbf{X}^{(i,j)},t = t_0) = 0$) then let at $t = t_0 + \Delta$ the vector become: ($C_{n(i)}(\mathbf{X}^{(i,j)},t = t_0 + \Delta) = 0\ C_{n'(j)}(\mathbf{X}^{(i,j)},t = t_0 + \Delta) = 1$). The system will keep beating at a specific frequency because it is entangled with the EM field.

Where is the transition state structure? We have to make an effort to see it. As we activate the reactant from $\mathbf{X}^{(i)}$ upwards to a neighborhood of $\mathbf{X}^{(i,j)}$ a force trying to bring the system down in a direction towards the attractor geometry (say to the left) sets up; use Hellmann-Feynman theorem with the D-PES to get it. As one approaches $\mathbf{X}^{(i,j)}$ the amplitude on the second channel starts increasing with a component pushing the system towards attractor $\mathbf{X}^{(j)}$ (say to the right). At the TS these two reciprocal forces compensate. The value of amplitudes and geometry determines something equivalent to a quantum transition state; a QTS is thence a linear superposition over excited vibronic states at a given energy shell. But we have something more.

If the electronic basis functions correspond to states both with equal parity then the transition amplitude is in first order zero; subtle second order effects may open couplings between such states. But, to first order, the linear superposition would stay put unless we find out other sources to couple in (i) and out (j) states; otherwise, only the relative phases change in time. Next electronic state having parity differing in one unit from the reactant and product states can open the transition integral in second order of perturbation theory. This latter, if it exists, may contribute with non-zero amplitude to form a quantum transition state coupling vibronic states belonging to reactant and product; in other words, a change of amplitudes can be detected. Of course such electronic state is already in the basis set because quantum mechanics implies all possibilities. We have omitted it in order to give a didactic view first. As a matter of fact the complete electronic spectra is there to our disposal including chromophoric residues that can come nearest the active site.

Remember the role we found the carbamylated lys-201 played in RubisCO. A first role was to procure a binding site for magnesium. This is a fundamental role, certainly. But we also saw the fine-tuning effect on the enolization step in our 5-C model [14]. This was a plain quantum effect elicited by increasing the quantum composite model.

The import for regulation of catalytic activity by enzyme domains can be a novelty in our way to think about catalytic sources.

Electromagnetic quantized fields play here a central role to construct reactive energy shells. Although at the end of a chemical process one can sense mechanical differences such as bond breaking/forming, the underlying level of understanding is quantum mechanics including light. One should remember now that in quantum mechanical terms a simple chemical picture is replaced by changes of quantum states amplitudes represented over basis states covering the complete spectra of both, the reacting subsystem and the protein.

10.5.3. Abstract BO Transition Structures

As a matter of computing fact, enzyme catalyzed reactions are represented with the help of the Hamiltonian $H_{op}(\mathbf{q}(\mathbf{X}),\mathbf{X})$. From present standpoint the proviso is that nodal envelope state holds. In standard packages the rule is to fill lowest energy MO first; this rule must be used very carefully. It is valid in searching stable minimum energy geometry. If the system can be related to two fragments at very large distances the rule works for dissociation but it should be discarded when looking in the direction of bond formation; this latter is not a classical mechanical event. In the present approach, the space electronic structure must show a nodal plane perpendicular to the vector connecting the fragments. Thus, start filling antibonding orbitals when appropriate. There is nothing strange with the procedure; bear in mind that one is constructing basis states and not mechanically representing electrons.

For computations carried out in a $\mathbf{q}(\mathbf{X})$-framework one continuously drive a nuclear parametrically dependent quantum state $\Psi(\mathbf{q};\mathbf{X})$ along energy constrained paths or reaction coordinate (RC), namely $\mathbf{q}(\mathbf{X}_{RC})$ and a variational procedure apply to get the lowest BO energy solution. But if solutions are sought for $\Psi(\mathbf{q};\mathbf{X})$ using the linear superposition Eq. (10-12) the result is the equivalent to the BO model with a caveat: operators coupling diverse diabatic states must be included to get amplitudes changes. This latter model is named as abstract BO model. The difference resides in the terms $C_{n(i)}(\mathbf{X})\ \Phi_i(\mathbf{q})$ that keep a trace to nodal patterns from the electronic base state while such information is lost for the standard adiabatic case. If in the standard $\mathbf{q}(\mathbf{X})$-framework one introduces the constraint consisting in keeping the internal NEs invariant then diabatic states can be elicited and crossings can be analyzed.

The TS-structures discussed in Section 10.2 were obtained with standard quantum chemical procedures using $\mathbf{q}(\mathbf{X})$- algorithms. A warning is now timely.

Some of the "spontaneous" changes" occurring during the calculation might well be artifacts due to a lack of nodal envelopes control.

10.5.3.1. One-Electron Basis Sets

The conceptual work described above concerns many-electron basis states. In practical cases, the construction of such functions is currently done within a one-electron basis set model.

The MOs are one-electron functions represented in LCAO or LCGO scheme. They result as solutions to a one-electron Hamiltonian; this latter obtains by defined procedures starting from the many-electron Hamiltonian. We need to examine the type of results one can get from the one-electron model.

Let C H O be a framework where one may sense hydrogen bonding. Let the initial state be label C–H⊥O, where a spin entangled 2-e state stands for a covalent bond (C–H) and ⊥ stands for a nodal plane perpendicular to the line C. . .O. The symbol (label) standing for product is C⊥H–O. In a path from R to P there is a difference of local NE. In very simple diabatic framework one has two basis states:

$$|\mathrm{C} - \mathrm{H}\perp\mathrm{O}> = |\mathrm{R}> \text{ and } |\mathrm{C}\perp\mathrm{H} - \mathrm{O}> = |\mathrm{P}>.$$

By calculating with a standard computing methods there is no control of NEs so that at the point where the NE includes the H-center locate a TS structure. For, as the RC is shifted from minimum $W^{(R)}$ towards $W^{(P)}$ hydrogen appear to be displaced from the minimum energy at reactant equilibrium configuration towards the "receptor" energy increases until NE pass through the hydrogen position. From this point onwards it is H-O bonding state that takes over in the algorithm (lower energy) until attaining the minimum energy geometry for product $W^{(P)}$. Such a calculation does not enforce local NE conservation; consequently it is not directly comparable to excited states in MO theory. The point at which the character of the NE changes would correspond to a (virtual) conical intersection in BO language.

By looking at energies it can easily be seen that HOMO C-H and LUMO H⊥O changes into LUMO C⊥H and HOMO H–O. The change ("reaction") is produced by a geometric change driven by the program with standard algorithm: i.e. filling orbitals of lowest energy to construct a one-determinant poly electronic function; this is not a quantum physical process. Introducing NE control simply means that at HOMO/LUMO inversion one disable automatic lowest energy orbital filling to impose NE pattern conservation. In this manner one would get diabatic-like PES that can be taken as model Marcus PE functions. This approach offers a path to *interpret* adiabatic results in terms of diabatic ones thereby opening for quantum mechanical descriptions in terms of the linear superposition principle. Thus, a multi-state reactivity scheme can hence be constructed in a natural manner.

Concerning actual 1-e models, it is not difficult to see the fundamental role played by parameter choice when constructing a one-electron Hamiltonian. The chemical reaction presentation relies on realistic HOMOs and LUMOSs and their crossings. The role of atomic basis sets becomes critical. Today, density functional

theory (DFT) via Kohn-Sham equations offers another framework to perform a type of parameter search allowing locations of HOMOs and LUMOs in a manner differing from Hartree-Fock schemes.

To sum up: within standard BO with $\mathbf{q}(\mathbf{X})$ algorithm one loses control over nodal patterns characterizing the initial attractor quantum state so that PESs change pattern as one moves along a reaction coordinate; chemical change obtains not as a chemical physics process but as geometry imposition executed by the practitioner; but, nevertheless, one gets invaluable geometry information at stationary points. Including NE pattern control one gets diabatic-like PES and consequently a QM-like representation in term of the linear superpositions principle (cf. Eq. (10-12). For an exact theory, the electronic configuration space is essentially independent of nuclear positions as required also by our diabatic framework. One gets a plus in terms of the analysis of symmetries, for example, rotation invariance, i.e. angular momentum, that is now considered from a unique origin associated to the selected I-frame. The concept of object fades away while the concept of quantum state occupies center stage. Because spin is not a space property, the preceding result must include angular momentum variables in a quantum mechanical setting, as we will see below.

10.6. ANGULAR MOMENTUM (SPIN) AND CHEMICAL REACTIVITY

In relativistic and non-relativistic limit quantum mechanics the total angular momentum $F = S + L + I = J + I$ is a conserved magnitude under inertial frame rotations. This is also true in semi-classic theory with a caveat: the quantum mechanical aspect of the positive charge background is neglected and this has a price [5].

Take again the QM description of quantum states for the hydrogen atom that are sustained by the material system constituted by one electron and one proton. Thus proton spin states must be included. For $L = 0$, the total spin state S+I for the hydrogen atom can add up to $F = 0$ and 1. This quantum system is a compound boson. The issue discussed in a preceding section should have a direct impact into the chemical reactivity.

To study the interaction (reaction) H + H as we do in quantum chemistry the system is represented as a two-electrons and +2 PCB system with the help of the Coulomb Hamiltonian; it is a typical *diradical* reaction. Or one can see the system as two bosons states weakly interacting. In the former case one misses a fundamental quantum element (nuclear spin). The rule imposing a nodal envelope once the two I-frame atoms are "merged" into a one-I-frame system in the semi-classic framework still holds.

To complete the semi-classic recipe, the electronic spin quantum state cannot be an eigenfunction of S^2 but a linear combination of spin singlet $|S = 0, M_S = 0\rangle = (\alpha(1)\beta(2) - \beta(1)\alpha(2))/\sqrt{2}$ and spin triplet $|S = 1, M_S = 0\rangle = (\alpha(1)\beta(2) + \beta(1)\alpha(2))/\sqrt{2}$, $|S = 1, M_S = 1\rangle = \alpha(1)\alpha(2)$ and $|S = 1, M_S = -1\rangle = \beta(1)\beta(2)$. Ordering the

spin basis set in a column vector: [α(1)β(2) β(1)α(2) α(1)α(2) β(1)β(2)] the spin quantum state for the entry channel representing a head-on collision where beam-1 is prepared in spin β(1) state and beam-2 in α(2) state, the row amplitudes for this situation reads as:

$$|\text{Beam - } \beta(1);\text{beam - } \alpha(2)> = (0\ 1\ 0\ 0)\ [\ \alpha(1)\ \beta(2)\ \ \beta(1)\ \alpha(2)\ \ \alpha(1)\ \alpha(2)\ \ \beta(1)\ \beta(2)]$$
$$\rightarrow (|S=1,M_S=0> - \ |S=0,M_S=0>)/\sqrt{2} = (|T_o> - \ |S>)/\sqrt{2} \tag{10-16}$$

This corresponds to a spin quantum state with non-zero amplitude at β(1)α(2) for a typical laboratory setup.

The type of electronics described above does not lead to chemical reaction even in the standard approach. Imposition of our NE constraint for the space part brings the triplet spin state to complete the basis state: $1s_a(1)1s_b(2)–1s_b(1)1s_a(2)$. The labels a and b denote two I-frames in real space; the products ab and ba signal possible assignments, that QM consider simultaneously to get a complete quantum space. The analysis presented so far indicates that it is the spin triplet that would show the amplitude different from zero. No mechanical formation of hydrogen molecule is hence allowed.

However, use of a computer package containing for instance CASSCF-PT2 scheme will produce a PES leading to formation of H_2. The reason is the choice made for an initial trial function imposing a singlet spin state so that the product reads; $N(1s_a(1)1s_b(2) + 1s_b(1)\ 1s_a(2))\ (\alpha(1)\beta(2) - \beta(1)\alpha(2))/\sqrt{2}$ where N is a normalization factor. This result contradicts experiment. Formation of molecular hydrogen at finite rate requires a catalyst [18].

The principle of spin conservation along a reaction coordinate for diradical reactions is not granted in the present approach. Even the most sophisticated computational methods based on a **q**(**X**) algorithm, that is to say, for all methods using the LCAO or LCGO schemes, would predict wrong entry channels leading to bond forming. This type of computation would miss the crossing of asymptotic states (that must be adapted as shown above) with bonding states. For a diradical state because of the anti-bonding state any closing displacement would yield a repulsive energy thereby setting up the source for a barrier at crossing with the bonded state.

Spin conservation rule should be seriously challenged in most cases. For, opening up an entangled spin state to produce another entangled state elsewhere this can be done within a full angular momentum space as we will see below.

10.6.1. Spin–Space Separation and Chemical Reactivity

The theoretic framework presented here uses many-electrons basis sets where spin and space are separated. In computational chemistry the use of spin-orbitals and Slater determinants complicates the analyses but simplifies the algorithms. The use of the general approach to discuss chemical reactivity is illustrated below.

Ethylene and dioxygen are isoelectronic. The latter is reactant in photorespiration catalyzed by RubisCO. The fragment –COH(2)–COH(3)– that is active element in the substrate RuBP during catalysis is closely related to ethylene. It is interesting to examine reactivity issues for these systems from the present perspective [19, 20]. The electronic ground state of oxygen is a triplet ($^3\Sigma_g^-$) while for ethylene singlet and triplet states play reactive roles to be examined below.

10.6.1.1. Ethylene

There are two *possible* nodal planes to set up the system: x, z-plane with y-axis perpendicular and y, z-plane with x-axis perpendicular to a planar PCB; the relative nodal planes are perpendicular, the choice is arbitrary but once it is done one must stick to it. The C–C axis is drawn along the z-axis, the origin of the I-frame in between. We speak of possible planes although in an object-like description only one will be realized. But in QM they contribute as possible (orthogonal) basis states [19, 20].

A reaction coordinate defined as a disrotatory displacement of each methylene fragments relates a PCB located in one plane to the other; Note that this is not a symmetry operation but a RC [20]. For, if one colors hydrogen atoms as green and blue at each methylene fragment and assign green (blue) in cis conformation then the disrotatory displacement by $\pi/4$ leads to a trans conformation for the color centers; the nodal plane perpendicular to x-axis becomes y-axis. The conformers are hence topologically different because of relative nodal plane disposition; groups attached to methylene will not "move" a lot under this type of displacement. One can then define two types of basis functions $^1\Phi_{\mathrm{cis(i)}}(\mathbf{q},\mathbf{X})$ and $^1\Phi_{\mathrm{tras(i)}}(\mathbf{q},\mathbf{X})$ with different nodal characteristic; spin multiplicity is signaled as usual. Check that the symmetry groups differ for this colored set up as it will do for actual cis/trans conformers.

Planar PCBs show minimum energy equilibrium configurations $^1\Phi_{\mathrm{cis(0)}}(\mathbf{q},\mathbf{X}^{\mathrm{cis}})$ and $^1\Phi_{\mathrm{trans(0)}}(\mathbf{q},\mathbf{X}^{\mathrm{trans}})$;where i = 0 indicates electronic ground state.

For spin triplet state planar $–CH_2$ fragments are out-of-plane at angles either $\pm\pi/2$ between the normal vectors to the fragments $–CH_2$; the equivalent to disrotatory angle is $\pi/4$ [20]. Spin triplet base state $^3\Phi_{\mathrm{k(0)}}(\mathbf{q}, \mathbf{X}^{(\pm\pi/2)})$ corresponds to a minimum energy PCB. In the exact model one would have:

$$\Psi(q,\mathrm{X}) = \mathrm{C}_{\mathrm{cis(i)}}{}^1\Phi_{\mathrm{cis(i)}}(q,\mathrm{X}) + \mathrm{C}_{\mathrm{trans(i)}}{}^1\Phi_{\mathrm{trans(i)}}(q,\mathrm{X}) + \mathrm{C}_{\mathrm{k(i)}}{}^3\Phi_{\mathrm{k(i)}}(q,\mathrm{X}) + \ldots \tag{10-17}$$

The amplitudes are just complex numbers while for the semi-classic Eq. (10-12) write the superposition as:

$$\Psi(q,\mathrm{X}) = \mathrm{C}_{\mathrm{cis(i)}}(\mathrm{X})^1\Phi_{\mathrm{cis(i)}}(q) + \mathrm{C}_{\mathrm{trans(i)}}(\mathrm{X})^1\Phi_{\mathrm{trans(i)}}(q) + \mathrm{C}_{\mathrm{k(i)}}(\mathrm{X})^3\Phi_{\mathrm{k(i)}}(q) + \ldots \tag{10-18}$$

Thus, a general electronic quantum state can be a mixing of different spin states but the basis states don't. External magnetic fields and gradients can put "in motion"

a quantum system as amplitudes change in time: $\Psi(\mathbf{q};\mathbf{X},t)$. Either Eq. (10-17) or (10-18) would put time dependence in the amplitudes, e.g. $C_{k(i)}(\mathbf{X},t)$ [5]. Although we are more used to think static situations, chemical change can be sensed via changes of quantum amplitudes as a function of time. It will depend upon injected energy (energy shell) to know those amplitudes that are set in motion [4, 5].

Experimentally, one senses the response of the quantum state $\Psi(\mathbf{q};\mathbf{X},t)$ sustained by the material system under study. This mean something simple: if a given base state show a non-zero amplitude in an otherwise normalized quantum state, the electronic spectra rooted at its energy eigenvalue can be activated when external sources of energy are focused onto the material system; the relative response intensity is given by the square modulus of the root sate amplitude. This is what we do when a chemical reaction is followed with adequate spectroscopic techniques. Sensing transient states corresponds to a time evolution where the amplitude at the base state for the "transient" is changing in time in a specific manner.

Chemical reactivity can also be sensed via $\Psi(\mathbf{q};\mathbf{X},t)$ with appropriate sensors. In what follows we discuss some reactivity issues.

An exact QM description of a cis-trans process would, e.g., starts by activating the vibronic states associated to $\Phi_{cis(i)}(\mathbf{q},\mathbf{X})$ until getting at those energy levels states from $\Phi_{trans(i)}(\mathbf{q}, \mathbf{X})$ that might come into quasi resonance. Activation implies external EM fields feeding or withdrawing energy. In the electric dipole model, direct coupling between these states is possible only via a third state showing parity different to cis and trans base states. This new state plays the role of quantum transition state as already discussed. It will be the light field the source prompting for "conformer" change [5, 19, 20].

The quantum state presented by Eq. (10-17) highlights a geometry dimension commensurate to a semi-classic QM scheme.

Thus, whenever a PCB attains an attractor arrangement determined by its diabatic electronic base function, the Coulomb field may range the energies of excited states. In principle, it is the diabatic electronic state that traps a PCB and the complete electronic spectra present excited states characterized by relative nodal planes distributions [3, 4]. In practice, we invert the procedure as befit our computing tools; that opens a possibility to determine minimum energy (stationary) state geometry.

Now, a disrotatory displacement of rigid, planar, carbene-fragments ($-CH_2$) is a reaction coordinate in a different guise. The displacement connects the plane y,z to plane x,z when the absolute value of the angle is $\pi/2$; bear in mind that the electronic state is not consequently changed. Instead of colors, put label 1(1), 2(1) to hydrogen atoms from one carbene fragment and 3(2), 4(2) to the other. Take for instance distances 1(1)–3(2) to be a generic cis configuration and 1(1)–4(2) for trans. The disrotatory displacement brings 1(1)–3(2) into trans configuration and 1(1)–4(2) into cis. A physical process changing cis- into trans-configuration will apparently relate the PCB from one plane to an orthogonal one keeping C–C along the same direction. Of course, experimentally these two "species" cannot be distinguished unless some label permits an experimental detection. Furthermore, we

know that a triplet spin state is to be found in between helping to open and close entanglements as it were.

From a computational viewpoint why do we bother with diabatic paths if we can simply get all what is required to describe conformational change in the adiabatic way? Answer: one may want to grasp intermediate basis states that may prompt for a change of quantum state along the reaction coordinate. Vibration excitation per se does not produce a chemical change; in the standard BO case with the operator $H_{op}(\mathbf{q}(\mathbf{X}),\mathbf{X})$ the change is induced by the algorithm $\mathbf{q}(\mathbf{X})$, it is not a quantum mechanical effect. But we can proceed by supplementing BO with some simple rules.

First, a simple qualitative method based on the number of nodal surfaces is introduced. For planar ethylene the number of nodal surface is $L = 1$ and we make it correspond to orbital angular momentum. The ground state total $J = L + S$ is 1.

Stopping disrotatory angle at $\pi/4$, the π-bond "disappears", but each carbene-fragment displays an orbital perpendicular to the corresponding plane and perpendicular to each other; this is equivalent to an orbital angular momentum $L = 2$. Displacement in the opposite direction one gets the normal to the planes in orthogonal directions. In this configuration, spin triplet state ($S = 1$) shows energy lower than for the spin singlet. In this bent-out shape with respect to the (ground state) singlet there corresponds minimum energy geometry; the C–C distance (covalent bond) relaxes in a spin triplet state. Thus, the attractors obtain from computations [5].

If one enforces spin $S = 0$ along the disrotatory coordinate and no NE constraint is ensured at $\pi/4$ we will find a potential energy maximum. Thence shifting up to $\pi/2$ one would locate the energy above the minimum this time located in another plane (orthogonal) with respect to the starting point at zero degrees. At crossing ($\pm\pi/4$) if there is a coupling one would end up at the other conformer on a different nodal plane. This sort of space displacement is useful when a molecular motor is to be designed; one has however to design the couplings to achieve that change because the result from the algorithm $\mathbf{q}(\mathbf{X})$ does not portray a physical process.

Back to NE constraint, at $\pi/4$ two nodal planes correspond to methylene planar fragment while C–C π-bond "disappear". One is working with electronuclear quantum base states not with substances. As noted above, this state corresponds to $L = 2$. Looking from the z-axis a d-orbital-like pattern emerges. In molecular spectroscopic nomenclature this may contain a delta (Δ) state, namely, $|M_L| = 2$.

As a matter of fact, at each end there is a free $\ell = 1$ basis functions so that according to angular momentum sum rules one should get states $L = 2, 1, 0$. Spin states $S = 1,0$. And total angular momentum $J = L + S = 3, 2, 1, 0$. We look at the basis states associated to the angular momentum.

We present some abstract basis states using the angular momentum scheme. The basis state $|\, J = 3, M_J = 3; L = 2, M_L = 2, S = 1, M_S = 1\rangle$ is associated to $\pi/4$ space configuration.

The ground basis configuration is associated to $|\, J = 1, M_J = 1; L = 1, M_L = 1, S = 0, M_S = 0\rangle$; also, $|\, J = 1, M_J = -1; L = 1, M_L = -1, S = 0, M_S = 0\rangle$ and $|\, J = 1, M_J = 1; L = 1, M_L = 0, S = 0, M_S = 0\rangle$.

The C-C axis can be used to define a quantum number similar to the diatomic Λ describing the projection of L along the molecular axis: e.g., $|M_L| = \Lambda = 0, 1, 2$ or Σ, Π, Δ states. The $\pi/4$ space configuration may cover $\Lambda = 2, 1, 0$.

Note that a direct transition from $|J = 1, M_J = 1\rangle$ to $|J = 3, M_J = 3\rangle$ is not allowed. At this level there is need of base states for $J = 2$. The subspace is spanned by $|J = 2, M_J = 2; L = 1, M_L = 1, S = 1, M_S = 1\rangle = {}^3\Pi_2$, $|J = 2, M_J = 2; L = 2, M_L = 2, S = 0, M_S = 0\rangle = {}^1\Delta_2$, $|J = 2, M_J = 0; L = 2, M_L = 0, S = 0, M_S = 0\rangle = {}^1\Sigma_2$, etc. Allowed transitions involve $J \rightarrow J + 1$ and $J \rightarrow J - 1$. This means couplings that would allow for changes of the amplitudes of a TQS. A spin-orbit operator would do the job and under resonance conditions a strong mixing would follow.

The basis states shown in general are not related to configuration space. Here, they are qualitatively based on relevant nodal planes. These may be global or local (methylene fragments). Because the theory is addressing quantum state, rotation invariance with the general angular momentum rules hold. Here, one focus on the set of basis states that might be relevant to discuss reactivity issues.

Let us write ethylene's time dependent quantum state [5, 19, 20] where only few base states are explicit:

$$\Psi(q,\mathrm{X,t}) = \mathrm{C_{cis(i)}}^1(\mathrm{X,t})^1\Phi_{\mathrm{cis(i)}}(q) + \mathrm{C_{trans(i)}}^1(\mathrm{X,t})^1\Phi_{\mathrm{trans(i)}}(q) + \mathrm{C_{k(i)}}(\mathrm{X,t})\Phi_{\mathrm{k(i)}}(q) + \mathrm{C_{cis(i)}}^3(\mathrm{X,t})^3\Phi_{\mathrm{cis(i)}}(q) + \mathrm{C_{trans(i)}}^3(\mathrm{X,t})^3\Phi_{\mathrm{trans(i)}}(q) + \ldots \tag{10-19}$$

The crux of the matter is that at the TQS the amplitudes different from zero would involve the terms with amplitudes $\mathrm{C_{cis(i)}}^1(\mathbf{X},\mathrm{t})$, $\mathrm{C_{cis(i)}}^3(\mathbf{X},\mathrm{t})$ and $\mathrm{C_{trans(i)}}^1(\mathbf{X},\mathrm{t})$. Bear in mind that a quantum state, in general, it is not a spin eigenfunction. For the sake of discussion all other base state amplitudes are left aside.

If the system is hold at a neighborhood of, say $\mathbf{X}^{\mathrm{TQS}}$ the system may show non-zero amplitude at the singlet and triplet components in Eq. (10-19). This is all what is needed to activate the system from an initial singlet state towards a (transient) triplet state reactant.

A conformational change would take the vibronic triplet state to couple to the vibronic state of the conformer with an activation energy lower than the one obtained by a crossing of the spin singlet state. So, a conformational change would activate ethylene via the diradical base states discussed above.

Furthermore, exothermic reactions between radical species related to a new system featuring a covalent bond must recur to a catalyst as it was the case with $H + H$ to yield $H_2({}^1\Sigma_g{}^+, v = 0)$ discussed above.

A solvent may act as catalyst. See for example the addition of carbon-centered radicals to alkenes in reference [21] where it is reported the gas phase reaction goes much slower than in solvent (cf. Table 3, page 1350 in [21]).

One of the most interesting results of the present QM approach is the non-conservation of spin angular momentum along a reaction coordinate. It is the sum $S + L$ that can be conserved. This outcome is more general than the example

under analysis. For a change of spin entanglement requires assistance from a corresponding change of orbital angular momentum; this is of course a way of speaking, what the statement implies is that you have to provide the basis states relevant to the process description so that the interplay of amplitudes can be registered as a sort of structural change. The transition rules are associated to the total angular momentum. It is the abstract electronic configuration space that allows use of general symmetry invariance of quantum states towards rotations of I-frames. The algorithm **q**(**X**) actually executes the L-changes unfortunately without appropriate control since spin is frozen.

10.6.1.2. Carbene Fragments: Basis States

It remains to discuss bond-knitting path from two carbene molecular states correlated to ethylene. This is an exothermic process (cf. Section 10.5.1). One may suspect the need for a catalyst in a manner similar to the one discussed in Section 10.5.1 albeit with more involved base states.

Each carbene shows L = 0,1 and S = 0,1 states. Thus, J = 4, 3, 2, 1, 0 for total angular momentum. Now, the two I-frame systems must be included into states for the one-I-frame case to make them commensurate to Ethylene. There must be now a nodal envelope separating the carbene states; this means a contribution of L = 1 plus the two nodal planes for carbene fragments yielding L = 3 is the lowest orbital angular momentum. Remember that this is a decision we make to complete the basis set with asymptotic states. Now we have the spin to be fixed. For S = 2 the base state is fully non-bonding. Next would come S = 1. For the case where both carbene fragments are triplets then J = 3 + 2 and one has to consider base states starting with J = 5, L = 3 and S = 2: | J = 5,M_J = 5; L = 3,M_L = 3, S = 2,M_S = 2> = $^5\Gamma_5$. The fragments in mixed spin states yield: | J = 4,M_J = 4; L = 3,M_L = 3, S = 1,M_S = 1> = $^3\Gamma_4$. Two fragments in singlet spin state yields J = 3 as ground state: i.e., gamma symmetry $^1\Gamma_3$ = |J = 3,M_J = 3; L = 3, M_L = 3, S = 0,M_S = 0>; also some components would show states $^3\Delta_3$ and $^5\Pi_3$.

Thus, the nodal envelope constraint introduces a term L = 1 compared to the independent fragments. This is a general property for diradical intermediates meaning with this that the uncoupled one spin electron states must be repulsive in their space part thereby leading to a spin triplet. The theoretical result obtained here is at variance with those standard based on the calculations with **q**(**X**) algorithm [22].

Now, because we want two methylene groups orthogonal to each other one must select L = 3. Amplitudes over base state $^5\Gamma_{J=5}$ would identify two carbene moieties in a spin quintuplet where no bonding elements are present. Ethylene ground state $^1\Sigma$ with J = 0 is not directly accessible to a colliding pair because spin entanglement must be reshuffled. Three transition amplitudes (steps) are required at least to bring amplitudes from one state to the other. First, spin flip connecting $^5\Gamma_{J=5}$ to $^3\Gamma_{J=4}$. This is also a matter of preparing the spin states by selecting M_S. Next may come in $^3\Delta_{J=3}$ followed by $^3\Pi_{J=2}$, base state $^1\Pi_1$ ending in $^1\Sigma$ (π^2 configuration). These are not the only intermediate base states. There is, for

example $^1\Delta_{J=2}$ standing for the twisted ethylene spin singlet. The system base states for $L = 2$ and $S = 1$ would correspond to single σ bond with singlet spin and two $-CH_2-$ radical fragments in a global triplet. The spin singlet state does as a $J = 2$ state while the triplet spin state is a $J = 3$ element: $^3\Delta_{J=3}$.

The base state $^1\Gamma_{J=3}$ would identify two carbene moieties combined into a singlet spin that do not enter the reactive space of ethylene because there is no global NE along C C; this may sound confusing, because when the standard $\mathbf{q}(\mathbf{X})$ algorithm is used one fill the lowest energy orbitals and consequently a downhill profile obtains; it goes without saying that such result looks as an artifact.

The base state $^3\Pi_{J=2}$ stands for an excited $\sigma \rightarrow \pi^*$ configuration. The $J = 2$ states are auxiliary in the sense of connecting transitions of $J = 3$ with $J = 1$. The base state $^1\Pi_{J=1}$ corresponds to the ground state of either cis or trans configurations which can be assigned to $M_L = \pm 1$; one π-orbital and the remaining being sigma orbitals. The base state $^1\Delta_{J=2}$ can be associated to $(\pi\sigma)^1(\pi^*)^1$ spin singlet excitation of a planar PCB; (π^*) would look like as d-orbital with 2 nodal planes. The state $^3\Delta_{J=3}$ can also be associated to configuration $(\pi\sigma)$ (π^*).

In a laboratory set up uncorrelated carbene fragments are not an eigenfunction of total S_{op}^2. The first step must be related to single bond ethylene state. Here, two electrons enter the picture with the space function presenting a nodal plane. Thus, the spin part is a triplet state. To couple this base state to $J = 2$ bonded states there is need for a TQS with $J = 3$, namely, $^3\Delta_{J=3}$. This correlates precisely to the $\pm\pi/4$ conformation of twisted ethylene.

Chemists look at chemical transformation in terms of hybridization change and bond displacements (arrows). For example, take the nodal plane with normal vector in the three-direction (z-axis), for each C-atom one can form quantum states as linear superpositions of orbitals p_x, p_y and s to get the sp^2 hybrids. Adding to these functions a p_z component, the resultant geometric shape would change. A similar result would obtain by using appropriate amplitudes for relevant basis states presented above.

The approach used here via total angular momentum labels permits mixing of space and spin aspects of basis states. Selection rules apply to J or F as the case might be. Furthermore, exothermic reactions between radical species leading to a new system with a covalent bond must recur to a catalyst. And of course a solvent would act as catalyst.

Reactions involving oxygenation of carbenes are exothermic and would require solvent to catalyze the process eluding explosive conditions. It can also be concluded that the reaction of two carbenes radicals to ethylene synthesis would require of a catalyst, a third body, to proceed at finite rate to the intermediate diradical (triplet). The radical centers must show a NE state or in other words they must be repulsive.

Connections to the carbon–carbon bond break in RubisCO can be discussed in similar terms. It is apparent that many of the spontaneous hydrogen transfer we found along reaction paths must be carefully reexamined. The structures of the saddle points of index one would remain.

10.7. PHOTORESPIRATION: DIOXYGEN

RubisCO is an oxygenase. Once Lysine-201 is carbamylated the substrate RuBP is taken into a dienol form as describe in Section 10.2 dioxygen can attack C(2) center in the twisted dienol form. We need now to explore those quantum base states for oxygen that might control reactivity.

This system and ethylene are iso-electronic while they obviously differ in PCB disposition. There are three low-lying electronic states of interest: ground state $O_2(^3\Sigma_g^-)$, next attractor is $a^1\Delta_g$ followed by $b^1\Sigma_g^+$; the excited state $A^3\Sigma_u^+$ state correlates to $O(^3P) + O(^1D)$ while the three low-lying ones do to $O(^3P) + O(^3P)$.

No direct electric dipole transition from ground state to these excited states is allowed. States mixing is carried out in a fairly subtle manner [23–25]. The first excited state to which second order transition is allowed is $O_2(B^3\Sigma_u^-)$; the absorption band lies between 175-and 200-nm wavelengths.

Our early work on TS for oxygenation spotted the possible coupling between the equivalent to $^3\Phi(L = 2)_{J\,=\,3}$ of dienol fragment and $O_2(^3\Sigma_g^-)$. We assumed that spin coupling would produce singlet state so that computations were carried out in this manner [26, 27] albeit no analysis of dioxygen electronic states was given by us. However, in order to open reactive channels a coupling between $O_2(^3\Sigma_g^-)$ and $b^1\Sigma_g^+$ must be ensured. Once this is done, the base state $a^1\Delta_g$ can be coupled via base state $b^1\Sigma_g^+$ allowing an opening of the chemical reactive so-called singlet oxygen O_2 ($a^1\Delta_g$). Therefore, the opening of oxygen chemical reactivity is a subtle quantum mechanical process where along the reaction path spin is not conserved.

This point can be elaborated a little further by including the two lowest lying singlet states of dioxygen, namely, $a^1\Delta_g$ and $b^1\Sigma_g^+$ [23]. Including the ground state the basis states form the column vector:

$$[|X^3\sum_{g,0}^{-}, v>|X^3\sum_{g,1}^{-}, v>|X^3\sum_{g,-1}^{-}, v>|\ a^1\Delta_g, v>|b^1\sum_{g}^{+},v>\ldots].$$

In general a quantum state omitting for the time being sums over vibration quantum numbers would look like as a row vector ($C_{10}\ C_{11}\ C_{1\text{-}1}\ C_2\ C_3\ \ldots$) [4, 5]:

$$|\Psi> = C_{10}(\Psi)\ |X^3\sum_{g0}^{-}, v> + C_{11}(\Psi)\ |X^3\sum_{g1}^{-}, v> + C_{1-1}(\Psi)\ |X^3\sum_{g-1}^{-},$$

$$v> + C_2(\Psi)\ |\ a^1\Delta_g, v> + C_3(\Psi)\ |b^1\sum_{g}^{+},v> + \ldots \tag{10-20}$$

Different basis states can be coupled via a certain number of physical operators. A spin-orbit coupling H_{SO} between $|\ b^1\Sigma_g^+,v>$ and $|X\ ^3\Sigma_{g0}^-, v>$ mixes these states [25] yielding a linear superposition for ground vibration states:

$$|\Psi_{X,0}\ v=0> = C_{10}(\Psi_{X,0})\ |X^3 \sum_{g0}^{-}, v=0> - C_3(\Psi_{X,0})|b^1 \sum_{g}^{+}, v=0> \tag{10-21}$$

$C_{10}(\Psi_{X,0}) \approx 1$ and $C_3(\Psi_{X,0}) = Z^*$ and the upper state reads

$$|\Psi_b,\ v=0> = C_{10}(\Psi_b)\ |X^3 \sum_{g0}^{-}, v=0> + C_3(\Psi_b)|b^1 \sum_{g}^{+}, v=0> \tag{10-22}$$

$C_3(\Psi_b) \approx 1$ and $C_{10}(\Psi_b) = Z$. The key result is that the ground state may borrow intensity from the b-excited state. From our perspective, as soon as vibration excitation is allowed the energy gaps between vibronic states diminish and mixing among such states increases.

The magnetic dipole operator $M_{op} = \beta\ (L_{op} + 2S_{op})$ coupling the spin multiplet from the ground state $|X\ ^3\Sigma_{g0}^-, v=0>$ to $|X\ ^3\Sigma_{g\pm 1}^-, v=0>$. The electric quadrupole moment operator can mix up the ground state to the delta state. All these gymnastics leads to a nice result: the ground state may show small amplitudes at base state $|a^1\Delta_g, v=0>$ thereby opening reactivity channel towards singlet spin state. Couplings can be enhanced by intermolecular interactions [25]. Our vibronic theory permits understand the quantum mechanical determinants for the increase in amplitude for singlet oxygen basis set.

Quantized surroundings such as an enzyme or aggregate thereof may withdraw or inject energy quanta towards or from the reactive quantum system. Eq. (10-20) is written as:

$$|\Psi> = C_{10}(\Psi)\ |X^3 \sum_{g0}^{-}, v_1> + C_{11}(\Psi)\ |X^3 \sum_{g1}^{-}, v_1> + C_{1-1}(\Psi)\ |X^3 \sum_{g-1}^{-},$$
$$v_1> + C_2(\Psi)\ |\ a^1\Delta_g, v_2> + C_3(\Psi)\ |b^1 \sum_{g}^{+}, v_3> + \ldots \tag{10.23}$$

The energy gaps between vibronic states can be modulated via collective modes from the global enzyme aggregate. In particular, the amplitude $C_2(\Psi)$ may grow so that the reactivity of singlet oxygen may show up. These results justify our calculations using singlet spin supermolecule to get TS structures [26, 27]. The intersystem crossing is achieved in subtle quantum mechanical manners that are difficult to get with simple levels of theory based on the use of $\mathbf{q}(\mathbf{X})$ algorithm. Yet one can make an educated guess as done in references [26, 27] and proceed with computations by checking the singlet/triplet energy gap as done in those works.

The activation process of RubiCO assisted by light [9] is critical to get activation at the substrate C(2) but also if dioxygen is fixed at the active site the strict selection rules are bypassed in ways described above. Activation towards carboxylation and oxygenation cannot be parted.

Understanding enzyme catalytic properties goes beyond the study of some chemical conversion steps. As pointed out by Inger Andersson [9], RubisCO's complex biosynthesis and multiple catalytic reactions are subjected to tight control involving light. RubisCO holoenzyme (so called form I) is fully active only when eight large (L) subunits and eight small (S) subunits form a complex $(L_2)_4(S_4)_2$ hexadecamer. Our contribution, albeit modest, must be put in perspective. The quantum formalism presented here including quantized EM radiation fields and used to construct a theoretical model of chemical reactions underlines the importance not only of energy conservation between the reactive subsystem and the catalytic machinery, it also emphasize spectroscopic aspects where finite transition amplitudes are determining reaction channel openings.

10.8. MORE LIGHT

In spite of the rather elementary use of quantum electrodynamics made here, the role of light, or more generally, electromagnetic radiation as an essential element of the understanding of chemical change has been documented.

More light can be shinned if a systematic use of this molecular quantum electrodynamics is made to describe catalysts sources. Even the apparently simple H + H system helped realizing the limitations one is confronted when standard quantum chemical computing schemes are used without due analyses.

Reactivity issues depend upon subtle couplings; a case in point concerns dioxygen lowest electronic states where surrounding effects are expected to modulate transition amplitudes as it happens in particular solvents [24]. Ethylene taken as model of alkenes in addition reactions of carbon-centered radicals is an interesting case. Perhaps the isomer effect can be described from the present perspective. For *Z*-1,2-disubstituded alkenes react more slowly than *E*-1,2-disubstituted; the nomenclature used in Organic chemistry assign cis- to *Z*- and trans- to *E*-. The QTS must be found in the region where the system is twisted about $\pi/4$ of the parent alkene. This state cannot be obtained with standard BO schemes as they freeze spin while a linear superposition form such as Eq. (10-19) mix spin states as we saw for dioxygen.

A necessary condition to activate an ethene fragment is production of non-zero amplitude at the triplet state multiplet; in particular at $\alpha\alpha$ or $\beta\beta$ base states. The physical mechanism can be spin-orbit couplings. In more chemical terms there will be carbon-centered radical at each methylene in orthogonal planes; vibration along the disrotatory reaction coordinate would open *Z*/*E* isomerization channels. Thus vibronic states of ethene-fragment $|^3\Delta_{MS \pm 1}, v>$ that we can think of sustained by a potential energy function to make contact with the curve crossing models although the setting is different. The addition reaction of carbon-centered radicals to triplet activated ethene fragment must proceed along the repulsive interaction between the asymptotic states as already discussed. The base states involved are described with chemical graphs as quantum labels. The collision (antibonding) base state

$|R^{\bullet} \perp {}^{\bullet}CHA\text{-}C^{\bullet}XY,v> = |1,v>$ and target radical $|2,v'> = |R\text{–}CHA\text{–}C^{\bullet}XY,v'>$ which seen from ethene a π-bond is replaced by a σ-bond. Note that planes for CHA and $C^{\bullet}XY$ must be perpendicular to each other; vibration number refers to their anti-symmetric fluctuations.

The quantum state given by Eq. (10-23) feature the basic elements of what we describe as bond making(BM)/breaking(BB):

$$|BM/BB> = C_{1v}(BM/BB)\ |1,v> + C_{2v'}(BM/BB)\ |2,v'> \qquad (10\text{-}24)$$

It is at crossing between the asymptotic dissociation states in the continuum, e.g. $|R^{\bullet} \perp {}^{\bullet}CHA\text{-}C^{\bullet}XY,v>$ with finite relative kinetic energy and ethene fragment state $|^3\Delta_{MS\ \pm\ 1}, v>$ where amplitudes can be modulated so one monitors the change of initial quantum state row vector $(C_{1v}(BM/BB)\ C_{2v'}(BM/BB)\ldots) = (1\ 0\ \ldots)$ to a resultant state $(C_{1v}\ C_{2v'}\ \ldots)$ where the sum $|\ C_{1v}|^2 + |\ C_{2v'}|^2$ is essentially equal to one. With the help of the semi-classic framework one can figure out the geometry that one can associate to the QTS. Case 1: $|\ C_{1v}|^2 > |\ C_{2v'}|^2$ characterizes early TSs in the old picture. Case 2: $|\ C_{1v}|^2 < |\ C_{2v'}|^2$ characterizes late TSs. Time dependence would add new features not discussed here.

The energy barrier concerns the base state $|R^{\bullet} \perp CHA = CXY,v'',k> = |0,v'',k>$ involving ethylene (ethene fragment) in the ground state with $k = 1,2$ labeling *Z* and *E* conformers. There must be a libration level v'' coming in a neighborhood of the crossing thereby determining the activation energy. One may have cases where *Z*- and *E*-states show equal zero-point energy. There are other cases where *Z*-PESs are deeper than *E*-PES; and vice-versa. The deeper the well is, the larger activation energy to get at the spin triplet PE. For 1,2-disubstituted alkenes it is common to find that *Z*-1,2-disubstituted alkenes react considerably slowly than *E*-1,2-disubstituted alkenes. This part of the kinetics refers to activation of ethene reactants.

Note than the density of states at crossing are larger for deep PES thereby modulating the relaxation rates from the triplet configuration into ground state of virtual ethene fragment.

A number of mechanistic issues are not discussed here. Concerning Rubisco, in references [26, 27] there are detailed correlations of experiment to theory. The examples given above serve to illustrate the line of thought. The picture differs radically with standard BO calculations. For, independently of the level of theory one plans to use, the result would reflect the algorithm $\mathbf{q}(\mathbf{X})$ and spin constraints. Yet characterizations of TSs offer unique tools to help identify reactive nodal envelopes; standard computations packages must be (can be) supplemented with NE to get model diabatic PE surfaces. Relevant quantum base states must be searched.

The electronic transition amplitude connecting ingoing to outgoing channels gives the principal factor for finite rates of change. Type and number of excited electronic states required to produce finite transition amplitudes enter as quantum effectors for the chemical mechanism. For practical applications there is need of accurate calculations of transition structures to fix geometry parameters and

relevant nodal envelopes patterns. Determination of stationary geometry structures would complete the lay out for this approach.

An important set of question marks has been raised: *are nodal envelope states discussed in Section 10.4.3 comparable to the excited states in MO theory? Then the treatment is quite similar to CI method? I understand that here the author is including nuclear coordinates as variables*. Note that solutions to $H_{op}(\mathbf{q}(\mathbf{X}),\mathbf{X})$ are not the same as $H_{op}(\mathbf{q}(\mathbf{X});\mathbf{X})$; the latter are used in MO theory. The use of the former operator is justified as a way to emphasize the no-go situation one gets at while trying to construct a consistent scheme. The model implies use of $H_{op}(\mathbf{q},\mathbf{X})$ where the electronic configuration space is fully independent from nuclear positions. It is only under such circumstance that one can use quantum mechanical formalism properly; the generalized diabatic scheme helps bridging these otherwise incompatible models. The approach works with general multi-electronic basis function, in principle. A CI method enters if one uses the one-electron function model to construct some of them. The concept of nodal envelope states is introduced to bridge two or more I-frame subsystems and put them in one I-frame case. Thus, when going back to the case $H_{op}(\mathbf{q}(\mathbf{X});\mathbf{X})$ some new constraints must be included in order to get a meaningful scheme. The selection of CI-configurations must include such constraints. Once we do that one come near a CI method.

Also, there is an issue concerning molecular shape and Pauling's hypothesis. A molecular species in the ground electronic state that displays a closed shell structure with centers that are not bonded would correspond to anti-bonding states. One may expect that an equilibrium geometry reflects maximum distances between anti-bonding centers thereby minimizing repulsive energy. A process bringing the structure bent out from equilibrium will work against anti-bonding states the work is pay by external sources; the relative energy difference between bonding and anti-bonding states goes down and the system becomes electronically activated. This rationale follows from the present theory. The effect may act synergistically with bond stretching factor.

A caveat: there is no geometry aspects related to a full quantum mechanical representation, only linear superpositions. Structural issues arise in semi-classic schemes only. Semi-classical schemes are models, not approximations to exact schemes. These two levels are not commensurate. The complementary use of such approaches produces a detailed picture of chemical change that no one separately could achieve.

A final important point, the theory of quantum catalysis sketched above implies energy conservation. So the hope of producing more energy that was put into the system is not warranted. Quantum processes are performed in an energy shell.

Readers might have noticed a non-standard handling of quantum mechanical concepts. One blend of quantum mechanics adapted to Physical Chemistry can be found at the direction indicated in ref.[28]. The mathematical apparatus is unchanged but we emphasize QM from within, namely, QM addressing all possibilities a system may have while experimentally a few of them might be expressed;

the non-zero amplitudes characterizing the behavior of the material system in laboratory experiments. For details see [28].

ACKNOWLEDGEMENTS

I dedicate this work to the memory of Prof. Carl-Ivar Brändén a great scientist who introduced us to the quest of enzyme catalytic sources by assigning the task of studying RubisCO's mechanism.
The author is most grateful to all coworkers participating in these enterprises: J Andrés, M. Oliva,V S Safont, V Moliner, H. Fidder for beautiful insights on RubisCO catalytic mechanisms; Prof. G A Arteca for working in developments of the diabatic theory and model applications. Figure 10-1 was adapted from Monica Oliva's thesis work, the present author is most grateful to her.

REFERENCES

1. Tapia O, Fidder H, Safont VS, Oliva M, Andres J (2002) Int J Quantum Chem 88:154–166
2. Arteca GA, Tapia O (2005) J Math Chem 37:389–408
3. Tapia O, Arteca GA (2004) Adv Quant Chem 47:273–289
4. Tapia O (2006) J Math Chem 39:637–669
5. Tapia O (2009) Adv Quant Chem 56:31–93
6. Kolos W, Wolniewicz L (1964) J Chem Phys 41:3663–3673
7. Tapia O, Cardenas R, Andres J, Colonna-Cesari F (1988) J Am Chem Soc 110:4046–4047
8. Jacob O, Cardenas R, Tapia O (1990) J Am Chem Soc 112:8692–8705
9. Andersson I (2008) J Exp Bot 59:1555–1568
10. Tapia O, Andres J (1992) Mol Eng 2:37–41
11. Tapia O, Andres J, Safont VS (1996) J Phys Chem 100:8543–8550
12. Moliner V, Andrés J, Oliva M, Safont VS, Tapia O (1999) Theoret Chem Acc 101:228–233
13. Oliva M, Safont VS, Andrés J, Tapia O (2001) Chem Phys Lett 340:391–399
14. Oliva M, Safont VS, Andrés J, Tapia O (2001) J Phys Chem A 105:9243–9251
15. Born M, Huang K (1954) Dynamical theory of crystal lattices. Clarendon, Oxford
16. Arteca GA, Rank JP, Tapia O (2008) Int J Quantum Chem 108:1810–1820
17. Wood RW (1922) Proc Roy Soc A 102:1–9
18. Kaplan J (1927) Phys Rev 30:639–643
19. Tapia O, Arteca GA (2003) Internet Elect J Mol Des 2:354–474
20. Tapia O, Polo V, Andrés J (2006) In: Julien J-P, Maruani J, Mayou D, Wilson S, Delgado-Barrio S (eds) Recent advances in the theory of chemical and physical systems. Springer, The Netherlands
21. Fischer H, Radom L (2001) Angew Chem Int Ed 40:1341–1371
22. Berson JA (1994) Science 266:1338–1339
23. Schweitzer C, Schmidt R (2003) Chem Rev 103:1685–1757
24. Schmidt R, Bodesheim M (1995) J Phys Chem 99:15919–15924
25. Minaev BF, Ågren H (1997) J Chem Soc Faraday Trans 93:2231–2239
26. Oliva M, Safont VS, Andrés J, Tapia O (2001) J Phys Chem A 105:4726–4736
27. Tapia O, Andrés J, Safont VS (1994) J Chem Soc Faraday Trans 90:2365–2374
28. Tapia O (2009) Quantum physical chemistry. E-book. To be found at e-address: http://www.uu.se/Personal/tapia-orlando/index.shtm

CHAPTER 11

MOLECULAR DYNAMICS SIMULATIONS: DIFFICULTIES, SOLUTIONS AND STRATEGIES FOR TREATING METALLOENZYMES

SÉRGIO F. SOUSA, PEDRO A. FERNANDES, AND MARIA JOÃO RAMOS
REQUIMTE, Departamento de Química, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 687, 4169-007 Porto, Portugal
e-mail: mjramos@fc.up.pt

Abstract: The application of molecular dynamics (MD) simulations is now firmly established as a strategy to help understanding the activity of biological systems, being routinely applied to investigate the structure, dynamics and thermodynamics of biological molecules and their complexes. Commonly available biomolecular force fields like AMBER, CHARMM, OPLS, and GROMOS contain sets of molecular mechanical parameters for the 20 natural amino acid residues and for a limited set of additional structural elements, allowing accurate MD simulations to be performed for a vast number of proteins and enzymes that are composed simply by such elements. For more diverse biological systems, such as those containing covalently bound metal atoms, no parameters are normally available to describe the specific interactions formed between the metal and the amino acid residues, limiting in practice the direct application of MD simulations to the study of these systems. This chapter presents the typical difficulties normally encountered when trying to run MD simulations on a metalloenzyme, and introduces common solutions and strategies to circumvent these problems, illustrating also the wide range of catalytic relevant properties that can be obtained from such simulations. The zinc metalloenzyme farnesyltransferase is used to illustrate these aspects.

Keywords: Metals, Metalloproteins, Simulations, Bonded-potential, Parameterization

11.1. INTRODUCTION

The existence of active-site metal ions is a feature that is common to a large number of biologically relevant enzymes. For many of these systems, the presence of the metal atoms is essential for the catalytic activity displayed. While metal atoms ensure a rich and functionally diverse chemistry, their presence brings increasing

P. Paneth and A. Dybala-Defratyka (eds.), *Kinetics and Dynamics*,
Challenges and Advances in Computational Chemistry and Physics 12,
DOI 10.1007/978-90-481-3034-4_11,

difficulties to a number of computational techniques normally applied in the study of biological systems. Among these are molecular dynamics (MD) simulations.

The dynamic properties of an enzyme have a profound effect upon its functional behaviour. MD simulations allow the study of the dynamic properties of a system. They enable the complex and dynamic processes that take place in biological systems to be analyzed in detail, and provide the ultimate detail concerning the individual particle motion as a function of time. MD simulations can provide particularly useful information regarding the conformational sampling of both the enzyme and of its substrates or inhibitors, and most notably on their interaction. In addition, MD simulations can enable the dynamic effect of the solvent in such processes to be evaluated atomistically. Other examples of application of MD simulations include the study of phenomena such as protein stability, molecular recognition, conformational changes, protein folding, and ion transport in biological systems [1, 2]. MD simulations are also frequently applied in drug design and in NMR and X-ray structure determination.

In the molecular dynamics (MD) approach, the evolution of a biomolecular system with time is determined. MD generates successive configurations of the system (trajectory) by integrating Newton's laws of motion.

$$\frac{d^2x_i}{dt^2} = -\frac{F_{x_i}}{m_i} \tag{11-1}$$

This equation determines the trajectory of a given particle of mass m_i along one coordinate (x_i), with F_{xi} representing the force acting on the particle in that direction. From the knowledge of the force on each atom, the acceleration of each atom in the system can be determined. By integrating Newton's equations of motion, one gets a trajectory describing the positions, velocities, and accelerations of the several particles that constitute the system, as they vary with time. The average values of a multitude of properties can then be determined from this trajectory. The method is deterministic, in the sense that once the positions and velocities of each particle are known, the state of the system can be predicted at any time, both in the future and in the past.

In MD, the potential energy of a system is given by a molecular mechanical force field. Molecular mechanical (MM) methods are based on a rather different paradigm from quantum mechanical schemes. Instead of trying to solve the electronic Schrödinger equation, MM methods bypass it, writing the energy of the system as a parametric function of the nuclear coordinates. MM methods neglect both electrons and the quantum aspects of the nuclear motion, and are based on classical Newtonian mechanics. They consider a rather simplified scheme of the interactions within a system. A "ball and spring" model is normally employed, in which the atoms are described as charged spheres of different sizes, whereas the bonds are described as springs with a different stiffness. The neglect of the concept of electron forecloses any direct study of processes involving the formation or breaking of chemical bonds.

The energy of the system is split into a sum of contributions from different processes, including the stretching of bonds, the opening and closing of angles, rotations around simple bonds, etc. Each of these contributing processes is described by an individual expression, parameterized for a given set of standard atoms types. Hence, a MM method is characterized not only by its functional form, but also by the corresponding parameters, the two of which form the single entity termed force field. The parameters involved are typically derived from experimental data or from calculations with higher level methods (e.g. HF, DFT) for small molecules. The accuracy of the parameterization protocol is of paramount importance to the reliability of the force field, and special care should be taken when calculating properties other than those included in the parameterization process.

11.2. BIOMOLECULAR FORCE FIELDS

The development of a general force field, able to yield accurate results for a plethora of chemically different compounds is a particularly hard, complex and ungrateful task. To obtain high accuracy calculations on such chemically different compounds, a careful parameterization of an extremely diverse and complete set of reference molecules is required. This is, in practice, an impossible mission. Hence, it is not surprising that currently available general force fields had to sacrifice accuracy for a wider applicability. Improved quality is normally achieved by developing specialized force fields, ensuring accurate calculations to be performed, albeit in a much more limited class of compounds.

The limited structural diversity, in terms of building blocks, that characterizes most biological systems of relevance, including proteins, carbohydrates, and nucleotides, renders the development of specialized force fields for each one of these large and important classes of biological macromolecules a particular interesting and valuable strategy, with an almost infinite number of applications, given the large number of combinations of the correspondent biological structural basic elements that can be found in nature.

AMBER, CHARMM, and OPLS are the three most popular molecular force fields devised to describe proteins. A common characteristic to these force-fields is that the potential energy function is a function of pairs of atoms (it is two-body additive). Most force-fields used in biological simulations apply the same form for the energy function, with harmonic terms for bonds and angles, Fourier series for torsions, and pairwise van der Waals and Coulombic interactions between atoms that are separated by three or more bonds. However, they are parameterized in conceptually different ways. Hence, individual parameters from different force fields should not be compared, as the parameterization scheme varies from force-field to force-field. Comparisons have to focus on the ability to reproduce observable data for a given system. Such a comparison has been conducted for the alanine tetrapeptide, using high-level quantum chemical calculations and considering 20 different force-fields [3]. Small inconsistencies in all of the force fields have been reported, and no force-field significantly outperformed the others in any aspect.

In this section, the basic principles associated to AMBER, CHARMM, and OPLS are presented and discussed.

11.2.1. AMBER

The Assisted Model Building and Energy Refinement (AMBER) force field was originally designed for the calculation of proteins and nucleic acids by Kollman and co-workers [4–8]. The designation "AMBER force field" actually refers to the energy function considered by the family of AMBER force fields. In its original formulation the AMBER force field considered an united atoms representation [4], i.e. the hydrogen atoms were assigned to the directly connecting non-hydrogen atoms, with the parameters of the latter including the effect of the first, greatly reducing the number of parameters required. An all-atom force field was later introduced [5], yielding significantly better results in the treatment of both proteins and nucleic acids. Presently, the family of AMBER force fields includes several parameterization schemes designed to treat peptides, proteins, and nucleic acids (ff94, ff96, ff98, ff99, ff99EP, ff02, ff02EP, ff03) [6, 8], the Generalized AMBER force field (GAFF) [7] which contains parameters for small organic molecules, and the GLYCAM force fields (GLYCAM2000, GLYCAM04, GLYCAM04EP) developed for carbohydrates.

In the AMBER force field, stretching and bending terms are calculated considering the harmonic approximation, using only squared terms, whereas torsional energies are determined by a Fourier series using the first six terms. Van der Waals interactions are approximated by a Lennard-Jones 12–6 formula, with the 1–4 interactions scaled by a factor of 0.5. A Coulombic formula was considered for the calculation of the electrostatic interactions, allowing the use of a constant or of a distance-dependent dielectric value. The way the atomic charge was calculated varied according to the force-field version considered. For most AMBER force fields, partial atomic charges have been determined for biological and organic molecules using the restrained electrostatic potential methodology [9, 10] at the HF/6-31G(d) level of theory. The AMBER energy function may be written as:

$$U(\vec{R}) = \sum_{bonds} K_b(b - b_0)^2 + \sum_{angles} K_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} \frac{V_n}{2}(1 + \cos(n\varphi - \delta)) + \sum_{nonbond} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\varepsilon_1 R_{ij}}\right] \quad (11\text{-}2)$$

where K_b, K_θ, and V_n are the bond, angle, and dihedral angle force constants, respectively; b, θ, and ϕ are the bond length, bond angle, and dihedral angle, with the subscript zero representing the equilibrium values for the individual terms; γ is the phase angle and takes values of either 0° or 180°; A_{ij} and B_{ij} represent respectively the van der Waals and London dispersion terms, q_i and q_j represent the partial atomic charges; ε is the dielectric constant that takes into account the effect of the medium. In this equation, the nonbonded terms are calculated for all atom pairs that are either separated by more than three bonds or not bonded.

11.2.2. CHARMM

The Chemistry at HARvard Macromolecular Mechanics (CHARMM) force field was initially designed for the study of proteins by Karplus and co-workers [11], and is based on a rather similar concept to that of AMBER. In fact, both were parameterized for high-quality computations on a limited set of molecules, and the mathematical formalism of the energy terms in CHARMM is also extremely similar to that of the AMBER force field. The main difference lies in the parameters employed and their derivation.

There are basically three CHARMM force fields designed for proteins. These are the united-atom CHARMM19, the all-atom CHARMM22 [12], and the dihedral potential corrected variant CHARMM22/CMAP [13]. The most frequently used CHARMM force field for protein simulations is CHARMM22. In this force field the atomic charges were derived from quantum chemical calculations at the 6–31G (d) level of the interactions between model compounds and water. CHARMM27 was designed for DNA, RNA, and lipids [14]. The combination of different CHARMM force fields is possible. For example, CHARMM22 and CHARMM27 may be combined for the study of systems involving protein–DNA interactions. Additional parameters for species such as NAD+, sugars, etc. may also be downloaded. These force field version numbers refer to the CHARMM version in which they were included for the first time. However, it is possible to use a more recent version of the CHARMM program with each one of these force fields. In addition, these force fields can also be used within several other molecular dynamics programs that support them.

The equation below presents the energy function considered in CHARMM. In this function, stretching and bending terms are calculated by a harmonic quadratic function, with the torsional term also described by a Fourier series. Van der Waals interactions are calculated by the Lennard-Jones 12–6 formula, whereas electrostatic interactions are described by a Coulomb equation. The CHARMM energy function also includes an Urey–Bradley term to describe 1,3-non-bonded interactions. This term is important for the in-plane deformations, as well as for separating symmetric and asymmetric bond stretching modes (e.g., in aliphatic molecules). The CHARMM energy function has the following form:

$$\begin{aligned} U(\vec{R}) = & \sum_{bonds} K_b(b-b_0)^2 + \sum_{UB} K_{UB}(S-S_0)^2 + \sum_{angles} K_\theta(\theta-\theta_0)^2 \\ & + \sum_{dihedrals} K_\chi(1+\cos(n\chi-\delta)) + \sum_{impropers} K_{imp}(\phi-\phi_0)^2 \\ & + \sum_{nonbond} \varepsilon\left[\left(\frac{R_{\min_{ij}}}{r_{ij}}\right)^{12} - \left(\frac{R_{\min_{ij}}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{\varepsilon_1 r_{ij}} \end{aligned} \tag{11-3}$$

where K_b, K_{UB}, K_θ, K_χ, and K_{imp} are the bond, Urey–Bradley, angle, dihedral angle, and improper dihedral angle force constants, respectively; b, S, θ, χ and φ are the

bond length, Urey–Bradley 1,3-distance, bond angle, dihedral angle, and improper torsion angle, respectively, with the subscript zero representing the equilibrium values for the individual terms. ε is the Lennard-Jones well depth and R_{min} is the distance at the Lennard-Jones minimum, q_i is the partial atomic charge, ε_1 is the effective dielectric constant, and r_{ij} is the distance between atoms i and j. The Lennard-Jones parameters between pairs of different atoms are obtained from the Lorentz–Berthelodt combination rules, in which ε_{ij} values are based on the geometric mean of ε_i and ε_j and $R_{min,ij}$ values are based on the arithmetic mean between $R_{min,i}$ and $R_{min,j}$ [12].

11.2.3. OPLS

The Optimized Potential for Liquid Simulations (OPLS) force field was designed for the study of proteins and nucleic acids by Jorgensen and co-workers [15–17]. Following the publication of an united atoms force field (OPLS-UA) [15], an improved all-atom version, termed OPLS-AA, was also published, yielding considerable better results [17]. This force field was based on the same type of basic principles illustrated for AMBER and CHARMM. The parameters and expressions for the stretching, bending and torsional terms are basically the same considered in AMBER. In fact, bond stretching and angle bending parameters were taken from the AMBER all-atom force field, whereas the torsional parameters were determined by fitting to rotational gas-phase energy profiles obtained from ab initio molecular orbital calculations at the RHF/6-31G* level of theory for more than 50 organic molecules and ions. The main difference lies in the way the parameterization of the non-bonding terms was carried out. In OPLS the nonbonded parameters were carefully developed from Monte Carlo statistical mechanics simulations by computing thermodynamic and structural properties for 34 pure organic liquids. Examples included alkanes, alkenes, alcohols, ethers, acetals, thiols, sulfides, disulfides, aldehydes, ketones, and amides. Hence it is not surprising that OPLS tends to yield better results than AMBER or CHARMM in simulations with explicit solvent or in cases where the non-bonding interactions are of particular importance. The energy function considered in the OPLS force field may be written as:

$$
\begin{aligned}
U(\vec{R}) = & \sum_{bonds} K_r (r - r_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 \\
& + \sum_{dihedrals} \left(\frac{V_1^i}{2} (1 + \cos(\varphi_i - f_i 1)) + \frac{V_2^i}{2} (1 - \cos(2\varphi_i + f_i 2)) \right. \\
& \left. + \frac{V_3^i}{2} (1 + \cos(3\varphi_i - f_i 3)) \right) \\
& + \sum_{nonbond} \left[4\varepsilon_{ij} \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j e^2}{r_{ij}}
\end{aligned}
\tag{11-4}
$$

where K_r, and K_θ are the bond, and angle force constants, respectively; r and θ are the bond length and bond angle, with the subscript zero representing the equilibrium values for the individual terms; ϕ is the dihedral angle, V_1, V_2, and V_3 are the coefficients in the Fourier series for each torsion, and f_i1, f_i2, and f_i3 are the correspondent phase angles; ε is the Lennard-Jones well depth, σ is the finite separation at which the Lennard-Jones potential is zero, q_i is the partial atomic charge, and r_{ij} is the distance between atoms i and j.

11.3. DIFFICULTIES IN TREATING A METALLOENZYME

From a computational point of view, metalloenzymes are particularly challenging. In fact, despite the rich chemistry that traditionally characterizes such systems, the presence of a metal atom renders particularly troublesome the use of some well established computational methodologies. In particular, the application of MD methodologies to metalloenzymes offers much greater difficulties than those normally presented by enzymes composed simply by standard amino acids residues. The computational treatment of metalloenzymes is therefore still, in general, a few steps behind that of the non-metalloenzymes. Hence, despite the very vivid chemistry that these enzymes traditionally present, many researchers still tend to avoid metalloenzymes, as to keep away from the technical and computational hurdles associated.

The biggest problem with performing MD simulation on metalloenzymes is the lack of accurate parameters in the corresponding force fields able to describe the interaction of the metal atom with the remaining of the protein. While several different strategies have been developed, the alternatives available typically involve making several non-trivial choices, which could bring important limitations to the study of a given problem. A good knowledge of the potential implications of such choices is hence paramount. In addition, the lack of well-established automated parameterization procedures for the metal coordination spheres in metalloenzymes, in contrast to what happens for substrate molecules for example, further complicates the treatment of Metalloenzymes in MD simulations. The user becomes the sole responsible for all the stages in the parameterization protocol.

11.4. PARAMETERIZATION STRATEGIES FOR METALLOPROTEINS

With the widespread application of MD simulations, three general approaches, at different levels of sophistification, have been developed to describe metal atoms and their interaction with the protein environment in molecular mechanical force fields. The first and the simpler of these approaches is normally referred to as the non-bonded model, and describes the metal ligand interactions through only electrostatic and van der Waals energy terms. A second scheme, termed bonded-model, includes also well-defined covalent bonds between the metal atom

and its ligands, thereby maintaining the metal's coordination number during the MD simulation. Finally, a third scheme, usually designated as the cationic dummy atom approach, has emerged during the past decade as an alternative to the two approaches indicated above, in an attempt to get the best of the two worlds. These three parameterization schemes, their advantages and pitfalls, are the subject of this section, summarized in Table 11-1. In addition, a brief discussion about the use of alternative simulation schemes, outside the classical MD simulation framework is also presented, with major emphasis being given to the use of combined QM/MM MD simulations.

11.4.1. The Non-Bonded Model Approach

To use a non-bonded potential is normally regarded as the first and most simple approximation to treat metal atoms in MD simulations. With this model, the metal–ligand interactions are described through only electrostatic and van der Waals energy terms, allowing in practice the formation and breakage of new interactions between metal and ligands [18].

The absence of direct and well defined coordination accounting for the metal–ligand interactions, greatly simplifies the number of parameters that have to be specified to handle the metal atom in MD simulations. However, such schemes have normally extreme difficulties in maintaining the overall topology of metal coordination sphere in MD simulations. For example, in Zn proteins, a long-standing problem with MD simulations, well described in the literature [19, 20], is that the typical four-ligand coordination of the metal atom is changed to six during the simulation. This is due to the fact that the van der Waals parameters normally available for Zn were actually parameterized for Zn(II) in aqueous solution, a medium in which this cation is hexacoordinated [21, 22].

Significant improvements can normally be obtained by a careful determination of van der Waals parameters, for both the metal atom and the directly coordinating atoms, and by recalculating the electrostatic parameters for the full complex. Nevertheless, most studies typically consider the amino acid ligands as retaining their charges as in the original biomolecular force field, with the positive charge of the cation (e.g. +1, +2, +3) totally concentrated in the metal atom [23]. While this is surely the simplest way to perform MD simulations in metalloproteins, as all the parameters can be taken as transferable, very dramatic distortions on the metal spheres are normally observed [23]. In fact, in addition to the difficulties in maintaining a reasonable coordination number for the metal complex, significant ligand swap tends to take place, with first coordination sphere neutral ligands being replaced by second and third coordination sphere anionic ligands, greatly distorting the enzyme structure around the active-site.

An easy trick to overcome such difficulties, when just a crude sampling of the dynamic properties of the global enzyme is required, and those regarding the metal coordination sphere can be sacrificed, is restraining the metal–ligand bond

Table 11-1 Summary of the three parameterization strategies normally employed to handle metal coordination sphere in metalloenzymes

	Non-bonded model approach	Bonded model approach	Cationic dummy atom approach
Main characteristic	– Metal atom defined only through electrostatic and van der Waals parameters – The metal formal charge is concentrated at the metal atom	– Bond, angle and dihedral parameters used, in addition to the electrostatic and van der Waals description	– Metal atom is substituted by a set of dummy atoms with partial charge around the metal nucleus, which impose a given coordination geometry to the complex – No bond, angle or dihedral parameters used to describe the metal coordination sphere
Advantages	– Much Simpler – Very small number of parameters to be defined – Allows in practice the formation and breakage of new interaction between the metal atom and ligands	– By far, the most accurate model in representing the metal coordination sphere and in describing the resulting electrostatics – Metal coordination sphere topology and coordination number maintained	– Much smaller number of parameters required than in the bonded model – Coordination geometry of the metal sphere maintained – Allows in practice the formation and breakage of new interaction between the metal atom and ligands
Disadvantages	– Topology and general characteristics of the metal coordination sphere extremely difficult to maintain – Ligand swap of first sphere neutral ligands with second- or third-sphere negative ligands	– Very large number of parameters have to defined, many of which are non-transferable – Requires in practice that each different metal coordination sphere to be specifically parameterized – Does not allow ligand exchange at the metal coordination sphere in the course of an MD simulation	– Ligand swap of first sphere neutral ligands with second- or third-sphere negative ligands – Only works in practice if all the first sphere ligands are anionic and the second sphere alternatives are neutral or positively charged.

distances [24] . Examples of metalloproteins for which the non-bonded model has been applied with success in performing MD simulations are the human carbonic anhydrase I [25], thermolysin [26], IMP-1 metallo-b-lactamase [27], and the histone deacetylase-like protein [28]. However, this approach greatly limits the potential advantages and insights that MD simulations could offer in the description of the region surrounding the metal coordination sphere.

11.4.2. The Bonded Model Approach

The bonded-model approach represents a particularly straightforward and effective way to retain the specific geometry of the metal–ligand coordination environment and of the electrostatics resulting from the metal–ligand interactions [29]. This model considers explicit bond and angle terms to represent the coordination sphere of the metal atom during the MD simulation. It is very effective in preserving the observed geometry of the metal coordination sphere, but is not appropriate for the study of systems in which the coordination number of the metal atom is expected to vary, or in which the metal–ligand atoms are exchanged in the course of a simulation (e.g. water molecules) [22].

In stable metal complexes this approach is by far the most accurate in representing metal coordination spheres and in describing the resulting electrostatics. However, it is also the most difficult one to implement. While the non-bonded model requires only molecular mechanical parameters for the electrostatics and non-bonded energy terms, with the bonded potential, parameters are also required for all bonds involving the metal atom (including equilibrium bond-lengths and force constants), all the angles that involve the metal atom, including those in which the metal is the central atom or a terminal atom (equilibrium bond-angles and force constants), and of course for all the corresponding dihedrals.

The reader should be aware that by simply changing the identity of a ligand at a metal coordination sphere causes many of the other molecular mechanical parameters also to change. Metal–ligand equilibrium bond-lengths and the corresponding force constants for example are very sensitive to the number and type of ligands present at the metal coordination sphere. Even more specific are the metal–ligand bond-angles for which the metal atom is central, which directly depend on pairs of ligands at the metal coordination sphere. Bond-angles for which the metal-atom is terminal can be sometimes taken as transferable between different metal complexes for the same combination metal–ligand, as they involve only the metal-atom and two-atoms in the ligand. Dihedral involving the metal atom are normally set to zero [30, 31] to simplify the number of parameters to define, a strategy that has been proved to be normally safe in the treatment of metal biological complexes [29, 32–36]. Van der Waals parameters are also frequently taken as transferable, while electrostatic parameters are always determined for each specific metal coordination sphere [30, 31].

Taken together, the specificity of each metal coordination sphere, in terms of identity of the central cation, oxidation state, spin state, ligands present, global

charge, etc., and the large multiplicity of possible metal coordination spheres arising from such combinations greatly limits both the development of a general set of parameters and the transferability of parameters between different metal spheres, leaving to the user the heavy burden of parameterization.

In spite of the difficulties inherent to this approach, several MD studies applying this generic methodology have been described in literature. Examples include the metalloenzymes human carbonic anhydrase II [29, 32, 37–39] , alcohol dehydrogenase [40, 41], yeast cytosine deaminase [33], histone deacetylase [34], Zn–beta-lactamase[35, 36, 42, 43], farnesyltransferase [44–49], urea amidohydrolase[50] , adenosine deaminase [51], Zinc and Copper binding Alzheimer's amyloid beta-peptide [52] , and the system Cu(II)-PHGGGWGQ octapeptide [53].

11.4.3. Cationic Dummy Atom Approach

A relatively simple and ingenious attempt to analyze metalloenzymes by using MD simulations is the use of the cationic dummy atom approach (CDAA), also called semi-bonded model. The CDAA method does not take into account covalent bonds or harmonic constraints at the metal complex, limiting in practice the number of parameters that have to be defined in relation to the more typical bonded-potential method. Instead, in this method the metal atom is substituted by a set of cationic dummy atoms placed around the metal nucleus, imposing a given coordination geometry to the complex, in agreement with the particular nature of the complex and of the metal atom. For example, for a standard Zn metalloenzyme, with a tetracoordinated tetrahedral coordination geometry, like the enzyme farnesyltransferase, the Zinc atom would be replaced by four cationic dummy atoms following a tetrahedral arrangement. The charge of the metal atom (e.g. 2+ for Zn^{2+}) would be distributed equally through these dummy atoms (four dummy atoms with a partial charge of +0.5 each for the example given), whereas van der Waals parameters and mass would be attributed solely to a pseudo-metal atom located at the centre of the referential and having no charge.

Examples of application of this methodology include the Mn(II) ions in Staphylococcal nuclease [54], several Zn(II) biological systems like carboxypeptidase A [55], carbonic anhydrase II [55], rubredoxin [55], farnesyltransferase [19], phosphotriesterase,[20] the IMP-1metallo-beta-lactamase [27, 56] and the zinc endopeptidase of botulinum neurotoxin serotype A [57]. More recently this approach has also been used to represent with success he two octahedrically coordinated bivalent Magnesium ions at the active site of the enzyme DNA polymerase β [58].

Success using this method is naturally limited by the drastic approaches considered [19, 20, 27, 55]. This method typically requires that the ligands in the first coordination sphere are deprotonated to negatively charged species not to lose contact with the dummy atoms introduced. In addition to this, the negatively charged second-sphere ligands normally have to be protonated as to avoid attack at the charged dummy atoms. In addition, to its conceptual simplicity the CADD

methods has the advantage not to impose a specific metal–ligand coordination, allowing ligand exchange at the metal coordination sphere in the course of a MD study, and can therefore be of great usefulness in the study of particular problems involving metalloenzymes, particularly when a negative protonation state can be reasonably assigned to the first sphere ligands.

11.4.4. Alternative Formulations

In addition to the three main MD strategies outlined above, which represent the mainstream approach in MD simulations and may be grouped under the generic designation of classical MD simulation, an alternative scheme has been applied in some studies to treat metalloenzymes: the combined quantum mechanics/molecular mechanics molecular dynamics (QM/MM MD).

The use of time-independent QM/MM methods in the study of metalloenzymes, is now well-established within the computational biochemistry field, particularly in computational enzymology studies [59–64]. The basic philosophy behind modern QM/MM methodologies has roots that trace back to the 1970s, in particular to the pioneer work by Warshel and Levit [65], that paved the way for the methodological developments that would come in the following decades. Generally speaking, QM/MM calculations admit that chemical systems of big dimensions, including the ones of biomolecular nature, can be divided into an electronically important region that requires a quantum chemical treatment (the quantum mechanical region or QM) and a surrounding region that admits a classical mechanical treatment (the molecular mechanical region, or MM). The QM region comprises the part of the system directly involved in a given chemical reaction under study, whereas the MM portion encompasses the remaining of the system, where events like the formation and breaking of bonds do not take place. QM/MM methods hence try to ally the accuracy of a QM method to the low computational cost characteristic of MM methodologies.

Combined QM/MM MD studies use the same underlying philosophy of a QM and a MM region of the time-independent QM/MM methods, but they tackle the problem from a dynamic view, and are emerging as a young but very promising alternative to the classic MD simulation methods. Generally these methods are used to sample the configuration space in order to calculate statistical thermodynamic ensemble averages, allowing free-energy differences to be estimated. Examples include reaction, activation and solvation free energies. In addition, these methods allow in principle that events such as ligand exchange and chemical reactions taking place at the metal coordination sphere to be analyzed. However, a drawback of these methods is the very high computational effort required for meaningful sampling, a difficulty that frequently causes that the QM part to be evaluated only at the semi-empirical level, with significant loss of accuracy. Examples of combined quantum mechanics/molecular mechanics molecular dynamics methods include the Born–Oppenheimer molecular dynamics [66, 67], the Car–Parrinello molecular

dynamics [68], and the QM/MM-FEP methods (or QM/MM-FE) [69–72]. A good introduction to such schemes may be found in the literature [61].

11.5. FARNESYLTRANSFERASE AS A TEST CASE

This section illustrates, in a step-by-step approach, the several tasks required to set up a classical MD simulation for a typical metalloprotein. Farnesyltransferase, a Zn enzyme, is used as an example.

11.5.1. The Target Protein

The enzyme Farnesyltransferase (FTase) is an important Zn enzyme that catalyzes the transfer of a 15-carbons isoprenoid farnesyl group from farnesyldiphosphate (FPP) to peptide substrates containing a characteristic carboxyl-terminal CAAX motif, where C is the cysteine that is farnesylated, A is an aliphatic amino acid, and X is the terminal amino acid, normally methionine, serine, alanine, or glutamine [73]. In the early 1990s, the finding that the farnesylation reaction was critical for the ability of oncogenic forms of Ras proteins to transform cells [74] turned the spotlight to this zinc enzyme, as mutated Ras proteins have been implicated in ca. 30% of all human cancer [75].

The first X-ray crystallographic structure of this enzyme was published in 1997, providing a first rough glimpse of the active site structure and of the metal coordination sphere [76]. A vast number of kinetics, mutagenesis, and most notably EXAFS studies [77–86] were published in the following years [77–86], together with new X-ray structures for the several intermediates formed during the catalytic mechanism of this enzyme (FTase–FPP, FTase–FPP–CAAX, FTase–product) [76, 87]. Even though these studies have provided a large body of experimental data regarding the catalytic mechanism of this enzyme, several inconsistencies and apparently contradicting experimental evidence arise [88]. Many of these doubts involved the very nature of the catalytically crucial Zn coordination sphere.

Some years ago, we embarked in a series of detailed quantum mechanical (QM) studies on the several Zn coordination spheres formed during the catalytic mechanism of FTase [89–92], trying to reconcile the apparently contradicting experimental evidence arising from the X-ray crystallographic structures [76, 87, 93–97], extended X-ray absorption fine structure (EXAFS) [84] results, and kinetic and mutagenesis data [80, 82, 98–104]. These computational studies allowed us to draw a detailed experimental/computational unified atomic level portrait of the several metal coordination spheres existing at each state in the catalytic mechanism of FTase (Figure 11-1), and served as the basis for subsequent mechanistic studies on the reaction processes connecting the several minima [44, 49, 105, 106].

