150
Structure and Bonding

For further volumes:
http://www.springer.com/series/430

Aims and Scope

The series *Structure and Bonding* publishes critical reviews on topics of research concerned with chemical structure and bonding. The scope of the series spans the entire Periodic Table and addresses structure and bonding issues associated with all of the elements. It also focuses attention on new and developing areas of modern structural and theoretical chemistry such as nanostructures, molecular electronics, designed molecular solids, surfaces, metal clusters and supramolecular structures. Physical and spectroscopic techniques used to determine, examine and model structures fall within the purview of *Structure and Bonding* to the extent that the focus is on the scientific results obtained and not on specialist information concerning the techniques themselves. Issues associated with the development of bonding models and generalizations that illuminate the reactivity pathways and rates of chemical processes are also relevant.

The individual volumes in the series are thematic. The goal of each volume is to give the reader, whether at a university or in industry, a comprehensive overview of an area where new insights are emerging that are of interest to a larger scientific audience. Thus each review within the volume critically surveys one aspect of that topic and places it within the context of the volume as a whole. The most significant developments of the last 5 to 10 years should be presented using selected examples to illustrate the principles discussed. A description of the physical basis of the experimental techniques that have been used to provide the primary data may also be appropriate, if it has not been covered in detail elsewhere. The coverage need not be exhaustive in data, but should rather be conceptual, concentrating on the new principles being developed that will allow the reader, who is not a specialist in the area covered, to understand the data presented. Discussion of possible future research directions in the area is welcomed.

Review articles for the individual volumes are invited by the volume editors.

In references *Structure and Bonding* is abbreviated *Struct Bond* and is cited as a journal.

Mihai V. Putz • D. Michael P. Mingos
Editors

Applications of Density Functional Theory to Biological and Bioinorganic Chemistry

With contributions by
M. Causá • P.K. Chattaraj • A. Chakraborty • M. D'Amore • A. Goursot • C. Garzillo • F. Gentile • E.S. Kryachko • A. de la Lande • S. Pan • A.M. Putz • M.V. Putz • R. Silaghi-Dumitrescu • D.R. Salahub • A. Savin • R. Zhang • Y. Zhang

Springer

Editors
Mihai V. Putz
Structural and Computational
Physical-Chemistry Laboratory
West University of Timisoara
Timişoara
Romania

D. Michael P. Mingos
Inorganic Chemistry Laboratory
University of Oxford
Oxford
United Kingdom

ISSN 0081-5993 ISSN 1616-8550 (electronic)
ISBN 978-3-642-32749-0 ISBN 978-3-642-32750-6 (eBook)
DOI 10.1007/978-3-642-32750-6
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012955471

Printed on acid-free paper

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

Preface

In the early twentieth century following the elucidation of the structure of atoms it became evident that atoms and molecules with even numbers of electrons were far more numerous than those with odd numbers of electrons. In 1916, G. N. Lewis provided the first comprehensive description of ionic and covalent bonds, when he postulated that atoms tend to hold an even number of electrons in their outer shells and a special stability was associated with eight valence electrons, which he speculated were arranged symmetrically at the eight corners of a cube. In 1919, I. Langmuir suggested that the structure of the periodic table could be rationalized using an extension of Lewis' postulates. In 1922, N. Bohr updated his model of the atom by assuming that certain numbers of electrons (for example 2, 8, and 18) corresponded to stable "closed shells." In 1926, Schrödinger established a wave mechanical description of the hydrogen atom which was subsequently extended to polyelectron atoms. Pauli was the first to realize that the complicated numbers of electrons in closed shells can be reduced to the simple rule of *one* per state, if the electron states are defined using four quantum numbers. For this purpose he introduced a new two-valued quantum number, identified by Goudsmit and Uhlenbeck as electron spin. The resulting Pauli Exclusion Principle states that no two electrons in a single atom can have the same four quantum numbers; if n, l, and m_l are the same, m_s must be different such that the electrons have opposite spins.

