

Theory and Applications of the Empirical Valence Bond Approach

Theory and Applications of the Empirical Valence Bond Approach

From Physical Chemistry to Chemical Biology

Edited by

Fernanda Duarte

Physical and Theoretical Chemistry Laboratory, University of Oxford, UK

Shina Caroline Lynn Kamerlin

Department of Cell and Molecular Biology, Uppsala University, Sweden

With a Foreword by Arieh Warshel

WILEY

This edition first published 2017
© 2017 John Wiley & Sons Ltd

Registered office:

John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

The right of the author to be identified as the author of this work has been asserted in accordance with the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by the UK Copyright, Designs and Patents Act 1988, without the prior permission of the publisher.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic books.

Designations used by companies to distinguish their products are often claimed as trademarks. All brand names and product names used in this book are trade names, service marks, trademarks or registered trademarks of their respective owners. The publisher is not associated with any product or vendor mentioned in this book.

Limit of Liability/Disclaimer of Warranty: While the publisher and authors have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. It is sold on the understanding that the publisher is not engaged in rendering professional services and neither the publisher nor the authors shall be liable for damages arising herefrom. If professional advice or other expert assistance is required, the services of a competent professional should be sought.

Library of Congress Cataloging-in-Publication Data

Names: Duarte, Fernanda, editor. | Kamerlin, Shina
Caroline Lynn - editor.

Title: Theory and applications of the empirical valence bond approach : from
physical chemistry to chemical biology / editors, Fernanda Duarte,
Shina Caroline Lynn Kamerlin ; with a foreword by Arieh Warshel.

Description: Chichester, West Sussex : John Wiley & Sons, Inc., 2017. |
Includes bibliographical references and index.

Identifiers: LCCN 2016035935 | ISBN 9781119245391 (cloth)

Subjects: LCSH: Valence (Theoretical chemistry) | Chemical processes.

Classification: LCC QD469 .T44 2017 | DDC 541/.224--dc23 LC record available at
<https://lccn.loc.gov/2016035935>

A catalogue record for this book is available from the British Library.

Cover image: Background - Getty Images
Center image – Courtesy of the author

9781119245391

Set in 10/12pt WarnockPro by SPi Global, Chennai, India

Contents

	List of Contributors	<i>xi</i>
	Foreword	<i>xiii</i>
	Acknowledgements	<i>xix</i>
1	Modelling Chemical Reactions Using Empirical Force Fields	1
	<i>Tibor Nagy and Markus Meuwly</i>	
1.1	Introduction	1
1.2	Computational Approaches	3
1.3	Molecular Mechanics with Proton Transfer	3
1.4	Adiabatic Reactive Molecular Dynamics	4
1.5	The Multi-Surface ARMD Method	6
1.6	Empirical Valence Bond	8
1.7	ReaxFF	9
1.8	Other Approaches	10
1.9	Applications	10
1.9.1	Protonated Water and Ammonia Dimer	10
1.9.2	Charge Transfer in $\text{N}_2 - \text{N}_2^+$	12
1.9.3	Vibrationally Induced Photodissociation of Sulfuric Acid	12
1.9.4	Proton Transfer in Malonaldehyde and Acetyl-Acetone	15
1.9.5	Rebinding Dynamics in MbNO	16
1.9.6	NO Detoxification Reaction in Truncated Hemoglobin (trHbN)	16
1.9.7	Outlook	18
	Acknowledgements	19
	References	19
2	Introduction to the Empirical Valence Bond Approach	27
	<i>Fernanda Duarte, Anna Pabis and Shina Caroline Lynn Kamerlin</i>	
2.1	Introduction	27
2.2	Historical Overview	28
2.2.1	From Molecular Mechanics to QM/MM Approaches	28
2.2.2	Molecular Orbital (MO) vs. Valence Bond (VB) Theory	29
2.3	Introduction to Valence Bond Theory	30
2.4	The Empirical Valence Bond Approach	32
2.4.1	Constructing an EVB Potential Surface for an $\text{S}_{\text{N}}2$ Reaction in Solution	33
2.4.2	Evaluation of Free Energies	36

2.5	Technical Considerations	38
2.5.1	Reliability of the Parametrization of the EVB Surfaces	38
2.5.2	The EVB Off-diagonal Elements	39
2.5.3	The Choice of the Energy Gap Reaction Coordinate	39
2.5.4	Accuracy of the EVB Approach For Computing Detailed Rate Quantities	40
2.6	Examples of Empirical Valence Bond Success Stories	40
2.6.1	The EVB Approach as a Tool to Explore Electrostatic Contributions to Catalysis: Staphylococcal Nuclease as a Showcase System	40
2.6.2	Using EVB to Assess the Contribution of Nuclear Quantum Effects to Catalysis	42
2.6.3	Using EVB to Explore the Role of Dynamics in Catalysis	42
2.6.4	Exploring Enantioselectivity Using the EVB Approach	43
2.6.5	Moving to Large Biological Systems: Using the EVB Approach in Studies of Chemical Reactivity on the Ribosome	44
2.7	Other Empirical Valence Bond Models	47
2.7.1	Chang-Miller Formalism	47
2.7.2	Approximate Valence Bond (AVB) Approach	47
2.7.3	Multistate Empirical Valence Bond (MS-EVB)	48
2.7.4	Multiconfiguration Molecular Mechanics (MCMM)	48
2.7.5	Other VB Approaches for Studying Complex Systems	49
2.8	Conclusions and Future Perspectives	50
	References	52

3 Using Empirical Valence Bond Constructs as Reference Potentials For High-Level Quantum Mechanical Calculations 63

Nikolay V. Plotnikov

3.1	Context	64
3.2	Concept	68
3.3	Challenges	69
3.3.1	Different Reference and Target Reaction Paths	69
3.3.2	Convergence of the Free Energy Estimates	70
3.4	Implementation of the Reference Potential Methods	71
3.4.1	Locating the Target Reaction Path	71
3.4.2	Low-accuracy Target Free Energy Surface from Non-equilibrium Distribution	71
3.4.3	Obtaining a Low-Accuracy Target Free Energy Surface from Free Energy Perturbation	72
3.4.4	Pre-Computing the Reaction Path	73
3.4.5	Reference Potential Refinement: the Paradynamics Model	74
3.4.6	Moving From the Reference to the Target Free Energy Surface at the TS Using Constraints on the Reaction Coordinate	74
3.4.7	High-Accuracy Local PMF Regions from Targeted Sampling	76
3.4.8	Improving Accuracy of Positioning the Local PMF Regions	77
3.5	EVB as a Reference Potential	77
3.5.1	EVB Parameter Refinement	80
3.5.2	EVB Functional Refinement	81
3.6	Estimation of the Free Energy Perturbation	82

3.6.1	Exponential Average	83
3.6.2	Linear Response Approximation (LRA)	84
3.6.3	Bennet's Acceptance Ratio	84
3.6.4	Free Energy Interpolation	85
3.7	Overcoming Some Limitations of EVB Approach as a Reference Potential	86
3.8	Final Remarks	86
	References	87
4	Empirical Valence Bond Methods for Exploring Reaction Dynamics in the Gas Phase and in Solution	93
	<i>Jeremy N. Harvey, Michael O'Connor and David R. Glowacki</i>	
4.1	Introduction	93
4.2	EVB and Related Methods for Describing Potential Energy Surfaces	94
4.3	Methodology	97
4.4	Recent Applications	100
4.4.1	Cl + CH ₄ in the Gas Phase	100
4.4.2	CN + <i>c</i> -C ₆ H ₁₂ (CH ₂ Cl ₂ Solvent)	102
4.4.3	CN + Tetrahydrofuran (Tetrahydrofuran Solvent)	103
4.4.4	F + CD ₃ CN (CD ₃ CN Solvent)	104
4.4.5	Diazocyclopropane Ring Opening	107
4.5	Software Implementation Aspects	108
4.5.1	CPU Parallelization Using MPI	109
4.5.2	GPU Parallelization	111
4.6	Conclusions and Perspectives	115
	References	117
5	Empirical Valence-Bond Models Based on Polarizable Force Fields for Infrared Spectroscopy	121
	<i>Florian Thauhay, Florent Calvo, Gilles Ohanessian and Carine Clavaguéra</i>	
5.1	Introduction	121
5.2	Infrared Spectra of Aspartate and Non-Reactive Calculations	123
5.2.1	Experimental Approach	123
5.2.2	Quantum Chemical Calculations	124
5.2.3	Finite Temperature IR Spectra Based on AMOEBA	126
5.2.3.1	The AMOEBA Force Field	126
5.2.3.2	Infrared Spectra From Molecular Dynamics Simulations	126
5.2.3.3	Role of the Multipoles	127
5.3	Empirical Valence-Bond Modeling of Proton Transfer	130
5.3.1	Two-State EVB Model	130
5.3.1.1	Implementation of EVB Model with AMOEBA	131
5.3.1.2	Coupling Between Diabatic States	131
5.3.2	Dynamics Under the EVB-AMOEBA Potential	133
5.3.3	Infrared Spectra with the EVB-AMOEBA Approach	136
5.4	Concluding Remarks	140
	Acknowledgements	140
	References	140

6	Empirical Valence Bond Simulations of Biological Systems	145
	<i>Avital Shurki</i>	
6.1	Introduction	145
6.2	EVB as a Tool to Unravel Reaction Mechanisms in Biological Systems	147
6.2.1	Hydrolysis of Organophosphate Compounds in BChE	147
6.2.2	Hydrolysis of GTP in Ras/RasGAP	150
6.3	EVB a Comparative Tool	152
6.3.1	Guided Reaction Paths	152
6.3.2	Studies of the Same Reaction in Different Environments	155
6.3.2.1	The Effect of Conformational Changes	155
6.3.2.2	Mutational Studies	156
6.4	EVB – A Sampling Tool	157
6.4.1	EVB – An Efficient Way to Run an Enormous Number of Calculations	157
6.4.2	EVB – An Efficient Way to Sample Conformations for Other QM/MM Approaches	159
6.4.2.1	Copper-Chaperones	159
6.4.2.2	Hybrid Ab Initio VB/MM Approach	161
6.4.2.3	EVB – An Efficient Reference Potential	161
6.5	EVB Provides Simple Yet Superior Definition of Reaction Coordinate	163
6.6	EVB – A Tool with Great Insight	164
6.7	Concluding Remarks	166
	Acknowledgements	166
	References	166
7	The Empirical Valence Bond Approach as a Tool for Designing Artificial Catalysts	173
	<i>Monika Fuxreiter and Letif Mones</i>	
7.1	Introduction	173
7.2	Proposals for the Origin of the Catalytic Effect	174
7.3	Reorganization Energy	177
7.4	Conventional In Silico Enzyme Design	179
7.5	Computational Analysis of Kemp Eliminases	183
7.6	Using the Empirical Valence Bond Approach to Determine Catalytic Effects	184
7.6.1	General EVB Framework	184
7.6.2	Computing Free Energy Profiles Within the EVB Framework	185
7.7	Computing the Reorganization Energy	186
7.8	E_{gap} : A General Reaction Coordinate and its Application on Other PES	187
7.9	Contribution of Individual Residues	189
7.10	Improving Rational Enzyme Design by Incorporating the Reorganization Energy	190
7.11	Conclusions and Outlook	191
	Acknowledgements	193
	References	193

8	EVB Simulations of the Catalytic Activity of Monoamine Oxidases: From Chemical Physics to Neurodegeneration	199
	<i>Robert Vianello and Janez Mavri</i>	
8.1	Introduction	199
8.2	Pharmacology of Monoamine Oxidases	200
8.3	Structures of MAO A and MAO B Isoforms	201
8.4	Mechanistic Studies of MAO	202
8.5	Cluster Model of MAO Catalysis	204
8.6	Protonation States of MAO Active Site Residues	211
8.7	EVB Simulation of the Rate Limiting Hydride–Abstraction Step for Various Substrates	215
8.8	Nuclear Quantum Effects in MAO Catalysis	218
8.9	Relevance of MAO Catalyzed Reactions for Neurodegeneration	221
8.10	Conclusion and Perspectives	223
	Acknowledgements	223
	References	224
	Index	233

List of Contributors

Florent Calvo

LiPhy, Université Grenoble I and CNRS
France

Carine Clavaguéra

LCM, CNRS, Ecole polytechnique
Université Paris Saclay
Palaiseau
France

Fernanda Duarte

Physical and Theoretical Chemistry
Laboratory
University of Oxford
United Kingdom

Monika Fuxreiter

MTA-DE Momentum Laboratory of Protein Dynamics
Department of Biochemistry and Molecular Biology
University of Debrecen
Hungary

David Glowacki

School of Chemistry
University of Bristol
United Kingdom

Jeremy Harvey

Department of Chemistry
KU Leuven
Belgium

Shina Caroline Lynn Kamerlin

Department of Cell and Molecular Biology
Uppsala University
Sweden

Janez Mavri

Laboratory for Biocomputing and Bioinformatics
National Institute of Chemistry
Ljubljana
Slovenia

Markus Meuwly

Department of Chemistry
University of Basel
Switzerland

Letif Mones

Department of Engineering
University of Cambridge
United Kingdom

Tibor Nagy

IMEC
RCNS
Hungarian Academy of Sciences
Budapest
Hungary

Michael O'Connor

School of Chemistry
University of Bristol
United Kingdom

Gilles Ohanessian

LCM, CNRS, Ecole polytechnique
Université Paris Saclay
Palaiseau
France

Anna Pabis

Department of Cell and Molecular Biology
Uppsala University
Sweden

Nikolay Plotnikov

Department of Chemistry
Stanford University
United States

Avital Shurki

Institute for Drug Research
School of Pharmacy
The Hebrew University of Jerusalem
Israel

Florian Thauway

LCM, CNRS
Ecole polytechnique
Université Paris Saclay
Palaiseau
France

Robert Vianello

Quantum Organic Chemistry Group
Ruđer Bošković Institute
Zagreb
Croatia

Foreword

Arieh Warshel

Department of Chemistry, University of Southern California, Los Angeles, USA

The EVB Approach as a Powerful Tool for Simulating Chemical and Biological Processes

The search for reliable yet practical approach for modeling reactions in condensed phases and enzymes led to the inception of the empirical valence bond (EVB) approach around 1980. The idea for this approach emerged from the realization that the use of molecular orbitals (MO) based hybrid quantum mechanical/molecular mechanical (QM/MM) approaches faces major problems when it comes to obtaining the proper asymptotic energetics for the autodissociation of water,^[1] while the corresponding valence bond (VB) representation provides an excellent way of imposing the correct physics on the system. This idea was initially formulated in 1980,^[2] using a simplified Langevin dipoles solvent model, which led to the need for a conceptual description of the response of the solvent to the different VB states. A much more rigorous coupling to the solvent was introduced in 1988,^[3] with an all-atom molecular dynamics treatment that included the free energy functional as well as a rigorous non-equilibrium solvation treatment. The main remaining fundamental problem therefore was the validation of the reasonable (but *ad hoc*) assumption about the transferability of the off-diagonal elements (that couple different resonance structures) between different phases; for example between vacuum to aqueous solution or an enzyme active site. This assumption has been numerically validated by means of constrained density functional theory (CDFT) studies.^[4] Overall, it appears that despite its seemingly oversimplified features, the EVB approach provides a very valid theoretical QM/MM framework that incorporates the environment in arguably the most physically meaningful way. Furthermore, the EVB approach can be systematically improved by the paradynamics approach,^[5] and by constraining it to reproduce experimental results in reference systems (while moving to other systems). The power of EVB is largely due to its “simple” orthogonal diabatic representation, as well as the assumption that the off-diagonal elements of the EVB Hamiltonian do not change significantly upon transfer of the reacting system from one phase to another.

Overall, therefore, despite unjustified criticism (see *e.g.*, ref. ^[6]), the EVB approach has become widely used with an increasing recognition of its potential as a very powerful way of simulating chemical processes in different environments. This book includes

chapters that consider different features of the EVB and its successful applications to complex chemical and biological problems. The different chapters presented in this book are briefly considered below.

In Chapter 1, *Nagy and Meuwly* describe reactive force field-based approaches for studies of bond breaking/making chemical reactions, including the EVB, ARMD, and MS-ARMD methods.

Particular emphasis is put on enabling investigations of the dynamics of such reactions. In this respect, we note that the EVB approach is arguably still the most powerful approach for studying the dynamics of reactions in the condensed phase, due to the consistent incorporation of the effect of solvent, which facilitates, among other special features, the consistent exploration of nonequilibrium solvation effects.

In Chapter 2, *Duarte et al.* provide a historical overview of the use of both MO and VB methods in the context of (bio)molecular modeling, introducing the basic theoretical aspects of both approaches. Particular emphasis is put on the EVB approach, following the overall theme of this book. This chapter exemplified the power of the EVB approach for studying challenging chemical processes in both the condensed phase and in enzymes. It concludes with an overview of further opportunities for utilizing the EVB framework, in combination with other approaches, for the study of enzymatic reactions.

In Chapter 3, *Nikolay Plotnikov* describes the paradynamics (PD) approach, showing how we can conveniently move from the EVB approach to high level *ab initio* surfaces. This method provides a very powerful way of obtaining the free energy surface for *ab initio* potentials, since the EVB presents an ideal reference potential for the *ab initio* surfaces.

In Chapter 4, *Harvey et al.* discuss the use of the EVB approach to exploring reaction dynamics in the gas and condensed phases. This chapter considers some of the relevant background and practical applications. The authors also discuss the ability of the EVB to explore short timescale dynamical effects, and discuss some applications, chosen to highlight the power of the method.

In considering the use of EVB in modeling dynamical effects, it is useful to add that the ability to explore not just short but also long timescale dynamical effects is particularly important in exploring the proposal that special “dynamical” effects play a major role in enzyme catalysis (*e.g.*, refs. [7,8]), which has become quite popular in recent years (*e.g.*, refs. [9–11]). However, a significant part of this popularity is a reflection of confusion with regards to the nature of dynamical effects. Combining the EVB approach with coarse-grained (CG) modeling allows one to explore the dynamical proposal that has been discussed in great length in several recent reviews.^[12–14] These reviews (and related works) have shown that enzyme catalysis is not due to dynamical effects, regardless of the definition used. In this respect we note that the recently developed approach^[15] has allowed us to use a CG model to simulate effective millisecond trajectories in the conformational and chemical coordinates, establishing that the conformational kinetic energy is fully randomized before it can be transfer to the chemical coordinate.^[15] Thus it had been determined that dynamical effects cannot be used to accelerate enzymatic reactions. It is also useful to note that the EVB approach is arguably the most effective approach for long timescale all atom simulations as it allows for the exploration of dynamical effect on quite long situation timescales with reasonable computational power.

In Chapter 5, *Thaunay et al.* describe the combination of the EVB approach with the AMOEBA polarizable force field, and demonstrate the performance of the resulting model in reproducing experimentally observed spectra. In this respect, we note that the EVB has originally been formulated with a polarizable force field considering both the induced dipoles of the solute and solvent.^[2]

In Chapter 6, Avital Shurki describes the applications of the EVB approach in studies of biological reactions. It is pointed out that the convenient and reliable calibration of the EVB approach provides a great advantage relative to other QM/MM approaches in terms of elucidating catalytic effects. Furthermore, the simplicity of the potential energy surface enables highly efficient sampling, which is important when particularly large systems or averages over considerably large conformational ensembles are of interest. Additionally, the EVB approach also provides simple definition of the reaction coordinate, which includes all the system's degrees of freedom. Finally, the method benefits from the valence bond character of the wavefunction, which includes easily accessible chemical insight. The review discusses the different capabilities of the method while highlighting the advantages of the method over other standard (MO based) QM/MM approaches.

In Chapter 7, *Fuxreiter and Mones* discuss the potential of the EVB approach in enzyme design, emphasizing in particular the use of the reorganization energy as a screening tool for predicting catalytic effects of enzymes. The ability to design effective enzymes presents one of the most fundamental challenges in biotechnology, and such ability would provide convincing manifestations of a general understanding of the origin of enzyme catalysis. A recent study^[16] explored the reliability of different simulation approaches in terms of their ability to rank different possible active-site constructs. This study demonstrated that the EVB approach is a practical and reliable quantitative tool in the final stages of computer aided-enzyme design, while other approaches were found to be comparatively less accurate, and mainly useful for the qualitative screening of ionized residues. The most obvious problem arises from the fact that current design approaches (*e.g.*, refs. ^[17–19]) are not based on modeling the chemical process in the enzyme active site. In fact, some approaches (*e.g.*, ref. ^[20]) use gas phase or small model cluster calculations, which then estimate the interaction between the enzyme and the transition state model, rather than the transition state binding free energy (or the relevant activation free energy). However, accurate ranking of the different options for enzyme design cannot be accomplished by approaches that cannot capture the electrostatic preorganization effect. Clearly, the ability of the EVB model to act as a quantitative tool in the final stages of computer-aided enzyme design is a major step towards the design of enzymes whose catalytic power is closer to native enzymes than the current generation of designer enzymes. It should be noted, however, that despite the temptation to use reorganization energies in the screening process there are many cases^[16] when it is essential to invest the additional computational time and to evaluate the full EVB free energy surfaces to obtain the relevant activation barriers.

In Chapter 8, *Vianello and Mavri* describe EVB simulations of the catalytic activity of monoamine oxidases (MAOs) in controlling neurodegeneration. The use of the EVB approach to study the reaction of MAOs appears to be very useful, and could, in principle, help to develop strategies for the prevention and treatment of neurodegeneration, including the design of irreversible MAO inhibitors.

Overall the book presents a compelling case for the general use of the EVB approach as a very effective computational and conceptual tool for studies of large complex (bio)chemical systems. This includes simulations of the reactivity of macromolecules and the modeling of general chemical processes in the condensed phase. This book demonstrates that the EVB approach provides a powerful way to connect the classical concepts of physical organic chemistry with the actual energetics of enzymatic reactions by means of computation. That is, when concepts such as Marcus' parabolas are formulated in a consistent microscopic way, they allow one to obtain quantitative linear free energy relationships in enzymes and in solution, which in turn allows one to quantify catalytic effects and to define them in terms of the relevant reaction free energies, reorganization energies and the preorganization of the enzyme active sites. Thus, we believe that the EVB method is probably the most powerful current simulation strategy as far as studies of chemical processes in the condensed phase in general and in enzymes in particular are involved. This ability is especially important in the exploration of the origin of enzyme catalysis, which, even in 2017, remains one of the Holy Grails of biochemistry.

Acknowledgements

This work was supported by NIH grants GM-24492 and GM-40283.

References

- 1 Warshel, A. (1979) Calculations of chemical processes in solutions. *Journal of Physical Chemistry*, **83**, 1640–1652.
- 2 Warshel, A. and Weiss, R.M. (1980) An empirical valence bond approach for comparing reactions in solutions and in enzymes. *Journal of the American Chemical Society*, **102**, 6218–6226.
- 3 Warshel, A., Sussman, F. and Hwang, J.-K. (1988) Evaluation of catalytic free energies in genetically modified proteins. *Journal of Molecular Biology*, **201**, 139–159.
- 4 Hong, G., Rosta, E. and Warshel, A. (2006) Using the constrained DFT approach in generating diabatic surfaces and off-diagonal empirical valence bond terms for modeling reactions in condensed phases. *Journal of Physical Chemistry B*, **110**, 19570–19574.
- 5 Plotnikov, N.V. and Warshel, A. (2012) Exploring, refining, and validating the paradiynamics QM/MM sampling. *Journal of Physical Chemistry B*, **116**, 10342–10356.
- 6 Kamerlin, S.C.L., Cao, J., Rosta, E. and Warshel, A. (2009) On unjustifiably misrepresenting the EVB approach while simultaneously adopting it. *Journal of Physical Chemistry B*, **113**, 10905–10915.
- 7 Careri, G., Fasella, P. and Gratton, E. (1979) Enzyme dynamics: The statistical physics approach. *Annual Review of Biophysics and Bioengineering*, **8**, 69–97.
- 8 Karplus, M. and McCammon, J.A. (1983) Dynamics of proteins: Elements and function. *Annual Review of Biochemistry*, **53**, 263–300.
- 9 Klinman, J.P. and Kohen, A. (2013) Hydrogen tunneling links protein dynamics to enzyme catalysis. *Annual Review of Biochemistry*, **82**, 471–496.

- 10 Henzler-Wildman, K.A. *et al.* (2007) Intrinsic motions along an enzymatic reaction trajectory. *Nature*, **450**, 838–844.
- 11 Bhabha, G. *et al.* (2011) A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science*, **332**, 234–238.
- 12 Villà, J. and Warshel, A. (2001) Energetics and dynamics of enzymatic reactions. *Journal of Physical Chemistry B*, **105**, 7887–7907.
- 13 Warshel, A. and Parson, W.W. (2001) Dynamics of biochemical and biophysical reactions: Insight from computer simulations. *Quarterly Review of Biophysics*, **34**, 563–670.
- 14 Olsson, M.H.M., Parson, W.W. and Warshel, A. (2006) Dynamical contributions to enzyme catalysis: Critical tests of a popular hypothesis. *Chemical Reviews*, **106**, 1737–1756.
- 15 Liu, H., Shi, Y., Chen, X.S. and Warshel, A. (2009) Simulating the electrostatic guidance of the vectorial translocations in hexameric helicases and translocases. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 7449–7454.
- 16 Roca, M., Vardi-Kilshtain, A. and Warshel, A. (2009) Toward accurate screening in computer-aided enzyme design. *Biochemistry*, **48**, 3046–3056.
- 17 Jiang, L. *et al.* (2008) De novo computational design of retro-aldol enzymes. *Science*, **319**, 1387–1391.
- 18 Rothlisberger, D. *et al.* (2008) Kemp elimination catalysts by computational enzyme design. *Nature*, **453**, 190–U194.
- 19 Lippow, S.M. and Tidor, B. (2007) Progress in computational protein design. *Current Opinion in Biotechnology*, **18**, 305–311.
- 20 Jiang, L. *et al.* (2008) De novo computational design of retro-aldol enzymes. *Science*, **319**, 1387–1391.

Acknowledgements

We would like to express our gratitude to all authors for their excellent contributions to this book, in order to make it a valuable resource for both experts and newcomers to the field. We would also like to thank Anna Pabis for her help with proofreading the book. Finally, despite the varied contributions to our edited volume, these comprise only a small fraction of the work of a much larger community. Therefore, we would like to dedicate this book to the great number of scientists that have contributed (and are continuing to contribute) to the development and renaissance of valence bond theory in its different flavors.

1

Modelling Chemical Reactions Using Empirical Force Fields

Tibor Nagy¹ and Markus Meuwly²

¹IMEC, RCNS, Hungarian Academy of Sciences, Budapest, Hungary

²Department of Chemistry, University of Basel, Switzerland

1.1 Introduction

Chemical reactions involve bond-breaking and bond-forming processes and are fundamental in chemistry and the life sciences in general. In many cases, mechanistic aspects of the reactions (“which reaction partners interact at which time with each other”) are of interest. However, many atomistic aspects in bond-breaking and bond-forming processes remain elusive by considering experimental data alone because “the reaction” itself is a transient process. The transition state is unstable and short-lived. Thus, the most interesting regions along a reaction path can not be investigated experimentally in a direct fashion. To shed light on such questions, theoretical and computational work has become invaluable to experimental efforts in understanding particular reaction schemes.

The computational investigation of a chemical or biological system requires models to compute the total energy of the system under investigation. There are two fundamentally different concepts to do that: either by solving the electronic Schrödinger equation, or by assuming a suitably defined empirical potential energy function. The first approach has been refined to a degree that allows one to carry out calculations with “chemical accuracy” – that is, accuracies for relative energies within 1 kcal/mol for the chemically bonded region and less accurately for transition state regions. Most importantly, a quantum chemical calculation makes no assumption on the bonding pattern in the molecule and is ideally suited to answer the question which atoms are bonded to one another for a particular relative arrangement of the atoms. To obtain realistic reaction profiles it is, however, necessary to carry out calculations at a sufficiently high level of theory, particularly in the region of the transition state. Through statistical mechanics and assuming idealized models of molecular motion such as rigid rotor or harmonic oscillator, average internal energies, enthalpies, and by including entropic effects, also 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, that is, small systems in the gas phase. This is due to the N^3 scaling of the secular determinants that need to be diagonalized, where N is the number of basis functions.

Alternative approaches to solving the electronic Schrödinger equation have been developed and matured to similar degrees. London’s work on the $H + H_2$ reaction for

which he used a 2×2 valence bond treatment^[1] is an early example for this. Further 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] Empirical relationships such as the one between bond length and bond order later became foundations to empirical descriptions of reactivity.^[8,9]

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

$$\begin{aligned} V_{\text{bond}} &= \sum K_b (r - r_e)^2 \\ V_{\text{angle}} &= \sum K_\theta (\theta - \theta_e)^2 \\ V_{\text{dihe}} &= \sum K_\phi (1 + \cos(n\phi - \delta)) \end{aligned} \quad (1.1)$$

of empirical force fields is thus not suitable to describe chemical reactions where chemical bonds are broken and formed. Here, 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

$$\begin{aligned} V_{\text{elstat}} &= \frac{1}{4\pi\epsilon_0} \sum \frac{q_i q_j}{r_{ij}} \\ V_{\text{vdW}} &= \sum \epsilon_{ij} \left[\left(\frac{R_{\text{min},ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{\text{min},ij}}{r_{ij}} \right)^6 \right] \end{aligned} \quad (1.2)$$

where the sums run over 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 $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$ and range $R_{\text{min},ij} = (R_{\text{min},i} + R_{\text{min},j})/2$ 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. 1.1 and 1.2 constitutes a minimal model for a force field (FF).

An important step to investigate reactions by simulation methods has been the introduction of mixed quantum mechanical/classical mechanics methods (QM/MM).^[18–20] 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 FF. The majority of applications of QM/MM methods to date use semiempirical (such as AM1, PM3,^[21] SCC-DFTB^[22,23]) or DFT methods. Typically, the QM part contains several tens of atoms. It should also be noted that studies of reactive processes in the condensed phase often employ energy evaluations along the pre-defined progression coordinates,^[21,24] that is, the system is forced to move along a set of more

or less well-suited coordinates. One of the main reasons why *ab initio* QM/MM calculations are not yet used routinely in fully quantitative studies is related to the fact that the energy and force evaluations for the QM region are 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 by combining them in suitable ways. They include RMD (Reactive Molecular Dynamics),^[25–27] EVB (Empirical Valence Bond)^[28] and its variants AVB (Approximate Valence Bond),^[29] and MCMM (Multiconfiguration Molecular Mechanics).^[30]

Force field-based treatments of chemical reactivity start from conventional FFs and employ the diabatic picture of electronic states to define reactant and product states.^[31] From a FF perspective, in a diabatic state the connectivity of the atoms does not change. Low-amplitude vibrations and conformational motion in these states can be efficiently described by conventional FFs. However, they yield very high potential energies, far from their equilibrium geometry, due to their functional form and parametrization. For example, force field evaluation of a chemical bond at its equilibrium geometry for the unbound state, in which the bonded term is replaced by electrostatic and van der Waals interactions, yields a very high energy for the unbound state due to van der Waals repulsion. This large energy difference can be exploited to define a dominant force field which is that with the lowest energy for almost all accessible configurations and makes the energy difference a useful coordinate. Other methods use geometric formulas to switch on and off interactions individually (e.g., ReaxFF). The various methods differ mainly in the choice of switching method and parameters.

The present chapter describes adiabatic reactive molecular dynamics (ARMD),^[27,32] its multi-surface variant (MS-ARMD)^[33–35] and molecular mechanics with proton transfer (MMPT).^[36] All three methods have been developed with the aim to combine the accuracy of quantum methods and the speed of FF simulations such that the processes of interest can be sampled in a statistically meaningful manner. This allows one to determine suitable averages, which can be then compared with experimental data. The chapter first discusses the three methods and briefly highlights similarities and differences to other methods, which are separately discussed in the present volume. Then, topical applications are presented and an outlook laid out future avenues.

1.2 Computational Approaches

In the following chapter 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 coordinate of a chemical reaction. Excluded from this discussion are nonadiabatic effects and quantum dynamics.

1.3 Molecular Mechanics with Proton Transfer

Molecular Mechanics with Proton Transfer (MMPT) is a parametrized method to follow bond breaking and bond formation between a hydrogen atom (or a proton) and

its donor and acceptor, respectively.^[36] The total interaction energy for the system with coordinates \mathbf{Q} is

$$V(\mathbf{Q}) = V_{\text{MM}}(\mathbf{q}) + V_{\text{PT}}(R, r, \theta), \quad (1.3)$$

where the proton transfer motif D-H-A with donor (D) and acceptor (A) is described by V_{PT} . This contribution is determined from quantum chemical calculations 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 dependence of the total potential energy on the remaining degrees of freedom of the system (\mathbf{q}) is given by a conventional force field V_{MM} . The resulting potential is called Molecular Mechanics with Proton Transfer (MMPT).^[36] 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 $f_{\text{sw}}(R, r) = 0.5(\tanh(2R(r - R/2)) + 1)$, whenever the migrating H attempts to transfer from donor to acceptor.^[36]

MMPT treats the proton transfer process with its full dimensionality while addressing three important aspects of the problem: speed, accuracy, and versatility. While speed and accuracy are rooted in the QM/MM formulation, the versatility of the approach is exploited by using the morphing potential method.^[37] 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).^[36] The potential energy surface (PES) of these three model systems, fitted to suitable 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

$$V_{\text{morph}}(R', r', \theta') = \lambda(R, r, \theta) V_{\text{orig}}(R, r, \theta), \quad (1.4)$$

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

1.4 Adiabatic Reactive Molecular Dynamics

In the ARMD^[27,38] simulation method, which is implemented in CHARMM^[39] (since v35b2), at least two parametrized PESs, V_1 for the reactant and V_2 for the product states, are considered. The adiabatic dynamics of the nuclei takes place on the lowest PES while the energy of the higher states is also determined. Whenever the energy of the current state equals that of a higher state, the simulation is restarted from a few fs ($t_s/2$) prior to the detected crossing and during time interval t_s (twice as long), called switching

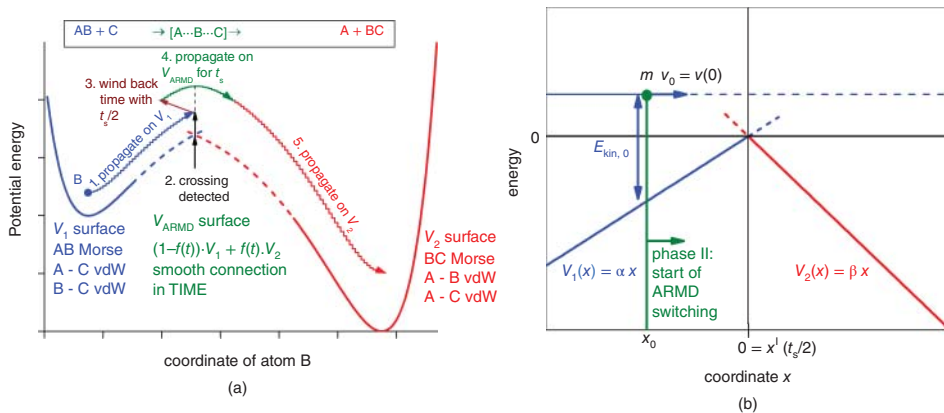


Figure 1.1 (a) The ARMD Method: Schematic Figure of the ARMD Simulation Method for a Collinear Reaction, Where Atom B is Transferred from Donor Atom A to Acceptor Atom C. During crossing the surfaces are switched in time and the Morse bond is replaced by van der Waals (vdW) interactions and vice versa. (b) Simple model for estimating energy violation in ARMD simulations. The system with mass m is approaching from the left on PES $V_1(x)$ (phase I). At $t = 0$ time it is at x_0 with velocity v_0 and kinetic energy $E_{\text{kin},0}$. After crossing is detected at $x = 0$ the time is rewound by $t_s/2$ and the dynamics is re-simulated while $V_1(x)$ is being switched to $V_2(x)$ in t_s (phase II). (See color plate section for the color representation of this figure.)

time, the PESs are mixed in different proportions by multiplying them with a suitable time-dependent smooth switching function $f(t)$ (e.g., a tanh function).^[27,38]

$$V_{\text{ARM D}}(x, t) = (1 - f(t))V_1(x) + f(t)V_2(x). \quad (1.5)$$

At the beginning of the mixing the system is fully in state 1 ($f(0) = 0$), while at the end it is fully in state 2 ($f(t_s) = 1$). The algorithm of ARMD is schematically shown for a collinear atom transfer reaction in Figure 1.1a.

As during surface crossing the ARMD potential energy is explicitly time-dependent, the total energy of the system can not be conserved in a strict sense. For large systems (e.g., proteins in solution) the total energy was found to be conserved to within ≈ 1 kcal/mol which is sufficient for most applications. This allowed successful application of ARMD simulation method to the investigation of rebinding dynamics of NO molecule in myoglobin^[27,38] the dioxygenation of NO into NO_3^- by oxygen-bound truncated hemoglobin.^[40]

However, for highly energetic reactions of small molecules in the gas phase this is not necessarily true. This was the case for vibrationally induced photodissociation of H_2SO_4 .^[35,41] If, however, several crossings between the states involved can take place or the course of the dynamics after the reaction is of interest – for example, for a final state analysis – energy conservation becomes crucial. The magnitude of energy violation ΔE for a simple 1D system (see Figure 1.1b) with effective mass m crossing between two linear potentials $V_1(x) = \alpha x$ and $V_2(x) = \beta x$ using a linear switching function $f(t) = t/t_s$ is:^[34]

$$\Delta E = \frac{\beta(\alpha - \beta)t_s^2}{24m}. \quad (1.6)$$

Hence, exact or nearly exact energy conservation, $\Delta E \approx 0$, can be achieved with ARMD (a) if the steepness of the two PESs along the trajectory during crossing are the same

($\alpha \approx \beta$), (b) if the second surface has a small slope ($\beta \approx 0$) in the crossing region thus accidental cancellation of violations can occur, (c) if the system has a large effective mass, which is often true for biomolecular systems, where both partners are heavy or the reaction is accompanied by the rearrangement of solvation shell involving many solvent molecules, (d) if the switching time is short, however, for $t_s \rightarrow 0$ the connection between the PESs will be unphysically sharp and thus fixed-stepsizes integrators fail to conserve energy.

ARMD involves two or multiple PESs defined by individual sets of force-field parameters. For macromolecular systems, the number of energy terms by which the PESs 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 states of interest with limited computational overhead.^[38] Because the FFs for the individual states are separately parametrized, they need to be related to each other by an offset Δ which puts the asymptotic energy differences between the states in the correct order.^[38]

1.5 The Multi-Surface ARMD Method

In the multi-surface (MS) variant of ARMD, the effective potential energy is also a linear combination of n PESs, however with coordinate-dependent weights $w_i(\mathbf{x})$, thereby the total energy is conserved during crossing.

$$V_{\text{MS-ARMD}}(\mathbf{x}) = \sum_{i=1}^n w_i(\mathbf{x}) V_i(\mathbf{x}) \quad (1.7)$$

The $w_i(\mathbf{x})$ are obtained by renormalizing the raw weights $w_{i,0}(\mathbf{x})$, which were calculated by using a simple exponential decay function of the energy difference between surface i and the minimum energy surface with over a characteristic energy scale ΔV (switching parameter).

$$w_i(\mathbf{x}) = \frac{w_{i,0}(\mathbf{x})}{\sum_{i=1}^n w_{i,0}(\mathbf{x})} \quad \text{where } w_{i,0}(\mathbf{x}) = \exp\left(-\frac{V_i(\mathbf{x})}{\Delta V}\right) \quad (1.8)$$

Only those surfaces will have significant weights, whose energy is within a few times of ΔV from the lowest energy surface. The performance of MS-ARMD is demonstrated for crossings of 1D and 2D surfaces in Figure 1.2. A smooth global surface is obtained everywhere, even in regions where more than two surfaces get close in energy.

The CHARMM^[39] implementation (available from v39a2) of MS-ARMD allows adding/removal and reparametrization of terms in any conventional force field, thus it can define new states and can join them into a reactive surface. Morse potentials and generalized Lennard-Jones potential (MIE potential^[42,43]) are also available in the implementation in order to improve the simultaneous description of PES regions close to the equilibrium and the crossing zone. Furthermore, as the energy of each force field is measured from its own global minimum, an additive constant has to be defined for bringing each force field to a common energy scale to reproduce reaction energies.

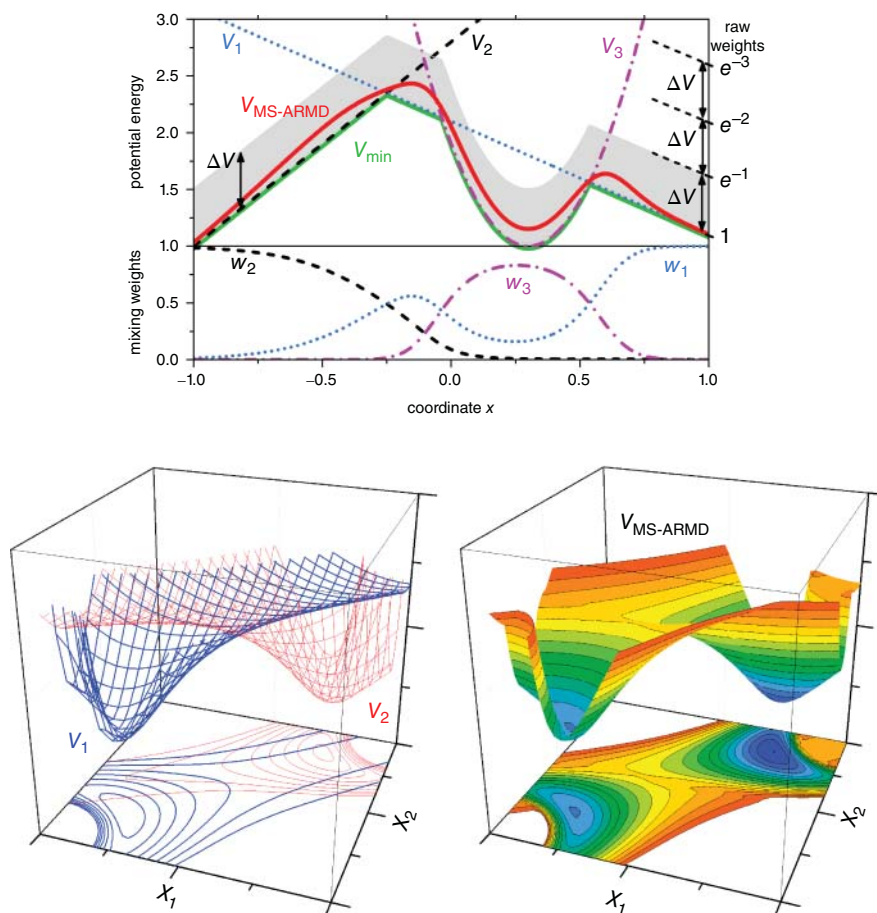


Figure 1.2 MS-ARMD Switching: The MS-ARMD Switching Method Applied in One and Two Dimensions to 3 and 2 Surfaces ($V_{1,2,3}$). The effective surface ($V_{MS-ARMD}$) always close to the lowest-energy surface (V_{min}), except for regions where other surfaces are within a few times ΔV (here = 0.5) in energy. Here, the algorithm switches smoothly among them by varying their weights ($w_{1,2,3}$; lower left panel of the top figure). (See color plate section for the color representation of this figure.)

Force fields separately optimized for reactant and product states sometimes predict an unrealistic, high-energy crossing point. According to MS-ARMD the transition point between two PESs has a weight of 0.5 from both contributing states. In order to adjust and reshape the barrier region to match energies obtained from electronic structure calculations, products of Gaussian and polynomial functions (GAPOs) ($k = 1, \dots, n_{ij}$) of the energy difference $\Delta V_{ij}(\mathbf{x}) = V_j(\mathbf{x}) - V_i(\mathbf{x})$ can be applied acting between any two surfaces (i and j).

$$\Delta V_{GAPO,k}^{ij}(\mathbf{x}) = \exp\left(-\frac{(\Delta V_{ij}(\mathbf{x}) - V_{ij,k}^0)^2}{2\sigma_{ij,k}^2}\right) \cdot \sum_{l=0}^{m_{ij,k}} a_{ij,kl} (\Delta V_{ij}(\mathbf{x}) - V_{ij,k}^0)^l \quad (1.9)$$

Here, $V_{ij,k}^0$ and $\sigma_{ij,k}$ denote the center and the standard deviation of the Gaussian function, respectively. Whenever the energy difference between the two PESs deviates from $V_{ij,k}^0$ more than a few times of $\sigma_{ij,k}$, the corresponding GAPO functions will be negligible provided that $V_{ij,k}^0$ and $\sigma_{ij,k}$ are small. The global MS-ARMD PES with this extension is a weighted sum of PESs and GAPO functions scaled with the sum of the weights of the two corresponding surfaces:

$$V_{\text{MS-ARMD}}(\mathbf{x}) = \sum_{i=1}^n w_i(\mathbf{x}) V_i(\mathbf{x}) + \sum_{i=1}^{n-1} \sum_{j=i+1}^n [w_i(\mathbf{x}) + w_j(\mathbf{x})] \sum_{k=1}^{n_{ij}} \Delta V_{\text{GAPO},k}^{ij}(\mathbf{x}) \quad (1.10)$$

The CHARMM implementation^[39] of MS-ARMD is a general tool for constructing global potential energy surfaces from empirical force fields for modelling chemical reactions in gas, surface and condensed phases, relevant to homogeneous, heterogeneous and enzymatic catalysis.

1.6 Empirical Valence Bond

One of the established methods to investigate chemical reactions based on empirical force fields is the empirical valence bond (EVB) method.^[28,44] EVB starts from the fact that valence bond states are suitable to distinguish between ionic and covalent resonance forms of a chemical bond which reflects chemical intuition. Since the environment of a chemical reaction primarily interacts through electrostatics with the reactive species, empirical force fields can be used to describe the resonant forms of the reactant and product states. For a bond-breaking reaction $AB \rightarrow A + B$, 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 structure ψ_3 is largely irrelevant and the process can be described by ψ_1 and ψ_2 .

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 location of an electron pair. 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.^[28,45] 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 $\left(\frac{-e^2}{r} + V_{nb}\right)$ 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 depending on a pre-defined reaction coordinate of the form $H_{ij} = A \exp(-\mu(r - r_0))$.^[44]

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.^[44] Alternative forms which also capture the shape and energetics of the potential energy surface around the transition state use generalized Gaussians.^[46] A comparison of different diabatic models has been recently given and provides a notion of the common features and the differences between various approaches.^[47–49] A useful comparison of the similarities and differences between the various methods can be found in the references.^[50] Applications of EVB include enzymatic reactions (for which it was originally developed^[51]), proton transfer processes, and the autodissociation of water.^[52] Furthermore, several extensions have been suggested to the original EVB method allowing its application to a wider class of problems.^[46,53,54]

1.7 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 chemical rates within an order of magnitude for reactants of 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, that is, $n_1 + n_2 = 1$. Or in the words of the authors, that "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 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. For a CC-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

$$E = E_{\text{bond}} + E_{\text{over}} + E_{\text{under}} + E_{\text{conj}} + E_{\text{val}} + E_{\text{pen}} + E_{\text{dihe}} + E_{\text{vdW}} + E_{\text{coul}} \quad (1.11)$$

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 the references.^[9]

Illustrative applications of ReaxFF range from the study of shock-induced chemistry in high energy materials^[56] to activation and dissociation of H₂ on platinum surfaces^[57] and the oxidation of nanoparticles on aluminum surface.^[58] Using ReaxFF and nonequilibrium MD simulations it was found that depending on the impact velocities cyclic-[CH₂N(NO₂)₃] decomposes into a variety of small molecules on the picosecond time scale or only into NO₂, both of which are consistent with experiments.^[56] Such simulations provide considerable insight into the time dependence of concentration changes of particular species.

1.8 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.^[47–50] 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).

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 (down-hill) dynamics to relax the system. Such an approach was employed to investigate the rebinding of CO in myoglobin.^[59] More generally, the approach is reminiscent of using an interpolating Hamiltonian as in the theory of electron transfer.^[60]

1.9 Applications

1.9.1 Protonated Water and Ammonia Dimer

Protonated water dimer has received substantial attention from both experiment^[61–64] and theory.^[65–69] The MMPT potential has been used to investigate the proton transfer dynamics and infrared spectroscopy of protonated water dimer.^[70] The MMPT potential allows to investigate the long-time (several 100 ps) bond-breaking and bond-formation dynamics and how this impacts the vibrational spectroscopy.^[70] MD simulations provide time series of coordinates and dipole moments which are then used to obtain vibrational spectra by Fourier transforming the dipole-dipole autocorrelation functions. For calculating the spectra, the 15 dimensional dipole moment surface of Huang *et al.*^[71] was used. 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 the O-H* coordinate involved in the O-H*-O asymmetric stretch is coupled to the O-O stretching and the O-H*O and HOH* bending coordinates.^[70] Here, H* is the transferring hydrogen atom. Vibrational excitations estimated from other calculations are well reproduced by the MD simulations with MMPT PES thus validating the approach.^[70] In particular, the MD simulations by Bowman and co-worker^[71] on a

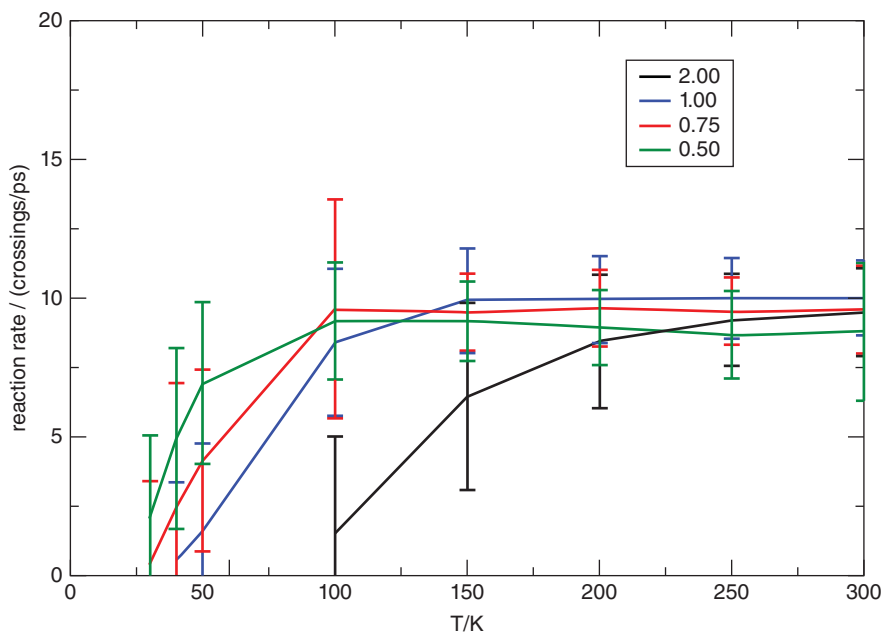


Figure 1.3 Proton Transfer Rates: Rates For Proton Transfer as a Function of T For $\text{NH}_4^+\text{-NH}_3$ Together with Error Bars. Rates are given for different energy scaling factors λE (proton-transfer barrier heights): $\lambda E = 2.0$ (black), 1.0 (blue), 0.75 (red), and 0.5 (green). For high temperatures the energy scaling has a smaller effect on the barrier than for low temperatures. This also explains why the rates agree within statistical fluctuations for $T \approx 300$ K. The figure is taken from the reference [36]. (See color plate section for the color representation of this figure.)

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 an MMPT potential.

Unlike protonated water dimer, the protonated ammonia dimer represents a symmetric double minima (SDM) potential. Earlier computational studies on protonated ammonia dimer include semiempirical calculations, EVB methods as well as approximate DFT (SCC-DFTB) simulations.^[72–75] Using a SDM potential for the MMPT part and standard force field parameters for the remaining degrees of freedom, MD simulations were carried out at various temperatures between 30 and 300 K. For the SDM potential, no proton transfer was seen below 40 K whereas between 40 K and 150 K a nearly linear increase in the rate is found, beyond which the rate remains constant, see Figure 1.3. To further characterize the rate dependence on the barrier separating the reactant and product, simulations were carried out on morphed PESs where the coordinate-independent morphing parameter λ in Eq. 1.4 was $\lambda = 0.5$, 0.75, and 2.0, respectively. 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 that a property common to all surfaces controls the maximum possible rate. Proton transfer probabilities of 8.8 to 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.^[36]

1.9.2 Charge Transfer in $N_2 - N_2^+$

In a recent study, the charge-transfer (CT) reaction $N_2^+ + N_2 \rightarrow N_2 + N_2^+$ between Coulomb-crystallized N_2^+ ions at mK temperatures and N_2 molecules from an 8 K beam was investigated by monitoring the change in spinrotational-state population of N_2^+ ions.^[33] The collision of the two interacting partners results in the electronically adiabatic formation of a vibrationally highly-excited $[N-N \cdots N-N]^+$ complex (the well depth is about 29 kcal/mol). After its decay the neutral N_2 leaves the ion trap, whereas the N_2^+ ions remain trapped and were found in rotationally excited states in the experiments.

Interpretation of the experimental findings required computational modelling of the collision process using “quasiclassical trajectory calculations” on an MS-ARMD PES. The PES was fitted accurately (RMSD = 1.4 kcal/mol) to 5565 of UCCSD/cc-pVTZ energies obtained from a global 6D scan. To achieve this accuracy, 24 force fields were needed for describing the three states (bound or complex, 2 unbound: charge-preserving and charge-transferring), which were obtained as follows. Within the complex, 4 connectivities ($N^1N^2-N^3N^4$, $N^1N^2-N^4N^3$, $N^2N^1-N^3N^4$, $N^2N^1-N^4N^3$) can be distinguished. However, at large deformations the complex becomes highly polarized (i.e., either $[N^1N^2](\delta^+) \cdots [N^3N^4](\delta^{++})$ or $[N^1N^2](\delta^{++}) \cdots [N^3N^4](\delta^+)$). Therefore, instead of using a single FF with a symmetric point-charge distribution for a given connectivity, two FFs with opposite polarization are required. Hence, altogether 8 FFs were needed for the 4 complex connectivities and 2 FFs correlating with them were necessary for each unbound state. For the parametrization of individual FFs, Coulomb interaction based on point charges (1-4 interaction in the bound state), Morse and Lennard-Jones potentials were used and within each of the three states, the FFs were related through permutation of the atom indices. The force fields were joined with the MS-ARMD method within the complex (8 FFs), the charge-preserving (2 FFs) and charge-transferring states (2 FFs). During complex formation, the active unbound state was smoothly connected to the bound state with a center of mass ($N_2^+ - N_2$) distance-dependent switching function in the range of 7.09 – 7.56 a_0 . During dissociation, upon reaching separation 7.09 a_0 , the dissociated state with lower potential energy was determined and the bound state was connected to that in the same manner. For an accurate fit of reference energies the flexibility of the global surface was increased by doubling the number of states (and adjustable parameters) to 24.

MS-ARMD simulations showed that the complex is formed up-to large impact parameters of $13a_0$, corresponding to large angular momentum. Upon complex formation (N_4^+) the system accelerates towards the energetically favoured linear arrangement which immediately induces vivid bending vibrations with magnitudes that depend on the impact parameter and the relative orientation of the diatomics. Usually, several rearrangements of the diatomics take place before its decay within 1 – 100 ps. The long average lifetime (10 ps) of the complex provides sufficient time for energy transfer from its overall rotation into bending and torsional vibrational modes, which eventually evolve into enhanced product state rotations after breaking of the central bond.

1.9.3 Vibrationally Induced Photodissociation of Sulfuric Acid

The MS-ARMD method was also applied to the vibrationally induced photodissociation of sulfuric acid (H_2SO_4),^[35,41] which can explain the anomalous enhancement of the

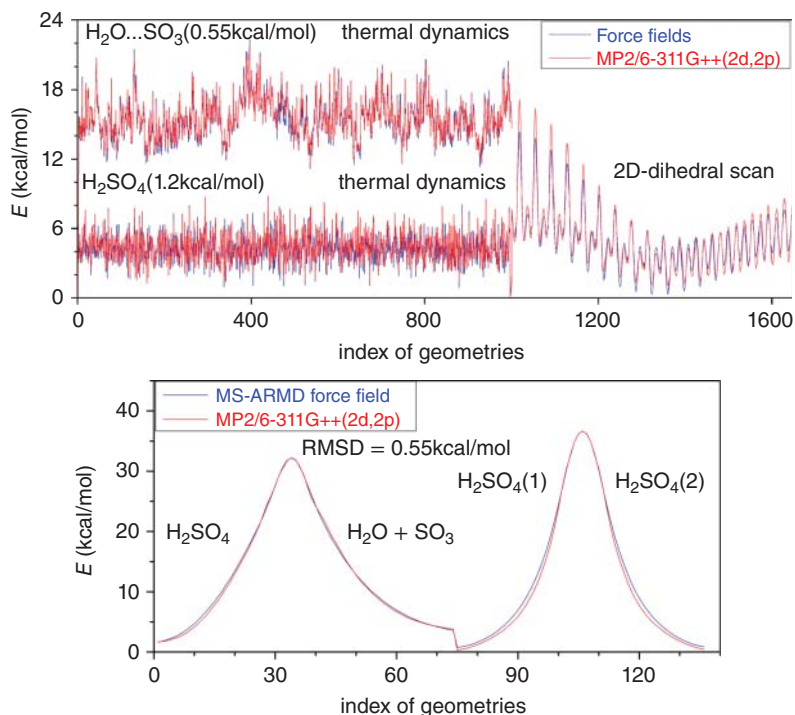


Figure 1.4 Reference MP2 Ab Initio and MS-ARMD Force Field Energies For States H_2SO_4 and $\text{H}_2\text{O} + \text{SO}_3$ and Along the Minimum Energy Paths For Water Elimination and For Intramolecular H-Transfer. (See color plate section for the color representation of this figure.)

polar stratospheric sulfate aerosol layer in springtime.^[76–79] Previous dynamical studies have shown that after significant vibrational excitation of the O-H stretching mode ($\nu_o = 4, 5, 6 \leftarrow 0$ corresponding to 38.6, 47.2 and 55.3 kcal/mol), H_2SO_4 can undergo intramolecular H-transfer ($V_{\text{barr}} = 32.2$ kcal/mol) and $\text{H}_2\text{SO}_4 \rightarrow \text{H}_2\text{O} + \text{SO}_3$ water elimination ($V_{\text{barr}} = 36.6$ kcal/mol).^[41,80]

To explore the ground-state potential energy surface of the system, calculations at MP2 / 6-311G++ (2d,2p) level were carried out for thousands of geometries taken from a 2D rigid dihedral scan of H_2SO_4 and from MD simulations at 300 K using previously developed FFs^[41] for H_2O , SO_3 , H_2SO_4 , and the $\text{H}_2\text{O} \dots \text{SO}_3$ vdW complex. New force fields were constructed using conventional FF terms, Morse and MIE potentials, and were fitted accurately (with RMSDs 0.02, 0.47, 1.23 and 0.55 kcal/mol) to MP2 energies, as shown in Figure 1.4.

Similarly, the minimum energy paths (MEPs) were determined and parameter ΔV and GAPO functions of the global MS-ARMD PES were optimized. Using 3 GAPOs with first-order polynomials for the H_2O elimination, and 2 GAPOs with second-order polynomials for the H-transfer reactions an accurate fit (RMSD = 0.55 kcal/mol) could be achieved, which is also shown in Figure 1.4.

After intramolecular H-transfer a sulfuric acid with different connectivity is obtained. Similarly, altogether, 8 H_2SO_4 and 4 $\text{H}_2\text{O} + \text{SO}_3$ states can be derived as both H atoms can be transferred and water elimination can involve any of the O atoms. Both sets

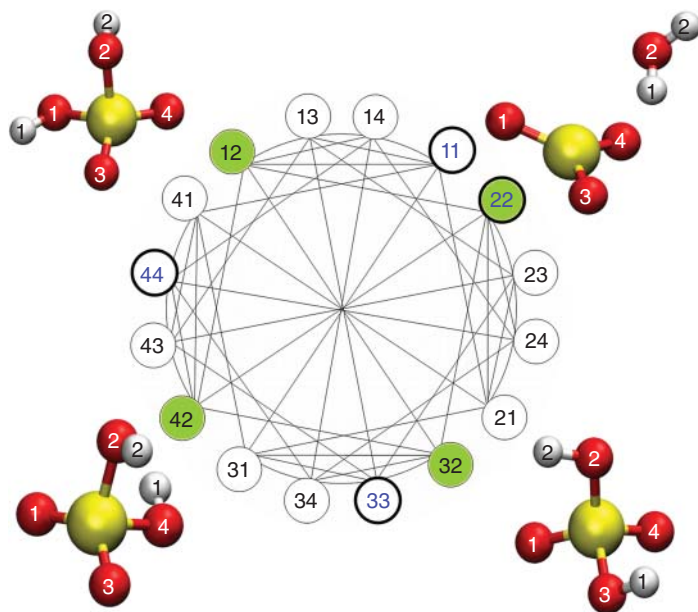


Figure 1.5 Reaction Network For H_2SO_4 : The Network of States (Circles) and Possible Reactions Between Them (Edges) For H_2SO_4 Represented as a Symmetric Graph. States labelled by two-digit numbers denoting the O indices of atoms to which atoms H1 and H2 bind. The four thick-line circles with identical digits indicates $\text{H}_2\text{O} + \text{SO}_3$ states, whereas the remaining 8 are H_2SO_4 states. Four states highlighted in green and given as a ball-and-stick representation were observed in a typical trajectory (12 \rightarrow 32 \rightarrow 12 \rightarrow 42 \rightarrow 22) showing three subsequent intramolecular H-transfers and water elimination at the end. (See color plate section for the color representation of this figure.)

of states are chemically equivalent to one another, therefore their FF and GAPO parametrizations need to be the same, but with permuted indices. The network of 16 states and possible 24-24 H-transfer and H_2O -elimination reactions between them can be represented as a graph (Figure 1.5).

Vibrational excitation by 4-6 quanta of thermalized (300 K) H_2SO_4 molecules was invoked by scaling the velocities along the OH-local mode.^[41,81] Then free dynamics was followed for several thousand trajectories until water elimination took place or at most for 1 ns. While several H-transfers and the final dissociation occurred, the total energy was conserved, which served as the validation of the MS-ARMD implementation in CHARMM.

Compared to previous studies,^[41,81] the MS-ARMD simulations were based on a more accurate PES which also enabled the competing intramolecular H-transfer process^[35] and thus lead to quite different lifetime distributions. The MS-ARMD description also allowed the analysis of products states, which showed distinct excitations that can help experimentalists to establish the vibrationally induced decay mechanism proposed by Vaida *et al.*^[79] Furthermore, kinetic analysis of the lifetime distributions lead to the conclusion that vibrationally induced photolysis rate drops significantly with decreasing altitude in the stratosphere due to the competing quenching process.

1.9.4 Proton Transfer in Malonaldehyde and Acetyl-Acetone

Proton Transfer in Malonaldehyde: Malonaldehyde (MA) has long served as a typical hydrogen transfer system to test and validate various computational approaches. Experimentally, the ground state tunneling splitting was determined to be 21.58314 cm^{-1} by different experiments with very high accuracy.^[82,83] Infrared spectra of MA have also been recorded at high resolution.^[84–87]

MMPT simulations with a generalization of the method to nonlinear H-bonds were carried out in order to determine the tunneling splittings for H- and D-transfer and to locate the position of the proton-transfer band in the infrared.^[88] Building on a harmonic bath averaged Hamiltonian (HBA) the effective reduced mass was chosen such as to reproduce the tunneling splitting for H-transfer. The effective reduced mass differs from the mass of the transferring hydrogen atom due to kinetic coupling in the system.^[88] However, the effective mass of the deuterated species is then determined by usual isotopic mass ratios which allows to validate the model because no new parameters are required. The computed tunneling splittings of 22.0 cm^{-1} and 2.9 cm^{-1} compare favourably with the experimentally determined ones which are 21.583 cm^{-1} and 2.915 cm^{-1} , respectively.^[82,83,89] The proton transfer mode exhibits a large red shifts relative to usual OH-stretching vibrations and is found at 1543 cm^{-1} .

Building on this MMPT potential a quantum mechanical treatment of the kinetic isotope effect (KIE) in MA was attempted. The KIE relates the rate constants for hydrogen and deuterium transfer via $\text{KIE} = k_{\text{H}}/k_{\text{D}}$. The KIE for the intramolecular hydrogen transfer in MA has not been determined experimentally. Combining a fully dimensional and validated PES^[88] based on molecular mechanics with proton transfer (MMPT)^[36] with quantum instanton (QI) path integral Monte Carlo (PIMC) simulation the primary H/D KIE on the intramolecular proton transfer in MA was found to be 5.2 ± 0.4 at room temperature.^[90] For higher temperatures, the KIE tends to 1, as required. Periodic orbit theory-based tunneling rate estimates and detailed comparisons with conventional transition state theory (CTST) at various levels suggest that the KIE in MA is largely determined by zero-point energy effects and that tunneling plays a minor role.

Proton Transfer in Acetyl-Acetone: Related to MA is acetyl-acetone (AcAc) but the proton transfer dynamics is far less well characterized. Although infrared and microwave spectra have been recorded, the symmetry of the ground state structure is still debated.^[91–96] Recent infrared experiments combined with atomistic simulations using an MMPT force field and quantum chemical calculations have considered the dynamics of the hydrogen transfer motion in AcAc.^[97] The morphed potential exhibits a barrier of 2.35 kcal/mol and was used in finite-temperature MD simulations from which the IR spectra was determined.

Experimentally, the fundamental OH-stretching band is observed as a broad band red-shifted relative to usual OH-stretching transitions by several hundred wavenumbers.^[97] The IR and power spectra from the MMPT simulations reproduce most experimentally recorded features and clearly assign the ν_{OH} proton transfer (PT) mode to the band experimentally observed in the $2000 - 3300 \text{ cm}^{-1}$ region. The location of this band was found to sensitively depend on the barrier for PT. Simulations with varying barrier heights and comparison of the power spectra with the experimental IR spectrum yield a barrier of $\approx 2.5 \text{ kcal/mol}$. This compares with a value of 3.2 kcal/mol from CCSD(T)

calculations and suggests that such an approach is meaningful to determine approximate barrier heights for proton transfer reactions which is difficult if not impossible from experiment alone.

1.9.5 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 O₂, 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, it becomes exponential at high temperature with a time scale of 100 ns, rebinding of NO is nonexponential at all temperature, with time constants of the order of tens of picoseconds.^[98,99]

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.^[38]

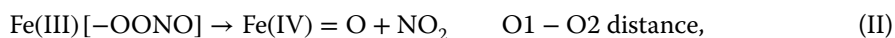
To visualize the crossing seam all observed crossing geometries are projected onto a plane containing the Fe-N distance and the angle formed by Fe and the ligand NO. The most probable iron-ligand distance lies around 3 \AA .^[38] The distribution is rather wide along the bond angle coordinate and the crossing seam is found to be rather insensitive towards the value of Δ , see Figure 1.6.

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 $\Delta = 65$ kcal/mol, the time constants are found to be 3.6 and 373 ps^[38] compared to the experimental (from ultrafast IR spectroscopy) value of 5.3 and 133 ps.^[99,100] 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.

1.9.6 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 NO₃⁻. The large second-order rate constant of $7.5 \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$ has been attributed to the existence of a continuous tunnel inside the protein which assists ligand migration.^[101,102] However, an atomistic understanding about the mechanism of the detoxification reaction had remained illusive.

ARMD was used to shed light on the reaction by dividing the overall reaction into following four steps:^[103]



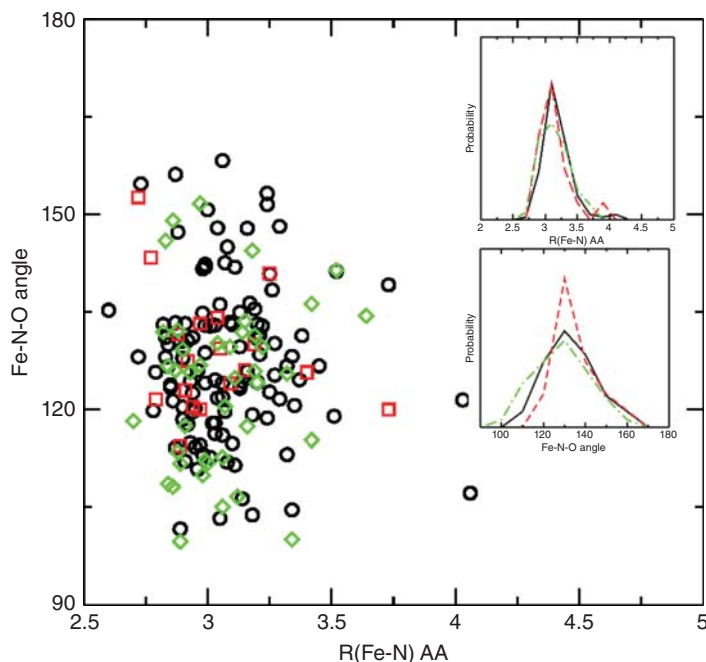


Figure 1.6 Crossing Seam For NO-Rebinding: The Crossing Geometries of NO Rebinding to Mb For Three Different Values of Δ (Red Squares $\Delta = 60$, Black Circles $\Delta = 65$, and Green Diamonds $\Delta = 70$ kcal/mol) Projected Onto the Distance Between Iron and the Nitrogen of the Ligand R_{FeN} 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 $\Delta = 60, 65,$ and 70 kcal/mol, respectively. The figure is taken from Reference [38]. (See color plate section for the color representation of this figure.)

The right-hand side column of the above reaction steps indicates the bond broken/formed bond in the reaction step. First, oxy-trHbN reacts with free NO and forms a peroxynitrite intermediate, which then undergoes homolytic fission. This is followed by the rebinding of free NO_2 to the oxo-ferryl species to form the heme-bound nitrate 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 steps was then studied by running multiple ARMD trajectories with a range of Δ values.^[40] 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 reaction rate.

For step II, however, no reactive events even on the nanosecond time scale were found.^[40] From DFT calculations it is known that this step involves a barrier of 6.7 kcal/mol.^[104] Umbrella sampling simulations with ARMD yielded a barrier of 12-15 kcal/mol which corresponds to timescales on the order of micro- to milliseconds. Since experimentally, the overall reaction is on the picosecond time scale, it is unlikely that the reaction occurs 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 nitrate complex.^[105,106]

To further corroborate this, ARMD simulations for the rearrangement reaction were carried out and found this process to occur within picoseconds, explaining the fast overall detoxification reaction.^[40]

1.9.7 Outlook

This outlook summarizes the methods presented in this chapter and describes potential improvements in investigating reactions using force fields.

The common feature of the EVB, ARMD and MS-ARMD methods is that they combining force fields by giving preference to the lowest energy surface, thus inherently the energy difference between the surfaces serves as a reaction coordinate. In EVB, the smooth switching is carried out by formally carrying out an adiabaticization of diabatic PESs, which is controlled by couplings between the states. These coupling terms have to decay when moving away from the dividing surface, which is assumed to depend on a predefined geometric reaction coordinate, whose determination, however, in complex reactions is not straightforward. Contrary to that, smooth switching in ARMD and MS-ARMD is carried out by time- and energy-dependent switching functions controlled by the switching time t_s or a switching parameter ΔV , respectively. These latter switching variables are one dimensional thus naturally serve as a reaction coordinate, whereas in the case of geometrical switching variables the determination of reaction coordinate is neither straightforward nor unambiguous. Furthermore, the potential energy and its gradient are analytic functions of the individual force field energies, whereas in EVB the matrix diagonalization allows efficient analytic evaluation only for a small number of surfaces due to the diagonalization involved.

Due to the explicit time-dependence of the ARMD Hamiltonian, the ARMD crossing is inherently a dynamical method and this is also the reason for violations of energy conservation observed during crossing in gas-phase reactions of highly-excited small molecules. On the contrary, in the MS-ARMD and EVB methods the PES is stationary and thus they can be used in *NVE* simulations. The ARMD method cannot describe crossings in regions properly where more than two surfaces are close in energy, whereas the MS-ARMD technique can be applied to the simultaneous switching among multiple PESs. Finally, in MS-ARMD, the height and the shape of the crossing region can be adjusted in a flexible manner, which makes it very similar in functionality to EVB, but without the need of referring to any geometrical reaction coordinate. This extension makes the CHARMM implementation of MS-ARMD method^[39] comparable to QM/MM methods in capability of modelling complex chemical reaction in gas, surface, and condensed phases albeit at the accelerated speed of a conventional force field simulation and with the only added effort of parametrization.

Future improvements of the ARMD and MS-ARMD methods include the more routine development of accurate force field parametrizations which still poses an obstacle to a more routine use of such approaches in all areas of chemistry and biophysics. For ARMD, which requires an asymptotic offset Δ between the states, an improvement could be achieved by making this parameter coordinate-dependent. On the other hand, this would introduce one or several geometric progression coordinates whose definition may be difficult. Nevertheless, a recent study of nitric oxide rebinding to Mb found that indeed, Δ is not constant but depends on the iron-out-of-plane position in this particular situation.^[107] Hence, a better reaction energetics and a more realistic modelling of the ligand-(re)binding dynamics is obtained from such improvements.

The current implementation of MS-ARMD requires the separate definition and evaluation of force fields for each possible product connectivity. This means several possibilities in solvent-phase or surface-phase simulations where the reaction can take place with any of the solvent molecules or at any of the surface sites. Even though the common part of all force fields are evaluated only once, the calculation and even the enumeration of all cases is unnecessary. In future developments, this will be constrained to a small number of momentary physically sensible connectivities by an automatic geometrical preselection.

At present, MMPT supports the transfer of one proton/hydrogen atom between one designated donor/acceptor pair. Combining this with the MS-ARMD philosophy will allow one to automatically determine the most probable H-bonding pattern and the dynamics involving them, which would bring the method much closer to QM/MM simulations.

Force field-based approaches to study chemical reactions is becoming more widespread as the functionalities are made available in commonly used atomistic simulation programs. They allow to study chemical reactivity on time scales relevant to the real processes and provide insight complementary to experiment provided that the underlying force fields are accurate. The fitting of reactive force fields remains a challenge but generalizing and simplifying this step will make such approaches valuable additions to the toolbox of computational and experimental chemists interested in chemical reactivity.

Acknowledgements

This work was supported by the Swiss National Science Foundation through grants 200021-117810, the NCCR MUST, and the University of Basel, which is gratefully acknowledged. We thank all coworkers mentioned in the references for their contributions to the methods and their applications discussed in the present work.

References

- 1 London, F.Z. (1929) Quantum mechanical interpretation of the process of activation. *Zeitschrift für Elektrochemie*, **35**, 552–555.
- 2 Eyring, H. and Polanyi, M.Z. (1931) Concerning simple gas reactions. *Zeitschrift für Physikalische Chemie. Abteilung B*, **12**, 279–311.
- 3 Sato, S. (1955) On a new method of drawing the potential energy surface. *Journal of Chemical Physics*, **23**, 592–593.
- 4 Sato, S. (1955) Potential energy surface of the system of three atoms. *Journal of Chemical Physics*, **23**, 2465–2466.
- 5 Ellison, F.O. (1963) A method of diatomics in molecules. 1. General theory and application to H₂O. *Journal of the American Chemical Society*, **85**, 3540–3544.
- 6 Pauling, L. (1960) *The nature of the chemical bond*, Cornell University Press, Ithaca, NY.
- 7 Pauling, L. (1932) The nature of the chemical bond. IV. The energy of single bonds and the relative electronegativity of atoms. *Journal of the American Chemical Society*, **54**, 3570–3582.

- 8 Johnston, H.S. and Parr, C. (1963) Activation energies from bond energies. I. Hydrogen transfer reactions. *Journal of the American Chemical Society*, **85**, 2544–2551.
- 9 van Duin, A.C.T., Dasgupta, S., Lorant, F. and Goddard, W.A. III, (2001) A reactive force field for hydrocarbons. *Journal of Physical Chemistry A*, **105**, 9396–9409.
- 10 Lifson, S. and Warshel, A. (1968) Consistent force field for calculations of conformations vibrational spectra and enthalpies of cycloalkane and n-alkane molecules. *Journal of Chemical Physics*, **49**, 5116–5129.
- 11 Levitt, M. and Lifson, S. (1969) Refinement of protein conformations using a macromolecular energy minimization procedure. *Journal of Molecular Biology*, **46**, 269–279.
- 12 Hwang, M.J., Stockfisch, T.P. and Hagler, A.T. (1994) Derivation of class II force fields: 2. Derivation and characterization of a class II force field, CFF93, for the alkyl functional group and alkane molecules. *Journal of the American Chemical Society*, **116**, 2515–2525.
- 13 Maple, J.R., Hwang, M.J., Stockfisch, T.P. *et al.* (1994) Derivation of class-II force-fields. 1. Methodology and quantum force-field for the alkyl functional-group and alkane molecules. *Journal of Computational Chemistry*, **15**, 162–182.
- 14 Brooks, B., Brucoleri, R., Olafson, B. *et al.* (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, **4**, 18–217.
- 15 Weiner, S.J., Kollman, P.A., Case, D.A. *et al.* (1984) A new force-field for molecular mechanical simulation of nucleic-acids and proteins. *Journal of the American Chemical Society*, **106**, 765–784.
- 16 Jorgensen, W.L. and Tirado-Rives, J. (1988) The OPLS potential functions for proteins-energy minimizations for crystals of cyclic-peptides and crambin. *Journal of the American Chemical Society*, **110**, 1657–1666.
- 17 Hermans, J., Berendsen, H.J.C., van Gunsteren, W.F. and Postma, J.P.M. (1984) A consistent empirical potential for water-protein interactions. *Biopolymers*, **23**, 1.
- 18 Warshel, A. and Levitt, M. (1976) Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology*, **103**, 227–249.
- 19 Alagona, G., Ghio, C. and Kollman, P. (1986) A simple-model for the effect of glu165- \rightarrow asp165 mutation on the rate of catalysis in triose phosphate isomerase. *Journal of Molecular Biology*, **191**, 23–27.
- 20 Bash, P.A., Field, M.J. and Karplus, M. (1987) Free-energy perturbation method for chemical-reactions in the condensed phase – a dynamical approach based on a combined quantum and molecular mechanics potential. *Journal of the American Chemical Society*, **109**, 8092–8094.
- 21 Claeysens, F., Ranaghan, K.E., Manby, F.R. *et al.* (2005) Multiple high-level QM/MM reaction paths demonstrate transition-state stabilization in chorismate mutase: correlation of barrier height with transition-state stabilization. *Chemical Communications*, **40**, 5068–5070.
- 22 Zhou, H.Y., Tajkhorshid, E., Frauenheim, T. *et al.* (2002) Performance of the AM1, PM3, and SCC-DFTB methods in the study of conjugated Schiff base molecules. *Chemical Physics*, **277**, 91–103.

- 23 König, P.H., Ghosh, N., Hoffmann, M. *et al.* (2006) Toward theoretical analysis of long-range proton transfer kinetics in biomolecular pumps. *Journal of Physical Chemistry A*, **110**, 548–563.
- 24 Cui, Q., Elstner, T. and Karplus, M. (2003) A theoretical analysis of the proton and hydride transfer in liver alcohol dehydrogenase (LADH). *Journal of Physical Chemistry B*, **106**, 2721–2740.
- 25 Li, H., Elber, R. and Straub, J.E. (1993) Molecular-dynamics simulation of NO recombination to myoglobin mutants. *Journal of Biological Chemistry*, **268**, 17908–17916.
- 26 Meuwly, M., Becker, O.M., Stote, R. and Karplus, M. (2002) NO rebinding to myoglobin: A reactive molecular dynamics study. *Biophysical Chemistry*, **98**, 183–207.
- 27 Nutt, D.R. and Meuwly, M. (2006) Studying reactive processes with classical dynamics: Rebinding dynamics in MbNO. *Biophysical Journal*, **90**, 1191–1201.
- 28 Warshel, A. and Weiss, R.M. (1980) An empirical valence bond approach for comparing reactions in solutions and in enzymes. *Journal of the American Chemical Society*, **102**, 6218–6226.
- 29 Grochowski, P., Lesyng, B., Bala, P. and McCammon, J.A. (1996) Density functional based parametrization of a valence bond method and its applications in quantum-classical molecular dynamics simulations of enzymatic reactions. *International Journal of Quantum Chemistry*, **60**, 1143–1164.
- 30 Kim, Y., Corchado, J.C., Villa, J. *et al.* (2000) Multiconfiguration molecular mechanics algorithm for potential energy surfaces of chemical reactions. *Journal of Chemical Physics*, **112**, 2718–2735.
- 31 Van Voorhis, T., Kowalczyk, T., Kaduk, B. *et al.* (2010) The diabatic picture of electron transfer, reaction barriers, and molecular dynamics. *Annual Review of Physical Chemistry*, **61**, 149–170.
- 32 Dayal, P., Weyand, S.A., McNeish, J. and Mosey, N.J. (2011) Temporal quantum mechanics/molecular mechanics: Extending the time scales accessible in molecular dynamics simulations of reactions. *Chemical Physics Letters*, **516**, 263–267.
- 33 Tong, X., Nagy, T., Reyes, J.Y. *et al.* (2012) State-selected ionmolecule reactions with Coulomb-crystallized molecular ions in traps. *Chemical Physics Letters*, **547**, 1–8.
- 34 Nagy, T., Yosa Reyes, J. and Meuwly, M. (2014) Multisurface adiabatic reactive molecular dynamics. *Journal of Chemical Theory and Computation*, **10**, 1366–1375.
- 35 Yosa Reyes, J., Nagy, T. and Meuwly, M. (2014) Competitive reaction pathways in vibrationally induced photodissociation of H₂SO₄. *Physical Chemistry Chemical Physics*, **16**, 18533–18544.
- 36 Lammers, S., Lutz, S. and Meuwly, M. (2008) Reactive force fields for proton transfer dynamics. *Journal of Computational Chemistry*, **29**, 1048–1063.
- 37 Meuwly, M. and Huston, J. (1999) Morphing ab initio potentials: A systematic study of Ne-HF. *Journal of Chemical Physics*, **110**, 8338–8347.
- 38 Danielsson, J. and Meuwly, M. (2008) Atomistic simulation of adiabatic reactive processes based on multi-state potential energy surfaces. *Journal of Chemical Theory and Computation*, **4**, 1083–1093.
- 39 Brooks, B.R. *et al.* (2009) CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry*, **30**, 1545–1614.

- 40 Mishra, S. and Meuwly, M. (2010) Atomistic simulation of NO dioxygenation in Group I truncated hemoglobin. *Journal of the American Chemical Society*, **132**, 2968.
- 41 Yosa, J. and Meuwly, M. (2011) Vibrationally induced dissociation of sulfuric acid (H_2SO_4). *Journal of Physical Chemistry A*, **115**, 14350–14360.
- 42 Mie, G. (1903) Zur kinetischen Theorie der einatomigen Körper. *Annals of Physics*, **11**, 657–697.
- 43 Kramer, C., Gedeck, P. and Meuwly, M. (2013) Multipole-based force fields from interaction energies and the need for jointly refitting all intermolecular parameters. *Journal of Chemical Theory and Computation*, **9** (3), 1499–1511.
- 44 Hong, G., Rosta, E. and Warshel, A. (2006) Using the constrained DFT approach in generating diabatic surfaces and off diagonal empirical valence bond terms for modeling reactions in condensed phases. *Journal of Physical Chemistry B*, **110**, 19570–19574.
- 45 Coulson, C.A. and Danielsson, U. (1954) Ionic and covalent contributions to the hydrogen bond. *Arkiv för Fysik*, **8**, 239–244.
- 46 Chang, Y.-T. and Miller, W.H. (1990) An empirical valence bond model for constructing global potential energy surfaces for chemical reactions of polyatomic molecular systems. *Journal of Physical Chemistry*, **94**, 5884–5888.
- 47 Valero, R., Song, L., Gao, J. and Truhlar, D.G. (2009) Perspective on diabatic models of chemical reactivity as illustrated by the gas-phase $\text{S}_{\text{N}}2$ reaction of acetate ion with 1,2-dichloroethane. *Journal of Chemical Theory and Computation*, **5**, 1–22.
- 48 Valero, R., Song, L., Gao, J. and Truhlar, D.G. (2009) Erratum: Perspective on diabatic models of chemical reactivity as illustrated by the gas-phase $\text{S}_{\text{N}}2$ reaction of acetate ion with 1,2-dichloroethane. *Journal of Chemical Theory and Computation*, **5**, 2191–2191.
- 49 Kamerlin, S.C.L., Cao, J., Rosta, E. and Warshel, A. (2009) On unjustifiably misrepresenting the EVB approach while simultaneously adopting it. *Journal of Physical Chemistry B*, **113**, 10905–10915.
- 50 Florian, J. (2002) Comments on molecular mechanics for chemical reactions. *Journal of Physical Chemistry A*, **106**, 5046–5047.
- 51 Warshel, A. (1984) Dynamics of enzymatic reactions. *Proceedings of the National Academy of Sciences USA*, **81**, 444–448.
- 52 Strajbl, M., Hong, G. and Warshel, A. (2002) Ab initio QM/MM simulation with proper sampling: "First principle" calculations of the free energy of the autodissociation of water in aqueous solution. *Journal of Physical Chemistry B*, **106**, 13333–13343.
- 53 Schmitt, U. and Voth, G. (1998) Multistate empirical valence bond model for proton transport in water. *Journal of Physical Chemistry B*, **102**, 5547–5551.
- 54 Schlegel, H.B. and Sonnenberg, J.L. (2006) Empirical valence-bond models for reactive potential energy surfaces using distributed Gaussians. *Journal of Chemical Theory and Computation*, **2**, 905–911.
- 55 Pauling, L. (1947) Atomic radii and interatomic distances in metals. *Journal of the American Chemical Society*, **69**, 542–553.
- 56 Strachan, A., van Duin, A.C.T., Chakraborty, D. *et al.* (2003) Shock waves in high-energy materials: The initial chemical events in nitramine RDX. *Physical Review Letters*, **91**, 098301–1.

- 57 Ludwig, J., Vlachos, D.G., van Duin, A.C.T. and Goddard, W.A. III, (2006) Dynamics of the dissociation of hydrogen on stepped platinum surfaces using the ReaxFF reactive force field. *Journal of Physical Chemistry B*, **110**, 4274–4282.
- 58 Vashishta, P., Kalia, R.K. and Nakano, A. (2006) Multimillion atom simulations of dynamics of oxidation of an aluminum nanoparticle and nanoindentation on ceramics. *Journal of Physical Chemistry B*, **110**, 3727–3733.
- 59 Zheng, D., Makarov, V. and Wolynes, P.G. (1996) Statistical survey of transition states and conformational substates of the sperm whale myoglobin-CO reaction system. *Journal of the American Chemical Society*, **118**, 2818–2824.
- 60 Marcus, R.A. (1993) Electron-transfer reactions in chemistry – theory and experiment. *Angewandte Chemie International Edition*, **32**, 1111–1121.
- 61 Asmis, K.R., Pivonka, N.L., Santambrogio, G. *et al.* (2003) Gas phase infrared spectrum of the protonated water dimer. *Science*, **299**, 1375–1377.
- 62 Fridgen, T.D., McMahon, G.B., MacAleese, L. *et al.* (2004) Infrared spectrum of the protonated water dimer in the gas phase. *Journal of Physical Chemistry A*, **108**, 9008–9010.
- 63 Moore, D.R., Oomens, J., van der Meer, L. *et al.* (2004) Probing the vibrations of shared, OH⁺O⁻-bound protons in the gas phase. *Computer Physics Communications*, **5**, 740–743.
- 64 Hammer, N.I., Diken, E.G., Roscioli, J.R. *et al.* (2005) The vibrational predissociation spectra of the H₅O₂⁺·RG_n (RG = Ar, Ne) clusters: Correlation of the solvent perturbations in the free OH and shared proton transitions of the Zundel ion. *Journal of Chemical Physics*, **122**, 244301.
- 65 Dai, J., Bacic, Z., Huang, X. *et al.* (2003) A theoretical study of vibrational mode coupling in H₅O₂⁺. *Journal of Chemical Physics*, **119**, 6571–6580.
- 66 Kaledin, M., Kaledin, A.L. and Bowman, J.M. (2006) Vibrational analysis of the H₅O₂⁺ infrared spectrum using molecular and driven molecular dynamics. *Journal of Physical Chemistry A*, **110**, 2933–2939.
- 67 Lobaugh, J. and Voth, G.A. (1996) The quantum dynamics of an excess proton in water. *Journal of Chemical Physics*, **104**, 2056–2059.
- 68 Cho, H.M. and Singer, S.J. (2004) Correlation function quantum Monte Carlo study of the excited vibrational states of H₅O₂⁺. *Journal of Physical Chemistry A*, **108**, 8691–8702.
- 69 Sauer, J. and Doebler, J. (2005) Gas-phase infrared spectrum of the protonated water dimer: molecular dynamics simulation and accuracy of the potential energy surface. *Computer Physics Communications*, **6**, 1706–1710.
- 70 Lammers, S. and Meuwly, M. (2007) Investigating the relationship between infrared spectra of shared protons in different chemical environments: A comparison of protonated diglyme and protonated water dimer. *Journal of Physical Chemistry A*, **111**, 1638–1647.
- 71 Huang, S., Braams, B.J. and Bowman, J.M. (2005) Ab initio potential energy and dipole moment surfaces for H₅O₂⁺. *Journal of Chemical Physics*, **122**, 044308.
- 72 Asada, T., Haraguchi, H. and Kitaura, K. (2001) Simulation studies of proton transfer in N₂H₇⁺ cluster by classical ab initio Monte Carlo and quantum wave packet dynamics. *Journal of Physical Chemistry A*, **105**, 7423–7428.
- 73 Li, G.-S., Costa, M.T.C.M., Millot, C. and Ruiz-Lopez, M.F. (1999) Effect of solvent fluctuations on proton transfer dynamics: a hybrid AM1/MM molecular dynamics simulation on the [H₃N–H–NH₃]⁺ system. *Chemical Physics*, **240**, 93–99.

- 74 Wang, Y. and Gunn, J.R. (1999) Computational study of structures and proton transfer in hydrogen-bonded ammonia complexes using semiempirical valence-bond approach. *International Journal of Quantum Chemistry*, **73**, 357–367.
- 75 Meuwly, M. and Karplus, M. (2002) Simulation of proton transfer along ammonia wires: An ab initio and semiempirical density functional comparison of potentials and classical molecular dynamics. *Journal of Chemical Physics*, **116**, 2572–2585.
- 76 Clement, F. and Ford, I.J. (1999) Gas-to-particle conversion in the atmosphere: I. Evidence from empirical atmospheric aerosols. *Atmospheric Environment – Part A*, **33**, 1352.
- 77 Larson, L.J., Kuno, M. and Tao, F.M. (2000) Hydrolysis of sulfur trioxide to form sulfuric acid in small water clusters. *Journal of Chemical Physics*, **112**, 8830–8838.
- 78 Donaldson, D., Frost, G., Rosenlof, K. *et al.* (1997) Atmospheric radical production by excitation of vibrational overtones via absorption of visible light. *Geophysical Research Letters*, **24**, 2651–2654.
- 79 Vaida, V., Kjaergaard, H.G., Hintze, P.E. and Donaldson, D.J. (2003) Photolysis of sulfuric acid vapor by visible solar radiation. *Science*, **299**, 1566–1568.
- 80 Miller, Y. and Gerber, R.B. (2006) Dynamics of vibrational overtone excitations of H₂SO₄, H₂SO₄-H₂O: Hydrogen-hopping and photodissociation processes. *Journal of the American Chemical Society*, **128**, 9594–9595.
- 81 Nguyen, P.N. and Stock, G. (2003) Nonequilibrium molecular-dynamics study of the vibrational energy relaxation of peptides in water. *Journal of Chemical Physics*, **119**, 11350–11358.
- 82 Baughcum, S.L., Duerst, R.W., Rowe, W.F. *et al.* (1981) Microwave spectroscopic study of malonaldehyde (3-hydroxy-2-propenal). 2. Structure, dipole-moment, and tunneling. *Journal of the American Chemical Society*, **103**, 6296–6303.
- 83 Firth, D.W., Beyer, K., Dvorak, M.A. *et al.* (1991) Tunable far infrared spectroscopy of malonaldehyde. *Journal of Chemical Physics*, **94**, 1812–1819.
- 84 Smith, Z. and Wilson, E.B. (1983) The infrared spectrum of gaseous malonaldehyde (3-hydroxy-2-propenal). *Spectrochimica Acta Part A*, **39**, 1117–1129.
- 85 Firth, D.W., Barbara, P.F. and Trommsdorf, H.P. (1989) Matrix induced localization of proton tunneling in malonaldehyde. *Chemical Physics*, **136**, 349–360.
- 86 Chiavassa, T., Roubin, R., Piazzala, L. *et al.* (1992) Experimental and theoretical-studies of malonaldehyde – vibrational analysis of a strongly intramolecularly hydrogen-bonded compound. *Journal of Physical Chemistry*, **96**, 10659–10665.
- 87 Duan, C. and Luckhaus, D. (2004) High resolution IR-diode laser jet spectroscopy of malonaldehyde. *Chemical Physics Letters*, **391**, 129–133.
- 88 Yang, Y. and Meuwly, M. (2010) A generalized reactive force field for nonlinear hydrogen bonds: Hydrogen dynamics and transfer in malonaldehyde. *Journal of Chemical Physics*, **133**, 064503.
- 89 Baughcum, S.L., Smith, Z., Wilson, E.B. and Duerst, R.W. (1984) Microwave spectroscopic study of malonaldehyde. 3. Vibration-rotation interaction and one-dimensional model for proton tunneling. *Journal of the American Chemical Society*, **106**, 2260.
- 90 Huang, J., Buchowiecki, M., Nagy, T. *et al.* (2014) Kinetic isotope effect in malonaldehyde from path integral Monte Carlo simulations. *Physical Chemistry Chemical Physics*, **16**, 204–211.

- 91 Johnson, M.R., Jones, N.H., Geis, A. *et al.* (2002) Structure and dynamics of the keto and enol forms of acetylacetone in the solid state. *Journal of Chemical Physics*, **116**, 5694–5700.
- 92 Lowrey, A.H., George, C., D’Antonio, P. and Karle, J. (1971) Structure of acetylacetone by electron diffraction. *Journal of the American Chemical Society*, **93**, 6399–6403.
- 93 Andreassen, A.L. and Bauer, S.H. (1972) Structures of acetylacetone, trifluoroacetylacetone, and trifluoroacetone. *Journal of Molecular Structure*, **12**, 381–403.
- 94 Srinivasan, R., Feenstra, J.S., Park, S.T. *et al.* (2004) Direct determination of hydrogen-bonded structures in resonant and tautomeric reactions using ultrafast electron diffraction. *Journal of the American Chemical Society*, **126**, 2266–2267.
- 95 Iijima, K., Ohnogi, A. and Shibata, S. (1987) The molecular structure of acetylacetone as studied by gas-phase electron diffraction. *Journal of Molecular Structure*, **156**, 111–118.
- 96 Caminati, W. and Grabow, J.U. (2006) The C_{2v} structure of enolic acetylacetone. *Journal of the American Chemical Society*, **128**, 854–857.
- 97 Howard, D., Kjaergaard, H., Huang, J. *et al.* (2015) Infrared and near infrared spectroscopy of acetylacetone and hexafluoroacetylacetone. *Journal of Physical Chemistry A*, **119** (29), 7980–7990.
- 98 Austin, R.H., Beeson, K.W., Eisenstein, L. *et al.* (1975) Dynamics of ligand-binding to myoglobin. *Biochemistry*, **14**, 5355–5373.
- 99 Kim, S., Jin, G. and Lim, M. (2004) Dynamics of geminate recombination of NO with myoglobin in aqueous solution probed by femtosecond mid-IR spectroscopy. *Journal of Physical Chemistry B*, **108**, 20366–20375.
- 100 Petrich, J.W., Lambry, J.C., Kuczera, K. *et al.* (1991) Ligand binding and protein relaxation in heme proteins: A room temperature analysis of NO geminate recombination. *Biochemistry*, **30**, 3975–3987.
- 101 Ouellet, H., Ouellet, Y., Richard, C. *et al.* (2002) Truncated hemoglobin HbN protects *Mycobacterium bovis* from nitric oxide. *Proceedings of the National Academy of Sciences USA*, **99**, 5902–5907.
- 102 Milani, M., Pesce, A., Ouellet, Y. *et al.* (2004) Heme-ligand tunneling in Group I truncated hemoglobins. *Journal of Biological Chemistry*, **279**, 21520–21525.
- 103 Bourassa, J.L., Ives, E.L., Marqueling, A.L. *et al.* (2001) Myoglobin catalyzes its own nitration. *Journal of the American Chemical Society*, **123**, 5142–5143.
- 104 Blomberg, L.M., Blomberg, M.R.A. and Siegbahn, P.E.M. (2004) A theoretical study of myoglobin working as a nitric oxide scavenger. *Journal of Biological Inorganic Chemistry*, **9**, 923–935.
- 105 Herold, S., Exner, M. and Nauser, T. (2001) Kinetic and mechanistic studies of the NO-mediated oxidation of oxymyoglobin and oxyhemoglobin. *Biochemistry*, **40**, 3385–3395.
- 106 Herold, S. (1999) Kinetic and spectroscopic characterization of an intermediate peroxynitrite complex in the nitrogen monoxide induced oxidation of oxyhemoglobin. *FEBS Letters*, **443**, 81–84.
- 107 Soloviov, M. and Meuwly, M. (2015) Reproducing kernel potential energy surfaces in biomolecular simulations: Nitric oxide binding to myoglobin. *Journal of Chemical Physics*, **143** (10), 105103.

2

Introduction to the Empirical Valence Bond Approach

Fernanda Duarte¹, Anna Pabis² and Shina Caroline Lynn Kamerlin²

¹Physical and Theoretical Chemistry Laboratory, University of Oxford, United Kingdom

²Department of Cell and Molecular Biology, Uppsala University, Sweden

2.1 Introduction

The award of the 2013 Nobel Prize in Chemistry to Martin Karplus, Michael Levitt and Arieh Warshel for “the development of multiscale models for complex chemical systems”^[1] has demonstrated how mature the field of biomolecular modeling has become. In particular, increases in computer power, including the availability of tailor-made massively parallelized computer architectures, together with specialized and more efficient computer algorithms, have allowed for spectacular advances in terms of both the size of molecular systems that can be studied and overall simulation timescales.^[2]

Despite the many advances in the field, the rigorous modeling of chemical reactions in condensed phase is still challenging. Key underlying issues for a correct description of these processes are: (1) an accurate and computationally efficient description of the bond breaking/forming processes and (2) proper modeling of the complex environment of the reaction, which involves efficient configurational sampling of the energy landscape. In principle, the use of high-level quantum mechanical (QM) approaches provides an accurate description of the electronic rearrangements during chemical processes. However, due to their very demanding computational cost, their use is still limited to relatively small systems consisting of up to hundreds of atoms. To address the second issue, an efficient sampling method is required. The use of molecular mechanics (MM) approaches, which are based on classical potentials, is extremely helpful as they allow inclusion of environmental effects (from either solvent molecules or enzymes) in a cost-efficient way. However, MM force fields are unable to describe the electronic changes taking place during a chemical reaction.

A solution to these challenges is the use of multilayer approaches, in which the interesting part of the system (usually where the chemical reaction takes place) is described at the electronic level by high-level QM approaches, while the rest of the system is represented by empirical force fields (or by a lower-level QM method). Multi-layer approaches have now become established state-of-the-art computational techniques for the modeling of chemical reactions in the condensed phase, including complex processes in organic chemistry, biochemistry, and heterogeneous catalysis,

among other examples. Over the past decade, a number of so-called combined quantum mechanical and molecular mechanical (QM/MM) methods have been implemented, using different approximations and interaction schemes.^[3]

Within the QM/MM methods, the primary models for representing the QM region can be grouped into molecular orbital (MO) and valence bond (VB)-based approaches. While MO-based approaches are currently the dominant computational method, VB-based approaches remain equally important conceptual tools for studying chemical mechanisms and reactivity. Herein, we will demonstrate that VB approaches, and in particular the empirical valence bond (EVB) method, provide a powerful way to incorporate the classical concepts of physical organic chemistry to the study of chemical reactions by means of computation. We will start this chapter with a historical overview of the use of both MO and VB methods in the context of (bio)molecular modeling, introducing the basic theoretical aspects of both approaches. Tying in with the overarching theme of the book, our focus will be on the EVB method, its applications to the study of biological processes, and the more recent developments and extensions of this approach.