The presently accepted view of the catalytic mechanism of FTase, displayed in Figure 11-1, admits that the enzyme exists at the resting state as an equilibrium of two states at a very close energetic proximity: a state in which the zinc ion is

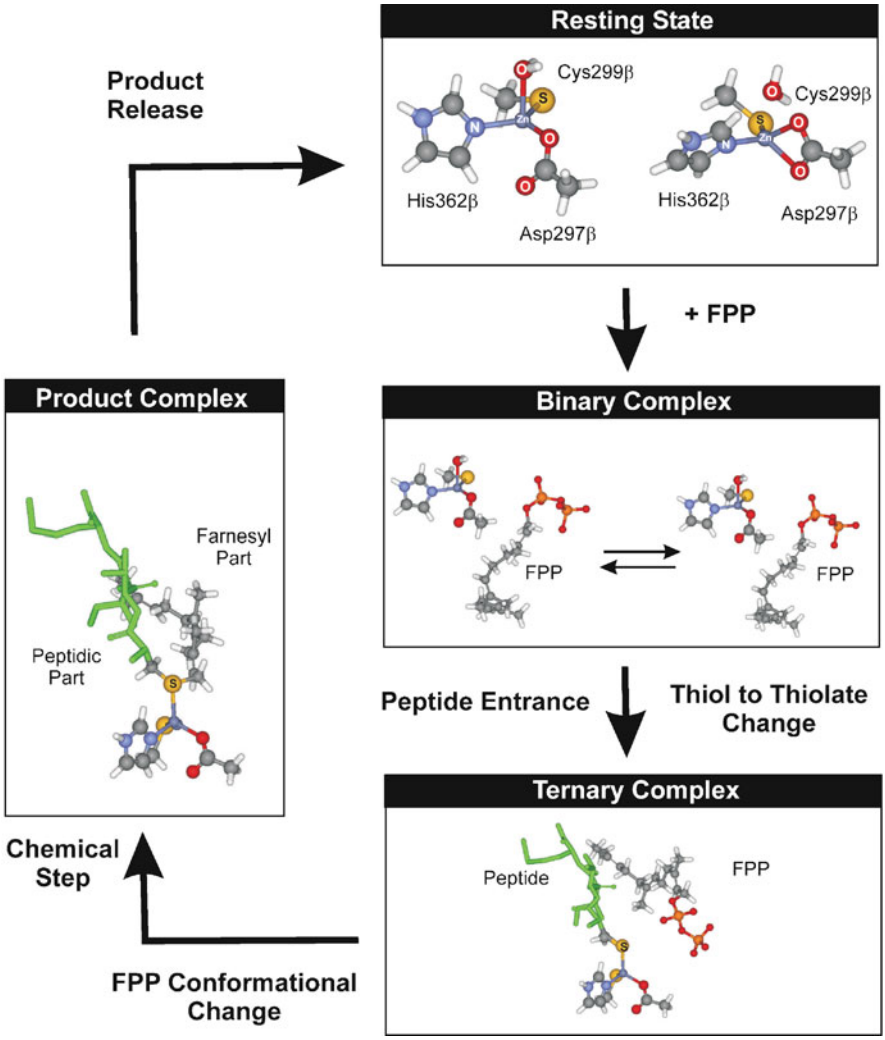


Figure 11-1. Schematic representation of the catalytic mechanism in the reaction catalyzed by the enzyme farnesyltransferase, illustrating the several Zn coordination spheres formed during catalysis

coordinated in a monodentate fashion to residue Asp297β, and further interacts with Cys299β, His362β, and with a water molecule, defining a distorted tetrahedron; and a second state where the zinc–water bond is replaced by a second interaction with Asp297β, which binds in a bidentate fashion, with the metal coordination sphere retaining a distorted tetrahedral geometry. These two states are in equilibrium through a change from a monodentate coordination mode to bidentate by Asp297β, in a process called carboxylate-shift [107]. The ability of the carboxylate group to rearrange in such a manner allows a constant or nearly

constant coordination number to be maintained throughout an entire catalytic pathway of a given enzyme, providing an efficient mechanism to lower the activation barriers for ligand exit or entrance processes, and has been studied in detail for several general zinc coordination environments [108].

FPP is the first substrate to bind FTase [101] and leaves the metal coordination sphere unchanged. Only then does the CAAX protein substrate binds [101], with the CAAX cysteine sulphur atom directly coordinating zinc [109] and locking the Asp297β residue into the monodentate state [89]. Experimental evidence [109, 110] and theoretical results [89] suggest that this cysteine residue initially coordinates the metal atom as a thiol, and only subsequently loses the proton to a base in the active site to form a more stable zinc-bound thiolate. This process leaves the two reactive atoms – the C1 atom of FPP and the zinc bound cysteine's sulphur – something like 8 Å away [95–97], a distance that must be reduced for the reaction to take place. The FPP substrate then suffers a major rearrangement through a rotation of the first two isoprenoid subunits (i.e. from C1 to C10) [82, 87, 91], bringing the two reactive atoms to a distance of around 5–5.5 Å at which the chemical step can take place.

The reaction then proceeds through an associative mechanism (S_N2 like) but with an unusual partially dissociative character, i.e. with the transition state intermediate displaying significant C1–O bond cleavage, and very modest C1–S bond formation, and with the C1 of FPP having carbocation character and the diphosphate leaving group having negative charge [82, 86, 105, 106, 111]. A magnesium ion is thought to play a critical role in this step by stabilizing the negative charge build-up on the diphosphate moiety and activating the leaving group [105, 110]. Globally, although magnesium is not absolutely required for FTase activity it has been shown to enhance the rate constant for product formation by 700-fold [86]. However, the exact binding site for this ligand is still the subject of debate [83, 88, 105, 110].

From this step results a farnesylated thioether product that is still coordinated (albeit weakly) to the zinc metal atom [87, 92, 112, 113]. The breaking of this bond has been shown to take place with a very low activation barrier that is compensated by a monodentate to bidentate change by Asp297β (carboxylate-shift) [92]. The farnesylated product then adopts a second binding conformation, occupying a hydrophobic groove near the active site [87, 114]. Finally, product dissociation, the rate limiting step, occurs – a process which has been shown to take place only in the presence of additional substrate molecules, with FPP being more effective in this regard than the CAAX peptide or protein substrates [101, 114].

11.5.2. Initial Strategies

While QM studies and in particular ONIOM (QM:QM) studies provided very valuable information regarding the nature of the several Zn coordination spheres formed during catalysis, some of the other problems involving the activity of FTase

required significantly larger models and a very different methodological setting. In particular, we were interested in performing a comparative analysis of several important molecular aspects for which are vital not only motion but also the conformational sampling of both enzyme and substrate as well as their interaction, and especially the effect of the solvent. Therefore we wanted to apply MD simulations to the several FTase catalytic states.

As in the case of many other metalloproteins, the difficulty in performing MD simulations in FTase lies in the treatment of the metal coordination sphere. In this particular case, this problem is further amplified by the fact that the several intermediate states differ in terms of their Zn coordination sphere (Figure 11-1).

Hence, the application of the more simple non-bonded potential seemed the best strategy to start with. However, preliminary tests quickly showed that such approach could not represent, not even roughly, the topology of the Zn coordination sphere. In addition, to the strong geometric distortion of the active site region, several unreasonable events were noticed. In particular, the His residue was quickly replaced by two Asp residue from the second coordination sphere, and the coordination number increased to six.

As position restraining artificially the metal-bound atoms was not a reasonable option for the type of studies that we intended to perform, we decided to try the cationic dummy atom approach, a method that had been previously applied several years before by Pang et al. in the study of FTase [19]. Using this approach the results were only slightly better than the ones using the non-bonded model. In fact, the tetracoordinated geometry was maintained, but the swap between first sphere neutral residues with second sphere negatively charged residues continued to take place in the course of the simulation. In their study, Pang et al had been able to circumvent this problem by assuming that the His362β residue was in the doubly deprotonated negatively charged state, and that the fourth position in the Zn coordination sphere was occupied by a hydroxide ion (instead of a water model or second carboxylate oxygen) [19].

These assumptions were not reasonable at the time that we started the simulations, taking into consideration the new findings on the catalytic mechanism of FTase that had taken place meanwhile. So, the application of the cationic dummy atom approach was also abandoned. The reader should however be aware, that depending on the particular problem at hand, both the cationic dummy atom approach and the non-bonded model can be of great usefulness to simulate metalloproteins. Only when these approaches fail should one consider moving up the ladder to use the more complex bonded potential.

11.5.3. Setting a Bonded Model Simulation

11.5.3.1. Atom Types

In this particular case, the new parameters for the Zn coordination sphere of FTase are intended to be an extension to the set of AMBER force fields [4, 6, 7]. For this

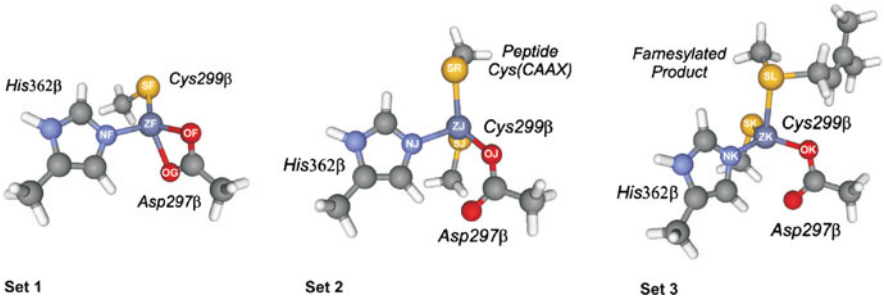


Figure 11-2. Schematic representation of the metal coordination spheres parameterized for FTase, illustrating the nomenclature used for the new atom types added and the type of models used on the quantum mechanical calculations

reason, atom types in conformity with this set of force fields were adopted for all the Zn-bound residues. However, to account for the specificity of the three different Zn coordination spheres, in terms of bond lengths, angles and charges, a total of 15 new atom types were also introduced, in three different sets. These are the Zn and the directly coordinated atoms in the three different environments, as indicated in Figure 11-2. As outlined above, Set 1 represents the active-site Zn coordination sphere of the enzyme resting state and binary complex (FTase–FPP). Set 2 represents the Zn complex in the ternary complex (FTase–FPP–peptide substrate), and Set 3 corresponds to the product complex (FTase–farnesylated product). Depending on the particular force field that the reader intends to use, similar notations and overall procedure may apply.

11.5.3.2. Bond Parameters

The three Zn coordination sphere considered have rather different coordination parameters, depending on the identity of the ligands at the metal sphere. This can be observed not only from the initial X-ray crystallographic structures, but also from the EXAFS [84] and quantum studies [89–92]. Hence, all the bonds involving Zn and all the angles where Zn was the central atom were computationally parameterized for each of the three models.

While some authors tend to use parameters for the metal–ligand bonds obtained from the literature, ideally selected from the most similar metal coordination spheres, we have found that such strategy normally compromises the description of the uniqueness of the metal coordination sphere. It represents, nevertheless, a superior approach to the use of a non-bonded potential or of the cationic dummy atom approach.

Bonding to transition metals is normally described by the same type of potentials used in general purpose force fields, with the harmonic potential being routinely used [30]. However, in some studies the weaker coordinative bonds have been found to require the use of anharmonic potentials for optimal description [30, 115].

For our particular case, we have used the AMBER force field expression, with the harmonic term for bonds given by:

$$E_{Bond} = \sum_{bonds} K_b(b - b_0)^2 \quad (11\text{-}5)$$

Two general approaches can be used to determine K_b e b_0 parameters: the use of experimental structure data and the use of theoretical methods. While the use of experimental structure data (typically X-ray information, from the Cambridge Structural Database) [116] is the preferred choice when general parameters for a enlarged set of similar compounds is searched and enough experimental information is available, it sometimes fails to capture the specificity of a given metal coordination sphere. Hence, for more characteristic cases and for fine-grained parameters the use of high-level theoretical methods is generally preferred. In this regard, density functional theory methods are normally used to generate a potential energy surface (PES), which is then fitted to a harmonic expression, from where K_b is derived.

An elegant method of generating the PES is to take the information from the first and second derivatives directly from the Hessian [117–121]. However, this approach is only valid when the coupling between different normal modes is relatively small for the different bonds that constitute the metal coordination sphere. The more common approach involves performing a systematic variation of each specific bond (SCAN) using a simplified model of the metal coordination sphere.

Initial models for the parameterization process are normally built from the crystallographic structures with the best resolution for the biological metal sphere that one intends to parameterize: in this case 1FT1 (Set1), 1JCR (Set2), and 1KZP (Set3) [76, 87, 96]. Conventional modelling of the amino acid side chains is typically used, that is, for the FTase examples given the zinc ligands aspartate, cysteinate, and histidine are modeled by acetate, methylthiolate, and methylimidazole, respectively, while the Zn-bound farnesylated product from Set 3 was modelled by the use of a methylthiolate bonded to the first isoprenoid sub-unit of FPP. The validity of this type of approaches has been demonstrated with success in the mechanistic study of FTase [89–92], and of several other different enzymes [122–129] and is a standard approach in computational enzymology [64, 130, 131]. Calculations were performed on the active-site models depicted in Figure 11-2, used for the three different sets of parameters using B3LYP/SDD with the Gaussian 03 software package. The performance of the B3LYP density functional and of the SDD pseudopotential basis sets for the treatment of Zn systems have been tested in detail at our group [132, 133], with this combination proving to be an excellent compromise between accuracy and computation time, particularly when comparing with the use of several other density functionals and all-electron basis sets evaluated in several different studies [132–136].

Generically, three types of SCANs can be applied: a rigid scan; free scan; and a semi-rigid scan. It is important to take into consideration that all the most common

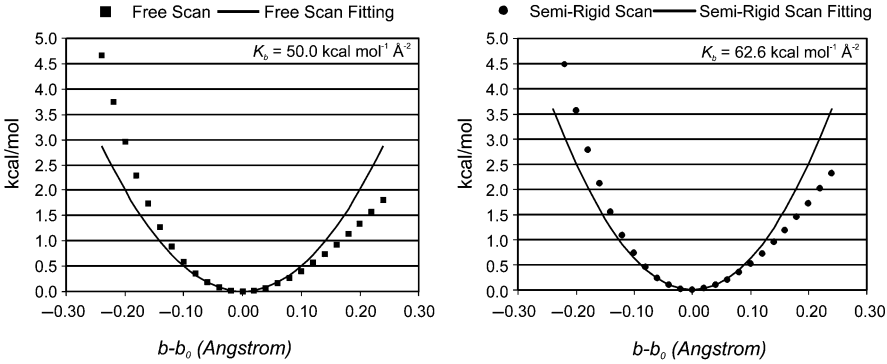


Figure 11-3. Potential energy surfaces for the Zn–Cys bond (parameter set 2) in the free and semi-rigid scans, illustrating also the corresponding harmonic fittings and force constants

biomolecular force fields assume the independent motions of the several bonds, angles, and torsions. In this regard, the free scan could give an inaccurate description of the metal–ligand bonds, as the PES will reflect not only the increase/decrease of a given bond-lengths but also the coordination sphere rearrangement to that change. This would imply for example, that in the specific case of FTase (parameter Set 2), when trying to sample the Zn–Cys bond, the increase of this bond could result in a monodentate to bidentate change by the neighbouring Asp residue, which would therefore contaminate the PES, leading to a non-harmonic behaviour, or at least to softened potential energy surfaces (see Figure 11-3).

The fully-rigid SCAN alternative, which involves constraining all atoms in the model, and simply varies the specific bond under evaluation, is extremely fast from a computational point of view, but gives also a very crude representation of the metal sphere. It is normally not pursued as it is not easily represented by a harmonic potential.

The strategy that we have found to work better, at least for the specific case of FTase, is the use of a semi-rigid scan. Such scans can be performed by gradually changing one of the parameters (positively and negatively) and re-optimizing the model, while keeping the new parameter value fixed and the remaining residues frozen, but both the metal position and the residue directly involved in the bond being evaluated fully free. For example, for the parameterization of the bond Zn–N (His) in model Set 1, both the aspartate and cysteine residues were kept frozen, whereas the zinc atom and the histidine residue were optimized for several Zn–N values around the equilibrium distance (Figure 11-3).

In principle, as the quadratic expression used to fit the PES is symmetric, one would only need to sample one dimension of the PES, i.e. positive for increased bond-lengths from the equilibrium geometry or negative for decreased bond-lengths. This is particularly true as harmonic potentials are known to be able to give a good representation only for relatively small distortions around the equilibrium geometry. However, in the particular case of Zn coordination spheres we have

Table 11-2 Values obtained for the bond term in the different parameter sets

Set 1			Set 2			Set 3		
Bonds	K_b (kcal mol^{-1} Å^{-2})	b_0 (Å)	Bonds	K_b (kcal mol^{-1} Å^{-2})	b_0 (Å)	Bonds	K_b (kcal mol^{-1} Å^{-2})	b_0 (Å)
ZF–OF	62.3	2.094	ZJ–OJ	88.3	1.999	ZK–OK	85.3	1.977
ZF–OG	43.8	2.159	ZJ–SJ	81.2	2.352	ZK–SK	108.3	2.280
ZF–NF	88.3	2.043	ZJ–NJ	62.6	2.129	ZK–NK	84.8	2.055
ZF–SF	115.8	2.258	ZJ–SR	69.3	2.373	ZK–SL	26.6	2.660

found that even for a maximum distance of 0.12 Å around the equilibrium geometry, the positive and negative faces of the PES can differ significantly, yielding very different K_b values. For this reason, the protocol that we recommend for metalloenzymes in general is to sample both sizes of the PES, and fit only later a harmonic potential to the full PES. For accurate sampling we recommend at least 6 points at each side of the PES, at 0.02 Å intervals (see Figure 11-3).

For the equilibrium bond-distances we choose to use the bond-lengths from the freely optimized quantum mechanical active site models, as the corresponding distances in the available X-ray structures (which have only modest quality) have been the subject of intense debate [84, 88, 90] and the EXAFS data, despite very accurate, presents only an average value for the existing Zn–O/N or Zn–S interactions [84] . In general however, both the use of high-quality X-ray structures from the Protein Data Bank [137] or from similar inorganic complexes in the Cambridge Structural Database [116] can be safe alternatives to pursue. A more empirical approach is to use tentative equilibrium bond-length values and to adjust these values by hand in order that the average values from the MD simulations are able to reproduce the experimental values for that system.

Table 11-2 illustrates the differences obtained for the different FTase parameters sets in terms of bonds, illustrating the intrinsic differences of the several metal coordination spheres considered.

11.5.3.3. Angle Parameters

For parameterization purposes, the angle parameters are typically divided into two main subsets: those corresponding to angles in which the metal atom is central; and those in angles in which the metal atom is terminal. While the first strongly depend on the specific ligand arrangement around the metal coordination sphere, the latter can be regarded as more insensitive to the surrounding and more specific of a given pair metal–ligand.

Hence, for FTase the angles for which Zn was a terminal atom were considered independent of the environment and hence transferable, and were obtained from other Zn enzymes in the literature [37, 40]. This strategy can significantly reduce the number of parameters to be determined to only the most crucial.

Table 11-3 Values obtained for the angle term in the different parameter sets

Set 1			Set 2			Set 3		
Angles	K_θ (kcal mol^{-1} rad^{-2})	Θ_0 (°)	Angles	K_θ (kcal mol^{-1} rad^{-2})	Θ_0 (°)	Angles	K_θ (kcal mol^{-1} rad^{-2})	Θ_0 (°)
OF–ZF–OG	318.9	63.5	OJ–ZJ–NJ	2.8	100.2	OK–ZK–NK	18.8	108.2
OF–ZF–NF	16.7	107.3	OJ–ZJ–SJ	23.0	114.0	OK–ZK–SK	15.2	127.6
OF–ZF–SF	16.5	126.6	OJ–ZJ–SR	24.1	115.1	OK–ZK–SL	24.6	92.3
OG–ZF–NF	19.4	97.0	SJ–ZJ–NJ	20.7	110.1	SK–ZK–NK	12.6	118.7
OG–ZF–SF	15.4	136.9	SJ–ZJ–SR	27.3	116.0	SK–ZK–SL	16.8	101.6
NF–ZF–SF	16.2	114.5	NJ–ZJ–SR	39.1	98.8	NK–ZK–SL	22.1	99.1

For the angles where Zn is the central atom the same type of principles illustrated for the treatment of the metal–ligand bond were applied, including the use of semi-rigid scans at the B3LYP/SDD level of theory. In this regard, for the parameterization of the (Cys)S–Zn–N(His) angle in model Set 1, for example, the aspartate residue was kept frozen, while the Zn atom, the cysteine and histidine residues were optimized for several S–Zn–N values around the equilibrium angle. The resulting PES were fitted by a least squares method, as exemplified for the bond term in the previous section, according to the harmonic expression given by the AMBER force field for this term.

$$E_{Angles} = \sum_{Angles} K_\theta(\theta - \theta_0)^2 \tag{11-6}$$

Table 11-3 presents the values obtained for the several angles, illustrating the numerical differences obtained for the different FTase metal coordination spheres considered.

11.5.3.4. Dihedral Parameters

Dihedral parameters are frequently among the most specific and difficult to define parameters in biomolecular force fields. According to the AMBER force field expression, the dihedrals are represented by periodic formulation written in terms of a cosine function given by:

$$E_{dihedrals} = \sum_{dihedrals} \frac{V_n}{2}(1 + \cos(n\varphi - \delta)) \tag{11-7}$$

In this equation Φ is the dihedral angle, i.e. the angle defined between the AB and CD bonds, whereas n is the correspondent multiplicity, i.e. a quantity that gives the number of minimum points in the function as the bond is rotated by 360°. V_n is the correspondent torsional force constant, while δ is the offset.

In addition to the increased complexity associated to this energy term, the large number of dihedrals associated to the addition of a single non-standard atom is also

a problem. For example, the first and most simple FTase coordination sphere – Sphere 1 – requires the definition of a total of 26 dihedrals involving the metal atom. Such high number of dihedrals renders the determination of dihedral parameters for specific metal coordination spheres normally impractical. While when describing conformationally flexible substrate molecules the dihedral parameters are normally of the uttermost importance for an accurate description of the molecule and are typically the subject of most of the parameterization effort, for metal centres that is not the case. In fact, a common approach when dealing with metal centres is that of neglecting all dihedrals which involve the metal atom. This happens because in typical tetrahedral or octahedral arrangements in enzymatic metal complexes, bonds and angles account for most of the energetic contribution in terms of the bonded terms in the force field expression. The values for the force constants associated to the dihedrals is normally almost meaningless. This approximation greatly simplifies the computational treatment of the system, and confers a certain degree of flexibility to the active-site, without compromising the accurate description of the first coordination sphere of the metal. This procedure has been used with success in the treatment of several different enzymes that have a metal atom covalently bonded [29, 32–36], and is normally viewed as a standard approximation when the bonded model is used with this type of systems [30, 31]. Hence, for the treatment of the Zn coordination sphere in all the FTase states considered, all the dihedral parameters involving the Zn–ligand interactions were set to zero.

11.5.3.5. Van der Waals Parameters

The van der Waals energy term describes the non-electrostatic attractions and repulsions between atoms that are not directly coordinated. These interactions are attractive at small distances, but fall rapidly to zero when the interacting atoms are separated by more than a few atoms. For very short distances, i.e. shorter than the sum of the van der Waals radii of two interacting atoms, this interaction is highly repulsive due to the overlap of the electronic density. The van der Waals energy term is normally approximated by a Lennard-Jones 12–6 potential [138, 139], which in the AMBER force field is represented as:

$$E_{vdw} = \sum_{\text{ij}} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} \right] \tag{11-8}$$

In this equation R_{ij} is the distance between the interacting atoms, whereas A_{ij} and B_{ij} are constants given by:

$$A_{ij} = \left(\frac{R_i^*}{2} + \frac{R_j^*}{2} \right)^{12} \sqrt{\varepsilon_i \varepsilon_j}$$
$$B_{ij} = 2 \left(\frac{R_i^*}{2} + \frac{R_j^*}{2} \right)^{6} \sqrt{\varepsilon_i \varepsilon_j} \tag{11-9}$$

in which R_i* and R_j* are the van der Waals radii of the atoms i and j, and ε_i and ε_j are the correspondent Lennard-Jones potential energy depths associated. In effect, A_{ij} determines the degree of "stickiness" of the van der Waals attraction, whereas B_{ij} determines the degree of "hardness" of the atoms. The parameter A_{ij} can be determined from atomic polarizability measurements or from theoretical calculations. B_{ij} can be obtained from crystallographic data, by fitting to reproduce average constant distances from crystallographic structures of different molecules.

The first term on the Lennard-Jones energy function accounts for repulsion, whereas the second term describes the long-range attractions. Even though the inverse sixth power dependence on the attractive term has a theoretical justification, for the inverse 12th power dependence of the repulsive term no theoretical justification exists. Instead, its use is justified only in terms of the computational efficiency it guarantees, as it can be easily computed by multiplying the term r^6 by itself. In fact, the repulsive part of the Lennard-Jones 12–6 potential is too strong to adequately describe the interactions between atoms at large distances. Other formulations have been proposed. In particular, the Lennard-Jones 9–6 potential has been shown to yield better results for organic compounds, albeit with a higher computational cost. Other popular alternative form is the Hill potential [140], but despite known drawbacks, the Lennard-Jones 12–6 potential continues to be used in many force-fields, particularly those devised to deal with large molecular systems, such as proteins.

One way to derive the van der Waals parameters for a given pair is to vary them in a molecular mechanics scan until the energies and geometries computed by using quantum methods are reproduced [23]. However, most biomolecular force fields (including AMBER) have these parameters broken down for each individual atom type. While special care is required in optimizing such parameters in case the simpler non-bonded potential or cationic dummy atom models are employed, for the application of a bonded atom approach (such as the one recommended in this chapter) the use of values taken from the most similar atoms in the force field or from other studies with similar enzymes is normally regarded as the standard approach. In fact, van der Waals parameters are commonly taken as transferable between roughly similar environments.

For FTase, the van der Waals parameters for the Zn atom in the three environments considered were taken from a set of studies with the enzyme alcohol dehydrogenase where a bonded potential was also used [40, 41, 141], while for the atoms directly coordinating to the Zn ion van der Waals parameters were assigned from the closest existing AMBER atom types in the standard AMBER force fields [4, 6]. These are summarized in Table 11-4.

11.5.3.6. Electrostatic Parameters

The electrostatic energy term describes the non-bonding interactions arising from the presence of atomic charges, i.e. from the internal distribution of electrons. From this, positively and negatively charged regions in the molecules result. The most

Table 11-4 Van der Waals parameters considered in the three parameter sets used for FTase

	Set 1	Set 2	Set 3	Van der Waals' parameters	
Element	Atom type	Atom type	Atom type	R_i* (Å)	ε_i (kcal mol^{-1})
Zn	ZF	ZJ	ZK	1.7000	0.0125
O	OF	OJ	OK	1.6612	0.2100
O	OG	–	–	1.6612	0.2100
N	NF	NJ	NK	1.8240	0.1700
S	SF	SJ	SK	2.0000	0.2500
S	–	SR	SL	2.0000	0.2500

common way to represent this effect is by considering a Coulomb expression. According to such scheme atomic charges are assigned to each individual atom, with the electrostatic interactions between different molecules or different parts of a same molecule being calculated from the sum of the interactions between pairs of atoms, computed from the Coulomb equation, as indicated below:

$$E_{elec} = \sum_{ij} \left[\frac{q_i q_j}{\varepsilon_1 R_{ij}} \right] \tag{11-10}$$

The electrostatic energy is then a function of the atomic charges (q_i and q_j), the interatomic distance (r_{ij}), and of the dielectric constant (ε), which accounts for the effect of the surrounding environment.

The main difference in the calculation of the electrostatic energy term between the several available force fields lies in the way the atomic charges are calculated. This can be performed quantum mechanically or semi-empirically. According to the method employed in its determination, several different types of atomic charges are available. Examples include Mulliken charges [142], Natural Bond Orbital (NBO) charges [143], Merz–Kollman charges [144, 145], Restrained ElectroStatic Potential (RESP) charges [9], DelRe charges [146], Gasteiger charges [147], Pullman charges [148], etc.

RESP charges are the default charge approach applied in the AMBER protein force fields. For consistency with the force field that one intends to use, the same type of charges should be applied to the metal atom. It is important to take into consideration that the use of formal charges on the metal ion (e.g. +2 for Zn(II)) would greatly overestimate the electrostatic interaction, and is therefore not usually pursued within a bonded-model approach.

The parameterization of the electrostatic interactions should include the metal and all the directly interacting ligands. Both the General amber force field (GAFF) and the Cornell et al. force field and based alternatives use RESP charges derived at the HF/6-31G(d) level of theory [6]. However, in systems involving metals B3LYP is normally regarded as a superior alternative and is commonly used [33, 42, 149–151], as HF/6-31G(d) generally underestimates much of the charge transfer

Table 11-5 RESP charges for the most relevant atoms in the three parameter sets considered at the B3LYP/6-311++G(3df,2pd) level of theory, illustrating the differences in terms of partial charges between the three Zn coordination spheres

Set 1		Set 2		Set 3	
Atom type	Charge	Atom type	Charge	Atom type	Charge
ZF	0.9091	ZJ	0.8019	ZK	0.4213
OF	−0.6476	OJ	−0.6620	OK	−0.4757
OG	−0.9126	SJ	−0.6497	SK	−0.5959
NF	−0.3587	NJ	−0.0218	NK	−0.0102
SF	−0.6025	SR	−0.6671	SL	−0.3269

effect involving the metal atom [33, 42]. A higher quality basis set is also normally employed to deal with this effect.

For FTase, RESP charges for the three Zn coordination spheres were derived at the B3LYP/6-311++G(3df,2pd) level. The models of the metal coordination spheres were energy-minimized with B3LYP/SDD starting from the X-ray crystallographic structures, an well established combination for the geometric characterization of Zn complexes [132]. Within the RESP methodology only the global charge of the system is restricted to be the net charge of the constituent groups. In fact, in the case of the Zn coordination spheres tested the following contribution were considered: +2 for Zn(II); 0 for His; and −1 for Cys and Asp. In addition, the partial charges for the C_α linking atoms were obtained from the correspondent amino acids in the AMBER ff99 force field and fixed during the RESP fit procedure, admitting that no significant charge alterations due to the metal–ligand interactions take place beyond the C_α atoms. Table 11-5 illustrates for the most relevant atoms the differences in terms of charges that were obtained for the different metal coordination spheres considered.

11.5.4. Validation and Application

The development of the molecular mechanical parameters for a given metalloenzyme is just part of the problem. In fact, one needs also to validate the new parameters added in the context of the biomolecular force field used, and ideally under the conditions in which the MD simulations is going to be performed.

While several different approaches could be outlined in this regard, the most common protocol involves comparing bond-distances, angles, and dihedrals taken from relatively small tentative MD simulations (or from full-length MD production runs) against the geometric parameters taken from X-ray crystallographic or NMR structures, or alternatively against spectroscopic data.

For FTase the three sets of parameters developed were applied to the four key states of the FTase catalytic pathway: FTase resting state, binary complex (FTase–FPP), ternary complex (FTase–FPP–peptide), and product complex (FTase–product). The

FTase resting state and binary complex share the same Zn coordination sphere with a bidentate aspartate residue [80, 84, 90], as parameterized for Set 1. For the FTase ternary complex the parameters described for Set 2 were used, while for the product complex the parameter Set 3 was employed.

The enzyme structures for the validation process were prepared from the crystallographic structures with the best resolution for each state: 1FT1 (Enzyme resting state) [76]; 1FPP (binary complex FTase–FPP) [93]; 1JCR (ternary complex FTase–FPP–peptide) [96]; and 1KZP (product complex) [87]. The AMBER 8.0 [152] molecular dynamics package was used in the entire process. Conventional protonation states for all amino acids at pH 7 were considered. All the hydrogens were added, and Na^+ ions were placed using the Leap program to neutralize the highly negative charges (ranging from −20 to −24) of the systems. The systems were then placed in rectangular boxes containing TIP3P water molecules, with a minimum distance of 15.0 Å of water molecules between the enzyme and the box side.

A 4-stage minimization protocol in which the constraints on the enzyme were gradually removed was applied prior to the MD simulation run [45]. Following a 50 ps equilibration procedure, 500 ps MD simulations were carried out at 300 K, using Berendsen temperature coupling [153] and constant pressure (1 atm) with isotropic molecule-based scaling. More details on the simulation protocol and on the results obtained can be found in the original reference [45].

One important aspect to take into consideration is the simulation time for the validation protocol. For this particular case, the 500 ps simulation time was considered large enough to adequately sample the structural variables that had been parameterized. In fact, the bond lengths and angles possess a fast dynamics and no slow conformational rearrangements are possible to occur in the parameterized region (neither are they expected, based on the available experimental structural data). Frequency calculations at the quantum level on the models of the several coordination spheres considered, performed at the B3LYP/SDD level, have rendered values for the bonds and angles parameterized in this study ranging from 400 cm^{-1} to around 30 cm^{-1}, implying frequency periods of between 0.1 (for bonds) and 1 ps (for the angles). This guarantees that in the simulation time performed each Zn parameter was sampled at least 500 times, with each bond typically sampled around 5,000 times. For metalloenzymes that have other metals, this simulation time can also in principle be taken as a first reference, although the variation of these parameters in longer simulations should also be checked.

Instant values for the bonds and angles parameterized collected at each 0.2 ps were analysed and compared with the corresponding values in all the X-ray crystallographic structures available (a total of 14) and with the results available from EXAFS [84] (Figure 11-4). Globally, a very good agreement was found for these coordination spheres in the MD simulations performed. The average bond-lengths obtained from the MD simulations exhibit a better agreement with the EXAFS data than the X-ray crystallographic results themselves. In addition, a consistent agreement with the values from the higher-resolution X-ray structures was also found, a notable result, particularly taking into consideration that the

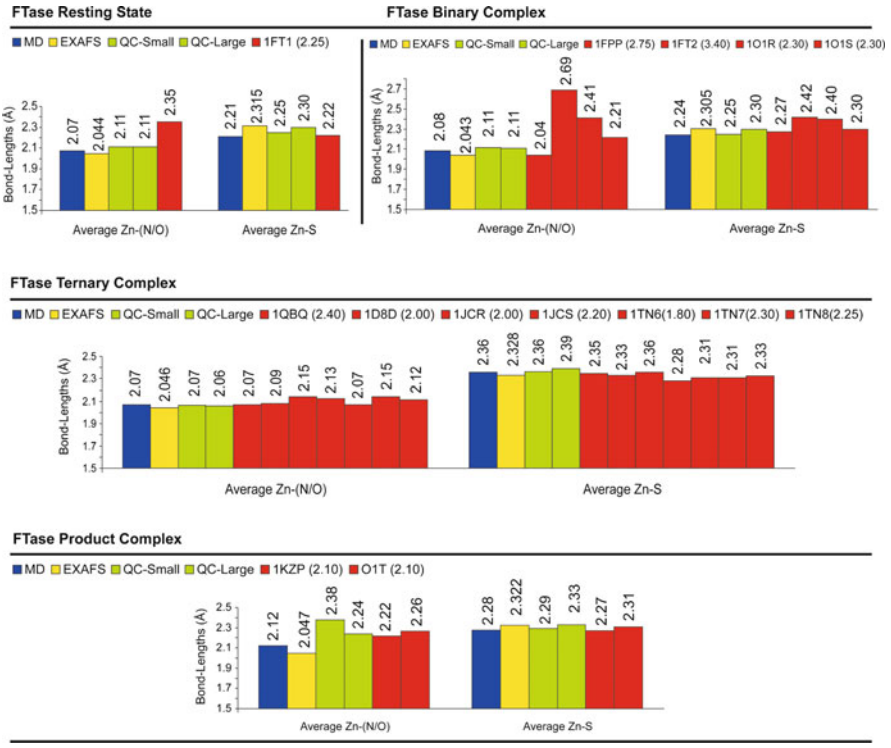


Figure 11-4. Comparison of the results obtained with the average bond-lengths obtained with the MD parameters developed for FTase against EXAFS data, quantum calculations (ranging from a total of 120–133 atoms) models (QC Small and QC Large), and 14 X-ray crystallographic structures (PDB code, resolution indicated in parentheses)

reference values chosen for the several bond-lengths were taken directly from first-principles quantum calculations and not from the experimental data (X-ray or EXAFS) that served as final comparison, or from any empirical fitting performed to improve the accuracy of the results. These observations, justify the use of these parameters with the AMBER force field in the study of FTase in more extensive MD simulations [45].

With the problem of the metal coordination sphere taken care of, with the addition and validation of the new parameters, a wide range of problems that can be studied by MD simulations fall into reach, taking advantage of the robustness of the well-established AMBER force fields in the study of this metalloenzyme, and taking into particular attention the effect of the solvent and of its interaction with the enzyme and substrates. In particular, we have used 10 ns simulation to analyze aspects such as the enzyme flexibility difference patterns on the several different states [46], the behavior of all the most conserved active-site residues [47], radial distribution functions of water molecules around the catalytically relevant Zn ions,

several catalytically relevant active-site distances and of the hydrogen-bonding patterns of the several substrates involved and their conformation [48]. Depending on the problem at hand many other problems can be evaluated with success.

11.6. SUMMARY

The application of MD simulations in the study of metalloproteins is still some steps beyond that of non-metallic biological systems. The lack of suitable parameters able to accurately describe the metal coordination spheres in the most common biomolecular force fields is perhaps the main problem beyond this undisputed reality.

Difficulties in treating metalloenzymes are further amplified by the absence of automated parameterization protocols able to accurately handle such systems in the most modern molecular dynamics software packages, features that are already commonly available for example for typical organic substrates and inhibitors. In metalloenzymes, the heavy burden of parameterizing the metal coordination sphere is hence left almost entirely to the user, with the intrinsic specificity of the coordination spheres greatly limiting parameter transferability.

Despite these difficulties several strategies, at different levels of sophistification, have been developed over the past 15 years. As presented in this chapter the bonded-model approach is clearly the most accurate strategy to handle such systems, with the cationic dummy atom approach and the non-bonded approach being in some cases interesting alternatives to be taken into account, particularly for preliminary work.

In spite of the inherent complexity and significant specificity of the bonded-model approach, several approximations can be now safely used, rendering the parameterization process less intensive and significantly simpler. While several different strategies can be safely pursued to obtain parameters for bonds and angles where the metal atom is central, angles for which the metal atom is terminal are typically taken as transferable and dihedrals involving the metal atom are routinely neglected. Parameter transferability is also a standard practice for van der Waals energy terms, while atomic charges are specific of each metal coordination sphere.

With a bonded-model force field criteriously set, MD simulation can be performed with basically the same level of accuracy of an MD simulation performed on non-metallic proteins (i.e. depending on the biomolecular force field considered), with the molecular dynamics studies on metalloproteins, particularly those that include explicit solvent, offering an even greater range of potential applications than those on the less “chemically-rich” non-metallic systems. Rather than a burden, metal sphere parameterization should be regarded as an opportunity for a detailed reflection about the metalloenzyme that one is studying, particularly on the chemistry surrounding the metal coordination sphere, providing an excellent occasion for linking a quantum mechanical description, with a molecular mechanics and molecular dynamics view, and with X-ray crystallographic and spectroscopic interpretation.

ACKNOWLEDGEMENTS

The authors thank the FCT (Fundação para a Ciência e a Tecnologia) for financial support (PTDC/QUI-QUI/103118/2008).

REFERENCES

1. Karplus M, McCammon JA (2002) Nat Struct Biol 9:646–652
2. Warshel A (2002) Acc Chem Res 35:385–395
3. Beachy MD, Chasman D, Murphy RB, Halgren TA, Friesner RA (1997) J Am Chem Soc 119:5908–5920
4. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S, Weiner P (1984) J Am Chem Soc 106:765–784
5. Weiner SJ, Kollman PA, Nguyen DT, Case DA (1986) J Comput Chem 7:230–252
6. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Fergunson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) J Am Chem Soc 117:5179–5197
7. Wang JM, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) J Comput Chem 25:1157–1174
8. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong GM, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang JM, Kollman P (2003) J Comput Chem 24:1999–2012
9. Bayly CI, Cieplak P, Cornell WD, Kollman PA (1993) J Phys Chem 97:10269–10280
10. Cieplak P, Cornell WD, Bayly CI, Kollman PA (1995) J Comput Chem 16:1357–1377
11. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) J Comput Chem 4:187–217
12. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) J Phys Chem B 102:3586–3616
13. MacKerell AD, Feig M, Brooks CL (2004) J Comput Chem 25:1400–1415
14. MacKerell AD, Banavali N, Foloppe N (2000) Biopolymers 56:257–265
15. Jorgensen WL, Tirado-Rives J (1988) J Am Chem Soc 110:1657–1666
16. Pranata J, Wierschke SG, Jorgensen WL (1991) J Am Chem Soc 113:2810–2819
17. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) J Am Chem Soc 118:11225–11236
18. Stote RH, Karplus M (1995) Proteins 23:12–31
19. Pang YP, Xu K, Yazal JE, Prendergas FG (2000) Protein Sci 9:1857–1865
20. Pang YP (2001) Proteins 45:183–189
21. Pang YP, Perola E, Xu K, Prendergast FG (2001) J Comput Chem 22:1750–1771
22. Sakharov DV, Lim C (2005) J Am Chem Soc 127:4921–4929
23. Vedani A, Huhta DW (1990) J Am Chem Soc 112:4759–4767
24. Li WF, Zhang J, Su Y, Wang J, Qin M, Wang W (2007) J Phys Chem B 111:13814–13821
25. Aqvist J, Warshel A (1992) J Mol Biol 224:7–14
26. Wasserman ZR, Hodge CN (1996) Proteins 24:227–237
27. Oelschlaeger P, Schmid RD, Pleiss J (2003) Protein Eng 16:341–350
28. Yan CL, Xiu ZL, Li XH, Li SM, Hao C, Teng H (2008) Proteins 73:134–149
29. Hoops SC, Anderson KW, Merz KM Jr (1991) J Am Chem Soc 113:8262–8270
30. Comba P, Remenyi R (2003) Coord Chem Rev 238–239:9–20
31. Zimmer M (1995) Chem Rev 95:2629–2649

32. Merz KM Jr, Banci L (1997) J Am Chem Soc 119:863–871
33. Yao L, Sklenak S, Yan H, Cukier RI (2005) J Phys Chem B 109:7500–7510
34. Park H, Lee S (2004) J Comput Aided Mol Des 18:375–388
35. Suarez D, Merz KM Jr (2001) J Am Chem Soc 123:3759–3770
36. Park H, Merz KM Jr (2005) J Med Chem 48:1630–1637
37. Merz KM Jr (1991) J Am Chem Soc 113:406–411
38. Lu D, Voth GA (1998) Proteins 33:119–134
39. Merz KM Jr (1996) J Phys Chem 100:17414–17420
40. Ryde U (1995) Proteins 21:40–56
41. Ryde U (1995) Protein Sci 4:1124–1132
42. Suarez D, Brothers EN, Merz KM Jr (2002) Biochemistry 41:6615–6630
43. Suarez D, Diaz N, Merz KM Jr (2002) J Comput Chem 23:1587–1600
44. Cui G, Wang B, Merz KM Jr (2005) Biochemistry 44:16513–16523
45. Sousa SF, Fernandes PA, Ramos MJ (2007) Theor Chem Acc 117:171–181
46. Sousa SF, Fernandes PA, Ramos MJ (2008) J Phys Chem B 112:8681–8691
47. Sousa SF, Fernandes PA, Ramos MJ (2008) Int J Quant Chem 108:1939–1950
48. Sousa SF, Fernandes PA, Ramos MJ (2009) Bioorg Med Chem 17:3369–3378
49. Cui G, Merz KM (2007) Biochemistry 46:12375–12381
50. Estiu G, Merz KM (2006) Biochemistry 45:4429–4443
51. Coi A, Tonelli M, Ganadu ML, Bianucci AM (2006) Bioorg Med Chem 14:2636–2641
52. Han DX, Wang HY, Yang P (2008) Biometals 21:189–196
53. Riihimaki ES, Martinez JM, Kloo L (2007) J Phys Chem B 111:10529–10537
54. Aqvist J, Warshel A (1990) J Am Chem Soc 112:2860–2868
55. Pang YP (1999) J Mol Model 5:196–202
56. Oelschlaeger P, Schmid RD, Pleiss J (2003) Biochemistry 42:8945–8956
57. Park JG, Sill PC, Makiyi EF, Garcia-Sosa AT, Millard CB, Schmidt JJ, Pang YP (2006) Bioorg Med Chem 14:395–408
58. Oelschlaeger P, Klahn M, Beard WA, Wilson SH, Warshel A (2007) J Mol Biol 366:687–701
59. van der Kamp MW, Mulholland AJ (2008) Nat Prod Rep 25:1001–1014
60. Mulholland AJ (2008) Biochem Soc Trans 36:22–26
61. Senn HM, Thiel W (2009) Angew Chem Int Ed 48:1198–1229
62. Bo C, Maseras F (2008) QM/MM methods in inorganic chemistry**.** Dalton Trans 2911–2919
63. Senn HM, Thiel W (2007) Curr Opin Chem Biol 11:182–187
64. Ramos MJ, Fernandes PA (2008) Acc Chem Res 41:689–698
65. Warshel A, Levitt M (1976) J Mol Biol 103:227–249
66. Herbert JM, Head-Gordon M (2005) Phys Chem Chem Phys 7:3269–3275
67. Barnett RN, Landman U (1993) Phys Rev B 48:2081–2097
68. Car R, Parrinello M (1985) Phys Rev Lett 55:2471–2474
69. Zhang YK, Liu HY, Yang WT (2000) J Chem Phys 112:3483–3492
70. Hu H, Yang WT (2008) Ann Rev Phys Chem 59:573–601
71. Rod TH, Ryde U (2005) J Chem Theor Comput 1:1240–1251
72. Rod TH, Ryde U (2005) Quantum mechanical free energy barrier for an enzymatic reaction. Phys Rev Lett 94:138302
73. Moores SL, Schaber MD, Mosser SD, Rands E, O'Hara MB, Garsky VM, Marshall MS, Pompliano DL, Gibbs JB (1991) J Biol Chem 266:14603–14610
74. Jackson JH, Cochrane CG, Bourne JR, Solski PA, Buss JE, Der CJ (1990) Proc Natl Acad Sci USA 87:3042–3046
75. Takai Y, Sasaki T, Matozaki T (2001) Physiol Rev 81:153–208

76. Park HW, Boduluri SR, Moomaw JF, Casey PJ, Beese LS (1997) Science 275:1800–1804
77. Dolence JM, Rozema DB, Poulter CD (1997) Biochemistry 36:9246–9252
78. Kral AM, Diehl RE, deSolms SJ, Williams TM, Kohl NE, Omer CA (1997) J Biol Chem 272:27319–27323
79. Trueblood CE, Boyartchuk VL, Rine J (1997) Proc Natl Acad Sci USA 94:10774–10779
80. Fu HW, Beese LS, Casey PJ (1998) Biochemistry 37:4465–4472
81. Wu Z, Demma M, Strickland CL, Radisky ES, Poulter CD, Le HV, Windsor WT (1999) Biochemistry 38:11239–11249
82. Pickett JS, Bowers KE, Hartman HL, Fu HW, Embry AC, Casey PJ, Fierke CA (2003) Biochemistry 42:9741–9748
83. Pickett JS, Bowers KE, Fierke CA (2003) J Biol Chem 278:51243–51250
84. Tobin DA, Pickett JS, Hartman HL, Fierke CA, Penner-Hahn JE (2003) J Am Chem Soc 125:9962–9969
85. Pais JE, Bowers KE, Fierke CA (2006) J Am Chem Soc 128:15086–15087
86. Huang CC, Hightower KE, Fierke CA (2000) Biochemistry 39:2593–2602
87. Long SB, Casey PJ, Beese LS (2002) Nature 419:645–650
88. Sousa SF, Fernandes PA, Ramos MJ (2005) J Biol Inorg Chem 10:3–10
89. Sousa SF, Fernandes PA, Ramos MJ (2005) J Mol Struct THEOCHEM 729:125–129
90. Sousa SF, Fernandes PA, Ramos MJ (2005) Biophys J 88:483–494
91. Sousa SF, Fernandes PA, Ramos MJ (2007) Proteins 66:205–218
92. Sousa SF, Fernandes PA, Ramos MJ (2007) J Comput Chem 28:1160–1168
93. Dunten P, Kammlott U, Crowther R, Weber D, Palermo R, Birktoft J (1998) Biochemistry 37:7907–7912
94. Long SB, Casey PJ, Beese LS (1998) Biochemistry 37:9612–9618
95. Long SB, Casey PJ, Beese LS (2000) Struct Fold Des 8:209–222
96. Long SB, Hancock PJ, Kral AM, Hellinga HW, Beese LS (2001) Proc Natl Acad Sci USA 98:12948–12953
97. Strickland CL, Windsor WT, Syto R, Wang L, Bond R, Wu Z, Schwartz J, Le HV, Beese LS, Weber PC (1998) Biochemistry 37:16601–16611
98. Pompliano DL, Rands E, Schaber MD, SD MOSSER, Anthony NJ, Gibbs JB (1992) Biochemistry 31:3800–3807
99. Pompliano DL, Schaber MD, Mosser SD, Omer CA, Shafer JA, Gibbs JB (1993) Biochemistry 32:8341–8347
100. Dolence JM, Cassidy PB, Mathis JR, Poulter CD (1995) Biochemistry 34:16687–16694
101. Furfine ES, Leban JJ, Landavazo A, Moomaw JF, Casey PJ (1995) Biochemistry 34:6857–6862
102. Yokoyama K, McGeady P, Gelb MH (1995) Biochemistry 34:1344–1354
103. Mathis JR, Poulter CD (1997) Biochemistry 36:6367–6376
104. Hightower KE, De S, Weinbaum C, Spence RA, Casey PJ (2001) Biochem J 360:625–631
105. Sousa SF, Fernandes PA, Ramos MJ (2009) Chemistry 15:4243–4247
106. Ho MH, De Vivo M, Dal Peraro M, Klein ML (2009) J Chem Theor Comput 5:1657–1666
107. Rardin RL, Tolman WB, Lippard SJ (1991) New J Chem 15:417–430
108. Sousa SF, Fernandes PA, Ramos MJ (2007) J Am Chem Soc 129:1378–1385
109. Hightower KE, Huang CC, Casey PJ, Fierke CA (1998) Biochemistry 37:15555–15562
110. Saderholm MJ, Hightower KE, Fierke CA (2000) Biochemistry 39:12398–12405
111. Dolence JM, Poulter CD (1995) Proc Natl Acad Sci USA 92:5008–5011
112. Huang CC, Casey PJ, Fierke CA (1997) J Biol Chem 272:20–23
113. Turek-Etienne TC, Strickland CL, Distefano MD (2003) Biochemistry 42:3716–3724
114. Tschantz WR, Furfine ES, Casey PJ (1997) J Biol Chem 272:9989–9993
115. Comba P, Zimmer M (1994) Inorg Chem 33:5368–5369

116. Allen FH (2002) Acta Crystallogr Sect B 58:380–388
117. Norrby PO, Liljefors T (1998) J Comput Chem 19:1146–1166
118. Maple JR, Hwang MJ, Stockfisch TP, Dinur U, Waldman M, Ewig CS, Hagler AT (1994) J Comput Chem 15:162–182
119. Halgren TA (1996) J Comput Chem 17:490–519
120. Dasgupta S, Yamasaki T, Goddard WA (1996) J Chem Phys 104:2898–2920
121. De Kerpel JOA, Ryde U (1999) Proteins 36:157–174
122. Siegbahn PEM (1998) J Am Chem Soc 120:8417–8429
123. Ryde U (1999) Biophys J 77:2777–2787
124. Fernandes PA, Ramos MJ (2003) Chem Eur J 9:5916–5925
125. Fernandes PA, Ramos MJ (2003) J Am Chem Soc 125:6311–6322
126. Lucas MF, Fernandes PA, Eriksson LA, Ramos MJ (2003) J Phys Chem B 107:5751–5757
127. Pereira S, Fernandes PA, Ramos MJ (2004) J Comput Chem 25:1286–1294
128. Pereira S, Fernandes PA, Ramos MJ (2004) J Comput Chem 25:227–237
129. Pereira S, Fernandes PA, Ramos MJ (2005) J Am Chem Soc 127:5174–5179
130. Himo F (2006) Theor Chem Acc 116:232–240
131. Leopoldini M, Marino T, Michelini MD, Rivalta I, Russo N, Sicilia E, Toscano M (2007) Theor Chem Acc 117:765–779
132. Sousa SF, Fernandes PA, Ramos MJ (2007) J Phys Chem B 111:9146–9152
133. Sousa SF, Carvalho ES, Ferreira DM, Tavares IS, Fernandes PA, Ramos MJ, Gomes JANF (2009) J Comput Chem 30:2752–2763
134. Amin EA, Truhlar DG (2008) J Chem Theor Comput 4:75–85
135. Steudel R, Steudel Y (2006) J Phys Chem A 110:8912–8924
136. Cramer CJ, Truhlar DG (2009) Phys Chem Chem Phys 11:10757–10816
137. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M (1977) J Mol Biol 112:535–542
138. Lennard-Jones JE (1924) Proc R Soc A 106:441–462
139. Lennard-Jones JE (1924) Proc R Soc A 106:463–477
140. Hill TL (1948) J Chem Phys 16:399–404
141. Ryde U (1996) J Comput Aided Mol Des 10:153–164
142. Mulliken RS (1962) J Chem Phys 36:3428–3439
143. Reed AE, Curtiss LA, Weinhold F (1988) Chem Rev 88:899–926
144. Besler BH, Merz KM, Kollman PA (1990) J Comput Chem 11:431–439
145. Singh UC, Kollman PA (1984) J Comput Chem 5:129–145
146. Delre G, Pullman B, Yonezawa T (1963) Biochim Biophys Acta 75:153–182
147. Gasteiger J, Marsili M (1980) Tetrahedron 36:3219–3228
148. Berthod H, Pullman A (1965) J Chim Phys 62:942–946
149. Banci L (2003) Curr Opin Chem Biol 7:143–149
150. Lecerof D, Fodje MN, Leon RA, Olsson U, Hansson A, Sigfridsson E, Ryde U, Hansson M, Al-Karadaghi S (2003) J Biol Inorg Chem 8:452–458
151. Gilboa R, Spungin-Bialik A, Wohlfahrt G, Schomburg D, Blumberg S, Shoham G (2001) Proteins 44:490–504
152. Case DA, Darden TA, Cheatham TE III, Simmerling CL, Wang J, Duke RE, Luo R, Merz KM, Wang B, Pearlman DA, Crowley M, Brozell S, Tsui V, Gohlke H, Mongan J, Hornak V, Cui G, Beroza P, Schafmeister C, Caldwell JW, Ross WS, Kollman PA (2004) AMBER 8. University of California, San Francisco, CA
153. Berendsen HC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) J Comput Phys 81:3684–3690

CHAPTER 12

QM/MM ENERGY FUNCTIONS, CONFIGURATION OPTIMIZATIONS, AND FREE ENERGY SIMULATIONS OF ENZYME CATALYSIS

HAIYAN LIU
School of Life Sciences, University of Science and Technology of China and Hefei National Laboratory of physical Sciences at the Micro Scales, Hefei, Anhui 230026, China
e-mail: hyliu@ustc.edu.cn

Abstract: Combined quantum mechanical/molecular mechanical (QM/MM) models have been established as efficient approaches to simulate chemical reactions in complex molecular systems including enzymes. The QM/MM Hamiltonian is defined based on partitioning a molecular system into a reactive center and its surrounding, namely, the QM and MM regions. How to properly treat the QM/MM interface, which involves both covalent and non-covalent interactions, has been one of the central focuses in QM/MM method development. Techniques for energy minimization and conformational sampling based on QM/MM Hamiltonians have also been continuously developed, with the goal to determine reaction paths and potential/free energy surfaces of complex molecular systems efficiently and reliably. Accompanying method development, there have been an increasing number of studies applying QM/MM to various enzyme systems and providing new insights into their mechanisms.

Keywords: QM/MM Hamiltonian, Transition state stabilization, Reaction path optimization, Conformation sampling, Enzyme mechanisms

12.1. ENZYME CATALYSIS AND QM/MM MODELING

12.1.1. Non-Covalent Contributions to Enzyme Catalysis

Enzymes are highly efficient catalysts of biochemical reactions. Unlike many other catalytic agents that function with the assists of harsh environmental factors such as higher temperatures, higher pressures or extreme pH, enzymes require only mild environmental conditions. This feature of enzyme catalysis can be largely attributed

P. Paneth and A. Dybala-Defratyka (eds.), *Kinetics and Dynamics*, Challenges and Advances in Computational Chemistry and Physics 12,
DOI 10.1007/978-90-481-3034-4_12,

to the fact that non-covalent molecular interactions, including van der Waals and electrostatics interactions, play critical roles in enzyme catalysis [1–3]. Non-covalent interactions, which are individually weak and are of the same order of magnitudes as k_BT (k_B refers to the Boltzmann constant and T the temperature) at room temperature, can play following roles effectively under ambient conditions. First, they are the major driving forces for the folding of the enzyme peptide chains into unique three-dimensional structures and for the formation of multi-chain protein complexes. These provide enzyme active pockets comprising functional groups to covalently participate in the respective chemical reactions [4]. Due to protein folding and complex formation, these groups are pre-organized at suitable positions and with appropriate orientations. Second, the non-covalent interactions between the chemically active reaction centers and the inert enzymatic surroundings can also contribute significantly to catalysis through transition state stabilizations [1, 2].

Now we focus on the transition state stabilization effects due to non-covalent interactions. The (virtual) thermodynamics cycle shown in Scheme 12-1 and the accompanying Eq. (12-1) provides a framework [2] to understand how such effects lower the activation barriers of enzymatic reactions relative to non-enzymatic ones.

$$\Delta\Delta G^{\ddagger} = \Delta G^{\ddagger}_{enz} - \Delta G^{\ddagger}_{aq} = \Delta G_{TS-binding} - \Delta G_{R-binding} \tag{12-1}$$

Here, R_{aq} and R_{enz} refer to the reactant states in aqueous solution and in an enzyme environment, respectively. The TS_{aq} and TS_{enz} refer to the respective transition states, and $\Delta G^{\ddagger}_{aq}$ and $\Delta G^{\ddagger}_{enz}$ refer to the respective free energies of activation [5]. The lowering of the free energy barrier in the enzyme relative to that in aqueous solution, $\Delta\Delta G^{\ddagger} = \Delta G^{\ddagger}_{enz} - \Delta G^{\ddagger}_{aq}$, is equal to the difference between the free energy changes $\Delta G_{TS-binding}$ and $\Delta G_{R-binding}$, the former being associated with transferring the transition state and the latter the reactant state from the solution to the enzyme environment [2, 3].

$$\begin{array}{ccc} & {\scriptstyle \Delta G^{\ddagger}_{aq}} & \\ R_{aq} & \rightarrow & TS_{aq} \\ \Delta G_{R-binding} \downarrow & & \downarrow \Delta G_{TS-binding} \\ & {\scriptstyle \Delta G^{\ddagger}_{enz}} & \\ R_{enz} & \rightarrow & TS_{enz} \end{array}$$

Scheme 12-1. The thermodynamics cycle of transition state stabilization

12.1.2. Modeling Non-Covalent Interactions in Enzyme Reactions by QM/MM

To investigate how non-covalent interactions synergize with chemical reactions during enzyme catalysis, theoretical models are needed to treat chemically active covalent interactions and chemically inert non-covalent interactions simultaneously and consistently. Both types of interactions are governed by quantum mechanical principles [6]. However, conventional quantum mechanical (QM) models are computationally too expensive to be applied to large systems such as an enzyme–substrate complex solvated by thousands of solvent molecules. For such large systems, computationally efficient molecular mechanics (MM) models or molecular force fields [7–10] have been established as good approximations to QM. Unfortunately, conventional MM models, excluding those "reactive" ones with specifically tailored functional forms and/or parameters to model specific reactions, are not suited for the modeling of reactions because they contain major approximations that are valid only for non-reactive processes.

Combined quantum mechanical/molecular mechanical (QM/MM) models have been proposed as a general solution to the problem of modeling chemical reactions in large molecular systems ([11, 12], for recent reviews, see, for examples, references [13–16]). In a QM/MM model, a molecular system is partitioned into two parts or two subsystems: the QM part corresponds to the reaction center, and the MM part corresponds to the enzyme and solvent surroundings [11]. The QM part involves both the atomic and the electronic degrees of freedom, while the MM part involves only the atomic degrees of freedom. Usually only the electrons in the QM part are treated quantum mechanically, while all the atomic degrees of freedoms, either in the QM part or in the MM part, are treated classically, moving on Born–Oppenheimer surfaces associated with specific electronic states [6] (usually the electronic ground states of the respective systems).

12.2. QM/MM AS POTENTIAL ENERGY MODELS

Two types of approaches have been proposed to define a total Hamiltonian for a hybrid QM/MM system. One has been called "mechanical embedding" and the other "electrostatic embedding" [13–15].

12.2.1. Mechanical Embedding QM/MM

In "mechanical embedding", the total Hamiltonian is defined as a combination of three Hamiltonians, including the QM Hamiltonian of the standalone QM subsystem, the MM Hamiltonian of the standalone QM subsystem, and the MM Hamiltonian of the entire system [17, 18],

$$\hat{H}^{total}_{mechanical-embedding}(\mathbf{R}^{QM}, \mathbf{R}^{MM}) = \hat{H}^{QM}(\mathbf{R}^{QM}) + [\hat{H}^{MM}(\mathbf{R}^{QM}, \mathbf{R}^{MM}) - \hat{H}^{MM}(\mathbf{R}^{QM})] \quad (12\text{-}2)$$

Here $\mathbf{R}^{QM}$ refers to QM atomic coordinates, and $\mathbf{R}^{MM}$ refers to MM atomic coordinates.

A major advantage of the "mechanical embedding" approach is that existing QM and MM models can be applied separately, the results directly combined without any needs for new models or parameters [17]. One drawback is that the polarization of the QM part by its surrounding environment is not included in QM modeling. Mechanical embedding models with more than two layers of QM and MM have also been introduced [18], the most inner layer modeled by a higher level QM model, the intermediate layer by a lower level QM model, and the most outer by MM. In such cases, polarization of the inner most layer (which corresponds to the reaction center) by the intermediate layer (which may correspond to the immediate environment of the reaction center) are treated by the lower-level QM. We refer to the literature [17, 18] for further discussions of the "mechanical embedding" approaches.

12.2.2. Electrostatic QM/MM

We will focus on the "electrostatic embedding" model [11, 12]. The total Hamiltonian of the system is partitioned into an QM component ($\hat{H}^{QM}$) and an MM component ($\hat{H}^{MM}$), representing the interactions within the QM part and within the MM part, respectively, as well as a QM/MM component ($\hat{H}^{QM/MM}$) representing interactions between the QM and MM parts,

$$\hat{H}^{total} = \hat{H}^{QM} + \hat{H}^{MM} + \hat{H}^{QM/MM} \quad (12\text{-}3)$$

The $\hat{H}^{QM}$ consists of kinetic and potential energies of QM nuclei and electrons as in conventional QM models. In particular, the electronic parts is given by

$$\hat{H}^{electron,QM} = \hat{T}^{e} + \hat{V}^{ee} + \hat{V}^{ne,QM} \quad (12\text{-}4)$$

where $\hat{T}^{e}$,$\hat{V}^{ee}$, and $\hat{V}^{ne,QM}$ are the electronic kinetic energy, electron-electron interactions and electron-QM nuclei interactions, respectively.

The $\hat{H}^{MM}$ in equation 12-3 consists of kinetic and potential energy terms for MM atoms, as in conventional MM models, the potential energy part comprising bonded interactions as well as non-bonded interactions including van der Waals and electrostatic terms.

The QM/MM interactions in Eq. (12-3) also include both van der Waals terms and electrostatic terms,

$$\hat{H}^{QM/MM} = \hat{H}^{QM/MM,vdw} + \hat{H}^{QM/MM,electrostatic} \quad (12\text{-}5)$$

The QM/MM van der Waals terms are usually modeled using Lennard-Jones interactions as in MM force fields [7–10],

$$\hat{H}^{QM/MM,vdw} = V^{QM/MM,vdw} = \sum_{i \in MM \text{ atoms}} \sum_{j \in QM \text{ atoms}} \left(\frac{C_{12}^{ij}}{r_{ij}^{12}} - \frac{C_6^{ij}}{r_{ij}^6} \right) \tag{12-6}$$

in which r_{ij} correspond to the distance between atoms i and j, and C_{12}^{ij}, C_6^{ij} are atom type specific Lennard-Jones parameters as in MM force fields (see also references [19, 20] for the choices of van der Waals parameters of QM/MM interactions).

In $\hat{H}^{QM/MM,electrostatic}$, the charge distribution within the MM part is usually modeled as atom-centered point charges [12] , the same as in most MM force field,

$$\hat{H}^{QM/MM,electrostatic} = \sum_{i \in MM \text{ atoms}} \sum_{j \in QM \text{ nuclei}} \frac{Q_i^{MM} Z_j}{r_{ij}} - \sum_{i \in MM \text{ atoms}} \sum_{j \in QM \text{ electrons}} \frac{Q_i^{MM}}{|\mathbf{R}_i^{MM} - \mathbf{r}_j|} \tag{12-7}$$

Here $Q_i^{MM}$is the point charge on, and $\mathbf{R}_i^{MM}$the position of, MM atom i. Z_j is the nuclear charge of QM nucleus j. $\mathbf{R}_i^{MM}$and $\mathbf{r}_j$ denote positions of MM atom i and QM electron j, respectively. Polarizable MM charge distributions can also be treated [11, 21–23] but were not widely adopted in QM/MM applications.

The electrostatic field generated by the MM charge distribution polarizes the electronic charge distribution within the QM part, resulting in the following total effective electronic Hamiltonian,

$$\hat{H}^{electron,effective} = \hat{H}^{electron,QM} - \sum_{i \in MM \text{atoms}} \sum_{j \in QM \text{electrons}} \frac{Q_i^{MM}}{|\mathbf{R}_i^{MM} - \mathbf{r}_j|} \tag{12-8}$$

After solving the electronic ground state $\psi_0^{electron}$ under this effective electronic Hamiltonian, an effective potential energy containing the QM and the QM/MM electrostatic interactions can be computed,

$$V^{QM+QM/MM,electrostatic}(\mathbf{R}^{QM}, \mathbf{R}^{MM}) = <\psi_0^{electron}|\hat{H}^{QM} + \hat{H}^{QM/MM,electrostatic}|\psi_0^{electron}> \tag{12-9}$$

Finally, the total potential energy of the QM/MM system is given by

$$V^{total}(\mathbf{R}^{QM}, \mathbf{R}^{MM}) = V^{MM}(\mathbf{R}^{MM}) + V^{QM/MM,vdw}(\mathbf{R}^{QM}, \mathbf{R}^{MM}) + V^{QM+QM/MM,electrostatic}(\mathbf{R}^{QM}, \mathbf{R}^{MM}) \tag{12-10}$$

12.2.3. QM/MM Partitioning and the Treatment of Boundaries

12.2.3.1. Partitioning Between QM and MM

One important question in the QM/MM model is how to partition the system into QM and MM parts, and how to treat the bonded-boundary between the two parts.

The first consideration is that QM modeling is computationally much more expensive than MM modeling, conventionally with computational cost scaling as $O(N^x)$ with $x \geq 3$ and N being the number of QM electrons or basis functions. It is thus computationally advantageous to keep the QM part as small as possible. A minimum QM part could include only all chemically active atoms, i.e., those participating in the forming or breaking covalent bonds or with changing electronic states in the course of the reaction. On the other hand, it is desired to have a QM part large enough so that adverse effects of the artificial partitioning of integral molecules into QM and MM parts are minimized.

One potentially consequential approximation in QM/MM is the cutting of covalent bonds between QM and MM parts [12, 24, 25]. Unavoidably, truncating the QM part at these boundaries has artificial influences on at least the chemically neighboring QM atoms. A major goal for the development of approximate treatments for bonded QM–MM boundaries is to minimize the influence that the partition can induce, or to limit the propagation of respective artifacts towards the more interior of the QM region, so that interesting properties of the concerned chemical reactions, such as reaction paths and energetic, can be faithfully obtained despite such truncations. Currently, several techniques are used to treat bonded QM/MM boundaries, all of them, in varying extents, still involve some ad hoc approximations (see below).

12.2.3.2. The Link Atom Approach

The earliest approach, which is still widely in use, is the so called "link atom" method [12, 25]. For each covalent QM/MM bond, an extra"link atom" (usually a hydrogen) is artificially introduced. The first role of the link atom is to fulfill the bonding requirement of the boundary QM atom in QM modeling. Its second role is to ensure proper bond angles between the cut QM–MM bond relative to other QM bonds linked to the boundary QM atom. To achieve this, the position of the link atom is constrained to be along the QM–MM bond, and in some implementations at a fixed bond length (e.g., 1 Å) away from the respective boundary QM atom [12]. In addition, all bonded terms involving the pair of boundary QM and MM atoms in a fully MM model, are retained in the QM/MM model as components of the total MM energy.

One major argument against the link atom approach is that the introduction of an extra non-physical atom could complicate the physical interpretation of the total energy. Although in the MM part the link atom can be treated as a ghost and makes no contribution to the total energy, in the QM part of the total energy, the physically unreal contribution of the link atom and its interactions with both the QM and MM charge distributions have simply been retained in the total energy as there is no proper way to remove them [12].

12.2.3.3. The Pseudo-Bond Approach

Several theoretically more stringent approaches have been proposed and are still under continuous development to avoid the introduction of link atoms [19, 26–30].

One of particular interest is the so called "pseudo-bond" approach developed for ab initio QM/MM modeling [26, 27]. In this approach, the boundary atom on the MM side of a QM/MM covalent bond is "shared" between the QM and the MM treatments. On the MM side, it is treated as a normal MM atom, however on the QM side, it is treated as a specially "designed" atom forming a "pseudo bond" with the corresponding QM-side boundary atom [26]. This "designed" QM atom is of single valence, with an effective nuclear potential parameterized so that key bonding properties of the respective real bond, such as bond length, bond angles, and charge shifts, are closely reproduced by the pseudo bond. The pseudo bond parameters, once developed for a certain atom and bond type, can be applied in a transferable way to define pseudo bonds in different systems [26, 27].

12.2.3.4. Treating Immediate Neighbors Between QM and MM

Besides the treatment of covalent bonds themselves, another not so well-defined approximation in the modeling of covalent QM–MM boundaries is the treatment of QM–MM interactions between atoms separated by only two or three covalent bonds. Treating MM atoms as point charges at such close distances to some of the QM atoms would result in spurious interactions between the point charges and the QM electrons.

At present, a common and admittedly ad hoc practice is to set the point charges on such MM atoms and additional atoms belonging to the same MM charge groups as these atoms (see below) to zero when performing the QM calculations [26]. That is, the charges Q_i^{MM} in Eqs. (12-7) and (12-8) is set to zero for all atoms i that are separated by only one or two bonds from any QM atoms or belong to the same MM charge groups of such atoms. The following factors have to be considered in this seemingly simple choice. First, it is not physical either to consider the interactions of these charges with the entire QM part because of their close distances to neighboring QM atoms, or to consider only the interactions of these MM point charges with electrons on QM atoms that are further away but not with electrons on neighboring QM atoms. Second, a conservation of total charges of the entire MM part is desirable when resetting point charges on these MM atoms. In most MM force fields, chemically neighboring atoms are organized into charge groups, so that the total charge in a charge group adds up to zero or appropriate integer values [7–10]. By setting the partial charges to zero on all the MM atoms belonging to the same neutral charge group, the conservation of the total charge is achieved. Other approaches, such as treating MM atoms not as point charges but as "smeared", continuous, and atom-centered distributions, have also been proposed [31, 32].

12.2.3.5. QM/MM Boundaries in the Modeling of Enzymatic Reactions

For the modeling of enzymatic reactions, the above approximations for the treatment of bonded QM/MM boundaries are actually not so poor, in most cases not causing any major difficulties or serious problems. For enzyme modeling, it is typical to include in the QM part, besides the substrate, entire side chains of amino acid residues containing functional groups that chemically participate in the reaction.

Thus, the most common situation of a bonded QM/MM boundary is that the MM and QM parts are divided at a C^{α}–C^{β} bond separating a side chain from the peptide backbone. Because of the chemical inertness of this C–C single bond, and also because that in most MM force fields charge groups have been defined within boundaries of amino acid residues [7–10], the previously described approximations in treating such QM/MM boundaries do not have large adverse effects as compared with other approximations in QM/MM models. There are of course systems, including both enzymatic and non-enzymatic ones, for which more rigorous treatments of the bonded QM/MM boundaries are desirable. To provide a general solution for such systems, further developments of the QM/MM methodology are required.

12.2.4. Long Range Electrostatic Effects

12.2.4.1. Finite Size Effects and Charge Scaling

Enzyme catalysis takes place in aqueous solutions, which are infinite systems from a microscopic point of view. So far, most QM/MM implementations have considered only finite systems. The neglecting of long range electrostatic effects, especially the long range solvent screening on electrostatic interactions, may introduce significant errors when the reaction of interest involves significant charge redistributions. In a QM/MM model, including solvent molecules only within a sphere of a limited size may lead to overestimation of the interactions between charged MM groups and the reacting QM part. This is due to discounting long range solvent screening effects. To mimic the effects of solvent screening, empirical scaling/neutralization of enzymatic charged groups based on continuum model calculations has been proposed [33, 34]. More rigorous approaches include the explicit treatment of solvent with periodic boundary conditions or the implicit treatment of solvent as a dielectric continuum.

12.2.4.2. Periodic QM/MM Systems

In several semi-empirical QM/MM implementations [35, 36], long range electrostatics under periodic boundary conditions has been included in the QM calculations through the Ewald summation [37, 38]. There, a trick is employed to avoid the mathematically complicated direct computations of the non-columbic-form Ewald short range interactions involving electronic charge distributions [35]. First, both the QM and MM charge distributions are treated as point charges, and the conventional point-charge based formulation is applied to obtain both the long range and short range Ewald contributions of all charge distributions [37, 38], including their effects on the single electron integrals. Then the pure-columbic-form interactions between short-range QM point charges are subtracted to remove the double counting of the electrostatic interactions within the QM part in a single unit cell by the Ewald summation, as these interactions have been treated as normal columbic integrals in the QM model. In theory, similar approaches could be adopted in first principle QM/MM models to include explicit solvent using periodic boundary conditions.

12.2.4.3. Incorporating QM/MM with Continuum Models

Another way of introducing long range solvent effects is to consider an additional layer of dielectric continuum model (CM) surrounding a finite QM/MM system, as in MM–CM and QM–CM models. The most successful CM model for macromolecular systems has been the finite-difference–Poission–Boltzmann (FDPB) model [39–41]. In this model, the space occupied by solute atoms (here the MM and QM subsystems) are described as filled with a low dielectric constant continuum, with non-zero charge distribution. The solvent region is modeled as a high dielectric continuum. The electrostatic potentials satisfy the Poission–Boltzmann (PB) equations which can be solved using numerical methods, the most commonly used for proteins being finite difference [39]. The results are total electrostatic potentials represented on discrete grid points in space. The potentials are the sums of those produced by the solute charges and by the polarized solute and solvent continuums (the CM reaction field).

An ab initio QM/MM/CM model combining QM/MM with FDBP may be constructed in the following manner. One can take the advantage that the QM subsystem is completely surrounded by the MM subsystem. Thus, for the CM part, the QM charge distribution may be safely approximated by point charges, for instance, electrostatic potential-fitting or ESP charges [42, 43]. Given the QM and MM charge distributions, current point-charge-based FDPB can be applied to solve for the CM reaction field, whose electrostatic potential acts back on the QM region. The basic theory of electrostatics tells that in the QM region, this CM reaction field-generated electrostatic field can be accurately reproduced using virtual point charges located on a surface enclosing the QM region. In practice this surface can be conveniently chosen in the MM region, and virtual charges on it readily optimized by fitting to the CM reaction field potentials in the QM region, provided as values on grid points by the FDPB solver. As point charge–electron interactions are readily treated in ab initio QM, this approach should be readily implemented in existing QM/MM programs without major efforts.

One potential technical problem can be caused by the numerical instability of earlier FDPB solvers, which leave the computed CM reaction field potential too sensitive to positions of the QM or MM point charges relative to the FDPB grid points. These may result in large numerical fluctuations in computed potential energies with respect to subtle coordinate variations of QM atoms, making energy gradient-guided configuration moves, such as geometry minimizations, meaningless. With the continued progress in the construction of more accurate and numerically stable FDPB solvers [44], this problem should be overcome.

12.3. OPTIMIZATION AND SAMPLING IN QM/MM CONFIGURATION SPACES

Along with developments of computational models to determine potential energies of molecular systems, a range of techniques to sample in the molecular configuration space have been developed (for a recent review focusing on classical models,

see reference [45]). QM/MM models have been established as powerful tools to describe reactive potential energy surfaces of complex molecular systems. There are also increasing demands on accompanying tools for efficient optimization, searching and sampling of the conformational space based on QM/MM.

12.3.1. Effects of System Sizes and Computational Characteristics of QM/MM

For smaller systems, the configuration spaces are of lower dimensions, thus easier to characterize than for larger systems. On the other hand, the potential energy models applied to small systems are often based on QM, which is far more expensive to compute than the MM-based energy functions frequently applied to large systems. Thus traditionally, the developments of configuration move algorithms for smaller systems have largely emphasized on the efficient determination of critical stationary points, such as energy minima, transition states, and minimum energy paths connecting them (see, for example, references [46, 47]). For larger systems, the vast conformational space renders the meaning of such configurations and paths to be of far less significance. Fluctuations and entropies play non-negligible roles. Thus configuration move algorithms for large systems have been developed with the strongest emphasis on equilibrium sampling and efficiently escaping from local free energy minima [45, 48, 49].

12.3.2. Optimization on QM/MM Potential Energy Surfaces

12.3.2.1. Meaning of Stationary Points and Minimum Energy Paths in QM/MM

The first type of QM/MM-oriented configuration move approaches can be categorized as potential-energy-centric and have mainly been applied to ab initio QM/MM potential energy functions [50, 51]. Starting from one or a few initial configurations, a limited number of stationary points and connecting paths can be derived on the QM/MM total potential energy surface. Usually, optimizations are "driven", i.e., constrained or restrained, in various ways, so that the resulting structures and their associated energies are most relevant for characterizing the concerned chemical reactions. For instances, the stationary points are desired to have their QM coordinates correspond to reactant (enzyme–substrate complex), transition states, possible intermediates, and product (enzyme–product complex), respectively. The paths connecting the stationary points are desired to be of minimum energies with respect to local variations in all degrees of freedom other than the reaction coordinates.

From a biophysical point of view, enzymes, like other biomolecules functioning at a finite temperature, explore a vast conformational space spanned by tens of thousands of degrees of freedom. Thus, unlike small molecules, the results of potential-energy-centric approaches, namely, a few stationary points and minimum energy paths, need to be interpreted with care [52]. However, the results should still

be meaningful providing the following conditions are met. First, the starting configurations should correspond to the active state of the enzyme, otherwise the model would not be able to reproduce properties of the actual catalysis [53, 54]. Second, at different stationary points along the connecting paths, the large number of environmental degrees of freedom must remain in the same microscopic state and have been optimized numerically to the same extent [50]. If these conditions are not met, the computed energy variations between different stationary points and along the reaction paths might only correspond to fluctuations in the environment, and therefore would be irrelevant to energetic variations associated with the reaction process. Third, the total free energy profile associated with the reaction should be dominated by the potential energy, and the neglected thermo-fluctuations of the environments should not have sizable contributions.

In general, prior judgments on whether the above conditions have been met are not feasible, simply because that available experimental information is rarely subtle or detailed enough for such judgments to be made (see, for example, related discussions in reference [14]). Fortunately, one can rely on posterior justifications in most cases. For example, we can safely assume that if an erroneous starting model, e.g., one that does not correspond to the catalytically active state of the respective enzyme, has been used, QM/MM optimizations would fail to reproduce the catalytic effects of the enzyme. Namely, one would expect largely overestimated activation barriers [51, 52]. On the other hand, if the environmental degrees of freedom had not been maintained in the same microstate when the system were optimized along a reaction path, there would be discontinuity in the total QM/MM energy as a function of the QM coordinates. Finally, after obtaining smooth and continuous QM/MM reaction paths, the QM/MM free energy approaches (see below) provide good estimations of the contributions of environmental thermo-fluctuations.

12.3.2.2. Iterative Optimizations of QM and MM Subsystems

To increase the computational efficiency in ab initio QM/MM based geometry optimizations, an iterative optimization scheme has been proposed [50]. In this scheme, QM and MM optimizations are performed separately, and alternated iteratively. Each QM or MM optimization step consists of micro-iterations optimizing the respective subsystems.

In QM micro-iterations, quasi-Newton optimization of the QM coordinates in redundant internal coordinates [55] is performed with a fixed MM environment. The property of the quasi-Newton method [56] to converge super-linearly towards stationary points efficiently reduces the number of micro-iterations or the number of expensive QM energy and force evaluations in QM optimization steps. On the other hand, the method needs approximate second order derivatives and frequent transformations between Cartesian and internal coordinates. These make the computation scales quadratic to cubically with the number of atoms. Quasi-Newton

optimization is thus appropriate only for the QM subsystem that contains a small number of atoms.

In MM micro-iterations, the QM part is frozen [50]. In addition, the QM charge distribution is approximated by atom-centered point charges [19, 50] being fitted to the electrostatic potential (ESP) [42, 43] and being kept constant. Thus, although each MM optimization step may take a large number of micro-iterations to converge due to the large number of MM atoms and the suboptimum optimization algorithms, only the inexpensive MM energy and force evaluations need to be repeated at each MM micro-iteration.

The above QM and MM optimization steps are alternated in the main iteration cycles. Each step is performed with the part of the system not being currently optimized frozen at its newly updated configuration from the previous micro-iteration. The cycles are repeated until convergence of both QM and MM configurations has been reached [50].

To take advantage of the optimization efficiencies in redundant coordinate space even for the larger MM part, a novel divide-and-conquer approach has been proposed [57]. The system is divided into molecular blocks, each block are optimized in turn, in the redundant internal coordinate space defined for the molecular block alone, augmented by coordinate components specifying the absolute position of the respective molecular block in space.

12.3.2.3. Reaction Path Optimizations and the Nudged Elastic Band (NEB) Method

The above iterative QM/MM optimization approach can be used to drive the QM/MM system along a chosen reaction coordinate in order to reach different stationary points. In such approaches, one relies on chemical intuitions or presumptions of reaction mechanisms to choose a proper reaction coordinate. For complex enzymatic reactions, it is often none trivial to define a reaction coordinate without making excessive assumptions about possible reaction mechanisms (for example, the order of forming and breaking different bonds, or of chemical changes and necessary conformational changes).

To overcome this difficulty, techniques to optimize the entire reaction path simultaneously or as a whole have been proposed. These include those employing higher-order optimization techniques and are thus appropriate for smaller systems [46, 47], and those suit for larger systems [58–62], such as the nudged elastic band (NEB) method [61, 62].

In the NEB approach, a reaction path is represented by a chain of replicas; each replica is associated with a single configuration in the high-dimensional configuration space of the entire QM/MM system. Replicas at the two ends of the chain are to be optimized to two different minima on the potential energy surface (for examples, minima corresponding to reactant and product, respectively), while internal replicas along the chain are to be optimized to the minimum energy path linking the two respective minima. To ensure that the configurations associated with neighboring

replicas vary continuously along the chain so that the chain can represent continuous configuration changes from one minimum to the other, the configurations of neighboring replicas are coupled to each other using elastic forces, so that in the configuration space, the replicas are retrained to an even distribution along the path,

$$\left|\mathbf{F}^{elastic}(\mathbf{R}_l, \mathbf{R}_{l-1}, \mathbf{R}_{l+1})\right| \propto |S(\mathbf{R}_l, \mathbf{R}_{l-1}) - S(\mathbf{R}_l, \mathbf{R}_{l+1})| \tag{12-11}$$

Here, $\mathbf{R}_l, \mathbf{R}_{l-1},$ and $\mathbf{R}_{l+1}$ refer to configurations of the lth replica and the replicas immediately before and after l along the chain of replicas, respectively. $\mathbf{F}^{elastic}$ and $\mathbf{S}$ denote the elastic forces acting on the replicas and the separation between two replicas in the configuration space respectively. The chain thus defined is reminiscent of an elastic band. When the replicas are evenly distributed, the elastic forces are zero.