The idea of shared electron pairs introduced by Lewis provided an effective qualitative picture of covalent bonding and it still forms the basis of the universal notation for chemical communication, but it was Heitler and London who in 1927 developed the first successful quantum mechanical expression for this bonding model. Initially they provided a description of the bonding in molecular hydrogen, but it was subsequently adapted to more complex molecules and its widespread applications were articulated with great conviction by Linus Pauling. An alternative molecular orbital description of chemical bonding originated from Burrau's description of the hydrogen molecule ion and this model was subsequently widely developed by Mulliken and Lennard-Jones. The electrons occupy molecular orbitals which are delocalized over the whole molecule and were filled according to the Aufbau Principle and assigned quantum numbers according to the Pauli

Exclusion Principle. The orbitals are calculated in a self-consistent fashion in a manner analogous to those developed previously for atomic orbitals and are based on linear combination of the atomic orbitals of the individual atoms. The number of molecular orbitals equals the number of atomic orbitals in the atoms being combined to form the molecule. A molecular orbital describes the behavior of one electron in the electric field generated by the nuclei and some average distribution of the other electrons. This approximation proved to be more amenable to computer programming than the valence bond model and was widely developed and used in increasingly less approximate forms from 1960 to 1990.

In the early 1970s, a new electronic structure approach emerged from the physics community and was described as density functional theory (DFT). The total energy of a molecule was expressed as a functional of the total electron density. Hohenburg and Kohn proved the unique relationship between electron density and energy and Kohn and Sham put forward a practical variational DFT approach. Although calculations in solid-state physics had been reported since the 1970s DFT was not considered accurate enough for calculations in quantum chemistry until the 1990s, when the approximations used in the theory were refined to more accurately describe the exchange and correlation interactions. Computational costs for ab initio DFT calculations are relatively low when compared to the valence bond and molecular orbital methods. DFT thus began to approach the goals of computational thermochemistry to calculate the energetic properties of chemical processes to an accuracy of 1 kcal mol^{-1}. The widespread acceptance of these methodologies by the chemical community led to Kohn and Pople sharing the Nobel Prize in Chemistry in 1998.

When in 2004 Volumes 112 and 113 of Structure and Bonding were devoted to the "Principles and Applications of Density Functional Theory in Inorganic Chemistry" the editors N. Kaltsoyanis and J.E. McGardy noted "It is difficult to overestimate the impact that Density Functional Theory has had on computational quantum chemistry over the last two decades. Indeed, this period has seen it grow from little more than a theoretical curiosity to become a central tool in the computational chemist's armory." In these volumes they described recent applications in inorganic and biochemistry and addressed key issues in spectroscopy, mechanistic studies, and magnetism.

As possibly the dominant discipline of the twenty-first century the biological sciences have assimilated analytical, conceptual, and computational techniques from the other natural sciences. The continuing need for interpreting the vast amount of new data from in vivo and in vitro experiments using causal and deterministic hypothesis requires a wide range of statistical and computational tools and algorithms. As a consequence bioinformatics and mathematical, physical, and chemical biology have flourished and been used to interpret complex natural biological phenomena and pharmaceutical/toxicological effects of chemicals to natural systems.

The universal implications of chemical interactions and more specifically the structure and bonding characteristics of biomolecules suggest that DFT may also play a crucial role *in cerebro* and *in silico* experiments. Establishing the molecular

basis of biological principles by means of quantum mechanical tools has become a realistic possibility given the current accuracy of DFT methods. The present volume opens with an authoritative review of the extensions of DFT (dispersion-corrected functionals, Born–Oppenheimer dynamics, hybrid with molecular mechanics, constrained, and interpretational) from chemical reactions to biochemical systems (containing over a hundred atoms, enzyme kinetics, etc.). The dispersion problem and the development of dispersion-corrected DFT, which may be used accurately to describe weakly bonded biological systems, are further formalized by specific density functional features in the second chapter. Computational models of DFT are used in the next chapter to exemplify the theoretical counterparts of the spectroscopic data to define the binding and activation energies of small molecules with high bioinorganic implications such as water, congeners of molecular oxygen, nitrogen oxides and oxyanions, sulfide, sulfur oxides and oxyanions, carbon dioxide, organic compounds, halogens, molecular hydrogen, and protons. The computational DFT approach as applied to the electronic localization functions and maximum probability domain analyses for modeling metal–porphyrins. These results suggest that the bonding is primarily ionic in porphyrins containing transition and non-transition metals. The last two chapters deal with the important problem of modeling toxicity phenomena using reactivity principles derived from DFT calculations. After introducing the connection between chemical structure and biological information by connecting the chemical reactivity with biological activity within the quantitative structure–activity relationship (QSAR) technique, the possible anticancer activity of two new metal–borane clusters is explored. It is further generalized by the last chapter which describes the full merging of the QSAR with logistic enzyme kinetics. This leads to a description of the mechanisms of chemical–biological interactions in chlorinated-PAHs by means of chemical reactivity principles derived from conceptual DFT.