2.2 Historical Overview

2.2.1 From Molecular Mechanics to QM/MM Approaches

Molecular mechanics (MM) calculations were first performed in the early 1960s with the arrival of digital computers in universities.^[4] The initial applications were the conformational analyses of cycloalkanes by several groups, including James Hendrickson at UCLA and Brandeis,^[5] Kenneth Wiberg at Yale,^[6] Norman Allinger at Georgia,^[7] and Shneior Lifson at the Weizmann Institute.^[8] In the late 1960s, Lifson and Warshel^[9] developed the first “consistent force field” (CFF), which formed the basis for the most common biomolecular force fields used today, that is, force fields such as AMBER, CHARMM, GROMOS, and OPLS.^[10] Initially, they focused on small molecules. However, in 1969 Levitt and Lifson extended the use of the CFF to perform the first energy minimizations of entire proteins, specifically myoglobin and lysozyme.^[11]

At around the same time, the first *ab initio* MO calculations on diatomic molecules began to appear and chemical applications for larger π -electron systems became possible at the semiempirical level, for example, by using the Pariser–Parr–Pople (PPP) method.^[12] However, calculations on chemical processes were still very rare, both due to the limitations in the available methodology as well as the corresponding limitations in hardware performance at the time. In 1971, Karplus and Honig implemented a hybrid approach to study the conformational energetics and electronic spectra for π -conjugated molecules.^[13] Within this approach, the π -system was described using the PPP SCF-CI calculations, while Buckingham terms were used to estimate the change in non-bonded interactions for different conformers. This approach was subsequently generalized by Warshel and Karplus,^[14] in which a CFF treatment of the σ -framework was merged with PPP calculations for the π -system. This hybrid approach was further improved and presented by Warshel and Levitt in their seminal work, *Theoretical Study of Enzymatic Reactions: Dielectric, Electrostatic and Steric Stabilization of the Carbonium Ion in the Reaction of Lysozyme*.^[15] In this work, all of the basic concepts of the QM/MM method

were introduced, including the partitioning of the system, the form of the potential energy function, and the interactions between the QM and MM parts.^[15] All these developments, which laid the foundations of the entire current field of QM/MM calculations, are part of the motivation for the 2013 Nobel Prize in Chemistry.^[1]

Ten years after Warshel and Levitt's seminal paper,^[15] Singh and Kollman^[16] devised a combined QM/MM approach, implemented in the QUEST program, that merged the Gaussian 80 UCSF package with the AMBER force field. They applied this method to study the methyl chloride/chloride exchange reaction in solution and the protonation of polyethers in the gas phase.^[16] In this work, Singh and Kollman introduced the notion of *junction dummy atoms* to saturate the free valencies of the QM atoms linked to MM atoms. This approach, later presented as the *link-atom approach*^[17] remains one of the most popular strategies in combined QM/MM applications today. Inspired by this work, Karplus and coworkers^[17] implemented a combined QM/MM model using semiempirical methods (MNDO and AM1) in the CHARMM program to study (bio)chemical reactions. Despite the advances made since 1970s, when the basic ideas of combined QM/MM methods were introduced, it was not until the 1990s that these approaches became widely used for the study of chemical and biomolecular systems.^[17,18]

2.2.2 Molecular Orbital (MO) vs. Valence Bond (VB) Theory

Equally important to the technical advances of both MM and combined QM/MM methods has been the development of more efficient and accurate QM approaches for describing the reactive region of the system. Among them, the most common choice, whether *ab initio* or semiempirical, have been Molecular Orbital (MO)-based approaches. However, as discussed below, this was not always the case. A more detailed historical account of VB theory can be found in the reference.^[19]

From 1927 to the mid-1950s, VB theory competed neck-and-neck with MO-based approaches and at times even dominated the field. This started with Heitler and London, who used VB theory to explain covalent bonding in the H₂ molecule.^[20] This was a major advance for chemistry in general, as it provided an explanation for one of the biggest problems in chemistry at the start of the last century, namely: how can neutral molecules be bonded?^[20] Further advances and refinements of VB models allowed VB theory to go from strength to strength, culminating with Marcus' use of VB theory to describe electron transfer in the mid-1950s,^[21] an achievement that ultimately led him to be awarded the 1992 Nobel Prize in Chemistry.^[22]

At the end of 1920s, closely after VB theory was established, Mulliken^[23] and Hund^[24] developed an alternative approach, called *molecular orbital theory*, where the many-electron wave function of a molecule was described by delocalized molecular orbitals. Mulliken received the Nobel Prize in Chemistry in 1966 for this contribution.^[25] However, it was the work of Hückel^[26] that gained MO theory its current popularity within the chemistry community. With these two alternative approaches available, the VB-MO rivalry began. As a result of this rivalry, the 1950s to 1980s were particularly challenging for VB theory. This was mainly due to two key reasons, one scientific and one pragmatic. The first one was a number of perceived failures of VB theory, which led the community to abandon it in favor of MO theory.^[27] Briefly, these can be centered on the hybridization of the O₂ molecule, which early VB theory would describe as

diamagnetic, but which is actually paramagnetic,^[28] as well as the “war over benzene,” which is eloquently described in the references [29, 30]. As demonstrated by Shaik^[27] these perceived failures of VB theory are in fact due to misuse and oversimplified interpretations of the VB approach, rather than the method itself. At a more pragmatic level, while correct programming of VB was extremely challenging, by 1950 one could already correctly program a model for benzene using MO theory.^[31]

Despite these difficult years for VB theory, from the 1980s onward, the use of VB-based methods in computer simulations has seen a renaissance, including both applications of the theory and developments for its computational implementation.^[32–35] This resurgence culminated in the indirect award of a second Nobel Prize for valence bond theory,^[1] which included Warshel’s empirical valence bond (EVB) approach.^[36] Today, the range of VB-based methods used is broad and involves a variety of techniques and approaches, including both *ab initio* and empirical variants. We recently provided an introductory generalized review on *How VB Theory can Help You Understand Your (Bio)chemical Reaction*.^[35] Here we will only briefly review the key aspects of VB theory and focus on the EVB approach and other semi-empirical incarnations of VB theory. For a more complete overview of different modern VB methods, we refer the reader to, for example, the work of Shaik, Hiberty and coworkers.^[19,27,32]

2.3 Introduction to Valence Bond Theory

In this section we will briefly discuss the main features of valence bond (VB) and molecular orbital (MO) theory, as well as their differences in the context of modeling chemical reactions. In the VB formalism the exact molecular wave function of a system is approximated as a linear combination of canonical (valence bond) chemical structures based on pure atomic orbitals (AOs) or hybrid atomic orbitals (HAOs).^[32] In classical VB theory, only the electrons that participate in the electronic reorganization during the bond making/breaking process are distributed in these orbitals. The wave function of the system is then optimized with respect to the structural coefficients, leading to a self-consistent field-type wave function.

In the case of the H₂ molecule, with its two AOs, χ_a and χ_b , the wave function is given as a linear combination of a covalent (H[•]–H[•]) and two ionic (H⁺ : H[–] and H : [–]H⁺) structures, see Figure 2.1. The covalent structure can be mathematically represented by the Heitler-London (HL) wave function:

$$\Phi_{HL} = |\chi_a\bar{\chi}_b| - |\bar{\chi}_a\chi_b| \quad (2.1)$$

where $|\chi_i\bar{\chi}_j|$ are the determinants of the form:

$$\begin{aligned} |\chi_a\bar{\chi}_b| &= \frac{1}{\sqrt{2}}[\chi_a(1)\alpha(1)\chi_b(2)\beta(2) - \chi_a(2)\alpha(2)\chi_b(1)\beta(1)] \\ |\bar{\chi}_a\chi_b| &= \frac{1}{\sqrt{2}}[\chi_a(1)\beta(1)\chi_b(2)\alpha(2) - \chi_a(2)\beta(2)\chi_b(1)\alpha(1)] \end{aligned} \quad (2.2)$$

that describes the two atomic spin-orbitals a and b , each having one electron (labeled 1 and 2); the bar on the AO description means a β spin, while no bar denotes an α

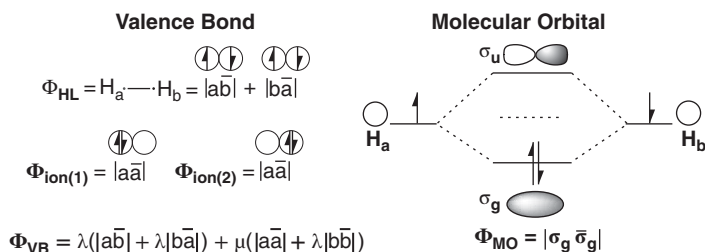


Figure 2.1 Valence Bond (VB) and Molecular Orbital (MO) Representations of the H_2 Molecule. In the VB description, the bonding of this molecule is represented by a covalent structure (also called a *Heitler-London (HL)* structure), as well as two ionic structures where one of the atoms posses two electrons and the other one none. In the MO description, the atomic orbitals mix to yield bonding (σ_g) and antibonding (σ_u) orbitals. Each orbital accommodates two electrons and the two electrons of the H_2 molecule fill the σ MO.

spin. Taking into account both the covalent (Eq. 2.1) and ionic contributions ($|\chi_a \bar{\chi}_a|$ and $|\chi_b \bar{\chi}_b|$), the VB wave function can be written as:

$$\Psi_{VB} = \lambda(|\chi_a \bar{\chi}_b| - |\bar{\chi}_a \chi_b|) + \mu(|\chi_a \bar{\chi}_a| + |\chi_b \bar{\chi}_b|) \quad \lambda > \mu \quad (2.3)$$

At the equilibrium bond distance, the bonding is predominantly covalent (about 75%), but it also includes some ionic component. As the bond is stretched, the weight of the ionic structures (μ) gradually decreases until a sufficiently long distance is reached, at which point the VB wave function is completely described by the HL wave function, such that at infinite distances, the molecule dissociates into two radicals.

In contrast, MO theory describes the state wavefunction of a given chemical system using delocalized molecular orbitals, which are generated as a linear combination of atomic orbitals. In MO theory, two types of molecular orbitals are defined: bonding (σ_g) and antibonding (σ_u) MOs. In the case of the H_2 molecule, one of the MOs (σ_g) is constructed by adding the mathematical functions for the two 1s atomic orbitals, while the antibonding orbital (σ_u) is formed by the difference between the two AOs, due to destructive interference:

$$\sigma = \chi_a + \chi_b ; \sigma^* = \chi_a - \chi_b \quad (2.4)$$

where normalization constants are omitted for simplicity. At the simplest MO level, the ground state of H_2 is described by the configuration σ^2 , in which the bonding σ_g orbital is doubly occupied. Expansion of the MO determinant into its AOs leads to:^[27]

$$\Psi_{MO} = |\sigma \bar{\sigma}| = (|\chi_a \bar{\chi}_b| - |\bar{\chi}_a \chi_b|) + (|\chi_a \bar{\chi}_a| + |\chi_b \bar{\chi}_b|) \quad (2.5)$$

The first half of Eq. 2.5 is the HL wave function of Eq. 2.1, while the remaining part is ionic. Eq. 2.5 provides an incorrect description of the dissociation process, as the wave function is always half-covalent and half-ionic, irrespective of the bonding distances. However, this can be corrected by including an appropriate configuration interaction (CI), describing the mixing of the ground state configuration (σ_g^2) with the di-excited one σ_u^2 , which is the same as σ_g^2 , but coupled with a negative sign:

$$\Psi^* = |\sigma_u \bar{\sigma}_u| = -(|\chi_a \bar{\chi}_b| - |\bar{\chi}_a \chi_b|) + (|\chi_a \bar{\chi}_a| + |\chi_b \bar{\chi}_b|) \quad (2.6)$$

In this way the MO-CI wave function is written as:

$$\begin{aligned}\Psi_{MO-CI} &= c_1|\sigma\bar{\sigma}| - c_2|\sigma_u\bar{\sigma}_u| \\ &= (c_1 + c_2)(|\chi_a\bar{\chi}_b| - |\bar{\chi}_a\chi_b|) + (c_1 - c_2)(|\chi_a\bar{\chi}_a| + |\chi_b\bar{\chi}_b|) \\ (c_1 + c_2) &= \lambda; (c_1 - c_2) = \mu\end{aligned}\quad (2.7)$$

Because c_1 and c_2 are variationally optimized, expansion of Ψ_{MO-CI} leads to the same description as Ψ_{VB} in Eq. 2.3, with $(c_1 + c_2) = \lambda$ and $(c_1 - c_2) = \mu$. From Eqs. 2.3 and 2.7, it is clear that both theories, when properly implemented, are correct and mutually transformable.^[37] Moreover, despite the differences in language used by each of them, both theories are exact in their more elaborate versions. This equivalence was recognized as early as 1935 by van Vleck and Sherman,^[38] but unfortunately not appreciated by the chemistry community, which is why VB is only recently observing the resurgence it deserves.^[33] Despite the simplicity of the model presented here, the MO-VB correspondence is general. In fact, any MO wave-function can be exactly transformed in a VB wave function, provided it is a spin-restricted eigenfunction.^[27,37] A systematic protocol for doing this transformation is provided by Shaik and coworkers in Chapter 4 of their book.^[27]

Today, VB theory is back and it comprises a variety of approaches, including methods that use purely localized orbitals (also named *classical* VB methods) and methods involving semi-localized orbitals.^[32] Among the former are the valence bond self-consistent field (VBSCF) method^[39] and its variants,^[40,41] while the second approach includes the generalized VB (GVB) method,^[42] the spin-coupled VB (SCVB) approach,^[43] and the molecular mechanics-valence bond (MMVB) method.^[44] Additionally, several other variants of VB approaches that facilitate the study of chemical reactions in solvents and/or protein have been proposed. These include the VB polarizable continuum model (VBPCM),^[45] the VB solvation model (VBSM),^[46] and the QM(VB)/MM method of Shurki *et al.*^[47,48] A major contribution to the field has been the empirical VB (EVB) method of Warshel *et al.*,^[36] from which several other approaches have been developed (see Section 2.8). In the next section we will discuss in detail the theory behind EVB and related methods in the context of modeling (bio)chemical reactions.

2.4 The Empirical Valence Bond Approach

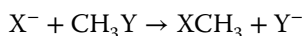
Despite significant advances in computational power, the use of higher-level QM approaches for representing the potential energy surface of complex (bio)chemical processes remains cumbersome. A strategy to address this problem is the empirical incarnation of valence bond theory (EVB). This approach is a semi-empirical QM/MM method that uses a fully classical description of the different VB configurations along a chemical reaction. The advantages of this approach are that it is fast, allowing for extensive conformational sampling, while carrying a tremendous amount of chemical information, allowing for a physically and quantitatively meaningful description of chemical reactivity.^[49]

As within a standard VB framework, EVB uses a set of VB configurations, which can involve covalent, ionic, or a mixture of bonding types, to describe the reactive system through the course of the chemical reaction. However, in this case, each VB state

corresponds to different bonding patterns of key energy minima (reactants, products, and any intermediates) instead of electronic configurations. In order to illustrate the basis of the method, we will examine the simple case of a S_N2 reaction as a didactic example, building on the model originally used by Warshel and coworkers.^[50,51] The same EVB treatment can be used for three or more states, as has been already shown by others.^[52–54]

2.4.1 Constructing an EVB Potential Surface for an S_N2 Reaction in Solution

For the simple case of an S_N2 reaction between a halide and halomethane (Figure 2.2), the reaction can be written as:



Depending on the identity of the nucleophile (X^-) and the leaving group (Y^-), the overall number of valence electrons may vary. However, independent of the specific system, there will always be at least four active electrons, that is, two electrons of the C–Y bond and the lone pair electrons of the X^- nucleophile, and three fragments, that is, the nucleophile (X), the CH_3 group (C), and the leaving group (Y). Hence, the chemical process can be described according to the general prescription of the four-electron three-orbital model.^[55]

The VB wavefunction for this system can be represented by a linear combination of six VB configurations resulting from this active space. Inspection of the bonding diagram in Figure 2.2a indicates that three of these configurations have high energy: *states 4* and *5* have an unfavorable juxtaposition of negative charges, while *state 6* has an almost completely broken bond as well as a carbanion. In a full treatment one would need to use all these resonance structures to recover the correlation energy. However, for the approximate empirical treatment, one can neglect these high-energy configurations,

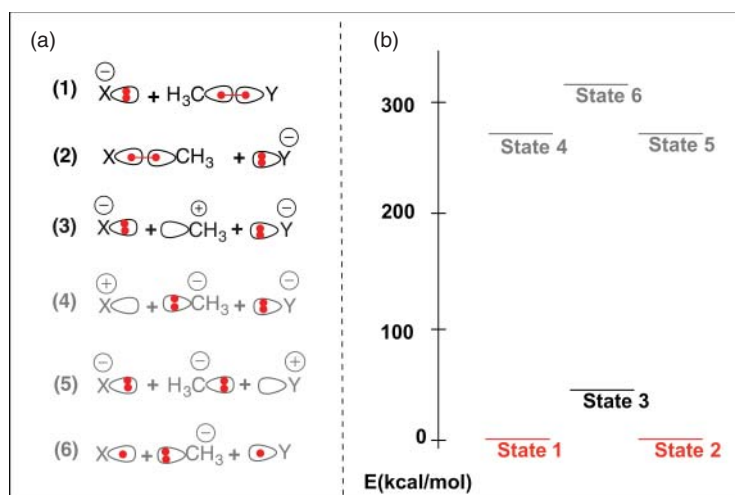


Figure 2.2 (a) Different VB Configurations Describing All Possible Independent Arrangements of Four Active Electrons Within Three Centers (X, CH₃, and Y). (b) Relative Approximate Energy Levels of the VB States; see Table 2.1 For the Estimation of These Energies. (See color plate section for the color representation of this figure.)

Table 2.1 *Back of the Envelope* Estimation of the Energies of the Valence-Bond States (kcal/mol) Used in Calculating the Energies of the States Shown in Figure 2.2. This table was originally presented in reference [50].

Parameter	Abbreviation ^{a)}	Value
Halogen (X) atom	IP_X	250
	EA_X	70
Methyl (CH_3) group	IP_{CH_3}	230
	EA_{CH_3}	≈ 0
C-X covalent bond	D_{C-X}	60
4.0 Å X-X bond	D_{X-X}	0
+-- Charge distribution	$V_{QQ}(+--)$	-60
-+- Charge distribution	$V_{QQ}(-+-)$	-180

Approximate Valence-Bond State Energies

$$E(1) = \text{reference state energy} = 0$$

$$E(2) = \text{reference state energy} = 0$$

$$E(3) = V_{QQ}(-+-) + IP_{CH_3} - EA_X + D_{C-X} = +40$$

$$E(4) = V_{QQ}(+--) + IP_X - EA_{CH_3} + D_{C-X} = +250$$

$$E(5) = V_{QQ}(+--) + IP_X - EA_{CH_3} + D_{C-X} = +250$$

$$E(6) = IP_X - EA_{CH_3} + D_{C-X} - D_{X-X} = +310$$

a) IP: Ionization potential; EA: Electron affinity; D: dissociation energy.

assuming that their contribution is small and that the error involved by doing so will be absorbed in the parametrization procedure. Using the three low-energy states the basis set for the electronic wavefunction can be defined as:

$$\begin{aligned}\Phi_1 &= [X^-C^--Y]\chi_1 \\ \Phi_2 &= [X^--CY^-]\chi_2 \\ \Phi_3 &= [X^-C^+Y^-]\chi_3\end{aligned}\tag{2.8}$$

where X, Y, and C designate atomic orbitals on the corresponding atoms, the χ 's are the wave functions of the inactive electrons moving in the field of the active electrons, Φ_1 and Φ_2 correspond to the reactant and product states, respectively, and Φ_3 is a Zwitterionic state, having two halide anions separated by a carbocation. Here, the usual EVB procedure would involve diagonalizing a 3×3 secular equation. However, since the energy of Φ_3 is higher than Φ_1 and Φ_2 , the three resonance states can be reduced to an effective two-state problem (ψ_1 and ψ_2), where one state is mostly Φ_1 (which describes the electronic structure of the reactant state) and the other is mostly Φ_2 (which describes the electronic structure of the product state):

$$\begin{aligned}\psi_1 &= \alpha_1\Phi_1 + \beta_1\Phi_3 \\ \psi_2 &= \alpha_2\Phi_1 + \beta_2\Phi_3\end{aligned}\tag{2.9}$$

The lowest electronic eigenvalue, that is, the potential energy surface, is then given by the solution of the 2×2 secular equation:

$$\begin{vmatrix} H_{11} - E_g & H_{12} - E_g S_{12} \\ H_{21} - E_g S_{21} & H_{22} - E_g \end{vmatrix} = 0 \quad (2.10)$$

where the matrix elements are given by $H_{ij} = \epsilon_{ij} = \langle \psi_i | H | \psi_j \rangle$ and $S_{ij} = \langle \psi_i | \psi_j \rangle$. These matrix elements can in principle be evaluated using standard quantum chemical methods. However, this evaluation is tedious and the assumptions made above will lead to errors in the matrix elements. In the EVB framework, no electronic matrix elements are actually calculated. Instead, H_{11} and H_{22} are described by empirical functions that incorporate bond stretching, as well as electrostatic and nonbonded interactions, in order to describe the potential surface of each reactant and product state. H_{11} and H_{22} are often referred to as *diabatic* potential surfaces, in contrast to the E_g itself, which is the *adiabatic* potential surface. Additionally, in the classical EVB approach the overlap integral, S_{ij} , is neglected and it is assumed that its effect is absorbed into the parametrization of H_{12} .

In our example of a bond breaking/forming process, the energy ϵ_i^0 of each individual VB state in gas phase (indicated by the superscript 0), can be approximated by potential functions of the form:

$$\begin{aligned} \epsilon_1^0 &= H_{11}^0 = \Delta M(b_1) + \frac{1}{2} \sum_m K_{\theta,m}^{(1)} (\theta_m^{(1)} - \theta_{m,0}^{(1)})^2 + U_{nb}^{(1)} + U_{inact}^{(1)} \\ \epsilon_2^0 &= H_{22}^0 = \Delta M(b_2) + \frac{1}{2} \sum_m K_{\theta,m}^{(2)} (\theta_m^{(2)} - \theta_{m,0}^{(2)})^2 + U_{nb}^{(2)} + U_{inact}^{(2)} + \alpha_2^0 \\ H_{12} &= A \exp\{-\mu(r_3 - r_{3,0})\} \end{aligned} \quad (2.11)$$

where $\Delta M(b_i)$ is a Morse-type potential function taken relative to its minimum value ($\Delta M(b_i) = M(b_i) - D$), $U_{nb}^{(i)}$ is the sum of nonbonded interaction in the given configuration, b_1 , b_2 , and r_3 are the X-C, C-Y, and X-Y distances, respectively, and θ_s are the X-C-H and H-C-Y angles defined by the covalent bonding arrangement for a given resonant structure. The subscript 0 in $r_{3,0}$ and $\theta_{m,0}$ refers to the equilibrium distance and angle, respectively. In this case no torsion parameters exist, otherwise the change in these parameters should also be considered. The potential $U_{inact}^{(i)}$ represents the interaction with the inactive part of the reactive system and is generally described by:

$$\begin{aligned} U_{inact}^{(i)} &= \frac{1}{2} \sum_{bonds} K_b^{(i)} (b - b_0^{(i)})^2 + \frac{1}{2} \sum_{angle} K_\theta^{(i)} (\theta - \theta_0^{(i)})^2 \\ &+ \sum_{torsions} K_\phi^{(i)} (1 + \cos(n^{(i)} \phi^{(i)} - \delta^{(i)}))^2 + U_{nb,inact}^{(i)} \end{aligned} \quad (2.12)$$

where the b s, θ s and ϕ s are, respectively, the bond lengths, bond angles, and dihedral angles in each state. Finally, α_2^0 is the energy difference between ψ_1 and ψ_2 with the fragments at infinite separation. It should be pointed out that α_2^0 can, in principle, be applied to either of the diabatic states in order to raise or lower the energy of the relevant parabola. However, irrespective of which state is chosen, once the reaction has been calibrated in one environment, it is absolutely crucial to apply it to the same state consistently.^[35] The key point about the functional form of ϵ_i^0 is that

the experimental properties of the reactants and products at infinite separation are reproduced exactly.^[50] The behavior at the transition-state region can then be obtained by adjusting the off-diagonal matrix elements, H_{12} , to fit ab initio calculations and/or experiments. One may also use different functional forms for the off-diagonal matrix elements, as shown by Chang and Miller,^[56] who use generalized Gaussians, or by Schlegel and Sonnenberg,^[57] who use quadratic polynomials combined with a spherical Gaussian. The most important aspect here is that the energetics and geometries of the transition-state region, where the coupling between the two adiabatic states is stronger, can be reasonably well represented.^[58]

Having obtained all the matrix elements of the Hamiltonian, the ground-state potential surface E_g of the system is obtained as the lowest eigenvalue of the 2×2 secular equation (Eq. 2.10):

$$(\epsilon_{11} - E_g)(\epsilon_{22} - E_g) - H_{12}^2 = 0 \quad (2.13)$$

This is a quadratic equation in E_g , applying the general solution for such equations ($ax^2 + bx + c = 0$) gives:

$$E_g = \frac{1}{2} \left[(\epsilon_1 + \epsilon_2) - \sqrt{(\epsilon_1 - \epsilon_2)^2 + 4H_{12}^2} \right] \quad (2.14)$$

The effect of the surroundings (solvent or enzyme, designated as S) on the reaction Hamiltonian is obtained by adding the corresponding energies to the diagonal matrix elements:

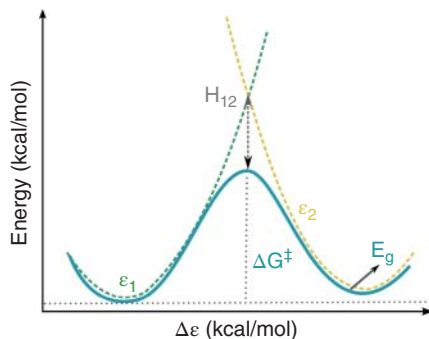
$$\begin{aligned} \epsilon_1^S &= H_{11} = H_{11}^0 + V_{rs}^1 + V_{ss} \\ \epsilon_2^S &= H_{22} = H_{22}^0 + V_{rs}^2 + V_{ss} \\ H_{12} &= H_{12}^0 \end{aligned} \quad (2.15)$$

where V_{rs}^i is the interaction potential between the solute atoms in the i^{th} VB configuration and the surroundings and V_{ss} is the surroundings–surroundings interaction potential. It should be noted that the gas-phase shift has been demonstrated to be phase-independent, and rigorous constrained density functional theory (CDFT) calculations have shown that the off-diagonal elements remain the same for any electrostatic environment.^[59] This is a powerful feature of the EVB approach, as it allows one to model the same chemical reaction in distinct chemical environments, using the same gas-phase shift and off-diagonal values, and thus directly observe how environmental changes affect the reaction. The nature of this analytical expression can best be understood from Figure 2.3. In this figure, the diabatic energy profiles of the reactant (ϵ_1) and product (ϵ_2) states represent a microscopic equivalent of the Marcus parabolae. The mixing of these diabatic states results in the ground-state potential surface E_g , where the stabilization resulting from the mixing of the two states at the activation barrier is given by H_{12} .

2.4.2 Evaluation of Free Energies

So far we have only considered how to obtain the potential energy surface for a given reaction. Although it is clearly of interest in itself, the objective is to calculate the corresponding free energy function, in particular the Gibbs free energy, which isothermal–isobaric (NpT) ensemble is closer to the real experimental conditions.

Figure 2.3 Relationship Between the Diabatic Parabolas ϵ_1 and ϵ_2 and the Adiabatic Ground State Energy E_g of a Hypothetical Two-State Reaction, Using an EVB Description. (See color plate section for the color representation of this figure.)



The strategy for this involves the free-energy perturbation (FEP) approach and the so-called umbrella sampling (U.S.) procedure.^[60] The perturbation formula^[61] for the free energy difference between two states is:

$$\delta G(1 \rightarrow 2) = -k_B T \ln \langle \exp\{-(\epsilon_2 - \epsilon_1)/k_B T\} \rangle_1 \quad (2.16)$$

here, ϵ_1 and ϵ_2 denote the potentials of the two *diabatic* states 1 and 2 and $\langle \rangle_1$ denotes an ensemble average obtained on the potential ϵ_1 . Although the expression in Eq. 2.16 is exact, it converges very slowly if states ϵ_1 and ϵ_2 are far away from each other in configurational space. However, by introducing a coupling parameter, λ_m ($0 \leq \lambda_m \leq 1$), the two states can be connected *via* a set of intermediate potentials that are sufficiently similar for allowing an adequate sampling between these two potentials.^[62] This is done by defining a *mapping potential* (ϵ_m) of the form:^[63]

$$\epsilon_m(\lambda_m) = \epsilon_1(1 - \lambda_m) + \epsilon_2\lambda_m \quad (2.17)$$

The subscript m in Eq. 2.17 is used to distinguish λ_m from the reorganization energy, λ_r , to be discussed in Chapter 7. Note that ϵ_1 has a minimum at the reactant geometry and ϵ_2 has a minimum at the product geometry so that as ϵ_m changes, the system is forced to move from the reactant state to the product state. The free energy increment $\delta G(\lambda_m \rightarrow \lambda_{m'})$ associated with the change of ϵ_m to $\epsilon_{m'}$ can be evaluated by:

$$\delta G(\lambda_m \rightarrow \lambda_{m'}) = -k_B T \ln \langle \exp\{-(\epsilon_{m'} - \epsilon_m)/k_B T\} \rangle_m$$

$$\Delta G(\lambda_n) = \delta G(\lambda_0 \rightarrow \lambda_n) = \sum_{m=0}^{m=n-1} \delta G(\lambda_m \rightarrow \lambda_{m'}) \quad (2.18)$$

where the average $\langle \rangle_m$ is evaluated in the potential surface ϵ_m . It must be emphasized that $\Delta G(\lambda_n)$ only represents the free energy associated with moving from ϵ_1 to ϵ_2 on the constraint potential ϵ_m and it is not sufficient for evaluating the activation free energy, ΔG^\ddagger , which reflects the probability of being at the transition state on the actual ground-state potential surface E_g . In order to obtain the free energy, $\Delta G(x)$, corresponding to the trajectories moving on the actual ground-state potential, E_g , one needs to define a reaction coordinate for the reaction of interest. A general reaction coordinate, x , can be defined in terms of the *energy gap*, $\Delta\epsilon$, between the two diabatic states ($\Delta\epsilon = \epsilon_1 - \epsilon_2$).^[51,63] The advantage of such a coordinate is that it allows for the projection of the full multidimensional space onto a single reaction coordinate, without assuming a pre-defined reaction path. In addition, it has been demonstrated that it can

greatly accelerate the convergence of free energy calculations.^[64] The corresponding free energy is given by:

$$\exp\{-\Delta g(X)/k_B T\} \cong \exp\{-\Delta G(\lambda_m)/k_B T\} \langle \exp\{-(E_g(X) - \varepsilon_m(X))/k_B T\} \rangle_m \quad (2.19)$$

This expression allows us to calculate the energy difference between the mapping potential ε_m and the ground-state potential E_g at each point of the MD trajectory and use the Boltzmann average of this difference to correct the free energy obtained on the mapping potential.^[58] The evaluation of $\Delta G(x)$ for an exchange reaction is described in Figure 2.3. Then, $\Delta G(x)$ can be used to determine the values of ΔG^\ddagger and ΔG^o , and adjust α_2^0 and H_{12} until the calculated and observed values of these free energies coincide with the experimental values.

2.5 Technical Considerations

Since the original work of Warshel and Weiss,^[36] more than 30 years ago, the EVB approach has been continuously developed and increasingly applied to the study of chemical and biochemical reactions.^[49,58,65–70] Additionally, as will be discussed in Section 2.8, several other approaches based on the EVB philosophy have since been formulated. However, the concerns about the reliability of this approach have also been raised.^[71] Therefore, in this section we will highlight key technical considerations that could be considered potential weaknesses of this approach and describe how they are addressed within an EVB framework.

2.5.1 Reliability of the Parametrization of the EVB Surfaces

The EVB approach is a semiempirical QM/MM method that relies on proper parameterization of the force fields describing the individual reacting states. Extensive work by Warshel and coworkers has focused on searching for an effective and reliable parameterization approach.^[72] Such parameterization can, in principle, be done based on either *ab initio* or experimental data. Reproducing exact gas-phase *ab initio* results has historically been one of the most popular alternatives for calibration of semi-empirical models,^[56] as they allow one to perform high-level electronic structure calculations at affordable computational cost. However, gas-phase *ab initio* calibration is neither a requirement nor the *only* option for obtaining quantitatively accurate EVB surfaces. In fact, they might be inaccurate when studying reactions in condensed phases, where polarization and screening effects can be difficult to parameterize. In such cases, the inclusion of either implicit or explicit solvation models is required in order to obtain physically meaningful results. Even more so, where available, the use of relevant experimental data is clearly the most meaningful way for obtaining accurately calibrated reaction surfaces. An example of this is the study of the first step in the reaction catalyzed by haloalkane dehalogenase (DhlA), where a key unresolved question has been the activation energy in water.^[72] Warshel and coworkers first tried to use *ab initio* calculations for the calibration of this reaction; however, the presence of negatively charged ions turned out to be challenging. Instead, the use of relevant experimental information, including

careful analysis of LFER results, proved to be the best alternative for parametrization.^[72] The accuracy of this parametrization has been later confirmed by quantitative ab initio FEP calculation on this system.^[73]

In conclusion, the use of ab initio data in the parametrization of EVB surfaces is not a particular feature of new EVB approaches nor a requirement for obtaining quantitatively accurate EVB surfaces. Since the 1980s, the EVB approach has used ab initio data as well as experimental information in the parametrization of EVB surfaces. Both sources of information, when carefully used, can provide important insights for the proper calibration of EVB surfaces.

2.5.2 The EVB Off-diagonal Elements

The power of the EVB is largely due to its *simple* orthogonal diabatic representation and the assumption that the off-diagonal elements of the EVB Hamiltonian do not change significantly when transferring the reacting system from one phase to another. This approximation has been rigorously validated using the frozen density functional theory (FDFT) and the constrained density functional theory (CDFT) models to examine the relationship between the diabatic and adiabatic surfaces, as well as the corresponding effective off-diagonal elements, in a prototype S_N2 reaction.^[59] In the FDFT/CDFT approach, the system is divided into subsystems and the Kohn–Sham formulation is applied to each subsystem separately. For the case studied in reference,^[59] the system was divided into two subsystems: Cl^-_A ion and the CH_3Cl_B , in order to obtain the diabatic energies of the reactant (ϵ_1) and product state (ϵ_2). The adiabatic ground-state energy (E_g) was then obtained by treating the whole system at the same level with a regular DFT calculation. Performing such calculations in the gas phase and in solution provides per definition the rigorous off-diagonal element:

$$H_{12} = \sqrt{(\epsilon_1 - E_g)(\epsilon_2 - E_g)} \quad (2.20)$$

The result of this relationship is shown in Figure 2.4 and demonstrates that the use of a solvent independent H_{12} is therefore an excellent assumption. This study demonstrated that using formally orthogonal diabatic surfaces, where the solvation of each diabatic state is included in the diabatic energies, is an excellent approximation.

2.5.3 The Choice of the Energy Gap Reaction Coordinate

As shown before (Eq. 2.17), the EVB U.S.-FEP mapping is a particularly effective mapping that uses the *energy gap* ($\Delta\epsilon = \epsilon_1 - \epsilon_2$)^[63] as a generalized reaction coordinate. This is a particularly powerful choice of reaction coordinate when dealing with multidimensional reaction coordinates, as it allows for the projection of the full multidimensional space onto a single reaction coordinate,^[50] without assuming a pre-defined reaction path, while at the same time accounting for the full environmental reorganization along the reaction coordinate. Additionally, it has also been demonstrated that the use of the *energy gap* as the reaction coordinate greatly accelerates the convergence of free energy calculations.^[64] Therefore, the EVB mapping approach and its associated energy gap reaction coordinate is an increasingly appreciated approach in the scientific community.^[64,74]

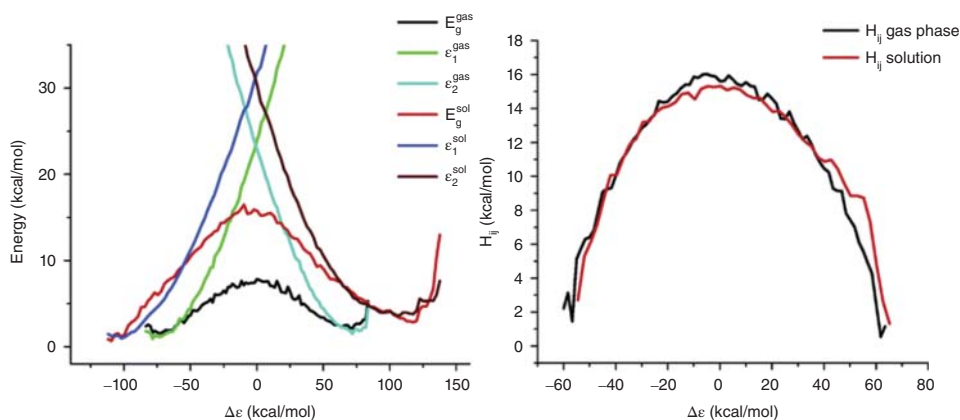


Figure 2.4 (left) Diabatic and Adiabatic FDFE Energy Profiles For the Reaction, $\text{Cl}^- + \text{CH}_3\text{Cl} \rightarrow \text{ClCH}_3 + \text{Cl}^-$, in Gas Phase and in Solution, Where the Reaction Coordinate is Defined as the Energy Difference Between the Diabatic Surfaces, $\Delta\epsilon = \epsilon_1 - \epsilon_2$. (Right) Plot of the H_{ij} of the reaction, $\text{Cl}^- + \text{CH}_3\text{Cl} \rightarrow \text{ClCH}_3 + \text{Cl}^-$, both in the gas phase and in solution.^[59] Reprinted with permission from reference [59]. Copyright 2006 American Chemical Society. (See color plate section for the color representation of this figure.)

2.5.4 Accuracy of the EVB Approach For Computing Detailed Rate Quantities

Finally, the reader may have some concerns^[71] about whether the EVB can reliably calculate detailed kinetic quantities such as kinetic isotope effects, which require accurate treatment of the underlying potential energy surface, as well as zero-point energies (ZPE) and quantum mechanical tunneling. The power of the EVB for performing such calculation has been demonstrated in numerous studies (by Warshel,^[75] Åqvist,^[76] Hammes-Schiffer,^[77] among others). In fact, as will be shown in Section 2.7, the EVB approach is one of the most effective ways of reproducing isotope effects and there are no reasons for concern over neither qualitative nor quantitative accuracy.

2.6 Examples of Empirical Valence Bond Success Stories

In the previous sections we have outlined the most important theoretical aspects to the modeling of chemical reactions using the EVB approach. In this section we will briefly describe how the EVB method has contributed to our understanding of the mechanisms and function of various biological systems and to the understanding of the molecular basis of enzymatic catalysis in general.

2.6.1 The EVB Approach as a Tool to Explore Electrostatic Contributions to Catalysis: Staphylococcal Nuclease as a Showcase System

The early work of Warshel and Levitt on the catalytic reaction of lysozyme suggested electrostatic interactions as being the most important factor for enzymatic rate enhancement,^[15] in contrast to other more popular contemporary proposals, such as ground-state destabilization.^[78] Since then, a growing number of experimental and

theoretical studies (see discussion in references [66, 79] among others) have identified electrostatic effects as being a key factor in enzyme catalysis. In particular, EVB studies have provided quantitative understanding of the contributions of these electrostatic effects to enzyme catalysis, helping also to discriminate between different catalytic proposals.

One of the earliest calculations of the absolute free-energy surfaces for enzymes was presented by Åqvist and Warshel, who studied the cleavage of nucleic acids by staphylococcal nuclease (SNase).^[80,81] This work, which combined the EVB method with FEP/MD simulations, showed that the large catalytic effect of this enzyme could be directly attributed to electrostatic interactions between the active-site Ca^{2+} ion and the transition state. In a subsequent study of metal ion substitution in the active site of SNase, the same authors were able to reproduce the free energy changes associated with the substitution of different metal centers in the SNase active site. This study demonstrates that SNase was optimized to use Ca^{2+} as the catalytic metal, as Ca^{2+} provides an optimum balance between the energetic cost of the proton-transfer step, the interaction of the metal with the formed OH^- group, and the stabilization of the accumulating negative charge on the 5'-phosphate group at the transition state (Figure 2.5). This work provided one of the first examples of applying computer simulation methods to examine the role of metal ions in enzymes in a semi-quantitative way, and demonstrated that their major catalytic effect is associated with large electrostatic contributions. The study also provided a good example of how the use of relevant experimental information provides a powerful baseline for the calibration of the energetics of the background reaction in aqueous solution.

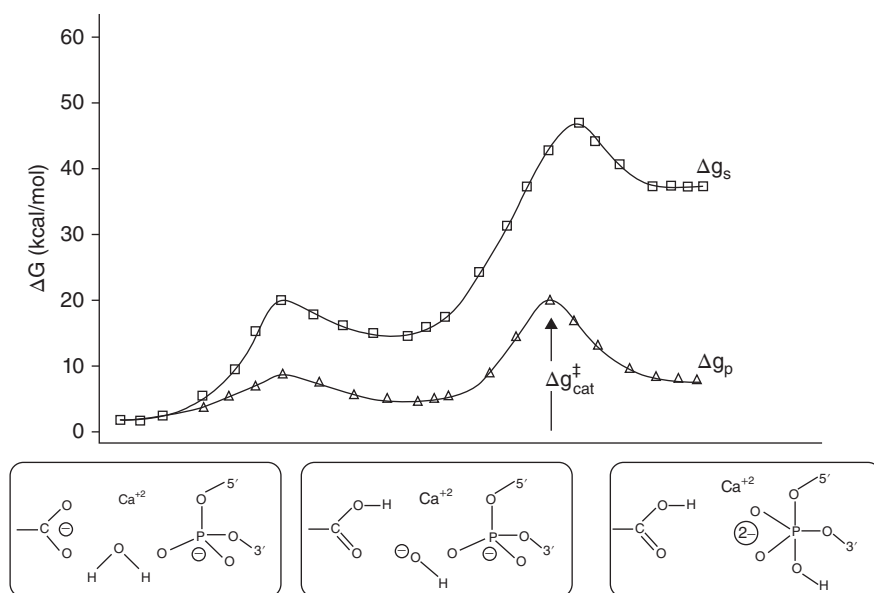


Figure 2.5 Calculated Free Energy Profiles For the Reaction of SNase and the Corresponding Reference Reaction in Solution. The lower inset represents the valence bond structures associated with each reaction step. Reprinted with permission from reference [66] (this figure was originally presented in reference [80]). Copyright 2006 American Chemical Society.

2.6.2 Using EVB to Assess the Contribution of Nuclear Quantum Effects to Catalysis

Nuclear quantum effects (NQE), particularly tunneling, play an important role in reactions involving transfer of light particles, such as in electron-transfer reactions and hydrogen-transfer reactions (H , H^+ , or H^-). However, the contribution of those effects to the total rate enhancement provided by enzymes has been highly debated.^[82,83] Experimentally, kinetic isotope effects (KIE) have been largely used as a proof of the manifestation of NQE in a given reactive process. However, clear evidence for the role of NQE in the enzymatic rate enhancement has been more difficult to obtain.

The calculation of NQE in enzymes using the EVB approach was first presented by Hwang and Warshel in their study of the hydride-transfer step in the catalytic reaction of lactate dehydrogenase (LDH).^[84,85] Their approach, which combined the classical EVB approach with the newly developed Quantum Classical Path (QCP) approach, provided grounds for using classical simulation methods in the evaluation of NQE in enzymatic reactions. The QCP method uses the Feynman's path integral (PI) centroid formulation, where each quantum particle is represented by a *ring* of quasiparticles subjected to a potential, to calculate the partition function at the transition state (TS) and reactant state (RS) of a given reaction. First, classical trajectories are generated along a collective reaction coordinate using an EVB potential in conjunction with FEP/US techniques. The nuclear quantum contribution is then determined for each configuration sampled at the RS and TS by performing PI calculations of the quantum mechanical partition function. Since calculating centroid probabilities in condensed-phase reactions is generally very challenging, the centroid of the beads are constrained to be at the position of the corresponding classical particle. The efficiency of this approach, which arises from the separation of the PI calculation from the classical MD simulation of the entire system, has been exploited in studies of quantum mechanical rate constants in solution^[85] and enzymatic active sites.^[76] These studies have greatly contributed to advancing our understanding of the role of NQE in enzyme catalysis. That is, it has been shown that even though NQE are present in reactions involving the movement of light-particles, the same NQE that occur in the enzyme-catalyzed reaction also occur in the corresponding reference reaction in solution. Therefore, while an inherent feature of the reaction itself, the NQE apparently do not make a significant contribution to the observed *catalytic effect* of the enzyme.^[75,82]

Following the original work of Warshel and coworkers, additional approaches have been introduced in studies of NQE in enzymatic reactions, such as the Molecular Dynamics with Quantum Transitions (MDQT) method of Hammes-Schiffer and coworkers^[86] and the Variational Transition State Theory (VTST) approach of Truhlar and coworkers.^[87] Like the QCP treatment, both methods allow one to include the protein motions in the classical region and thus provide a fully microscopic way of exploring vibrationally assisted nuclear tunneling.^[75]

2.6.3 Using EVB to Explore the Role of Dynamics in Catalysis

One of the proposals for the rationalization of the enormous catalytic power of enzymes is that dynamics plays a significant role in catalysis.^[88,89] Here, the concept of *dynamical effect* refers to the ability of enzymes to promote a certain vibrational mode along the reaction coordinate. As this hypothesis cannot be directly tested, its analysis has mainly been done through indirect considerations, such as the observed effects

of mutations on the catalytic activity of enzymes^[90] or by means of computational studies.

Theoretical studies addressing this hypothesis are computationally challenging, as they require extensive sampling of the full enzymatic models on the relevant time scales. Warshel *et al.* used the EVB approach^[36] in combination with coarse-grained (CG) models^[91] to investigate the experimentally observed effects of mutations on the catalysis of dihydrofolate reductase (DHFR) – one of systems most commonly used to investigate the dynamical proposal.^[92] This combined approach allowed them to compute free-energy landscapes as a function of both conformational flexibility and chemical step of the catalysis.^[92] By using CG models, the authors achieved sufficient sampling of the conformational coordinate at an affordable computational cost. Then, from structures generated from CG-based conformational landscape, they calculated the catalytic chemical step using the EVB approach. Finally, by combining the activation energy calculated at different conformations of the enzyme, a full catalytic landscape was obtained. Similarly to previous studies^[93], the calculations showed that conformational motions are fully randomized before they can be transferred to the chemical coordinate, and therefore dynamical effects cannot be used to accelerate chemical reactions. Instead, the largest contribution to the catalytic effect was found to come from the electrostatic preorganization.^[92]

Two key concepts of EVB that make it a particularly powerful choice for exploring dynamical effects in catalysis are the *energy gap* reaction coordinate and the *reorganization energy* (see Section 2.4.2 of this Chapter and Chapter 7). As mentioned before, the *energy gap* reaction coordinate is especially useful for examining conformational space, since it allows for the projection of multidimensional space onto a single reaction coordinate, without predefining any reaction path. At the same time, the calculation accounts for the full environmental reorganization along the reaction coordinate. The *reorganization energy*, on the other hand, allows one to precisely probe the determinants of chemical step of the catalysis, which is also particularly relevant when examining the contribution of conformational changes to the enzyme's catalytic ability.

2.6.4 Exploring Enantioselectivity Using the EVB Approach

During the last decade the use of enzymes to catalyze enantioselective reactions has experienced rapid growth and several efforts have been made to optimize these natural systems for industrial purposes.^[94] In particular, the development of both experimental and computational protein engineering tools has allowed for tremendous advances in the field. However, quantifying and manipulating the molecular aspects that govern enzyme selectivity still remains a significant challenge for current computational approaches. This is in part due to the very small energy differences that define selectivity, as well as the tight interplay between steric and electronic effects, making it necessary to use an approach that is both very precise and also allows extensive conformational sampling.^[95–97]

In this regard, the EVB approach has been shown to be a powerful tool for computer-guided design of enantioselective enzymes.^[95,97] For example, Fruschicheva *et al.*^[95] studied the enantioselective hydrolysis of α -substituted esters by both wild-type *Candida antarctica* lipase A (CALA) as well as a number of different mutant forms of this enzyme. Using the EVB approach for effective sampling and

free-energy calculations, in combination with several screening strategies, the authors were able to reproduce the overall observed enantioselectivity and to explore the effect of different mutations that switch the catalytic activities between the *R* and *S* enantiomers. Additionally, it was found that faster strategies for identifying mutations that could guide directed-evolution experiments, although useful, require further refinement if they are supposed to be used in a more quantitative manner. This work was later extended to CALB,^[97] which shows a switch in enantioselectivity from *R*- to *S*-preference upon mutating the active site tryptophan into the smaller alanine residue (W104A). Through a combination of the EVB and the restraint release (RR)^[98] approaches, the authors rationalized the experimentally observed switch in selectivity, which were found to be primarily steric and entropic in origin.^[97] These works have highlighted the importance of proper sampling to reliably rationalize enantioselectivity in enzymes, in particular when large conformational effects, due to mutations, are involved.

More recently, we have used the EVB approach to establish the principles underlying selectivity in potato epoxide hydrolase 1 (StEH1).^[99] This enzyme is also biocatalytically important as it catalyzes the ring opening of epoxide substrates to yield the corresponding vicinal diols. Using a combined approach, based on computational, biochemical, and structural characterization of the hydrolysis of trans-stilbene oxide by StEH1, we provided a unifying molecular model that, for the first time, explained the pH-dependent catalysis observed for this system, its regiopreference, and changes in activity and selectivity upon mutation. Additionally, our computational results suggested that this substrate regioselectivity is determined not at the first (alkylation) step of the reaction, but rather in the subsequent hydrolytic step, where a lack of electrostatic stabilization of one of the benzyl rings leads to a preference for the intermediate formed by attack at one of the carbon atoms. In addition, our results indicated the involvement of a previously neglected residue, H104, which is electrostatically linked to the general base H300. In combination, these insights obtained through the EVB approach together with experimental work allowed us to understand the main factors driving catalysis and regioselectivity in this enzyme, as well as identifying targets for subsequent enzyme design efforts.

2.6.5 Moving to Large Biological Systems: Using the EVB Approach in Studies of Chemical Reactivity on the Ribosome

Apart from numerous studies on the origins of enzymatic catalysis and enzymatic reaction mechanisms in general, there exist other notable examples of the application of the EVB methodology to study biological systems, including the work on elucidating the mechanisms involved in ribosomal translation processes.^[70,100–103]

For example, the possible catalytic mechanisms for ribosomal peptidyl transfer were first explored using the EVB approach by Trobro and Åqvist,^[100,101] who considered the different possible reaction mechanisms shown in Figure 2.6. Extensive simulations performed on a model of the reactive peptidyl transfer system with aminoacylated CCA fragments bound in the A- and P-sites indicated that the reaction does *not* proceed via general acid-base catalysis, as was previously hypothesized.^[104,105] Rather, out of

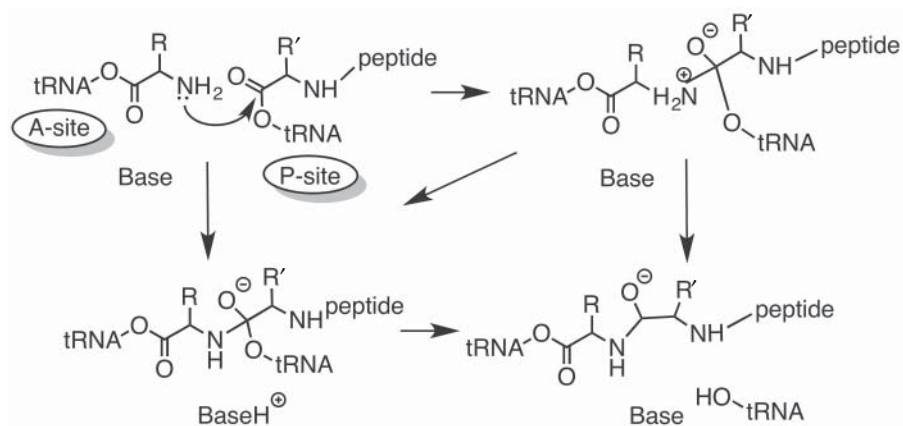


Figure 2.6 Possible Pathways For Ribosomal Peptidyl Transfer Involving the Aminolysis of the P-Site Peptidyl-tRNA Ester Bond by the Amino Group of the A-Site. A presumed ribosomal group acting as a base in the reaction is indicated. Adapted from reference [101]. Copyright (2005) National Academy of Sciences, U.S.A.

three possible pathways examined, the most feasible pathway involved the attack of the A-site α -amino group on the P-site ester, followed by an intramolecular proton shuttle proceeding via the P-site adenine O2' oxygen. The simulations allowed the authors to identify the groups involved in the peptidyl transfer reaction. Additionally, based on the comparison with the uncatalysed reaction, it has been also suggested that the catalytic effect in ribosomal catalysis of peptidyl transfer is purely entropic and stems from the decrease of the solvent reorganization energy rather than proper positioning of the substrate.^[106,107] These findings have been further discussed in subsequent work by Trobro and Åqvist,^[100] which focused on the analysis of the predictions for the mechanisms of peptidyl transfer and peptidyl-tRNA hydrolysis obtained from crystal structures and molecular simulations. Apart from drawing vital conclusions on the reactions of interest, these studies have demonstrated how computer simulations, and EVB in particular, can be used to examine structure-function relationships in biomolecules that are difficult to obtain from experimental work alone.

The details of the peptide bond formation in the ribosome have been also investigated by Sharma *et al.*,^[102] who specifically focused on the nature of the overall catalysis and its implications for the understanding of the catalytic factors in modern enzymes. Specifically, a systematic EVB study comparing the substrate-assisted peptidyl transfer in water and in the ribosome allowed the authors to examine the catalytic effect of the latter on the studied reaction. One of the main conclusions of this work was that electrostatic effects play a major role in reducing the activation barrier for the peptide bond formation catalyzed by ribosome, shedding light on the principles of the catalytic action of this ancient enzyme.

Another important subprocess in protein synthesis, namely guanosine triphosphate (GTP) hydrolysis catalyzed by the elongation factor (EF-Tu), has been also widely

studied using EVB. Adamczyk and Warshel explored the GTPase reaction of EF-Tu by performing extensive simulations on the wild type and several mutant forms of EF-Tu in an isolated form and in a complex with the ribosome.^[103] The EVB model employed in the study helped to advance the understanding of the long-distance EF-Tu activation by the ribosome and the origins of its catalytic effect, which, in the view of the authors' findings, is associated with an overall allosteric transition to a preorganized active form of EF-Tu, with the indirect effect of the ribosome playing the major role in the activation of the GTPase hydrolysis. It has been also found that the conserved His84 residue, although having a critical influence on the GTPase's catalysis by occupying a pivotal position in the active EF-Tu/ribosome/aa-tRNA complex, is not likely to act as a general base in the reaction. This finding was further supported by an EVB study of the activation of EF-Tu and Ras GTPase activating protein,^[70] which has also indicated that His84 does not serve as a base in the reaction, and that the origin of the catalytic effect is the transition state stabilisation through electrostatic allosteric effects. In a more recent contribution, Åqvist and Kamerlin^[108,109] have also employed the EVB approach to study various competing pathways of the GTP hydrolysis by the activated EF-Tu/ribosome complex. The simulations revealed several key features underlying the observed catalytic effect, focusing in particular on the role of the protonation state of His84 and the importance of the catalytic loop conformation for the GTPase activity of the EF-Tu/ribosome complex. It has been demonstrated that GTP hydrolysis is governed by the electrostatic interactions within the reaction center, where the net charge effect is the driving force for the reaction and leads to the stabilization of the transition state. The origins of the exceptional catalytic efficiency of EF-Tu, which is the fastest known GTPase, has been further investigated by performing extensive EVB calculations at various temperatures to determine the activations enthalpy (ΔH^\ddagger) and entropy (ΔS^\ddagger) for the EF-Tu catalyzed GTP hydrolysis.^[109] The catalytic effect of EF-Tu was shown to lie in the remarkably large positive entropy term ($T\Delta S^\ddagger$), which is specific to the reaction mechanism used by the translational GTPases.

In this section we have illustrated only a few out of many examples of how the EVB approach has been successfully applied to studying enzyme catalysis. Over the years, the EVB method has proven to be an effective tool for understanding the reaction mechanisms of various biological systems and played a key role in exploring the origins of enzyme catalysis. In particular, the EVB approach has been used to quantify the molecular basis for enzymatic catalysis and to address and verify various proposals for the key factors underlying the enormous catalytic power of enzymes.

As discussed above, the EVB studies have played a key role in demonstrating that enzyme catalysis is driven by electrostatics, that is, that the electrostatic preorganization of the enzyme active site is the origin of the catalytic power of enzymes. At the same time, it has been shown that nuclear quantum effects, particularly electron tunneling, proposed as key factors behind the rate acceleration in enzymes, have no contribution to catalysis, as the same effects appear in reactions taking place in solution. Finally, the EVB studies have proven that there is no link between dynamical effects and enzyme catalysis, inferring that the rate acceleration observed in enzymes is not related to their flexibility and structural dynamics.

Undoubtedly, the EVB method has not only contributed to improving our understanding of the mechanisms of various enzymes, but also, most importantly, this approach has played a key role in building the foundations of our overall understanding of the

factors driving enzyme catalysis at the molecular level. Owing to its powerful features and numerous successful applications, it continues to be used today in exploring enzymatic reaction mechanisms and understanding how enzymes function.

2.7 Other Empirical Valence Bond Models

In the last decades, new variations of the original EVB approach have been developed. Here, we will briefly review the main features of each method (Table 2.2). In brief, the main differences among them are (1) the approach used to evaluate the off-diagonal term H_{ij} , (2) the level of theory at which parametrization is done, and (3) the detail of description of the global shape of the potential energy surface (PES). For a more complete overview of these approaches and their usage, we refer the reader to the original works.

2.7.1 Chang-Miller Formalism

One of the first variants based on the original EVB method^[36] was the EVB approach of Chang and Miller,^[56] where the square of the off-diagonal element, H_{12}^2 , is represented by a generalized Gaussian positioned at or near the transition state (TS):

$$H_{12}^2(\mathbf{r}) = \mathbf{A} \exp\left(\mathbf{B} \cdot \Delta\mathbf{r} - \frac{1}{2} \Delta\mathbf{r} \mathbf{C} \Delta\mathbf{r}\right) \quad (2.21)$$

Here, $\Delta\mathbf{r} = \mathbf{r} - \mathbf{r}_{TS}$, where \mathbf{r}_{TS} is the TS geometry, and \mathbf{A} , \mathbf{B} (a vector), and \mathbf{C} (a matrix) are parameters chosen to yield a desired barrier height, transition state geometry, and transition state Hessian, respectively.^[56] In order to avoid divergence problems caused by negative exponents in \mathbf{C} ^[115,116] and also in order to improve the accuracy of the fit, more elaborate functions of the interaction matrix elements have been developed. They include work by Minichino and Voth,^[117] Truhlar and coworkers,^[114] and Schelegel and Sonnenberg.^[57] These methods, even though elegant, are limited to cases where accurate ab initio calculations are feasible.

2.7.2 Approximate Valence Bond (AVB) Approach

The AVB approach of Grochowski and McCammon^[112,118] is similar to the EVB approach, but it uses ab initio rather than empirical parameterization of the off-diagonal matrix element, as well as a different parameterization strategy. In the particular case of a two-state system, the off-diagonal matrix element is expressed as follows:

$$H_{12} = \sqrt{(\epsilon_1 - E^0)(\epsilon_2 - E^0)} \quad (2.22)$$

where E^0 , which is a function of the atomic positions (and therefore limited to the TS), can directly be determined by fitting its analytical representation to quantum-mechanical and experimental energy data. Like in the EVB approach, H_{12} is assumed to be independent of the environment and such an effect is included into the diagonal terms. The effect of the surroundings is added as a perturbation expansion of the AVB ground state. Additionally, when modeling proton dynamics, the time-dependent Schrödinger equation for the proton motion is integrated in parallel with the classical

Table 2.2 Taxonomy of the EVB and Related Methods. This table is partially reproduced from reference [110].

Method	VB ^[111]	EVB ^[36]	Miller-EVB ^[56]	AVB ^[112]	Extended EVB ^[113]	MS-EVB ^[53]	MC-MM ^[114]
Year	1954	1980	1990	1996	1997	1998	2000
H_{ij} function	Exponential	Exponential	Exponential	Constant	Exponential	General	General
H_{ii} function	Morse	Morse + MM	Morse + MM	Morse + MM	Morse + MM	Morse + MM	MM
Inclusion of Solvent	No	In H_{ii}	No	In H_{ii}	In H_{ii}	In H_{ii} and H_{ij}	In H_{ii}

dynamics for other atoms.^[118] The AVB approach, which was originally incorporated into the Gromos 96 MD package, has been successfully applied in studies of large biomolecular systems.^[112,118] However, for new systems it requires extensive parameterization, which limits its functionality.^[119]

2.7.3 Multistate Empirical Valence Bond (MS-EVB)

The MS-EVB by Schmitt and Voth^[53] was specifically designed to model proton transport in water. The approach extends the two-state EVB method by introducing an *exchange charge distribution* in the off-diagonal terms to mimic transition dipole moments. In contrast to the original EVB approach, the off-diagonal term depends explicitly on the solvent configuration. That is, the coupling between state i and j has the following form:

$$H_{ij} = (V_{const}^{ij} + V_{ext}^{ij}) * A(R_{o-o}, q, w) \quad (2.23)$$

where V_{const}^{ij} is a constant coupling term, V_{ext}^{ij} denotes the electrostatic interactions between the exchange charge distribution and the remaining water molecules, and $A(R_{o-o}, q, w)$ is a product of functions of the distance between the two oxygens in the $H_5O_2^+$ dimer ($R_{o-o} = |\vec{R}_{O1} - \vec{R}_{O2}|$), the distance between the positions of the transferring proton ($q = |\vec{r}_3 - 0.5(\vec{R}_{O2} + \vec{R}_{o2})|$), and w (the O-H-O bond angle within the dimer). This model has provided an excellent representation of bulk transport of protons and it has been extended to numerous biological and chemical systems.^[54,120]

2.7.4 Multiconfiguration Molecular Mechanics (MCM)

The MCM approach of Truhlar and coworkers^[114,121] is almost identical to the EVB approach of Warshel, except for the choice of the off-diagonal term, which is obtained following the Chang-Miller formalism in combination with the Shepard interpolation scheme.^[122] In this case, unlike the Chang-Miller model, H_{12} is being fit not only at the TS, but also at a sequence of reference points, $\mathbf{q}(k)$ (called *Shepard points*), such that the lowest eigenvalue of the multiconfigurational $H_{12}(\mathbf{q})$ matrix reproduces a quadratic expansion around each point.^[114] Once the quadratic expansion of $H_{12}(\mathbf{q})$ is completed for all the Shepard points, $H_{12}(\mathbf{q})$ can be evaluated at other points by means of Shepard

interpolation as a linear combination of the quadratic expansions around these $\mathbf{q}(k)$ points:

$$H_{12}(\mathbf{q}) = \sum_{k=1}^M W(\mathbf{q}) H_{12}^{\text{mod}}(\mathbf{q}, k) \quad (2.24)$$

where $W(\mathbf{q})$ are normalized weights and $H_{12}^{\text{mod}}(\mathbf{q}, k)$ is a modified quadratic function,^[114] as defined in reference.^[114] Once the interpolation is constructed, $H_{12}(\mathbf{q})$ and its derivatives are available analytically at any desired geometry, following Eq. 2.24.

As one can see from the above procedure, the accuracy of an MCMM surface depends on the accuracy of the Shepard data points obtained by QM or QM/MM calculations used as an input. In principle, the inclusion of more Shepard points provides a better description of the PES. However, it comes at a higher computational cost.

2.7.5 Other VB Approaches for Studying Complex Systems

In addition to the above empirical VB approaches, there has been an increasing interest in combining *ab initio* VB calculations with approaches that allow description of the effect of the environment in chemical process. Here we will briefly mention the *ab initio* VB molecular mechanics (VB/MM) methodology,^[47] its variant the density-embedded VB/MM (DE-VB/MM) approach^[123] (which includes electrostatic instead of mechanical embedding), the molecular orbital VB (MOVB),^[124,125] and the VB with effective fragment potential (VBEFP).^[126] For a detailed description of each approach, we refer the reader to the original works.

- Molecular orbital valence bond method (MOVB/MM).** This method is based on the construction of diabatic states using the block localized wavefunction (BLW) approach followed by configuration interaction calculations to obtain the adiabatic potential energy surface.^[124] As in EVB and its extensions, a calibration procedure is adopted. Thus, the off-diagonal matrix elements are adjusted by a scaling parameter or by shifting them to produce the experimental or *ab initio* barrier heights. In addition, an extension of this method called effective Hamiltonian-MOVB (EH-MOVB) also introduced a diabatic coupling scaling factor to uniformly scale the *ab initio* off-diagonal matrix element H_{12} such that the computed energy of activation from the EH-MOVB method is adjusted to be in exact agreement with the target value, either directly from experiment or from high-level *ab initio* calculations. The MOVB approach has been used to study a number of nucleophilic substitution reactions and proton transfer processes both in the gas phase and in aqueous solution.^[124,125]
- Ab Initio Valence Bond Molecular Mechanics (VB/MM):** The VB/MM method, introduced by Shurki *et al.*^[47] is a QM/MM method that combines the *ab initio* VB method with MM. As in the classical EVB approach, the diagonal matrix elements of the Hamiltonian represent the energy of the corresponding i^{th} diabatic state. However, in contrast to EVB, the overlap integral for the isolated part of the reactive system (S_{ij}) is not negligible and therefore has to be considered when calculating the off-diagonal term. In this case, H_{ij} is given by:^[47]

$$H_{ij} = \beta_{ij} + \frac{1}{2}(H_{ii} + H_{jj})S_{ij} \quad H_{ii} = H_{jj} \quad (2.25)$$

where β_{ij} is the reduced resonance integral. Within the VB/MM framework it is assumed that both S_{ij} and β_{ij} are independent of the environment. The DE-VB/MM approach, which considers electrostatic instead of mechanical embedding, has been used to demonstrate that mechanical embedding used in VB/MM can in fact account for most of the effect of the surroundings at a much lower computational cost compared to electrostatic embedding.^[123] The method has been used to study reactions in gas and solvent phases, as well as enzymatic processes.^[47,48]

- **VB with Effective Fragment Potential (VBEFP).** Within the VBEFP approach,^[126] the QM part of system is treated within the ab initio VB scheme (VBSCF) and environment is accounted by the effective fragment potential (EFP) method, which is a polarized force field approach developed by Day *et al.*^[127] The Hamiltonian, therefore, has the following form:

$$H_{VBEFP} = H_{VB} + H_{EFP} \quad (2.26)$$

where H_{VB} is the Hamiltonian of the isolated reactive part and H_{EFP} is the EFP represented by a set of one-electron potentials. The method has been used to study hydrated metal–ligand complexes and $n \rightarrow \pi^*$ vertical excitation of formaldehyde and acetone in aqueous solution.^[126,128]

In this section we have briefly presented some of the methods that use the (E)VB framework to study chemical reactions in complex systems. We note that there are many additional VB- based methods, both ab initio and semiempirical, not discussed in this review. For detailed reviews on these methods we refer the reader to references [32, 35, 129]. Additionally, a list of popular simulation packages for performing (E)VB simulations is provided in Table 2.3. These developments in both methodologies and computational implementations of (E)VB methods demonstrate the explosion of interest in valence bond approaches. VB theory has demonstrated to be a conceptually powerful and particularly efficient tool to describe reactivity in complex systems, including reactions in the condensed phase and bio(molecules). It is expected that, in parallel with the growth in computational power, the developments of E(VB) approaches will continue to increase, providing more accurate and efficient approaches that will allow us to address some of the challenges still present when modeling reactivity in the condensed phase.

2.8 Conclusions and Future Perspectives

Recent decades have seen tremendous advances in computational power and in methods for studying biochemical reactions. Hybrid molecular orbital QM/MM (MO-QM/MM) approaches are the most popular computational tool for studying chemical reactivity in biomolecular systems. However, with the renaissance of valence bond (VB) theory, the use of QM/MM methods in which the QM part is based on VB methods has become increasingly popular in the field.

Table 2.3 Examples of Available Software For Performing (E)VB Calculations as of January 2016.

Program	Capabilities ^{a)}	Website	Comments	Reference
MOLARIS-XG	EVB, FEP, AC	http://laetro.usc.edu/software.html	Available for purchase from the University of Southern California.	Warshel and coworkers ^[130]
Q	EVB, FEP, LIE	http://xray.bmc.uu.se/~aqwww/q/	Free for academic use. Available upon request from the authors.	Åqvist and coworkers ^[131]
V2000	VBSCF, BOVB, VB CI, SCVB, CASVB ^{b)} , GVB	http://www.scinetec.com	Integrated into GAMESS	McWeeny and coworkers ^[132]
TURTLE	VBSCF	http://tc5.chem.uu.nl/ATMOL/turtle/turtle_main.html	Integrated into GAMESS-UK	van Lenthe <i>et al.</i> ^[133]
VM/MM	VB/MM DE-VB/MM	Available upon request avitalsh@ekmd.huji.ac.il	Interface program that communicates between XMVB and MOLARIS	Shurki and coworkers ^[47,48]
MS-EVB	MS-EVB		In house implementation in LAMMPS MD package ^[134]	Voth and coworkers ^[135]
AMBER	Distributed Gaussian EVB	http://ambermd.org/	AMBER license is required for GPU version	Case and coworkers ^[136]
XMVB	VBSCF, BOVB, VB CI, VBPT2, DFVB, VBPCM, VBEP, VBEP/PCM	http://fcc.xmu.edu.cn/xmvp/index.html	Integrated into GAMESS	Wu and coworkers ^[137]

a) VB self-consistent field (VBSCF), breathing orbital VB (BOVB), VB configuration interaction (VB CI), VB perturbation theory (VBPT2), density functional VB (DFVB), VB polarizable continuum model (VBPCM), VB with effective fragment potential (VBEPF), combined VB effective fragment potential polarizable continuum model (VBEPF/PCM), spin-coupled VB (SCVB), complete active space VB (CASVB), generalized VB (GVB), density-embedded VB/MM (DE-VB/MM), empirical VB (EVB), adiabatic charging (AC), free energy perturbation (FEP), linear interaction energy (LIE).

b) CASVB is also integrated in the MOLPRO and MOLCAS packages.

In this chapter, we have built on the concepts presented in the introductory chapter to briefly familiarize the reader with the general theoretical background of VB theory. We have described different aspects of the theory, including its historical development and relationship with MO theory, as well as its basic concepts and general application for the study of chemical reactions. Our focus has been on the empirical incarnation of VB theory, that is, the EVB approach developed by Warshel and coworkers. Through illustrative examples we have demonstrated that EVB is a powerful tool for studying chemical processes both in the condensed phase and in enzymes. Finally, we have discussed some technical aspects that could be considered potential weaknesses of this approach and outline how they are addressed within an EVB framework.

Currently, many research groups are actively developing and applying different approaches based on the EVB philosophy. These recent developments, which included new methodologies and program packages, have greatly expanded the scope of problems that can be studied with EVB, enabling applications from small to large-scale molecular systems. Some of these approaches will be discussed in detail in the following chapters and here we have only provided a brief summary of them.

It is expected that future work may see greater use of methods that combine both *ab initio* and semiempirical VB more intimately to MM approaches. Such developments would allow one to combine the conceptual power of VB with the development of computational approaches for the study of enzymatic reactions.

References

- 1 The Nobel Foundation. (2013) Nobel Prize in Chemistry – Advanced Information. Nobelprize.org.
- 2 Dror, R.O., Dirks, R.M., Grossman, J.P. *et al.* (2012) Biomolecular simulation: A computational microscope for molecular biology. *Annual Review of Biophysics*, **41**, 429–452.
- 3 Senn, H.M. and Thiel, W. (2009) QM/MM methods for biomolecular systems. *Angewandte Chemie International Edition*, **48** (7), 1198–1229.
- 4 Jorgensen, W.L. (2013) Foundations of biomolecular modeling. *Cell*, **155** (6), 1199–1202.
- 5 Hendrickson, J.B. (1961) Molecular geometry. I. Machine computation of the common rings. *Journal of the American Chemical Society*, **83** (22), 4537–4547.
- 6 Wiberg, K.B. (1965) A scheme for strain energy minimization. Application to the cycloalkanes. *Journal of the American Chemical Society*, **87** (5), 1070–1078.
- 7 Allinger, N.L. (1977) Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms. *Journal of the American Chemical Society*, **99** (25), 8127–8134.
- 8 Bixon, M. and Lifson, S. (1967) Potential functions and conformations in cycloalkanes. *Tetrahedron*, **23** (2), 769–784.
- 9 Lifson, S. and Warshel, A. (1968) Consistent force field for calculations of conformations, vibrational spectra, and enthalpies of cycloalkane and n-alkane molecules. *Journal of Chemical Physics*, **49** (11), 5116–5129.

- 10 Jorgensen, W.L. and Tirado-Rives, J. (2005) Potential energy functions for atomic-level simulations of water and organic and biomolecular systems. *Proceedings of the National Academy of Sciences of the United States of America*, **102** (19), 6665–6670.
- 11 Levitt, M. and Lifson, S. (1969) Refinement of protein conformations using a macromolecular energy minimization procedure. *Journal of Molecular Biology*, **46** (2), 269–279.
- 12 (a) Pariser, R. and Parr, R.G. (1953) A semi-empirical theory of the electronic spectra and electronic structure of complex unsaturated molecules. *Journal of Chemical Physics*, **21** (3), 466–471.
(b) Pople, J.A. (1953) Electron interaction in unsaturated hydrocarbons. *Transactions of the Faraday Society*, **49**, 1375–1385.
- 13 Honig, B. and Karplus, M. (1971) Implications of torsional potential of retinal isomers for visual excitation. *Nature*, **229** (5286), 558–560.
- 14 Warshel, A. and Karplus, M. (1972) Calculation of ground and excited-state potential surfaces of conjugated molecules. 1. Formulation and Parametrization. *Journal of the American Chemical Society*, **94** (16), 5612–5625.
- 15 Warshel, A. and Levitt, M. (1976) Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology*, **103** (2), 227–249.
- 16 Singh, U.C. and Kollman, P.A. (1986) A combined ab initio quantum-mechanical and molecular mechanical method for carrying out simulations on complex molecular-systems – applications to the $\text{CH}_3\text{Cl} + \text{Cl}^-$ exchange-reaction and gas-phase protonation of polyethers. *Journal of Computational Chemistry*, **7** (6), 718–730.
- 17 Field, M.J., Bash, P.A. and Karplus, M. (1990) A combined quantum-mechanical and molecular mechanical potential for molecular-dynamics simulations. *Journal of Computational Chemistry*, **11** (6), 700–733.
- 18 Gao, J.L. and Xia, X.F. (1992) A priori evaluation of aqueous polarization effects through Monte-Carlo QM-MM simulations. *Science*, **258** (5082), 631–635.
- 19 Shaik, S. and Hiberty, P.C. (2004) Valence bond theory, its history, fundamentals, and applications: A primer. *Reviews in Computational Chemistry*, **20**, 1–100.
- 20 Heitler, W. and London, F. (1927) Wechselwirkung neutraler atome und homöopolare bindung nach der quantenmechanik. *Zeitschrift für Physik*, **44**, 455–472.
- 21 Marcus, R.A. (1964) Chemical and electrochemical electron-transfer theory. *Annual Review of Physical Chemistry*, **15**, 155–196.
- 22 Marcus, R.A. (1993) Electron-transfer reactions in chemistry – theory and experiment (Nobel lecture). *Angewandte Chemie International Edition*, **32** (8), 1111–1222.
- 23 Mulliken, R.S. (1928) The assignment of quantum numbers for electrons in molecules I. *Physical Review*, **32** (2), 0186–0222.
- 24 Hund, F. (1931) Zur Frage der chemischen Bindung. *Zeitschrift für Physik*, **73**, 1.
- 25 Mülliken, R.S. (1967) Spectroscopy molecular orbitals and chemical bonding. *Science*, **157** (3784), 13–24.
- 26 Hückel, E. (1931) Quantum contributions to the benzene problem. *Zeitschrift für Physik*, **70** (3-4), 204–286.

- 27 Shaik, S.S. and Hiberty, P.C. (2008) *A chemist's guide to valence bond theory*, John Wiley & Sons, New Jersey.
- 28 Lennard-Jones, J.E. (1929) The electronic structure of some diatomic molecules. *Transactions of the Faraday Society*, **25**, 0668–0685.
- 29 Brush, S.G. (1999) Dynamics of theory change in chemistry: Part 1. The benzene problem 1865-1945. *Studies in History and Philosophy of Science Part A*, **30** (1), 21–79.
- 30 Brush, S.G. (1999) Dynamics of theory change in chemistry: Part 2. Benzene and molecular orbitals, 1945-1980. *Studies in History and Philosophy of Science Part A*, **30** (2), 263–302.
- 31 Parr, R.G., Craig, D.P. and Ross, I.G. (1950) Molecular orbital calculations of the lower excited electronic levels of benzene, configuration interaction included. *Journal of Chemical Physics*, **18** (12), 1561–1563.
- 32 Wu, W., Su, P., Shaik, S. and Hiberty, P.C. (2011) Classical valence bond approach by modern methods. *Chemical Reviews*, **111** (11), 7557–7593.
- 33 Braida, B., Derat, E., Humbel, S. *et al.* (2012) The valence bond workshop in Paris: The phoenix rises from the ashes or, has a love story with MO-based theories begun? *ChemPhysChem*, **13** (18), 4029–4030.
- 34 Usharani, D., Lai, W.Z., Li, C.S. *et al.* (2014) A tutorial for understanding chemical reactivity through the valence bond approach. *Chemical Society Reviews*, **43** (14), 4968–4988.
- 35 Shurki, A., Derat, E., Barrozo, A. and Kamerlin, S.C.L. (2015) How valence bond theory can help you understand your (bio)chemical reaction. *Chemical Society Reviews*, **44** (5), 1037–1052.
- 36 Warshel, A. and Weiss, R.M. (1980) An empirical valence bond approach for comparing reactions in solutions and in enzymes. *Journal of the American Chemical Society*, **102** (20), 6218–6226.
- 37 Hiberty, P.C. and Shaik, S.S. (2014) Bridging Cultures, in *The chemical bond: Chemical bonding across the periodic table* (eds G. Frenking and S.S. Shaik), WILEY.
- 38 Van Vleck, J.H. and Sherman, A. (1935) The quantum theory of valence. *Reviews of Modern Physics*, **7** (3), 167–228.
- 39 Vanlenthe, J.H. and Balintkurti, G.G. (1980) The valence-bond SCF (VBSCF) method – synopsis of theory and test calculation of OH potential-energy curve. *Chemical Physics Letters*, **76** (1), 138–142.
- 40 Wu, W., Song, L.C., Cao, Z.X. *et al.* (2002) Valence bond configuration interaction: A practical ab initio valence bond method that incorporates dynamic correlation. *Journal of Physical Chemistry A*, **106** (11), 2721–2726.
- 41 Hiberty, P.C., Flament, J.P. and Noizet, E. (1992) Compact and accurate valence bond functions with different orbitals for different configurations – application to the 2-configuration description of F₂. *Chemical Physics Letters*, **189** (3), 259–265.
- 42 Goddard, W.A., Dunning, T.H., Hunt, W.J. and Hay, P.J. (1973) Generalized valence bond description of bonding in low-lying states of molecules. *Accounts of Chemical Research*, **6** (11), 368–376.
- 43 Cooper, D.L., Gerratt, J. and Raimondi, M. (1991) Applications of spin-coupled valence bond theory. *Chemical Reviews*, **91** (5), 929–964.

- 44 Bernardi, F., Olivucci, M., Ragazos, I.N. and Robb, M.A. (1992) A new mechanistic scenario for the photochemical transformation of ergosterol: An MC-SCF and MM-VB study. *Journal of the American Chemical Society*, **114** (21), 8211–8220.
- 45 Song, L.C., Wu, W., Zhang, Q.N. and Shaik, S. (2004) VBPCM: A valence bond method that incorporates a polarizable continuum model. *Journal of Physical Chemistry A*, **108** (28), 6017–6024.
- 46 Su, P., Wu, W., Kelly, C.P. *et al.* (2008) VBSM: A solvation model based on valence bond theory. *Journal of Physical Chemistry A*, **112** (50), 12761–12768.
- 47 Shurki, A. and Crown, H.A. (2005) Hybrid ab initio VB/MM method – a valence bond ride through classical landscapes. *Journal of Physical Chemistry B*, **109** (49), 23638–23644.
- 48 Sharir-Ivry, A., Shnerb, T., Strajbl, M. and Shurki, A. (2010) VB/MM protein landscapes: A study of the S_N2 reaction in haloalkane dehalogenase. *Journal of Physical Chemistry B*, **114** (6), 2212–2218.
- 49 Kamerlin, S.C.L. and Warshel, A. (2011) The empirical valence bond model: Theory and applications. *WIREs Computational Molecular Science*, **1** (1), 30–45.
- 50 Hwang, J.K., King, G., Creighton, S. and Warshel, A. (1988) Simulation of free-energy relationships and dynamics of S_N2 reactions in aqueous-solution. *Journal of the American Chemical Society*, **110** (16), 5297–5311.
- 51 Warshel, A. (1991) *Computer modeling of chemical reactions in enzymes and solutions*, Wiley, New York.
- 52 Warshel, A. and Russell, S. (1986) Theoretical correlation of structure and energetics in the catalytic reaction of trypsin. *Journal of the American Chemical Society*, **108** (21), 6569–6579.
- 53 Schmitt, U.W. and Voth, G.A. (1998) Multistate empirical valence bond model for proton transport in water. *Journal of Physical Chemistry B*, **102** (29), 5547–5551.
- 54 Maupin, C.M., Wong, K.F., Soudackov, A.V. *et al.* (2006) A multistate empirical valence bond description of protonatable amino acids. *Journal of Physical Chemistry A*, **110** (2), 631–639.
- 55 Coulson, C.A. and Danielsson, U. (1954) Ionic and covalent contributions to the hydrogen bond. Part 1. *Arkiv för Fysik*, **8** (3), 239–244.
- 56 Chang, Y.T. and Miller, W.H. (1990) An empirical valence bond model for constructing global potential-energy surfaces for chemical-reactions of polyatomic molecular-systems. *Journal of Physical Chemistry*, **94** (15), 5884–5888.
- 57 Schlegel, H.B. and Sonnenberg, J.L. (2006) Empirical valence-bond models for reactive potential energy surfaces using distributed Gaussians. *Journal of Chemical Theory and Computation*, **2** (4), 905–911.
- 58 Åqvist, J. and Warshel, A. (1993) Simulation of enzyme-reactions using valence-bond force-fields and other hybrid quantum-classical approaches. *Chemical Reviews*, **93** (7), 2523–2544.
- 59 Hong, G.Y., Rosta, E. and Warshel, A. (2006) Using the constrained DFT approach in generating diabatic surfaces and off diagonal empirical valence bond terms for modeling reactions in condensed phases. *Journal of Physical Chemistry B*, **110** (39), 19570–19574.

- 60 Torrie, G.M. and Valleau, J.P. (1974) Monte-Carlo free-energy estimates using non-Boltzmann sampling – application to subcritical Lennard-Jones fluid. *Chemical Physics Letters*, **28** (4), 578–581.
- 61 Zwanzig, R.W. (1954) High-temperature equation of state by a perturbation method. 1. Nonpolar gases. *Journal of Chemical Physics*, **22** (8), 1420–1426.
- 62 Chipot, C. and Pohorille, A. (2007) *Free energy calculations: Theory and applications in chemistry and biology*, Springer, Berlin; New York.
- 63 Warshel, A. (1982) Dynamics of reactions in polar-solvents – semi-classical trajectory studies of electron-transfer and proton-transfer reactions. *Journal of Physical Chemistry*, **86** (12), 2218–2224.
- 64 Mones, L., Kulhanek, P., Simon, I. *et al.* (2009) The energy gap as a universal reaction coordinate for the simulation of chemical reactions. *Journal of Physical Chemistry B*, **113** (22), 7867–7873.
- 65 Warshel, A., Hwang, J.K. and Åqvist, J. (1992) Computer-simulations of enzymatic-reactions – examination of linear free-energy relationships and quantum-mechanical corrections in the initial proton-transfer step of carbonic-anhydrase. *Faraday Discussions*, **93**, 225–238.
- 66 Warshel, A., Sharma, P.K., Kato, M. *et al.* (2006) Electrostatic basis for enzyme catalysis. *Chemical Reviews*, **106** (8), 3210–3235.
- 67 Bjelic, S. and Åqvist, J. (2006) Catalysis and linear free energy relationships in aspartic proteases. *Biochemistry*, **45** (25), 7709–7723.
- 68 Mones, L., Tang, W.-J. and Florián, J. (2013) Empirical valence bond simulations of the chemical mechanism of ATP to cAMP conversion by anthrax edema factor. *Biochemistry*, **52** (15), 2672–2682.
- 69 Rychkova, A., Mukherjee, S., Bora, R.P. and Warshel, A. (2013) Simulating the pulling of stalled elongated peptide from the ribosome by the translocon. *Proceedings of the National Academy of Sciences of the United States of America*, **110** (25), 10195–1020.
- 70 R. Prasad, B., Plotnikov, N.V., Lameira, J. and Warshel, A. (2013) Quantitative exploration of the molecular origin of the activation of GTPase. *Proceedings of the National Academy of Sciences of the United States of America*, **110** (51), 20509–20514.
- 71 Valero, R., Song, L.C., Gao, J.L. and Truhlar, D.G. (2009) Perspective on diabatic models of chemical reactivity as illustrated by the gas-phase S_N2 reaction of acetate ion with 1,2-dichloroethane. *Journal of Chemical Theory and Computation*, **5** (1), 1–22.
- 72 Olsson, M.H.M. and Warshel, A. (2004) Solute solvent dynamics and energetics in enzyme catalysis: The S_N2 reaction of dehalogenase as a general benchmark. *Journal of the American Chemical Society*, **126** (46), 15167–15179.
- 73 Rosta, E., Klahn, M. and Warshel, A. (2006) Towards accurate ab initio QM/MM calculations of free-energy profiles of enzymatic reactions. *Journal of Physical Chemistry B*, **110** (6), 2934–2941.

- 74 Mones, L. and Csanyi, G. (2012) Topologically invariant reaction coordinates for simulating multistate chemical reactions. *Journal of Physical Chemistry B*, **116** (51), 14876–14885.
- 75 Olsson, M.H.M., Parson, W.W. and Warshel, A. (2006) Dynamical contributions to enzyme catalysis: Critical tests of a popular hypothesis. *Chemical Reviews*, **106** (5), 1737–1756.
- 76 Feierberg, I., Luzhkov, V. and Åqvist, J. (2000) Computer simulation of primary kinetic isotope effects in the proposed rate-limiting step of the glyoxalase I catalyzed reaction. *Journal of Biological Chemistry*, **275** (30), 22657–22662.
- 77 Hatcher, E., Soudackov, A.V. and Hammes-Schiffer, S. (2004) Proton-coupled electron transfer in soybean lipoxygenase. *Journal of the American Chemical Society*, **126** (18), 5763–5775.
- 78 Blake, C.C.F., Johnson, L.N., Mair, G.A. *et al.* (1967) Crystallographic studies of activity of hen egg-white lysozyme. *Proc. R. Soc. London, Ser. B*, **167** (1009), 378–388.
- 79 Fried, S.D., Bagchi, S. and Boxer, S.G. (2014) Extreme electric fields power catalysis in the active site of ketosteroid isomerase. *Science*, **346** (6216), 1510–1514.
- 80 Åqvist, J. and Warshel, A. (1989) Calculations of free-energy profiles for the staphylococcal nuclease catalyzed reaction. *Biochemistry*, **28** (11), 4680–4689.
- 81 Åqvist, J. and Warshel, A. (1990) Free-energy relationships in metalloenzyme-catalyzed reactions – calculations of the effects of metal-ion substitutions in staphylococcal nuclease. *Journal of the American Chemical Society*, **112** (8), 2860–2868.
- 82 Kamerlin, S.C.L. and Warshel, A. (2010) An analysis of all the relevant facts and arguments indicates that enzyme catalysis does not involve large contributions from nuclear tunneling. *Journal of Physical Organic Chemistry*, **23** (7), 677–684.
- 83 Kohen, A. and Klinman, J.P. (1999) Hydrogen tunneling in biology. *Chemistry & Biology*, **6** (7), R191–R198.
- 84 Hwang, J.K., Chu, Z.T., Yadav, A. and Warshel, A. (1991) Simulations of quantum-mechanical corrections for rate constants of hydride-transfer reactions in enzymes and solutions. *Journal of Physical Chemistry*, **95** (22), 8445–8448.
- 85 Hwang, J.K. and Warshel, A. (1993) A quantized classical path approach for calculations of quantum-mechanical rate constants. *Journal of Physical Chemistry*, **97** (39), 10053–10058.
- 86 Billeter, S.R., Webb, S.P., Agarwal, P.K. *et al.* (2001) Hydride transfer in liver alcohol dehydrogenase: quantum dynamics, kinetic isotope effects, and role of enzyme motion. *Journal of the American Chemical Society*, **123** (45), 11262–11272.
- 87 Gao, J.L. and Truhlar, D.G. (2002) Quantum mechanical methods for enzyme kinetics. *Annual Review of Physical Chemistry*, **53**, 467–505.

- 88 Epstein, D.M., Benkovic, S.J. and Wright, P.E. (1995) Dynamics of the dihydrofolate-reductase folate complex – catalytic sites and regions known to undergo conformational change exhibit diverse dynamical features. *Biochemistry*, **34** (35), 11037–11048.
- 89 Schnell, J.R., Dyson, H.J. and Wright, P.E. (2004) Structure, dynamics, and catalytic function of dihydrofolate reductase. *Annual Review of Biophysics and Biomolecular Structure*, **33**, 119–140.
- 90 Bhabha, G., Lee, J., Ekiert, D.C. *et al.* (2011) A dynamic knockout reveals that conformational fluctuations influence the chemical step of enzyme catalysis. *Science*, **332** (6026), 234–238.
- 91 Levitt, M. and Warshel, A. (1975) Computer-simulation of protein folding. *Nature*, **253** (5494), 694–698.
- 92 Adamczyk, A.J., Cao, J., Kamerlin, S.C.L. and Warshel, A. (2011) Catalysis by dihydrofolate reductase and other enzymes arises from electrostatic preorganization, not conformational motions. *Proceedings of the National Academy of Sciences of the United States of America*, **108** (34), 14115–14120.
- 93 Pislakov, A.V., Cao, J., Kamerlin, S.C.L. and Warshel, A. (2009) Enzyme millisecond conformational dynamics do not catalyze the chemical step. *Proceedings of the National Academy of Sciences of the United States of America*, **106** (41), 17359–17364.
- 94 Reetz, M.T. (2011) Laboratory evolution of stereoselective enzymes: A prolific source of catalysts for asymmetric reactions. *Angewandte Chemie International Edition*, **50** (1), 138–174.
- 95 Frushicheva, M.P. and Warshel, A. (2012) Towards quantitative computer-aided studies of enzymatic enantioselectivity: The case of *Candida antarctica* Lipase A. *ChemBioChem*, **13** (2), 215–223.
- 96 Barrozo, A., Borstnar, R., Marloie, G. and Kamerlin, S.C.L. (2012) Computational protein engineering: Bridging the gap between rational design and laboratory evolution. *International Journal of Molecular Sciences*, **13** (10), 12428–12460.
- 97 Schopf, P. and Warshel, A. (2014) Validating computer simulations of enantioselective catalysis; reproducing the large steric and entropic contributions in *Candida antarctica* Lipase B. *Proteins*, **82** (7), 1387–1399.
- 98 Strajbl, M., Sham, Y.Y., Villà, J. *et al.* (2000) Calculations of activation entropies of chemical reactions in solution. *Journal of Physical Chemistry B*, **104** (18), 4578–4584.
- 99 Amrein, B.A., Bauer, P., Duarte, F. *et al.* (2015) Expanding the catalytic triad in epoxide hydrolases and related enzymes. *ACS Catalysis*, **5** (10), 5702–5713.
- 100 Trobro, S. and Aqvist, J. (2006) Analysis of predictions for the catalytic mechanism of ribosomal peptidyl transfer. *Biochemistry*, **45** (23), 7049–7056.
- 101 Trobro, S. and Aqvist, J. (2005) Mechanism of peptide bond synthesis on the ribosome. *Proceedings of the National Academy of Sciences of the United States of America*, **102** (35), 12395–12400.
- 102 Sharma, P.K., Xiang, Y., Kato, M. and Warshel, A. (2005) What are the roles of substrate-assisted catalysis and proximity effects in peptide bond formation by the ribosome? *Biochemistry*, **44** (34), 11307–11314.

- 103 Adamczyk, A.J. and Warshel, A. (2011) Converting structural information into an allosteric-energy-based picture for elongation factor Tu activation by the ribosome. *Proceedings of the National Academy of Sciences of the United States of America*, **108** (24), 9827–9832.
- 104 Nissen, P., Hansen, J., Ban, N. *et al.* (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science*, **289** (5481), 920–930.
- 105 Muth, G.W., Ortoleva-Donnelly, L. and Strobel, S.A. (2000) A single adenosine with a neutral pK_a in the ribosomal peptidyl transferase center. *Science*, **289** (5481), 947–950.
- 106 Moore, P.B. and Steitz, T.A. (2003) After the ribosome structures: How does peptidyl transferase work? *RNA*, **9** (2), 155–159.
- 107 Hansen, J.L., Schmeing, T.M., Moore, P.B. and Steitz, T.A. (2002) Structural insights into peptide bond formation. *Proceedings of the National Academy of Sciences of the United States of America*, **99** (18), 11670–11675.
- 108 Aqvist, J. and Kamerlin, S.C.L. (2015) The conformation of a catalytic loop is central to GTPase activity on the ribosome. *Biochemistry*, **54** (2), 546–556.
- 109 Åqvist, J. and Kamerlin, S.C.L. (2015) Exceptionally large entropy contributions enable the high rates of GTP hydrolysis on the ribosome. *Scientific Reports*, **5**, 15817.
- 110 Florian, J. (2002) Comment on molecular mechanics for chemical reactions. *Journal of Physical Chemistry A*, **106** (19), 5046–5047.
- 111 Coulson, C.A. and Danielsson, U. (1954) Ionic and covalent contributions to the hydrogen bond. Part 2. *Arkiv för Fysik*, **8** (3), 245–255.
- 112 Grochowski, P., Lesyng, B., Bala, P. and McCammon, J.A. (1996) Density functional based parametrization of a valence bond method and its applications in quantum classical molecular dynamics simulations of enzymatic reactions. *The International Journal of Quantum Chemistry*, **60** (6), 1143–1164.
- 113 Vuilleumier, R. and Borgis, D. (1997) Molecular dynamics of an excess proton in water using a non-additive valence bond force field. *Journal of Molecular Structure*, **437**, 555–565.
- 114 Kim, Y., Corchado, J.C., Villa, J. *et al.* (2000) Multiconfiguration molecular mechanics algorithm for potential energy surfaces of chemical reactions. *Journal of Chemical Physics*, **112** (6), 2718–2735.
- 115 Chang, Y.T., Minichino, C. and Miller, W.H. (1992) Classical trajectory studies of the molecular dissociation dynamics of formaldehyde – $H_2CO \rightarrow H_2 + CO$. *Journal of Chemical Physics*, **96** (6), 4341–4355.
- 116 Anglada, J.M., Besalu, E., Bofill, J.M. and Crehuet, R. (1999) Prediction of approximate transition states by Bell-Evans-Polanyi principle: I. *Journal of Computational Chemistry*, **20** (11), 1112–1129.
- 117 Minichino, C. and Voth, G.A. (1997) Potential energy surfaces for chemical reactions: An analytical representation from coarse grained data with an application to proton transfer in water. *Journal of Physical Chemistry B*, **101** (23), 4544–4552.
- 118 Bala, P., Grochowski, P., Nowinski, K. *et al.* (2000) Quantum-dynamical picture of a multistep enzymatic process: Reaction catalyzed by phospholipase A(2). *Biophysical Journal*, **79** (3), 1253–1262.

- 119 Walewski, L., Bala, P., Elstner, M. *et al.* (2004) Fast QM/MM method and its application to molecular systems. *Chemical Physics Letters*, **397** (4-6), 451–458.
- 120 Sumner, I. and Voth, G.A. (2012) Proton transport pathways in [NiFe]-hydrogenase. *Journal of Physical Chemistry B*, **116** (9), 2917–2926.
- 121 Lin, H., Zhao, Y., Tishchenko, O. and Truhlar, D.G. (2006) Multiconfiguration molecular mechanics based on combined quantum mechanical and molecular mechanical calculations. *Journal of Chemical Theory and Computation*, **2** (5), 1237–1254.
- 122 Ischtwan, J. and Collins, M.A. (1994) Molecular-potential energy surfaces by interpolation. *Journal of Chemical Physics*, **100** (11), 8080–8088.
- 123 Sharir-Ivry, A., Crown, H.A., Wu, W. and Shurki, A. (2008) Density embedded VB/MM: A hybrid ab initio VB/MM with electrostatic embedding. *Journal of Physical Chemistry A*, **112** (11), 2489–2496.
- 124 Mo, Y.R. and Gao, J.L. (2000) Ab initio QM/MM simulations with a molecular orbital-valence bond (MOVB) method: Application to an S_N2 reaction in water. *Journal of Computational Chemistry*, **21** (16), 1458–1469.
- 125 Mo, Y.R. and Gao, J.L. (2000) An ab initio molecular orbital-valence bond (MOVB) method for simulating chemical reactions in solution. *Journal of Physical Chemistry A*, **104** (13), 3012–3020.
- 126 Ying, F.M., Chang, X., Su, P.F. and Wu, W. (2012) VBEFP: A valence bond approach that incorporates effective fragment potential method. *Journal of Physical Chemistry A*, **116** (7), 1846–1853.
- 127 Day, P.N., Jensen, J.H., Gordon, M.S. *et al.* (1996) An effective fragment method for modeling solvent effects in quantum mechanical calculations. *Journal of Chemical Physics*, **105** (5), 1968–1986.
- 128 Huang, J., Ying, F.M., Su, P.F. and Wu, W. (2014) VBEFP/PCM: A QM/MM/PCM approach for valence-bond method and its application for the vertical excitations of formaldehyde and acetone in aqueous solution. *Science China Chemistry*, **57** (10), 1409–1417.
- 129 Shurki, A. and Sharir-Ivry, A. (2014) Valence bond-based hybrid quantum mechanics molecular mechanics approaches and proper inclusion of the effect of the surroundings. *Israel Journal of Chemistry*, **54** (8-9), 1189–1204.
- 130 Warshel, A. MOLARIS-XG: Version 9.13 http://laetro.usc.edu/doc/theory_molaris_9.11.pdf.
- 131 Marelus, J., Kolmodin, K., Feierberg, I. and Åqvist, J. (1998) Q: A molecular dynamics program for free energy calculations and empirical valence bond simulations in biomolecular systems. *Journal Molecular Graphics Modelling*, **16** (4-6), 213–225.
- 132 Li, J.B. and McWeeny, R. (2002) VB2000: Pushing valence bond theory to new limits. *The International Journal of Quantum Chemistry*, **89** (4), 208–216.
- 133 van Lenthe, J.H., Dijkstra, F. and Havenith, W.A. (2002) TURTLE – a Gradient VBSCF Program Theory and Studies of Aromaticity, in *Valence bond theory* (ed. D.L. Cooper), Elsevier, Amsterdam, The Netherlands.
- 134 Plimpton, S. (1995) Fast parallel algorithms for short-range molecular-dynamics. *Journal of Computational Physics*, **117** (1), 1–19.