At the start of an NEB calculation, the coordinates of two initial end-state replicas are needed. There is however, no need to specify any reaction coordinates to drive the system from one end state to the other. The coordinates of internal replicas are initialized, possibly through interpolation between the end-state replicas [46, 61].

During the path optimization, the two end-state replicas are minimized under the system forces (i.e., forces associated with the physical energy, here the QM/MM potential energy). The internal replicas move according to both the system forces and the elastic forces from neighboring replicas. The replicas move along system forces perpendicular to the tangent of the chain, as well as along elastic forces parallel to the tangent of the chain. When the chain has moved onto a minimum energy path, the following conditions should be met: the system forces on the replica should be parallel to the path (or the perpendicular components of the system force should be zero); the elastic forces should be zero (or the replicas are equally spaced). A quenched velocity Verlet algorithm can be employed to achieve the minimization of forces with linear convergence [61]. A super-linear convergence approach has also been proposed [63].

The inter-replica spacing S in Eq. (12-11) is usually defined using the Cartesian coordinates of all atoms. This, however, causes problems for enzymes, which contain both very rigid (or hard) degrees of freedom, whose variations are associated with large energy changes, and very floppy (or soft) degrees of freedom, whose variations are associated with small energy changes [64]. Usually, the chemical coordinates varying during the reaction process are hard coordinates, while the environment contains a large number of soft coordinates, which are in a sense spectators of the reaction process. Defining the inter-replica spacing in NEB without discriminating between the two types of coordinates may lead the replicas to distribute along the soft coordinates but to collapse into separated minima along the hard coordinates. The chain of replicas would not converge onto the actual reaction path, which comprises continuous variations in the hard chemical coordinates [64].

The NEB approach can be adapted for complex enzymatic systems, by selecting only chemically relevant hard coordinates to define the inter-replica spacing [64],

which can include bond lengths that are changing between the bonded and non-bonded states during the reaction. A covalent bond, if present at the beginning of the reaction but broken in the middle of the reaction, can switch from a hard chemical coordinate into a soft spectator coordinate. A non-linear transformation of the bond length can be used as composing coordinates to define inter-replica spacing so that, only when a bond length is changing in the chemical reaction range, its variation contributes to the inter-replica separation. This adapted NEB scheme has been successfully applied to explore minimum energy pathways of several enzymatic reactions without having to presume reaction coordinates [64, 65].

12.3.3. Free Energies and Sampling in QM/MM Configuration Spaces

A second type of configuration move approach can be categorized as free-energy-centric. As mentioned above, at finite temperatures, the enzyme may explore a vast number of conformations where all possible conformations of the enzyme form a statistical mechanical ensemble. Free energies of structural ensembles instead of potential energies of single structures should be considered to understand catalysis.

12.3.3.1. The Basic Concept of Potential of Mean Force

Given a one-dimensional coordinate $\boldsymbol{Q}(\mathbf{R})$,which has been expressed as a function of the Cartesian coordinates R, its distribution in an equilibrium ensemble at temperature T can be written as

$$P(Q) = \frac{Z(Q)}{Z} \tag{12-12}$$

in which [67]

$$Z(Q) = \int \delta(Q(\mathbf{R}) - Q) \exp\left[-\frac{V(\mathbf{R})}{k_B T}\right] d\mathbf{R} \tag{12-13}$$

and

$$Z = \int Z(Q) dQ \tag{12-14}$$

Here, δ is the Dirac delta function, k_B is the Boltzmann constant, and $V(\mathbf{R})$ is the potential energy function. Z is the overall configuration integral, while Z(Q) is the configuration integral over all coordinates orthogonal to $Q(\mathbf{R})$ at the section $Q(\mathbf{R}) = Q$.

The corresponding free energy profile or potential of mean force (PMF) can be defined according to Z(Q) or P(Q),

$$\begin{aligned} G(Q) &= -k_B T \ln Z(Q) \\ &= -k_B T \ln P(Q) + \text{constant} \end{aligned} \tag{12-15}$$

If Q is a reaction coordinate associated with a chemical reaction, the minima on the free energy profile along Q correspond to the reactant, intermediates or the product. The maxima in between the minima are respective transition states. Transition state theory states that the reaction rate is proportional to the probability that the system is at the transition state [5],

$$k \propto \exp^{-\frac{G(Q^{\neq})-G(Q_R)}{k_B T}} \tag{12-16}$$

Here, $Q^{\neq}$ and Q_{R} correspond to locations of the transition state and reactant state along the reaction coordinate, respectively.

Obtaining the desired PMF from equilibrium MD simulations is possible only in principle but not in practice. In room temperature simulations of chemical reactions, distributions in high free energy regions as well as transitions over high free energy barriers would not be sampled, or only sampled poorly in any realistically possible amount of simulation time.

12.3.3.2. Determining PMFs Through Umbrella Sampling

Umbrella sampling [66] is the most widely applied statistical mechanics technique to obtain PMFs along one or a few pre-specified reaction coordinates. An umbrella or biasing potential as a function of the reaction coordinates can be included as an additional term in the potential energy function. This is usually a restraining potential that forces the system to stay in a region around given reaction coordinate values in simulations. The distributions of configurations and the corresponding potential of mean forces are biased in these simulations. Knowing the exact biasing potentials, the biases can be removed and the unbiased potential of mean forces restored,

$$\begin{aligned} G(Q) &= -k_B T \ln \frac{Z(Q)}{Z^{biased}(Q)} * Z^{biased}(Q) \\ &= -k_B T \ln \frac{\int \delta(Q(\mathbf{R}) - Q) \exp[-\frac{V(\mathbf{R})}{k_B T}] d\mathbf{R}}{\int \delta(Q(\mathbf{R}) - Q) \exp[-\frac{V(\mathbf{R}) + V^{bias}(Q(\mathbf{R}))}{k_B T}] d\mathbf{R}} - k_B T \ln Z^{biased}(Q) \\ &= G^{biased}(Q) - V^{bias}(Q) \\ &= -k_B T \ln P^{biased}(Q) - V^{bias}(Q) + \text{constant} \end{aligned} \tag{12-17}$$

Here $V^{bias}(Q)$ is the umbrella or biasing potential. The superscript biased denotes quantities associated with the biased ensemble. The "constant" term is also for the biased ensemble and thus varies between differently biased simulations.

The restored PMF from the distribution of Q in the biased simulation, up to an unknown constant, however, covers only the reaction coordinate region well-sampled under the respective biasing potential. To achieve sampling in different regions along the reaction coordinates, including those of higher free energies, the

positions of restraints are shifted in different simulations, so that the simulations with different biasing potentials collectively cover all regions of interest in the reaction coordinate space. Configuration distributions in different simulations are combined based on the fact that the PMFs restored from different biased simulations must agree with each other at places where the sampling overlaps. In particular, the so-called weighted histogram analysis method (WHAM) has been proposed as a technique to combine simulations with different biasing potentials and derive a single PMF with minimal statistical uncertainties in different regions of the reaction coordinate space [67, 68].

There are two limitations on current umbrella sampling techniques. The first is that reaction coordinates need to be pre-specified. Depending on the system this can be either simple or highly complex. One solution is to use reaction path optimization techniques to provide the reaction coordinates needed. Another limitation is the computational demand. Typically, umbrella sampling requires hundreds of thousands of molecular dynamics or Monte Carlo simulation steps to achieve equilibrium. At each step, not only the MM but also QM calculation of energies and forces need to be performed. In most applications involving QM/MM and umbrella sampling, semi-empirical QM/MM models have been used [69–72]. As computer power increases, applications using umbrella sampling on ab initio QM/MM potential energy surfaces have emerged [73] and are expected to increase.

12.3.3.3. QM/MM Free Energy Perturbations

As the QM part consists of only a small number of atoms, it can be a reasonable approximation to consider only key configurations of the QM part while treating thermo fluctuations of only the MM part by statistical mechanics sampling. The key QM configurations can be stationary points and minimum energy reaction paths determined in locally minimized MM environments using the previously described potential-energy-centric approaches. As an approximation, one can ignore variations in the QM part induced by deviations of the MM environment from the local MM minima. Thus the state of the QM part, including its configuration and charge distribution, can be frozen during sampling of the MM part. The QM/MM free energy perturbation (FEP) approach does exactly this. In this method, the QM part is perturbed to different state along the reaction path to obtain free energy changes in different MM thermo ensembles, each associated with a different QM state [50, 74],

$$G(\mathbf{R}_A^{QM}) - G(\mathbf{R}_B^{QM}) \approx (V_A^{QM} - V_B^{QM}) + \left[-k_B T \ln < \exp\left(-\frac{H^{QM/MM}(\mathbf{R}_A^{QM}, \mathbf{R}^{MM}) - H^{QM/MM}(\mathbf{R}_B^{QM}, \mathbf{R}^{MM})}{k_B T} \right) >_{B,MM} \right] \tag{12-18}$$

Here, subscripts A and B refer to different QM configuration and charge states. The $<\ldots>_{B,MM}$ refers to the average over the MM configuration ensemble sampled with the QM configuration and charge distribution fixed at state B. Sampling with the

QM fixed at state A can also be carried out and the above formula applied in the backward direction to check whether the differences between the QM states are small enough for a reliable application of the free energy perturbation formula. The above approach has employed the free energy perturbation theory [75] in molecular simulations, and is related to early works in which the QM parts have been perturbed between different stationary points obtained from gas phase modeling [76–78].

We note that in QM/MM FEP, the QM states have been optimized in environments whose coordinates remain in the same microstate but have been locally relaxed to accommodate changes in the QM parts. Thus, environment variations accompanying the proceeding of the QM part along the reaction path, as well as the QM responses to such variations, have been largely included. The missing effects are QM fluctuations as well as QM responses to MM thermo-fluctuations with the QM fixed at a location along the reaction path.

In the above approach, the QM part has been frozen at potential energy minima or on minimum energy paths with the environment remaining in a single microstate (but relaxed to accommodate different QM states). A more consistent approach would be to identify and freeze the QM part at free energy minima or on minimum free energy paths, with the MM environment averaged over thermodynamics ensembles in equilibrium with different QM minima. The QM/MM free energy minimization approach has extended the QM/MM FEP method in this direction [16, 79]. In this approach, MM configurations are sampled in the presence of frozen QM states by molecular dynamics simulations. The QM states are optimized in environments corresponding to averages of the sampled MM configurations.

12.4. APPLYING QM/MM TO ENZYMATIC SYSTEMS

12.4.1. Practical Issues

12.4.1.1. Choosing a QM/MM Model

The application of QM/MM and related configuration optimization and sampling techniques to an enzymatic system is still not a trivial task and involves many stages. The first step is to choose an appropriate level of model based on properties of the system and questions to be asked. Regarding the QM level of theory, semi-empirical models, including those based on molecular orbital theory [80–82] and on density functional theory [83], are computationally efficient. Their combination with MM models [12, 36, 84, 85] can allow for dynamical simulations on the complete QM/MM potential energy surface. On the other hand, the accuracy of commonly used semi-empirical models is highly system specific [81, 82]. For new applications, comparisons with higher level theories must be investigated to assess the accuracy and appropriateness of the method on a case by case basis.

Reaction specific parameterization of a semi-empirical QM model [86–88], which is, in a sense, related to the empirical valence bond (EVB) approach

[1, 89], is a viable solution. This approach, however involves substantial model development efforts and the results are not transferable between different chemical reactions.

Ab initio methods, including molecular orbital theory and density functional theory models, can be improved in a systematic manner to achieve the required chemical accuracy. Commonly employed models perform more or less consistently across different systems. Such models, however, allow far less extensive exploration of the configuration space, especially for the QM part. With current computer power, it is still too expensive to perform equilibrium sampling on an ab initio QM/MM surface of enzymes routinely.

Besides models solving the electronic problem by the self-consistent-field theory, the first-principle Car–Parrinello approach in which the electronic state variables are treated by virtual degrees of freedom and varies "on the fly" with nuclei coordinates in molecular dynamics simulations [90], are also adopted for the simulations of reactive biological systems [91, 92].

Multistage approaches, with models at different levels of theory applied at different stages, can be applied to problems which require both extensive exploration of the conformational space and chemical accuracy of energetics and structures. For example, one can first apply the computationally efficient semi-empirical QM/MM models together with pathway optimization techniques to explore different reaction pathways without having to presuming any reaction coordinates. The results provide the basis for the definition of reaction coordinates and initial configurations for subsequent ab initio QM/MM and QM/MM FEP modeling [65, 93]. In a reversed order, lower level reaction specific models (e.g. reaction specific parameterization of semi-empirical QM [85–87] or reactive force field [94]) can be derived by fitting to structures, energies and charge distributions of key QM configurations on the ab initio QM/MM surface. Such a model may be reapplied to achieve simultaneous sampling of thermo-fluctuations of both the reaction center and the enzyme/solvent surroundings.

12.4.1.2. Defining QM and MM Subsystems

After choosing a (initial) potential energy model, the molecular system to be modeled needs to be defined and partitioned into QM and MM parts. Using a larger QM part reduces influences of the treatment of QM/MM boundaries, but requires more computational resources. The different approaches of treating QM/MM bonded boundaries, although of varying theoretical stringency, are mostly of equal performance in practice: If handled with care, all should be reasonable and not the limiting approximations as compared with other components of the modeling process.

12.4.1.3. Setting up Initial Configurations

The preparatory set up of the starting configurations is one of the most tedious and error-prone step in a QM/MM application, being especially critical when a potential

energy-centric approach is adopted. Whenever possible, starting models have been constructed from experimental structures of enzymes in complex with transition-state like inhibitors. Experimental data lacks direct information about the position of hydrogen atoms and ionic states of titrable groups, such as the side chains of Glu, Asp, His, Lys and Arg residues. These need to be built into the starting structure either using software tools [41, 95, 96] or manually. An erroneously assigned ionic state would lead to an incorrect charge distribution in the system. Thus this issue needs to be handled with some extra care. It is not uncommon in QM/MM modeling that the protonation states of particular titrable groups are still uncertain after thorough considerations of available theoretical and experimental information, and that different possibilities have to be considered in parallel and the subsequent results compared.

The modeling of solvation is also an important issue when setting up the system. Explicit solvent models are usually used. As mentioned before, long range solvation effects may play roles on processes involving significant charge reorganizations. For such problems, the application of CM/QM/MM models is promising but has not been fully established. In most cases, solvent molecules are added at random initial positions, except for those inherited from the starting experimental structures. Molecular dynamics or Monte Carlo simulations are often applied to equilibrate the complete system consisting of enzyme, substrate and solvent. The equilibrated configurations provide starting points for subsequent applications of the various configuration move or sampling algorithms. For the optimization-based approaches, it has become common practice to start separate QM/MM calculations from different snapshots extracted from equilibrium sampling [97]. The results provide partial information about how key QM configurations and paths respond to changes of the MM environment between a limited number of different microsates.

12.4.2. Learning How Enzymes Work Through QM/MM Modeling

12.4.2.1. Structures, Energetic, and Transition State Stabilization

For many systems, properly optimized and posteriorly justified (based on structure and energetic results) stationary points and minimum paths on the potential energy surface can already provide the most essential pictures about enzymatic mechanism. These include the inclusion or exclusion of functional groups that play chemical roles, changes in structures and charge distributions along the reaction paths, the locations of intermediates and transition states along the paths, as well as the potential energy barriers associated with each transition state.

The differential interactions between the reaction center and the environment along the reaction path provide key insights into how enzymes achieve catalysis [1]. For instance, the differential interactions between the rate-limiting transitions state and the reactant can represent direct contributions of non-convalent interactions to the lowering of the activation barrier. These contributions can be decomposed and attributed to different environmental groups or amino acid residues. Such results,

when jointly analyzed with three dimensional positioning of these residues, provide understanding of how, through the folding of a one-dimensional polypeptide chain into specific three-dimensional structures, an enzyme active site has been pre-organized, both in terms of short range specific interactions and of long range electrostatic interactions, to selectively stabilize a transition state. For example, QM/MM investigation of enolase [74] have found that by arranging environmental charged residues in special orientations relative to the reactant, the enzyme provides specifically oriented electrostatic fields in the three dimensional space. These fields assist a catalytic metal ion in a first reaction step which shifts electrons towards a metal ligand group, while counter-balancing the resisting effects of the metal ion on a second reaction step which shifts electrons in an opposite direction.

12.4.2.2. Roles of Enzyme Dynamics

There is increasing evidence that not only statistical structures but also conformation dynamics play critical roles in enzyme catalysis. The results from optimization-based QM/MM are meaningful only when the enzyme has been framed in an active microstate, either in the previous experimental process (e.g., through the binding to a transition state analogue) or through relaxation in the equilibrium simulations. There will thus be increasing interest in identifying such active microstates, investigating how they equilibrate with less active conformational states and how the transitions between different microstates, or conformational dynamics, interplays with the chemical process.

Currently, techniques such as umbrella sampling and QM/MM FEP can be applied to investigate the coupling between chemical processes and thermo-fluctuations in the conformational space. In such applications, the principal reaction coordinates are mostly chemical ones. The sampling of environmental configurations has thus been limited to those rapidly reachable from the starting configurations through local thermo-fluctuations. For instance, it has been found that QM/MM umbrella sampling performed on the same enzymatic systems using starting structures prepared from different experimental structures give different activation barriers, and the origin of the discrepancy could been traced to different conformational states of the experimental structures [53, 54]. Such results indicate that some conformational coordinates do play important roles in enzyme catalysis, and that, at present, it is still a challenge to identify such conformational coordinates and to explore efficiently on QM or QM/MM potential energy surfaces. A number of techniques to enhance sampling on reaction potential energy surfaces, and related to techniques developed for MM energy surfaces, have been proposed [98–100].

In summary, synergy between developments of energetic models and configuration move or sampling algorithms has provided powerful tools whose applications have generated substantial insights into the catalytic power of enzymes. We expect such developments to be continued in the future, with novel and deeper understandings of how enzymes really work. Eventually, such understandings can be integrated with other theoretical and experimental approaches, allowing rational

approaches toward the improvements of existing enzymes and the creation of new enzymes that have never existed in nature before.

ACKNOWLEDGEMENTS

The support from Chinese Natural Science Foundation with grant number 30970560 is gratefully acknowledged.

REFERENCES

1. Warshel A (1991) Computer modeling of chemical reactions in enzymes and solutions. Wiley, New York
2. Wolfenden R, Snider MJ (2001) Acc Chem Res 34:938–945
3. Garcia-Viloca M, Gao J, Karplus M, Truhlar DG (2004) Science 303:186–195
4. Zhang X, Houk KN (2005) Acc Chem Res 38:379–385
5. Glasstone S, Laidler KJ, Eyring H (1941) The theory of rate processes. McGraw-Hill, New York
6. Atkins P, Friedman R (2000) Molecular quantum mechanics, 4th edn. Oxford University Press, New York
7. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) J Am Chem Soc 117:5179–5197
8. MacKerell AD Jr et al (1998) J Phys Chem B 102:3586–3616
9. Scott WRP, Hunenberger P, Tironi IG, Mark AE, Billeter SR, Fennen J, Torda AE, Huber T, Krueger P, van Gunsteren WF (1999) J Phys Chem A 103:3596–3607
10. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) J Am Chem Soc 118:11225–11236
11. Warshel A, Levitt M (1976) J Mol Biol 103:227–249
12. Field MJ, Bash PA, Karplus M (1990) J Comput Chem 11:700–733
13. Senn HM, Thiel W (2009) Angew Chem Int Ed 48:1198–1229
14. Gao J, Ma S, Major DT, Nam K, Pu J, Truhlar DG (2006) Chem Rev 106:3188–3209
15. Lin H, Truhlar DG (2007) Theor Chem Acc 117:185–199
16. Hu H, Yang W (2008) Annu Rev Phys Chem 59:573–601
17. Maseras F, Morokuma K (1995) J Comput Chem 16:1170–1179
18. Svensson M, Humbel S, Froese RDJ, Matsubara T, Sieber S, Morokuma K (1996) J Phys Chem 100:19357–19363
19. Murphy RB, Philipp DM, Friesner RA (2000) J Comput Chem 21:1442–1457
20. Riccardi D, Li G, Cui Q (2004) J Phys Chem B 108:6467–6478
21. Thompson MA (1996) J Phys Chem 100:14492–14507
22. Gao J (1997) J Comput Chem 18:1061–1071
23. Geerke DP, Thiel S, Thiel W, van Gunsteren WF (2007) J Chem Theory Comput 3:1499–1509
24. Bakowies D, Thiel W (1996) J Phys Chem 100:10580–10594
25. Rod TH, Ryde U (2005) J Chem Theory Comput 1:1240–1251
26. Zhang Y, Lee T-S, Yang W (1999) J Chem Phys 110:46–54
27. Zhang Y (2006) Theor Chem Acc 116:43–50
28. Assfeld X, Rivail JL (1996) Chem Phys Lett 263:100–106
29. Murphy RB, Philipp DM (2000) Chem Phys Lett 321:113–120
30. Gao J, Amara P, Alhambra C, Field MJ (1998) J Phys Chem A 102:4714–4721
31. Eichinger M, Tavan P, Hutter J, Parrinello M (1999) J Chem Phys 110:10452–10467

32. Amara P, Field MJ (2003) Theor Chem Acc 109:43–52
33. Dinner AR, Lopez X, Karplus M (2003) Theor Chem Acc 109:118–124
34. Cui Q, Karplus M (2003) Adv Protein Chem 66:315–372
35. Nam K, Gao J, York DM (2005) J Chem Theory Comput 1:2–13
36. Liu H, Elstner M, Kaxiras E, Fraunheim T, Hermans J, Yang W (2001) Proteins Struct Funct Genet 44:484–489
37. Ewald PP (1921) Ann Phys 64:253–287
38. Darden T, York DM, Pedersen L (1993) J Chem Phys 98:10089–10092
39. Gilson MK, Honig B (1988) Proteins Struct Funct Genet 4:7–18
40. Honig B (1995) Science 268:1144–1149
41. Bashford D (2004) Front Biosci 9:1082–1099
42. Singh UC, Kollman PA (1984) J Comput Chem 5:129–145
43. Besler BH, Merz KM, Kollman PA (1990) J Comput Chem 11:431–439
44. Zhou YC, Feig M, Wei GW (2008) J Comput Chem 29:87–97
45. Christen M, van Gunsteren WF (2008) J Comput Chem 29:157–166
46. Farkas O, Schlegel HB (2002) Phys Chem Chem Phys 2002 4:11-15
47. Schlegel HB (2003) J Comput Chem 24:1514–1527
48. Huber T, Torda AE, van Gunsteren WF (1994) J Comput Aided Mol Des 8:695–708
49. Grubmuller H (1995) Phys Rev E 52:2893–2906
50. Zhang Y, Liu H, Yang W (2000) J Chem Phys 112:3483–3492
51. Hall RJ, Hindle SA, Burton NA, Hillier IH (2000) J Comput Chem 21:1433–1441
52. Klaehn M, Braun-Sand S, Rosta E, Warshel A (2005) J Phys Chem B 109:15645–15650
53. Bhattacharyya S (2005) Biochemistry 44:16549–16563
54. Garcia-Viloca M, Alhambra C, Truhlar DG, Gao J (2002) J Am Chem Soc 124:7268–7269
55. Pulay P, Fogarasi G (1992) J Chem Phys 96:2856–2860
56. Ayala PY, Schlegel HB (1997) J Chem Phys 107:375–384
57. Billeter SR, Turner AJ, Thiel W (2000) Phys Chem Chem Phys 2:2177–2186
58. Elber R, Karplus M (1987) Chem Phys Lett 139:375–380
59. Czerminski R, Elber R (1990) Int J Quant Chem 24:167–186
60. Fischer S, Karplus M (1992) Chem Phys Lett 194:252–261
61. Henkelmaan G, Jonsson H (2000) J Chem Phys 113:9978–9985
62. Henkelmaan G, Uberuaga BP, Jonsson H (2000) J Chem Phys 113:9901–9904
63. Chu J-W, Trout BL, Brooks BR (2003) J Chem Phys 119:12708–12717
64. Xie L, Liu H, Yang W (2004) J Chem Phys 120:8039–8052
65. Zhao Z, Liu H (2008) J Phys Chem B 112:13091–13100
66. Torrie GM, Valleau JP (1977) J Comput Phys 23:187–199
67. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM (1992) J Comput Chem 13:1011–1021
68. Kumar S, Rosenberg JM, Bouzida D, Swendsen RH, Kollman PA (1995) J Comput Chem 16:1339–1350
69. Liu H, Mueller-Plathe F, van Gunsteren WF (1996) J Mol Biol 261:454–469
70. Liu H, Mueller-Plathe F, van Gunsteren WF (1995) J Chem Phys 102:1722–1730
71. Wu N, Mo Y, Gao J, Pai EF (2000) Proc Natl Acad Sci U S A 97:2017–2022
72. Garcia-Viloca M, Truhlar DG, Gao J (2003) Biochemistry 42:13558–13575
73. Hu P, Wang SL, Zhang Y (2008) J Am Chem Soc 130:16721–16728
74. Liu H, Zhang Y, Yang W (2000) J Am Chem Soc 122:6560–6570
75. Zwanzig R (1954) J Chem Phys 22:1420–1426
76. Jorgensen WL, Ravimohan C (1985) J Chem Phys 83:3050–3054

77. Chandrasekhar J, Smith SF, Jorgensen WL (1985) J Am Chem Soc 107:154–163
78. Donini O, Darden T, Kollman PA (2000) J Am Chem Soc 122:12270–12280
79. Hu H, Lu Z, Yang W (2007) J Chem Theory Comput 3:390–406
80. Dewar MJS, Thiel W (1977) J Am Chem Soc 99:4899–4907
81. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) J Am Chem Soc 107:3902–3909
82. Stewart JJP (1989) J Comput Chem 10:209–220
83. Elstner M, Porezag D, Jungnickel G, Elsner J, Haugk M, Frauenheim T, Suhai S, Seifert G (1998) Phys Rev B 58:7260–7268
84. Cui Q, Elstner M, Kaxiras E, Frauenheim T, Karplus M (2001) J Phys Chem B 105:569–585
85. Elstner M, Frauenheim T, Suhai S (2003) THEOCHEM 632:29–41
86. Gonzalez-Lafont A, Truong TN, Truhlar DG (1991) J Phys Chem 95:4618–4627
87. Cui Q, Karplus M (2002) J Phys Chem B 106:1768–1798
88. Ridder L, Rietjens IMCM, Vervoort J, Mulholland AJ (2002) J Am Chem Soc 124:9926–9936
89. Aqvist J, Warshel A (1993) Chem Rev 93:2523–2544
90. Car R, Parrinello M (1985) Phys Rev Lett 55:2471–2474
91. Colombo MC, Gossens C, Tavernelli I, Rothlisberger U (2006) In: Naidoo KJ, Brady J, Field MJ, Gao J, Hann M (eds) Modelling molecular structure and reactivity in biological systems. Royal Society of Chemistry, Cambridge
92. Rothlisberger U, Carloni P (2006) In: Ferrario M, Ciccotti G, Binder K (eds) Computer simulations in condensed matter systems: from materials to chemical biology, vol 2. Springer, Berlin
93. Dong M, Liu H (2008) J Phys Chem B 112:10280–10290
94. Lu Z, Yang W (2004) J Chem Phys 121:89–100
95. Davis ME, Madura JD, Luty BA, McCammon JA (1991) Comput Phys Commun 62:187–197
96. Madura JD, Briggs JM, Wade RC, Davis ME, Luty BA, Ilin A, Antosiewicz J, Gilson MK, Bagheri B, Scott LR, McCammon JA (1995) Comput Phys Commun 91:57–95
97. Zhang Y, Kua J, McCammon JA (2003) J Phys Chem B 107:4459–4463
98. Bussi G, Laio A, Parrinello M (2006) Phys Rev Lett 96:090601
99. Ensing B, De Vivo M, Liu ZW, Moore P, Klein ML (2006) Acc Chem Res 39:73–81
100. Yang W, Bitetti-Putzer R, Karplus M (2004) J Chem Phys 120:9450–9453

CHAPTER 13

COMPUTATIONAL MODELING OF BIOLOGICAL SYSTEMS: THE LDH STORY

SILVIA FERRER[1], SERGIO MARTÍ[1], VICENT MOLINER[1,2,3], AND IÑAKI TUÑÓN[3]
[1]*Departament de Química Física i Analítica, Universitat Jaume I, 12071 Castelló, Spain,*
[2]*Institute of Applied Radiation Chemistry, Technical University of Lodz, 90-924 Lodz, Poland,*
[3]*Departament de Química Física, Universitat de València, 46100 Burjasot, Spain*

Abstract: Lactate dehydrogenases, LDH, catalyzed reaction has been used in this chapter as a conductor wire to present the evolution and difficulties on computing methods to model chemical reactions in enzymes, since the early calculations based at semiempirical level carried out in gas phase to the recent sophisticated simulations based on hybrid Quantum Mechanical/Molecular Mechanics Dynamics (QM/MM MD) schemes. LDH catalyzes the reversible transformation of pyruvate into lactate. The chemical step consists in a hydride and a proton transfer from the cofactor (NADH) and a protonated histidine (His195), respectively. This fact has generated a lot of controversy about the timing of both transfers in the active site, as well as the role of the different aminoacids of the active site and problems related with the flexibility of the protein. We herein show how an adequate method within a realistic model, taking into account the pKa of the titratable aminoacids, the flexibility of the protein, the size of the MM and QM region or the level of theory used to describe the QM region, must be used to obtain reliable conclusions. Finally, and keeping in mind the size of the system under study, it has been demonstrated the need of performing statistical simulations to sample the full conformational space of all states involved in the reaction, that allow getting free energies and averaged properties directly compared with experimental data.

Keywords: Lactate dehydrogenase, LDH, Enzyme catalysis, Molecular mechanisms, Molecular modeling of biomolecules, Hybrid quantum mechanic/molecular mechanics, Hybrid molecular dynamics, Potential energy surfaces, Transition state characterizations, Potential of mean force, Statistical simulations

P. Paneth and A. Dybala-Defratyka (eds.), *Kinetics and Dynamics*, Challenges and Advances in Computational Chemistry and Physics 12,
DOI 10.1007/978-90-481-3034-4_13,

13.1. INTRODUCTION

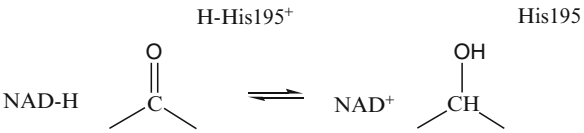
Since the early nineteenth century there were observations that certain substances influenced the progress of a chemical reaction without being consumed and hence apparently not being affected by this reaction. The term catalysis remained heavily debated [1] until around 1900, when Wilhelm Ostwald proposed its valid definition in terms of the concepts of chemical kinetics [2, 3]. In living systems, where cell metabolism takes place at room temperature and atmospheric pressure, chemical reactions are accelerated by means of a very special kind of macromolecules called enzymes; biological catalysts that allow organisms to carry out biological reactions with time scales compatible with life. Nevertheless, the precise origins of their catalytic power still remain unresolved after more than a century of research [4–6]. In order to known how enzymes work, Quantum Chemistry, the discipline in which quantum mechanics (QM) principles are used to rationalize and predict chemical behaviour and, in particular, to study chemical reaction processes, can be used. Quantum Chemistry developed rapidly during the second half of the twentieth century and it was fuelled largely by breathtaking improvements in computer technology and, by 1998, when the Nobel prize for Chemistry was awarded jointly to Walter Kohn and John Pople. Nowadays, it is widely held that computational chemistry has matured into a useful partner of experiment in almost all areas of chemistry [7].

Nonetheless, significant obstacles remain preventing the routine use of QM to provide quantitative understanding of complicated chemical systems [8]. The problems raised because describing processes that involve bond-breaking and bond-forming, charge transfer, and/or electronic excitation, require a QM treatment which, due to the very demanding computational cost, their application is still limited to relatively small systems. On the other side, models based on classical Molecular Mechanical (MM) force fields, widely used in molecular dynamics (MD) simulations of large systems, are unable to describe a chemical reaction. Algorithms that combine QM and MM (QM/MM) [9] provide the way to sort out the problem; the solute molecule or the reaction center relevant to the formation and breaking of chemical bonds is described quantum mechanically, while the solvent and protein environments are treated by a MM force field [10, 11].

An additional difficulty of modelling large molecular systems is found since in enzymatic environments exploration of a potential energy surface (PES) is not enough to get magnitudes directly comparable to experiment. The PES contains a myriad of stationary structures mainly due to the great number of possible conformations accessible to the enzyme and the solvent molecules. Furthermore, enzymes are usually very flexible molecular systems and in fact, backbone movements are, in many cases, responsible of some of the steps of the full catalytic process. Thus, a statistical ensemble of minima and transition structures must be explored to properly define the reactant state (RS), transition state (TS) and product state (PS). Free energies associated to the transformation between them can be

extracted using different techniques applied to molecular simulations [8]. These free energies profiles can be obtained, for instance, as a Potential of Mean Force (PMF) appearing along a particular distinguished reaction coordinate [12–14], or by means of the Free Energy Perturbation method (FEP) [10, 11]. Averaged properties can be derived from the dynamics obtained on the maximum and the minima regions of the PMF profile, characterizing in this way the three states involved in the reaction; RS, TS and PS. These QM/MM-MD calculations are very expensive and nowadays restricted to semiempirical hamiltonians [15] or empirical valence bond (EVB) methods [16, 17]. The final result is that new computational protocols have to be used to predict reliable results comparable with experimental data. Advances in modelling potential energy surfaces, in statistical mechanical sampling, and in dynamics will be required [18].

The aim of this chapter is to present some of the most common computational protocols designed to study biological complex molecular processes such as the ones catalyzed by the enzymes, from gas phase calculations to the last QM/MM schemes that include the dynamics of the full system. Lactate dehydrogenases (LDHs) offer an excellent opportunity for the analysis of computational strategies applied to the study of enzyme catalysis. LDHs are a family of 2-hydroxy acid oxidoreductases that simultaneously and stereoespecifically interconvert pyruvate to L-lactate and the nicotinamide adenine dinucleotide cofactor from the NADH form to NAD^+ [19–24]. This reaction, known as anaerobic homolactic fermentation, is an important pathway for regenerating NAD^+, enabling the continuation of glycolysis. Most mammal express several isozymes of LDH (up to five different forms in humans) and they have been also found in a range of species such as dogfish, *Bacillus stearothermophilus, Bifidobacterium longum* and *Plasmodium falciparum*. A high level of structural similarity is preserved among all these LDHs. In these enzymes binding of cofactor is followed by the substrate in an ordered process that includes the enclosure of the active site by a mobile loop of the protein. This last step is the rate-limiting one in the wild type enzyme [24]. In this movement several key residues are brought to close contact with the substrate, including an invariant arginine (Arg109) that is believed to stabilize the transition state (TS). The chemical step of the enzymatic process involves a hydride transfer (HT) from the dihydronicotinamide ring of NADH to the carbonyl C atom of pyruvate and a proton transfer (PT) to the carbonyl O atom from a N atom of the protonated His195 residue, as initially proposed by Holbrook et al. from exhaustive experimental investigations (see Scheme 13-1) [19, 21–24].



Scheme 13-1. Schematic representation of the chemical reaction catalyzed by LDH

13.2. GAS PHASE CALCULATIONS

Historically, first insights into enzymatic reaction mechanisms were obtained through gas phase calculations, using standard programs of quantum chemistry to obtain PESs. Thus, for example, in the early 1990s, the reaction catalyzed by LDH was studied in our group by means of semiempirical Hamiltonians (AM1 and PM3) in gas phase [25–27].

Our first theoretical calculations carried out in gas phase predicted a concerted but synchronous mechanism with a TS describing a PT in an advanced stage of the reaction and a hydride almost in between the donor and acceptor carbon atoms. Another conclusion that could be derived from these calculations was that both hydrogen transfers took place in almost perpendicular planes.

These results were in accordance with previous studies carried out also in gas phase by Krechl and Kuthan [28], Wilkie and Williams [29, 30] and by Ranganathan and Gready [31], but in contradiction with empirical valence bond calculations performed by Warshel et al., which pointed out to the opposite way [32]. Thus, the timing of the proton and hydride transfer was still unclear and in fact both, stepwise and concerted mechanisms could not be discarded from evidences obtained by gas phase calculations. These contradictory conclusions begged the question: what influence does the full protein environment have upon the reaction mechanism?

13.3. INCLUSION OF ENVIRONMENT EFFECTS

13.3.1. Cluster Models

The properties of stationary structures of enzymatic processes can be different from those obtained in gas phase calculations because, obviously, the interaction effects with the environment are not considered in the latter. A more realistic picture of enzyme catalyzed reactions can be obtained including a small part of the active center into the calculations. Following with the example of the LDH, it is interesting to point out how, although some of the studies mentioned above were carried out in gas phase, the effect of the protein was already introduced, although in a quite preliminary way. For instance, the work of Williams [29, 30] evaluated the distance effect of a negative point charge on the His195, thus mimicking at some extent the effect of Asp168 in modulating the pKa of the proton donor residue. The substrate binding role of Arg171 was modelled in our calculations by the inclusion of a 1-methylguanidinium ion, thus fixing the pyruvate molecule [19, 21–24], while Ranganathan and Gready included some fragments representing key active site residues like Arg171, Val138, Arg109 and Asp168 [31]. The problem is that in this *cluster* or *supermolecule* models, the optimised structures do not necessary fit into the enzyme active site and computing artefacts can be obtained. A common strategy is to anchor some key atoms of the enzyme to their crystallographic positions and

then optimising the rest of the coordinates of the model. Nevertheless, this is an approximate solution that presents several deficiencies. First, the result will be dependent on the initial X-ray structure that, in many cases, is far from the real structure of the protein-substrate complex in the TS. Second, the long-range effects on the nuclear and electronic polarisation of the chemical system are not included in the calculations. Third, and probably the most dramatic deficiency, the enzyme flexibility is not properly taken into account. And finally, the computational cost of these calculations rapidly increases, as more atoms of the environment are explicitly included.

Nevertheless, the *cluster models* is still broadly used with models with up to 200 atoms, placed in a dielectric cavity and treated at the highest possible QM level, which so far has mainly meant hybrid density functional theory (DFT) [33]. The most common application of these models is on systems where the chemical reaction implies complex electronic states as the ones appeared in enzymes containing transition metals where more than one electronic state is involved and ab initio methods are required to describe the reaction. Nevertheless, due to the computational cost, such studies have embraced topics devoted to the modelling of the first coordination sphere at the metallic cations up to the development of biomimetic, or bioinspired, catalytic systems [34, 35]. But, in these cases, as discussed above, the effect of the rest of the protein and/or solvent environment is not included in the calculations and, moreover, one of the requisites of a proper modelling of an enzyme, the flexibility of the system, is not taken into account.

13.3.2. QM/MM Methods

Shortly after those first studies based on gas phase calculations, there was a major breakthrough that included the enzyme in the calculations by a combination of QM Hamiltonians and MM force fields; namely QM/MM methods. Paradoxically, although this approach pioneered proposed by Warshel and Levitt in the late 1970s [9], it was ignored for long time, probably due to problems in the implementations. In these methods, a small portion of the system is described by QM, the region where the most important chemical changes (i.e. bond breaking and forming processes) take place, while the rest of the system is described by means of MM potentials. In this way a very large number of atoms can be explicitly considered in the calculations. The division of the full system in two subsystems may require cutting a covalent bond. Then, the problem of fulfilling the valence of the quantum atom placed in the QM/MM boundary has been treated by means of different techniques such as adding hydrogen atoms or frozen orbitals [36]. The combination of these two levels, by means of the appropriated coupling terms, allows obtaining the wave function of the quantum subsystem, and thus any related property, under the influence of the environment [12–14].

The existence of this QM/MM potential energy function, which combines reliability and computational efficiency, is not the only requirement to deal with

chemical reactions. To describe such a process we should be able to locate and characterize the set of stationary structures (RS, TS, PS and possible intermediates) that defines a particular reaction mechanism. This is complicated by the large dimensionality of the PES when the environment is explicitly included. In a first approximation to follow a chemical process, a distinguished geometrical coordinate can be chosen and energy minimizations would be carried out for different values of this coordinate. Thus, the minima on a potential energy profile corresponded with reactants, possible intermediates or products, while the maximum would represent an approximate TS of the reaction. Obviously, this procedure cannot be always convenient as far as the true reaction coordinate can be different from the selected distinguished one. Most of the programs using hybrid calculations used to contain only algorithms to locate energy minima. The direct location of a saddle point of index one is much more difficult and was not included in commercial package of programs.

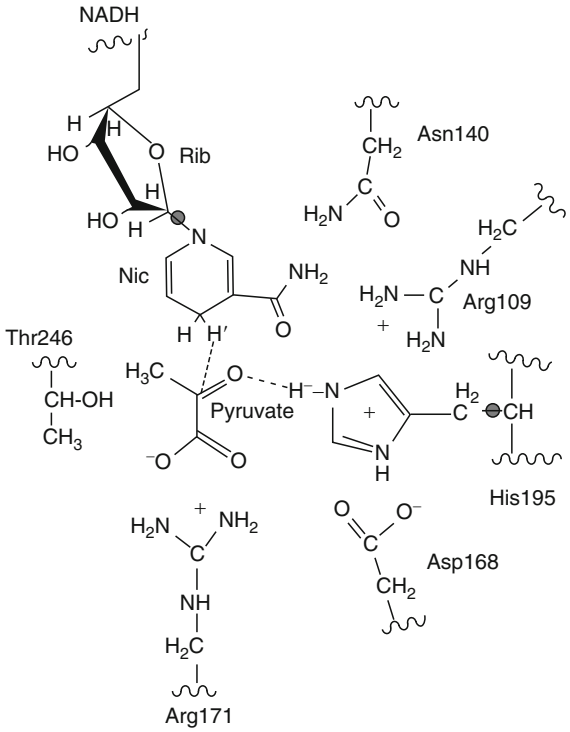
Realistic simulations of chemical reactions require the use not only of methods capable of describing accurately the energy of molecules undergoing bonding changes within a particular chemical environment, but also of methods capable of exploring topographical features of significance on energy hypersurfaces spanning several thousand degrees of freedom. Hybrid QM/MM techniques show much promise for the first task, but existing computer codes are many times inadequate for the second. Application of these methods to real chemical problems demands tools for location and characterisation of saddle points, intrinsic reaction coordinates, hessians and vibrational frequencies for very large flexible systems. Nevertheless, the high dimensionality of the surface prevents the calculation of the full Hessian matrix and thus the proper location and characterization of saddle points of index one.

In the late 1990s, we carried out the first QM/MM study where a TS structure was located and characterized as a first order saddle point for a fully flexible molecular system on the reaction catalyzed by the LDH [37]. In order to locate the TS a new algorithm, GRACE [38], was employed with an explicit hessian for 147 core atoms (QM subsystem + 6 residues in the active-site region) by central finite-differencing of the gradient vector: the single negative eigenvalue of the resultant hessian corresponded to the force constants associated to the transition vector. The remaining atoms of the MM enzyme environment were continually relaxed to a rms gradient $<10^{-3}$ kcal mol^{-1} Å^{-1}. TS optimisation in the core was continued until no element of the gradient vector of the entire system (core plus environment) was larger than 10^{-2} kcal mol^{-1} Å^{-1}. The intrinsic reaction coordinate (IRC) was computed in both directions from each saddle point to confirm each as being the expected TS; energy minimisations from a point along each IRC path yielded structures for the reactant, intermediate, and product. The stationary structure obtained in this way could be characterized as a true stationary structure (all the first derivatives are equal to zero) and having the correct number of imaginary frequencies in the Hessian matrix of the core (i.e. zero for minima and only one for transition state). From this structure the transition vector and

reaction paths can be defined and used to get relevant knowledge about the reaction mechanism.

The molecular model and the level of theory employed to describe the QM region were limited by the state of the art in computational capabilities. Thus, it was impractical to include the entire protein in the calculations and so, after several tests, it was considered a system formed by pseudo-sphere of a total amount of 1,900 atoms; active site, bulk protein, crystal and solvent water molecules altogether. The pyruvate molecule, the nicotinamide ring of the NADH and imidazol ring of the His195 was treated quantum mechanically by means of AM1 semi-empirical hamiltonian, while the rest of the protein and the crystal and solvent water molecules were treated classically by means of the CHARMM24b force field [39]. Two link atoms were needed, since the QM/MM boundary intersected covalent bonds of protein backbone and cofactor. In Scheme 13-2, a schematic representation of the active site of LDH is depicted.

Six different starting structures (arbitrarily selected from a classical molecular dynamics trajectory for the enzyme-substrate complex) lead to six different TSs [38]. Although the essential features of these TSs are invariant, the relative dispositions of active-site residues differ quite significantly. The analysis of these TS structures revealed differences in the interactions established between the



Scheme 13-2. Representation of the active site of the LDH. Link atoms, as used in ref 37, are indicated as grey dots

substrate and the invariant Arg109. In particular it was found that the TS-stabilising role of this residue could also be alternatively played by Asn140. Nevertheless, the resulting transition vectors for the located TSs of the LDH-catalysed reaction involved motion of both the transferring hydride and proton, indicating a concerted mechanism. However, PT was much further advanced than HT. This was remarkably similar to the TSs founded in a gas-phase AM1 supermolecule calculation for a *N*-methyl-1,4-dihydronicotinamide–pyruvate–(methylguanidinium)–methylimidazolium complex [25–27].

These QM/MM studies revealed that the TSs for this reaction appeared to be a robust entity, the essential features of whose structure are invariant to the nature of its environment. Moreover, capabilities of the employed method showed that several active site residues, notably Arg109 and Asn140, displayed considerable mobility over the range of structures spanned by the energy surface which gave evidence that the reported TSs were only one subset of a family of structures of similar energy which differ in regard to conformations of amino acid side chains. Although other authors as Truhlar and coworkers had already pointed out this conclusion for molecular systems in condensed media [40], these results were one of the first QM/MM studies that demonstrate how the transition state for the enzymatic reaction must be represented as an average of the properties of these many, nearly degenerate TS structures. This insight emerged only as a consequence of the flexible model of the active site employed in this study. As we will discuss later on, subsequent studies showed that it is also possible to find different families of stationary structures separated by large conformational barriers. Each of these families would correspond to different reaction paths that can be followed depending on the particular conformation adopted by the enzyme in the Michaelis complex. This was already observed by Yeung and co-workers, in the 1990s, who claimed to have monitored differences in the chemical reactivity of individual LDH molecules [41, 42]. Single molecules were trapped inside femtolitre-sized reaction vessels (pores in membranes or nanoscopic silica vials manufactured by photolithography) when filled with solutions containing very low concentrations of enzyme together with excess lactate and NAD+. By means of a laser-based optical microscope and a charge-coupled device detector, it was possible to follow the production of fluorescent NADH in more than a hundred individual reactors simultaneously. It was found that activity of individual electrophoretically pure enzyme molecules could vary by a factor of three or four, and that these activities remained unchanged over reasonably long times. Even when their microenvironments were identical, individual enzyme molecules displayed different rates of reaction. These observations were interpreted as evidence for subtly distinct conformations of the LDH molecules. In fact, this was the interpretation of our first QM/MM simulations on LDH started from initial configurations obtained by equilibrium molecular dynamics simulations, which generated different minimum energy paths with noticeable different potential energy barriers [38]. Nevertheless, it must be pointed out that the potential energy barriers obtained within this AM1/CHARMM/TIP3P method appeared to be overestimated probably due, mostly, to

the limitations of the semiempirical Hamiltonian. This was confirmed in subsequent studies at higher level of theory (see below).

Almost simultaneously to the publication of our first QM/MM results on the LDH catalyzed reaction, Ranganathan and Gready published a QM/MM study on the same system and obtained a different molecular mechanism; a stepwise mechanism where hydride transfer took place previous to the proton transfer [43]. They used the same AM1 method for the QM region but AMBER [44] force field for the entire dogfish LDH sub-unit ternary complex. Moreover, it is important to point out that a relatively small number of atoms were allowed to move in their simulation. In contrast, in our first QM/MM study on LDH we used the CHARMM22 force field for a truncated *B. stearothermophilus* LDH (1,900 atom) but all atoms were free to move during optimizations [37].

In order to overcome the possible limitations of our computational model we repeated the same computational protocol as in our previous study but using a larger model which included all atoms of the monomeric *B. stearothermophilus* LDH subunit, plus NADH, pyruvate, and water molecules within a 20 Å radius ball of water centred on His195 [45]. All atoms were free to move in the optimisations and TS searches were done within a larger QM region containing also the ribose moiety. The amazing result was that reaction paths and TSs for *both* the stepwise HT/PT and the concerted PT/HT mechanisms were localized (see Figure 13-1) [45].

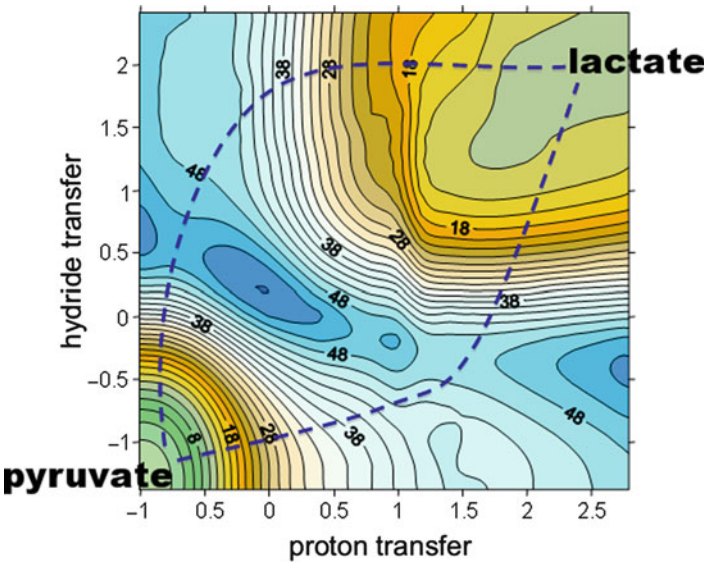


Figure 13-1. AM1/MM PES for the reduction of pyruvate to lactate catalyzed by LDH, obtained with a large QM region and the solvated monomeric *B. stearothermophilus* LDH subunit. Proton transfer and hydride transfer coordinates correspond to the anti-symmetric combination of the forming and breaking bonds for each transfer (Data from [45])

There was no fundamental disagreement between our results and the ones obtained by Ranganathan and Gready [43]: the apparent discrepancy raised simply because different parts of the hypersurface had been explored in each of the earlier studies. Flexible QM/MM modelling, allowing the enzymatic environment to change while the system advances along the reaction coordinate, could embrace the alternative mechanisms previously described.

Another common problem that can rise in theoretical studies of enzyme reaction mechanisms is that the most expensive computational simulation may give meaningless results if an erroneous charge is selected for a single amino acid of the protein. Properties of enzymes are pH dependent because their efficacy depends on the protonation state of their ionisable residues. Thus any theoretical analysis of protein structure or reactivity requires an adequate titration of all ionisable residues. Assignment of the protonation states of such residues is often based on the pK_a of the corresponding aminoacids in aqueous solution. However, this practice may introduce significant artifacts into simulations because standard pK_a values of ionisable groups are shifted by local protein environments [46, 47]. We therefore improved our computational protocol with the recalculation of the pK_a values using the "cluster method" [46, 47] as implemented by Field and co-workers [48], according to which each titratable residue in the protein is perturbed by the electrostatic effect of the protein environment. When the pK_a values of the ionisable residues were recalculated under the effect of the protein environment, as explained above, the protonation state of several ionisable groups at pH = 7 was found to have changed. In particular, Glu199 and His82 became protonated, thereby changing their charges to neutral and positive, respectively [49]. The effect of this change in only two of the residues on the resulting PES is drastic.

All previous QM/MM calculations were made using a semiempirical AM1 Hamiltonian that presents, as mentioned above, serious deficiencies for obtaining at least accurate energies. We more recently tackled the problem of improving the quality of the PES by means of corrections to the quantum description [50]. From our results, obtained using corrections to the AM1/MM energy surface calculated at the MP2/6-31G* level, we arrived at some conclusions about the timing of the hydride transfer and proton transfer components in the chemical step catalyzed by LDH that had not been possible to state definitively up to then. A concerted mechanism was obtained where the hydride transfer was more advanced than the proton transfer [49]. In addition, we demonstrated that reaction mechanisms and barrier heights of an enzymatic reaction can be wrong if molecular model and computing method are not good enough.

Finally, in order to set up a realistic molecular model for enzyme simulations, it has been recently shown by Swiderek and Paneth [51] that quaternary structure of LDH has important bearings on the structure of the active site and that models that comprise all amino acids in the vicinity of the active site but missing this structural information might lead to erroneous results.

13.4. STATISTICAL SIMULATIONS

Once clearly stated the protocol to obtain accurate QM/MM PESs, we must come back to the point that enzymes are huge molecular systems. In this sense, the stationary points of large molecular systems on a PES are not unique since the calculated PES is a projection in 3 dimensions of a 3N-5 (3N-6 geometrical variables plus the energy) hypersurface. The global process involves not only two chemical events (the proton and the hydride transfers) but also conformational changes of the enzyme that can determine the differences in the reactivity of different single molecules. A valley in the potential energy landscape of the enzyme contains an ensemble of nuclear configurations that can be explored by means of a long-time molecular dynamics trajectory. As explained before, the landscape within each valley is extremely rugged, with a huge number of stationary points. There exist multiple local stationary points (RS, TS and PS) with slight geometrical differences. The interconversions among all the low energy configurations within the same valley usually consist of relatively easy local conformational changes and are quite fast at physiological temperatures. In these conditions, a local equilibrium assumption within a given reactant valley can be adopted and then it is possible to get an ensemble average over paths, so leading to a single free energy barrier which determines the catalytic rate constant within the framework of the transition state theory. Thus, the accurate determination of transition state properties requires statistical averaging over many possible stationary structures corresponding to RS, PS or TS. Simulations can be carried out using Monte Carlo or Molecular Dynamics methods obtaining detailed information from which averaged and thermodynamical properties can be derived [8]. In particular, free energies associated to the transformation from the RS to the TS (the activation free energy) and to the PS (the reaction free energy) can be extracted using different techniques applied to molecular simulations [10].

13.4.1. Free Energy Perturbation (FEP)

The firsts attempts to obtain free energy profiles for the pyruvate to lactate transformation catalyzed by LDH was carried out in our lab much more recently [52]. We first tried to use the umbrella sampling technique and to obtain the free energy barrier in terms of a Potential of Mean Force (PMF). This technique allows, in principle, for a correct sampling of all the degrees of freedom of the system except, obviously, the distinguished reaction coordinate. Thus, a proper selection of this coordinate is critical to obtain meaningful free energy profiles. The problem in the case of LDH is the fact that two transfers are simultaneously involved in a single chemical event and then it is difficult to define an adequate geometrical reaction coordinate. We tried to trace the free energy profiles using umbrella sampling with different definitions of the reaction coordinate of the type:

$$RC = \alpha_1 r_{CnicH'} - \alpha_2 r_{CpyrH'} + \alpha_3 r_{NhisH''} - \alpha_4 r_{OpyrH''} \qquad (13\text{-}1)$$

However, the results were quite discouraging. While we were able to control the behaviour of the hydride transfer (monitoring the corresponding transfer coordinate we observed a soft and continuous variation as the reaction coordinate was changed), the proton transfer invariably took place suddenly, being transferred from His195 to pyruvate in a short time interval of a single window. Attempts to control the variation of the proton transfer coordinate changing the values of the α_1 weights appearing in the definition of the global reaction coordinate always failed. The free energy profiles obtained in this way displayed unexpectedly high reaction free energy barriers (corresponding to exploration of high energy regions of the PES) and discontinuous variations along the reaction coordinate. We then decided to employ a different strategy based on the use of the free energy perturbation (FEP) technique. QM-FEP method, developed by Kollman and coworkers [10, 11], is one of the most popular techniques to get reaction free energy profiles for enzymatic reactions. In this approach the reaction path is obtained for a gas phase model of the active site. Coordinates and charges of the chemical system are afterwards used into purely classical simulations where only the changes in the environment are sampled. FEP is then used to get the reaction and activation free energies as the sum of the gas phase reaction energy and environment free energy contributions. The main advantage of this approach is that high-level quantum methods can be used for the gas phase calculations. The main drawbacks are: (i) the lack of the chemical system flexibility contribution to the free energy and (ii) during the simulation of the reaction path, the environment is not incorporated in the calculation. A similar but more coupled scheme is obtained if the reaction path is determined in the active site by means of QM/MM iterative optimizations along a reaction path [53].

In our approach [52] the reaction path is an obtained as an IRC traced from a TS structure located in the active site and considering the influence of the full enzymatic environment. The configurations for which the free energy difference on LDH catalyzed reaction were estimated correspond to those structures obtained along the IRC calculation and are thus characterized by a single coordinate:

$$s = \left[\sum_{i \in QM} m_i \left((x_i - x_{0i})^2 + (y_i - y_{0i})^2 - (z_i - z_{0i})^2 \right) \right]^{1/2} \tag{13-2}$$

where x_{0i}, y_{0i}, z_{0i} are the Cartesian coordinates of the QM atoms in the transition state structure while x_i, y_i and z_i are the coordinates of a structure belonging to the IRC traced from this transition state structure, and m_i are the masses of the atoms. Within this treatment the free energy relative to the reactant can be expressed as a function of the s coordinate as:

$$\begin{aligned} \Delta G_{FEP}(s^{\mathrm{R}} \rightarrow s^j) &= \Delta E_{QM}^{0,R \rightarrow j} + \Delta G_{QM/MM}^{R \rightarrow j} \\ &= \left(E_{QM}^0(s^{\mathrm{j}}) - E_{QM}^0(s^R) \right) \\ &\quad - k_B T \sum_{i=R}^{i=\mathrm{j}-1} \ln \left\langle \exp \beta \left(\mathrm{E}_{\mathrm{QM/MM}}(s^{\mathrm{i}+1}) - \mathrm{E}_{\mathrm{QM/MM}}(s^{\mathrm{i}}) \right) \right\rangle_{MM,i} \end{aligned} \tag{13-3}$$

where E^0_{QM} is the gas-phase energy of the QM subsystem, including the MP2 two-dimensional correction depending on the R_1 and R_2 coordinates, k_B is the Boltzmann's constant, T the temperature, and $\beta = \frac{1}{k_B T}$. The QM/MM interaction contribution to the free energy difference between two different values of s is obtained by averaging the QM/MM interaction energy (including the polarization energy) over all the MM coordinates of the system obtained for a particular value of the s coordinate. Then we added a term to include the contribution of the vibration of all the QM coordinates (except s) to the free energy profile, assuming a quantum harmonic treatment for the vibrational modes of the QM subsystem in the field created by the rest of the enzyme. In this way the chemical system flexibility is also considered. Then, the quasiclassical free energy profile can be obtained adding the contribution of the QM vibrational degrees of freedom as:

$$G(s^{\mathrm{j}}) - G(s^R) = \Delta E^{0,R\to j}_{QM} + \Delta G^{R\to j}_{QM/MM} + \Delta G^{j}_{vib,QM} \tag{13-4}$$

where

$$\Delta G^{j}_{vib,QM} = \sum_{x=1}^{3N-7}\left[\frac{1}{2}h\nu^{j}_{x} + k_B T \ln\left(1 - e^{-\frac{h\nu^{j}_{x}}{k_B T}}\right)\right] - \sum_{x=1}^{3N-6}\left[\frac{1}{2}h\nu^{R}_{x} + k_B T \ln\left(1 - e^{-\frac{h\nu^{R}_{x}}{k_B T}}\right)\right] \tag{13.5}$$

The frequencies were obtained projecting out the contribution of the reaction coordinate in the Hessian for all the structures but the reactant. Thus, the vibrational term was obtained after summation over the 3N-6 highest-frequency vibrational modes (being N the number of QM atoms) in the reactant structure and 3N-7 for the rest [10]. The six lowest-frequency vibrational modes were assumed to correspond to the relative motion of the MM system with respect to the QM system, which is already included in the second term of the second member of Eq. (13-4). The quantitative importance of this term is undeniable since it can amount to more than 3 kcal mol^{-1} [52]. The activation free energy is then obtained as:

$$\Delta G^{\ddagger} = \max_j\left(G(s^j) - G(s^R)\right) \tag{13-6}$$

These results indicated that the two transfers (HT and PT) take place in a single concerted process although the degree of asynchronicity may depend on the particular reaction path followed.

We computed FEPs for the pyruvate to lactate transformation starting at four different TS structures located from snapshots of a QM/MM MD simulation carried out in LDH. Four different profiles displaying different free energy barriers were obtained, as observed in Figure 13-2. The existence of different reaction paths, with different free energy barriers, is in accordance with our previous conclusions obtained from internal energy optimizations of different PESs. Effectively, each of these paths led to different RS showing distinctive features related to the

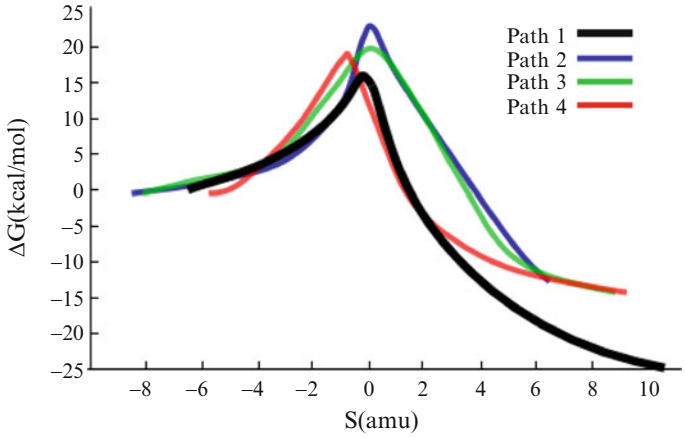


Figure 13-2. Free energy profiles for the pyruvate to lactate transformation catalyzed by LDH, obtained for four reaction paths studied starting from different reactant valleys (Adapted from [52])

orientation of some residues in the active site. Our finding of several different, well characterized reaction paths, with its own individual reaction rate constant, agrees with the dynamical disorder of the enzymes exhibiting fluctuations of catalytic rate constants, as shown by single-molecule enzymatic studies [41, 42]. We will comment further on this point below.

13.4.2. Potential of Mean Force (PMF)

Free energies profiles can be also obtained as a PMF associated to a particular reaction coordinate [8]. Selection of this reaction coordinate should be based on the exploration of the PES including the environment, or even better, on IRCs traced down to the corresponding products and reactants valleys from transition structures located and characterized in the enzyme active site.

The main advantage of PMF technique is that contributions of all the degrees of freedom of the system, except the reaction coordinate, are naturally considered and then it provides a magnitude more easily comparable to the activation free energy. The PMF is in fact related to the normalized probability of finding the system at a particular value of the chosen coordinate by Eq. (13-7):

$$W(\xi) = C - kT \ln \int \rho(\mathbf{r}^N)\delta(\xi(\mathbf{r}^N) - \xi)d\mathbf{r}^{N-1} \tag{13-7}$$

The activation free energy can be then expressed as [54]:

$$\Delta G^{\ddagger}(\xi) = W(\xi^{\ddagger}) - \left[W(\xi^R) + G_{\xi}(\xi^R)\right] \tag{13-8}$$

where the superscripts indicate the value of the reaction coordinate at the RS and TS and $G_{\xi}(\xi^R)()$is the free energy associated with setting the reaction coordinate to

a specific value at the reactant. Normally this last term makes a small contribution [55] and the activation free energy is directly estimated from the PMF change between the maximum of the profile and the reactant's minimum:

$$\Delta G^{\ddagger}(\xi) \approx W(\xi^{\ddagger}) - W(\xi^{R}) = \Delta W^{\ddagger}(\xi) \tag{13-9}$$

The selection of the reaction coordinate is trivial when the mechanism can be driven by a single internal coordinate or a simple combination (as the anti-symmetric combination of two interatomic distances). However this is not the case for the LDH catalyzed reaction, were several distances must be considered. Any attempt to use a single internal coordinate or a combination of them failed to produce a smooth free energy profile. Instead we were compelled to obtain a much more computationally demanding two dimensional PMF (2D-PMF) using two coordinates: the anti-symmetric combination $r_{CnicH'} - r_{CpyrH'}$ defining the HT, and the anti-symmetric combination $r_{NhisH''} - r_{OpyrH''}$ corresponding to the PT. The 2D-PMF is related to the probability of finding the system at particular values of these two coordinates:

$$W(\zeta_1, \zeta_2) = C' - kT \ln \int \rho(\mathbf{r}^N)\delta(\zeta_1(\mathbf{r}^N) - \xi_1)\delta(\zeta_2(\mathbf{r}^N) - \xi_2)d\mathbf{r}^{N-2} \tag{13-10}$$

To estimate the activation free energy from this quantity one dimensional PMF profiles mat be recovered tracing a maximum probability reaction path on the 2D-PMF surface and integrating over the perpendicular coordinate.

The 2D-PMFs corresponding to the reaction in aqueous solution, in the native enzyme and in the mutated form were obtained using the weighted histogram analysis method (WHAM) combined with the umbrella sampling approach [56, 57]. Simulations were performed at different values of ζ_2 (61 simulations in a range from −1.5 Å to 1.5 Å with an umbrella force constant of 1,255 kJ·mol^{-1}·A^{-1}) for each value of ζ_1 (61 simulations in a range from −1.5 Å to 1.5 Å with a force constant of 1,255 kJ·mol^{-1}·A^{-1},). Consequently, there are 3,721 windows per 2D-PMF. In each window, 10 ps of relaxation were followed by 20 ps of production with a time step of 0.5 fs. The values of the variables sampled during the simulations were then pieced together to construct a full distribution function from which the 2D-PMFs were obtained. A AM1/MM energy function corrected by means of single-point MP2/6-31G* calculations was employed for these simulations.

The resulting MP2-corrected 2D-PMF is presented in Figure 13-3. As observed, the reaction is described as a concerted mechanism with a free energy barrier of 26 kcal·mol^{-1}. It must be pointed out that during the 2D-PMF calculations all the degrees of freedom, except the selected reaction coordinates are being sampled, while using the FEP technique the degrees of freedom of the chemical system are only sampled in an uncoupled way using a harmonic approximation. Nevertheless, the obtained reaction path coincides with one of the FEP curves presented in Figure 13-2.

Thus, in spite of the statistical methodology employed to extract free energy changes, a precise value for the free energy barrier for the LDH catalyzed reaction cannot be reached. The origin of this a priori limitation is in the fact that many

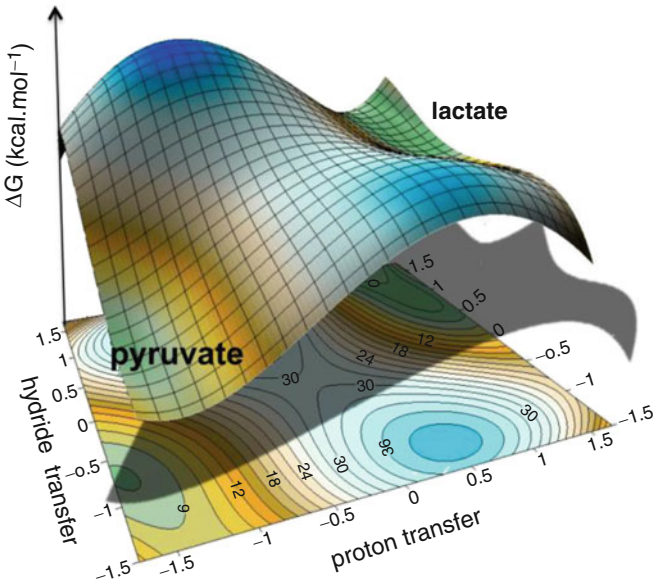


Figure 13-3. MP2 corrected AM1/MM 2D PMF for the reduction of pyruvate to lactate catalyzed by LDH. Proton transfer and hydride transfer coordinates correspond to the anti-symmetric combination of the forming and breaking bonds for each transfer

different conformational valleys can be present in huge flexible molecular systems, as already pointed out. This feature of enzyme systems provokes the requisite of large scale conformational search to arrive at a proper averaged value for geometries or energetics in these catalyzed reactions.

13.5. LARGE SCALE CONFORMATIONAL CHANGES AND AVERAGED KINETIC PROPERTIES

In addition to local changes occurring within a valley, which can be conveniently sampled during FEP or PMF calculations, more global conformational changes may also exist in proteins, in such a way that the interconversion rates among the corresponding conformational substrates (or valleys) can be much slower than the catalytic rate constants. All the stationary points of a given reactant valley would share a distinct global conformational characteristics. In these conditions molecular dynamics simulations started in a given valley are in practice unable to jump to another valley, because the free energy barrier for the conformational change is too high. Then while a local equilibrium exists inside each valley, a global equilibrium including all the valleys is very unlikely. The existence of relatively stable reactant valleys (or long-lived conformational substrates) in comparison with the time scale of the enzymatic reactions is the required molecular basis to understand the

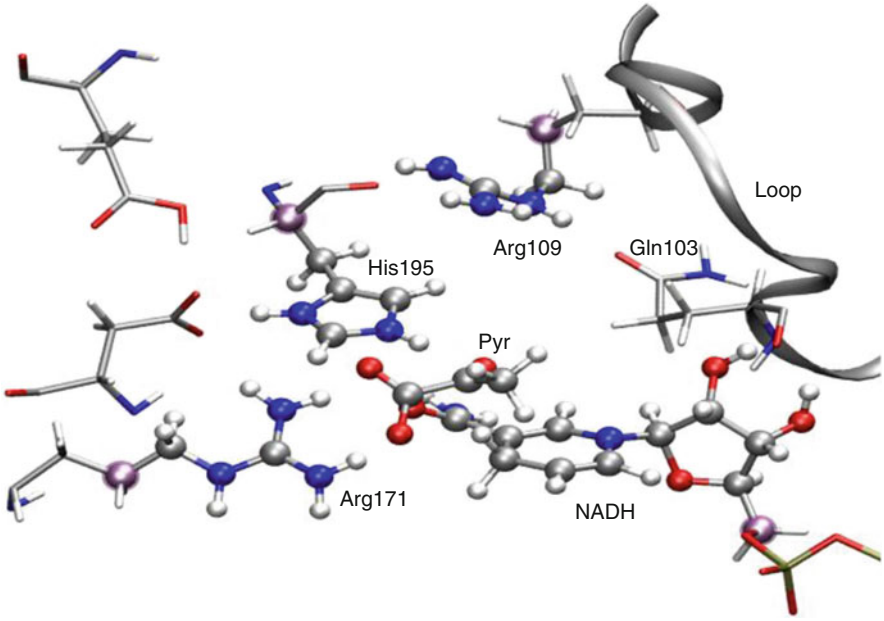


Figure 13-4. Snapshot of the reactants in LDH showing the positioning of the active loop that closes the active site

dynamic disorder observed in single molecular studies of some enzyme mechanisms [58–61] including LDHs [41, 42]. Slow transitions among different conformational substrates (reactant valleys) of the enzyme having different characteristics may cause fluctuations in the observed catalytic rate constant. Everything happens as if each reactant valley acts as an independent enzyme with a different rate constant, as commented above.

The four reactant valleys obtained during the FEPs calculations on the LDH catalyzed reaction were explored in detail by means of relatively long stochastic boundary molecular dynamics simulations [52]. We found some differences among these complexes related to the positioning of the flexible loop closing the active site (see Figure 13-4). This is translated to a different enzyme-substrate pattern of interactions, a different number of water molecules in the active site and other differences related to the key residue Arg109.

As presented before, the four reaction valleys analyzed displayed significant differences in the calculated rate constants (see Figure 13-2), so reproducing the experimental dispersion of reaction rate constants in single-molecule enzyme studies. A simple arithmetic average of the four individual reaction rate constants gives 696 s^{-1}, a value very close to the experimental macroscopic lower limit estimation. However, averaging among different reactant complexes to obtain the macroscopic rate constant is a question without an obvious answer. To obtain a definitive answer one would need to trace the free energy profiles connecting them. Our estimation

displays significant differences but cannot be taken as conclusive. Anyway, taking into account that the movement of the loop is in fact the rate-limiting step in the global reaction we can imagine a scenario where the reactant complexes interconversion is slower that the chemical reaction and then global equilibrium cannot be assumed. We have also calculated kinetic isotope effects for each of the four reaction valleys. Our calculated values [52] range between 3.47 and 5.05 in good agreement with experimental determinations for this enzyme [62], LDH mutants and other dehydrogenase enzymes where hydrogen transfer is rate limiting [19, 24].

13.6. CONCLUSIONS

LDH catalyzed reaction has been used in this chapter as a conductor wire to present the evolution and difficulties on computing methods to model chemical reactions in enzymes, since the early calculations based at semiempirical level carried out in gas phase to the recent sophisticated simulations based on hybrid Quantum Mechanical/Molecular Mechanics Dynamics (QM/MM MD) schemes. LDH catalyzes the reversible transformation of pyruvate into lactate. The chemical step consists in a hydride and a proton transfer from the cofactor (NADH) and a protonated histidine (His195), respectively. This fact has generated a lot of controversy about the timing of both transfers in the active site, as well as the role of the different aminoacids of the active site and problems related with the flexibility of the protein.

Our last findings, obtained from FEP and 2D PMF calculations by means of high level MP2/AM1/MM statistical simulations within a fully flexible molecular model, reveal the existence of several different, well characterized reactant valleys, each one acting as an independent enzyme, with its own individual reaction rate constant, agrees with the dynamical disorder of the enzymes exhibiting fluctuations of catalytic rate constants, as shown by single-molecule enzymatic studies. The insights obtained from the theoretical study of this enzyme clearly show that we are just starting to understand the whole complexity and richness of the dynamic/kinetic behaviour of enzymes. Clearly, further developments of not only computational tools but also of adequate theoretical frameworks are needed to achieve this goal. In this sense, we strongly think that LDH provides an excellent example to be analyzed, because it seems that as deeper studies are carried out more and more surprises are revealed.

ACKNOWLEDGEMENTS

We thank the Spanish Ministry *Ministerio de Ciencia e Inovación* for project CTQ2009-14541, Universitat Jaume I - BANCAIXA Foundation for projects P1·1B2005-13, P1·1B2005-15 and P1·1B2005-27, and Generalitat Valenciana for *Prometeo*/2009/053 project. We are also grateful to Prof. Ian H. Williams and Prof.

J. Andrés for fruitful discussions. The authors also acknowledge the Servei d´ Informatica, Universitat Jaume I for generous allotment of computer time. V. Moliner would like to thank the Spanish Ministry *Ministerio de Ciencia e Innovación* for traveling financial support, project PR2009-0539.

REFERENCES

1. Ertl G, Gloyna T (2003) Zeitschrift Fur Physikalische Chemie-Int J Res Phys Chem Chem Phys 217 (10):1207–1219
2. Ostwald W (1902) Phys Z 3:313–322
3. Ostwald W (1910) Ann Naturphil 9:1
4. Corma A (2004) Cat Rev Sci Eng 46(3–4):369–417
5. Garcia-Viloca M, Gao J, Karplus M, Truhlar DG (2004) Science 303(5655):186–195
6. Warshel A, Sharma PK, Kato M, Xiang Y, Liu HB, Olsson MHM (2006) Chem Rev 106 (8):3210–3235
7. Truhlar DG (2008) J Am Chem Soc 130(50):16824–16827
8. Field MJ (2007) A practical introduction to the simulation of molecular systems. Cambridge University Press, Cambridge
9. Warshel A, Levitt M (1976) J Mol Biol 103(2):227–249
10. Kollman P (1993) Chem Rev 93(7):2395–2417
11. Kollman PA, Kuhn B, Donini O, Perakyla M, Stanton R, Bakowies D (2001) Acc Chem Res 34 (1):72–79
12. Marti S, Andres J, Moliner V, Silla E, Tuñón I, Bertran J, Field MJ (2001) J Am Chem Soc 123 (8):1709–1712
13. Marti S, Roca M, Andres J, Moliner V, Silla E, Tuñón I, Bertran J (2004) Chem Soc Rev 33 (2):98–107
14. Marti S, Andres J, Moliner V, Silla E, Tuñón I, Bertran J (2008) Chem Soc Rev 37(12):2634–2643
15. Gao JL, Truhlar DG (2002) Annu Rev Phys Chem 53:467–505
16. Villa J, Warshel A (2001) J Phys Chem B 105(33):7887–7907
17. Warshel A (1998) J Biol Chem 273(42):27035–27038
18. Lin H, Truhlar DG (2007) Theor Chem Acc 117(2):185–199
19. Clarke AR, Wigley DB, Chia WN, Barstow D, Atkinson T, Holbrook JJ (1986) Nature 324 (6098):699–702
20. Hart KW, Clarke AR, Wigley DB, Chia WN, Barstow DA, Atkinson T, Holbrook JJ (1987) Biochem Biophys Res Commun 146(1):346–353
21. Clarke AR, Atkinson T, Holbrook JJ (1989) Biochem Sci 14:145–148
22. Badcoe IG, Smith CJ, Wood S, Halsall DJ, Holbrook JJ, Lund P, Clarke AR (1991) Biochemistry 30 (38):9195–9200
23. Deng H, Zheng J, Clarke A, Holbrook JJ, Callender R, Burgner JW (1994) Biochemistry 33 (8):2297–2305
24. Clarke AR, Wigley DB, Barstow DA, Chia WN, Waldman ADB, Hart KW, Atkinson T, Holbrook JJ (1987) Biochem Soc Trans 15(1):152–153
25. Andres J, Moliner V, Krechl J, Silla E (1993) Bioorg Chem 21(3):260–274
26. Andres J, Moliner V, Safont VS (1994) J Chem Soc Faraday Trans 90(12):1703–1707
27. Andres J, Moliner V, Krechl J, Silla E (1995) J Chem Soc Perkin Trans 2(7):1551–1558
28. Krechl J, Kuthan J (1988) Theochem J Mol Struct 47:239–244

29. Wilkie J, Williams IH (1992) J Am Chem Soc 114(13):5423–5425
30. Wilkie J, Williams IH (1995) J Chem Soc Perkin Trans 2(7):1559–1567
31. Ranganathan S, Gready JE (1994) J Chem Soc Farad Trans 90(14):2047–2056
32. Yadav A, Jackson RM, Holbrook JJ, Warshel A (1991) J Am Chem Soc 113(13):4800–4805
33. Siegbahn PEM, Himo F (2009) J Biol Inorg Chem 14(5):643–651
34. de la Lande A, Gerard H, Moliner V, Izzet G, Reinaud O, Parisel O (2006) J Biol Inorg Chem 11 (5):593–608
35. de la Lande A, Parisel O, Gerard H, Moliner V, Reinaud O (2008) Chem Eur J 14(21):6465–6473
36. Alhambra C, Corchado J, Sanchez ML, Garcia-Viloca M, Gao J, Truhlar DG (2001) J Phys Chem B 105(45):11326–11340
37. Moliner V, Turner AJ, Williams IH (1997) Chem Commun 14:1271–1272
38. Turner AJ, Moliner V, Williams IH (1999) Phys Chem Chem Phys 1(6):1323–1331
39. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) J Comput Chem 4(2):187–217
40. Cramer CJ, Truhlar DG (1999) Chem Rev 99(8):2161–2200
41. Xue QF, Yeung ES (1995) Nature 373(6516):681–683
42. Tan WH, Yeung ES (1997) Anal Chem 69(20):4242–4248
43. Ranganathan S, Gready JE (1997) J Phys Chem B 101:5614–5618
44. Weiner SJ, Kollman PA, Nguyen DT, Case DA (1986) J Comput Chem 7(2):230–252
45. Moliner V, Williams IH (2000) Chem Commun 19:1843–1844
46. Gilson MK (1993) Proteins Struct Funct Genet 15(3):266–282
47. Antosiewicz J, McCammon JA, Gilson MK (1994) J Mol Biol 238(3):415–436
48. Field MJ, David L, Rinaldo D. Personal Communication
49. Ferrer S, Silla E, Tuñón I, Oliva M, Moliner V, Williams IH (2005) Chem Commun 47:5873–5875
50. Marti S, Moliner V, Tuñón I (2005) J Chem Theor Comput 1(5):1008–1016
51. Swiderek K, Paneth P (2010) J Phys Chem B 114(9):3393–3397
52. Ferrer S, Tuñón I, Marti S, Moliner V, Garcia-Viloca M, Gonzalez-Lafont A, Lluch JM (2006) J Am Chem Soc 128(51):16851–16863
53. Zhang YK, Liu HY, Yang WT (2000) J Chem Phys 112(8):3483–3492
54. Schenter GK, Garrett BC, Truhlar DG (2003) J Chem Phys 119(12):5828–5833
55. Roca M, Moliner V, Ruiz-Pernia JJ, Silla E, Tuñón I (2006) J Phys Chem A 110(2):503–509
56. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM (1992) J Comput Phys 13 (8):1011–1021
57. Torrie GM, Valleau JP (1977) J Comput Phys 23(2):187–199
58. Kou SC, Cherayil BJ, Min W, English BP, Xie XS (2005) J Phys Chem B 109(41):19068–19081
59. Smiley RD, Hammes GG (2006) Chem Rev 106(8):3080–3094
60. Lu HP, Xun LY, Xie XS (1998) Science 282(5395):1877–1882
61. Yang H, Luo GB, Karnchanaphanurach P, Louie TM, Rech I, Cova S, Xun LY, Xie XS (2003) Science 302(5643):262–266
62. Seravalli J, Huskey WP, Schowen KB, Schowen RL (1994) Pure Appl Chem 66(4):695–702

CHAPTER 14

ENZYME DYNAMICS AND CATALYSIS: INSIGHTS FROM SIMULATIONS

John D. McGeagh, Adrian J. Mulholland
Centre for Computational Chemistry, School of Chemistry, University of Bristol, Bristol, BS8 1TS, UK
e-mail: adrian.mulholland@bristol.ac.uk

Abstract: Molecular modelling and simulation are making increasingly important contributions to the study of the structure and function of biological macromolecules. An area of particular current interest and debate is that of enzyme catalysis, and the role of protein dynamics in enzyme-catalysed reactions. Simulations allow enzyme catalytic mechanisms and protein dynamics to be investigated at the atomic level (e.g. with combined quantum mechanics/molecular mechanics (QM/MM) calculations and atomistic molecular dynamics simulations). This level of detailed analysis is beyond what is currently possible in experiments for reactions in enzymes, and simulations therefore have a crucial role to play in testing hypotheses and aiding in the interpretation of experimental data. Biomolecular simulations will therefore be crucial in resolving current controversies about the role of protein dynamics in enzyme catalysis. In this chapter we describe some recent simulations which have contributed to an understanding of enzyme mechanism, dynamics and function.