- 135 Yamashita, T., Peng, Y.X., Knight, C. and Voth, G.A. (2012) Computationally efficient multiconfigurational reactive molecular dynamics. *Journal of Chemical Theory and Computation*, **8** (12), 4863–4875.
- 136 Case, D.A., Betz, R.M., Botello -Smith, W. *et al.* (2016) *AMBER 2016*, University of California, San Francisco.
- 137 Chen, Z., Ying, F., Chen, X. *et al.* (2014) XMVB 2.0: A new version of Xiamen valence bond program. *The International Journal of Quantum Chemistry*, **115**, 731–737.

3

Using Empirical Valence Bond Constructs as Reference Potentials For High-Level Quantum Mechanical Calculations

Nikolay V. Plotnikov

Department of Chemistry, Stanford University, United States

This chapter describes the paradynamics (PD) approach – a computational algorithm for evaluating the free energy changes in chemical processes at a target Quantum Mechanical (QM) level using sampling with a simplified reference potential. This method aims to reduce the high computational cost of the free energy calculations arising from an extensive sampling with the target potential. While reduction of the computational cost often means reduction in accuracy, the PD approach avoids the loss of accuracy while computing activation barriers. Here, the PD approach is formulated to closely reproduce activation free energies obtained from the Potential of Mean Force (PMF) which involves extensive sampling with the target QM potential. In fact, the most recent formulation of the PD approach involves computing the PMF (also known as the free energy surface) locally, at the Reactant State (RS) and at the Transition State (TS), and positioning these regions relative to the reference free energy shifts. These reference free energy levels are obtained with the direct PMF approach from sampling with a reference potential. It is, therefore, imperative to understand how the PMF is computed before proceeding to the understanding of the PD model.

The goal of this chapter is to introduce the reader to the key concepts of the PD approach: use of free energy perturbation and of thermodynamic cycles methods, refinement of the reference potential to mimic behavior of the target QM potential for a system of interest (to improve convergence of the free energy perturbation), and to identify location of the reactants and of the transition state on the target low-accuracy free energy surface. A secondary objective is to show how the PD approach is related to and derived from more general approaches of statistical thermodynamics, such as free energy estimation algorithms, the umbrella sampling approach and self-consistent force-fields. The aim of this chapter is to provide a good starting point for researchers interested in application or adaptation of this and related algorithms in order to reduce computational time for screening libraries of compounds with physics-based models or for validating proposals of rational design. The chapter assumes basic knowledge of statistical thermodynamics and molecular modeling techniques. Some of the key concepts are reviewed in the text with appropriate references to dedicated books given alongside.

3.1 Context

Equilibrium and rate constants for chemical processes are determined by the corresponding free energy changes.^[1–5] Computational models can substitute more expensive experiments if they can predict these changes in a quantitative and timely fashion. Particularly promising are *ab initio* models based on the first principles of physics, since they deliver chemical accuracy for gas-phase experiments.^[6] However, application of these models to the condensed phase becomes very computationally expensive, because of the steeply increasing cost of each energy evaluation with the system size and due to an extensive sampling, which is required to capture energetically relevant solvent fluctuations. These methods, nevertheless, proved to be extremely insightful in elucidating multiple fundamental problems of chemical biology (as covered in the reviews cited above and by others^[7,8]), and can greatly facilitate solving practical problems in life sciences, such as the design of enzyme inhibitors^[9,10] and artificial enzymes.^[11,12] Still, the high cost prevents routine application of *ab initio* quantum mechanical and hybrid quantum mechanical/molecular mechanical (QM/MM) free energy methods to a broad spectrum of problems in chemical biology, particularly with high-level electronic structure methods.

Chemical processes of biological significance take place in complex polar environments and involve interactions between thousands of atoms, due to the long-range nature of electrostatic forces.^[13] However, in many processes of interest, such as the solvation of small molecules, enzyme-substrate binding, photo-excitations, and chemical reactions in condensed phases, one can often identify a reaction center, or a solute – a group of atoms, where the most significant changes in electronic density take place. This allows the use of QM/MM^[14] potential energy surfaces for computational modeling. In this hybrid multi-level formalism, the solute is described by QM methods, while the solvent (whether it is water, protein or nucleic acid chains) is described using empirically parameterized MM force fields. Use of the QM/MM description of the potential energy surface significantly reduces the computational cost (compared to a QM description of the full system). However, in any case, describing interactions between electrons in the QM (sub) system remains by far the most computationally expensive part of the simulations.^[15]

A statistical thermodynamic definition of the free energy of an electronic state is given^[16] through the partition function, Q , which describes interactions between N particles with a model potential energy surface, E :

$$Q(N, V, T) \equiv \frac{1}{\Lambda^{3N} N!} \int d\mathbf{r}^N \exp[-\beta E(\mathbf{r}^N)] \quad (3.1)$$

$$F = -\beta^{-1} \ln Q(N, V, T) \quad (3.2)$$

Here, $\beta^{-1} = k_B T$ (k_B is the Boltzmann constant and T is the absolute temperature); $\Lambda = \sqrt{\frac{\beta \hbar^2}{2\pi m}}$ is the thermal de Broglie wavelength; \mathbf{r} is the coordinate vector; and N is the number of particles. Consequently, different models for the potential energy surface will correspond to different free energy values. The quality of the potential, in turn, determines the quality of the theoretical (and quantitative) description of transitions between the final states, or to the transition state of a chemical process. In practice, the use of *ab initio* QM potentials involves a higher computational cost, which can easily become

computationally prohibitive for an adequate configurational sampling. Yet, these potentials are needed to adequately describe changes in electronic density of species involved in chemical processes and to reach chemical accuracy.^[6] However, one can represent a high-level target potential (E_{TGT}) as a low-level reference potential (E_{REF}) with an added perturbation:

$$E_{TGT} = E_{REF} + (E_{TGT} - E_{REF}) \quad (3.3)$$

Altering the force-law in Eq. (3.1) causes changes in the free energy. This computational technique of switching the model potential is called free energy perturbation (FEP).^[17,18]

The overall computational cost of calculating QM/MM free energy changes is further increased by: 1) The extensive sampling needed to capture energetically relevant fluctuations in combined solvent-solute degrees of freedom; and 2) the presence of high barriers in the energy landscape, which, in turn, requires long simulation times or enhanced sampling algorithms. To minimize the high computational cost of ab initio QM/MM free energy modeling and related tasks, Warshel and coworkers have developed a method,^[19-25] which is now called the Paradynamics (PD) approach.^[24] In this approach, the target QM free energy change is computed perturbatively using a thermodynamic cycle, which is schematically illustrated below (Figure 3.1).

Configurational sampling in the PD approach is predominantly performed with a computationally *cheap* reference potential, which allows for significant reduction in the amount of computer time, which is required to calculate the activation free energy barrier at a target QM level, compared to when the sampling is performed with the target potential directly. A limited amount of sampling with E_{TGT} is performed at RS and at TS, while evaluating the free energy perturbation of moving from the reference to the target free energy surface (Figure 3.2). In other words, this method minimizes the number of configurations evaluated with expensive electronic structure methods while exploring the reaction free energy surface. The computational gain achieved with this algorithm depends on the cost difference between the reference and the target potential. Not surprisingly, the empirical valence bond^[26] (EVB) potential has been one

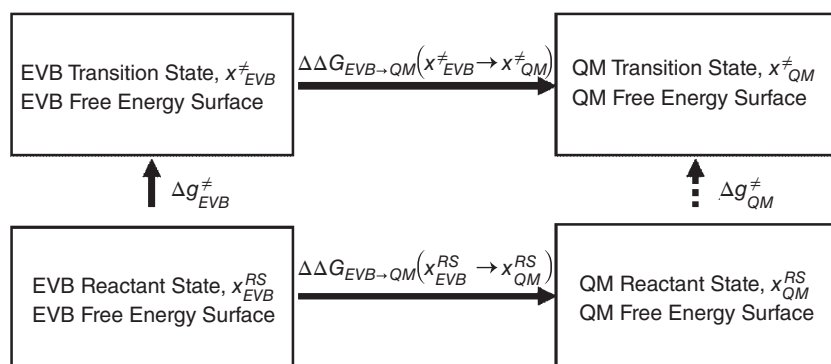


Figure 3.1 The Thermodynamic Cycle Used in Calculating the Ab Initio QM/MM Activation Free Energy. The QM/MM activation free energy (denoted here as QM and shown by a vertical dashed arrow) is calculated as the sum of the EVB barrier (vertical solid arrow) evaluated by the EVB FEP/US method and the perturbations between the QM and EVB surfaces at the RS and TS (horizontal solid arrows), which are evaluated using the LRA approach. Reprinted with permission from reference [24]. Copyright 2011 American Chemical Society.

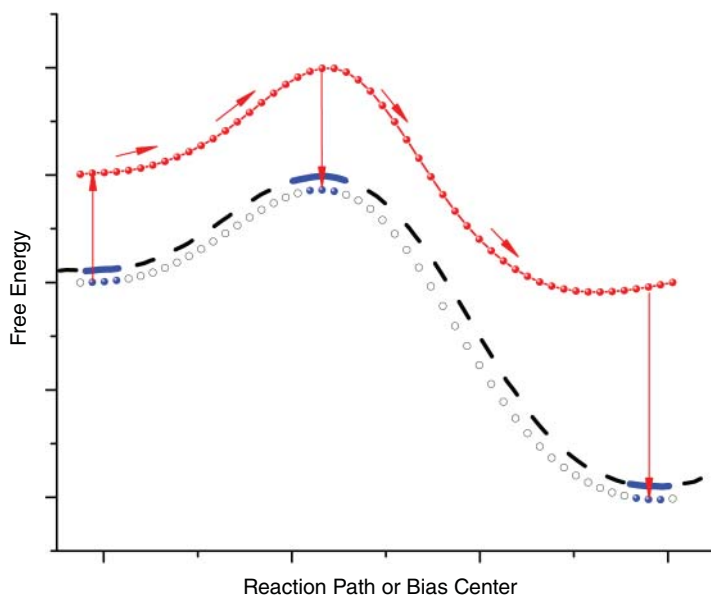


Figure 3.2 Thermodynamic Cycle Used in Calculation of the Activation Free Energy by Positioning Regions of the Target Free Energy Surface (Blue Solid Line) at the Final States and at the Transition State Relative to the Reference Free Energy Shifts (Red Dots). Black empty dots show the target free energy shifts which are computed with a potential of mean force (PMF); the complete target free energy profile (from the PMF) is shown with a dashed black line. The red arrows show the thermodynamic detour that is an alternative to computing the full target PMF. Note that red vertical arrows correspond to the free energy changes of moving between the reference and the target potentials with the same bias. Adapted from reference [35]. (See color plate section for the color representation of this figure.)

of the most practical choices for the reference potential, since its computational cost is only (roughly) n times higher of that for the classical force field, where n is the number of EVB states.

While one could assume that the reaction path is the same on both target and reference free energy surfaces, this is not necessarily correct, as the two paths can also be substantially different, in which case even convergent free energy perturbations lead to incorrect estimates of the activation energy (Figure 3.3). The main approach to address this problem involves constructing a reasonable approximation for the target free energy surface to locate the minimum free energy path of interest. The PD method incorporates a number of computational techniques to achieve that and to improve the convergence of the perturbative treatment. For example, self-consistent refinement^[27] of the EVB reference potential is used to accelerate convergence of the free energy perturbation while moving from the reference EVB to the target free energy surface; this refinement is combined with the linear response approximation^[28] (LRA) to efficiently evaluate this perturbation. The PD model and other related reference potential strategies, such as those presented in references^[29–33] were recently formulated in terms of the umbrella sampling^[34] technique in order to established the connection between the PMF and the PD approach, as well as to reach the accuracy of the PMF approach with the PD model in a systematic way. Since the PMF approach is a traditional method for calculating the

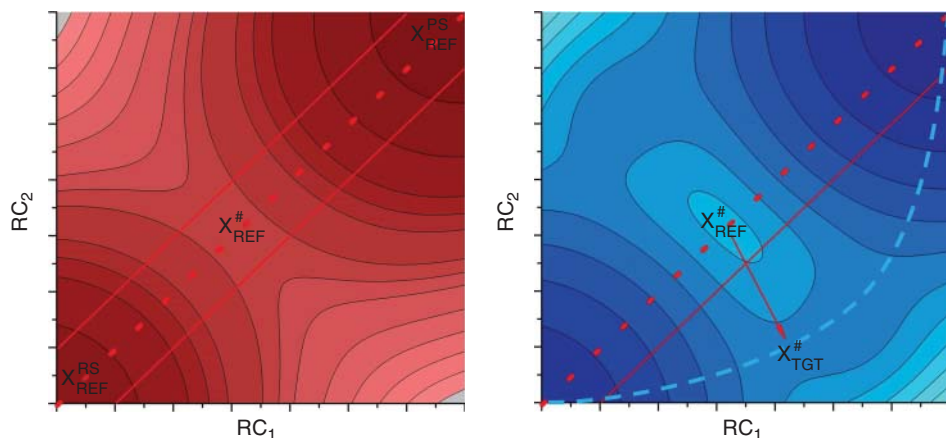


Figure 3.3 Challenges in Application of the PD Approach. The first challenge is the existence of a difference between the reaction paths on the reference (RED) and on the target (BLUE) free energy surfaces. The reference reaction path (dashed red line) is concerted while the target reaction path (dashed blue line) is dissociative (or associative). If sampling is performed only along the reference path – the explored region (between two red solid lines) does not contain the target transition state (TS). Thus, the free energy change of moving to the target potential at the reference TS does not provide the actual correction for the target free energy barrier. Secondly, moving from the reference TS to the target TS while computing the free energy perturbation (red arrow) will lead to a slow convergence since not only the force law changes, but also the system moves along the reaction surface. (See color plate section for the color representation of this figure.)

activation free energies using the direct sampling with the target potential, the ability to reproduce the PMF results with the PD approach is critical for the practical significance and validation of the latter. In order to reach this task, several reference potential strategies were combined in a multi-step algorithm. That is, at first, a low-accuracy multidimensional target free energy surface is computed to locate the RS and the TS regions. Next, high-accuracy target free energy surface (i.e., PMF) is constructed locally at these regions. Finally, accuracy of positioning these PMF regions relative to the reference free energy levels is systematically improved.^[35]

The rest of the chapter is organized as following. The concept section introduces the basic ideas of both the PD and PMF approaches. Here, again, the two are considered together, because we are trying to get the barrier from the PMF for a fraction of its computational cost by applying the PD method! In the next section, two main practical challenges of using the PD approach instead of the PMF are presented, as well as some ways to solve them. Then, the statistical thermodynamics formulation is given for a general reference potential strategy. This section is followed by discussing technicalities of using the EVB method as a reference potential, in particular the parametric and functional EVB refinement strategies and formulation of the TS constraint to estimate the free energy perturbation while moving to the target free energy surface. The problem of computing the free energy perturbation with different estimators concludes the methodological part of the chapter. This last section discusses efficiency of the free energy estimators in the regime, which is likely to be encountered in the reference potential approach – when the overlap with the target potential is small or poor.

3.2 Concept

To accurately compute the activation barrier on a reaction path one has to establish the free energy difference between the TS and the RS, in other words, to position the corresponding regions relative to each other on the free energy scale. This difference can reach tens of thermal units of energy, $k_B T$ (for macroscopic system with N_A particles at 298 K, $RT = k_B T \cdot N_A = 0.593 \text{ kcal} \cdot \text{mol}^{-1}$). However, all existing free energy algorithms can accurately and reliably estimate free energy differences on the order of only few $k_B T$. Consequently, the activation barrier is computed by connecting the TS to the RS (or the product state) on the free energy scale in a step-wise fashion, for instance by moving along the reaction path. Within this approach, all steps connecting the TS to the RS should be computed with equally high desired accuracy, since the error is dependent on the largest error made in the computation of previous steps. Discretization of the reaction path in the FEP-based PMF approach^[36,37] is achieved using the multi-step approach by gradually moving from the RS towards the TS (Figure 3.2). This also allows one to efficiently sample the elevated regions of the reaction free energy surface. That is, the original potential (e.g., E_{TGT}) is modified by introducing a bias, w_m , (e.g., a harmonic constraint), centered at a particular value of reaction coordinate, thus creating a biased potential, T_m :

$$T_m = E_{TGT} + K(\xi - \xi_m^0)^2 \equiv E_{TGT} + w_m \quad (3.4)$$

Here ξ designates a chosen reaction coordinate; ξ_m^0 is the center of the harmonic bias; and K is the force constant. Note, that the same construction, which is based on the reference potential will provide a biased reference potential, which is further referred to as R_m . In order to properly sample the full range of the reaction coordinate, a set of harmonic biases is created by changing the mapping (coupling) parameter (λ_m) incrementally from 0 to 1:

$$\xi_m^0 = (1 - \lambda_m)\xi_{LB}^0 + \lambda_m\xi_{UB}^0 \quad (3.5)$$

Here ξ_{LB}^0 and ξ_{UB}^0 correspond to the values of the selected ξ at the lower and the upper bounds, respectively. Generalization of Eq. (3.4) to the case of a multi-dimensional reaction coordinate is straightforward and involves introducing a harmonic constraint for each component of the reaction coordinate.

The unbiased probability distribution of the reaction coordinate is related to the configurational averages computed with the modified (biased) target potential, T_m , through the umbrella sampling formula:^[34]

$$\langle \delta(\xi - x) \rangle_{E_{TGT}} = \frac{\langle \delta(\xi - x) \exp[-\beta(E_{TGT} - T_m)] \rangle_{T_m}}{\langle \exp[-\beta(E_{TGT} - T_m)] \rangle_{T_m}} = \frac{\langle \delta(\xi - x) \exp[\beta w_m(\xi)] \rangle_{T_m}}{\langle \exp[-\beta(E_{TGT} - T_m)] \rangle_{T_m}} \quad (3.6)$$

Note that here and further $\langle \dots \rangle_E$ denotes a time (or ensemble) average computed with a potential E . Rewriting this equation in terms of the free energy^[36] changes and in the case of the harmonic bias of Eq. (3.4) gives three terms:

$$\Delta g_{TGT}(x) = \Delta G(E_{TGT} \rightarrow T_m) - \beta^{-1} \ln \langle \delta(\xi - x) \rangle_{T_m} - w_m(x) \quad (3.7)$$

The first right-hand side term of Eq. (3.7) corresponds to the denominator of Eq. (3.6). It is the free energy penalty for introducing a bias, which is also known as the free energy shift for a biased potential, T_m . This shift can be calculated by any FEP method with a multi-step scheme (see Estimation of the Free Energy Perturbation Section). The last two terms of Eq. (3.7) correspond to the numerator of Eq. (3.6), and include the biased probability distribution of the reaction coordinate minus the bias. In fact, this is the local free energy surface, which is positioned relative to other local regions using the corresponding free energy shift.

The total computational cost for computing the PMF is proportional to the number of independent trajectories in the multi-step free energy perturbation. Naturally, the larger the space of interest (in particular, its dimensionality) the more trajectories one has to propagate in order to explore the corresponding free energy surface. However, if the locations of the RS and the TS regions are known, even approximately, the relative position of these regions can be established at a fraction of the computational cost, and to the same degree of accuracy, as with the direct sampling. The solution is simple: first, the reaction free energy surface is explored with a computationally cheap reference potential; secondly, the vertical free energy perturbation is computed from the reference to the target free energy surface at the RS and at the TS regions, and the target free energy surface is computed locally (Figure 3.2).

Thus, the free energy shifts computed with the target potential are simply replaced by the free energy shifts computed with the reference potential, augmented with the free energy change of moving to the target potential with the same bias. The free energy is a function of state, and its difference between two states is determined solely by those states, independently of the transformation path. In the PD model, the transformation path used to calculate the free energy change is different: a computationally cheaper route is taken by moving from the RS to the PS along the reference free energy surface instead of directly moving on the target free energy surface. However, moving between the two surfaces incurs a computational cost, which has to be lower than computing the PMF directly in order for the method to be efficient.

3.3 Challenges

While the PD scheme, with a 1D reaction coordinate, can already reduce the computational cost by up to two orders of magnitude,^[24] compared to direct PMF calculations corresponding to Eq. (3.7), obtaining quantitatively reliable estimates for the ab initio QM/MM activation free energy with these approaches can be challenging.^[25,38] The first challenge involves identifying the location of the RS and TS (or the target path) on the target free energy surface, while the second challenge is obtaining convergent estimates for moving from the reference to the target free energy surface. The following paragraphs illustrate solutions to these problems.

3.3.1 Different Reference and Target Reaction Paths

In principle, the target reaction path is a priori unknown and approximating it with a reference counterpart can lead to incorrect estimates, since the reference free energy

surface can be substantially different from the target free energy surface (Figure 3.3). Consider the following example: a reaction path on the reference free energy surface corresponds to a concerted mechanism, while the reaction path on the target free energy surface corresponds to a different (e.g., a dissociative or associative) mechanism. In this case, the position of the transition state (TS) will be completely different between the two free energy surfaces, and evaluating the perturbation while moving to the target potential, at the TS, will not provide a correct estimate of the activation free energy for the actual target path. This situation can be encountered, for example, when computing the minimum free energy path for phosphate monoester hydrolysis,^[39,40] where different methods or computational protocols can predict associative, concerted and dissociative mechanisms.

3.3.2 Convergence of the Free Energy Estimates

The second challenge involves obtaining convergent estimates for moving from the reference to the target free energy surfaces at particular values of the reaction coordinate. Convergence in FEP calculations is fast if the change in energy depends evenly on many degrees of freedom, but it is slow if the dependence is strong on few particular degrees of freedom.^[35,41] Examples of the latter case include moving along the reaction path in addition to altering the force law (see Figure 3.3) or moving across barriers in other degrees of freedom (e.g., conformational transition or change in coordination). A simple solution to this problem involves freezing the reacting fragments at a specific configuration (obtained from a pre-determined reaction path) when moving from one surface to the other, which is equivalent to evaluation of a higher-level-of-theory solvation free energy for a given solute configuration.^[1,2,30,31,42] However, some disadvantages of this approach include missing the solute entropic effects and reliance on relevance of the pre-determined path (see the previous paragraph on some possible pitfalls of this assumption).

A related problem concerns obtaining the least-error free energy estimate for a given sample size or performing the minimum sampling to compute the free energy change below a given error. This question will be discussed in details in the section “Estimation of the Free Energy Perturbation”, where a particular emphasis is made on the free energy interpolation algorithms, which provide better results for two non-overlapping ensembles of the energy gap between the target and the reference potential. At any rate, while one can leverage emerging advances in development of free-energy estimators,^[43] the most important question (as far as the method efficiency is concerned) is how to minimize the number of bridging ensembles in the multi-step FEP approach. Naturally, every step in the multi-step formulation decreases the efficiency of the PD approach over direct PMF calculations.

Finally, sampling with a biased reference potential provides a non-Boltzmann distribution for the target potential. Recovery of the Boltzmann distribution (reweighting) requires, strictly speaking, removal of the effect of both biases on the reaction coordinate and due to the reference potential. Reweighting, in fact, was an early implementation of the method.^[19,21] However, it was found that obtaining convergent free energies is problematic in a naïve implementation, due to poor overlap between the reference and the target potentials. This, in turn, leads to the regime where the umbrella sampling becomes inefficient.^[34] Computational approaches for overcoming these challenges are discussed in the next section.

3.4 Implementation of the Reference Potential Methods

Computing the activation free energy barrier at a target level of theory requires locating the minimum free energy path. An alternative task, however, might consist of quantifying possible reaction pathways.^[39,44] In general, if several mechanistic options are possible – the free energy barriers along the corresponding paths have to be computed. These competing pathways can be formulated and explored in terms of some multidimensional order parameter – such as breaking/forming bonds in More O’Ferrall–Jencks^[45] plots (Figure 3.3); this particular case of great importance in chemistry is considered in the next section. In other cases, clarifying reaction mechanisms requires more elaborated schemes based on empirical chemical knowledge,^[44] which, however, can often be reduced to a collection of 1D and 2D surfaces. The question of locating the minimum energy path is, however, not unique to the PD approach. It is highly relevant, since the minimum-energy paths are generally different between two levels of theory. What is important, therefore, is to ensure that the target free energy path is sampled. Thus, the main focus is to sample with the RP those degrees of freedom along which larger energy changes occur. Some of the suggested solutions to this problem involve: 1) Performing sampling with the RP in some extended space, which guarantees to contain the target path, and perturbatively computing the low-accuracy target surface; 2) Pre-defining the target reaction path and biasing sampling with the RP along it, and 3) Matching the reaction paths by refining the RP along the pre-determined target path. The key thing to keep in mind is that we need to find the RS and TS in order to compute the FEP and to position those regions of the target free energy surface. The only way to find the target reaction path is to compute the full target free energy surface. This is done, however, at a low accuracy or with a reduced dimensionality, for example, with a continuum solvent, in order to reduce the computational cost. These possible scenarios are described in the next section.

3.4.1 Locating the Target Reaction Path

3.4.2 Low-accuracy Target Free Energy Surface from Non-equilibrium Distribution

The first approach for locating the target path involves computing the target free energy surface at low-accuracy, but in an extended space, which contains all possibilities for the minimum energy path.^[35,46] A particularly important example in chemistry is the More O’Ferrall–Jencks plot, which describes the possible available mechanisms for a chemical reaction as a function of a geometric parameter. The RP potential is first used to sample the complete 2D surface, and then the target potential energy is evaluated for a set of points at each simulation window. This data are used to construct an approximation of the target free energy surface, from which the target minimum energy path is identified.

The first approach for processing these points involves a reweighting strategy, such as umbrella sampling.^[19] That is, the sampling is performed with a set of biased reference potentials, $R_m = E_{REF} + w_m$ (compare to Eq. (3.4)), and the target distribution is recovered using:

$$\langle \delta(\xi - x) \rangle_{E_{TGT}} = \frac{\langle \delta(\xi - x) \exp[-\beta(E_{TGT} - R_m)] \rangle_{R_m}}{\langle \exp[-\beta(E_{TGT} - R_m)] \rangle_{R_m}} \quad (3.8)$$

The free energy change of introducing a combined bias to the target potential corresponds to the denominator of Eq. (3.8). This free energy change can be represented as the sum of two free energy terms:

$$\Delta G(E_{TGT} \rightarrow R_m) = \Delta G(E_{TGT} \rightarrow E_{REF}) + \Delta G(E_{REF} \rightarrow R_m) \quad (3.9)$$

where $\Delta G(E_{TGT} \rightarrow E_{REF})$ does not depend on the bias ξ_m^0 and the dependence on the reaction coordinate is integrated out. Therefore, it is just a constant free energy shift, which can be subtracted from the free energy shifts for all mapping potentials, R_m . The second term is accurately evaluated from sampling with the RP using the multi-step FEP approach (i.e., by computing the free energy function of the reaction coordinate,^[36,47] or the potential of mean force^[37]). The estimated free energy surface from multiple mapping potentials can be further improved by requiring smoothness of the reaction surface, and by applying some appropriate interpolation technique.

Thus, evaluation of the numerator of Eq. (3.8) is the most challenging part of the calculation in terms of accuracy for approximating Eq. (3.6). Re-writing the numerator of Eq. (3.8) shows that with a poor overlap in the phase space the free energy estimate diverges due to the large energy gap:

$$\langle \delta(\xi - x) \rangle_{E_{TGT}} = \frac{\langle \delta(\xi - x) \exp[\beta w_m(\xi)] \exp[-\beta \Delta E] \rangle_{R_m}}{\langle \exp[-\beta(E_{TGT} - R_m)] \rangle_{R_m}} \quad (3.10)$$

In other words, configurations generated with the reference potential are, mostly, unlikely configurations for the target potential. In which case, the sampling obtained with the reference potential is not representative for the target potential. As a result, the convergence of the free energy estimate with Eq. (3.10) is slow, and the corresponding estimate is of low accuracy. However, obtaining this estimate doesn't require sampling with the target potential, only evaluating energy for a selected set of configurations. Furthermore, the resulting low-accuracy free energy surface, while quantitatively unreliable, can be used to identify the location of the RS and of the TS.

3.4.3 Obtaining a Low-Accuracy Target Free Energy Surface from Free Energy Perturbation

An alternative approach for computing the low-accuracy target free energy surface from sampling with a reference potential in a high-dimensional space with the reweighting approach of Eq. (3.6) is based on approximating the target free energy shifts (blue empty circles in Figure 3.2) by adding a perturbative correction (red vertical arrows) to the reference free energy shifts (red solid circles). That is, another thermodynamic cycle is introduced in computing the Eq. (3.6). Note, that while moving from the reference to the target free energy surface at a particular value of the reaction coordinate we are switching between the biased potentials that govern interaction in the system (also known as the force law). This correction is computed with the FEP approach from sampling generated with the reference potential only. The FEP exponential average estimator can also be approximated with its linear expansion (see "Estimation of the Free Energy Perturbation" below). Then, the free energy change corresponding to the denominator in Eq. (3.6) (the free energy of introducing the bias to the target potential), is computed via the following thermodynamic cycle:

$$\Delta G(E_{TGT} \rightarrow T_m) = \Delta G(E_{TGT} \rightarrow E_{REF}) + \Delta G(E_{REF} \rightarrow R_m) + \Delta G(R_m \rightarrow T_m) \quad (3.11)$$

where the first two terms were discussed in the previous section and the last term is the FEP exponential average for moving from the biased reference potential to the biased target potential:

$$\Delta G(R_m \rightarrow T_m) = -\beta^{-1} \ln \langle \exp[-\beta \Delta E] \rangle_{R_m} \approx \langle \Delta E \rangle_{R_m} \quad (3.12)$$

where $\Delta E = T_m - R_m = E_{TGT} - E_{REF}$ if the bias is the same in R_m and T_m . This method of computing the perturbation is very similar to a number of other reference-potential based strategies,^[32,48,49] where it is used as the final free energy estimate for the target path. Mathematically, this approach also corresponds to approximating the numerator of Eq. (3.6) with the reference potential distribution:

$$\langle \delta(\xi - x) \rangle_{T_m} \approx \langle \delta(\xi - x) \rangle_{R_m} \quad (3.13)$$

The final approximation of Eq. (3.6) is:

$$\langle \delta(\xi - x) \rangle_{E_{TGT}} \approx \frac{\exp[\beta w_n(x)] \langle \delta(\xi - x) \rangle_{R_m}}{\exp[-\beta \Delta F(E_{REF} \rightarrow R_m)] \langle \exp[-\beta \Delta E] \rangle_{R_m}} \quad (3.14)$$

While the convergence of Eq. (3.14) is limited by the convergence of the free energy perturbation in the denominator, the accuracy of the final estimate is also affected by the assumption of Eq. (3.13), which can be further improved using:

$$\langle \delta(\xi - x) \rangle_{T_m} = \frac{\langle \delta(\xi - x) \exp[-\beta \Delta E] \rangle_{R_m}}{\langle \exp[-\beta \Delta E] \rangle_{R_m}} \quad (3.15)$$

Then, substitution of Eq. (3.15) into Eq. (3.14) yields:

$$\langle \delta(\xi - x) \rangle_{E_{TGT}} \approx \frac{\exp[\beta w_n(x)] \langle \delta(\xi - x) \exp[-\beta \Delta E] \rangle_{R_m}}{\exp[-\beta \Delta F(E_{REF} \rightarrow R_m)] \langle \exp[-\beta \Delta E] \rangle_{R_m}^2} \quad (3.16)$$

Free energy shifts from multiple potentials can be further smoothed to minimize the overall error. While the error of these estimates is usually too high for quantitative purposes, the topology of the target surface is likely to be qualitatively correct, which allows one to locate the target path.^[35]

3.4.4 Pre-Computing the Reaction Path

Another approach to deal with the mismatch between the reference and the target reaction paths involves pre-computing the target reaction path. In this case, the target reaction path is defined in terms of solute degrees of freedom, but with a simplified solvation model.^[50] This can be done using one of the energy minimization methods^[51] for locating the minimum energy path or by performing a potential energy scan. Then, this path is used as an approximation to the path with an atomistic solvent model. It, however, needs to be defined only in terms of the reaction coordinate, by letting all other degrees of freedom relax. At the next step, sampling with the RP is biased to the identified reaction path. An additional refinement procedure in this region can be used to maximize the overlap with the TP (see more below on the refinement procedure).

Conceptually, this approach is similar to the previous two: the target reaction path is identified on a coarse target free energy surface. While in the previous two examples the coarsening is done in terms of accuracy, in this example the coarsening is achieved by a simplified (continuum) representation of solvent. Once the target path is located, sampling with the reference potential can be biased along this path.

3.4.5 Reference Potential Refinement: the Paradynamics Model

The PD model introduced another way to circumvent difficulties arising due to the mismatch between the target and the reference reaction paths. It employs an iterative refinement procedure, which brings the reference potential and the reaction path close to the target counterparts. Here, we only introduce this concept, while details of the implementation are presented below. The refinement of the reference potential starts along a guess for the target path, which can either be pre-computed or approximated by the reference path. In any case, with every iteration the PD refinement increases the overlap between the two potential and minimizes the mismatch between the paths. The proximity of two paths ensures that the target reaction path is sampled with the reference (e.g., EVB) potential. A limited sampling with the target potential is used to improve convergence of the free energy estimates and to provide a framework for the refinement procedure. The error in evaluating Eq. (3.6) is reduced by performing a limited sampling with the target potential while computing the free energy change from the end-point distributions within a two-step LRA approach:

$$\Delta G_{LRA} = \frac{1}{2}(\langle T_m - R_m \rangle_{R_m} + \langle T_m - R_m \rangle_{T_m}) \quad (3.17)$$

The LRA approach is applied to compute the free energy perturbation while switching from the reference potential to the target potential (with the same bias at the RS and the TS, e.g.,^[22,25]). With the same bias, Eq. (3.17) becomes:

$$\Delta G_{LRA} = \frac{1}{2}(\langle \Delta E \rangle_{R_m} + \langle \Delta E \rangle_{T_m}) \quad (3.18)$$

Then, with the approximation of Eq. (3.18), the free energy perturbation reproduces the difference between two free energy functions,^[25] which are computed locally from the limited sampling. That is:

$$\Delta G(R_m \rightarrow T_m) \cong \Delta g_{TGT}(x = \xi_{TGT}^\ddagger) - \Delta g_{REF}(x = \xi_{REF}^\ddagger) \quad (3.19)$$

However, if the reference reaction path differs from the target path, that is, $\xi_{REF}^\ddagger \neq \xi_{TGT}^\ddagger$, application of Eq. (3.18) will cause convergence problems, since during the perturbation process the system will move along the reaction path. Therefore, if the target TS is known, it is more efficient to choose the closest mapping potential, R_m , compute the vertical perturbation with Eq. (3.18), and then move towards the closest to the target TS mapping potential, T_{m+n} .^[22] Finally, the accuracy of evaluating the free energy cost of moving to the target free energy surface can be systematically improved with a multi-step FEP procedure, which requires more local sampling with the target potential,

3.4.6 Moving From the Reference to the Target Free Energy Surface at the TS Using Constraints on the Reaction Coordinate

Formulation of constraints on the reaction coordinate for moving from the reference free energy surface to the target free energy surface is one of the key prerequisites for obtaining the convergent estimates for the free energy perturbation. This subsection describes a protocol for computing the free energy perturbation at a particular value of the reaction coordinate, for example, at the TS.

The free energy perturbation of moving between the reference and the target free energy surfaces ($\Delta G(R_m \rightarrow T_m)$) (red vertical arrows in Figure 3.2) has to converge much

faster than the free energy perturbation of moving to the TS along the reaction path on the target free energy surface, ($\Delta G(E_{TGT} \rightarrow T_m)$), shown with empty blue circles in Figure 3.2. Otherwise, the reference potential approach becomes impractical compared to PMF calculations. As was noted above, the free energy change of altering the force law converges faster when the energy difference is distributed among many degrees of freedom (such as solvent polarization) rather than strongly depends on few particular degrees of freedom (for example, on the reaction coordinate). Therefore, in computing the free energy perturbation of moving from one potential to another, the degrees of freedom on which the energy difference depends strongly and unequally between the two potentials should be constrained/frozen. For instance, the perturbation has to be vertical relative to the reaction coordinate (Figure 3.2 and Eq. (3.18)), and it should not involve overcoming high barriers in the transverse degrees of freedom or any other chemical events (such as changes in the coordination number). The vertical perturbation means that the bias to the target potential and to the reference potential should be the same. This is achieved by setting the same bias center and the force constant parameters in Eq. (3.4) for R_m and T_m . This bias should be sufficiently strong to contain sampling within the desired value of the reaction coordinate, which is typically close to the bias center.

For the EVB method this formulation is more sophisticated due to a particular solute-solvent coupling. Here, one important distinction between EVB and molecular orbital-based QM/MM methods is the natural reaction coordinate. In EVB it corresponds to the energy gap, also called the global reaction coordinate (see the section on EVB below and Chapter 2 of this book), whereas for ab initio QM/MM it is the nuclear reaction coordinate. While it is possible to compute the EVB free energy profile vs. the nuclear reaction coordinate (Figure 3.4), imposing constraints on the nuclear reaction coordinate within the EVB approach is inefficient, since the solvent is relatively insensitive to it. In the EVB method, the solvent has to be polarized and moved towards the TS,^[25] since the EVB charges in the condensed phase are artificially strongly dependent on the existing solvent configuration and insensitive to the nuclear configuration (see Eq. (1.23)). In ab initio (or semi-empirical) QM/MM, on the other hand, the solute electronic density is highly sensitive to the nuclear (solute) reaction coordinate and the solvent is polarized towards it by new QM charges. This minor inconvenience requires a special formulation of constraints when moving from the EVB surface to the ab initio QM/MM surface.^[25]

For the EVB potential, a more efficient approach to contain sampling within its TS (where $E_1 = E_2$) involves using one of approximations^[52] for the EVB mapping potential, in addition to a harmonic constraint on the nuclear reaction coordinate used on the target potential. One formulation for the EVB mapping (biased) potential is given by:

$$E_{EVB}(x_{EVB}^\ddagger) \approx 0.5E_1 + 0.5E_2 - H_{12} \quad (3.20)$$

While this approximation closely follows the EVB adiabatic potential at the TS, the off-diagonal element in the EVB approach is often chosen by calibration while post-processing the calculations. Thus, the presence of the off-diagonal element in the biased potential can be undesirable. Note that Eq. (3.20) is equal to the adiabatic EVB at the TS (*cf.* Eq. (3.36)). The use of this expression is equivalent to constraining the eigenvector components to be equal (so that both EVB diabatic states equally contribute to the adiabatic state). In another approximation for the mapping potential,

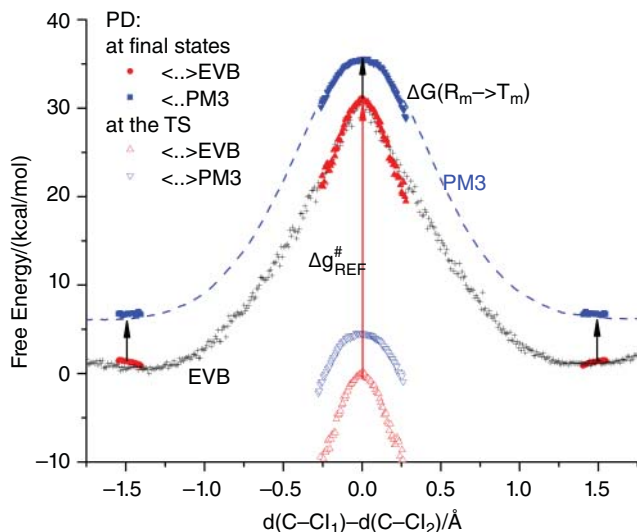


Figure 3.4 PM3/MM Activation Free Energy Barrier For $\text{MeCl}+\text{Cl}^-$ S_N2 Reaction Computed with the Paradyamics Model^[25] Using EVB as a Reference Potential (Which was Parametrically Refined^[24]). The dashed blue line is the PM3 target free energy surface generated with the PMF approach. To compute the activation barrier with the PD approach, the EVB profile is computed first (black "+" symbols). Next, local free energy functions are computed at the TS (empty blue flipped triangles for PM3 and empty red triangles for EVB) and at final states (solid blue squares for PM3 and red circles for EVB). The relative positions of the local free energy functions are computed using the LRA estimator (black arrows). The EVB reference free energy functions are aligned to match the full EVB free energy profile (red triangles with black "+" symbols), thus the local PM3 free energy function at the TS is positioned relative to the ones at the final states (flipped blue triangles). Adapted with permission from reference [25]. Copyright 2012 American Chemical Society. (See color plate section for the color representation of this figure.)

the off-diagonal element in Eq. (3.20) is omitted. Finally, the mapping potential can be the adiabatic EVB surface with a harmonic constraint on the energy gap.

3.4.7 High-Accuracy Local PMF Regions from Targeted Sampling

Once constraints on the TS are promptly formulated, and any other degrees of freedom that influence energy particularly strongly are frozen, the question of convergence of the free energy perturbation becomes a technical issue, which can be formulated to give a systematically improved estimate. Using the limited sampling with the target potential at the RS and TS regions, one can evaluate the numerator of Eq. (3.6), $\langle \delta(\xi - x) \exp[-\beta(E_{TGT} - T_m)] \rangle_{T_m}$, at the same level of accuracy as with the direct sampling (in calculating the full PMF). With the denominator evaluated using Eq. (3.11), the only difference in computing the reaction barrier from direct PMF calculations is that the denominator is computed using the thermodynamic cycle of the PD model (Figure 3.2). Thus, the overall efficiency of this approach is determined by convergence of Eq. (3.11). For a reference potential retaining an adequate physical description of changes in the electron density of the reacting fragments, the corresponding structural changes in the system, upon switching to the TP biased at the same region of the free energy surface are smaller than changes caused by moving the system on the target

surface from the RS to TS. Furthermore, the perturbation of Eq. (3.11) should converge with fewer simulation windows (trajectories propagated with the target potential) than the perturbation associated with the first term of Eq. (3.7). The computational cost of calculating the activation barrier (or the reaction free energy change), relative to conventional PMF approaches, will be smaller and the difference in the cost between the PMF and the PD approach will increase as the dimensionality of the reaction coordinate increases.

3.4.8 Improving Accuracy of Positioning the Local PMF Regions

In practice, better estimates for the energy difference between the RS and the TS regions of the target free energy surface can be obtained by averaging over perturbations for several mapping potentials, which are biased to sample the respective regions using the procedure described below. That is, the perturbation is computed for several adjacent bias centers (shown with solid blue circles in Figure 3.2). This operation also extends the region of the target free energy surface computed locally.^[35] These simulation windows are processed with Eq. (3.6) and, thus, are connected via the target free energy shifts, $\Delta G(T_m \rightarrow T_{m+i})$. On the other hand, several LRA free energy estimates for moving from the reference to the target potential, $\Delta G(R_i \rightarrow T_i)$, are available for each bias location, ξ_{m+i}^0 . These estimates are related:

$$\Delta G(R_m \rightarrow R_{m+i}) + \Delta G_{LRA}(R_{m+i} \rightarrow T_{m+i}) = \Delta G_{LRA}(R_m \rightarrow T_m) + \Delta G(T_m \rightarrow T_{m+i}) \quad (3.21)$$

Therefore, one can compute the average free energy perturbation over n adjacent simulation windows:

$$\overline{\Delta G_{LRA}(R_m \rightarrow T_m)} = \frac{1}{n} \sum_{i=1}^n [\Delta G(R_m \rightarrow R_{m+i}) + \Delta G_{LRA}(R_{m+i} \rightarrow T_{m+i}) + \Delta G(T_{m+i} \rightarrow T_m)] \quad (3.22)$$

and use this improved estimate to position the corresponding region of target free energy surface.

Systematic accuracy improvement of the positioning of the local free energy surface regions involves multi-step FEP. Additional simulation windows can be easily introduced by scaling the perturbation of Eq. (3.3) until Eq. (3.11) converges. Once $\Delta G(R_m \rightarrow T_m)$ is converged, the estimate of the activation free energy barrier is of the same accuracy as the one computed with Eq. (3.7) using the PMF approach.

3.5 EVB as a Reference Potential

The EVB potential is very advantageous to be used as a reference potential due to its low computational cost. A detailed technical description of the EVB potential can be found in Chapter 2 of this book, and in dedicated review articles and books.^[13,53] Arguably, one of the main advantages of the EVB description of a chemical reaction is the natural relationship to valence bond theory and the concept of resonance structures, which are intuitive to most chemists. That is, essentially every resonance structure can be described by

a corresponding EVB diabatic state, and a set of EVB states forms an active space for the configuration interaction treatment. The EVB adiabatic potential is obtained by mixing EVB diabatic surfaces with the corresponding coupling elements.^[26,52] In practice, the number of relevant states is often limited (to a good approximation) to only the reactants and the products states, however a great deal of development has been made in the multi-state EVB models.^[54] From a computational point of view, the main advantage of the EVB approach is its low computational cost (compared to other methods for describing chemical processes). This is due to the fact that within the classical EVB approach, diabatic states are described with a specialized force-field. A notable difference from the classical MM force-fields is the description of chemical bonds with a Morse potential, incorporating bond-angle coupling and a description of van der Waals interaction with a Buckingham-type potential. In any case, the computational cost of the EVB treatment is only (roughly) n times higher of that for the classical force field, where n is the number of EVB states. This allows addressing very computationally demanding problems such as estimating the effect of mutations on the reaction free energy profile^[55] or extracting entropic contributions.^[56]

Another advantage of the EVB approach is its ability to describe chemical processes in the condensed phase. First, it provides a simple yet robust way of training the EVB potential on available experimental results for simulations in the condensed phase. The most frequently adjusted EVB parameters are the gas-phase shifts and the coupling element described in Chapter 2, for which a number of empirical functional forms have been suggested^[13,57] (see also the summary in Table 2.1). This parametric refinement, which is often called EVB calibration, lets one calibrate the EVB potential using available experimental and/or high-level *ab initio* information for reactions in vacuum or in solvent. The derived parameters are transferrable to reactions in proteins without further recalibration. This is possible since the solvent-solute interactions are included only in the diagonal elements of the EVB matrix.

Lastly, the EVB approach gives a microscopic interpretation of the Marcus parabolas,^[58] which in turn provides valuable theoretical insight into the thermodynamics of enzymatic catalysis, and a molecular interpretation of linear free-energy relationships. The energy gap between the EVB diabatic states is also a microscopically defined global reaction coordinate, which includes both solvent and solute degrees of freedom, and is capable of naturally capturing the non-equilibrium solvation effect.

Diagonal $H_{ii} = \langle i|H|i \rangle$ and off-diagonal $H_{ij} = \langle i|H|j \rangle$ elements of the EVB configuration interaction (CI) matrix are approximated using analytical expressions. This makes the EVB method extremely fast even compared to semi-empirical methods, but it requires empirical parameterization. Each EVB diagonal element is approximated by the analytical energy function for the i -th VB diabatic state $|i\rangle$:

$$E_i = E_{\text{bonded}} + E_{\text{nonbonded}} + \alpha_i \quad (3.23)$$

The first term of Eq. (3.23) describes bonded interaction within the QM region and the bonded QM/MM coupling; the second term accounts for electrostatic and Van der Waals interactions within the QM region and the corresponding nonbonded QM/MM coupling; the third term is the gas phase shift of the diabatic state.

The first term of Eq. (3.23) contains the terms describing bond stretching, angle bending and deformation of torsional (and improper torsional) angles:

$$E_{\text{bonded}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{tors}} \quad (3.24)$$

Breaking and forming bonds are described using Morse potentials:

$$E_{bond} = \sum_{bonds} D(1 - e^{-\beta(r-r_0)})^2 \quad (3.25)$$

Where D is the potential well depth; r_0 is the bond equilibrium distance and b is the parameter controlling the steepness of the well.

Angles deformations are described using the harmonic potential with a force constant K_Θ and the equilibrium value Θ_0 :

$$E_{angle} = \sum_{angles} K_\Theta(\Theta - \Theta_0)^2 \cdot \Phi_{coupl} \quad (3.26)$$

where Φ_{coupl} is the coupling element which is introduced when an angle formed by atoms i, j, k exists only in one of the EVB states due to breaking/forming bond between atoms i, j or j, k :

$$\Phi_{coupl} = \begin{cases} 1, & r \leq r_0 \\ 2e^{-\beta(r-r_0)} - e^{-2\beta(r-r_0)}, & r > r_0 \end{cases} \quad (3.27)$$

Torsional angles deformations are accounted for by:

$$E_{tors} = \sum_{tors} K_\phi(1 + \cos(n\phi - \phi_0))\Phi_{coupl} \quad (3.28)$$

where Φ_{coupl} (Eq. (3.27)) is the coupling element which is introduced for a torsional angle formed by atoms i, j, k, l which exists only in one of the EVB states due to breaking/forming bond between atoms i, j or k, l atoms.

In the case of a non-polarizable force field, the second term of Eq. (3.23) incorporates electrostatic and Van der Waals interactions:

$$E_{nonbonded} = E_{elec} + E_{vdW} \quad (3.29)$$

Electrostatic interactions consist of those within the QM region and of the interactions between the QM and MM regions. They are described with a classical Coulomb law:

$$E_{elec} = \sum_{QM} \sum_{i < j} C \frac{Q_i Q_j}{r_{ij}} f(r_{ij}) + \sum_{QM} \sum_{MM} C \frac{Q_i q_j}{r_{ij}} \quad (3.30)$$

here C is a constant that depends on the units of energy used (332 for kcal/mol); Q is the charge of a QM atom and q is the charge on the MM atom; $f(r_{ij})$ is the screening function defined as:

$$f(r_{ij}) = 1 - e^{-\alpha r_{ij}^2} \quad (3.31)$$

Van der Waals interactions within the EVB region are described by either a Lennard-Jones potential or a Buckingham potential:

$$E_{vdW} = \sum_{QM} \sum_{i < j} \left[C_{ij} e^{-\beta_{ij} r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right] + \sum_{QM} \sum_{MM} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right] \quad (3.32)$$

The calibration of the EVB method, and transferability of parameters between different phases is possible since the effect of solvent is included only in Hamiltonian diagonal elements.

The off-diagonal elements can be conveniently approximated by Gaussian functions:

$$H_{ij} = Ae^{-\alpha\xi^2} \quad (3.33)$$

where ξ is the reaction coordinate defined as the difference between the forming and breaking bond lengths.

After diagonalization of the EVB CI matrix the lowest eigenvalue is identified:

$$E_{EVB} = \langle \psi | \mathbf{H} | \psi \rangle \quad (3.34)$$

The corresponding normalized eigenvector $\langle \psi | = \langle c_1 \dots c_n |$ has components $c_1, c_2, \dots, c_i, \dots, c_n$. Nuclear derivatives are calculated using Hellman-Feynman theorem:^[59]

$$\frac{\partial}{\partial x} E_{EVB} = \langle \psi | \left[\frac{\partial}{\partial x} \mathbf{H} \right] | \psi \rangle \quad (3.35)$$

For example, for the 2 x 2 EVB Hamiltonian Eq. (3.34) and Eq. (3.35) become:

$$E_{EVB} = c_1^2 E_1 + c_2^2 E_2 + 2c_1 c_2 H_{12} \quad (3.36)$$

$$\frac{\partial}{\partial x} E_{EVB} = c_1^2 \frac{\partial}{\partial x} E_1 + c_2^2 \frac{\partial}{\partial x} E_2 + 2c_1 c_2 \frac{\partial}{\partial x} H_{12} \quad (3.37)$$

The EVB approach considered above is based on a very intuitive picture of diabatic states, which are closely related to the concept of resonance structures. Due to the analytical form of empirical approximations the computational cost of the EVB method is very low compared to standard QM/MM methods, which allows one to perform extensive configurational sampling and to achieve quantitatively reliable results, which are determined by the quality of the EVB parameterization.

3.5.1 EVB Parameter Refinement

EVB parameters have been mostly derived by fitting to experimental data and/or high-level QM calculations. However, EVB parameters can be fitted to any target potential. The chosen set of EVB parameters will determine the reference reaction path. As was already mentioned in the discussion above, one of the approaches for locating the target reaction path involves pre-computing it with a simplified solvent model. In that case, the reference potential can either be forced to follow the target path with a bias, or refined to improve the overlap with the target potential. The refinement can involve parametric fitting^[24] of the reference potential to the target potential energies, derivatives or partial charges. The EVB refinement procedure can leverage emerging advances in force-field optimizations schemes based on force-matching.^[60,61] An alternative strategy^[25] involves functional refinement of the RP by adding the difference between the reference and the target potentials along the reaction coordinate, which is approximated by Gaussian functions. In recent works,^[32,48] an implicit refinement strategy was used by choosing a suitable semi-empirical reference potential. A similar approach can easily be used with EVB parameters by creating parametric libraries optimized to specific QM methods.

Parametric refinement brings the EVB potential closer to the target potential and improves the convergence of the FEP calculations. Since the perturbation is computed at both the TS and at the RS, the PD refinement procedure is often performed on

configurations generated around these critical points. The EVB parameters are refined by seeking the minimum of the least square penalty function:^[24]

$$\begin{aligned} \mathfrak{S}(\mathbf{p}, \mathbf{r}) = & k_1 \sum_{i=1}^N (E_i^{EVB}(\mathbf{p}, \mathbf{r}) - E_i^{QM}(\mathbf{r}))^2 \\ & + k_2 \sum_{i=1}^N \sum_{j=1}^3 \left(\frac{\partial}{\partial x_j} E_i^{EVB}(\mathbf{p}, \mathbf{r}) - \frac{\partial}{\partial x_j} E_i^{QM}(\mathbf{r}) \right)^2 \end{aligned} \quad (3.38)$$

by either using a simplified Newton-Rhaphson approach where one EVB parameter, p^i , is refined at a time, with the result of $k + 1$ -th iteration given by:

$$p_{k+1}^i = p_k^i - g[\partial \mathfrak{S} / \partial p_i] / [\partial^2 \mathfrak{S} / \partial p_i^2]_k \quad (3.39)$$

or by refining the vector of EVB parameters, \mathbf{p} , in the optimal steepest descent approach with the result of $(k + 1)$ -th iteration given by:

$$\mathbf{p}_{k+1} = \mathbf{p}_k - s \cdot \text{grad} \mathfrak{S}(\mathbf{p}_k) \quad (3.40)$$

One might want to restrict the boundaries of the parametric space in order to avoid overfitting, and to stay within physically meaningful values. Note that parameters here include those used to describe bonding interactions, the nonbonded van der Waals interactions, and the coupling element for a given set of EVB charges, which, in turn, are usually derived by fitting the electrostatic potential of a molecule in an implicit solvent. While the refinement scheme given above is for EVB parameters, it can be generalized to semi-empirical RPs by fitting parameters to a particular level of theory.^[62] Thus, the PD approach can be applied to any reference potential.

3.5.2 EVB Functional Refinement

An alternative to parametric refinement is the functional refinement scheme.^[25] This approach utilizes the data obtained from the potential energy scan, if it was performed in implicit solvent to locate the target path (see pre-computing the reaction path). First, charges computed for the final states are used as charges for corresponding EVB diabatic states (see Eq. (3.30)). Second, the potential energy scan (after subtracting the solvation effect) is fitted with a set of k Gaussian functions:

$$\Gamma_{TGT} = \sum_k A_k \exp(-\alpha_k (\xi - \xi_k^0)^2) \quad (3.41)$$

See Figure 3.5. This can be done, for instance, by the least-square minimization of the difference between the fit and the potential energy scan, E_{TGT} , represented by N points along the reaction path:

$$\mathfrak{S}(\{\alpha_k\}, \{A_k\}) = \sum_{i=1}^N (\Gamma_{TGT}(\xi_i) - E_{TGT}(\xi_i))^2 \quad (3.42)$$

Similarly, the reference potential (e.g., EVB) is evaluated for the same set of geometries and with updated charges for EVB diabatic states. The resulting energy profile is fitted using the same procedure with a set of Gaussians, Γ_{REF} . In the next step, the difference between the two fitted functions is used to refine the original reference potential:

$$E'_{REF} = E_{REF} - \Gamma_{REF} + \Gamma_{TGT} \quad (3.43)$$

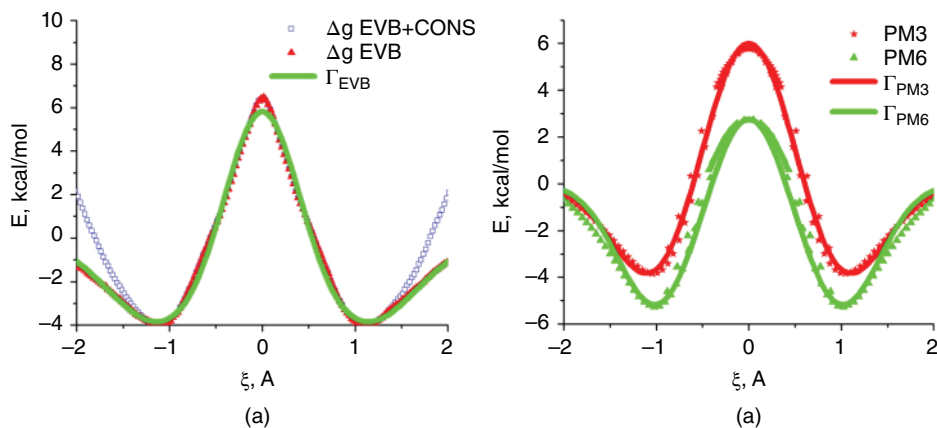


Figure 3.5 Fitting PMFs with Gaussian Functions. (A) Γ_{EVB} (green line) is a fit to the EVB gas-phase PMF (red triangles), which was obtained by subtracting contributions of the EVB distance constraints on the RC values from the actual EVB free energy profile (blue squares). (B) Γ_{PM3} and Γ_{PM6} (green line) are fits of potential energy scans obtained with semiempirical PM3 and PM6 hamiltonians. Reprinted with permission from reference [25]. Copyright 2012 American Chemical Society. (See color plate section for the color representation of this figure.)

The aim of this refinement procedure is to bring the reference potential close to the target potential in the region of interest along the reaction path (this procedure can be easily generalized to multiple dimensions). The refined reference potential is more likely to follow the target reaction path defined by the chosen reaction coordinate. When EVB is used as a reference potential, the fitting procedure is performed only for intramolecular interactions in order to transfer parameters to other phases (see the next section).

The solute-solvent coupling in EVB is included in the EVB diagonal elements.^[52] Thus, with the functional refinement derived for the intramolecular interactions within the solute, the EVB potential can be represented (see Eq. (3.36)):

$$E'_{\text{EVB}} = E_{\text{INTRA}} + \Delta\Gamma_{\text{INTRA}} + c_1^2 E_{1,\text{sS}}(\mathbf{Q}_{\text{RS}}^{\text{TGT}}) + c_2^2 E_{2,\text{sS}}(\mathbf{Q}_{\text{PS}}^{\text{TGT}}) \quad (3.44)$$

Where $\Delta\Gamma_{\text{INTRA}} = \Gamma_{\text{TGT}} - \Gamma_{\text{EVB}}$ refines only the intramolecular part of the EVB potential. And the solute-solvent electrostatic interactions are refined by using a new vector of EVB charges, which is derived for the reactant and product with the target potential with a continuum solvent model. This procedure allows to refine the EVB reference potential to be used with different target potentials in the condensed phase (Figure 3.6).

3.6 Estimation of the Free Energy Perturbation

One of the main computation tools in the PD approach is the FEP, which is used to estimate the free energy change while moving from the reference to the target free energy surface. FEP as was already mentioned earlier involves altering the force law. That is, the process of moving between the free energy surfaces consists of switching the potential that describes interactions in a system. As was also mentioned earlier, once the target high-accuracy PMF is computed locally at the RS and at the TS regions, the accuracy becomes limited by the convergence of the FEP in Eq. (3.11). Thus, it becomes critical to use the best available free energy estimator in the FEP approach. Below a number of

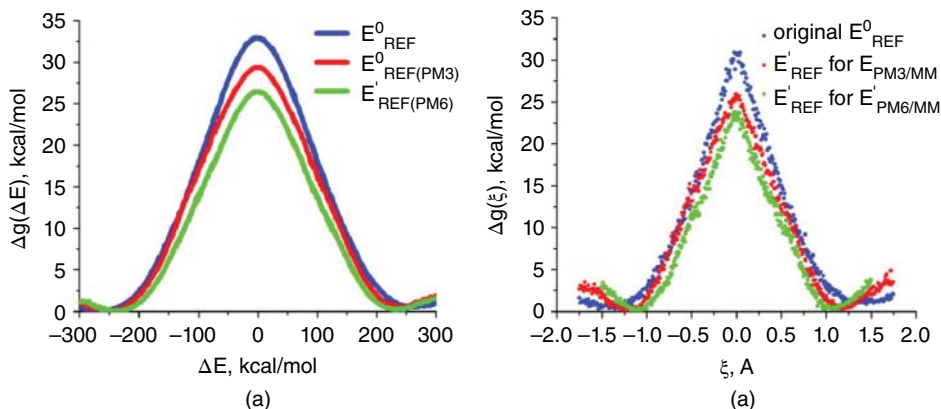


Figure 3.6 Demonstrating the Refinement of the EVB Reference Potential in Condensed Phases, Using the Correction Potential $-\Gamma_{EVB} + \Gamma_{TARGET}$ Represented by Gaussians And a Vector of New EVB Charges Derived With the Target Potential. The free energy profiles obtained by the EVB FEP/US approach along the EVB energy gap (A) and along the nuclear RC (B). (blue) the original EVB reference potential with parameters refined for the gas phase PM3; (red) the EVB reference potential refined for PM3/MM target potential; (green) the EVB reference potential refined for PM6/MM target potential. Reprinted with permission from reference [25]. Copyright 2012 American Chemical Society. (See color plate section for the color representation of this figure.)

such estimators is presented, and they can be implemented concurrently. In fact, the difference between the free energy estimators can be used to check the quality of data, and to estimate the error, since in the limit of infinite sampling all free energy estimators converge to the same answer.

3.6.1 Exponential Average

The free energy change of moving from the reference potential to the target potential is formulated in terms of the FEP approach.^[17,18] Computing this perturbation involves evaluating the exponential average of the energy gap between these two potentials. For instance, the free energy of moving from the reference potential to the target potential $E_{REF} \rightarrow E_{TGT}$ is equivalent to adding the perturbation, $\Delta E = E_{TGT} - E_{REF}$, to the reference potential (Eq. (3.3)). Then the classical FEP approach provides the exponential average free energy estimate:

$$\Delta F = -\beta^{-1} \ln \frac{\int \exp(-\beta \Delta E) \exp(-\beta E_{REF}) d\mathbf{x}^N}{\int \exp(-\beta E_{REF}) d\mathbf{x}^N} \equiv -\beta^{-1} \ln \langle \exp(-\beta \Delta E) \rangle_{E_{REF}} \quad (3.45)$$

This configurational average can be computed over time or over an ensemble using molecular dynamics or Monte Carlo simulations. Recall, that $\langle \dots \rangle_{E_i}$ indicates that the sampling is performed with the potential E_i , while obtaining the Helmholtz free energy for the canonical (NVT) ensemble. This formulation can be easily generalized to the Gibbs free energy (NPT) and other thermodynamic ensembles.^[63] Below we assume that the mechanical contribution to the free energy change is negligible and the Gibbs and the Helmholtz free energy changes are equal: $\Delta G = \Delta F$.

The free energy change for the reverse process can be computed by taking the configurational average with the TP:

$$-\Delta G = -\beta^{-1} \ln \langle \exp(-\beta \Delta E) \rangle_{E_{TGT}} \quad (3.46)$$

Computing the corresponding average with the target potential (e.g., in Eq. (3.46)) is computationally more expensive than computing it with the reference potential. While Eq. (3.45) provides an efficient free energy estimate for sampling obtained from a single distribution, obtaining quantitatively more reliable estimates requires sampling from both end-point distributions. While the average of Eqs. (3.45) and (3.46) can be used to obtain such an estimate,^[47] another way of improving the FEP estimate is to imagine the existence of an intermediate point between the two potentials, such that $E_{INT} = 0.5E_{REF} + 0.5E_{TGT}$, and apply the FEP twice. Then

$$\Delta G(E_{REF} \rightarrow E_{TGT}) = \Delta G(E_{REF} \rightarrow E_{INT}) + \Delta G(E_{INT} \rightarrow E_{TGT}) \quad (3.47)$$

The two terms on the right hand side are computed using:^[64]

$$\Delta G(E_{REF} \rightarrow E_{TGT}) = -\beta^{-1} \ln \langle \exp(-0.5\beta\Delta E) \rangle_{E_{REF}} + \beta^{-1} \ln \langle \exp(0.5\beta\Delta E) \rangle_{E_{TGT}} \quad (3.48)$$

3.6.2 Linear Response Approximation (LRA)

It was demonstrated^[41] that, in the context of the various free energy calculations^[28,65] and for the PD approach in particular,^[22,25] a more reliable estimate of the corresponding free energy change of altering the force law is obtained from the both end-point distributions (i.e., from sampling with the reference and with the target potentials). In the reference potential scheme, the LRA approach gave the same accuracy as multi-step perturbative schemes with various estimators. The LRA approach of Eq. (3.17) can be derived^[66] by double expansion of Eqs. (3.45) and (3.46) into power series in perturbation, truncating^[18] to the linear term and taking the average.

However, such formal derivation creates a false impression on limited applicability of this approach to only small perturbations, $\Delta E \ll k_B T$. First, the LRA, Eq. (3.17) can be derived by applying the trapezoid rule to the Thermodynamic Integration (TDI) approach,^[35,64] after taking the analytical derivative of the free energy with respect to the perturbation parameter λ (defined by $E_m = (1 - \lambda_m)E_{REF} + \lambda_m E_{TGT}$):

$$\Delta G_{TDI} = \int_0^1 \frac{\partial G}{\partial \lambda} d\lambda = \sum_{m=1}^{n-1} \frac{1}{2} \left(\left\langle \frac{\partial G}{\partial \lambda} \right\rangle_{E_m} + \left\langle \frac{\partial G}{\partial \lambda} \right\rangle_{E_{m+1}} \right) \Delta \lambda \quad (3.49)$$

Second, the derivation of this approach based on Marcus parabolas^[28] reveals its true computational power as a physics-based recipe for free-energy interpolation (see the dedicated section below).

3.6.3 Bennet's Acceptance Ratio

Bennet^[41] has shown that the least-error free energy estimate, given two sufficiently overlapping distributions of the energy gap between the target potential and the reference potential, can be obtained by solving these two equations self-consistently:

$$\begin{aligned} \Delta G_{BAR}^i &= \beta^{-1} \ln \frac{\langle f[\beta(C^{(i)} - \Delta E)] \rangle_T}{\langle f[\beta(\Delta E - C^{(i)})] \rangle_R} + C^{(i)} \\ C^{(i+1)} &= \Delta G_{BAR}^i + \beta^{-1} \ln \frac{n_T}{n_R} \end{aligned} \quad (3.50)$$

Where f is the Fermi function:

$$f[x] = \frac{1}{1 + e^x}$$

And n_T and n_R are number of points from the reference and target distributions.

A good overlap between distributions of the energy gap is achieved in the multi-step free energy perturbation schemes. In this case, using the acceptance ratio or weighted histogram method^[67,68] can be the most optimal choice for minimizing the relative free energy compared to other estimators,^[69] even though the absolute error in the limit of adequate sampling might be negligible. In contrast, when moving from one potential to another in the reference potential approach, a problem can be a poor or no overlap between two energy gap distributions. In this case, the free energy interpolation schemes (or the curve-fitting method^[41]) has an advantage over the extrapolation methods.

3.6.4 Free Energy Interpolation

In the free energy interpolation scheme (also called the curve-fitting method^[41]), the free energy functions of the energy gap are fitted with polynomials, which are extrapolated across the gap. However, if the system response is linear to the perturbation, the functional form of this polynomial is a parabola. By further assuming the equal curvature of two parabolas (or equal width of two Gaussian distributions, which means the same variance of the energy gap) one arrives to the LRA method.^[35] The advantage of knowing the functional form of the polynomial a priori is that it allows for a straightforward implementation (numerical solution versus graphical solution) thus minimizing the risk of overfitting.

Two free distributions are related to the free energy shifts that can be shown from the umbrella sampling equation:

$$-\frac{1}{\beta} \ln p_{TGT}(x) = -\frac{1}{\beta} \ln p_{REF}(x) + x - \Delta G(E_{REF} \rightarrow E_{TGT}) \quad (3.51)$$

Assuming the linear response approximation in Eq. (3.51) and recognizing that the distribution of the energy gap corresponding to a parabola is a Gaussian, one can not only recover the LRA formula but even go beyond the approximation of the equal curvature.^[35]

$$\begin{aligned} \Delta G(E_{REF} \rightarrow E_{TGT}) &= \frac{\mu_{REF} + \mu_{TGT}}{2} - \frac{\sigma_{TGT}^2}{4\beta} (\mu_{REF} - \mu_{TGT})^2 \\ &+ \frac{\sigma_{REF}^2}{4\beta} (\mu_{TGT} - \mu_{REF})^2 - \frac{1}{\beta} \ln \left(\frac{\sigma_{REF}}{\sigma_{TGT}} \right) \end{aligned} \quad (3.52)$$

Where $\mu_i = \langle \Delta E \rangle_i$ and $\sigma_i = (\langle \Delta E^2 \rangle_i - \langle \Delta E \rangle_i^2)$.

This allows for correction of the parabolic polynomials of unequal curvature, or uneven variance of the energy gap between the two distributions. One can also compute both distributions of the energy gap (from sampling with the RP and with the TP) and inspect their overlap and functional forms to choose the most appropriate free energy estimator. Computing the free energy change with several estimators, or evaluating the perturbation in a multi-step way, can also be used to assess the error and convergence of the perturbation.

3.7 Overcoming Some Limitations of EVB Approach as a Reference Potential

Using EVB as a reference potential has certain technical subtleties, which are often compensated by its low computational cost. First, the EVB free energy surface is usually mapped and expressed as a function of the energy gap between two diabatic states (Figure 3.6A). Ab initio QM/MM free energy surfaces, on the other hand, are mapped and expressed in nuclear reaction coordinates. Mapping the EVB free energy surface along the nuclear reaction coordinate in the condensed phase is, however, inefficient since adiabatic EVB charges in the condensed phase show a strong dependence on the current solvent polarization and weak dependence on the solute configuration. This is easily overcome by mapping along the global reaction coordinate and forcing solvent to move to the transition state configuration. The EVB sampling, obtained this way, can be used to construct a free energy profile along the nuclear reaction coordinate (Figure 3.4 and 3.6B). Ab initio QM/MM methods don't automatically provide diabatic states, and defining and mapping along the global reaction coordinate is problematic, even though some attempts to use this approach have been made.^[70] The free energy surfaces in nuclear reaction coordinates can be used to provide a detailed quantitative insight into reactions mechanisms (as those are often defined in terms of nuclear coordinates like associative, dissociative, and concerted).

Constrained density functional theory^[71–73] (CDFT) approach allows one to obtain the diabatic states from first principles, which can then be used to parameterize EVB diabatic states. It combines the frozen density approach^[74] with valence bond theory and provides an electronic structure description of VB diabatic states in terms of developed DFT functionals. Furthermore, it lets the coupling elements be estimated, thus giving a full configuration-interaction matrix,^[75,76] the elements of which can be used in the refinement procedure.

Since EVB mapping along the global reaction coordinate only samples the minimum reaction path, one has to make sure that EVB is parameterized and follows the ab initio QM/MM reaction path. Alternatively, computing local PMF on the target free energy surface from the reference TS to the target TS is necessary.^[22] This in turn requires computing the reaction path from energy minimization technique with an implicit solvation model and reliance on its relevance. Alternatively, this problem can be overcome by exploring alternative pathways with corresponding EVB diabatic state.

3.8 Final Remarks

In recent years, a great deal of progress has been made towards development of sampling schemes, based on the reference potential approach, which resulted in various simulation protocols proposed by different workers. The EVB potential, due to its low computational cost, allows to overcome the high-computational cost of calculating the free-energy changes for chemical processes in the condensed phase. The PD approach provides a physically rigorous way to perturbatively improve the level of theoretical description. This chapter is not an exhaustive review, which an interested reader can

find elsewhere,^[33] but a summary of the key concepts of the field.

The main objective of this chapter is to present the key elements of the PD strategy, which allows one to compute the QM free energy barrier at the same level of accuracy as the PMF approach, but at a fraction of its computational cost. In order to achieve this, several elements are critical. First, one has to identify the target reaction path (at least RS and TS). This is achieved by constructing a low-accuracy (coarse) target free energy surface. Once the target path is identified, the target PMF (high-accuracy target free energy surface) is computed locally at the RS and at the TS from a limited sampling with the target potential. Finally, free energy levels of local PMF regions are established relative to the reference free energy surface using the free energy perturbation approach, which is formulated for switching from the reference to the target potential at a constrained value of the reaction coordinate. Thus, the accuracy of the target activation free energy estimate is limited by convergence of the free energy perturbation, which can be systematically improved in a multi-step fashion and by choosing an optimal free energy estimator. Refinement of the reference potential can significantly improve quality of the low-accuracy target free energy surface and the convergence of the free energy perturbation.

It should be noted that the PD method is not a specialized approach, which specifically works with the EVB method. In fact, recent advances in the PD method have focused on developing a rigorous statistical thermodynamics formalism for using sampling with two arbitrary levels of theory in free energy calculations and related tasks. Recent advances have been made in studying reactions in water and in proteins with semi-empirical PM3^[77] and PM6^[62] reference potentials.^[25,35] Furthermore, recent developments of density functional tight-binding methods^[78,79] opened up new opportunities for applications of these methods as reference potentials in PD-related approaches.

References

- 1 Kamerlin, S.C., Haranczyk, M. and Warshel, A. (2009) Progress in ab initio QM/MM free-energy simulations of electrostatic energies in proteins: Accelerated QM/MM studies of pKa, redox reactions and solvation free energies. *Journal of Physical Chemistry B*, **113** (5), 1253–1272.
- 2 Hu, H. and Yang, W. (2008) Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods. *Annual Review of Physical Chemistry*, **59**, 573–601.
- 3 Gao, J., Ma, S., Major, D.T. *et al.* (2006) Mechanisms and free energies of enzymatic reactions. *Chemical Reviews*, **106** (8), 3188–3209.
- 4 Acevedo, O. and Jorgensen, W.L. (2014) Quantum and molecular mechanical Monte Carlo techniques for modeling condensed-phase reactions. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **4** (5), 422–435.
- 5 Senn, H.M. and Thiel, W. (2009) QM/MM methods for biomolecular systems. *Angewandte Chemie International Edition*, **48** (7), 1198–1229.

- 6 Helgaker, T., Jørgensen, P. and Olsen, J. (2000) Calibration of the Electronic-Structure Models, in *Molecular electronic-structure theory*, John Wiley & Sons, Ltd, pp. 817–883.
- 7 Friesner, R.A. and Guallar, V. (2005) *Ab initio* quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis. *Annual Review of Physical Chemistry*, **56** (1), 389–427.
- 8 van der Kamp, M.W. and Mulholland, A.J. (2013) Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry*, **52** (16), 2708–2728.
- 9 Shaw KE, Woods CJ, Mulholland AJ, Abraham DJ. (2003) *QM and QM/MM. Approaches to evaluating binding affinities. Burger's medicinal chemistry and drug discovery*. (John Wiley & Sons, Inc.).
- 10 Rathore, R.S., Reddy, R.N., Kondapi, A.K. *et al.* (2012) Use of quantum mechanics/molecular mechanics-based FEP method for calculating relative binding affinities of FBPase inhibitors for type-2 diabetes. *Theoretical Chemistry Accounts*, **131** (2), 1–10.
- 11 Frushicheva, M.P., Mills, M.J., Schopf, P. *et al.* (2014) Computer aided enzyme design and catalytic concepts. *Current Opinion in Chemical Biology*, **21**, 56–62.
- 12 Kries, H., Blomberg, R. and Hilvert, D. (2013) De novo enzymes by computational design. *Current Opinion in Chemical Biology*, **17** (2), 221–228.
- 13 Warshel, A. (1991) *Computer modeling of chemical reactions in enzymes and solutions*, John Wiley & Sons, New York.
- 14 Warshel, A. and Levitt, M. (1976) Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology*, **103** (2), 227–249.
- 15 Kamerlin, S.C.L., Vicatos, S., Dryga, A. and Warshel, A. (2011) Coarse-grained (multiscale) simulations in studies of biophysical and chemical systems. *Annual Review of Physical Chemistry*, **62** (1), 41–64.
- 16 Frenkel, D. and Smit, B. (2001) *Understanding molecular simulation: From algorithms to applications*, Academic Press.
- 17 Zwanzig, R. (1954) High temperature equation of state by a perturbation method. I. Nonpolar gases. *Journal of Chemical Physics*, **22** (8), 1420.
- 18 Landau, L.D. and Lifshitz, E.M. (1969) *Statistical physics*, Pergamon Press, Oxford.
- 19 Luzhkov, V. and Warshel, A. (1992) Microscopic models for quantum mechanical calculations of chemical processes in solutions: LD/AMPAC and SCAAS/AMPAC calculations of solvation energies. *Journal of Computational Chemistry*, **13**, 199–213.
- 20 Bentzien, J., Muller, R.P., Florián, J. and Warshel, A. (1998) Hybrid ab initio quantum mechanics/molecular mechanics calculations of free energy surfaces for enzymatic reactions: The nucleophilic attack in subtilisin. *Journal of Physical Chemistry B*, **102** (12), 2293–2301.
- 21 Muller, R.P. and Warshel, A. (1995) Ab initio calculations of free energy barriers for chemical reactions in solution. *Journal of Physical Chemistry*, **99**, 17516–17524.
- 22 Rosta, E., Klahn, M. and Warshel, A. (2006) Towards accurate ab initio QM/MM calculations of free-energy profiles of enzymatic reactions. *Journal of Physical Chemistry B*, **110** (6), 2934–2941.

- 23 Rosta, E., Haranczyk, M., Chu, Z.T. and Warshel, A. (2008) Accelerating QM/MM free energy calculations: Representing the surroundings by an updated mean charge distribution. *Journal of Physical Chemistry B*, **112**, 5680–5692.
- 24 Plotnikov, N.V., Kamerlin, S.C.L. and Warshel, A. (2011) Paradynamics: An effective and reliable model for ab initio QM/MM free-energy calculations and related tasks. *Journal of Physical Chemistry B*, **115** (24), 7950.
- 25 Plotnikov, N.V. and Warshel, A. (2012) Exploring, refining, and validating the paradynamics QM/MM sampling. *Journal of Physical Chemistry B*, **116** (34), 10342–10356.
- 26 Warshel, A. and Weiss, R.M. (1980) An empirical valence bond approach for comparing reactions in solutions and in enzymes. *Journal of the American Chemical Society*, **102** (20), 6218–6226.
- 27 Lifson, S. and Warshel, A. (1968) Consistent force field for calculations of conformations, vibrational spectra, and enthalpies of cycloalkane and n-alkane molecules. *Journal of Chemical Physics*, **49** (11), 5116–5129.
- 28 Lee, F.S., Chu, Z.-T., Bolger, M.B. and Warshel, A. (1992) Calculations of antibody-antigen interactions: Microscopic and semi-microscopic evaluation of the free energies of binding of phosphorylcholine analogs to McPC603. *Protein Engineering*, **5** (3), 215–228.
- 29 Polyak, I., Benighaus, T., Boulanger, E. and Thiel, W. (2013) Quantum mechanics/molecular mechanics dual Hamiltonian free energy perturbation. *Journal of Chemical Physics*, **139** (6), 064105.
- 30 Claeysens, F., Harvey, J.N., Manby, F.R. *et al.* (2006) High-accuracy computation of reaction barriers in enzymes. *Angewandte Chemie, International Edition in English*, **45** (41), 6856–6859.
- 31 Rod, T.H. and Ryde, U. (2005) Quantum mechanical free energy barrier for an enzymatic reaction. *Physical Review Letters*, **94** (13), 138302.
- 32 Hou, G. and Cui, Q. (2013) Stabilization of different types of transition states in a single enzyme active site: QM/MM analysis of enzymes in the alkaline phosphatase superfamily. *Journal of the American Chemical Society*, **135** (28), 10457–10469.
- 33 Duarte, F., Amrein, B.A., Blaha-Nelson, D. and Kamerlin, S.C. (2015) Recent advances in QM/MM free energy calculations using reference potentials. *Biochimica et Biophysica Acta (BBA) – General Subjects*, **1850** (5), 954–965.
- 34 Torrie, G.M. and Valleau, J.P. (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, **23** (2), 187–199.
- 35 Plotnikov, N.V. (2014) Computing the free energy barriers for less by sampling with a coarse reference potential while retaining accuracy of the target fine model. *Journal of Chemical Theory and Computation*, **10** (8), 2987–3001.
- 36 King, G. and Warshel (1990) Investigation of the free-energy functions for electron-transfer reactions. *Journal of Chemical Physics*, **93** (12), 8682–8692.
- 37 Roux, B. (1995) The calculation of the potential of mean force using computer simulations. *Computer Physics Communications*, **91** (1–3), 275–282.
- 38 Heimdal, J. and Ryde, U. (2012) Convergence of QM/MM free-energy perturbations based on molecular-mechanics or semiempirical simulations. *Physical Chemistry Chemical Physics*, **14** (36), 12592–12604.

- 39 Plotnikov, N.V., Prasad, B.R., Chakrabarty, S. *et al.* (2013) Quantifying the mechanism of phosphate monoester hydrolysis in aqueous solution by evaluating the relevant ab initio QM/MM free-energy surfaces. *Journal of Physical Chemistry B*, **117** (42), 12807–12819.
- 40 Prasad, B.R., Plotnikov, N.V. and Warshel, A. (2013) Addressing open questions about phosphate hydrolysis pathways by careful free energy mapping. *Journal of Physical Chemistry B*, **117** (1), 153–163.
- 41 Bennett, C.H. (1976) Efficient estimation of free energy differences from Monte Carlo data. *Journal of Computational Physics*, **22** (2), 245–268.
- 42 Valiev, M., Yang, J., Adams, J.A. *et al.* (2007) Phosphorylation reaction in cAPK protein kinase-free energy quantum mechanical/molecular mechanics simulations. *Journal of Physical Chemistry B*, **111** (47), 13455–13464.
- 43 Chipot, C. and Pohorille, A. (2007) *Free energy calculations*, Springer, p. 2007.
- 44 Bora, R.P., Plotnikov, N.V., Lameira, J. and Warshel, A. (2013) Quantitative exploration of the molecular origin of the activation of GTPase. *Proceedings of the National Academy of Sciences of the United States of America*, **110** (51), 20509–20514.
- 45 Jencks, W.P. (1985) A primer for the Bema Hapothle. An empirical approach to the characterization of changing transition-state structures. *Chemical Reviews*, **85** (6), 511–527.
- 46 Ensing, B., Laio, A., Parrinello, M. and Klein, M.L. (2005) A recipe for the computation of the free energy barrier and the lowest free energy path of concerted reactions†. *Journal of Physical Chemistry B*, **109** (14), 6676–6687.
- 47 Kato, M. and Warshel, A. (2005) Through the channel and around the channel: Validating and comparing microscopic approaches for the evaluation of free energy profiles for ion penetration through ion channels. *Journal of Physical Chemistry B*, **109** (41), 19516–19522.
- 48 Polyak, I., Benighaus, T., Boulanger, E. and Thiel, W. (2013) Quantum mechanics/molecular mechanics dual Hamiltonian free energy perturbation. *Journal of Chemical Physics*, **139** (6): 064105.
- 49 Lonsdale, R., Hoyle, S., Grey, D.T. *et al.* (2012) Determinants of reactivity and selectivity in soluble epoxide hydrolase from quantum mechanics/molecular mechanics modeling. *Biochemistry*, **51** (8), 1774–1786.
- 50 Klamt, A. (2011) The COSMO and COSMO-RS solvation models. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **1** (5), 699–709.
- 51 Henkelman, G. and Jónsson, H. (2000) Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *Journal of Chemical Physics*, **113** (22), 9978–9985.
- 52 Hwang, J.K., King, G., Creighton, S. and Warshel, A. (1988) Simulation of free energy relationships and dynamics of SN2 reactions in aqueous solution. *Journal of the American Chemical Society*, **110** (16), 5297–5311.
- 53 Kamerlin, S.C.L. and Warshel, A. (2011) The empirical valence bond model: Theory and applications. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **1** (1), 30–45.
- 54 Day, T.J.F., Soudackov, A.V., Čuma, M. *et al.* (2002) A second generation multistate empirical valence bond model for proton transport in aqueous systems. *Journal of Chemical Physics*, **117** (12), 5839–5849.

- 55 Frushicheva, M.P., Cao, J., Chu, Z.T. and Warshel, A. (2010) Exploring challenges in rational enzyme design by simulating the catalysis in artificial kemp eliminase. *Proceedings of the National Academy of Sciences of the United States of America*, **107** (39), 16869–16874.
- 56 Schopf, P., Mills, M.J.L. and Warshel, A. (2015) The entropic contributions in vitamin B12 enzymes still reflect the electrostatic paradigm. *Proceedings of the National Academy of Sciences of the United States of America*, **112** (14), 4328–4333.
- 57 Chang, Y.T. and Miller, W.H. (1990) An empirical valence bond model for constructing global potential energy surfaces for chemical reactions of polyatomic molecular systems. *The Journal of Physical Chemistry*, **94** (15), 5884–5888.
- 58 Marcus, R.A. (1968) Theoretical relations among rate constants, barriers, and Broensted slopes of chemical reactions. *The Journal of Physical Chemistry*, **72** (3), 891–899.
- 59 Feynman, R.P. (1939) Forces in molecules. *Physical Review*, **56** (4), 340–343.
- 60 Ercolessi, F. and Adams, J.B. (1994) Interatomic potentials from first-principles calculations: The force-matching method. *Europhysics Letters*, **26** (8), 583.
- 61 Maurer, P., Laio, A., Hugosson, H.W. *et al.* (2007) Automated parametrization of biomolecular force fields from quantum mechanics/molecular mechanics (QM/MM) simulations through force matching. *Journal of Chemical Theory and Computation*, **3** (2), 628–639.
- 62 Stewart, J.P. (2007) Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *Journal of Molecular Modeling*, **13** (12), 1173–1213.
- 63 Lu, N. and Kofke, D.A. (2001) Accuracy of free-energy perturbation calculations in molecular simulation. I. Modeling. *Journal of Chemical Physics*, **114** (17), 7303–7311.
- 64 Hummer, G. and Szabo, A. (1996) Calculation of free-energy differences from computer simulations of initial and final states. *Journal of Chemical Physics*, **105** (5), 2004–2010.
- 65 Villà, J. and Warshel, A. (2001) Energetics and dynamics of enzymatic reactions. *Journal of Physical Chemistry B*, **105** (33), 7887–7907.
- 66 Leach, A.R. (2001) *Molecular modelling: Principles and applications*, Pearson Education.
- 67 Ferrenberg, A.M. and Swendsen, R.H. (1989) Optimized Monte Carlo data analysis. *Physical Review Letters*, **63** (12), 1195–1198.
- 68 Kumar, S., Rosenberg, J.M., Bouzida, D. *et al.* (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry*, **13** (8), 1011–1021.
- 69 Shirts, M.R. and Pande, V.S. (2005) Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration. *Journal of Chemical Physics*, **122** (14), 144107.
- 70 Mones, L., Kulhánek, P., Simon, I. *et al.* (2009) The energy gap as a universal reaction coordinate for the simulation of chemical reactions. *Journal of Physical Chemistry B*, **113** (22), 7867–7873.
- 71 Hong, G., Strajbl, M., Wesolowski, T.A. and Warshel, A. (2000) Constraining the electron densities in DFT method as an effective way for ab initio studies of metal-catalyzed reactions. *Journal of Computational Chemistry*, **21** (16), 1554–1561.

- 72 Kaduk, B., Kowalczyk, T. and Van Voorhis, T. (2012) Constrained density functional theory. *Chemical Reviews*, **112** (1), 321–370.
- 73 Wesolowski, T., Muller, R.P. and Warshel, A. (1996) *Ab initio* frozen density functional calculations of proton transfer reactions in solution. *Journal of Physical Chemistry*, **100**, 15444–15449.
- 74 Wesolowski, T. and Warshel, A. (1994) *Ab initio* free energy perturbation calculations of solvation free energy using the frozen density functional approach. *Journal of Physical Chemistry*, **98**, 5183–5187.
- 75 Hong, G., Rosta, E. and Warshel, A. (2006) Using the constrained DFT approach in generating diabatic surfaces and off diagonal empirical valence bond terms for modeling reactions in condensed phases. *Journal of Physical Chemistry B*, **110** (39), 19570–19574.
- 76 Wu, Q., Cheng, C.-L. and Van Voorhis, T. (2007) Configuration interaction based on constrained density functional theory: A multireference method. *Journal of Chemical Physics*, **127** (16), 164119.
- 77 Stewart, J.J.P. (1989) Optimization of parameters for semiempirical methods. II. Applications. *Journal of Computational Chemistry*, **10**, 221–264.
- 78 Lu, X., Gaus, M., Elstner, M. and Cui, Q. (2015) Parametrization of DFTB3/3OB for magnesium and zinc for chemical and biological applications. *Journal of Physical Chemistry B*, **119** (3), 1062–1082.
- 79 Gaus, M., Lu, X., Elstner, M. and Cui, Q. (2014) Parameterization of DFTB3/3OB for sulfur and phosphorus for chemical and biological applications. *Journal of Chemical Theory and Computation*, **10** (4), 1518–1537.

4

Empirical Valence Bond Methods for Exploring Reaction Dynamics in the Gas Phase and in Solution

Jeremy N. Harvey¹, Michael O'Connor² and David R. Glowacki²

¹Department of Chemistry, KU Leuven, Belgium

²School of Chemistry, University of Bristol, United Kingdom

4.1 Introduction

In this chapter, we will review the use of empirical valence bond (EVB)^[1] methods for studying the dynamics of chemical reactions. The focus will be on recent work that we have done in our groups studying reactions in solution, but we will also include examples of other work on gas-phase reactions. By *reaction dynamics*, we mean the detailed way in which atoms move during, prior to, and after a bond-making or breaking event.

Dynamics can be studied experimentally, for example, through the determination of state-dependent cross-sections and angular distribution functions. Huge progress in this field was obtained using the crossed-molecular beam approach, which was instrumental in the award of the 1986 Nobel prize in Chemistry to Dudley Herschbach, Yuan T. Lee and John Polanyi.^[2] Indirect information about the atomic motions during the chemical reaction is obtained from the observed distribution among energy levels for the products, as a function of the collision energy and of the initial rotational and vibrational states of the reactants. Also important is the spatial distribution of the products for a particular initial collision geometry.

Because the information provided by such experiments is somewhat *indirect*, theory has always played an important role in understanding reaction dynamics, by providing simulated models for the chemical reaction events. Such simulations are of various types. In the simplest case, with which we will mostly be concerned here, the atomic motions are assumed to occur classically on a potential energy surface as defined within the Born-Oppenheimer approximation. Multiple sets of initial conditions (atomic positions and velocities) are chosen based on some classical interpretation of the reaction conditions, and Newton's equations of motion are then integrated numerically. Reactive trajectories are monitored, the final positions and velocities mapped onto the experimental observables, and comparison is made to experiment. In more sophisticated methods, quantum effects on the nuclear motions are introduced, for example, through a classical representation of the nuclear zero-point energy^[3] or through use of quantum mechanical laws of motion in place of Newton's laws.^[4,5] Also, the existence of multiple electronic states and motion of the system from one state to another may be treated. In all these cases, it is necessary to be able to evaluate the potential energy (and its gradient) for many different arrangements of the atoms making up the reacting system.

One obvious way to do this is to solve the electronic Schrödinger equation for the corresponding nuclear coordinates. In the context of classical simulations of nuclear motions, this leads to the direct dynamics approach, or to the related Car-Parrinello simulation approach.^[6] However, these approaches are rather computationally demanding, especially for extended systems, since electronic structure methods typically scale badly with the number of atoms. In this respect, it should be noted that development of new computer architectures and of efficient algorithms is increasingly making such methods applicable to treat dynamics in solution. However, another challenge with such techniques is that the electronic structure method used always involves approximations, and thereby one is not necessarily guaranteed a potential energy surface of the required accuracy for the chemical question at hand. It is not straightforward to tune electronic structure methods to correctly predict important quantities such as barrier heights, reaction energy, and so on.

It is against this background – that is, accuracy balanced against efficiency – that we elected to use EVB methods in our recent work investigating the dynamics of solution-phase reactions. Over the last few years, ultrafast experimental pump-probe techniques, where a chemical reaction is initiated with one short laser pulse, followed by a spectroscopic analysis of the reaction mixture using a second fast pulse have yielded new insights into the dynamics of a number of reactions (for a recent overview, see^[7]). Like the crossed-beam and other gas-phase dynamical studies mentioned above, these approaches yield valuable but indirect information about the elementary reaction behaviour – usually in the form of a transient spectral band. Hence theory can play an important role in helping to interpret the experimental results. This requires accurate simulation of the reactions in solution, and we elected to use EVB for this given the accuracy that it can achieve with a relatively low computational cost.

4.2 EVB and Related Methods for Describing Potential Energy Surfaces

There are many different varieties of the EVB method, as discussed in Chapter 2. They have found widespread use in studying a number of problems. In the present chapter, we want to focus only on those studies where the EVB approach has been used to study reaction dynamics, taken in the narrow sense of quantum or classical studies of nuclear motion during chemical reactions, in both non-equilibrium and equilibrium regimes. We thus do not include studies of *rate constants* in the framework of various versions of transition state theory, or the related problem of characterizing potentials of mean force. We do note however that the borderline between advanced versions of transition state theory and ‘reaction dynamics’ is ill-defined.^[8] Also, given the table of contents of the present volume, we explicitly do not set out to discuss the work that has been performed by others on the use of EVB to study the kinetics of biochemical reactions, or the dynamics of proton transport in aqueous solutions, as these topics will be covered by several other chapters.

Using this narrow definition of ‘reaction dynamics’ – that is, the detailed atomic motions that take place as part of a chemical reaction – it turns out that (to our knowledge) there have been rather few studies using EVB. This is an interesting observation in itself, and we start by considering the methods that *have* been used for this purpose.

Classical or quantum studies of reaction dynamics for reactions involving small numbers of atoms have tended to use other methods to represent the potential energy surface, especially various forms of global fit of the energy surface to multiple *ab initio* energies.^[9] For systems with a small number of atoms, this is a reasonable approach, since it is possible to calculate accurate *ab initio* energies, to do so at the many points needed to survey the surface, and to fit the resulting large number of points using a flexible and accurate (but complex) set of analytical basis functions. However, for larger systems, all three aspects become more challenging, and fitting global *ab initio* potential energy surfaces is no longer a viable option. As a result, many studies of reaction dynamics tend toward so-called “direct dynamics,” where the potential energy is calculated at each time step using electronic structure theory – usually involving a computationally inexpensive method like semiempirical theory, but also Hartree-Fock or density functional theory (DFT). Apart from choice of the atomic orbital basis set, these methods have the benefit of needing no fitting, and are able to represent a “model chemistry” that encompasses a range of possible reactions within the system under study. This includes reaction pathways which are known in advance, and also those which are unknown. However, electronic structure theory is computationally demanding, especially for more accurate methods, and affordable approaches can be very inaccurate indeed, as discussed for some examples below. For the purposes of running dynamics, one has the liberty to select among a number of related methods that which best reproduces some known features of the potential energy surface. Semiempirical methods afford some flexibility insofar as it is possible to modify their parameterization in order to obtain a better agreement with benchmark values, giving rise to the so-called specific reaction parameter (SRP) semiempirical methods.^[10,11] There have similarly been attempts to formulate versions of DFT which allow one to enact a diabaticization into separate reactant and product states, with tuneable coupling interactions to bring barrier heights into better agreement with more accurate approaches.^[12]

The EVB approach represents a good alternative choice in constructing reactive potentials for systems that are too large to be exhaustively mapped out using electronic structure methods. It effectively relies on molecular mechanics force fields as the “basis functions” for representing chemical reactions, and such force fields are known to be able to capture the broad features of the potential energy surface around a given energy minimum. Since EVB is a fitted method, it can also be adjusted to reproduce exactly (or near-exactly) the energies of key stationary points on the potential energy surface. The only major disadvantage from the point of view of dynamics is that EVB potentials typically only treat reactive channels that have been explicitly included; it is not possible (usually) to “discover” new reactive channels when running dynamics, as happens with direct dynamics or in some cases with fitted potentials.

One elegant study of reaction dynamics was carried out by Chang, Minichino and Miller in 1992.^[13] The EVB method was used to construct a reactive potential for the CH₂O system, describing the minimum energy configuration, the transition state (TS), and the CO + H₂ dissociation region (neither the HCOH hydroxycarbene secondary minimum nor the TS leading to it was described). The surface was parameterized by requiring the local Hessian at the TS to match the quantum chemical Hessian. The quantum mechanical method used was either second-order Møller-Plesset perturbation (MP2) or coupled-cluster (CCSD), together with a medium-sized basis set. These electronic structure approaches yield fair agreement with the best experimental estimates

of barrier heights and reaction energy, but the EVB surfaces were ultimately tuned so as to fit the experimental values.

The fitted EVB surfaces were used to propagate quasiclassical trajectories for dissociation. Based on the target experimental observations, the initial conditions were chosen to represent CH_2O formed by internal conversion of specific S1 photoexcited states – that is, with high levels of initial rovibrational excitation. The trajectories were initiated close to the dissociation TS, in order to increase the efficiency of obtaining dissociated products, thereby making it possible to have meaningful comparisons with experimentally measured translational, vibrational and rotational energy of the H_2 and CO products. Especially with the CCSD-fitted surface, the agreement with experiment is overall rather good.

Beyond Chang *et al.*'s approach to modelling this reaction, several other quasiclassical trajectory studies have been performed.^[14–18] For example, direct dynamics studies have been performed using Hartree-Fock,^[14] MP2,^[16] DFT^[16] and SRP Hamiltonians.^[15] The semiempirical methods do not yield very good agreement with experiment; arguably they are even less good than the EVB approach, prompting the authors of that study to write that even reparameterized semiempirical Hamiltonians “might be well suited for locally interpolating ab initio data, but not for quantitatively describing global potential energy surfaces.”^[15] The MP2 and DFT studies perform reasonably well, more or less on a par with EVB. With Hartree-Fock, the relative energy of the TS is incorrect (overestimated by 26 kcal/mol), so it is necessary to express the average translational energy release as a proportion of the available energy in order to make a meaningful comparison with experiment. The EVB approach does not require this scaling, since by construction, the energetics on the surface are very good.

The benchmark potential energy surface for this system is one generated by Bowman and co-workers, which has formed the basis for a number of quasiclassical trajectory studies.^[17,18] Along with careful experimental analysis, these studies have led to the discovery of a novel mode of dissociation for formaldehyde, in which one C-H bond nearly breaks, and the corresponding H atom travels (or “roams”) around the molecule prior to “abstracting” the other H atom en route to H_2 formation. This minor pathway, which is characterized by nascent H_2 with significant internal vibrational energy excitation, was not discovered in the EVB or direct dynamics studies, in part no doubt because the corresponding regions of the potential energy surfaces were not accurate enough, and also because dissociation was initiated in those studies at the standard TS, rather than at the minimum. It is nevertheless notable that the much earlier EVB study is qualitatively reasonably accurate for describing the dynamic behaviour along the main pathway. Presumably, re-parametrization of the EVB surface could be carried out if desired to include a description of those regions of the PES which are important for the “roaming” pathway.

Another dynamical study of a gas-phase chemical reaction using an EVB-type surface was performed by Tishchenko and Truhlar,^[19] for the $\text{OH} + \text{H}_2 \rightarrow \text{H}_2\text{O} + \text{H}$ reaction. This too is a good test, since the dynamical behaviour of this reaction has been extensively studied both experimentally and theoretically. In fact, there has been even more work on this reaction than on the formaldehyde dissociation. In this case, the “EVB” potential energy surface used was in fact generated using a modification of the multi-configuration molecular mechanics (MCOMM) procedure.^[20] It was noted in the paper that the study of dynamics, as opposed to rate constants with transition state theory, provides a more stringent test of potential energy surface accuracy – an

observation which is certainly in line with our own work and discussed in further detail below. Tishchenko and Truhlar were more focussed on testing the accuracy of the potential energy representation scheme rather than on comparisons to experiment, and it was found that the MCMM approach is in fact very accurate.

To close this section, we again reiterate that there seem to be relatively few studies of reaction dynamics using EVB potential energy surfaces. For systems with small numbers of atoms, this is easy to understand, since more accurate methods exist which are able to describe several different regions of the energy surface in a balanced way, enabling very subtle aspects of dynamics to be discovered. The example of formaldehyde dissociation and roaming is a case in point. To a certain extent, this is also understandable for larger systems, in that “direct dynamics” methods have the huge advantage of being “off-the-shelf” techniques that can be applied immediately, without the need for precious coworker time to be invested in fitting of surfaces beyond that which is already baked into the basis sets.

However, we believe that the many advantages of EVB methods mean that they should be used more frequently for dynamical studies, especially for larger systems, where only qualitative or semi-quantitative accuracy is needed. A significant strength of EVB models of reactive potentials in complex systems is the fact that they build on the significant work (i.e., tens of thousands of man-hours) within the chemical molecular dynamics community that has been invested into the parameterization of molecular mechanics force fields. In what follows, we outline recent work we have recently undertaken in applying the EVB methodology to describe non-equilibrium reaction dynamics in the gas phase, and also in solution – that is, for solutes which are embedded in both weakly and strongly coupled solvents. Much of this work has been motivated by close collaboration with experimental colleagues carrying out ultrafast time-resolved infrared spectroscopy. As a result, careful comparisons with time-resolved experimental data have been an important aspect in our use of the EVB method, providing good insight into the performance of EVB methods in non-equilibrium regions of the potential energy surface in condensed phases. As our work with the EVB method has expanded to treat increasingly complex systems (particularly the solution phase dynamics results in strongly coupled solvents) by a growing user base, we have been forced to think carefully about the computational efficiency, transparency, and usability of our EVB software implementations and their ability to exploit the power of modern parallel computer architectures. For this reason, we have included within this chapter a section discussing the EVB algorithms that we have implemented as well as those which we are testing, along with measures of their performance. Readers who are interested only in the chemical applications should feel free to skip the section on algorithms; however, more expert readers may be interested in the lessons that we have learned as we have undertaken this work.

4.3 Methodology

The most common approach to EVB reactive dynamics involves a pseudo-Hamiltonian matrix $\mathbf{H}(\mathbf{q})$, which is constructed from two diabatic basis functions – a reactant (R) and product (P) function – yielding a 2×2 matrix with the following structure:

$$\mathbf{H}(\mathbf{q}) = \begin{bmatrix} V_R + \epsilon_R & H_{12} \\ H_{12} & V_P + \epsilon_P \end{bmatrix} \quad (4.1)$$

where V_R and V_P are expressions for the potential energies of the reactant and product diabatic states at a given structure, obtained from fairly standard MM force fields. The ϵ_R and ϵ_P values in Eq. (4.1) are constant diagonal energy shifts, usually chosen so as to reproduce the known exo or endo-thermicity of the reaction in question. H_{12} is an off-diagonal element that couples the reactant and product basis functions. The off-diagonal coupling elements are usually taken to be simple functions of the nuclear coordinates – in our work, typically a Gaussian function of one or two internuclear distances. \mathbf{H} is then diagonalized, that is, one finds the matrix \mathbf{U} such that:

$$\mathbf{D} = \mathbf{U}^T \mathbf{H} \mathbf{U} \quad (4.2)$$

where \mathbf{D} is a diagonal matrix, whose elements λ are the eigenvalues of \mathbf{H} . \mathbf{U} is a matrix containing the corresponding eigenvectors, \mathbf{U}_i . The adiabatic ground state energy is taken as λ_0 , the lowest eigenvalue of \mathbf{H} ; the corresponding eigenvector \mathbf{U}_0 contains the coefficients describing the extent to which each diabatic basis state (or force field) contributes to the state with energy λ_i . Application of the Hellman-Feynman relation then gives a matrix of Cartesian atomic forces \mathbf{F} :

$$\mathbf{F} = -\frac{d\mathbf{D}}{d\mathbf{q}} = -\mathbf{U}^T \frac{d\mathbf{H}}{d\mathbf{q}} \mathbf{U} \quad (4.3)$$

which contains vectors \mathbf{F}_i for the gradients of those adiabatic states which have corresponding eigenvalues in \mathbf{D} . The \mathbf{F}_0 vector contains those forces which correspond to the lowest eigenvalue λ_0 , and is used for dynamics propagation on the adiabatic ground state.

To obtain an EVB potential, the diagonal forcefields are first compared to ab initio results (or experiment) for the corresponding stable species, and if necessary, small adjustments are made in order to reproduce the structures or perhaps vibrational frequencies. Then, the behavior of the diagonal forcefields (shifted appropriately to take reaction energy into account) along the reaction coordinate is assessed, and

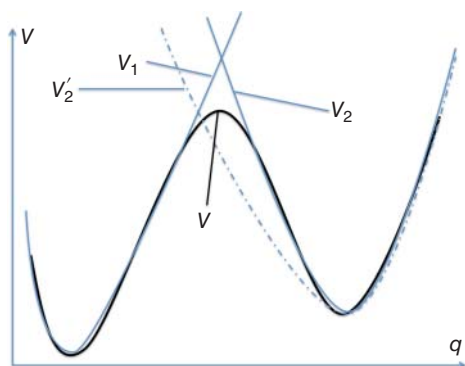


Figure 4.1 Schematic Representation of an EVB Reactive Potential Energy Surface, V , Obtained From Two Diabatic States V_1 and V_2 Corresponding to Reactant and Product States. V is very similar to V_1 for values of the coordinates lying in the reactant region, and to V_2 in the product region. An example of an inappropriate product diabatic state function, V'_2 , that lies lower than the target potential energy for some values of q , is also shown as the dash-dotted line (---). (See color plate section for the color representation of this figure.)

compared to ab initio data. As shown in Figure 4.1, the effect of diagonalization of the EVB pseudo-Hamiltonian can only be to *lower* the energy relative to the lowest of the diagonal state energies. Hence if one or more of the diagonal states is significantly lower in relative energy than the ab initio data in one section of the reaction coordinate, as for the case of V'_2 shown in the schematic Figure 4.1, then it will be impossible to get a good fit of the potential energy surface. Note also that in Figure 4.1, V_1 and V_2 are shown to cross each other rather close (both in terms of energy and in terms of the reaction coordinate) to the maximum of the ab initio energy profile. Figure 4.1 highlights an ideal scenario for obtaining a good fit to the ab initio data in the EVB model. Appropriate choice of the dependence of the off-diagonal coupling matrix element on coordinates can however in principle lead to a good fit even when the crossing occurs away from the maximum.

In this respect, it should be noted that along a reaction coordinate with bond-making and/or breaking, one or both of the diagonal states becomes severely anti-bonding in the region corresponding to the minimum of the other state. Consider a case where the states involve atom or group transfer, with V_1 corresponding to an X-Y...Z connectivity and V_2 corresponding to X...Y-Z connectivity. For a structure at which V_1 is low in energy, because $r(X-Y)$ is small and $r(Y-Z)$ is large, V_2 will be high in energy due to unfavourable bond-stretching (large $r(Y-Z)$, far from equilibrium), and due to unfavourable close approach of two atoms that are considered to be non-bonded in this state (X and Y). The very steep behavior of typical non-bonded potentials (such as the Lennard-Jones 12-6 potential used in many forcefields) with respect to interatomic distance means that this second factor is typically dominant, accounting (in the case considered) for a very high relative energy for V_2 . We have found that it is often necessary to turn off non-bonded energy terms (of the type Y-Z for V_1 and X-Y for V_2) to obtain accurate EVB potentials.

Once it has been established that the diagonal forcefields produce an acceptable description, the off-diagonal terms can be fitted to the target ab initio data, usually through application of a non-linear least-squares protocol. In many cases, even a constant off-diagonal element can produce qualitatively reasonable results. In order to avoid persistent coupling away from the intended coupling region, the off-diagonal element can be obtained as a Gaussian function (or a sum of two or more Gaussian functions) of one or more of the key interatomic distances, and this is the approach we have mostly used to date. For a typical "one-dimensional" Gaussian function (i.e., one which depends on one interatomic distance, or on the difference between two such distances), there are three parameters to be fitted: the amplitude, centre, and width. This means that only a small amount of ab initio data is needed in order to obtain a numerically meaningful fit. Even with a sum of two Gaussian functions, there are only six parameters to be fitted. Typically, we use data along the minimum energy reaction path, as well perhaps as a few points displaced from the reaction path (e.g., for an atom abstraction reaction $X-Y \cdots Z \rightarrow X \cdots Y-Z$, a grid of energies for different values of $r(X-Y)$ and $r(Y-Z)$ could be used). A typical fitting set in our work might contain of the order of 50-200 structures and relative energies, though of course more data can always be incorporated into the fitting set. We note that using higher-dimensional Gaussian functions for the coupling matrix element is possible,^[13] but this requires caution as the corresponding element can remain quite large even quite far from the reference structure.

Some of the applications discussed in the remainder of this chapter concern reaction dynamics in the gas phase; in this case, the *ab initio* calculations used to fit the EVB potential are of course carried out in vacuum. However, our primary focus – and the applications which initially led us to explore the EVB methodology – concern reactions in solution. Some of the reactions which we have studied involve relatively non-polar solutes and relatively weakly-coupled solvents.^[21,22] In such cases, it has seemed reasonable to us to fit the EVB potential in the gas phase, based on gas-phase *ab initio* data. The solute-solvent terms are then simply added in the appropriate way to the diagonal potential energy terms within the framework set up in the corresponding force-field, with the overall diagonal element obtained as a sum of “solute” (meaning the reacting sub-system) terms, solvent terms, and solute-solvent terms. The system was then treated with periodic boundary conditions, and a cut-off for interactions at long range. In a typical forcefield approach, the intra-solvent terms will be the same for all diagonal elements, though obviously the solute terms are very different. Even the solute-solvent terms may be somewhat different, due to different non-bonding interactions (electrostatic, where used, and van der Waals). We have assumed that the off-diagonal terms are not modified by the solvent, an approximation that has been shown elsewhere to be accurate.^[23]

We have also studied systems with stronger solute/solvent coupling. One of the systems, which is described in further detail below, features a more polar solvent (acetonitrile), together with a fairly polar product (hydrofluoric acid, obtained from fluorine atoms by hydrogen atom abstraction from solvent). In this system, we faced the problem of needing to describe the solvent-product interaction, and we found that one convenient way to do so was by significantly expanding the EVB matrix to include states for which proton transfer has occurred between the hydrofluoric acid and the solvent.

4.4 Recent Applications

In this section, we describe some recent applications of EVB to generate reactive potential energy surfaces for the study of reaction dynamics in both solution phase and in the gas phase. The work discussed is from our groups in Bristol and Leuven.

4.4.1 Cl + CH₄ in the Gas Phase

A first simple example is the gas-phase reaction of chlorine atoms with methane to yield HCl and methyl radical. This reaction has been extensively studied experimentally, focussing on such aspects as the distribution of HCl rotational states formed. It has also been studied computationally in several groups. From the point of view of this chapter, this is not a particularly challenging system for which to construct an EVB potential: at the energy regime we are interested in, there is only one reaction possible, and the number of atoms involved is very small. Nevertheless, this system provides a good test of the accuracy of the EVB approach as we use it, since results can be compared both to experiment and to more accurate theoretical approaches. In the latter respect, we emphasize in particular the very accurate coupled-cluster based potential developed by Czako and Bowman.^[24] This is based on over 16,000 CCSD(T) energies extrapolated to the infinite basis set limit, and also includes corrections for spin-orbit coupling. The *ab initio*

data are fitted with a mean accuracy of 0.2 kcal/mol in the low-energy regime where reaction occurs. The *ab initio* calculations and the dynamical studies can be used to compare experiment and theory for this system. For example, it is useful to compare the predicted distribution of rotational states of the product HCl with the experimentally observed distribution.

To model this system, we constructed a 5×5 EVB Hamiltonian, allowing for abstraction of each of the four hydrogen atoms of the methane.^[25] Because the potential energy expressions are identical by symmetry, and only two (at most) of the diagonal potential energies are at reasonable energy at any given structure, the effort involved in fitting is similar to that used for a simpler 2×2 case. The calculations were performed using an in-house code specifically designed to carry out EVB calculations, together with the VENUS code^[26] to drive the quasiclassical trajectories. Because the potential energy routines were written specifically for this and related gas-phase problems, it was possible to write efficient serial code, that is, we did not use the MPI or other parallelization frameworks mentioned in later sections. Individual bond stretching, bending, and torsional terms, as well as non-bonded van der Waals terms, were computed, and assembled as needed for the different diagonal energy states. The off-diagonal terms were only introduced between the reactant state and each of the product states. In each case, these were expressed as Gaussian functions of the C-H and H-Cl distances for the respective hydrogen atom.

Only a small amount of *ab initio* data was necessary in order to obtain a reasonably accurate EVB surface. In fact, two surfaces were constructed, with the first using only 300 *ab initio* calculations for the reactive surface within the minimum energy path region, that is, with C_{3v} symmetry corresponding to collinear approach of the chlorine atom to one of the C-H bonds. In the second approach, a set of 500 points were sampled from reactive trajectories, and *ab initio* energies computed at these points. In this case, **no** structures with C_{3v} symmetry were considered, since such points are not visited during classical trajectories using random initial coordinates and velocities. For both surfaces, the reactant and product state molecular mechanics expressions were fitted to further *ab initio* data for the CH_4 , CH_3 and HCl molecules. Both surfaces yielded good fits to the corresponding *ab initio* data, with mean unsigned fitting errors of 0.3 and 0.6 kcal/mol respectively.

We also computed the structure of the transition state saddle-point, and found that it matched that obtained with MP2 and CCSD(T) very well. Similarly, the vibrational frequencies at the transition state were computed and compared to those obtained with CCSD(T). The agreement was again found to be very good, with frequencies on both EVB fitted surfaces differing by about 6% from the *ab initio* values on average, and by less than 25% for all modes. This is impressive, given that the EVB parameterization based on collinear input has no *ab initio* input concerning the distorted structures lying away from the minimum energy, collinear, path. It is to be noted that others^[13] have used more complex expressions for the off-diagonal coupling terms that can *guarantee* exact matching of the *ab initio* Hessian at a number of points – but this has not been used here. This means that the bending frequencies at the TS are described reasonably well only thanks to a reasonable admixture of terms such as H-C-H bending terms in the reactant state. The fact that these provide an adequate description of the TS region is a sign that the pseudo-Hamiltonian formulation of EVB incorporates the correct “physics” and hence has some predictive power.

The EVB potentials do not return perfect results for the reaction dynamics, but they do provide very good results. The HCl product is predicted to be rotationally cold, albeit slightly less so than observed either in experiment or in trajectories with the best fully ab initio potential.^[24] The slight discrepancy appears to be due to slight inaccuracies in the treatment of the potential away from C_{3v} symmetry in the transitional region between the TS and products. The predicted reaction cross-section is in good agreement with the best theory.

4.4.2 CN + *c*-C₆H₁₂ (CH₂Cl₂ Solvent)

Our first EVB potential was developed for the reaction of the cyano radical with cyclohexane, $CN + c-C_6H_{12} \rightarrow HCN + c-C_6H_{11}$ in CH₂Cl₂ solvent. This reaction was carried out by experimental colleagues using an ultra-fast pump-probe set-up, wherein the nascent HCN was detected using time-resolved infra-red spectroscopy, which could discriminate between different HCN vibrational states. Remarkably, a large fraction of the nascent HCN product appeared to form vibrationally hot (i.e., in a $\nu = 1$ excited vibrational quantum state), and then undergo subsequent cooling by energy transfer to solvent, both of which could be monitored using time-resolved spectroscopy.^[22] Surprised by the observations of relatively long-lived excited vibrational quantum states in the products, our experimental colleagues asked us whether the conclusions they had reached based on their spectral assignments could be reproduced by a molecular dynamics model.

Initially, we planned to use direct dynamics, and a gas-phase model system. This was based on previous good experience with direct dynamics methods. However, this reaction is not well suited to this approach: test calculations using high-level ab initio methods for the related abstraction of a secondary hydrogen atom from propane showed a highly exothermic reaction, with essentially no barrier, but instead a submerged plateau in the reactant region. Van der Waals attraction on first approach of reactants leads to a drop in energy, and then chemical reaction occurs with essentially no barrier. Using a variety of cheaper ab initio, DFT and semiempirical methods, we obtained a potential energy barrier that was much too high with some methods; with other methods, we failed to obtain any barrier, or indeed the plateau indicated by the benchmarks, giving energies along the reaction path that were too low. There were also problems with the calculated reaction energy, and some difficulties due to multi-reference character of the wavefunction were found in simple correlated ab initio methods. Overall, even considering the wide range of DFT functionals we tested, no suitable method could be found. The combination of wanting to get a low (or no) barrier with a plateau, a long-range attractive interaction, and a reasonable account of the exothermicity proved too demanding. The best functional considered, BB1K, was only of marginally acceptable accuracy.

Essentially, it was these difficulties with on-the-fly direct dynamics approaches that prompted us to investigate the suitability of the EVB method. Using the same benchmark CCSD(T) ab initio data, the MMFF forcefield was modified to provide an acceptable description of the reactant and product states. Then, a simple 2×2 EVB model [as in Eq. (4.1)] was constructed and fitted to the ab initio data. A good fit was obtained, and this potential was subsequently used to model the dynamics of reactive events both in the gas phase and in dichloromethane solution. The details of the simulations are of less importance in the present context, and the reader is referred to the original papers for

this aspect.^[21,22,27] The key point here is that the efficiency of the EVB potential allowed us to examine the reaction dynamics in a solvent box in a well-converged way. Two hundred and fifty reactive trajectories were sampled, based on different equilibrated initial conditions, and using a simulation system comprised of the CN radical, the cyclohexane co-reactant, and 125 dichloromethane solvent molecules, in a periodically-repeated cubic box of length 23.7 Å. Each simulation involved an equilibration phase (200 ps, 0.5 fs timestep, 4×10^5 steps) and a production phase (200 ps, 0.1 fs timestep, 2×10^6 steps). In total, this makes 6×10^8 propagation steps – which is what is required in order to obtain a converged model for the reactive dynamics.

The availability of an accurate and efficient reactive potential enabled us to carry out extensive sampling and also enabled us to perform various numerical “experiments,” in which we varied the initial conditions in order to explore the effect on the dynamics. One of the most fascinating microscopic insights that came out of the MD modelling had a direct impact on the interpretation of the experimental results: we noted that the rate of vibrational cooling of the vibrationally hot HCN seemed to be nearly an order of magnitude faster at very short times (i.e., in the immediate wake of reaction) compared to longer times. We hypothesized that this arose because transfer occurs faster to the cyclohexyl radical co-product than it did within the bulk CH_2Cl_2 solvent itself. Additional simulations were run in which the HCN and cyclohexyl radical were constrained to lie close to one another, and indeed the rate of energy transfer was extremely high – making it much faster than the energy transfer that occurred for HCN in bulk CH_2Cl_2 solvent. In interpreting their data, our experimental colleagues had reasonably assumed that energy transfer from the nascent HCN had a rate identical to that which occurs in bulk CH_2Cl_2 solvent. Our EVB MD simulations showed that the energy transfer rate out of the hot HCN is in fact time-dependent – very fast at short times, and significantly slower at long times. We used our conclusions to re-analyze the experimental data, and we found that a time-dependent energy transfer rate led to significantly better fits of the experimental data at short times, providing strong evidence for the nascent HCN product emerging from the reaction even hotter than our experimental colleagues had initially guessed.^[27] These extensive simulations, which furnished considerable microscopic insight into the measurements that could not have otherwise been obtained experimentally, were largely possible using the EVB approach. In retrospect, it is actually serendipitous that our initial investigations using direct dynamics were partly unsuccessful, since it encouraged us to explore alternative approaches in the first place.

4.4.3 CN + Tetrahydrofuran (Tetrahydrofuran Solvent)

In a closely related study, our experimental colleagues in the group of Andrew Orr-Ewing also examined the reaction of CN radical with tetrahydrofuran (THF) solvent (and its perdeuterated form), using the same UV pump/IR probe approach.^[28] In this reaction, experimental measurements of the vibrational state of the product HCN and DCN was more difficult due to overlap of vibrational bands, but it was still possible to get insight into the dynamics and subsequent vibrational cooling. Similar to the cyclohexane reaction, *ab initio* calculations were carried out, and used as a basis to construct an EVB potential describing the hydrogen atom transfer. Both abstraction from the 2- and 3- positions of THF were considered in the *ab initio* calculations, and the more exothermic reaction was found to be the one adjacent to the ether oxygen in

the 2 position. However, both reactions are very exothermic, and have very small or no potential energy barriers. The EVB model was constructed for abstraction in the 3-position.

In the case of HCN formation in cyclohexane discussed above, the timescale for vibrational cooling of HCN in the simulations was in fair agreement with the experimentally measured one, with the initially developed, MMFF-based EVB potential. With CN in THF, this was not the case, apparently due to poor description of the spectral overlap of the C-H stretch with the vibrational spectrum of the THF. Using an adjusted second-generation potential, tailored to attempt to describe the effect of solvent interaction on the C-H stretch, much better results were obtained.

4.4.4 F + CD₃CN (CD₃CN Solvent)

This case has been mentioned briefly above in the context of our description of methodology. Again, the study was motivated by collaboration with experimental colleagues studying the reaction using pump-probe experimental methods.^[29] Whereas our previous investigations were aimed at understanding thermal reaction dynamics of solutes in relatively weakly coupled solvents, this system was chosen to investigate solute dynamics where the solvent coupling was considerably stronger. UV photolysis of xenon difluoride gives rise to fluorine atoms, which then react with the solvent to produce DF. The D—F bond is much stronger than the reactant C—D bond, and therefore the reaction is considerably exothermic. Similar to gas-phase reactions of fluorine atoms, the nascent DF therefore has considerable vibrational excitation, with essentially zero product in the $\nu=0$ vibrational state, and the bulk of the population in either $\nu = 2$ or $\nu = 3$. The fate of the DF was then probed over the ps time scale using a probe IR laser pulse. The difference with respect to the cyano radical reactions (already mentioned) is that DF is quite polar, and interacts quite strongly with acetonitrile solvent molecules, as well as the CD₂CN by-product, forming a strong hydrogen bond. It was anticipated that these features would lead to interesting dynamics both for the atom abstraction reaction and the post-reaction structural adjustments and energy redistribution.

Accordingly, we constructed an EVB model for the reaction.^[29,30] As described previously, this proved more challenging than in the earlier cases mentioned here, as we required a potential that would describe both the atom abstraction (this proved relatively easy to achieve) and the hydrogen-bonding in product. This second feature is described reasonably well at least with respect to the structure and energy of the equilibrium structure through the electrostatics and van der Waals terms in the product state forcefield. Also, since DF is such a weak acid, no deuteron transfer is expected in the process, and indeed no evidence for it was observed experimentally. In fact, our CCSD(T)-F12 ab initio calculations show that there is no secondary potential energy surface minimum corresponding to a CD₃CND⁺---F⁻ arrangement of the nuclei. However, for high vibrational excitation of the DF, partial transfer to the solvent can occur. The molecular mechanics force-field does not describe this feature very well, yielding energies which are much too high for the corresponding structures compared to the ab initio results, by failing to describe the significant anharmonicity in the potential energy surface, coupling the D—F bond stretch to the CD₃CN—DF hydrogen bond.

In principle, description of this feature in the product region does not require an EVB treatment. A carefully constructed single-state molecular mechanics potential energy

surfaces could instead be used, likely requiring a polarizable forcefield to describe the required potential energy surface features. However, we elected to use an EVB approach as described earlier, adding extra EVB states corresponding to $\text{CD}_3\text{CND}^{+---}\text{F}^-$ “states” $\text{CD}_3\text{CND}^{+---}\text{F}^-$ ‘states’ of the system. This is effectively an alternative strategy for describing the significant polarizability of the DF and the acetonitrile upon close approach to one another. Specifically for this system, we treated a fluorine radical embedded in n solvent molecules with a matrix having a dimension of $(n + 2) \times (n + 2)$, with the following structure:

$$\mathbf{H} = \begin{bmatrix} V_1 + \varepsilon_1 & H_{12} & 0 & 0 & \cdots & 0 \\ H_{12} & V_2 + \varepsilon_2 & H_{23} & H_{24} & \cdots & H_{24} \\ 0 & H_{23} & V_3 + \varepsilon_3 & 0 & \cdots & 0 \\ 0 & H_{24} & 0 & V_4 + \varepsilon_4 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & H_{24} & 0 & 0 & \cdots & V_{n+2} + \varepsilon_{n+2} \end{bmatrix} \quad (4.4)$$

where the diagonal elements V_1 and V_2 correspond to the respective energies of the reactant ($\text{F} + n\text{CD}_3\text{CN}$) and product ($\text{DF} + \text{CD}_2\text{CN} + (n-1)\text{CD}_3\text{CN}$) states, respectively. Diagonal element V_3 treats a state involving proton transfer from the product DF to the CD_2CN co-product (*i.e.*, $\text{F}^- + \text{CD}_2\text{CND}^+ + (n-1)\text{CD}_3\text{CN}$), while elements $V_4 \dots V_{n+2}$ correspond to the states in which the DF transfers its proton to all possible solvent molecules (*i.e.*, $\text{F}^- + \text{CD}_3\text{CND}^+ + \text{CD}_2\text{CN} + (n-2)\text{CD}_3\text{CN}$). The off-diagonal matrix element H_{12} (modelled using a two-dimensional ellipsoidal Gaussian function of the C—D and D—F distances for the transferring deuterium atom) is responsible for coupling together reactant and product diabatic states i and j . The matrix elements between state 2 and states 3 to $n + 2$ are one-dimensional Gaussian functions, and describe the mixing of deuteron-transfer character to the product state.

The abstraction and post-reaction potentials, shown in Figure 4.2 with comparisons to the CCSD(T)-F12 results, reproduce very well the solvatochromatic shift of the infra-red stretching frequency observed between gas-phase DF, and DF within an equilibrium CD_3CN solvation environment. It also provides an excellent description of the experimentally observed abstraction dynamics, and again enables efficient reaction simulation. MD simulations using this EVB model predict a DF product formed with a very high vibrational excitation, on the order of $\nu = 2 - 3$, in essentially quantitative agreement with experimental observations. The subsequent vibrational cooling that takes place as DF’s excess energy leaks into the solvent is also observed to occur on the same timescale as that measured experimentally. Once again, the EVB approach is key to being able to sample enough dynamics to obtain well-converged statistical treatment of the reaction. Many hundreds of simulations were performed, each involving many hundreds of thousands of timesteps, which is affordable given the short time required to evaluate the EVB potential, despite its complexity in this case.

As in the case of the cyano radical reaction, one of the other benefits of the EVB approach is that it is affordable enough to enable simulation with modified physics, which provide insight into the origin of the phenomena observed. In the present case, two sets of modified simulations were key to unravelling two competing effects that influence the appearance of the DF’s transient IR spectrum. The first effect is vibrational cooling, which brings the DF from a very anharmonic region of the potential

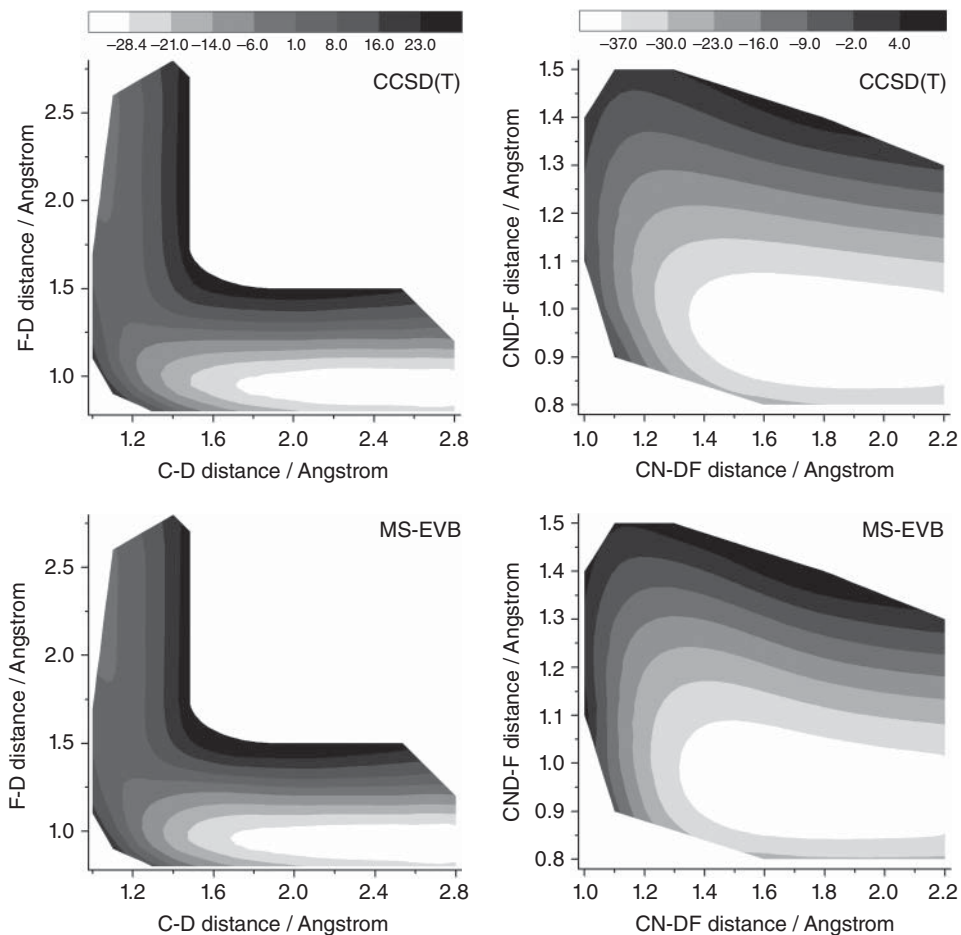


Figure 4.2 Comparison Between CCSD(T)-F12 and Multi-State EVB (MS-EVB) Potential Energy Surfaces For the $F + CD_3CN$ Reaction, Along the Reaction Profile (Left) and For the Solvent-DF Hydrogen Bonding Region (Right).

energy curve into a harmonic region of its potential energy curve, and thereby leads to an increase in the vibrational frequency (manifest as a spectral blue-shift in the transient IR spectrum of DF).

The other effect is hydrogen bond formation. In the reactants, the F atom is neutral, and interacts only weakly with solvent, perhaps through two-centre three-electron bonding with the nitrile nitrogen lone pair. Upon reaction, DF is formed, but the D is initially oriented towards the carbon atom from which it has just been abstracted. There is no possibility for hydrogen bonding to appear during the abstraction step – the nitrogen of the CD_2CN co-product radical is not in an orientation that allows it to participate in H-bonding, and neither are the other solvent molecules. Motions of the DF rapidly allow it to form H bonds, on the sub-ps timescale, and this in turn leads to a significant loosening of the D—F bond, and an associated red shift in the transient IR spectra.

Simulations with modified physics furnished the microscopic insight that allowed us to understand these red and blue shifts separately. The first key part of this analysis was the ability to carry out reactive trajectory simulations of a rather special variety – that is, modified so as to remove the excess vibrational energy of the nascent DF by suitable damping of the atomic velocities at the moment of DF formation. Hence the nascent DF is formed in the correct post-reaction solvation environment (i.e., without hydrogen bonding), but is vibrationally cold. These trajectories (with DF in post-reaction solvation environment, but with its vibrational excitation damped) were critical in examining non-equilibrium post-reaction solvation dynamics. They showed a significant *red* shift that accompanies relaxation of DF's post-reaction solvation environment in the picoseconds following reaction – providing a clear link between the observed transient DF spectral profile and the ultrafast onset of the hydrogen bond following reaction.

The second key part of our spectral analysis for this reaction involved simulations in which a DF molecule was first allowed to reach an equilibrium solvation environment, and then underwent a modification of its atomic velocities so as to inject an amount of vibrational energy similar to that yielded by the reactive trajectories. These simulations effectively amounted to a Franck-Condon excitation of DF within an equilibrium solvation environment. These simulations of non-equilibrium vibrational relaxation dynamics showed that the DF – which was both hydrogen bonded *and* vibrationally excited – did indeed cool down, at roughly the same rate as observed in the full non-equilibrium reactive dynamics simulations, and that relaxation was linked to a significant blue-shift in the calculated IR stretching band.

Using these two sets of additional simulations to analyze the full non-equilibrium reactive dynamics simulations was possible owing to the efficiency of the EVB potential. They also provided valuable microscopic insight into various reactive and solvation dynamics phenomena (and their associated timescales). Experimentally, and in the full simulations, only a modest shift in the DF infra-red absorption spectrum is observed for the nascent product, which the simulations allow us to attribute to a cancellation of the blue-shift and red-shift observed following DF formation.

4.4.5 Diazocyclopropane Ring Opening

Our final example concerns a much more complex reaction: ring-opening of a diazo-cyclopropane, through loss of nitrogen to form a carbene, which undergoes C—C bond cleavage to yield an allene (Figure 4.3). This reaction was chosen by our collaborator, Barry Carpenter, as a possible candidate for solvent-induced enantioselective amplification. The reactant has a plane of symmetry, and is hence non-chiral, but the product allene can exist as either of two enantiomers (both shown in Figure 4.3). The hypothesis (at this stage only a hypothesis, since the reaction has not been carried out experimentally) is that carrying out the ring-opening in a chiral solvent might lead to observation of the product in a chiral form, enriched in one or other of the enantiomers. This would be a dynamical effect,^[31] since the minimum energy path from reactants, over a transition state for N₂ loss, and then formation of product, starts out completely symmetric, then “splits” at a valley-ridge inflexion point^[32] into two equivalent but chiral paths leading to the two enantiomers of product.

This suggestion was initially tested through some simulations using a simple model potential,^[33] then through direct dynamics simulations with DFT.^[34] In the latter calculations, a chirally-arranged set of point charges were used to simulate the effect of a

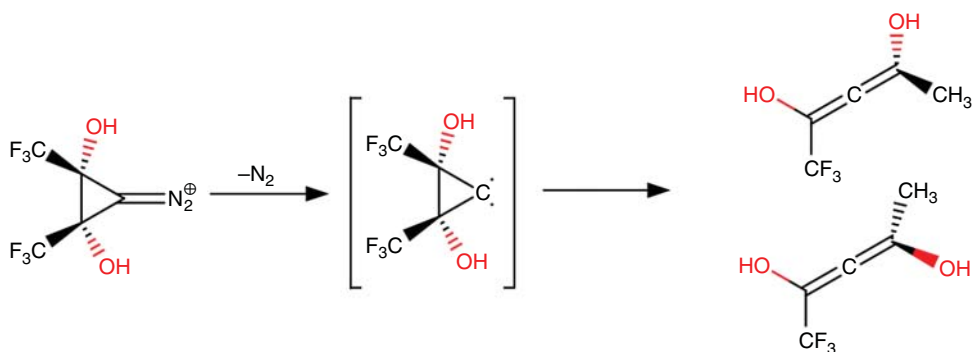


Figure 4.3 Ring-Opening of a Diazocyclopropane, Leading to Enantiomeric Products. (See color plate section for the color representation of this figure.)

chiral solvent.^[34] However, it was clearly desirable to be able to model the reaction using a full solute + solvent model. Accordingly, an EVB model was constructed, and used to explore the dynamics in explicit solvent.^[35] Intriguingly, it does appear that with some chiral solvents, a very significant excess of one enantiomer of product over the other should be obtained – and this is being further investigated.

This final example shows that the EVB approach is not limited to atom-abstraction reactions – the fact that the earlier examples we have studied have been of this type is due to the nature of the projects that our experimental collaborators have worked on as we have endeavoured to push the limits of the EVB methodology for non-equilibrium reaction dynamics, rather than an intrinsic limitation of the EVB protocol.

4.5 Software Implementation Aspects

Throughout the course of our EVB studies of fundamental reaction dynamics, increasing numbers of colleagues, collaborators, post-docs, and students have gotten involved, and there has been a steady increase in the efforts to extend the methodology to systems of increasing complexity (e.g., solution phase). Therefore, we have found that it is in our interest to build EVB frameworks which are able to benefit from the substantial research effort that has been invested in popular MD software packages and force fields developed over the last few decades, with which many workers across chemistry already have some familiarity. Some established simulation codes have EVB functionality (e.g., AMBER), but it has not been of the type that we wished to use, nor has it been sufficient to handle cases where one wishes to include large numbers of states. In our groups, we have developed three different types of code. The first – perhaps the most “bespoke” – is for the simple gas-phase $\text{Cl} + \text{CH}_4$ system, in which we have simply coded all aspects of the calculation of the diagonal and off-diagonal matrix elements of the EVB matrix, together with the diagonalization and gradient computation.^[25] Second, for the more complex solution-phase simulations, we have instead chosen to adapt existing codes to carry out EVB calculations. Third, we have recently added EVB functionality to accelerated force field routines that we have been developing over the last few years, and which are designed to run interactively for real-time steering by human users on multi-core

mobile and laptop architectures.^[36] Whereas the first code uses fairly traditional serial programming, the latter two codes have relied heavily on hardware-adapted parallelization algorithms, both as a strategy that enables existing force field codes to carry out EVB calculations in a fashion that is relatively transparent, and also to ensure computational efficiency. In the rest of this section, we give an overview of the parallelization strategies that we have implemented to accelerate our EVB MD simulations on modern computer architectures – that is, the latter two classes of code mentioned in the previous paragraph.

It needs pointing out that (by the standards of modern biomolecular simulations) the solution-phase chemical dynamics applications discussed in this chapter which we have carried out to date are rather small in terms of the number of atoms involved in each simulation: typically a few hundred. Nevertheless, they incur non-trivial computational effort, for the following reasons: (1) because we have undertaken detailed comparisons with time-resolved infrared spectroscopy measurements, we need to obtain results over a wide range of initial conditions and ensure that we have converged the simulation statistics, meaning that many trajectories are required, and (2) when large numbers of EVB states are involved – for example, in the case of F in CD₃CN solvent, the computational cost increases significantly. In an effort to speed up our EVB molecular dynamics simulations and increase their efficiency for future studies, we have experimented with different parallelization strategies over the past couple years. The first strategy, which we outline below, is based on the message-passing interface (MPI) and exploits multi-core computer processing unit (CPU) architectures. We have also investigated strategies designed to run efficiently on single-instruction-multiple-data (SIMD) stream processing architectures like graphical processing units (GPUs).

4.5.1 CPU Parallelization Using MPI

As discussed above, our production simulations for F + CD₃CN in CD₃CN solvent, which described the reactive dynamics of an F radical embedded in 62 solvent molecules, yielded a 64 × 64 $\mathbf{H}(\mathbf{q})$ matrix. The decision to utilize a 64-state matrix was determined through consideration of the number of CPU cores which we could reasonably exploit on the architectures available to us, the minimum size of the simulation required to quench DF without unduly heating the bath, and the fact that our computational resources consisted of 8-core CPU nodes. These simulations relied on the parallelized dynamics propagation strategy schematically illustrated in Figure 4.4. The propagation algorithm works by instructing each diabatic state to calculate its energy and forces in parallel as a separate thread within an MPI framework. The master thread then: (i) gathers together the results for each state to construct the diagonal matrix elements for $\mathbf{H}(\mathbf{q})$; (ii) calculates the off-diagonal matrix elements of \mathbf{H} ; (iii) undertakes the diagonalization required to solve Eq. (4.2), yielding the λ_0 eigenvalue of the \mathbf{D} matrix and its corresponding eigenvector \mathbf{U}_0 ; and (iv) solves Eq. (4.3) to yield the Hellman-Feynman force vector \mathbf{F}_0 . Once this is done, \mathbf{F}_0 and λ_0 are then dispatched to each MPI thread, overwriting the force and energy data held on each thread from the previous timestep. Each thread then propagates forward a single dynamical timestep, with the identical forces and energies ensuring that the new geometry on each thread is identical. Each thread then carries out its own energy and force calculations, the results of which are specific to the connectivity of the particular diabatic state. This

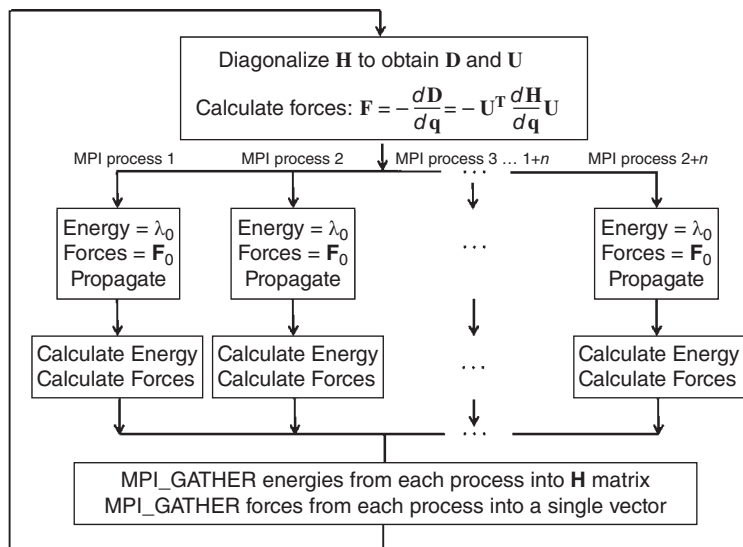


Figure 4.4 MPI State-Parallelized EVB Propagation Scheme.

parallelized propagation strategy, which we henceforth refer to as *state-parallelism*, scales nearly linearly in the number of EVB states so long as: (1) a large enough multi-core architecture is available across which to parallelize the MPI threads, and (2) the cost for any given thread to carry out a force field calculation is significantly larger than what it costs the master thread to diagonalize the Hamiltonian matrix and calculate the Hellman-Feynmann forces at each time step. Figure 4.5 shows the scaling that we see for our MPI-parallelized CHARMM implementation up to 64 cores (i.e., eight 8-core nodes), tested on the 64-state F + CD₃CN systems described above, along with analysis of the timings spent on particular computational tasks. For very large matrices, diagonalization will eventually emerge as the computational bottleneck; however, we have yet to arrive at this point for any of the examples discussed above.

Since beginning the work described in this chapter, we have implemented the state-parallelized MPI scheme shown in Figure 4.4 (and benchmarked in Figure 4.5) in both CHARMM and TINKER, allowing us access to the wide range of flexible force field methods and related molecular modeling algorithms available within both of these programs. Figure 4.5 shows that we get scaling that is close to linear as one increases the number of cores used to construct and diagonalize the 64-state $\mathbf{H}(\mathbf{q})$. The fact that we have been able to implement the Figure 4.4 scheme in well-known simulation programs highlights another key aspect of the MPI parallelization described above: because it essentially involves several very similar replicas of the same executable, relatively little modification is required to the underlying simulation code itself. The resulting MPI-parallelized EVB code rather requires only top level modifications, and ends up looking rather like a sort of parallel wrapper, in which the routines for calculating energy and forces are modified so as to include the appropriate MPI_GATHER commands, a matrix diagonalization, and a Hellman-Feynmann multiplication routine. To run an EVB calculation using this framework, the user has only to run the appropriate MPI executable, specifying the number of MPI threads to launch, the state topology for each

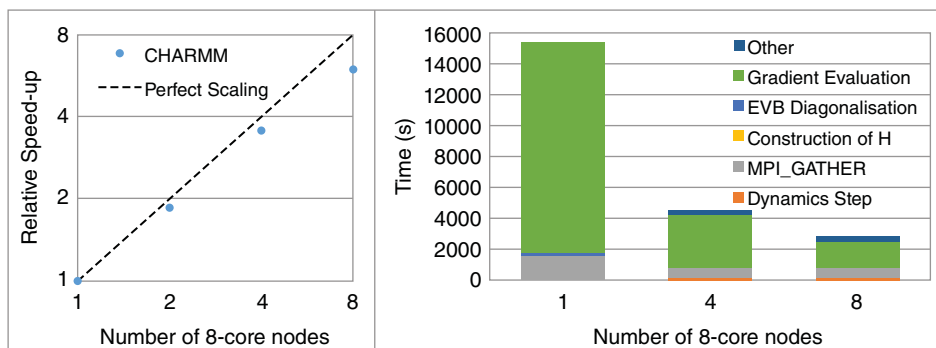


Figure 4.5 Left-Hand Panel Shows the Scaling of the State Parallelized EVB Scheme in Figure 4.4 (Tested For $F + CD_3CN$) Versus the Number of 8-Core CPU Nodes Utilized, up to a Maximum of 64 CPUs. The right-hand panel shows a decomposition of the relative timings required to carry out various computational tasks during the MS-EVB runs for the $F + CD_3CN$. (See color plate section for the color representation of this figure.)

thread, the functional form and parameters for the coupling elements, and the diagonal offsets.

4.5.2 GPU Parallelization

Modern molecular mechanics simulation codes are increasingly making use of GPUs for force field evaluation. For example, AMBER, GROMACS, and NAMD all have GPU implementations of their force-field routines, which allow the fast evaluation of force fields for very large systems, and which we henceforth refer to as *force-field parallelism*. Molecular simulation code parallelized on GPUs often ends up looking rather different than CPU-parallelized code like that discussed above, as a result of the differences in their underlying hardware architectures. CPUs are suited to heavy or complex processing of no more than a few parallel tasks: they excel at fast sequential execution of programs with complex logic. GPUs, on the other hand are designed to exploit massive parallelism in which (typically) thousands of lightweight threads (each with a small memory space) are run on thousands of cores designed to carry out efficient parallel floating-point arithmetic. They excel at multi-threaded, fine-grained, data parallel computing.

The observation that data-parallelism (i.e., performing the same operation on a large amount of data) is prevalent in many scientific algorithms including molecular dynamics, is in fact what has led to the widespread adoption of GPU-accelerated routines in scientific applications. GPUs have their origins in graphical rendering applications, where a single thread is often tasked with evaluating a relatively simple operation on a specific pixel or small subset of pixels (in an image which is potentially composed of millions of pixels). Scientific software applications often have a structure which is not too different – for example, force field evaluation potentially requires millions of identical calculations, each of which is largely independent of any another. The current trend in CPU architectures is also one of increased parallelism: this includes increasing the number of cores within a CPU, and also hyper-threading and vector parallelism within a core. Devoting time to optimally exploit parallelism in modern algorithms will be beneficial both now and in the future, as parallel computer architectures continue to evolve.

The simplest way to exploit the parallel capabilities of GPU accelerated molecular mechanics codes in efficiently evaluating EVB forces is essentially to stick with the state-parallel framework in Figure 4.4, organizing tasks such that a given state-specific force evaluation (each of which occurs on a separate MPI thread in Figure 4.4) is executed on a GPU rather than a CPU. For most force fields, the terms contributing to a given diagonal element V_i are expressed as follows:

$$V_i = V_{\text{nonbond}} + V_{\text{bond}} + V_{\text{angle}} + V_{\text{torsion}}$$

For a single state that includes N atoms, evaluation of V_{nonbond} is a formally $O(N^2)$ operation that involves computing non-bonded interaction between all pairs of atoms, while other terms typically scale linearly. The overall scaling of the evaluation of energy and gradients, in big-O notation, for a system with M states, is then $O(MN^2)$. Assuming that GPU-accelerated routines for fast force calculations within a particular state are available, then the most obvious strategy for GPU-accelerated parallelism involves sequentially evaluating V_i for each state, but with the internal force field terms for each V_i computed in parallel on the GPU.

The two parallel approaches discussed so far – *state parallelism* and *force field parallelism* – highlight the two factors at play in achieving high performance. State parallelism performs best where there are a large number of states, each with relatively few atoms. An example of this case is the F + CD₃CN work described above, where less than a thousand atoms were simulated, and sequential evaluation of the force fields was not a massively limiting factor. The force field parallelism approach would perform poorly in this case, because the number of atoms N needs to be sufficiently large to fully saturate the large number of stream processors typically available on the GPU, and sequentially computing a large number of states becomes the limiting factor on performance. Conversely, when the number of states is small, and N is sufficiently large, then the force field parallelism approach takes advantage of the large degree of parallelism offered for the computation of V_i on the GPU. The latter case is more typical of most EVB applications one finds within the literature.

Below we outline a hybrid approach incorporating aspects of both *state parallelism* and *force field parallelism*, which we have implemented on the GPU in a custom package designed to carry out real-time user guided reactive MD.^[36] While there are a number of programming languages designed to expose the parallelism of the GPU, we have used the Open Computing Language (OpenCL), which provides a hardware portable framework for writing high performance programs on GPUs and multi-core CPUs. In OpenCL, the basic unit of work is referred to as a work-item, which may be thought of as a lightweight thread performing an operation on a unique portion of data. The code for a specific work-item (i.e., a C function) is referred to as a *kernel*. Work-items are grouped into *work-groups*, each of which is computed independently of every other *work-group*. Expressing the task in this way allows one to exploit significant parallelism in accomplishing the overall task, because each work-group can be enqueued independently. Work-items with a specific work-group will typically perform the same operation on a different piece of data; this structure allows for further parallelization in the form of vector operations, which perform the same task on a contiguous array of data, and lead to considerable computational acceleration. This method for expressing parallelism works well on GPUs, where large groups of threads or vector operations are computed simultaneously.

In our hybrid approach, we constructed an OpenCL kernel for each of the force field terms included in V_i . As described above, the force field for a given state may be computed in parallel, and each term contributing to V_i may be also be computed in parallel. This provides an ample amount of parallelism which can be used to accelerate EVB force field evaluation for several coupled states; however, the parallelism is fundamentally *two-dimensional*, involving both the state index and the force field term index. OpenCL is particularly well suited to this task, because it contains within it the notion of multi-dimensional work-items – that is, a work-item may be expressed in terms of a particular state i and also to a particular term within state V_i . For example, in the calculation of V_{nonbond} a work-item will compute the Lennard-Jones force for a particular atom in a particular state. The parallelism within a state is performed on a per-atom basis – that is, a work-item for a particular atom in a particular state includes a loop for Lennard-Jones interactions between that atom and the other $N-1$ atoms in the system. This parallelization strategy involves a certain amount of redundant computation (e.g., there are $N(N-1)$ Lennard-Jones calculations rather than $N(N-1)/2$ calculations). This may seem wasteful, but for GPU architectures this actually turns out to be the more efficient strategy owing to the fact that it allows for coalesced memory access to atomic data, which is the greatest concern in achieving high-speed access to device memory. With this parallel strategy, the number of work-items is then $M \times N$, which for large N and/or M generates a significant number of lightweight threads that can be executed in parallel. This feature is what makes the approach well suited to exploiting massively parallel architectures such as GPUs. For efficiency in both computation and memory access, the forces acting on an atom for a particular state are also accumulated during the evaluation of each term in V_i .

Once the state energies and forces are computed, the state energies are copied from the GPU to the host and \mathbf{H} is constructed and diagonalized using LAPACK routines. The eigenvalues and eigenvectors of \mathbf{H} are then obtained, and the forces acting on the system may be computed. Off-diagonal terms in the \mathbf{H} matrix are presently calculated sequentially on the CPU; parallelizing their computation on the GPU would require a separate kernel for each type of off-diagonal term, and this work is in progress.

Figure 4.6 provides a schematic that illustrates the overall hybrid parallel approach that we have implemented to accelerate the EVB method on GPUs, highlighting how the notion of a multi-dimensional work-item allows one to create sufficient lightweight threads to saturate the GPU cores, and thereby most effectively exploit the parallelism available on the architecture. We benchmarked our implementation on an interactive MD simulation^[36] of OH embedded in a bath of methane (CH_4) molecules, which is able to undergo one of atmospheric chemistry's most prototypical reaction sequences: $\text{CH}_4 + \text{OH} \rightarrow \text{CH}_3 + \text{H}_2\text{O}$. Because the hydrogen atoms on each methane molecule are able to undergo abstraction by the OH radical, an increase in the number of methane molecules in the system increased the number of EVB states. The benchmarking was run on a 2014 Macbook Pro with a relatively modest NVIDIA GTX 570M GPU (336 CUDA cores). Figure 4.7 shows results obtained for two different GPU acceleration approaches – a force-field only parallelization, in which the force fields for different states are calculated sequentially, and the hybrid strategy in which parallelization is implemented across both the states and the force fields. Figure 4.7 shows timings for the construction of H using the two different approaches, and also for the diagonalization of H . The force-field only parallelization strategy exhibits poor performance

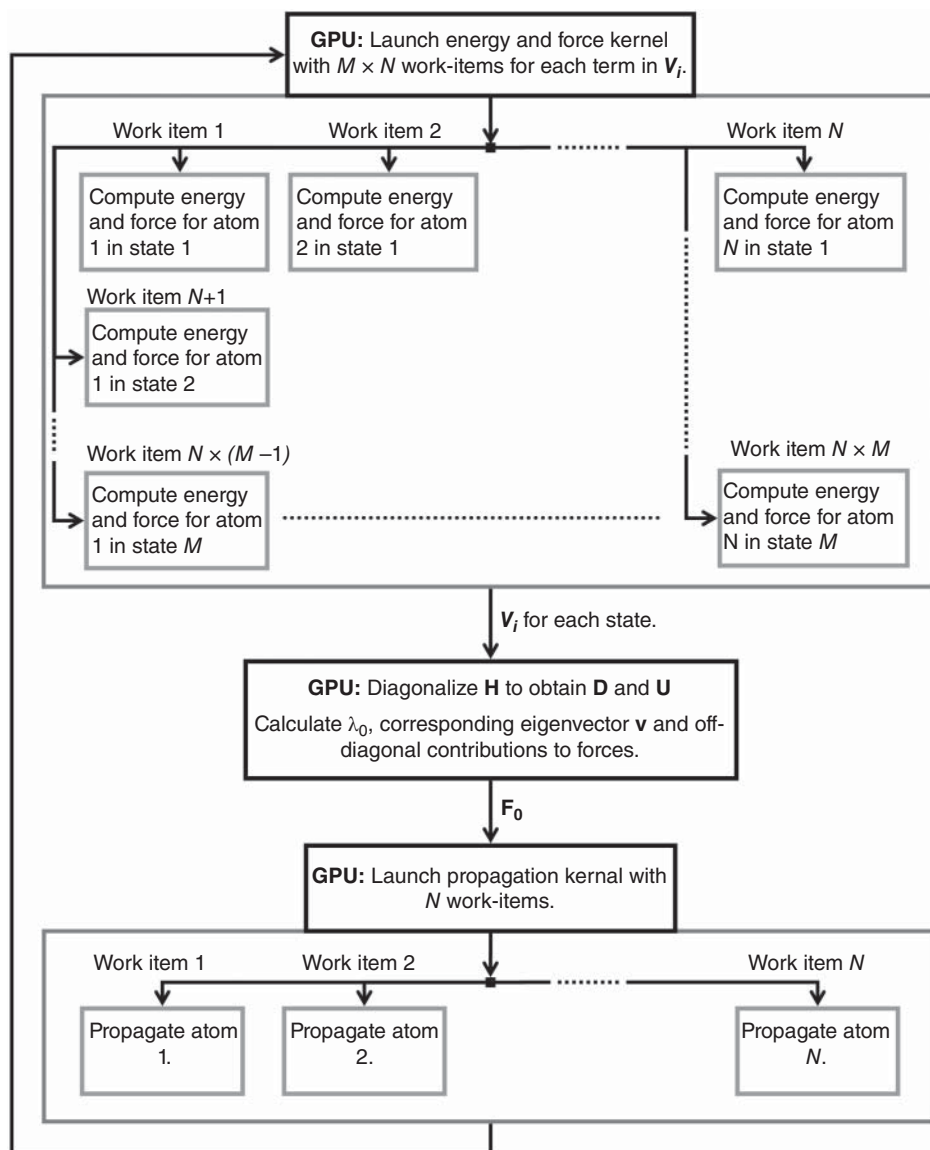
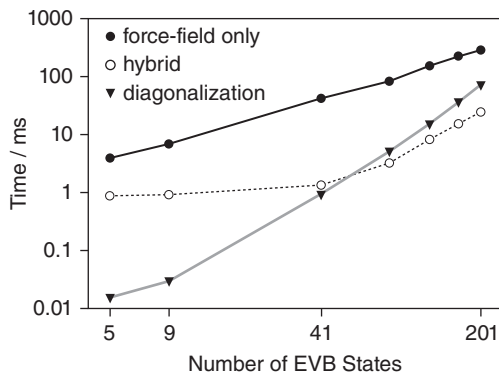


Figure 4.6 Illustration of the GPU Accelerated EVB Hybrid Parallel Framework.

for even just one OH and one CH₄ molecule, and the hybrid parallelization strategy gives a significant speed-up. Up to approximately 41 states, the hybrid implementation does not saturate the parallel cores of the GPU, and therefore scales sub-linearly. After the saturation point, the quadratic scaling of the nonbonded term is evident (with over an order of magnitude speedup compared to the force-field only parallel implementation!). Beyond 50 states, diagonalization of H , which scales as $O(N^3)$, begins to take longer than the force evaluation. The Figure 4.7 benchmark-tests utilized a serial LAPACK diagonalization algorithm; a parallelized implementation should exhibit better

Figure 4.7 Computational Scaling of the Time Required to Construct the H Matrix, Comparing the Force-Field Only Parallelization Approach with the Hybrid Parallelization Approach. The time required for matrix diagonalization is also included. Benchmarked on CH₄ + OH system on a 2014 Macbook Pro with an NVIDIA GTX 750M GPU.



performance. Because condensed-phase EVB systems typically consist of many atoms and relatively few states, the cubic scaling of matrix diagonalization does not typically become a bottleneck.

4.6 Conclusions and Perspectives

In this chapter, we have provided an overview of the use and development of EVB potentials for exploring reaction dynamics. We first compared the EVB approach to other methods to describe reactive potential energy surfaces, especially global fits based on extensive *ab initio* data, and “direct” or “on the fly” dynamics where the potential is calculated using DFT or other cheap electronic structure methods at each time step. Globally fitted potentials are able to be highly accurate, and for small systems, EVB will not generally be competitive with them for accuracy – though we discuss one example, the reaction between Cl and methane, where an EVB potential is at least similarly accurate as a global fit. For larger systems, especially for describing dynamics in condensed phases, EVB will likely carry on as an attractive option, and is much more efficient than direct dynamics. This enables a more systematic exploration of reaction dynamics, including modifying the physics of the system to attempt to unravel the impact of different effects on the observed behaviour. We also reviewed some other applications of EVB to reaction dynamics studies – though there are relatively few such studies.

Next, we described briefly the way in which we have developed EVB potentials for studies of reaction dynamics, based on *ab initio* calculations, and fitting of relatively simple functional forms for off-diagonal coupling matrix elements. We then reviewed a number of recent applications we have made of the EVB approach to describing reaction dynamics. A simple system, abstraction of hydrogen atom from methane by chlorine, is an example of a gas-phase reaction that has been studied by many other theoretical approaches. The EVB potential we developed is of good accuracy, in terms of its ability to both reproduce the reference *ab initio* data, and its ability to predict dynamical observables that agree with experiment. We next covered three other atom abstraction reactions, for abstraction of hydrogen atoms by the cyano radical from cyclohexane or tetrahydrofuran, and by the fluorine atom from acetonitrile. In each case, we carried out simulations in solution, using periodic boundary conditions within the EVB framework.

The high efficiency of the EVB framework allowed us to carry out extensive simulations, well converged in terms of the observables that were monitored by our experimental collaborators. For the reactions of cyano radical, relatively non-polar species are involved and simple 2×2 EVB models were adequate to describe the dynamics. For fluorine atom reaction in acetonitrile, the product hydrogen fluoride forms strong hydrogen bonds with the solvent, and we found that a more complex EVB model was needed, in which additional EVB states were included to describe partial proton transfer. Finally, we discuss an example of a more complex reaction, ring-opening of a diazocyclopropane in a chiral solvent.

We also discussed the algorithmic progress that we have made in developing efficient EVB codes which are adapted to modern parallel architectures. The most obvious parallel strategy (e.g., that which we have implemented in CHARMM and TINKER) utilizes *state parallelism*, and involves using MPI to parallelize force fields on different states across different CPU cores. Another approach involves *hybrid parallelism*: this is a parallel two-dimensional strategy across both the state index and force field term index. Our efforts in this algorithmic and software domain are ongoing. For example, we are investigating strategies for parallelizing across multiple GPUs on separate compute nodes. The most straightforward way to accomplish this is to use a hybrid MPI/GPU parallel framework – for example, in the case where one has access to multiple GPU compute nodes, each node is assigned the task of calculating GPU-accelerated force fields for a particular state. This will require a distributed memory approach, involving data reductions of the energies and forces across nodes in order to achieve scalable performance. We are also experimenting with identifying matrix elements of \mathbf{H} with extremely high energies (and hence contribute nearly nothing to the final eigenvector), so that they can be screened from inclusion in H , thereby reducing the total cost of force evaluation, and the cost of diagonalization.

Overall, the EVB method, which has had much success in modeling other aspects of chemical reactivity, seems very well suited to exploring reaction dynamics of quite complex systems in solution. It is efficient, so that extensive classical (or quantum, though this has not been done here) simulation is possible. Yet it is also fairly easy to parameterize EVB potentials based on limited ab initio data.

Molecular mechanics force fields continue to be improved – techniques for fitting to non-standard systems are becoming more available, accuracy is being improved, and important effects such as polarizability are now being treated increasingly frequently. As EVB is a method that builds upon molecular mechanics, this also means that EVB potentials will continue to improve. The approach we have used is but one member of the family of empirical methods based on pseudo-Hamiltonian valence-bond techniques, and the many other formalisms discussed in this volume will also continue to be successful.

In closing, it is also worth pointing out that the study of chemical reaction dynamics has traditionally focused on simple gas-phase reactions. This is not so much because only these reactions display interesting dynamical effects – but rather because these are the systems that are the most amenable to study using spectroscopic methods that are sensitive to the timescales involved. Ultrafast spectroscopy is increasingly used to study dynamics in more complex systems, and theory is often an essential partner for such studies, as the confounding effect of solvent often means that less precise experimental is available than for simpler gas-phase systems. We believe that dynamical effects

will start to become investigated for systems that are of genuine relevance to synthetic chemistry – and EVB-based theoretical methods will be ideally placed to form a strong partnership with experiment in such investigations.

References

- 1 Warshel, A. and Weiss, R.M. (1980) An empirical valence bond approach for comparing reactions in solutions and in enzymes. *Journal of the American Chemical Society*, **102** (20), 6218–6226.
- 2 The Nobel Prize in Chemistry 1986. Nobelprize.org. Nobel Media AB 2014. Web. 27 Jul 2015. Available at: http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1986/ [Accessed August 2016].
- 3 Bonnet, L. and Rayez, J.C. (1997) Quasiclassical trajectory method for molecular scattering processes: Necessity of a weighted binning approach. *Chemical Physics Letters*, **277** (1–3), 183–190.
- 4 Althorpe, S.C. (2004) The plane wave packet approach to quantum scattering theory. *International Reviews in Physical Chemistry*, **23** (2), 219–251.
- 5 Habershon, S., Manolopoulos, D.E., Markland, T.E. and Miller, T.F. (2013) Ring-polymer molecular dynamics: Quantum effects in chemical dynamics from classical trajectories in an extended phase space. *Annual Review of Physical Chemistry*, **64** (64), 387–413.
- 6 Car, R. and Parrinello, M. (1985) Unified approach for molecular dynamics and density-functional theory. *Physical Review Letters*, **55** (22), 2471–2474.
- 7 Orr-Ewing, A.J. (2015) Dynamics of bimolecular reactions in solution. *Annual Review of Physical Chemistry*, **66** (66), 119–141.
- 8 Liu, Y.P., Lu, D.H., Gonzalez-Lafont, A. *et al.* (1993) Direct dynamics calculation of the kinetic isotope effect for an organic hydrogen-transfer reaction, including corner-cutting tunneling in 21 dimensions. *Journal of the American Chemical Society*, **115** (17), 7806–7817.
- 9 Braams, B.J. and Bowman, J.M. (2009) Permutationally invariant potential energy surfaces in high dimensionality. *International Reviews in Physical Chemistry*, **28** (4), 577–606.
- 10 Greaves, S., Rose, R., Abou-Chahine, F. *et al.* (2011) Quasi-classical trajectory study of the dynamics of the $C1 + CH_4 \rightarrow HC1 + CH_3$ reaction. *Physical Chemistry Chemical Physics*, **13** (23), 11438–11445.
- 11 Troya, D. and Weiss, P.J. (2006) Ab initio and direct quasiclassical-trajectory study of the $C1 + CH_4 \rightarrow HC1 + CH_3$ reaction. *Journal of Chemical Physics*, **124** (7), 074313.
- 12 Cembran, A., Song, L., Mo, Y. and Gao, J. (2009) Block-localized density functional theory (BLDFT), diabatic coupling, and their use in valence bond theory for representing reactive potential energy surfaces. *Journal of Chemical Theory and Computation*, **5** (10), 2702–2716.
- 13 Chang, Y.T., Minichino, C. and Miller, W.H. (1992) Classical trajectory studies of the molecular dissociation dynamics of formaldehyde $H_2CO \rightarrow H_2 + CO$. *Journal of Chemical Physics*, **96** (6), 4341–4355.

- 14 Chen, W., Hase, W.L. and Schlegel, H.B. (1994) Ab initio classical trajectory study of $\text{H}_2\text{CO} \rightarrow \text{H}_2 + \text{CO}$ dissociation. *Chemical Physics Letters*, **228** (4), 436–442.
- 15 Peslherbe, G.H. and Hase, W.L. (1996) Semiempirical MNDO, AM1, and PM3 direct dynamics trajectory studies of formaldehyde unimolecular dissociation. *Journal of Chemical Physics*, **104** (20), 7882–7894.
- 16 Li, X., Millam, J.M. and Schlegel, H.B. (2000) Ab initio molecular dynamics studies of the photodissociation of formaldehyde, $\text{H}_2\text{CO} \rightarrow \text{H}_2 + \text{CO}$: Direct classical trajectory calculations by MP2 and density functional theory. *Journal of Chemical Physics*, **113** (22), 10062–10067.
- 17 Townsend, D., Lahankar, S., Lee, S. *et al.* (2004) The roaming atom: Straying from the reaction path in formaldehyde decomposition. *Science*, **306** (5699), 1158–1161.
- 18 Lahankar, S.A., Chambreau, S.D., Townsend, D. *et al.* (2006) The roaming atom pathway in formaldehyde decomposition. *Journal of Chemical Physics*, **125** (4), 044303.
- 19 Tishchenko, O. and Truhlar, D.G. (2009) Efficient global representations of potential energy functions: Trajectory calculations of bimolecular gas-phase reactions by multiconfiguration molecular mechanics. *Journal of Chemical Physics*, **130** (2), 024105.
- 20 Kim, Y., Corchado, J.C., Villa, J. *et al.* (2000) Multiconfiguration molecular mechanics algorithm for potential energy surfaces of chemical reactions. *Journal of Chemical Physics*, **112** (6), 2718–2735.
- 21 Glowacki, D.R., Orr-Ewing, A.J. and Harvey, J.N. (2011) Product energy deposition of CN+ alkane H abstraction reactions in gas and solution phases. *Journal of Chemical Physics*, **134** (21), 214508.
- 22 Greaves, S.J., Rose, R.A., Oliver, T.A. *et al.* (2011) Vibrational quantum-state-specific reaction dynamics of H atom abstraction by CN radical in solution. *Science*, **331** (6023), 1423–1426.
- 23 Hong, G., Rosta, E. and Warshel, A. (2006) Using the constrained DFT approach in generating diabatic surfaces and off diagonal empirical valence bond terms for modeling reactions in condensed phases. *Journal of Physical Chemistry B*, **110** (39), 19570–19574.
- 24 Czakó, G. and Bowman, J.M. (2011) Dynamics of the reaction of methane with chlorine atom on an accurate potential energy surface. *Science*, **334** (6054), 343–346.
- 25 Hornung, B., Harvey, J.N., Preston, T.J. *et al.* (2015) Empirical valence bond theory studies of the $\text{CH}_4 + \text{Cl} \rightarrow \text{CH}_3 + \text{HCl}$ reaction. *Journal of Physical Chemistry A*, **119**, 9590–9598.
- 26 Lourderaj, U., Sun, R., Kohale, S.C. *et al.* (2014) The VENUS/NWChem software package. Tight coupling between chemical dynamics simulations and electronic structure theory. *Computer Physics Communications*, **185** (3), 1074–1080.
- 27 Glowacki, D.R., Rose, R.A., Greaves, S.J. *et al.* (2011) Ultrafast energy flow in the wake of solution-phase bimolecular reactions. *Nature Chemistry*, **3** (11), 850–855.
- 28 Rose, R., Greaves, S., Abou-Chahine, F. *et al.* (2012) Reaction dynamics of CN radicals with tetrahydrofuran in liquid solutions. *Physical Chemistry Chemical Physics*, **14** (30), 10424–10437.

- 29 Dunning, G., Glowacki, D., Preston, T. *et al.* (2015) Vibrational relaxation and microsolvation of DF after F-atom reactions in polar solvents. *Science*, **347** (6221), 530–533.
- 30 Glowacki, D.R., Orr-Ewing, A.J. and Harvey, J.N. (2015) Non-equilibrium reaction and relaxation dynamics in a strongly interacting explicit solvent: F + CD₃CN treated with a parallel multi-state EVB model) *Journal of Chemical Physics*, **143**, 044120.
- 31 Rehbein, J. and Carpenter, B.K. (2011) Do we fully understand what controls chemical selectivity? *Physical Chemistry Chemical Physics*, **13** (47), 20906–20922.
- 32 Quapp, W., Hirsch, M. and Heidrich, D. (1998) Bifurcation of reaction pathways: The set of valley ridge inflection points of a simple three-dimensional potential energy surface. *Theoretical Chemistry Accounts*, **100** (5–6), 285–299.
- 33 Collins, P., Carpenter, B.K., Ezra, G.S. and Wiggins, S. (2013) Nonstatistical dynamics on potentials exhibiting reaction path bifurcations and valley-ridge inflection points. *Journal of Chemical Physics*, **139** (15), 154108.
- 34 Carpenter, B.K. (2014) Effect of a chiral electrostatic cavity on product selection in a reaction with a bifurcating reaction path. *Theoretical Chemistry Accounts*, **133** (8), 1–8.
- 35 Carpenter, B.K., Harvey, J.N. and Glowacki, D.R. (2015) Prediction of enhanced solvent-induced enantioselectivity for a ring opening with a bifurcating reaction path. *Physical Chemistry Chemical Physics*, **17** (13), 8372–8381.
- 36 Glowacki, D.R., O'Connor, M., Calabró, G. *et al.* (2014) A GPU-accelerated immersive audio-visual framework for interaction with molecular dynamics using consumer depth sensors. *Faraday Discussions*, **169**, 63–87.

5

Empirical Valence-Bond Models Based on Polarizable Force Fields for Infrared Spectroscopy

Florian Thauhay¹, Florent Calvo², Gilles Ohanessian¹ and Carine Clavaguéra¹

¹LCM, CNRS, Ecole polytechnique, Université Paris Saclay, Palaiseau, France

²LiPhy, Université Grenoble Alpes and CNRS, France

5.1 Introduction

Proton transfer (PT) through hydrogen bond networks plays a major role in chemical and biochemical processes.^[1] For instance, a major part of biochemical reactions operates under controlled pH conditions, because the protonation state of the chemical groups involved is essential. Furthermore, proton transport in biological media and through membranes governs cellular activity to a large extent. In addition, organization at the molecular level strongly influences physical and chemical properties of the solvent but also the solute itself. This influence can be indirect through the environmental conditions being imposed, but also more direct through the hydrogen bond network and the possible exchange of mobile protons. In recent years the number of works focused on the study of hydrogen bond networks and their capacity for charge transportation has dramatically increased.^[2–4]

Infrared (IR) spectroscopy is a powerful tool to probe the structure of hydrogen-bonded compounds, especially biomolecules. In proteins, the so-called fingerprint range (i.e., 1000–2000 cm⁻¹) is particularly informative. The amide I and II bands (corresponding to the C=O stretching and CNH bending modes, respectively) found in this region are very sensitive to hydrogen bond networks. Furthermore, charged groups such as NH₃⁺ and CO₂⁻ make strong hydrogen bonds that also give rise to characteristic bands.^[5] The best appraisal of the intrinsic impact of hydrogen bonds on vibrational frequencies can be gained in the gas phase, whereby in absence of the environment the local effects can all be appreciated. While conventional absorption spectroscopy cannot be directly used in gas phase, specific techniques such as infrared multiple photon dissociation (IRMPD) have been developed for compounds in vacuo. The IRMPD method has been found to be particularly suited to probe the structures of gaseous ions,^[6–8] providing a wealth of insight into the intrinsic properties of biomolecular ions in their local environment.

Despite the large amount of experimental data available for infrared signatures of biomolecules in the gas phase, theoretical calculations remain an essential tool to assist interpretation of spectra as they provide fundamental knowledge about the role and the relative magnitude of hydrogen bonds in such species. Usually, energetic and spectroscopic properties are obtained applying quantum mechanical (QM) methods on static

conformations. These calculations are typically (but not always) based on density functional theory (DFT), which allows systems of biological interest such as peptides of small proteins to be studied at a relatively low computational cost. However, in many situations the agreement between experimental and computed spectra is poor and corrections to account for anharmonicities (which can be noticeably important at room temperature) have to be included.^[9–13] Anharmonicities are particularly important in proton transfer, and therefore the conventional QM approach has to be overtaken in order to include dynamical effects. Several computational approaches have been previously developed to address the spectroscopic manifestations of PT.^[14–22]

Early attempts at modeling PT in molecular systems include the pioneering work of Goldblum,^[23] which relies on highly parametrized potential energy surfaces. In this work, the semi-empirical MNDO method was adapted to consider proton migration between two OH⁻ anions, using an analytical function of the two states to weight the proton evolution along the interoxygen distance.^[23] A similar two-state function was also used in the so-called “hydrogen dynamics” or HYDYN method, which allows PT processes to be modeled both with nonpolarizable and polarizable force fields. The method was used to study structural and dynamical properties of small protonated water clusters, such as Eigen and Zundel complexes.^[24] This multistate description has become widespread with the development of the empirical valence-bond (EVB) model by Warshel and coworkers.^[25,26] The EVB approach has been extended by several groups for the study of proton transport in water, most notably by Voth and coworkers^[14] and by Borgis *et al.*^[21] Using molecular dynamics (MD) simulations based on a multistate EVB (MS-EVB) model, Voth *et al.* interpreted a broad absorption as a characteristic infrared signature of the strong hydrogen bond.^[15] More recently, this approach has been extended to incorporate nuclear quantum effects in the description of proton diffusion, highlighting the concept of presolvation in which multiple weak hydrogen bonds on the hydronium ion facilitate PT.^[16] The use of EVB for exploring reaction dynamics is reviewed in Chapter 4 of this book.^[27]

With the development of first-principle molecular dynamics methods, either of the Born-Oppenheimer (BOMD) or Car-Parrinello (CPMD) types, it has become possible to explicitly model PT processes in molecular systems. These approaches combine electronic structure methods with a classical or quantum mechanical description of nuclear motion. CPMD simulations have notably been used to study PT in protonated dialanine^[17] and protonated nicotine in water.^[28] Proton transfer dynamics in the (HCO₃⁻)₂ dimer, and the corresponding vibrational spectrum, have also been studied by CPMD and path-integral molecular dynamics simulations.^[18] Quantum dynamical methods have also been employed to treat smaller size systems containing only a few tens of degrees of freedom. For instance, the multiconfiguration time-dependent Hartree (MCTDH) approach has been applied to model the Zundel cation and was found to reproduce fairly well its experimental IR spectrum,^[19] despite many Fermi resonances. The MCTDH method was also successfully applied to model the vibrational spectrum of malonaldehyde.^[20]

These calculations, although very accurate, are computationally expensive and therefore limited in size and time scales. This has motivated the use of semi-empirical methods to address the behavior of PT in water clusters using approaches such as AM1 or PM3.^[29,30] Semi-empirical methods have also been used to describe the diabatic states in EVB models.^[31] Alternatively, fully QM methods on parametrized potential energy

surfaces have been applied to model PT processes in (bio)molecules. Hybrid approaches combining classical MD and QM derived proton hopping have also been developed to simulate proton transport in extended systems.^[32] QM/MM methods have been applied to model an excess proton in water and enzymatic activity.^[33,34] Finally, reactive force fields such as ReaxFF^[35,36] (for details see Chapter 1), have also been applied to treat PT reactions in the context of heterogeneous catalysis.^[37]

In this chapter, we describe an accurate computational framework for modeling the IR spectra of biomolecules undergoing proton transfer in the gas phase. Briefly, our model combines the AMOEBA (atomic multipole optimized energetics for biomolecular simulation) polarizable force field,^[38,39] as implemented in the Tinker program,^[40] with an EVB approach. AMOEBA includes an explicit self-consistent polarization term and a multipole expansion treatment of the electrostatic effects,^[38] both ingredients appearing necessary to reproduce accurate IR spectra of anharmonic systems, especially those bearing electrically charged sites.^[41,42]

We first present the experimental motivation behind the development of our framework, which is focused on the aspartate molecule, and proceed by showing the limitations of standard QM approaches, as well as the simulations with non reactive force fields, for reproducing IR spectra. Subsequently we introduce the ingredients of the force field that allow proton exchange between the two anionic sites to be accounted for within the EVB framework. The success of the model for reproducing the experimentally measured spectrum is discussed in the light of the underlying dynamics of the proton. The dynamical and spectroscopic pictures are related to one another and evolve jointly as a function of temperature. Possible extensions and perspectives of the EVB-AMOEBA approach in the context of IR spectroscopy of biomolecules in the gas phase are finally discussed.

5.2 Infrared Spectra of Aspartate and Non-Reactive Calculations

Aspartate and glutamate are amino acids bearing a carboxylic acid function in their side chain, which at biological pH is deprotonated. As a single unit, these molecules have two carboxylate anions and a single proton that can be bound to either carboxylic group or bridging between them in a sharing configuration (Figure 5.1). Thus, they are potential candidates for a dynamically shared proton configuration in the gas phase. This was first demonstrated by Oomens *et al.*^[43] using IRMPD. Herein, we take this case to illustrate our implementation of the EVB methodology, which we have devised for simulating infrared spectra of gaseous ions. However, before discussing the performance of the EVB approach, we show how standard QM approaches compare to experimental measurements.

5.2.1 Experimental Approach

Oomens *et al.* measured the IR spectra of different amino acids using IRMPD.^[43] In the fingerprint region, sharp differences between deprotonated Trp, Tyr, Phe, Cys and Ser on one hand, and Asp and Glu on the other hand were observed.^[43] In the former case, they were assigned to carboxylate signatures near 1630 and 1330 cm^{-1} , arising from the antisymmetric and symmetric OCO stretching modes, respectively. In the second case,

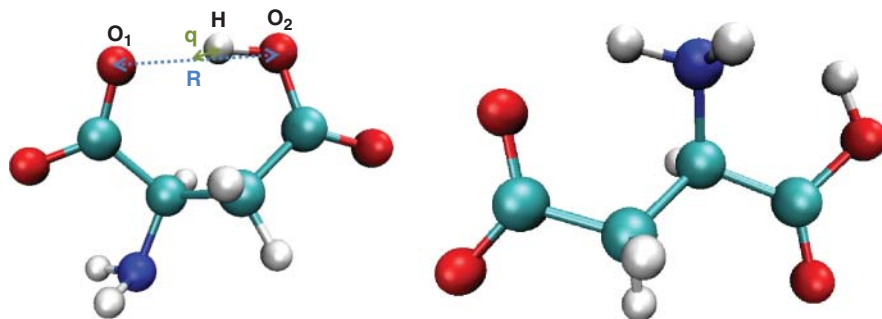


Figure 5.1 Most Stable Structures of Aspartate: Bent Conformation (Left) with Definitions of the R and q Distances, and Linear Conformation (Right). (See color plate section for the color representation of this figure.)

the Asp and Glu spectra were found to be dominated by a very broad absorption feature extending from ca. $1300\text{--}1700\text{ cm}^{-1}$, which was attributed to dynamical proton sharing between the two carboxylate groups. This unusually broad band was not reproduced by harmonic spectrum calculations. Using the same IRMPD technique, we recorded these spectra again at room temperature in the $1100\text{--}1800\text{ cm}^{-1}$ range (Figure 5.2, top), and the relatively similar experimental setup at CLIO (Centre Laser Infrarouge d’Orsay, France).

The IR spectra were recorded using a free-electron laser (FEL) coupled to a 7T FT-ICR mass spectrometer and obtained by plotting the fragmentation efficiency, F_{eff} as a function of wavelength. F_{eff} is defined as $F_{\text{eff}} = -\log[I_p/(I_p + \sum I_{\text{frag}})]$, in which I_p and I_{frag} are the parent and fragment ion intensities, respectively. By optimizing carefully experimental parameters, we were able to record a better resolved spectrum (for experimental details please, see references^[43,44]). The spectrum displayed in Figure 5.2 is clearly different from those of other deprotonated amino acids, with a main broad feature extending from 1470 to 1680 cm^{-1} , which is likely to be a signature of C=O or OCO stretching when a proton is shared between the two anions. On the red side of this band, a shoulder appears near 1450 cm^{-1} .

5.2.2 Quantum Chemical Calculations

Most of our finite temperature simulations rely on the use of the AMOEBA polarizable force field,^[38,39] which is parametrized on ab initio data. To extend this approach within the EVB framework for the study of PT processes, we aim to follow the same lines, that is, to parametrize the coupling term between the diabatic states involved in the proton sharing using electronic structure data. It is important to keep in mind that the intrinsic performance of the QM approach should be also assessed with respect to experiments and within its limits of feasibility, where assumptions such as the harmonic approximation for both the potential energy and the dipole moment surface are used. The static IR spectra of Asp were thus calculated for two different conformers using DFT and the modern M06 functional, as well as at the post-Hartree-Fock MP2 level using the Gaussian09 software package.^[45]

Among the two conformers considered for Asp, the bent conformer is the most stable structure at all levels of theory tested here (by 7.4 kcal/mol at the MP2/aug-cc-pVTZ

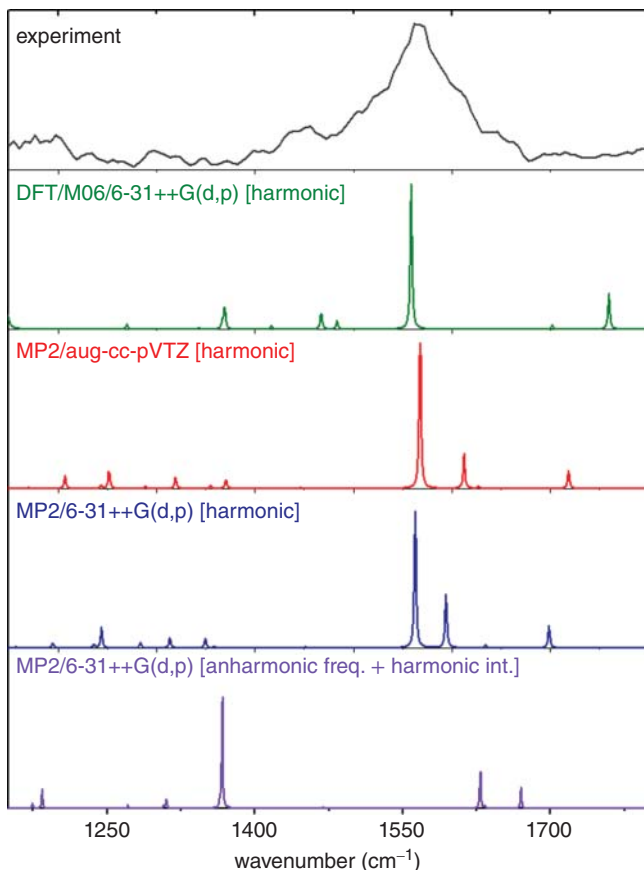


Figure 5.2 IRMPD Spectrum Measured For Aspartate and Spectrum Predicted by Static QM Calculations Using Different Methods, at the Harmonic Levels of DFT/M06/6-31++G(d,p), MP2/aug-cc-pVTZ and MP2/6-31++G(d,p), and Using Anharmonic Frequencies at the MP2/6-31++G(d,p) Level but with Harmonic Intensities.

level of theory). For this conformer, two different isomers exist, which differ in the location of the proton (see Figure 5.1). In the most stable isomer, the proton is bound to the side chain, that is, an oxygen of the COO^- group farthest away from the amine. The geometry slightly depends on the level of theory, the OO distance varying from 2.46 Å at the DFT/M06/6-31++G(d,p) level to 2.44 Å at the MP2/aug-cc-pVTZ level, the shortest distance of the proton to an oxygen being equal to 1.08 Å for both methods.

As usual after the computation of the IR spectra,^[46,47] the harmonic frequencies were scaled by a factor of 0.98 for MP2/aug-cc-pVTZ, 0.96 for MP2/6-31++G(d,p) and 1.03 for M06/6-31++G(d,p) to reproduce the intensity maximum of the experimental spectrum (Figure 5.2). As can be seen from Figure 5.2 (top), the C=O stretching modes in the harmonic spectra cannot account for the broad band experimentally observed. Furthermore, there is no computed frequency at the MP2 level that corresponds to the experimental band at 1450 cm^{-1} . These results are not improved when calculating anharmonic frequencies at the MP2/6-31++G(d,p) level of theory. Finally, the anharmonic intensities predict

very intense bands around 1300 and 1400 cm^{-1} , which overshadow the rest of the spectrum. This means that QM approaches, even when anharmonicity is included, cannot reproduce the signature of the shared proton experimentally observed in aspartate.

5.2.3 Finite Temperature IR Spectra Based on AMOEBA

5.2.3.1 The AMOEBA Force Field

A direct way to include both anharmonicities and dynamical effects in the computed IR spectrum consists of performing MD simulations. The AMOEBA polarizable force field was selected for its ability to reproduce various physicochemical properties in both the gas and condensed phases. It has been successfully used to reproduce structures and relative energies of peptides,^[38,48,49] and also to model the hydration process of various multicharged ions.^[50–53] One of the main differences with first-generation force fields such as AMBER or CHARMM lies in its more realistic description of the nonbonded intermolecular terms (*vide infra*):

$$\Delta E_{\text{int}} = \Delta E_{\text{vdw}} + \Delta E_{\text{elec}} + \Delta E_{\text{pol}}$$

In particular, the electrostatic interactions (E_{elec}) originate from the interaction of atomic multipoles (up to quadrupoles) derived from QM calculations^[38] in order to ensure a greater accuracy than the Coulomb interaction based on point charges used in standard force fields and valid at first order only. Another valuable feature of AMOEBA is the inclusion of many-body polarization effects (E_{pol}) in which the induced dipoles are treated self-consistently.^[38,54] The combination of such a refined electrostatic term and an explicit treatment of polarization effects leads to a good reproduction of the electrostatic potential and an accurate description of the hydrogen bonds both in their strength and directionality.^[41,48] The van der Waals term (E_{vdw}) takes into account exchange-repulsion and dispersion interactions between pairs of nonbonded atoms using a buffered 14-7 potential,^[55] in contrast to the 12-6 potential usually employed in standard force fields. The 14-7 form has the advantage of reproducing simultaneously series of *ab initio* results in the gas phase and liquid properties on noble gases and diatomic species. In addition, the bonding terms, consisting in bond stretchings, angle bendings, torsions and stretch-bend couplings, are all derived from the MM3 force field.^[56] Except for multipoles, we used the AMOEBAbio-09 set of parameters available in TINKER 7.^[40]

5.2.3.2 Infrared Spectra From Molecular Dynamics Simulations

As mentioned before, theoretical IR spectra are most often obtained from harmonic QM calculations. *A posteriori*, and as was actually shown in Figure 5.2, some scaling factors can be applied to partially account for anharmonicities and temperature effects. However, a more rigorous way to include these effects is to perform simulations at finite temperature MD using potential energy and dipole moment surfaces that are either classically parameterized^[41,57,58] or use an explicit description of electronic structure.^[10,12] From the trajectories the IR vibrational spectrum $I(\omega)$ is obtained by Fourier transforming the dipole moment time autocorrelation function (DACF), a standard harmonic weight being applied to yield a quantity comparable to an absorption intensity:^[10]

$$I(\omega) \propto \omega [1 - \exp(-\beta \hbar \omega)] \int e^{-i\omega t} \langle \vec{\mu}(t) \cdot \vec{\mu}(0) \rangle dt,$$

where we have denoted $\beta = 1/k_B T$ with k_B the Boltzmann constant, $\vec{\mu}(t)$ the dipole moment vector at time t , and $\langle \cdot \rangle$ an average over the time origin.

This methodology has recently been shown to yield accurate IR spectra for model peptides when compared against experimental measurements.^[41] In this example, the DACF procedure was performed as follows:

- All trajectories are initiated with an equilibration phase in the canonical ensemble to reach the desired temperature, equilibration being reached using a Nosé-Hoover thermostat. The equations of motion are integrated using the velocity Verlet algorithm;
- At the end of the equilibration period, atomic coordinates are stored and three independent simulations of 200 ps each are performed at constant total energy, again using the velocity Verlet integrator. The dipole moment vector $\vec{\mu}(t)$ is recorded every time step (0.1 fs) along the trajectory;
- The time autocorrelation function $\langle \vec{\mu}(t) \cdot \vec{\mu}(0) \rangle$ is calculated from the time series of $\vec{\mu}(t)$ and Fourier transformed to yield the infrared spectrum in the 1000–2000 cm^{-1} frequency range;
- The final IR spectrum is averaged over the contributions from independent trajectories (in this case three).

5.2.3.3 Role of the Multipoles

A key aspect of AMOEBA is the use of a set of multipoles for the electrostatic contribution to the potential energy. In many applications, such as configurational sampling, MD at thermal equilibrium or for determining solvation energies, the dependence of the computed properties on the multipolar component is usually disregarded. However, in the context of IR spectroscopy, this contribution can be important in the calculation of the dipole moment and the IR frequencies. Furthermore, the multipolar description improves the calculation of the infrared intensities, as was notably shown in the case of the Ace-Phe-NH₂ peptide.^[42]

The original version of AMOEBA suitable for biomolecules and employed here (AMOEBA_{bio-09})^[38] includes a standard multipolar set, which we denote as O. In our study we have considered new multipoles, extracted from ab initio calculations at the MP2/aug-cc-pVTZ level of theory on the MP2/cc-pVTZ geometry. The multipoles were obtained from the ab initio electron density using the distributed multipole analysis method.^[59] Furthermore, atomic dipoles and quadrupoles were refined on the electrostatic potential while keeping the charges fixed.^[39] In this optimization process, the density was evaluated around each atom with a fixed radius r usually taken as 0.65 Å for all atoms except hydrogen, for which either $r = r_H = 0.325$ Å or $r_H = 0.31$ Å have been recommended.^[41,48,60] Using these values, we have thus obtained two alternative sets of multipoles for the aspartate molecule in its most stable configuration, which we denote as set I and set II. Likewise, two corresponding sets were obtained for the other configuration in which the proton has transferred to the other carboxylate. Only one isomer exists at the MP2 level (Figure 5.1), protonation of the other site leading to the transfer back to the initial site. However, the two isomers differ when Asp is modelled using a non reactive force field because two distinct sets of parameters are required.

In order to get an energetic reference property that should be reproduced by the force field, we used the difference in MP2 single-point energies between the AMOEBA equilibrium configurations of the two isomers. The performance of these multipolar sets

Table 5.1 Relative Energies of Aspartate at the MP2/aug-cc-pVTZ Level and with the AMOEBA Force Field Using the Original Multipole Set (O) and Two Newly Proposed Versions I and II Differing by the Value of r_H Where the Reference Electrostatic Potential is Evaluated Around Each Hydrogen Atom During the Adjustment Process.

Multipolar set	original (O)	new (I)	new (II)
$r_H(\text{\AA})$	0.65	0.325	0.31
$\Delta E(\text{kcal/mol})$			
MP2/aug-cc-pVTZ	-2.29	-2.30	-2.05
AMOEBA	-11.09	-1.31	-1.15

against MP2 data is illustrated in Table 5.1, and the IR spectra obtained at 300 K with the AMOEBA force field and the three sets of multipoles are represented in Figure 5.3. Overall, the two spectra with reparametrized multipoles are very similar to each other, but they differ in the 1200–1600 cm^{-1} wavelength range from the spectrum obtained using the original set of multipoles. While these results confirm the importance of the multipolar electrostatic description on IR spectra, they do not allow for a clear distinction between the reparametrized sets I and II based on experimental comparison.

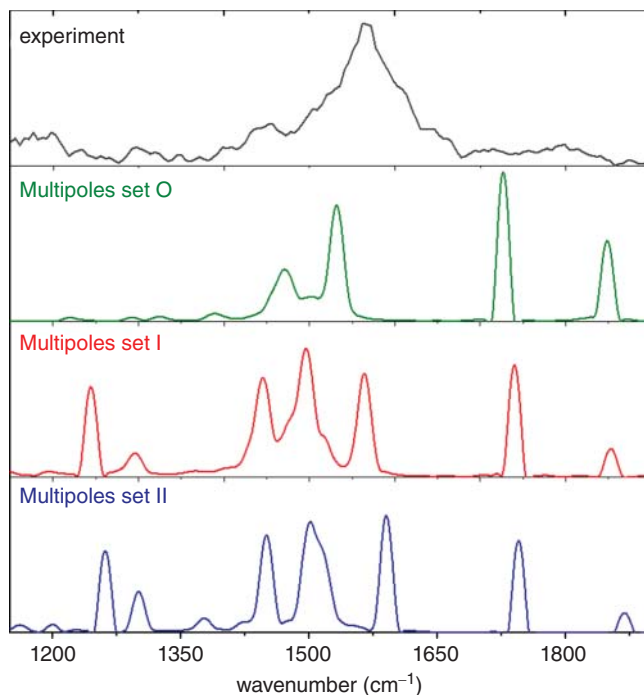


Figure 5.3 Non-Reactive AMOEBA Spectra For the Aspartate Molecule Obtained at 300 K For the Three Sets of Multipoles O, I, and II as Defined in the Text.

In the absence of spectroscopic discrimination, we turn to energetics and notice that the original set of multipoles leads to a larger error in the relative stabilities between the two isomers. In contrast, the two new sets of multipoles significantly improve this situation. Following previous studies and based on this result, we chose the set II of multipole parameters, using a value $r_H = 0.31 \text{ \AA}$ (Table 5.1).

The IR spectra obtained from AMOEBA for the most stable conformer are clearly unable to account for the experimental features in this spectral range. In particular, the C=O stretches remain excessively resolved and shifted to the blue below 1600 cm^{-1} and above 1750 cm^{-1} . Before invoking proton sharing, it is useful to consider additional possible causes for such discrepancies, namely the presence of the other isomer differing in the protonated site, as well as temperature effects that are notorious in broadening and shifting the vibrational peaks.^[61] AMOEBA simulations with multipole set II were thus repeated starting from the other conformer and at different temperatures (100, 200 and 400 K). The calculated spectra are shown in Figure 5.4. The consideration of another

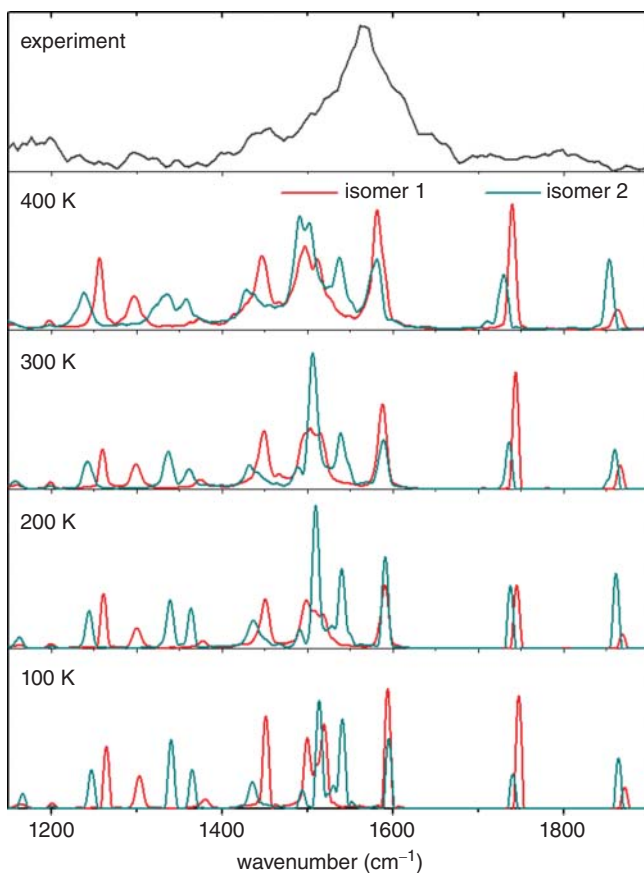


Figure 5.4 Finite Temperature Spectra Obtained From the Non-Reactive AMOEBA Model with Multipole Set II and For the Two Isomers Differing in the Protonated Site, at Four Temperatures. The experimental spectrum is shown in the upper panel.

isomer and the inclusion of temperature effects are both insufficient to reach experimental agreement.

5.3 Empirical Valence-Bond Modeling of Proton Transfer

The failure of both static QM methods and classical MD using a non-reactive force field to reproduce the experimental IR spectra suggests that alternative approaches are necessary in order to account for the essential missing feature, that is, the dynamical signature of proton sharing between the two carboxylate sites. In principle, such a property should be computed using ab initio MD^[17] or classical but reactive force fields.^[37] However, given the computational cost or the extensive parameterization used in these approaches, the EVB framework^[15,26] combined with the AMOEBA polarizable force field was chosen instead. The approach turns out to be highly accurate to reproduce the missing spectral characteristics, giving thus a great insight into their dynamical origin.

5.3.1 Two-State EVB Model

The systems under consideration here are amino acids with two carboxylic groups sharing a proton (Figure 5.1). The EVB methodology was used to represent the ground state energy of the delocalized system as the linear combination of two diabatic valence states ψ_i , each of them i having the proton only between oxygens of the two COO^- groups facing each other:

$$\Psi = g_1\psi_1 + g_2\psi_2 \quad (5.1)$$

Strictly speaking, the molecules contain four carboxylate oxygen atoms that correspond to four possible protonation sites. However, MD simulations indicate that the rotation of a COO^- group is a rare event considering the timescale of the simulations. This mere observation is confirmed by evaluating the barrier for rotation, which for the present force field is found to be 13 kcal/mol, in the same range than ab initio values. Therefore, a two-state description can be safely used.

AMOEBA is used to describe the potential energy V_i of each diabatic state in which the proton is covalently bound to either of the two oxygens participating to the exchange, and only hydrogen-bonded to the other oxygen. In matrix representation, the 2×2 Hamiltonian can be written with V_{12} being the coupling between the two diabatic states allowing proton sharing. For the present 2-state model the lowest eigenvalue of the Hamiltonian explicitly reads

$$\begin{aligned} V_{\text{EVB}} &= \frac{1}{2}[V_1 + V_2 - \sqrt{(V_1 - V_2)^2 + 4V_{12}^2}] \\ &= \frac{1}{2}(V_1 + V_2 - D), \\ \text{with } D &= [(V_1 - V_2)^2 + 4V_{12}^2]^{1/2} \end{aligned} \quad (5.2)$$

Similarly, the excited state is given by $\frac{1}{2}(V_1 + V_2 + D)$. The components (g_1, g_2) of the eigenvector corresponding to the EVB state on the (V_1, V_2) basis are as follows:

$$\begin{pmatrix} g_1 \\ g_2 \end{pmatrix} = \begin{pmatrix} [1 + (V_{\text{EVB}} - V_1)^2/V_{12}^2]^{-1/2} \\ (1 - g_1^2)^{1/2} \end{pmatrix} \quad (5.3)$$

with the eigenvector associated to the excited state being $(-g_2, g_1)$.