Keywords: Quantum mechanics, Molecular mechanics, QMMM, NAMD, Biomolecular simulation, Molecular modelling, Molecular dynamics, Conformational changes, Functional dynamics, Computational enzymology

14.1. INTRODUCTION

Proteins play crucial roles in virtually all biological processes from immune protection to transmission of nerve impulses. One vital role of proteins is to provide catalysts for biochemical reactions: such proteins are termed enzymes (a wider definition of enzyme is any biological catalyst, thus including ribozymes). Enzyme action is crucial for all forms of life, as biosynthesis and metabolism would otherwise be impossible. Understanding enzyme action is fundamentally and practically important. Enzymes are remarkable molecular devices that

P. Paneth and A. Dybala-Defratyka (eds.), *Kinetics and Dynamics*, Challenges and Advances in Computational Chemistry and Physics 12,
DOI 10.1007/978-90-481-3034-4_14,

provide enormous catalytic power and specificity. To understand fully how an enzyme functions requires knowledge of its complete catalytic cycle, from binding of the substrates through to release of the products and all intermediate steps, in particular how chemical transformations occur in these biological macromolecules. As with all proteins, enzymes have complex dynamics, and there has been a lot of debate about the possible role of protein dynamics in catalysis [1, 2].

Biomolecular simulations, with the aid of advances in computer power, are being applied increasingly and can potentially provide detailed insights into biological processes at the atomic level [3–9]. Simulations can complement experiments on enzyme reactions by allowing transient species such as transition states and reaction intermediates to be studied, as well as allowing proposed reaction mechanisms to be probed [3, 10, 11]. Recent applications of biomolecular simulation include studies of protein folding and conformational changes [6], association of proteins with small molecules [12] or other proteins [13], structure-based drug design [14], ion binding [15], transport across membranes [16] and the modelling and analysis of enzyme mechanisms [17]. In this chapter the focus is on atomistic simulation of enzymes, including studies of enzyme mechanism, and of enzyme dynamics.

14.2. CHALLENGES IN BIOMOLECULAR SIMULATION

Simulating biological macromolecules in atomic detail is challenging because of their large size, and the need often to include at least a representative part of their environment, be it the solvent, membrane, and any cofactors which may be bound to a protein. For an enzyme simulation to be reliable, an accurate structure of the protein is required. Experimental methods, such as X-ray Crystallography and NMR, have provided detailed insights into the three-dimensional structure of enzymes [18]. These experimental structures can be used as starting points in biomolecular simulations. The most widely used experimental technique for studying protein structure is X-ray crystallography. This technique requires a well-ordered crystal which can be used to determine positions of heavy atoms, with structures ranging from low resolution (where perhaps only an outline of the protein shape may be obtained) to higher resolution (e.g. less than 1.5 Å) where most (ordered) atomic positions can be determined, at least for heavy atoms [18]. However, it must be remembered that, even in high resolution structures, there can be considerable uncertainty due to the dynamic nature of proteins which can cause conformational variability. For example, Pentikäinen et al. [6] performed molecular dynamics simulations of scavenger decapping enzyme (DcpS), which indicate that the crystallographically observed symmetric structure is only a minor conformation in solution. These results indicate that in solution this enzyme undergoes significant structural changes.

Other considerations in preparing a molecular model for simulation include possible alternative conformations for some groups, choice of protonation states

for acidic and basic groups (hydrogen atoms are usually not observed in X-ray structures because of their low electron density), and the identification of atoms in histidine and glutamine sidechains. The approximately symmetric nature of histidine and glutamine sidechains means identification of the correct rotamer from crystallographic electron density can be difficult. A modeller should ensure that the correct conformers of these residues are used in the starting structure. Hydrogen bonding networks can be judged by eye, but there are programs available, such as WHATIF [19], which position polar hydrogen atoms in protein structures by optimizing the total hydrogen bond energy. This process can also help to identify the protonation states of ionisable residues such as histidine, aspartate and glutamate residues and suggest if residues should be 'flipped' to compensate for crystallographic misassignments. Likely protonation states can be identified, for example, using the program PROPKA [20] which assigns pK_a values to residues, though it remains very difficult to assign pK_as for titratable groups in proteins reliably. A modeller must also be aware of any experimental techniques utilised in the crystallization process, as mutations are sometimes used to aid crystallization. It may be necessary to alter the crystal structure to accommodate such alterations in the construction of a model of the wild type enzyme.

The number of structures of biological macromolecules, as determined by X-ray crystallography or NMR studies, is large and ever increasing. A challenge for a modeller is to attempt to make use of this wealth of information. The field of biomolecular simulations is currently evolving and, as yet, is not at the stage where quantitative predictions can routinely be made [21]. Computational predictions should therefore be validated by comparison to (or prediction of) experimental results: such validations are best carried out for well characterized systems, with a lot of experimental (and preferably modelling or simulation) data to compare to. In the area of enzyme reaction mechanism modelling, a prototypical system is chorismate mutase which has been used for method development and which has become an enzyme of choice to address the question of how enzymes catalyse reactions [7].

One important experimental observable is the rate of an enzyme reaction (usually an apparent rate for the overall reaction, but sometimes rates for individual steps can be measured). If a single step is rate determining, it may either be a chemical step or a conformational change (e.g. related to substrate binding or product release). This can complicate comparison with calculation results unless a rate-determining step is known. Any proposed mechanism should have an energy barrier which is consistent with the measured rate, within the range of error of the modelling technique (and the experimental results). Techniques such as site-directed mutagenesis can help in the identification of residues involved in substrate binding and catalysis, but the results can be misleading due to mutations inducing subtle changes in enzyme structure [22] or mechanism [23]. Catalytic mechanisms can be proposed using chemical and structural knowledge. However due to the complexity of enzymes, several mechanisms will almost inevitably be possible.

Once a suitable starting structure has been obtained an appropriate simulation or modelling method must be selected which balances computational expense with

accuracy, delivering a reliable result in a reasonable amount of time. The first atomistic molecular dynamics simulation of a protein utilised an early molecular mechanics (MM) approach for the bovine pancreatic trypsin inhibitor. These simulations showed that, over a short (9 ps) molecular dynamics (MD) run, there were significant structural fluctuations. This led the authors to conclude that the atomic displacements were dominated by collisions with neighbouring atoms, much like a liquid [24]. Since this pioneering early work, computational power has increased exponentially, allowing for longer simulation timescales and bigger systems to be probed in biomolecular simulations while also allowing the inclusion of a significant part of the local environment – be it solvent, membrane or cofactors [25, 26]. During the same time, the force fields describing biomolecular structure and interactions have been refined to provide results more consistent with experiment [9, 27].

There are important limitations to MM force fields, for example, electrostatic interactions are typically represented by including a point charge on each atom in current biomolecular MM force fields [28]. This simple point charge model cannot capture the full electrostatic properties of a molecule. Also electronic polarization is usually not included, i.e. the atomic charges are invariant, they do not change in response to changes in the molecular environment or conformation. Such a simple model makes MM calculations fast, but also imposes limitations. MM–MD simulations can now be run for long enough timescales to provide insight into conformational changes of enzymes [6]. However, questions regarding the catalytic mechanism cannot be addressed by standard MM methods. For this, a method capable of modelling electronic structure must be used. Methods based on quantum mechanics (QM) – quantum chemical – approaches can be used to study reactions in molecular systems. Quantum chemical techniques include ab initio molecular orbital methods or approaches based on density functional theory (DFT): these can currently deal practically with molecular systems containing of the order of tens of atoms (DFT methods can treat larger systems than correlated ab initio calculations. This prevents their routine application to large systems such as biological macromolecules (often upwards of 10,000 atoms).

One of the most significant advances in the field of modelling enzyme-catalysed reactions has been the development of combined quantum mechanics/ molecular mechanics (QM/MM) methods [29, 30]. The basic idea is very simple in that a small region of the enzyme, usually the active site and substrate(s), is modelled using a QM (quantum chemical, molecular electronic structure) method and the surrounding region is modelled with an MM method. The QM region allows for bond making and/or breaking and electronic rearrangements to occur in the active site region, with the effect of the surroundings included by modelling part (or all, depending on the system size) of the enzyme, and its environment, using MM.

A technique that is becoming increasingly important in biomolecular simulations is the Car–Parinello (or ab initio) MD method. First proposed some 25 years ago [31] the application involves the combination of molecular dynamics with density functional theory (DFT). In effect the electronic degrees of freedom are included as dynamical variables. This approach, being based on a quantum

mechanical (electronic structure) method, allows for MD simulations with the possibility of reactions taking place as the dynamics progress. Currently the use of this approach is limited by several factors. Due to the computational expense of using a QM method in an MD simulation, long timescales are out of reach (particularly for large systems) and hence reactions which occur over long timescales are unobtainable. However the potential advantages of this method are evident as the inclusion of electronic structure methods provides the ability to describe bond making and breaking.

At the other end of the spectrum of biomolecular simulations, in terms of computational expense, size of simulation system and timescale, are 'coarse-grained' methods [9]. These are being developed to provide a route to longer timescales in biomolecular simulations than are possible with atomistic (e.g. MM or QM/MM) representations. Coarse-graining works by grouping atoms together to reduce the number of explicit interaction sites [32]. Coarse-graining techniques were introduced very early in the history of biomolecular simulation. In the 1970s Levitt and Warshel introduced a model for globular proteins using just two coarse-grained particles per residue: one was centred on the α-carbon and one represented the side chain atoms [33, 34]. This model was shown to be effective at reducing the computational demand for simulations of proteins whilst also giving useful results; it was able to refold bovine pancreatic trypsin inhibitor from a denatured state [33]. By reducing the number of interaction sites, coarse-grained simulations are computationally less expensive then the all-atom models. In addition to this the coarse-grained simulations remove the stiffest degrees of freedom from the model, e.g. the carbon–hydrogen bond vibrational modes, therefore allowing for longer time steps which aids the increase in speed of simulations. Coarse-grained models have developed significantly since the pioneering work of Levitt and Warshel, for example leading to forcefields which show good agreement with all-atom simulations and experimental data [35]. Schemes to couple together different levels of simulation (e.g. coarse-grained and atomistic) are also being developed: this area of multi-level modelling is a highly active field [36].

14.3. PROTEIN DYNAMICS AND ENZYME CONFORMATIONAL CHANGES

An important concept, which molecular dynamics simulations have significantly aided the development of, is that proteins are not rigid but undergo structural fluctuations. Most motions are "random", but some fluctuations have functional importance (e.g. in forming reactive conformations [37]). There is increasing evidence to suggest that some important motions are inherent features of some enzymes' structures [2, 6]. Enzyme structure should therefore be considered as an ensemble of conformational states rather than a static structure.

Many enzymes undergo large conformational changes as part of their catalytic cycles, e.g. changing from an 'open' (unliganded) to a 'closed' (substrate bound)

form. Such conformational changes can involve the motion of one protein domain relative to another, but are not limited to this type of structural change. Experimental techniques can give crystal structures of an unbound and bound state, and NMR gives very useful information on protein dynamics and structure (especially for smaller proteins), but it is hard to study conformational changes directly. However Wolf-Watz et al., who performed an NMR study on hyperthermophilic and mesophilic homologues of adenylate kinase showed that a large-scale protein conformational change is the rate-determining step for enzyme reaction turnover [38]. The opening of the nucleotide-binding lids was found to be rate-limiting for both enzymes and the reduced activity of the hyperthermophilic homologue at mesophilic temperatures was found to be caused solely by a slower lid-opening rate. The results suggest that, in many enzymes, evolution has fine-tuned the chemical steps to be very fast and thus often it is large-scale conformational changes which limit the catalytic turnover rate. Pathways of conformational change can be generated by computational methods such as the *Dyndom* program [39], for example: this program takes two crystal structures – e.g. one in the bound and one in the unbound form – and, using a clustering algorithm, identifies domains from clusters of rotation vectors. This analysis also defines so-called hinge axes around which the domains move relative to each other. Molecular dynamics simulations can also potentially probe such transitions, although conformational changes may occur on longer timescales than are routinely accessible in current simulations of large proteins. Due to this, regions of low probability/high free energy are sampled rarely, making convergence of free energy barriers (or differences) for rare events extremely difficult. Techniques such as steered, driven, replica-exchange, locally enhanced sampling or digitally filtered molecular dynamics may help in accelerating changes, but could run the risk of following an unrealistic path.

Simulations and experiments suggest that enzymes have an intrinsic ability to sample conformations that meet functional requirements, that is to say, for example, that enzymes typically sample both their open and closed forms, irrespective of the presence of substrate, succeeded by the stabilization of one form (usually the closed) upon binding of the substrate [40], these motions are often termed "functionally relevant motions" [41].

Due to the ever-increasing timescales that (MM) molecular dynamics simulations can achieve, it has been necessary to develop advanced analytical techniques in order to visualise conformational changes on a global scale. Essential dynamics (ED) is fast becoming a critical tool in the analysis of MD trajectories [42]. In effect, ED is principal component analysis as applied to MD simulations. The theory is to separate the essential degrees of freedom, comprising of slow – global motions, from the "noise" (vibrations and rotations) of the simulation. This results in two or three principal components with highly correlated motion, these motions can often describe up to 75% of the total data within a trajectory and can aid in the analysis of enzyme dynamics to decipher if an enzyme is sampling functionally relevant conformations. The prominence of ED within the modelling community has driven the development of applications, such as interactive essential dynamics

(IED) [43] and Dynamite [44] which integrate with the widely used visualisation program visual molecular dynamics (VMD) [45] allowing the slow modes of enzyme motion to be visualised.

14.3.1. Scavenger Decapping Enzyme (DcpS)

Pentikäinen et al. have observed interesting conformational sampling effects, of potential functional relevance, in MD simulations of the scavenger decapping enzyme (DcpS) [6]. DcpS is a dimeric enzyme with two active sites, which catalyses the cap hydrolysis following mRNA degradation, a major step of gene regulation occurring at the level of mRNA turnover [46–48]. Eukaryotic mRNAs contain, at their 5′ end, a cap structure containing a guanosine methylated at the 7 position which protects them from 5′–3′ exonucleases. It also acts as recognition site factors involved in pre-RNA splicing, RNA transport and translation initiation. Interestingly, DcpS only catalyses cap removal of short chain mRNA oligomers (10–20 residues) leaving long capped mRNA molecules (over 100 residues) untouched, although binding can occur [49, 50]. This suggests that DcpS has evolved to prevent premature decapping of mRNA molecules.

DcpS has two distinct domains, the N- and C-terminal domains (Figure 14-1). The active site is comprised of residues from each domain. The unliganded crystal structure shows a very symmetrical structure with two open active sites. However Chen et al. [51], on observing the m^7GDP bound and the *apo* crystal structures, noticed a distinct difference in structure from the unbound to bound forms. The highly symmetric form, observed in the unliganded crystal structure differed significantly from the asymmetric structure of the m^7GDP bound form. From the crystallographic *B*-factors, Chen et al. proposed that the N-terminal domain in *apo*-DcpS displayed inherent flexibility [51].

Pentikäinen et al. performed atomistic 20 ns MD simulations which showed a strikingly asymmetric structure with the two C-terminal domains acting as one, in a cooperative manner [6]. The motions were observed by tracking a distance between the Cα atoms of Asp111 and Trp175 on the N- and C-terminal domains respectively. The motions showed the two subunits differed significantly from the crystal structure, with one domain forming an open structure while the other closes (Figure 14-2). These findings clearly demonstrate cooperativity between the two subunits and also show that functionally relevant conformations are sampled in the absence of the ligand for DcpS. Simulations suggest that the closure of both active sites is structurally impossible and that the domain-swapped dimer structure has evolved to give these striking dynamic properties.

Understanding how protein conformational changes relate to enzyme catalysis is a long-standing goal for enzymologists. Over the short period during which computational enzymology has existed, significant advances have been made. In particular the development and refinement of forcefields as well as the development of methods, such as QM/MM, which allow enzyme reactions to be investigated.

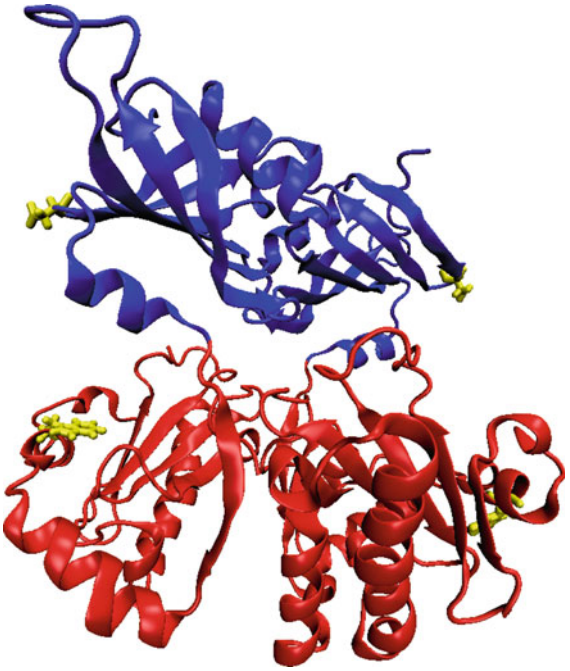


Figure 14-1. The crystal structure of human scavenger decapping enzyme (DcpS) (PDB code 1XML [51]) showing both N- and C-terminal domains, shown in *blue* and *red* respectively. The residues between which the distances are measured by Pentikäinen et al. [6] are shown by the *yellow highlighted residues*.

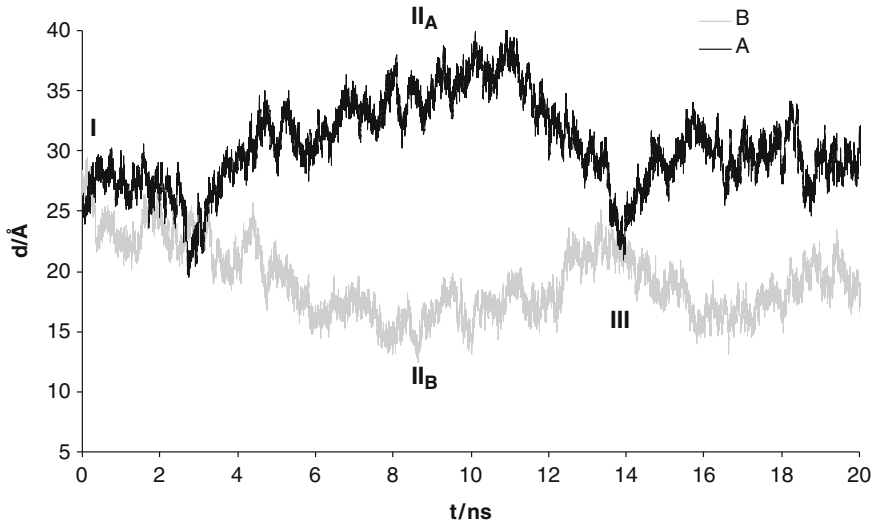
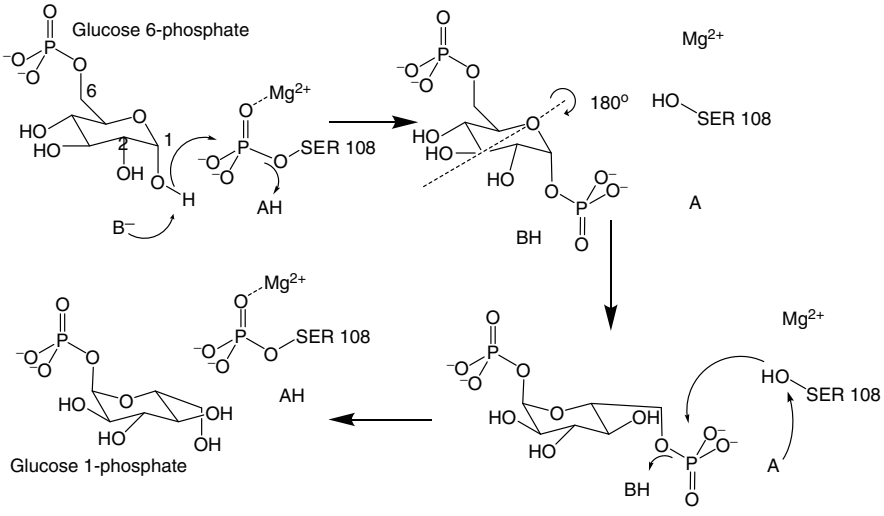


Figure 14-2. The distance between the N- and C-terminal domains of subunits A (*dark grey line*) and B (*light grey line*) during a molecular dynamics simulation of scavenger decapping enzyme (DcpS). Active site A is open at the state II_A and the active site B is closed at II_B with the (approximately) symmetric conformation regained at III [6]

These advances have enabled enzymes to be probed in greater depth. Conformational changes have long been known to be related to substrate binding and product release [52] but just how, or if, these motions contribute to enzyme reactions is largely unknown. It has been proposed that the free-energy landscape for enzyme catalysis can be thought of as 'mountain ranges' where there are multiple reaction coordinates with multiple ensembles of conformations [53].

14.3.2. Phosphomannomutase/Phosphoglucomutase

An enzyme which relies on a major conformational change triggered by the formation of an intermediate is phosphomannomutase/phosphoglucomutase (PMM/PGM). Once a phosphoryl group has been transferred from the enzyme to the substrate, the substrate is believed to be reorientated in a 180° flip to allow for a phosphoryl group to be transferred back to the enzyme (Scheme 14-1). In similar enzymes from the α-D-phosphohexomutase family, a ping-pong mechanism is used where the intermediate is released into bulk solution before binding again for the second phosphoryl transfer step [54]. However PMM/PGM re-orientates the 1,6-bisphosphorylated sugar intermediate without dissociation from the enzyme, i.e. this a processive enzyme. The formation of the bisphosphorylated sugar intermediate may create unfavourable interactions which distort the global motions of the enzyme away from its previous minimum, allowing conformational space



Scheme 14-1. The proposed mechanism of transformation of glucose-6-phosphate to glucose-1-phosphate as catalysed by phosphomannomutase/phosphoglucomutase (PMM/PGM): this reaction is believed to proceed via a 180° rotation of the bisphosphorylated intermediate before the enzyme removes a phosphate group, thus reactivating the enzyme [55]

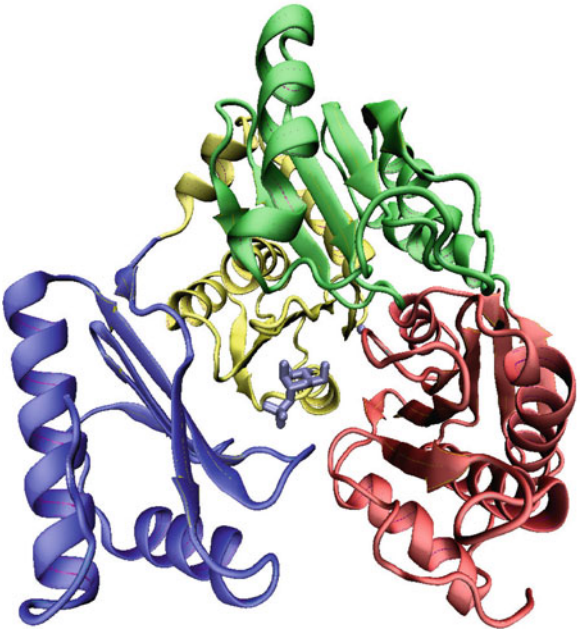


Figure 14-3. The enzyme phosphomannomutase/phosphoglucomutase (PMM/PGM) with the four domains shown by different colours. It has been proposed that a small movement in domain 4 (shown in *red*), when bound to the bisphosphorylated intermediate, creates a 'half open' conformation of the enzyme, which accommodates the 180° rotation of the intermediate without dissociation [55]

representing the slight opening of a subunit to be sampled. This opening would provide an increase in the active site cleft which would then provide the necessary space to reorient the intermediate. Naught et al. proposed that a small movement of domain 4 would create a 'half-open' conformer, thereby expanding the active site enough to allow a 180° rotation whilst preventing its dissociation [55]. The same authors also showed that one in every 15 catalytic cycles resulted in the dissociation of the bisphosphorylated intermediate [55]. The loss of this intermediate inhibits the enzyme and so an intermediate species must be re-bound to complete the reaction and reactivate the enzyme (Figure 14-3).

14.4. ENZYME CATALYSIS

Enzymes can achieve rate accelerations as large as a factor of 10^{19} over the same reaction in solution [56]. Probably the most influential theory for how enzymes achieve such large rate accelerations was originally proposed by Pauling in 1948: in modern terms, he proposed that enzymes are adapted to bind the transition state specifically and thereby lower the activation energy [57]. The reduction of activation energy increases the number of reactant molecules with enough energy to overcome the energy barrier and proceed to products. A schematic free energy

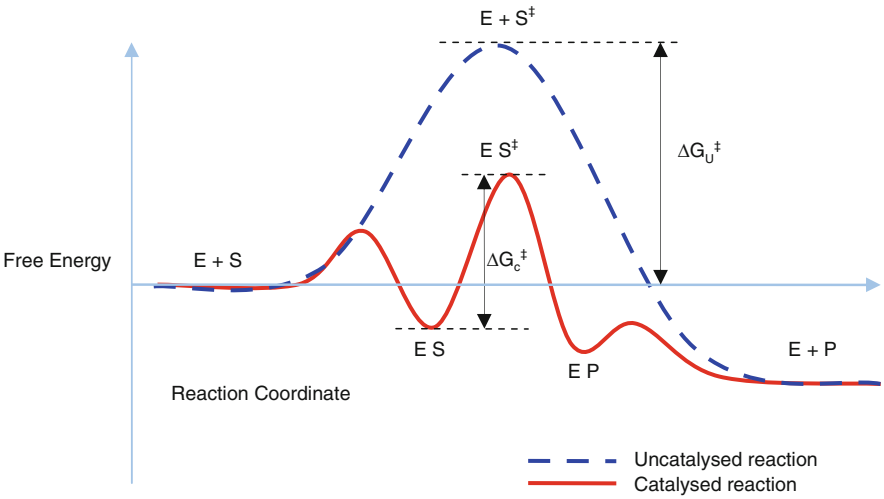


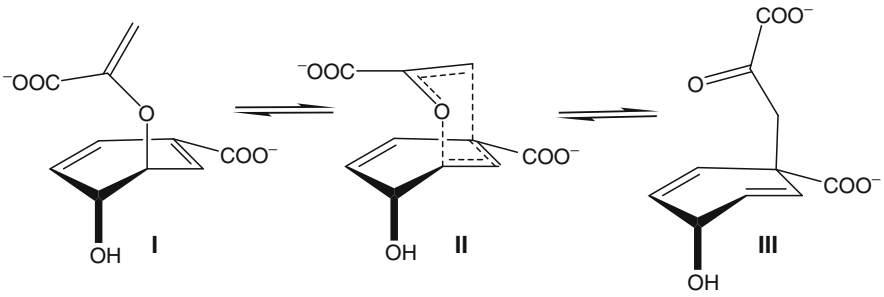
Figure 14-4. Schematic free energy profiles for an enzyme-catalysed reaction in comparison to the equivalent uncatalysed reaction in solution, proceeding from the enzyme (E) and substrate (S), through to the enzyme and product (P) in solution

reaction profile showing an enzyme-catalysed reaction and for the equivalent uncatalysed reaction in solution is shown in Figure 14-4 where E corresponds to enzyme, S to substrate and P to product. It has been shown by Warshel et al. that electrostatic effects make the largest contribution to catalysis (comparing enzyme-catalysed reactions to equivalent 'reference' reactions in solution) [58]. In particular, enzyme active sites generally provide a pre-organized polar environment that stabilizes the transition state and/or unstable intermediates. In contrast, in water (the nonenzymic reaction), the reorganization energy is significantly higher [59].

14.4.1. Chorismate Mutase

Chorismate mutase has proved a useful enzyme in the development of theories of catalysis. It catalyses the Claisen rearrangement of chorismate to prephenate (Scheme 14-2), as a key step in the Shikimate pathway for biosynthesis of aromatic amino acids in plants, fungi and bacteria. This enzyme, therefore, provides a possible target for the development of new herbicides, fungicides and antibiotics.

Chorismate mutase is an ideal model system for the investigation of enzyme catalysis as it is the only known example of a biocatalysed pericyclic reaction that also occurs via an identical mechanism in solution. This allows for direct comparisons to be made between the reactions in the enzyme and in solution. The reaction catalysed by this enzyme has been widely studied by experimental [60–62] and computational [63–69] techniques and due to the large amounts of data available, the enzyme has been used for testing various computational modelling methods [70–72].



Scheme 14-2. The Claisen rearrangement of chorismate to prephenate catalysed by the enzyme chorismate mutase. The chorismate structure (I), the transition state structure (II) and the prephenate product (III) are shown

Guildford et al. carried out experiments and eliminated many plausible mechanistic explanations; they found the kinetic data to be consistent with a non-dissociative mechanism suggesting that the rate acceleration might be due to general base catalysis [73]. However after the release of an enzyme–transition state analogue [74] complex the active site was observed to be devoid of any suitable residues to act as a base for catalysis. Mutation experiments [75–77] identified several residues as important for catalysis: replacement of Arg90 (in the *Bacillus subtilis* enzyme) with a glycine or alanine residue resulted in a reduction in k_{cat} by a factor of 10^5. Quantum chemical studies to investigate the interaction energies of these key residues showed that Arg90 provides the dominant stabilizing interactions of the transition state relative to the substrate [70].

Chorismate mutase catalyses the reaction without covalent binding, without the necessity of bringing together several substrates or providing an acid or base for catalysis, and exactly how it achieves catalysis is intriguing. Early QM/MM calculations indicated that stabilization of the transition state in the enzyme is important [63]. This has been confirmed in more recent studies [7, 70, 78–80]. QM/MM and empirical valence bond calculations indicate that transition state stabilization (rather than substrate preorganization) is the dominant factor in chorismate mutase catalysis [65, 81]. The structure of the active site is adapted to bind chorismate, but to bind the transition state more tightly. Copley et al. found that the reactive conformation of chorismate, the di-axial conformer, is present 20% of the time in solution. These authors suggested that chorismate mutase selects the reactive conformer from solution which readily proceeds to product [82]. Contrary to most opinion and detailed simulations, Bruice et al. suggested that catalysis is almost entirely due to a so-called ‘near attack’ conformation (NAC) being selectively bound by the enzyme and that transition state stabilization does not play a significant role in catalysis [83]. This suggestion has been discredited in part due to the inability to uniquely define a NAC (various different definitions of a NAC in chorismate mutase have been used in different studies). A NAC is a structure which is chosen by some arbitrary criteria, and resembles a transition state; it is not clear

that this is a useful concept. QM/MM free energy perturbation calculations found the free energy cost of forcing the conformation of chorismate in solution into the reactive conformation to be 3.8–4.6 kcal mol^{-1} [84]; calculations with empirical valence bond methods found a similar result, around 5 kcal mol^{-1} [65]. This is significantly smaller than the experimentally determined $\Delta\Delta^{\ddagger}G$ (difference in activation free energies between the enzymic and uncatalysed reaction of 9.1 kcal mol^{-1}. This strongly indicates that conformational effects alone cannot account for catalysis by chorismate mutase; transition state stabilization is vital [84].

Claeyssens et al. [68] calculated a B3LYP/6-31(*d*)/CHARMM27 QM/MM energy barrier (12.0 kcal mol^{-1}), averaged over several profiles, in excellent agreement with the experimentally determined activation enthalpy ($\Delta^{\ddagger}H = 12.7$ kcal mol^{-1}). The calculations showed that CM stabilizes the transition state by an average of 4.2 kcal mol^{-1} more than it stabilizes the reactant. When the values of transition state stabilization and conformational effects (3.8–5 kcal mol^{-1}) [84] are summed they give similar values to the experimentally observed catalytic rate ($\Delta\Delta^{\ddagger}G = 9.1$ kcal mol^{-1}), therefore these two factors together account for catalysis by CM. The conformational effects probably arise because of TS stabilization by the enzyme, i.e. the same residues that stabilize the TS (e.g. Arg90 in *Bacillus subtilis* CM) also stabilize the substrate in a reactive conformation.

14.5. ENZYME REACTION MECHANISMS

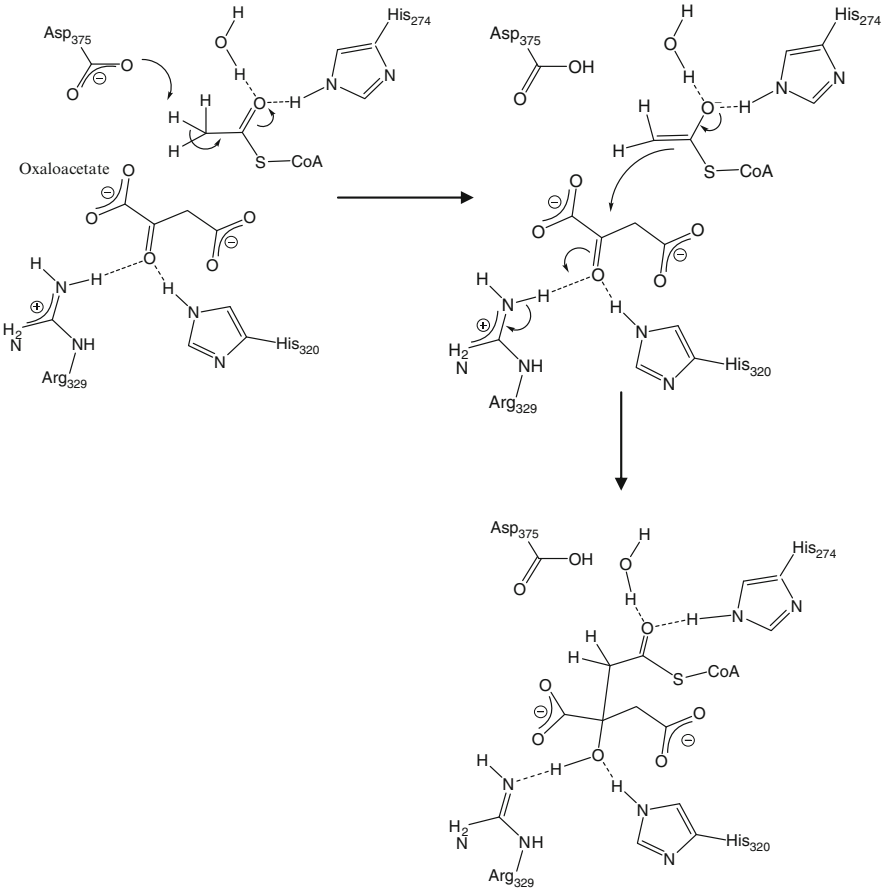
Enzyme reaction mechanisms are often initially proposed using chemical and structural knowledge. Proposed mechanisms can be investigated using techniques such as site directed mutagenesis to identify residues essential to catalysis. However, as mentioned above (see Section 14.2), the results can be misleading due to mutations inducing subtle changes in enzyme structure [22] or mechanism [23]. Therefore theoretical calculations can provide a unique insight into reaction mechanisms – using the naturally occurring sequence to investigate mechanisms without any structural changes.

Combined QM/MM techniques can be used to investigate reaction mechanisms, by treating the reacting region with QM and the surrounding environment with MM the effects of the protein environment can be taken into account. Early QM/MM calculations were performed in the 1970s [30, 85] QM/MM techniques have since developed significantly and it is becoming increasingly apparent that these techniques can be utilised to gain significant insight into enzyme catalysed reactions.

14.5.1. Citrate Synthase

Another enzyme whose reaction has been subject to a large amount of experimental and computational study is citrate synthase (CS). CS catalyses the first step in the

citric acid cycle, the conversion of oxaloacetate (OAA) to citrate using acetyl-coenzyme A (acetyl-CoA). CS functions in many conditions by first binding OAA to its active site followed by acetyl-CoA. A proton is then abstracted from the terminal group of acetyl-CoA, followed by condensation with OAA and finally hydrolysis to form citrate and CoA (see Scheme 14-3). The structure of the intermediate, after proton abstraction, is difficult to determine experimentally. Proposals of an enolate, enol (with proton abstraction from His274) or enolic form (with a low-barrier hydrogen bond, proton shared between His274 and the acetyl-CoA oxygen) have been put forward [86–90]. Early QM/MM calculations at the (AM1/CHARMM level of theory) indicated that the most energetically favourable intermediate was the enolate [4]. This conclusion has been supported by high



Scheme 14-3. The proton abstraction [17] and condensation steps catalysed by citrate synthase. Recently QM/MM modelling has suggested that Arg329 acts as the acid to form the citryl-CoA intermediate [96]

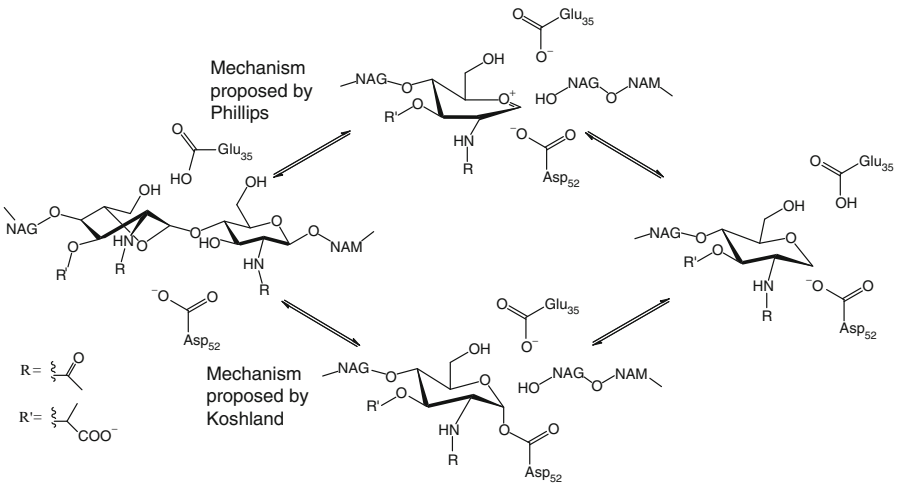
level ab initio, QM/MM (MP2/6-31+G(d)//6-31G(d)/CHARMM) modelling [17, 91]. The high level calculations showed that the enolate is sufficiently stabilized by hydrogen bond donation from His274 and a crystallographically observed water molecule. Both hydrogen bonds were found to be normal, i.e. not of the low barrier type, ruling out the enolic form [17].

The enolate intermediate undergoes condensation with OAA to form citryl-CoA. Fourier transform infrared spectroscopy (FTIR) [92] and NMR [93] studies have shown that the OAA carbonyl, to which condensation occurs, is polarized when bound to the enzyme. This has been proposed to be of importance to catalysis. Modelling using QM/MM techniques, with different degrees of substrate polarization, has shown that the effect of substrate polarization on the reaction barrier is small [94] providing evidence to propose that catalysis is primarily due to the stabilization of the acetyl-CoA enolate.

High-level QM/MM modelling has led to an unusual proposal for the mechanism of the condensation step. The formation of the (kinetically competent) citryl-CoA intermediate involves the enzyme acting as an acid to protonate (what was originally) the carbonyl carbon of oxaloacetate. It had been suggested that His320 was the donating group [95], but simulations have shown this histidine is likely to be neutral, making this unlikely [96]. High level ab initio QM/MM modelling (at the MP2/aug-cc-pVDZ//B3LYP/CHARMM27 level) results suggest that Arg329 acts as the acid in this condensation reaction [96]. Other studies investigating conformational changes [97] also suggest that this same residue could be involved in the closure process, by interaction of Arg329 with oxaloacetate causing a shift in helix 328–341 relative to the large domain resulting in closure of the subunit. This has led to speculation that the chemical change at Arg329 (i.e. its deprotonation) may result in a change in dynamics or conformation that causes the opening of the subunit releasing the product. The question that remains is how the enzyme regains its active state – this must involve proton transfer from Arg329 to Asp375, over 6 Å apart - which would allow the enzyme to turn over and enable the catalytic cycle to repeat.

14.5.2. Hen Egg White Lysozyme

Lysozyme, a natural anti-bacterial agent, was first discovered in human mucus in 1922 by Alexander Fleming [98]. A similar type of lysozyme is also found in hen egg white [99] (termed hen egg white lysozyme, HEWL), and it is this form that is most commonly studied. HEWL was the first enzyme to have its three-dimensional structure resolved, via X-ray crystallography [100], and the first enzyme to have a detailed chemical mechanism suggested, based on structural data, for its catalytic action [101]. HEWL functions by cleaving the glycosidic bond between sugar units in bacterial cell walls; the bacterial cell is then quickly destroyed by osmotic pressure as water enters the cell.



Scheme 14-4. The proposed catalytic mechanisms for HEWL, proposed by Phillips [101] (via an oxocarbenium ion intermediate) and the generic mechanism for glycosides as proposed by Koshland [105] (via a covalent intermediate)

The active site of lysozyme is in a deep cleft which has six subsites, this enables it to bind a NAG_A–NAM_B–NAG_C–NAM_D–NAG_E–NAM_F portion of the polysaccharide with glycolysis occurring at the glycosidic bond joining NAG_C and NAM_D (the subscript indicates the subsite in the original nomenclature) [102]. In the original mechanism of Phillips et al. [101], the NAM sugar at the D-site was proposed to be distorted (i.e. not in lowest energy chair conformation) when bound to the enzyme [103, 104]; susbstrate distortion is often suggested for Koshland-type mechanisms also (see below).

There have been several mechanisms proposed for the reaction in HEWL, with most debate centring on two general proposals (see Scheme 14-4): one proceeding via an ionic intermediate (Phillips [101]) and the other with an aspartate residue (Asp52) covalently bound to the NAM_D ring (Koshland [105]). The Phillips mechanism was widely held to be correct, often given in textbooks [106]. However the identification of a covalent glycosyl-enzyme intermediate in a mutant HEWL enzyme [107] supported the Koshland type mechanism. Recent QM/MM studies have shown that the wild type enzyme also proceeds via a covalent intermediate [8]; simulations have the advantage of being able to study the reaction of the wild type enzyme with its natural substrate, which is typically not possible, e.g. in experimental structural studies.

The question remains as to the distortion of the NAM_D ring and its relationship to catalysis. Phillips proposed that distortion would occur to adopt a conformation similar to that of the oxocarbenium transition state [101]. A trisaccharide substrate bound – as NAM_B–NAG_C–NAM_D – HEWL was published by Strynadka and James as a high resolution (1.5 Å) X-ray crystallographic structure [103]. This structure

showed the NAM_D unit distorted to a 'sofa' type conformation. The necessity and nature of substrate distortion at this sub-site is still a matter of discussion [108, 109].

14.5.3. Aromatic Amine Dehydrogenase

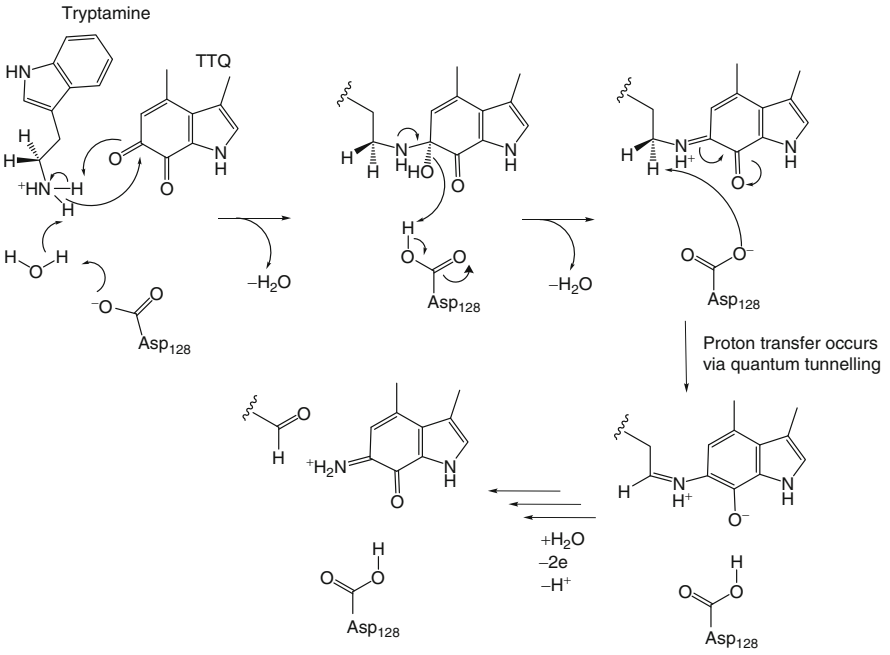
A lively current debate in enzymology centres on whether enzymes have adapted to use quantum mechanical tunnelling to assist in enzyme catalysis [110–115]. Tunnelling is the quantum effect of a small particle, e.g. an electron or proton, passing through a classically forbidden energy barrier. In effect this means that a particle with a small mass is able to 'jump' from one heavy atom to another without proceeding over the energy barrier. Experimental studies of kinetic isotope effects (KIE) have demonstrated the involvement of quantum tunnelling in some enzyme-catalysed reactions. However atomic-level understanding of the factors involved is necessary to analyse to what extent tunnelling may enhance the rate of enzyme catalysed reactions, and its contribution to catalysis.

Masgrau et al. [111, 112] have studied quantum tunnelling effects in the enzyme aromatic amine dehydrogenase using QM/MM variational transition state theory calculations with small curvature tunnelling corrections [116]. Aromatic amine dehydrogenase catalyses the transfer of a proton from tryptamine (bound to a tryptophan tryptophylquinone (TTQ) cofactor) to an aspartate residue (see Scheme 14-5). Masgrau et al. concluded that the proton is transferred to the aspartate carboxylate oxygen that is initially further away from the proton on tryptamine. They show that the proton and oxygen approach each other, up to a distance of 1.6 Å, after which the proton tunnels a distance of 0.6 Å through the barrier. The energetic benefit of tunnelling and other quantum effects (compared to a purely classical process) in this example was estimated to be 3.1 kcal mol^{-1} [111]. In an equivalent 'reference' reaction in solution, some tunnelling would also occur, meaning that the catalytic contribution of tunnelling is likely to be smaller than this.

In a study of the reaction of lipoxygenase (with empirical valence bond methods) the effect of tunnelling, in the enzyme, was compared to that in solution and was shown to be similar for both lipoxygenase and other enzymes [5]. This indicates that, although tunnelling effects can be large, they can be similar in enzyme and solution reactions and so do not necessarily contribute to enzyme catalysis.

14.6. CONCLUSIONS

Biomolecular simulation is providing detailed insights into how enzymes function at the atomic level. Here we have highlighted the ability of molecular modelling to discriminate between mechanisms, identify catalytic interactions and test hypotheses. QM/MM studies are showing themselves to be valuable in the understanding of catalysis by biomolecular systems. Classical MD studies are also proving



Scheme 14-5. An overview of the reaction mechanism of aromatic amine dehydrogenase with tryptamine [117]. Once the complex has formed between tryptamine and the TTQ cofactor, proton abstraction by aspartate128 occurs, which is dominated by quantum tunnelling [111, 112]

invaluable for investigating protein dynamics. It is clear that dynamics must play a role in enzyme catalytic cycles, in particular for substrate binding and product release but also, as shown for PMM/PGM (see Section 14.3.2), during other steps of catalytic cycles, which may have a direct impact on the overall rate of reaction. However it is still a matter of current debate as to the effect of dynamics on enzyme catalysis.

The field of computational enzymology is advancing at a significant pace and is having a growing impact in many areas. For example, modelling of enzyme reactions has the potential to provide information relevant to drug design and development [11]. Another area which computational enzymology has the potential to affect is protein design, e.g. in the design of novel protein catalysts for the synthesis of variants of natural products and the development of thermostable enzymes for industrial process applications. As highlighted above there are still unanswered questions, in particular as to the role of dynamics in enzyme-catalysed reactions. Biomolecular simulations will be important in resolving controversies surrounding this area and contribute increasingly to the understanding of enzyme mechanism, dynamics and function.

REFERENCES

1. Olsson MHM, Parson WW, Warshel A (2006) Chem Rev 106:1737–1756
2. Hammes-Schiffer S, Benkovic SJ (2006) Annu Rev Biochem 75:519–541
3. Mulholland AJ (2008) Biochem Soc Trans 36:22–26
4. Mulholland AJ, Richards WG (1997) Proteins: Struct Funct Genet 27:9–25
5. Olsson MHM, Siegbahn PEM, Warshel A (2004) J Am Chem Soc 126:2820–2828
6. Pentikäinen U, Pentikäinen OT, Mulholland AJ (2008) Proteins: Struct Funct Bioinform 70:498–508
7. Ranaghan KE, Ridder L, Szefczyk B, Sokalski A, Hermann JC, Mulholland AJ (2003) Mol Phys 101:2695–2714
8. Bowman AL, Grant IM, Mulholland AJ (2008) Chem Commun 4425–4427
9. Van Der Kamp MW, Shaw KE, Woods CJ, Mulholland AJ (2008) J R Soc Interface 5:S173–S190
10. Van Der Kamp MW, Mulholland AJ (2008) Nat Prod Rep 25:1001–1014
11. Mulholland AJ (2005) Drug Discov Today 10:1393–1402
12. Guvench O, MacKerell AD Jr (2009) Curr Opin Struct Biol 19:56–61
13. Motiejunas D, Gabdoulline R, Wang T, Feldman-Salit A, Johann T, Winn PJ, Wade RC (2008) Proteins: Struct Funct Bioinform 71:1955–1969
14. Kang L, Li H, Zhao X, Jiang H (2008) J Mat Chem 46:182–198
15. Luzhkov VB, Åqvist J (2000) Protein Struct Mol Enz 1481:360–370
16. Liu Z, Xu Y, Tang P (2006) J Phys Chem B 110:12789–12795
17. Van Der Kamp MW, Perruccio F, Mulholland AJ (2007) J Mol Graph Model 26:676–690
18. Flack HD, Bernardinelli G (2008) Chirality 20:681–690
19. Hooft RWW, Sander C, Vriend G (1996) Proteins: Struct Funct Genet 26:363–376
20. Li H, Robertson AD, Jensen JH (2005) Proteins: Struct Funct Bioinform 61:704–721
21. van Gunsteren WF, Bakowies D, Baron R, Chandrasekhar I, Christen M, Daura X, Gee P, Geerke DP, Glättli A, Hünenberger PH, Kastenholz MA, Oostenbrink C, Schenk M, Trzesniak D, van der Vegt NFA, Yu HB (2006) Angew Chem Int Ed 45:4064–4092
22. Karplus M, Evanseck JD, Joseph D, Bash PA, Field MJ (1992) Faraday Discussions 93:239–248
23. Nickbarg EB, Davenport RC, Petsko GA, Knowles JR (1988) Biochemistry 27:5948–5960
24. McCammon JA, Gelin BR, Karplus M (1977) Nature (London) 267:585–590
25. Khalili-Araghi F, Gumbart J, Wen P, Sotomayor M, Tajkhorshid E, Schulten K (2009) Curr Opin Struct Biol 19:128–137
26. Jójárt B, Kiss R, Viskolcz B, Keseru GM (2008) J Chem Inf Model 48:1199–1210
27. MacKerell AD Jr (2001) In: Watanabe M (ed) Computational biochemistry and biophysics. Marcel Dekker, Inc., New York, pp 7–38
28. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) J Comput Chem 4:187–217
29. Senn HM, Thiel W (2007) Curr Opin Chem Biol 11:182–187
30. Warshel A, Levitt M (1976) J Mol Biol 103:227–249
31. Car R, Parrinello M (1985) Phys Rev Lett 55:2471–2474
32. Whitehead L, Edge CM, Essex JW (2001) J Comput Chem 22:1622–1633
33. Levitt M, Warshel A (1975) Nature 253:694–698
34. Warshel A, Levitt M (1976) J Mol Biol 106:421–437
35. Sansom MSP, Scott KA, Bond PJ (2008) Biochem Soc Trans 36:27–32
36. Woods CJ, Mulholland AJ (2008) In: Hinchliffe A (ed) Chemical modelling: applications and theory, vol 5. Specialist Periodical Reports. Royal Society of Chemistry, London

37. Lodola A, Mor M, Zurek J, Tarzia G, Piomelli D, Harvey JN, Mulholland AJ (2007) Biophys J 92: L20–L22
38. Wolf-Watz M, Thai V, Henzler-Wildman K, Hadjipavlou G, Eisenmesser EZ, Kern D (2004) Nature 11:945–949
39. Hayward S, Kitao A, Berendsen HJC (1997) Proteins: Struct Funct Genet 27:425–437
40. Bahar I, Chennubhotla C, Tobi D (2007) Curr Opin Struct Biol 17:633–640
41. Kitao A, Go N (1999) Curr Opin Struct Biol 9:164–169
42. Hall BA, Kaye SL, Pang A, Perera R, Biggin PC (2007) J Am Chem Soc 129:11394–11401
43. Mongan J (2004) J Comput Aided Mol Des 18:433–436
44. Barrett P, Hall B, Noble M (2004) Acta Crystallogr Sect D: Biol Cryst 60:2280–2287
45. Humphrey W, Dalke A, Schulten K (1996) J Mol Gra 14:33–38
46. Caponigro G, Parker R (1996) Microbiol Rev 60:233–249
47. Mitchell P, Tollervey D (2001) Curr Opin Cell Biol 13:320–325
48. Wilusz CJ, Wormington M, Peltz SW (2001) Nat Rev Mol Cell Biol 2:237–246
49. Liu H, Rodgers ND, Jiao X, Kiledjian M (2002) EMBO J 21:4699–4708
50. Liu S, Jiao X, Liu H, Gu M, Lima CD, Kiledjian M (2004) RNA 10:1412–1422
51. Chen N, Walsh MA, Liu Y, Parker R, Song H (2005) J Mol Biol 347:707–718
52. Koshland DE (1958) Proc Natl Acad Sci 44:98–104
53. Benkovic SJ, Hammes GG, Hammes-Schiffer S (2008) Biochemistry 47:3317–3321
54. Regni C, Schramm AM, Beamer LJ (2006) J Biol Chem 281:15564–15571
55. Naught LE, Tipton PA (2005) Biochemistry 44:6831–6836
56. Wolfenden R, Snider MJ (2001) Acc Chem Res 34:938–945
57. Pauling L (1948) Am Sci 36:51–58
58. Warshel A, Sharma PK, Kato M, Xiang Y, Liu H, Olsson MHM (2006) Chem Rev 106:3210–3235
59. Marcus RA (1965) J Chem Phys 43:679–701
60. Cload ST, Liu DR, Pastor RM, Schultz PG (1996) J Am Chem Soc 118:1787–1788
61. Kast P, Asif-Ullah M, Hilvert D (1996) Tetrahedron Lett 37:2691–2694
62. Kienhofer A, Kast P, Hilvert D (2003) J Am Chem Soc 125:3206–3207
63. Lyne PD, Mulholland AJ, Richards WG (1995) J Am Chem Soc 117:11345–11350
64. Lee YS, Worthington SE, Krauss M, Brooks BR (2002) J Phys Chem B 106:12059–12065
65. Strajbl M, Shurki A, Kato M, Warshel A (2003) J Am Chem Soc 125:10228
66. Crespo A, Scherlis DA, Marti MA, Ordejon P, Roitberg AE, Estrin DA (2003) J Phys Chem 107:13728–13736
67. Eletsky A, Kienhofer A, Hilvert D, Pervushin K (2005) Biochemistry 44:6788–6799
68. Claeyssens F, Ranaghan KE, Manby FR, Harvey JN, Mulholland AJ (2005) Chem Commun 40:5068–5070
69. Ruiz-Pernia JJ, Silla E, Tunon I, Marti S, Moliner V (2004) J Phys Chem B 108:8427–8433
70. Szefczyk B, Mulholland AJ, Ranaghan KE, Sokalski WA (2004) J Am Chem Soc 126: 16148–16159
71. Crespo A, Marti MA, Estrin DA, Roitberg AE (2005) J Am Chem Soc 127:6940–6941
72. Ishida T, Fedorov DG, Kitaura K (2005) J Phys Chem B 110:1457–1463
73. Guilford WJ, Copley SD, Knowles JR (1987) J Am Chem Soc 109:5013–5019
74. Haynes MR, Stura EA, Hilvert D, Wilson IA (1994) Proteins: Struct Funct Genet 18:198–200
75. Cload ST, Liu DR, Pastor RM, Schultz PG (1996) J Am Chem Soc 118:1787–1788
76. Kast P, Hartgerink JD, Asif-Ullah M, Hilver D (1996) J Am Chem Soc 118:3069–3070
77. Kast P, Asif-Ullah M, Jiang N, Hilvert D (1996) Proc Natl Acad Sci USA 93:5043–5048
78. Marti S, Andres J, Moliner V, Silla E, Tunon I, Bertran J, Field MJ (2001) J Am Chem Soc 123:1709–1712

79. Ranaghan KE, Ridder L, Szefczyk B, Sokalski A, HJ C, Mulholland AJ (2004) Org Biomol Chem 2:968–980
80. Marti S, Andres J, Moliner V, Silla E, Tunon I, Bertran J (2000) J Phys Chem B 104:11308–11315
81. Marti S, Andres J, Moliner V, Silla E, Tunon I, Bertran J (2003) J Am Chem Soc 126:311–319
82. Copley SD, Knowles JR (2002) J Am Chem Soc 109:5008–5013
83. Bruice TC, Benkovic SJ (2000) Biochemistry 39:6267–6274
84. Ranaghan KE, Mulholland AJ (2004) Chem Commun 1238–1239
85. Warshel A, Karplus M (1972) J Am Chem Soc 94:5612–5625
86. Gerlt JA, Kreevoy MM, Cleland WW, Frey PA (1997) Chem Biol 4:259–267
87. Gerlt JA, Gassman PG (1993) J Am Chem Soc 115:11552–11568
88. Cleland W, Kreevoy MM (1994) Science 264:1887–1890
89. Warshel A, Papazyan A (1996) Proc Natl Acad Sci U S A 93:13665–13670
90. Alagona G, Ghio C, Kollman PA (1995) J Am Chem Soc 117:9855–9862
91. Mulholland AJ, Lyne PD, Karplus M (2000) J Am Chem Soc 122:534–535
92. Kurz LC, Drysdale GR (1987) Biochemistry 26:2623–2627
93. Kurz LC, Ackerman JJH, Drysdale GR (1985) Biochemistry 24:452–457
94. Van Der Kamp MW, Perruccio F, Mulholland AJ (2007) Proteins: Struct Funct Bioinform 69:521–535
95. Karpusas M, Branchaud B, Remington SJ (2002) Biochemistry 29:2213–2219
96. Van Der Kamp MW, Perruccio F, Mulholland AJ (2008) Chem Commun 1874–1876
97. Daidone I, Roccatano D, Hayward S (2004) J Mol Biol 339:515–525
98. Fleming A (1922) Proc R Soc Lond B 93:306–317
99. Canfield RE (1963) J Biol Chem 238:2698–2707
100. Blake CCF, Koenig DF, Mair GA, North CT, Phillips DC, Sarma VR (1965) Nature 206:757–761
101. Phillips DC (1966) Sci Am 215:78–90
102. Davies GJ, Wilson KS, Henrissat B (1997) Biochem J 321:557–559
103. Strynadka NCJ, James MNG (1991) J Mol Biol 220:401–424
104. Hadfield AT, Harvey DJ, Archer DB, MacKenzie DA, Jeenes DJ, Radford SE, Lowe G, Dobson CM, Johnson LN (1994) J Mol Biol 243:856–872
105. Koshland DE (1953) Biol Rev 28:416–436
106. Stryer L (2006) Biochemistry. WH Freeman & Co, New York
107. Vocadlo DJ, Davies GJ, Laine R, Withers SG (2001) Nature 412:835–838
108. Karlsen S, Hough E (1995) Acta Crystallogr Sect D 51:962–978
109. Smith BJ (1997) J Am Chem Soc 119:2699–2706
110. Ball P (2004) Nature 431:396–397
111. Masgrau L, Ranaghan KE, Scrutton NS, Mulholland AJ, Sutcliffe MJ (2007) J Phys Chem B 111:3032–3047
112. Masgrau L, Roujeinikova A, Johannissen LO, Hothi P, Basran J, Ranaghan KE, Mulholland AJ, Sutcliffe MJ, Scrutton NS, Leys D (2006) Science 312:237–241
113. Liang ZX, Lee T, Resing KA, Ahn NG, Klinman JP (2004) Proc Natl Acad Sci USA 101:9556–9561
114. Kohen A, Klinman JP (1999) Chem Biol 6:191–198
115. Hammes-Schiffer S (2002) Biochemistry 41:13335
116. Gao JL, Truhlar DG (2002) Annu Rev Phys Chem 53:7755–7760
117. Masgrau L, Basran J, Hothi P, Sutcliffe MJ, Scrutton NS (2004) Arch Biochem Biophys 428:41–51

CHAPTER 15

TRANSPORT MECHANISM IN THE *ESCHERICHIA COLI* AMMONIA CHANNEL AMTB: A COMPUTATIONAL STUDY

YIRONG MO[1,2], ZEXING CAO[1], AND YUCHUN LIN[2,3]
[1]*Department of Chemistry, The State Key Laboratory for Physical Chemistry of Solid States, Center for Theoretical Chemistry, Xiamen University, Xiamen, Fujian 361005, P.R. China,*
[2]*Department of Chemistry, Western Michigan University, Kalamazoo, MI 49008, USA,*
[3]*Present address: Department of Chemical Engineering, University of California, Berkeley, Berkeley, California 94720, USA*
e-mail: yirong.mo@wmich.edu; zxcao@xmu.edu.cn; yuchun.lin@berkeley.edu

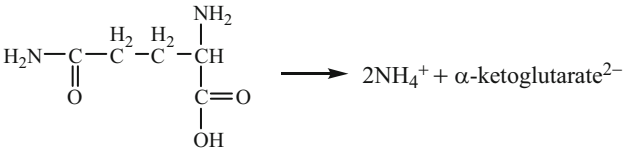
Abstract: Computational approaches at various levels have been used to elucidate the mechanism of the ammonium/ammonia transport process through the *Escherichia coli* AmtB membrane protein. Molecular dynamics (MD) simulations at the classical molecular mechanical (MM) level confirmed that only NH_3 can transport through the highly hydrophobic AmtB channel. Thus, NH_4^+, which is predominant in solution, must deprotonate before crossing the channel. Significantly, conformational analyses revealed that in the end of the recruitment vestibule, there is a hydrogen bond wire between NH_4^+ and the carboxylate group of Asp160 via two water molecules. Thus, Asp160 is most likely the proton acceptor from NH_4^+. This explains the high conservation of Asp160 in Amt proteins and why the D160A mutant would completely quench the activity of AmtB. The proposed deprotonation mechanism was further examined by the combined QM/MM methods. Computations at both QM(DFT)/MM and QM(PM3)/MM levels concur that the proton transfer starts from a lose of a proton from a nearby water to Asp160 to form a hydroxide anion in the intermediate state, followed by a proton transfer from NH_4^+ to the hydroxide ion through a water molecule.

Keywords: Molecular dynamics simulation, Combined QM/MM, Ammonium transport protein, Deprotonation mechanism

P. Paneth and A. Dybala-Defratyka (eds.), *Kinetics and Dynamics*, Challenges and Advances in Computational Chemistry and Physics 12,
DOI 10.1007/978-90-481-3034-4_15,

15.1. OVERVIEW

Life processes often involve the movement of ions, small molecules, or even macromolecules like proteins in or out of cells. In other words, the transported species must pass through biological membranes. Membrane transport proteins (or transporters in short), which are integral membrane proteins, are the facilitators for this kind of transport. The ammonium transporter (Amt)[1], methylammonium permease (Mep) [2–4] and the homologues of Amt in animals Rhesus (Rh) [5–7] membrane proteins form a super Amt/Mep/Rh family which is ubiquitous in all domains of life and responsible for the transmembrane transport of ammonium (NH_4^+) which is involved in the fundamental nitrogen metabolism [8]. As such, the transport of ammonium across cellular membranes is of high biological relevance. For instance, in bacteria and plants, ammonium is an important nutrient that is taken from surroundings such as air to provide a source of nitrogen for amino acid synthesis, whereas for humans, ammonium is toxic at high concentration and produced predominantly in the cells of the proximal tubules in the renal cortex from glutamine (Eq. 15-1) and secreted into the urine as a result of the ammonium/ammonia transport processes and consequently plays a critical role in the maintenance of our body's acid–base balance (pH regulation) through the regulation of renal ammonia excretion [9].



(Gln)

(15-1)

As the uptake and secretion of ammonium is mediated by the Amt/Mep/Rh family of membrane proteins, so far, over 200 genes belonging to this family have been discovered. For bacteria and plants, ammoniums are transported into cells, in contrast, for human and animals, ammoniums are transported out of cells. In either direction, the transported species must go across the cellular membrane. The question now is how the transport occurs. It has been known that ammonium transport can occur by direct NH_4^+ transport or by a combination of H^+ and NH_3 transport [9]. Apparently, passive NH_4^+ transport can be driven by a concentration gradient or an electrical potential gradient. On the other hand, NH_3, like water and carbon dioxide, is moderately lipid soluble and is thought to penetrate cell membrane mainly by diffusion [10]. At high concentration of NH_3/NH_4^+, passive membrane permeation of neutral NH_3, as shown in Figure 15-1a, may be effective enough. At low concentration, however, a specific transport system (Figure 15-1b) is required. This transport system is called ammonia/ammonium channels which exist in the Amt/Mep/Rh family membrane proteins.

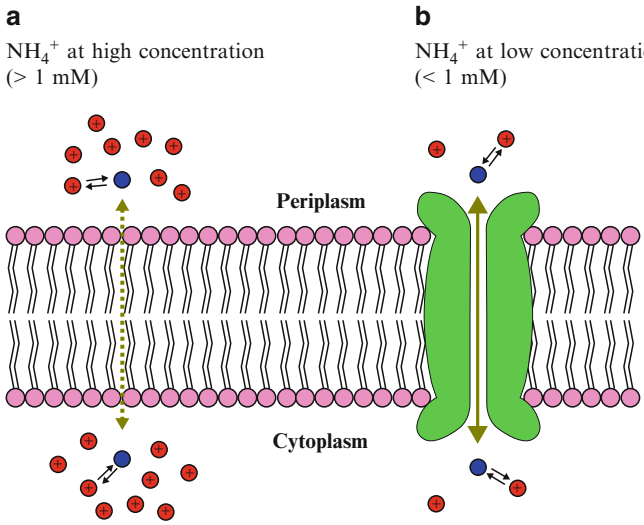


Figure 15-1. Schemes for the transport mechanisms at different concentrations, where the red dots refer to NH_4^+, blue dots are NH_3 and the green parts highlight the channel

Although there has been a controversy over the ammonium transport mechanism, it had often been thought that the transported species through the membrane proteins was ionic ammonium and the channels were therefore ion channels [11–16]. But this predominant view has been challenged by Soupene et al. who claimed that Amt and Rh proteins facilitate the transports of neutral NH_3 and CO_2, respectively [17–22]. Unfortunately, it is difficult to measure ammonium directly, and its transport behavior can only be inferred from its inhibitory effect on methylamine uptake [11, 12] or the measurement of ion current in oocytes [6, 13, 22]. It has been found that the accumulation of methylammonium (labeled with ^{14}C) in whole cells depends on Amt/Mep proteins, and thus it is inferred that Amt/Mep proteins mediate the uptake of NH_4^+ [4]. However, evidence shows that Amt/Mep proteins can increase the rate of equilibrium of ammonia across the cytoplasmic membrane, and thus are channels for ammonia [18, 19]. Significantly, the recent X-ray crystallographic structures of *E. coli* AmtB with and without ammonia or methylammonia clearly demonstrated that the 20 Å long AmtB channel is predominantly hydrophobic and it is the neutral NH_3 rather than the positively charged NH_4^+ that crosses the hydrophobic channel [23–26]. This is further endorsed by in vivo data which show that AmtB function is independent of either the membrane potential or the intracellular ATP pool [27]. As a consequence, much work remains to be done to elucidate the transport mechanism, deprotonation mechanism and selectivity for the AmtB gas channel [28–31]. But we emphasize that this gas channel mechanism may not be universal and at least for certain plant Amt

homologues and Rh proteins, net ammonium ion conductance and thus electrogenic transport is strongly suggested [13, 16, 32]. Apparently, comparable studies between gas channels and ion channels are highly valuable. In this chapter, however, we will focus on the *E. coli* AmtB channel whose crystal structures are available and use computational techniques to explore the unsolved issues in order to provide new insights into its structure and functions at the atomistic level and enrich our mechanistic understanding for ammonium/ammonia transport proteins in general.

15.2. EXPERIMENTAL EVIDENCES ON *ESCHERICHIA COLI* AMTB CHANNEL

Ammonia can exist in two forms in aqueous solution, namely neutral ammonia (NH_3) and positively charged ion ammonium (NH_4^+) with the equilibrium constant (pK_a) for the following conversion being 9.25

$$NH_4^+ \rightleftharpoons NH_3 + H^+ \tag{15-2}$$

Thus, at ambient condition and the ionic strength of blood plasma, the form of NH_4^+ is predominant.

E. coli AmtB is an archetypal member of the Amt family and contains 406 amino acids which construct 11 transmembrane α-helices with a N-terminal periplasmic extension and a C-terminal cytoplasmic extension. In the native cytoplasmic membrane, AmtB forms a stable trimeric structure as shown in Figure 15-2, and retains this structure when purified and reconstituted in the presence of lipids

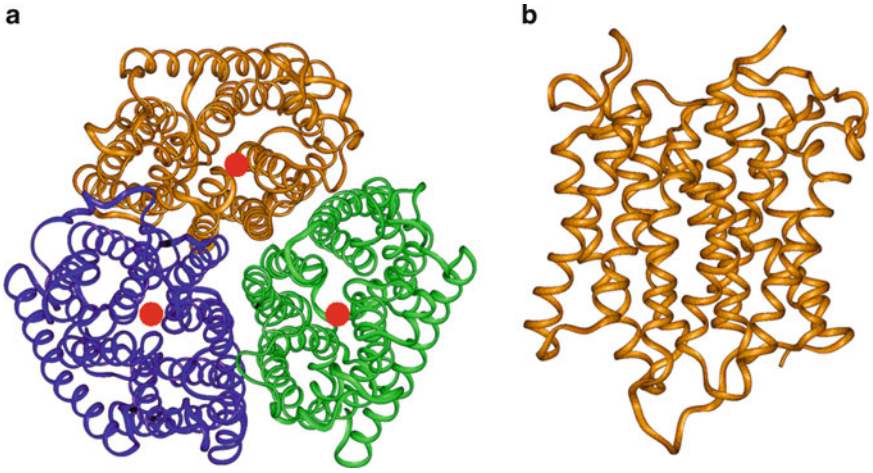


Figure 15-2. (**a**) Top-view of the ribbon structure of the AmtB trimer, where the red dots denote the channel positions; (**b**) Side-view of the ribbon structure of the AmtB monomer with 11 transmembrane (TM) helices

[33, 34]. The transport rate is estimated to be $10\sim10^4$ ammonium molecules per second per channel, compared with the diffusion limit $10^8\sim10^9$ in an open channel [24]. Whereas studies suggested that the cytoplasmic C-terminus, disordered in the AmtB structures, might mediate the co-operativity among the three subunits of AmtB, the X-ray crystallographic structures complexed with ammonia or methylammonia have shown that the narrow channels are located in the middles of monomers, which are bound together by hydrophobic forces (see Figure 15-2a) [23, 24]. The recent structure of Amt-1 from *Archaeoglobus fulgidus* resembles the AmtB structure [35]. A similar concept of channel has also been proposed for human Rh proteins [21, 36, 37]. The X-ray structure reveals a recruitment vestibule for NH_4^+ which involves Phe103, Phe107, Trp148, Phe215 and Ser219 as plotted in Figure 15-3. These amino acids contain aromatic rings and can bind NH_4^+ via π-cation interactions. π-cation interactions are known to play a key role in numerous biological recognition processes and have been extensively studied [38–43].

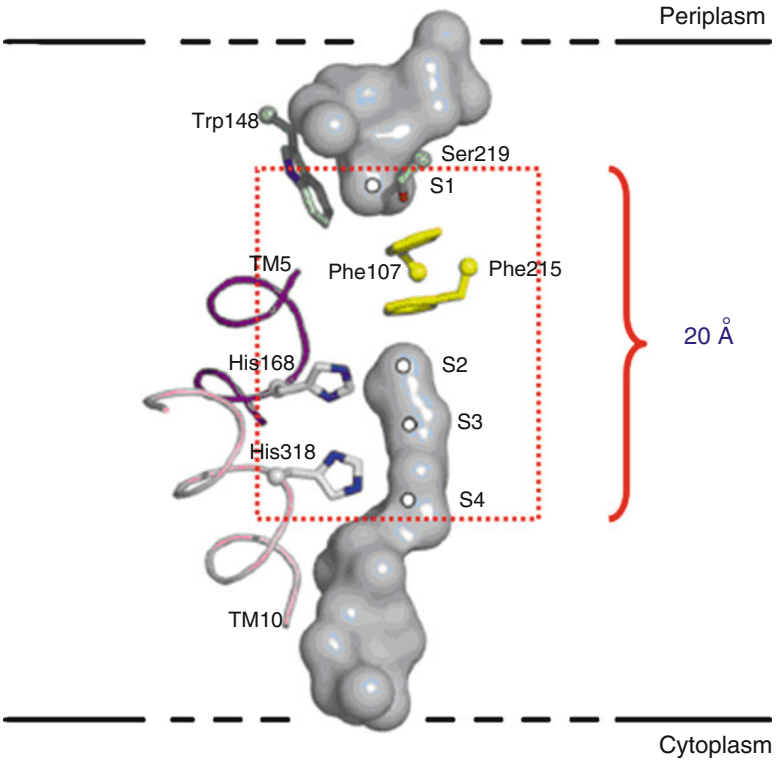


Figure 15-3. Important residues in the ammonia conducting channel which consists of four distinct parts S1-4 (Adapted from [46] with permission from Journal of Structural Biology. Copyright © 2007 Elsevier)

The sequence alignments of Amt and Rh proteins [36, 44, 45], however, reveal the conservation of Trp148 and Ser219 in Amt proteins which may help attract NH_4^+ and responsible for the higher affinity for NH_4^+ in Amt proteins than in Rh proteins. Interestingly, the crystal structures suggest that the phenyl rings of Phe107 and Phe215 (as shown in yellow in Figure 15-3) also serve as gates for the channel which is highly hydrophobic and located at the center of each monomer of the trimeric AmtB. Thus, it is of general interests to explore the mechanism of gate-opening, which is a dynamical process and hard to be captured experimentally. Since there is no significant conformational change between AmtB structures with or without ammonium [23, 24], it must be the side chains that control the gate movement and the subsequent ammonia transport. Furthermore, the high conservation of these two "dynamic" gates throughout the whole Amt/Rh family implies a common gating mechanism which remains to be elucidated. In addition, the substrate specificity to the gating mechanism is another interesting issue for further studies.

As the conducted species has now been confirmed to be ammonia for at least *E. coli* AmtB, the next question is where and how the initial ammonium deprotonates. No direct experimental evidence is available so far. There are two highly conserved residues in the mid-membrane center of the pathway, His168 and His318, whose adjacent imidazole rings are arranged such that a lateral hydrogen bond is formed between their ND1 nitrogens. It has been speculated that these two histidines may serve as proton acceptors for ammonium ions [23, 24, 47], and the essential role of these twin histidines for Amt protein activity has just been confirmed by site-directed mutagenesis experiments [48]. However, the subsequent release of protons from these histidines back to the periplasmic phase seems problematic as there is no apparent proton transfer route from the inner channel to the external solution particularly with the gates (Phe107 and Phe213) closed. Javelle et al. studied fourteen mutants of AmtB where one or both of His168 and His318 were replaced with other polar or non-polar residues [48]. They found that all these variants with the exception of H168E which retains 25% activity, are inactive in the conduction of methylamine (note not ammonia). Importantly, crystallographic study indicated that the structures of these mutants are not affected by the substitution. Thus, the roles of these two histidines must be functional.

While the hydrophobic channel obviously disfavors the physiologically important cations Na^+ and K^+ whose concentrations are much higher than that of ammonium due to the extremely high energetic cost of desolvation, it is intriguing whether they can occupy the binding site as they can also form π-cation bonds with surrounding aromatic rings. Experiments showed that there is no significant inhibition of the Amt/Mep/Rh proteins by Na^+ and K^+ [12, 22], which is further verified by adiabatic free energy calculations [24]. However, it is an interesting question whether water can go through the channel, considering that oxygen is only slightly more electronegative than nitrogen and H_2O and NH_3 are of comparable dipole moments and sizes.

Although extensive experimental studies have been conducted so far as introduced in the above, questions regarding the selection, gating, protonation/deprotonation and

transport mechanisms in AmtB remain to be answered and computational studies have the potential to provide supplementary and interpretive information for experimental data. Integrated computational and experimental investigations can potentially lead to the resolution of these questions.

15.3. COMPUTATIONAL METHODS

Whereas crystallographic structures can provide clear yet static pictures about the three-dimensional spatial arrangement of atoms, little knowledge is available about the transient conformations in the biological processes. But accumulating evidence shows that dynamics plays a critical role in the functions of proteins [49, 50]. Computational studies have the potential to correlate the structures, dynamics and functions of proteins, thus are able to complement experimental information [51–53]. Molecular dynamics simulations of membrane proteins at the atomic level can reveal the dynamic behavior of membrane proteins in biological processes and provide insights into the physical mechanisms and the physical principles underlying these processes [54–62]. Membrane proteins which have been computationally simulated include ion channels (e.g., potassium, calcium and chloride channels), aquaporins and bacteriorhodopsin. Following the procedure detailed by Woolf and Roux [63–65], we can build computational models for the AmtB membrane protein and conduct free energy simulations to explore the ammonium/ammonia transport mechanism.

An AmtB monomer is composed of 406 amino acids with around 3,000 non-hydrogen atoms. Because the binding power and specificity of this transmembrane protein largely comes from the overall system and solvation environment, it is crucial to consider the whole system in calculations to probe the transport mechanism. High level quantum mechanical (QM) calculations without considering the whole protein cannot offer information about the role of the part of the protein that is not modeled [66]. As a matter of fact, since most chemical reactions and biological processes occur in condensed states, the impact of environmental effects on the properties and reactivity of substances has been an intensive research subject for computational chemistry. QM methods are state-of-the-art in the treatment of small molecules and chemical reactions in the gas phase, while MM methods have enjoyed remarkable successes in modeling biosystems in solution. But difficulties remain in MM methods for the treatment of chemical reactions that involve the forming/breaking of covalent bond. As such, pure MM methods are appropriate to probe the transportation of NH_4^+/NH_3 across the AmtB channel, but they cannot be used to investigate the NH_4^+ deprotonation mechanism. For the latter, a plausible way to take the advantages of both the accuracy of QM methods and the efficiency of MM methods is to combine QM and MM methods such that a small part of the whole system is treated explicitly by a QM method, while the remaining majority is approximated by a MM method [67, 68]. During the past decade, combined QM/MM methods (Figure 15-4) represent one of the most significant advances in

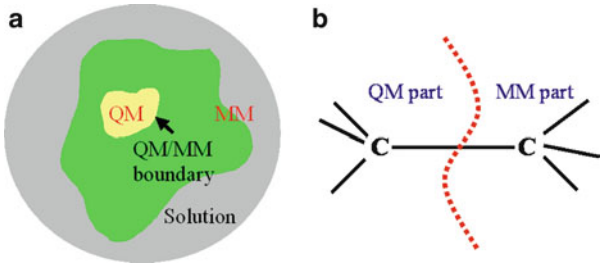


Figure 15-4. (**a**) Illustration of the combined QM/MM method in the study of a biological system (in green); (**b**) Boundary issue in the partition of QM and MM regions

computational chemistry [51, 69–80]. Molecular dynamics simulations with combined QM/MM methods can provide the details related to the catalytic activities and reaction mechanisms.

The accuracy of the combined QM/MM methods mainly relies on three factors: (1) the partition of QM and MM parts for the targeted complex system, (2) the quality of the empirical potential functions, and (3) the level of the QM computation. Often the separation of the QM and MM regions involves the breaking of one or more covalent bonds. Thus, the appropriate treatment of the QM-MM connection is centrally important to the accuracy of combined QM/MM methods [69, 81–87]. For the second factor, various force fields have been proposed and used to model the dynamic properties of biomolecules such as protein and DNA [88–90]. Most of the force fields assume static charge distribution for atoms. Polarizable force fields can improve the computational results [91–93], but there are few theoretical or experimental data to guide the parameterization of polarizable force fields. Finally, a key issue in combined QM/MM methods is the level of QM computations. Due to high computational costs, combined QM/MM calculations usually put the QM portion at semi-empirical level (such as AM1 [94] or PM3 [95]), but DFT methods are gradually being adopted in combined QM/MM methods [85, 96]. An attractive method is the self-consistent charge density functional tight binding (SCC-DFTB) method [97, 98].

Based on the available computational tools and demanding, we conducted the study of the ammonium/ammonia transport mechanism through the *E. coli* AmtB membrane protein at three different stages. At first, classical MM molecular dynamics simulations at the atomic level were performed with CHARMM software [99] and CHARMM22 force field [88] to probe the ammonium/ammonia transport across the AmtB membrane protein and importantly, identify the most probable site for the deprotonation of NH_4^+ [28]. Our simulations showed that NH_4^+ can be easily recruited into the periplasmic recruitment vestibule. In the end of this recruitment vestibule, the movements of the sidechains of Phe107 and Trp148 eventually create a slot for the initially buried carboxylate group of Asp160 to become exposed to the bulk solvent, and a hydrogen bond wire between NH_4^+ and the carboxylate group of Asp160 via two water molecules was clearly observed. At the second stage,

single-point combined QM(DFT)/MM geometry optimizations were performed with reduced models to examine the NH_4^+ deprotonation mechanism where Asp160 is the proton acceptor [100]. These computations were carried out with the ChemShell package [101] integrating the TURBOMOLE [102] and DL-POLY [103] programs. Interestingly, the QM(DFT)/MM optimizations not only endorsed our hypothesis that Aps160 is the proton acceptor from NH_4^+ via two water molecules, but also located an intermediate state where a proton is first lost from water to the carboxylate group of Asp160, followed by a proton transfer from NH_4^+ to the hydroxide anion. This extraordinary mechanism was further supported by the most recent third-stage QM(PM3)/MM molecular dynamics simulations [104].