The trajectories on the EVB surface are generated by propagating the equations of motion using the forces deriving from the potential V_{EVB} :

$$\frac{\partial V_{\text{EVB}}}{\partial x} = \frac{1}{2} \left(\frac{\partial V_1}{\partial x} + \frac{\partial V_2}{\partial x} \right) - \frac{1}{2D} \left[(V_1 - V_2) \left(\frac{\partial V_1}{\partial x} - \frac{\partial V_2}{\partial x} \right) + 4V_{12} \frac{\partial V_{12}}{\partial x} \right], \quad (5.4)$$

which is equivalent to the expression obtained from the Hellmann-Feynman theorem:

$$\frac{\partial V_{\text{EVB}}}{\partial x} = g_1^2 \frac{\partial V_1}{\partial x} + g_2^2 \frac{\partial V_2}{\partial x} + 2g_1 g_2 \frac{\partial V_{12}}{\partial x}. \quad (5.5)$$

Finally, physical observables such as the dipole moment $\vec{\mu}$ are obtained by linear combination over their values on the diabatic states, e.g.

$$\vec{\mu}_{\text{EVB}} = g_1^2 \vec{\mu}_1 + g_2^2 \vec{\mu}_2. \quad (5.6)$$

5.3.1.1 Implementation of EVB Model with AMOEBA

The two-state EVB model used here relies on three ingredients V_1 , V_2 and V_{12} all based on analytical representations of the corresponding energies. For V_1 and V_2 , the valence terms are derived from the MM3 force field and are expressed using the following form:^[56]

$$U_{\text{short-range}} = U_{\text{bond}} + U_{\text{angle}} + U_{b\theta} + U_{\text{oop}} + U_{\text{torsion}}, \quad (5.7)$$

where the various terms on the right-hand side denote the bond stretching, bending, bond-bending coupling, out-of-plane bending, and dihedral torsion terms, respectively. In the present work, special attention was paid to the stretching potential between the proton and its covalently bound oxygen, a term which is expanded to fourth order as:

$$U_{\text{bond}}(\ell) = K_s(\ell - \ell_0)^2 [1 + C(\ell - \ell_0) + Q(\ell - \ell_0)^2], \quad (5.8)$$

and for which K_s is the stretching constant, C and Q two factors for the cubic and quartic terms, respectively. In conventional MM3 and AMOEBA these two terms adopt values chosen to match the variations of a Morse potential with the same equilibrium distance, frequency and well depth, and read $C = -2.55 \text{ \AA}^{-1}$ and $Q = (7/12)2.55^2 \text{ \AA}^{-2}$ independently of the nature of the atoms involved in the bond.^[38] This single-parameter limitation is problematic for the present situation, in which the covalent bond experienced by the sharing proton is more anharmonic than a regular bond. Hence we have considered giving more flexibility to the potential by adjusting the parameters C and Q in the specific case of the OH bonds involving the mobile proton.

5.3.1.2 Coupling Between Diabatic States

The off-diagonal coupling term V_{12} is used to represent the mixing between the two diabatic structures that is necessary to allow for proton exchange. A broad variety of expressions have been proposed in the literature to model this term, usually motivated by computational efficiency. Typically, the coupling potential for a proton shared by two atomic sites depends on the distance R between these two sites^[14,62,63] and the distance q of the proton^[62,63] to the center of the line separating these sites (Figure 5.1),

$$q = \left\| \vec{r}_{\text{H}} - \frac{1}{2}(\vec{R}_{\text{O}_1} + \vec{R}_{\text{O}_2}) \right\|.$$

In the case of protonated water, Vuilleumier and Borgis used a simple form for $V_{12} = A \exp(-\alpha R - \gamma q^2)$, with A , α and γ three parameters.^[62] For the present problem, this expression turned out to be inaccurate at distances $q > 4 \text{ \AA}$, and a more suitable form was found by still employing both R and q variables as well but a slightly different form for the off-diagonal term,

$$V_{12} = A \frac{\exp(-\alpha R)}{1 + \gamma q^2}, \quad (5.9)$$

with A , α and γ to be adjusted.

Compared to other forms available in the literature such as those advocated by Voth and coauthors,^[15,63] the present form for the coupling is deliberately simple, for two reasons. First, we mainly rely on the accuracy of the AMOEBA force field to describe the main part of the proton motion. Indeed, the role of the coupling is not to govern the simulation, but just to connect the two valence states together. It can be noted that this form does not depend on the $\widehat{O_1 O_2 H}$ angle, which could be necessary to account for the deviation from linearity of the hydrogen bond. To some extent we expect the presence of the multipoles to account for a satisfactory orientation of the bond. However, in order to further justify the expression above we have performed MP2/cc-pVTZ calculations of the $\widehat{O_1 O_2 H}$ angle around the proton-sharing configurations, varying the distance R and evaluating this angle without constraining the geometry except along this coordinate. The results of these calculations, shown in Figure 5.5, indicate that the angle is always below 10° in the relevant range of O-O distance in the 2.4-2.9 \AA range, which corresponds to an usual hydrogen bond, thereby providing a better justification for the use of R and q only as the useful coordinates.

A second justification for the relatively simple form of the off-diagonal EVB coupling elements is to be found in the hopefully better transferability to chemically similar

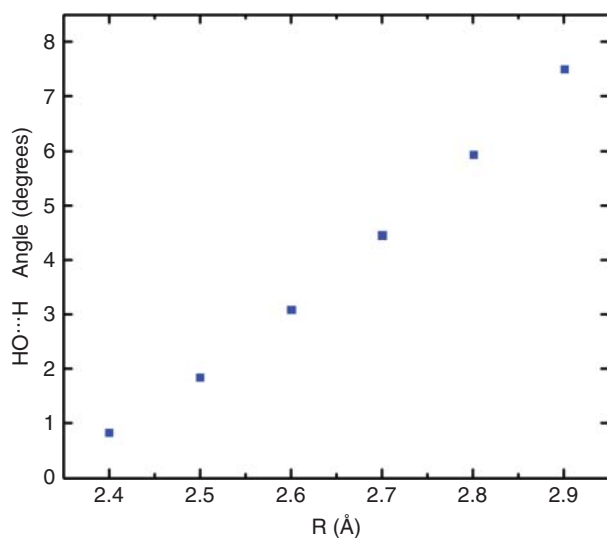


Figure 5.5 Optimal $\widehat{O_1 O_2 H}$ Angle Between the Vectors $\overrightarrow{O_1 O_2}$ and $\overrightarrow{O_2 H}$ as a Function of the $O_1 O_2$ Distance, as Obtained From MP2/cc-pVTZ Calculations Without Constraint Except on $O_1 O_2$ During the Geometry Optimization.

systems without having to go through further parametrization. We also note that no additional properties such as dissociation energies or pKa, as previously considered by several authors^[14,25,64] were included here in the adjustment process.

The parameters of the EVB model were optimized to reproduce ab initio energy surfaces initially obtained for the adipic acid $(\text{CH}_2)_4(\text{COOH})_2$, a molecule similar to aspartate,^[65] and found to be satisfactorily transferable to aspartate. The ab initio calculations employed MP2 theory and the cc-pVTZ basis set, they were limited to ranges in R and q relevant to PT between two nearby oxygen atoms, leading to $A = 420$ kcal/mol, $\alpha = 0.745 \text{ \AA}^{-1}$ and $\gamma = 5.35 \text{ \AA}^{-2}$. The EVB parameters we obtained by standard least-square fitting of the EVB-AMOEBAs potential energy surface against QM calculations, using a zero-temperature Monte Carlo procedure in the space of parameters and periodically decreasing the maximum amplitude of the random moves.

A slight further adjustment was made in the case of aspartate based on reference MP2/cc-pVTZ calculations performed by constrained geometry optimizations for each R distance, followed by a scan to fix the position of the hydrogen, leading to 96 energy points. The same geometries were kept to compute the AMOEBA energies with both V_1 and V_2 potentials for the two protonation sites, respectively. Based on these calculations the parameter A was changed to $A = 370$ kcal/mol. Moreover, the cubic parameter C of the U_{bond} term was adjusted to -2.60 \AA^{-1} in order to achieve a better curvature of the surface for larger R distances.

The performance of the EVB model for aspartate can be gauged from Figure 5.6 where several potential energy profiles are shown following the R and r_{OH} coordinates. The level of agreement is overall very satisfactory, the EVB potential appearing capable of reproducing the relative energies of the two minima and the energy barrier along the proton transfer coordinate to within 4 kcal/mol at 2.7 \AA .

This EVB model was implemented in Tinker^{7[40]} and the scheme in Figure 5.7 presents the procedure. The main challenge was to have two force fields coexist at each step of the dynamics while limiting the additional computational cost with respect to the original implementation of AMOEBA. In order to avoid I/O bottlenecks, the data necessary to compute V_1 and V_2 energies and the corresponding forces are stored in memory in the initial step. A supplementary cost is due to the parameter attribution on each atom. At each time step of the dynamics, V_1 , V_2 and their gradients are computed, followed by the calculation of the EVB coupling term and its gradient. Then, velocities and accelerations are computed to generate the next step, and the dipole moment is recorded. For the next step, the order of the two isomers forming the two states is reversed to limit the initialization procedure to one state only. For a 50 ps simulation, the EVB-AMOEBAs simulation is approximately 1.6 times longer than the standard AMOEBA simulation.

5.3.2 Dynamics Under the EVB-AMOEBAs Potential

MD simulations have been repeated with the form of the EVB potential and the associated parameters described above. The time evolution of several properties is depicted in Figure 5.8 along a typical trajectory at 300 K. In this trajectory, the proton is mainly localized over either of the carboxylate sites, as shown by the corresponding EVB weight g_i being close to 1. The analysis of the trajectory highlights that this occurs about 95% of the time. Occasionally, the proton can be trapped between the two sites, as occurs in the portion between 6 and 8 ps (Figure 5.8a). Proton sharing is manifested by the two

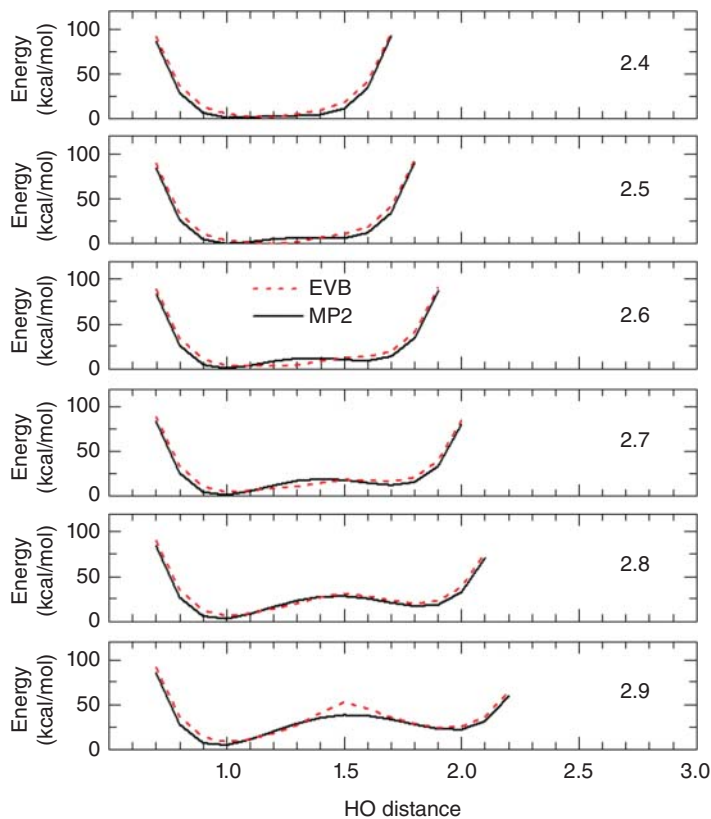


Figure 5.6 MP2/cc-pVTZ and EVB-AMOEBA Potential Energy Surfaces Obtained Along the Coordinates R and r_{OH} Causing PT Between the Two Carboxylate Sites. The Value of R is Given as an Inset in each Panel (in Å).

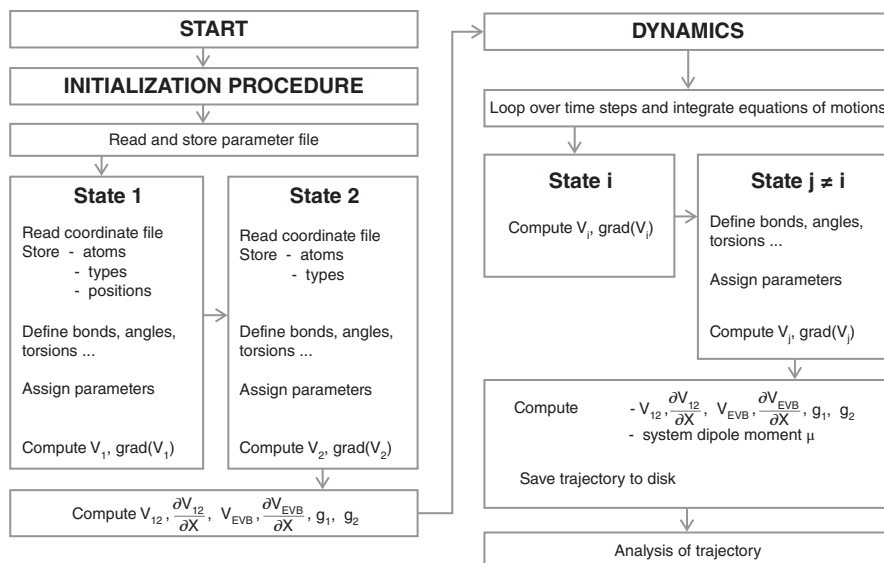


Figure 5.7 Implementation of the EVB Model in Tinker 7.

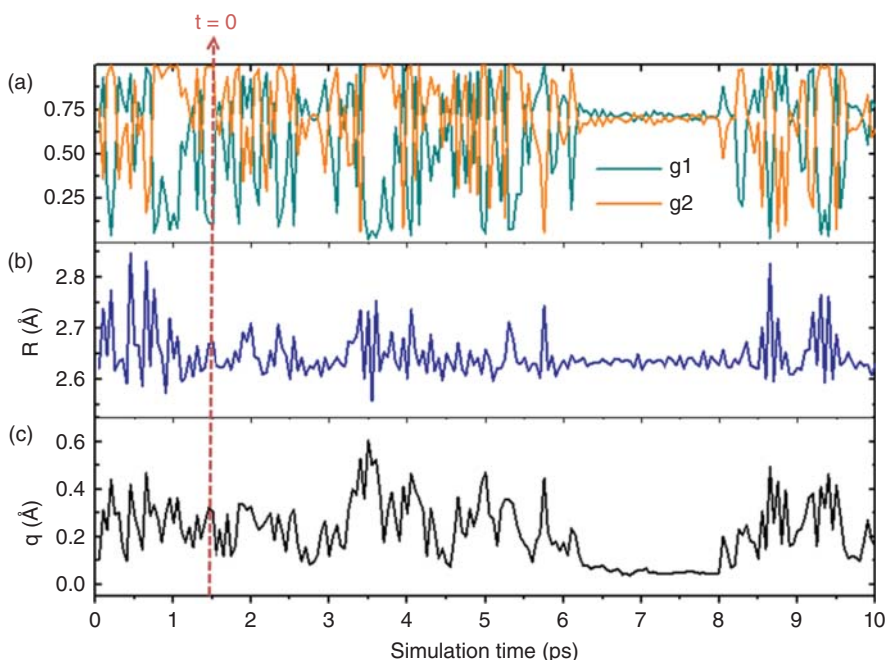


Figure 5.8 10-ps EVB-AMOEBa Simulation at 300 K, (a) Evolution of g_1 (Green), g_2 (Orange), (b) R and q Distances (in Å). Short-time averaging over 50 fs windows has been used to improve visual quality. (See color plate section for the color representation of this figure.)

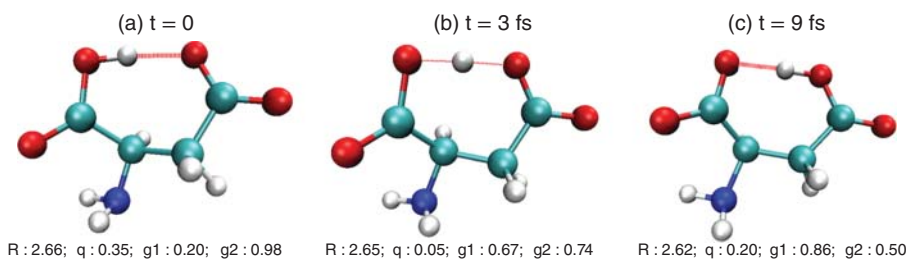


Figure 5.9 Proton Exchange During an EVB-AMOEBa MD Trajectory. The geometrical parameters R and q , the EVB weights g_1 and g_2 and the time frames are indicated. (See color plate section for the color representation of this figure.)

coefficients g_i being both close to $1/\sqrt{2}$, and under such conditions the geometric variables R and q appear to experience much smoother variations, which indicates a much more rigid conformation with the proton lying near the middle of the OO bond (small q values). The mechanism of PT is also visualized in Figure 5.9 where three snapshots of the molecule before, during and after PT are shown, as occurring after about 1.5 ps in Figure 5.8.

Looking in more details into how the geometrical parameters vary during such a reactive trajectory, the distance R between the protonated sites fluctuates between 2.5 and 2.9 Å and the proton exchange appears to be fast as the g_i coefficients oscillate between 0 and 1, unless the proton is locked in the shared position. The distributions of the R and q

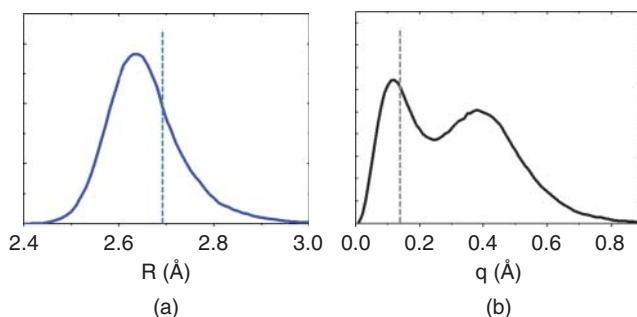


Figure 5.10 Distance Distributions For (a) R and (b) q , as Accumulated from 10 Simulations of 50 ps Each at 300 K.

distances visited during 10 EVB trajectories of 50 ps each are represented in Figure 5.10a and b, respectively. The O-O distance is spread out over and below the static equilibrium value of 2.68 Å, such a spreading being of course caused by the finite temperature of the simulation. More interestingly, the average R distance appears at a lower value closer to 2.62 Å and also corresponding to the distance of the proton-shared configuration in Figure 5.8b. The distribution of q distances appears bimodal, which is the expected behavior in presence of proton sharing. Here the distribution extends on both sides of the equilibrium which only corresponds to the proton being on the energetically lowest carboxylate site.

5.3.3 Infrared Spectra with the EVB-AMOEBa Approach

Before discussing the spectroscopic manifestations of proton sharing as predicted by the present EVB-AMOEBa model, we consider a simpler situation in which proton transfer cannot operate, that is, where the molecule lies in its linear (extended) conformation with the two carboxylate sites opposite to each other. In this case the coupling element of the EVB matrix is extremely small and the two basis functions are effectively uncoupled, the proton being stuck at its initial site. The infrared spectrum obtained at 100 K for the linear conformer of aspartate with the EVB model and the non reactive AMOEBa model applied here with the same set of multipoles are barely distinguishable from each other (data not shown). More importantly, and even considering the colder conditions of these calculated spectra, they deviate qualitatively from the experimentally measured spectrum. Incidentally this allows us to discard the linear conformer from further spectroscopic consideration.

The IR spectrum obtained for the bent conformer from the EVB-AMOEBa method under the conditions discussed in the previous section is shown in Figure 5.11 together with the experimental spectrum. We have also reported the spectra obtained from the same EVB-AMOEBa approach but with the other two sets of multipole parameters denoted as O and I, all at 300 K and using 10 independent trajectories of 50 ps each. The main change in comparison with the non-reactive spectra is that CO modes become closer, thus providing a clear signature of proton exchange with a broad band similar to the experimental one. However, the spectral features clearly exhibit differences depending on the choice of the multipolar set, and the equilibrium distributions of R and q distances obtained with the three parametrizations have been represented in Figure 5.12

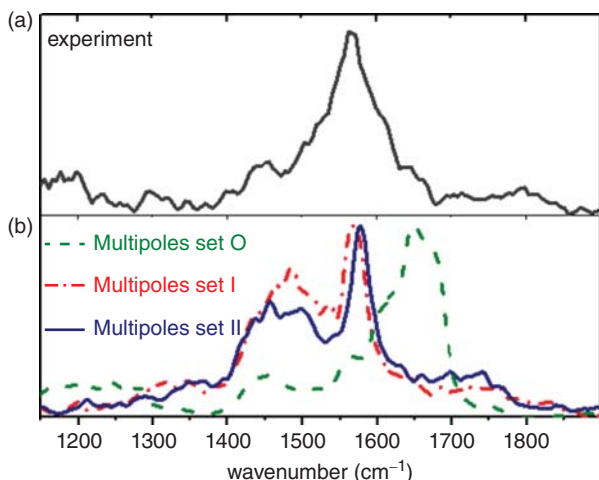


Figure 5.11 (a) Experimental (Top) and (b) Calculated IR Spectra at 300 K, as Obtained From the EVB-AMOEBa Approach with the Three Multipole Parameter Sets O, I, and II.

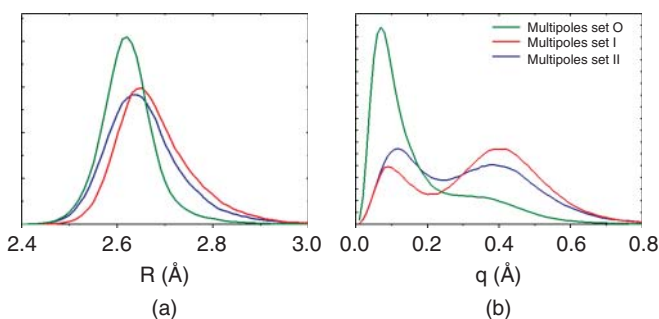


Figure 5.12 Equilibrium Distributions of (a) R and (b) q , as Obtained From the EVB-AMOEBa Simulations at 300 K Using the Three Multipolar Sets O, I, and II. (See color plate section for the color representation of this figure.)

to assist interpretation of these results. The spectrum obtained with the original set of multipoles from AMOEBaBio-09 predicts a CO stretching band excessively shifted to the blue. The distribution of q distances is largely unimodal, the shoulder on the larger values indicating minor PT.

In contrast, the multipolar sets I and II obtained from taking $r_H = 0.325 \text{ \AA}$ and $r_H = 0.31 \text{ \AA}$, respectively, in the adjustment process predict a much more balanced bimodal distribution in q , suggesting a much stronger importance of proton-exchange configurations. Both spectra match very well the intensity maximum near 1575 cm^{-1} and give also a broad contribution around 1450 cm^{-1} which is undoubtedly present in the measured spectrum although with lower relative intensity.

Having validated the present EVB-AMOEBa model for spectroscopic characterization at room temperature, we have investigated the dynamics of the system under other conditions. Temperature plays a key role in PT, and in the classical description of nuclear motion it is expected to facilitate such transfer and possibly to promote proton sharing.

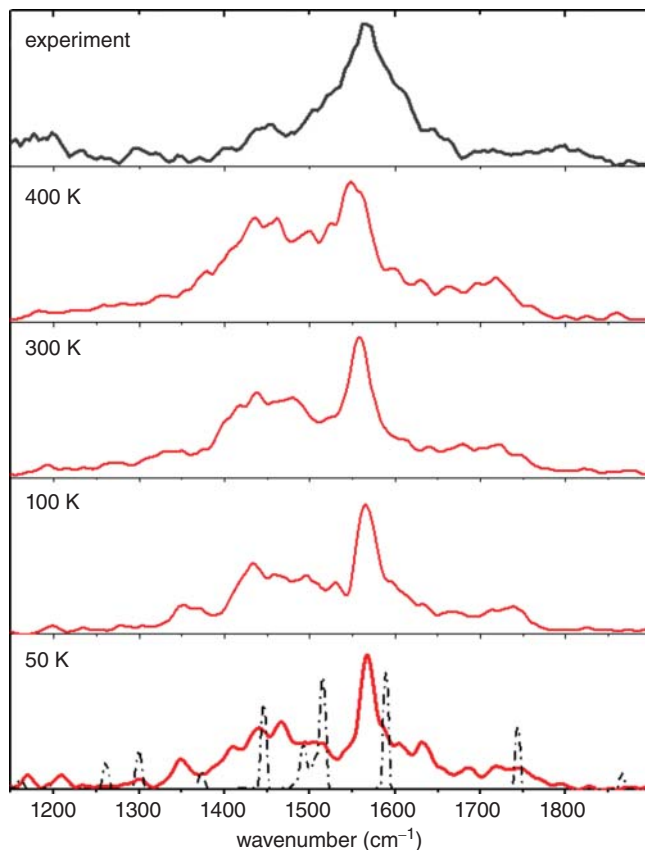


Figure 5.13 EVB-AMOEBAs Spectra as a Function of Temperature Between 50 and 400 K, with the Experimental Spectrum Being Reported Again in the Upper Panel. The non-reactive AMOEBA spectrum at low temperature is also superimposed in dot-dashed black lines.

In Figure 5.13 we have represented the IR spectra obtained from the EVB-AMOEBAs model with multipole set II at 50, 100, 300 and 400 K, all from 10 trajectories of 50 ps each. As anticipated, the spectra at low temperature appear much more resolved, with the peaks increasingly broadening and shifting as temperature increases.

At low temperature $T < 100$ K, the proton is shared between the two sites at $R = 2.6$ Å and small values of q , which can be directly visualized on the distributions of these distances in Figure 5.14a and b. To confirm that the proton is indeed shared by the two carboxylate sites we have also calculated the distributions of the EVB weights g_1 and g_2 . These are shown in Figure 5.14c and d, and clearly exhibit highly probable regions where g_1 and g_2 are both close to $1/\sqrt{2}$, indicating nearly equal delocalization on both sites.

A transition to unequally shared proton on both sides of the barrier occurs already at 100 K, where the distribution of q becomes bimodal, the EVB weights localizing on more extreme values close to 0 and 1. The spectroscopic signature of this transition is only

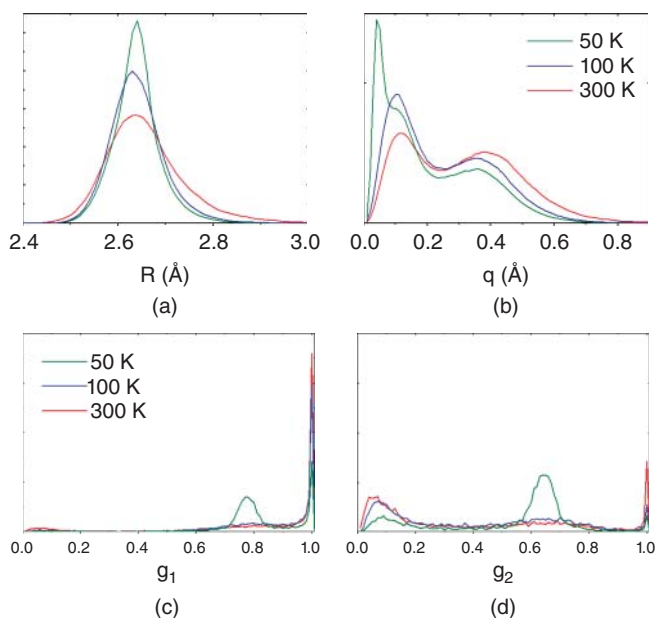


Figure 5.14 Equilibrium Distributions of (a) R and (b) q Obtained at 50, 100 and 300 K, and Distributions of the EVB Weights (c) g_1 and (d) g_2 , as Obtained at 50, 100 and 300 K. (See color plate section for the color representation of this figure.)

quite modest in Figure 5.13, the direct temperature effects being notably difficult to disentangle from the pure contributions of proton hopping versus proton sharing. At this temperature the EVB-AMOEBAs spectrum significantly differs from the non-reactive spectra obtained with AMOEBAs only, and it could not be accounted for either by superimposing the non-reactive spectra obtained with the proton on each of the sites.

As temperature increases, the proton remains in a mixed state where it constantly hops between the two carboxylate sites. The contributions to the C=O stretch modes are shifted to the red until they reach the experimental position at 300 K, before showing excessive shift and broadening above this temperature. These results show that the dynamical transfer of the proton on both sides of the barrier is essential in reproducing the experimental broadening of the IR spectrum in this range. Furthermore, and although such spectral features are intimately related to the motion of the proton between the two carboxylate sites, they do not directly provide evidence for the nature of this proton motion. Only through a detailed analysis of the trajectories, we can interpret the proton behavior as evolving from an equally shared configuration stabilized by resonance mechanism at low temperature to a constant hopping motion between the carboxylate sites through an intermediate barrier at 100 K and above. Such a qualitative change in the motion of the mobile proton originates from the thermal motion of the molecular backbone: as temperature increases, the R distance enlarges (Figure 5.14a), which results in a stronger barrier to proton delocalization (Figure 5.6) and a change in the nature of proton motion.

5.4 Concluding Remarks

In this chapter, we discussed our implementation of a two-state EVB model for the computation of IR spectra, in particular for systems containing a signature of proton motion in flexible molecules. In this model we used the AMOEBA polarizable force field to describe the potential energy of the diabatic states and a simple form of the coupling potential, following here previous works by Vuilleumier and Borgis for protonated water.^[62] As a representative example, the EVB-AMOEBA model was applied to aspartate deprotonated amino acid, for which an experimental gas-phase IR spectrum was newly recorded. The experimental spectrum provides a clear spectroscopic signature of a dynamically shared proton between two carboxylate groups, which was not possible to reproduce by standard QM and MD approaches. A careful parameterization from ab initio calculations, used to obtain the multipoles for the AMOEBA potential and the off-diagonal coupling term in the EVB model, was necessary to get a consistent dynamical picture of the proton sharing and an accurate spectral signature that agrees with experiments.

The work described in this chapter represents the first step in the extension of the polarizable AMOEBA force field to address reactivity in relation with characteristic spectroscopic features. Beyond this first successful assessment, further developments are in progress in order to extend the methodology to a general two-state model in which the protonation sites may differ, and subsequently to develop a multi-state EVB model. While there is no fundamental restriction to the application of polarizable force fields as a building block of a multistate EVB model, technicalities and the significant computational cost associated with the multipolar expansion will require efficient parallelization algorithms in order to address time scales compatible with low-frequency motions and slow rearrangements. The ultimate goal will be to extend the force field capabilities so they include proton transfer reactions in peptides and in protein active sites in relation with experimental infrared signatures. Finally, the model will be used to study ionic chemistry in large water droplets in the presence of acidic or basic impurities such as sulfuric acid, and their infrared spectroscopy.

Acknowledgements

FT is grateful to Ecole polytechnique for PhD fellowship. Implementation of EVB model was supported by travel grants from GDR 3533 EMIE funded by CNRS. Use of the FT-ICR mass spectrometer at Orsay was supported by the "Spectrométrie de masse FT-ICR à très haut champ" research infrastructure funded by the Chemistry Institute of CNRS. We are grateful to Edith Nicol for carrying out the IRMPD experiments on deprotonated Asp and to the CLIO team for support.

References

- 1 Marx, D. (2006) Proton transfer 200 years after von Grotthuss: Insights from ab initio simulations. *ChemPhysChem*, 7 (9), 1848–1870.

- 2 Tuckerman, M.E., Marx, D. and Parrinello, M. (2002) The nature and transport mechanism of hydrated hydroxide ions in aqueous solution. *Nature*, **417** (6892), 925–929.
- 3 Wraight, C.A. (2006) Chance and design – proton transfer in water, channels and bioenergetic proteins. *Biochimica et Biophysica Acta (BBA) – Bioenergetics*, **1757** (8), 886–912.
- 4 Knight, C. and Voth, G.A. (2012) The curious case of the hydrated proton. *Accounts of Chemical Research*, **45** (1), 101–109.
- 5 Semrouni, D., Balaj, O., Calvo, F. *et al.* (2010) Structure of sodiated octa-glycine: IRMPD spectroscopy and molecular modeling. *Journal of the American Society for Mass Spectrometry*, **21**, 728–738.
- 6 Maitre, P., Caër, S.L., Simon, A. *et al.* (2003) Ultrasensitive spectroscopy of ionic reactive intermediates in the gas phase performed with the first coupling of an IR FEL with an FTICR-MS. *Nuclear Instruments and Methods in Physics Research*, **507** (1-2), 541–546.
- 7 MacAleese, L. and Maitre, P. (2007) Infrared spectroscopy of organometallic ions in the gas phase: From model to real world complexes. *Mass Spectrometry Reviews*, **26** (4), 583–605.
- 8 Polfer, N.C. and Oomens, J. (2009) Vibrational spectroscopy of bare and solvated ionic complexes of biological relevance. *Mass Spectrometry Reviews*, **28** (3), 468–494.
- 9 Gageot, M.P., Vuilleumier, R., Sprik, M. and Borgis, D. (2005) Infrared spectroscopy of N-methylacetamide revisited by ab initio molecular dynamics simulations. *Journal of Chemical Theory and Computation*, **1** (5), 772–789.
- 10 Gageot, M.P. (2010) Theoretical spectroscopy of floppy peptides at room temperature. A DFTMD perspective: gas and aqueous phase. *Physical Chemistry Chemical Physics*, **12** (14), 3336–3359.
- 11 Ingrosso, F., Monard, G., Hamdi Farag, M. *et al.* (2011) Importance of polarization and charge transfer effects to model the infrared spectra of peptides in solution. *Journal of Chemical Theory and Computation*, **7** (6), 1840–1849.
- 12 Rossi, M., Blum, V., Kupser, P. *et al.* (2010) Secondary structure of Ac-Ala_n-LysH⁺ polyalanine peptides (n=5, 10, 15) in vacuo: Helical or not? *Journal of Physical Chemistry Letters*, **1** (24), 3465–3470.
- 13 Thomas, M., Brehm, M., Fligg, R. *et al.* (2013) Computing vibrational spectra from ab initio molecular dynamics. *Physical Chemistry Chemical Physics*, **15** (18), 6608–6622.
- 14 Lobaugh, J. and Voth, G.A. (1996) The quantum dynamics of an excess proton in water. *Journal of Chemical Physics*, **104** (5), 2056–2069.
- 15 Schmitt, U.W. and Voth, G.A. (1999) The computer simulation of proton transport in water. *Journal of Chemical Physics*, **111** (20), 9361–9381.
- 16 Biswas, R., Tse, Y.L.S., Tokmakoff, A. and Voth, G.A. (2016) Role of presolvation and anharmonicity in aqueous phase hydrated proton solvation and transport. *Journal of Physical Chemistry B*, **120** (8), 1793–1804.
- 17 Marinica, D.C., Grégoire, G., Desfrancois, C. *et al.* (2006) Ab initio molecular dynamics of protonated dialanine and comparison to infrared multi-photon dissociation experiments. *Journal of Physical Chemistry A*, **110** (28), 8802–8810.

- 18 Dopieralski, P.D., Latajka, Z. and Olovsson, I. (2010) Proton-transfer dynamics in the HCO₃ dimer of KHCO₃ from Car-Parrinello and path- integrals molecular dynamics calculations. *Acta Crystallographica Section B*, **66** (2), 222–228.
- 19 Vendrell, O., Gatti, F. and Meyer, H.D. (2007) Dynamics and infrared spectroscopy of the protonated water dimer. *Angewandte Chemie International Edition*, **46** (36), 6918–6921.
- 20 Hammer, T. and Manthe, U. (2011) Intramolecular proton transfer in malonaldehyde: Accurate multilayer multi-configurational time-dependent Hartree calculations. *Journal of Chemical Physics*, **134** (22), 224305.
- 21 Vuilleumier, R. and Borgis, D. (1999) Transport and spectroscopy of the hydrated proton: A molecular dynamics study. *Journal of Chemical Physics*, **111** (9), 4251–4266.
- 22 Marciante, M. and Calvo, F. (2014) Modelling infrared action spectra of protonated water clusters. *Molecular Simulation*, **40** (1-3), 176–184.
- 23 Goldblum, A. (1988) Calculation of proton transfers in hydrogen bonding interactions with semi-empirical MNDO/H. *Journal of Molecular Structure: THEOCHEM*, **179** (1), 153–163.
- 24 Wolf, M.G. and Groenhof, G. (2014) Explicit proton transfer in classical molecular dynamics simulations. *Journal of Computational Chemistry*, **35** (8), 657–671.
- 25 Warshel, A. and Weiss, R.M. (1980) An empirical valence bond approach for comparing reactions in solutions and in enzymes. *Journal of the American Chemical Society*, **102** (20), 6218–6226.
- 26 Warshel, A. (2002) Molecular dynamics simulations of biological reactions. *Accounts of Chemical Research*, **35** (6), 385–395.
- 27 Harvey, J.N., O'Connor, M., Glowacki, D.R. (2017) Theory and applications of the empirical valence bond approach: from physical chemistry to chemical biology (John Wiley & Sons, Ltd.). chap. Empirical Valence Bond Methods for Exploring Reaction Dynamics in the Gas Phase and in Solution.
- 28 Gaigeot, M.P., Cimas, A., Seydou, M. *et al.* (2010) Proton transfer from the inactive gas-phase nicotine structure to the bioactive aqueous-phase structure. *Journal of the American Chemical Society*, **132** (51), 18067–18077.
- 29 Wang, S., MacKay, L. and Lamoureux, G. (2014) Development of semiempirical models for proton transfer reactions in water. *Journal of Chemical Theory and Computation*, **10** (8), 2881–2890.
- 30 Wu, X., Thiel, W., Pezeshki, S. and Lin, H. (2013) Specific reaction path Hamiltonian for proton transfer in water: Reparameterized semiempirical models. *Journal of Chemical Theory and Computation*, **9** (6), 2672–2686.
- 31 Vener, M., Rostov, I., Soudackov, A. and Basilevsky, M. (2000) Semiempirical modeling free energy surfaces for proton transfer in polar aprotic solvents. *Chemical Physics*, **254** (2-3), 249–265.
- 32 Lill, M.A. and Helms, V. (2001) Molecular dynamics simulation of proton transport with quantum mechanically derived proton hopping rates (Q-HOP MD). *Journal of Chemical Physics*, **115** (17), 7993–8005.
- 33 Pezeshki, S. and Lin, H. (2015) Adaptive-partitioning QM/MM for molecular dynamics simulations: 4. Proton hopping in bulk water. *Journal of Chemical Theory and Computation*, **11** (6), 2398–2411.

- 34 Riccardi, D., Yang, S. and Cui, Q. (2010) Proton transfer function of carbonic anhydrase: Insights from QM/MM simulations. *Biochimica et Bio-physica Acta (BBA) – Proteins and Proteomics*, **1804** (2), 342–351.
- 35 van Duin, A.C.T., Dasgupta, S., Lorant, F. and Goddard-III, W.A. (2001) ReaxFF: A reactive force field for hydrocarbons. *Journal of Physical Chemistry A*, **105** (41), 9396–9409.
- 36 Nagy, T., Meuwly, M. (2017) Theory and applications of the empirical valence bond approach: from physical chemistry to chemical biology (John Wiley & Sons, Ltd.), chap. Modelling Chemical Reactions Using Empirical Force Fields.
- 37 van Duin, A.C.T., Zou, C., Joshi, K. *et al.* (2014) A Reaxff Reactive force-field for proton transfer reactions in bulk water and its applications to heterogeneous catalysis, in *Computational catalysis*, The Royal Society of Chemistry, Chapter 6 pp. 223–243.
- 38 Ponder, J.W., Wu, C., Ren, P. *et al.* (2010) Current status of the AMOEBA polarizable force field. *Journal of Physical Chemistry B*, **114** (8), 2549–2564.
- 39 Ren, P., Wu, C. and Ponder, J.W. (2011) Polarizable atomic multipole- based molecular mechanics for organic molecules. *Journal of Chemical Theory and Computation*, **7** (10), 3143–3161.
- 40 Ponder, J.W. (2015) TINKER – Software Tools for Molecular Design (version 7). Available at: <http://dasher.wustl.edu/tinker>.
- 41 Semrouni, D., Sharma, A., Dognon, J.P. *et al.* (2014) Finite temperature infrared spectra from polarizable molecular dynamics simulations. *Journal of Chemical Theory and Computation*, **10** (8), 3190–3199.
- 42 Thauay, F., Dognon, J.P., Ohanessian, G. and Clavaguéra, C. (2015) Vibrational mode assignment of finite temperature infrared spectra using the AMOEBA polarizable force field. *Physical Chemistry Chemical Physics*, **17** (39), 25968–25977.
- 43 Oomens, J., Steill, J.D. and Redlich, B. (2009) Gas-phase IR spectroscopy of deprotonated amino acids. *Journal of the American Chemical Society*, **131** (12), 4310–4319.
- 44 Balaj, O., Semrouni, D., Steinmetz, V. *et al.* (2012) Structure of sodiated polyglycines. *Chemistry - A European Journal*, **18**, 4583–4592.
- 45 Frisch, M.J., Trucks, G.W., Schlegel, H.B. *et al.* (2009) Gaussian 09 Revision D.01. Gaussian Inc., Wallingford, CT.
- 46 Merrick, J.P., Moran, D. and Radom, L. (2007) An evaluation of harmonic vibrational frequency scale factors. *Journal of Physical Chemistry A*, **111** (45), 11683–11700.
- 47 Semrouni, D., Clavaguéra, C., Dognon, J.P. and Ohanessian, G. (2010) Assessment of density functionals for predicting the infrared spectrum of sodiated octa-glycine. *Journal of Mass Spectrometry*, **297** (1-3), 152–161.
- 48 Rasmussen, T.D., Ren, P., Ponder, J.W. and Jensen, F. (2007) Force field modeling of conformational energies: Importance of multipole moments and intramolecular polarization. *International Journal of Quantum Chemistry*, **107** (6), 1390–1395.
- 49 Semrouni, D., Ohanessian, G. and Clavaguéra, C. (2010) Structural, energetic and dynamical properties of sodiated oligoglycines: Relevance of a polarizable force field. *Physical Chemistry Chemical Physics*, **12** (14), 3450–3462.

- 50 Grossfield, A., Ren, P. and Ponder, J.W. (2003) Ion solvation thermodynamics from simulation with a polarizable force field. *Journal of the American Chemical Society*, **125** (50), 15671–15682.
- 51 Piquemal, J.P., Perera, L., Cisneros, G.A. *et al.* (2006) Towards accurate solvation dynamics of divalent cations in water using the polarizable amoeba force field: From energetics to structure. *Journal of Chemical Physics*, **125** (5), 054511.
- 52 Marjolin, A., Gourlaouen, C., Clavaguéra, C. *et al.* (2012) Toward accurate solvation dynamics of lanthanides and actinides in water using polarizable force fields: From gas-phase energetics to hydration free energies. *Theoretical Chemistry Accounts*, **131**, 1–14.
- 53 Semrouni, D., Isley, W.C.I., Clavaguéra, C. *et al.* (2013) Ab initio extension of the AMOEBA polarizable force field to Fe²⁺. *Journal of Chemical Theory and Computation*, **9** (7), 3062–3071.
- 54 Ren, P. and Ponder, J.W. (2002) Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. *Journal of Computational Chemistry*, **23** (16), 1497–1506.
- 55 Halgren, T.A. (1992) The representation of van der Waals (vdW) interactions in molecular mechanics force fields: Potential form, combination rules, and vdW parameters. *Journal of the American Chemical Society*, **114** (20), 7827–7843.
- 56 Allinger, N.L., Yuh, Y.H. and Lii, J.H. (1989) Molecular mechanics. The MM3 force field for hydrocarbons. 1. *Journal of the American Chemical Society*, **111** (23), 8551–8566.
- 57 Fanourgakis, G.S. and Xantheas, S.S. (2008) Development of transferable interaction potentials for water. V. Extension of the flexible, polarizable, Thole-type model potential (TTM3-F, v. 3.0) to describe the vibrational spectra of water clusters and liquid water. *Journal of Chemical Physics*, **128** (7), 074506.
- 58 Schultheis, V., Reichold, R., Schropp, B. and Tavan, P. (2008) A polarizable force field for computing the infrared spectra of the polypeptide backbone. *Journal of Physical Chemistry B*, **112** (39), 12217–12230.
- 59 Stone, A.J. (1981) Distributed multipole analysis, or how to describe a molecular charge distribution. *Chemical Physics Letters*, **83**, 233–239.
- 60 Semrouni, D., Cramer, C.J. and Gagliardi, L. (2014) AMOEBA force field parameterization of the azabenzenes. *Theoretical Chemistry Accounts*, **134**, 1–11.
- 61 Marx, D. and Hutter, J. (2009) *Ab initio molecular dynamics: Basic theory and advanced methods*, Cambridge University Press, Cambridge.
- 62 Vuilleumier, R. and Borgis, D. (1997) Molecular dynamics of an excess proton in water using a non-additive valence bond force field. *Journal of Molecular Structure*, **436-437**, 555–565.
- 63 Maupin, C.M., Wong, K.F., Soudackov, A.V. *et al.* (2006) A multistate empirical valence bond description of protonatable amino acids. *Journal of Physical Chemistry A*, **110** (2), 631–639.
- 64 Schlegel, H.B. and Sonnenberg, J.L. (2006) Empirical valence-bond models for reactive potential energy surfaces using distributed Gaussians. *Journal of Chemical Theory and Computation*, **2** (4), 905–911.
- 65 Thauay, F. (2016) Ph.D thesis, Palaiseau, France.

6

Empirical Valence Bond Simulations of Biological Systems

Avital Shurki*

*Institute for Drug Design, School of Pharmacy, The Lise Meitner-Minerva Center for Computational Quantum Chemistry,
The Hebrew University of Jerusalem, Israel*

6.1 Introduction

Understanding reactivity in biological systems and in particular in enzymes is one of the most fundamental problems in biology. In recent years, molecular modeling and computational methods have become increasingly relevant for the understanding of these biological processes at a molecular level. In particular, hybrid quantum-mechanics/molecular-mechanics (QM/MM) approaches have played an important role for understanding enzyme-catalyzed reactions, with QM methods describing the electronic process, for example, the actual reaction, and the MM method describing the rest of the system.^[1–5]

The empirical valence bond (EVB) is one such method.^[1,6–9] It involves a linear combination of diabatic states that represent stable species along the reaction (e.g., reactants, products). These diabatic states are described by MM potentials and represent the diagonal matrix elements of the systems' Hamiltonian. In order to obtain the adiabatic ground state energy, which is the solution of the secular equations, the coupling between these states is required. In the EVB approach, the coupling term is approximated by a simple function, which is based on empirical parameters. This coupling term, in turn, introduces the ability to describe the cleavage/formation of bonds, which is crucial for description of chemical reactions. In other words, this coupling term is, in fact, what turns the method into a QM/MM method, despite the use of pure classical states for the description of the diabatic states.

Like any other QM/MM method, the EVB approach is particularly designed to study chemical reactivity and can unravel the relevant reaction mechanism for any given reaction. This chapter will focus on different abilities of the EVB, highlighting particularly the various advantages of the approach.

One advantage involves the comparative aspect inherent to any EVB study. The EVB approach relies on calibration of empirical parameters to describe the energetics of a reference reaction, which can be the reaction in either vacuum, solution, or in a wild-type (WT) enzyme. This parameterization, which is intrinsic to the reaction, remains unchanged in different environments, allowing one to directly compare the

*Affiliated with the David R. Bloom Center for Pharmacy at the Hebrew University

effect of the enzyme on the reaction under study, as well as eliminating any systematic computational errors between the reference and studied system, (e.g., WT and mutant enzymes). This parameterization, therefore, guarantees reproducing the catalytic effect of enzymes with high accuracy and even more so, obtaining reliable quantitative predictions for the catalytic effects of enzymes with modified active sites.

Another advantage involves the method's efficiency. Biological systems are large, containing thousands of atoms and a reliable evaluation of the energetics often requires extensive sampling of the configurational space involved. Despite the rapid increase in the computational power, many methods still face a sampling problem, leaving many important problems where sampling is crucial out of scope. In contrast to the computationally demanding QM methods, the efficiency of the EVB is comparable to that of MM approaches. As a result, its sampling capabilities are higher by a few orders of magnitude compared with other *ab-initio*-based QM/MM approaches. Therefore, the EVB addresses problems with particularly high complexity and serves as a sampling tool for other QM/MM methods.

A third advantage involves the definition of the reaction coordinate as the energy gap between the VB states involved.^[6] In order to calculate reaction barriers within a biological system, usually one has to drive the system along a pre-chosen reaction coordinate. Despite its importance, there is no clear definition of the reaction coordinate and its selection is often arbitrary and depends on the user. Most reaction coordinates are geometric and involve one or several coordinates of the reacting fragments. However, often these reaction coordinates ignore the presence of the environment and are therefore deficient, particularly within a biological system. The definition of the reaction coordinate within the EVB approach, as the energy gap between the two diabatic states, ($\chi = \epsilon_1 - \epsilon_2$) is a simple definition which involves inclusion of all the systems degrees of freedom. As such, this reaction coordinate is highly advantageous as it allows better convergence and enables inclusion of non-equilibrium effects when evaluating the reaction barrier.

Finally, a great advantage of the EVB approach is the fact that it is based on the VB theory. VB theory, sometimes referred to as the more simplified and pictorial resonance theory, proved to be very useful and important for understanding bonding and reactivity in chemistry due to the simple and unique insights it provides.^[10–12] One example for the insights, which are unique to VB theory, is the demonstration that the barrier for any reaction originates from the crossing between relevant VB states (the diabatic states) along the reaction coordinate. Furthermore, the resonance structures/diabatic-states composing the VB wavefunction – which are often equivalent to the well-known Lewis structures – are chemically interpretable and serve as the thinking tools of chemists. Thus, rather than describing the transition state (TS) by a delocalized wavefunction, one can relate it to a linear combination of chemically well defined diabatic states with known weights. Hence, the fact that EVB uses these VB structures as its building blocks turns the EVB approach into a great tool to study reactivity and contributes to the understanding of chemistry.

All these advantages turn the EVB approach into a very powerful tool for studies of chemical reactivity and this chapter will try to demonstrate it in a variety of complex biological systems. Hence the first part of this chapter will demonstrate application of the EVB for unraveling the reaction mechanism in two cases: the hydrolysis of organo-phosphate compounds and in particular echthiophate in mutants of BChE^[13,14]

and the hydrolysis of GTP in G-proteins and in particular in Ras/RasGAP.^[15,16] The second part will focus on the advantage of the EVB as a comparative method. This will be demonstrated by studies of GTP hydrolysis. In these studies, a guided reaction path was used, leading to a deficient description of the mechanism. Nevertheless, the conclusions, which are based on a comparative study of the same reaction mechanism in solution and in the protein, are still valid.^[15,17] The advantage of the comparative nature of the approach will also be highlighted in studies of the same reaction in relatively similar environments, such as different conformations in the studies of the engineered monomeric chorismate mutase,^[18] or different mutants in the case of haloalkane dehalogenase (DhIA).^[19] The third part will emphasize the efficiency of the EVB method and its unique sampling abilities, for example in the calculation of the entropy contributions within biological systems, such as citrate synthases,^[20] trypsin^[21] or in the corresponding reactions in solution.^[22] The use of these sampling abilities will also be illustrated in cases where EVB provides configurations for ab-initio QM/MM calculations, as was the case in the studies of copper-chaperones,^[23,24] and ab-initio VB/MM calculations of DhIA.^[19] Finally, it will be shown by the usage of EVB as a reference potential.^[25–30] The fourth part will discuss the advantage of using the energy gap as the reaction coordinate. This will be established by comparing studies of the identity S_N2 reaction $\text{CH}_3\text{Cl} + \text{Cl}^-$ in solution, using two different reaction coordinates.^[31] Finally, in the last part, it will be demonstrated how EVB can be used as a great tool for providing better insight into reactivity. This will be exemplified by the insights it gives on the role of various residues in DhIA.^[32]

It is noted that this chapter tried to give the flavor of the biological problems EVB can solve with an emphasis on its advantages. However, it could not, of course, cover the countless biological problems solved to date by the EVB. The reader is, therefore, referred to various additional reviews that discuss these and many other complex biological problems studied by the EVB approach.^[33–44]

6.2 EVB as a Tool to Unravel Reaction Mechanisms in Biological Systems

Understanding enzyme catalysis requires detailed characterization of the mechanisms involved. Often, several mechanisms can be proposed for a certain reaction and to find the one that governs the kinetics, one has to reliably describe the relevant parts of the system's potential energy surface and quantitatively predict the activation free energies. These quantitative predictions often require QM/MM calculations. Such calculations of all possible alternative reaction mechanisms place high demands on computer resources. The availability of an efficient QM/MM method such as the EVB is clearly an advantage in such cases. Here, two examples where the EVB approach served to reveal the reaction mechanism will be described.

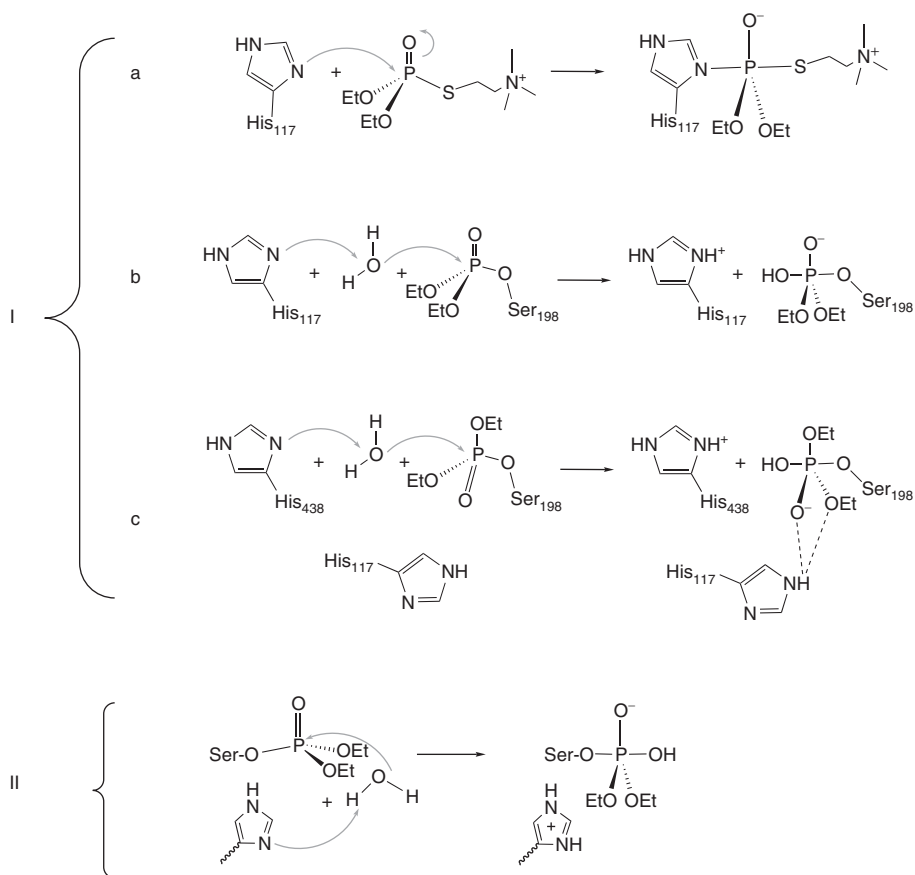
6.2.1 Hydrolysis of Organophosphate Compounds in BChE

Acetylcholinesterase (AChE) is a central enzyme in the nerve system. It hydrolyzes acetylcholine, thus allowing the next nerve impulse to pass. Organophosphate esters (OP) are highly toxic compounds due to their irreversible covalent inhibition of AChE. Since OPs are in large use as pesticides and may serve as nerve agents in warfare, it is

important to find a way to hydrolyze them in the body. Butyrylcholinesterase (BChE) is a serine esterase which, similarly to AChE, is capable of hydrolyzing acetylcholine. WT BChE is covalently inhibited by OPs; however, its G117H mutant was shown to be able of hydrolyzing certain OPs and catalytically protect against their toxicity. The rate of hydrolysis, though, is not sufficient for therapeutic use, and the role of the G117H mutation, which is important for development of an improved mutant, was not clear. The first step to achieve that goal involved establishing the structure of the G117H-BChE mutant, which was not available at that time. Amitay *et al.* started from the available WT-structure and replaced glycine at position 117 to histidine. They then took advantage of both the fact that the mutation does not impair hydrolysis of ACh and the efficiency of the EVB calculations, to validate the structure. This was done by simulating acetylcholine acetylation within G117H-BChE, where the mechanism is considered to be common knowledge, and reproducing the catalytic effect of that step.^[13]

Having a reliable structure Amitay *et al.* focused on the understanding of the hydrolysis mechanism of OP compounds within this mutant. They studied the first step in the dephosphorylation reaction of echothiophate – an OP compound that G117H-BChE is capable of hydrolyzing. Various mechanisms were proposed for the hydrolysis of that step, mainly differing in the role of the mutated histidine, His117. That is, while in some mechanisms His117 non-covalently stabilizes various compounds along the reaction coordinate, in others His117 is actually chemically involved. More specifically, one of the proposed mechanisms suggested His117 to serve as a nucleophile, with formation of a phosphoryl-histidine intermediate (Scheme 6.1-Ia). A second mechanism proposed a nucleophilic attack of a water molecule activated by His117 that serves as a general base in the first step of dephosphorylation (Scheme 6.1-Ib). Finally, a third mechanism suggested that the general base role in the dephosphorylation step is played by a different histidine (His 438) with diverse roles proposed, in that case, to His117 including, for example, facilitating deprotonation of His438 by electron withdrawal together with distortion of either the active site or the pentacoordinated TS, or perturbing the hydrogen bond between His438 and the OP (Scheme 6.1-Ic).^[45] The first mechanism was ruled out based on the structure of phosphorylated G117H BChE-mutant. This structure revealed a phosphorylated Ser198, thus ruling out the possibility that His117 acts as a nucleophile, in which case phosphorylated His117 should have been observed. For the remaining two mechanisms, different arguments were given against various roles of His117, but none of them disproved the mechanisms. Therefore, the last two mechanisms were examined. This was done by quantitatively calculating the catalytic effect of G117H BChE on the first step of echothiophate dephosphorylation, with either His438 or His117 serving as a general base and comparing the reaction profiles in the enzyme to the one obtained in solution and is common to both mechanisms (Scheme 6.1-II).^[14]

Various orientations of water attack are possible, resulting with pentacoordinated intermediates with different groups in the apical positions. Furthermore, both a concerted and stepwise mechanisms were considered. Finally, different ionization schemes including key residues in the vicinity of the reactive fragments were considered. All these scenarios increase the number of profiles to calculate even further and therefore the computational sources needed. In most QM/MM methods, calculation of each of these profiles is independent. However, within the EVB approach parameterization of the potential energy surface is done only once for each (sometimes hypothetical) uncatalyzed reaction mechanism. The simulations of the catalyzed reaction in the



Scheme 6.1 I. Different mechanisms proposed for echothiophate. In each mechanism only the first step that involves His117 is shown. a. His117 serves as the nucleophile in the phosphorylation step. b. His117 serves as a general base in the dephosphorylation step. c. His348 serves as the general base and His117 is not chemically involved but rather stabilizes, for example, the intermediate. II. The reference reaction in solution for the first step in the dephosphorylation with histidine acting as a general base.

mutated enzyme then follow using the same parameters, only each time the identity of the histidine that serves as the general base, and the ionization scheme differs. Parameterization is computationally demanding. Therefore, the fact that it should be done only once per reaction mechanism, considerably reduces the computational effort required for EVB calculations of the same reaction in different environments. The concerted mechanism with His117 acting as a general base provided the best agreement with the experimental data thus, suggesting that to be the mechanism of the first step in the dephosphorylation of echothiophate and possibly other OP compounds.^[14]

The efficiency of the EVB approach also allows one to simply examine the factors that are important for the catalysis, thus, clarifying the details of the mechanism even further. For example, for the preferred mechanism, the energy profiles were recalculated with His438- positively charged, both with and without explicit protonation. From these results it was possible to conclude that the positive charge on histidine is not sufficient

for catalysis and that the hydrogen bond interaction is essential to stabilize the transition state (TS).^[14]

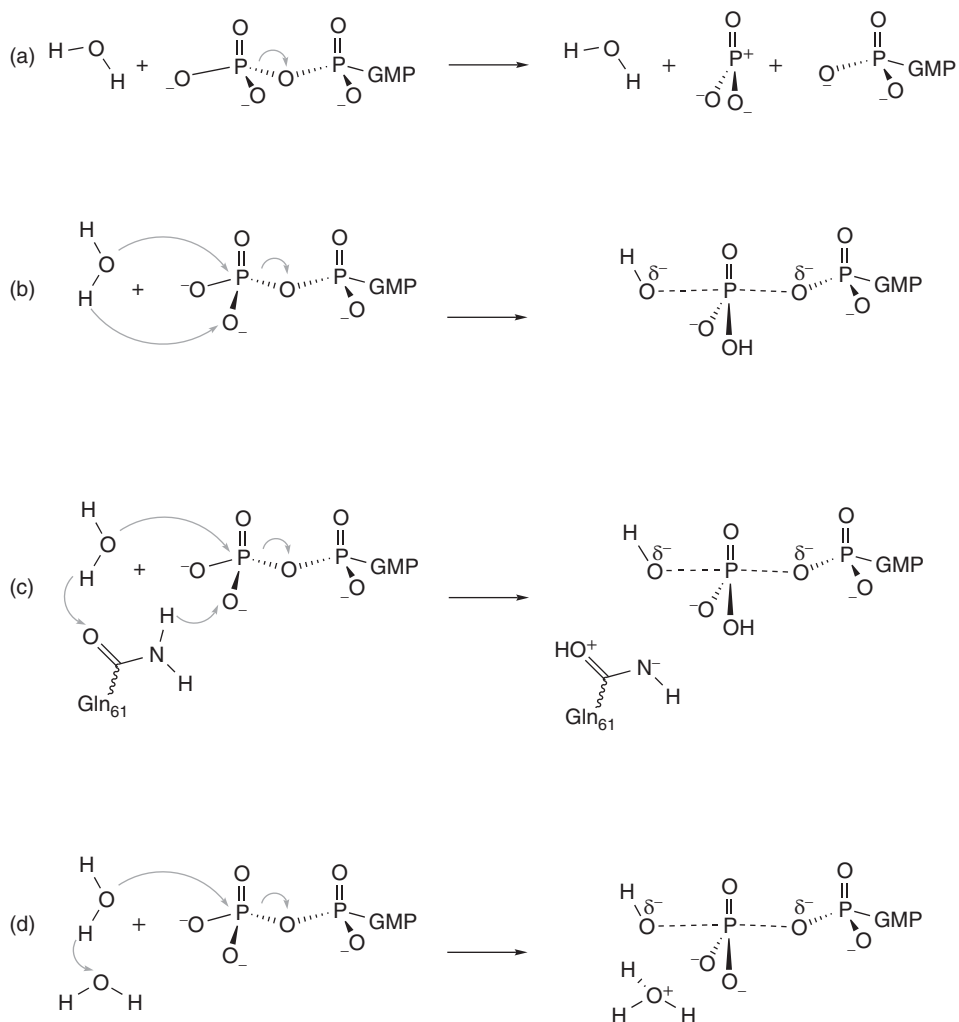
6.2.2 Hydrolysis of GTP in Ras/RasGAP

Ras is a molecular switch that through its GTPase activity controls a wide variety of signaling processes, including cell differentiation, proliferation and apoptosis. Ras functions by cycling between two states. In the Guanosine triphosphate (GTP)-bound state, Ras is active and can interact with downstream effectors, initiating various signaling cascades. In the Guanosine diphosphate (GDP)-bound state, Ras loses its ability to bind to and activate effector proteins. The intrinsic GTPase reaction, which returns Ras-GTP to its inactive GDP-bound form, is rather slow. However, it is accelerated by orders of magnitude upon interaction with a helper protein called GTPase activating protein (GAP). Several point mutations in Ras are associated with cancer and are found in 30% of human tumors. Most common oncogenic mutations are located at three positions on Ras: 12, 13, and 61. The oncogenic Ras mutants are not able to convert GTP to GDP and hence are locked in the “on” state, where they constantly activate cell cycle progression and division. Therefore, understanding the mechanistic details of GTP hydrolysis, within the protein and its complex with GAP, is highly relevant for the elucidation of the role of Ras mutants in tumor formation.

Different mechanisms have been suggested in the literature (see Scheme 6.2 for a partial list of proposed mechanisms).^[16,46] They include both a dissociative mechanism, with the phosphate dissociating prior to the nucleophilic water attack (Scheme 6.2a), as well as an associative mechanism, with formation of the new Nu(O)-P $_{\gamma}$ bond simultaneous to the breaking of the phosphate-GDP bond (Scheme 6.2b-d).^[46,47] In addition to these mechanistic possibilities, the identity of the general base in the reaction has also been a major source of debate; with some studies suggesting a conserved Gln61 residue as a general base, while others have pointed to a substrate assisted catalysis, with GTP acting as the general base.^[46,48,49] Furthermore, in the last case, the proton transfer (PT) process can take place in different ways, either directly from attacking nucleophilic water molecule to the γ -phosphate oxygen (e.g., reference^[15], Scheme 6.2b), or through Gln61 or additional water molecules serving as proton shuttle (Scheme 6.2c and 6.2d respectively).^[16,50] Most of these possibilities can be described as either stepwise or concerted and should be studied both in solution and within the enzyme. The computational demand for studying such complex scenarios is further increased when considering not only wild-type but also mutant forms of the enzyme either isolated or in complex with GAP.

QM/MM studies have often focused on one particular mechanism, due to the computational cost of calculating the full surface. However, studying only the relevant guided parts, may lead to wrong conclusions. Furthermore, these studies have been carried out at different levels of theory, which makes their comparison problematic.

Warshel and coworkers^[15,16,47,48,51,52] have used the EVB in order to efficiently and quantitatively explore the full PES, including all the different possible mechanisms. Their study ruled out the role of Gln61 as the general base in Ras, showing that the energy of the relevant PT step in Ras is higher than the actual observed barrier for the GTPase reaction in solution.^[51] They then demonstrated, based on energy considerations that GTP is the most likely general base, obtaining reasonable agreement



Scheme 6.2 The first step of different suggested mechanisms for GTP hydrolysis. a) dissociative mechanism. b) c) and d) present an associative mechanism with GTP as a general base while in c) and d) PT takes place with Gln61 (in c) or additional water molecule (in d) serving as a proton shuttle. (b)-(d) present only the concerted path, though stepwise options were also suggested. GMP stands for guanosine monophosphate.

with the experimental data.^[48] These results were further supported by a linear free energy relationship (LFER) study of the reaction rate within Ras versus the pKa of the γ -phosphate in GTP.^[49] The associative/dissociative character of the reaction was tested in a subsequent study on Ras, RasGAP, and some of their mutants. Here, using EVB, the effect of GAP was found to be similar on both the associative and the dissociative mechanisms.^[47] Hence, it was concluded that the PES in solution dictates the mechanism in the enzymes. The experimental data in solution is limited and does not provide sufficient information to dictate the exact shape of the PES.

Therefore, Warshel and coworkers constructed QM(ai)/MM surfaces of the reaction in solution.^[52] They then used the shape of that surface to calibrate the full EVB PES in RasGAP.^[16] The reaction in solution identified a TS (referred to here as TS1) with a significant barrier, which was common to various different mechanisms and involved the approach of a water molecule to the phosphate group and shift of the substrate negative charge towards the leaving group moiety. This TS1 was found as the first TS for different mechanisms (e.g., Scheme 6.2c and 6.2d) and depending on the mechanism, it was followed by one or several additional TSs that resulted from the PT step. The different mechanisms were all studied within the RasGAP. One mechanism involved a direct PT from the nucleophilic water to the GTP. Other mechanisms involved PT to a second water molecule or to Gln61 that serve in that case as a proton shuttles and transfer the proton to the GTP either in concerted or stepwise mechanisms.

The calculations demonstrated that the main effect of the enzyme is the electrostatic stabilization of TS1, where most of the charge-shift towards the leaving group occurs. In addition, it was shown that in RasGAP the reaction is likely to involve either water or Gln61 assisted mechanisms, but no catalysis was observed in these PT steps.^[16] Calculations of the two water mechanism were repeated in various mutants of RasGAP to examine the effect of mutations of Gln61 and hence, its corresponding role. The mutations were shown to reduce the stabilization of TS1 and had almost no influence on the PT step. It is noted in this regard, that there are other suggestions (e.g., reference [53]). Yet, the focus of this chapter is to describe the abilities of the EVB rather than solving the hydrolysis mechanism of GTP. The reader is referred to a recent review on that topic.^[46]

6.3 EVB a Comparative Tool

An EVB calculation involves calibration of the parameters that define the reactive part of the system. A reference reaction is first calculated and calibrated to reproduce experimental/or best available estimated data. Then, using the same parameters the reaction is calculated in the environment of interest (e.g., reference reaction in solution followed by calculation within an enzyme). This process guarantees that one obtains reliable results since the calculations account only for changes in the PES due to changes in the environment, thus, minimizing the other sources of error. This approach is based on a comparative study, and therefore has several clear advantages in particular in cases of guided reaction paths or studies of the same reaction in different environments.^[38]

6.3.1 Guided Reaction Paths

Often, guided reaction paths are used when studying reaction mechanisms. The results in such cases may be misleading as only part of the system has been explored rather than the full potential energy surface. The advantage of a comparative EVB study in such cases, as will be shown in the following example, is that it gives the best possible results within the constraints of the calculations. That is, the calculations capture in such a case only part of the reality, however, since it is a comparative study, the reality of the reference reaction, where calibration took place, is also partial. As a result, the conclusions may not be wrong, despite the fact that they are based on partial information.

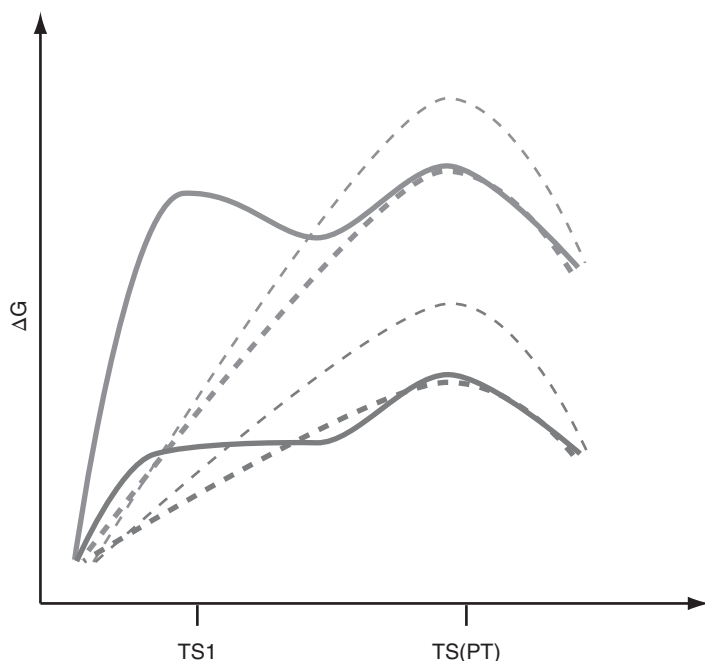
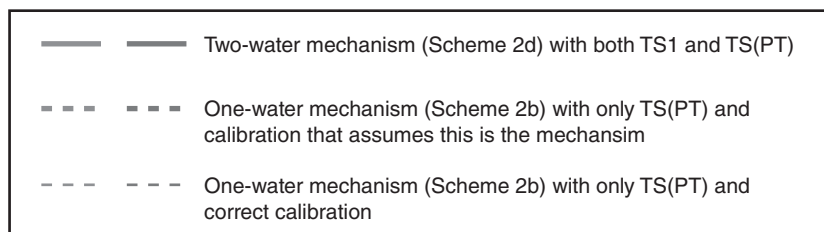
An example for that is the study of the role of Gln61 in Ras/RasGAP system, which is described at length in the Section 6.2.2. An early study of that system considered only

one water molecule and the GTP as reactive fragments (Scheme 6.2b). Clearly, in such a case, where neither a second molecule nor Gln61 were considered as reactive, the mechanism where Gln61 serves as a proton shuttle (Scheme 6.2c), or the two-water mechanism where a second water molecule assists the PT (Scheme 6.2d) could not be captured. These studies reproduced the overall trend of the experimental catalytic effects of Ras and the RasGAP complex relative to the uncatalyzed reaction, while using an associative mechanism with GTP as a general base.^[15] Furthermore, the effects of several real mutations including Q61L, Q61A and Q61N Gln (where N Gln is the isosteric nitro analogue of Gln) as well as the mutation of Gln61 to its nonpolar analogue were studied and demonstrated good agreement with the experimental data. The mutation of Gln61 to its nonpolar analogue showed a very small anticatalytic effect, indicating that Gln61 does not directly act by electrostatic stabilization of the substrate. Instead, it has an allosteric effect, keeping a preorganized catalytic configuration of the active site of the RasGAP complex.

Subsequent studies on this system^[16,52] demonstrated that the hydrolysis mechanism requires the assistance of either a second water molecule or Gln61 to serve as a proton shuttle. This implied, that the calibration of the reference reaction in solution using one water molecule, directly transferring the proton to the substrate, was somewhat deficient, leading to an underestimation in the activation barriers in both the reference reaction and in the protein. See Scheme 6.3 for the description of the profile with one and two water molecules. Still, these studies also demonstrated that the effect of the protein is to stabilize the first TS, which involves the approach of the nucleophilic water to the substrate and a shift of the negative charge towards the leaving group (TS1 in section 6.2.2). The following PT step, on the other hand, was not catalyzed by the protein.^[16] Therefore, it became clear that the conclusions derived based on the one water molecule mechanism are still valid. That is, the deficiency was in the description of the PT in the reference reaction, which was described as one step rather than two steps. The catalytic effect of the protein was shown not to involve the PT but rather the preceding step where the second water molecule does not play a role. This first step was considered in the deficient description as part of the PT step, and therefore the effect of the mistake diminished and the remaining of the conclusions continued to be applicable. This was also proven in that later study.^[16] Hence, the conclusions regarding the role of Gln61, which was earlier demonstrated to be allosteric, were shown to still hold even in the case of a mechanism that involves the assistance of a second water molecule.

Scheme 6.3 further clarifies that point. In the earlier calculation,^[15] a second water molecule was missing in the description of the reference reaction in solution. Hence, as seen in Scheme 6.3, the resulting wrong calibration for the reference reaction (presented by bold dashed lines in Scheme 6.3), led to wrong conclusions regarding the profile shape and the barrier height of the direct PT mechanism in the protein, which was estimated to be lower than its actual value (presented by plain dashed lines in Scheme 6.3). The study however, is comparative and the catalytic effect does not involve the PT step, but a prior step which in the deficient description is considered as part of the PT step. Therefore, all other considerations continue to be valid as the picture obtained by the deficient description, is very similar to the true picture.^[16]

Comparable insights were obtained for the hydrolysis reaction of GTP within the activated (ribosome-bound) form of elongation factor-Tu (EF-Tu) and the role of His84. Similar to the RasGAP system, initial calculations, in this case, also involved direct PT



Scheme 6.3 EVB free energy profiles of the first steps in the hydrolysis of GTP in solution and in the active site of RasGAP. Only the steps that lead to a PT to the substrate (GTP) are shown. The profiles of the uncatalyzed reaction and within the RasGAP complex are in blue and green, respectively. Plain lines display the profiles of the mechanism that involves a second water molecule, serving as a proton shuttle. Dashed lines, on the other hand, exhibit the profiles of the mechanism that involved direct PT from the nucleophilic water molecule. Here, bold and plain dashed lines correspond to the deficient (based on the earlier study) and true calibration, respectively. Activation barriers and reaction free energies are given with respect to the energy of their corresponding reactants regardless of the mechanism used. (See color plate section for the color representation of this figure.)

from the nucleophilic water molecule and His84 was concluded to have an allosteric role.^[17] Later it was shown that a more correct description involves assistance of a second water molecule.^[16] Yet, the conclusions regarding the role of His84 in the catalysis were proven to still be valid despite the different mechanism.^[16]

It is noted in this regards, that these results are relevant also for other GTPases and ATPases, where a direct PT mechanism was assumed and calibrated to reproduce experimental data. Yet, it is likely to believe, that the conclusions reached in these studies (e.g., regarding the role of the helical domain in Gs_{α} ^[54] or the ATP hydrolysis within ATPase^[55]) are still valid.

6.3.2 Studies of the Same Reaction in Different Environments

The advantage of a comparative study increases with the similarity of the relevant PESs. Therefore, the advantage becomes even more pronounced when the effect of a conformational change within the protein or the effect of various mutations on the reaction is of interest. Here, any inaccuracy in the description of the parent PES (e.g., the PES of either the parent conformation or the WT-enzyme) is likely to appear in its entirety also in the description of the subsequent systems (i.e., other studied conformations or respective mutants) unless it concerns the particular change (the particular conformation or residue which has been mutated). Therefore looking at the relative values, the effect of that inaccuracy is not apparent. Here, two of such examples will be discussed, one that involves studies of conformational changes and a second one that focuses on mutational effects.

6.3.2.1 The Effect of Conformational Changes

Most studies of enzyme catalysis suggest that a prerequisite for proper catalysis is an exquisitely preorganized active-site structure, which often translates into a particular fold along with specific residue orientations. Yet, several recent studies identified catalysis within partially folded enzymes.^[56,57] For example, the catalytic efficiency of a recently engineered monomeric chorismate mutase was shown to be similar to that of the naturally occurring dimer even though it has properties of an intrinsically disordered molten globule, (i.e., an ensemble of poorly packed and rapidly interconverting conformers).^[56] Furthermore, when it binds a TS analogue it becomes more ordered but retains a unique flexibility. This example, triggered an in-depth investigation of the relationship between protein folding and catalysis in this case.^[18] Warshel and coworkers combined a coarse-grained model and the EVB approach to study the catalytic behavior of this system in different folding states.^[18] The folding landscape of both the monomer and the dimer were explored as a function of radius of gyration (R_g) and contact order (CO) and provided the probability to be at each conformation along the surface. Conformations from three different regions of this surface were, then considered. Region I involved native conformations (NMR and X-Ray for the monomer and dimer, respectively). Region II involved near-native conformations, taken from the minimum free energy surface of the simplified model while using small constraint on the distance between the catalytic residues and the ligand to be near the native distance. Region III involved conformations far from the minimum of the simplified model, with the same small restraint on the distance between the catalytic residues and the ligand (see Figure S1 in SI of reference^[18]). Explicit models were generated for conformations taken from the simplified model (regions II and III). All models were then relaxed and served as starting conformations for the calculations of the catalytic power in the monomer and in the dimer.^[18]

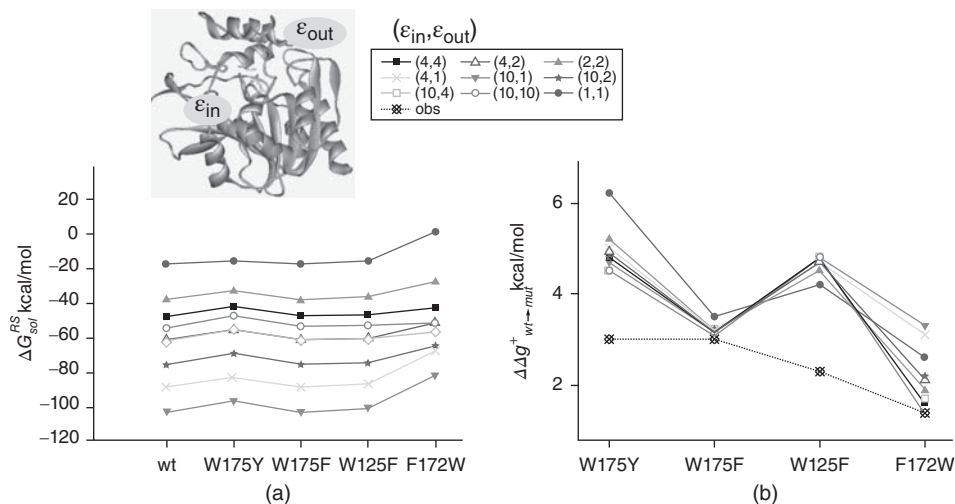
Overall the calculated results reproduced the observed experimental trends. A shallow folding landscape was obtained. Catalytic conformations were found both in regions I and II. Furthermore, the probability of finding low barriers in region II was high in the monomer but rather low in the dimer. It was shown that the shallow landscape enables the monomer fast interconversion between different preorganized catalytic conformations, without paying significant preorganization free energy.^[18]

Using the EVB approach is probably the only way to conduct such a study cost-effectively and reliably to date. Its efficiency enables extensive calculations of

reaction profiles on many different conformations, which is required to properly describe the entire folding landscape. Furthermore, the calibration process within the EVB, that essentially reproduces the observed barrier in one system, guarantees reliability of the results due to the comparative character of the approach. The differences between calculated reaction barriers in different conformations necessarily represent only the effect of the conformational change. Using alternative QM/MM approaches in such a study (let alone the fact that it would be computationally unfeasible with most approaches) could result in differences that originate in an artifact of changes in the energetics of the intrinsic reaction.

6.3.2.2 Mutational Studies

Mutational studies by the EVB approach, often nicely reproduce the effect of the mutation due to the comparative character of the method (e.g., references^[58–61]). An example, that emphasizes the advantage of a comparative approach in studies of mutational effects, is a recent study of the first S_N2 step within Dh1A to explore various aspects of the linear response approximation (LRA) approach.^[19] In the enzyme, the carboxylic group of Asp124 attacks the substrate, a haloalkane, leading to a substitution reaction which is the first step in the conversion of haloalkanes into alcohols. Thus, in this study, different dielectric constants ϵ , were used in different parts of the protein as scaling for the electrostatic interactions. Here, ϵ_{in} and ϵ_{out} served as scaling factors for ionizable residues within the interior of the protein (inner) and those located on the protein surface (outer), respectively. The effect of different scaling schemes on the solvation energy of the reacting fragments (including the haloalkane and the functional group of Asp124) and on the overall barrier was examined. Scheme 6.4a demonstrates the solvation energies of the reactive fragments in RS, ΔG_{sol}^{RS} , within WT-Dh1A as well



Scheme 6.4 The effect of the scaling scheme of the electrostatic interactions with the ionizable residues on a. the solvation energy of the reactive fragments, ΔG_{sol}^{RS} and b. the relative catalytic activity of the mutants, $\Delta\Delta G_{wt \rightarrow mut}^{\ddagger}$. Each graph corresponds to a different set of scaling factors, (ϵ_i, ϵ_o) , for the inner and outer ionizable residues, respectively. Dashed line represents the observed experimental estimations. (See color plate section for the color representation of this figure.)

various mutants. In addition, the relative catalytic ability of these mutants, $\Delta\Delta g_{wt \rightarrow mut}^\ddagger$ is demonstrated (Scheme 6.4b). In both cases the results are shown using the different scaling factors, $(\epsilon_{in}, \epsilon_{out})$ used.

From the scheme it is clear that even though different scaling schemes resulted in different absolute values (in a range of over 100 kcal/mol), they all provided proper trends when different mutants were compared with the *WT*-enzyme results. In other words, proper trends are obtained regardless of the scaling scheme due to the comparative character of the approach.^[19]

6.4 EVB – A Sampling Tool

The calibration process within the EVB approach requires special attention and may be time consuming. However, since the parameters are intrinsic to the reaction, once calibration is obtained it is very simple to calculate the same reaction in different conditions (e.g., different temperatures, different initial conformations) and different environments (e.g., solution, *WT*-protein and its mutants). These calculations are equivalent to MM as each one of the VB states is described by an analytical force field (FF). These calculations are, therefore, also highly efficient. Here, several examples where the calculations take advantage of this feature of the EVB approach will be demonstrated.

6.4.1 EVB – An Efficient Way to Run an Enormous Number of Calculations

The calculation of the energy profile of a reaction within a biological system requires very large sampling due to the large number of degrees of freedom involved. Using various techniques (some of which will be discussed in the next section) along with the fast development of computers, some QM/MM approaches are now able of calculating free energies of activation. However, calculations of some properties, such as activation entropy, still require particularly large sampling, which makes the use of most standard QM/MM approaches impossible. In such cases the EVB approach provides an excellent and efficient solution.

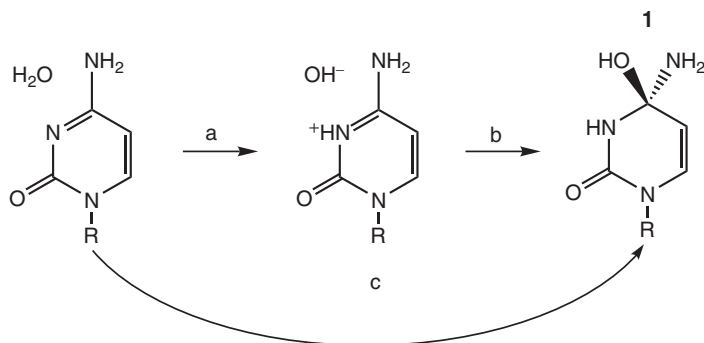
Two examples of the use of EVB in such context are the recent studies by Åqvist and coworkers on the cold adaptation of citrate synthases^[20] or trypsin enzymes.^[21] The goal of these studies was to understand how these enzymes survive in cold temperatures and still stay stable and active. Cold-active enzymes are known to exhibit reduced activation enthalpy and entropy compared to their mesophilic and thermophilic homologues, which makes their reaction rates less sensitive (approaching linear dependence) to temperature. However, the origin of this effect was not clear. In order to answer this question the authors computationally obtained Arrhenius plots trying to reproduce the experimental rate-temperature dependence and decomposed the overall activation energy into enthalpic and entropic components. The accuracy in such calculations is highly important.

In the study of trypsin, for example,^[21] Arrhenius plots were obtained by calculating the activation free energy profiles at eight different temperatures. In order to obtain converged results at least 100 independent FEP calculations, each of 500 ps length, were used, resulting in over 400 ns in total for each point in the Arrhenius plot. This was done for salmon and bovine trypsin, representing cold- and warm-active trypsins

respectively, and for the WT and several mutant forms. The experimental reaction rates as well as the thermodynamic activation parameters were reproduced by the calculations at impressively high accuracy. It was shown that the reduction in the activation enthalpy originates from interactions within the outer part of the protein. That is, the origin for the exceptionally low enthalpy along with the low and negative entropy is not associated with the active site, but rather with the protein surface, which retains certain softness. Hence, while the enzyme has a relatively rigid core, its outer region is relatively soft and flexible thus, creating a softer potential energy surface. This conclusion was further supported by computational predictions on the effects of various mutations in the protein surface which were verified.^[21] Clearly this amount of computations and the resulting conclusion could not have been carried out with any other QM/MM approach but the EVB.

A second example involved the hydrolytic deamination of cytidine and 5,6-dihydrocytidine in aqueous solution.^[22] Here, various different mechanisms were explored for the formation of the tetrahedral intermediate, **1** in Scheme 6.5, which is rate limiting in this deamination. These include a stepwise mechanism, with protonation of the N3 nitrogen followed by hydroxide ion attack on the C4 carbon (steps a and b in Scheme 6.5), or a concerted mechanism where the two former steps occur simultaneously, and formation of **1** due to the water attack is obtained in one step (step c in Scheme 6.5). In the latter case, the involvement of additional water molecules that participate in the reaction was also suggested, and various possibilities including participation of either two or three water molecules were suggested.

Calculations in various approaches including both QM with varying solvent models as well as EVB, showed that both the concerted and the stepwise mechanisms have very similar activation energies. Therefore, to examine what is the mechanism that actually dictates the reaction rate in this case, further calculations were performed to obtain the exact enthalpy–entropy partitioning of the activation energy. This again, required calculation of Arrhenius plots for each one of the mechanisms separately, which in turn involved tremendous amount of calculations due to the accuracy requirement. Thus, in the current system that yielded about 76 ns of simulation time for each reaction step at each one of the seven different temperatures used and for each substrate (cytidine and 5,6-dihydrocytidine). Finally, the results for the concerted mechanism with three water



Scheme 6.5 Formation of the tetrahedral intermediate, **1**, in the hydrolytic deamination of cytidine. a and b represent the stepwise mechanism whereas c stands for the concerted mechanism.

molecules showed excellent agreement with the experimental thermodynamic activation parameters, suggesting that this mechanism is the main player.^[22]

It is noted here, that since activation free energies were available from both experiments and DFT calculations, two sets of calibration schemes of the EVB potentials were used, exactly reproducing the experimental and calculated data. The overall energetics as well as the thermodynamic properties, following these two schemes, were very similar, suggesting that the energetic is insensitive to the calibration scheme. This is actually a direct result of the comparative character of the EVB and highlights this advantage of the EVB.

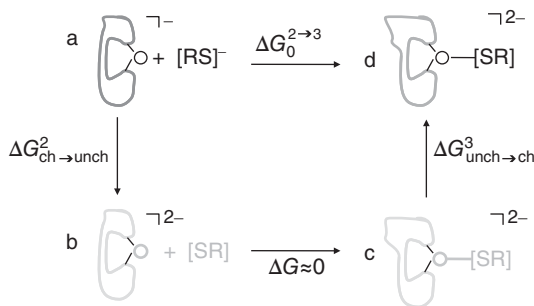
This approach, calculating thermodynamic parameters and comparing to experiment, is suggested as a means to make mechanistic predictions.^[22] Yet, one has to remember that extensive conformational sampling is a prerequisite and is currently feasible for biological systems only with the EVB approach. This is nicely illustrated in a recent work of GTP hydrolysis in EF-Tu, where entropy contributions were shown to enable the high rates of GTP hydrolysis on the ribosome.^[62]

6.4.2 EVB – An Efficient Way to Sample Conformations for Other QM/MM Approaches

Evaluation of free energy surfaces with reliable QM(ai)/MM approaches (where ai stands for ab-initio) encounter very serious sampling problems as they require major computational resources. An elegant solution to this, is the use of the EVB approach as a reference potential which allows collection of representative ensemble of conformations which are then calculated with higher-level QM(ai)/MM approaches. This allows to efficiently sample the conformational space at a much lower computational cost, while giving a relatively simple solution for cases where usage of QM(ai)/MM is of interest. This approach, originally developed by Warshel and coworkers,^[25–28] has several variations as discussed in the three examples below.

6.4.2.1 Copper-Chaperones

Copper ions, which are essential elements in living systems, are highly toxic if they are not tightly regulated, and excess of copper is associated with a variety of neurodegenerative diseases. Controlled cellular copper transport is achieved through copper binding proteins named copper-chaperones. These proteins tightly bind Cu(I) and forbid undesired chemistry, yet they can efficiently release Cu(I) ions at the designated locations. Understanding the Cu(I)-binding modes to copper-chaperones is therefore of high importance. It was found that Cu(I) binding within Atox1 (the human protein) is achieved through two conserved cysteine residues in a low (2-3) coordination number due to possible binding of a third (external) ligand. The experimental observation of coordination of two in the Atox1 *vs.* coordination of three in yeast and bacteria homologues was puzzling, and required validation and better understanding. Ansbacher *et al.* established a computational method to predict the coordination number of copper in the copper-chaperones. They utilized the density functional theory, using PBE0 functional with the Pulay modified 6-31+G* basis set for Cu(I) and the 6-31+G* basis set for the remaining elements. This level of theory was shown to be a good compromise between accuracy and efficiency, predicting reliably coordination number



Scheme 6.6 The two step (simultaneous) thermodynamic cycle used to calculate the free energy difference between the di- and tricoordinated Cu(I), $\Delta G_0^{2 \rightarrow 3}$. RS^- represents the third (external) thiolato ligand. Compounds whose partial charges are set to zero are colored gray. Different colors of the proteins present the different protein conformation ensemble, which is adjusted to the charge distribution. (See color plate section for the color representation of this figure.)

of small Cu(I) complexes.^[63] It was used therefore, for subsequent QM(ai)/MM study of Cu(I) coordination within Atox1. The possibility of binding of a third external ligand, was explored and the goal was to energetically discriminate between di- and tricoordinated Cu(I) within the protein.^[23]

Coordination within a protein complex can differ from the pattern found in small Cu(I) complexes, and therefore free energy differences (and not only potential energy differences) have to be evaluated. This was done by using the thermodynamic cycle shown in Scheme 6.6.^[23] It was based on assessing the solvation free energy of Cu(I) and the third thiolato ligand in both the di- and tricoordinated states, either simultaneously or in steps (i.e., evaluation of the ligands' solvation followed by the solvation of Cu(I)). Here, solvation of a compound refers to its interaction with the remaining system, while accounting for the reorganization required, and was calculated as the free energy for setting its partial charges to zero. These free energies are calculated by the LRA approach,^[64] where considerable sampling is required. For example, $\Delta G_{a \rightarrow b}$ in Scheme 6.6 is given by the following equation:

$$\Delta G_{a \rightarrow b} = \frac{1}{2} [\langle U_b - U_a \rangle_a + \langle U_b - U_a \rangle_b] \quad (6.1)$$

where $\langle \rangle_i$ designates an average over an ensemble of conformations of a dicoordinated complex (i.e., when the thiolato ligand is not bound to Cu(I)) collected from trajectories which are propagated on the potential energy surface U_i , where Cu(I) and the thiolato ligand are polar, (U_a) and when their respective partial charges are set to zero, U_b . Furthermore, for the results to be reliable, the ensemble has to be large enough, and properly sample the PES of the relevant states. Such sampling using the QM(ai)/MM approach would be computationally costly and the EVB served in this case to sample conformations. Thus, for each state (a, b, c or d in Scheme 6.6) the coordinates of the initially QM(ai)/MM optimized reactive part were kept fixed and its corresponding electronic structure (charge distribution) was set to reproduce the QM(ai) partial charges. The rest of the system was subjected to relaxation using MD simulations where different protein conformations were collected. Finally, single point calculations were performed at the QM(ai)/MM level using protein conformations that were obtained by the EVB approach.^[23]

The results obtained by the two step (simultaneous) cycle were found to agree with the results obtained by the multiple step cycle (where uncharging of Cu(I) and the ligand was done in steps), suggesting the methodology is reliable.^[23] Calculations showed that there are two different conformations referred to as “in” and “out” differing by their coordination preference. Thus, the “out” conformations exhibit higher preference for di-coordinated Cu(I) within Atox1 than the “in” conformations.^[24] Following these results a unique explanation for the coordination and metal transfer enigma of copper in Atox1 was suggested. It is noted that if MD simulations were to use QM(ai)/MM calculations, without the assistance of the EVB in sampling conformations, these results would have required significantly more computer resources.

6.4.2.2 Hybrid Ab Initio VB/MM Approach

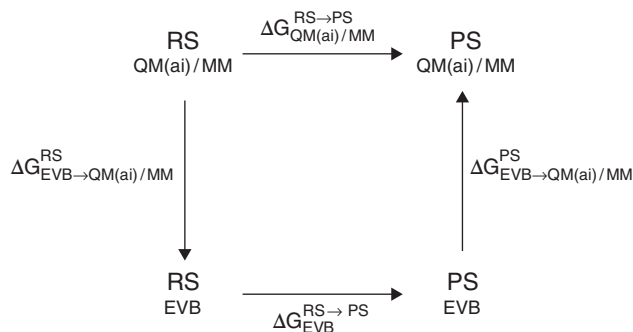
The hybrid ab-initio VB/MM is a unique approach which similar to other QM(ai)/MM approaches combines ab-initio VB with MM.^[32,65–69] Similar to the EVB approach, the wavefunction of the VB/MM is a linear combination of different VB structures and it, therefore, contains unique chemical information (see also section 6.6). This approach, however, is highly demanding and in the current state of computer power it faces significant efficiency disadvantage compared with EVB. Furthermore, MD calculations at most of the VB/MM levels are not possible currently, due to the difficulty to obtain the corresponding derivatives (currently derivatives may be obtained for only limited ab-initio VB based methods, for example, reference ^[70]).

Transformation of the VB/MM wavefunction into an EVB wavefunction is, on the other hand, straightforward and the corresponding PESs are expected to be very similar. Therefore, in that case, using the EVB as a sampling tool is naturally expected. Thus, the VB/MM approach is based on single point VB/MM calculations on conformations which are collected on EVB surfaces, while the EVB wavefunction is dictated by the VB/MM approach. These symbiotic relations are beneficial for the two methods which are in fact complementary. EVB provides efficiency, while the VB/MM enables simple inclusion of as many VB structures as one wishes to the calculation, with minimal effort, resulting with more accurate chemical information on the system.^[32,65,68,71] Various systems were studied successfully using this approach, and as will be demonstrated in section 6.6, it has an added value due to the special chemical insight that it provides.

6.4.2.3 EVB – An Efficient Reference Potential

A different solution to the serious sampling problem of standard QM(ai)/MM methods is the use of the EVB potential as a reference potential, an approach recently referred to as paradynamics (PD).^[25–30] A detailed description of PD can be found in Chapter 3 of this book. Here, the EVB does not only provide coordinates of the protein conformations which are later evaluated by the QM(ai)/MM approach, but rather performs the free energy calculation as well. The idea is that the PESs of the EVB and the QM(ai)/MM levels are reasonably similar. Hence, the EVB Hamiltonian is used to estimate the relative free energy and then correct it with the difference in free energy between the QM/MM and the EVB surfaces.^[25–28]

The method uses a thermodynamic cycle, as shown in Scheme 6.7. The free energy change, between the two states, RS and PS (representing, e.g., reactants and products, respectively as in Scheme 6.7) is first estimated using the EVB potential, giving $\Delta G_{EVB}^{RS \rightarrow PS}$. This estimate is then corrected to the QM(ai)/MM value by calculating the free energy of moving from the reference (EVB) to the corresponding QM(ai)/MM potentials at



Scheme 6.7 The thermodynamic cycle used to calculate the QM(ai)/MM free energy difference between the reactants, RS, and the products PS states, $\Delta G_{QM(ai)/MM}^{RS \rightarrow PS}$

the two respective states, RS and PS, leading to $\Delta G_{EVB \rightarrow QM(ai)/MM}^{RS}$ and $\Delta G_{EVB \rightarrow QM(ai)/MM}^{PS}$ respectively. In principle, proper evaluation of these corrections would lead to the exact QM(ai)/MM free energy difference, $\Delta G_{QM(ai)/MM}^{RS \rightarrow PS}$. However, these corrections are usually approximated, for the sake of efficiency, resulting therefore with only an approximated value of $\Delta G_{QM(ai)/MM}^{RS \rightarrow PS}$. That is, originally, these differences were evaluated by a single-step free energy perturbation, where configurations were collected only from trajectories on the reference (EVB) surface.^[25–27] These works, however, encountered difficulties in getting converging results whenever the differences between the QM(ai)/MM surface and the reference potential are too large.

An improved version, involves a. refining the EVB surface by considering several gas-phase model systems and fitting the EVB surfaces to the QM(ai)/MM surfaces, and b. calculating the corrections, $\Delta G_{EVB \rightarrow QM(ai)/MM}^X$ ($X=RS, PS$), using the LRA approach.^[28] This version was shown to provide reasonable estimates of the QM/MM free energy, $\Delta G_{QM(ai)/MM}^{RS \rightarrow PS}$, even when the QM(ai)/MM and the simplified EVB potentials are quite different.^[28] Additional improvement, extended the use of the method which was designed for calculations of reaction free energies and enabled in addition, obtaining converging QM(ai)/MM activation free energies while considering the effect of the substrate motion. This involved applying a constraint which fixes the TS at a given value of the reaction coordinate allowing therefore to use the LRA at the TS.^[29] An improvement that related only to calculations of solvation free energies involves representing the solute environment by an average solvent potential, which is then added to the solute Hamiltonian. This approach was shown to reduce the computational cost by up to 3 orders of magnitude for QM(ai)/MM-FEP calculations of solvation free energies when the solute structure is kept fixed during the simulation.^[72,73]

The idea is to have a reference potential that is as close as possible to the QM(ai)/MM surface, since this is the main factor in accelerating the convergence of the LRA approach, and hence the overall calculation. Therefore, the recent progress of the PD approach involved an automated iterative refinement procedure for fitting the EVB potential to the QM(ai)/MM potential, thus reducing the difference between the two potentials.^[30]

It is inevitable to comment about the metadynamics,^[74] in this respect, as both approaches (metadynamics and paradynamics) utilize simplified reference potentials to

enhance sampling of QM(ai)/MM surfaces. Yet, while metadynamics creates a flattened potential that enables faster sampling, the advantage of paradynamics is that the potential is similar to the potential of interest and includes great amount of chemical information. Furthermore, it was shown that paradynamics is far more cost-effective as it requires considerably less quantum calculations.^[30]

On a final note, it is stressed here that despite the efficiency gained by using the EVB as a reference potential or using the PD potential, these treatments still involve extensive calculations. Hence, if EVB alone can be used, it seems it would be the preferred option.

6.5 EVB Provides Simple Yet Superior Definition of Reaction Coordinate

The quality of the sampling process is at least as (if not more) important as its quantity. In other words, if the sampling process does not focus on relevant regions of the PES, it leads to serious convergence problems and may even provide improper results regardless of the ensemble size. Here, the choice of the reaction coordinate (RC) is of crucial importance. If a certain degree of freedom is missing in the description of the RC despite its involvement in the reaction, it would lead to hysteresis due to uncontrolled sampling of that degree of freedom. This problem, which is relevant for any system, increases with the system's size.^[6]

In biological systems the RC should include degrees of freedom not only of the reacting fragments but also of the environment (i.e., protein and/or solvent) since conformational changes or changes in solvent penetration to the active site, may considerably affect the TS (e.g., reference^[75]). Thus, a RC that may be fine for the reaction in the gas-phase, may not be suitable when describing the same reaction in solution or in an enzyme.

This is a difficult problem, as one has to define a multi-dimensional system using one or two coordinates. In most methods, to date, the choice of the RC, albeit resulting by usage of collective variables, is still arbitrary and does not include all the degrees of freedom of the system (i.e., the reactive part and its surroundings (e.g., references^[76–80]). Using such RCs requires very long relaxations for the environment to accommodate changes in the reactive part, and, in addition, they fail to account for the effect of non-equilibrium solvation on the reaction barrier.^[81–84]

The advantage of the EVB approach over other approaches when considering sampling is not only a quantitative advantage, due to its efficiency, but also a qualitative advantage, due to the unique definition of its RC. Here the RC, referred to as ϵ_{gap} , is expressed as the difference between the energy of the diabatic resonance states describing the system. That is

$$\epsilon_{gap} = \epsilon_{11} - \epsilon_{22} \quad (6.2)$$

where ϵ_{11} and ϵ_{22} are the energies of the diabatic resonance states representing the reaction. These are electronic structures that do not depend on the geometrical coordinates. Thus, one does not have to decide which coordinates to include and which to keep aside. In addition, the energy of these resonance structures accounts for the interactions of the

whole system, thus, capturing all aspects of the reaction including both chemical and environmental changes as well as their respective coupling.^[6]

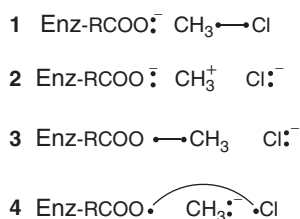
A recent study by Fuxreiter and coworkers demonstrated that ϵ_{gap} is by far the best RC known to date and suggested to use it as a universal RC for modelling chemical reactions.^[31] In their study they showed that ϵ_{gap} can be used for simulations with any quantum Hamiltonian. More importantly, they compared the results obtained by using a standard geometrical RC and ϵ_{gap} . The reaction they studied was the identity nucleophilic substitution reaction of methyl chloride with a chloride ion, which is symmetric and thermoneutral. In the gas-phase, accurate profiles, symmetry and barriers were obtained when using the canonical RCs. However, in solution, ϵ_{gap} results in a better profile that (contrary to the geometrical RC) still keeps the symmetry and leads to an activation barrier which is closer to the experimental data.^[31] Furthermore, the results suggest that the TS region is not uniformly sampled when the geometrical RC is used, and that ϵ_{gap} outperforms the geometrical RC in TS localization as well. Examination whether better TS localization with ϵ_{gap} is due to higher computational cost, revealed a serious hysteresis with the geometrical RC in that case, while calculations with ϵ_{gap} were shown to increase the efficiency of the calculations. In summary, it was concluded that ϵ_{gap} which is an outcome of the EVB approach, provides better sampling at a lower cost and can be applied to any computational level of calculation.^[31]

6.6 EVB – A Tool with Great Insight

The advantages of the EVB over other MO based QM/MM approaches, in particular the efficiency of EVB, may gradually fade away with increasing computer power. Yet, a special value of VB based QM/MM approaches including the EVB is the unique chemical insight they provide. This advantage has only recently been explicitly demonstrated within the combined VB/MM – EVB study,^[32] and its use is expected to increase and improve with time. The valence bond structures contain chemical information, which simplifies our understanding of the system's wavefunction. This information, which is easily accessible for any chemist, provides the method great capabilities that are not possible with MO based methods. Here, one such example, that demonstrates how the method facilitates understanding of the role of particular residues in catalysis, will be discussed.

Catalysis and enzyme design are among the most fascinating and challenging aspects of chemical and biological research. Here understanding the contribution of specific residues to catalysis is an essential step in the fine-tuning process of enzyme design.^[58,85–91] Applying the linear response approximation (LRA) using electrostatic interactions was shown to be useful to estimate these contributions in systems where this approximation is valid and catalysis is governed by electrostatics. In a recent study of Dh1A, this ability of the LRA was further developed by combining it with EVB sampling of a VB/MM wavefunction. As a result, the effect of various known mutations on catalysis was successfully explained.^[32]

This study focused on the first $S_{\text{N}}2$ step in the conversion of haloalkanes into alcohols and used methyl chloride as a model system with substituting the chloride by Asp124



Scheme 6.8 The different valence bond structures that govern the S_N2 reaction within DhIA. Enz stands for the enzyme, RCOO corresponds to residue Asp124.

from the enzyme. The contributions of particular residues to catalysis in both WT and its various known mutants were compared identifying residues for which biggest differences are observed. Due to the additive character of the LRA, these residues dictate the overall change in catalysis within the mutants. In most cases the leading residue was actually the one mutated, as would be expected from mutations that have little effect on the overall structure of the enzyme. To fully understand the effect of the mutation, the contributions of these residues to reactants, transition state as well as to the overall catalysis were decomposed into their different VB components. The VB wavefunction which is governed by four VB structures (Scheme 6.8), was obtained from VB/MM calculations and the respective weights of the different VB structures were used for further EVB/LRA calculations.^[32]

This analysis provides clear chemical picture for the origin of the change in catalysis. For example, the catalytic contribution of Trp175 in WT-DhIA was shown to be mainly due to stabilization of VB structures **2** and **3** (Scheme 6.8), which is higher in the TS. These two structures share a negatively charged Cl^- ion. It is suggested therefore, that the catalytic contribution of Trp175 originates from its interaction with Cl^- . The partial negative charge on this atom grows along the reaction leading to a higher stabilization in the TS compared with the reactants and therefore to a catalytic effect of Trp175.^[32] This result, which is solely based on calculations, is in agreement with observed data, and this residue is indeed known to be highly important for the stabilization of the leaving chloride ion.

Once the chemical origin of the contribution of each residue to catalysis is fully understood it is possible to suggest mutations that are expected to enhance/diminish/alter the effect in a way that would lead to an overall improved catalysis provided that the overall structure of the protein undergoes minimal changes due to the mutation. This strategy was shown to be successful in predicting point mutations that led to computationally enhanced catalysis in that system, where the VB character of the method and its unique chemical insights greatly facilitated the choice of proper mutations.^[32] Thus, for example Glu56 which was shown to have an anticatalytic contribution due to its destabilizing interactions with structures **2** and **3** and thus their common Cl^- ion, was computationally mutated to glutamine. This mutation converted the destabilizing interactions into stabilizing interactions, and as a result, the anticatalytic into a catalytic contribution, enhancing therefore the overall reaction rate.^[32]

These insights, which are readily available within VB-based approaches, provide additional advantages to the EVB approach.

6.7 Concluding Remarks

In summary, this chapter covered several examples that highlight the various advantages of the EVB approach over other MO-based QM/MM approaches in studies of complex biological problems. This included in particular its ability to conduct a comparative study, thus reducing mistakes, its efficiency that enables to confront problems which are out of the scope of many other computational approaches, its clever definition of the reaction coordinate which results in better sampling both qualitatively and quantitatively, and its readily available chemical insight which contributes to better understanding of the system. The EVB has many other aspects that were not discussed here and the reader is referred to other chapters of the book for a complementary overview of the EVB approach and its capabilities.

Acknowledgement

This research was supported by the ISRAEL SCIENCE FOUNDATION (grant No. 1337/13). I would also like to thank the Alex Grass Center for Drug Design and Synthesis of Novel Therapeutics for financial support.

References

- 1 Warshel, A. and Weiss, R.M. (1980) An empirical valence bond approach for comparing reactions in solutions and in enzymes. *Journal of the American Chemical Society*, **102** (20), 6218–6226.
- 2 Field, M.J., Bash, P.A. and Karplus, M. (1990) A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *Journal of Computational Chemistry*, **11**, 700–733.
- 3 Svensson, M., Humbel, S., Froese, R. *et al.* (1996) Oniom: A multilayered integrated MO-MM method for geometry optimizations and single point energy predictions. A test for diels-alder reactions and Pt(P(*t*-Bu)₃)₂+H₂ oxidative addition. *Journal of Chemical Physics*, **100**, 19357–19363.
- 4 Lin, H. and Truhlar, D.G. (2007) QM/MM: What have we learned, where are we, and where do we go from here? *Theoretical Chemistry Accounts*, **117** (2), 185–199.
- 5 Senn, H.M. and Thiel, W. (2009) QM/MM methods for biomolecular systems. *Angewandte Chemie International Edition*, **48** (7), 1198–1229.
- 6 Warshel, A. (1991) *Computer modeling of chemical reactions in enzymes and solutions*, John Wiley & Sons.
- 7 Warshel, A. and Florian, J. (2004) *Empirical valence bond and related approaches*, John Wiley & Sons, Ltd.
- 8 Aqvist, J. and Warshel, A. (1993) Simulation of enzyme reactions using valence bond force fields and other hybrid quantum/classical approaches. *Chemical Reviews*, **93**, 2523–2544.
- 9 Shurki, A. and Warshel, A. (2003) Structure/function correlations of proteins using mm, QM/MM and related approaches; methods, concepts, pitfalls and current progress. *Advances in Protein Chemistry*, **66**, 249–312.

- 10 Shaik, S. and Shurki, A. (1999) Valence bond diagrams and chemical reactivity. *Angewandte Chemie International Edition Eng*, **38** (5), 586–625.
- 11 Shurki, A. (2006) Valence bond – rebirth of the phoenix or relic from the stone age. *Theoretical Chemistry Accounts*, **116** (1-3), 253–261.
- 12 Ushararni, D., Lai, W., Li, C. *et al.* (2014) A tutorial for understanding chemical reactivity through the valence bond approach. *Chemical Society Reviews*, **43**, 4968–4988.
- 13 Amitay, M. and Shurki, A. (2009) The structure of G117H mutant of butyrylcholinesterase: Nerve agents scavenger. *Proteins-Structure Function and Bioinformatics*, **77** (2), 370–377.
- 14 Amitay, M. and Shurki, A. (2011) Hydrolysis of organophosphate compounds by mutant butyrylcholinesterase: A story of two histidines. *Proteins-Structure Function and Bioinformatics*, **79** (2), 352–364.
- 15 Shurki, A. and Warshel, A. (2004) Why does the ras switch “Break” by oncogenic mutations? *Proteins*, **55**, 1–10.
- 16 Prasad, R.B., Plotnikov, N.V., Lameira, J. *et al.* (2013) Quantitative exploration of the molecular origin of the activation of GTPase. *Proceedings of the National Academy of Sciences USA*, **110** (51), 2050–20514.
- 17 Adamczyk, A.J. and Warshel, A. (2011) Converting structural information into an allosteric-energy-based picture for elongation factor tu activation by the ribosome. *Proceedings of the National Academy of Sciences USA*, **108** (24), 9827–9832.
- 18 Roca, M., Messer, B., Hilvert, D. *et al.* (2008) On the relationship between folding and chemical landscapes in enzyme catalysis. *Proceedings of the National Academy of Sciences USA*, **105** (37), 13877–13882.
- 19 Sharir-Ivry, A., Varatharaj, R. and Shurki, A. (2015) Challenges within the linear response approximation when studying enzyme catalysis and effects of mutations. *Journal of Chemical Theory and Computation*, **11**, 293–302.
- 20 Bjelic, S., Brandsdal, B.O. and Aqvist, J. (2008) Cold adaptation of enzyme reaction rates. *Biochemistry*, **47** (38), 10049–10057.
- 21 Isaksen, G.V., Aqvist, J., Brandsdal, B.O. (2014). Protein surface softness is the origin of enzyme cold-adaptation of trypsin. *PLOS Computational Biology*, **10** (8), e1003813.
- 22 Kazemi, M. and Aqvist, J. (2015) Chemical reaction mechanisms in solution from brute force computational arrhenius plots. *Nature Communications*, **6** (7293), 1–7.
- 23 Ansbacher, T. and Shurki, A. (2012) Predicting the coordination number within copper chaperones: Atx1 as case study. *Journal of Physical Chemistry B*, **116** (15), 4425–4432.
- 24 Ansbacher, T., Chourasia, M. and Shurki, A. (2013) Copper-chaperones with dicoordinated Cu(I) – unique protection mechanism. *Proteins: Structure Function and Bioinformatics*, **81**, 1411–1419.
- 25 Muller, R.P. and Warshel, A. (1995) Ab initio calculations of free energy barriers for chemical reactions in solution. *Journal of Physical Chemistry*, **99**, 17516–17524.
- 26 Bentzien, J., Muller, R.P., Florian, J. *et al.* (1998) Hybrid ab initio quantum mechanics/molecular mechanics calculations of free energy surfaces for enzymatic reactions: The nucleophilic attack in subtilisin. *Journal of Physical Chemistry B*, **102**, 2293–2301.

- 27 Wesolowski, T., Muller, R.P. and Warshel, A. (1996) Ab initio frozen density functional calculations of proton transfer reactions in solution. *Journal of Physical Chemistry*, **100**, 15444–15449.
- 28 Strajbl, M., Hong, G. and Warshel, A. (2002) Ab-initio QM/MM simulation with proper sampling: “First principle” Calculations of the free energy of the auto-dissociation of water in aqueous solution. *Journal of Physical Chemistry B*, **106**, 13333–13343.
- 29 Rosta, E., Klahn, M. and Warshel, A. (2006) Towards accurate ab initio QM/MM calculations of free-energy profiles of enzymatic reactions. *Journal of Physical Chemistry B*, **110** (6), 2934–2941.
- 30 Plotnikov, N.V., Kamerlin, S.C.L. and Warshel, A. (2011) Paradynamics: An effective and reliable model for ab initio QM/MM free-energy calculations and related tasks. *Journal of Physical Chemistry B*, **115**, 7950–7962.
- 31 Mones, L., Kulhanek, P., Simon, I. *et al.* (2009) The energy gap as a universal reaction coordinate for the simulation of chemical reactions. *Journal of Physical Chemistry B*, **113** (22), 7867–7873.
- 32 Sharir-Ivry, A., Rajapandian, V. and Shurki, A. (2015) Valence bond and enzyme catalysis: A time to breakdown and a time to buildup. *Journal of Chemistry – A European Journal*, **21**, 7159–7169.
- 33 Kamerlin, S.C.L. and Warshel, A. (2010) The EVB as a quantitative tool for formulating simulations and analyzing biological and chemical reactions. *Faraday Discussions*, **145**, 71–106.
- 34 Kamerlin, S.C.L. and Warshel, A. (2010) At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? *Proteins*, **78**, 1339–1375.
- 35 Kamerlin, S.C.L. and Warshel, A. (2010) An analysis of all the relevant facts and arguments indicates that enzyme catalysis does not involve large contributions from nuclear tunneling. *Journal of Physical Organic Chemistry*, **23**, 677–684.
- 36 Kamerlin, S.C.L. and Warshel, A. (2011) The empirical valence bond model: Theory and applications. *Wiley Interdisciplinary Reviews-Computational Molecular Science*, **1** (1), 30–45.
- 37 Olsson, M.H.M., Parson, W.W. and Warshel, A. (2006) Dynamical contributions to enzyme catalysis: Critical tests of a popular hypothesis. *Chemical Reviews*, **106** (5), 1737–1756.
- 38 Warshel, A., Sharma, P.K., Kato, M. *et al.* (2006) Electrostatic basis for enzyme catalysis. *Chemical Reviews*, **106** (8), 3210–3235.
- 39 Liu, H. and Warshel, A. (2009) Tunnelling does not contribute significantly to enzyme catalysis, but studying temperature dependence of isotope effects is useful, in *Quantum tunnelling in enzyme-catalysed reactions* (eds R.K. Allemann and N.S. Scrutton), Royal Society of Chemistry.
- 40 Jorn, R., Savage, J. and Voth, G.A. (2012) Proton conduction in exchange membranes across multiple length scales. *Accounts of Chemical Research*, **45** (11), 2002–2010.
- 41 Layfield, J.P. and Hammes-Schiffer, S. (2014) Hydrogen tunneling in enzymes and biomimetic models. *Chemical Reviews*, **114** (7), 3466–3494.
- 42 Sharir-Ivry, A. and Shurki, A. (2014) Valence-bond based hybrid quantum mechanics molecular mechanics approaches and proper inclusion of the effect of the surrounding. *Israel Journal of Chemistry*, **54**, 1189–1204.

- 43 Glowacki, D.R., Orr-Ewing, A.J. and Harvey, J.N. (2015) Non-equilibrium reaction and relaxation dynamics in a strongly interacting explicit solvent: $F+CD_3CN$ treated with a parallel multi-state evb model. *Journal of Chemical Physics*, **143** (4), 044120-1–044120-18.
- 44 Hanoian, P., Liu, C.T., Hammes-Schiffer, S. *et al.* (2015) Perspectives on electrostatics and conformational motions in enzyme catalysis. *Accounts of Chemical Research*, **48** (2), 482–489.
- 45 Millard, C.B., Lockridge, O. and Broomfield, C.A. (1995) Design and expression of organophosphorus acid anhydride hydrolase activity in human butyrylcholinesterase. *Biochemistry*, **34** (49), 15925–15933.
- 46 Kamerlin, S.C.L., Sharma, P.K., Prasad, R.B. *et al.* (2013) Why nature really chose phosphate. *Quarterly Review of Biophysics*, **46** (1), 1–132.
- 47 Glennon, T.M., Villa, J. and Warshel, A. (2000) How does GAP catalyze the GTPase reaction of Ras? A computer simulation study. *Biochemistry*, **39**, 9641–9651.
- 48 Schweins, T., Langen, R. and Warshel, A. (1994) Why have mutagenesis studies not located the general base in ras p21. *Nature Structural Biology*, **1** (7), 476–484.
- 49 Schweins, T., Geyer, M., Scheffzek, K. *et al.* (1995) Substrate-assisted catalysis as a mechanism for GTP hydrolysis of p21^{ras} and other GTP-binding proteins. *Nature Structural Biology*, **2** (1), 36–44.
- 50 Grigorenko, B.L., Rogov, A.V. and Nemukhin, A.V. (2006) Mechanism of triphosphate hydrolysis in aqueous solution: QM/MM simulations in water clusters. *Journal of Physical Chemistry B*, **110**, 4407–4412.
- 51 Langen, R., Schweins, T. and Warshel, A. (1992) On the mechanism of guanosine triphosphate hydrolysis in ras p21 proteins. *Biochemistry*, **31**, 8691–8696.
- 52 Plotnikov, N.V., Prasad, B.R., Chakrabarty, S. *et al.* (2013) Quantifying the mechanism of phosphate monoester hydrolysis in aqueous solution by evaluating the relevant ab initio QM/MM free-energy surfaces. *Journal of Physical Chemistry B*, **117** (42), 12807–12819.
- 53 Khrenova, M.G., Grigorenko, B.L., Kolomeisky, A.B. *et al.* (2015) Hydrolysis of guanosine triphosphate (GTP) by the ras-gap protein complex: Reaction mechanism and kinetic scheme. *Journal of Physical Chemistry B*, **119**, 12838–12845.
- 54 Shnerb, T., Lin, N. and Shurki, A. (2007) What is the role of the helical domain of $G_{s\alpha}$ in the GTPase reaction? *Biochemistry*, **46** (38), 10875–10885.
- 55 Strajbl, M., Shurki, A. and Warshel, A. (2004) Converting conformational changes to electrostatic energy in molecular motors: The energetics of atp synthase. *Proceedings of the National Academy of Sciences USA*, **100** (25), 14834–14839.
- 56 Vamvaca, K., Vogeli, B., Kast, P. *et al.* (2004) An enzymatic molten globule: Efficient coupling of folding and catalysis. *Proceedings of the National Academy of Sciences USA*, **101** (35), 12860–12864.
- 57 Bemporad, F., Gsponer, J., Hopearuoho, H.I. *et al.* (2008) Biological function in a non-native partially folded state of a protein. *EMBO Journal*, **27** (10), 1525–1535.
- 58 Frushicheva, M.P., Cao, J., Chu, Z.T. *et al.* (2010) Exploring challenges in rational enzyme design by simulating the catalysis in artificial kemp eliminase. *Proceedings of the National Academy of Sciences USA*, **107** (39), 16869–16874.
- 59 Frushicheva, M.P. and Warshel, A. (2012) Towards quantitative computer-aided studies of enzymatic enantioselectivity: The case of candida antarctica lipase a. *ChemBioChem*, **13**, 215–223.

- 60 Singh, M.K., Chu, Z.T. and Warshel, A. (2014) Simulating the catalytic effect of a designed mononuclear zinc metalloenzyme that catalyzes the hydrolysis of phosphate triesters. *Journal of Physical Chemistry B*, **118**, 12146–12152.
- 61 Xiang, Y., Oelschlaeger, P., Florian, J. *et al.* (2006) Simulating the effect of DNA polymerase mutations on transition-state energetics and fidelity: Evaluating amino acid group contribution and allosteric coupling for ionized residues in human pol beta. *Biochemistry*, **45** (23), 7036–7048.
- 62 Aqvist, J., L., K.S.C. (2015). Exceptionally large entropy contributions enable the high rates of GTP hydrolysis on the ribosome. *Scientific Reports*, **5**, 15817.
- 63 Ansbacher, T., Srivastava, H.K., Martin, J.M.L. *et al.* (2010) Can DFT methods correctly and efficiently predict the coordination number of copper(I) complexes? A case study. *Journal of Computational Chemistry*, **31** (1), 75–83.
- 64 Sham, Y.Y., Chu, Z.T., Tao, H. *et al.* (2000) Examining methods for calculations of binding free energies: LRA, LIE, PDL-D-LRA, and PDL-D/S-LRA calculations of ligands binding to an HIV protease. *Proteins*, **39**, 393–407.
- 65 Shurki, A. and Crown, H.A. (2005) Hybrid ab initio VB/MM method – a valence bond ride through classical landscapes. *Journal of Physical Chemistry B*, **109** (49), 23638–23644.
- 66 Sharir-Ivry, A., Crown, H.A., Wu, W. *et al.* (2008) Density embedded VB/MM: A hybrid ab initio VB/MM with electrostatic embedding. *Journal of Physical Chemistry A*, **112**, 2489–2496.
- 67 Sharir-Ivry, A. and Shurki, A. (2008) VB/MM – the validity of the underlying approximations. *Journal of Physical Chemistry B*, **112** (39), 12491–12497.
- 68 Sharir-Ivry, A., Shnerb, T., Strajbl, M. *et al.* (2010) VB/MM protein landscapes: A study of the S_N2 reaction in haloalkane dehalogenase. *Journal of Physical Chemistry B*, **114**, 2212–2218.
- 69 Shurki, A., Derat, E., Barrozoc, A. *et al.* (2015) How valence bond theory can help you understand your (bio)chemical reaction. *Chemical Society Reviews*, **44**, 1037–1052.
- 70 Mo, Y. (2003). Geometrical optimization for strictly localized structures. *Journal of Chemical Physics*, **119** (3), 1300–1306.
- 71 Sharir-Ivry, A. and Shurki, A. (2008) A VB/MM view of the identity S_N2 valence-bond state correlation diagram in aqueous solution. *Journal of Physical Chemistry A*, **112** (50), 13157–13163.
- 72 Rosta, E., Haranczyk, M., Chu, Z.T. *et al.* (2008) Accelerating QM/MM free energy calculations: Representing the surroundings by an updated mean charge distribution. *Journal of Physical Chemistry B*, **112**, 5680–5692.
- 73 Kamerlin, S.C.L., Haranczyk, M. and Warshel, A. (2009) Progress in ab initio QM/MM free energy simulations of electrostatic energies in proteins: Accelerated QM/MM studies of pKa, redox reactions and solvation free energies. *Journal of Physical Chemistry B*, **113**, 1253–1272.
- 74 Laio, A. and Parrinello, M. (2002) Escaping free-energy minima. *Proceedings of the National Academy of Sciences USA*, **99** (20), 12562–12566.
- 75 Fothergill, M., Goodman, M.F., Petruska, J. *et al.* (1995) Structure-energy analysis of the role of metal ions in phosphodiester bond hydrolysis by DNA polymerase i. *Journal of the American Chemical Society*, **117** (47), 11619–11627.
- 76 Weinan, E. and Vanden-Eijnden, E. (2010) Transition-path theory and path-finding algorithms for the study of rare events. *Annual Review of Physical Chemistry*, **61**, 391–420.

- 77 Dellago, C., Bolhuis, P.G. and Chandler, D. (1999) On the calculation of reaction rate constants in the transition path ensemble. *Journal of Chemical Physics*, **110**, 6617–6625.
- 78 Roston, D. and Kohen, A. (2010) Elusive transition state of alcohol dehydrogenase unveiled. *Proceedings of the National Academy of Sciences USA*, **107** (21), 9572–9577.
- 79 Henkelman, G. and Uberuaga, B.P. (2000) A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *Journal of Chemical Physics*, **113** (22), 9901–9904.
- 80 Doron, D., Kohen, A. and Major, D.T. (2012) Collective reaction coordinate for hybrid quantum and molecular mechanics simulations: A case study of the hydride transfer in dihydrofolate reductase. *Journal of Chemical Theory and Computation*, **8**, 2484–2496.
- 81 Gertner, B.J., Wilson, K.R. and Hynes, J.T. (1989) Nonequilibrium solvation effects on reaction-rates for model S_N2 reactions in water. *Journal of Chemical Physics*, **90**, 3537–3558.
- 82 Chuang, Y.Y. and Truhlar, D.G. (1999) Nonequilibrium solvation effects for a polyatomic reaction in solution. *Journal of the American Chemical Society*, **121**, 10157–1016.
- 83 Villa, J. and Warshel, A. (2001) Energetics and dynamics of enzymatic reactions. *Journal of Physical Chemistry B*, **105**, 7887–7907.
- 84 Olsson, M.H.M., Mavri, J. and Warshel, A. (2006) Transition state theory can be used in studies of enzyme catalysis: Lessons from simulations of tunnelling and dynamical effects in lipoxxygenase and other systems. *Philosophical Transactions of the Royal Society B-Biological Sciences*, **361** (1472), 1417–1432.
- 85 Marti, S., Andres, J., Silla, E. *et al.* (2007) Computer-aided rational design of catalytic antibodies: The 1f7 case. *Angewandte Chemie International Edition*, **46** (1-2), 286–290.
- 86 Bash, P.A., Field, M.J., Davenport, R.C. *et al.* (1991) Computer simulation and analysis of the reaction pathway of triosephosphate isomerase. *Biochemistry*, **30**, 5826–5832.
- 87 Perakyla, M. and Kollman, P.A. (1997) A simulation of the catalytic mechanism of aspartylglucosaminidase using ab initio quantum mechanics and molecular dynamics. *Journal of the American Chemical Society*, **119**, 1189–1196.
- 88 Funke, S.A., Otte, N., Eggert, T. *et al.* (2005) Combination of computational pre-screening and experimental library construction can accelerate enzyme optimization by directed evolution. *Protein Engineering, Design and Selection*, **18** (11), 509–514.
- 89 Dinner, A.R., Blackburn, G.M. and Karplus, M. (2001) Uracil-DNA glycosylase acts by substrate autocatalysis. *Nature*, **413**, 752–754.
- 90 Szefczyk, B., Mulholland, A.J., Ranaghan, K.E. *et al.* (2004) Differential transition-state stabilization in enzyme catalysis: Quantum chemical analysis of interactions in the chorismate mutase reaction and prediction of the optimal catalytic field. *Journal of the American Chemical Society*, **126**, 16148–16159.
- 91 Wong, K.F., Watney, J.B. and Hammes-Schiffer, S. (2004) Analysis of electrostatics and correlated motions for hydride transfer in dihydrofolate reductase. *Journal of Physical Chemistry B*, **108**, 12231–12241.

7

The Empirical Valence Bond Approach as a Tool for Designing Artificial Catalysts

Monika Fuxreiter¹ and Letif Mones^{2,3}

¹MTA-DE Momentum Laboratory of Protein Dynamics, Department of Biochemistry and Molecular Biology, University of Debrecen, Hungary

²Department of Engineering, University of Cambridge, United Kingdom

³Mathematics Institute, University of Warwick, United Kingdom

7.1 Introduction

At the molecular level, enzymatic catalysis is a complex phenomenon involving a fine-tuned set of interactions between the substrate and the enzyme all along the chemical event, which results in several orders of magnitude increase of the catalytic rate constant (k_{cat}) as compared to the corresponding reference reaction in solution.

The active site of the enzyme includes residues, which are involved in the chemical reaction either directly by forming covalent bonds or indirectly via electrostatic interactions. Although the active site residues stand out in catalysis due to their geometric proximity to the substrate, other distant regions of the enzyme can also provide significant contributions. The impact of the whole enzymatic environment – including residues that are located further from the active center – are less trivial to assess. The proposal, that the enzyme acts as a “super solvent” of the reaction has gradually gained recognition and is now widely accepted.^[1] Therefore, atomistic simulations of reactions should not be limited to the sole investigation of the active site using gas phase simulations or cluster models;^[2] instead, adequate models will need to consider the effect of the entire enzymatic/solvent dynamics.

In the last 30 years, owing to the remarkable improvement of simulation approaches, such as hybrid quantum mechanics/molecular mechanics (QM/MM) methods and sampling approaches, several enzymatic reactions have been investigated and their mechanisms have been revealed in atomistic details. In most cases, the interpretation of these multiscale simulations has been primarily based on transition state theory (TST),^[3] which relates the activation free energy to the reaction kinetic rate, enabling direct comparison of the simulation-computed free energy profile to experimental results:

$$k = \frac{\kappa}{\beta h} e^{-\beta \Delta F^\ddagger} \quad (7.1)$$

Here, k is the rate constant, κ is the transmission coefficient, h is the Planck constant, $\beta = \frac{1}{k_{\text{B}} T}$ is the inverse temperature (with Boltzmann constant k_{B} and temperature T) and ΔF^\ddagger is the activation (Helmholtz) free energy.

Formally, the activation barrier in Eq. 7.1 can be reduced in the enzymatic environment either by destabilizing the ground state (GSD) or stabilizing the transition state (TSS).^[4,15] Over the years, several proposals, based on either structural and/or computational studies, have been put forward in order to understand the origin of enzyme catalysis. These include (i) the steric strain hypothesis,^[5] (ii) dynamical effects,^[6,7] (iii) correlated modes,^[8] (iv) near attack conformations (NAC),^[9] (v) entropy effects of the proximity of reactants,^[10] and (vi) desolvation effects^[11] (see Table 7.1). As it will be detailed below, the catalytic effect stems from optimized interactions with the transition state rather than from unfavorable contacts with the ground state. In addition to the understanding gained about several biologically relevant reactions, computational techniques have made it also possible to examine individual catalytic factors quantitatively. This can serve as a basis for enzyme engineering or design to achieve desired activities. Using simulation approaches it has been found^[4] that although each of the listed proposals can contribute to the catalytic power of enzymes, the extent of the individual contributions of these factors can vary from enzyme to enzyme and most of their effect is moderate (only few kcal·mol⁻¹; < 10⁴-fold increase in k_{cat}). Instead, the major catalytic effect has been associated to the electrostatic preorganization of the active site,^[12] which can lead to $\sim 10^7 - 10^8$ fold increase in k_{cat} .^[13] This observation is consistent with the prevalent concept that the stabilization of the transition state is primarily responsible for the catalytic effect.^[4,14,15]

Identifying and understanding those factors, which drive enzyme catalysis, is important not only to explain the molecular mechanism, but also to predict the effect of specific mutations. This connects simulations of enzymatic catalysis to rational enzyme design. In this chapter, we present an overview of the current status of *in silico* enzyme design and show how the Empirical Valence Bond (EVB) method can be applied as a quantitative tool in current and future applications.

7.2 Proposals for the Origin of the Catalytic Effect

To reduce the activation barrier of a given reaction in an enzyme, as compared to the uncatalyzed reaction in solution, the following equation should be considered:

$$\Delta\Delta F_{w \rightarrow p}^\ddagger = \Delta F_p^\ddagger - \Delta F_w^\ddagger = \Delta\Delta H_{w \rightarrow p}^\ddagger - T\Delta\Delta S_{w \rightarrow p}^\ddagger \quad (7.2)$$

where F^\ddagger , H^\ddagger and S^\ddagger are the activation free energy, enthalpy and entropy, respectively, of the reaction in water solution (w) and enzymatic environment (p) and T is the temperature. According to Eq. 7.2, the catalytic effect could be exerted by optimizing either the enthalpic or the entropic terms. In Table 7.1 we summarize different proposals for the origin of the catalytic effect.

It is straightforward to believe that enzyme active sites bring reactants into proximity, which greatly reduces the activation entropy (*proximity effect*).^[16] This idea is derived from organic chemical reactions, in which neighboring substituents are utilized to increase catalytic rates.^[17] Comparing the rates of mono- versus bimolecular gas-phase reactions also support the *proximity effect* and *entropy* as the main catalytic factor(s). In solution, however, the situation is considerably different. The concentration of water in bulk water is 55M, which sets an upper limit for the “effective concentration” of the reactants. Hence the maximum gain in the catalytic effect can be estimated

Table 7.1 Theories Describing the Origin of Enzymatic Catalysis. Opposing reference corresponds to the study, which argues against the given hypothesis.

Theory	Definition	Example	Reference	Opposing Reference
Proximity effect	Reactants are closer to each other in the enzyme active site than in solution (see also chelate effects).		[16,10]	[20]
Entropic effect	The configurational space of the enzyme bound substrate is significantly reduced as compared to that in solution.	cytidine deaminase	[19,10]	[20,33,34]
Steric strain	The substrate conformation is constrained.	lysozyme	[5]	[21]
Near attack conformation (NAC)	Reactant configurations with close interaction distances are preferred by the enzyme.	chorismate mutase	[9]	[22]
Orbital steering	Angular displacement from linear overlap in the transition state of the enzyme-substrate complex.	ribozyme	[18]	
Quantum tunneling	Tunneling effects for light atoms, especially hydrogen transfer increase the preexponential factor.	glucose oxidase	[35]	[36]
Dynamic effects	Dynamic contributions significantly modulate the preexponential factor (e.g., via coherent motions, nonequilibrium solvation).	pancreatic trypsin inhibitor, dihydrofolate reductase, adenylate kinase	[37,6,38]	[39]
Correlated modes	Large correlated backbone motions are required for catalysis.	dihydrofolate reductase	[40]	[41]
Low-barrier hydrogen bonds	Enzymes form low-barrier (partially covalent) hydrogen bonds with charged substrates.	chymotrypsin, serine proteases	[29,42]	[43]
Medium effect	Nonpolar environment accelerates the reaction.	orotidine 5'-monophosphate decarboxylase	[11]	[32]
Desolvation	Enzymes work similarly to gas-phase reactions using hydrophobic active sites.	acetylcholine esterase	[30,31]	[13]
Electrostatic preorganization	Enzyme dipoles form a preorganized environment, which establishes optimized electrostatic interactions with the transition state.	>50 enzymes (lysozyme, chorismate mutase, staphylococcus nuclease, acetylcholine esterase, orotidine 5'-monophosphate decarboxylase)	[4]	

as $-\beta^{-1} \ln(V_w/V_p)$, which provides $\sim 1.5 \text{ kcal mol}^{-1}$ reduction in the activation free energy depending on the volume of the active site.^[1] The real *entropic effect* originates from decreasing the conformational freedom of the substrates within the active center (e.g., in *orbital steering*).^[16,18,19] This could also be quantified by comparing the size of the configurational space of the substrate in enzyme and in solution^[1] using well-defined potential energy surfaces. As proteins are flexible entities, they cannot constrain their substrates considerably.^[20] Strain is distributed over many interatomic contacts between the enzyme and the substrate, which could be characterized by low spring constants.^[5] Thus, proteins cannot exert much force by distorting their substrates.^[21] It has been also shown by Warshel and coworkers that favorable electrostatic interactions provide a major contribution to the *near attack conformation* (NAC), as well.^[22] *Orbital steering* is a variant of reactant destabilization hypotheses, which also deals with constraining the configurational space,^[18] but again it cannot provide large catalytic effects.^[1] The importance of dynamic effects – that is, the enzyme promotes certain vibrational modes, which are relevant for catalysis – is still intensively discussed, albeit the concept dates back at least three decades. For a variety of enzymes (e.g., in dihydrofolate reductase,^[8] cyclophilin A,^[23] purine nucleoside phosphorylase,^[24] triose phosphate isomerase^[25]) connection between motions of different timescales were concluded as the main catalytic factor. Systematic studies of these effects by EVB showed that they do not have major impact on shaping the energy landscape of the reaction (the chemical step), rather they influence substrate binding. Correlated motions can increase the transmission coefficient and thus accelerate the reaction, but these motions also exist in solution.^[26] Similarly, correlated motions can increase the efficiency of hydrogen tunneling, which, however, do not differ significantly between catalyzed and uncatalyzed reactions.^[27,28] The impact of other exotic effects, such as partially covalent low-barrier hydrogen bonds^[29] can be examined by studying the pH dependence of the enzymatic reactions. This proposal as well as the desolvation hypothesis^[30,31] cannot reproduce the proper ionization state of the catalytic groups and can lead to anti-catalysis.^[13,32] The same argument holds for the medium effect, where the enzyme is proposed to provide a hydrophobic, gas-phase like medium.

To explain the catalytic effect of enzymes, it is crucial to define the proper reference reaction in solution and consider the same mechanism. The only proposal, which can quantitatively explain how enzymes work and produce catalytic rate constants comparable to the experimental ones is electrostatic preorganization.^[4] Accordingly, enzymes have evolved to provide optimal solvation with the transition state of the reaction, which is mostly achieved by electrostatic interactions. The energy to preorganize enzyme dipoles is invested during the folding process. In uncatalyzed reactions in solution, the dipoles of the solvent have to reorient constantly to follow charge redistributions along the reaction. In enzymes, the optimal environment is “pre-set” such that it significantly decreases the reorganization energy (see calculations below), which provides the largest contribution to the catalytic effect. In acetylcholine esterase, for example, the decrease in reorganization energy is 10 kcal mol^{-1} out of the 15 kcal mol^{-1} reduction in the activation barrier.

7.3 Reorganization Energy

Reorganization energy was introduced by Marcus in his renowned theory of electron transfer reactions.^[44,45] According to the original concept, the reorganization energy of an elementary reaction step is the free energy required for the hypothetical transition of the system between its product and reactant equilibrium states, while both are in the electronic valence state of the product. Figure 7.1 illustrates the meaning of this concept, that is, the reorganization energy is a free energy difference accompanying the vertical transition between the diabatic reactant and product free energy curves at the reactant state:

$$\lambda = F_{\text{PS}}(\xi_{\text{RS}}) - F_{\text{PS}}(\xi_{\text{PS}}) \quad (7.3)$$

where F_{PS} is the free energy profile calculated on the diabatic product state, ξ is the reaction coordinate, and the subscripts RS and PS designate the reactant and product states, respectively.

In general, the reorganization energy can be decomposed into two parts:

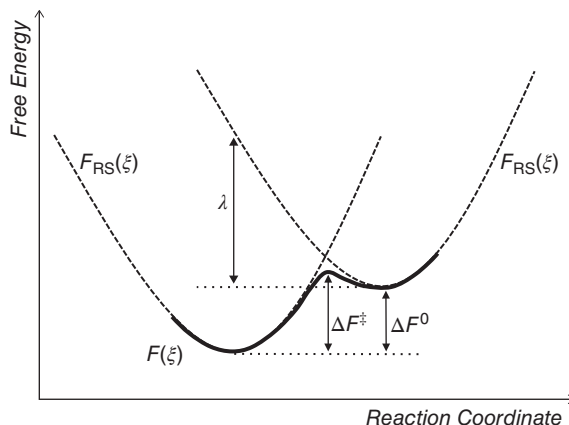
$$\lambda = \lambda_{\text{in}} + \lambda_{\text{out}} \quad (7.4)$$

where λ_{in} and λ_{out} are the inner- and outer-sphere (usually medium or solvent) reorganization energies. The former corresponds to the reorganization energy accompanying the structural changes of the reactants, whereas the latter is due to the reorganization of the environment.

If the diabatic free energy profiles are described by two parabolas with the same curvature, then the intersection of the curves represents the free energy of the transition state. Within the framework of these approximations, Marcus theory introduces a very useful relationship between the activation free energy (ΔF^\ddagger), the reaction free energy (ΔF^0), and the reorganization energy (λ):

$$\Delta F^\ddagger \approx F_{\text{RS}}(\xi_{\text{TS}}) - F_{\text{RS}}(\xi_{\text{RS}}) = \frac{(\Delta F^0 + \lambda)^2}{4\lambda} \quad (7.5)$$

Figure 7.1 Illustration of the Definition of the Reorganization Energy Concept on a Hypothetical Free Energy Profile.



Here, F_{RS} is the diabatic free energy profile on the reactants surface. Marcus theory was later extended by Warshel and coworkers^[1,13] to reactions in solutions and enzymatic environments, taking also into account the effect of the coupling between the diabatic states. This results in the modified Marcus equation:^[1,46]

$$\Delta F^\ddagger \approx \frac{(\Delta F^0 + \lambda)^2}{4\lambda} - \langle H_{12} \rangle_{\text{TS}} + \frac{\langle H_{12} \rangle_{\text{RS}}^2}{\Delta F^0 + \lambda} \quad (7.6)$$

where $\langle H_{12} \rangle$ denotes the average value of the off-diagonal element at the corresponding reactant or transition state.

Equations 7.5 and 7.6 provide a qualitative guide to understanding how the activation barrier can be reduced in enzymes as compared to the reference reaction in aqueous solution. In general, reducing the reaction free energy (ΔF^0) and/or the reorganization energy (λ) leads to an increase in catalytic efficiency. We must note that due to the dependence of ΔF^\ddagger on λ and ΔF^0 , the relationship is more complicated (see Eq. 7.5 and 7.6). Schematically, enzymes can change the relative position of the diabatic free energy functions both vertically and horizontally and the combination of the two results in the ultimate decrease of the activation barrier. The two extreme cases are presented in Figure 7.2.

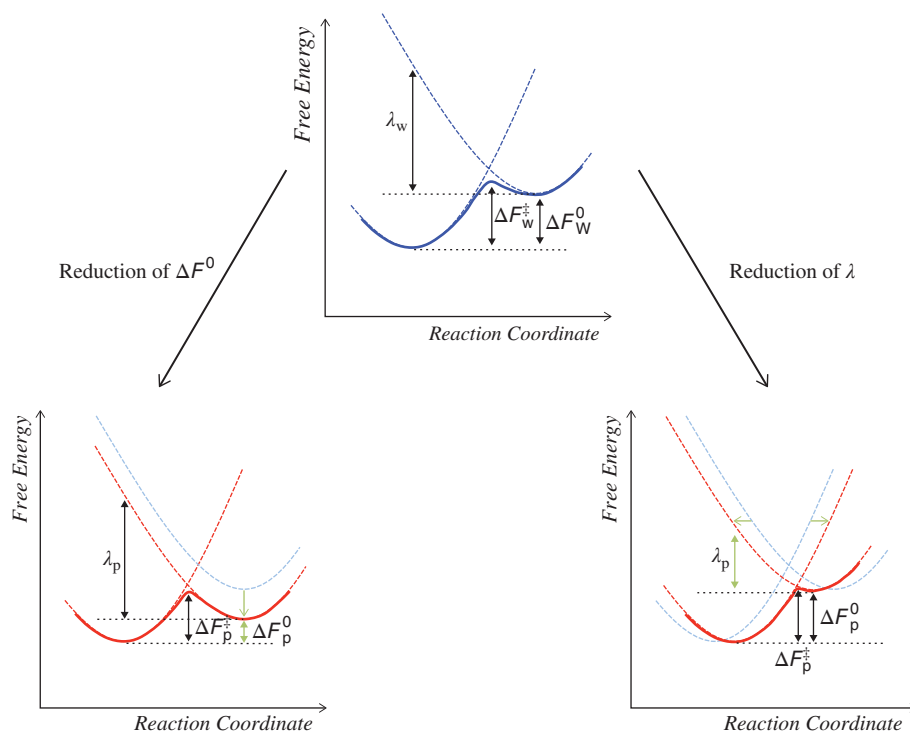


Figure 7.2 Schematic Representation of the Two Extreme Cases of How Enzymes Reduce Catalytic Barriers by Shifting the Relative Position of the Diabatic States Vertically to Reduce ΔF^0 and Horizontally to Reduce λ . Reprinted from reference [47]. Copyright (2014), with permission from Elsevier. (See color plate section for the color representation of this figure.)

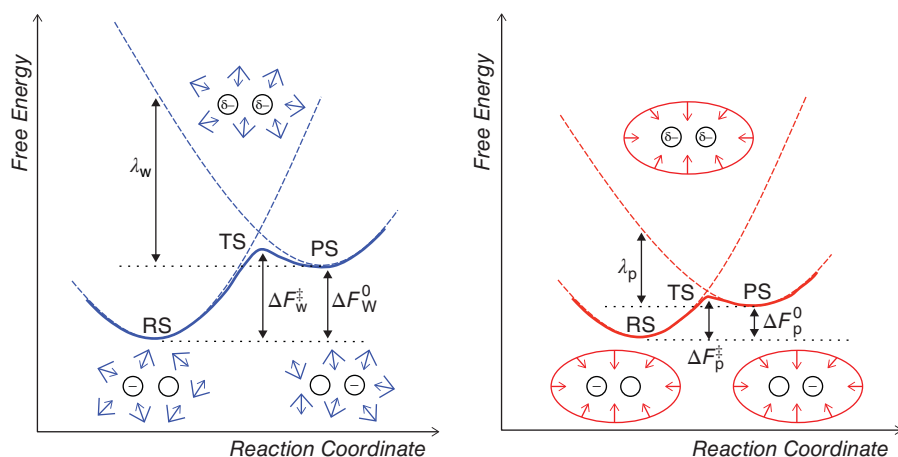


Figure 7.3 Geometrical Representation of the Preorganization Effect in Water Solution (Left) and Enzymatic Environment (Right). Reprinted from reference [47]. Copyright (2014), with permission from Elsevier. (See color plate section for the color representation of this figure.)

From a chemical point of view, the significant difference between the solvent reorganization energies in water solution and the enzymatic environment can be explained by the change in the orientation of the dipoles in these two environments. In water solution, the solvent dipoles have to follow the rearrangement of the reactant dipoles as the reaction proceeds, which requires significant reorganization. In contrast, in enzymes the dipoles are already preorganized to optimally interact with the substrate throughout the reaction (Figure 7.3).

Several quantitative studies of enzymatic reactions demonstrated the importance of reorganization energy in lowering the activation barrier. For instance, investigation of the diabatic free energy profiles of a two-step catalytic reaction catalyzed by lactate dehydrogenase (LDH) revealed that the enzyme significantly lowers the solvent reorganization energy of the intermediate state as compared to the solution (26 kcal mol^{-1}), which led to the reduction of the activation barrier by $\sim 6 \text{ kcal mol}^{-1}$.^[48] In the case of acetylcholinesterase (AChE), a $\sim 40 \text{ kcal mol}^{-1}$ decrease of the reorganization energy was observed for the rate determining nucleophilic attack step as compared to the reference water reaction.^[13] This effect provided the majority of the total catalytic effect by reducing the activation barrier by $\sim 10 \text{ kcal mol}^{-1}$. Feierberg and Åqvist carried out a series of simulations on enzymatic keto-enol isomerization reactions and found that the reduction of the reorganization energy contributed to the TS stabilization by 30%, 34% and 40% for the glyoxalase I (GlxI), triosephosphate isomerase (TIM) and ketosteroid isomerase (KSI), respectively.^[49] All these studies and many others^[4] demonstrate that the reorganization energy is the major source of the catalytic effect of enzymes.

7.4 Conventional In Silico Enzyme Design

De novo enzyme design, to engineer novel enzymes that catalyze reactions for which no natural catalysts exist, is an intensively developing field.^[50] This process consists

of two main steps: i) *in silico* design to provide a computational model^[51,52] and ii) experimental optimization of the designed variant. The computational models usually have moderate catalytic activities. Thus subsequent experimental optimization is required, which in most cases is carried out using the directed evolution technique, including random mutagenesis and natural selection^[53,54] (*in vitro* step). *In vitro* optimizations can improve the catalytic rate constant by two to three orders of magnitude.

The first stage can be further divided into the following steps:^[50] i) Determination of the transition state (TS) geometry and position of key functional groups to obtain the maximal stabilization of the TS, using gas phase *ab initio* quantum calculations. This small, theoretical model of the catalytic reaction is called *theozyme*.^[55] This step limits the number of applicable functional groups of the model (in most cases three or four), as complicated theozymes would prevent straightforward scaffold selection; ii) Then an appropriate scaffold, which can optimally embed the designed active center (theozyme), is selected based on available crystal structures using geometrical constraints for an optimal fit with the TS model and the surrounding functional groups; iii) Finally, the scaffold candidates are ranked based on geometry, calculated interaction energy with the theozyme, and binding energy with the substrate and optionally, additional active site optimizations are carried out (Figure 7.4).

This computational design process primarily optimizes the local *medium* by selecting the corresponding functional groups and their tight packing (i.e., proximity effect). Medium effect however provides only limited contribution to catalysis, as it has been

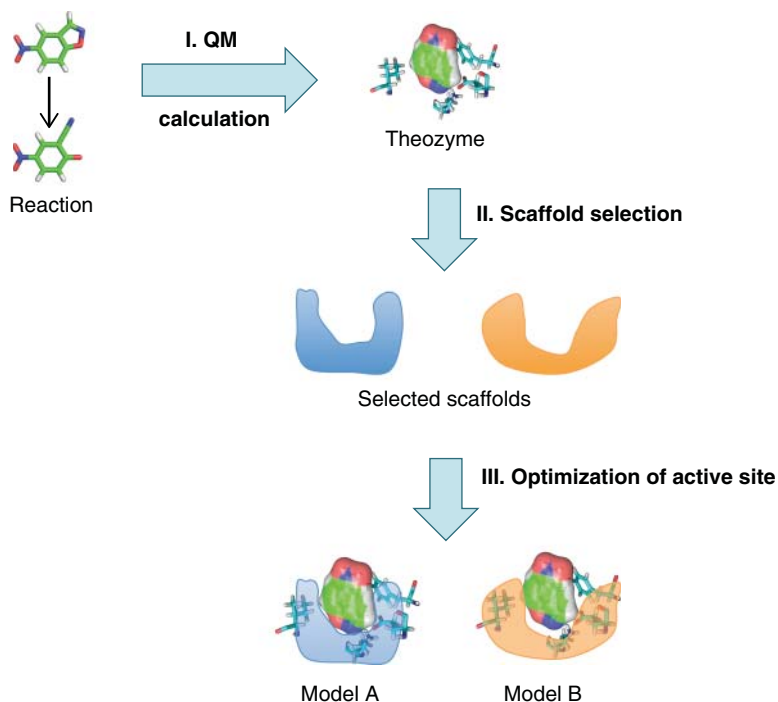


Figure 7.4 Standard Protocol of *In Silico* Enzyme Design. (See color plate section for the color representation of this figure.)

shown earlier.^[56,57] Increasing the *packing* within and around the active site might constrain the reactants to certain conformations (entropic effects, NAC), but at the same time also optimizes electrostatic interactions by optimizing the orientation of the corresponding functional groups.^[58] The latter effect accounts for decreasing the activation barrier.^[4] In case of the highly stereoselective and substrate specific biomolecular Diels-Alder reaction, for example, the active site pocket of the design was tightly filled by hydrophobic residues.^[58] Mutational analysis, however, showed that this nonspecific hydrophobic pocket did not contribute to the reduction of the activation barrier (no *medium* effect), and the major catalytic effect arose from the close packing, which resulted in an optimal orientation of the key functional groups. Another aspect of the *packing* effect is its influence on the structure and subtle electrostatic interactions at the active site. In the case of an active retro-aldolase design, the tight packing led to an optimal hydrogen-bonding network, perturbation of which hampered catalysis.^[56] It is instructive to note that this hydrogen-bonded network also involved crystallographic water molecules. Simultaneous mutation of water-coordinating residues increased the mobility of the water molecules and, therefore, decreased the preorganization cost at the active center, which led to a significant $\sim 10^3$ -fold reduction of the catalytic activity.^[59]

The limited size of the theozyme and the requirement for the tight fit also decreases the structural *flexibility* of the active site. “Wobbling” indeed can be unfavorable by rearranging the interaction network at the active site.^[59] On the other hand, correlated motions could increase the transmission factor and facilitate accessing the active site.^[60–62] A series of examples illustrate that increased *packing* optimizes *preorganization of the active site* rather than optimizes catalysis by *desolvation*.^[47]

To illustrate how computer-aided design process work, we select the Kemp elimination as a model reaction. The enzyme catalysts developed for this man-made reaction are one of the most successful examples of computer-aided enzyme design.^[63,61,64] The overall net reaction of the Kemp elimination is the conversion of 5-nitrobenzisoxazole to cyanophenol^[65,63] (Figure 7.5). The reaction requires a general base (:B), which deprotonates the nitrobenzisoxazole. Proton transfer from the carbon atom takes place in a concerted manner with the cleavage of the nitrogen-oxygen bond, which triggers ring-opening and leads to the formation of the product.

Design of the active site was carried out using density functional theory (DFT) calculations, where the relative orientation of three key residues around the reactant were optimized.^[63] For the general base two models were considered: i) a carboxyl group using glutamate or aspartate and ii) an imidazole group of histidine coupled to the carboxyl group of an aspartate (His-Asp dyad). Several hydrogen-bond donors were used to interact with the phenolic oxygen atom of the substrate (including Lys, Arg, Ser, Tyr, His, and

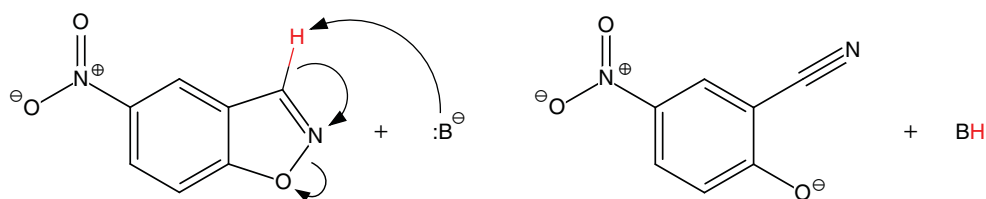


Figure 7.5 Reaction Catalyzed by Kemp Elimination Catalysts.

Table 7.2 Scaffolds and Key Residues of the Most Effective Kemp Eliminase Designs.

Theozyme	Scaffold	General base	H-bond donor	π -stacking stabilizer	k_{cat} (s^{-1})
KE07 ^[63]	TIM barrel (1THF)	Glu (E101)	Lys (K222)	Trp (W50)	0.018
KE70 ^[63]	TIM barrel (1JCL)	His-Asp diad (H17-D45)	Ser (S138)	Tyr (Y48)	0.16
KE59 ^[63]	α/β barrel (1A53)	Glu (E230)	Ser (S179, S210)	Trp (W109)	0.29
HG3 ^[61]	TAX (1GOR)	Asp (D127)	Thr (T265)	Trp (W44)	0.68

water). Finally, the aromatic moiety of the substrate was stabilized by π -stacking using a tryptophan. Based on these criteria, several theozymes have been constructed. Scaffolds were selected using the Rosetta program,^[66] the most successful 3D structure prediction method today. The most important features of the Kemp eliminase models discussed in this chapter are presented in Table 7.2.

The resulted enzyme structures were based on either a TIM barrel scaffold (e.g., KE07 and KE70 variants, PDB codes 1THF and 1JCL, respectively) or an α/β barrel scaffold (e.g., KE59 variant, PDB code 1A53). High-resolution crystallographic structure information on the KE07 model showed an excellent match with the designed structure, indicating the accuracy of the scaffold construction. However, the catalytic activity of the most promising candidates (KE07, KE59, and KE70) was found to be relatively low, with $k_{\text{cat}} / K_{\text{M}}$ values of $10^1 - 10^2 \text{ M}^{-1}\text{s}^{-1}$. The increase of the enzymatic catalytic rate constant, as compared to the uncatalyzed reaction ($k_{\text{cat}} / k_{\text{uncat}}$), was also found to be moderate, only three to four orders of magnitude in contrast to >10 orders of magnitude for natural enzymes.

Directed evolution experiments were subsequently conducted to improve the catalytic activity of the designed models. After seven rounds of in vitro evolution, the catalytic efficiency of the KE07 enzyme converged, resulting in a 116-fold improvement of $k_{\text{cat}} / K_{\text{M}}$ relative to the originally designed KE07 model.^[67] Out of the total 8 mutations found during the evolution experiments, three active site hydrophobic residues were substituted by polar ones, which increased the $\text{p}K_{\text{a}}$ of the general base and improved the preorganization of the polar groups within the active site. In addition, surface mutations (K19 and K146) increased the structural stability of the enzyme. In the case of KE70, nine rounds of random mutagenesis increased catalytic efficiency ($k_{\text{cat}} / K_{\text{M}}$) by ~ 400 -fold relative to the original model, mostly owing to optimizing electrostatic interactions.^[68] Amongst the KE series, KE59 was found to be the most effective variant, with a $k_{\text{cat}} = 9.5 \text{ s}^{-1}$ and a corresponding 2000-fold increase in catalytic activity after 13 rounds of directed evolution.^[69] Here, the mutations were proposed to increase the hydrophobicity of the active site, resulting in an elevated $\text{p}K_{\text{a}}$ of the catalytic base (E230).

An alternative iterative approach has also been applied to obtain a highly effective catalyst of the Kemp elimination reaction, using a similar initial model (HG1) as the KE59 design.^[61] The contacts in HG1 were essentially the same as those found with the catalytic antibody 34E4.^[70] In this model, a glutamate or aspartate residue was used

as the general base, an aromatic residue was used to establish π -stacking interaction with the aromatic moiety of the substrate, and serine/threonine/tyrosine residues were used as the H-bonding donor. A modified xylanase of *Thermoascus aurantiacus* (TAX),^[71] with seven mutations based on the repacking of the active site, was used as scaffold and its xylan binding pocket served as potential active site. However, HG1 showed no Kemp elimination activity. X-ray crystallography and molecular dynamic (MD) simulations revealed that the inactivity of the first-generation design was due to the presence of solvent and the looseness of key residues at the active site. Subsequent iteration with structural and dynamical characterization resulted in *tighter packing* of the key residues and *increased hydrophobicity* of the active site. HG2 employed an aspartate (D127) as the general base, a tryptophan (W44) as the π -stacking residue and a serine (S265) as H-bond donor. Additional MD analysis suggested that a threonine residue as hydrogen-bond donor could provide better packing and improved catalytic activity. This third generation design (HG3) did indeed lead to a threefold increase of $k_{\text{cat}} / K_{\text{M}}$ ($425 \text{ M}^{-1}\text{s}^{-1}$) and to a considerable improvement in rate ($k_{\text{cat}} / k_{\text{uncat}} = 5.9 \times 10^5$). A further, 17 rounds of directed evolution led to an enzyme (HG3.17), which catalyzed the Kemp reaction with a k_{cat} of 700 s^{-1} and 6×10^6 fold rate acceleration as compared to the uncatalyzed reaction.^[64] The performance of this variant is in reasonable agreement with that of naturally evolved enzymes.

In the last case, the higher catalytic activity was assigned to medium effect and increased hydrophobicity, which raises the $\text{p}K_{\text{a}}$ of the catalytic base.^[72] However, pH-rate profiles for the designed and evolved enzymes were similar, which argues against *medium* effects. Instead, an alternative substrate conformation in the evolved enzyme, which allowed a tighter packing and improved stabilization of the negatively charged transition state, was suggested. This study provides an excellent example of how *dynamical effects* and *preorganization* are important in order to improve the evolvability of the initial designs.

As illustrated by many examples, the moderate efficiency of the *in silico* enzyme designs might be related to the lack of the major catalytic factor: the reorganization energy.

7.5 Computational Analysis of Kemp Eliminases

To underscore this point, reorganization energy has been studied in the evolved Kemp variants.^[73] EVB simulations on several models of the KE07 showed an excellent correlation between the computed and experimental activation barriers,^[73,74] corroborating that the initial KE07 design does not lead to significant increase of the rate constant as compared to the reference reaction in water. The catalytic effect was due to the entropic effect resulting from placing the corresponding functional groups (reactant, general acid, and base) in the same cage.

It was found that within the margin of error the reorganization energy of the initial design does not differ from that of the corresponding reference reaction in water solution. This is in a good agreement with both experimental and theoretical studies.^[75] A significant decrease of $27.4 \text{ kcal mol}^{-1}$ of the reorganization energy, however, was observed between the wild type and the evolved R7 1/3H variant, which indicates that directed evolution actually optimized the electrostatic preorganization. Thus, structural

and mechanistic studies of the Kemp variants also revealed that the reorganization energy is optimized by in vitro evolution.^[73]

In the following sections we will show how the EVB framework was used to draw the conclusions above and, more generally, how it can be applied for computing the catalytic effect and reorganization energy of enzymes.

7.6 Using the Empirical Valence Bond Approach to Determine Catalytic Effects

7.6.1 General EVB Framework

A detailed description of the empirical valence bond (EVB) approach can be found elsewhere,^[1,76,77] including Chapter 2 of this book. Here, for convenience, we re-summarize only the most important features of the method, which are exploited for computer-aided enzyme design.

Within the EVB approach, in its simplest form, the system is represented by two resonance structures that correspond to the reactant and product valence states. The ground state potential energy surface of an elementary reaction step is then calculated as the lower eigenvalue of a 2×2 Hamiltonian:

$$\mathbf{H} = \begin{pmatrix} H_{11} & H_{12} \\ H_{12}^* & H_{22} \end{pmatrix} \quad (7.7)$$

where H_{11} and H_{22} correspond to the potential energy functions of the reactant and product states, respectively and H_{12} is the coupling between the two states. In this case, the ground state reaction energy function (i.e., the lower eigenvalue of \mathbf{H}) can be written as:

$$E_g = \frac{H_{11} + H_{22}}{2} - \frac{1}{2} \sqrt{(H_{11} - H_{22})^2 + 4H_{12}^2} \quad (7.8)$$

It is important to note that although EVB applies a quantum mechanical formalism, the individual elements in Eq. 7.7 can be and are in practice usually computed using classical molecular mechanical approaches. The bond breaking and formation steps are described by special energy terms, as will be described below. In other words, the EVB formalism follows the QM framework, but technically it merits the simplicity and speed of molecular mechanics. Taken together, the EVB approach is a good tool to construct a reactive force field using interatomic potentials for the pure valence states and an exponential or more complex function (see Chapters 2 and 4) for approximating the off-diagonal elements:

$$H_{ii} = V_{\text{qq}}^i + V_{\text{qs}}^i + V_{\text{ss}}^i + \alpha^i \quad (7.9)$$

$$H_{ij} = \delta \exp[-\{\mu(r_{\text{ab}} - r_{\text{ab}}^0) + \rho(r_{\text{ab}} - r_{\text{ab}}^0)^2\}] \quad (7.10)$$

where i and j designate the diabatic valence states, V_{qq} is the (bonded and nonbonded) potential energy of the EVB “quantum region”, that is, those atoms that change their atom types and force field parameters during the chemical process, V_{qs} is the interaction

energy between the quantum region and solvent, and V_{ss} is the solvent-solvent interaction energy. α is an offset parameter that is responsible to adjust the surfaces of the different valence states to reproduce gas phase energetics. In the off-diagonal element δ , μ , ρ and r_{ab}^0 are parameters and r_{ab} is a characteristic distance between the atoms of the bond to be formed/broken during the reaction.

It is important to note that in the EVB method, the off-diagonal elements depend solely on the interaction of valence states and not on the environment around the reactants, that is, their functional form is the same in gas phase, (aqueous) solution, and an enzyme active site. This approximation also gives the possibility to calibrate the parameters by reproducing the free energy profile (i.e., activation barrier and reaction free energy) of some reference reactions. Usually, the reaction in water solution is used as a reference state, but in specific cases gas phase reactions can also be employed. For studying the effect of mutations that do not alter the reaction mechanism on activities, the reaction in the wild type enzyme can also serve as a reference.

7.6.2 Computing Free Energy Profiles Within the EVB Framework

Simulation of the reference reaction aims to create the ground state potential E_g by parameterizing the empirical parameters used in Eqs. 7.9 through 7.10. To this end, the reaction is carried out on a series of mapping potentials (E_m), which are simple linear combination of the diagonal elements according to the conventional stratification procedure of the Free Energy Perturbation theory:^[78]

$$E_m^\gamma = (1 - \gamma)H_{11} + \gamma H_{22} \quad \gamma : 0 \rightarrow 1 \quad (7.11)$$

where γ is an order parameter that is used to transform the system from the reactant state ($\gamma = 0$) to the product state ($\gamma = 1$). Extensive molecular dynamics simulations are carried out on each mapping potential (window), and the free energy differences between the consecutive windows are computed ($\delta F_{k \rightarrow k+1}$). The free energy profile of the ground state reaction potential energy surface then can be obtained by applying the Zwanzig formula:^[78]

$$F(\xi) = \left\langle \sum_{k=0}^{i-1} \delta F_{k \rightarrow k+1} - \beta^{-1} \ln \langle \delta(\xi - \xi(\mathbf{r})) \exp\{-\beta[E_g - E_m^\gamma]\} \rangle_\gamma \right\rangle_i \quad (7.12)$$

where the outer and inner angle brackets indicate an average over all and the individual mapping potentials, respectively and $\xi(\mathbf{r})$ is a genuine reaction coordinate. The parameters are calibrated iteratively by using Eq. 7.12. The resulting parameters can be employed to simulate the reaction on E_g using the reaction coordinate directly (i.e., computing and applying its bias on-the-fly during the dynamics).^[79] Owing to code implementation and historical reasons, however, the same formalism is applied as for the reference reaction.

The EVB method has been successfully applied to study many enzymatic reactions,^[4] to reproduce the experimentally observed reaction rates,^[73–75] and identify major catalytic factors, which cannot be assessed experimentally. EVB has also been applied to study other theoretical aspects of catalysis, for example, the connection between folding and chemical energy landscapes.^[80]

7.7 Computing the Reorganization Energy

Computing the reorganization energy using the EVB formalism is straightforward. According to Eq. 7.3, the reorganization energy is equal to the free energies difference evaluated on the diabatic product resonance state at the reactant and product states. Similarly to Eq. 7.12, the free energy profile of the product valence state can be computed as:

$$F_{\text{PS}}(\xi) = \left\langle \sum_{k=0}^{i-1} \delta F_{k \rightarrow k+1} - \beta^{-1} \ln \langle \delta(\xi - \xi(\mathbf{r})) \exp\{-\beta[H_{22} - E_m^{\gamma}]\} \rangle_{\gamma} \right\rangle_i \quad (7.13)$$

Note that the only difference compared to Eq. 7.12 is the use of H_{22} instead of E_g .

As the reorganization energy is evaluated based on a hypothetical diabatic state, which is a mathematical construction to describe the reaction, its experimental determination is not feasible. Using theoretical approaches, however, it is possible to calculate the reorganization energy in complex systems. A possible approach is to compute the inner (reactants) and outer (surroundings) reorganization energies separately (according to Eq. 7.4). There are standard methods to calculate the inner-sphere reorganization energy. For example, in the case of the most widely used Nelsen's four-point method^[81,82] the oxidants and reductants are separated and the inner reorganization energy is computed based on four energy values as

$$\lambda_{\text{in}}^{\text{Nelson}} = [E(D^+|D) - E(D^+|D^+)] + [E(A^-|A) - E(A^-|A^-)] \quad (7.14)$$

where $E(X|Y)$ is the energy of state X calculated at the equilibrium structure of state Y and A/A^- and D/D^+ denote the acceptor and donor species in their neutral and charged states, respectively. In a more sophisticated approach a constrained DFT method (CDFT)^[83] is applied. This method also approximates the points of the corresponding free energy curves by simple energy calculations but this time the reactants are not separated:

$$\lambda_{\text{in}}^{\text{CDFT}} = E(D^+A^-|DA) - E(D^+A^-|D^+A^-) \quad (7.15)$$

where the same notation is used as before. This method's advantage is that it requires only two constrained DFT geometry optimizations.

The outer part of the reorganization energy has to be computed separately, which is a more complicated task due to the non-equilibrium polarization effect of the environment.^[84,85] Alternatively, the two components of the reorganization energy can be computed simultaneously. In general, computing the total reorganization energy in solution and enzymatic environment requires: i) a chemically accurate potential energy surface (PES) for modeling the reaction, ii) a reaction coordinate that distinguishes the reactant and product states and finds the proper pathway between them, iii) identifying the value of this reaction coordinate at the reactant and product states, iv) a method, which is able to provide the PES of the product diabatic state, and v) extensive sampling to obtain statistical averages. It is very difficult to meet all these requirements. For example, the previously mentioned CDFT^[83] approach has been combined with the hybrid QM/MM to provide diabatic PES for the entire system, but this method is still relatively expensive for performing extensive dynamics. So far, the empirical valence bond approach seems to be the most suitable method for computing reorganization

energies.^[48,86–88] Within this framework, the calculation of both the adiabatic and diabatic states is straightforward and an excellent general reaction coordinate can be defined (E_{gap}), which outperforms the widely applied simple geometric variables.^[79,89]

7.8 E_{gap} : A General Reaction Coordinate and its Application on Other PES

Both Eqs. 7.12 and 7.13 require a reaction coordinate $\xi(\mathbf{r})$ along the corresponding free energy profiles. EVB offers the opportunity to apply an excellent reaction coordinate, termed as energy gap (E_{gap}).^[90] E_{gap} , which was originally introduced by Marcus,^[44,45] is an energy based collective variable defined as the difference between the instantaneous energy values of the diabatic states:

$$E_{\text{gap}}(\mathbf{r}) = H_{11}(\mathbf{r}) - H_{22}(\mathbf{r}) \quad (7.16)$$

Unlike simple geometrical collective variables, which are mostly based on chemical intuition and focus only on the geometry of the reactants, E_{gap} contains all relevant degrees of freedom of the system, including those of the environment.^[79] Thus, in the EVB formalism, any relevant mode, even far from the active site, could be considered as part of the reaction coordinate via the application of E_{gap} . This is an important difference because considering only a limited set of degrees of freedom can lead to an inaccurate TS, and consequently wrong energetics (and so free energies). For example, as it was demonstrated,^[79] even for a simple nucleophilic substitution reaction in water solution, the geometrical coordinates such as interatomic distance or difference in different bond lengths failed to catch the correct TS ensemble, whereas E_{gap} precisely identified the TS.

Since the diabatic energies are computed in EVB on the fly, E_{gap} is generally applied in such simulations. It is also important to stress that E_{gap} is not limited to the EVB framework, rather, it can be used in any quantum chemistry Hamiltonian^[79] and in combination with any available (but preferably adaptive) potential of mean force technique. In these simulations, the molecular dynamics calculation is performed on a higher-level energy surface (e.g., DFT), but the atomic forces are modified according to an E_{gap} dependent bias, which is calculated within the EVB framework.^[79] In other words, the final atomic forces (\mathbf{f}) used for the propagation of the dynamics are computed based on the following concise form:

$$\mathbf{f} = -\nabla E_{\text{QM}} + F_{\text{EVB}}(E_{\text{gap}}(\mathbf{r}))\nabla E_{\text{gap}}(\mathbf{r}) \quad (7.17)$$

The first term in the right-hand side of Eq. 7.17 is the force vector calculated from the underlying QM (or QM/MM) potential, whereas the second term is the bias and $F_{\text{EVB}}(E_{\text{gap}}(\mathbf{r}))$ is the E_{gap} dependent biasing force that can be estimated by either as a derivative of some potential (e.g., Metadynamics^[91]) or directly (e.g., Adaptive Biasing Force^[92]). The peculiarity of this technique is that the computation of the collective variable itself requires potential energy functions (provided by EVB) that differ from the underlying one.

Such a combination gives the possibility to use this generalized reaction coordinate even when the investigation of the enzymatic reaction requires a more accurate potential energy surface (i.e., ab initio QM/MM) than EVB. Although the quality of the EVB

ground state energy surface is determined by force field parameters, which are used to construct the diabatic potential energy surfaces, these parameters do not need to be optimized to obtain a reasonable reaction coordinate. Varying the EVB force field parameters within a reasonable limit, the distinct E_{gap} coordinates still result in the same free energy profiles on the higher-level energy surface.^[89]

Recently, new variants of E_{gap} have been introduced to cope with multistate reactions simulated on higher-level energy surfaces.^[89] In such multistate reactions, at least two chemically equivalent topologies are applicable in the reactant and/or the product states. Typical examples are proton transfer reactions involving water molecule (having two equivalent protons) or carboxylate group (having two equivalent acceptor atoms). For such reactions, conventional E_{gap} would fail since undesired side reactions occur and alter the true free energy profile. From a technical point of view, E_{gap} is a single topology coordinate, that is, the MM potentials H_{11} and H_{22} depend on the atomic indexes and not the chemical identity. This prevents E_{gap} to properly describe valence states that have many alternative but chemically identical connectivities. To overcome this problem, new E_{gap} based collective variables were introduced that are invariant to these chemically equivalent topologies. For example, the minimum energy multistate energy difference (*MinMSED*) coordinate computes all possible reactant and product states' energy value and selects the lowest ones from both sets:

$$\text{MinMSED}(\mathbf{r}) = \min\{H_{11}^i(\mathbf{r})\}_i^{n_{\text{R}}} - \min\{H_{22}^j(\mathbf{r})\}_j^{n_{\text{P}}} \quad (7.18)$$

where the indexes i and j run over all reactant (n_{R}) and product (n_{P}) states, respectively. Although *MinMSED* is a continuous function, its derivatives are not because of the minimum functions. This could cause some numerical instability^[89] for some PMF methods, so an alternative variant was also introduced whose derivatives are also continuous. The energy-weighted multistate energy difference (*EwMSED*) coordinate is based on a Boltzmann-weighted linear combination of all possible reactant and product states:

$$\text{EwMSED}(\mathbf{r}, \tilde{\beta}) = \frac{\sum_i^{n_{\text{R}}} e^{-\tilde{\beta}H_{11}^i(\mathbf{r})} H_{11}^i(\mathbf{r})}{\sum_i^{n_{\text{R}}} e^{-\tilde{\beta}H_{11}^i(\mathbf{r})}} - \frac{\sum_j^{n_{\text{P}}} e^{-\tilde{\beta}H_{22}^j(\mathbf{r})} H_{22}^j(\mathbf{r})}{\sum_j^{n_{\text{P}}} e^{-\tilde{\beta}H_{22}^j(\mathbf{r})}} \quad (7.19)$$

and $\tilde{\beta}$ is a fictitious inverse temperature for the Boltzmann-weighting. It is worth noting that although in principle both *MinMSED* and *EwMSED* require the computation of the energy of all possible reactant and product topologies, the total computation time can be significantly reduced by filtering out the chemically least sensible topologies, using simple geometric criteria. As it has been shown for the symmetrical proton-transfer reaction between a hydronium ion and water molecule (including 10-10 equivalent reactant and product states), both new variants are able to provide correct free energy profiles.^[89]

Although evaluating free energy profiles of mutated enzymes by EVB is significantly less time-consuming than using ab initio QM/MM models, it still requires extensive geometry optimization, relaxation, and sampling. Therefore, calculating free energy profiles of enzymes with random mutations and then ranking them according to the calculated barrier is not an efficient strategy to optimize enzymes. Instead, we propose an alternative approach, which is based on evaluating the individual contribution of residues to the reorganization energy and to the catalytic effect.

7.9 Contribution of Individual Residues

In general, resultant total free energies are not additive and hence, the activation free energy cannot be exactly decomposed into individual contributions. On the other hand, breaking down the catalytic effect into individual terms and quantifying the contributions of residues is essential to designing mutations to improve enzymatic activity. Here, we describe an approach to determine contributions to the activation free energy and the reorganization energy, which can be used for screening in model selection during the design process. It must be emphasized, that owing to the coupling between different interactions (many-body effects), individual contributions should always be considered as approximate values.

The electrostatic interaction energy between atoms of a given residue in the enzymatic environment with the atoms of the QM region is given by:

$$E_{\text{el}}^{\gamma}(n) = \sum_i^{N_{\text{QM}}} \sum_j^{N_n} \frac{q_i^{\gamma} q_j}{r_{ij} \epsilon_{ij}} \quad (7.20)$$

where $E_{\text{el}}^{\gamma}(n)$ is the electrostatic interaction energy of the n th residue with the QM region that is in the state of γ according to the mapping potential (Eq. 7.11). N_{QM} and N_n are the numbers of atoms of the n th residue and the QM region, respectively. $q_i^{\gamma} = (1 - \gamma)q_{i1} + \gamma q_{i2}$ is the approximate point charge of the i th atom of the QM region at mapping state γ and q_j is the point charge of the j th atom of the residue. ϵ_{ij} is the dielectric constant for the specific interaction. The value of ϵ_{ij} depends on the actual simulation conditions,^[73] and can vary between 1 to 20. Now we can define the individual electrostatic contribution to the catalytic effect as:

$$\Delta F_{\text{el}}^{\ddagger}(n) = \langle E_{\text{el}}^{\gamma_{\text{TS}}}(n) \rangle_{\text{TS}} - \langle E_{\text{el}}^{\gamma=0}(n) \rangle_{\text{RS}} \quad (7.21)$$

where γ_{TS} indicates the mapping potential that produces the most configurations around the transition state and the angle brackets refer to an average on the corresponding mapping potential. For the reorganization contribution, we introduce a similar equation but this time all charges of the reactants are considered in their product valence state:

$$\Delta F_{\text{el}}^{\text{reorg}}(n) = \langle E_{\text{el}}^{\gamma=1}(n) \rangle_{\text{RS}} - \langle E_{\text{el}}^{\gamma=1}(n) \rangle_{\text{PS}} \quad (7.22)$$

In case of the KE07 design, electrostatic contributions indicated that replacement of Ala9 by a polar residue (serine or asparagine) can stabilize the negatively charged oxygen atom at the transition state.^[75] The full free energy profile of the mutant enzyme, which was derived from EVB simulations, did not support the prediction. This was also corroborated by independent EVB studies on the wild type KE07 and its evolved variant, where electrostatic interactions were found to provide minor contributions to the catalysis.^[73] Besides the moderate change of electrostatic contribution to the catalytic effect, it was also found that some mutated residues destabilized the transition state (Figure 7.6a). Altogether, these observations indicated that laboratory evolution did not optimize the electrostatic contribution to the TS.

Further analysis of the mutated residues revealed a significant and favorable total change in the reorganization energy contribution.^[73] This suggested that the optimization of the reorganization energy was the driving force during the directed evolution. To test this hypothesis an extensive screening was carried out for all residues within

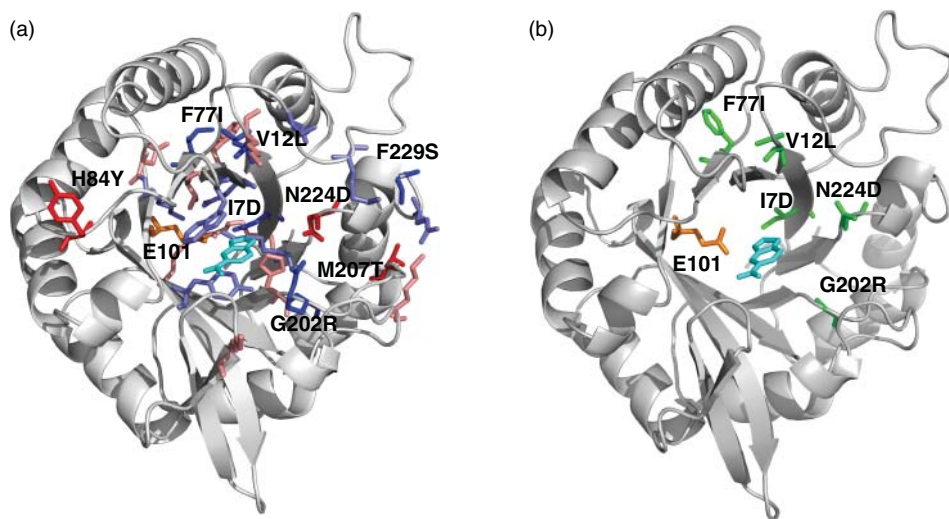


Figure 7.6 (a) The Effect of Directed Evolution on the Electrostatic Contribution of Residues, Which Significantly Contribute to Catalysis (Favorable – Light Blue and Unfavorable – Light Red) and Residues that Were Replaced During the In Vitro Evolution (Favorable – Dark Blue and Unfavorable – Dark Red). (b) Mutations of KE07 design that occurred during directed evolution and were also predicted based on their contributions to the reorganization energy (green). The substrate (cyan) and key residue Glu101 (orange) are also displayed in both cases. Reprinted from reference [47]. Copyright 2014, with permission from Elsevier. (See color plate section for the color representation of this figure.)

12 Å of the donor carbon atom of the substrate. Mutations were carried out for each residue using 5 amino acids with different chemical characters (alanine, phenylalanine, serine, glutamic acid and lysine). After a short geometry optimization, the reorganization energy contributions of the in silico variants were determined and compared to the KE07 design. Mutations with favorable or neutral reorganization energy contributions predicted by this algorithm were in accord with the active variants that were isolated during direct evolution (Figure 7.6b).^[73]

7.10 Improving Rational Enzyme Design by Incorporating the Reorganization Energy

To incorporate the reorganization energy into the scaffold selection step of the computer-aided enzyme design, an efficient screening procedure based on this quantity has to be developed. To this end, Eq. 7.13, which applies to the construction of the full free-energy profile, needs to be simplified. An economic way to estimate reorganization energy is to use the linear response approximation (LRA),^[94] which requires configurations only at the reactant and product states:

$$\begin{aligned} \lambda &= \langle H_{22} - H_{11} \rangle_{\text{RS}} + \frac{1}{2} (\langle H_{22} - H_{11} \rangle_{\text{RS}} + \langle H_{22} - H_{11} \rangle_{\text{PS}}) \\ &= \frac{1}{2} (\langle H_{22} - H_{11} \rangle_{\text{RS}} - \langle H_{22} - H_{11} \rangle_{\text{PS}}) \end{aligned} \quad (7.23)$$

Indeed, the LRA approach has been successfully applied to many systems.^[93,95]

In silico enzyme design strategies primarily focus on the TS stabilization and lack dynamic effects. As it has been demonstrated in references^[61,64], dynamic effects can be critical for increasing catalytic efficiency. Enzymatic preorganization has been shown to be the key catalytic factor, which is also optimized by directed evolution. If the design, however, lacks considerable flexibility or it is based on optimizing the hydrophobic effect, random mutations and functional pressure cannot improve the design significantly, as has been previously observed.^[50] Hence, incorporation of the reorganization energy in ranking the enzyme models can be pivotal for the initial design and subsequent laboratory optimization of the catalytic activity. Based on these findings, we recently suggested a new computational procedure,^[47] which includes reorganization energies of the whole enzyme and the contributions of individual residues. The flowchart of the suggested procedure is shown in Figure 7.7. The first two steps of this procedure follow the conventional design. Ab initio calculations are employed to obtain the theozyme and to determine the reaction mechanism and TS structure. Based on both steric and electrostatic complementarity a high-throughput scaffold search is carried out using the TS binding energy. The potential scaffolds are then ranked based on their computed global reorganization energy, which is an additional step to the conventional strategy. The selected variants can be further optimized by mutating residues based on the calculation of the individual contributions to the reorganization energy. This process can provide an evolvable enzyme model, where directed evolution has more room to improve the enzymatic activity, making it comparable to that of natural enzymes.

7.11 Conclusions and Outlook

In this chapter, we described how the EVB approach can be implemented in computer-assisted enzyme design to quantify the enzyme features that are the major determinants of the catalytic effect. The EVB approach is a powerful and efficient tool to compute free energy profiles, using a potential energy surface, which is parametrized to experimental data. The EVB approach is computationally more economical than ab initio QM/MM-methods to compare enzymatic reactions of wild type and mutant enzymes, as it eventually applies a reactive force field built from molecular mechanical terms. Another advantage of the EVB approach is the excellent reaction coordinate, E_{gap} , defined within its framework, which – unlike geometrical variables – takes into account the environmental degrees of freedom to compute the PME. This results in more efficient and precise TS localization as well as enhanced sampling as compared to conventional geometric variables. E_{gap} can also be applied in high-level ab initio QM/MM methods as a collective variable.

The diabatic state calculation in the EVB approach is also useful to obtain the reorganization energy, the major factor in enzymatic catalysis. Calculation of this quantity is not as straightforward by other methods because they either address only the inner-sphere reorganization energy or are computationally more expensive.

We have shown that the reorganization energy is optimized by directed evolution of KE07 models and only variants with favorable or negligible contributions to the reduction of reorganization energy have been selected. These observations offer an opportunity to implement the reorganization energy as a feature to be considered

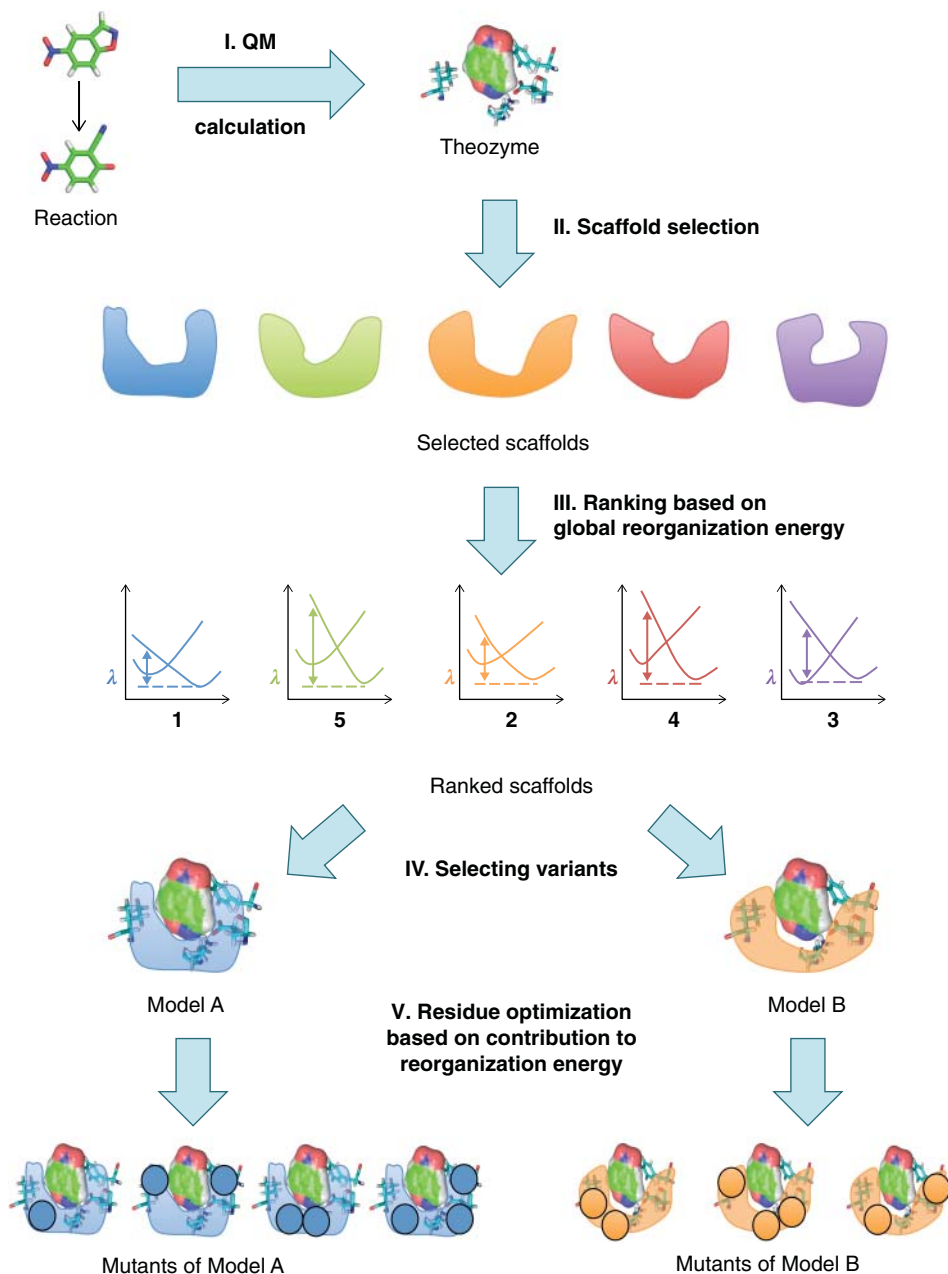


Figure 7.7 Key Steps of the Recently Suggested New Computational Enzyme Design Including the Computation of the Reorganization Energy. Reprinted from reference [47]. Copyright 2014, with permission from Elsevier. (See color plate section for the color representation of this figure.)

the computer-assisted design. Candidates could be ranked based on the reduction of the reorganization energy, while determining the individual contribution of residues to the reorganization energy can be exploited to refine computational models by selected *in silico* random mutations. This methodology has the potential to not only improve experimental evolvability of the designed models, but also to consider different enzymatic mechanisms with more favorable free energy profiles due to the whole enzymatic environment and dynamical effects.

Although the EVB approach is the most straightforward method that could be applied for these purposes, the precision and efficiency of calculations of the reorganization energy can be improved. That is, there is a long way to go in collaboration with experimental laboratories to test the new strategy and develop a computationally efficient approach for *in silico* enzyme design.

Acknowledgements

This work was supported by the Momentum program (LP2012-41) of the Hungarian Academy of Sciences (MF).

References

- 1 Warshel, A. (1991) *Computer modeling of chemical reactions in enzymes and solutions*, John Wiley & Sons, New York.
- 2 Siegbahn, P.E.M. and Himo, F. (2011) The quantum chemical cluster approach for modeling enzyme reactions. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **1** (3), 323–336.
- 3 Eyring, H. (1935) The activated complex in chemical reactions. *Journal of Chemical Physics*, **3** (2), 107–115.
- 4 Warshel, A., Sharma, P.K., Kato, M. *et al.* (2006) Electrostatic basis for enzyme catalysis. *Chemical Reviews*, **106** (8), 3210–3235.
- 5 Blake, C.C.F., Johnson, L.N., Mair, G.A. *et al.* (1967) Crystallographic studies of the activity of hen egg-white lysozyme. *Proceedings of the Royal Society of London B: Biological Sciences*, **167** (1009), 378–388.
- 6 Cameron, C.E. and Benkovic, S.J. (1997) Evidence for a functional role of the dynamics of glycine-121 of *Escherichia coli* dihydrofolate reductase obtained from kinetic analysis of a site-directed mutant. *Biochemistry*, **36** (50), 15792–15800.
- 7 Eisenmesser, E.Z., Bosco, D.A., Akke, M. *et al.* (2002) Enzyme dynamics during catalysis. *Science*, **295** (5559), 1520–1523.
- 8 Agarwal, P.K., Billeter, S.R., Rajagopalan, P.T.R. *et al.* (2002) Network of coupled promoting motions in enzyme catalysis. *Proceedings of the National Academy of Sciences of the United States of America*, **99** (5), 2794–2799.
- 9 Hur, S. and Bruice, T.C. (2003) The near attack conformation approach to the study of the chorismate to prephenate reaction. *Proceedings of the National Academy of Sciences of the United States of America*, **100** (21), 12015–12020.

- 10 Page, M.I. and Jencks, W.P. (1971) Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect. *Proceedings of the National Academy of Sciences of the United States of America*, **68** (8), 1678–1683.
- 11 Lee, J.K. and Houk, K.N. (1997) A proficient enzyme revisited: The predicted mechanism for orotidine monophosphate decarboxylase. *Science*, **276** (5314), 942–945.
- 12 Štrajbl, M., Shurki, A., Kato, M. *et al.* (2003) Apparent NAC effect in chorismate mutase reflects electrostatic transition state stabilization. *Journal of the American Chemical Society*, **125** (34), 10228–10237.
- 13 Fuxreiter, M. and Warshel, A. (1998) Origin of the catalytic power of acetylcholinesterase: Computer simulation studies. *Journal of the American Chemical Society*, **120** (1), 183–194.
- 14 Pauling, L. (1946) Molecular architecture and biological reactions. *Chemical & Engineering News Archive*, **24** (10), 1375–1377.
- 15 Pauling, L. (1948) Nature of forces between large molecules of biological interest. *Nature*, **161**, 707–709.
- 16 Jencks, W.P. (1986) *Catalysis in chemistry and enzymology*, Dover Publications, New York.
- 17 Bruice, T.C. (1976) Some pertinent aspects of mechanism as determined with small molecules. *Annual Review of Biochemistry*, **45** (1), 331–374.
- 18 Bruice, T.C., Brown, A. and Harris, D.O. (1971) On the concept of orbital steering in catalytic reactions. *Proceedings of the National Academy of Sciences of the United States of America*, **68** (3), 658–661.
- 19 Blow, D. (2000) So do we understand how enzymes work? *Structure*, **8** (4), R77–R81.
- 20 Villà, J., Štrajbl, M., Glennon, T.M. *et al.* (2000) How important are entropic contributions to enzyme catalysis? *Proceedings of the National Academy of Sciences of the United States of America*, **97** (22), 11899–11904.
- 21 Warshel, A. and Levitt, M. (1976) Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology*, **103** (2), 227–249.
- 22 Shurki, A., Štrajbl, M., Villà, J. *et al.* (2002) How much do enzymes really gain by restraining their reacting fragments? *Journal of the American Chemical Society*, **124** (15), 4097–4107.
- 23 Eisenmesser, E.Z., Millet, O., Labeikovsky, W. *et al.* (2005) Intrinsic dynamics of an enzyme underlies catalysis. *Nature*, **438** (7064), 117–121.
- 24 Saen-oon, S., Quaytman-Machleder, S., Schramm, V.L. *et al.* (2008) Atomic detail of chemical transformation at the transition state of an enzymatic reaction. *Proceedings of the National Academy of Sciences of the United States of America*, **105** (43), 16543–16548.
- 25 Kempf, J.G., Jung, J.-y., Ragain, C. *et al.* (2007) Dynamic requirements for a functional protein hinge. *Journal of Molecular Biology*, **368** (1), 131–149.
- 26 Schwartz, S.D. and Schramm, V.L. (2009) Enzymatic transition states and dynamic motion in barrier crossing. *Nature Chemical Biology*, **5** (8), 551–558.
- 27 Cha, Y., Murray, C.J. and Klinman, J.P. (1989) Hydrogen tunneling in enzyme reactions. *Science*, **243** (4896), 1325–1330.
- 28 Cui, Q. and Karplus, M. (2002) Quantum mechanics/molecular mechanics studies of triosephosphate isomerase-catalyzed reactions: Effect of geometry and tunneling on

- proton-transfer rate constants. *Journal of the American Chemical Society*, **124** (12), 3093–3124.
- 29 Frey, P.A., Whitt, S.A. and Tobin, J.B. (1994) A low-barrier hydrogen bond in the catalytic triad of serine proteases. *Science*, **264** (5167), 1927–1930.
 - 30 Wolfenden, R. (1983) Waterlogged molecules. *Science*, **222** (4628), 1087–1093.
 - 31 Dewar, M.J.S. and Dieter, K.M. (1988) Mechanism of the chain extension step in the biosynthesis of fatty acids. *Biochemistry*, **27** (9), 3302–3308.
 - 32 Warshel, A., Štrajbl, M., Villà, J. *et al.* (2000) Remarkable rate enhancement of orotidine 5'-monophosphate decarboxylase is due to transition-state stabilization rather than to ground-state destabilization[†]. *Biochemistry*, **39** (48), 14728–14738.
 - 33 Snider, M.J., Gaunitz, S., Ridgway, C. *et al.* (2000) Temperature effects on the catalytic efficiency, rate enhancement, and transition state affinity of cytidine deaminase, and the thermodynamic consequences for catalysis of removing a substrate “anchor”[†]. *Biochemistry*, **39** (32), 9746–9753.
 - 34 Kazemi, M. and Åqvist, J. (2015) Chemical reaction mechanisms in solution from brute force computational Arrhenius plots. *Nature Communications*, **6**, 7293.
 - 35 Kohen, A. and Klinman, J.P. (1999) Hydrogen tunneling in biology. *Chemistry & Biology*, **6** (7), R191–R198.
 - 36 Hwang, J.K., Chu, Z.T., Yadav, A. *et al.* (1991) Simulations of quantum mechanical corrections for rate constants of hydride-transfer reactions in enzymes and solutions. *Journal of Physical Chemistry*, **95** (22), 8445–8448.
 - 37 Karplus, M. and McCammon, J.A. (1983) Dynamics of proteins: Elements and function. *Annual Review of Biochemistry*, **52** (1), 263–300.
 - 38 Henzler-Wildman, K.A., Thai, V., Lei, M. *et al.* (2007) Intrinsic motions along an enzymatic reaction trajectory. *Nature*, **450** (7171), 838–844.
 - 39 Warshel, A. (1984) Dynamics of enzymatic reactions. *Proceedings of the National Academy of Sciences of the United States of America*, **81** (2), 444–448.
 - 40 Miller, G.P. and Benkovic, S.J. (1998) Deletion of a highly motional residue affects formation of the Michaelis complex for *Escherichia coli* dihydrofolate reductase. *Biochemistry*, **37** (18), 6327–6335.
 - 41 Warshel, A., Parson, W. and W. (2001) Dynamics of biochemical and biophysical reactions: Insight from computer simulations. *Quarterly Reviews of Biophysics*, **34** (04), 563–679.
 - 42 Cleland, W.W. and Kreevoy, M.M. (1994) Low-barrier hydrogen bonds and enzymic catalysis. *Science*, **264** (5167), 1887–1890.
 - 43 Warshel, A. and Papazyan, A. (1996) Energy considerations show that low-barrier hydrogen bonds do not offer a catalytic advantage over ordinary hydrogen bonds. *Proceedings of the National Academy of Sciences of the United States of America*, **93** (24), 13665–13670.
 - 44 Marcus, R.A. (1964) Chemical and electrochemical electron-transfer theory. *Annual Review of Physical Chemistry*, **15** (1), 155–196.
 - 45 Marcus, R.A. (1965) On the theory of electron-transfer reactions. VI. Unified treatment for homogeneous and electrode reactions. *Journal of Chemical Physics*, **43** (2), 679–701.
 - 46 Warshel, A., Hwang, J.K. and Åqvist, J. (1992) Computer simulations of enzymatic reactions: Examination of linear free-energy relationships and quantum-mechanical

- corrections in the initial proton-transfer step of carbonic anhydrase. *Faraday Discussions*, **93**, 225–238.
- 47 Fuxreiter, M. and Mones, L. (2014) The role of reorganization energy in rational enzyme design. *Current Opinion in Chemical Biology*, **21**, 34–41.
 - 48 Yadav, A., Jackson, R.M., Holbrook, J.J. *et al.* (1991) Role of solvent reorganization energies in the catalytic activity of enzymes. *Journal of the American Chemical Society*, **113** (13), 4800–4805.
 - 49 Feierberg, I. and Åqvist, J. (2002) Computational modeling of enzymatic keto-enol isomerization reactions. *Theoretical Chemistry Accounts*, **108** (2), 71–84.
 - 50 Hilvert, D. (2013) Design of protein catalysts. *Annual Review of Biochemistry*, **82** (1), 447–470.
 - 51 Bolon, D.N. and Mayo, S.L. (2001) Enzyme-like proteins by computational design. *Proceedings of the National Academy of Sciences of the United States of America*, **98** (25), 14274–14279.
 - 52 Kaplan, J. and DeGrado, W.F. (2004) De novo design of catalytic proteins. *Proceedings of the National Academy of Sciences of the United States of America*, **101** (32), 11566–11570.
 - 53 Arnold, F.H. (2001) Combinatorial and computational challenges for biocatalyst design. *Nature*, **409** (6817), 253–257.
 - 54 Jäckel, C., Kast, P. and Hilvert, D. (2008) Protein design by directed evolution. *Annual Review of Biophysics*, **37** (1), 153–173.
 - 55 Tantillo, D.J., Jiangang, C. and Houk, K.N. (1998) Theozymes and compuzymes: Theoretical models for biological catalysis. *Current Opinion in Chemical Biology*, **2** (6), 743–750.
 - 56 Jiang, L., Althoff, E.A., Clemente, F.R. *et al.* (2008) De novo computational design of retro-aldol enzymes. *Science*, **319** (5868), 1387–1391.
 - 57 Lassila, J.K., Baker, D. and Herschlag, D. (2010) Origins of catalysis by computationally designed retroaldolase enzymes. *Proceedings of the National Academy of Sciences of the United States of America*, **107** (11), 4937–4942.
 - 58 Siegel, J.B., Zanghellini, A., Lovick, H.M. *et al.* (2010) Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science*, **329** (5989), 309–313.
 - 59 Wang, L., Althoff, E.A., Bolduc, J. *et al.* (2012) Structural analyses of covalent enzyme–substrate analog complexes reveal strengths and limitations of de novo enzyme design. *Journal of Molecular Biology*, **415** (3), 615–625.
 - 60 Kiss, G., Röthlisberger, D., Baker, D. *et al.* (2010) Evaluation and ranking of enzyme designs. *Protein Science*, **19** (9), 1760–1773.
 - 61 Privett, H.K., Kiss, G., Lee, T.M. *et al.* (2012) Iterative approach to computational enzyme design. *Proceedings of the National Academy of Sciences of the United States of America*, **109** (10), 3790–3795.
 - 62 Ruscio, J.Z., Kohn, J.E., Ball, K.A. *et al.* (2009) The influence of protein dynamics on the success of computational enzyme design. *Journal of the American Chemical Society*, **131** (39), 14111–14115.
 - 63 Rothlisberger, D., Khersonsky, O., Wollacott, A.M. *et al.* (2008) Kemp elimination catalysts by computational enzyme design. *Nature*, **453** (7192), 190–195.
 - 64 Blomberg, R., Kries, H., Pinkas, D.M. *et al.* (2013) Precision is essential for efficient catalysis in an evolved Kemp eliminase. *Nature*, **503** (7476), 418–421.