15.4. COMPUTATIONAL STUDIES

15.4.1. Molecular Dynamics at the Molecular Mechanical Level

15.4.1.1. Computational Model

AmtB is a stable homotrimer in the native cytoplasmic membrane and retains this structure when purified and reconstituted in the presence of lipids [33, 34]. Whereas recent studies suggests that the cytoplasmic C-terminus, disordered in the AmtB structures, might mediate the co-operativity among the three subunits of AmtB, X-ray crystallographic structures have shown that the narrow channels are located in the middles of monomers, which are bound together by hydrophobic forces (see Figure 15-2a). In other words, the channels in the trimer might function individually due to the weak hydrophobic interaction among monomers. As such, we used one monomer instead of the AmtB trimer in our simulations in order to reduce the computational costs and embedded the AmtB monomer in a phospholipid bilayer constructed with DMPC (1,2-dimyristoyl-*sn*-glycero-3-phosphocholine) to imitate the membrane environment. The X-ray structure of AmtB at an atomic resolution of 1.35 Å determined by Khademi et al. was used (Protein Data Bank ID code 1U7G) [23]. The three mutated residues (F68S, S126P and K255L) and engineered Met residues (S atoms were replaced by Se) in 1U7G were changed back to their native states. Hydrogen positions of the protein were incorporated using the HBUILD facility in CHARMM [99]. The protonated states of histidine residues were determined based on their individual micro-environments, e.g., both His168 and His318 were found to be neutral, but the proton was attached to ND1 in the imidazole ring for His168 whereas the proton was on NE2 for His318. Water slabs were added to both the top and bottom of the phospholipid bilayer. The complete model for computations with unit cell dimensions of $60 \times 60 \times 92$ $Å^3$ (Figure 15-5a) consisted of about 35,000 atoms, including the AmtB, 72 DMPC molecules, 6,414 water molecules, as well 2–3 chloride ions to neutralize the whole system. This model was further extended with periodic boundary condition (PBC) during simulations. The electrostatic interactions were computed with no truncation using

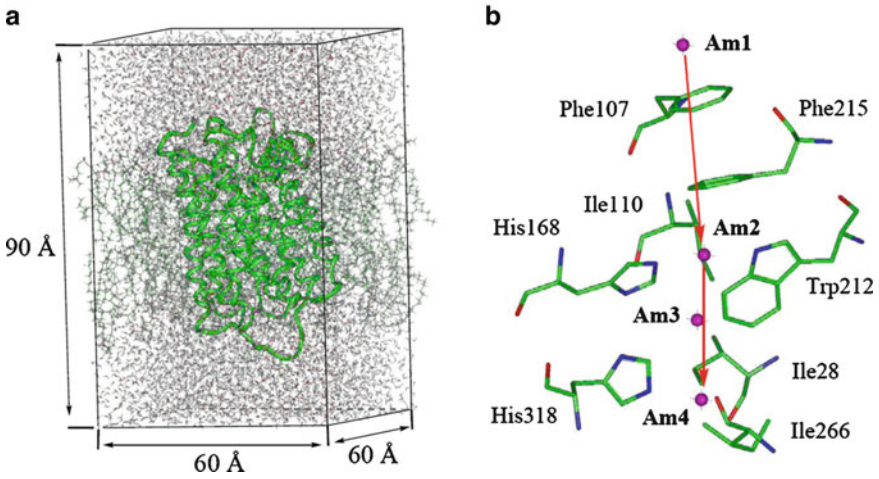


Figure 15-5. (**a**) AmtB membrane protein model (unitary cell) for simulations from side view; (**b**) Important residues around the channel in AmtB and the definition of the reaction coordinate. Am1-Am4 are the ammonium/ammonia species identified in the crystal structure

the particle mesh Ewald (PME) algorithm [105]. The trajectories were generated with a time step of 0.2 fs at constant pressure and temperature (300 K).

As all trajectories are plotted with respect to certain reaction coordinate (mostly geometric coordinates), here we chose a site within the channel as the reference point. Apart from an ammonium ion (Am1) in the recruitment vestibule, Khademi et al. observed three weak electron density peaks within the hydrophobic channel which were interpreted to be partially occupied ammonia molecules (Am2, Am3 and Am4 in Figure 15-5b) [23]. These four ammonium/ammonia sites are roughly in line. Thus, we took the position of the last ammonia Am4 as the reference site to describe the movement of ammonium/ammonia and plotted the corresponding energy profiles. It should be noted, however, that the pathway of ammonium/ammonia was by no means constrained to be linear in all our simulations.

Simulations were carried out with CHARMM [99], and the all-atom empirical potential energy function for protein [88] and phospholipids [106] was used. The TIP3P potential was used for modeling the water molecules [107]. The list of nonbonded interactions was truncated at 13 Å and the van der Waals and electrostatic interactions was smoothly switched off in the range of 11–12 Å. The umbrella sampling technique was adopted to generate the free energy profile or potential of mean force (PMF) along the transduction trajectory. Simulation windows were separated by around 0.4 Å, i.e., there were 50 independent simulations for the transport of ammomium/ammonia of 20 Å distance. For each window, we run 50 ps simulation to bring the whole system to an equilibrium state, followed by another 50∼100 ps simulation to produce dynamics data for further analyses. In the simulations, all of the bonds with hydrogen atoms in the MM part were constrained with the SHAKE algorithm [108].

15.4.1.2. Energy Profiles for the Transduction of Ammonium

To fully understand the whole process of the ammonium/ammonia transport across AmtB, we started with a state where the ammonium ion was completely solvated by aqueous solution and had no direct interactions with the membrane protein, and performed free energy molecular dynamics simulations using the umbrella sampling technique for the transduction of ammonium at the classical molecular mechanical level based on the simplified model with a monomeric form of AtmB in water box (Figure 15-5a). We chose a site which was outside the protein and had a distance of ~21 Å with the reference point for NH_4^+. Initial structural analysis showed that at this state, a solvent shell composed of 4–5 water molecules forms hydrogen bonds with the ammonium ion. As the solvation free energy of NH_4^+ in water is around 80 kcal/mol [109, 110], NH_4^+ reaches its energy minimum at this state. Starting from this configuration, umbrella sampling simulations were conducted to move the ion to the channel gates and the subsequent energy profile would reveal the favorable binding sites for NH_4^+.

Figure 15-6 showed the energy profile for the transport of NH_4^+. The energy file revealed two binding sites for the ammonium ion, ranging from $-17.1 \sim -15.5$ Å and $-12.2 \sim -9.2$ Å, respectively. We note that the first binding site is consistent with the X-ray crystal structure where an ammonium ion is captured (−16.3 Å). Our simulations showed that the entrance of NH_4^+ into the periplasmic recruitment vestibule (the first binding site) requires only 3.1 kcal/mol, which is consistent with the results of Javelle et al. who examined the temperature dependence of methylammonium uptake by AmtB from 4 to 37°C and determined an activation energy of 1.6 kcal/mol [27]. The very low barrier confirms AmtB as a channel rather than a transporter. For comparison, the activation barrier of transporters usually

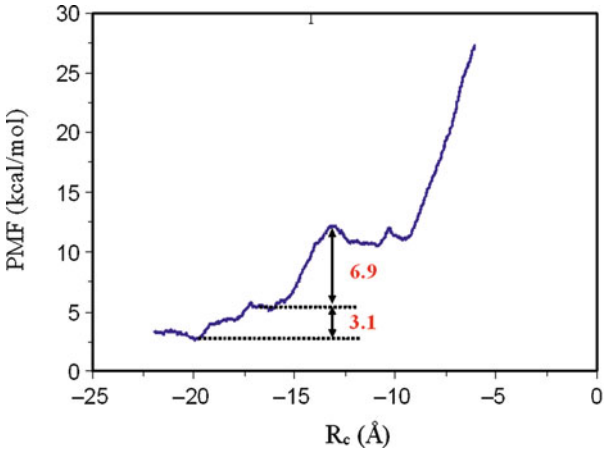


Figure 15-6. Free energy profiles for the passage of ammonium ion across the E. coli AmtB membrane protein

ranges from 11 to 27 kcal/mol [111]. Structural analyses revealed that the initial access of NH_4^+ to the first binding pocket involves Phe107, Trp148, Phe161, and Ser219. NH_4^+ loses one water molecule in its first hydration shell, but the energetic loss is mostly compensated by the hydrogen bond with the backbone carbonyl oxygen of Phe161 as well as the π-cation interaction with Trp148. As a consequence, the overall solvation energy reduction is only 3.1 kcal/mol. With the further movement of NH_4^+ towards the channel, Ser219 plays its role by forming a strong hydrogen bond with NH_4^+. In the matter of fact, in the range from −17 Å to −16 Å, the free energy changes little. Nevertheless, the interesting finding in this range is that the first gate of the channel starts to open dynamically (see next subsection).

The second binding site (−12.2 ∼ −9.2 Å) is located in the beginning of the channel and NH_4^+ has already entered into the channel. One water molecule was observed to follow NH_4^+ into the channel due to the strong hydrogen bonding interaction between them. While Phe107 has already returned to its original position and closed the channel, Phe215 starts to open its gate as its phenyl ring, together with Trp212, form a binding pocket for NH_4^+ via stabilizing π-cation interactions. Apart from one hydrogen bond with water, the ion forms hydrogen bonds with surrounding residues. Interestingly, we also observed a noticeable barrier (1.3 kcal/mol) in the middle of the second binding region, which is thus divided into two areas of similar energies. Structural analysis showed that this small barrier results from the switching of hydrogen bonds. Compared with the first binding region, there is a considerable energy loss for the system at the second binding region by 5.6 kcal/mol and the barrier between these two sites is about 6.9 kcal/mol (Figure 15-6). This is a little lower than the estimated electrostatic barrier 10 kcal/mol by Zheng et al. [24], who also suggested that NH_4^+ may lose its proton at this site to His168. Energetically, it seems probable, although it is certainly more favorable for NH_4^+ to deprotonate at the first binding site as the barrier from the first to the second binding site is about 6.9 kcal/mol. But one concern for this deprotonation mechanism at the second binding site is how the proton can be shipped out of this channel. Also it should be pointed out that there is no water found in the crystal structure, suggesting no appropriate hydrogen bond chain via water molecules to facilitate the transduction of proton out of the channel. Besides, the protonated His168 is still around NH_3 and there will be a strong electrostatic attraction between them. Moving NH_3 down the channel thus will be energetically very costly. Thus, it is more likely that NH_4^+ deprotonates at a periplamic binding site. However, we note that most recent site-directed mutagenesis experiment confirmed the essential role of His168 and His318 in the substrate conductance [48].

After the second binding site, a sharp increase of energy is observed if the ammonium ion keeps moving down the hydrophobic channel, largely due to the high desolvation penalty and the lack of stabilizing hydrogen bonds with the channel. This strongly confirms that ammonium ions can not go through the very hydrophobic channel and the conducted species thus must be neutral ammonia molecules. Obviously, thus computational finding is in an excellent agreement with the experimental evidences.

15.4.1.3. Gating Mechanism

Crystal structures of AmtB manifest two gates in the entrance of the channel, including the phenyl rings of Phe107 and Phe215, as shown in Figures 15-3 and 15-5b. In both structures with and without ammonia or methyl ammonia, these two gates are in parallel and block the channel [23]. Apparently, the passage of ammonium/ammonia through the gates must be a dynamical and fast process, and it would be of interests to study the related gating mechanism.

Our MM MD simulations on the AmtB-NH_4^+ complex showed that when the ammonium ion is far away from the binding vestibule, no apparent change of the positions of the two phenyl rings of Phe107 and Phe215 could be found. Interestingly, with the approaching of NH_4^+ to the first gate (Phe107), we observed a rotation (or more appropriately, opening) of the phenyl ring of Phe107. Figure 15-7 animated this process by superimposing 11 snapshots along the movement of the ammonium ion from the outside to the inside of the binding vestibule. It is clear that the gating-opening is not a gradual but rather a dynamical process. There are only two states for Phe107, namely "closed" (perpendicular to the channel) or "open" (parallel to the channel). When the ammonium ion gets close to Phe107, the latter will choose the "open" state, otherwise it will stay in the "closed" state. This gate-opening process does not require additional energy (by comparing with the free energy profile) and thus is a very fast process.

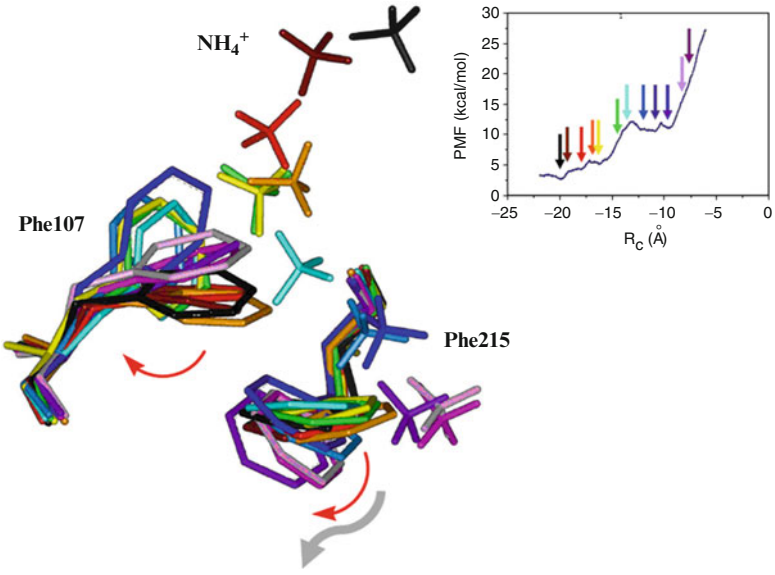


Figure 15-7. The dynamical opening of the gate gates by the ammonium ion, where colors refer to different reaction coordinates as shown in the upright corner

The continuing movement of NH_4^+ gradually pushed the second gate, namely the phenyl ring of Phe215, open. However, this process accompanies a high energy cost as demonstrated by the energy profile (Figure 15-6). This energy cost should mainly come from the loss of hydrogen bonds with water molecules which are shut out of the channel due to the narrowness and hydrophobicity of the channel. Thus, we conclude that NH_4^+ is not responsible for the opening of the second gate of Phe215, which will be opened by the neutral NH_3 as discussed later.

15.4.1.4. Specific Role of Asp160

There are two highly conserved aspartates, Asp160 and Asp310 in the Amt proteins of bacteria, fungi and plants. Whereas there is no evidence yet to show that they actively participate in the ammonia transporting process, their roles are generally believed to be structural as their acidic side chains form hydrogen bonds with main chain N-H groups (residues 163–165 and 314–316, respectively). These hydrogen bonding interactions can fix the backbone of the helical structure preceding His168 and His318 [24]. Although structural modeling of the AmtB protein suggests Asp160 as a potential candidate for an initial ammonium binding site on the periplasmic face of the membrane [112], X-ray structures clearly rule out this possibility. Mutation of Asp160 to Ala160 completely disables the transporting capability, while the D160E mutant retains 71% of the activity of the wild type [113]. It has been shown that Asp160 is a helix-capping residue for the fifth helix of AmtB and its carboxylate group forms hydrogen bonds with Thr165, Gly164 and Gly163 at the N-terminal end of the fifth helix. Although it is an ionic residue, Asp160 itself is not accessible to any solvent molecule because the indole ring of Trp148 is interposed between Asp160 and the binding vestibule for the ammonium ion and thus shields Asp160 from the solvent [23]. Based on the crystal structure, Asp160 seems playing a structural role. But we note that sometimes static structures can hardly recognize the exact role of certain amino acids as X-ray crystallographic experiments usually are unable to capture the structures in motion.

Of significance, our simulation results revealed that in the periplasmic binding vestibule, the dynamical flipping of the phenyl ring of Phe107 is accompanied by a slight rotation and shifting of the indole ring of Trp148. Meanwhile, the carboxylate group of Asp160 rotates and subsequently re-orients the backbone carbonyls of Gly163 and Ala162. All of these collective movements lead to a slot for Asp160, which is originally buried inside the protein, to become exposed to the bulk solvent. As a consequence, we clearly observed a hydrogen bond wire between NH_4^+ and the carboxylate group CO_2^- of Asp160 through two water molecules. Figure 15-8 depicts the changes of the surroundings of the carboxylate group of Asp160 with the approaching of ammonium ion from bulk solvent to the first binding vestibule by superimposing two snapshots with the ammonium ion outside and inside of the binding vestibule. It can be seen that the carboxylate group re-orients to point to the solvent side, and one water molecule enters into the cavity with the shifting of Trp148, Gly163 and Ala162 (not shown) to bind with the carboxylate group.

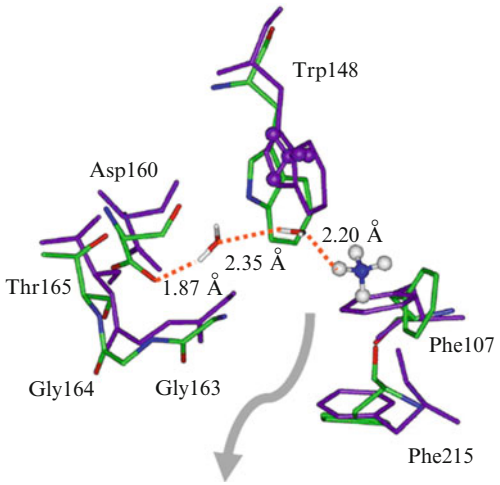


Figure 15-8. The superimposition of two snapshots at Rc = −20.7 Å (in purple) and −15.6 Å (in multi-colors). The hydrogen bond chain between Asp160 and NH4+ via water molecules is depicted in dashed red line

This water molecule connects to another solvent molecule which is in the first hydration shell of the ammonium ion. Therefore, we hypothsize that the deprotonation process occurs in the first binding site, and particularly, *Asp160 may be the initial proton acceptor or at least the driving force for the deprotonation of ammonium ions*. This can reasonably explain the sequence conservation and extreme functional importance of this residue. We note that Bostick and Brooks proposed that the most plausible proton acceptor is water [31]. But even in the latter proposal, we believe that Asp160 should serve as the driving force (catalyst). Moreover, in our proposal that Asp160 is the proton acceptor, water is still the ultimate proton carrier as the proton will transfer back from Asp160 to solvent, which is a typical ionization process for aspartatic acid at normal pH values.

15.4.1.5. Homology Model of the D160A Mutant

To clarify the role of Asp160, we also performed molecular dynamics simulations on the D160A homology mutant and compared its ammonium transport capability with that of native AmtB. After performing 100 ps simulations for equilibrium, the final mutant configuration was superimposed with the wild-type protein and the RMSD is 1.16 Å, which suggested the high similarity of both structures (see Figure 15-9). But our focal point was the surroundings of Asp160. In AmtB, Asp160 was a helix-capping residue for the fifth helix and its negatively charged carboxyl group is inaccessible to the solvent and used to fix the backbone of the helical structure by forming hydrogen bonds with backbone N-H groups of residues 163–165 [23, 24, 28]. Its main chain carbonyl group, however, directs to the solvent

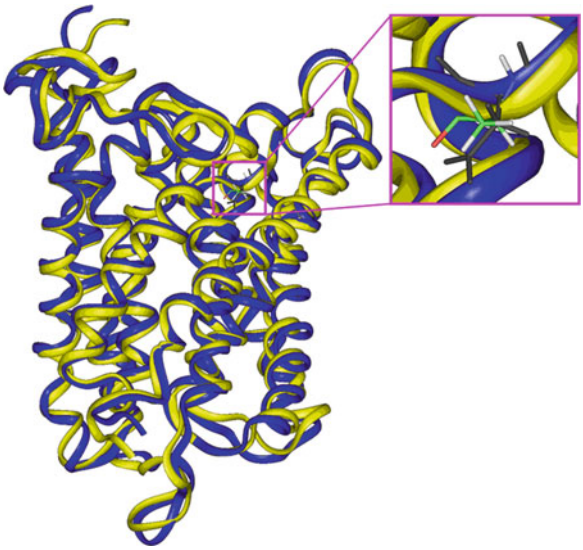


Figure 15-9. The orientations of Asp160 in AmtB (in black) and Ala160 in the D160A mutant (in color) after superimposing these two protein structures (AmtB in blue and D160A mutant in yellow)

and forms hydrogen bonds with water molecules. With the removal of the carboxyl group, the main chain carbonyl group of Ala160 in the D160A mutant flips to the position originally taken by the carboxyl group of Asp160 (Figure 15-9) to form hydrogen bonds with the main chain N–H groups of Gly163 and Gly164 to sustain the framework of the fifth helix. As a consequence, *no remarkable structural difference between AmtB and its D160A variant is observed*. Of course, this requires further confirmation from X-ray crystallographic experiments.

Subsequent PMF simulations along the binding and transduction trajectory of NH_4^+ through the D160A channel (Figure 15-10) revealed that Asp160 is responsible for the recognition and binding of NH_4^+ in AmtB. Luzhkov et al. found that the mutation of Asp160 to Asn destabilizes the bound NH_4^+ by 10 kcal/mol [29]. Albeit a significant loss of binding and a barrier, NH_4^+ can remain in the periplasmic binding vestibule ($R_c = -17.5 \sim -15.5$ Å) of the D160A mutant stably, as demonstrated by Figure 15-10. However, there is no intracellular binding site in D160A. Although the negatively charged carboxyl group of Asp160 is buried and about 8 Å away from the transported ammonium ion, our analysis showed that the high binding energy for NH_4^+ in the binding site of AmtB is significantly contributed from the negatively charged carboxyl group of Asp160, in both terms of direct electrostatic interaction and enhanced hydrogen bonding with the backbone carbonyl group of Ala162 and the amine group of Gly163. This explains the occurrence of the intracellular binding site in AmtB but none in the D160A mutant, as well as the high conservation of Asp160 in the Amt proteins and why the D160A mutant would completely lose the transport capability.

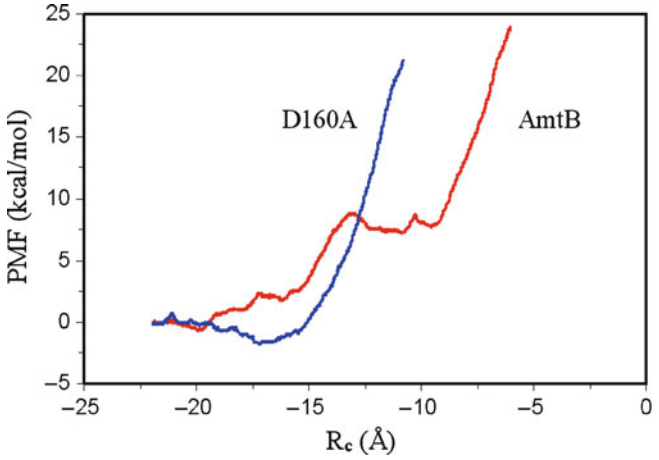


Figure 15-10. Potential of mean force (PMF) profiles for the passage of ammonium across the AmtB (in red) and the D160A mutant (in blue) proteins. The free energies at the starting configurations are referenced as zero

15.4.1.6. The Transduction of Ammonia

As we hypothesized that ammonium ions lose their protons in the periplasmic ion-binding vestibule to the carboxylate group of Asp160 through water molecules (though the protons will return to water ultimately), we took one proton away in the simulations at this stage from NH_4^+ and attempted to derive the free energy profile for the conduction of neutral ammonia through the channel starting from this site. We chose an equilibrium configuration after 100 ps simulation for the NH_4^+-AmtB complex at $R_c = -15.8$ Å and removed the proton pointing toward the carboxylate group of Asp160. The subsequent 100 ps simulation for the NH_3-AmtB complex perturbs the amino acid orientations surrounding the neutral ammonia due to the dramatic reduction of hydrogen bonds of substrate (NH_3) with Ser219 and water molecules compared with the ammonium ion. Figure 15-11 shows the superimposition of two respective snapshots of NH_4^+ and NH_3 in the binding vestibule (at $R_c = -15.8$ Å). In accord with the very low solvation energy of ammonia (4.3 kcal/mol [114]) compared with the very high value of ammonium ion (~80 kcal/mol [109, 110]), the water molecules around NH_3 tends to move away apparently due to the greatly weakened hydrogen bonds, and ammonia re-orientates to make the nitrogen lone pair points toward the solvent side to form a hydrogen bond with solvent molecules. Notably, the water molecule just above the phenyl ring of Phe215 (see Figure 15-11) loses the binding to NH_3 but it is still trapped in the vestibule. Due to the hydrophobic nature of the deep site formed by Ala162, Phe215 and Phe107, water is difficult to reside in this pocket stably as the favorable water–water interactions can slightly stabilize the system. Thus, with the deprotonation of NH_4^+, the trapped water molecule has a tendency to move back the bulk

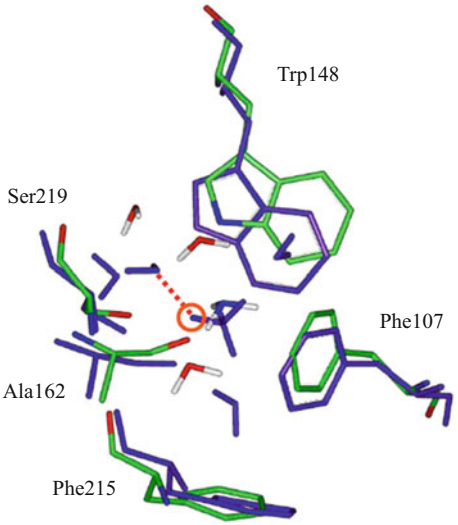


Figure 15-11. Snapshots before (in purple) and after (in multi-color) the deprotonation of ammonium ion at the binding vestibule (Rc = −15.6 Å). The transferred proton is highlighted with a circle in orange

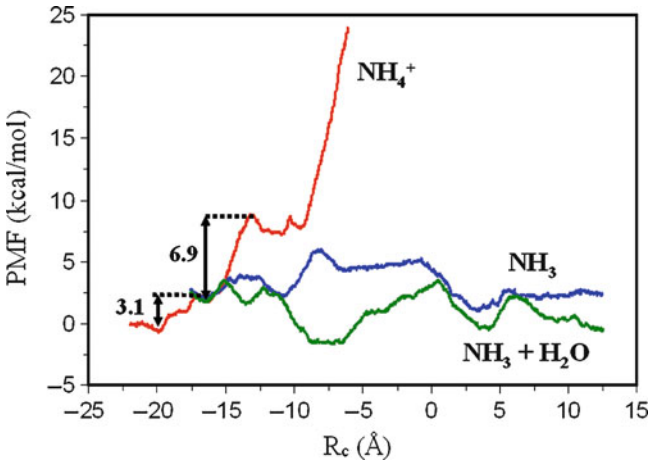


Figure 15-12. Free energy profiles for the passage of ammonium ion across the E. coli AmtB membrane protein

solvent. However, our test simulations starting from this site but with different equilibrium conformations suggest that this trapped water may go through the channel coupled with the ammonia before being able to get out. As a consequence, we performed complete simulations on two trajectories, one with NH_3 alone, and the other with both NH_3 and H_2O through the AmtB channel, whose free energy profiles are plotted in Figure 15-12. Note that no constraint was imposed in the

simulations on the NH_3–H_2O case, i.e., H_2O moves with NH_3. For comparison, the energy profile for the transduction of NH_4^+ was also plotted.

The passage of ammonia through the channel is relatively smooth and the biggest bump is only 3.6 kcal/mol. When the ammonia travels alone, the first energy minimum is found at around −11.1 Å, due to the favorable hydrogen bond between NH_3 and His168. The energy maximum appears at −8.3 Å, and may result from the environmental effect (i.e., the electrostatic field imposed by the surrounding protein environment) as we did not observe any near-range interactions except the hydrogen bond with His168. After a long movement with the breaking of the hydrogen bond with His168, NH_3 starts to stabilize slightly again at 3.8 Å by forming a hydrogen bond with His318. A further transduction without noticeable barrier makes the ammonia exit the channel at $R_c = 8$ Å where Phe31 holds the gate. However, we did not observe the rotation of the phenyl ring of Phe31. Instead, the ammonia gets out of the channel around the phenyl ring. Since the passage of ammonia through AmtB is a freely reversible process, and a specific base (Asp160) is required for the deprotonation of the ammonium on the periplasmic side, there must also be a specific base for deprotonation on the cytosolic side for transport in the reverse (and an acid for the forward transport).[39] We examined the structure and found that at the exit ($R_c = 9$ Å), the ammonia connects via two water molecules to the carboxylate group of Asp313 and the ammonium group of Lys303 which may play roles in the protonation/deprotonation of ammonia/ammonium on the cytosolic side.

The trajectory of the NH_3–H_2O complex is slightly different due to the involvement of the water. For instance, the first minimum comes early at −13.5 Å due to the NH_3...H_2O...His168 hydrogen bridge. In contrast to the maximum state for the first trajectory at −8.3 Å, the NH_3+H_2O complex finds the minimum state at this site. Configurational analyses indicated that apart from the different orientations of the conducted species, there are no other apparent discrepancies. We therefore surmise the difference originates from the environment effect. Both energy profiles reaches minima together at 3.8 Å as at this moment the water in the second case has already slipped out of the channel. In the simulations, we were unable to observe any other external water coming into the channel from either the top or the bottom. This is different from the simulations of the transduction of the ammonium ion where we observe water molecules coming from the bottom due to the strong electrostatic attraction from the cation.

Conformational analyses along the transduction of NH_3 clearly demonstrated that the second gate, namely the phenyl ring of Phe215, is opened by NH_3. Figure 15-13 animated this process by superimposing 10 snapshots along the movement of the neutral ammonia from the binding vestibule where the ammonium ion loses a proton to the inside AmtB channel. Much similar to the first gate, the gating-opening is a dynamical process, as there are only two states for Phe215, namely “closed” or “open”. When ammonia approaches o Phe215, the latter chooses the “open” state. As soon as ammonia passes the gate, Phe215 shuts the door by rotating its phenyl ring.

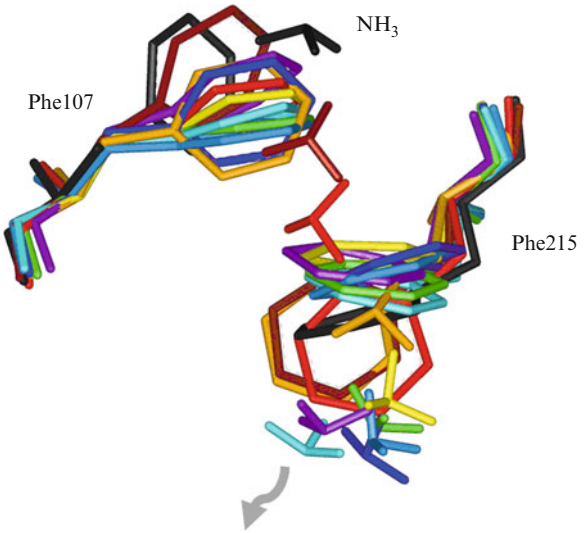


Figure 15-13. The dynamical opening of the gate Phe215 by the ammonia, where colors refer to different reaction coordinates

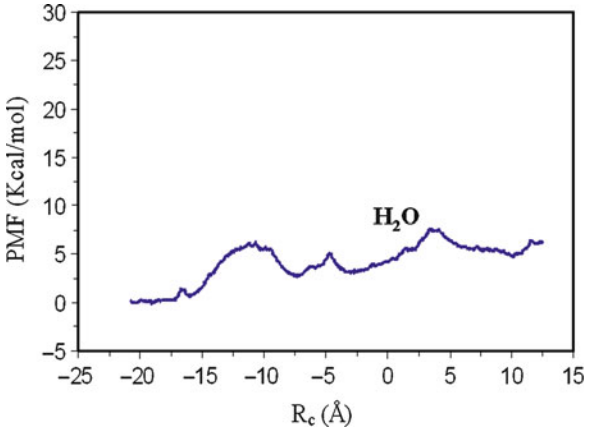


Figure 15-14. Free energy profiles for the passage of water across the E. coli AmtB membrane protein

15.4.1.7. The Transduction of Water

Due to the similarity between H_2O and NH_3 in terms of size and dipole moments, it is intriguing whether the AmtB channel can facilitate the passage of water molecules. To answer this question, we computed the free energy profile for the passage of water through the ammonia transport channel. Similar to the calculations of AmtB with ammonia/ammonium, we initiated the MD simulations at $R_c = -20.7$ Å (which is the initial site for NH_4^+). Figure 15-14 shows the free energy profile for

the transduction of a water molecule through the AmtB channel. Interestingly, there is no apparent periplasmic binding site for water, in contrast to the case of ammonium. In addition, water requires about 6 kcal/mol to pass the two gates, which rotate to allow the water through. Since the ammonium ion needs only 3.1 kcal/mol to enter the periplasmic binding vestibule, the binding site outside the AmtB channel favors the ammonium ions over water molecules. Once the water molecule enters the channel, however, the energy trajectory is relatively smooth with only low barriers which can be easily overcome at room temperature, as the hydrophobic channel can incorporates the hydrogen bond acceptor not only ammonia but also water using weak interactions with C–H bonds [23, 24]. The two highly conserved residues in the mid-membrane center of the pathway, His168 and His318, can interact with water a little more strongly than with ammonia. We note that Khademi et al. investigated the possible water conductivity by AmtB by the measurement of osmotic permeability[115] and the vesicle shrinkage and swelling [116]. Both experiments indicated no additional water conductivity in AmtB proteoliposomes. The structure of AmtB with NH_4^+ also showed no ordered water in between the hydrophobic constricted regions of the channel [23]. But based on our free energy computations, it seems that water can move through the AmtB channel. Obviously, further studies are needed in this regard to verify the role of AmtB in water transportation.

15.4.2. Combined QM(DFT)/MM Studies

The above MM molecular dynamics simulations not only confirmed that the transported species is NH_3 not NH_4^+, but also pinpointed the exact role of Asp160 which is highly conserved in Amt proteins, i.e., the most probable proton acceptor from NH_4^+, as illustrated in Figure 15-15. So far there is still no convincing experimental evidence to either support or disapprove this NH_4^+ deprotonation mechanism. In the following, we continued the computational studies by performing combined QM

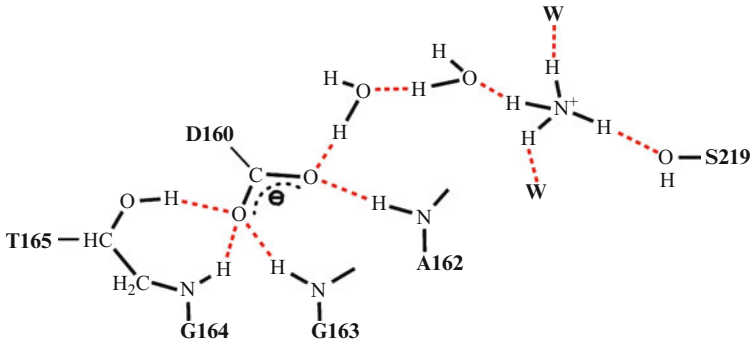


Figure 15-15. Our proposed NH_4^+ deprotonation mechanism (the hydrogen bonds are highlighted in red)

(DFT)/MM computations to examine the geometries and energetic in the proton transfer from NH_4^+ to Asp160 via two water molecules.

15.4.2.1. Computational Model

To examine our claim that Asp160 is most probably the proton acceptor from the ammonium ion through two water molecules, we performed single-point combined QM/MM geometry optimizations on the snapshot partially shown in Figure 15-8 where the QM is at the DFT level [100]. Instead of using a unit cell model which requires much more computational resources, at this stage we simply took the monomer out and put it into a water sphere of 30 Å as shown in Figure 15-16a. The QM/MM calculations consist of a QM region with 149 QM atoms (involving selected residues within 5 Å around NH_4^+, see Figure 15-16b) and 13106 MM atoms. In the QM/MM geometry optimizations, the QM region and 1774 MM atoms (defined by including all atoms within 13 Å around NH_4^+) were allowed to relax, while the remaining MM atoms were constrained. The QM part was treated at the B3LYP/6-31G(d) level, and the MM part was described by the CHARMM22 force field [88]. An electronic embedding scheme [117] was adopted, and the ChemShell package [101] integrating the TURBOMOLE [102] and DL-POLY [103] programs was employed to carry out the QM/MM calculations.

15.4.2.2. Geometries and Energetics in the NH_4^+ Deprotonation Process

Similar to the proton transfer paths in many biochemical processes such as the intramolecular proton transfer between zinc-bound water and His64 with the aid of water in carbonic anhydrase II (CAII) [118–122], the deprotonation of NH_4^+ to

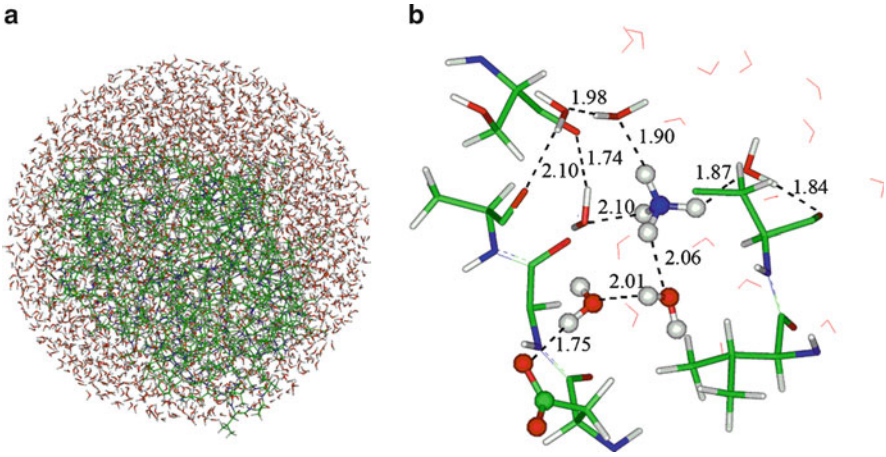


Figure 15-16. (**a**) Spherical model for combined QM/MM geometry optimizations; (**b**) The QM part includes the surrounding residues and six water molecules. The focused proton transfer path is highlighted in a ball-stick style

Asp160 can occur via a sequential hopping of protons through water molecules analogous to the von Grotthuss mechanism explaining the anomalously high proton mobility in water [123], or via a stepwise mechanism where the proton is first transferred from NH_4^+ to the adjacent water, then to the next water, and finally to the carboxylate group. However, it is also possible that the negative rather than the positive charge migrates. In this case, the water molecule close to the carboxylate group first loses a proton to the latter, and then another water molecule loses a proton to the hydroxide ion before abstracting the proton from NH_4^+. This alternative to the Grotthuss mechanism for a proton transfer in solution and CAII was investigated by Riccardi and coworkers very recently [124]. As a preliminary study, here we first optimized the reactant and product structures, and then used the linear synchronous transit approach to generate a reaction path by geometric interpolation between reactant and product, which provided the starting point for the subsequent transition state search.

Our single-point optimizations revealed that NH_4^+ is stabilized in the binding vestibule by a hydrogen bond network around it in AmtB. In particular, there is a hydrogen bond wire between NH_4^+ and the carboxylate group CO_2^- of Asp160 mediated by two water molecules as shown in Figure 15-15. This is in agreement with our molecular mechanical MD simulations (see Figure 15-8) [28]. Figure 15-17 showed the optimal geometries of the initial (reactant) state, an intermediate state and the pre-final state. Of significance, our calculations resulted in a stepwise mechanism. If we restrict the proton transfer process with a concerted mechanism, the barrier would be as high as ~30 kcal/mol. Interestingly, in the stepwise route, Asp160 in the initial conformation (Asp-NH_4^+, as shown in Figure 15-17) accepts one proton, leading to a hydroxyl anionic conformation (AspH-NH_4^+) with a ZPE-corrected barrier of 7.2 kcal/mol. The generation of a hydroxide ion (OH^-) in this step seems very unusual, as the hydrolysis of water molecules normally occurs at a transition metal ion center where the subsequent binding of OH^- to the metal ion significantly lowers the reaction barrier and stabilizes the system.

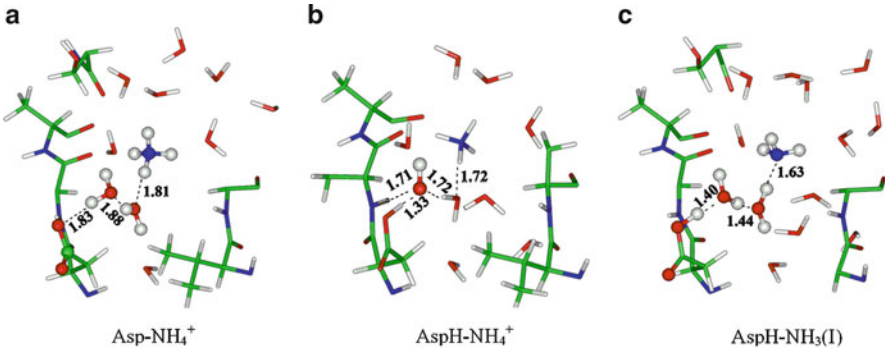


Figure 15-17. Combined QM/MM geometry optimization results: (**a**) Initial state; (**b**) Intermediate state; (**c**) Pre-final state. The intermediate structure was found by the linear synchronous transit approach

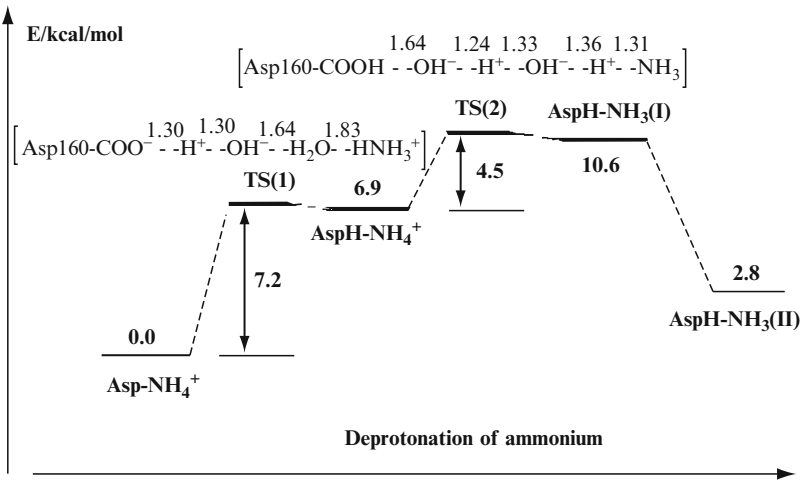


Figure 15-18. Relative energy profile with zero-point energy corrections for the deprotonation of NH_4^+ to Asp160 mediated by two water molecules

In the present case, the hydroxide ion is stabilized by three hydrogen bonds, namely Asp-COOH···OH^-, HOH···OH^-, and Phe161-NH···OH^-, and as a consequence, the AspH-NH_4^+ conformation is less stable than Asp-NH_4^+ by only 6.9 kcal/mol. Tuckerman et al. have demonstrated using BLYP-based MD simulations that similar conformations of hydrated OH^- with a threefold oxygen coordination play a crucial role for the fast transport of OH^- in aqueous solution [125, 126]. The subsequent concerted proton transfer from NH_4^+ to OH^- via one water molecule yields a metastable species AspH-NH_3(I) with a barrier of 4.5 kcal/mol. With the relaxation of the surrounding hydrogen bond network, AspH-NH_3(I) evolves into the final conformation AspH-NH_3(II) (structure not shown, but see Figure 15-18) which is stabilized by 7.8 kcal/mol compared with AspH-NH_3(I) but slightly higher in energy than the initial conformation Asp-NH_4^+ by 2.8 kcal/mol. The overall proton transfer process has a Gibbs free energy change ΔG of -1.8 kcal/mol at 298.15 K. Figure 15-18 depicts the predicted relative energy profile for the overall proton transfer process.

The structural environment of NH_4^+ at the binding vestibule, as shown in Figure 15-16b, exhibits other possibilities of proton transfer from NH_4^+ to the carbonyl group of residues Trp148, Phe161, and Ser219 through hydrogen bond wires. However, our calculations showed that these proton transfer processes are endothermic by more than 26 kcal mol^{-1}, and these amino acids are thus less likely to be the proton acceptor for NH_4^+. In conclusion, our high-level combined QM/MM computations, though limited with a reduced model, demonstrated that energetically the ammonium ion is indeed able to lose its proton to the carboxylate group of Asp160 via two water molecules.

15.4.3. Combined QM(PM3)/MM Molecular Dynamics Simulations

After performing combined QM(DFT)/MM geometry optimizations of key configurations with a reduced spherical model which endorsed our NH_4^+ deprotonation hypothesis, we moved forward and accomplished the extensive QM(PM3)/MM molecular dynamics simulations with the initial cubic model (Figure 15-5a) [104].

Computationally, localized proton transfers have been well studied [51, 75, 127–131], but long-range proton transfers in biochemical processes are more challenging to quantify since the exact transfer pathways and rate-limiting factors are difficult to identify due to the involvement of many residues and solvents [121, 132–143]. One major difficulty in the theoretical studies of long-range proton transfers is the definition of a reaction coordinate to characterize the progress of long-range proton transfer [122]. Whereas geometric coordinates (such as bond distance changes) are appropriate to study chemical processes in the gaseous phase, for reactions in the condensed phases there are collective solvent reorganizations which should be considered in the reaction coordinates, particularly in the charge-transfer or proton-transfer reactions where the protein and solvent undergo significant reorganization [144]. A promising solution is to define reaction coordinates for long-range proton transfers by considering the movement of the excess charge center [122, 145, 146]. For our current case where only two water molecules are involved, however, a linear combination of donor-proton and acceptor-proton distances might be effective [122].

Alternatively, reaction coordinates can be defined in terms of energy. Warshel proposed the empirical valence bond (EVB) method and defined the energy gap between the reactant state and product state as a reaction coordinate [140, 147–149]. Voth and others have extended the simple EVB ideas to modeling proton transfer reactions in aqueous systems with multistate EVB configurations [150, 151]. We have developed a MOVB method which combines the block-localized wavefunction (BLW) method [152, 153] with explicit solvent molecules. The solvent molecules are treated molecular mechanically, and the dynamics of chemical reactions in solution can be studied by performing Monte Carlo simulations. The QM-MOVB/MM method has been applied to three nucleophilic substitution reactions in aqueous solution [154–156]. The proton transfer process in Figure 15-15 can be described as a resonance of two states, or

$$-CO_2^- \cdots H_2O \cdots H_2O \cdots NH_4^+(I) \leftrightarrow -CO_2H \cdots H_2O \cdots H_2O \cdots NH_3(II)$$

While each state can be defined with one BLW, the energy difference between the two states can be used to define the reaction coordinate. In the present study, however, we chose geometric coordinates and used a linear combination of donor-proton and acceptor-proton distances to define the reaction coordinate. But the proton transfer from NH_4^+ to Asp160 involves three hydrogen bonds which may break or form sequentially (as the well-known von Grotthuss mechanism[123]) or step-wisely. In either way, the departure of a proton from NH_4^+ initiates the whole process. Thus, we chose the distance difference between the bonds N-H1 and

O1-H1 as the reaction coordinate in the combined QM(PM3)/MM molecular dynamics simulations (see Figure 15-19).

In the present QM(PM3)/MM simulations, the QM part consists of the ammonium ion, two water molecules and the carboxyl group of Asp160 where the QM/MM boundary atom Cγ (Figure 15-19) was treated with the generalized hybrid orbital (GHO) method [86]. In total, there are 17 QM atoms.

Figure 15-20 shows the energy profile along the proton transfer process, where the negative R_C corresponds to the reactant state and positive R_C to the product state. The computations demonstrated that the overall reaction barrier is only 7.7 kcal/mol which is in good agreement with the value of 7.2 kcal/mol derived from the QM(DFT)/MM single-point optimizations as shown in Figure 15-18 [100]. The NH_4^+ deprotonation process is exothermic due to the combination of the positive and negative charges. Conformational analyses showed that at the reactant state ($R_C = -0.85 \sim -0.75$ Å, and a snapshot is shown in Figure 15-21a), the

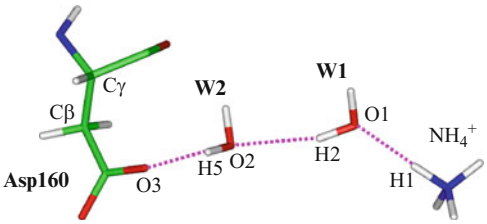


Figure 15-19. Atomic labels for the definition of the reaction coordinate for the deprotonation pathway via a hydrogen bond wire between ammonium and the carboxyl group of Asp160 through two water molecules

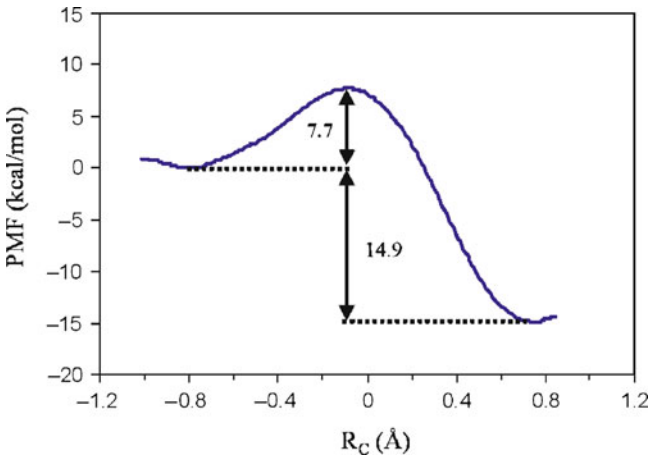


Figure 15-20. Potential of mean force (PMF) profile for the process of proton transfer from the ammonium ion to the nearest water molecule which links to the carboxyl group of Asp160 via another water molecule

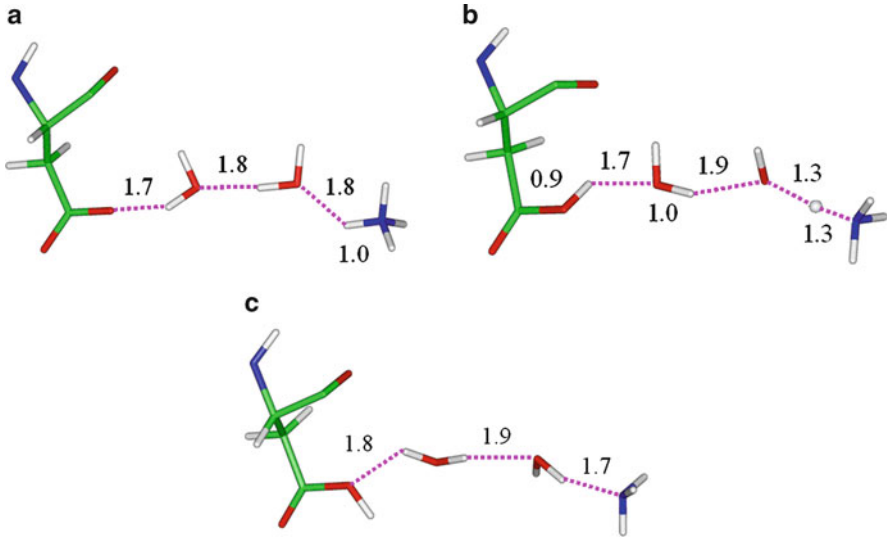


Figure 15-21. Snapshots for the structural changes of the reactive part along the deprotonation of NH_4^+ in the combined QM/MM molecular dynamics simulations: (**a**) reactant state ($R_C = -0.8$ Å); (**b**) transition state ($R_C = -0.1$ Å); (**c**) product state ($R_C = 0.7$ Å)

ammonium ion forms a hydrogen bond network with three water molecules as well as the hydroxyl group of Ser219, while the carboxyl group of Asp160 interacts with the hydroxyl group of Thr165 and one water molecule. The ideal hydrogen-bond wire from NH_4^+ to Asp160 is observed. The subsequent proton transfer experiences a transition state at $R_C = -0.09$ Å where the transferred proton H1 lingers in the midst of the donor and acceptor. Figure 15-21b shows a geometry at the transition state. Remarkably, the negatively charged carboxylate ion of Asp160 has already been neutralized by a proton (H5) transfer from the adjacent water molecule (W2), which has simultaneously got a proton (H2) from W1 to maintain its neutral structure. As a consequence, the water molecule (W1) in the first hydration shell of NH_4^+ is deprotonated and becomes a hydroxide anion in the transition state. This is actually consistent with the QM(DFT)/MM single-point optimizations and MD simulations [100], which also suggested the existence of a hydroxide ion in the intermediate state except that the hydroxide ion is in the W2 position.

Since the present reaction coordinate only considers the transfer of H1 and thus does not reflect the movement of H2 and H5, we traced the distance changes for hydrogen bonds O3-H5 and O2-H2 along the reaction coordinate. Figure 15-22 plotted the fluctuations of these bond distances. While the changes for the bonds O1-H1 and N-H1 are smooth as their difference is used as the reaction coordinate, there are dramatic changes for the bonds O2-H2 and O3-H5 at $R_C = -0.6$ Å which is still quite close to the reactant state rather than to the transition state. When R_C is less than -0.6 Å, we observed large fluctuations of these hydrogen bond distances which fall into two groups, one is at around 1.8–2.0 Å and the other varies at 2.3–3.2 Å.

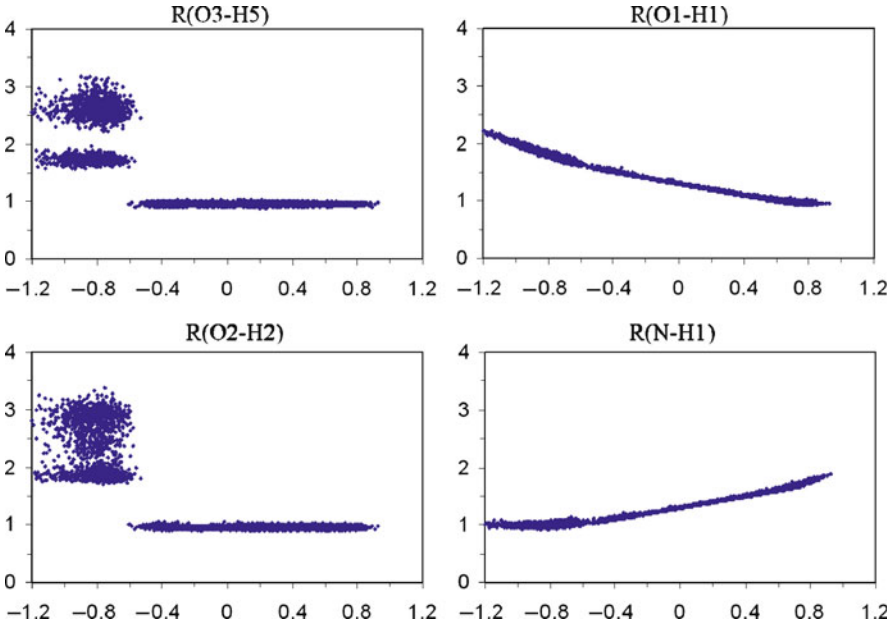


Figure 15-22. Fluctuations of bond distances along the reaction coordinate which is defined as the bond difference between R(N–H1) and R(O1–H1)

These two groups actually reflect the flipping/rotation of two O–H bonds in both water molecules. Once R_C moves over −0.6 Å, however, the O2–H2 and O3–H5 bonds fluctuates little at 1.0 Å, implying strong chemical bonds. Thus, the proton transfers along the hydrogen bond wire W1-W2-Asp160 are sequential with very little energetic costs, and the net outcome is the generation of a hydroxide ion as an intermediate. In other words, the barrier (7.7 kcal/mol) mostly results from the breaking of the N–H1 bond and the formation of the O1–H1 bond. The very low barrier from the combined QM(PM3)/MM molecular dynamics simulations once again supported the hypothesis that NH_4^+ loses a proton to Asp160 via two water molecules.

15.5. SUMMARY

Our extensive molecular dynamic simulations not only supported the experimental evidences that the neutral ammonia is the transported species, but also showed that the ammonium ion may lose its proton in the binding vestibule. Furthermore, simulations revealed a perfect wire of hydrogen bonds from NH_4^+ to the carboxylate group of Asp160 via water molecules in the end of the binding vestibule, suggesting that Asp160 plays a critical functional role in AmtB. Subsequent combined QM/MM computations where the QM part were treated at both the DFT and PM3 levels led

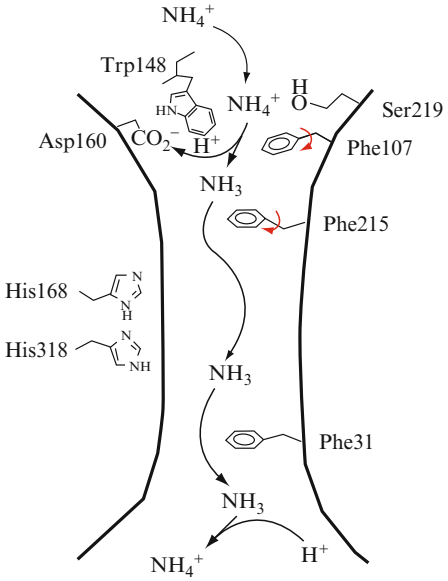


Figure 15-23. Illustration of the NH_4^+/NH_3 transport mechanism in AmtB

to consistent results and further showed that the proton transfer process occurs via an unconventional mechanism by migrating the negative charge (instead of positive charge in conventional mechanisms) from the carboxyl group of Asp160 to NH_4^+ via two water molecules, which can be illustrated as $-CO_2^-\cdots H_2O\cdots H_2O\cdots NH_4^+ \rightarrow -COOH\cdots H_2O\cdots OH^-\cdots NH_4^+ \rightarrow -COOH\cdots H_2O\cdots H_2O\cdots NH_3$. In conclusion, the overall NH_4^+/NH_3 transport mechanism in AmtB was illustrated in Figure 15-23. We note that the present role of Asp160 is still a hypothesis derived from the computational studies, further experimental study thus is essential to examine this NH_4^+ deprotonation hypothesis, and the exact roles of His168 and His318 etc. are yet to be elucidated.

REFERENCES

1. Jayakumar A, Hwang SJ, Fabiny JM, Chinault AC, Barnes EMJ (1989) J Bacteriol 171:996–1001
2. Marini AM, Vissers S, Urrestarazu A, Andre B (1994) EMBO J 13:3456–3463
3. Ninnemann O, Jauniaux JC, Frommer WB (1994) EMBO J 13:3464–3471
4. von Wirén N, Gazzarrini S, Gojon A, Frommer WB (2000) Curr Opin Plant Biol 3:254–261
5. Marini AM, Matassi G, Raynal V, Andre B, Cartron JP, Cherif-Zahar B (2000) Nature Genet 26:341–344
6. Westhoff CM, Ferreri-Jacobia M, Mak DOD, Foskett JK (2002) J Biol Chem 277:12499–12502
7. Le Van KC, Colin Y, Cartron JP (2006) Blood Rev 20:93–110
8. von Wirén N, Merrick M (2004) Trends Curr Genet 9:95–120
9. Knepper MA, Packer R, Good DW (1989) Physiol Rev 69:179–249

10. Lande MB, Donovan JM, Zeidel ML (1995) J Gen Physiol 106:67–84
11. Kleiner D (1985) FEMS Microbiol Rev 32:87–100
12. Marini AM, Soussi-Boudekou S, Vissers S, Andre B (1997) Mol Cell Biol 17:4282–4293
13. Ludewig U, von Wirén N, Frommer WB (2002) J Biol Chem 277:13548–13555
14. Ludewig U (2004). Electroneutral ammonium transport by basolateral rhesus B glycoprotein. J. Physiol. 559: 751–759.
15. Nakhoul NL, Dejong H, Abdulnour-Nakhoul SM, Boulpaep EL, Hering-Smith K, Hamm LL (2005) Am J Physiol Renal Physiol 288:F170–F181
16. Mayer M, Dynowski M, Ludewig U (2006) Biochem J 396:431–437
17. Soupene E, He L, Yan D, Kustu S (1998) Proc Natl Acad Sci USA 95:7030–7034
18. Soupene E, Lee H, Kustu S (2002) Proc Natl Acad Sci USA 99:3926–3931
19. Soupene E, Ramirez RM, Kustu S (2001) Mol Cell Biol 21:5733–5741
20. Meier-Wagner J, Nolden L, Jakoby M, Siewe R, Kramer R, Burkovski A (2001) Microbiology 147:135–143
21. Ripoche P, Bertrand O, Gane P, Birkenmeier C, Colin Y, Cartron JP (2004) Proc Natl Acad Sci USA 101:17222–17227
22. Bakouh N, Benjelloun F, Hulin P, Brouillard F, Edelman A, Cherif-Zahar B, Planelles G (2004) J Biol Chem 279:15975–15983
23. Khademi S, O'Connell J, Remis J, Robles-Colmenares Y, Miericke LJW, Stroud RM (2004) Science 305:1587–1594
24. Zheng L, Kostrewa D, Berneche S, Winkler FK, Li XD (2004) Proc Natl Acad Sci USA 101:17090–17095
25. Khademi S, Stroud RM (2006) Physiology 21:419–429
26. Khademi S, Stroud RM (2006) In: Grisshammer R, Buchanan SK (eds) Structural biology of membrane proteins. Cambridge, Roy Soc Chem, pp 212–234
27. Javelle A, Thomas G, Marini AM, Kramer R, Merrick M (2005) Biochem J 390:215–222
28. Lin Y, Cao Z, Mo Y (2006) J Am Chem Soc 128:10876–10884
29. Luzhkov VB, Almloef M, Nervall M, Aaqvist J (2006) Biochemistry 45:10807–10814
30. Liu YM, Hu XH (2006) J Phys Chem A 110:1375–1381
31. Bostick DL, Brooks CL III (2007) PLoS Comput Biol 3:231–246
32. Mayer M, Schaaf G, Mouro I, Lopez C, Colin Y, Neumann P, Cartron JP, Ludewig U (2006) J Gen Physiol 127:133–144
33. Blakey D, Leech A, Thomas GH, Coutts G, Findlay K, Merrick M (2002) Biochem J 364:527–535
34. Conroy MJ, Jamieson SJ, Blakey D, Kaufmann T, Engel A, Fotiadis D, Merrick M, Bullough PA (2004) Embo Reports 5:1153–1158
35. Andrade SLA, Dickmanns A, Ficner R, Einsle O (2005) Proc Natl Acad Sci USA 102:14994–14999
36. Conroy MJ, Bullough PA, Merrick M, Avent ND (2005) Brit J Haematol 131:543–551
37. Lupo D, Li XD, Durand A, Tomizaki T, Cherif-Zahar B, Matassi G, Merrick M, Winkler FK (2007) USA 104:19303–19308
38. Dougherty DA, Stauffer DA (1990) Science 250:1558–1560
39. Burley SK, Petsko GA (1986) FEBS Lett 203:139–143
40. Scrutton NS, Raine ARC (1996) Biochem J 319:1–8
41. Ma JC, Dougherty DA (1997) Chem Rev 97:1303–1324
42. Gallivan JP, Dougherty DA (1999) Proc Natl Acad Sci USA 96:9459–9464
43. Mo Y, Subramanian G, Gao J, Ferguson DM (2002) J Am Chem Soc 124:4832–4837
44. Callebaut I, Dulin F, Bertrand O, Ripoche P, Mouro I, Colin Y, Mornon JP, Cartron JP (2006) Transfus Clin Biol 13:70–84

45. Huang CH, Peng JB (2005) Natl Acad Sci USA 102:15512–15517
46. Javelle A, Lupo D, Li X-D, Chami M, Ripoche P, Merrick M, Winkler FK (2007) J Struct Biol 158:472–481
47. Winkler FK (2006) Pfluegers Arch Eur J Physiol 451:701–707
48. Javelle A, Lupo D, Zheng L, Li X-D, Winkler FK, Merrick M (2006) J Biol Chem 281:39492–39498
49. Eisenmesser EZ, Bosco DA, Akke M, Kern D (2002) Science 295:1520–1523
50. Eisenmesser EZ, Millet O, Labeikovsky W, Korzhnev DM, Wolf-Watz M, Bosco DA, Skalicky JJ, Kay LE, Kern D (2005) Nature 438:117–121
51. Benkovic SJ, Hammes-Schiffer S (2003) Science 301:1196–1202
52. Garcia-Viloca M, Gao J, Karplus M, Truhlar DG (2004) Science 303:186–195
53. Agarwal PK (2005) J Am Chem Soc 127:15248–15256
54. Bernèche S, Roux B (2001) Nature 414:73–77
55. de Groot BL, Grubmüller H (2001) Science 294:2353–2357
56. Tajkhorshid E, Nollert P, Jensen MØ, Miercke LJW, O'Connell J, Stroud RM, Schulten K (2002) Science 296:525–530
57. Hansson T, Oostenbrink C, van Gunsteren WF (2002) Curr Opin Struct Biol 12:190–196
58. Amaro R, Tajkhorshid E, Luthey-Schulten Z (2003) Proc Natl Acad Sci USA 100:7599–7604
59. Lehnert U, Xia Y, Royce TE, Goh C-S, Liu Y, Senes A, Yu H, Zhang ZL, Engelman DM, Gerstein M (2004) Q Rev Biophys 37:121–146
60. Roux B, Allen T, Bernèche S, Im W (2004) Q Rev Biophys 37:15–103
61. Gumbart J, Wang Y, Aksimentiev A, Tajkhorshid E, Schulten K (2005) Curr Opin Struct Biol 15:423–431
62. Karplus M, Kuriyan J (2005) Proc Natl Acad Sci USA 102:6679–6685
63. Pastor RW, Venable RM, Karplus M (1991) Proc Natl Acad Sci USA 88:892–896
64. Venable RM, Zhang YH, Hardy BJ, Pastor RW (1993) Science 262:223–226
65. Woolf TB, Roux B (1994) Proc Natl Acad Sci USA 91:11631–11635
66. Warshel A, Florian J (1998) Proc Natl Acad Sci USA 95:5950–5955
67. Warshel A, Levitt M (1976) J Mol Biol 103:227–249
68. Field MJ, Bash PA, Karplus M (1990) J Comput Chem 11:700–733
69. Dapprich S, Komaromi I, Byun KS, Morokuma K, and Frisch MJ (1999) A new ONIOM implementation in Gaussian98. Part I. The calculation of energies, gradients, vibrational frequencies and electric field derivatives. J Struct Mol (THEOCHEM) 462:1–21
70. Monard G, Merz KM (1999) Acc Chem Res 32:904–911
71. SIQCMfL M (2000) J Comput Chem 21:1419–1588
72. Zhang YK, Liu HY, Yang WT (2000) J Chem Phys 112:3483–3492
73. Cui Q, Guo H, Karplus M (2002) J Chem Phys 117:5617–5631
74. Field MJ (2002) J Comput Chem 23:48–58
75. Gao JL, Truhlar DG (2002) Ann Rev Phys Chem 53:467–505
76. Truhlar DG, Gao JL, Alhambra C, Garcia-Viloca M, Corchado J, Sanchez ML, Villa J (2002) Acc Chem Res 35:341–349
77. Gordon MS, Schmidt MW (2003) In Computational Science – Iccs 2003, Pt Iv, Proceedings 2660:75–83
78. Monard G, Prat-Resina X, Gonzalez-Lafont A, Lluch JM (2003) Int J Quantum Chem 93:229–244
79. Warshel A (2003) Ann Rev Biophys Biomol Struct 32:425–443
80. Rega N, Iyengar SS, Voth GA, Schlegel HB, Vreven T, Frisch MJ (2004) J Phys Chem B 108:4210–4220
81. Reuter N, Dejaegere A, Maigret B, Karplus M (2000) J Phys Chem A 104:1720–1735

82. Thery V, Rinaldi D, Rivail J-L, Maigret B, Ferenczy GG (1994) J Comput Chem 15:269–282
83. Zhang YK, Lee TS, Yang WT (1999) J Chem Phys 110:46–54
84. Kairys V, Jensen JH (2000) J Phys Chem A 104:6656–6665
85. Murphy RB, Philipp DM, Friesner RA (2000) Chem Phys Lett 321:113–120
86. Gao JL, Amara P, Alhambra C, Field MJ (1998) J Phys Chem A 102:4714–4721
87. Pu JZ, Gao JL, Truhlar DG (2004) J Phys Chem A 108:632–650
88. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) J Phys Chem B 102:3586–3616
89. Kollman PA, Merz KM (1990) Acc Chem Res 23:246–252
90. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) J Phys Chem B 105:6474–6487
91. Lamoureux G, Roux B (2003) J Chem Phys 119:3025–3039
92. Rick SW, Stuart SJ, Berne BJ (1994) J Chem Phys 101:6141–6156
93. Banks JL, Kaminski GA, Zhou RH, Mainz DT, Berne BJ, Friesner RA (1999) J Chem Phys 110:741–754
94. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) J Am Chem Soc 107:3902–3909
95. Stewart JJP (1989) Method J Comput Chem 10:209–220
96. Hong GY, Strajbl M, Wesolowski TA, Warshel A (2000) J Comp Chem 21:1554–1561
97. Elstner M, Porezag D, Jungnickel G, Elsner J, Haugk M, Frauenheim T, Suhai S, Seifert G (1998) Phys Rev B 58:7260–7268
98. Cui Q, Elstner M, Kaxiras E, Frauenheim T, Karplus M (2001) J Phys Chem B 105:569–585
99. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) J Comput Chem 4:187–217
100. Cao Z, Mo Y, Thiel W (2007) Angew Chem Int Ed 46:6811–6815
101. Sherwood P, de Vries AH, Guest MF, Schreckenbach G, Catlow CRA, French SA, Sokol AA, Bromley ST, Thiel W, Turner AJ, Billeter S, Terstegen F, Thiel S, Kendrick J, Rogers SC, Casci J, Watson M, King F, Karlsen E, Sjovoll M, Fahmi A, Schafer A, Lennartz C (2003) Mol Struct – THEOCHEM 632:1–28
102. Ahlrichs R, Bär M, Haser M, Horn H, Kölmel C (1989) Chem Phys Lett 162:165–169
103. Smith W, Forester TR (1996) Mol Graph 14:136–141
104. Lin Y, Cao Z, Mo Y (2009) J Phys Chem B 113:4922–4929
105. Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG (1995) J Chem Phys 103:8577–8593
106. Schlenkrich M, Brickmann J, MacKerell AD Jr, Karplus M (1996) In: Merz KM Jr, Roux B (eds) Biological membranes: a molecular perspective from computation and experiment. Boston, Birkhäuser, pp 31–81
107. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) J Chem Phys 79:926–935
108. van Gunsteren WF, Berendsen HJC (1977) Mol Phys 34:1311–1327
109. Pearson RG (1986) J Am Chem Soc 108:6109–6114
110. Cao Z, Lin M, Zhang Q, Mo Y (2004) J Phys Chem A 108:4277–4282
111. Liu G, Hinch B, Davatol-Hag H, Lu Y, Powers M, Beavis AD (1996) J Biol Chem 271:19717–19723
112. Thomas GH, Mullins JG, Merrick M (2000) Mol Microbiol 37:331–344
113. Javelle A, Severi E, Thornton J, Merrick M (2004) J Biol Chem 279:8530–8538
114. Marten B, Kim K, Cortis C, Friesner RA, Murphy RB, Ringnalda MN, Sitkoff D, Honig B (1996) J Phys Chem 100:11775–11788

115. Priver NA, Rabon EC, Zeidel ML (1993) Biochemistry 32:2459–2468
116. Fu D, Libson A, Miercke LJW, Weitzman C, Nollert P, Krucinski J, Stroud RM (2000) Science 290:481–486
117. Bakowies D, Thiel W (1996) Phys Chem 100:10580–10594
118. Eriksson AE, Jones TA, Liljas A (1988) Proteins: Struct Funct Genet 4:274–282
119. Tu C, Silverman DN, Forsman C, Jonsson B-H, Lindskog S (1989) Biochemistry 28:7913–7918
120. Lu D, Voth GA (1998) J Am Chem Soc 120:4006–4014
121. Toba S, Colombo G, Merz KM (1999) J Am Chem Soc 121:2290–2302
122. König PH, Ghosh N, Hoffmann M, Elstner M, Tajkhorshid E, Frauenheim T, Cui Q (2006) J Phys Chem A 110:548–563
123. de Grotthuss CJT (1806) Ann Chim Phys 58:54–73
124. Riccardi D, König P, Prat-Resina X, Yu HB, Elstner M, Frauenheim T, Cui Q (2006) J Am Chem Soc 128:16302–16311
125. Tuckerman ME, Chandra A, Marx D (2006) Acc Chem Res 39:151–158
126. Tuckerman ME, Marx D, Parrinello M (2002) Nature 417:925–929
127. Hwang J-K, Warshel A (1996) J Am Chem Soc 118:11745–11751
128. Hammes-Schiffer S (2002) Biochemistry 41:13335–13343
129. Cui Q, Karplus M (2002) J Am Chem Soc 124:3093–3124
130. Kiefer PM, Hynes JT (2003) J Phys Chem A 107:9022–9039
131. Liang Z, Klinman JP (2004) Curr Opin Struct Biol 14:648–655
132. Nagle JF, Morowitz HJ (1978) Proc Natl Acad Sci USA 75:298–302
133. Nagle JF, Mille M (1981) J Chem Phys 74:1367–1372
134. Lu D, Voth GA (1998) Protiens Struct Funct Genet 33:119–134
135. Sham YY, Muegge I, Warshel A (1999) Proteins: Struct Funct Genet 36:484–500
136. Pomès R, Roux B (2002) Biophys J 82:2304–2316
137. Wu Y, Voth GA (2003) Biophys J 85:864–875
138. Popovic DM, Stuchebrukhov AA (2004) J Am Chem Soc 126:1858–1871
139. Bondar A-N, Fischer S, Smith JC, Elstner M, Suhai S (2004) J Am Chem Soc 126:14668–14677
140. Schutz CN, Warshel A (2004) J Phys Chem B 108:2066–2075
141. Frank RAW, Titman CM, Pratap JV, Luisi BF, Perham RN (2004) Science 306:872–876
142. Fisher Z, Prada JAH, Tu C, Duda D, Yoshioka C, An H, Govindasamy L, Silverman DN, McKenna R (2005) Biochemistry 44:1097–1105
143. Braun-Sand S, Burykin A, Chu ZT, Warshel A (2005) J Phys Chem B 109:583–592
144. Schenter GK, Garrett BC, Truhlar DG (2001) J Phys Chem B 105:9672–9685
145. Pomès R, Roux B (1998) Biophys J 75:33–40
146. Chakrabarti N, Tajkhorshid E, Roux B, Pomès R (2004) Struct Biol Biochem 12(1):65–74
147. Warshel A, Weiss RM (1980) J Am Chem Soc 102:6218–6226
148. Hwang JK, Warshel A (1996) J Am Chem Soc 118:11745–11751
149. Shurki A, Warshel A (2003) Adv Prot Chem 66:249–313
150. Schmitt UW, Voth GA (1998) J Phys Chem B 102:5547–5551
151. Vuilleumier R, Borgis D (1998) Chem Phys Lett 284:71–77
152. Mo Y, Peyerimhoff SD (1998) J Chem Phys 109:1687–1697
153. Mo Y, Song L, Lin Y (2007) J Phys Chem A 111:8291–8301
154. Mo Y, Gao J (2000) J Comput Chem 21:1458–1469
155. Mo Y, Gao J (2000) J Phys Chem A 104:3012–3020
156. Gao J, Garcia-Viloca M, Poulsen TD, Mo Y (2003) Adv Phys Org Chem 38:161–181

CHAPTER 16

CHALLENGES FOR COMPUTER SIMULATIONS IN DRUG DESIGN

HANNES G. WALLNOEFER[1,2], THOMAS FOX[1], AND KLAUS R. LIEDL[2]
[1]*Computational Chemistry, Department of Lead Discovery, Boehringer Ingelheim Pharma GmbH & Co. KG, 88397 Biberach, Germany*
[2]*Institute of General, Inorganic and Theoretical Chemistry, University of Innsbruck, Innrain 52a 6020 Innsbruck, Austria*

Abstract: Many computational methods have become standard techniques in modern drug discovery. However, approaches which employ explicit molecular dynamics simulations still are restricted to special applications, as their extensive computational requirements make it difficult to obtain results within the necessary time scale of industrial drug development projects. Moreover, a high expertise is needed to analyze and interpret the enormous amount of resulting data. Nevertheless, both the increasing computational power, and theawareness that it is important to use not static, but flexible models of biomolecules to represent the properties of the system of interest, have brought computer simulations back into the focus of interest: they are the most straightforward method to include flexibility into the *in silico* description of molecules. Here, state-of-the-art methods, applications, and arising challenges of molecular dynamics simulations to support drug discovery are discussed.

Keywords: Simulations, Molecular dynamics, flexibility, Structure-based drug design, Virtual screening, Docking, Pharmacophore modelling, Free energy prediction, MM/PBSA, MM/GBSA, Linear interaction energy, FEP/TI, Water in proteins, GPCR simulations

16.1. INTRODUCTION

Structure-based drug design (SBDD) has become an essential part in the early stages of the drug discovery process. For several drugs on the market, like HIV protease inhibitors [1, 2] or neuramidase inhibitors [3, 4], computational methods were an integral part in the discovery of drug molecules. Moreover, an increasing number of drug candidates in clinical testing were developed with the aid of *in silico*

P. Paneth and A. Dybala-Defratyka (eds.), *Kinetics and Dynamics*, Challenges and Advances in Computational Chemistry and Physics 12,
DOI 10.1007/978-90-481-3034-4_16,

methods. Examples are HIV integrase inhibitors [5, 6], hepatitis C protease inhibitors [7, 8] and β-secretase inhibitors [9].

Most of the protein structures, which are the basis of SBDD approaches, were derived from X-ray crystallography. This technique gives a static picture of the protein or the protein–ligand complex, which allows the investigation of the forces and interactions in the system of interest. However, in the last few years the view has changed drastically: now it has become clear that flexibility of the receptor protein plays an important role in ligand interaction and protein function. Experimental methods hardly provide information about flexibility, such as protein domain motions, structural ensembles, alternative conformations, or entropic contributions. Here, computational methods offer the possibility to provide important insights which give additional and/or complementary information to experimental data.

We note that for a compound to become a successful drug, many other aspects besides the affinity to the primary target are important. These features are often summarized under the keywords absorption, distribution, metabolism, excretion, and toxicology [10–15]. However, it is still common practice that adequate affinity, or, even more precisely, the enthalpic part to it, is the primary optimization parameter in a drug discovery project. Medicinal chemists barely consider entropic contributions in the design of potential drug candidates. As significant resources have to be allocated to the experimental determination of the binding affinity, a reliable and accurate *in silico* prediction is highly desirable.

Early on structural flexibility was recognized as an integral part of protein function. Haemoglobin, the first available protein crystal structure [16–19], can adopt two structurally different conformations, named “tense” and “relaxed”, dependent on its oxygenation state. Recently, two additional conformations were reported [20]. Other examples (prominent are e.g. adenylate synthase [21–23], or nuclear receptors [24, 25]) also show that proteins have both an intrinsic possibility and the need to undergo conformational transitions in order to carry out their functions [26, 27]. Even the first drug that was designed using SBDD, the ACE inhibitor Captopril, targets a highly flexibly protein. Here, one has to assume – as in many other cases were ligand design was based on a static structural model without any flexibility – the success was due to some unpredictable induced-fit effects [28, 29].

All above shows that a static structure obtained from X-ray crystallography of the isolated protein is only a part of the whole picture; protein flexibility is crucial for the correct representation of a biomolecule.

Different theoretical models for the conformational behavior of a receptor during ligand binding and the related free energy landscape [30] have been developed. The Koshland–Nemethy–Filmer (KNF) induced-fit model proposes that conformational changes upon ligand binding are initiated at the binding site. Then they are propagated through the protein structure. Thereby, the conformation of the receptor in the complex is not adopted until a ligand binds [31]. Alternatively, the Monod–Wyman–Changeux (MWC) model suggests that a protein can adopt two conformations which are differently populated. The less frequently populated

conformation is stabilized upon ligand binding and thus becomes the dominant one [32]. A more recent adaptation of the MWC model proposes that the protein exists in more than two conformations. In the apo-form it samples many conformations, most of them barely populated [33, 34]. The binding of the ligand changes the free energy landscape in a way that the conformation of the protein in the complex becomes more favorable [30, 34]. In this model, allosteric inhibitors modify the free energy surface such that the active state is less favorable.

In general, protein flexibility upon ligand binding can be classified into three different types: (i) rigid proteins with ligand-induced side chain motions; (ii) flexible proteins performing large motions via hinge regions when the ligand binds; (iii) unstable proteins which become stable when the ligand binds [28].

The protein data bank (PDB) [35], a freely accessible depository of protein structures, currently provides structural data for about 60,000 biomolecules. Most of the structures result from X-ray crystallography; many of those have been measured under unbiological conditions, e.g. very low temperature. Moreover, flexible proteins are usually harder to crystallize. As a consequence, rigid proteins (class i) are overrepresented in the PDB. Genomic and proteomic projects have shown, however, that the other two protein classes (flexible and unstable) are a significant part of the proteome [36].

Despite the importance of a flexible or at least multi-conformational representation of structural conditions in the receptor, a static picture was and still is the standard in SBDD [37, 38]. Consequently, only ligand binding to the available conformation can be addressed. However, the free energy landscape of a protein is a complex hyper surface with a multitude of minima of different depths and curvatures [39–43]. Furthermore, this free energy surface is strongly influenced by the environment conditions such as e.g. ionic strength, pH, or temperature; the interaction with surrounding molecules such as solvent, membrane, or other proteins in the cell also has profound influence on the energy surface and, thus, the conformational preferences. For an accurate representation of the complex, or for thermodynamic predictions such as binding equilibria, not just the one conformation obtained in an X-ray structure, but all conformational states and the probabilities of their populations have to be taken into account [40, 44].

The consequence for drug design is the urgent need to incorporate conformational flexibility into the structural models. A single X-ray conformation of a receptor represents a particular local minimum of the energy hyper surface and, hence, limits the chemical space of putative binders that can be identified by computational approaches. To improve this situation, it would be necessary to map the accessible conformational space of the receptor.

However, the introduction of flexible structures or at least ensembles of structures, which mimic the flexibility, is not straightforward.