- 65 Kemp, D.S. and Casey, M.L. (1973) Physical organic chemistry of benzisoxazoles. II. Linearity of the Broensted free energy relation for the base-catalyzed decomposition of benzisoxazoles. *Journal of the American Chemical Society*, **95** (20), 6670–6680.
- 66 Zanghellini, A., Jiang, L., Wollacott, A.M. *et al.* (2006) New algorithms and an in silico benchmark for computational enzyme design. *Protein Science*, **15** (12), 2785–2794.
- 67 Khersonsky, O., Röthlisberger, D., Dym, O. *et al.* (2010) Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the KE07 series. *Journal of Molecular Biology*, **396** (4), 1025–1042.
- 68 Khersonsky, O., Röthlisberger, D., Wollacott, A.M. *et al.* (2011) Optimization of the in-silico-designed Kemp eliminase KE70 by computational design and directed evolution. *Journal of Molecular Biology*, **407** (3), 391–412.
- 69 Khersonsky, O., Kiss, G., Röthlisberger, D. *et al.* (2012) Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59. *Proceedings of the National Academy of Sciences of the United States of America*, **109** (26), 10358–10363.
- 70 Debler, E.W., Ito, S., Seebeck, F.P. *et al.* (2005) Structural origins of efficient proton abstraction from carbon by a catalytic antibody. *Proceedings of the National Academy of Sciences of the United States of America*, **102** (14), 4984–4989.
- 71 Lo Leggio, L., Kalogiannis, S., Eckert, K. *et al.* (2001) Substrate specificity and subsite mobility in *T. aurantiacus* xylanase 10A. *FEBS Letters*, **509** (2), 303–308.
- 72 Hu, Y., Houk, K.N., Kikuchi, K. *et al.* (2004) Nonspecific medium effects versus specific group positioning in the antibody and albumin catalysis of the base-promoted ring-opening reactions of benzisoxazoles. *Journal of the American Chemical Society*, **126** (26), 8197–8205.
- 73 Labas, A., Szabo, E., Mones, L. *et al.* (2013) Optimization of reorganization energy drives evolution of the designed Kemp eliminase KE07. *Biochimica et Biophysica Acta (BBA) – Proteins and Proteomics*, **1834** (5), 908–917.
- 74 Frushicheva, M.P., Cao, J. and Warshel, A. (2011) Challenges and advances in validating enzyme design proposals: The case of Kemp eliminase catalysis. *Biochemistry*, **50** (18), 3849–3858.
- 75 Frushicheva, M.P., Cao, J., Chu, Z.T. *et al.* (2010) Exploring challenges in rational enzyme design by simulating the catalysis in artificial Kemp eliminase. *Proceedings of the National Academy of Sciences of the United States of America*, **107** (39), 16869–16874.
- 76 Warshel, A. and Weiss, R.M. (1980) An empirical valence bond approach for comparing reactions in solutions and in enzymes. *Journal of the American Chemical Society*, **102** (20), 6218–6226.
- 77 Warshel, A. and Weiss, R.M. (1981) Empirical valence bond calculations of enzyme catalysis. *Annals of the New York Academy of Sciences*, **367** (1), 370–382.
- 78 Zwanzig, R.W. (1954) High-temperature equation of state by a perturbation method. I. Nonpolar gases. *The Journal of Chemical Physics*, **22** (8), 1420–1426.
- 79 Mones, L., Kulhánek, P., Simon, I. *et al.* (2009) The energy gap as a universal reaction coordinate for the simulation of chemical reactions. *The Journal of Physical Chemistry B*, **113** (22), 7867–7873.

- 80 Roca, M., Messer, B., Hilvert, D. *et al.* (2008) On the relationship between folding and chemical landscapes in enzyme catalysis. *Proceedings of the National Academy of Sciences of the United States of America*, **105** (37), 13877–13882.
- 81 Nelsen, S.F., Blackstock, S.C. and Kim, Y. (1987) Estimation of inner shell Marcus terms for amino nitrogen compounds by molecular orbital calculations. *Journal of the American Chemical Society*, **109** (3), 677–682.
- 82 Rauhut, G. and Clark, T. (1993) Electron-transfer reactions: AM1 and ab initio studies on self-exchange in *p*-diaminobenzene systems. *Journal of the American Chemical Society*, **115** (20), 9127–9135.
- 83 Wu, Q. and Van Voorhis, T. (2006) Direct calculation of electron transfer parameters through constrained density functional theory. *The Journal of Physical Chemistry A*, **110** (29), 9212–9218.
- 84 Liu, Y.-P. and Newton, M.D. (1995) Solvent reorganization and donor/acceptor coupling in electron-transfer processes: Self-consistent reaction field theory and ab initio applications. *The Journal of Physical Chemistry*, **99** (33), 12382–12386.
- 85 Ren, H.-S., Ming, M.-J., Ma, J.-Y. *et al.* (2013) Theoretical calculation of reorganization energy for electron self-exchange reaction by constrained density functional theory and constrained equilibrium thermodynamics. *The Journal of Physical Chemistry A*, **117** (33), 8017–8025.
- 86 Åqvist, J. and Fothergill, M. (1996) Computer simulation of the triosephosphate isomerase catalyzed reaction. *Journal of Biological Chemistry*, **271** (17), 10010–10016.
- 87 Warshel, A. (1998) Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. *Journal of Biological Chemistry*, **273** (42), 27035–27038.
- 88 Villà, J. and Warshel, A. (2001) Energetics and dynamics of enzymatic reactions. *The Journal of Physical Chemistry B*, **105** (33), 7887–7907.
- 89 Mones, L. and Csányi, G. (2012) Topologically invariant reaction coordinates for simulating multistate chemical reactions. *The Journal of Physical Chemistry B*, **116** (51), 14876–14885.
- 90 Warshel, A. (1982) Dynamics of reactions in polar solvents. Semiclassical trajectory studies of electron-transfer and proton-transfer reactions. *The Journal of Physical Chemistry*, **86** (12), 2218–2224.
- 91 Laio, A. and Parrinello, M. (2002) Escaping free-energy minima. *Proceedings of the National Academy of Sciences of the United States of America*, **99** (20), 12562–12566.
- 92 Darve, E. and Pohorille, A. (2001) Calculating free energies using average force. *The Journal of Chemical Physics*, **115** (20), 9169–9183.
- 93 Muegge, I., Tao, H. and Warshel, A. (1997) A fast estimate of electrostatic group contributions to the free energy of protein-inhibitor binding. *Protein Engineering*, **10** (12), 1363–1372.
- 94 Lee, F.S., Chu, Z.-T., Bolger, M.B. *et al.* (1992) Calculations of antibody-antigen interactions: Microscopic and semi-microscopic evaluation of the free energies of binding of phosphorylcholine analogs to McPC603. *Protein Engineering*, **5** (3), 215–228.
- 95 Burykin, A., Kato, M. and Warshel, A. (2003) Exploring the origin of the ion selectivity of the KcsA potassium channel. *Proteins: Structure, Function, and Bioinformatics*, **52** (3), 412–426.

8

EVB Simulations of the Catalytic Activity of Monoamine Oxidases: From Chemical Physics to Neurodegeneration

Robert Vianello¹ and Janez Mavri²

¹Computational Organic Chemistry and Biochemistry Group, Ruđer Bošković Institute, Zagreb, Croatia

²Laboratory for Biocomputing and Bioinformatics, National Institute of Chemistry, Ljubljana, Slovenia

8.1 Introduction

Biogenic amines are a large group of naturally occurring biologically active compounds, most of which act as neurotransmitters – endogenous chemicals that allow the transmission of signals from a neuron to target cells across synapses. There are five established amine neurotransmitters: the three catecholamines (dopamine, norepinephrine and adrenaline), and histamine and serotonin. These substances are active in regulating many centrally mediated body functions, including behavioral, cognitive, motor and endocrine processes, and can cause adverse symptoms when they are out of balance.^[1] Despite immense functional importance, we are currently far from understanding the complex cellular and molecular actions of biogenic amines, the timing of their release and the full spectrum of processes that are influenced by them. An important prerequisite to fully realize the actions of biogenic amines is to study, at the molecular level, the mechanisms of both the catalytic activity and inhibition of their metabolic enzymes, and the way these small molecules interact with larger systems such as transporters and receptors. The present chapter focuses on the utilization of modern multi-scale methods of computational biochemistry – from quantum mechanics to combined QM/MM approaches within the Empirical Valence Bond framework – in order to elucidate the catalytic mechanism of a common metabolic enzyme, monoamine oxidase (MAO), responsible for regulating the concentration of biogenic and dietary amines in various parts of the body.

Brain monoaminergic systems have been extensively implicated in the etiology and course of various neurodegenerative disorders, including Alzheimer's disease, Parkinson's disease, and Huntington's disease, causing problems with movement (ataxias) or mental functioning (dementias).^[1,2] Although perceived as the diseases of the elderly, these disorders can have a much earlier onset, starting even before the age of 40, leading to long-term treatment and substantial financial burden for the public healthcare systems.

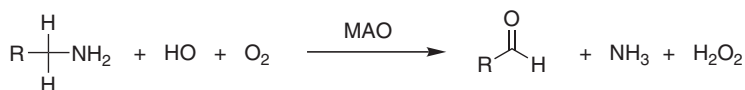
In spite of this, there is, at present, no effective treatment able to substantially slow down or even stop the deterioration of neurons resulting in impaired brain function.

All compounds employed nowadays exhibit a range of adverse effects, and drugs tend to address only symptoms rather than the causes of the dysfunction. Early diagnosis is close to impossible and, because it often occurs too late, treatment to mitigate the effects of the illness remains limited. In view of this background, the development of novel compounds for psychiatric manifestations in neurodegenerative disorders is not only of scientific interest to advance understanding of the brain at a systems level, but also fundamental to improving the management of symptoms, the therapeutic compliance and the quality of life of patients.

Elucidating the precise atomistic details about the catalytic activity of MAO enzymes is of paramount importance for understanding the chemistry of neurodegeneration on molecular level. The knowledge gained in this area would facilitate the research on other members of the large family of flavoenzymes and would, beside general biochemistry and physiology, be of significant value for designing and preparing novel effective MAO inhibitors as transition state analogues, which are potential clinical drugs for the treatment of depression, Parkinson and Alzheimer diseases.^[3]

8.2 Pharmacology of Monoamine Oxidases

Monoamine oxidase (MAO, E.C. 1.4.3.4) is a mitochondrial outer membrane-bound flavoenzyme that catalyzes the oxidative deamination of a broad range of biogenic and dietary amines into their corresponding imines, thus playing a critical role in the degradation of monoamine neurotransmitters in the central and peripheral nervous systems.^[2] Formed imines are then non-enzymatically hydrolyzed to the final carbonyl compounds and ammonia. The enzyme itself is regenerated to its active form by molecular oxygen, O₂, which is in turn reduced to hydrogen peroxide, H₂O₂, according to the overall equation (Scheme 8.1):



Scheme 8.1 Overall Oxidative Deamination of Biogenic and Dietary Amines Catalyzed by the MAO Enzyme. Amines are Enzymatically Converted to the Corresponding Imines, Which Leave the MAO Active Site and are Non-Enzymatically Hydrolyzed to Aldehydes.

Subsequently, the aldehyde intermediate is rapidly metabolized, usually by oxidation via the enzyme aldehyde dehydrogenase to the corresponding acid, or, in some circumstances, to the alcohol or glycol by the enzyme aldehyde reductase. MAOs operate using the flavin adenine dinucleotide (FAD) cofactor, which is, in contrast to the majority of other flavoenzymes, covalently bound to a cysteine through an 8 α -thioether linkage (Figure 8.1).^[4] During the catalytic reaction, FAD is reduced to FADH₂ by accepting two protons and two electrons from the substrate. Although having around 70% sequence identities and a conserved pentapeptidic sequence (Ser-Gly-Gly-Cys-Tyr) that binds the identical FAD cofactor,^[4] both the A and the B isoforms of the enzyme differ on the basis of their substrate affinities and inhibitor sensitivities,^[3] but it is assumed that they act by the same mechanism. The two isoforms, encoded by separate genes located on the X chromosome, are expressed in different amounts in all tissues. In humans, MAO A predominates in the gastrointestinal tract, placenta and heart, whereas MAO B

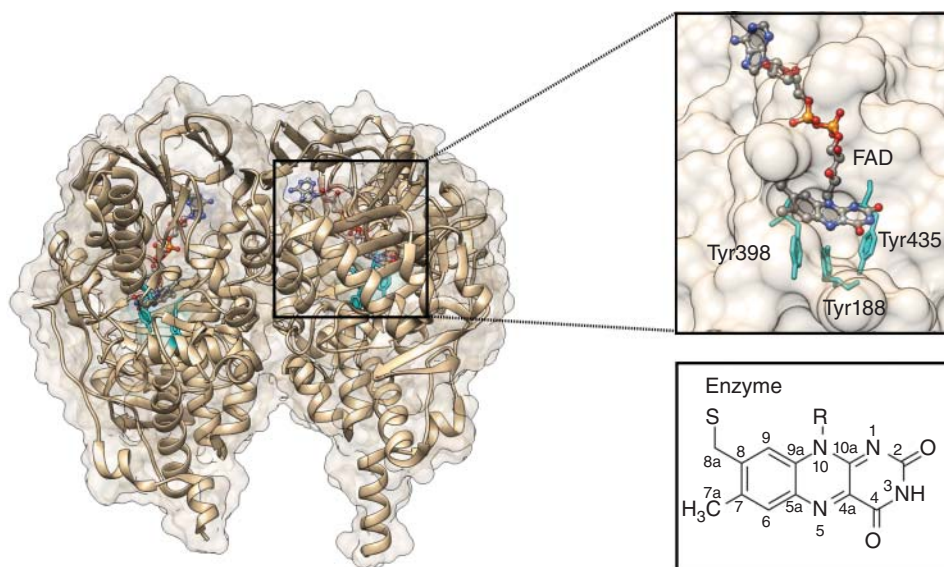


Figure 8.1 Position of the Active Site Within the MAO B Crystal Structure (2XFN.pdb) Indicating the Flavin Cofactor (FAD) and Three Tyrosine Residues that all Form the “Aromatic Cage” Structural Feature. Atom numbering of the FAD fragment is shown in the bottom-right box. (See color plate section for the color representation of this figure.)

predominates in platelets and glial cells in the brain. Each isoform is present in specific subsets of neurons, where they metabolize neurotransmitters, and both are found in liver, where biogenic amines are rapidly metabolized to less bioactive forms for excretion. Under normal physiologic conditions, noradrenalin and serotonin are the preferred substrates of MAO A, whereas β -phenylethylamine is the preferred substrate of MAO B. Both forms of the enzyme metabolize dopamine, albeit with different efficiency. Inhibitors that act mainly on MAO A are used in the treatment of depression, due to their ability to raise serotonin concentrations, whereas MAO B inhibitors decrease dopamine degradation and improve motor control in patients with Parkinson disease.^[3] Inhibition of MAOs has a notable neuroprotective effect, since the MAO catalyzed reactions yield neurotoxic products such as hydrogen peroxide and aldehydes.^[5–8] The cloning of two separate cDNAs encoding two isoforms of MAO^[9] provided the basis for a range of important discoveries, thereby allowing the elucidation of their biological roles and development of inhibitors. However, despite tremendous research efforts devoted to MAOs over several decades, neither the catalytic nor the inhibition mechanisms of MAO have been unambiguously established.

8.3 Structures of MAO A and MAO B Isoforms

After successful heterologous overexpression and purification of recombinant human MAO in yeast,^[10,11] the three-dimensional structures of human MAO A and MAO B have been solved at a resolution of 2.2 Å and 1.65 Å, respectively.^[12–14] The X-ray structures of both human MAOs showed that the enzyme is membrane bound, attached

to the mitochondrial outer membrane, with the active site binding and decomposing amines in the cytosol, and its α -helix transmembrane motif located at the C-terminus. The active-site cavities are reached from the flavin-binding site at the core to the surface of the protein and are mainly hydrophobic. The FAD cofactor binding site, which is also highly conserved between the two enzymes, lies at the end of a long tunnel leading from the outside of the protein close to the membrane surface. The tunnel is generally considered to be hydrophobic, ending in an aromatic cage near the flavin where two tyrosines align the substrate towards the C4–N5 region of the flavin (Figure 8.1). Mutational studies of these residues in MAO B^[15,16] have shown that even though none of these residues is essential to catalysis, the affinity for and turnover of substrates is altered in the mutant enzymes. For example, the K_m for benzylamine increases by more than 10-fold in the Tyr435Phe mutant.^[15,16] The evidence reveals a clear role for these tyrosines in correctly aligning the substrate for catalysis, perpendicular to the flavin N5 atom and on the *re* face of the isoalloxazine ring. These residues also exert a dipole effect on the substrate, that can make the amine more susceptible to oxidation.^[15,16] Therefore, key features for substrate-positioning in the active site are proximity and orientation relative to the N5–C4a region of the flavin ring (Figure 8.1).

8.4 Mechanistic Studies of MAO

The general reaction of flavin amine oxidases, including MAO, can be divided into two half-reactions. In the reductive half-reaction, a hydride equivalent is transferred from the substrate to the flavin, thus reducing it to FADH₂, whereas the oxidative half-reaction involves the oxidation of the reduced flavin back to FAD by molecular oxygen, producing H₂O₂ (Scheme 8.1).

The chemical mechanism of the reductive half-reaction of flavoprotein amine oxidases has been the source of controversy and debate.^[17] Oxidation of an amine substrate by an amine oxidase necessarily involves the removal of two protons and two electrons as the carbon–nitrogen single bond is converted to a double bond. Three-dimensional structures of MAO A and MAO B isoforms,^[12–14] together with extensive kinetic and spectroscopic studies on mutant enzymes,^[14–16] have led researchers to propose three possible catalytic scenarios for MAO^[17]: (a) the direct hydride mechanism, (b) the radical mechanism, and (c) the polar nucleophilic mechanism (Figure 8.2). Studies on deuterated substrate analogues have suggested that the rate-limiting step is the cleavage of a carbon-hydrogen bond vicinal to the amino group^[18] and, hence, the catalytic proposals differ in the nature of the hydrogen being transferred, namely a hydride (H⁻) in (a), a hydrogen atom (H[•]) in (b), and a proton (H⁺) in (c),^[19,20] commonly by the flavin N5 atom. In contrast, the other hydrogen from the amino N–H moiety is generally proposed to be abstracted as a proton.^[21] Thus, establishing the MAO catalytic mechanism necessarily requires the knowledge of the timing of the removal of hydrogens from both the carbon and the nitrogen atoms.

The possibility of the hydride mechanism for MAOs (Figure 8.2a) was based on their structural similarities to flavoprotein D–amino acid oxidases (DAAO), for which both kinetic measurements^[22,23] and related calculations^[23] suggested a hydride or a single electron transfer. Additionally, deuterium and ¹⁵N kinetic isotope effect studies of flavin amine oxidases including DAAO,^[22] tryptophan-2-monooxygenase (TMO),^[24]

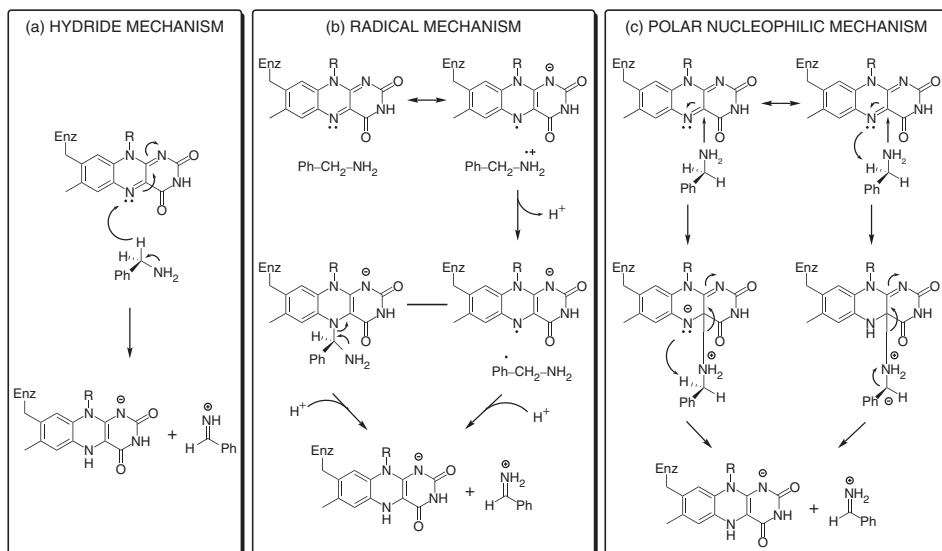


Figure 8.2 Three Mechanisms For MAO Catalytic Activity Proposed in the Literature, Initiated by Either (a) Abstraction of a Hydride Anion, (b) Removal of a Hydrogen Atom or (c) Deprotonation (H^+ abstraction), All From the Substrate α -Carbon Atom. The final products from all three mechanisms are the protonated iminium cation and semi-reduced flavin $FADH^-$.

and *N*-methyltryptophan oxidase (MTOX),^[23] are most consistent with a hydride transfer mechanism. However, Erdem *et al.*^[25] assumed that a hydride mechanism was unlikely, because in MAO it would be associated with a barrier too high to be readily crossed. In addition, a combined ^{15}N and deuterium isotope effects demonstrated that the C–H bond cleavage is not concerted with the rehybridization of the substrate amino group,^[17] which seem to rule out the feasibility of a concerted hydride transfer mechanism.

According to Silverman and coworkers^[26,27] the radical mechanism is initiated by a single-electron transfer from the substrate amine to the flavin, producing an aminium radical cation and a flavin semiquinone as transient intermediates. The lowered pK_a of the α -CH bond of the aminium radical cation can result in either a stepwise^[26] or concerted^[28] deprotonation and a second electron transfer producing reduced flavin and the iminium ion (Figure 8.2b). The main evidence in support of the radical mechanism is the observation that both MAOs are inactivated by cyclopropylamine analogues with subsequent ring opening, a process characteristic of radical reactions.^[29] An argument against the radical mechanism has been the failure of any laboratory to detect semiquinone species during turnover in a stopped-flow monitored reduction or using radical traps (see later). The chemistry of the cyclopropylamine probes and resulting products supported the view that single electron transfer is possible at least with these compounds, but the lack of inactivation by, and a ring-opened product from *trans*-2-phenyl(aminomethyl)cyclopropane suggested that other mechanisms must be possible.^[30] In a modified single electron transfer mechanism, the involvement of protein-based radicals has also been proposed but questioned.^[31,32] Mutation of the substrate-orienting tyrosines and isotopic labeling of the tyrosines in MAO A, and the

fungal counterpart MAO B, provide evidence for radicals delocalized away from the active site.^[33]

The experiments by Edmondson and coworkers,^[19] as well as related electron paramagnetic resonance studies^[34] and stopped-flow kinetic determinations,^[35] failed to provide any evidence for radical intermediates, and no influence of the magnetic field on the kinetics of enzyme reduction was observed.^[36] On the other hand, Taft correlation studies of Miller and Edmondson with benzylamines showed that attaching the electron-withdrawing groups to the substrate *para*-position increases the rate of the reaction in both human^[37] and rat^[20] MAO A, implying negative charge build-up on the substrate α -carbon atom, thus suggesting that proton transfer is an integral part of the rate limiting step. This led authors to propose the polar nucleophilic mechanism for MAO A^[37] (Figure 8.2c), originally formulated by Hamilton.^[38] This mechanism involves the creation of a highly energetic substrate-flavin adduct which then decomposes to the protonated imine, with proton abstraction concerted with either the adduct formation or the product formation. The crucial issue related to this mechanism is what moiety on the enzyme would be strong enough base to perform this task, since the pK_a of a benzyl proton is expected to be around 25.^[39] Structural analysis of both MAO isoforms showed there are no active site basic residues that could act as proton acceptors. Also, no direct evidence for a stable amine-flavin adduct has been found experimentally. A study on human MAO B, however, showed an inverse Taft correlation,^[40] supporting the hydride transfer mechanism, which led the authors to propose different mechanisms for two isozymes: H^+ transfer in MAO A and H^- transfer in MAO B, although, the same authors had previously ruled out H^- transfer in human MAO B, based on the nitrogen secondary kinetic isotope effects.^[17] Based on the similarities in the continuous wave electronic paramagnetic resonance spectrum between MAO and *D*-amino acid oxidase, determined to operate by the H^- transfer mechanism,^[41] Kay *et al.* suggested a hydride transfer mechanism should be re-examined for MAO.^[31,32]

Therefore, it is clear that, despite the widespread use of MAO inhibitors, the mechanism of MAO catalysis is not unambiguously determined which calls for the rationalization using advanced computational techniques.

8.5 Cluster Model of MAO Catalysis

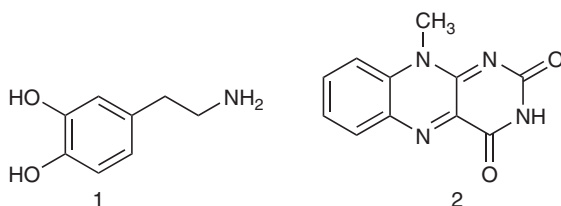
The past century has seen substantial advances, both in computational techniques and in our understanding of how enzymes really work. The use of quantum chemical methods to address enzymatic reaction mechanisms has become a booming area in enzymology.^[42–45] Nevertheless, studying an enzyme at atomic resolution is a computationally demanding task. In this case, the complexity of the system prevents resorting to pure quantum-mechanical (QM) methods to treat the entire system, whereas the phenomena under study cannot be accurately represented by molecular-mechanical (MM) methods (since MM methods are in their traditional functional form unable to describe chemical reactions). Currently, there are two popular approaches to describe enzymatic processes: the quantum mechanics-only (QM-only),^[46–48] which uses a small but carefully selected cluster model of the active site, and the multiscale quantum mechanics/molecular mechanics (QM/MM) method,^[49,50] which uses a layer-based approach using the entire protein. Both methods have been successfully applied to the

study of various classes of enzymes, and in many cases similar results and conclusions have been obtained.^[42,43]

In the QM-only approach, commonly also called the cluster approach,^[46–48] a model of the active site is designed on the basis of available crystal structures. Density functional theory methods are most frequently used for the calculation of the geometries and energies of all stationary points along the reaction pathways. The missing steric and electrostatic effects from the remaining part of the protein are considered by two simple procedures. The steric effects imposed by the protein matrix are taken into account by applying position constraints to certain key atoms at the periphery of the cluster model, while electrostatic effects are modeled by the dielectric cavity method, usually with a dielectric constant of 4. Recent work even demonstrated that the solvation effects saturate with increasing the size of a cluster and that the particular choice of dielectric constant is then no longer of much concern.^[51] On the basis of extensive calculations for a large number of enzymes, the error of the cluster approach for modeling metalloenzymes has been assessed by Siegbahn and coworkers to be less than 5 kcal/mol.^[48,52]

The first computational study on the complete mechanism of MAO catalysis was performed by our groups,^[53] employing DFT methodology within the cluster model of the enzyme. The starting point for our calculations was the high-resolution (1.6 Å) X-ray structure of MAO B complexed with 2-(2-benzofuranyl)-2-imidazoline (PDB code 2XFN).^[12,13] We truncated the enzyme to the cofactor flavin moiety (isoalloxazine group 2, Figure 8.3) and three tyrosine side-chains (*p*-hydroxytoluenes of Tyr188, Tyr398 and Tyr435), which all form the mentioned hydrophobic “aromatic cage,”^[15,16] an important structural feature of MAO enzymes (Figure 8.1). Previously, we calculated the pK_a values of the Tyr residues with bound dopamine, employing the full dimensionality of the protein,^[54] and obtained an upward shift to 13.0–14.7 (10.1 in aqueous solution). This clearly confirmed the hydrophobic nature of the active site, which indicates that gas-phase calculations on truncated MAOs are reliable. Also, we showed that the pK_a value of bound dopamine changes to only 8.8 (8.9 in aqueous solution), a result of stabilizing cation– π interactions with the tyrosine side-chains.^[54] This implies that dopamine binds to the MAO active site as a protonated monocation, but the free-energy cost to deprotonate it to the bulk, being as low as 1.9 kcal mol⁻¹, allows it to enter the chemical step either as an ionized or neutral molecule, which led us to consider both alternatives. Based on both crystal structures^[12,13] and our MD simulations^[54], which indicate the presence of few water molecules in the active site, our model also included four crystal waters (HOH₂₁₅₇, HOH₂₁₈₁, HOH₂₃₂₉ and HOH₂₃₇₂). It turned out that two of these waters are chemically involved in catalysis. We manually placed dopamine 1 (Figure 8.3) within the so formed cluster, resulting in the initial stationary-point (SP) complexes (Figure 8.4). The system was modeled at the (CPCM)/M06–2X/6–311++G(2df,2pd)//(CPCM)/M06–2X/6–31+G(d) level of

Figure 8.3 Structures of Dopamine (1) and Isoalloxazine Moiety (2) of the FAD Cofactor Used in the Cluster Model of the MAO Active Site.



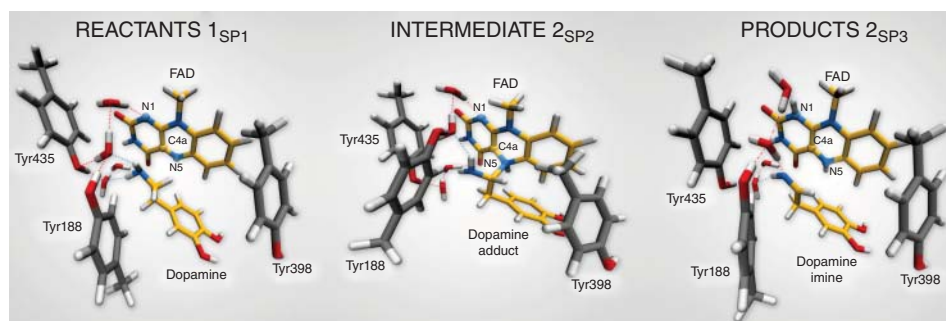


Figure 8.4 Structures of Relevant Stationary Points For the Newly Proposed Two-Step MAO Catalytic Hydride Mechanism For the Degradation of Dopamine 1. (See color plate section for the color representation of this figure.)

theory employing the M06–2X functional of Zhao and Truhlar, which has been shown to be particularly successful in treating non-bonding interactions and accurately reproducing thermodynamic and kinetic parameters.^[55–58] To account for the polarization effects caused by the rest of the enzyme, we included a conductor-like polarizable continuum model (CPCM)^[59] with a dielectric constant of $\epsilon = 4$, taking the rest of the parameters for pure water, as employed by Siegbahn, Himo and their coworkers to elucidate the catalytic mechanism of a large variety of enzymes.^[46–48]

For 1_{SP1} (Figure 8.4) we first considered the single-electron radical mechanism (Figure 8.2b), which would initially create a biradical system either in the singlet or triplet electronic states. (CPCM)/UM06–2X/6–31+G(d) calculations for the triplet state led to a biradical system $54.2 \text{ kcal mol}^{-1}$ higher in energy than 1_{SP1} . For the singlet state, we employed “symmetry-broken” open-shell calculations, defining dopamine and the rest of the system as two different fragments each having an unpaired electron of opposite spin, which resulted in a stable wavefunction with both the energy and electron distribution identical to that in 1_{SP1} . Overall, this suggests that the singlet-state biradical is non-existent, whereas the free-energy cost to generate the triplet is too high for an efficient catalysis. In addition, our calculations yielded the ionization energy for isolated dopamine **1** of $148.2 \text{ kcal mol}^{-1}$ and an electron affinity for the flavin **2** of $78.5 \text{ kcal mol}^{-1}$. The latter, however, increases to $87.3 \text{ kcal mol}^{-1}$ when the structure of the whole MAO cluster is considered, still leaving a gap of around 60 kcal mol^{-1} between these two quantities, thus revealing a thermodynamic imbalance for a feasible electron-transfer process. This agrees well with the experimentally observed mismatch between the oxidation/reduction potentials of the FAD cofactor, which is too low (-0.2 V)^[60] for it to be an effective oxidant of the neutral amine (around $+1.0$ – 1.5 V).^[45] Furthermore, on top of the fact that all X-ray analyses have confirmed that in MAOs there are no classical radical initiators to start the reaction and that there is no experimental evidence for a radical intermediate, our results suggest that it is very unlikely that a radical pathway is feasible and we did not consider it further.

The polar nucleophilic mechanism is another alternative for the amine oxidation and involves proton abstraction from the α -carbon atom as the rate-limiting step (Figure 8.2c).^[37] The crucial issue relating to this mechanism is what moiety on the enzyme would be a strong enough base to perform this task, because the pK_a of a benzyl proton is expected to be around 25.^[39] Structural analysis of both MAO isoforms

shows there are no basic active-site residues that could act as proton acceptors.^[13,62] Edmondson and coworkers upheld their arguments by stating that in MAOs the flavin is bent by around 30° from planarity about the N5–N10 axis,^[62] which enhances the basicity of the N5 atom and depletes the electron density on the C4a atom, thus facilitating substrate–flavin complex formation with the former flavin site making subsequent proton abstraction possible (Figure 8.2c). We feel that, even if all of these effects are operational, it would still be insufficient to downshift the substrate pK_a value by around 10–15 units in order to make the α -CH bond acidic enough for an efficient catalysis. Moreover, inspection of $\mathbf{1}_{SP1}$ shows that the flavin ring is not bent, but almost perfectly planar with the C10a–N10–N5–C5a dihedral angle of 174.3° and the N(dopamine)···C4a(flavin) bond length as long as 2.827 Å. Furthermore, a relaxed-geometry scan of the latter bond, by compressing it with 0.1 Å increments, showed no indication of the formation of a stable complex. In addition, NBO charges on the flavin C4a and N5 sites and the substrate N atom in $\mathbf{1}_{SP1}$ are 0.13, –0.35 and –0.97|e|, respectively, which demonstrate that they do not change much from the values found in isolated flavin **2** and dopamine **1** (0.10, –0.34 and –0.93|e|). This all suggests that neutral amines do not exhibit the necessary nucleophilicity to readily add to the flavin C4a position, in agreement with the fact that no direct evidence for a stable amine–flavin adduct has been found experimentally. Taken all together, these results also led us to rule out this mechanism as feasible.

The geometry of stationary structure $\mathbf{1}_{SP1}$ (Figure 8.4) suggests that the following two pathways are possible. The substrate amino group is connected through two active-site water molecules to the flavin N1 atom, which is the most basic position within the isoalloxazine moiety.^[53] This implies that the substrate could be first activated by amino deprotonation to the flavin N1 site. Also, the α -C(substrate)···N5 bond length in $\mathbf{1}_{SP1}$ is 3.198 Å, being sufficiently short to suggest that the substrate is properly oriented for a direct α -CH abstraction.

Deprotonation of the neutral substrate amino group is feasible, but the process is associated with a large barrier of 37.3 kcal mol⁻¹ (Figure 8.5a).^[53] The transition-state structure $\mathbf{1}_{TS1}$ has one imaginary frequency of 270i cm⁻¹ representing proton transfer to the flavin N1 atom assisted by two water molecules via the de Grotthuss mechanism.^[63] Such a high energy requirement is rationalized by the large difference in the pK_a values between proton donor and acceptor sites. For the flavin N1–H deprotonation pK_a was measured to be around 7.0,^[64] whereas amine deprotonation typically has a pK_a of around 35.^[39] Upon proton removal, the anionic substrate turns nucleophilic and covalently binds to the flavin C4a atom, accompanied by a reduction in the corresponding N(substrate)···C4a distance from 4.063 ($\mathbf{1}_{SP1}$) and 1.550 Å ($\mathbf{1}_{TS1}$) to 1.508 Å in $\mathbf{1}_{SP2}$. The formed complex facilitates subsequent H⁻ abstraction from the α -CH group, requiring only 13.2 kcal mol⁻¹ to arrive at the transition-state structure $\mathbf{1}_{TS2}$, characterized by one imaginary frequency of 871i cm⁻¹. Abstraction of the H⁻ is concerted with the loosening of the N(substrate)···C4a bond from 2.439 Å in $\mathbf{2}_{TS2-N}$ to 3.505 Å in the final products $\mathbf{1}_{SP3}$, being reduced flavin FADH₂ and neutral imine. This pathway is associated with collective activation energy of 44.6 kcal mol⁻¹, which is too high to be feasible for the MAO enzyme activity. In an analogous pathway, only starting from the substrate with the protonated amino group, the flavin N1 deprotonation of the –NH₃⁺ group becomes easier requiring only 20.7 kcal mol⁻¹ ($\nu_{imag} = 980i$ cm⁻¹). The subsequent α -hydride abstraction by the flavin N5 atom costs an additional 24.4 kcal mol⁻¹

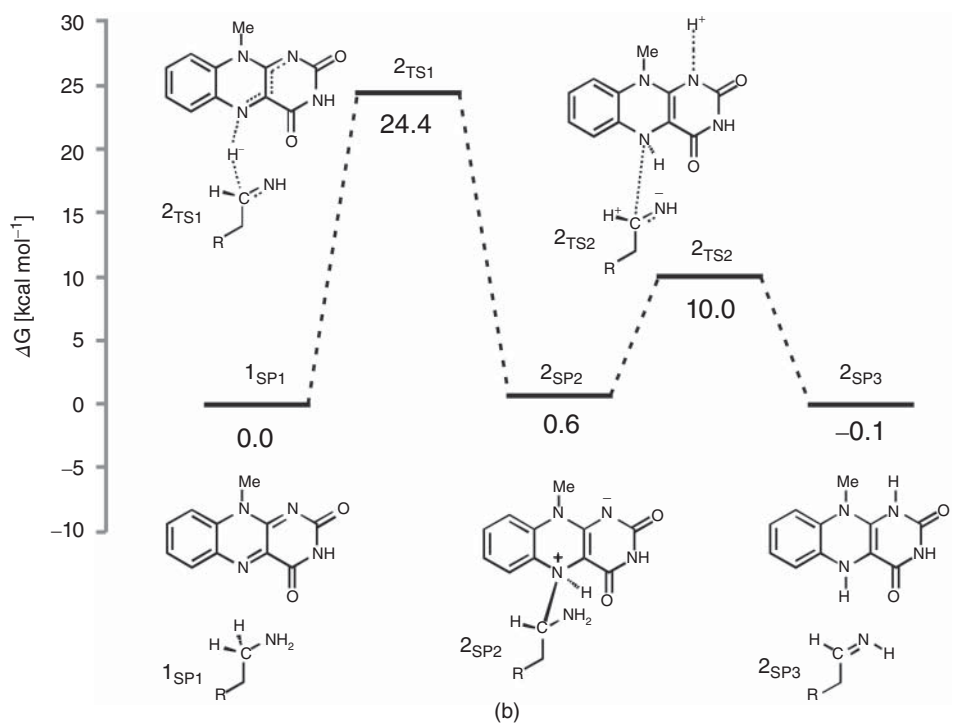
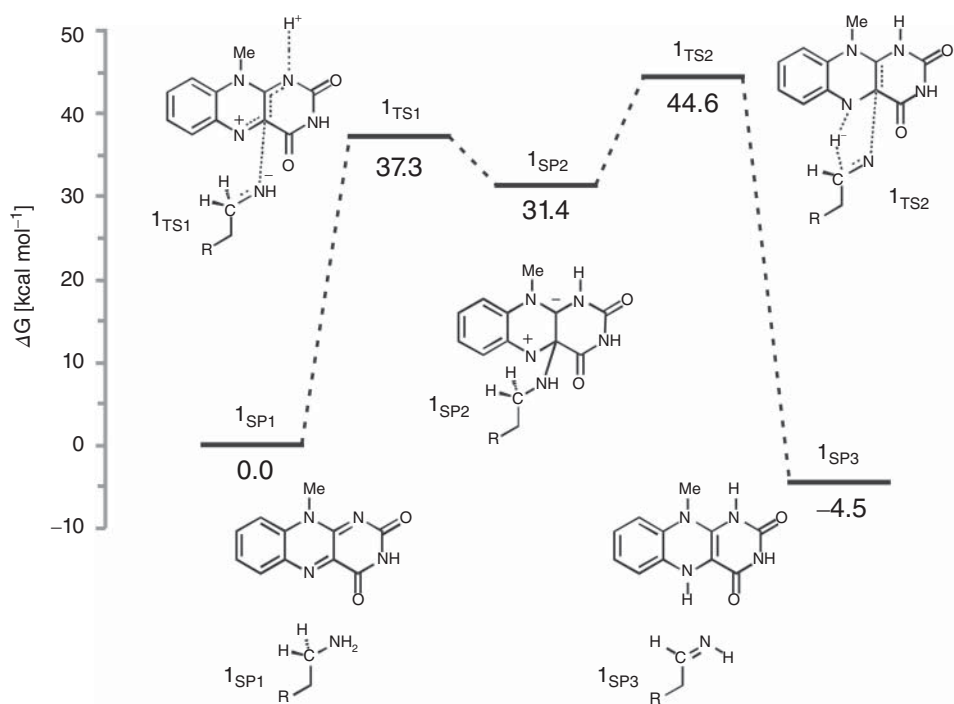
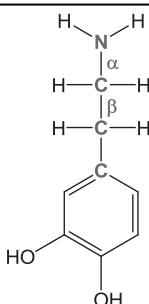


Figure 8.5 Free-Energy Profiles For the Reaction Initiated by Either Amino Deprotonation (a), and a Direct Hydride Abstraction (b).

to add up to an overall reaction free-energy barrier of 43.1 kcal mol⁻¹, yielding protonated imine, R-CH₂CH₂=NH₂⁺. Although the collective activation barrier for this process is slightly lower than with the unionized substrate, it is still too high for this mechanism to be plausible.

According to Figure 8.2a, the traditional notion of the hydride abstraction should (a) generate positive charge on the substrate α -carbon atom and (b) be concerted with the substrate amino nitrogen atom rehybridization from sp³ to sp². The first process would seem to contrast the aforementioned positive Taft correlation^[37] that suggested negative charge development on the stated carbon atom, whereas the second process was suggested as unlikely by Matsson and coworkers after measuring the ¹⁵N KIE for the MAO-B oxidation of benzylamine.^[17] However, due to H⁻ abstraction, the depletion of electron density on the α -carbon atom is completely outperformed by the strong electron-donating ability of the vicinal amino group, as evidenced by a decrease in both the corresponding N(amino)-C(ff) bond length and the nitrogen atomic charge from 1.466 Å and 0.97|e| in **1**_{SP1} to 1.354 Å and -0.79|e| in **2**_{TS1}. As a result, the charge on the α -carbon atom changes from -0.25|e| in **1**_{SP1} to -0.11|e| in **2**_{TS1}, surprisingly preserving a significant portion of the negative charge (Table 8.1). More importantly, the charge on the dopamine β -carbon atom, which is, in terms of *para*-substituent effects, analogous to the benzylamine α -carbon atom, even increases from -0.49|e| (**1**_{SP1}) to -0.50|e| (**2**_{TS1}). Taken all together, this gives us strong evidence to rationalize why this reaction is facilitated by the electron-withdrawing *para*-substituents on the aromatic ring, in firm agreement with the work of Miller and Edmondson.^[37] With regards to the substrate nitrogen sp³-sp² rehybridization, our results demonstrate that it is not concerted, but takes place after the α -hydride is removed. A useful and practical measure of the degree of nitrogen sp³-sp² hybridization can be obtained from the nitrogen degree of pyramidalization (DPN) defined elsewhere.^[65] Our calculations reveal that DPN for the dopamine nitrogen atom is 37.4% in isolated **1** and 36.5% in **1**_{SP1}, values typical of primary amines. These change to 4.2 and 13.9% in **2**_{TS1} and **2**_{SP2}, respectively. It follows that in the transition-state structure **2**_{TS1}, the substrate still pertains 4.2% of its nitrogen pyramidalization, which is then, surprisingly, even increased to 13.9% following H⁻ removal. The latter occurs, because, after the flavin accommodates H⁻, its N5 atom becomes sp³-hybridized with excess negative charge, having enough nucleophilicity to

Table 8.1 Evolution of Atomic Charges During the C(α)-H Hydride Abstraction Reaction From Dopamine to Flavin as Obtained with the NBO Approach at the (CPCM)/M06-2X/6-31G(d) Level of Theory

System	Atom/Group	Isolated	Reactants (1 _{SP1})	TS (2 _{TS1})	Intermediate (2 _{SP2})
	N(amino)	-0.93	-0.97	-0.79	-0.90
	α -C	-0.26	-0.25	-0.11	0.16
	β -C	-0.50	-0.49	-0.50	-0.52
	C1(phenyl)	-0.06	-0.04	-0.06	-0.07
	dopamine	0.00	-0.03	0.31	0.43
	N5(flavin)	-0.34	-0.35	-0.50	-0.50
	N1(flavin)	-0.63	-0.68	-0.71	-0.70
	flavin	0.00	0.01	-0.29	-0.35

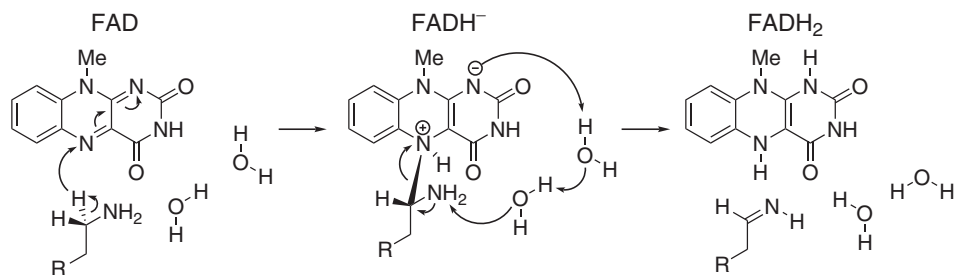


Figure 8.6 Complete Two-Step Mechanism For MAO Catalyzed Amine Degradation. The first step involves H^- abstraction from the substrate to form the flavin–substrate adduct, which then decomposes to the final products, namely neutral imine and fully reduced flavin, FADH_2 , a reaction promoted by amine deprotonation facilitated by two water molecules.

form a covalent bond with the thus formed cationic substrate (Figure 8.4 and Figure 8.6). Macheroux and coworkers^[64] measured the $\text{p}K_{\text{a}}$ value of the fully hydrogenated flavin $\text{N}(5)\text{-H}_2$ moiety to be $\text{p}K_{\text{a}} = 4$, which indicates that the $\text{N}(5)\text{-H}^-$ group possesses sufficient basicity to act as a base. The calculated substrate–flavin interaction free energy in $2_{\text{SP}2}$ is as high as $27.7 \text{ kcal mol}^{-1}$, which, together with the corresponding $\text{N}5\text{-C}(\text{ff})$ bond length of only 1.703 \AA , suggests that the complex formed is rather strong. In $2_{\text{SP}2}$, the dopamine nitrogen atom maintains its amine -NH_2 features, which restores its pyramidalization back to 13.9% leaving its planarization only for the upcoming step, thus strongly tying in with the experimental results.^[17] This all gives a convincing indication that α -hydride abstraction is not the only and last step during MAO catalysis. For product release and subsequent enzyme regeneration, another chemical reaction must take place. This idea represents an important advantage over other mechanistic proposals that all advise protonated imine as the final product of one-step catalysis (Figure 8.2). We disagree with this latter suggestion for the following two reasons. First, it would be difficult for a protonated imine to leave the active site, because on its way out it would strongly bind to the “aromatic cage” through favorable cation– π interactions.^[54] Secondly, it is well established that the final imine hydrolysis to aldehydes occurs non-enzymatically outside the MAO.^[66,67] However, the protonated product would immediately be hydrolyzed by the nearest water molecule within the enzyme, because in solution this reaction readily proceeds with the protonated imine under acidic conditions.^[39]

The next step involves amino group deprotonation by the flavin N1 atom with an activation free energy of $9.4 \text{ kcal mol}^{-1}$ (Figure 8.5b), being concerted with the weakening of the adduct $\text{N}5(\text{flavin})\cdots\text{C}\alpha(\text{dopamine})$ bond. The transition-state structure $2_{\text{TS}2}$ ($\nu_{\text{imag}} = 933i \text{ cm}^{-1}$) again describes a de Grotthuss-type proton transfer proceeding with the two active-site water molecules, but has both dopamine amino protons still close to the nitrogen atom with N-H bond lengths of 1.032 and 1.063 \AA . This yields a nitrogen DPN value of 0.1% and represents the stage at which full nitrogen $\text{sp}^3\text{-sp}^2$ rehybridization occurs. Upon deprotonation, the system is stabilized by $10.1 \text{ kcal mol}^{-1}$ to $2_{\text{SP}3}$, making the whole reaction energetically feasible and yielding the neutral trans-imine and the fully reduced flavin (FADH_2) as the final products (Figure 8.4 and Figure 8.6). It has to be strongly emphasized that the presence of the acidic N-H bond enables the completion of MAO turnover and explains why many alkyl- and arylamines

change from being MAO substrates to MAO inhibitors upon *N,N*-dimethylation.^[68] The fact that dopamine is converted into a neutral imine is significant, because this suggests it will predominantly remain unprotonated in the hydrophobic active site, based on consideration of the pK_a values of similar unconjugated imines, which are, as a rule, found to be below the physiological value of 7.4 (e.g., the pK_a value of $\text{Me}_2\text{C}=\text{N}-\text{Me}$ is 5.5),^[39] ensuring that the neutral product could go past the "aromatic cage" on its release from the active site. This, however, does not rule out the possibility that, when it is finally liberated from the enzyme, the imine formed could be protonated, because, during its departure, changes in the environment could make the protonation feasible. This would then fully agree with Edmondson *et al.*,^[66] who showed that the protonated *p*-(dimethylamino)benzylimine is the form that is released from the enzyme. Also, the fact that flavin is fully reduced to FADH_2 enables an important prerequisite for MAO regeneration by molecular oxygen to revert flavin into its oxidized form (FAD) by creating hydrogen peroxide (H_2O_2), a reaction for which two hydrogen atoms are required (Scheme 8.1).

In concluding this section, let us emphasize that presented results have convincingly demonstrated the feasibility of the hydride transfer mechanism (Figure 8.6). In recent years, there have been several additional computational studies showing its prevailing energetic feasibility in MAO,^[69–71] or some other flavoenzymes.^[72,73] Furthermore, in what follows, we will demonstrate that when we moved from the QM-only cluster model towards including the full enzyme structure, via the Empirical Valence Bond QM/MM approach, the calculated activation free energy for the MAO B catalyzed degradation of dopamine drops down to $\Delta G^\ddagger = 16.1$ kcal/mol,^[74] being in excellent agreement with the available experimental value of 16.5 kcal/mol,^[19,20] thus providing a strong support for our mechanistic picture.

8.6 Protonation States of MAO Active Site Residues

Computational studies have shown that the bulk of the catalytic power of any enzyme originates from the electrostatic preorganization of the active site.^[75] The essence of this fact is that, in water, solvent molecules must reorient during the course of a reaction due to polarization induced by a changing charge distribution. On the other hand, this energetic penalty is much smaller in an enzyme, as the enzyme provides an electrostatic environment that has evolved to optimally solvate the transition state and requires much less reorganization energy to adapt.^[75] A change in the protonation states of ionizable residues results in an altered electrostatic potential pattern in the enzyme, which manifests on the rate of enzymatic reactions and its pH-dependence. The electrostatic potential of the enzyme active site is a fluctuating scalar field and cannot be directly measured experimentally. However, there are few experimentally accessible quantities that are closely related to it. The pK_a values of ionizable residues in the enzyme active site are, for example, an excellent tool for probing the electrostatic environment, since pK_a values are very sensitive to any structural and electronic change. The main contribution to the pK_a values is offered by the free energy of the solvation of the charged species in the (de)protonation reaction. Due to the similarities between MAO A and B enzymes (70% homologous and the overall RMSD between the X-ray structures is only 0.554 Å for human isoforms),^[14] we expect that their electrostatic potential pattern in

the active site will be similar and therefore the corresponding residues will have a similar pK_a . It was suggested in the literature that the MAO active site is hydrophobic since it is composed of the aromatic and apolar residues, including two tyrosines and the FAD cofactor, forming together an aromatic cage.^[15,16] Change of the pK_a value is also a measure of hydrophobicity: the hydrophobic environment does not favor the presence of charged species giving rise to significant shifts in the pK_a values. Despite the fact that monoamines are predominantly present as monocations at a physiological pH of 7.4, a hydrophobic active site has been proposed to favor unprotonated substrates and, furthermore, every catalytic proposal to date,^[17] including ours,^[53] has agreed that the substrate must be neutral for the reaction to take place. Active site pK_a values are difficult to determine experimentally and, similarly, whereas experimental pH rate profiles can provide tremendous insight, it can be hard to conclusively determine the identity of residues whose protonation state is being affected. For MAO, some attempts were performed to determine pK_a values of some ionizable groups experimentally. However, they were mainly limited to the surface residues.^[76–78] The work of Scrutton and coworkers^[79] reported the pH-dependence of MAO A kinetic parameters by stopped-flow and steady state techniques, as well as the H/D isotope substitution, and the authors concluded that the substrates, including benzylamine, kynuramine and phenyl-ethyl amine bind in the protonated form, while deprotonation is required for chemical step. Moreover, the authors demonstrated that the pH-dependence of the kinetic isotope effect decreases from approximately 13 to 8 with increasing pH, leading to the assignment of this catalytically important deprotonation to that of the bound amine substrate. The strong H/D kinetic isotope effect dependence gives evidence that at low pH the substrate deprotonation interferes with the rate-limiting step. Moreover, the same authors suggested that the pK_a values of the studied substrates are decreased for about 2 pK_a units when transferring the substrate from aqueous solution to the enzyme.

In order to get more insight into the nature of the MAO active sites, we calculated the pK_a values of key ionizable residues in the active sites of both isoforms.^[54,80] Converged solvation free energy calculations were performed using the MOLARIS program package together with an all-atom representation of the solvated enzymes. In line with the general consensus that dopamine enters the chemical step in a neutral form, we manually docked a neutral dopamine molecule into the active site in a way suitable for the catalytic step. We selected only one substrate on purpose in order to avoid the effects associated with ligands of different sizes on the pK_a values. pK_a calculations were performed using the semimacroscopic protein dipole/Langevin dipole approach of Warshel and coworkers in its linear response approximation version (PDLD/S–LRA).^[81,82] The essence of the PDLD/S–LRA pK_a calculation is to convert the problem of evaluating a pK_a in a protein to an evaluation of the change in solvation energy associated with moving the charge from water to the protein (Figure 8.7). To do so, one must consider a thermodynamic cycle described by the following equation:

$$\Delta G^p(AH_p \rightarrow A_p^- + H_w^+) = \Delta G^w(AH_w \rightarrow A_w^- + H_w^+) + \Delta G_{sol}^{w \rightarrow p}(A^-) - \Delta G_{sol}^{w \rightarrow p}(AH) \quad (8.1)$$

where p and w denote protein and water, respectively. This equation can be rewritten for each ionizable residue i , as:

$$pK_{a,i}^p = pK_{a,i}^w - \frac{\bar{q}_i}{2.3RT} \Delta \Delta G_{sol}^{w \rightarrow p}(AH_i \rightarrow A_i^-) \quad (8.2)$$

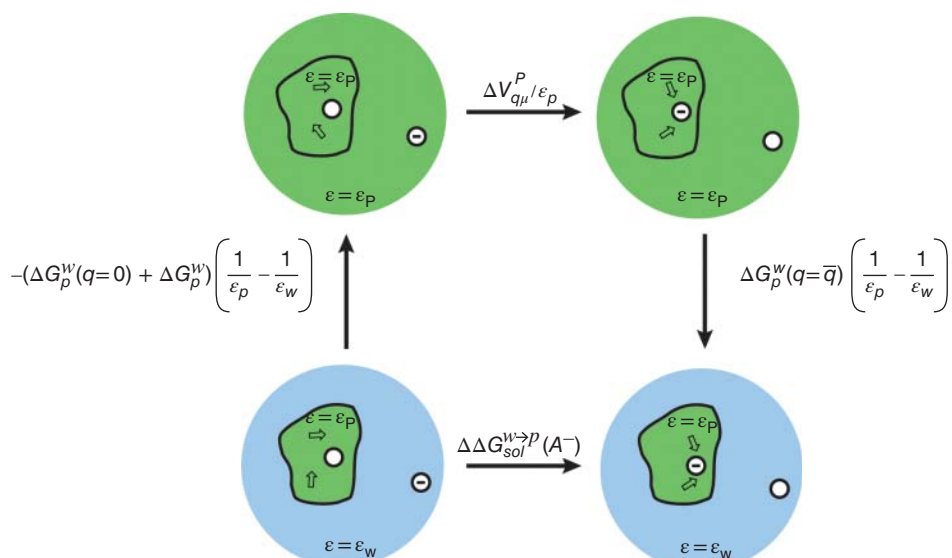


Figure 8.7 The PDLDS–LRA Thermodynamic Cycle For the Evaluation of $\Delta\Delta G^{W\rightarrow P}_{sol}(A^{-})$ Within a Protein (w and p Indicate Water and Protein, Respectively). The cycle involves the change of the dielectric constant of the solvent around the protein from ϵ_w to ϵ_p , moving the charge from the solvent to the protein, changing back to the solvent dielectric and uncharging the ionized group inside the protein. The energy contribution of each step is indicated in the figure (see text). For a detailed discussion see references ^[81,82]. (See color plate section for the color representation of this figure.)

where the $\Delta\Delta G$ term consist of the last two terms of the previous equation, q_i is the charge of the ionized form of the given residue, for an acid

$$\bar{q}_i = -1[q(AH) = 0, q(A^{-}) = -1] \quad (8.3)$$

and for a base

$$\bar{q}_i = +1[q(AH) = +1, q(A^{-}) = 0] \quad (8.4)$$

The pK_a calculations are thus reduced to two free energy calculations in addition to the experimental value in aqueous solution. The first simulation calculates the free energy associated with the transfer of a neutral residue from protein to water and the other calculates the free energy cost of transferring the charged residue from water to protein. In addition to these two values one must know only the experimental pK_a in aqueous solution to determine the free energy associated with ionizing the residue in water. In practice one calculates reversible work for mutation of a charged residue or a substrate to the neutral species in the enzyme and in aqueous solution. Together with the experimental pK_a value in aqueous solution it is possible to calculate the pK_a value in enzyme via the thermodynamic cycle. The pK_a calculations are, thus, reduced to two free energy calculations in addition to the experimental pK_a value in aqueous solution (Figure 8.7).

The calculated pK_a values are sensitive to the applied external dielectric constant during the simulations. The choice of the correct dielectric constant to describe the protein interior is a very complex issue and has been the subject of heated debates over the years.^[83] In our work, we employed $\epsilon = 10 - 16$, however, due to the focus on the relative difference between pK_a values in the corresponding MAO A and MAO B active sites,

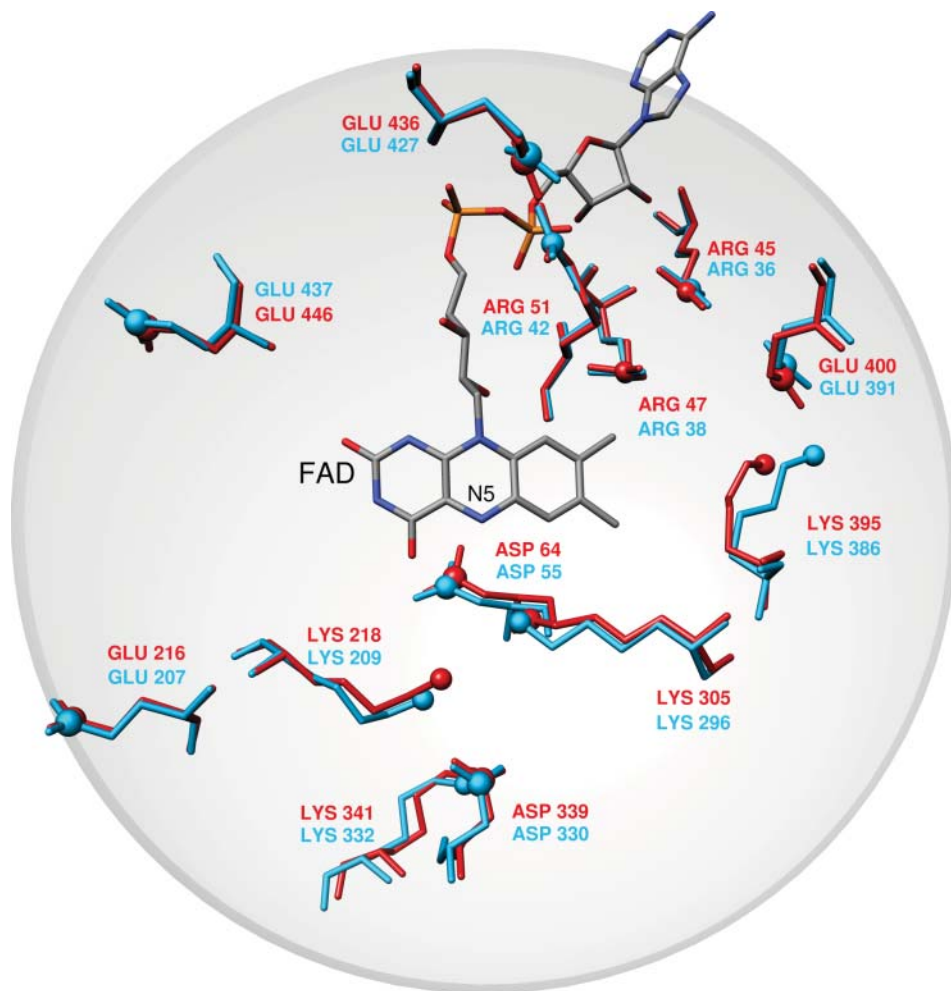


Figure 8.8 Superposition of FAD and Ionizable Residues Surrounding the Active Site (MAO A is Red, MAO B is Blue). (See color plate section for the color representation of this figure.)

the choice would not change the qualitative picture and is, thus, of lesser importance. All PDL/S-LRA calculations were performed using the ENZYMIK force field and the MOLARIS simulation package.^[84] Specifically, we have examined the pK_a values of four tyrosine residues that are part of the so-called aromatic cage and a Lys residue close to the reacting atoms of FAD (Figure 8.8). The average absolute differences in pK_a values, between the two isoforms (MAO A/B), are 0.05, 0.07, 0.12, 0.75, and 1.23 for the pairs Tyr69/Tyr60, Lys305/Lys296, Tyr407/Tyr398, Tyr444/Tyr435, and Tyr197/Tyr188, respectively. These results clearly demonstrate that, for both isozymes, the pK_a values for the studied Tyr and Lys residues are very similar (they differ 1.23 pK_a units at maximum), which together with the already mentioned overall agreement in the structure of the isoforms (Figure 8.8)^[14] provides strong evidence that the electrostatic potential pattern in the active sites of both isozymes is very closely matched. Since enzymes work by

preorganized electrostatics, the same electrostatic environment cannot be at the same time suitable for optimal solvation of the transition state with a positive and a negative charge build-up, as would be the case in the hydride and polar nucleophilic mechanisms. It is therefore very unlikely that MAO A and MAO B would work by different chemical mechanisms on the same family of substrates.^[40] Our results on the pK_a value of the bound dopamine ($pK_a = 8.8$) suggest the latter is practically unchanged compared to the corresponding value in aqueous solution ($pK_a = 8.9$), strongly indicating that the MAO active sites are not hydrophobic when the substrate is bound.^[54] This might appear to contradict the results by Scrutton and coworkers which reported the pK_a shift of the substrates for about 2 pK_a units within MAO active site,^[79] but the discrepancy can be, among other factors, explained by the differences in the sizes of diverse substrates. Practically unchanged dopamine pK_a relative to aqueous solution is attributed to favorable cation- π interactions between the $-\text{NH}_3^+$ group on dopamine and aromatic moieties, which provide a stabilizing effect to the charged fragment.^[54]

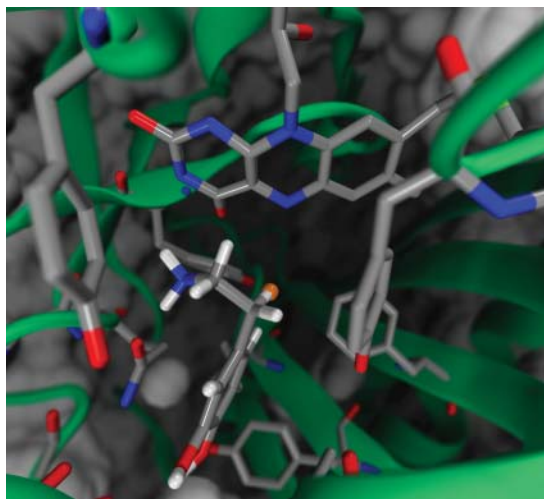
8.7 EVB Simulation of the Rate Limiting Hydride-Abstraction Step for Various Substrates

In this section, we extend our static cluster model study of the rate-limiting step in MAO B to a full enzyme model,^[74] using the EVB approach, which allows extensive sampling at a reduced computational cost, while providing a physically meaningful description of the catalytic process. EVB relies on a well-defined reference state. The most natural choice for it is the uncatalyzed reaction in aqueous solution, since biology occurs in this medium. However, considering the lack of direct experimental observables in aqueous solution, our calculated reference state for the MAO B catalyzed reaction is the corresponding hydride transfer in the gas phase, as this is the state for which one is most likely to obtain reliable estimates of the activation barrier using accurate QM approaches. For the gas phase reference reaction, between lumiflavin and dopamine, we used the optimized geometries of the reactants and transition state from the cluster study (Figure 8.4). The gas phase energetics of the reaction were calculated using the M06-2X density functional^[55] in conjunction with the 6-31+G(d,p) basis set. For a full-enzyme simulations, atomic coordinates were obtained from the high resolution (1.6 Å) crystal structure of MAO B obtained from the Protein Data Bank (PDB code 2XFN) and the dopamine substrate was manually docked into the active site. The charges of all relevant structures were fitted to the electrostatic potential according to the CHelpG scheme and calculated at the (CPCM)/B3LYP/6-311G(d,p) level. EVB calculations were performed in vacuo, representing the reference state, aqueous solution, and in the MAO B enzyme, to examine the effect of different environments on the reaction barrier. For all EVB calculations the same EVB region was used, consisting of lumiflavin moiety and dopamine (FAD and dopamine in the enzyme), see Figure 8.3. For the corresponding reaction in solution and in the enzyme, the system was solvated in a spherical water droplet with a radius of 20 Å (centered in the reactive region) subject to surface-constrained all-atom solvent (SCAAS) boundary conditions.^[85] The spherical droplet was embedded in a 3 Å cubic grid of Langevin dipoles with a radius of 22 Å, which was in turn placed in a continuum with the dielectric constant of water. A cut-off of 10 Å was used for the calculation of long-range interactions and the neighbor list was updated every 30 steps.

All systems were subsequently relaxed with a time step size of 1 fs and gradually heated for 50 ps at 30 K, 100 ps at 100 K and 500 ps at 300 K. The system obtained in that way was considered to be relaxed and was mutated from a reactant to product diabatic states in 51 mapping frames, each consisting of a molecular dynamics simulation of 30 ps, using a 1 fs time step, for a total simulation time of 1.53 ns. The EVB simulation of the reactive trajectory was performed 10 times starting from different initial configurations, giving rise to ten independent free energy profiles. All EVB calculations were performed using the standard EVB free energy perturbation/umbrella sampling (EVB–FEP/US) procedure^[86] and the MOLARIS simulation package in combination with the ENZY MIX force field.^[84]

In order to study enzyme catalysis, it is highly desirable to have the experimental kinetic data for the reference reaction in aqueous solution. However, to the best of our knowledge, no such are available for the aqueous solution reaction between the flavin and dopamine (or any other monoamine or suitable model system). Some studies have shown that flavins react with amines and alcohols in water on a timescale of days, but no further kinetic data were provided, therefore, experimental value for the activation free energy of the reference reaction cannot be deduced.^[87,88] As a result, we parameterized our EVB Hamiltonian to reproduce the reliable M06–2X/6–31+G(d,p) energetics in vacuo ($\Delta G_{\text{gas}}^{\ddagger}$ and ΔG_{gas}^0). The system was subsequently moved to aqueous solution and MAO B active site using the same parameter set (EVB off-diagonal term plus the gas phase shift). This is a valid approximation, due to the demonstrated phase-independence of the EVB off-diagonal (H_{ij}) coupling term.^[89] While it is straightforward to obtain $\Delta G_{\text{gas}}^{\ddagger}$, the calculation of ΔG_{gas}^0 is complicated by the fact that, in the gas phase, the resulting FADH[−] anion and dopamine cation form an adduct immediately upon hydride transfer (see previous sections). In order to circumvent this issue, we selected the point in the IRC profile where the structure of FADH[−] moiety matches the geometry of the gas-phase optimized FADH[−] prior to adduct formation. In addition, the IRC profile shows a shoulder in the potential energy surface around the same point, thus suggesting a valid transient intermediate.^[74] Using this setup, the gas-phase calculations gave $\Delta G_{\text{gas}}^{\ddagger} = 32.8 \text{ kcal mol}^{-1}$ and $\Delta G_{\text{gas}}^0 = 24.9 \text{ kcal mol}^{-1}$, which were used to parameterize the EVB gas-phase shift and a coupling parameter. In aqueous solution, the activation free energy is still high, $26.5 \text{ kcal mol}^{-1}$, nevertheless, lowered by $6.3 \text{ kcal mol}^{-1}$ relative to the gas phase. Even more pronounced is the effect of the aqueous environment to the free energy of reaction, which is lowered to $4.2 \text{ kcal mol}^{-1}$. Including the full enzyme dimensionality reduced both the barrier to $\Delta G_{\text{MAO}}^{\ddagger} = 14.2 \text{ kcal mol}^{-1}$ and the reaction free energy to $\Delta G_{\text{MAO}}^0 = -0.6 \text{ kcal mol}^{-1}$.^[74] The reduction in the activation barrier relative to the aqueous solution is $12.3 \text{ kcal mol}^{-1}$, which corresponds to a rate-enhancement of more than nine orders of magnitude. The overall reaction is not excessively exergonic, allowing for the reverse reaction in the case of product overproduction. Experimental studies revealed that MAO substrates are bound to the active site in the protonated form, while they enter the reaction only if they are neutral. We calculated the $\text{p}K_{\text{a}}$ value for dopamine placed at the MAO B active site and we have shown that it is not significantly changed relative to its value in the aqueous solution.^[54] The reversible work necessary for dopamine deprotonation can be calculated then analytically and is assumed $1.9 \text{ kcal mol}^{-1}$. When this correction is added to the calculated barrier of $14.2 \text{ kcal mol}^{-1}$, the final activation free energy is $\Delta G_{\text{MAO}}^{\ddagger} = 16.1 \text{ kcal mol}^{-1}$. The latter value is in excellent agreement with the experimentally determined value of 16.5

Figure 8.9 Structure of MAO B Active Site with Neutral Dopamine. The active site residues are shown in gray, FAD prosthetic group in orange, dopamine in light blue, and Lys296 in violet. The orientation of the substrate facilitates the positioning of the pro-*R* hydrogen atom for the hydride transfer. (See color plate section for the color representation of this figure.)



kcal mol^{-1} ,^[19,20] thus strongly supporting the proposed hydride transfer mechanism. A snapshot from the simulation of MAO B active site with the reactive dopamine molecule is shown in Figure 8.9.

We have also observed that, for the efficient catalysis, the substrate amino group must be oriented towards the O4 atom of the flavin moiety to allow for hydrogen bonding. There is considerably more space on the O4 side of the active site cavity compared to the other, which allows for better accommodation of the substrate and proper positioning for the hydride transfer. MAO preferentially cleaves the pro-*R* hydrogen atom and in the mentioned configuration the pro-*R* hydrogen atom of the dopamine $\alpha - \text{CH}_2$ moiety is aligned in a favorable direction to enable the hydride transfer (Figure 8.9).

The Lys296 residue is part of the FAD–H₂O–Lys296 motif conserved in many flavin-dependent oxidases, and we speculated that it might be important in the MAO catalysis. We have performed our calculations with both the neutral and protonated form of Lys296 to quantify the effect of the ionization state of this residue on the hydride transfer reaction. Our results show that the protonation state of Lys296 does not affect the reductive half-reaction of MAO B and that the barrier for the reaction is virtually unchanged at around $14.0 \text{ kcal mol}^{-1}$.^[74] One may speculate that the protonation state of Lys296 may be relevant for the oxidative half-reaction, what is in accordance with the conclusion for mouse polyamine oxidase.^[90]

Following the same methodology, we also addressed the catalytic step of the MAO A catalyzed decomposition of noradrenaline.^[91] We showed that MAO A lowers the activation barrier by $14.3 \text{ kcal mol}^{-1}$ relative to the same reaction in aqueous solution. Taking into account the deprotonation of noradrenaline prior to the hydride transfer reaction, the activation barrier in the enzyme is calculated to be $20.3 \pm 1.6 \text{ kcal mol}^{-1}$, being in reasonably good agreement with the correlated experimental value of $16.5 \text{ kcal mol}^{-1}$.^[91] Overall, the results presented here offer a strong support that both MAO A and MAO B isoforms function by the same hydride transfer mechanism. To strengthen our conclusion, we considered the effect of point mutations on the activation free energy.^[91] The choice of the investigated point-mutations was prompted by the experimental work of Edmondson and coworkers,^[15,16] who expressed the mentioned mutant

enzymes of MAO B in *Pichia pastoris*, whereas the same attempt to do so for MAO A resulted in enzymes found to be unstable upon membrane extraction. Our results show that the pK_a uncorrected values for the activation free energy are as follows (in kcal mol⁻¹): WT MAO A (18.7), Tyr444Phe (18.2), Tyr444Leu (19.3) and Tyr444Trp (20.2), which confirmed the functional importance of the probed "aromatic cage" Tyr444 residue. Interestingly, Tyr444Phe mutation slightly lowered the activation free-energy, which indicates that the role of the mutated tyrosine residue is predominantly exerted through its aromatic moiety associated with linear quadrupole, rather than through the hydroxyl –OH group, which is a significant observation. Another very interesting aspect is provided with His and Glu mutants regarding their protonation forms. It turns out that Tyr444His(0) mutant, where His residue is unionized, gives an enzyme with 7 kcal mol⁻¹ higher activation free-energy than WT (25.7 kcal mol⁻¹), which is even further increased by another 9 kcal mol⁻¹ to $\Delta G^\ddagger = 34.9$ kcal mol⁻¹ in Tyr444His(+) upon protonating the histidine residue. This observation goes perfectly well with the idea that there is a positive charge build-up on the substrate in the transition state, thus strongly supporting the idea that during reaction the departing hydrogen is abstracted as an anion (hydride, H⁻). Therefore, positively charged species near the active site have an anti-catalytic effect as indicated by these results. To further test this concept, we investigated two additional Tyr444Glu mutants corresponding to unionized and negatively charged glutamic acid residues. For the neutral form, Tyr444Glu(0), the barrier is higher than for WT, assuming $\Delta G^\ddagger = 23.8$ kcal mol⁻¹, but is significantly reduced to $\Delta G^\ddagger = 21.7$ kcal mol⁻¹, when Glu residue is deprotonated to –CH₂–CH₂–COO⁻.^[91] The induced negative charge in this active site residue stabilizes the cationic transition state, which lowers the barrier. Taken all together, the results obtained for both Tyr444His and Tyr444Glu mutants and their different protonation forms provide confirmation that MAO enzymes operate through the direct hydride transfer mechanism.

8.8 Nuclear Quantum Effects in MAO Catalysis

The quantum-mechanical nature of nuclei motion in chemical reactions is experimentally evident in kinetic isotope effect (KIE) that is, by definition, the ratio of the rate constants for the species involving various isotopomers. If the isotopomer atoms are directly involved in the making and breaking of chemical bonds, this is termed primary KIE, otherwise it is denoted secondary KIE. The most pronounced are always the KIE values involving H/D isotopes, because of the relative differences in masses among the corresponding isotopes.

Interesting proposals were given for the role of tunneling and dynamical effects in enzyme catalysis. These proposals were mainly based on the experimental detection and careful quantization of hydrogen (radical, proton, or hydride) tunneling in enzymatic reactions. Several independent groups gathered evidence that room temperature nuclear tunneling occurs in several enzymes, especially those involving C–H bond activation.^[92–95] The donor-acceptor distance, which influences the shape of the potential for hydrogen transfer, played a special role in these proposals, in accordance with previous suggestions that steric strain and barrier compression can help catalysis.^[96] The rationale behind assigning a special relevance for catalysis to the modes involving donor-acceptor distance is that, for the compressed case, the barrier for the hydrogen

transfer just gets narrower at the preserved height allowing tunneling. The experimental results were interpreted by the vibronic formula approach^[97] and by a theory of electron transfer-coupled hydrogen transfer.^[98] In particular, recent studies argued that the promoting mode proposal is consistent with pressure effects on enzymatic reactions that support the idea of the vibrationally enhanced catalysis. By critical recompilation of the experimental and multiscale simulation data we have demonstrated that serious inconsistencies exist in the evidence to support these hypotheses.^[99] Tunneling reflected in H/D KIE decreases upon compression and external pressure does not lead to the applicable compression of the free energy surface. Moreover, pressure experiments do not provide actual evidence for the vibrationally enhanced catalysis.^[100] Finally, the temperature dependence of the entropy change in hydrogen transfer reactions is shown to reflect simple electrostatic effects.^[101] Hydrogen transfer reactions in enzymology can be explained by transition state theory and the concept of promoting modes is not necessary. It is enough to know the probability density for the reactive system at the transition state and equilibrium. Speaking about the computational methods to calculate the KIE, the method of choice is path integration, where each quantum atom is represented by a necklace of beads. The method yields correct ensemble averages, whereas probability densities are different for H and D giving rise to different activation free energies. Tunneling is not a dynamical phenomenon and, technically speaking, KIE can also be calculated by Monte Carlo method.

The rate limiting step of MAO catalyzed reaction involves a hydride transfer and, as a light particle, its motion obeys the laws of quantum rather than classical mechanics. Tunneling in MAO-catalyzed reactions was addressed by several experimental studies. Husain and coworkers^[102] studied bovine liver MAO B with benzylamine and deduced kinetic parameters spectrophotometrically from UV/VIS detection of an aldehyde signal as a function of time at 25°C. The H/D KIE turned out to be dependent on the oxygen level, and, under the oxygen and benzylamine saturation, the observed H/D KIE was 6.4–6.7, whereas at low oxygen levels the value increased to 8.7. Following from this, Walker and Edmondson^[19] applied advanced data processing techniques to study a series of *para* and *meta* substituted benzylamines with bovine MAO B at 25°C. The H/D KIE value for unsubstituted benzylamine was between 8.2–10.1, depending on the level of oxygen, while for substituted derivatives, the H/D KIE values ranged between 6.5 for *p*-Br substitution and 14.1 for *m*-Cl derivative. For the recombinant human liver MAO A with *para*-substituted benzylamines, Miller and Edmondson^[37] reported H/D KIE values in the range of 6–13. Scrutton and coworkers studied pH dependence of KIE in recombinant human liver MAO A with benzylamine and the value decreases from approximately 13 to 8 with increasing pH value.^[79] Such significant pH dependence of H/D KIE values strongly supports the idea that only the neutral substrate will enter the reaction. Wang and Edmondson addressed tunneling in a rat MAO A for a series of *para*-substituted benzylamines,^[20] and concluded that the H/D KIE values are pH-independent and range from 7 to 14, demonstrating a rate-limiting α -CH bond cleavage step in catalysis. Unfortunately, no experimental data are available for dopamine in any of the mentioned studies.

Here we report the results of the path integration calculated H/D KIE of the rate-limiting step of dopamine decomposition catalyzed by MAO B.^[103] We decided to proceed with dopamine, because of its immense importance in neuroscience and because we have developed simulation protocol for this reaction for classical treatment

of nuclear motion. Path integration was employed on the full enzyme dimensionality in the form of quantum classical path (QCP) method with an EVB potential energy surface.^[104] The QCP approach is based on the isomorphism between the nuclear wavefunction and the ring of quasiparticles that are propagated on the "quantum mechanical" potential U_{qm} . The quantum mechanical free-energy barrier is calculated using the following potential:

$$U_{qm} = \sum_{k=1}^p \frac{1}{2p} M \Omega^2 \Delta x_k^2 + \frac{1}{p} U(x_k) \quad (8.5)$$

Here $\Delta x_k = x_{k+1} - x_k$, where $x_{p+1} = x_1$. The last equality ensures that the necklace is closed. Moreover, $\Omega = p / \hbar \cdot \beta$, whereas M is the mass of the quantum atom, where U is the actual potential used in the simulation. The total quantum mechanical partition function can then be obtained by running classical trajectories of the quasiparticles with the potential U_{qm} and $\beta = 1 / k_B T$. The probability of being at the transition state is approximated by a probability distribution of the center of mass of the quasiparticles (the centroid) rather than the classical single point. At the temperature values approaching zero, the quantum correction to free energy reduces to the contribution of the zeroth vibrational level and matches the zero point energy. At the finite temperature values contributions from all the excited states are included. Traditional "on-the-fly" path integral simulations are demanding because of the poor equilibration between quantum and classical particles. The QCP approach offers an effective and rather simple way to overcome this problem by propagating classical trajectories on the classical potential energy surface of the reacting system and using the atom positions to generate the centroid position for the quantum mechanical partition function. This treatment is based on the finding that the quantum mechanical partition function can be expressed as:

$$Z_{qm}(\bar{x}) = Z_{cl}(\bar{x}) \left\langle \left\langle \exp \left\{ - \left(\frac{\beta}{p} \right) \sum_k U(x_k) - U(\bar{x}) \right\} \right\rangle_{fp} \right\rangle_U \quad (8.6)$$

where \bar{x} is the centroid position, $\langle \cdot \cdot \rangle_{fp}$ designates an average over the free particle quantum mechanical distribution obtained with the implicit constraint that \bar{x} coincides with the current position of the corresponding classical particle, and $\langle \cdot \cdot \rangle_U$ designates an average over the classical potential U . The quantum mechanical free energy correction calculated from the partition function reads $A_{QM} = -1 / \beta \cdot \ln(Z)$.

In order to evaluate the QCP correction to the activation free energy, it is necessary to perform simulation for the transition state and the reactant minimum. Both simulations are performed for H and D, respectively. Since the wave function is more delocalized for hydrogen than for deuterium, the corresponding A_{QM} values are different. The procedure was as follows: Initially, we used exactly the same protocol as for calculation of the reaction profile with the classical treatment, followed by running additional 600 ps of MD simulation at the reactant and transition states. We quantized nuclear motion for the dopamine methylene group next to the amino group and the N5 atom of the flavin moiety using 18 beads. In this way, the motion of four atoms was quantized corresponding to 12 degrees of freedom. 100 ps of QCP simulation were performed both for H and D derivatives. When performing calculations for the D isotopomer, both hydrogen atoms of the methylene group were replaced by D in order to facilitate comparison

with the experiment. The error was estimated by using 10 different starting points that were 1 ps apart. Our computational strategy evaluates the relevant nuclear quantum corrections and gives the H/D kinetic isotope effect of 12.8 ± 0.3 ,^[103] being in agreement with the experimental values in the range of 6.4–14.1 for MAO B with structurally related benzylamines.^[19] The calculated H/D KIE gives additional very strong evidence that MAO enzymes operate through our hydride transfer mechanism.^[53,74] The H/D KIE values for the enzymes where the contribution to the rate constant comes only from the zero point energy of the reactant well are between 3–8. The elevated value for MAO B indicates that non-negligible contribution comes from the tunneling effects. Our calculations showed once more that the QCP approach is a reliable computational tool for determining the quantum mechanical contributions to the activation free energies even in the case when significant tunneling contributions to the rate constant are present.

8.9 Relevance of MAO Catalyzed Reactions for Neurodegeneration

MAO catalyzed oxidative deamination reactions produce hydrogen peroxide as a by-product (Scheme 8.1) giving rise to several reactive oxygen species (ROS) that are responsible for oxidative stress. Unless cleared from the body by enzymes such as catalase or glutathione peroxidase, H_2O_2 molecules easily undergo Fenton-type chemistry to give OH^\bullet radicals. Therefore, MAO inhibition is an important strategy for the prevention and treatment of neurodegeneration. On the other hand, H_2O_2 production from the non-enzymatic metal ions mediated oxidative deamination of dopamine and noradrenaline seems to be particularly problematic, since it is not restricted to the mitochondrial membrane, where H_2O_2 scavenging enzymes are located in significant quantities.

A gross scheme of neurodegeneration on the molecular level is based on two pathways.^[8] Firstly, reactive species oxidize heavy atom ions, which enhances their interaction with α -synuclein (αSyn), thus promoting its folding to the beta form and giving rise to insoluble amyloid plaques. The latter prevents the function of vesicular transport leading to gradual neuronal death. In the second pathway, radical species, OH^\bullet in particular, react with the methylene groups of the apolar part of the lipid bilayer of either the cell or mitochondrial wall, resulting in membrane leakage followed by dyshomeostasis, loss of resting potential and neuron death. Unlike other cells, neurons are non-replicating and do not regenerate, thus the brain is, despite massive redundancy, sensitive to the loss of function if too many neurons cease to exist. A gradual loss of neurons clinically manifests as neurodegenerative diseases, such as Alzheimer's, Parkinson's and Huntington's disease. Additional complication is that initial damage of neurons triggers inflammatory response that produces even more ROS resulting in perpetual neuron damages. The incidence of neurodegeneration rapidly increases with age. There are estimates that in the western societies, the probability for neurodegenerative disorders, mainly Alzheimer's disease, at the age of 85 is about 50%, which increases to as much as 90% at the age of 95. It has to be mentioned that MAO is not the only source of ROS. Electron transfer chain is even a richer resource, but it cannot be inhibited. In this section we present a brief overview of potential strategies

targeted at neurodegeneration, while for a recent and more detailed review the reader can refer to reference^[8].

The central nervous system (CNS) is vulnerable to oxidative stress and neurodegeneration is its direct consequence. The main reason for the oxidative stress is that human brain consumes about ten times more oxygen than is the average over all other tissues, which is directly linked to the high-energy consumption of neural signal transduction. Human brain represents about 2% of the body weight and consumes 20% of the oxygen. In addition, neurons have large surface to volume ratio and therefore there is higher probability for the cellular membrane damage than it is for other cells. There are many dietary components that can not only scavenge ROS but also influence some of the biochemical events (signal transduction, stress protein synthesis, glycation and toxin generation) associated with neurodegenerative pathologies, thereby either ameliorating the risks or slowing down the progression of the disease. In general, food containing sulfur rich compounds, such as garlic, onion, and avocados, are also good options. Bilirubin is a very efficient ROS scavenger,^[105] which provides a possible explanation for a low incidence of cardiovascular diseases and, to a certain extent, neurodegeneration in patients with Gilbert–Meulengracht syndrome. A promising strategy for the prevention and, to a certain extent, treatment of neurodegeneration is the administration of curcumin, an essential ingredient of curry, which has recently been demonstrated to have significant neuroprotective potential.^[106] Interestingly, the prevalence of Alzheimer's disease in India among adults aged between 70 and 79 years is 4.4 times lower than in the United States.^[107] Green tea and coffee drinking seems to have neuroprotective potential. Catechins found in green tea can penetrate the hematoencephalic barrier and they act as metal chelating agents and ROS scavengers.^[108]

The ROS production originating from inflammatory processes in the central neural system can be blocked at the level of the arachidonic acid cascade with one of the COX-2 selective non-steroidal anti-inflammatory drugs (NSAID). Ibuprofen seems to be a first choice, because of its low ulceration potential.^[109] However, it remains a challenge to balance the benefit of NSAID administration with its unwanted side effects and to give recommendations for administering NSAID in the context of neuroprotection. It is probably not recommendable to administer NSAID to patients with no signs of neurodegenerative diseases, particularly as long-term therapies.

One can lower ROS production through the MAO pathway by inhibiting MAO with one of the irreversible MAO B inhibitors, such as selegiline and rasagiline, for which we suggested an operational chemical mechanism of the inhibition reaction.^[110,111] MAO A inhibitors, such as clorgyline, seem to be less appropriate because of their psychoactive properties. An unknown substance(s) in tobacco smoke also irreversibly inhibit MAO by up to 60%, suggesting that sporadic smoking in low quantities, for example, one cigar per week, could be beneficial for the prevention of neurodegeneration,^[112] whereas balancing its potential in the development of neoplasia and cardiovascular diseases. Interestingly, nicotine per se is a reversible inhibitor of MAO,^[113] whereas its metabolite nornicotine binds to the arginine side chain in α Syn, thus preventing conformational change to its β -form.^[114] Novel, promising strategies for the design of drugs used in the treatment of neurodegeneration are recently appearing. Weinreb and coworkers are working toward pleiotropic MAO inhibitors that simultaneously show potential for the iron ion chelation to be used in treating Alzheimer's disease.^[115]

An interesting strategy is the development of multi-target drugs possessing both cholinesterase and MAO-inhibitory activity like Ladostygil.^[116]

In this section, we have presented several known chemical mechanisms of neurodegeneration on the molecular level with particular emphasis on reactions involving the activity of MAO enzymes. One can view neurodegeneration as the interplay of several chemical reactions with complex kinetics. We hope that our results in this field will offer new insight into the features of molecular events linked to neurodegeneration, paving the way toward new strategies and rational drug design for the prevention and treatment of these debilitating diseases.^[8]

8.10 Conclusion and Perspectives

In this contribution we gave an overview of monomine oxidases, enzymes that catalyze the oxidative deamination of biogenic neurotransmitters and cardio- and vasoactive amines in various parts of body. We demonstrated using the QM-only cluster model that the rate-limiting first step for monoamine oxidases represents a hydride transfer from the methylene group next to the amino moiety to the cofactor flavin N5 atom. This is followed by the substrate amino group deprotonation to the flavin N1 site, which creates a fully reduced flavin, FADH₂, and neutral imine. Critical comparison between the structure and pK_a values of the ionizable groups of the active centers of MAO A and MAO B isozymes gives strong evidence that both enzymes operate by the same chemical mechanism. By using the empirical valence bond QM/MM approach we showed that MAO B lowers the barrier of the hydride transfer reaction for dopamine degradation by 12.3 kcal mol⁻¹ relative to the reference reaction in aqueous solution, corresponding to a rate-enhancement of more than 9 orders of magnitude. The barrier for the enzymatic reaction starting from the deprotonated substrate is 14.2 kcal mol⁻¹. Taking into account the free energy cost of dopamine deprotonation in the active site prior to the enzymatic reaction, the reaction barrier is 16.1 kcal mol⁻¹, in excellent agreement with the available experimental value of 16.5 kcal mol⁻¹. We calculated H/D kinetic isotope effect for MAO B catalyzed decomposition of dopamine that is in agreement with experimental values. In conjunction with additional experimental and computational work, the data presented here improve the understanding of the mechanism of the catalytic activity of MAO, as well as a large family of flavoenzymes, which can allow for the design of novel and improved MAO B inhibitors^[2,3] as transition-state analogues for antiparkinsonian and neuroprotective use.

Acknowledgements

We would like to thank Prof. Arieh Warshel (University of Southern California, USA) and Prof. Lynn Kamerlin (University of Uppsala, Sweden) for many stimulating discussions concerning multiscale modeling of enzyme reactions. We would like to acknowledge many precious contributions of Dr. Matej Repič, Miha Purg, Matic Poberžnik, Matic Pavlin and Dr. Rok Borštnar, without which our computational studies of monoaminergic systems would not be on the same level. R. V. gratefully acknowledges the European Commission for an individual FP7 Marie Curie Career

Integration Grant (contract number PCIG12–GA–2012–334493). J.M. would like to thank the Slovenian Research Agency for the financial support in the framework of the programme group P1–0012. Part of this work was supported by COST Action CM1103. The authors thank Dr. Matej Repič and Mag. Biochem. Aleksandra Maršavelski for assistance with the preparation of graphic artworks.

References

- 1 Di Giovanni, G., Di Matteo, V. and Esposito, E. (2008) *Serotonin-dopamine interaction: Experimental evidence and therapeutic relevance*, Elsevier, Amsterdam.
- 2 Ramsay, R.R. (2012) Monoamine oxidases: The biochemistry of the proteins as targets in medicinal chemistry and drug discovery. *Current Topics in Medicinal Chemistry*, **12**, 2189–2209.
- 3 Youdim, M.B.H., Edmondson, D.E. and Tipton, K.F. (2006) The therapeutic potential of monoamine oxidase inhibitors. *Nature Reviews Neuroscience*, **7**, 295–309.
- 4 Miller, J.R. and Edmondson, D.E. (1999) Influence of flavin analogue structure on the catalytic activities and flavinylation reactions of recombinant human liver monoamine oxidases A and B. *Journal of Biological Chemistry*, **274**, 23515–23525.
- 5 Klinman, J.P. (2007) How do enzymes activate oxygen without inactivating themselves? *Accounts of Chemical Research*, **40**, 325–333.
- 6 Gadda, G. (2012) Oxygen activation in flavoprotein oxidases: The importance of being positive. *Biochemistry*, **51**, 2662–2669.
- 7 Edmondson, D.E. (2014) Hydrogen peroxide produced by mitochondrial monoamine oxidase catalysis: Biological implications. *Current Pharmaceutical Design*, **20**, 155–160.
- 8 Pavlin, M., Repič, M., Vianello, R. and Mavri, J. (2016) The chemistry of neurodegeneration: Kinetic data and their implications. *Molecular Neurobiology*, **53**, 3400–3415.
- 9 Bach, A.W., Lan, N.C., Johnson, D.L. *et al.* (1988) cDNA cloning of human liver monoamine oxidase A and B: Molecular basis of differences in enzymatic properties. *Proceedings of the National Academy of Sciences of the United States of America*, **85**, 4934–4938.
- 10 Li, M., Hubalek, F., Newton-Vinson, P. and Edmondson, D.E. (2002) High-level expression of human liver monoamine oxidase A in *Pichia pastoris*: Comparison with the enzyme expressed in *Saccharomyces cerevisiae*. *Protein Expression and Purification*, **24**, 152–162.
- 11 Newton-Vinson, P., Hubalek, F. and Edmondson, D.E. (2000) High-level expression of human liver monoamine oxidase B in *Pichia pastoris*. *Protein Expression and Purification*, **20**, 334–345.
- 12 Binda, C., Newton-Vinson, P., Hubalek, F. *et al.* (2002) Structure of human monoamine oxidase B, a drug target for the treatment of neurological disorders. *Nature Structural Biology*, **9**, 22–26.
- 13 De Colibus, L., Li, M., Binda, C. *et al.* (2005) Three-dimensional structure of human monoamine oxidase A (MAO A): Relation to the structures of rat MAO A and human MAO B. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 12684–12689.

- 14 Son, S.Y., Ma, J., Kondou, Y. *et al.* (2008) Structure of human monoamine oxidase A at 2.2 Å resolution: The control of opening the entry for substrates/inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, **105**, 5739–5744.
- 15 Li, M., Binda, C., Mattevi, A. and Edmondson, D.E. (2006) Functional role of the “aromatic cage” in human monoamine oxidase B: Structures and catalytic properties of Tyr435 mutant proteins. *Biochemistry*, **45**, 4775–4784.
- 16 Akyuz, M.A., Erdem, S.S. and Edmondson, D.E. (2007) The aromatic cage in the active site of monoamine oxidase b: Effect on the structural and electronic properties of bound benzylamine and p-nitrobenzylamine. *Journal of Neural Transmission*, **114**, 693–698.
- 17 MacMillar, S., Edmondson, D.E. and Matsson, O. (2011) Nitrogen kinetic isotope effects for the Monoamine Oxidase B-catalyzed oxidation of benzylamine and (1,1-²H₂)benzylamine: Nitrogen rehybridization and CH bond cleavage are not concerted. *Journal of the American Chemical Society*, **133**, 12319–12321.
- 18 Klinman, J.P. and Matthews, R.G. (1985) Calculation of substrate dissociation constants from steady-state isotope effects in enzyme-catalyzed reactions. *Journal of the American Chemical Society*, **107**, 1058–1060.
- 19 Walker, M.C. and Edmondson, D.E. (1994) Structure-activity relationships in the oxidation of benzylamine analogs by bovine liver mitochondrial monoamine oxidase B. *Biochemistry*, **33**, 7088–7098.
- 20 Wang, J. and Edmondson, D.E. (2011) ²H kinetic isotope effects and pH dependence of catalysis as mechanistic probes of rat monoamine oxidase A: Comparisons with the human enzyme. *Biochemistry*, **50**, 7710–7717.
- 21 Harris, C.M., Pollegioni, L. and Ghisla, S. (2001) pH and kinetic isotope effects in D-amino acid oxidase catalysis: Evidence for a concerted mechanism in substrate dehydrogenation via hydride transfer. *European Journal of Biochemistry*, **268**, 5504–5520.
- 22 Kurtz, K.A., Rishavy, M.A., Cleland, W.W. and Fitzpatrick, P.F. (2000) Nitrogen isotope effects as probes of the mechanism of D-amino acid oxidase. *Journal of the American Chemical Society*, **122**, 12896–12897.
- 23 Ralph, E.C., Hirschi, J.S., Anderson, M.A. *et al.* (2007) Insights into the mechanism of the flavoprotein-catalyzed amine oxidation from nitrogen isotope effects of the reaction of N-methyltryptophan oxidase. *Biochemistry*, **46**, 7655–7664.
- 24 Ralph, E.C., Anderson, M.A., Cleland, W.W. and Fitzpatrick, P.F. (2006) Mechanistic studies of the flavoenzyme tryptophan 2-monoxygenase: Deuterium and ¹⁵N kinetic isotope effects on alanine oxidation by an L-amino acid oxidase. *Biochemistry*, **45**, 15844–15852.
- 25 Erdem, S.S., Karahan, O., Yildiz, I. and Yelekci, K. (2006) A computational study on the amine-oxidation mechanism of monoamine oxidase: Insight into the polar nucleophilic mechanism. *Organic & Biomolecular Chemistry*, **4**, 646–658.
- 26 Silverman, R.B. (1995) Radical ideas about monoamine oxidase. *Accounts of Chemical Research*, **28**, 335–342.
- 27 Lu, X., Ji, H. and Silverman, R.B. (2002) in *Flavins and flavoproteins* (Eds: (eds S. Chapman, R. Perham and N.S. Scrutton), University Park Press, Baltimore, pp. 817–830.

- 28 Vintém, A.P.B., Price, N.T., Silverman, R.B. and Ramsay, R.R. (2005) Mutation of surface cysteine 374 to alanine in monoamine oxidase A alters substrate turnover and inactivation by cyclopropylamines. *Bioorganic & Medicinal Chemistry*, **13**, 3487–3495.
- 29 Silverman, R.B. (1983) Mechanism of inactivation of monoamine oxidase by trans-2-phenylcyclopropylamine and the structure of the enzyme-inactivator adduct. *Journal of Biological Chemistry*, **258**, 14766–14769.
- 30 Fitzpatrick, P.F. (2010) Oxidation of amines by flavoproteins. *Archives of Biochemistry and Biophysics*, **493**, 13–25.
- 31 Kay, C.W.M., El Mkami, H., Molla, G. *et al.* (2007) Characterization of the covalently bound anionic flavin radical in monoamine oxidase a by electron paramagnetic resonance. *Journal of the American Chemical Society*, **129**, 16091–16097.
- 32 Rigby, S.E.J., Hynson, R.M.G., Ramsay, R.R. *et al.* (2005) A stable tyrosyl radical in monoamine oxidase A. *Journal of Biological Chemistry*, **280**, 4627–4631.
- 33 Dunn, R.V., Munro, A.W., Turner, N.J. *et al.* (2010) Tyrosyl radical formation and propagation in flavin dependent monoamine oxidases. *ChemBioChem*, **11**, 1228–1231.
- 34 Tan, A., Glantz, M.D., Piette, L.H. and Yasunobu, K.T. (1983) Electron spin-resonance analysis of the FAD in bovine liver monoamine oxidase. *Biochemical and Biophysical Research Communications*, **117**, 517–523.
- 35 Nandigama, R.K. and Edmondson, D.E. (2000) Structure–activity relations in the oxidation of phenethylamine analogues by recombinant human liver monoamine oxidase A. *Biochemistry*, **39**, 15258–15265.
- 36 Miller, J.R., Edmondson, D.E. and Grissom, C.B. (1995) Mechanistic probes of monoamine oxidase B catalysis: Rapid-scan stopped flow and magnetic field independence of the reductive half-reaction. *Journal of the American Chemical Society*, **117**, 7830–7831.
- 37 Miller, J.R. and Edmondson, D.E. (1999) Structure–activity relationships in the oxidation of para-substituted benzylamine analogues by recombinant human liver monoamine oxidase A. *Biochemistry*, **38**, 13670–13683.
- 38 Hamilton, G.A. (1971) Proton in biological redox reactions. *Progress in Bioorganic Chemistry*, **1**, 83–157.
- 39 Smith, M. and March, J. (2001) *March's advanced organic chemistry: Reactions, mechanisms and structure*, 5th edn, Wiley, New York.
- 40 Orru, R., Aldeco, M. and Edmondson, D.E. (2013) Do MAO A and MAO B utilize the same mechanism for the C–H bond cleavage step in catalysis? Evidence suggesting differing mechanisms. *Journal of Neural Transmission*, **120**, 847–851.
- 41 Umhau, S., Pollegioni, L., Molla, G. *et al.* (2000) The X-ray structure of D-amino acid oxidase at very high resolution identifies the chemical mechanism of flavin-dependent substrate dehydrogenation. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 12463–12468.
- 42 Ramos, M.J. and Fernandes, P.A. (2008) Computational enzymatic catalysis. *Accounts of Chemical Research*, **41**, 689–698.
- 43 Shaik, S., Cohen, S., Wang, Y. *et al.* (2010) P450 enzymes: Their structure, reactivity, and selectivity-modeled by QM/MM calculations. *Chemical Reviews*, **110**, 949–1017.

- 44 Lonsdale, R., Ranaghan, K.E. and Mulholland, A.J. (2010) Computational enzymology. *Chemical Communications*, **46**, 2354–2372.
- 45 Carvalho, A.T.P., Barrozo, A., Doron, D. *et al.* (2014) Challenges in computational studies of enzyme structure, function and dynamics. *Journal of Molecular Graphics and Modelling*, **54**, 62–79.
- 46 Siegbahn, P.E.M. and Borowski, T. (2006) Modeling enzymatic reactions involving transition metals. *Accounts of Chemical Research*, **39**, 729–738.
- 47 Himo, F. (2006) Quantum chemical modeling of enzyme active sites and reaction mechanisms. *Theoretical Chemistry Accounts*, **116**, 232–240.
- 48 Siegbahn, P.E.M. and Himo, F. (2011) The quantum chemical cluster approach for modeling enzyme reactions. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **1**, 323–336.
- 49 Senn, H.M. and Thiel, W. (2009) QM/MM methods for biomolecular systems. *Angewandte Chemie International Edition*, **48**, 1198–1229.
- 50 van der Kamp, M.W. and Mulholland, A.J. (2013) Combined quantum mechanics/molecular mechanics (QM/MM) methods in computational enzymology. *Biochemistry*, **52**, 2708–2728.
- 51 Liao, R.-Z., Yu, J.-G. and Himo, F. (2011) Quantum chemical modeling of enzymatic reactions: the case of decarboxylation. *Journal of Chemical Theory and Computation*, **7**, 1494–1501.
- 52 Siegbahn, P.E.M. (2006) The performance of hybrid DFT for mechanisms involving transition metal complexes in enzymes. *Journal of Biological Inorganic Chemistry*, **11**, 695–701.
- 53 Vianello, R., Repič, M. and Mavri, J. (2012) How are biogenic amines metabolized by monoamine oxidases? *European Journal of Organic Chemistry*, **2012**, 7057–7065.
- 54 Borštnar, R., Repič, M., Kamerlin, S.C.L. *et al.* (2012) Computational study of the pK_a values of potential catalytic residues in the active site of monoamine oxidase B. *Journal of Chemical Theory and Computation*, **8**, 3864–3870.
- 55 Zhao, Y. and Truhlar, D.G. (2008) The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theoretical Chemistry Accounts*, **120**, 215–241.
- 56 Zhao, Y. and Truhlar, D.G. (2011) Density functional theory for reaction energies: Test of meta and hybrid meta functionals, range-separated functionals, and other high-performance functionals. *Journal of Chemical Theory and Computation*, **7**, 669–676.
- 57 Cheong, P.H.-Y., Legault, C.Y., Um, J.M. *et al.* (2011) Quantum mechanical investigations of organocatalysis: Mechanisms, reactivities, and selectivities. *Chemical Reviews*, **111**, 5042–5137.
- 58 Bell, A.T. and Head-Gordon, M. (2011) Quantum mechanical modeling of catalytic processes. *Annual Review of Chemical and Biomolecular Engineering*, **2**, 453–477.
- 59 Cossi, M., Rega, N., Scalmani, G. and Barone, V. (2003) Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *Journal of Computational Chemistry*, **24**, 669–681.

- 60 Newton-Vinson, P. and Edmondson, D.E. (1999) in *Flavins and flavoproteins*(Eds: (eds S. Ghisla, P. Kroneck, P. Macheroux and H. Sund), Agency for Scientific Publications, Berlin.
- 61 Hull, L.A., Davis, G.T., Rosenblatt, D.H. and Mann, C.K. (1969) Oxidations of amines. VII. Chemical and electrochemical correlations. *Journal of Physical Chemistry*, **73**, 2142–2146.
- 62 Binda, C., Li, M., Hubalek, F. *et al.* (2003) Insights into the mode of inhibition of human mitochondrial monoamine oxidase B from high-resolution crystal structures. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 9750–9755.
- 63 de Grotthuss, C.J.T. (1806) Sur la décomposition de l'eau et des corps qu'elle tient en dissolution à l'aide de l'électricité galvanique. *Annali Di Chimica*, **58**, 54–73.
- 64 Macheroux, P., Ghisla, S., Sanner, C. *et al.* (2005) Reduced flavin: NMR investigation of N(5)-H exchange mechanism, estimation of ionisation constants and assessment of properties as biological catalyst. *BMC Biochemistry*, **6**, 26–36.
- 65 Coles, M.P., Aragón-Sáez, P.J., Oakley, S.H. *et al.* (2009) Superbasicity of a bis-guanidino compound with a flexible linker: A theoretical and experimental study. *Journal of the American Chemical Society*, **131**, 16858–16868.
- 66 Edmondson, D.E., Bhattacharyya, A.K. and Walker, M.C. (1993) Spectral and kinetic studies of imine product formation in the oxidation of p-(N,N-dimethylamino)benzylamine analogs by monoamine oxidase B. *Biochemistry*, **32**, 5196–5202.
- 67 Woo, J.C.G. and Silverman, R.B. (1995) Monoamine oxidase B catalysis in low aqueous medium. Direct evidence for an imine product. *Journal of the American Chemical Society*, **117**, 1663–1664.
- 68 Ding, C.Z., Lu, X., Nishimura, K. and Silverman, R.B. (1993) Transformation of monoamine oxidase-B primary amine substrates into time-dependent inhibitors. Tertiary amine homologs of primary amine substrates. *Journal of Medicinal Chemistry*, **36**, 1711–1715.
- 69 Akyüz, M.A. and Erdem, S.S. (2013) Computational modeling of the direct hydride transfer mechanism for the MAO catalyzed oxidation of phenethylamine and benzylamine: ONIOM (QM/QM) calculations. *Journal of Neural Transmission*, **120**, 937–945.
- 70 Atalay, V.E. and Erdem, S.S. (2013) A comparative computational investigation on the proton and hydride transfer mechanisms of monoamine oxidase using model molecules. *Computational Biology and Chemistry*, **47**, 181.
- 71 Zapata-Torres, G., Fierro, A., Barriga-González, G. *et al.* (2015) Revealing monoamine oxidase B catalytic mechanisms by means of the quantum chemical cluster approach. *Journal of Chemical Information and Modeling*, **55**, 1349–1360.
- 72 Karasulu, B. and Thiel, W. (2015) Amine oxidation mediated by N-methyltryptophan oxidase: Computational insights into the mechanism, role of active-site residues, and covalent flavin binding. *ACS Catalysis*, **5**, 1227–1239.
- 73 Kopacz, M.M., Heuts, D.P.H.M. and Fraaije, M.W. (2014) Kinetic mechanism of putrescine oxidase from *Rhodococcus erythropolis*. *FEBS Journal*, **81**, 4384–4393.
- 74 Repič, M., Vianello, R., Purg, M. *et al.* (2014) Empirical valence bond simulations of the hydride transfer step in the monoamine oxidase B catalyzed metabolism of dopamine. *Proteins: Structure, Function, and Bioinformatics*, **82**, 3347.

- 75 Warshel, A., Sharma, P.K., Kato, M. *et al.* (2006) Electrostatic basis for enzyme catalysis. *Chemical Reviews*, **106**, 3210–3235.
- 76 Ramsay, R.R. (1997) Mechanistic study of monoamine oxidase: Significance for MAO A and MAO B in situ. *Voprosy Meditsinskoi Khimii*, **43**, 457–470.
- 77 Ramsay, R.R., Dunford, C. and Gillman, P.K. (2007) Methylene blue and serotonin toxicity: Inhibition of monoamine oxidase A (MAO A) confirms a theoretical prediction. *British Journal of Pharmacology*, **152**, 946–951.
- 78 Ramsay, R.R., Koerber, S.C. and Singer, T.P. (1987) Stopped-flow studies on the mechanism of oxidation of *N*-methyl-4-phenyl-tetrahydropyridine by bovine liver monoamine oxidase B. *Biochemistry*, **26**, 3045–3050.
- 79 Dunn, R.V., Marshall, K.R., Munro, A.W. and Scrutton, N.S. (2008) The pH dependence of kinetic isotope effects in monoamine oxidase A indicates stabilization of the neutral amine in the enzyme–substrate complex. *FEBS Journal*, **275**, 3850–3858.
- 80 Repič, M., Purg, M., Vianello, R. and Mavri, J. (2014) Examining electrostatic pre-organization in monoamine oxidases A and B by structural comparison and pK_a calculations. *Journal of Physical Chemistry B*, **118**, 4326–4332.
- 81 Kato, M. and Warshel, A. (2006) Using a charging coordinate in studies of ionization induced partial unfolding. *Journal of Physical Chemistry B*, **110**, 11566–11570.
- 82 Sham, Y.Y., Chu, Z.T., Tao, H. and Warshel, A. (2000) Examining methods for calculations of binding free energies: LRA, LIE, PDL–LRA, and PDL/S–LRA calculations of ligands binding to an HIV protease. *Proteins: Structure, Function, and Bioinformatics*, **39**, 393–407.
- 83 Schutz, C.N. and Warshel, A. (2001) What are the dielectric “constants” of proteins and how to validate electrostatic models? *Proteins: Structure, Function, and Bioinformatics*, **44**, 400–417.
- 84 Lee, F.S., Chu, Z.T. and Warshel, A. (1993) Microscopic and semimicroscopic calculations of electrostatic energies in proteins by the POLARIS and ENZYMIK programs. *Journal of Computational Chemistry*, **14**, 161–185.
- 85 Warshel, A. and King, G. (1985) Polarization constraints in molecular dynamics simulation of aqueous solutions: The surface constraint all atom solvent (SCAAS) model. *Chemical Physics Letters*, **121**, 124–129.
- 86 Kamerlin, S.C.L. and Warshel, A. (2011) The empirical valence bond model: Theory and applications. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **1**, 30–45.
- 87 Brown, L.E. and Hamilton, G.A. (1970) Model reactions and a general mechanism for flavoenzyme-catalyzed dehydrogenations. *Journal of the American Chemical Society*, **92**, 7225–7227.
- 88 Kim, J.M., Bogdan, M.A. and Mariano, P.S. (1993) Mechanistic analysis of the 3-methylflavin-promoted oxidative deamination of benzylamine. A potential model for monoamine oxidase catalysis. *Journal of the American Chemical Society*, **115**, 10591–10595.
- 89 Hong, G., Rosta, E. and Warshel, A. (2006) Using the constrained DFT approach in generating diabatic surfaces and off diagonal empirical valence bond terms for modeling reactions in condensed phases. *Journal of Physical Chemistry B*, **110**, 19570–19574.

- 90 Henderson Pozzi, M. and Fitzpatrick, P.F. (2010) A lysine conserved in the monoamine oxidase family is involved in oxidation of the reduced flavin in mouse polyamine oxidase. *Archives of Biochemistry and Biophysics*, **498**, 83–88.
- 91 Poberžnik, M., Purg, M., Repič, M. *et al.* (2016) Empirical valence bond simulations of the hydride-transfer step in the monoamine oxidase A catalyzed metabolism of noradrenaline. *Journal of Physical Chemistry B*, in print. DOI: 10.1021/acs.jpccb.6b09011.
- 92 Klinman, J.P. (2015) Dynamically achieved active site precision in enzyme catalysis. *Accounts of Chemical Research*, **48**, 449–456.
- 93 Klinman, J.P. (2010) Enzyme dynamics: Control of active-site compression. *Nature Chemistry*, **11**, 907–909.
- 94 Hay, S., Pudney, C.R., McGrory, T.A. *et al.* (2009) Barrier compression enhances an enzymatic hydrogen-transfer reaction. *Angewandte Chemie International Edition*, **121**, 1480–1482.
- 95 Dzierlenga, M.W., Antoniou, D. and Schwartz, S.D. (2015) Another look at the mechanisms of hydride transfer enzymes with quantum and classical transition path sampling. *Journal of Physical Chemistry Letters*, **6**, 1177–1181.
- 96 a Jencks, W.P. (1987) *Catalysis in chemistry and enzymology*, Dover, New York; b Phillips, D.C. (1966) The 3-dimensional structure of an enzyme molecule. *Scientific American*, **215**, 78–90.
- 97 Klinman, J.P. (2009) An integrated model for enzyme catalysis emerges from studies of hydrogen tunneling. *Chemical Physics Letters*, **471**, 179–193.
- 98 Hammes-Schiffer, S. (2010) Introduction: Proton-coupled electron transfer. *Chemical Reviews*, **110**, 6937–6938.
- 99 Kamerlin, S.C.L., Mavri, J. and Warshel, A. (2010) Examining the case for the effect of barrier compression on tunneling, vibrationally enhanced catalysis, catalytic entropy and related issues. *FEBS Letters*, **584**, 2759–2766.
- 100 Hay, S., Johannissen, L.O., Sutcliffe, M.J. and Scrutton, N.S. (2010) Barrier compression and its contribution to both classical and quantum mechanical aspects of enzyme catalysis. *Biophysical Journal*, **98**, 121–128.
- 101 Liu, H. and Warshel, A. (2007) Origin of the temperature dependence of isotope effects in enzymatic reactions: The case of dihydrofolate reductase. *Journal of Physical Chemistry B*, **111**, 7852–7861.
- 102 Husain, M., Edmondson, D.E. and Singer, T.P. (1982) Kinetic studies on the catalytic mechanism of liver monoamine oxidase. *Biochemistry*, **21**, 595–600.
- 103 Mavri, J., Matute, R., Chu, Z.T. and Vianello, R. (2015) Path integral simulation of the H/D kinetic isotope effect in monoamine oxidase B catalyzed decomposition of dopamine. *Journal of Physical Chemistry B*, **120**, 3488–3492.
- 104 Hwang, J.K. and Warshel, A. (1993) A quantized classical path approach for calculations of quantum mechanical rate constants. *Journal of Physical Chemistry*, **97**, 10053–10058.
- 105 Joshi, M., Billing, B.H. and Hallinanm, T. (1995) Investigation of the role of reactive oxygen species in bilirubin metabolism in the Gunn rat. *Biochimica et Biophysica Acta – General Subjects*, **1243**, 244–250.
- 106 Trujillo, J., Granados-Castro, L.F., Zazueta, C. *et al.* (2014) Mitochondria as a target in the therapeutic properties of curcumin. *Archiv der Pharmazie*, **347**, 873–884.

- 107 Mishra, S. and Palanivelu, K. (2008) The effect of curcumin (turmeric) on Alzheimer's disease: An overview. *Annals of Indian Academy of Neurology*, **11**, 13–19.
- 108 Mandel, S., Amit, T., Reznichenko, L. *et al.* (2006) Green tea catechins as brain-permeable, natural iron chelators-antioxidants for the treatment of neurodegenerative disorders. *Molecular Nutrition & Food Research*, **50**, 229–234.
- 109 Gao, X., Chen, H., Schwarzschild, M.A. and Ascherio, A. (2011) Use of ibuprofen and risk of Parkinson disease. *Neurology*, **76**, 863–869.
- 110 Borštnar, R., Repič, M., Kržan, M. *et al.* (2011) Irreversible inhibition of monoamine oxidase B by the antiparkinsonian medicines rasagiline and selegiline: A computational study. *European Journal of Organic Chemistry*, **32**, 6419–6433.
- 111 Pavlin, M., Mavri, J., Repič, M. and Vianello, R. (2013) Quantum-chemical approach to determining the high potency of clorgyline as an irreversible acetylenic monoamine oxidase inhibitor. *Journal of Neural Transmission*, **120**, 875–882.
- 112 Fowler, J.S., Volkow, N.D., Wang, G.J. *et al.* (1996) Brain monoamine oxidase A inhibition in cigarette smokers. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 14065–14069.
- 113 Pavlin, R. and Sket, D. (1993) Effect of cigarette smoke on brain monoamine oxidase activity. *Farmaceutski Vestnik*, **44**, 185–192.
- 114 Dickerson, T.J. and Janda, K.D. (2003) Glycation of the amyloid β -protein by a nicotine metabolite: A fortuitous chemical dynamic between smoking and Alzheimer's disease. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 8182–8187.
- 115 Weinreb, O., Mandel, S., Bar-Am, O. and Amit, T. (2011) Iron-chelating backbone coupled with monoamine oxidase inhibitory moiety as novel pluripotential therapeutic agents for Alzheimer's disease: A tribute to Moussa Youdim. *Journal of Neural Transmission*, **118**, 479–492.
- 116 Weinstock, M., Bejar, C., Wang, R.H. *et al.* (2000) TV3326, a novel neuroprotective drug with cholinesterase and monoamine oxidase inhibitory activities for the treatment of Alzheimer's disease. *Journal of Neural Transmission. Supplementa*, **60**, 157–169.

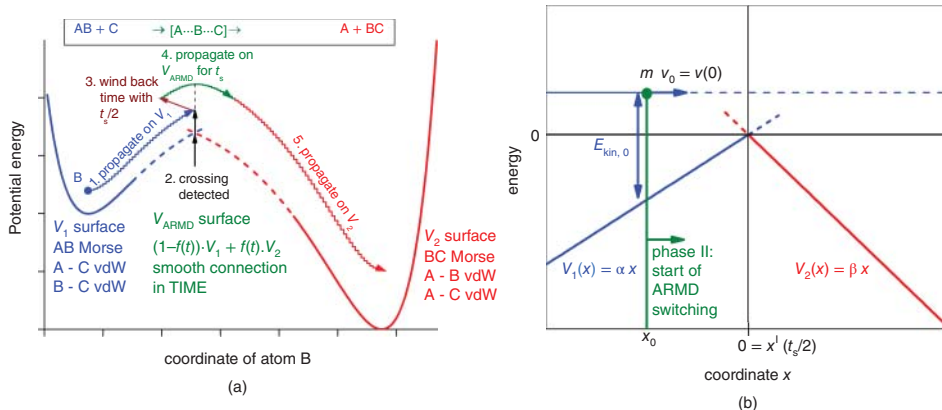


Figure 1.1 (a) The ARMD Method: Schematic Figure of the ARMD Simulation Method For a Collinear Reaction, Where Atom B is Transferred From Donor Atom A to Acceptor Atom C. During crossing the surfaces are switched in time and the Morse bond is replaced by van der Waals (vdW) interactions and vice versa. (b) Simple model for estimating energy violation in ARMD simulations. The system with mass m is approaching from the left on PES $V_1(x)$ (phase I). At $t = 0$ time it is at x_0 with velocity v_0 and kinetic energy $E_{kin,0}$. After crossing is detected at $x = 0$ the time is rewound by $t_s/2$ and the dynamics is re-simulated while $V_1(x)$ is being switched to $V_2(x)$ in t_s (phase II).

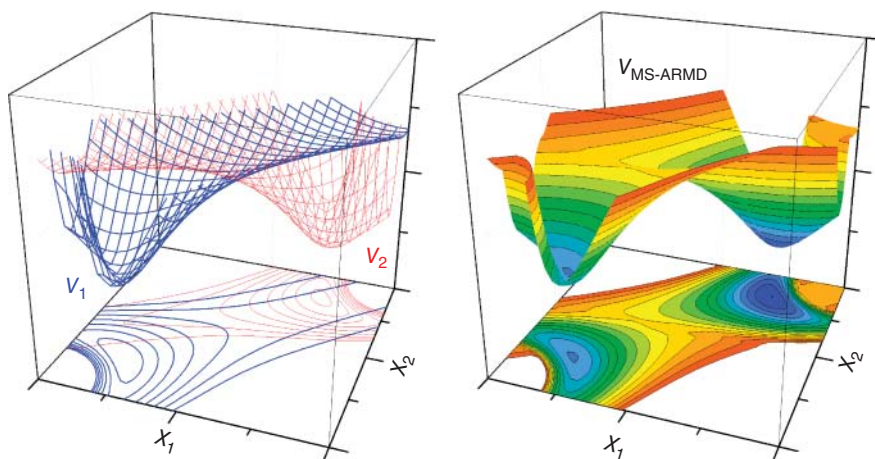
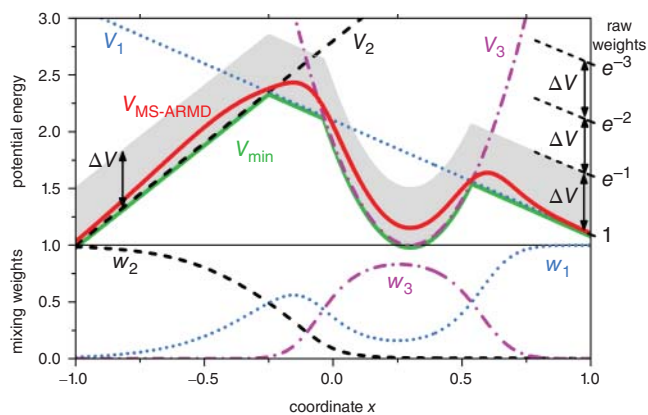


Figure 1.2 MS-ARMD Switching: The MS-ARMD Switching Method Applied in One and Two Dimensions to 3 and 2 surfaces ($V_{1,2,3}$). The effective surface is ($V_{\text{MS-ARMD}}$) always close to the lowest-energy surface (V_{min}), except for regions where other surfaces are within a few times ΔV (here = 0.5) in energy. Here, the algorithm switches smoothly among them by varying their weights ($w_{1,2,3}$; lower left panel).

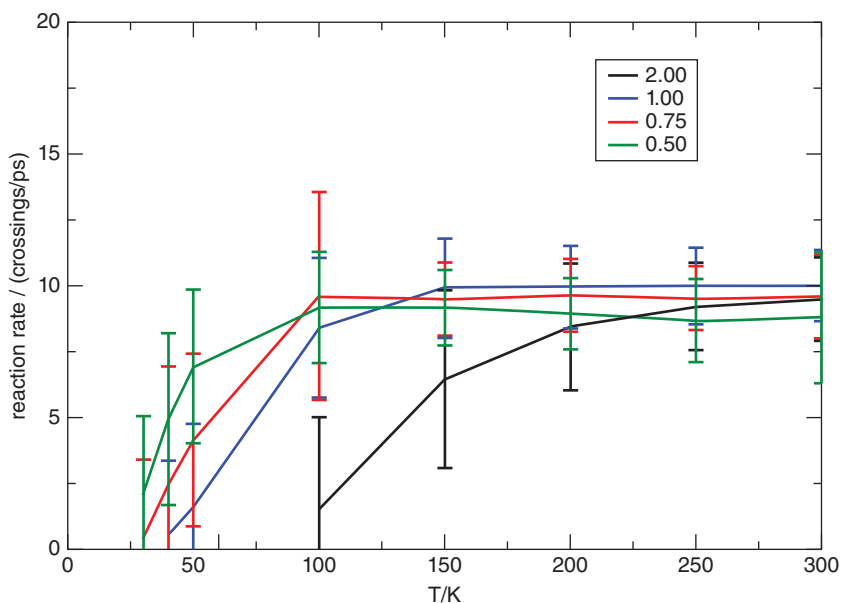


Figure 1.3 Proton Transfer Rates: Rates For Proton Transfer as a Function of T For $\text{NH}_4^+ \cdot \text{NH}_3$ Together with Error Bars. Rates are given for different energy scaling factors λE (proton-transfer barrier heights): $\lambda E = 2.0$ (black), 1.0 (blue), 0.75 (red), and 0.5 (green). For high temperatures the energy scaling has a smaller effect on the barrier than for low temperatures. This also explains why the rates agree within statistical fluctuations for $T \approx 300$ K. The figure is taken from the references.^[36]

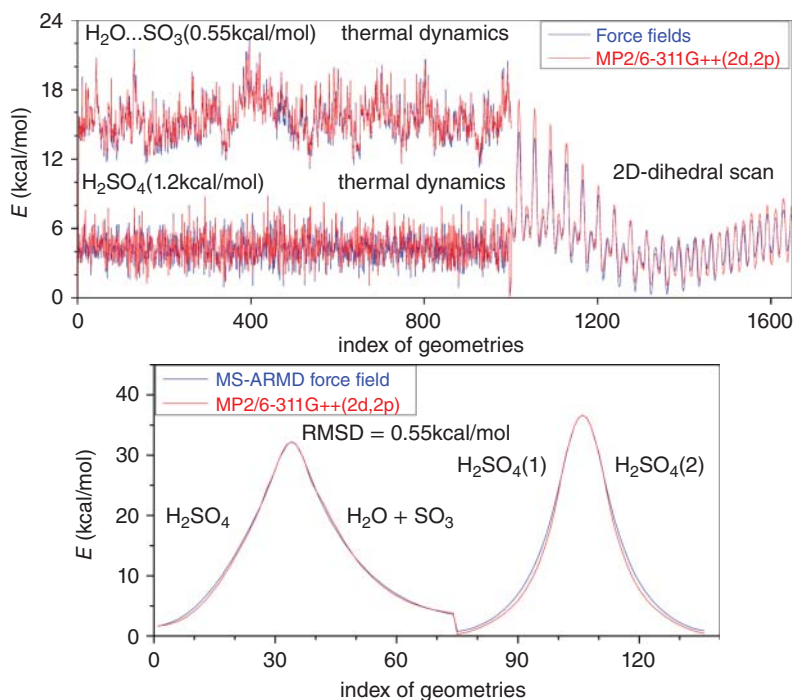


Figure 1.4 Reference MP2 Ab Initio and MS-ARMD Force Field Energies For States H_2SO_4 and $\text{H}_2\text{O} + \text{SO}_3$ and Along the Minimum Energy Paths For Water Elimination and For Intramolecular H-Transfer.

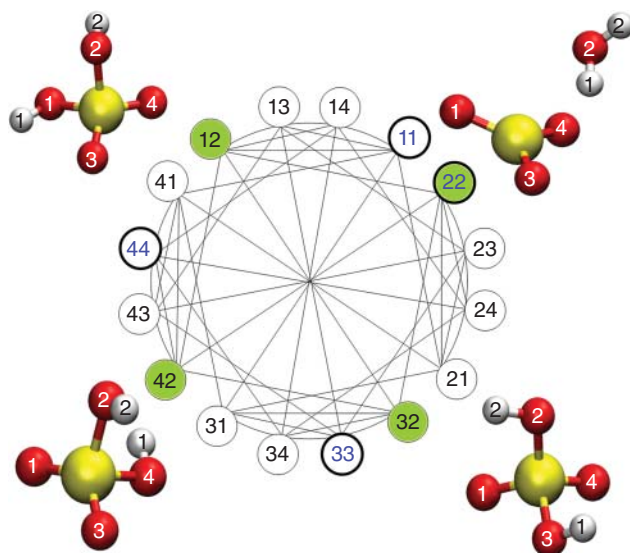


Figure 1.5 Reaction Network For H_2SO_4 : The Network of States (Circles) and Possible Reactions Between Them (Edges) For H_2SO_4 Represented as a Symmetric Graph. States labelled by two-digit numbers denoting the O indices of atoms to which atoms H1 and H2 bind. The four thick-line circles with identical digits indicates $\text{H}_2\text{O} + \text{SO}_3$ states, whereas the remaining 8 are H_2SO_4 states. Four states highlighted in green and given as a ball-and-stick representation were observed in a typical trajectory ($12 \rightarrow 32 \rightarrow 12 \rightarrow 42 \rightarrow 22$) showing three subsequent intramolecular H-transfers and water elimination at the end.

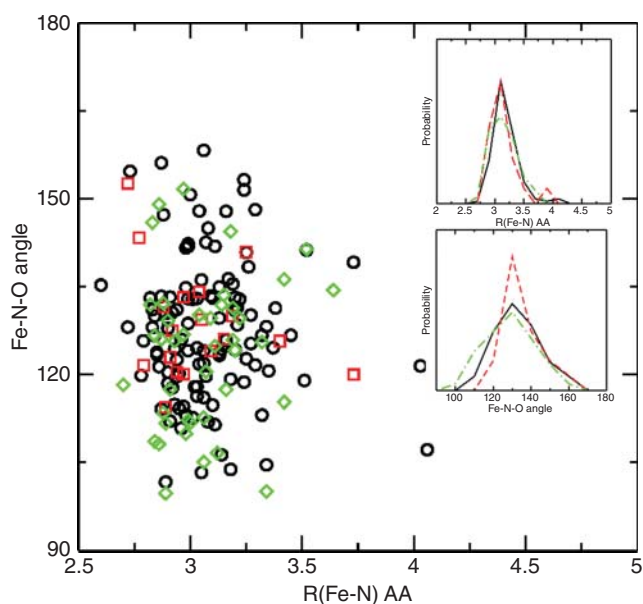


Figure 1.6 Crossing Seam For NO-Rebinding: The Crossing Geometries of NO Rebinding to Mb For Three Different Values of Δ (Red Squares $\Delta = 60$, Black Circles $\Delta = 65$, and Green Diamonds $\Delta = 70$ kcal/mol) Projected Onto the Distance Between Iron and the Nitrogen of the Ligand R_{FeN} 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 $\Delta = 60, 65$, and 70 kcal/mol, respectively. The figure is taken from Ref.^[38].

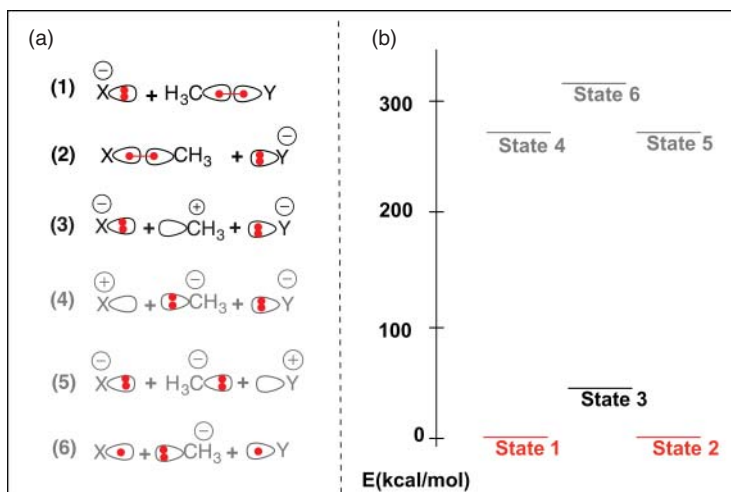


Figure 2.2 (a) Different VB Configurations Describing All Possible Independent Arrangements of Four Active Electrons Within Three Centers (X, CH₃, and Y). (b) Relative Approximate Energy Levels of the VB States; see Table 2.1 For the Estimation of These Energies.

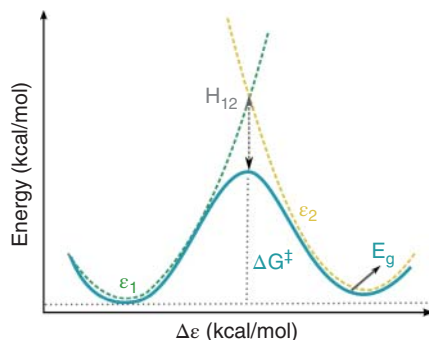


Figure 2.3 Relationship Between the Diabatic Parabolas ϵ_1 and ϵ_2 and the Adiabatic Ground State Free Energy E_g of a Hypothetical Two-State Reaction, Using an EVB Description.

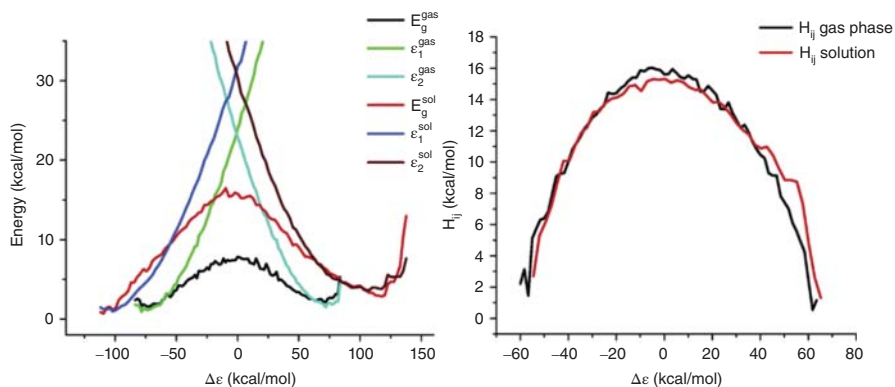


Figure 2.4 (left) Diabatic and Adiabatic FDFD Energy Profiles For the Reaction, $\text{Cl}^- + \text{CH}_3\text{Cl} \rightarrow \text{ClCH}_3 + \text{Cl}^-$, in Gas Phase and in Solution, Where the Reaction Coordinate is Defined as the Energy Difference Between the Diabatic Surfaces, $\Delta\epsilon = \epsilon_1 - \epsilon_2$. (Right) Plot of the H_{ij} of the reaction, $\text{Cl}^- + \text{CH}_3\text{Cl} \rightarrow \text{ClCH}_3 + \text{Cl}^-$, both in the gas phase and in solution.^[59] Reprinted with permission from reference^[59]. Copyright 2006 American Chemical Society.

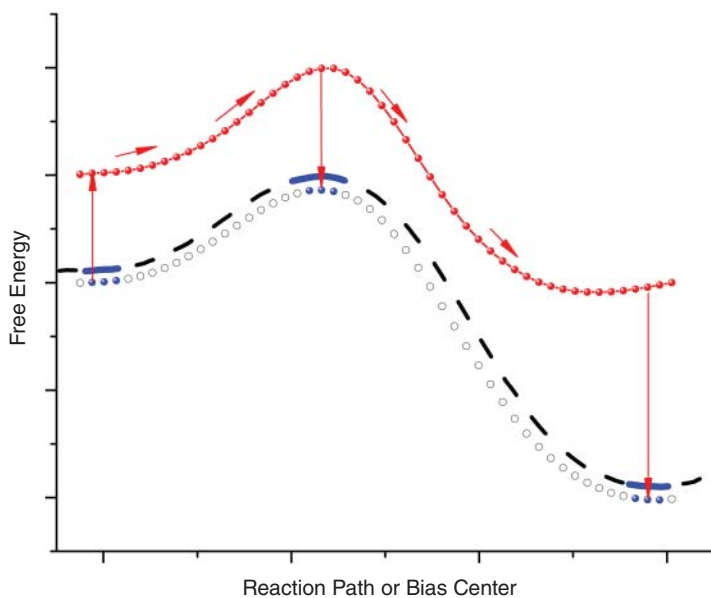


Figure 3.2 Thermodynamic Cycle Used in Calculation of the Activation Free Energy by Positioning Regions of the Target Free Energy Surface (Blue Solid Line) at the Final States and at the Transition State Relative to the Reference Free Energy Shifts (Red Dots). Black empty dots show the target free energy shifts which are computed with a potential of mean force (PMF); the complete target free energy profile (from the PMF) is shown with a dashed black line. The red arrows show the thermodynamic detour that is an alternative to computing the full target PMF. Note that red vertical arrows correspond to the free energy changes of moving between the reference and the target potentials with the same bias. Adapted from reference^[35].

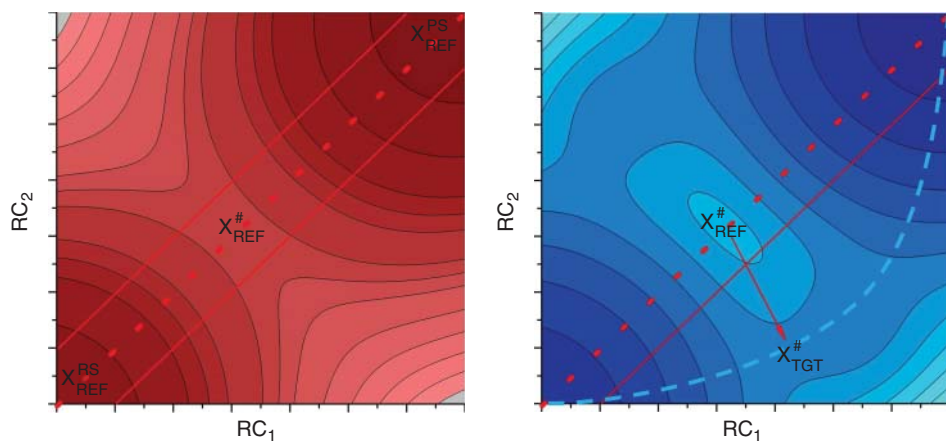


Figure 3.3 Challenges in Application of the PD Approach. The first challenge is the existence of a difference between the reaction paths on the reference (RED) and on the target (BLUE) free energy surfaces. The reference reaction path (dashed red line) is concerted while the target reaction path (dashed blue line) is dissociative (or associative). If sampling is performed only along the reference path – the explored region (between two red solid lines) does not contain the target transition state (TS). Thus, the free energy change of moving to the target potential at the reference TS does not provide the actual correction for the target free energy barrier. Secondly, moving from the reference TS to the target TS while computing the free energy perturbation (red arrow) will lead to a slow convergence since not only the force law changes, but also the system moves along the reaction surface.

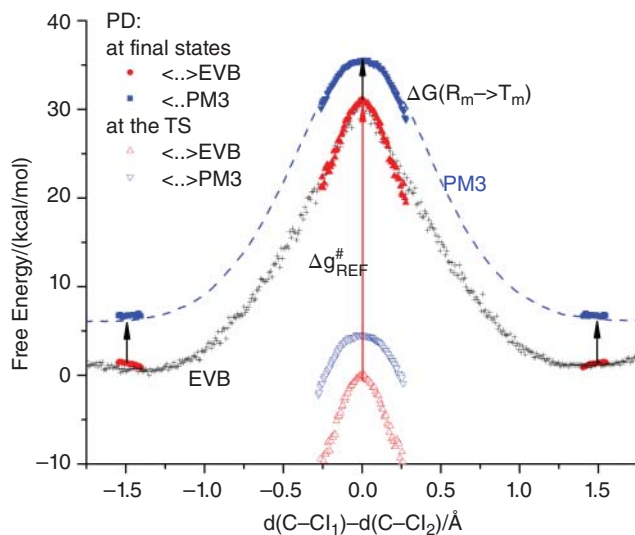


Figure 3.4 PM3/MM Activation Free Energy Barrier For $\text{MeCl} + \text{Cl}^- \text{S}_{\text{N}}2$ Reaction Computed with the Paradynamics Model^[25] Using EVB as a Reference Potential (Which was Parametrically Refined^[24]). The dashed blue line is the PM3 target free energy surface generated with the PMF approach. To compute the activation barrier with the PD approach, the EVB profile is computed first (black "+" symbols). Next, local free energy functions are computed at the TS (empty blue flipped triangles for PM3 and empty red triangles for EVB) and at final states (solid blue squares for PM3 and red circles for EVB). The relative positions of the local free energy functions are computed using the LRA estimator (black arrows). The EVB reference free energy functions are aligned to match the full EVB free energy profile (red triangles with black "+" symbols), thus the local PM3 free energy function at the TS is positioned relative to the ones at the final states (flipped blue triangles). Adapted with permission from reference^[25]. Copyright 2012 American Chemical Society.

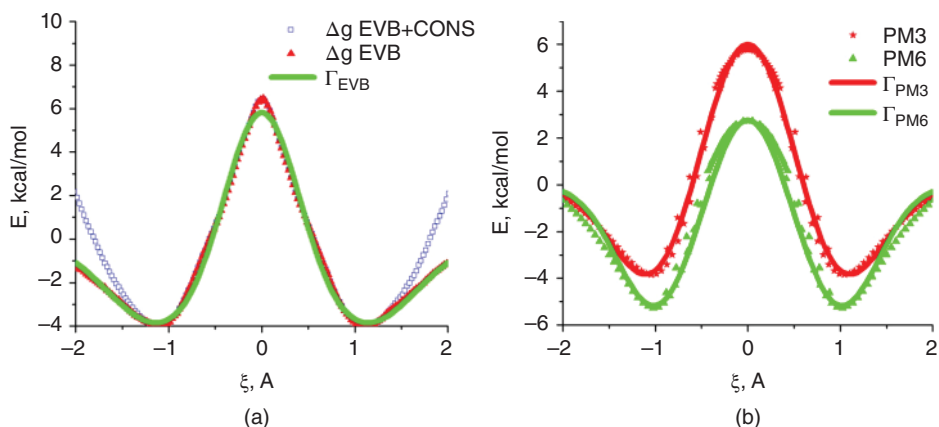


Figure 3.5 Fitting PMFs with Gaussian Functions. (a) Γ_{EVB} (green line) is a fit to the EVB gas-phase PMF (red triangles), which was obtained by subtracting contributions of the EVB distance constraints on the RC values from the actual EVB free energy profile (blue squares). (b) Γ_{PM3} and Γ_{PM6} (green line) are fits of potential energy scans obtained with semiempirical PM3 and PM6 hamiltonians. Reprinted with permission from reference^[25]. Copyright 2012 American Chemical Society.

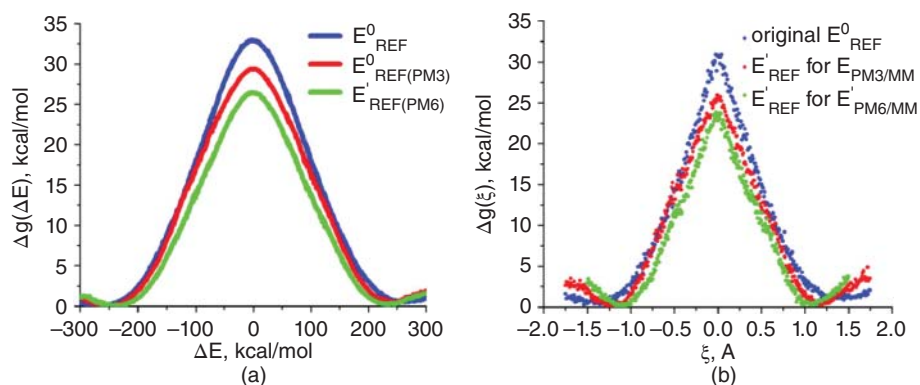


Figure 3.6 Demonstrating the Refinement of the EVB Reference Potential in Condensed Phases, Using the Correction Potential $-\Gamma_{EVB} + \Gamma_{TARG}$ Represented by Gaussians And a Vector of New EVB Charges Derived With the Target Potential. The free energy profiles obtained by the EVB FEP/US approach along the EVB energy gap (a) and along the nuclear RC (b). (blue) the original EVB reference potential with parameters refined for the gas phase PM3; (red) the EVB reference potential refined for PM3/MM target potential; (green) the EVB reference potential refined for PM6/MM target potential. Reprinted with permission from reference^[25]. Copyright 2012 American Chemical Society.

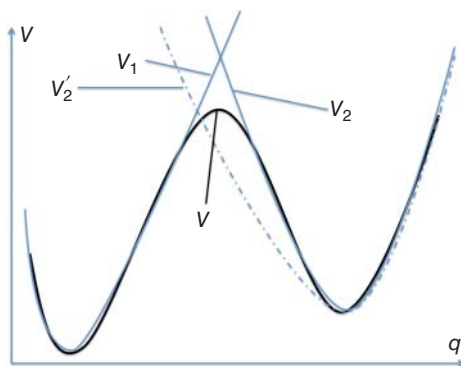


Figure 4.1 Schematic Representation of an EVB Reactive Potential Energy Surface, V , Obtained From Two Diabatic States V_1 and V_2 Corresponding to Reactant and Product States. V is very similar to V_1 for values of the coordinates lying in the reactant region, and to V_2 in the product region. An example of an inappropriate product diabatic state function, V'_2 , that lies lower than the target potential energy for some values of q , is also shown as the dash-dotted line (-----).

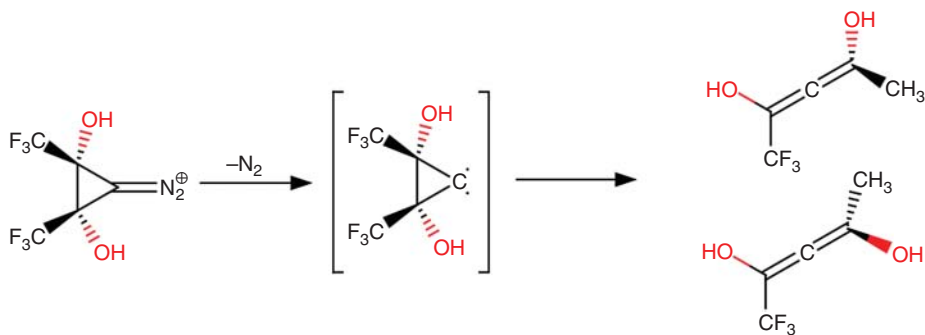


Figure 4.3 Ring-Opening of a Diazocyclopropane, Leading to Enantiomeric Products.

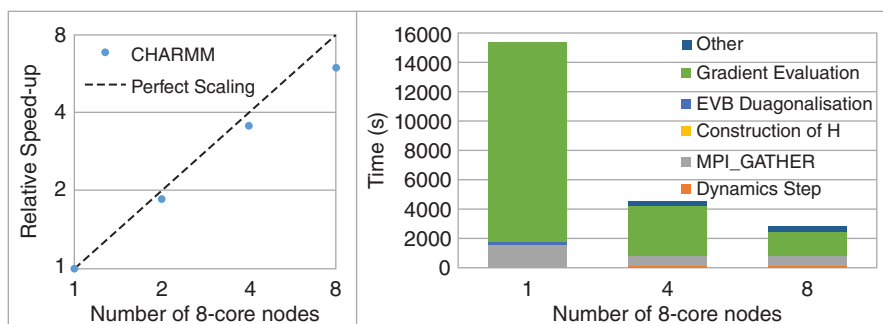


Figure 4.5 Left-Hand Panel Shows the Scaling of the State Parallelized EVB Scheme in Figure 4.4 (Tested For F + CD₃CN) Versus the Number of 8-Core CPU Nodes Utilized, up to a Maximum of 64 CPUs. The right-hand panel shows a decomposition of the relative timings required to carry out various computational tasks during the MS-EVB runs for the F + CD₃CN.

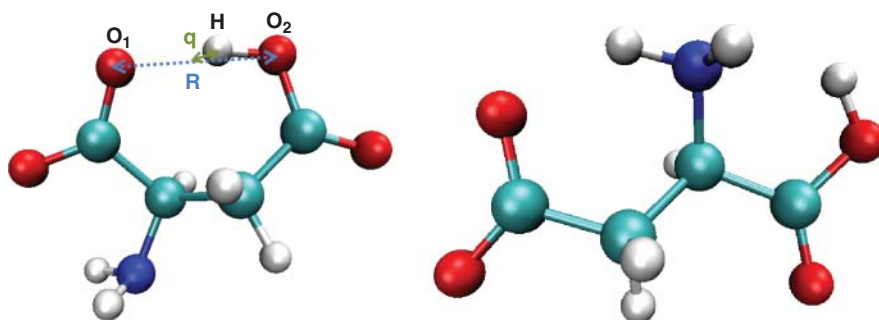


Figure 5.1 Most Stable Structures of Aspartate: Bent Conformation (Left) with Definitions of the R and q Distances, and Linear Conformation (Right).

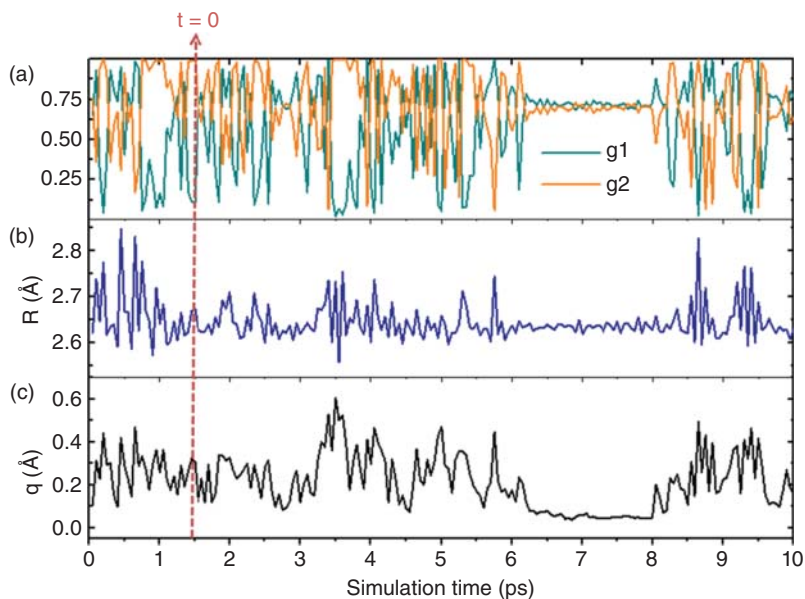


Figure 5.8 10-ps EVB-AMOEBA Simulation at 300 K, (a) Evolution of g_1 (Green), g_2 (Orange), (b) R and q Distances (in Å). Short-time averaging over 50 fs windows has been used to improve visual quality.

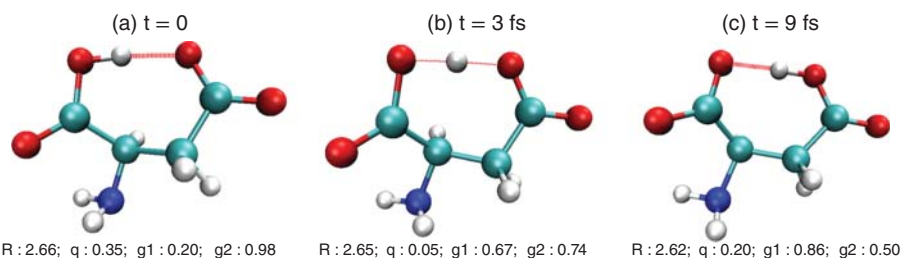


Figure 5.9 Proton Exchange During an EVB-AMOEBA MD Trajectory. The geometrical parameters R and q , the EVB weights g_1 and g_2 and the time frames are indicated.

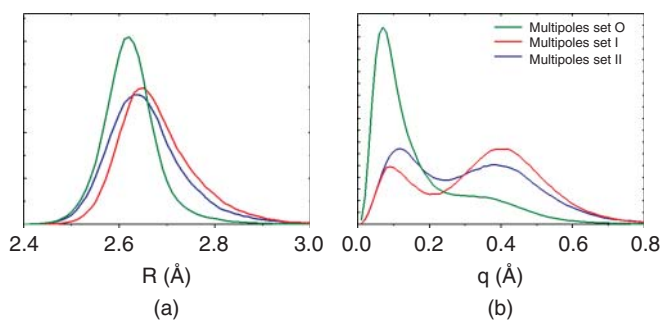


Figure 5.12 Equilibrium Distributions of (a) R and (b) q , as Obtained From the EVB Simulations at 300 K Using the Three Multipolar Sets O, I, and II.

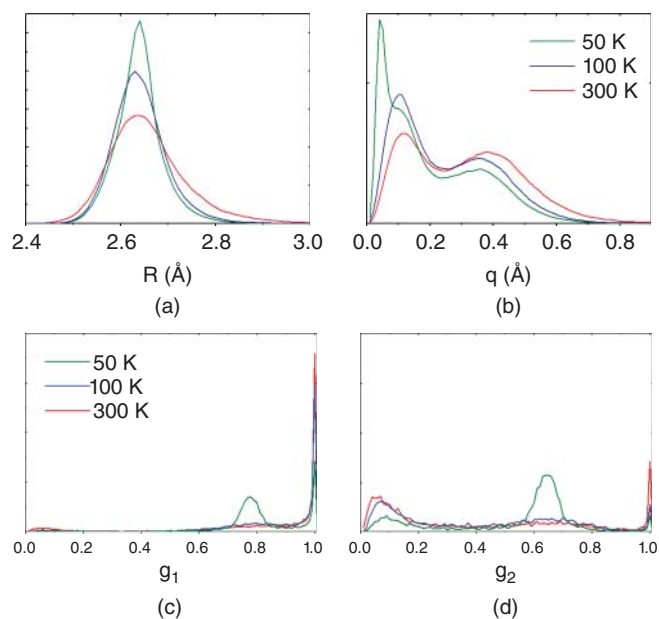
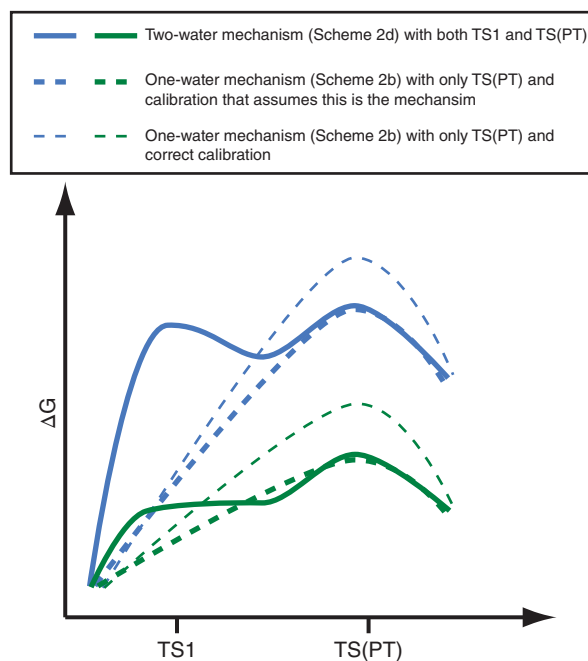
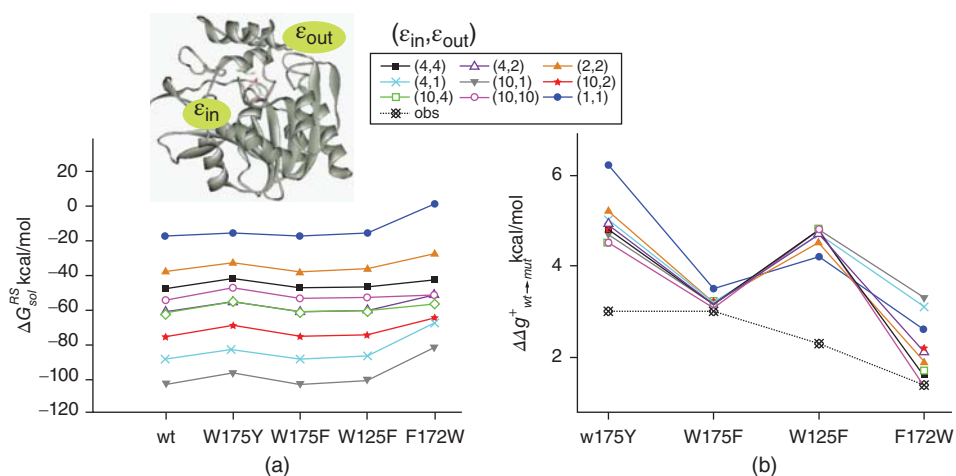


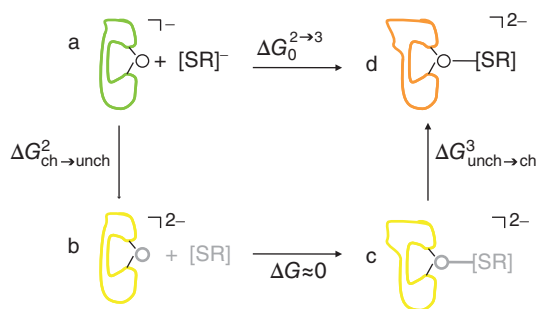
Figure 5.14 Equilibrium Distributions of (a) R and (b) q Obtained at 50, 100 and 300 K, and Distributions of the EVB Weights (c) g_1 and (d) g_2 , as Obtained at 50, 100 and 300 K.



Scheme 6.3 EVB free energy profiles of the first steps in the hydrolysis of GTP in solution and in the active site of RasGAP. Only the steps that lead to a PT to the substrate (GTP) are shown. The profiles of the uncatalyzed reaction and within the RasGAP complex are in blue and green, respectively. Plain lines display the profiles of the mechanism that involves a second water molecule, serving as a proton shuttle. Dashed lines, on the other hand, exhibit the profiles of the mechanism that involved direct PT from the nucleophilic water molecule. Here, bold and plain dashed lines correspond to the deficient (based on the earlier study) and true calibration, respectively. Activation barriers and reaction free energies are given with respect to the energy of their corresponding reactants regardless of the mechanism used.



Scheme 6.4 The effect of the scaling scheme of the electrostatic interactions with the ionizable residues on a. the solvation energy of the reactive fragments, ΔG_{sol}^{RS} and b. the relative catalytic ability of the mutants, $\Delta\Delta g_{wt \rightarrow mut}^+$. Each graph corresponds to a different set of scaling factors, (ϵ_i, ϵ_o) , for the inner and outer ionizable residues, respectively. Dashed line represents the observed experimental estimations.



Scheme 6.6 The two step (simultaneous) thermodynamic cycle used to calculate the free energy difference between the di- and tricoordinated Cu(I), $\Delta G_0^{2 \rightarrow 3}$. SR^- represents the third (external) thiolato ligand. Compounds whose partial charges are set to zero are colored gray. Different colors of the proteins present the different protein configuration ensemble, which is adjusted to the charge distribution.

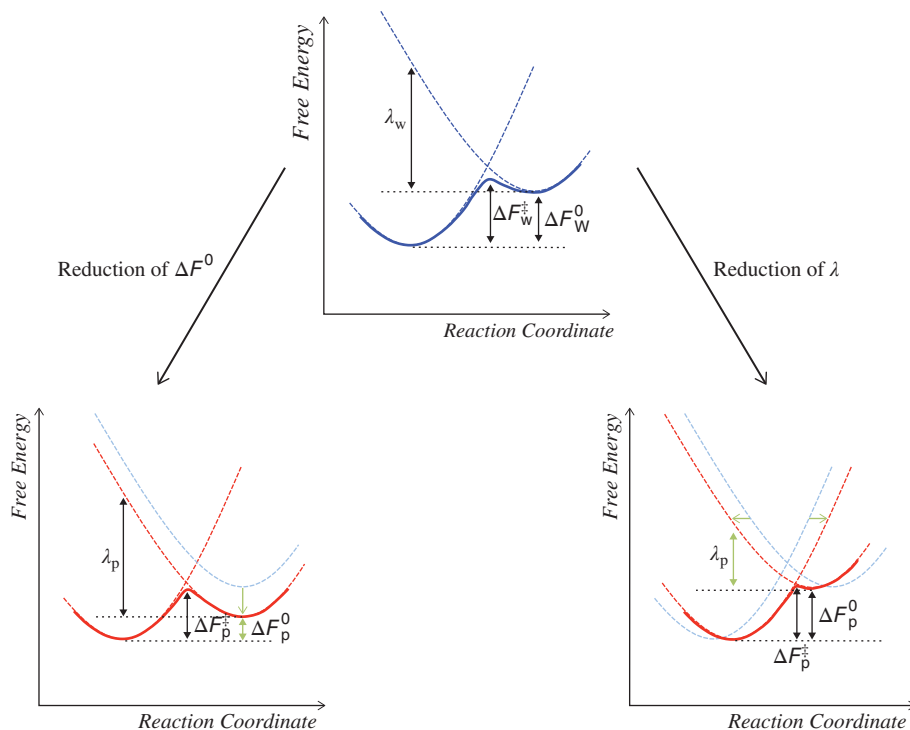


Figure 7.2 Schematic Representation of the Two Extreme Cases of How Enzymes Reduce Catalytic Barriers by Shifting the Relative Position of the Diabatic States Vertically to Reduce ΔF^0 and Horizontally to Reduce λ . Reprinted from reference^[47]. Copyright (2014), with permission from Elsevier.

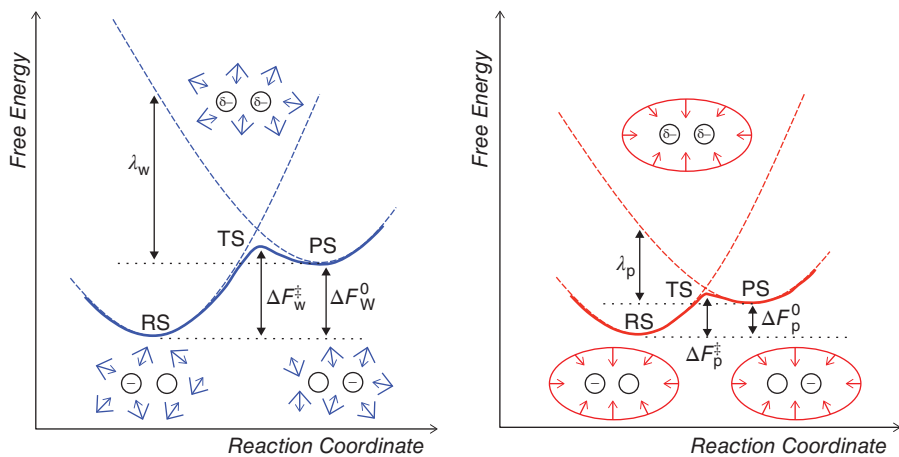


Figure 7.3 Geometrical Representation of the Reorganization Effect in Water Solution (Left) and Enzymatic Environment (Right). Reprinted from reference^[47]. Copyright (2014), with permission from Elsevier.

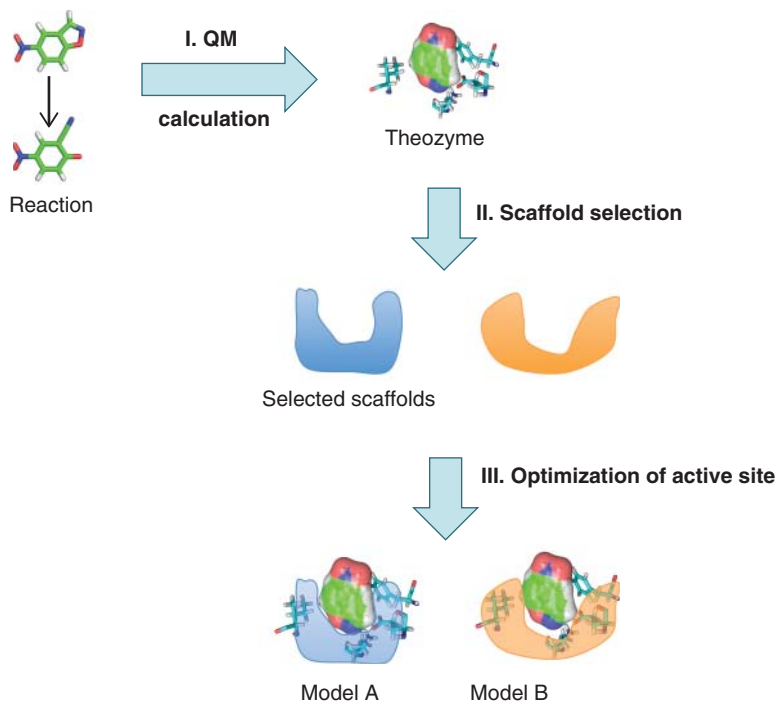


Figure 7.4 Standard Protocol of In Silico Enzyme Design.

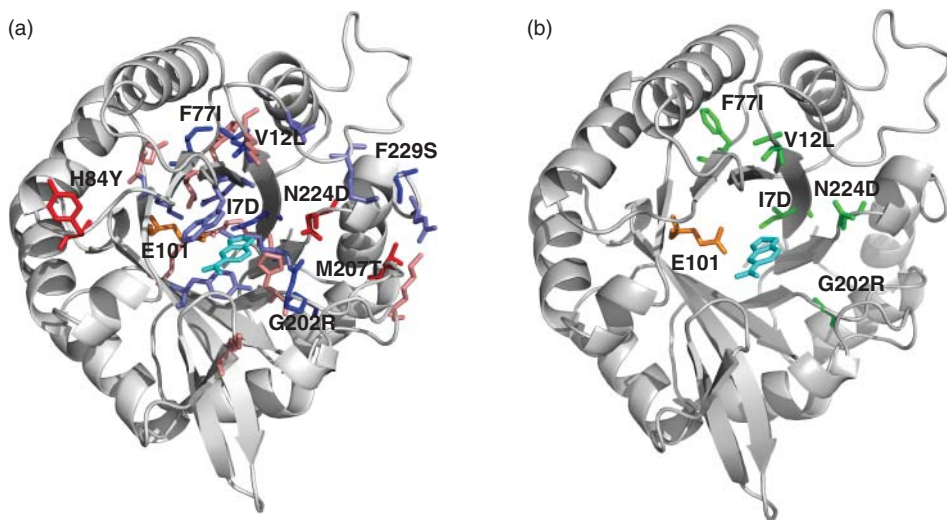


Figure 7.6 (a) The Effect of Directed Evolution on the Electrostatic Contribution of Residues, Which Significantly Contribute to Catalysis (Favorable –Light Blue and Unfavorable –Light Red) and Residues that Were Replaced During the In Vitro Evolution (Favorable –Dark Blue and Unfavorable –Dark Red). (b) Mutations of KE07 design that occurred during directed evolution and were also predicted based on their contributions to the reorganization energy (green). The substrate (cyan) and key residue Glu101 (orange) are also displayed in both cases. Reprinted from reference⁴⁷¹. Copyright 2014, with permission from Elsevier.

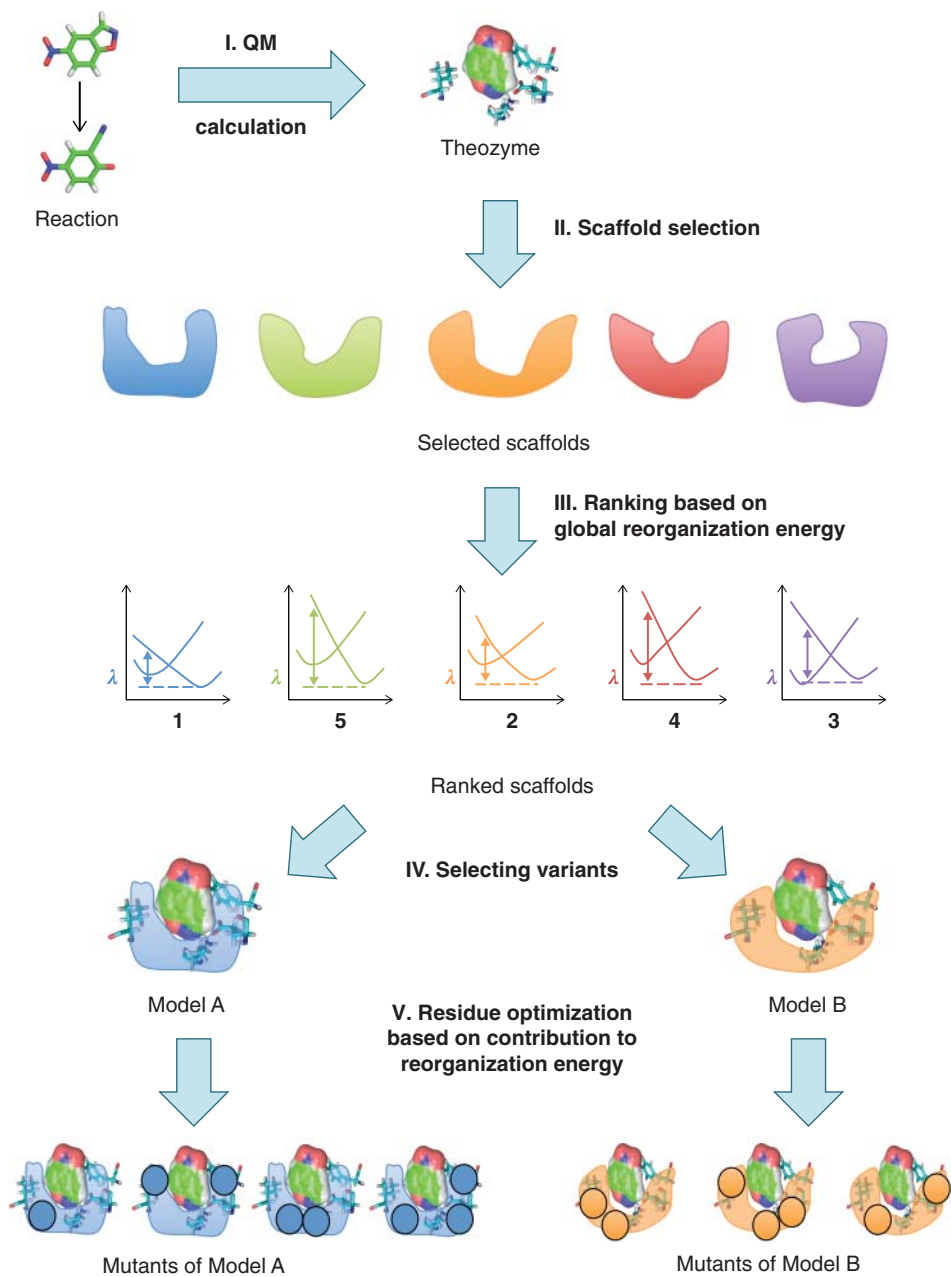


Figure 7.7 Key Steps of the Recently Suggested New Computational Enzyme Design Including the Computation of the Reorganization Energy. Reprinted from reference^[47]. Copyright 2014, with permission from Elsevier.

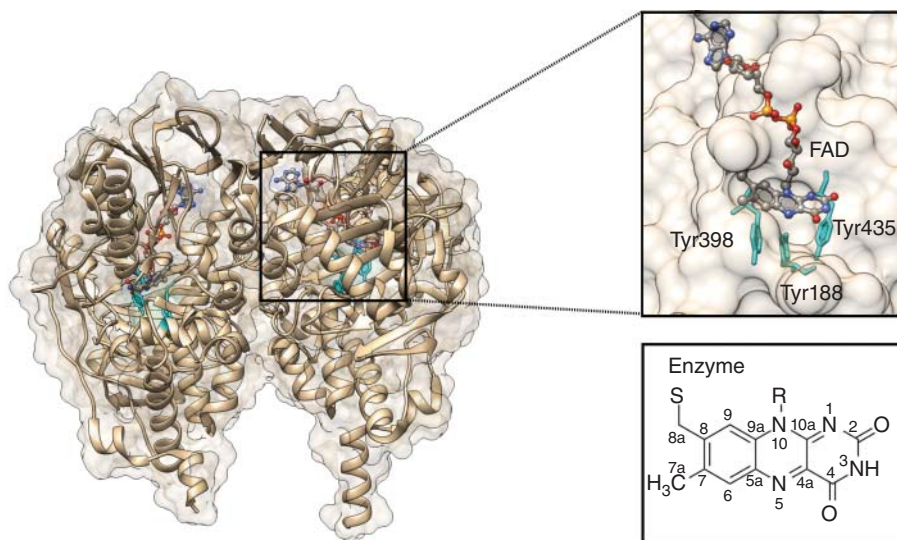


Figure 8.1 Position of the Active Site Within the MAO B Crystal Structure (2XFN.pdb) Indicating the Flavin Cofactor (FAD) and Three Tyrosine Residues that all Form the “Aromatic Cage” Structural Feature. Atom numbering of the FAD fragment is shown in the bottom-right box.

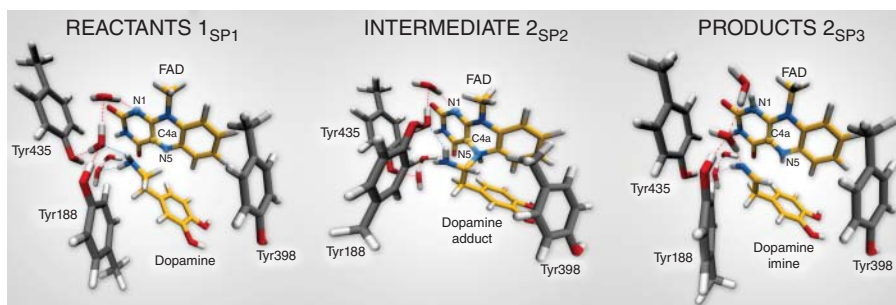


Figure 8.4 Structures of Relevant Stationary Points For the Newly Proposed Two-Step MAO Catalytic Hydride Mechanism For the Degradation of Dopamine 1.

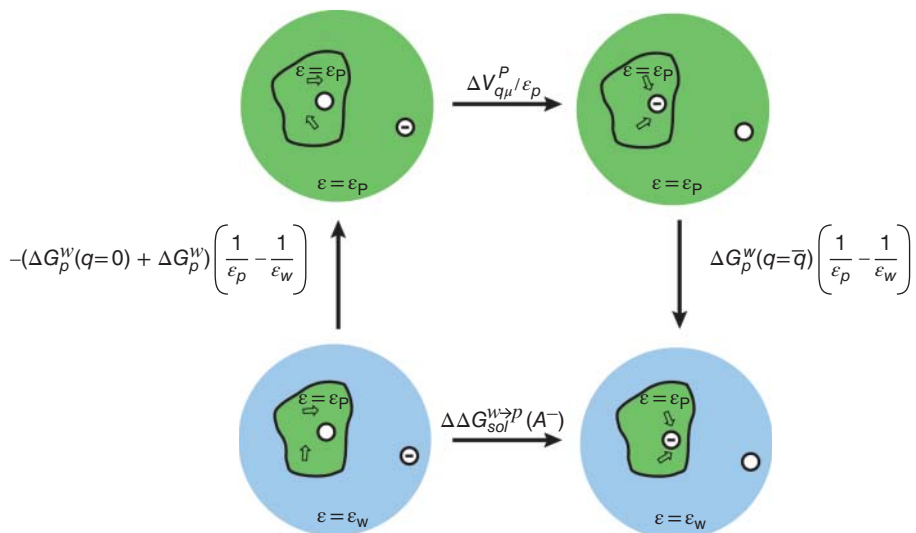


Figure 8.7 The PDL D/S-LRA Thermodynamic Cycle For the Evaluation of $\Delta\Delta G_{sol}^{w \rightarrow p}(A^-)$ Within a Protein (w and p Indicate Water and Protein, Respectively). The cycle involves the change of the dielectric constant of the solvent around the protein from ϵ_w to ϵ_p , moving the charge from the solvent to the protein, changing back to the solvent dielectric and uncharging the ionized group inside the protein. The energy contribution of each step is indicated in the figure (see text). For a detailed discussion see references^[81,82].

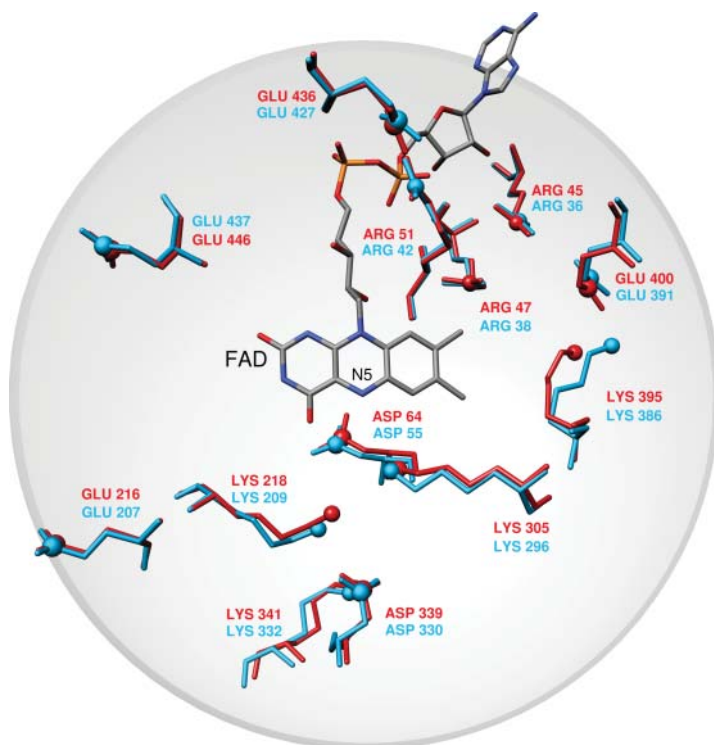


Figure 8.8 Superposition of FAD and Ionizable Residues Surrounding the Active Site (MAO A is Red, MAO B is Blue).

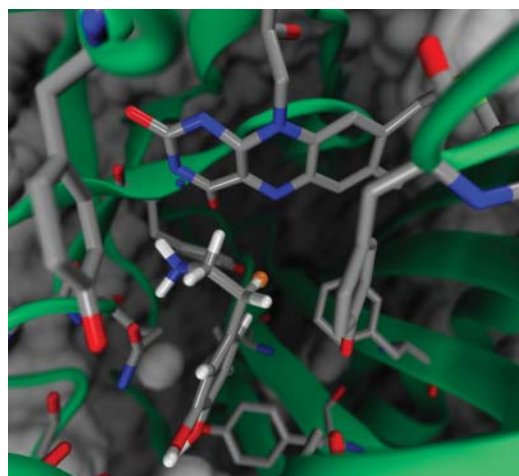


Figure 8.9 Structure of MAO B Active Site with Neutral Dopamine. The active site residues are shown in gray, FAD prosthetic group in orange, dopamine in light blue, and Lys296 in violet. The orientation of the substrate facilitates the positioning of the pro-R hydrogen atom for the hydride transfer.

Index

a

- ab initio valence bond molecular orbital method (VB/MM) 49–50, 75, 161
- acetyl-acetone (AcAc), proton transfer 15–16
- acetylcholinesterase (AChE) 147–148, 179
- activation entropy 174
- activation free energy barrier 76
- active sites of enzymes 173
- adiabatic potential surfaces 35, 40
- adiabatic reactive molecular dynamics (ARMD) 3, 4–6
 - applications
 - myoglobin (Mb) and nitric oxide (NO) rebinding dynamics 16
 - nitric oxide (NO) detoxification in truncated hemoglobin (trHbN) 16–18
 - outlook 18–19
- adipic acid 133
- AMBER software 28, 111
- ammonia dimer 11
- approximate valence bond (AVB) method 3, 47–48
- artificial catalysts designed using EVB 173–174
 - computational analysis of Kemp eliminases 183–184
 - contributions of individual residues 189–190
 - EVB determination of catalytic effects
 - free energy profile computation 185
 - general framework 184–185
 - general reaction coordinate 187–188
 - in silico enzyme design 179–183
 - key steps 192
 - origin of catalytic effect 174–175
 - outlook 191–193
 - reorganization energy 177–179
 - calculation 186–187
 - enzyme design improvement 190–191
- aspartate, IR spectra 123
 - AMOEBA finite temperature force field 126
 - molecular dynamics simulations 126–127
 - multipoles, role of 127–130
 - experimental approach 123–124
 - quantum chemical calculations 124–126
 - stable structures 124
- asymmetric double minimum (ADM)
 - optimized structure 4
- atomic multipole optimized energetics for biomolecular simulation (AMOEBA) 123
 - finite temperature IR spectra
 - force field 126
 - molecular dynamics simulations 126–127
 - multipoles, role of 127–130
- atomic orbitals (AOs) 30

b

- Bennet's acceptance ratio 84–85
- biological systems, EVB simulations 145–147, 166
- comparative tool uses 152

- biological systems, EVB simulations
(*continued*)
 conformational changes, effects of
 155–156
 guided reaction paths 152–154
 mutational studies 156–157
 same reaction in different
 environments 155–157
 insight 164–165
 reaction coordinate definition 163–164
 reaction mechanisms 147
 butyrylcholinestase (BChE) hydrolysis
 147–150
 guanosine diphosphate (GDP)
 hydrolysis 150–152
 sampling tool uses 157
 calculations, large number of
 157–159
 copper-chaperones 159–161
 efficient reference potential 161–163
 efficient sample configurations
 159–163
 hybrid ab initio VB/MM approach
 161
 block localized wavefunction (BLW) 49
 bond energy bond order (BEBO) potential
 9
 Born–Oppenheimer approximation 93
 Born–Oppenheimer molecular dynamics
 (BOMD) 122
 Buckingham potential 79
 butyrylcholinestase (BChE) EVB simulation
 147–150
- C**
Candida antarctica lipase A (CALA)
 43–44
 Car–Parrinello molecular dynamics
 (CPMD) 122
 Car–Parrinello simulation 94
 catalysis
 dynamics 42–43
 electrostatic contributions 40–41
 nuclear quantum effect (NQE)
 contributions 42
 catalytic effect of enzymes 42
 central nervous system (CNS) 222
- Chang–Miller formalism 47
 charge-transfer (CT) reactions
 $N_2-N_2^+$ 12
 CHARMM method 4, 6, 8, 18, 28, 110
 coarse-grained (CG) models 43
 cold-active enzymes 157
 computational analysis of Kemp eliminases
 183–184
 conductor-like polarizable continuum
 model (CPCM) 206
 configuration interaction (CI) 31–32
 conformational changes 155–156
 consistent force field (CFF) 28
 constrained density functional theory
 (CDFT) 36, 39, 86, 186
 contact order (CO) 155
 conventional transition state theory (CTST)
 15
 copper-chaperones 159–161
 correlated modes 175
 Coulomb law 79
 coupled-cluster (CCSD) method 95, 96
 coupling parameter 68
 cyclopropylamine 203
 cytidine deamination 158
- d**
 D-amino acid oxidases (DAAOs)
 202–203
 degree of pyramidalization (DPN) 209
 density function theory (DFT) 95, 96
 enzyme active sites 181–182
 desolvation 175
 diabatic potential surfaces 35, 40
 diatomics-in-molecules (DIM) theory 2
 Diels–Alder reactions 181
 dihydrofolate reductase (DHFR) 43
 dipole moment time autocorrelation
 function (DACF) 126–127
 directed evolution 182
 dynamic effects 175
 dynamics in catalysis 42–43
- e**
 echothiophate 149
 effective fragment potential (EFP) method
 50

- effective Hamiltonian-molecular orbital valence bond method (EH-MOVb) 49
- electronic Schrödinger equation 1–2
- electrostatic contributions to catalysis 40–41
- electrostatic preorganization 175, 181, 183
- elongation factor (EF-Tu) 45–46
- empirical force field methods 1–3
 - applications
 - ammonia dimer 11
 - charge transfer in $N_2-N_2^+$ 12
 - myoglobin (Mb) and nitric oxide (NO) rebinding dynamics 16
 - nitric oxide (NO) detoxification in truncated hemoglobin (trHbN) 16–18
 - protonated water 10–11
 - sulfuric acid, vibrationally induced photodissociation 12–14, 15–16
 - computational approaches 3
 - adiabatic reactive molecular dynamics (ARMD) 4–6
 - empirical valence bond (EVB) method 8–9
 - molecular mechanics with proton transfer (MMPT) 3–4
 - multi-surface adiabatic reactive molecular dynamics (MS-ARMD) 6–8
 - ReaxFF method 9–10
 - other approaches 10
 - outlook 18–19
 - empirical valence bond (EVB) approach 27–28, 32–33
 - artificial catalyst design 173–174
 - computational analysis of Kemp eliminases 183–184
 - contributions of individual residues 189–190
 - EVB determination of catalytic effects 184–185
 - general reaction coordinate 187–188
 - in silico enzyme design 179–183
 - key steps 192
 - origin of catalytic effect 174–175
 - outlook 191–193
 - reorganization energy 177–179
 - reorganization energy calculation 186–187
 - reorganization energy enzyme design improvement 190–191
- biological system simulations 145–147, 166
- butyrylcholinestase (BChE) hydrolysis 147–150
- calculations, large number of 157–159
- comparative tool uses 152–157
- conformational changes, effects of 155–156
- copper-chaperones 159–161
- efficient reference potential 161–163
- efficient sample configurations 159–163
- guanosine diphosphate (GDP) hydrolysis 150–152
- guided reaction paths 152–154
- hybrid ab initio VB/MM approach 161
- insight 164–165
- mutational studies 156–157
- reaction coordinate definition 163–164
- reaction mechanisms 147–152
- same reaction in different environments 155–157
- sampling tool uses 157–163
- free energy evaluation 36–38
- future perspectives 50–52
- historical overview
 - MM to QM/MM approaches 28–29
 - MO vs. VB theory 29–30
- monoamine oxidase (MAO) simulation 199–200
 - cluster model 204–211
 - isoform structures 201–202
 - mechanistic studies 202–204
 - nuclear quantum effects 218–221
 - perspectives 223
 - pharmacology of MAOs 200–201
 - protonation states 211–215
 - rate limiting hydride-abstraction step 215–218

- empirical valence bond (EVB) approach
 - (*continued*)
 - relevance for neurodegeneration 221–223
 - other models 47
 - ab initio valence bond molecular orbital method (VB/MM) 49–50
 - approximate valence bond (AVB) approach 47–48
 - Chang–Miller formalism 47
 - molecular orbital valence bond method (MOVb/MM) 49
 - multiconfiguration molecular mechanics (MCMM) 48–49
 - multiple empirical valence bond (MS-EVB) approach 48
 - summary table 48
 - valence bond with effective fragment potential (VBEFP) 50
 - potential surface construction 33–36
 - proton transfer (PT) modeling 130
 - adiabatic state coupling 131–133
 - dynamics 133–136
 - IR spectra 136–139
 - model implementation 131
 - two-state model 130–131
 - reaction dynamics 93–94
 - Cl + CH₄ gas phase 100–102
 - CN + *c*-C₆H₁₂ solution 102–103
 - CN + tetrahydrofuran (THF) solution 103–104
 - CPU parallelization using MPI 109–111
 - diazocyclopropane ring opening 107–108
 - F + CD₃CN solution 104–107
 - GPU parallelization using MPI 111–115
 - methodology 97–100
 - perspectives 115–117
 - potential energy surfaces (PESs) 94–97
 - recent applications 100–108
 - software aspects 108–115
 - reference potential 65–66, 77–80, 161–163
 - functional refinement 81–82
 - limitations and solutions 86
 - parametric refinement 80–81
 - S_N2 reactions 33–36
 - software available 51
 - successful applications 40
 - dynamics in catalysis 42–43
 - electrostatic contributions to catalysis 40–41
 - enantioselectivity 43–44
 - nuclear quantum effect (NQE) contributions to catalysis 42
 - ribosome chemical selectivity 44–47
 - technical considerations 38
 - accuracy in rate quantity computation 40
 - energy gap reaction coordinate 39–40
 - off-diagonal elements 39
 - surface parameterization reliability 38–39
 - valence bond (VB) theory 30–32
 - empirical valence bond (EVB) method 3, 8–9
 - outlook 18–19
 - enantioselectivity 43–44
 - energy gap 37, 39, 43
 - entropic effect 175, 176
 - enzymes
 - active sites 173
 - structural flexibility 181
 - in silico design 179–183
 - origin of catalytic effect 174–175
 - ENZYMIX force field 214, 216
 - exchange charge distribution 48
 - exponential average 83–84
- f**
- Fermi function 85
 - flavin adenine dinucleotide (FAD) 200, 211
 - cofactor binding site 202
 - force field parallelism 112
 - force fields (FFs) 3
 - free energy change 72
 - free energy evaluation 36–38, 64
 - free energy interpolation 85
 - free-electron laser (FEL) 124

- free-energy perturbation (FEP) approach
37, 70
- Bennet's acceptance ratio 84–85
 - estimation 82–83
 - exponential average 83–84
 - free energy interpolation 85
 - linear response estimation (LRA) 84
 - low-accuracy target free energy surface
72–73
 - reaction coordinate constraints 74–75
- frozen density functional theory (FDFT)
39
- g**
- Gaussian and polynomial functions
(GAPOs) 7–8, 13
- Gaussian functions 99
- generalized VB (GVB) method 32
- Gibbs free energy 36–37, 83
- Gilbert–Meulengracht syndrome 222
- GROMACS software 111
- GROMOS software 28
- ground state destabilization (GSD) 174
- GTPase activating protein (GAP) 150
- guanosine diphosphate (GDP) hydrolysis,
EVB simulation 150–152
- guanosine triphosphate (GTP) 45–46
- gyration, radius of 155
- h**
- harmonic bias 68
- Hartree–Fock theory 95, 96
- Heitler–London (HL) wave function
30–31
- Hellman–Feymann theorem 80, 98, 131
- force vector 109
- Helmholtz free energy 83
- Haemoglobin
- nitric oxide (NO) detoxification in
truncated hemoglobin (trHbN)
16–18
- hybrid atomic orbitals (HAOs) 30
- hybrid parallelism 116
- HYDYN method 122
- i**
- infrared multiple photon dissociation
(IRMPD) 121–123, 140
- aspartate and non-reactive calculations
123
- AMOEBA force field 126
 - experimental approach 123–124
 - molecular dynamics simulations
126–127
 - multipoles, role of 127–130
 - quantum chemical calculations
124–126
- EVB modeling of proton transfer (PT)
130
- diabatic state coupling 131–133
 - dynamics 133–136
 - model implementation 131
 - spectra 136–139
 - two-state model 130–131
- k**
- Kemp elimination 181–182
- computational analysis 183–184
- kernel 112
- ketosteroid isomerase (KSI) 179
- kinetic isotope effect (KIE) 15, 42,
218–220
- l**
- lactate dehydrogenase (LDH) 42, 179
- LAPACK routines 113, 114
- least square penalty function 81
- Lennard–Jones potential 2, 6, 79
- calculation 113
- linear free energy relationship (LFER) 151
- linear response approximation (LRA) 66,
156, 164
- enzyme design improvement 190–191
 - PD model 74
- linear response estimation (LRA) 84
- London–Eyring–Polanyi (LEP) surface 2
- London–Eyring–Polanyi–Sato (LEPS)
surface 2
- low-barrier hydrogen bonds 175
- m**
- malonaldehyde (MA), proton transfer 15
- mapping potential 37, 68
- Marcus equation 178
- Marcus parabolas 78, 84

- mechanics-valence bond (MMVB) method 32
- medium effect 175, 180–181, 183
- message-passing interface (MPI) 109–110
 CPU parallelization 109–111
 GPU parallelization 111–115
- N*-methyltryptophan oxidase (MTOX) 203
- minimum energy paths (MEPs) 13
- modelling chemical reactions 1–3
 applications
 ammonia dimer 11
 charge transfer in $N_2-N_2^+$ 12
 myoglobin (Mb) and nitric oxide (NO)
 rebinding dynamics 16
 nitric oxide (NO) detoxification in
 truncated hemoglobin (trHbN)
 16–18
 proton transfer in malonaldehyde and
 acetyl-acetone 15–16
 protonated water 10–11
 sulfuric acid, vibrationally induced
 photodissociation 12–14
- computational approaches 3
 adiabatic reactive molecular dynamics
 (ARMD) 4–6
 empirical valence bond (EVB) method
 8–9
 molecular mechanics with proton
 transfer (MMPT) 3–4
 multi-surface adiabatic reactive
 molecular dynamics (MS-ARMD)
 6–8
 ReaxFF method 9–10
 other approaches 10
 outlook 18–19
- MOLARIS program package 212, 214, 216
- Molecular Dynamics with Quantum
 Transitions (MDQT) 42
- molecular mechanical (MM) approach 27
- molecular mechanics with proton transfer
 (MMPT) 3–4
 protonated water 10–11
- molecular orbital (MO) approach 28,
 31–32
 historical overview 29–30
- molecular orbital theory 29
- molecular orbital valence bond method
 (MOVB/MM) 49
- Møllet–Plesset perturbation (MP2) method
 95
- monoamine oxidases (MAOs), EVB
 simulation 199–200
 cluster model 204–211
 isoform structures 201–202
 mechanistic studies 202–204
 nuclear quantum effects 218–221
 perspectives 223
 pharmacology of MAOs 200–201
 protonation states 211–215
 rate limiting hydride-abstraction step
 215–218
 relevance for neurodegeneration
 221–223
- More O’Ferrall–Jencks plot 71
- Morse potentials 6, 79
- multiconfigurational molecular mechanics
 (MCMM) 10, 48–49
- multi-configuration molecular mechanics
 (MSMM) procedure 96–97
- multiconfiguration time-dependent Hartree
 (MCTDH) approach 122
- multi-layer approaches 27–28
- multiple empirical valence bond (MS-EVB)
 approach 48
- multi-surface adiabatic reactive molecular
 dynamics (MS-ARMD) 3, 6–8
 applications
 acetyl-acetone (AcAc), proton transfer
 15–16
 charge transfer in $N_2-N_2^+$ 12
 malonaldehyde (MA), proton transfer
 15
 sulfuric acid, vibrationally induced
 photodissociation 12–14
 outlook 18–19
- mutational studies 156–157
- myoglobin (Mb) and nitric oxide (NO)
 rebinding dynamics 16
- n**
- NAMD software 111
- near attack conformation (NAC) 175, 176,
 181

- neurodegeneration 221–223
 Newton–Raphson approach 81
 non-steroidal anti-inflammatory drugs (NSAIDs) 222
 nuclear quantum effect (NQE)
 contributions to catalysis 42
- O**
 Open Computing Language (OpenCL)
 112, 113
 OPLS software 28
 orbital steering 175, 176
 organophosphate esters (OPs) 147–148
- P**
 packing 181
 paradynamics (PD) approach 63, 86–87
 challenges 69
 convergence of free energy estimates 70
 different reference and target reaction paths 69–70
 concept 68–69
 context 64–67
 EVB as reference potential 77–80
 functional refinement 81–82
 limitations and solutions 86
 parametric refinement 80–81
 FEP estimation 82–83
 Bennet’s acceptance ratio 84–85
 exponential average 83–84
 free energy interpolation 85
 linear response estimation (LRA) 84
 reference potential methods 71
 high-accuracy local PMF regions 76–77
 improving accuracy in local PMF regions 77
 low-accuracy target free energy surface from free energy perturbation 72–73
 low-accuracy target free energy surface from non-equilibrium distribution 71–72
 PD model 74
 pre-computing reaction path 73
 reaction coordinate constraints 74–76
- Pariser–Parr–Pople (PPP) method 28
 Potential of Mean Force (PMF) 63, 66–67, 69, 82, 87
 high-accuracy in local regions 76
 improving accuracy in local regions 77
 polarizable force fields for IR spectroscopy 121–123, 140
 aspartate and non-reactive calculations 123
 AMOEBA force field 126
 experimental approach 123–124
 molecular dynamics simulations 126–127
 multipoles, role of 127–130
 quantum chemical calculations 124–126
 EVB modeling of proton transfer (PT) 130
 diabatic state coupling 131–133
 dynamics 133–136
 model implementation 131
 spectra 136–139
 two-state model 130–131
 potato epoxide hydrolase 1 (StEH1) 44
 potential energy surfaces (PESs) 4, 5–8, 47, 64
 reaction dynamics 94–97
 propagation algorithm 109
 proton transfer (PT) 121, 122, 153
 EVB modeling 130
 diabatic state coupling 131–133
 dynamics 133–136
 IR spectra 136–139
 model implementation 131
 two-state model 130–131
 protonated water 10–11
 proximity effect 174, 175
 pseudo-Hamiltonian matrix 97–98, 99
- Q**
 QM(VB)/MM method 32
 Quantum Classical Path (QCP) 42, 220–221
 quantum mechanical (QM) approach 27
 quantum mechanical/classical mechanics (QM/MM) methods, mixed 2–3, 28

quantum mechanical/classical mechanics
 (QM/MM) methods, mixed
 (*continued*)
 computational methods 64–65
 historical overview 28–29
 quantum tunneling 175, 218–219
 quenching dynamics 10
 QUEST program 29

r

radius of gyration 155
 rate constants 94
 Reactant State (RS) 42, 63, 65, 72
 PD model 74
 reactive oxygen species (ROS) 221
 reaction coordinate (RC) definition
 163–164
 reaction dynamics, EVB methods 93–94
 methodology 97–100
 perspectives 115–117
 potential energy surfaces (PESs) 94–97
 recent applications 100
 Cl + CH₄ gas phase 100–102
 CN + *c*-C₆H₁₂ solution 102–103
 CN + tetrahydrofuran (THF) solution
 103–104
 diazocyclopropane ring opening
 107–108
 F + CD₃CN solution 104–107
 software aspects 108–109
 CPU parallelization using MPI
 109–111
 GPU parallelization using MPI
 111–115
 reactive molecular dynamics (RMD)
 method 3
 ReaxFF method 9–10, 123
 reorganization energy 43, 177–179
 calculation 186–187
 enzyme design improvement 190–191
 ribosome chemical selectivity 44–47

s

S_N2 reactions
 empirical valence bond (EVB) approach
 33–36

scaffolds 182
 screening function 79
 semiempirical methods 95
 Shepard points 48–49
 single-instruction-multiple-data (SIMD)
 stream 109
 singlet states 206
 specific reaction parameter (SRP) method
 95
 spin-coupled VB (SCVB) approach 32
 staphylococcal nuclease (SNase) 40–41
 state parallelism 112, 116
 steric strain 175
 sulfuric acid, vibrationally induced
 photodissociation 12–14
 surface-constrained all-atom solvent
 (SCAAS) 215
 symmetric double minimum (SDM)
 optimized structure 4
 ammonia dimer 11
 symmetric single minimum (SSM)
 optimized structure 4

t

theozyme reaction 180
Thermoascus aurantiacus (TAX) 183
 Thermodynamic Integration (TDI)
 approach 84
 TINKER 110
 Transition State (TS) 42, 63, 65, 70, 72
 PD model 74
 transition state stabilization (TSS) 174
 transition state theory (TST) 173
 trypsin 157–158
 tryptophan-2-monooxygenase (TMO)
 202–203
 two-dimensional parallelism 113

u

umbrella sampling formula 68

v

valence bond (VB) theory 30–32
 historical overview 29–30
 valence bond self-consistent field (VBSCF)
 method 32

valence bond with effective fragment
potential (VBEFP) method
50
Variational Transition State Theory
(VTST) 42
VB polarizable continuum model
(VBPCM) 32
VB solvation model (VBSM) 32

W

water, protonated 10–11
wild-type (WT) enzymes 145
butyrylcholinestase (BChE) 148
work groups 112

Z

zero-point energy (ZPE) 40