On the experimental side, flexibility can be addressed by several techniques. NMR (nuclear magnetic resonance) spectroscopy [45] can in principle observe dynamics and conformational changes of protein motions in time scale from subnano to micro- and milliseconds [46–49]. From multi-dimensional experimental

spectra, geometrical constraints can be extracted, which are used to derive several low-energy conformations. Together, these snapshots can represent the dynamic behavior of the system. However, NMR is limited to small biomolecules, although new techniques push these limits from 20 to 30 kDa [29] to proteins of 80–100 kDa [50–52], or systems in membranes [53]. Another interesting new NMR method determines the structure of a complex in solution by examining the changes in chemical ^{1}H NMR shifts when the ligand binds to the receptor[54]. Generally, higher magnetic field intensities and new pulse sequences improve the quality of the structures and thus increase the number of conformers and the size range of accessible proteins. An interesting recent example for a NMR study which addresses the dynamics of the system was performed on a fatty acid binding protein [55, 56]. Via ^{15}N-^{1}H experiments it was shown that the dynamics of the system dramatically decrease when a fatty acid binds. We note in passing, that in many cases the data obtained from NMR experiments can only be interpreted with the help of computational methods such as MD simulations.

The most prominent experimental method for structure derivation remains X-ray crystallography. Usually, the resulting structures are averages over space and time, and are a poor representation of the dynamic behavior. Some recent developments try to address dynamics. The main approaches are time-resolved measurements on several crystals or Laue diffraction for a fast collection of data and the better estimation of the atomic oscillations around their positions. This leads to more precise analysis of the positional displacements (B-factors). For example, anisotropic (ellipsoids) instead of isotropic (spheres) models for the atoms [57, 58]. This provides knowledge about direction and magnitude of the atomic motions leading to a good estimation of the dynamic behavior of the receptor. Large and small motions are accessible [59, 60]. Measuring several structures following a reaction path could result in a movie of the dynamics [59, 61, 62]. The limiting factor is obviously the time to measure one structure. Faster events cannot be resolved.

Other important methods which allow the collection of data about protein dynamics are fluorescence spectroscopy [63], spin label paramagnetic resonance [64], small angle X-ray scattering [65], UV/VIS absorption, and Raman spectroscopy [66].

However, in most cases experimental limitations and costs prevent the generation of extensive ensembles covering the important conformational space. Here computer simulations are well suited to extend the experimentally determined structures to generate such ensembles.

In this review, we will illustrate the application of computer simulations to molecular modeling with a focus on its practicability for drug design. We will start by shortly reviewing the principles of molecular dynamics simulations. An overview of the role in drug design is followed by a more detailed description of its application. The main topics are the generation of representative structural ensembles for further investigations, specially in virtual screening methods. The calculation of Gibbs Free Energy of binding for protein–ligand complexes has become an important application, and both the underlying principles and some interesting applications will be highlighted. Finally, prominent examples for the

direct analysis and interpretation of the observed dynamics of biomolecules and its role for their function are given.

16.2. MD SIMULATIONS

Originally developed to study fluids, the application of molecular dynamics (MD) simulations to biologically relevant receptor–ligand systems started in the late 1970s. The first simulation of a protein was described by Karplus, Gelin, and McCammon in 1977 [67]. They performed a 9 ps simulation of the bovine pancreatic trypsin inhibitor containing 886 atoms *in vacuo*. Recent MD simulations cover up to several microseconds and the system size ranges up to millions of atoms (a complete virus [68–72] or ribozyme [73]). Moreover, while early simulations were limited to an implicit representation of the solvent if any [74, 75], today explicit water molecules as solvent is standard.

In most implementations, MD simulates the evolution of a system of atoms on the time scale via the numerical solution of Newton's second law of motion:

$$m_i \frac{d^2 r_i}{dt^2} = -\nabla_i [V(r_1, r_2, ..., r_n)] \tag{16-1}$$

Here m_i corresponds to the mass of atom i, r_i are its Cartesian coordinates, t the time, and V is the potential energy of the system. V is usually evaluated by a molecular mechanics force field, but also combined quantum mechanics/molecular mechanics approaches are available and have recently been reviewed [76–80]. Starting from a configuration of particles, the differential equation is numerically integrated to move the particles to new positions and to get new velocities at these new positions.

During the simulations, at certain intervals the coordinates of the system are stored. This so called trajectory can then be analyzed to investigate the dynamic behavior and evolution of the structure. With an appropriate statistic sampling also a thermodynamic evaluation of the observed events can be performed.

Alternatively to solving the Newton's second law of motion, one can use Metropolis Monte Carlo (MC) simulations to investigate the conformational behavior of a system. Here, a new system conformation is generated by random changes of internal coordinates and then its energy is evaluated. If the new conformation is in an energetically reasonable range, according to the Metropolis criterion, it is accepted and used as starting point for the next Monte Carlo step.

A number of practical issues need to be considered when using MD simulations.

As the starting structure is derived either directly or indirectly (e.g. via homology modeling) from experiment, all the experimental limitations (size, solubility, time of measurement, crystalizibility) apply and may reduce the quality of the structure. Most often the metrics used to assess the quality of a structure (in X-ray the resolution and the R-value) are based on the whole protein and not only on the region of primary interest. Therefore, a detailed validation of the structure, preferentially by

investigating the experimental electron density is necessary. Moreover, as X-ray structures do not provide the location of hydrogen atoms, both the protonation states of amino acid side chains (most prominently histidines) and the orientation of amide side chains and water molecules has to be carefully determined by an analysis of hydrogen bond networks. Very often not a complex structure with the ligand of interest is the starting point, but an apo-structure or a holo-structure complexed with another ligand. This may induce severe artefacts as different ligands often induce conformational changes in the receptor which may be hard to predict. Here the assumption is that the system adjusts itself during the equilibration or the first part of the simulation, if the structural perturbation is not too large.

Another major drawback is the lack of standardization. Most MD programs (amber [81], CHARMm [82], Gromacs [83], NAMD [84]) use their own file formats and varying methodologies, and not all force fields (amber [81], CHARMM [82], Gromos [85], OPLS [86]) can be combined with every simulation engine. This makes a direct comparison of results difficult, although in some cases qualitatively similar results could be obtained.

In all-atom MD simulations both large-scale motions and ligand–protein interactions can be investigated. However, most simulations are limited to few nanoseconds, and it is unlikely that all important minima are sampled during this time. Several techniques can be applied to check the behavior of the system. Essential dynamics filters the important motions from the rest [87]. Cosine analysis of eigenvectors estimate real motions from noise [88, 89]. Calculated B-factors, root mean square deviation (RMSD) from the starting structure, and residue-wise correlations can show if new conformational space is reached or just sampling in close vicinity of the starting has occurred.

16.3. THE ROLE OF SIMULATIONS IN THE DRUG DISCOVERY PROCESS

There are several possible applications for molecular dynamics simulations in drug design. In the first phase of a drug discovery project, the aim is the identification of interesting compounds (often called leads) which interact with the biological target in a desired way. Often, this lead discovery is achieved by an extensive experimental testing of a huge number of compounds (several hundred thousand or even millions), called high-throughput screening (HTS). The *in silico* pendant is called 'Virtual Screening' (VS), where large databases of chemical entities are searched for potentially active compounds. This can be achieved by ligand-based methods which search for similar shape to known actives or for similar positioning of interaction features. Alternatively, structure-based approaches like docking search for a reasonable interaction pattern of the compound with the binding site. Here, one tries to establish a putative binding mode of the ligand, and so called scoring functions estimate the free energy of binding by several empirically derived terms.

In all cases, an assessment is needed to distinguish between "good" and "bad" compounds; this assessment can be either done on a single conformer or on a representative, physically well defined ensemble. Then, simulations are a means to obtain such an ensemble.

Following their discovery, the lead compounds are optimized to improve their affinity, selectivity towards unwanted off-targets, and their general physicochemical and pharmacokinetic profile. If the structure of the target is known, molecular dynamics simulations can be used during this lead optimization phase to predict the free energy of binding of the compounds to their target. An important aspect is a better understanding of the interactions, dynamics, and induced-fit effects of ligands in the binding site. Assessing the importance of water molecules, which can act as mediators of interactions, can stabilize parts of the protein structure, or can be replaced by the ligand is another fundamental task.

In the following, we discuss these aspects in more detail.

16.4. VIRTUAL SCREENING AND MD SIMULATIONS

16.4.1. Pharmacophore Modelling

A well established method of virtual screening is pharmacophore modeling. A pharmacophore is defined as "a set of structural features in a molecule that is recognized at a receptor site and is responsible for that molecule's biological activity" [90]. Typical structural features are H-bond acceptors and donors, positively or negatively charged atoms, hydrophobic groups, or aromatic rings. In the ligand-based approach a pharmacophore model is derived by using several active ligands in several conformations each. All ligand conformations are aligned in order to find a common chemical functionality arrangement. This is usually done if no receptor structure is available. The structure-based approach directly transforms a protein–ligand complex structure or an apo-structure into a pharmacophore. All interactions are mapped into a 3D model. The generated models can be used to screen compound libraries for substances fitting the interaction pattern represented by the pharmacophore.

Here, MD simulations can be used to generate representative conformational ensembles for ligand, receptor, or both. Some illustrative examples are highlighted below.

16.4.1.1. Ligand-Based Approach

A ligand-based dynamic pharmacophore study was applied to anti-complement peptide inhibitors [91]. Large conformational ensembles for four active and four inactive analogues of the 13-residue cyclic peptide compstatin were generated by 5 ns of MD simulations. For every single snapshot, a pharmacophore was generated. To discriminate between active and inactive compounds, the arrangement of the functional groups was analyzed. Distances, angles and dihedrals between the pharmacophoric points were calculated and transformed into probability distributions

for the whole ensemble. Additionally, two-dimensional angle-distance distributions were generated. Then the probability distributions for actives and inactives were used to discriminate between them. A mixed success was reported: the results were highly dependent on the selection of training set, MD protocol and the type and selection of the pharmacophore points. However, with a sufficient optimization of the training sets and the descriptors a selective classification seemed possible.

The same method was earlier applied to δ-opioid agonists and antagonists [92]. Whereas 1D-probability distributions were not able to distinguish between agonists and antagonists, the more abstract 2D probability distributions resulted in a good separation of the biological effect of the molecules. As a major advantage of this method, no conformer alignment is necessary as the MD generated conformations are directly transformed into pharmacophores from which the descriptors are deduced.

Another very successful combination of conformational sampling via MD simulation and ligand-based pharmacophore modeling was recently described by Deng et al. [93]. They report a computer-aided optimization of quinoxalinehydrazines as anti-cancer agents. Starting from a highly potent drug candidate, they used MD simulations for conformational space sampling. Then six intramolecular geometric descriptors (two intramolecular distances, three torsions, and the angle between aromatic ring planes) were used to cluster the resulting snapshots and to derive a multi-conformer ligand-based pharmacophore. Searching a database of 350.000 small organic in-house molecules with this pharmacophore yielded 968 hits. From this, a structurally diverse subset of 35 compounds was selected for testing. Seventeen of them showed a cytotoxicity of IC_{50} below 10 μM. Similar approaches were used in other cases [94, 95].

16.4.1.2. Structure-Based Approach

One of the first studies with dynamic pharmacophore models was performed by Carlson et al. [96] to find inhibitors of HIV-1 integrase. As this is a highly flexible target, the application of a flexible model is crucial. Snapshots from a MD simulation of the protein were overlaid and conserved regions were identified. The identification of the pharmacophoric points was performed with test molecules. Small molecules, which mimic certain types of interactions (e.g. methanol for H-bonding), were placed into representative snapshots from the simulations and minimized to identify important interaction points in the binding site. Using water in a validation study, many of the crystal waters with low B-factors could be reproduced. A comparison with a pharmacophore model derived with the same procedure from a single X-ray structure highlights the importance for including flexibility: None of 59 known inhibitors fit to the static model. The dynamic model on the other side identified 15 of the 20 most active (more than 50 % of inhibition at a concentration of 1 mM) compounds. A subsequent search in a public available compound database (ACD) using several sub-models of the dynamic pharmacophore identified 45 new potential HIV-1 integrase inhibitors. Eleven of these compounds were active in the biochemical assay.

In a follow-up study, the original model was modified by introducing the originally present Mg^{2+} ion [97]. The conformational sampling of the protein was performed by Metropolis Monte Carlo (MC) simulation. A set of 22 potential inhibitors were selected from a database, and testing identified five ligands with an activity (IC_{50}) below 100 μM.

The importance to use representative ensembles in pharmacophore modeling was also highlighted by experimental studies. Dihydrofolate reductase (DHFR) is an interesting test case. Differences in the flexibility in DHFR from human(HU), *Pneumocystis carinii(PC), and Candida albicans(CA)* are already obvious by comparing X-ray structures of the different species. The average crystal structure RMSD values for the C_α atoms in *CA* DHFR is low with 0.34 Å, indicating a rigid structure. In contrast, the RMSD values for *PC* and *HU* X-ray structures are larger (1.69 and 1.58 Å), showing higher flexibility [98].

Seven X-rays for the *HU* DHFR, eight for *CA* DHFR and nine for *PC* DHFR were minimized and flooded with 1,000 probe molecules to identify important interaction points. The probes were benzene and ethane for aromatic and hydrophobic interactions, and methanol for the H-bonding interactions. The resulting probe sites were clustered and transformed into pharmacophore models. For validation, three databases were prepared. One containing potent and specific inhibitors, one weaker, and one non-active molecules. Often it is argued that including flexibility into pharmacophore models may reduce their specificity. Here, however, the screening for the two flexible examples, *HU* and *PC*, perfectly pre-ranked selective-active inhibitors. The model for the relatively rigid *CA* DHFR did not reach the selectivity for high-actives, but could also separate actives from inactives. This shows that for DHFR a pharmacophore model with flexibility included can identify very potent inhibitors. Further approaches using experimentally derived structural ensembles (X-ray [98] or NMR [99]) confirm the importance for flexibility in pharmacophore modeling. Often, MD simulations were successfully used to produce representative structural ensembles for that purpose.

16.4.2. Docking

Docking is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex. In SBDD, docking is usually used to place a ligand into the binding site of a protein. The strength and quality of the interaction between ligand and protein is then assessed by a scoring function. Often, subsequent optimization further improves the interaction.

The development of docking over the years highlights the steady improvement of the method to include flexibility into the binding mode prediction: Docking evolved from rigid receptors and rigid ligands [100] over rigid receptor and flexible ligands [101–104] to the current state-of-the-art with flexible receptors and flexible ligands [105–111].

One of the first approaches to incorporate flexibility was developed by Jones et al. [112] in 1995. They used a genetic algorithm for binding pose optimization which allowed reorientation of both hydrogen atoms and lone pairs in hydrogen bonding groups of the receptor. The docking studies targeted dihydrofolate reductase, L-arabinose binding protein, and influenza neuramidase. Later, rotations of all torsional angles around side chain bonds were allowed, combined with sampling techniques such as the Metropolis algorithm [113]. Docking approaches with rotamer libraries for most side chains were reported [114, 115]. Modern concepts to include the dynamics use sampling of predefined side chains rotamers and terminal hydrogen optimization [110], protein conformers extracted from several X-ray structures [106], mean-field optimization to rotate whole parts of the ligand and the protein to avoid van der Waals overlaps [116], constrained geometric simulations [117], elasticity network models[105], or docking to structure ensembles obtained from different methods like molecular dynamics simulations, NMR, or X-ray [109, 118–120]. An interesting study showed the importance for flexibility in computational methods. Kuhn et al. [121] docked ligands in their bound conformations into the corresponding apo-structures. Only 9 of 42 thrombin and 9 of 15 glutathione-S-transferase inhibitors could be docked without steric clashes. This shows that using a representative structural ensemble, generated by MD in silico [6, 122–124], is important.

With an MD simulation a greater conformational space is sampled, which increases the probability to find additional structures, while docking into rigid X-ray structures often only returns structural classes which are close to that of the ligand of the X-ray complex. In principle, if all low-energy conformations of the receptor could be generated, and a scoring function could perfectly predict the affinity, the prediction of the correct complex geometry would be a simple optimization problem. However, in reality, no complete or even perfectly representative ensemble is accessible nor is a perfect scoring function. This is especially noticeable when the free energy differences are small or when entropic contributions are important. Therefore, docking is often used as a pre-filter to reduce the number of molecules which are investigated in a more detailed manner. The suggested binding modes can then be used as a basis for a more accurate interaction evaluation. To overcome the problems associated with docking into a single protein structure, several approaches were suggested.

One possible approach to “ensemble docking” would be to dock the ligand into representative snapshots of the MD run. Some docking methods use interaction grid maps to represent the receptor. Once calculated for several types of interactions, they can be used for every ligand, making the docking procedure very efficient. If the grid maps of representative conformations are combined, this mimics docking into several protein conformations. Several ways to combine the grids have been described: energy and a geometric weighted [125], or simple mean, grid point minima, and Boltzmann-weighted combinations [109]. The combination of several grids leads to a significant gain in speed, however, structural diversity is limited, and the non-physical intermediate grid maps sometimes lead to artificial poses.

An additional problem is that the relative energies of the protein snapshots usually are not taken into account, and even high-energy or unphysical protein conformations are not adequately penalized.

Another approach to combine several receptor structures is "in-situ cross docking" [126, 127]. Here, a simultaneous optimization of the ligand in several binding site conformations is performed, thus selecting the best binding site conformation for the ligand. This method was applied to aldose reductase. It is a very well characterized system with a very flexible behavior [128]. The binding site consists of a rigid part next to the catalytic center and a very flexible region in the so called specificity site [129–131]. The flexibility is mainly due to side chain movement and a backbone stretching. Three different binding modes with two subtypes each were observed, also indicating the flexibility of the binding site [132, 133]. As starting point the structure in complex with Tolrestat was chosen. Redocking of Tolrestat reproduced the X-ray binding pose. However, docking two structurally similar ligands lead to totally different binding modes, which were later experimentally confirmed with X-ray analysis. One ligand adopted a binding pose which was already known from other ligands, while the other ligand stabilized a completely different protein conformation which is characterized by a closed specifity site and an opening of an ionic salt bridge.

The "relaxed complex method" approach was introduced by McCammon and co-workers. Here, several receptor conformations are generated with MD simulation (for enhanced sampling, techniques like accelerated MD [134] can be applied), then small ligand libraries are automatically docked into the different binding site conformations. Finally, a scoring with e.g. MM/PBSA (see below) is performed. An interesting application used FKBP-12 (FK506-binding protein) [122–124]. Starting from an X-ray structure, an MD simulation was performed. Subsequent analysis showed that most of the protein is rigid during the simulation, however, several active site residues were found to be part of flexible regions. Due to the variations of the amino acid positions during the simulation, the surface area of the binding site varies by almost 200 $Å^2$. Subsequently, three ligands were docked with AutoDock into 200 simulation snapshots and different binding poses were identified. Estimates for free energy of binding of the various binding poses showed that the average docking free energies for the ligand can vary by several kcal/mol, corresponding to a 100–1,000-fold difference in affinity. This obviously would obfuscate any attempt to distinguish strong from weak binders. Only if the "correct" receptor conformation for every ligand can be provided for the docking algorithm, an accurate ranking of the molecules can be expected from virtual screening.

Another example is a study on HIV-1 integrase inhibitors [6]. At the time of the study only one crystal structure of moderate quality and with unresolved loops of the protein was available. Hence, drug design was complicated by only partial knowledge about the interactions of the ligands with the binding site. Further in its biological function HIV-1 integrase binds to DNA and the structure of this complex was completely unknown. In addition, it was argued that ligand positions were influenced by crystal packing effects. A 2 ns MD simulation was run on the crystal structure with ligand 5CITEP bound. After redocking of the ligand into snapshots of

the MD simulations, two binding modes were identified. One reproduced the crystal structure conformation. The new "flipped" conformation was preferred in most of the snapshots, as in this case a new subpocket (called 'trench') opened in the binding site, which could be occupied by parts of the ligand. Subsequently, a series of ligands was docked into ten structural variations of the binding site. One was the crystal structure, two snapshots with closed trench and seven with open trench. All compounds showed preferred binding to the open trench conformations over the X-ray structure. The energy gain was up to 2 kcal/mol corresponding to a 100-fold increased affinity. It was concluded that the MD simulations identified a second, more populated conformational minimum for the receptor: although the X-ray conformation is favored under crystallization conditions, both the prevalence during the simulation and the better binding of several ligands support the preference for the "new" conformation in solution. Later this was further confirmed by X-ray structures with other ligands. The discovery of the new binding subsite was the basis for the development of the first clinically approved HIV-1 integrase inhibitor raltegravir [135].

A very challenging area for computer aided drug design is the investigation of protein-protein interfaces. A very complicated conformational space has to be covered, because of the large surface areas which are involved. Also several induced-fit effects can occur (for a current review see Keskin et al. [136]).

A interesting docking/MD study was performed on 25 protein–protein complexes [137]. Heterodimer protein-protein structures were generated with docking and post-processed with MD simulations. Starting with the bound conformation from the docking, which biased the MD simulation in the required direction, the X-ray complex structures could be reproduced. Starting a simulation with the apo-structures did not lead to the conformation as observed in the crystal structure. Hence, the conformational space that would have to be examined is too large and the energy barriers too high to cover all important conformations within a MD simulation. Here, docking provided a starting structure close enough to the bound conformation that the limited sampling during the simulation was sufficient to find the experimental structure.

A general problem in docking is the energetic assessment of the binding poses obtained ('scoring'). Usually the scoring function is a linear combination of van der Waals (vdW), electrostatic, solvation, and hydrogen bonding contributions, which is derived by fitting to experimental ligand binding data. Despite significant progress in this area [138, 139], current scoring functions do not include (or mimic only crudely) polarization, a rigorous treatment of solvation effects, or entropy contributions. Often, a minimization is performed after the initial docking to remove steric and conformational problems. However, this minimization only moves into the next local minimum which might be very different from the global one. This problem is exacerbated on complex energy surfaces like those for protein–protein interactions. Here, MD simulations can be used to scan a larger part of the energy surface.

An example for this procedure is a work on 41 proteins forming protein-protein complexes [140]. Here, 22 docking poses, generated by a rigid docking approach, were used as starting points for MD simulation runs. The snapshots of the

simulations were clustered and a representative of each cluster was further analyzed. Compared to structures which were obtained by simple minimization of the docking poses, the poses relaxed by MD performed significantly better in terms of e.g. fraction of natural contacts or RMSD to experimental structure.

16.5. PREDICTION OF GIBBS FREE ENERGY OF BINDING

Understanding and quantifying the binding in biological recognition is one of the key goals of MD simulations. The Gibbs Free Energy (G) is of fundamental importance for every system, because it will evolve in the direction of the greatest G descent until equilibrium is reached ($\Delta G = 0$). Every change in the system is related to the equation:

$$\Delta G = \Delta H - T\Delta S \tag{16-2}$$

Here ΔH and ΔS are the changes in enthalpy and entropy at a given temperature T. In a biochemical ligand-receptor binding:

$$R + L <=> RL \tag{16-3}$$

R is the receptor and L the ligand, the standard binding free energy is:

$$\Delta G = \Delta G^0 - RT \ln([R][L]/[RL]) \tag{16-4}$$

At equilibrium:

$$\Delta G^0 = RT \ln\left([R]_{eq}[L]_{eq}\Big/[RL]_{eq}\right) = RT \ln(K_d) \tag{16-5}$$

K_d is the dissociation constant and the brackets indicate concentrations. ΔG^0 is the free energy at standard conditions (T = 298 K, p = 1 atm, all concentrations are 1 M). In case of ligand binding, often not K_d, but inhibition related data like K_i (inhibition constant calculated from an experiment) or IC_{50} (ligand concentration necessary for a 50% inhibition of a an enzyme) are reported. They have no well defined relation to binding affinity, but are affected by the exact experimental conditions of the biochemical assay. This often makes a direct comparison from different sources difficult [141], although in a series of ligands with IC_{50} values determined with the same method a comparison is reasonable. One way to estimate the affinity from IC_{50} is the Cheng-Prusoff relation [142]. The experimental determination of these values can be done by calorimetry, surface plasmon resonance, or fluorescence-based methods.

MD simulations have been used successfully for the prediction of free energy. They are able to generate thermodynamic ensembles and quite accurate energy evaluations. The force field energy value itself is empirically originated and therefore without any physical meaning. Due to practical reasons, usually only relative free energies are calculated, absolute free energy calculations are rare (for examples see [143–149]).

Several techniques were developed: TI (thermodynamic integration) [150] and FEP (free energy perturbation) [151] observe the change of the force field Hamiltonian during a simulation while mutating the ligand in the complex and in

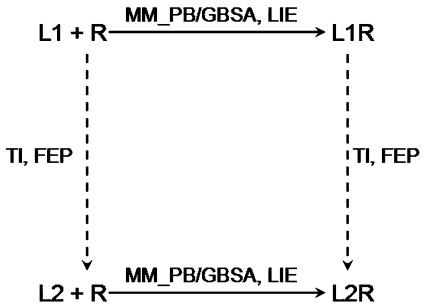


Figure 16-1. Thermodynamic cycle for two ligands (L1 and L2) binding to a receptor (R). The fundamental difference between MM/PB(GB)SA, LIE and TI, FEP is the direction. Because of practical reasons, the latter ones follow the structural transformation of ligand 1 into ligand 2 in the complex and in the unbound state, MM/PB(GB)SA try to evaluate the free energy of the binding process directly. The dashed line for Ti and FEP indicate that many unphysical intermediates have to be simulated

water. This way, one avoids the tedious direct simulation of the binding process. Other methods build a statistical model which connects the descriptors obtained from MD simulations with experimental values. This model can then be used to predict the affinities for new ligands. A prominent example is LIE (linear interaction energies) [152] which uses force field energies as descriptors. MM/PBSA (molecular mechanics / Poisson Boltzmann Surface Area) and MM/GBSA (molecular mechanics/Generalized Born Surface Area) post-process the MD generated ensemble and combine force field energies with solvation and entropy contributions. Figure 16-1 shows the thermodynamic cycle for the binding of two ligands. While MM/PB(GB)SA and LIE try to evaluate the binding energies for the ligands, TI and FEP estimate the differences between the ligands in bound and unbound state. The dashed lines for the TI/FEP direction indicate that many MD simulations with unphysical intermediates are necessary in that direction, while in the other direction only one is needed. The most important aspects of the methods and some of the important issues in their application are described in the following. For more details on the methodology and summary of applications we refer to several recent reviews [153–157].

16.5.1. MM/PB(GB)SA

The basic idea of the MM/PBSA (molecular mechanics Poisson-Boltzmann Surface Area) and MM/GBSA (molecular mechanics Generalized Born Surface Area) approaches is to split the free energy G into several contributions, which are determined separately:

$$G = <H_{MM}> - \left(<G_{pol}> + <G_{npol}>\right) - T<S> \qquad (16\text{-}6)$$

The first term is the enthalpy $<H_{MM}>$, which corresponds to the *in vacuo* force field energy. The solvation energy is split two parts: the polar solvation contribution

$< G_{pol} >$ is estimated by a numerical solution of the Poisson-Boltzmann (PB) equation [158]. Alternatively the more approximate and faster Generalized Born (GB) scheme [159–161] can be used. In the following we will concentrate on MM/PBSA as this is more standard. The non-polar solvation energy $< G_{npol} >$ is considered as proportional to the solvent accessible surface area (SASA). Finally, the entropic contribution $- T< S >$ can be obtained from normal mode analysis.

Each of the above terms results from averaging over a set of representative conformations. Thus, in principle this evaluation has to be done for ensembles of the complex, receptor, and unbound ligand. These can be generated separately [162–164] or, alternatively, extracted from one simulation of the complex [160, 165, 166]. Generating a complete or at least representative ensemble for all three constituents has proven to be difficult and time consuming. Using only one simulation for both complex and isolated protein often keeps the sampling in the most interesting part of conformational space and leads to faster converged results [167–169].

Finally, calculation of the difference between complex, protein and ligand gives the estimated free energy of binding:

$$\Delta G_{bind} = G_{complex} - G_{receptor} - G_{ligand} \tag{16-7}$$

16.5.1.1. Electrostatic Contribution ΔG_{pol}

A good representation of the electrostatic solvation is important, because the interactions of the solute with the solvent are very prominent, and the screening of the Coulomb interactions of ligand and receptor can influence the free energy dramatically [170–172]: screening of the interaction of two partial charges with the same sign (positive or negative) is favorable for binding, while the screening of opposite charges is unfavorable. Despite its long-range nature, usually the electrostatic contribution is only evaluated within a certain cut-off radius for computational efficiency. The solvation electrostatic energy is calculated by means of the Poisson-Boltzmann equation, which has to be calculated numerically, because of the arbitrary shape of the solute molecules [173, 174]. In most cases, the less complicated linearized Poisson-Boltzmann equation is sufficient, only highly charged solutes require the more exact and complicated non-linearized equation [172, 175]. Despite all approximations, the obtained results often correlate well with experiment, which has been attributed to error cancellation between bound and unbound state [176, 177]. Several software packages are available for the solution of the PB equation [81, 158, 161, 170, 178–182].

One of the main problems in calculating the solvation energy stems from the necessity to divide the system into solute and solvent. Both of them must be associated with different dielectric constants ε. For the solvent usually the value of water ($\varepsilon = 78$) is used. The dielectric constant for the solute (ε_s) is much more difficult to determine. Different values were applied in different studies [183–186] with the reasoning that not one single value is correct for all problems, because it

depends on the structure and flexibility of the receptor. Especially for the estimation of protein side chain pK_a values, ε_S was varied [187–190]. A lower ε_S means a weaker screening between solute and solvent and a higher desolvation penalty for charges that become buried upon ligand binding. Although these two effects partly compensate each other, in effect ε_S is a parameter, which would have to be fitted for every receptor using experimental values [177]. In most publications, the value of ε_S was around 4 is chosen.

An important aspect is the definition of the interface between solute and solvent, respectively between ε_S and ε_W as this has a huge impact on the results [191–193]. The solvent accessible surface area (SASA) or the molecular surface are possible borders. For the atomic radii "Born Radii" are often applied [194–198].

In most applications, the ionic strength of the solvent is not taken into account. While this is valid for protein–ligand complexes, it was shown that it is important for the highly charged DNA systems where the high number of counter ions result in high ionic strength [177, 199–201].

16.5.1.2. Non-polar Interactions ΔG_{npol}

The origin of these interactions is the tendency for apolar groups to reduce the solvent (water) exposed surface to a minimum [202, 203], which is also true upon receptor-ligand binding [204–207]. It is an entropically driven hydrophobic effect, which is not included in the -TΔS part of Eq. (16-6). The resulting effect is an energy penalty for the transfer of apolar groups from an apolar medium to the biological solvent water which can be correlated with the solvent accessible surface area [196, 208, 209]. For drug candidates, normally small organic molecules, the surface to volume ratio is very high and hence, the surface and surface effects become even more important [210]. Also a correlation between the SASA and the entropy of hydration was shown [211]. The correlation coefficient applied in the MM/PBSA calculation is usually called microscopic surface tension γ [199, 205, 212]; usually values between 20 and 33 cal mol^{-1} $Å^{-2}$ are used [208, 209].

16.5.1.3. Entropic Contributions

Calorimetric experimental studies have demonstrated the crucial role of the entropic contribution to the free energy of binding. It can be unfavorable as well as favorable [213–215]. Unfavorable contributions arise from the loss of degrees of freedom. Besides the translational and rotational degrees of freedom [169, 216–218], also intramolecular motions in the ligand can be lost [218–220]. Negative contributions leading to a gain in free energy may arise from the six new vibrations in the formed complex [201, 216] as well as from changes in the overall vibrational spectrum, changes in the surrounding water structure, or water molecules which are liberated upon ligand binding. The experimental determination of the size of the entropic contributions is difficult, theoretical estimations from rigid body calculations give values of 2–18 kcal mol^{-1} at 298 K [217, 218, 221]. In many publications the entropic part is neglected, resulting in much too favorable ΔG values. One reason for this

approximation is the assumption that ligands with similar mass should have similar entropic contributions to the binding energy [205, 222–225]. However, the validity of this approximation has been questioned [201, 219]. To estimate the entropy out of ensembles from MD simulation a normal mode analysis [160, 162, 226] can be performed, however at a high computational cost, which may be an additional reason that this part is often neglected [168, 227]. Often only a limited sphere around the ligand is investigated [160, 162, 226], but this can lead to unphysical, artificial vibrational modes associated with errors of ± 5 kcal mol^{-1} [162]. As an alternative to normal mode analysis, quasi-harmonic analysis has been suggested [228].

16.5.1.4. Applications

A large number of reviews have summarized the many applications of the MM/PBSA or MM/GBSA method to the estimation of binding energies [153–157]. We mention the application of MM/PBSA to single conformations e.g. to postprocess docking poses [223, 229–232], however, interaction energies are very sensitive to conformational changes [169, 190, 233, 234]. The electrostatic solvation energy from the PB equation, varies significantly with conformational changes [234], as does the Coulomb energy from the force field. However, these effects are anticorrelated [169, 222, 235], leading to smaller fluctuations in the overall electrostatic energy than for the individual components. Thus, a single structure hardly represents the energetics of a system correctly. In praxis, quite short simulation times are sufficient to cover the fluctuation. Most of them are 0.5 ns or shorter [164, 200, 236, 237], as longer sampling has not improved the results [168].

A number of studies applied continuum MD to generate a representative ensemble [235, 238, 239]. However, it is generally accepted that explicit water is important. Often, individual water molecules have important roles: they screen electrostatic interactions, mediate hydrogen bonding, or are integral part of complexes. Such water molecules need to be included into the post-MD MM/PBSA calculation explicitly [165]. For the other waters continuum solvent models include all effects for the solvation energy estimation, presumably more completely than a calculation with explicit solvent [239–242].

16.5.2. LIE Approach

In the LIE (linear interaction energy) approach, the free energy of binding is calculated as a linear combination of the electrostatic and van-der-Waals interaction energies [152, 243, 244]. LIE has its basics in the linear response theory which states that the environment reacts linearly to the electrostatics of a small molecule. The basic LIE equation is:

$$\begin{aligned}\Delta G_{\text{bind}} = {} & \alpha(< V_{\text{vdW}} >_{\text{bound}} - \Delta < V_{\text{vdW}} >_{\text{free}}) \\ & + \beta(\Delta < V_{\text{elec}} >_{\text{bound}} - \Delta < V_{\text{elec}} >_{\text{free}}) + \gamma \end{aligned} \tag{16-8}$$

The brackets indicate that ensemble averages (the mean value over the representative snapshots) are used, the subscripts bound and free refer to values obtained for the ligand in the receptor and in solution, respectively. Although the parameter for the electrostatic interaction (β) is 0.5 by theory [152, 244–246], usually the values for α and β are obtained by fitting to the experimental free binding energies for a training data set. The necessity for such a training set of several ligands with a known free energy of binding is the main differences of LIE to other free energy methods. In addition, LIE is a ligand-centered method as the only interactions taken into account are those of the ligand with its environment. Intramolecular energies are neglected for ligand and receptor as is the desolvation energy of the receptor. These contributions usually are integrated into the parameters α and β [243, 247], but also can be treated explicitly [248, 249].

A weakness of LIE is that the receptor is investigated only in the complexed form, which means that no protein reorganization upon binding is taken into account, which has been suggested to lead to poor performance in some cases [250]. The main difference of LIE to most empirical scoring functions is the ensemble averaging and the inclusion of the interactions with the surrounding solvent for the unbound ligand.

Similar to MM/PBSA, the method just needs the endpoints of the binding process for the free energy estimation, the ligand in solution and the complex in solution, for which thermodynamic representative ensembles have to be generated [152, 243, 244]. It was also proposed to reduce this to a single simulation of the complex state [251]. As every ligand is treated separately, a broad structural diversity in the ligands is possible [252].

One of the critical points in LIE applications is the treatment of the long-ranging electrostatic interaction [243], as electrostatic contributions are a prominent part in Eq. (16-8). As the ligand is simulated both in water and in the receptor, the net charge in the two simulations often is different. This is even more problematic, if only electrostatic interactions within a certain cut-off range are taken into account, as then the net charge for the protein simulation may depend on the chosen cut-off [242, 253, 254]. To address these issues, one can selectively neutralize the receptor and apply a continuum solvent model beyond the hydration sphere for long-range electrostatics [242, 250, 252–256]. However, even without these improvements reasonable results were reported [245, 257–259]. Counter ions for neutralization often lead to a slow convergence of the results [254]. One publication proposed to replace the electrostatic energy by a simple H-bond [260].

In an early study on endothiapepsin inhibitors [152], the optimization for α gave a value of 0.16. This parameter set ($\alpha = 0.16$, $\beta = 0.5$, $\gamma = 0$) then gave reasonable results in several additional test cases [242, 253, 261].

Later, an extensive validation study on 18 protein–ligand complexes [246] showed that the optimal value for β is not always 0.5, but largely depends on the nature of the ligand. While charged ligands should have a β of 0.5, for neutral ligands it depends on the number of hydroxyl groups: OH-groups mimic the

interactions with water, thus damping the electrostatic interaction, and consequently, the optimal β is reduced. This new parameter set was tested on eight different thrombin inhibitors [252] and performed very good with a mean unsigned binding energy error of 0.68 kcal mol^{-1}. To achieve this high accuracy, γ was adjusted to a value of −2.916 kcal mol^{-1}. The same parameters gave reasonable results for dihydrofolate reductase [262], but performed poorly, with a maximum unsigned error of 4.39 kcal mol^{-1} and an average unsigned error of 1.24 kcal mol^{-1} on FKBP-12 [258]. An optimization of the parameters improved the maximum unsigned error to 1.12 kcal mol^{-1} and the average unsigned error to 0.47 kcal mol^{-1}. From these studies it emerged that no generally applicable model can be found due to the dependence on the force fields, polarity, flexibility and the structure of the system.

Despite the problems associated with LIE applications, many studies [227, 245, 249, 253, 258, 259, 263–266] were published where the errors between calculated and experimental binding energies were about 1 kcal mol^{-1}.

Further extensions of Eq. (16-8) introduced the change in the solvent accessible surface area (SASA) upon binding as an additional descriptor [245, 249, 258, 267, 268]. However, as the SASA is highly correlated with $V_{vdW\_Lig,}$ the influence on the results was small [243, 244]. The application of continuum solvent models (GB on BACE and HIV-1 protease [239] and PB on HIV-1 protease and reverse transcriptrase [230, 251]) resulted in similar errors (mean unsigned errors from 0.9 to 1.1 kcal mol^{-1}) as with explicit waters. Johnson & Johnson Pharmaceutical Research and Development applied it to β-secretase with a mean unsigned error of 0.8 kcal mol^{-1} [254].

Another extension of the LIE approach is the calculation of the binding energy as a linear combination of physico-chemical descriptors ε which were derived from MD trajectories [230, 251, 263, 265]:

$$\Delta G_{bind} = \Sigma c_i \varepsilon_i + \gamma \tag{16-9}$$

This so called "extended linear response" (ELR) method is a QSAR-like (quantitative structure-activity relationship) approach. The descriptors include H-bond count, ligand SASA (total, apolar, aromatic, polar), receptor SASA, presence or absence of special functional groups, or the number of rotatable bonds. In contrast to many QSAR applications, only physically meaningful parameters which are related to the binding process are used [248, 255, 269]. However, also typical QSAR problems like cross-correlations and the predictivity of the models have to be considered [248, 249, 269]. Chemometric methods like partial least squares (PLS) can reduce the dimensionality of the model, but this is often connected with a loss in interpretability [249]. The performance of the method is similar to the classical LIE approach, correlation coefficients of maximally 0.8 and errors around 1 kcal mol^{-1} were published. Although the results seem to be more stable, this is counterbalanced by the loss in interpretability.

16.5.3. FEP/TI

Free energy perturbation (FEP) and Thermodynamic Integration (TI) are the most rigorous approaches for free energy prediction (for detailed reviews see [153–157]). They are based on Kirkwood's [150] and Zwanzig's [151] work on continuous coupling parameters for fluid systems. The free energy change of a system from an initial state to an end state is calculated. As in biological systems the two states are very dissimilar in terms of energy and conformation, this change cannot be performed in a single step. To ensure a sufficient and converged sampling for the structural transformation, several simulations with unphysical intermediates between the two endpoints are necessary to obtain a sufficient overlap between the individual states to enable a reliable evaluation of the free energy changes. With every step the properties of the starting point are slowly altered towards the end point of the mutation. Introducing a coupling parameter λ, the properties (χ) of every intermediate (i) in the simplest case can be defined as a combination of the properties of the initial (1) and the end (2) state:

$$\chi_{\mathrm{i}} = \lambda_{\mathrm{i}}\chi_2 + (1 - \lambda_{\mathrm{i}})\chi_1 \tag{16-10}$$

For instance, a value of 0.2 for λ means that the system has 80% of the properties of the starting point and 20% of the end point of the mutation. The number of λ values depends on the complexity of the mutation, the larger the differences between start and end point, the more intermediate states are necessary.

For every λ value, four simulations [270] (start and end point both for the uncomplexed as well as for the complexed ligand) have to be performed, making FEP and TI computationally very demanding, despite some techniques to reduce the cost [271].

The difference between FEP and TI is the free energy evaluation. In FEP it is calculated as the sum of the individual free energy differences of all steps:

$$\Delta G = \sum_{i=0}^{k-1} \Delta G_{\lambda_i \to \lambda_{i+1}} \tag{16-11}$$

with:

$$\Delta G_{\lambda_i \to \lambda_{i+1}} = -kT \ln \left\langle e^{-\frac{V(\lambda_{i+1}) - V(\lambda_i)}{kT}} \right\rangle_i \tag{16-12}$$

The brackets indicate the ensemble means for every intermediate state i, k is the Boltzmann constant and T the temperature. Contrary to this summation over finite differences, thermodynamic integration applies an integration over λ:

$$\Delta G = \int_0^1 \left\langle \frac{\partial V(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \tag{16-13}$$

This integral formulation leads to an evaluation of the current ensemble, while in FEP a neighboring, incorrect ensemble is considered.

As free energy is a state function, FEP or TI has no defined direction. All states, initial, start, and intermediates, are path-independent. Thus, the mutation from state (1) to (2) would lead to the same result (neglecting the sign) as the mutation from state (2) to state (1). In praxis, however, insufficient sampling leads to differences in forward and backward mutation. Fortunately, this provides the possibility to validate the quality of the data and make an estimation of the minimal error [272], and often the final free energy is calculated as the mean of both directions:

$$\Delta G(\lambda_1 \rightarrow \lambda_2) = (\Delta G(\lambda_1 \rightarrow \lambda_2) - \Delta G(\lambda_2 \rightarrow \lambda_1))/2 \qquad (16\text{-}14)$$

The first FEP calculation for a molecular system was published in 1985 [270]. The hydration energies for methanol and ethane were calculated with a direct mutation of the initial state into the end state.

One of the most prominent problems hampering these rigorous free energy calculation methods is insufficient sampling. To address that, methods to enhance sampling were recently introduced. Lawrenz et al. [273] applied an independent trajectory thermodynamic integration study to the H5N1 avian influenza virus neuraminidase inhibitor Peramivir. It had been shown earlier that for the sampling of an equilibrium state of a system a combination of multiple independent simulations is possible [274–276]. When this approach was applied to the intermediate states of a free energy calculation, where an extensive sampling is necessary [277], the combination of several independent TI calculations resulted in higher accuracy. In addition, the simulations gave insight into key interactions of the inhibitor with the viral enzyme and important contributions to the binding could be identified. Again, the tremendous importance for extensive sampling could be shown.

Another publication by Fajer et al. [278] applied a combination of replica exchange and accelerated MD for improved sampling in a TI calculation. Accelerated MD employs an extra potential, which is added to the force field Hamiltonian which modifies the energy landscape to overcome usual energy thresholds [134, 279]. Afterwards a reweighing of the Hamiltonian is performed so that the original potential can be reproduced. In TI calculations, which are very sensitive to correct statistical ensemble averages [280], one has to take care that the simulation is not too much accelerated as then it is difficult to achieve an accurate reweighting [281]. Additionally, a replica-exchange method was applied. The change of Hamiltonians of highly accelerated and less accelerated simulations leads to a better reweighing while keeping the enhanced sampling of accelerated MD.

Because of the high computational costs of TI and FEP, usually they are applied in the lead optimization phase. An example which somewhat mimics a in a drug discovery project, aims to optimize inhibitors for the human neutrophil elastase [282]. The aim was to improve the binding affinity of cinnamic acid ester derivatives. No crystal structure was available, but a proposed binding mode for bornyl caffeate was available as starting point. Starting in μM activity range, TI

calculations were performed for a variety of potentially active bornyl caffeate derivatives, but only for one ((3,4,5-trihydroxy)-cinnamate) a higher affinity was predicted. Subsequent biological testing confirmed the affinity gain for the new compound, although only qualitatively but not quantitatively. Nevertheless, the prediction of free energy differences for different derivatives of a scaffold and a resulting prioritization of compounds to be synthesized and tested, seems to be the perfect application for FEP and TI in a drug development process.

16.6. ELUCIDATION OF STRUCTURAL FUNCTION USING SIMULATIONS

In addition to the generation of structural representative ensembles and direct calculation of thermodynamic observables, MD simulations provide direct insight into structural and dynamical features of a biological system. Conclusions can directly be drawn on the molecular level from changes in the dynamics of the atomic motions. Two illustrative examples are presented in the following. The first considers G-protein coupled receptors (GPCRs) which are experimentally very difficult to access and show the complementarity of simulations and experimental data. As a second example, the investigation of important water molecules in biological systems with simulations is presented.

16.6.1. GPCRs

GPCRs have an immense importance for drug design, because more than 50% of all sold drugs target GPCRs [283]. They are membrane proteins with seven transmembrane helices which are connected with flexible loops of differing lengths. Their function is to promote extra-cellular signals (e.g. presence of a ligand) through the cell membrane into the intracellular region, and starting a cascade of secondary processes. The structural exploration of GPCRs is difficult. The main reason is the problem to obtain sufficient amounts of protein material. Moreover, even if enough protein is available, GPCRs are very difficult to crystallize or to solve. As a consequence, only few high-resolution structures are available, very contrary to their high importance. The first X-ray structure of bovine rhodopsin was solved in 2000 [284, 285]. The first human GPCR structure of the β_2-adrenogen receptor in complex with the partial antagonist Carazolol was resolved in 2007 with a resolution of 2.4 Å [285–287]. In 2008, the structures of β_1- and β_2-adrenogen receptor and an A_{2A}-adenosine receptor were solved [288].

Biophysical investigations showed that GPCRs exist in several conformations, and that the conformation that is stabilized upon ligand binding does already exist before [285, 289, 290]. The influence of ligands, proteins, the lipid membrane, and mutations crucially change the conformational behavior [285, 291].

Both the multitude of possible conformations and the lack of structural information severely hamper the understanding of the mechanism GPRC activation and

signal transduction. Here, both supramolecular modeling and MD simulation were performed to match the limited experimental data.

Fanelli et al. [285, 292] developed a protocol of comparative or ab initio modeling, ligand-protein and protein-protein docking followed by MD simulations and analysis to get more insight into the function of GPCRs. Very striking are the differences between free and bound state. Reducing the degrees of freedom with implicit membrane models and intrahelical restraints, the most prominent motions and structural changes were identified and investigated. Representative structures for ligand and mutation induced activation and receptor induced guanosine diphosphate release gave important input for the understanding of GPCR function. Simulations of GPCRs with activating and inactivating mutations or activating and inactivating ligand binding in the G protein binding part of the receptor domain of glycoprotein subfamily suggest that those indicate a communication with a distal cytosolic receptor domain [285, 292–294] situated near the conserved "E/DRY" motif. The communication induces a perturbation of the interaction pattern of the E/DRY arginine. Consequently, a higher solvent accessible surface area (SASA) than in the inactive state can be observed [285, 292, 293, 295]. This was shown with serotonine receptor [296], melanine-concentrating hormone receptor [297], and thromboxane A2 receptor. In the case of the TXA2 receptor, the higher SASA leads to more favorable binding of the C-terminal Gq_{α} region of the G protein at the intracellular loop 3. This interaction between G protein and receptor favors the formation of a GDP exit route between parts of the helices and the GTPase-like domain [295]. To our knowledge, this is the first example for comparative MD analysis to explain the communication between agonist binding and GDP release over a distance of about 70 Å.

This communication is bilateral, meaning that also a perturbation of the cytosolic domain induces structural changes in the extra-cellular receptor as shown for the κ opioid receptor [298]. Similarly, mutations on different sites of the receptor helix bundle may induce a constitutive receptor activation [285, 292–294].

Another interesting and intensively studied structural feature of GPCRs is the so called "ionic lock", formed by contacts between helices 3 and 6. It was suggested to form a molecular switch for receptor activation [299], closed in the inactive state, its opening should be tightly connected with the activation process. However, in the crystal structure of the inactive β_2-adrenergic receptor the ionic lock was surprisingly found broken, raising questions about the role and conformational behavior of this structural feature.

MD simulations of several variations of the protein (mutations; with or without ligand) resulted in over 10 μs of trajectory [300]. In all simulations of the wild type protein, the ionic lock closed within the first several hundred nanoseconds forming the expected structure. However, the ionic lock did not stay closed, but showed a dynamic breaking and re-forming behavior with a preference for the closed conformation. A previous MD simulation study of β_2-adrenergic receptor used a homology model based on the rhodopsin crystal structure [301]. Starting with a closed lock, during the 10 ns of simulation, one opening event was observed which reproduced

the crystal structure published later. It was shown that the inactive state of the β_2-adrenergic receptor does not require a closed ionic lock, as it was suggested before, but a clear preference can be observed.

16.6.2. Water

Water molecules play important roles for folding, stability, and function of biomolecules [302–305]. MD simulations with explicit solvent are the most accurate method to investigate the dynamic behavior of water molecules. In typical simulation solvent boxes, many thousand water molecules are integrated, and often it needs suitable statistical approaches to identify and extract the important and interesting water molecules from the enormous amount of data from a simulation trajectory.

For water molecules in proteins several quantities can be determined. Similar to water molecules in crystal structures, calculated B-factors are a measure for the mobility – or better the positional fluctuations – of the water molecules. A low B- factor usually indicates a tightly bound water molecule [306], which should be considered during the analysis of protein-ligand interactions. We note, that in addition, the B factors of surrounding atoms or residues can crucially depend on a water molecule as well.

Besides the B-factor, the time a individual water molecule stays at a defined position relative to the protein is another very important metric. A high residence time usually indicates strong binding, which indicates whether a water molecule could be replaced by a functional group in the ligand, or if it is energetically more favorable to include the water as an H-bond mediator in the binding mode of a potential drug molecule [307]. The solvent accessible surface area has been shown to be highly correlated to the residence time of a water molecule [308], another study noted that a small SASA indicates a weakly bound water molecule [309].

For the calculation of the residence time the motions of ligand and protein (small scale side chain motions and larger domain motions) have to be considered [310]. Often, the superposition of individual protein snapshots can cause problems [311]. An easy and intuitive way to determine residence times is the calculation of the distance between a protein atom and the water molecule [312]. If the distance is lower than a cut-off the water is still on the “same” place. This introduces a bias for a certain region. Another approach is define a grid and use the grid points to calculate the distances to water molecules [306].

Radial density functions (RDF) for the characterization of surrounding water molecules have been applied, however, this destroys the knowledge about individual water molecules. In addition, hardly any protein has a spherical shape which significantly complicates the interpretation of such RDFs.

So far, we have concentrated on the dynamic behavior of water molecules at interesting protein sites. Especially in drug design the energetics of water in the binding site are highly interesting. Tightly bound waters usually occupy hydrophilic voids and form several hydrogen bonds, while weakly bound water molecules often

are placed in hydrophobic cavities with hardly any H-bonding partner. Barillari et al. [313] used the double decoupling method, and replica exchange thermodynamic integration in Monte Carlo simulations to estimate the free energy of binding of a conserved water molecule as -6.2 kcal mol^{-1}. The displacement of a water molecule by a ligand molecule delivers -3.7 kcal mol^{-1}.

Another method to estimate the free energy contribution of water molecules is described by Abel et al. [314] for the blood coagulation factor Xa (fXa). Here, a 10 ns MD simulation of a fixed protein in explicit solvent is performed to analyze the dynamics of the solvent in the binding site. Afterwards the water molecules are clustered to obtain 43 "hydration spheres" distributed in the binding cavity. Based on the inhomogeneous solvation theory [315], the enthalpy and entropy of binding for each of these sites was calculated. This way, for each hydration site it was determined if it was enthalpically or entropically favorable or unfavorable when a water is displaced by a drug molecule at that site. With this information, 31 pairs of congeneric fXa ligands were compared, and for each pair the relative binding energy was assumed to stem exclusively from the displacement of water molecules in the binding site. Inserting the values obtained from the analysis above, an overall correlation of $R^2 = 0.81$ between the predicted and experimentally determined affinities could be achieved. However, expanding this approach from the evaluation of pairs to a series of ligands was less successful.

16.7. PERSPECTIVE

MD simulations are often titled as *in silico* experiments and they should be treated as such. Because of the non-deterministic behavior a reasonable statistic has to be applied. Although the asymptote of reality of the Bezdek curve (Figure 16-2) is reached [316], we still expect an increasing importance for drug design. The main reason for this is the steady increase in computational power, which still follows the Moore's law [317]. The physical basis of MD simulations is very solid and theoretically the results should be accurate, if a sampling of all physically relevant states is possible provided the force field adequately describes the system of interest. A main problem is still to reach areas of conformational space, which are only accessible by overcoming entropic and enthalpic barriers, as at room temperature these states are very unlikely to be sampled. Here the increased computer power and new sampling techniques like replica exchange MD (REMD) [318–321] are expected to dramatically improve the current status. Also the steady development of software speed and special hardware, like the application of graphical chips or special machines like Anton [322], a specially designed supercomputer with a CPU arrangement comparable to the parallelization algorithm, boosts the calculations.

Some of the current limitations result from force field inaccuracies, e.g. the missing description of polarization in most force fields today, and quantum mechanically based interactions such as electron correlation are also often not adequately represented in the force field energies. A major problem is still the

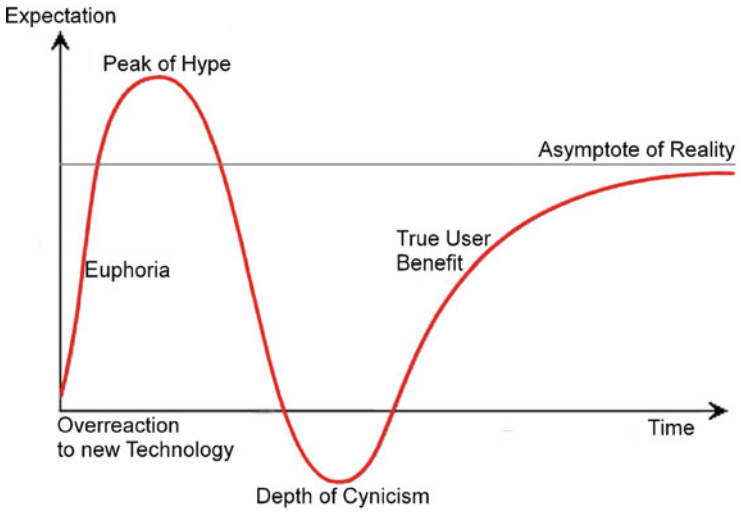


Figure 16-2. The Bezdek curve reflects the development of a newly introduced technique. After an immense euphoria in the beginning, the peak of hype is reached. The limitations of a method start to appear and lead to a descent in expectations. After a fall to an unrealistic cynicism, slowly the true applicability becomes clearer and finally, reaches a realistic status

derivation of force field terms for functional groups present in common drug-like molecules. While protein and DNA force fields have reached a high level of accuracy, the description of small molecules containing non-amino acid-like functionality is still far from optimal. The same is true for simulations in non-aqueous environments such as cell membranes for membrane-bound proteins.

MD simulations are currently the best method to describe and to include flexibility into the molecular models which are used in the drug design process, and many applications have shown that flexibility can be imperative to adequately describe the biological target. In principle, MD calculations are feasible in the time lines of a drug development project, and as the few examples we have presented above have shown, they may offer interesting insight which is difficult or impossible to gain otherwise. However, it is often difficult to assess scope and – more importantly – limitations of MD simulations, especially for newer and more advanced methods. Only rarely unsuccessful applications are published, usually only success stories make it into scientific journals which makes it difficult to obtain a clear picture about the limits of a new methodology. In addition, compared to other tools in computer aided molecular design, e.g. docking programs, the user interfaces of most MD programs are not in a state that they allow an easy setup or analysis of the simulations. Recently, first steps along this way were reported [323, 324].

Nevertheless, we are confident that the steady experimental feedback, be it biological testing or structure determination, will be both the cause and motivation to further improve MD methods and to guarantee reliable models of the biological system.

REFERENCES

1. Greer J, Erickson JW, Baldwin JJ et al (2002) J Med Chem 37:1035–1054
2. Jorgensen WL (2004) Science 303:1813–1818
3. Chand P, Kotian PL, Dehghani A et al (2001) J Med Chem 44:4379–4392
4. Gubernator K, Böhm H-J (1998) structure-based ligand design. Wiley-VCH Verlag GmbH, Weinheim
5. Hazuda DJ, Anthony NJ, Gomez RP et al (2004) Proc Natl Acad Sci USA 101:11233–11238
6. Schames JR, Henchman RH, Siegel JS et al (2004) J Med Chem 47:1879–1881
7. Liverton NJ, Holloway MK, McCauley JA et al (2008) J Am Chem Soc 130:4607–4609
8. Thomson JA, Perni RB (2006) Curr Opin Drug Disc Dev 9:606–617
9. Stauffer SR, Stanton MG, Gregro AR et al (2007) Bioorg Med Chem Lett 17:1788–1792
10. Hop CECA, Cole MJ, Davidson RE et al (2008) Curr Drug Metab 9:847–853
11. Jamei M, Marciniak S, Feng KR et al (2009) Expert Opin Drug Metab Toxicol 5:211–223
12. Nogueira RC, Oliveira-Costa JF, De Sa MS et al (2009) Curr Drug Targets 10:291–298
13. Selick HE, Beresford AP, Tarbit MH (2002) Drug Disc Today 7:109–116
14. van de Waterbeemd H, Gifford E (2003) Nat Drug Disc 2:192–204
15. Wang J, Hou T (2009) Annu Rep Comput Chem 5:101–103
16. Bolton V, Perutz M (1970) Nature 228:551–552
17. Fermi G, Perutz M, Shaanan B et al (1984) J Mol Biol 175:159–174
18. Kendrew J, Bodo GR, Dintzis H et al (1958) Nature 181:662–666
19. Watson H (1969) Prog Stereochem 4:299–333
20. Safo MK, Abraham DJ (2005) Biochemistry 44:8347–8359
21. Wolf-Watz M, Thai V, Henzler-Wildman K et al (2004) Nat Struct Mol Biol 11:945–949
22. Vonrhein C, Schlauderer GJ, Schulz GE (2009) Structure 3:483–490
23. Henzler-Wildman KA, Lei M, Thai V et al (2007) Nature 450:913–916
24. Hu X, Lazar MA (2000) Trends Endocrinol Metab 11:6–10
25. Hellal-Levy C, Fagart J, Souque A et al (2000) Mol Endocrinol 14:1210–1221
26. Henzler-Wildman K, Kern D (2007) Nature 450:964–972
27. Bahar I, Chennubhotla C, Tobi D (2007) Curr Opin Struct Biol 17:633–640
28. Cozzini P, Kellogg GE, Spyrakis F et al (2008) J Med Chem 51:6237–6255
29. Lee GM, Craik CS (2009) Science 324:213–215
30. Tsai CJ, Ma B, Nussinov R (1999) Proc Natl Acad Sci USA 96:9970–9972
31. Koshland DE, Nemethy G, Filmer D (2002) Biochemistry 5:365–385
32. Monod J, Wyman J, Changeux JP (1965) J Mol Biol 12:88–118
33. Weber G (2002) Biochemistry 11:864–878
34. Tsai CJ, Kumar S, Ma B et al (1999) Protein Sci 8:1181–1190
35. Berman HM, Westbrook J, Feng Z et al (2000) Nucl Acid Res 28:235–242
36. Tompa P (2002) Trends Biochem Sci 27:527–533
37. Zheng Q, Kyle DJ (1997) Drug Disc Today 2:229–234
38. Walters WP, Stahl MT, Murcko MA (1998) Drug Disc Today 3:160–178
39. Ma B, Kumar S, Tsai CJ et al (1999) Protein Eng 12:713–720
40. Freire E (1998) In: Richards FM (ed) Advances in protein chemistry – linkage thermodynamics of macromolecular interactions. Academic, San Diego, CA
41. Onuchic JN (1997) Proc Natl Acad Sci USA 94:7129–7131
42. Shoemaker B, Wang J, Wolynes P (1997) Proc Natl Acad Sci USA 94:777–782
43. Rejto PA, Freer ST (1996) Prog Biophys Mol Biol 66:167–196

44. Brem R, Dill KA (1999) Protein Sci 8:1134–1143
45. Boehr DD, Dyson HJ, Wright PE (2006) Chem Rev 106:3055–3079
46. Wüthrich K (2003) J Biomol NMR 27:13–39
47. Mulder FAA, Mittermaier A, Hon B et al (2001) Nat Struct Mol Biol 8:932–935
48. Ishima R, Torchia DA (2000) Nat Struct Mol Biol 7:740–743
49. Eisenmesser EZ, Millet O, Labeikovsky W et al (2005) Nature 438:117–121
50. Horst R, Wider G, Fiaux J et al (2006) Proc Natl Acad Sci USA 103:15445–15450
51. Grishaev A, Tugarinov V, Kay L et al (2008) J Biomol NMR 40:95–106
52. Sprangers R, Velyvis A, Kay LE (2007) Nat Meth 4:697–703
53. Liang B, Tamm LK (2007) Proc Natl Acad Sci USA 104:16140–16145
54. Cioffi M, Hunter CA, Packer MJ et al (2008) J Med Chem 51:2512–2517
55. Ramos CRR, Figueredo RCR, Pertinhez TA et al (2003) J Biol Chem 278:12745–12751
56. Pertinhez T, Sforça M, Alves A et al (2004) J Biomol NMR 29:553–554
57. Vitkup D, Ringe D, Karplus M et al (2002) Proteins 46:345–354
58. Merritt E (1999) Acta Cryst Sect D 55:1109–1117
59. Schmidt A, Lamzin V (2007) Cell Mol Life Sci 64:1959–1969
60. Bossi RT, Aliverti A, Raimondi D et al (2002) Biochemistry 41:8807–8818
61. Ringe D, Petsko GA (1985) Prog Biophys Mol Biol 45:197–235
62. Bourgeois D, Royant A (2005) Curr Opin Struct Biol 15:538–547
63. Somogyi B, Lakos Z, Szarka A et al (2000) J Photochem Photobiol B: Biol 59:26–32
64. Hubbell WL, Cafiso DS, Altenbach C (2000) Nat Struct Mol Biol 7:735–739
65. Lipfert J, Doniach S (2007) Annu Rev Biophys Biomol Struct 36:307–327
66. Katona G, Carpentier P, Niviere V et al (2007) Science 316:449–453
67. McCammon JA, Gelin BR, Karplus M (1977) Nature 267:585–590
68. Freddolino PL, Liu F, Gruebele M et al (2008) Biophys J 94:L75–L77
69. Klepeis JL, Lindorff-Larsen K, Dror RO et al (2009) Curr Opin Struct Biol 19:120–127
70. Maragakis P, Lindorff-Larsen K, Eastwood MP et al (2008) J Phys Chem B 112:6155–6158
71. Perez A, Luque FJ, Orozco M (2007) J Am Chem Soc 129:14739–14745
72. Romo TD, Grossfield A, Pitman MC et al (2009) Biophys J 96:340a–340a
73. Vaiana AC, Sanbonmatsu KY (2009) J Mol Biol 386:648–661
74. McCammon JA, Wolynes PG, Karplus M (2002) Biochemistry 18:927–942
75. van Gunsteren WF, Karplus M (1981) Nature 293:677–678
76. Lin H, Truhlar DG (2007) Theor Chem Acc 117:185–199
77. Gleeson MP, Gleeson D (2009) J Chem Inf Model 49:1437–1448
78. Senn HM, Thiel W (2009) Angew Chem Int Ed 48:1198–1229
79. Bakowies D, Thiel W (1996) J Phys Chem 100:10580–10594
80. Gogonea V, Suβrez D, Avd V et al (2001) Curr Opin Struct Biol 11:217–223
81. Case DA, Darden TE, Cheatham TE et al (2008) AMBER 10. University of California, San Francisco, CA
82. Brooks BR, Bruccoleri RE, Olafson BD et al (1983) J Comput Chem 4:187–217
83. Van Der Spoel D, Lindahl E, Hess B et al (2005) J Comput Chem 26:1701–1718
84. Phillips JC, Braun R, Wang W et al (2005) J Comput Chem 26:1781–1802
85. Berendsen HJC, Postma J, van Gunsteren W et al (1981) In: Pullman B (ed) Intermolecular Forces. D.Reidel, Dordrecht, The Nederlands
86. Jorgensen WL, Maxwell DS, Tirado-Rives J (1996) J Am Chem Soc 118:11225–11236
87. Mongan J (2004) J Comput Aided Mol Des 18:433–436
88. Hess B (2000) Phys Rev E 62:8438
89. Hess B (2002) Phys Rev E 65:031910

90. Gund P (1977) Prog Mol Subcell Biol 5:117–143
91. Mallik B, Morikis D (2005) J Am Chem Soc 127:10967–10976
92. Bernard D, Coop A, MacKerell AD (2003) J Am Chem Soc 125:3101–3107
93. Deng J, Taheri L, Grande F et al (2008) Chem Med Chem 3:1677–1686
94. Meagher KL, Carlson HA (2004) J Am Chem Soc 126:13276–13281
95. Meagher KL, Lerner MG, Carlson HA (2006) J Med Chem 49:3478–3484
96. Carlson HA, Masukawa KM, Rubins K et al (2000) J Med Chem 43:2100–2114
97. Deng J, Sanchez T, Neamati N et al (2006) J Med Chem 49:1684–1692
98. Bowman AL, Lerner MG, Carlson HA (2007) J Am Chem Soc 129:3634–3640
99. Damm KL, Carlson HA (2007) J Am Chem Soc 129:8225–8235
100. Kuntz ID, Blaney JM, Oatley SJ et al (1982) J Mol Biol 161:269–288
101. Goodsell DS, Olson AJ (1990) Proteins 8:195–202
102. Morris GM, Goodsell DS, Huey R et al (1996) J Comput Aided Mol Des 10:293–304
103. Morris GM, Goodsell DS, Halliday RS et al (1998) J Comput Chem 19:1639–1662
104. Rarey M, Kramer B, Lengauer T et al (1996) J Mol Biol 261:470–489
105. Cavasotto CN, Abagyan RA (2004) J Mol Biol 337:209–225
106. Claussen H, Buning C, Rarey M et al (2001) J Mol Biol 308:377–395
107. Huey R, Morris GM, Olson AJ et al (2007) J Comput Chem 28:1145–1152
108. Jain A (2007) J Comput Aided Mol Des 21:281–306
109. Österberg F, Morris GM, Sanner MF et al (2002) Proteins 46:34–40
110. Verdonk ML, Cole JC, Hartshorn MJ et al (2003) Proteins 52:609–623
111. Wei BQ, Weaver LH, Ferrari AM et al (2004) J Mol Biol 337:1161–1182
112. Jones G, Willett P, Glen RC (1995) J Mol Biol 245:43–53
113. Schnecke V, Swanson CA, Getzoff ED et al (1998) Proteins 33:74–87
114. Leach AR (1994) J Mol Biol 235:345–356
115. Schaffer L, Verkhivker GM (1998) Proteins 33:295–310
116. Zavodszky MI, Kuhn LA (2005) Protein Sci 14:1104–1114
117. Zavodszky MI, Lei M, Thorpe MF et al (2004) Proteins 57:243–261
118. Cavasotto CN, Kovacs JA, Abagyan RA (2005) J Am Chem Soc 127:9632–9640
119. Kallblad P, Dean PM (2003) J Mol Biol 326:1651–1665
120. Nabuurs SB, Wagener M, de Vlieg J (2007) J Med Chem 50:6507–6518
121. Kuhn LA (2008) In: Stroud RM, Finer-Moore J (eds) Computational and structural approaches to drug discovery: ligand-protein interactions. RSC Publishing, Cambridge
122. McCammon JA (2005) Biochim Biophys Acta 1754:221–224
123. Lin JH, Perryman AL, Schames JR et al (2002) J Am Chem Soc 124:5632–5633
124. Lin J-H, Perryman AL, Schames JR et al (2003) Biopolymers 68:47–62
125. Knegtel RMA, Kuntz ID, Oshiro CM (1997) J Mol Biol 266:424–440
126. Macchiarulo A, Nobeli I, Thornton JM (2004) Nat Biotech 22:1039–1045
127. Sotriffer CA, Dramburg I (2005) J Med Chem 48:3122–3125
128. Klebe G, Krämer O, Sotriffer CA (2004) Cell Mol Life Sci 61:783–793
129. Sotriffer CA, Krämer O, Klebe G (2004) Proteins 56:52–66
130. Steuber H, Zentgraf M, Gerlach C et al (2006) J Mol Biol 363:174–187
131. Steuber H, Zentgraf M, Podjarny A et al (2006) J Mol Biol 356:45–56
132. Da Settimo F, Primofiore G, La Motta C et al (2005) J Med Chem 48:6897–6907
133. Zentgraf M, Steuber H, Koch C et al (2007) Angew Chem Int Ed 46:3575–3578
134. Hamelberg D, Mongan J, McCammon JA (2004) J Chem Phys 120:11919–11929
135. Egbertson MS (2007) Curr Top Med Chem 7:1251–1272
136. Keskin O, Gursoy A, Ma B et al (2008) Chem Rev 108:1225–1244

137. Król M, Tournier AL, Bates PA (2007) Proteins 68:159–169
138. Gottschalk KE, Neuvirth H, Schreiber G (2004) Protein Eng 17:183–189
139. Duan Y, Reddy BVB, Kaznessis YN (2005) Protein Sci 14:316–328
140. Smith GR, Sternberg MJE, Bates PA (2005) J Mol Biol 347:1077–1101
141. Burlingham BT, Widlanski TS (2003) J Chem Educ 80:214–218
142. Yung-Chi C, Prusoff WH (1973) Biochem Pharmacol 22:3099–3108
143. Cheluvaraja S, Meirovitch H (2008) J Chem Theor Comput 4:192–208
144. Cheluvaraja S, Mihailescu M, Meirovitch H (2008) J Phys Chem B 112:9512–9522
145. Deng Y, Roux B (2009) J Phys Chem B 113:2234–2246
146. Meirovitch H, Cheluvaraja S, White RP (2009) Curr Protein Pept Sci 10:229–243
147. Mihailescu M, Meirovitch H (2009) J Phys Chem B 113:7950–7964
148. Miyamoto S, Kollman PA (1993) Proteins 16:226–245
149. Zhang X, Mamonov AB, Zuckerman DM (2009) J Comput Chem 30:1680–1691
150. Kirkwood JG (1935) J Chem Phys 3:300–313
151. Zwanzig RW (1954) I Nonpolar Gases J Chem Phys 22:1420–1426
152. Åqvist J, Medina C, Samuelsson JE (1994) Protein Eng 7:385–391
153. Lee MS, Olson MA (2006) Biophys J 90:864–877
154. Kollman P (1993) Chem Rev 93:2395–2417
155. Huang N, Jacobson MP (2007) Curr Opin Drug Disc Dev 10:325–331
156. Gilson MK, Zhou HX (2007) Annu Rev Biophys Biomol Struct 36:21
157. Foloppe N, Hubbard RE (2006) Curr Med Chem 13:3583–3608
158. Honig B, Nicholls A (1995) Science 268:1144–1149
159. Still WC, Tempczyk A, Hawley RC et al (1990) J Am Chem Soc 112:6127–6129
160. Bryce RA, Hillier IH, Naismith JH (2001) Biophys J 81:1373–1388
161. Bashford D, Case DA (2000) Annu Rev Phys Chem 51:129–152
162. Kuhn B, Kollman PA (2000) J Med Chem 43:3786–3791
163. Gouda H, Kuntz ID, Case DA et al (2003) Biopolymers 68:16–34
164. Barril X, Gelpi JL, Lopez JM et al (2001) Theor Chem Acc 106:2–9
165. Masukawa KM, Kollman PA, Kuntz ID (2003) J Med Chem 46:5628–5637
166. Bea I, Cervello E, Kollman PA et al (2001) Comb Chem High Throughput Screen 4:605–611
167. Cox SR, Williams DE (1981) J Comput Chem 2:304–323
168. Pearlman DA (2005) J Med Chem 48:7796–7807
169. Swanson JMJ, Henchman RH, McCammon JA (2004) Biophys J 86:67–74
170. Gilson MK, Honig B (1988) Proteins 4:7–18
171. Scarsi M, Caflisch A (1999) J Comput Chem 20:1533–1536
172. Sharp KA (1994) Curr Opin Struct Biol 4:234–239
173. Holst M, Saied F (1993) J Comput Chem 14:105–113
174. Warwicker J, Watson HC (1982) J Mol Biol 157:671–679
175. Honig B, Nicholls A (1995) Science 268:1144–1149
176. Gilson MK, Sharp KA, Honig BH (1988) J Comput Chem 9:327–335
177. Shen J, Wendoloski J (1996) J Comput Chem 17:350–357
178. Baker NA, Sept D, Joseph S et al (2001) Proc Natl Acad Sci USA 98:10037–10041
179. Bashford D, Gerwert K (1992) J Mol Biol 224:473–486
180. Im W, Beglov D, Roux B (1998) Comput Phys Commun 111:59–75
181. Ma C, Baker NA, Joseph S et al (2002) J Am Chem Soc 124:1438–1442
182. Madura JD, Briggs JM, Wade RC et al (1995) Comput Phys Commun 91:57–95
183. Gilson MK, Honig BH (1986) Biopolymers 25:2097–2119
184. Sharp KA, Honig B (1990) Annu Rev Biophys Biophys Chem 19:301–332

185. Simonson T, Perahia D (1995) Proc Natl Acad Sci USA 92:1082–1086
186. Simonson T, Perahia D (1995) Comput Phys Commun 91:291–303
187. Antosiewicz J, McCammon JA, Gilson MK (1994) J Mol Biol 238:415–436
188. Demchuk E, Wade RC (1996) J Phys Chem 100:17373–17387
189. Demchuk E, Genick UK, Woo TT et al (2000) Biochemistry 39:1100–1113
190. Foloppe N, Sagemark J, Nordstrand K et al (2001) J Mol Biol 310:449–470
191. Grant JA, Pickup BT, Nicholls A (2001) J Comput Chem 22:608–640
192. Richards FM (1977) Annu Rev Biophys Bioeng 6:151–176
193. Swanson JMJ, Mongan J, McCammon JA (2005) J Phys Chem B 109:14769–14772
194. Nina M, Beglov D, Roux B (1997) J Phys Chem B 101:5239–5248
195. Nina M, Im W, Roux B (1999) Biophys Chem 78:89–96
196. Sitkoff D, Sharp KA, Honig B (1994) J Phys Chem 98:1978–1988
197. Swanson JMJ, Adcock SA, McCammon JA (2005) J Chem Theory Comput 1:484–493
198. Wang J, Morin P, Wang W et al (2001) J Am Chem Soc 123:5221–5230
199. Froloff N, Windemuth A, Honig B (1997) Protein Sci 6:1293–1301
200. Huo S, Wang J, Cieplak P et al (2002) J Med Chem 45:1412–1419
201. Schwarzl SM, Tschopp TB, Smith JC et al (2002) J Comput Chem 23:1143–1149
202. Tanford C (1978) Science 200:1012–1018
203. Blokzijl W, Engberts JBFN (1993) Angew Chem Int Ed 32:1545–1579
204. Gao J, Qiao S, Whitesides GM (1995) J Med Chem 38:2292–2301
205. Hunenberger PH, Helms V, Narayana N et al (1999) Biochemistry 38:2358–2366
206. Janin J, Chothia C (1978) Biochemistry 17:2943–2948
207. Scarsi M, Majeux N, Caflisch A (1999) Proteins 37:565–575
208. Hermann RB (1972) J Phys Chem 76:2754–2759
209. Reynolds JA, Gilbert DB, Tanford C (1974) Proc Natl Acad Sci USA 71:2925–2927
210. Sims PA, Wong CF, McCammon JA (2003) J Med Chem 46:3314–3325
211. Shen J, Wendoloski J (1995) Protein Sci 4:373–381
212. Sharp KA, Nicholls A, Fine RF et al (1991) Science 252:106–109
213. Doyle ML (1997) Curr Opin Biotechnol 8:31–35
214. Leavitt S, Freire E (2001) Curr Opin Struct Biol 11:560–566
215. Weber PC, Salemme FR (2003) Curr Opin Struct Biol 13:115–121
216. Finkelstein AV, Janin J (1989) Protein Eng 3:1–3
217. Murray CW, Verdonk ML (2002) J Comput Aided Mol Des 16:741–753
218. Page MI, Jencks WP (1971) Proc Natl Acad Sci USA 68:1678–1683
219. Gidofalvi G, Wong CF, McCammon JA (2002) J Chem Educ 79:1122–1126
220. Sirockin F, Sich C, Improta S et al (2002) J Am Chem Soc 124:11073–11084
221. Searle MS, Williams DH (1992) J Am Chem Soc 114:10690–10697
222. Bonnet P, Bryce RA (2004) Protein Sci 13:946–957
223. Foloppe N, Fisher LM, Howes R et al (2005) J Med Chem 48:4332–4345
224. Steinbrecher T, Case DA, Labahn A (2006) J Med Chem 49:1837–1844
225. Zhou Z, Madura JD (2004) Proteins 57:493–503
226. Kollman PA, Massova I, Reyes C et al (2000) Acc Chem Res 33:889–897
227. Hou T, Guo S, Xu X (2002) J Phys Chem B 106:5527–5535
228. Karplus M, Kushick JN (1981) Macromolecules 14:325–332
229. Ferrara P, Gohlke H, Price DJ et al (2004) J Med Chem 47:3032–3047
230. Huang D, Caflisch A (2004) J Med Chem 47:5791–5797
231. Kuhn B, Gerber P, Schulz-Gasch T et al (2005) J Med Chem 48:4040–4048
232. Zou X, Sun Y, Kuntz ID (1999) J Am Chem Soc 121:8033–8043

233. Barril X, Aleman C, Orozco M et al (1998) Proteins 32:67–79
234. Woods CJ, King MA, Essex JW (2001) J Comput Aided Mol Des 15:129–144
235. Rizzo RC, Toba S, Kuntz ID (2004) J Med Chem 47:3065–3074
236. Donini OAT, Kollman PA (2000) J Med Chem 43:4180–4188
237. Kuhn B, Kollman PA (2000) J Med Chem 43:3786–3791
238. Mardis KL, Luo R, Gilson MK (2001) J Mol Biol 309:507–517
239. Zhou R, Friesner RA, Ghosh A et al (2001) J Phys Chem B 105:10388–10397
240. Resat H, McCammon JA (1996) J Chem Phys 104:7645–7651
241. Simonson T, Archontis G, Karplus M (2002) Acc Chem Res 35:430–437
242. Åqvist J (1996) J Comput Chem 17:1587–1597
243. Åqvist J, Marelius J (2001) Comb Chem High Throughput Screen 4:613–626
244. Åqvist J, Luzhkov VB, Brandsdal BO (2002) Acc Chem Res 35:358–365
245. Jones-Hertzog DK, Jorgensen WL (1997) J Med Chem 40:1539–1549
246. Hansson T, Marelius J, Åqvist J (1998) J Comput Aided Mol Des 12:27–35
247. Essex JW, Jorgensen WL (1995) J Comput Chem 16:951–972
248. Pierce AC, Jorgensen WL (2001) J Med Chem 44:1043–1050
249. Wall ID, Leach AR, Salt DW et al (1999) J Med Chem 42:5142–5152
250. Wang J, Dixon R, Kollman PA (1999) Proteins 34:69–81
251. Zoete V, Michielin O, Karplus M (2003) J Comput Aided Mol Des 17:861–880
252. Ljungberg KB, Marelius J, Musil D et al (2001) Eur J Pharm Sci 12:441–446
253. Hansson T, Åqvist J (1995) Protein Eng 8:1137–1144
254. Tounge BA, Reynolds CH (2003) J Med Chem 46:2074–2082
255. Rizzo RC, Tirado-Rives J, Jorgensen WL (2001) J Med Chem 44:145–154
256. Svab I, Alexandru D, Vitos G et al (2004) J Cell Mol Med 8:551–562
257. Dey I (1999) J Biomol Struct Dyn 16:891–900
258. Lamb ML, Tirado-Rives J, Jorgensen WL (1999) Bioorg Med Chem 7:851–860
259. Smith RH Jr, Jorgensen WL, Tirado-Rives J et al (1998) J Med Chem 41:5272–5286
260. Pierce KL, Premont RT, Lefkowitz RJ (2002) Nat Rev Mol Cell Biol 3:639–650
261. Hulten J, Bonham NM, Nillroth U et al (1997) J Med Chem 40:885–897
262. Graffner-Nordberg M, Kolmodin K, Åqvist J et al (2001) J Med Chem 44:2391–2402
263. Singh P, Mhaka AM, Christensen SB et al (2005) J Med Chem 48:3005–3014
264. Paulsen MD, Ornstein RL (1996) Protein Eng 9:567–571
265. Chen J, Wang R, Taussig M et al (2001) J Org Chem 66:3021–3026
266. Wang W, Wang J, Kollman PA (1999) Proteins 34:395–402
267. Chipot C (2003) J Comput Chem 24:409–415
268. Carlson HA, Jorgensen WL (1995) J Phys Chem 99:10667–10673
269. Rizzo RC, Udier-Blagovic M, Wang DP et al (2002) J Med Chem 45:2970–2987
270. Jorgensen WL, Ravimohan C (1985) J Chem Phys 83:3050–3054
271. Jorgensen WL, Tirado-Rives J (2005) Proc Natl Acad Sci USA 102:6665–6670
272. Bennett CH (1976) J Comput Phys 22:245–268
273. Lawrenz M, Baron R, McCammon JA (2009) J Chem Theor Comput 5:1106–1116
274. Adcock SA, McCammon JA (2006) Chem Rev 106:1589–1615
275. Zagrovic B, van Gunsteren WF (2006) J Chem Theor Comput 3:301–311
276. Fujitani H, Tanida Y, Ito M et al (2005) J Chem Phys 123:084108–5
277. Reinhardt WP, Miller MA, Amon LM (2001) Acc Chem Res 34:607–614
278. Fajer M, Hamelberg D, McCammon JA (2008) J Chem Theor Comput 4:1565–1569
279. Woods CJ, Essex JW, King MA (2003) J Phys Chem B 107:13703–13710
280. de Oliveira CA, Hamelberg D, McCammon JA (2008) J Chem Theor Comput 4:1516–1525

281. Hamelberg D, de Oliveira CA, McCammon JA (2007) J Chem Phys 127:155102–155109
282. Steinbrecher T, Hrenn A, Dormann KL et al (2008) Bioorg Med Chem 16:2385–2390
283. Lundstrom K (2006) Curr Protein Pept Sci 7:465–470
284. Palczewski K (2006) Annu Rev Biochem 75:743
285. Fanelli F, De Benedetti PG (2005) Chem Rev 105:3297–3351
286. Cherezov V, Rosenbaum DM, Hanson MA et al (2007) Science 318:1258–1265
287. Rasmussen SGF, Choi HJ, Rosenbaum DM et al (2007) Nature 450:383–387
288. Warne T, Serrano-Vega MJ, Baker JG et al (2008) Nature 454:486–491
289. Kenakin T (2002) Nat Rev Drug Discov 1:103–110
290. Cremades N, Sancho J, Freire E (2006) Trends Biochem Sci 31:494–496
291. Barnett-Norris J, Lynch D, Reggio PH (2005) Life Sci 77:1625–1639
292. Fanelli F, De Benedetti PG (2006) J Comput Aided Mol Des 20:449–461
293. Feng X, Muller T, Mizrachi D et al (2008) Endocrinol 149:1705–1717
294. Angelova K, Fanelli F, Puett D (2008) Mol Endocrinol 22:126–138
295. Raimondi F, Seeber M, De Benedetti PG et al (2008) J Am Chem Soc 130:4310–4325
296. Seeber M, De Benedetti PG, Fanelli F (2003) J Chem Inf Model 43:1520–1531
297. Vitale RM, Pedone C, De Benedetti PG et al (2004) Proteins 56:430–448
298. Yan F, Mosier PD, Westkaemper RB et al (2008) Biochemistry 47:1567–1578
299. Ballesteros JA, Jensen AD, Liapakis G et al (2001) J Biol Chem 276:29171–29177
300. Dror RO, Arlow DH, Borhani DW et al (2009) Proc Natl Acad Sci USA 106:4689–4694
301. Wolf S, Böckmann M, Höweler U et al (2008) FEBS lett 582:3335–3342
302. Hamelberg D, McCammon JA (2004) J Am Chem Soc 126:7683–7689
303. Lu YP, Yang CY, Wang SM (2006) J Am Chem Soc 128:11830–11839
304. Zhang L, Hermans J (1996) Proteins 24:433–438
305. Olano LR, Rick SW (2004) J Am Chem Soc 126:7991–8000
306. Lounnas V, Pettitt BM (1994) Proteins 18:133–147
307. Bizzarri AR, Cannistraro S (2002) J Phys Chem B 106:6617–6633
308. Luise A, Falconi M, Desideri A (2000) Proteins 39:56–67
309. Otting G, Liepinsh E, Wüthrich K (1991) Science 254:974–980
310. Henchman RH, McCammon JA (2002) J Comput Chem 23:861–869
311. Huang HC, Jupiter D, Qiu M et al (2008) Biopolymers 89:210–219
312. Brunne RM, Liepinsh E, Otting G et al (1993) J Mol Biol 231:1040–1048
313. Barillari C, Taylor J, Viner R et al (2007) J Am Chem Soc 129:2577–2587
314. Abel R, Young T, Farid R et al (2008) J Am Chem Soc 130:2817–2831
315. Lazaridis T (1998) The Journal of Physical Chemistry B 102:3531–3541
316. Van Drie J (2007) J Comput Aided Mol Des 21:591–601
317. Moore GE (1965) Electronics 38:114–117
318. Mitsutake A, Sugita Y, Okamoto Y (2001) Biopolymers 60:86–123
319. Okabe T, Kawata M, Okamoto Y et al (2001) Chem Phys Lett 335:435–439
320. Berg BA, Neuhaus T (1992) Phys RevLett 68:9–12
321. Kubitzki MB, de Groot BL (2007) Biophys J 92:4262–4270
322. Shaw, DE, Deneroff, MM, Dror, RO et al (2007) Anton, a special-purpose machine for molecular dynamics simulation. Proceedings of the 34th annual international symposium on computer architecture, San Diego, California
323. Miller BT, Singh RP, Klauda JB et al (2008) J Chem Inf Model 48:1920–1929
324. Stjernschantz E, Marelius J, Medina C et al (2006) J Chem Inf Model 46:1972–1983

CHAPTER 17

INTERPRETATION OF KINETIC ISOTOPE EFFECTS IN ENZYMATIC CLEAVAGE OF CARBON-HYDROGEN BONDS

WILLEM SIEBRAND AND ZORKA SMEDARCHINA
Steacie Institute for Molecular Sciences, National Research Council of Canada, Ottawa K1A 0R6, Canada,
e-mail: willem.siebrand@nrc-cnrc.gc.ca

Abstract: A model is presented which relates kinetic isotope effects and their temperature dependence to physical parameters governing enzymatic carbon–hydrogen cleavage, a reaction that typically exhibits a large isotope effect indicative of tunneling. The model aims to replace the Arrhenius equation, which is not valid for tunneling reactions, and the one-dimensional Bell model, which excludes skeletal vibrations. Cast in the form of three user-friendly analytical equations, it applies directly to the observed rate constants and effective activation energies, for both adiabatic and nonadiabatic hydrogen transfer. The approach has the dual aim of establishing criteria to probe whether a set of data can be assigned to a single rate-limiting tunneling step and to find numerical values for the proton or hydrogen transfer distance and for properties of the tunneling mode as well as skeletal modes involved in this step. It is applied to several enzymatic (model) reactions, including free-radical reactions catalyzed by cofactor B_{12} and phenylalanine hydroxylase, and electron-assisted proton-transfer reactions catalyzed by soybean lipoxygenase-1 and methylamine dehydrogenase.

Keywords: Enzymatic reactions, proton tunneling, CH-bond cleavage, deuterium isotope effects

17.1. INTRODUCTION

In many enzymatic CH-cleavage reactions the CH bond is broken by transfer of the hydrogen to an acceptor, usually an oxygen or other carbon atom. If, upon replacement of hydrogen by deuterium (or tritium), the rate of such a reaction is strongly reduced, it is evident that transfer occurs by quantum-mechanical tunneling and that this tunneling is, by itself or in combination with other steps, rate-limiting. The

P. Paneth and A. Dybala-Defratyka (eds.), *Kinetics and Dynamics*,
Challenges and Advances in Computational Chemistry and Physics 12,
DOI 10.1007/978-90-481-3034-4_17,

question to be addressed in this contribution is whether it is possible to go beyond these obvious conclusions and draw additional information from the magnitude of a kinetic isotope effect (KIE) and its temperature dependence. Elsewhere [1] we showed that there is no apparent common pattern among the various enzymatic CH-cleavage reactions for which data are available. Given the complexity of enzymatic processes, direct attempts to calculate these properties run into many problems. To be at least partly realistic, elaborate calculations are required that nevertheless cannot avoid arbitrary cut-offs. There are always too many variables and therefore it is unlikely that a particular fit yields unambiguous information on structure or mechanism. The problem is compounded if the data include several reactions involving minor variations in enzyme or substrate. Obviously, the successful reproduction of an observed KIE of a single enzyme-substrate combination at a single temperature will tell us more about the theoretical methods used than about the enzymatic system studied. Moreover, the implied assumption that the reported kinetic data pertain to a kinetically isolated single step in the enzymatic reaction sequence may not even be valid.

Experimentally, the hydrogen transfer rate constant is usually measured as the initial rate of product formation under excess substrate conditions. This follows from the Michaelis-Menten equation, which in its simplest form is given by

$$k_{t\to 0} = k_{\mathrm{cat}}[\mathrm{SH}]/([\mathrm{SH}] + K_{\mathrm{M}}) \tag{17-1}$$

where [SH] is the substrate concentration and $K_{\mathrm{M}} = (k_{-1} + k_{\mathrm{cat}})/k_1$ is the Michaelis constant, k_1 and k_{-1} being the association and dissociation rate constants, respectively, of the reactive complex formed by enzyme and substrate, and k_{cat} is the hydrogen transfer rate constant.

For most chemical reactions the rate constant $k(T)$ depends exponentially on the temperature, as expressed by the Arrhenius equation

$$ln\, k(T) = ln\, A - E_a/k_B T \tag{17-2}$$

where E_a is the activation energy and $k_B T$ the thermal energy. Pictorially, it describes the reaction as a transition over a potential-energy barrier with height E_a, the frequency factor A being the rate at which the barrier is attacked by collisional, diffusional or vibrational motions. Evidently, this picture is not valid for light particles such as protons that need not go over the barrier but can tunnel through it. Nevertheless, the rate of tunneling reactions, in particular those observed in biological systems, is generally reported in terms of Arrhenius parameters. Although for such reactions these parameters are no longer constants, they appear to be constant if measured over the narrow temperature interval to which biological systems are usually constrained. They are also isotope-dependent, although the potential energy and thus the barrier is mass-independent. Obviously, they have lost their original physical meaning: E_a is no longer the barrier height, i.e., the energy of the transition state, and A is no longer the attack frequency characteristic of the reaction coordinate. Without further modeling of the tunneling process, these parameters are quite meaningless.

The fact that, depending on its energy, the tunneling particle may penetrate the barrier at different points and follow paths of different lengths undoubtedly complicates the picture. Moreover, the barrier itself is not static, as in the Bell model [2], but subject to the vibrations of the molecular skeleton. However, these complications are offset by the existence of different hydrogen isotopes; their different masses leave the barrier essentially untouched but greatly affect their ability to penetrate it. This allows us to analyze tunneling reactions to the same extent as classical reactions, although the methods are different and, admittedly, more complicated. However, the very complexity of enzymatic processes implies that even the most sophisticated methods currently in use can only deal with an isolated step in the reaction chain and can therefore produce only approximate answers. In this contribution, we take advantage of this situation by simplifying the treatment of the tunneling step to the point where it can be used routinely for the interpretation and presentation of tunneling data. In contrast to the Arrhenius method and its adaptations based on the Bell model, we start from premises that are fundamentally correct, even if the approximations introduced to achieve tractability will undoubtedly cause inaccuracies. However, these approximations are not fundamental to the method and can be gradually removed if more information becomes available and if the kinetic data warrant it. Our approach, even in its simplest form, will provide answers on the role of skeletal vibrations in the tunneling reaction, answers that cannot be obtained from the one-dimensional Arrhenius/Bell approach. This role is a fundamental aspect of the specificity and efficiency of enzymes.

The need to include skeletal modes explicitly in the modeling follows, firstly, from the fact that such modes can modulate the barrier width and thus the tunneling distance, and, secondly, from the fact that they, rather than the high-frequency CH- and CD-stretching vibrations of the tunneling coordinate, govern the temperature dependence of the tunneling process. This realization is far from new [3–8], but earlier implementations rely on elaborate calculations that are overly complex and thus unattractive, and/or parameters that are not easy to guess. The new aspect of our modeling is that it is analytical and leads to formulas that apply directly to the observed rate constants. A full description together with application to a specific set of kinetic data is published elsewhere [8]. The application is to reactions catalyzed by soybean lipoxygenase-1, studied by Klinman and coworkers [5, 6]; it includes not only an analysis of the rate constants and their temperature and isotope effects, but also an analysis of the effect of six mutations, leading to several new results. Questions addressed, apart from the basic question whether the data are representative of a single rate-limiting tunneling step, include the magnitude of the tunneling distance, the properties of the skeletal modes involved in the tunneling, and, in general, the mechanism of the tunneling reaction.

Two such mechanisms have been found to be operative for enzymatic CH cleavage. The first and presumably older involves a free radical that abstracts the hydrogen by valence forces. Such a mechanism is well suited to hydrogen transfer between two carbon centers since it is operative over large distances. However, its long range turns it into a ‘brute force’ mechanism that runs the risk of causing

collateral damage, a problem that is avoided by the evolution of a more gentle ionic mechanism involving an atom or group that withdraws an electron from the CH bond to initiate a three-center process in which this bond donates an electron to one receptor and a proton to another. This mechanism, usually referred to as electron-assisted or nonadiabatic proton transfer, uses van der Waals rather than valence forces and involves shorter transfer distances to a proton receptor such as an oxygen center that engages in hydrogen bonding. As we show in this contribution, the analysis we propose can determine, on the basis of the observed KIE and its temperature dependence, which mechanism prevails in a given enzymatic reaction.

In the next section we briefly review the model and quote the equations to be used in the analysis of the observed kinetic data. These equations together with standard parameter values are used to determine to what extent the tunneling step can be identified as rate-limiting. If the data pass this test, qualitative and quantitative information can be obtained on the mechanism of the tunneling reaction and the structure of the reaction center. In subsequent sections the model is applied to a representative set of kinetic data on enzymatic systems and models thereof, including soybean lipoxygenase-1 (SLO1) [5, 6], coenzyme B_{12} [9, 10], primary amine dehydrogenases [11, 12], and phenylalanine hydroxylase [13].

17.2. MODEL

To replace the Arrhenius Eq. (17-2), which, as pointed out above, is not valid for tunneling reactions, we use an equation derived in [8] from Fermi's Golden Rule, which in its simplest form reads:

$$ln\, k(T) \simeq ln\left[\frac{2\pi}{\hbar}J^2\varrho(E_f)\right] - \frac{r_0^2}{2a_0^2 + A^2(T)} - \Delta E^{\text{eff}}/k_B T \qquad (17\text{-}3)$$

The derivation applies explicitly to tunneling, primarily to adiabatic transfer, but, as shown in [8], it can also be used for nonadiabatic reactions provided it is applied to relative rather than absolute rate constants. The form of the first term on the rhs follows immediately from the Golden Rule. J is an electronic coupling term and $\varrho(E_f)$ is the density of final states; both parameters are assumed to be isotope- and temperature-independent and will not concern us further. The second term represents the behavior of the vibrations directly involved in the transfer. The parameter r_0 denotes a particular value of the tunneling coordinate r, namely the equilibrium transfer distance, defined as the separation of the two equilibrium positions of the tunneling particle, H or D. The parameter a_0 (c.q. a_{eff}) represents the harmonic (c.q. anharmonic) zero-point amplitude of the XH-stretching vibration along the tunneling coordinate r, where X denotes C, O, or whatever atom carries the tunneling particle. If X in the final state differs from X in the initial state, we take an average value for a_{eff}. Thus for $A(T) = 0$, the second term represents the exponent of an overlap integral between zero-point vibrations in a static

double-well potential formed by two equivalent parabolas displaced by r_0. If an excited vibrational level is involved, there will be additional terms, but these are nonexponential and tend to be unimportant for relative rate constants.

The term $A(T)$ represents the effect of those skeletal vibrations that 'promote' or 'gate' tunneling. These modes have components R parallel to the tunneling coordinate r and are symmetric relative to the dividing plane between the initial (reactant) and final (product) states. They effectively shorten the tunneling distance [3, 14]. Finally, ΔE^{eff} is a term that includes any endothermicity, presumed to be small in typical enzymes, as well as contributions from those skeletal modes that are antisymmetric relative to the dividing plane of the reaction; since they have the same symmetry as the tunneling mode, they can contribute to the barrier. Thus they hinder tunneling similar to the Marcus-type exponent describing polar-solvent reorganization. We will assume that the temperature- and isotope-dependence of this term is weak and can be neglected. As the notation emphasizes, we assume that only the skeletal amplitudes are temperature-dependent at biologically relevant temperatures. For a critical discussion of the various approximations used to derive Eq. (17-3) we refer to [8]. To relate the CH and CD modes, we set $m^{\mathrm{D}} = 2m^{\mathrm{H}} = 2$, so that $(a_{\mathrm{eff}}^{\mathrm{H}})^2 = \sqrt{2}(a_{\mathrm{eff}}^{\mathrm{D}})^2$.

Since many of the parameters in Eq. (17-3) are irrelevant to our purpose, we simplify the notation by keeping only those that are needed and expressing all lengths in units $a_{\mathrm{eff}}^{\mathrm{H}}$. The new parameters

$$L = ln[\frac{2\pi}{\hbar} J^2 \varrho(E_f)]; \quad \rho = \frac{r_0^2}{2(a_{\mathrm{eff}}^{\mathrm{H}})^2}; \quad z(T) = \frac{A^2(T)}{2(a_{\mathrm{eff}}^{\mathrm{H}})^2}; \quad B(T) = \Delta E^{\mathrm{eff}}/k_B T \tag{17-4}$$

reduce Eq. (17-3) to

$$ln\, k^{\mathrm{H}}(T) = L - \frac{\rho}{1+z(T)} - B(T), \quad ln\, k^{\mathrm{D}}(T) = L - \frac{\rho\sqrt{2}}{1+z(T)\sqrt{2}} - B(T) \tag{17-5}$$

In this notation $z(T)$ and $z(T)\sqrt{2}$ are the contributions of the skeletal modes to the transfer relative to the contribution of the H and D tunneling modes, respectively. This contribution is both isotope- and temperature-dependent and turns out to be a key parameter in the analysis. It enters the expression for the KIE, denoted by $\eta = k^{\mathrm{H}}(T)/k^{\mathrm{D}}(T)$ in the form

$$ln\, \eta = \frac{\rho(\sqrt{2}-1)}{[1+z(T)][1+z(T)\sqrt{2}]} \tag{17-6}$$

Hence within the model the KIE is fully determined by the transfer distance, represented by ρ, and the contributions of the skeletal modes to the transfer, represented by $z(T)$ and $z(T)\sqrt{2}$. If the atoms between which the tunneling takes place are known, or can be reduced to two possibilities, as is often the case, the transfer distance r_0 and thus the transfer parameter ρ can be estimated. Substitution

in Eq. (17-6) will typically give rise to a range of $z(T)$ values, with which the problem of the temperature dependence can then be tackled.

For this we require a specific form for the skeletal-modes amplitude $A(T)$. Elsewhere [14] we have shown that in simple cases we can deal with the skeletal vibrations $\Lambda(\mathbf{R})$ that modulate the equilibrium transfer distance r_0 in Eq. (17-3) by combining them into an *effective* mode, whose properties can be evaluated if a full force field calculation is available. Here we follow the common procedure of fitting the effective frequency and reduced mass of this mode to the observed kinetic parameters. Assuming that it is harmonic with an equilibrium position $R = R_0$, a frequency Ω and a reduced mass M, we derive values for these parameters from the observed temperature dependence of the transfer rate constants. In this approach the squared amplitude of the effective mode takes the form

$$A^2(T) = (\hbar/M\Omega)coth(\hbar\Omega/2k_BT) \tag{17-7}$$

or, in dimensionless notation,

$$z(\xi) = \frac{\hbar/M\Omega}{2(a_{\text{eff}}^{\text{H}})^2}coth\xi \tag{17-8}$$

where $\xi = \hbar\Omega/2k_BT$ measures the vibrational energy in terms of the thermal energy and thus represents a scaled inverse temperature. The distribution $z(\xi)$ replaces z(T) when the skeletal modes are represented by a single effective mode. Since this representation may tend to reduce skeletal-mode support, we expect that $z(\xi) \leq z(T)$.

Differentiating Eq. (17-5) with respect to the inverse temperature, and noting that $dT^{-1} = T^{-1}dlnT^{-1}$, we obtain in dimensionless notation

$$-\frac{dlnk^{\text{H}}(T)}{dlnT^{-1}} \equiv \frac{E_a^{\text{H}}}{k_BT} = \rho\frac{z(\xi)F(\xi)}{[1+z(\xi)]^2} + B(T) \tag{17-9}$$

$$-\frac{dlnk^{\text{D}}(T)}{dlnT^{-1}} \equiv \frac{E_a^{\text{D}}}{k_BT} = \rho\frac{2z(\xi)F(\xi)}{[1+z(\xi)\sqrt{2}]^2} + B(T) \tag{17-10}$$

The scaled activation energies can be read directly off conventional Arrhenius plots since $dlnT^{-1} = TdT^{-1}$. The function $F(\xi) = \xi(coth\xi - tanh\xi)$, which varies smoothly from 0 to 1, as illustrated in Figure 17-1, depends only on the ratio $\hbar\Omega/k_BT$. It represents the dependence of the model on the assumption that the effect of skeletal vibrations on the transfer dynamics can be approximated by that of a single effective harmonic oscillator with frequency Ω collinear with the tunneling mode. It is very useful because, due to its limited range of permissible values, it provides a test for the compatibility of the data with the model. Apart from the upper limit $F(\xi) = 1$, there is a lower limit set by the physically acceptable range of values for the skeletal-mode frequency Ω.

Equations (17-9) and (17-10) express the dependence of the activation energies on the parameter ξ and thus on the temperature, information that cannot be obtained

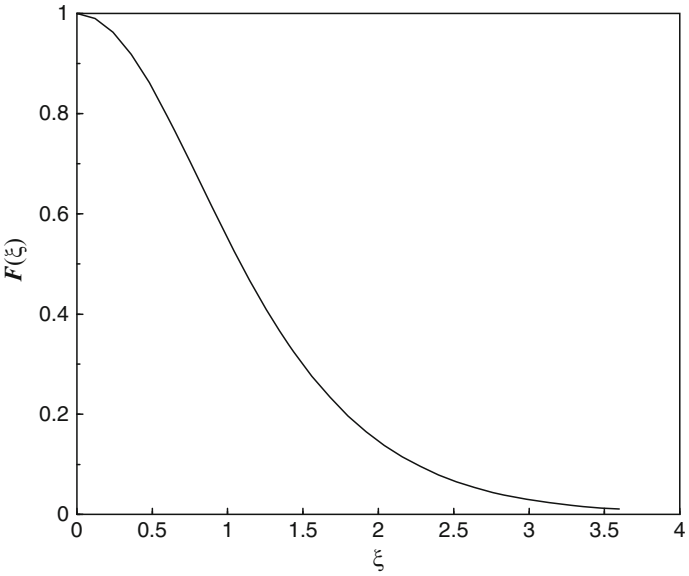


Figure 17-1. Plot of the function $F(\xi) = \xi(coth\xi - tanh\xi)$

from the Arrhenius formula. Since both equations, considered as functions of $z(\xi)$, depend also on $F(\xi)$ and on $B(T)$, it is often more convenient not to use them directly but to transform to their ratio

$$\frac{E_a^{\mathrm{D}} - \Delta E^{\mathrm{eff}}}{E_a^{\mathrm{H}} - \Delta E^{\mathrm{eff}}} = 2\left[\frac{1 + z(\xi)}{1 + z(\xi)\sqrt{2}}\right]^2 \tag{17-11}$$

and their difference

$$\frac{dln\eta}{dlnT^{-1}} \equiv \frac{\Delta E_a}{k_B T} = z(\xi)F(\xi)\rho\left\{\frac{2}{[1 + z(\xi)\sqrt{2}]^2} - \frac{1}{[1 + z(\xi)]^2}\right\} \tag{17-12}$$

where $\Delta E_a = E_a^{\mathrm{D}} - E_a^{\mathrm{H}}$. Using Eq. (17-6) we can eliminate ρ and obtain an equation that depends only on $z(\xi)$ and $F(\xi)$:

$$\frac{\Delta E_a / k_B T}{ln\eta} = \frac{z(\xi)F(\xi)}{\sqrt{2} - 1}\left[2\frac{1 + z(\xi)}{1 + z(\xi)\sqrt{2}} - \frac{1 + z(\xi)\sqrt{2}}{1 + z(\xi)}\right] \tag{17-13}$$

Thus Eqs. (17-11) and (17-13) have each only one extra parameter, ΔE^{eff} or $F(\xi)$. Moreover, these equations provide two new compatibility tests. From Eq. (17-11) it follows that $E_a^{\mathrm{H}} \leq E_a^{\mathrm{D}} \leq 2E_a^{\mathrm{H}}$, and from Eq. (17-13) that $\Delta E_a / k_B T \leq 2ln\eta$. A disadvantage of Eqs. (17-11) and (17-13) relative to Eqs. (17-9) and (17-10) is that they may be more sensitive to the accuracy of the measured activation energies. Hence in practice it is prudent to use all four equations in the analysis.

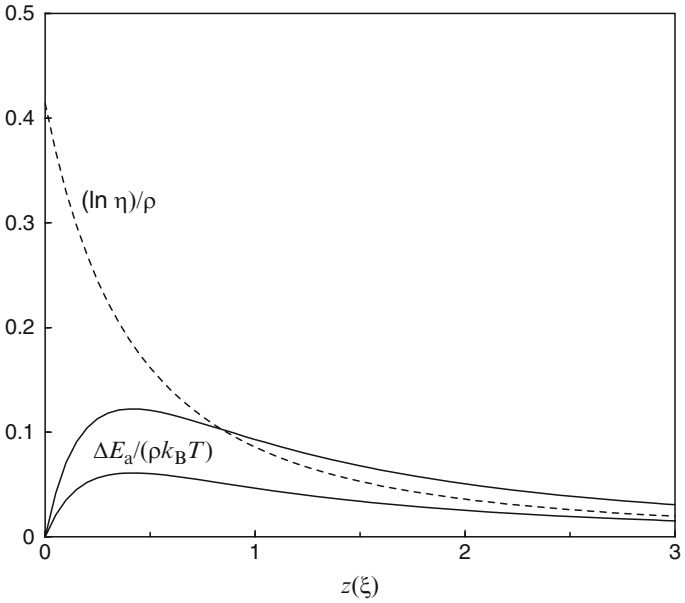


Figure 17-2. Plot of $\Delta E_a/(\rho k_B T)$ (solid lines) and $(ln\,\eta)/\rho$ (broken line) against $z(\xi)$ for $F(\xi) = 1$ (*top*) and 0.5 (*bottom*)

In Figure 17-2, we have plotted $\Delta E_a/(\rho k_B T)$ from Eq. (17-12) and $(ln\eta)/\rho$ from Eq. (17-6) against $z(\xi)$ for $F(\xi) = 1$ and 0.5. In principle, the graph allows us to deduce both ρ, and thus the transfer distance r_0, and $z(\xi)$, and thus the contribution of the effective mode to the tunneling, from the KIE and its temperature dependence, provided $F(\xi)$, which depends solely on the effective-mode frequency, can be estimated. If instead ρ can be estimated, we can deduce both r_0 and $F(\xi)$. Of course such estimates require reliable input values.

With Eqs. (17-6) and (17.9–17.13) we have the tools needed to test whether the temperature- and isotope-dependent rate constants k_{cat} of a given enzymatic reaction is compatible with the assumption that the rate is determined by a single tunneling step. If so, we can use the data to extract information on the vibrations governing the rates; if not, there is evidence that the rate is limited by more than one step. If the interfering slow steps are isotope-independent, they may only interfere with hydrogen and not with deuterium transfer, which is generally much slower. In that case the analysis should be mainly based on the deuterium data; an example of such a situation is treated in [8].

The treatment is readily extended to ratios of rate constants involving the same isotope but slightly different substrates or enzymes such as those obtained by mutations. In [8] this approach was applied to six mutants of the enzyme SLO1; the results of this analysis are summarized in Section 17.4.

17.3. PHYSICAL PARAMETERS

Two key parameters of the model are the equilibrium proton transfer distance r_0, i.e. the distance between the proton equilibrium positions in the initial and final state, and the zero-point amplitude of the proton in the two states, $a_{\text{eff}}^{\text{H}}$. In practice, the transfer distance can be reasonably estimated from standard bond lengths and van der Waals radii, and the zero-point amplitude from standard XH-stretching potentials. Together they determine the distance parameter ρ defined by Eq. (17-4). For instance, the van der Waals radius of a donating methyl or methylene group is about 2 Å and that of an accepting oxygen atom about 1.4 Å. The standard CH-bond length is about 1.1 Å and the standard OH-bond length about 1.0 Å.

In the case of nonadiabatic transfer of the proton between a carbon and an oxygen center, we expect that hydrogen bonding will reduce the C···O distance and increase the OH-bond length. In the initial state we expect little hydrogen bonding, but in the final state, which contributes on an equal footing in the Golden Rule approach, the ionic structure may lead to strong bonding. Therefore we estimate the transfer distance (in Å) to be in the range $1.0 \leq r_0 \leq 1.2$. The zero-point amplitude a_0 for a harmonic oscillator with $m = 1$ and $\omega = 3{,}000\ \text{cm}^{-1}$ equals 0.105 Å. Taking into account that the oscillator will be anharmonic and that the anharmonicity will increase with increasing hydrogen bonding, we estimate $0.11 < a_{\text{eff}} < 0.14$ Å. Combining these estimates, we arrive at $25 \leq \rho \leq 50$.

While hydrogen bonding reduces the tunneling distance, which favors transfer, it also tightens the C. . .O binding and thus increases the frequency Ω of the effective mode, which hinders its thermal excitation and thus has the opposite effect. Since the enzymes under investigation are the product of a long evolutionary process, the physical parameters governing the reaction are expected to be optimized such that at biologically relevant temperatures the contributions of the tunneling mode and the promoting mode are comparable, which leads to the estimate $z(\xi) \sim 1$. For the mass of the promoting mode we expect a relatively low value that may approach the lower limit $M \sim 1$; since the promoting mode is an effective mode this is not an absolute limit. The frequency of the C···H···O vibration is expected to be low in the initial state in which there will be little hydrogen bonding, but may be quite high in the final state, which is ionic. We estimate that the average of these values will remain below 400 cm^{-1}, which implies $F(\xi) \geq 0.5$.

For free radical reactions corresponding to neutral hydrogen transfer between methyl and/or methylene groups, we combine the van der Waals radius of about 2.0 Å for these groups with the CH-bond length of about 1.08 Å to obtain an equilibrium transfer distance $1.8 \leq r_0 \leq 1.9$ Å. A similar value, namely 1.88 Å is obtained from data for the abstraction of a hydrogen atom by a methyl radical from the methyl group of methanol in methanol glass; using esr techniques, Doba et al. [15] measured a distance of 2.96 Å between the radical carbon and the nearest methyl proton. Since there will be no significant hydrogen bonding, the anharmonicity will be that of a standard CH bond; on that basis we estimate $a_{\text{eff}} \simeq 0.11$ Å. These

estimates indicate that for neutral hydrogen transfer between two carbon centers ρ should be about 150. The absence of significant hydrogen bonding between the carbon centers also implies that the effective-mode frequency will be low in terms of the thermal energy. As a result, the effective-mode function $F(\xi)$ may be replaced by its maximum value $F(\xi) = 1$ for these transfers.

17.4. APPLICATION TO LIPOXYGENASE-1

The study of CH cleavage by soybean lipoxygenase-1 and six mutants by Klinman et al. [5, 6] represents the most complete experimental study presently available in the field of enzyme kinetics. Since a detailed analysis of these kinetic data was reported elsewhere [8], we restrict ourselves here to a brief summary to place it in the context of the other systems for which kinetic data are available. The observed rate constants and their temperature dependence are collected in Table 17-1. Of the mutations, the first four occur at the same site located at some distance from the reaction center [5, 6], while the last two occur at different and closer sites [5]. In all cases mutants differ only by an alkyl chain from the wild-type (WT) enzyme and from each other. The WT enzyme and the first two mutants have the same rate constants at $T = 298$ K and thus the same values for the parameters ρ and $z(298)$. It may appear therefore that these mutations have no effect on the transfer; however, the observed activation energies, which, admittedly have relatively large uncertainties, contradict this. The following two mutants have somewhat smaller rate constants, although the mutations are very similar as well as on the same site. Contrary to the suggestion of Meyer et al. [6], our analysis indicates that there are no prominent changes in the transfer distance. The final two mutants show a strongly reduced activity, but their KIEs are similar to those of the others. Here the lower activity implies an increase in the transfer distance, and thus in ρ, but the similar KIE indicates that the increased distance is accompanied by an increase in the contribution of the skeletal modes, $z(298)$. Specifically it indicates that r_0 must be

Table 17-1 Rate constants (in s^{-1}) and activation energies at $T = 298$ K of soybean lipoxygenase-1 and six mutants calculated by linear regression from the kinetic data reported as Supporting Information in [5, 6]

Enzyme	$\ln k^H$	$\ln k^D$	$\ln \eta$	E_a^H/k_BT	E_a^D/k_BT	$\Delta E_a/k_BT$
Linoleic/SLO1	5.7 ± 0.2	1.3 ± 0.3	4.4 ± 0.5	2.8 ± 0.5	5.3 ± 1.2	2.4 ± 1.7
Mutant I553A	5.7 ± 0.1	1.2 ± 0.4	4.5 ± 0.5	3.3 ± 0.7	8.6 ± 1.0	5.3 ± 1.7
Mutant I553L	5.8 ± 0.1	1.3 ± 0.3	4.5 ± 0.4	0.7 ± 1.2	6.7 ± 0.5	6.0 ± 1.7
Mutant I553V	4.3 ± 0.2	0.0 ± 0.4	4.3 ± 0.6	4.1 ± 0.9	8.3 ± 1.1	4.2 ± 2.0
Mutant I553G	4.1 ± 0.1	-1.2 ± 0.3	5.3 ± 0.4	0.0 ± 0.7	7.3 ± 0.7	7.3 ± 1.4
Mutant L546A	0.4 ± 0.2	-4.4 ± 0.4	4.8 ± 0.6	5.6 ± 0.4	9.3 ± 0.8	3.7 ± 1.2
Mutant L754A	-1.7 ± 0.3	-6.4 ± 0.4	4.6 ± 0.6	6.2 ± 0.8	8.9 ± 1.7	2.7 ± 2.5

Table 17-2 Model parameters calculated in [8] from the data listed in Table 17-1. Values of ρ in parentheses are chosen estimates

Enzyme	ρ	$z(298)$	$F(\xi)$	ξ	Ω/cm^{-1}	$_{M/m}{}^{\text{H}}$
Lin./SLO1	(25)	0.44 ± 0.06	0.63 ± 0.10	0.87 ± 0.15	350 ± 80	14 ± 4
Mut. I553A	(25)	0.43 ± 0.06	~ 1	Small	Small	Large
Mut. I553L	(25)	0.42 ± 0.06	0.81 ± 0.07	0.56 ± 0.05	220 ± 30	32 ± 3
Mut. I553V	31 ± 2	0.60 ± 0.08	0.76 ± 0.10	0.65 ± 0.10	260 ± 50	17 ± 4
Mut. I553G	26 ± 2	0.38 ± 0.06	0.87 ± 0.10	0.45 ± 0.06	180 ± 40	52 ± 10
Mut. L546A	37 ± 3	0.65 ± 0.10	0.88 ± 0.10	0.44 ± 0.05	180 ± 50	31 ± 8
Mut. L754A	53 ± 4	1.0 ± 0.2	0.59 ± 0.15	0.93 ± 0.20	370 ± 80	6 ± 2

roughly proportional to the amplitude $A(T)$ of the effective mode, which supports the notion that this relationship is governed by hydrogen bonding in the final state of the transfer reaction.

Turning to the activation energy, we observe large experimental uncertainties and a loss of regularity. Application of Eqs. (17.9–17.13) leads to the conclusion that the data are not fully compatible with the model, a conclusion supported by the observation of very small and even vanishing activation energies for H, as opposed to D, transfer, which is not acceptable for room-temperature tunneling. It seems likely that these low values are due to contamination of the corresponding rates by isotope-independent slow steps in the reaction sequence. Such contaminations have indeed been observed in other reactions catalyzed by lipoxygenase [16]. It is associated with fast tunneling steps, which appear to have evolved so as to reach a rate approaching that of other steps in the reaction. To complete the analysis we have therefore ignored the offending E_a^{H} values of the active enzymes. In Table 17-2 we summarize the quantitative results taken from [8].

17.5. APPLICATION TO FREE RADICAL TRANSFER

Most of the kinetic data on transfer of hydrogen transfer by a free-radical mechanism involve cobalamine (Cbl), a cobalt-centered complex which can release a ligand in the form of a free radical. A typical example is coenzyme B_{12} in which the ligand is 5′-deoxyadenosyl (Ado). The radical thus formed abstracts a hydrogen atom from a carbon-hydrogen bond of the substrate, a process characterized by a large and strongly temperature-dependent KIE. The hydrogen abstraction step in methylmalonyl-coenzyme A mutase (MMCoAM) has been studied by Chowdhury and Banerjee [9]; the kinetic data relevant for our purpose are listed in Table 17-3 (Reaction 1). Since the transfer presumably occurs between a methylene group and a methylene radical, we expect the distance parameter ρ to much larger than in the case of SLO1, as discussed in Section 17.3. In view of the implied large transfer

Table 17-3 Kinetic parameters (T in K, k in s^{-1}) used in the analysis of hydrogen atom transfer

Nr.	Reactant	T	ΔT	$\ln k^{H}$	E_a^H/k_BT	$\ln \eta(T)$	$\Delta E_a/k_BT$	Reference
1	MMCoAM	293	15	5.3	31.8	3.8	6.9	9
2	(MeO)AdoCbl	353	40	–	–	2.5	3.5	10
3	NeopentylCbl	292	30	–	–	3.6	5.3	10
4	PheH	306	60	–	–	2.4	3.9	13
5	TyrH	306	60	–	–	2.6	4.7	13

distance, we expect the effective-mode frequency to be low and hence we set $F(\xi) \sim 1$. Furthermore, we assume that the reaction is not endothermic and involves little reorganization, in keeping with the accepted mechanism which suggests that the tunneling step is nonpolar and thermoneutral. Based on these assumptions, we can obtain values for the model parameters ρ and $z(293)$ from Eq. (17-6) and either Eqs. (17-9) and (17-10) or Eqs. (17-11) and (17-13). Although these two pairs of equations are equivalent, the first choice is the best in this case, since it is less dependent on the accuracy of the measured input parameters. The results, $z(\xi) = 2.66$ and $\rho = 160$, and $z(\xi) = 2.75$ and $\rho = 168$, are in excellent mutual agreement, thus supporting the validity of our assumptions. Using the default value $a_0 = 0.105$ Å, we derive an equilibrium transfer distance $r_0 = 1.9$ Å, a value in close agreement with the distance of 1.88 Å measured by Doba et al. for hydrogen transfer between a methyl radical and methanol; it also agrees with the value estimated from the van der Waals radius of 2.0 Å for methyl together with a CH-bond length of 1.08 Å. The results again confirm the assumptions $\Delta E_0 + E_R \sim 0$ and $F(\xi) \sim 1$, implying a frequency $\Omega \leq 100\ cm^{-1}$ for the effective mode. It is worth noting that if we had used Eqs. (17-11) and (17-12) instead of (17-9) and (17-10), we would have obtained two conflicting values for $z(\xi)$ and thus r_0, which can be traced back to the relative inaccuracy of ΔE_a, a value obtained as a small difference between two large numbers. Overall, we conclude that the model provides a good fit to the data, which confirms the rate-limiting character of the tunneling step and its adiabatic transfer mechanism.

A similar reaction, but without the enzyme, has been studied by Doll and Finke [10], whose substrate was the solvent, viz. ethylene glycol; the relevant kinetic data, listed in Table 17-3 as Reactions 2 and 3, are derived from product distributions. Strictly speaking, our equations, which neglect thermal excitation of CH and CD stretching modes, are not applicable to Reaction 2, which involves AdoCbl (as well as MeO-AdoCbl which yields similar results), because of the high temperature. We therefore consider only Reaction 3, in which Ado is replaced by neopentyl, which separates faster from Cbl. If we use Eq. (17-13), set $F(\xi) \sim 1$, and assume a thermoneutral reaction with minimal reorganization, we obtain $z(\xi) \simeq 2.4$ and, after substitution in Eq. (17-6), $\rho \simeq 130$, implying an equilibrium transfer distance

$r_0 = 1.7$ Å. This estimate indicates that the tunneling step is not enhanced by the presence of the enzyme, a conclusion similar to that of Doll and Finke [10].

Reactions 4 and 5 in Table 17-3 concern hydrogen transfer in phenylalanine hydroxylase; the listed kinetic data were measured by Pavon and Fitzpatrick [13] from product distributions. Application of Eq. (17-6) with $\rho \simeq 150$ yields $z(306) \simeq 3.4$ for Reaction 4 and 3.3 for Reaction 5, values that are similar to but somewhat larger than those for Reactions 1 and 3. The implied large transfer distances justify the assumption that $F(\xi) \sim 1$. Together with the relatively large values of $\Delta E_a/k_BT$, they indicate that the reaction proceeds by means of a free radical mechanism. Substitution of the observed values of $\Delta E_a/k_BT$ in Eq. (17-13) leads to $z(\xi) = 3.7$ and 13 for Reaction 4 and 5, respectively. Upon substitution in Eq. (17-6), the first value yields $\rho \simeq 170$, corresponding to a transfer distance $r_0 \simeq 1.9$ Å, similar to that found for coenzyme B_{12}. However, the second value is obviously much too large. In the absence of rate constants and activation energies, we have no means to trace back the cause of this discrepancy. Restricting ourselves therefore to Reaction 4, we propose that the model provides an acceptable fit to the data and suggests that the rate-limiting step corresponds to tunneling of a neutral hydrogen atom between two carbon centers.

17.6. APPLICATION TO METHYLAMINE DEHYDROGENASE

Scrutton and coworkers [11, 12] have studied the oxidative demethylation of primary amines to form aldehydes and ammonia by methylamine dehydrogenase (MADH) and also by aromatic amine dehydrogenase (AADH), which shows much the same behavior. In these reactions a proton is abstracted by an active-site base from an iminoquinone intermediate while an electron is transferred to a tryptophylquinone cofactor. Most of these reactions exhibit KIEs between 5 and 20 at 298 K with a very weak temperature dependence, although the rate constants show activation energies of more than 10 kcal/mol. The authors assumed that the proton-transfer step is rate-determining. However, they report that the kinetics do not follow the Michaelis-Menten equation; specifically the association and dissociation rate constants of the enzyme-substrate complex, denoted by k_1 and k_{-1} in Eq. (17-1), are subject to KIEs of the same general magnitude as k_{cat}. Our earlier efforts [1, 7] to fit the data to a tunneling model were unsuccessful.

Our present analysis starts with the basic substrate methylamine [11], for which a KIE of 16 ($ln\eta = 2.8$) was observed. Application of Eq. (17-6) to the estimated range of transfer distances, represented by $25 \leq \rho \leq 50$, yields $0.76 \leq z(298) \leq 1.43$. These numbers are comparable to those for SLO1 and indicate that $F(\xi)$ will be smaller than unity. Substitution in Eq. (17-11) together with the observed activation energies $E_a^{\text{H}}/k_BT = 17.6$ and $E_a^{\text{D}}/k_BT = 18.2$ yields effective endothermicities in the range $16.6 \geq E^{\text{eff}}/k_BT \geq 15.9$. These numbers are suspect, however, because they give rise to very small activation energies, not only for H transfer (less than $2k_BT$, as in SLO1 reactions, but also for D transfer (less than

$2.5k_BT$. If we substitute the deduced range of $z(T)$ values, interpreted as $z(\xi)$, in Eq. (17-13), we obtain a range of values of $F(\xi)$ of $0.21 - 0.17$, implying an effective-mode frequency Ω well above the proposed limit of 400 cm^{-1}, corresponding to $F(\xi) \sim 0.5$. Use of Eqs. (17-9) and (17-10) instead of (17-11) and (17-12) confirms that there is no combination of acceptable values for E^{eff} and the temperature dependence of the KIE that can reproduce the experimental data. It follows that these data are not representative of the properties of a single tunneling step. This is likely to be related to the observation that the kinetics does not follow the Michaelis-Menten equation (17-1).

Much the same holds for the data reported for a number of AADH reactions [12]. However, the MADH reaction in which methylamine is replaced by ethanolamine [12] is an exception. The latter reaction exhibits the same KIE as the methylamine reaction, implying the same range of $z(298)$ values. While it has nearly the same activation energy for H transfer (17.6 against 17.9 in units of k_BT), its activation energy for D transfer is considerably larger (20.2 against 18.5 in units of k_BT). As a result the value of E^{eff} is much reduced, namely to the range $11.7 \geq E^{\text{eff}}/k_BT \geq 8.6$, implying acceptable values for the activation energies. If their difference is substituted in Eq. (17-13), it yields acceptable values for $F(\xi)$, namely a range of 0.96–0.74, corresponding to a range of 100–280 cm^{-1} for the effective-mode frequency Ω. Hence it appears that a single tunneling step determines the rate of the reaction of ethanolamine, in contrast to that of methylamine. This may be related to the fact that the ethanolamine reaction is considerably slower and thus less likely to suffer interference from isotope-independent steps.

17.7. DISCUSSION

It has been our purpose to show that an investigator of enzymatic CH-cleavage reactions, who is likely to encounter large kinetic isotope effects, need not be limited to the conclusion that the reaction proceeds by quantum-mechanical tunneling. With the use of the proper tools, a good deal of additional information can be gathered about the mechanism of the reactions and the vibrations governing their rates. In this contribution we offer such a tool in the form of an analytical model that is simple to use and can be applied directly to the data in the form in which they are traditionally reported. Our model is meant to replace the Arrhenius equation, which is manifestly invalid for tunneling reactions, and the Bell model, which overlooks the effect of skeletal vibrations on tunneling. To demonstrate the usefulness of the model, we have applied it to a representative set of enzymatic reactions for which KIEs and their temperature dependence are available.

From our analysis of these reactions, we conclude that not all of them are compatible with the model. While the model in the form used here is very basic, it is structured in such a way that incompatibility is readily recognized and can be traced back to (a specific subset of) the data. As the applications show, the data recognized as problematic tend to suffer from other problems as well. In general,

incompatibility serves as a warning that the kinetics of these reactions may not be governed by a single rate-limiting tunneling step, but could, e.g., suffer from interference from slow isotope-independent steps. If the data pass the compatibility test, they can offer information on such quantities as the hydrogen transfer distance, the endothermicity, if any, the contribution of skeletal modes to the transfer, the properties of these modes as well as of the tunneling mode, the absence or presence of hydrogen bonding, and the effect of mutations, if any. To derive these properties from the observed kinetic parameters, some empirical input is required such as the van der Waals radii of the atomic centers between which the hydrogen is transferred and relevant bond lengths, but even with a minimum of input, information can be extracted that cannot be obtained from the Arrhenius equation and the Bell model. None of this requires any computational effort, although the availability of theoretical input or the use of more elaborate versions of the model, where the data warrant it, is likely to improve the accuracy of the results.

REFERENCES

1. Siebrand W, Smedarchina Z (2006) In: Kohen A, Limbach H-H (eds) Isotope effects in chemistry and biology. CRC Press, Baton Rouge
2. Bell RP (1980) The tunnel effect in chemistry. Chapman & Hall, London
3. Siebrand W, Wildman TA, Zgierski MZ (1984) J Am Chem Soc 106:4083–4089; ibid 4089–4095
4. Kuznetsov AM, Ulstrup J (1999) Can J Chem 77:1085–1096
5. Knapp MJ, Rickert KW, Klinman JP (2002) J Am Chem Soc 124:3865–3874
6. Meyer MP, Tomchick DR, Klinman JP (2008) Proc Natl Acad Sci USA 105:1146–1151
7. Siebrand W, Smedarchina Z (2004) J Phys Chem 108:4185–4195
8. Siebrand W, Smedarchina Z (2010) J Phys Org Chem 23: issue 7
9. Chowdhury S, Banerjee R (2000) J Am Chem Soc 122:5417–5418
10. Doll KM, Finke RG (2003) Data Inorg Chem 42:4849–4857
11. Basran JB, Sutcliffe MJ, Scrutton NS (1999) Biochemistry 38:3218–3222
12. Basran JB, Patel S, Sutcliffe MJ, Scrutton NS (2001) J Biol Chem 276:6234–6242
13. Pavon JA, Fitzpatrick PF (2005) J Am Chem Soc 127:16414–16415
14. Smedarchina Z, Siebrand W (1993) Phys Chem 170:347–358
15. Doba T, Ingold KU, Reddoch AH, Siebrand W, Wildman TA (1984) J Chem Phys 8:6622–6630
16. Jacquot C, Wecksler AT, Mcginley CM, Segraves EN, Holman TR, van der Donk WA (2008) Biochemistry 47:7295–7304

CHAPTER 18

TUNNELING TRANSMISSION COEFFICIENTS: TOWARD MORE ACCURATE AND PRACTICAL IMPLEMENTATIONS

RUBÉN MEANA-PAÑEDA AND ANTONIO FERNÁNDEZ-RAMOS
Faculty of Chemistry, Department of Physical Chemistry, University of Santiago de Compostela, 15706 Santiago de Compostela, Spain
e-mail: qf.ramos@usc.es

Abstract: The accurate evaluation of quantum effects is of great importance in many reaction processes. Variational transition state theory with multidimensional tunneling is the natural choice for the study of these reactions, because it incorporates quantum effects through a multiplicative transmission coefficient and it can deal with large systems. Currently, the main approximation used for taking into account tunneling is the small-curvature approximation, mainly because the large curvature and the least-action approximations are computationally very demanding and their use it is usually associated to small systems. Here we describe two algorithms based on splines under tension, which allow the evaluation of these two transmission coefficients for large systems. The analysis of kinetic isotope effects on a model reaction show that the least-action transmission coefficient should be used instead of the more inexpensive, but probably less accurate small-curvature transmission coefficient.

Keywords: Variational transition state theory, Multidimensional tunneling, Transmission coefficient, Semiclassical methods, Least-action path, Small-curvature tunneling, Large-curvature tunneling, Direct dynamics, Spline under tension

18.1. INTRODUCTION

The contribution of the theoretical methods to the understanding of complex reaction mechanisms relies on both, the accuracy of the electronic structure calculations and the reliability of the dynamics models. The electronic structure calculations provide us with the potential energy surface and the thermochemistry of the

P. Paneth and A. Dybala-Defratyka (eds.), *Kinetics and Dynamics*,
Challenges and Advances in Computational Chemistry and Physics 12,
DOI 10.1007/978-90-481-3034-4_18,

reaction, whereas the dynamics models provide us with the kinetic parameters (for instance, thermal rate constants). There are reactions for which the thermochemical calculations are sufficient to disregard a given proposed mechanism, without proceeding further, i.e., it is enough to know the relative stability of all the stationary points (reactants, intermediates and transition states) to decide the reaction path. This is usually the case for reactions in which classical dynamics effects (recrossing) and/or quantum dynamics effects (tunneling) are not significant. Besides, if we need to evaluate the thermal rate constants, conventional transition state theory (CTST) [1, 2], which only requires information at reactants and the transition state, would be adequate. On the other hand, there are systems with a small number of atoms (about six or seven) for which it is now possible to evaluate accurately the thermal rate constants by quantum dynamical methods. However, there is a large number of chemical reactions which do not fall in any of these categories, i.e., they are too big to be treated by quantum dynamics and they present dynamical effects which cannot be handled by CTST. For instance, this is the case for most of the proton transfer reactions.

Proton transfer reactions are probably the most important group of processes that take place in Chemistry and Biology [3, 4]. Protons, because of their light weight, can undergo tunneling, i.e. they can penetrate quite easily through classically forbidden potential energy barriers [5–9]. Therefore, tunneling is a quantum mechanical effect that cannot be taken into account by classical methods (like for instance, classical trajectories). On the other hand, the applicability of quantum dynamics is reduced to small systems, so we have to find alternative methodologies able to deal with tunneling in large systems. Treatments based on semiclassical methods, such as the Wentzel–Kramers–Brillouin (WKB) approximation [10], meet those requirements, they can be used for large systems and include quantum mechanical tunneling using classical concepts. In the WKB approximation tunneling is considered as the penetration through a barrier with negative kinetic energy so the momentum and the action are imaginary. Usually the error introduced by semiclassics regarding to quantum calculations is about 15%.

Variational transition state theory [11–21] with multidimensional corrections for tunneling (VTST/MT) is a semiclassical (WKB like) theory based on CTST, which incorporates both classical dynamics effects (recrossing) and quantum mechanical tunneling. It has been tested against quantum mechanical calculations for several atom–diatom reactions and for the polyatomic $H + CH_4 \rightarrow H_2 + CH_3$ hydrogen abstraction reaction [22]. In general, the comparison showed a very good agreement, so it makes VTST/MT one of the most suitable candidates to study proton transfer reactions. It is not our intention to explain all the aspects related to VTST, which can be looked up in a recent review [21], so hereafter we focus mainly our attention onto the tunneling effect, since for most of the proton transfer reactions is a crucial phenomena.

The most common version of VTST is canonical variational transition state theory (CVT) [14, 23], in which the rate constants are calculated considering

a fixed-temperature ensemble. The CVT/MT thermal rate constant, $k^{\mathrm{CVT/X}}$, is given by

$$k^{\mathrm{CVT/X}}(T) = \kappa^{\mathrm{CVT/X}}(T) k^{\mathrm{CVT}}(T) \tag{18-1}$$

where $k^{\mathrm{CVT}}(T)$ is the CVT thermal rate constant that minimizes the one-way flux from reactants to products through trial dividing surfaces that cross the reaction path at several points. The reaction path is chosen as the minimum energy path (MEP) in isoinertial coordinates [24, 25], scaled to a reduced mass of μ, and the signed distance along this path is labeled as s. By convention $s = 0$ indicates the location of the transition state, whereas $s<0$ and $s>0$ correspond to the reactant and product sides, respectively. The $k^{\mathrm{CVT}}(T)$ rate constant is 'quasiclassical' because involves quantized vibrational partition functions in its evaluation. $\kappa^{\mathrm{CVT/X}}(T)$ is the transmission coefficient, which gives the tunneling contribution in the approximation X to the final rate constant. All the $3N$-7 normal modes perpendicular to the reaction coordinate ($3N$-6 in linear molecules), being N the number of atoms, are quantized (through the vibrational partition functions) and, therefore, the transmission coefficient takes into account the quantum effects associated to the reaction-coordinate motion. The transmission coefficients accompanying Eq. (18-1) are typically evaluated by any of the following approximations: zero-curvature tunneling (ZCT) [24], which neglects the coupling between the reaction coordinate and the transverse modes; small-curvature tunneling (SCT) [26–31], which incorporates such coupling, but considering that tunneling effect is not too large; large-curvature tunneling (LCT) [19, 21, 30, 32–37], which considers straight trajectories as the tunneling paths and it was specially designed for reactions with important quantum effects; the microcanonically optimized multidimensional tunneling (μOMT) [36], which is the best compromise, at every tunneling energy between SCT and LCT and; the least-action tunneling (LAT) [38–40], which samples a given family of paths between the ZCT and LCT paths with the objective of seeking the path that minimizes the imaginary action. Very recently we have extended the range of application of the LAT method to polyatomic reactions [40]. A more detailed description of these methods is presented in Section 18.2.

The LCT and LAT transmission coefficients are computationally quite demanding, which makes SCT the most practical choice for large systems. This technical problem has led to a misuse of the SCT approximation in some cases. An important group of proton transfer processes, which certainly would benefit from a greater availability of the LCT and LAT transmission coefficients, are the enzymatic reactions, because there are many enzymes for which the rate-determining step for reaction is the proton transfer between the enzyme and the substrate. One of the main experimentally determined kinetic parameters to get insight into the mechanism of the enzymatic catalysis is the analysis of intrinsic isotope effects (KIEs). In general, the isotopic substitution is carried out on the transferred atom, so the KIE is the ratio between the thermal rate constant for transfer of the root species (hydrogen) and that for transfer of the heavier isotope (deuterium or tritium). Nowadays it is quite established that quantum effects are of great importance in

enzymatic proton transfer [41–45], due mainly to the observation of abnormally large KIEs. This is usually considered one of the signatures of tunneling, since protons are lighter and pass through the barrier easier than the heavier isotopes. In general, although not always [46–49], we expect tunneling to be more important for hydrogen than for deuterium [50]. We also expect LCT to be a better approximation for the hydrogen transfer than SCT, at least in some cases. Pang et al. [51] found that tunneling dominates the proton transfer reaction from nicotinamide adenine dinucleotide (NADH) to enzyme-bound flavin mononucleotide (FMN) in the flavoprotein morphinose reductase. They also found that LCT described tunneling more adequately than SCT for the proton transfer, whereas SCT was more adequate for the deuteron transfer. This example shows the importance of making LCT and LAT computationally more accessible. This issue will be discussed in Section 18.3.

Finally in Section 18.4 we discuss briefly with an example the influence of the choice of the tunneling transmission coefficient over the computed KIEs.

18.2. TUNNELING TRANSMISSION COEFFICIENTS

In VTST, the tunneling transmission coefficient [24, 52, 53] is evaluated as the ratio between the semiclassical adiabatic ground state (SAG) probability and the quasiclassical probability:

$$\kappa^{\mathrm{SAG}}(T) = \frac{\int_0^\infty dE P^{\mathrm{SAG}}(E)\exp(-\beta E)}{\int_{V_{\mathrm{a}}^{\mathrm{AG}}}^\infty dE P^{\mathrm{C}}(E)\exp(-\beta E)} \tag{18-2}$$

where $P^{\mathrm{C}}(E)$ is the classical probability, which equals zero below $V_{\mathrm{a}}^{\mathrm{AG}}$ (the maximum of the vibrationally adiabatic potential) and unity otherwise, so the transmission coefficient can be written as:

$$\kappa^{\mathrm{SAG}}(T) = \beta\exp(\beta V_{\mathrm{a}}^{\mathrm{AG}})\int_0^\infty dE P^{\mathrm{SAG}}(E)\exp(-\beta E) \tag{18-3}$$

As shown in Eqs. (18-2) and (18-3), the tunneling transmission coefficient is evaluated by using an effective potential that in the first approximation, is vibrationally adiabatic and harmonic with the further approximation [15] that the vibrationally adiabatic potential curves of all the vibrational excited states have the same shape as the ground-state vibrationally adiabatic potential curve, $V_{\mathrm{a}}^{\mathrm{G}}(s)$, so all the tunneling probabilities are evaluated with this potential, which is given by

$$V_{\mathrm{a}}^{\mathrm{G}}(s) = V_{\mathrm{MEP}}(s) + \frac{\hbar}{2}\sum_m \omega_m(s) \tag{18-4}$$

where $\omega_m(s)$ is the frequency of one of the $3N$-7 generalized normal modes at s. The semiclassical adiabatic probability of the ground state for tunneling $P^{\mathrm{SAG}}(E)$ is given by the semiclassical-WKB relation

$$P^{\mathrm{SAG}}(E) = \frac{1}{1+\exp[2\theta(E)]} \tag{18-5}$$

where $\theta(E)$ is the imaginary action integral:

$$\theta(E) = \hbar^{-1} \int_{\tilde{s}_0}^{\tilde{s}_1} ds\{2\mu_{\text{eff}}(s)(V_{\text{a}}^{\text{G}}(s) - E)\}^{1/2} \tag{18-6}$$

where $\mu_{\text{eff}}(s)$ is the effective mass of the tunneling motion, and $\tilde{s}_0$ and $\tilde{s}_1$ are the classical turning points at a given tunneling energy, E, in the reactant and product valleys, respectively. Both turning points obey the resonance condition:

$$V_{\text{a}}^{\text{G}}(\tilde{s}_0) = V_{\text{a}}^{\text{G}}(\tilde{s}_1) = E \tag{18-7}$$

The transmission coefficient does not include only tunneling but also the non-classical reflection above the barrier. Thus, the semiclassical probability for the whole range of energies is given by

$$P^{\text{SAG}}(E) = \begin{cases} 0, & E < E_0 \\ \{1 + \exp[2\theta(\text{E})]\}^{-1}, & E_0 \le E \le V_{\text{a}}^{\text{AG}} \\ 1 - \text{P}^{\text{SAG}}(2V_{\text{a}}^{\text{AG}} - E), & V_{\text{a}}^{\text{AG}} \le E \le 2V_{\text{a}}^{\text{AG}} - E_0 \\ 1, & 2V_{\text{a}}^{\text{AG}} - E_0 < E \end{cases} \tag{18-8}$$

where E_0 is the lowest energy at which it is possible to have tunneling.

When Eq. (18-6) is evaluated within the ZCT [24] approximation, there is no coupling between the reaction coordinate and the transverse modes and therefore the tunneling path follows the MEP. From the classical point of view the MEP is the most favorable path but the longest, because usually it involves a substantial motion of the heavy atoms which readjust their positions. However, the best tunneling path is the one that minimizes the imaginary action (maximizes the tunneling probability), i.e., the one with the best compromise between energy and length of the path. Currently, this transmission coefficient is calculated only for comparative purposes, because it has been shown that seriously underestimates the tunneling contribution [54, 55]. Marcus and Coltrin [26, 27] showed that the coupling between the reaction coordinate and the transverse modes shortens the tunneling path increasing the tunneling probability. They showed, that for the collinear $H + H_2$ reaction, the coupling curves the reaction path in such a way that the least-action path is obtained. The coupling can be incorporated as an effective mass into Eq. (18-6). Unfortunately the method could only handle cases in which the reaction path curvature was small, because when the reaction path curvature is large there is a singularity in the reaction path Hamiltonian due to the breakdown of the natural collision coordinates. Truhlar and coworkers [28–31] modified the method (i.e. the effective mass) developing an analytical expression that avoids the singularity and extends its range of application to polyatomic reactions. This method is called the centrifugal-dominant small-curvature semiclassical adiabatic ground-state (CD-SCSAG) approximation or in abbreviated form the small-curvature tunneling (SCT) approximation. It should be noticed that the effective mass obtained by SCT is always smaller than (considers coupling) or equal to (neglects coupling) the mass of the transferred particle.

For a reaction of the type A + BC → AB + C, where A, B and C are atoms or groups of atoms, the reaction path curvature is a function of the skew angle, which in isoinertial coordinates (coordinates in which the kinetic energy consists of diagonal square terms) is given by

$$\beta = \cos s^{-1}\left(\frac{m_A m_C}{(m_A + m_B)(m_B + m_C)}\right)^{1/2} \tag{18-9}$$

The skew angle is close to zero when the masses of A and C are much larger than the mass of B and it is close to 90° when B has a much larger mass than A and C. Tunneling effects are more important for small skew angles because a light particle is being transferred between two heavy atoms (i.e. heavy–light–heavy system) [56]. In those cases tunneling may be dominated by paths that lie very far from the MEP and the adiabatic approximation may breakdown. The large-curvature ground-state tunneling (LCT) method [19, 21, 30, 32–37] considers the extreme case in which tunneling is dominated at every tunneling energy by paths that are straight trajectories between the two classical turning points. The geometries along a given straight path are given by

$$\mathbf{x}[\xi, \tilde{s}_0] = \mathbf{x}(\tilde{s}_0) + \frac{\xi}{\xi_P}[\mathbf{x}(\tilde{s}_1) - \mathbf{x}(\tilde{s}_0)] \tag{18-10}$$

where $\mathbf{x}(\tilde{s}_0)$ and $\mathbf{x}(\tilde{s}_1)$ are the MEP geometries at the classical turning points. The parameter ξ indicates the progress along the straight path. The length of the path is given by

$$\xi_P = |\mathbf{x}(\tilde{s}_1) - \mathbf{x}(\tilde{s}_0)|. \tag{18-11}$$

The evaluation of the action integrals along these paths requires not only information of the potential valley around the MEP, which can be treated within the adiabatic approximation, but also information about the reaction swath on the concave-side of the MEP, which is vibrationally nonadiabatic. The vibrationally adiabatic region in the reactants and products sides is labeled as I and III, respectively, whereas the nonadiabatic region is labeled as II (see Figure 18-1). The vibrationally adiabatic potential is obtained in such a way that the vector $\mathbf{x}[\xi, \tilde{s}_0] - x(\tilde{s}_0)$ is perpendicular to the gradient at that s value, i.e.,

$$\{\mathbf{x}[\xi, \tilde{s}_0] - \mathbf{x}(\tilde{s}_0)\}\frac{d\mathbf{x}(\tilde{s}_0)}{ds} = 0 \tag{18-12}$$

If it is not possible to find a geometry along the MEP which satisfies Eq. (18-12), it means that there is no projection for that geometry of the tunneling path onto the modes perpendicular to the reaction path in the interval $[\tilde{s}_0, \tilde{s}_1]$. When this happens going from reactants to products, the nonadiabatic region in the reactants side starts at $\xi = \xi_I$, being ξ_I the last value for which there is projection onto the MEP. That point on the MEP is labeled as $s_I[\xi, \tilde{s}_0]$. In the same way, the nonadiabatic region in products side starts at $\xi = \xi_{III}$, being ξ_{III} the last value for which $s_{III}[\xi, \tilde{s}_0]$ exists. If there is an overlap between the adiabatic regions and the nonadiabatic region does

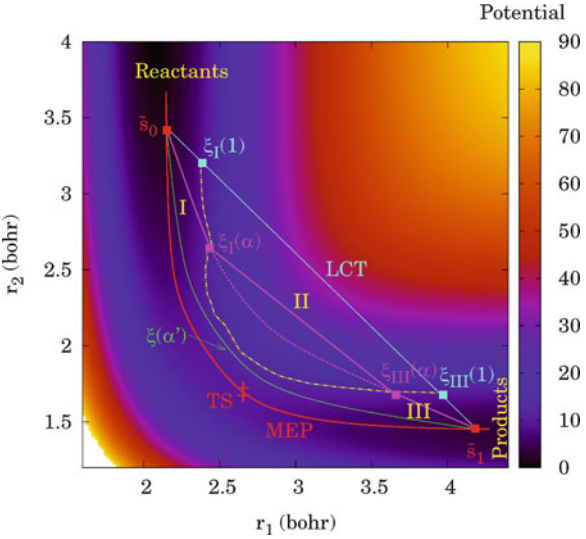


Figure 18-1. Generic contour plot of the PES of a bimolecular reaction as a function of to distances r_1 and r_2. The dashed-dotted yellow line divides the adiabatic region (labeled as I for reactants and III for products) from the non-adiabatic region (labeled as II). Four different tunneling paths are plotted: (i) The path with $\alpha = 0$, (red solid line) corresponds to the MEP. (ii) The path with $\alpha = 0.12$ (green solid line, labeled as $\xi(\alpha')$) corresponds to a curved path passing through a region, which is completely vibrationally adiabatic. (iii) The path with $\alpha = 0.50$ (magenta solid line, labeled as $\xi(\alpha)$) corresponds to a curved path which crosses the nonadiabatic region with boundaries given by $\xi_I(\alpha)$ in the reactants side and by $\xi_{III}(\alpha)$ in the products side. This path does not follow the curved path (magenta dotted line) but the straight path (magenta solid line) through the nonadiabatic region. (iv) The path with $\alpha = 1$, (cyan solid line) corresponds to the LCT. The boundaries of the nonadiabatic region are given by ξ_I (1) in the reactants side and by ξ_{III} (1) in the products side

not exists, i.e., $\xi_{III} < \xi_I$, the vibrationally adiabatic potential in the interval $[\xi_{III}, \xi_I]$ is evaluated as:

$$\min\left\{V_a^G[s_I(\xi);\tilde{s}_0], V_a^G[s_{III}(\xi);\tilde{s}_0)]\right\} \tag{18-13}$$

Each of the two s_i, $i = I, III$ values needed for the evaluation of the vibrationally adiabatic potentials $V_a^G[s_i(\xi);\tilde{s}_0)]$ is obtained from Eq. (18-12).

The nonadiabatic region has boundaries ξ_I and ξ_{III} with regions I and III, respectively. Therefore, region I is located in the interval $0 \leq \xi < \xi_I$, region II in the interval $\xi_I \leq \xi \leq \xi_{III}$, and region III in the interval $\xi_{III} < \xi \leq \xi_P$. Besides of the condition (i), given by Eq. (18-12), we consider that a given geometry is in the adiabatic region when (ii) all the generalized normal mode coordinates are within their vibrational turning points, (iii) the geometry $\mathbf{x}[\xi, \tilde{s}_0]$ lies within the single-valued region of the reaction path coordinates, and (iv) the adiabatic potential should be greater than or equal to the effective potential at the boundary of the nonadiabatic region, in other case the nonadiabatic region is extended until this condition is met. On the other hand, if the adiabatic potential is smaller than the

effective potential the difference is due to anharmonicity, so the effective potential is modified with a nonquadratic correction. Condition (iv) entails the difference between the large curvature ground-state version 3 (LCG3) and large curvature ground-state version 4 (LCG4) [21, 37] methods. The previous LCG3 method usually involved abrupt changes in the effective potential at the boundaries between the adiabatic and nonadiabatic regions, leading to spuriously large transmission coefficients at low temperatures [57–59]. The effective potential in the LCG4 approximation is given by

$$V_{\mathrm{eff}}^{\mathrm{II}}(\xi,\tilde{s}_0) = V\{\mathbf{x}[\xi,\tilde{s}_0]\} + V_{\mathrm{corr}}^{\mathrm{I}}(\xi_{\mathrm{I}},\tilde{s}_0) + V_{\mathrm{anh}}^{\mathrm{I}}(\xi_{\mathrm{I}},\tilde{s}_0) + \frac{\xi-\xi_{\mathrm{I}}}{\xi_{\mathrm{III}}-\xi_{\mathrm{I}}}[V_{\mathrm{corr}}^{\mathrm{III}}(\xi_{\mathrm{III}},\tilde{s}_0) - V_{\mathrm{corr}}^{\mathrm{I}}(\xi_{\mathrm{I}},\tilde{s}_0) + V_{\mathrm{anh}}^{\mathrm{III}}(\xi_{\mathrm{III}},\tilde{s}_0) - V_{\mathrm{anh}}^{\mathrm{I}}(\xi_{\mathrm{I}},\tilde{s}_0)] \tag{18-14}$$

The potentials $V_{\mathrm{corr}}^{i}(\xi_i,\tilde{s}_0),\ i=\mathrm{I},\mathrm{III}$ correct for the zero-point energy in the modes that are still within their turning points. The potentials $V_{\mathrm{anh}}^{i}(\xi_i,\tilde{s}_0)$ incorporate anharmonic nonquadratic corrections to the effective potential, i.e.,

$$V_{\mathrm{anh}}^{i}(\xi_i,\tilde{s}_0) = V_{\mathrm{a}}^{\mathrm{G}}[(s_i(\xi);\tilde{s}_0)] - V_{\mathrm{eff}}^{\mathrm{II}}(\xi_i,\tilde{s}_0),\ i=\mathrm{I},\mathrm{III} \tag{18-15}$$

The LCG4 imaginary action integral at every tunneling energy along the straight paths given by Eq. (18-10) is:

$$\theta(\tilde{s}_0) = \frac{(2\mu)^{1/2}}{\hbar}\left[\int_0^{\xi_{\mathrm{I}}} d\xi \left\{V_{\mathrm{a}}^{\mathrm{G}}[(s_{\mathrm{I}};\tilde{s}_0)] - V_{\mathrm{a}}^{\mathrm{G}}(\tilde{s}_0)\right\}^{1/2} \cos\chi_0\{\tilde{s}_0,\hat{\boldsymbol{\eta}}[\xi,\tilde{s}_0]\} + \int_{\xi_{\mathrm{I}}}^{\xi_{\mathrm{III}}} d\xi \left\{V_{\mathrm{eff}}^{\mathrm{II}}(\xi,\tilde{s}_0) - V_{\mathrm{a}}^{\mathrm{G}}(\tilde{s}_0)\right\}^{1/2} + \int_{\xi_{\mathrm{III}}}^{\xi_{\mathrm{P}}} d\xi \left\{V_{\mathrm{a}}^{\mathrm{G}}[(s_{\mathrm{III}};\tilde{s}_0)] - V_{\mathrm{a}}^{\mathrm{G}}(\tilde{s}_0)\right\}^{1/2} \cos\chi_1\{\tilde{s}_1,\hat{\boldsymbol{\eta}}[\xi,\tilde{s}_1]\}\right], \tag{18-16}$$

The two cosines of Eq. (18-16), $\cos\chi_i\{\tilde{s}_i,\hat{\boldsymbol{\eta}}[\xi,\tilde{s}_i]\}$ for $i=0,1$ are obtained as the dot products between the unit vectors $\hat{\boldsymbol{\eta}}[\xi,\tilde{s}_i] = \frac{\mathbf{x}(\tilde{s}_1)-\mathbf{x}(\tilde{s}_0)}{\xi_{\mathrm{P}}}$. and the unit vectors tangent to the MEP at s_{i}, i.e.,

$$\cos\chi_i\{\tilde{s}_i,\hat{\boldsymbol{\eta}}[\xi,\tilde{s}_i]\} = \hat{\boldsymbol{\eta}}[\xi,\tilde{s}_i]\frac{\mathrm{d}\mathbf{x}(\tilde{s}_i)}{\mathrm{d}s},\ i=0,1 \tag{18-17}$$

The tunneling amplitude of the LCG4 path initiated at the reactants turning point $\tilde{s}_0$ is approximated using a primitive semiclassical expression

$$T_{\mathrm{tun}}^{\mathrm{LCG4}}(\tilde{s}_0) = T_{\mathrm{tun}}^{\mathrm{LCG4}}(\tilde{s}_1) = \exp[-\theta(\tilde{s}_0)] \tag{18-18}$$

The LCG4 primitive probability, $P_{\mathrm{prim}}^{\mathrm{LCG4}}(E)$, is obtained from the tunneling amplitude of the previous Eq. (18-18) plus the contribution due to the vibrational motion perpendicular to the reaction coordinate along the incoming $T_0(E)$ and outgoing $T_1(E)$ trajectories at tunneling energy E

$$P_{\text{prim}}^{\text{LCG4}}(E) = |T_0(E) + T_1(E)|^2 + \left(\frac{\cos\chi_0\{\tilde{s}_0, \hat{\boldsymbol{\eta}}[\xi, \tilde{s}_0]\} + \cos\chi_1\{\tilde{s}_1, \hat{\boldsymbol{\eta}}[\xi, \tilde{s}_1]\}}{2}\right)^2 \times \exp[-2\theta(\tilde{s}_0)] \tag{18-19}$$

The tunneling amplitude of the incoming trajectory along the reaction coordinate is $\exp[-\theta(\tilde{s}_0)]\cos\chi_0\{\tilde{s}_0, \hat{\boldsymbol{\eta}}[\xi, \tilde{s}_0]\}$ and that of the outgoing trajectory is $\exp[-\theta(\tilde{s}_0)]\cos\chi_1\{\tilde{s}_1, \hat{\boldsymbol{\eta}}[\xi, \tilde{s}_1]\}$. To enforce macroscopic reversibility, the expression used in Eq. (18-19) is the average of the two tunneling amplitudes and the amplitude due to all the vibrational degrees of freedom perpendicular to the reaction coordinate is averaged using $T_0(E)$ and $T_1(E)$. The expressions for $T_0(E)$ and $T_1(E)$ are given in [21].

The primitive probability of Eq. (18-19) has to be uniformized prior to its use in Eq. (18-8) because it can be greater than one, so it is enforced to go to 1/2 at the maximum of the vibrational adiabatic potential V_{a}^{AG} by the expression:

$$P^{\text{LCG4}}(E) = \left\{1 + \frac{1}{2}\frac{\left[P_{\text{prim}}^{\text{LCG4}}(V_{\text{a}}^{\text{AG}})\right]^{-1} - 1}{P_{\text{prim}}^{\text{LCG4}}(V_{\text{a}}^{\text{AG}})} P_{\text{prim}}^{\text{LCG4}}(E)\right\} \times \frac{1}{1 + \left[P_{\text{prim}}^{\text{LCG4}}(E)\right]^{-1}} \tag{18-20}$$

In the LCG4 method the tunneling path starts and ends as vibrationally adiabatic but during the tunneling process the potential can be nonadiabatic and tunneling may end up in vibrationally excited states. The contribution of those final vibrationally excited states (considering always the reaction in the exoergic direction) is included by finding all accessible vibrational states of a projected mode defined as linear combination of all the generalized normal modes coupled to the tunneling path [30, 33, 35, 36, 60].

The LCG4 approximation is a particular case of a more general method that minimizes the action integral for a family of α dependent paths for which the MEP is located at $\alpha = 0$ and the LCT path at $\alpha = 1$ (hereafter we indicate explicitly the dependence with α of the progress variable along the straight path and use $\xi(1)$ instead of ξ). This method is called least-action tunneling (LAT) and since it is based on the LCG4 approximation is called least-action tunneling ground-state version 4 (LAG4) approximation. The family of paths used for the search of the least-action at every tunneling energy is given by

$$\mathbf{x}[\alpha, \xi(\alpha), \tilde{s}_0] = (1 - \alpha)\mathbf{x}[0, \xi(0), \tilde{s}_0] + \alpha\mathbf{x}[1, \xi(1), \tilde{s}_0] \tag{18-21}$$

where $\mathbf{x}[0, \xi(0), \tilde{s}_0]$ and $\mathbf{x}[1, \xi(1), \tilde{s}_0]$ are the geometries along the MEP and along the straight path, respectively. It should be noticed that the progression variable ξ depends on the value of the α parameter.

The LAG4 imaginary action integral at every tunneling energy and for each α value along the paths of Eq 18-21 is:

$$\theta(\alpha,\tilde{s}_0)=\frac{(2\mu)^{1/2}}{\hbar}\left[\int_0^{\xi_{\rm I}(\alpha)} d\xi(\alpha)\left\{V_{\rm a}^{\rm G}[s_{\rm I}(0,\xi(0));\tilde{s}_0]-V_{\rm a}^{\rm G}(\tilde{s}_0)\right\}^{1/2}\right.$$
$$\times\cos\chi_0\{\tilde{s}_0,\hat{\boldsymbol{\eta}}[\alpha,\xi(\alpha),\tilde{s}_0]\}+\int_{\xi_{\rm I}(\alpha)}^{\xi_{\rm III}(\alpha)} d\xi(\alpha)\left\{V_{\rm eff}^{\rm II}(\alpha,\xi(\alpha),\tilde{s}_0)-V_{\rm a}^{\rm G}(\tilde{s}_0)\right\}^{1/2}$$
$$\left.+\int_{\xi_{\rm III}(\alpha)}^{\xi_{\rm P}(\alpha)} d\xi(\alpha)\left\{V_{\rm a}^{\rm G}[s_{\rm III}(0,\xi(0));\tilde{s}_0]-V_{\rm a}^{\rm G}(\tilde{s}_0)\right\}^{1/2}\cos\chi_1\{\tilde{s}_1,\hat{\boldsymbol{\eta}}[\alpha,\xi(\alpha),\tilde{s}_1]\}\right], \tag{18-22}$$

Equation (18-22) is similar to Eq. (18-16) with the difference that the action integral is not restricted to the straight paths.

The effective potential of Eq. (18-22) of the LAG4 method is the same as that of the LCG4 method, with the difference that now the geometries at which the potential is evaluated are functions of α and of the progress variable ξ, which also depends on α. Therefore the effective potential is given by

$$V_{\rm eff}^{\rm II}(\alpha,\xi(\alpha),\tilde{s}_0)=V\{\mathbf{x}[\alpha,\xi(\alpha),\tilde{s}_0]\}+V_{\rm corr}^{\rm I}(\alpha,\xi_{\rm I}(\alpha),\tilde{s}_0)+V_{\rm anh}^{\rm I}(\alpha,\tilde{s}_0)+\frac{\xi(\alpha)-\xi_{\rm I}(\alpha)}{\xi_{\rm III}(\alpha)-\xi_{\rm I}(\alpha)}$$
$$\times\left[V_{\rm corr}^{\rm III}(\alpha,\xi_{\rm III}(\alpha),\tilde{s}_0)-V_{\rm corr}^{\rm I}(\alpha,\xi_{\rm I}(\alpha),\tilde{s}_0)+V_{\rm anh}^{\rm III}(\alpha,\tilde{s}_0)-V_{\rm anh}^{\rm I}(\alpha,\tilde{s}_0)\right] \tag{18-23}$$

The potentials $V_{\rm corr}^{i}(\alpha,\xi_i(\alpha),\tilde{s}_0),\ i=\mathrm{I},\mathrm{III}$ correct for the zero-point energy in the modes that are still within their turning points and the potentials $V_{\rm anh}^{i}(\alpha,\tilde{s}_0)$ incorporate anharmonic nonquadratic corrections to the effective potential in the same way as in the LCG4 method.

As shown in Figure 18-1, in the case that the whole path is adiabatic, which it may happen for small values of α, the path given by Eq. (18-21) is the reference path between the two classical turning points. However, if there is a nonadiabatic-region (region II), the path along regions I ($0\le\xi(\alpha)<\xi_{\rm I}(\alpha)$) and III ($\xi_{\rm III}(\alpha)<\xi(\alpha)\le\xi_{\rm P}(\alpha)$) is evaluated along the reference path given by Eq. (18-21), whereas the path along region II ($\xi_{\rm I}(\alpha)\le\xi(\alpha)\le\xi_{\rm III}(\alpha)$) is built as a straight path between the boundaries, where the geometries are given by

$$\mathbf{x}[\alpha,\xi(\alpha),\tilde{s}_0]=\mathbf{x}[\alpha,\xi_{\rm I}(\alpha),\tilde{s}_0]+\frac{\xi(\alpha)-\xi_{\rm I}(\alpha)}{\xi_{\rm III}(\alpha)-\xi_{\rm I}(\alpha)}\left(\mathbf{x}[\alpha,\xi_{\rm III}(\alpha),\tilde{s}_0]-\mathbf{x}[\alpha,\xi_{\rm I}(\alpha),\tilde{s}_0]\right) \tag{18-24}$$

The LAG4 path is the optimum path from the family of paths given by Eq. (18-21), i.e., it is the one that minimizes the imaginary action integral of Eq. (18-22). The optimum value of α at every tunneling energy, which we label as $\tilde{\alpha}$, is found through a search in which different initial $\theta(\alpha_i,\tilde{s}_0)$ values are calculated for a given number K of equally spaced points ($i=1,\ldots,K$). We have found out that $K=11$ is a good starting guess. In the grid of K points we search for

the values of the imaginary action which are minima, i.e., the values $\theta(\alpha_j, \tilde{s}_0)$ with $j = 1, \ldots, K_m$, being K_m the number of minima. If there is only one minimum ($K_m = 1$), we limit the search to the interval $[\alpha_{i-1}, \alpha_{i+1}]$, where $\alpha_j = \alpha_i$ is inside the interval. From the three initial values $\{\alpha_{i-1}, \alpha_j, \alpha_{i+1}\}$, two new values are obtained as $\alpha_L = \frac{1}{2}[\alpha_i - \alpha_{i-1}]$ and $\alpha_R = \frac{1}{2}[\alpha_{i+1} - \alpha_i]$ and two new values of the imaginary action are calculated $\theta(\alpha_L, \tilde{s}_0)$ and $\theta(\alpha_R, \tilde{s}_0)$. Of the set of five resulting α values $\{\alpha_{i-1}, \alpha_L, \alpha_j, \alpha_R, \alpha_{i+1}\}$, we take the three new values of α from the values of the imaginary actions $\{\theta(\alpha_{i-1}, \tilde{s}_0), \theta(\alpha_L, \tilde{s}_0), \theta(\alpha_j, \tilde{s}_0), \theta(\alpha_R, \tilde{s}_0), \theta(\alpha_{i+1}, \tilde{s}_0)\}$. There are the following possibilities: (a) if $\theta(\alpha_L, \tilde{s}_0) < \theta(\alpha_R, \tilde{s}_0)$ there are two possibilities: (a1) if $\theta(\alpha_L, \tilde{s}_0) < \theta(\alpha_j, \tilde{s}_0)$ then the new set of three values of α is $\{\alpha_{i-1}, \alpha_L, \alpha_j\}$; (a2) if $\theta(\alpha_L, \tilde{s}_0) > \theta(\alpha_j, \tilde{s}_0)$ then $\{\alpha_L, \alpha_j, \alpha_R\}$; (b) if $\theta(\alpha_L, \tilde{s}_0) > \theta(\alpha_R, \tilde{s}_0)$ there are also two possibilities: (b1) if $\theta(\alpha_R, \tilde{s}_0) < \theta(\alpha_j, \tilde{s}_0)$ then $\{\alpha_j, \alpha_R, \alpha_{i+1}\}$; (b2) if $\theta(\alpha_R, \tilde{s}_0) > \theta(\alpha_j, \tilde{s}_0)$ then $\{\alpha_L, \alpha_j, \alpha_R\}$. The new three values of α substitute the initial values and the process is repeated. Convergence is achieved when two successive loops give differences in α smaller than 10^{-6} or differences in the imaginary action smaller than 5×10^{-5}. Usually the number of iterations is smaller than 15. If there is more than one minima ($K_m > 1$), the above iterative process is repeated for each of them. Of all the obtained local minima, we choose the absolute minimum, i.e. the lowest value of the imaginary action, which is labeled as $\theta(\tilde{\alpha}, \tilde{s}_0)$.

The next step is to calculate the LAG4 primitive probabilities, which are obtained in the same way than the LCG4 ones, but in this case considering the optimized path with $\alpha = \tilde{\alpha}$ instead of the path with $\alpha = 1$. The evaluation of the LAG4 transmission coefficients involves an extra computational cost when compared with the evaluation of the LCG4 ones, because, at every tunneling energy, we have to search for the best tunneling path.

In principle, an approximate way of obtaining better transmission coefficients than the SCT and LCT ones, but avoiding the explicit evaluation of the LAT transmission coefficients, is to built a transmission coefficient which, at every tunneling energy, chooses the largest between the SCT and LCT tunneling probabilities (the smallest between the SCT and LCT imaginary action integrals). This approximation is called microcanonical optimized multidimensional tunneling (μOMT) method [36]. The μOMT tunneling probabilities are, therefore,

$$P^{\mu\text{OMT}}(E) = \max_E \begin{cases} P^{\text{SCT}}(E) \\ P^{\text{LCT}}(E) \end{cases} \tag{18-25}$$

where $P^{\text{SCT}}(E)$ and $P^{\text{LCT}}(E)$ are the SCT and LCT probabilities evaluated within the CD-SCSAG and LCG4 approximations, respectively.

18.3. PRACTICAL IMPLEMENTATION OF THE LCG4 AND LAG4 METHODS

One of the main problems for the applicability of LCG4 and LAG4 methods to large systems is their computational cost. Both methods use the harmonic approximation to build the potential along the tunneling path in the adiabatic region from the

information along the MEP. In the nonadiabatic region, extrapolation from the MEP is not possible, so we have to compute the values of the effective potential given by Eq. (18-14) in LCG4 and by Eq. (18-23) in LAG4, respectively. The procedure for reducing the computer time is different for LCG4 than for LAG4, so we discuss them separately.

The LCG4 transmission coefficients are calculated by numerical integration of Eq. (18-2), which involves the evaluation of about 80 tunneling energies; at every of these tunneling energies the calculation, also by numerical integration, of the imaginary action integral (and therefore of the tunneling probability), requires about 180 single-point energy calculations of the potential along the straight path, involving a total of 14,400 points (hereafter we use 'point' as an abbreviated form of 'single-point-energy'). Of that large amount of points, a number of them, which in general may oscillate between 1,000 and 5,000 (depending on the system), are in the nonadiabatic region. At every tunneling energy the nonadiabatic region along the straight path has boundaries $\xi_{\rm I}(1)<\xi(1)<\xi_{\rm III}(1)$ as shown in Figure 18-1. One way of reducing the computational cost is to evaluate the energies at selected values of $\xi(1)$ along the nonadiabatic region of the straight path and to interpolate the remaining energies. We have found that the fitting of the calculated points to a spline-under-tension function [61, 62] gives very good results, i.e., the calculation of 9 or 11 points at every tunneling energy leads to errors, when compared to the full-LCG4 calculations, of about 3%. This algorithm is called interpolated large curvature tunneling in one dimension (ILCT1D) [63] and it reduces the computer time about five times compared to the full calculation (See Table 2 of reference 63).

A much cheaper and accurate way of evaluating LCG4 transmission coefficients is by interpolation, of not only the energies along the straight paths, but also of the tunneling energies, i.e., all the energies are evaluated within a two-dimensional grid. As for the ILCT1D algorithm, we perform interpolation in the progress variable $\xi(1)$ at every tunneling energy, but we also perform interpolation in the tunneling energies. We called this algorithm interpolated large curvature tunneling in two dimensions(ILCT2D) [64]. The ILCT2D algorithm works as follows: (i) Generation of a $M \times N$ two-dimensional grid at M tunneling energies E_i, with $i = 1, \ldots, M$, in such a way that $E_{i=1}$ is the lowest energy at which tunneling is possible, i.e., E_0 and $E_{i=\rm M}$ is the maximum of the vibrationally adiabatic potential, i.e., $E_M = V_{\rm a}^{\rm AG}$. For every of the E_i tunneling energies we select $\xi_j(1)$ points along the tunneling path, $(j = 1, \ldots, N)$, taking into account that $\xi_1(1) = 0$ and that $\xi_N(1) = \xi_P(1)$, being $\xi_{\rm P}(1)$ the length of the straight path. (ii) Calculation of each of the $(E_i, \xi_j(1))$ points of the grid. (iii) Transformation of the previous grid with points $(E_i, \xi_j(1))$ in a unit-square grid $\left(\overline{E}_i, \bar{\xi}_j(1)\right)$ by performing the scaling:

$$\overline{E}_i = \frac{E_i - E_0}{V_{\rm a}^{\rm AG} - E_0} \quad \text{and} \quad \overline{\xi}_j(1) = \frac{\xi_j(1)}{\xi_P(1)} \tag{18-26}$$

(iv) The grid of points is fitted to a two-dimensional spline under tension following the procedure described by Renka and Cline [65]. (v) The effective potential of Eq. (18-14) is obtained from the fit.

To illustrate how the ILCT2D method performs, we have apply it to the $H + CH_4 \rightarrow H_2 + CH_3$ hydrogen abstraction reaction using the Jordan and Gilbert analytical potential energy surface (JG-PES) [66]. Figure 18-2 shows how the effective potential, which is delimited by the MEP and the straight path at the lowest tunneling energy, is transformed into a unit-square grid. The grid was built with $M = 9$ tunneling energies and $N = 11$ points along each tunneling path; of these 99 points many of them fall in the adiabatic region, so additional calculations are needed only for those points which are inside the nonadiabatic region. It should be noticed that the top side of the square grid is the maximum of the vibrationally adiabatic potential, whereas the left and right sides correspond to the MEP in the reactants and products sides, respectively. The bottom side of the square corresponds to the lowest tunneling energy.

Table 18-1 lists both, the full-LCT and the ILCT2D transmission coefficients, at temperatures in the interval 200–400 K. Comparison between both transmission coefficients shows, that even at the lowest temperature the error is smaller than 1%. Taking into account that the grid was built with $M = 9$ tunneling energies and $N = 11$ points along each tunneling path and that some of these 99 points fall in the adiabatic region, the ILCT2D algorithm is about 30 times faster than the full LCG4 calculation. This huge reduction in computer time shows that LCG4 calculations could became feasible for large systems.

The evaluation of LAG4 transmission coefficients is much more expensive than that of LCG4 transmission coefficients, since at each tunneling energy, the imaginary action has to be evaluated until the optimum value is found. If the LCG4 calculations take about 80 (tunneling energies) $\times 180$ (points along $\xi(1)) = 14,400$, the LAG4 calculation needs that number of computations for each cycle of optimization of α (a close number is 3×10^5). It is possible to apply modified versions of the previous algorithms to the LAG4 method, although the construction of a two dimensional grid in this case involves to choose N_A values α_k, $(k = 1, \ldots, N_A)$, with the condition that $\alpha_1 = 0$ is the MEP and $\alpha_{N_A} = 1$ is the straight path. For each of those α_k paths, we have to select $\xi_j(\alpha_k)$ points along the tunneling path, $(j = 1, \ldots, N)$, where $\xi_1(\alpha_k) = 0$ and $\xi_N(\alpha_k) = \xi_P(\alpha_k)$ (the total length of the tunneling path). Therefore a point inside the grid has coordinates $(\alpha_k, \xi_j(\alpha_k))$. The next step would be to transform this grid into a unit square and search for the value of α which minimizes the action at this tunneling energy. The construction of the two-dimensional grid and the search for $\tilde{\alpha}$ has to be repeated for every of the tunneling energies.

We find the above procedure very complicated, so instead we present here a different algorithm. Initially a number M of tunneling energies E_i $i = 1, \ldots, M$ is chosen with the boundary conditions that $E_{i=1}$ corresponds to the lowest tunneling energy E_0, whereas $E_{i=M}$ corresponds to the latest tunneling energy for which there is a nonadiabatic region along the straight path (here it is different than for the ILCT2D method, for which $E_M = V_a^{AG}$). At every of these tunneling energies a search of $\tilde{\alpha}$ is carried out as indicated in Section 18.2. The difference with the full calculation is that, in this case, the effective potential is evaluated only at some

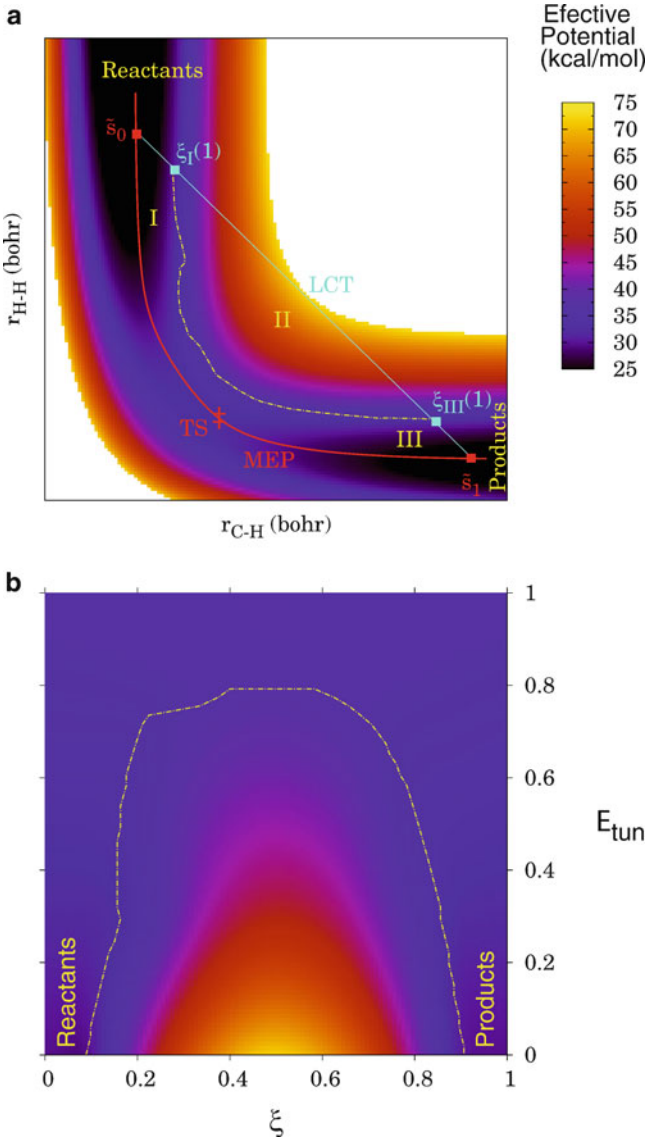


Figure 18-2. (a) Plot of the effective potential for the $H + CH_4 \rightarrow H_2 + CH_3$ hydrogen abstraction reaction; r_{C-H} is the distance between the carbon atom and the abstracted CH_4 hydrogen atom and r_{H-H} is the distance between the hydrogen atom and the abstracted CH_4 hydrogen atom. The solid yellow dotted-dashed line indicates the boundaries between the adiabatic and the non-adiabatic regions. The cyan solid line represents the straight path at a given tunneling energy with classical turning points given by $\tilde{s}_0$ and $\tilde{s}_1$ in the reactants and the products side, respectively. The reaction swath is partitioned into the adiabatic region of reactants (labeled as I), adiabatic region of products (labeled as III), and the nonadiabatic region (labeled as II and with boundaries given by a yellow dotted-dashed line). (b) Unit-square effective potential surface obtained from the (a) PES. The bottom side of the square is the unitary

Table 18-1 Several transmission coefficients obtained for the JG-PES

T(K)	SCT	LCG4	ILCT2D	μOMT	LAG4	DILAT
200	18.74	13.66	13.75	18.76	27.59	27.37
250	6.27	5.11	5.13	6.28	7.34	7.31
300	3.54	3.04	3.05	3.54	3.74	3.74
400	2.02	1.84	1.84	2.03	2.01	2.01

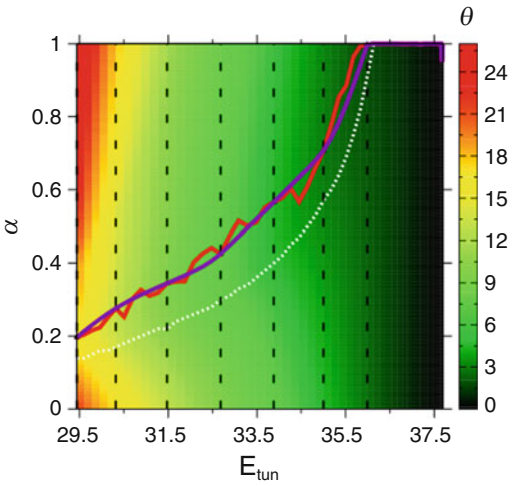


Figure 18.3. Two-dimensional projection showing the variation of the imaginary action integral with the α parameter at every tunneling energy. The red solid line indicates the values of $\tilde{\alpha}$ at every tunneling energy obtained by the full-LAG4 calculation, whereas the violet solid line corresponds to the values of $\tilde{\alpha}$ obtained by DILAT method. The vertical black dashed lines indicate the tunneling energy at which the DILAT method computes the imaginary action prior to the interpolation procedure. The white dotted line indicates the boundaries between the adiabatic and nonadibatic regions

selected $\xi_j(\alpha)$ points ($j = 1, \ldots, N_B$) within the nonadiabatic region of the tunneling path, i.e., the boundaries are $\xi_1(\alpha) = \xi_{\mathrm{I}}(\alpha)$ and $\xi_{N_B}(\alpha) = \xi_{\mathrm{III}}(\alpha)$. The effective potential at every value of α is obtained by a one-dimensional fit to a spline under tension. This process is repeated for each of the E_i tunneling energies, so we end up with a set of M values of $\tilde{\alpha}_i$ and $\theta(\tilde{\alpha}_i)$. Now there is the possibility of fitting the $\tilde{\alpha}_i$ parameters or the imaginary actions $\theta(\tilde{\alpha}_i)$ to a one-dimensional spline under tension. We have found out that the fitting of the imaginary actions to $M = 7$ number of tunneling energies gives very good results. Figure 18-3 shows the very

Figure 18-2 (continued) line given in (a) by the LCT path. The sides of the square represent the vibrationally adiabatic potential along the MEP. The top side of the square is the maximum of the vibrationally adiabatic potential

good agreement between the least imaginary actions obtained by the full-LAG4 calculation (depicted as a black line) and those obtained by the interpolated algorithm (depicted as a green line). The algorithm uses two one-dimensional interpolations, so we call the method double interpolated least-action tunneling (DILAT). In general the number of points that have to be evaluated is about 2,000, which means a reduction of about 150 times regarding to the full-LAG4 calculation. A comparison between the full calculation and the DILAT algorithm is carried out in Table 18-1. Although the DILAT algorithm may need further testing, the difference between both methods is smaller than 1%, which is a very encouraging result.

18.4. TRANSMISSION COEFFICIENTS AND KIES

In this section we analyze the effect that the approximations used for computing tunneling contributions have on the KIEs. As mentioned in the Introduction, this issue is of great relevance because the KIEs are one of the most important parameters for the elucidation of the reaction mechanisms in enzymes. Here we discuss a number of abstraction reactions with several isotopic substitutions of the type $CH_3X + Y \rightarrow CH_3 + XY$, where X = H or D and Y = H, D or T, i.e.,

$$CH_4 + H \rightarrow CH_3 + H_2 \quad (R1)$$
$$CH_3D + H \rightarrow CH_3 + DH \quad (R2)$$
$$CH_4 + D \rightarrow CH_3 + HD \quad (R3)$$
$$CH_3D + D \rightarrow CH_3 + D_2 \quad (R4)$$
$$CH_4 + T \rightarrow CH_3 + HT \quad (R5)$$
$$CH_3D + T \rightarrow CH_3 + DT \quad (R6)$$

All the calculations are based on the JG-PES. Table 18-2 shows that the LAG4 transmission coefficients are always the largest for the hydrogen transfer reaction independent of which is the acceptor atom, whereas the SCT transmission coefficients are always the largest for the deuterium transfer. In principle one would expect the LAG4 transmission coefficients to be the largest for any isotope, although is difficult to decide which of the two methods (SCT or LAG4) is more accurate if the curvature is not too large.

The skew angles given by Eq. (18-9) and listed in Table 18-3 range from small curvature for reaction R2 to large curvature for reaction R5. However, the LCG4 transmission coefficients always underestimate tunneling with respect to the LAG4 transmission coefficients independently of the reaction path curvature, although the LCG4 values are larger than the SCT values for the hydrogen abstractions. The μOMT transmission coefficient takes into account that LCG4 is larger for the hydrogen abstraction and that SCT is larger for the deuterium abstraction, so it seems to solve the problem. However, the KIEs calculated by the different tunneling approximations, and listed in Table 18-4, show that the μOMT transmission

Table 18-2 Transmission coefficients computed with the SCT, LCG4 and LAG4 approximations for reactions R1 to R6 using the JG-PES

	κ_H				κ_D			
T(K)	SCT	LCG4	μOMT	LAG4	SCT	LCG4	μOMT	LAG4
		$CH_4 + H$				$CH_3D + H$		
200	18.74	13.66	18.76	27.59	8.19	4.75	8.19	5.40
250	6.27	5.11	6.28	7.34	3.54	2.47	3.54	2.55
300	3.54	3.04	3.54	3.74	2.34	1.81	2.34	1.83
400	2.02	1.84	2.03	2.01	1.59	1.37	1.59	1.38
		$CH_4 + D$				$CH_3D + D$		
200	6.71	9.82	9.82	16.06	6.23	4.03	6.23	5.14
250	3.36	4.25	4.25	5.70	3.12	2.29	3.12	2.51
300	2.33	2.74	2.74	3.32	2.18	1.73	2.18	1.80
400	1.61	1.77	1.77	1.96	1.54	1.34	1.54	1.36
		$CH_4 + T$				$CH_3D + T$		
200	3.64	8.24	8.24	12.49	4.65	3.70	4.65	4.56
250	2.25	3.73	3.73	4.82	2.63	2.22	2.63	2.42
300	1.74	2.47	2.47	2.93	1.95	1.71	1.95	1.79
400	1.36	1.66	1.66	1.82	1.45	1.34	1.45	1.36

Table 18-3 Reactions studied with the JG-PES. The skew angle given by Eq. (18.9) is given in the second column. The largest transmission coefficients for each of the reactions at $T =$ 300 K is indicated in the third column

Reaction	β	κ
$CH_4 + H \rightarrow CH_3 + H_2$	46.8	LAG4
$CH_3D + H \rightarrow CH_3 + DH$	57.2	μOMT(SCT)[a]
$CH_4 + D \rightarrow CH_3 + HD$	37.8	LAG4
$CH_3D + D \rightarrow CH_3 + D_2$	48.4	μOMT(SCT)
$CH_4 + T \rightarrow CH_3 + HT$	33.1	LAG4
$CH_3D + T \rightarrow CH_3 + DT$	43.4	μOMT(SCT)

[a]The μOMT and the SCT transmission coefficients are equivalent

coefficients seriously underestimate tunneling when compared with the LAG4 ones, because SCT gives values which are too large for the deuterium transfer and LCG4 gives values which are small for hydrogen transfer. Thus, the μOMT KIEs give values which are between the calculated SCT and the LCG4 KIEs. The SCT and μOMT KIEs only coincide when the acceptor is a hydrogen atom, as the mass of the acceptor increases both KIEs get more different, i.e., the SCT KIEs become too small. The LAG4 KIEs are always larger than KIEs obtained by the other approximations. Although this example is not conclusive, from the LCG4,

Table 18-4 KIEs due to tunneling (κ_H/κ_D) computed with different tunneling approximations using the JG-PES

	κ_H/κ_D			
T(K)	SCT	LCG4	μOMT	LAG4
		$CH_4 + H/CH_3D + H$		
200	2.29	2.88	2.29	5.11
250	1.77	2.07	1.77	2.88
300	1.51	1.68	1.51	2.04
400	1.27	1.34	1.27	1.46
		$CH_4 + D/CH_3D + D$		
200	1.08	2.44	1.58	3.12
250	1.08	1.86	1.36	2.27
300	1.07	1.58	1.26	1.84
400	1.05	1.33	1.15	1.45
		$CH_4 + T/CH_3D + T$		
200	0.78	2.23	1.77	2.74
250	0.85	1.68	1.42	1.99
300	0.90	1.45	1.27	1.64
400	0.94	1.24	1.14	1.33

μOMT and SCT approximations, it seems that the latter is the approximation which underestimates the quantum KIEs by the largest amount. This result is disappointing taking into account that the SCT is the most applied approximation for the study of enzymatic reactions.

In summary, the results obtained for this series of reactions show that SCT should be used with care. A small transmission coefficient does not guarantee that tunneling is well handle by the SCT approximation (see for instance reaction R5 in Table 18-2). Of course, there are reactions for which SCT should work very well as it is the case for heavy atom reorganization reactions at low temperatures [67, 68] or for hydrogen transfer reactions with very wide barriers [69]. However, there are hydrogen transfer reactions presenting very large KIEs [70–73], for which the results obtained with SCT are doubtful. We recommend the use of the DILAT algorithm as alternative for the study of large systems.

REFERENCES

1. Eyring H (1935) J Chem Phys 3:107–115
2. Evans MG, Polanyi M (1935) Trans Faraday Soc 31:875–894
3. Kohen A, Limbach HH (eds) (2006) Isotope effects in chemistry and biology. CRC Press, Boca Raton, FL
4. Hynes JT, Schowen RL, Klinman JP, Limbach HH (eds) (2007) Hydrogen-transfer reactions. Wiley-VCH, Weinheim

5. Gamow G (1928) Z Phys 51:204
6. Eckart C (1930) Phys Rev 35:1303–1309
7. Wigner E (1932) Z Phys Chem B19:203–216
8. Bell RP (1933) Proc Roy Soc A 139:466–474
9. Johnston HS (1966) Gas phase reaction rate theory. Ronald Press, New York
10. Kemble EC (1937) The fundamental principles of quantum mechanics with elementary applications. Dover, New York
11. Wigner E (1937) J Chem Phys 5:720–725
12. Horiuti J (1938) Bull Chem Soc Jpn 13:210–216
13. Keck JC (1967) Adv Chem Phys 13:85–121
14. Garrett BC, Truhlar DG (1979) J Chem Phys 70:1593–1598
15. Garrett BC, Truhlar DG (1980) Acc Chem Res 13:440–448
16. Pechukas P (1981) Annu Rev Phys Chem 32:159–177
17. Truhlar DG, Hase WL, Hynes JT (1983) J Phys Chem 87:2664–2682
18. Truhlar DG, Garrett BC (1984) Annu Rev Phys Chem 35:159–189
19. Truhlar DG, Isaacson AD, Garrett BC (1985) In: Baer M (ed) Theory of chemical reaction dynamics. CRC Press, Boca Raton, FL
20. Truhlar DG, Garrett BC, Klippenstein SJ (1996) J Phys Chem 100:12771–12800
21. Fernández-Ramos A, Ellingson BA, Garrett BC, Truhlar DG (2007) Rev Comput Chem 23:125–232
22. Bowman JM, Wang D, Huang X, Huarte-Larrañaga F, Manthe U (2001) J Chem Phys 114:9683–9684
23. Garrett BC, Truhlar DG (1979) J Phys Chem 83:1052–1079
24. Truhlar DG, Kupperman A (1971) J Am Chem Soc 93:1840–1851
25. Fukui K, Kato S, Fujimoto H (1975) J Am Chem Soc 97:1–7
26. Marcus RA (1966) J Chem Phys 45:4493–4499
27. Marcus RA, Coltrin ME (1977) J Chem Phys 67:2609–2613
28. Skodje RT, Truhlar DG, Garrett BC (1981) J Phys Chem 85:3019–3023
29. Skodje RT, Truhlar DG, Garrett BC (1982) J Chem Phys 77:5955–5976
30. D-h Lu, Truong TN, Melissas VS, Lynch GC, Liu Y-P, Garrett BC, Steckler R, Isaacson AD, Rai SN, Hancock GC, Lauderdale JG, Joseph T, Truhlar DG (1992) Comput Phys Commun 71:235–262
31. Liu YP, Lynch GC, Truong TN, D-h Lu, Truhlar DG (1993) J Am Chem Soc 115:2408–2415
32. Garrett BC, Truhlar DG, Wagner AF, Dunning TH Jr (1983) J Chem Phys 78:4400–4413
33. Garrett BC, Abusalbi N, Kouri DJ, Truhlar DG (1985) J Chem Phys 83:2252–2258
34. Garrett BC, Joseph T, Truong TN, Truhlar DG (1989) Chem Phys 136:271–293
35. Truong TN, D-h Lu, Lynch GC, Liu Y-P, Melissas VS, Stewart JJP, Steckler R, Garrett BC, Isaacson AD, González-Lafont A, Rai SN, Hancock GC, Joseph T, Truhlar DG (1993) Comput Phys Commun 75:143–159
36. Liu Y-P, D-h Lu, González-Lafont A, Truhlar DG, Garrett BC (1993) J Am Chem Soc 115:7806–7817
37. Fernández-Ramos A, Truhlar DG (2001) J Chem Phys 114:1491–1496
38. Garrett BC, Truhlar DG (1983) J Chem Phys 79:4931–4938
39. Lynch GC, Truhlar DG, Garrett BC (1989) J Chem Phys 90:3102–3109
40. Meana-Pañeda R, Truhlar DG, Fernández-Ramos A (2010) J Chem Theory Comput 6:6–17
41. Pu J, Gao J, Truhlar DG (2006) Chem Rev 106:3140–3169
42. Marcus RA (2006) Phil Trans R Soc Lond B 361:1445–1455
43. Klinman JP (2006) Phil Trans R Soc B 361:1323–1331
44. Nagel ZD, Klinman JP (2006) Chem Rev 106:3095–3118
45. Wong KY, Richard JP, Gao J (2009) J Am Chem Soc 131:13963–13971

46. Smedarchina Z, Siebrand W, Fernández-Ramos A, Cui Q (2003) J Am Chem Soc 125:243–251
47. Tautermann CS, Loferer MJ, Voegele AF, Liedl KR (2004) J Chem Phys 120:11650–11657
48. Smedarchina Z, Siebrand W (2005) Chem Phys Lett 410:370–376
49. Zhang Y, Lin H (2009) J Phys Chem A 113:11501–11508
50. Dybala-Defratyka A, Paneth P, Banerjee R, Truhlar DG (2007) Proc Natl Acad Sci USA 104:10774–10779
51. Pang J, Hay S, Scrutton NS, Sutcliffe MJ (2008) J Am Chem Soc 130:7092–7097
52. Hirschfelder JO, Wigner E (1939) J Chem Phys 7:616–628
53. Kuppermann A (1979) J Phys Chem 83:171–187
54. Truhlar DG, Kuppermann A (1972) J Chem Phys 56:2232–2252
55. Allison TC, Truhlar DG (1998) In: Thompson DL (ed) Modern methods for multidimensional dynamics computations in chemistry. World Scientific, Singapore, pp 618–712
56. Babamov VK, Marcus RA (1981) J Chem Phys 74:1790–1798
57. Kim Y (1996) J Am Chem Soc 118:1522–1528
58. Kim Y (1998) J Phys Chem A 102:3025–3036
59. Loerting T, Liedl KR, Rode BM (1998) J Am Chem Soc 120:404–412
60. Kreevoy MM, Ostovic D, Truhlar DG, Garrett BC (1986) J Phys Chem 90:3766–3774
61. Renka RJ (1987) SIAM J Sci Stat Comput 8:393–415
62. Renka RJ (1993) ACM Trans Math Software 19:81–94
63. Fernández-Ramos A, Truhlar DG, Corchado JC, Espinosa-Garcia J (2002) J Phys Chem A 106:4957–4960
64. Fernández-Ramos A, Truhlar DG (2005) J Chem Theory Comput 1:1063–1078
65. Cline AK, Renka RJ (1984) Rocky Mountain J Math 14:223–237
66. Jordan M, Gilbert R (1995) J Chem Phys 102:5669–5682
67. Zuev PS, Sheridan RS, Albu TV, Truhlar DG, Hrovat DA, Borden WT (2003) Science 299:867–870
68. Datta A, Hrovat DA, Borden WT (2008) J Am Chem Soc 130:6684–6685
69. Kim Y, Marenich AV, Zheng J, Kim KH, Kołodziejska-Huben M, Rostkowski M, Paneth P, Truhlar DG (2008) J Chem Theory Comput 5:59–67
70. Tejero I, Gonzalez-Garcia N, Gonzalez-Lafont A, Lluch JM (2007) J Am Chem Soc 129:5846–5854
71. Masgrau L, Ranaghan KE, Scrutton NS, Mulholland AJ, Sutcliffe MJ (2007) J Phys Chem B 111:3032–3047
72. Peles DN, Thoburn JD (2008) J Org Chem 73:3135–3144
73. Wu A, Mader EA, Datta A, Borden WT, Mayer JM (2009) J Am Chem Soc 131: 11985–11997

CHAPTER 19

INTEGRATING COMPUTATIONAL METHODS WITH EXPERIMENT UNCOVERS THE ROLE OF DYNAMICS IN ENZYME-CATALYSED H-TUNNELLING REACTIONS

LINUS O. JOHANNISSEN[1], SAM HAY[2], JIAYUN PANG[1], MICHAEL J. SUTCLIFFE[1], AND NIGEL S. SCRUTTON[2]
[1]*School of Chemical Engineering and Analytical Sciences, University of Manchester, UK*
[2]*Faculty of Life Sciences, University of Manchester, UK*
e-mail: nigel.scrutton@manchester.ac.uk

Abstract: We review the role of dynamics in enzyme catalysed H-tunnelling reactions with particular focus on the integration of computational methods with experimental and numerical modelling studies. We show that H-tunnelling requires compressive motion along the H-transfer coordinate and these reactions can be modelled successfully using vibrationally-coupled H-tunnelling models in which barrier compression is driven by fast motions within the enzyme–substrate complex.

Keywords: Aromatic amine dehydrogenase, Bovine serum oxidase, Dihydrofolate reductase, Flavin, Horse liver alcohol dehydrogenase, Kinetic isotope effect, Molecular dynamics, Morphinone reductase, Multidimensional tunnelling correction, Nicotinamide, Promoting vibration, Soybean lipoxygenase, Transition state theory

19.1. INTRODUCTION

Enzymes achieve unparalleled rate enhancements of up to 21 orders of magnitude [1]. This means that many of the reactions they catalyse would have half-times exceeding the age of the universe in solution [2]. The standard view of enzyme catalysis is rooted in transition state theory (TST), the classical description of chemical reactions, wherein a reactant must overcome a potential energy barrier and pass through the transition state to form the corresponding product. By stabilising the transition state, the enzyme lowers the barrier and therefore speeds up the reaction. The general idea – the complementarity principle – is that an enzyme

P. Paneth and A. Dybala-Defratyka (eds.), *Kinetics and Dynamics*, Challenges and Advances in Computational Chemistry and Physics 12,
DOI 10.1007/978-90-481-3034-4_19,

active site binds more strongly to the transition state than the reactant – for example by the proximity of appropriately charged amino acids if the transition state has different electrostatic properties than the reactant.

Despite the appeal and success of this elegant model, transition state complementarity, which offers a fundamentally static view of enzyme catalysis, is not the complete picture, and the role of dynamics is becoming increasingly recognised [3–5].The idea of the "fluctuating enzyme" was proposed by Careri over 30 years ago [6]. In 1983 Karplus and McCammon first suggested that protein motions can affect the chemistry in the active site – in their study by optimising the donor–acceptor distances to enhance the probability of electron tunnelling [7]. Over a decade ago, researchers working on H-transfer reactions (where H = proton, hydrogen atom or hydride) began to notice that experimentally observed kinetic isotope effects (KIEs) in enzymes could not be explained without incorporating small-scale fluctuations that modulate the reaction barrier [8, 9]. In fact, experimental observation of significant tunnelling contributions in many H-transfer reactions lies behind much of the current drive to understand the role of dynamics in enzyme-catalysed reactions [10–14]. This chapter will explore some of the combined experimental and computational studies that have helped shape our current understanding of enzymatic H-tunnelling reactions, and discuss some of the computational techniques that allow us to go beyond current experimental limits.

19.2. H-TUNNELING REACTIONS AS PROBES OF DYNAMICS

Tunnelling is the quantum mechanical transition through a classically forbidden energy state. At the quantum level, a particle is described by its wavefunction, the square of which corresponds to its probability function. For a particle near an insurmountable energy barrier, there is a non-zero overlap between the wavefunctions of the particle on either side of the barrier (Figure 19-1), which corresponds to the probability that the particle will spontaneously appear on the other side – i.e. tunnel through the barrier. The magnitude of the wavefunction overlap is very sensitive (and related inversely) to the tunnelling distance and the mass of the transferring particle, leading to potentially very large isotope effects.

Traditionally, the magnitude of the KIE was used as an indication of tunnelling. In classical TST the mass of the transferring isotope is unimportant, but when corrections are made for the zero-point energy (ZPE) vibrations of the system in so-called quasi-classical treatments (defined and reviewed in detail in reference [15]), a KIE arises from the differences in ZPE between the transferring isotopes. This imposes a limit of ~7 on the KIE of H-transfers at room temperature [16], and therefore, H-transfers with KIEs significantly elevated beyond this value must proceed with some degree of tunnelling. Initial efforts to account for a tunnelling contribution incorporated a tunnelling correction into quasi-classical TST, using the Bell correction model for tunnelling below the top of the barrier [16]. However, by

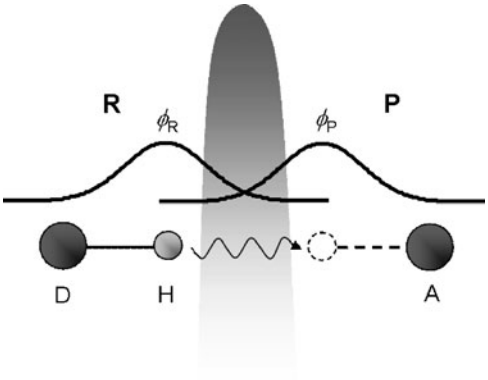


Figure 19-1. Schematic of H-transfer occurring by tunnelling. An insurmountable energy barrier separates the reactant, **R**, from the product, **P**. $\phi_{H,R}$ and $\phi_{H,P}$ are the wavefunctions for H (proton, hydrogen atom or hydride) bound to the donor (D) and acceptor (A) atoms, respectively. The square of the overlap between $\phi_{H,R}$ and $\phi_{H,P}$ is the probability that H will tunnel through the barrier

the late 1990s it became apparent that this model is not always adequate when a significant degree of tunnelling is involved.

Unusual kinetic data reported by Grant and Klinman in the reaction catalysed by bovine serum oxidase (BSAO) [17] made it clear that a new conceptual framework was required. The observed activation energy, which sets an upper limit to the classical barrier height, was ~56 kJ mol^{-1}, while the zero point energy of a C–H bond is ~36 kJ mol^{-1}. This means that the vibrational quanta are comparable to the barrier height and therefore that bond cleavage cannot be treated classically. As such, quantum mechanical tunnelling was invoked as part of the reaction mechanism. However, tunnelling over typical enzymatic H-transfer distances would lead to hugely inflated KIEs compared to those observed experimentally. Consequently, atomic fluctuations are required to achieve lower tunnelling distances and reasonable KIEs. This led to the Vibrationally Enhanced Ground State Tunnelling (VEGST) model [9], which successfully explained the temperature-dependent KIE in BSAO. However, this model is too simple to account for the temperature-independent KIEs observed experimentally in a number of enzyme systems since publication of the original BSAO study.

Currently, there are several physical models for vibrationally coupled H-tunneling [18–21]. Conceptually, these are all very similar to the model of Kuznetsov and Ulstrup [21] which employs the Franck–Condon principle that tunnelling of a light particle (such as an electron or proton) only occurs at heavy nuclear configurations where the reactant and product energies are degenerate. When reactant and product energies are degenerate, the probability of tunnelling depends on the wavefunction overlap between the reactant and product. Vibrations that bring the system to such a configuration – the transition state – are similar to the rearrangements required to achieve electron tunnelling in the Marcus theory of electron transfer [22]. These are

vibrations along the solvent coordinate, so-called because they mainly involve reorganisation of solvent to achieve the required electrostatic environment, but it also involves reorganisation of the enzyme system. Vibrationally coupled models also employ very rapid vibrations that are coupled to the orthogonal H-transfer coordinate that modulate the tunnelling distance at the transition state by decreasing the donor–acceptor distance. These have been given many names, including promoting vibrations, promoting motions and gating; in this chapter, the term "promoting vibration" will be employed. Within these models, these two classes of vibrations are treated as independent coordinates, although in reality it is unlikely they are so neatly distinguishable. For example, a vibration along the H-transfer coordinate might affect degeneracy; similarly, vibrations along the solvent coordinate could affect the donor–acceptor separation. Nevertheless, the model is conceptually useful, and has been very successful for modelling and rationalising experimental observations [23–27]. In its simplest form, the equation for the rate of tunnelling according to this model can be written as:

$$k_{tun} = \frac{1}{2\pi}|V_{EL}|^2\sqrt{\frac{4\pi^3}{\lambda k_B T\hbar^2}}\exp\left\{\frac{-\Delta G^{\ddagger}}{k_B T}\right\} \times |\langle\phi_R|\phi_P\rangle|^2 \tag{19-1}$$

This is essentially the Marcus equation for the rate of electron transfer, with an added Franck–Condon nuclear overlap term that represents the probability of H-tunnelling at the transition state. V_{EL} is the electronic coupling between reactants and products, λ the reorganisation energy, $\Delta G^{\ddagger}$ the activation free energy, T the temperature and $\phi_{R/P}$ the H wavefunction in the reactant and product; the Franck–Condon term is the square of the wavefunction overlap between the reactant and product states.

The majority of the temperature-dependence of the reaction rates arises from $\Delta G^{\ddagger}$, which depends on the reorganisation energy, λ, and the driving force for the reaction, ΔG^0, according to the Marcus relation:

$$\Delta G^{\ddagger} = \frac{(\Delta G^0 + \lambda)^2}{4\lambda} \tag{19-2}$$

Within this formalism, the kinetic isotope effect (KIE) arises mainly from the difference in the wavefunction overlap of the transferring particle: lighter particles have more diffuse wavefunctions and therefore larger wavefunction overlap for a given inter-atomic separation. The wavefunction overlap and isotope effect are very sensitive to the tunnelling distance. One feature of these models is that the temperature-dependence of the KIE – in other words the difference in activation energies (ΔE_a) from an Arrhenius or activation enthalpies ($\Delta\Delta H^{\ddagger}$) from an Eyring plot – depends on the relative contribution of promoting vibrations that modulate the tunnelling distance. A strongly temperature-dependent KIE suggests a significant contribution from promoting vibrations, since these give rise to a temperature-dependent Franck–Condon term. On the other hand, a reaction dominated by Marcus-like

dynamics will have an essentially temperature-independent KIE since $\Delta G^{\ddagger}$ is essentially isotope independent – the energy required to achieve quantum degeneracy does not depend on the isotope being transferred (note that the contribution of vibrationally excited states means that ΔG^{0} and, to some extent, λ are not entirely isotope independent, but this has little effect on the temperature-dependence of the KIE [29]).

The magnitude of the temperature-dependence of the KIE has therefore been used as a measure of the importance of promoting vibrations in many enzymatic H-transfer reactions, and important insight can be obtained from modelling the experimental data using vibrationally-coupled H-tunnelling models. We will now discuss some of these studies which have been critical in shaping our current understanding of vibrationally-coupled enzymatic H-tunnelling reactions.

19.2.1. Hydrogen Atom Transfer in Soybean Lipoxygenase-1

Studies of the rate-limiting hydrogen atom (H•) abstraction step in the reaction catalysed by soybean lipoxygenase-1 (SLO-1) and its mutants have been important in establishing the importance of promoting vibrations in enzymes. The wild-type (WT) enzyme exhibits a weak temperature dependence on the KIE ($\Delta E_a = 3.8$ kJ mol^{-1}), and this becomes significantly stronger when certain critical amino acids are mutated to smaller residues (L56A, L754A and I553A; $\Delta E_a = 7.9$, 7.9 and 16.7 kJ mol^{-1}, respectively) [23]. The kinetics were simulated using the Kuznetsov and Ulstrup (K–U) model [21], wherein the promoting vibration is modelled as a harmonic oscillator which causes the tunnelling distance r_{tun} to deviate from the equilibrium tunnelling distance at the transition state, r_0. The force constant k_X for this harmonic oscillator, which relates to its frequency ω_X and effective mass m_X by Hooke's law, governs the energy E_X required to deviate from r_0 by a distance r_X along the promoting vibration coordinate X, to achieve a given tunneling distance r_{tun}:

$$r_{tun} = r_0 - r_X \tag{19-3}$$

$$X = r_X \sqrt{m_X \omega_X / \hbar} \tag{19-4}$$

$$E_X = \tfrac{1}{2} \hbar \omega_X X^2 \tag{19-5}$$

The nuclear wavefunction overlap at temperature T is the integral of the overlap terms for all possible values of r_{tun}:

$$\langle \phi_\mu | \phi_v \rangle = \int_{r_{tun}=0}^{r_0} \langle \phi_\mu | \phi_v \rangle_{r_{tun}} \exp(-E_X / k_B T) dX \tag{19-6}$$

Contribution from excited states is also taken into account, so that Eqs. (19-1) and (19-2) become a summation over reactant and product vibrational states μ and v:

$$k_{CEP} = \sum_{\mu} P_{\mu} \sum_{v} \frac{1}{2\pi} |V_{EL}|^2 \sqrt{\frac{4\pi^3}{\lambda RT\hbar^2}} \exp\left\{\frac{-\Delta G^{\ddagger}_{\mu v}}{RT}\right\} \left|\langle \phi_{\mu} | \phi_{v} \rangle\right|^2 \tag{19-7}$$

$$\Delta G^{\ddagger}_{\mu v} = \frac{\left(\Delta G^0 + \Delta E^{vib}_{\mu v} + \lambda\right)^2}{4\lambda} \tag{19-8}$$

P_{μ} is the Boltzmann probability for the reactant vibrational state μ and $\Delta E^{vib}_{\mu v}$ is the energy difference between states μ and v. The Franck–Condon term is calculated by using a quantum harmonic oscillator to describe the breaking and forming bonds of the transferring particle, and depends on the frequency (ω) and mass (m) of the oscillator, and on the tunneling distance. For ground state tunneling (μ, $v = 0$), the Franck–Condon term is:

$$\langle \phi_0 | \phi_0 \rangle = \exp\left(-m\omega {r_{tun}}^2 / 2\hbar\right) \tag{19-9}$$

The Franck–Condon terms for other vibrational states are given in reference [21].

Simulating the rates for the SLO-1 mutants using the K–U model required larger r_0 values, with concomitant lower frequencies ω_X. This suggests that the mutations perturb the active site geometry, causing longer equilibrium donor–acceptor well separations that require more donor–acceptor compression to achieve tunnelling. The decrease in the frequencies of the promoting vibration in the mutants relative to WT is consistent with substrate fluctuating in a larger active site pocket. The models also reproduced the increased E_a of two of the mutants – L56A and L754A – although it also suggested an increased E_a for I553A, which is in fact equal to the WT value. Note that this presumably highlights the difficulty in accounting for changes in ΔG^0 and λ in the mutants – and enzyme reactions in general – and not a failing of the model (ΔG^0 and λ were assumed to remain constant). This study was later extended to involve multiple mutants of I553, and the temperature-dependence of the KIE was found to increase as the size of the mutated residue side chain decreases, which is consistent with the previous discussion [30].

One shortcoming of these SLO-1 simulations is that they predict unreasonably short equilibrium donor–acceptor distances [23, 28]. The K–U model treats the hydrogen wavefunctions as quantum harmonic oscillators – by using a more accurate quantum Morse oscillator instead, the data were later reproduced with longer distances [25]. Another shortcoming is that the K–U model treats the hydrogen wavefunctions of the reactant and product as identical. Simulations for WT SLO-1 that took account of the differences between the reactant and product wavefunctions and used a more complete quantum mechanical treatment where λ and ΔG^0 depend on the donor–acceptor separation and vibrational state have also been carried out [28]. Using molecular dynamics simulations to determine the frequency for donor–acceptor fluctuations, this method was able to reproduce the magnitude and temperature-dependence of KIEs without fitting any parameters to

the experimental data [29]. While there is disagreement as to the magnitude of the force constants for the donor–acceptor fluctuations involved, these studies confirm the importance of such fluctuations in decreasing the tunnelling distance and hence increasing the rate, and demonstrate that they give rise to the majority of the temperature-dependence of the KIE.

19.2.2. Hydride Transfer in Morphinone Reductase

Hydride transfer from nicotinamide coenzyme NADH to the flavin cofactor (flavin mononucleotide; FMN) in the reaction catalysed by morphinone reductase (MR) has proved to be a very useful model system that has given detailed insight into the role of dynamics in enzymatic H-transfers. This reaction exhibits a strongly temperature-dependent KIE, with $\Delta\Delta H^{\ddagger} \sim 8$ kJ mol^{-1}, indicative of H-transfer assisted by promoting vibrations [31]. It had previously been demonstrated that KIEs for non-tunnelling reactions are pressure-independent [32, 33], so to further study the degree of tunnelling and the nature of promoting vibrations in this enzyme, the simultaneous pressure and temperature dependencies of the KIE were measured [27]. The rates of H-transfer were found to increase with pressure, which could be explained by pressure-induced donor–acceptor compression, which leads to increased tunnelling probabilities. On the other hand, the KIE was found to increase with pressure, and the K–U model predicts that if all other parameters remain the same (or if no promoting vibration is present) a decrease in tunnelling distance will lead to a *decrease* in the KIE. Furthermore, the temperature-dependence of the KIE was almost invariant with pressure ($d\Delta\Delta H^{\ddagger}/dp \sim 0$), and according to the K–U model a decrease in r_0 leads to a decrease in $\Delta\Delta H^{\ddagger}$. By systematically varying the equilibrium tunnelling distance, the force constant of the promoting vibration and the temperature while solving for the experimental KIE and $\Delta\Delta H^{\ddagger}$ the pressure-temperature dependence was successfully modelled with a promoting vibration that increases in frequency as the transfer distance is compressed with pressure (Figure 19-2b and c). Specifically, within this framework the promoting vibration requires a fairly low frequency (~ 100 cm^{-1} at 298 K and 1 bar), so that a significant reduction in tunnelling distance can account for the small KIE of ~4–8. This study demonstrates the effectiveness of pressure as a complementary probe to temperature for H-tunnelling reactions, and confirms the need for a promoting vibration in MR [27].

19.2.2.1. High-Pressure Molecular Dynamics Confirm the Effect of Pressure

The pressure–temperature dependence of the KIE of MR suggests that the equilibrium tunnelling distance r_0 decreases with pressure, with a concomitant increase in the frequency of the promoting vibration. To test this hypothesis, the strength of the charge-transfer (CT) complex between the hydride donor and acceptor groups has

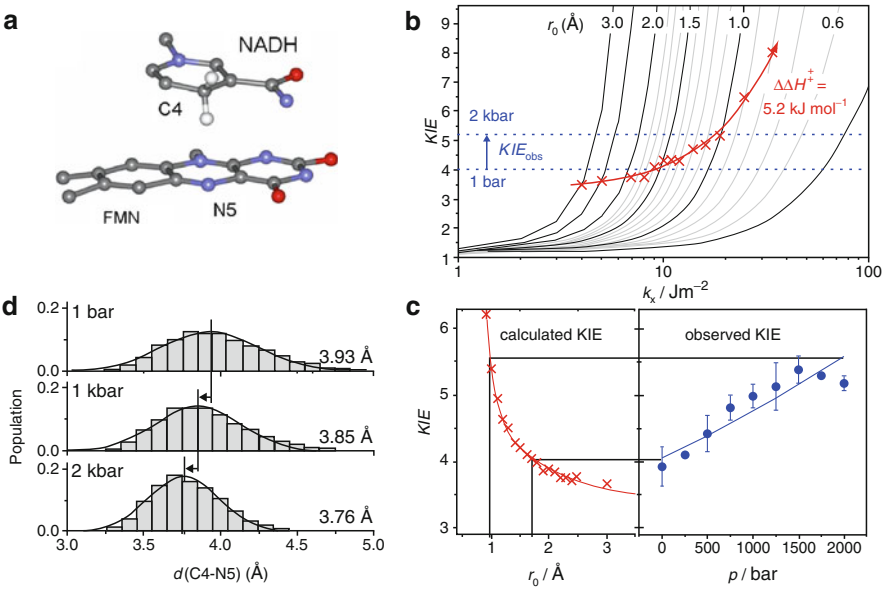


Figure 19-2. Modelling pressure-dependence of the KIE in MR. (**a**) Hydride transfer occurs from C4 of NADH to N5 of FMN. (**b**) The change in KIE with pressure was successfully modeled using a promoting vibration whose force constant k_X, and therefore frequency, increases as the equilibrium tunnelling distance r_0 decreases with pressure. The red crosses are the modeled KIEs when the temperature-dependence of the KIEs, $\Delta\Delta H^{\ddagger}$, was fixed to the experimentally observed value of 5.2 kJ mol^{-1}. (**c**) Correlation between the modeled and observed KIEs: the calculated KIEs (when $\Delta\Delta H^{\ddagger} = 5.2$ kJ mol^{-1}) versus r_0 and the measured KIEs versus pressure. Within the Kuznetsov and Ulstrup model, the observed increase in KIE is consistent with a total decrease in r_0 of ~0.7 Å. (**d**) MD simulations qualitatively confirm the trend of donor–acceptor compression with pressure and a concomitant decrease in the magnitude of donor–acceptor fluctuations, which suggests an increased frequency (Adapted from [27, 34])

been measured [34]. The hydride transfer from NADH to FMN occurs between two π-stacked aromatic systems (Figure 19-2a), which form a CT complex, whose strength can be monitored by absorbance spectroscopy. The strength of the π–π interaction is very sensitive to the distance between the ring systems, and therefore the measured increase in CT absorbance with pressure suggests donor–acceptor compression, although the relative orientation of the rings is also important. The specific effect of pressure on the system over the experimental range of 1 bar–2 kbar was further estimated by constant pressure molecular dynamics (MD) simulations at 1 bar and 1 kbar. Analysis revealed a roughly Gaussian distribution of donor–acceptor distances (Figure 19-2d), which becomes narrower and shifts to lower distances as the pressure increases. Since the variance of the donor–acceptor distance is directly related to the frequency of donor–acceptor fluctuations [29], this supports the K–U prediction that pressure causes an increase in the frequency of the promoting vibration.

19.2.2.2. Further Insights from Molecular Modelling of the H-Transfer

The precise mechanism of hydride transfer – the degree of tunnelling and the reaction path – has been determined computationally [35] using variational transition state theory with ensemble-averaged multidimensional tunnelling corrections (EA-VTST/MT) [36–40]. This method uses umbrella sampling to sample many structures along the reaction coordinate to build a free energy barrier and an ensemble of transition state structures to calculate the tunnelling coefficient, which describes the percentage of transfers that occur via tunnelling. This revealed that, despite the low KIE that is not elevated above the quasi-classical limit, 99% of H-transfers occur via tunnelling, which enhances the rate 100-fold relative to the classical rate for the same free energy barrier. This is the first time that such a "deep tunnelling" mechanism has been observed for hydride transfer. Comparing the reactant, transition state and product geometries revealed a significant heavy-atom rearrangement prior to tunnelling, which is consistent with the need for a promoting vibration. Specifically, an initial rearrangement of the donor nicotinamide ring relative to the acceptor isoalloxazine appears to be required to achieve the correct H-transfer geometry, which is followed by a donor–acceptor compression of ~0.2 Å. The rearrangement can be ascribed to active site reorganization, and the subsequent compression to a promoting vibration. From the magnitude of the compression, the frequency of the promoting vibration was estimated as ~50 cm^{-1} from the K–U model; note that while this is a qualitative assessment, it does confirm that a fairly soft (i.e. low-frequency) promoting vibration is required for hydride transfer in MR.

19.3. COMPUTATIONAL TECHNIQUES FOR ATOMISTIC ANALYSIS OF PROMOTING VIBRATIONS

Direct experimental observation of promoting vibrations is currently not feasible. Since they are coupled to the H-transfer coordinate, they must occur on sub-picoseconds time-scale, i.e. with frequency > ~30 cm^{-1}, and to be significantly populated at physiological temperatures they cannot significantly exceed the available thermal energy, k_BT ~220 cm^{-1}. This places them in the THz frequency region, which is very difficult to observe using conventional spectroscopic techniques, particularly in complex systems like enzymes. On the other hand, a multitude of computational techniques have been developed for achieving atomic-scale descriptions of these vibrations.

19.3.1. Spectral Density Analysis Reveals a Promoting Vibration in Horse Liver Alcohol Dehydrogenase

A computational technique for detecting promoting vibrations that decrease the donor–acceptor distance was published in 2001 [41], based on the principles that a

vibration coupled to the reaction coordinate must vanish at the transition state. The relative velocities between the donor and acceptor atoms were obtained from molecular dynamics simulations of the reactant, transition and product states and projected onto the donor–acceptor inter-atomic axis. The donor–acceptor spectral density, a deconvolution of the frequencies involved in modulating the donor–acceptor distance, was then calculated by Fourier transforming the velocity auto-correlation function of this projected velocity. The diagnosis for a promoting vibration is a strong peak in the spectral density for donor–acceptor compression in the reactant and product, which almost disappears in the transitions state. Using this method, a $\sim$150 cm^{-1} promoting vibration that compresses the cofactor–substrate distance was identified in the hydride transfer of the reaction catalysed by horse liver alcohol dehydrogenase [42]. Spectral densities for the relative motion between a valine on the cofactor-side of the active site and the substrate also exhibit a strong peak at this frequency. This was interpreted in terms of this amino acid participating in the promoting vibration, pushing the cofactor towards the substrate. However, this result is also consistent with a vibration of the substrate towards the cofactor, which decreases both the donor–acceptor distance and the valine–acceptor distance. To determine which scenario is correct, it would be necessary to determine whether the cofactor or substrate (or both) gives rise to the promoting vibration. Nevertheless, when this residue was replaced with a smaller alanine, the $\sim$150 cm^{-1} donor–acceptor peak, while still present, was significantly reduced in magnitude [42], consistent with mutagenesis studies on this enzyme [43].

19.3.2. Spectral Densities Coupled with Digital Filtering of Atomic Motions Reveal a Complicated Picture in Aromatic Amine Dehydrogenase

The spectral density approach was later applied to the rate-limiting proton transfer in the oxidative deamination of tryptamine catalysed by aromatic amine dehydrogenase (AADH) [44, 45], which involves the transfer of a proton from an iminoquinone intermediate to an aspartate base (Figure 19-3a). Calculations using variational transition state theory with multidimensional tunnelling (VTST/MT) [36–40] have previously shown that this reaction occurs mainly by tunnelling, and that the H-transfer coordinate prior to tunnelling is defined by a rotation of the donor methylene group towards the acceptor oxygen atom, coupled with a small repositioning of the acceptor oxygen [44]. The KIE for this proton transfer is not measurably temperature dependent (within experimental error and the accessible temperature range), which suggests that it is not assisted by a promoting vibration. However, in-depth computational analysis has revealed that this experimental probe is not sufficiently precise [45].

The H-transfer coordinate prior to tunnelling is dominated by motion of the donor C1, so a promoting vibration should involve motion of this atom. The spectral density for C1 motion shows a pronounced peak at 165 cm^{-1} (Figure 19-3b), but

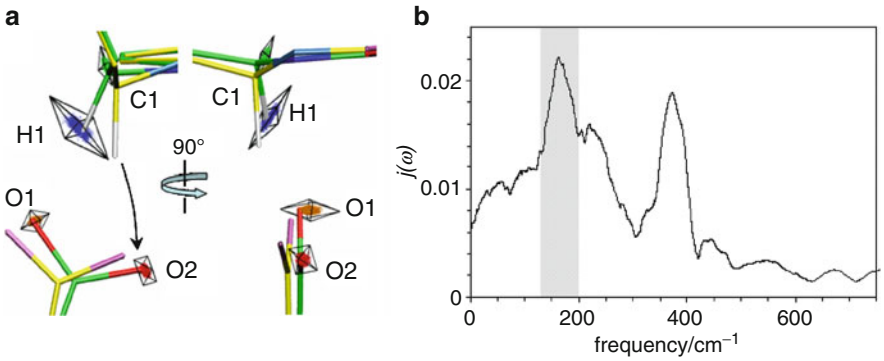


Figure 19-3. Proton transfer in AADH is assisted by the principal vibration of the donor group. (**a**) Reactant (*green carbon atoms*) and representative structure for tunnelling (RTE, *yellow carbon atoms*) from VTST/MT calculations, with superimposed atomic coordinates for 165 cm^{-1} vibrations; the *diamond shapes* represent the principal components of the vibrations. (**b**) Spectral density for the motion of C1; the *shaded region* represents the frequency range used for analysing the motions corresponding to the 165 cm^{-1} vibration (Adapted from reference [45])

the spectral density for donor–acceptor compression does not, which indicates that this vibration is not responsible for the largest donor–acceptor compression. However, the H-transfer is significantly non-linear, so in this case vibrations along the transfer coordinate may not correspond to the biggest donor–acceptor compression. The donor–acceptor spectral density was found to exhibit a shoulder at 165 cm^{-1}, suggesting that this C1 vibration does compress the C1–O2 distance, and is therefore a viable candidate for a promoting vibration.

To determine the precise nature of this vibration, a method of digital filtering by frequency deconvolution was applied to the atomic motions of the donor, acceptor and transferring proton [44, 45]. This method was initially developed for filtering out high-frequency vibrations from MD simulations to reveal domain motions and low frequency "breathing" modes [46], but was used here to reveal motions within a specific frequency window: a Fourier transformation was performed on the atomic velocities, the resulting spectrum was set to zero for all values outside of the desired frequency range, and an inverse Fourier transformation was performed. The resulting velocities were then converted to atomic coordinates, so that the motion within the chosen frequency range could be visualised.

From VTST/MT calculations, the probability of proton tunnelling as a function of reaction coordinate was obtained, from which the geometry where the majority of tunnelling occurs was identified – this is the structure at the representative tunnelling energy (RTE) along the minimum energy path (MEP, the potential energy profile along the reaction coordinate) [44]. The vectors connecting the atomic positions in the reactant structure to those in the RTE structure – the R-to-RTE vectors – therefore represent the reaction coordinate leading to tunnelling. Comparing these vectors to the motion within a $\pm$ 30 cm^{-1} window around the 165 cm^{-1} frequency for the major vibration of the C1 revealed that this vibration is

strongly coupled to the R-to-RTE vectors for C1 and transferring proton, H1 (Figure 19-3a). The O2 spectral density has a shoulder at ~165 cm^{-1}, and the O2 motion within the ± 30 cm^{-1} frequency range also strongly corresponds to the R-to-RTE vector for O2 [45]. Furthermore, it was shown that the motions of C1 and O2 within this frequency range are strongly correlated, indicating that the major vibration of C1 couples to O2 (by forming a hydrogen bond when sufficiently close) and drives the system along the H-transfer coordinate towards the tunnelling geometry. Numerical modelling was used to reconcile this result with the apparent temperature-independence of the KIE: while a small temperature-dependence was found to arise as a result of the promoting vibration, this falls within the experimental error for the small temperature-range accessible experimentally for studying enzyme kinetics [45].

The digital filtering technique was also applied to the principal vibration of O2, at 100 cm^{-1}, which corresponds to major donor–acceptor compression. This was found to decrease the donor–acceptor distance sufficiently for the coupling between C1 and O2 to take place, thus serving as a necessary "pre-organisation" of the active site [45].

19.3.3. Potential Energy Scans Reveal the Effect of the Promoting Vibration on Barrier Scaling in AADH

A very popular computational technique for modeling enzymatic reactions is the aforementioned VTST method [36–40]. This samples many structures at consecutive values along the reaction coordinate to build a free energy barrier. This has the advantage of implicitly accounting for the dynamics of the system (since free energy takes into account all degrees of freedom), but this can also be a disadvantage since it is then difficult to explicitly study these dynamics: each point along the reaction coordinate corresponds to an ensemble of structures, so the specific vibrations that lead from one point to the next cannot be discerned.

An alternative approach for determining the effect of promoting vibrations on the progress of the reactant along the reaction coordinate was used for the proton transfer in AADH [47]. This involved running a QM/MM MD simulation and performing potential-energy scans for the proton transfer for structures sampled from the MD simulation at 1 ps intervals. A harmonic constraint was applied to the non-reacting (MM) region, i.e. everything apart from the iminoquinone and catalytic Asp, so that any structural changes that occur during the MD simulations were maintained during the successive energy minimisations of the potential-energy scans. Motion along the reaction coordinate should not be restricted, but VTST calculations previously revealed that this was governed by motion of atoms within the region treated using QM – the iminoquinone and Asp (see Section 19.3.2). In this way, over 200 potential energy barriers were obtained. These revealed that the sampled structures did not usually correspond to a minimum along the reaction coordinate, but that thermal fluctuation corresponding to the previously identified promoting vibration was pushing the system part-way up the barrier. This

promoting vibration causes a significant decrease in the "effective activation energy", i.e. the barrier height remaining after thermal fluctuations along the reaction coordinate [47].

Comparison with a modified form of AADH (i.e. the isolated active-site containing small subunit) revealed how the enzyme gives rise to this promoting vibration [47]. The isolated small subunit is catalytically active, but catalysis is three orders of magnitude slower than the native form of the enzyme [48]. The active site of the isolated small subunit is exposed to solvent, so that the iminoquinone can adopt conformations not accessible in the native enzyme (although the reactive groups are

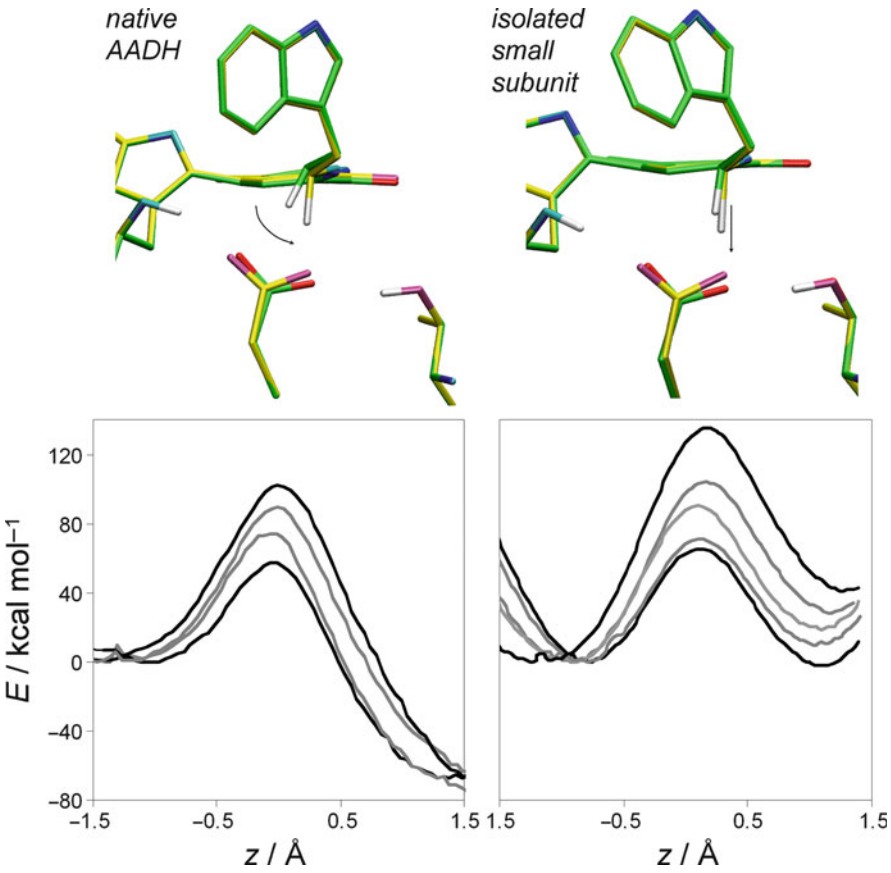


Figure 19-4. Proton transfer in native AADH versus the isolated small subunit. *Top*: A rotation of the donor groups leads to the representative structure for tunneling (RTE structure) in the native enzyme, while the same O–H compression in the isolated small subunit requires significant C1–H1 stretching. *Bottom*: Representative potential energy scans, spanning the full range of observed activation energies, with the reaction coordinate $z = d(\text{C1–H1}) - d(\text{O2–H1})$. The stiffer C1–H1 stretching in the small subunit leads to a narrower reactant well, which allows less fluctuation along the reaction coordinate (Adapted from [47])

still shielded from solvent). From the QM/MM MD simulations it appeared that the donor–acceptor geometry is in fact more favourable to proton transfer in the small subunit: the average (and the minimum) distance between the proton and acceptor oxygen is slightly shorter with a more linear C–H–O angle. This seems in direct contradiction to the slower rate, but, potential energy scans on structures sampled at 1 ps intervals from these simulations have much narrower (steeper) reactant wells, with far more restricted motion along the reaction coordinate, so that the sampled structures in this case did not lie far from the minimum. This is because the small subunit reactant requires significant C–H stretching to achieve the same O–H compression as the rotation of the donor group in the native enzyme, leading to narrower reactant wells in the potential energy scans (Figure 19-4). Therefore, the native enzyme constrains the geometry of the iminoquinone in such a way that the reaction coordinate becomes coupled to the vibration of the donor group.

19.4. THE ROLE OF LONG-RANGE COUPLED MOTIONS

Ever since the importance of donor–acceptor fluctuations in enzymatic H-transfer reactions was recognised, the possibility that enzymes might have evolved long-ranged coupled motions to promote the catalysed reaction has been discussed. This could manifest itself as rapid vibrations which give rise to a promoting vibration or as slower motions which bring the system to a configuration from which H-transfer is more likely, and computational studies on different enzyme systems suggest both might be possible.

19.4.1. Coupled Motions of Different Timescales in DHFR

The hydride transfer between dihydrofolate and NADPH in dihydrofolate reductase (DHFR) has been extensively studied, both experimentally and computationally, by a number of groups. The first evidence that enzyme dynamics might affect the rate of H-transfer came from measurements of the NADPH fluorescence lifetime in a series of mutants [49]. The fluorescence lifetimes were found to increase as the rate of catalysis decreases for most of the mutants, while those mutations that principally altered binding had no effect on the lifetimes. These results suggest that dynamic motions of the flexible M20 loop, which closes over the active site and is perturbed by these mutations, might lead to faster fluorescence decay and enhanced H-transfer rates. If the low-frequency open-close motions of the loop were coupled to catalysis, then the rate should be affected by solvent viscosity, but adding glycerol did not alter the rate. This suggests the closure of the loop is not rate-determining, and that catalysis is instead coupled to high-frequency intramolecular vibrations [49].

Later, cross-correlation analysis was performed on molecular dynamics simulations of three complexes in the catalytic cycle of DHFR. Of particular interest was the strong anti-correlation found between residues ~40–80 and the M20 and FG

loops (residues 14–24 and 116–125 respectively) in the reactant complex, which disappears in the post-hydride-transfer complexes, suggesting a potential link between hydride transfer and dynamics [50]. Site-directed mutagenesis of the highly conserved distal G121 and M42 residues, which are about 19 Å apart, showed that these residues directly affect the H-transfer rate. The double mutation showed a non-additive effect, which indicates that these residues are somehow interacting and sensitive to changes in the volume of each other's side-chain, which can be explained by their involvement in a network of coupled motions. Such a network was later identified by computational modelling of the hydride transfer [51], using a hybrid QM/MM technique for MD simulations that includes the dynamics of the entire solvated enzyme as well as the nuclear quantum effects of the transferring hydride [52]. In an approach similar to the VTST method, simulations were carried out at different regions of the reaction coordinate to adequately sample structures and build a free energy profile for the H-transfer. From these simulations, several structural rearrangements were found to couple to the reaction coordinate on the millisecond timescale, suggesting a network of coupled motions involving five residues that are conserved across 36 species of DHFR. Note that these slow rearrangements do not correspond to a promoting vibration as defined in this chapter (although they are termed "promoting motions" in the original study), but rather to a pre-organisation step. Real-time dynamical trajectories initiated at the transition state and propagated forwards and backwards in time (which can be done by inverting the atomic velocities) suggested that a much faster, femtosecond-timescale G121-M42 compression is also required for H-transfer. However, while this compression appears to be a prerequisite for the transfer, there is no evidence that this is part of a promoting vibration (indeed, there is no apparent correlation between this compression and the donor–acceptor distance [51]) and that this too constitutes pre-organisation.

19.4.2. A Proposed Conserved Network of Vibrations in HLADH

The previously discussed spectral density approach has been extended to study the coupling between non-catalytic residues and the promoting vibration, by investigating which residues have motion in the right direction and with the right frequency [53]. To determine the coupling between the promoting vibration and a given amino acid, the velocities for its centre of mass were projected onto the vector between it and the donor or acceptor atom (depending on which side of the active side the residue lies), then further projected onto the donor–acceptor axis. The resulting vector therefore represents the component of the velocity both towards the donor or acceptor atom and along the donor–acceptor axis. The time-correlation function between the resulting vector and the donor–acceptor relative velocity projected onto the donor–acceptor axis was then calculated and Fourier transformed. The resulting spectral decomposition shows the vibrational frequencies involved in pushing a given residue against the donor or acceptor along the

donor–acceptor axis. A strong peak in this spectral decomposition at the frequency of the previously identified promoting vibration suggests that the residue is coupled to the promoting vibration. This method was used to identify eight potential residues with at least one atom within 10 Å of the donor or acceptor atoms as strongly coupled to the promoting vibration of HLADH (see Section 19.3.1) [53].

To determine whether this network is conserved across other proteins, a promoting vibration "generating sequence" was defined, containing the eight amino acids in order separated by spacers of defined length, but with unspecified amino acids [54]. From a database of 10^6 protein sequences, 44 were found to contain this sequence, which, despite their diverse origins, are all alcohol dehydrogenases. Interestingly, on altering slightly the length of the spacer regions no matches were found in the database, suggesting that the positioning of the eight amino acids is essential, as would be expected if they were vibrationally coupled [54]. Note, however, that while this is consistent with a conserved network of vibrations, sequence similarity across a series of alcohol dehydrogenases is not surprising and does not constitute conclusive evidence.

19.4.3. A Small-Scale, Local Promoting Vibration in AADH

The spectral density approach was also extended to study the potential role of an extended network in the promoting vibration for proton transfer in AADH [26]. In this case, since the 165 cm^{-1} promoting vibration is not the vibration most strongly coupled to donor–acceptor compression, but rather a vibration of the donor carbon along the non-linear reaction coordinate (see Section 19.3.2), the criterion used was coupling to the donor carbon at this frequency. All heavy atoms that never moved further than 12 Å from the donor carbon were considered in this study. The strength of the coupling C for a specific frequency range for an atom A was calculated based on the spectral densities for both the velocity autocorrelation function for A and the time-correlation function of A and the donor carbon:

$$C(f_1 - f_2) = C_\tau(f_1 - f_2) \times \left[1 - \frac{|V(f_1 - f_2) - C_\tau(f_1 - f_2)|}{\max[V(f_1 - f_2), C_\tau(f_1 - f_2)]}\right] \tag{19-10}$$

where $C_\tau(f_1 - f_2)$ and $V(f_1 - f_2)$ are the areas under the time-correlation and velocity autocorrelation spectral densities, respectively, in the $f_1 - f_2$ frequency range, divided by the area in the same frequency range for the donor carbon spectral density – in this case a window of 135–195 cm^{-1} was used (the same as for the digital filtering discussed in Section 19.3.2). This equation is necessary because the time-correlation function is calculated from the dot-product of the vectors for the two atoms, and therefore a given value of $C_\tau(f_1 - f_2)$ can correspond to either a strong vibration in the $f_1 - f_2$ range in a different direction to the donor carbon, or a weaker vibration in the same direction. In fact, using purely the time-correlation spectral densities was found to lead to false positives. A strong coupling requires that both $C_\tau(f_1 - f_2)$ and $V(f_1 - f_2)$ are fairly strong, and the square-bracketed term in Eq. (19-10) ensures that

the magnitude of $C(f_1 - f_2)$ decreases as the difference between $V(f_1 - f_2)$ and $C_\tau(f_1 - f_2)$ increases. The denominator serves to normalise $C(f_1 - f_2)$ to 1 [26].

Using Eq. (19-10), strong couplings were obtained for atoms directly bonded to the donor carbon, which validates the method. However, no atom in surrounding amino acids is strongly coupled to the promoting vibration. Even using a low cut-off of 0.5, only 13% of the atoms in the selection have $C(f_1 - f_2)$ values above 0.5, and these appear randomly scattered around the active site. There is therefore no indication of a network of atoms that might give rise to the promoting vibration of this H-transfer reaction. In other words, the 165 cm^{-1} vibration of the donor methylene group appears to be inherent to the iminoquinone intermediate. This was confirmed by a DFT frequency calculation performed on the isolated iminoquinone constrained to its active site geometry, where a vibration corresponding to a rotation of the methylene was found at 164 cm^{-1} (when its geometry was not constrained, a similar vibration was also found, but with a shift in frequency of about 10 cm^{-1}). While this analysis does not preclude the possibility of networks of motions at other timescales, this seems unlikely since MD simulations revealed no strongly anti-correlated motions [26].

19.5. DISCUSSION AND FUTURE PERSPECTIVES

The importance of fluctuations coupled to the reaction coordinate – here termed promoting vibrations – have become widely accepted through studies like those discussed in this chapter. These significantly increase the tunnelling probability and decrease the KIE relative to a fixed tunnelling distance scenario. More contentious is the role of promoting vibrations as part of the catalytic effect, i.e. to the rate enhancement relative to the same reaction in solution: computational studies on certain enzymatic H-transfers and their corresponding uncatalysed reactions suggest that similar compression effects occur in solution [55, 56]. On the other hand, enzyme active sites are highly organised, with reacting moieties significantly constrained. Efficient chemistry requires facile motion along the reaction coordinate, and promoting vibrations may offer a mechanism for achieving the best compromise. To date, studies of promoting vibrations have focused on their effect on enzyme kinetics, and not on their relation to uncatalysed reactions. Further studies are therefore required to fully determine to what extent promoting vibrations are involved in the catalytic effect.

The concept of long-range motions is also an unresolved issue. Computational studies on a range of systems suggest that different processes might be involved. The computational analysis of DHFR suggests that long-range motions are required for hydride transfer, but these appear to be pre-requisites for H-transfer rather than vibrations directly coupled to the H-transfer coordinate. On the other hand, a network of residues coupled to a rapid promoting vibration was identified in HLADH, but no such network was identified in AADH. While the degree of involvement of internal enzyme motions may vary from system to system, further

computational work and new experimental probes are clearly required to fully understand the link between dynamics and kinetics in H-transfer reactions.

ACKNOWLEDGEMENTS

Work in the authors' laboratories is funded by the UK Biotechnology and Biological Sciences Research Council. NSS is a BBSRC Professorial Research Fellow.

REFERENCES

1. Lad C, Williams NH, Wolfenden R (2003) Proc Natl Acad Sci USA 100:5607
2. Wolfenden R, Snider MJ (2001) Acc Chem Res 34:938
3. Nagel ZD, Klinman JP (2006) Chem Rev 106:3095
4. Schwartz SD, Schramm VL (2009) Nat Chem Biol 5:551
5. Sutcliffe MJ, Scrutton NS (2006) Phys Chem Chem Phys 8:4510
6. Careri G (1974) In: Mintz S, Widmayer SM (eds) Quantum Statistical Mechanics in the Natural Sciences. New York, Plenum
7. Karplus M, McCammon JA (1983) Annu Rev Biochem 52:263
8. Cha Y, Murray CJ, Klinman JP (1989) Science 243:1325
9. Bruno WJ, Bialek W (1992) Biophys J 63:689
10. Antoniou D, Schwartz SD (1997) Proc Natl Acad Sci USA 94:12360
11. Kohen A, Klinman JP (1999) Chem Biol 6:R191
12. Sutcliffe MJ, Scrutton NS (2000) Phil Trans R Soc Ser A 358:367
13. Sutcliffe MJ, Scrutton NS (2000) Trends Biochem Sci 25:405
14. Ball P (2004) Nature 431:396
15. Pu J, Gao J, Truhlar DG (2006) Chem Rev 106:3140
16. Bell RP (1980) The tunnel effect in chemistry. Chapman & Hall, London
17. Grant KL, Klinman JP (1989) Biochemistry 28:6597
18. Soudackov AV, Hatcher ER, Hammes-Schiffer S (2004) Effects of proton donor-acceptor vibrational motion on proton-coupled electron transfer in solution and proteins. Abstracts of Papers of the American Chemical Society 228, U254
19. Soudackov AV, Hammes-Schiffer S (2005) Dynamical effects of proton donor-acceptor mode and solvent in nonadiabatic proton-coupled electron transfer. Abstracts of Papers of the American Chemical Society 229, U767
20. Antoniou D, Schwartz SD (2001) J Phys Chem B 105:5553
21. Kuznetsov AM, Ulstrup J (1999) Can J Chem 77:1085
22. Marcus RA, Sutin N (1985) Biochim Biophys Acta 811:265
23. Knapp MJ, Rickert K, Klinman JP (2002) J Am Chem Soc 124:3865
24. Knapp MJ, Klinman JP (2002) Eur J Biochem 269:3113
25. Meyer MP, Klinman JP (2005) Chem Phys 319:283
26. Johannissen LO, Hay S, Scrutton NS et al (2007) J Phys Chem B 111:2631
27. Hay S, Sutcliffe MJ, Scrutton NS (2007) Proc Natl Acad Sci USA 104:507
28. Hatcher E, Soudackov AV, Hammes-Schiffer S (2004) J Am Chem Soc 126:5763
29. Hatcher E, Soudackov AV, Hammes-Schiffer S (2007) J Am Chem Soc 129:187
30. Meyer MP, Tomchick DR, Klinman JP (2008) Proc Natl Acad Sci USA 105:1146
31. Basran J, Harris RJ, Sutcliffe MJ et al (2003) J Biol Chem 278:43973

32. Northrop DB (2002) Biochim Biophys Acta 1595:71
33. Isaacs NS (1984) In: Buncel E, Lee CC (eds) Isotope effects in organic chemistry. Elsevier, London
34. Hay S, Pudney CR, McGrory TA et al (2009) Angew Chem Int Ed Engl 48:1452
35. Pang J, Hay S, Scrutton NS et al (2008) J Am Chem Soc 130:7092
36. Alhambra C, Corchado J, Snchez ML et al (2001) J Phys Chem B 105:11326
37. Truhlar DG, Isaacson AD, Garret BC (1985) In: Baer M (ed) Theory of Chemical Reaction Dynamics. CRC Press, Boca Raton, FL
38. Gao J, Truhlar DG (2002) Annu Rev Phys Chem 53:467
39. Truhlar DG, Gao JL, Garcia-Viloca M et al (2004) Int J Quant Chem 100:1136
40. Fernandez-Ramos A, Ellingson BA, Garrett BC, Truhlar DG (2007) Rev Comput Chem 23:125
41. Caratzoulas S, Schwartz SD (2001) J Chem Phys 114:2910
42. Caratzoulas S, Mincer JS, Schwartz SD (2002) J Am Chem Soc 124:3270
43. Bahnson BJ, Colby TD, Chin JK et al (1997) Proc Natl Acad Sci USA 94:12797
44. Masgrau L, Roujeinikova A, Johannissen LO et al (2006) Science 312:237
45. Johannissen LO, Hay S, Scrutton NS et al (2007) J Phys Chem B 111:2631
46. Sessions RB, Dauber-Osguthorpe P, Osguthorpe DJ (1989) J Mol Biol 210:617
47. Johannissen LO, Scrutton NS, Sutcliffe MJ (2008) J R Soc Interface 5(Suppl 3):S225
48. Hothi P, Lee M, Cullis PM et al (2008) Biochemistry 47:183
49. Farnum MF, Magde D, Howell EE et al (1991) Biochemistry 30:11567
50. Radkiewicz JL, Brooks CL (2000) J Am Chem Soc 122:225
51. Agarwal PK, Billeter SR, Rajagopalan PT et al (2002) Proc Natl Acad Sci USA 99:2794
52. Agrawal PK, Billeter SR, Hammes-Schiffer S (2002) J Phys Chem B 106:3283
53. Mincer JS, Schwartz SD (2003) J Phys Chem B 107:366
54. Mincer JS, Schwartz SD (2003) J Prot Res 2:437
55. Olsson MH, Parson WW, Warshel A (2006) Chem Rev 106:1737
56. Olsson MH, Mavri J, Warshel A (2006) Philos Trans R Soc Lond B Biol Sci 361:1417

INDEX