Overall the volume provides a coherent exposition of the application of DFT to various biological and bioinorganic chemical systems. We hope that it will encourage the DFT community in further refining and extending the electronic models to complex and correlated biological–chemical systems and interactions in the years to come.

We thank the contributors to this volume for the consistent efforts they have made in writing high-class scientific reviews and for providing the readers with a broad perspective which has revealed the widespread uses of DFT in interpreting biological and bioinorganic systems. MVP acknowledges the research and editing facilities provided for the present volume by the Romanian Education and Research Ministry within the project CNCS-UEFISCDI-TE-16/2010-2013. MVP and DMPM sincerely thank the Springer team and in particular Marion Hertel, Ursula Gramm, Elizabeth Hawkins, and Tanja Jaeger for professionally supervising the production of the **Structure and Bonding** series in general and of this volume in particular.

Timişoara, Romania — Mihai V. Putz
Oxford, UK — D. Michael P. Mingos

Contents

List of Contributors

Mauro Causà Dipartimento di Chimica, Università di Napoli Federico II, Napoli, Italy

Arindam Chakraborty Department of Chemistry, Centre for Theoretical Studies, Indian Institute of Technology, Kharagpur, India

Pratim K. Chattaraj Department of Chemistry, Centre for Theoretical Studies, Indian Institute of Technology, Kharagpur, India

Maddalena D'Amore Dipartimento di Chimica, Università di Napoli Federico II, Napoli, Italy

Aurélien de la Lande University of Calgary, Calgary, AB, Canada

Carmine Garzillo Dipartimento di Chimica, Università di Napoli Federico II, Napoli, Italy

Francesco Gentile Dipartimento di Chimica, Università di Napoli Federico II, Napoli, Italy

Annick Goursot University of Calgary, Calgary, AB, Canada

Eugene Serge Kryachko Bogolyubov Institute for Theoretical Physics of the National Academy of Sciences of Ukraine, Kiev, Ukraine

Sudip Pan Department of Chemistry, Centre for Theoretical Studies, Indian Institute of Technology, Kharagpur, India

Ana-Maria Putz Institute of Chemistry Timişoara of the Romanian Academy, 24 Mihai Viteazul Bld., Timişoara 300223, Romania

Mihai V. Putz Department of Biology-Chemistry, West University of Timişoara, Timişoara, Romania

Dennis R. Salahub University of Calgary, Calgary, AB, Canada

Andreas Savin Dipartimento di Chimica, Università di Napoli Federico II, Napoli, Italy

Radu Silaghi-Dumitrescu Department of Chemistry, Babes-Bolyai University Romania, Cluj-Napoca, Romania

Rui Zhang University of Calgary, Calgary, AB, Canada

Yue Zhang University of Calgary, Calgary, AB, Canada

Struct Bond (2013) 150: 1–64
DOI: 10.1007/978-3-642-32750-6_1

Recent Progress in Density Functional Methodology for Biomolecular Modeling

Dennis R. Salahub, Aurélien de la Lande, Annick Goursot, Rui Zhang, and Yue Zhang

Abstract Density Functional Theory (DFT) has become the workhorse of applied computational chemistry. DFT has grown in a number of different directions depending on the applications concerned. In this chapter, we provide a broad review of a number of DFT and DFT-based methods, having in mind the accurate description of biological systems and processes. These range from pure "cluster" DFT studies of the structure, properties, and reactions of biochemical species (such as enzymatic catalysts) using either straight DFT or dispersion-corrected functionals (DFT-D), to Born–Oppenheimer-DFT dynamics of systems containing up to a hundred atoms or more (such as glycero-lipids), to hybrid DFT/Molecular Mechanical Molecular Dynamics methods which include protein and solvent environments (for enzymes or ion channels, for example), to constrained-DFT (working within the Marcus framework for electron-transfer reactions), to Interpretational-DFT (which provides the interpretational benefits of the Kohn–Sham DFT methodology).

D.R. Salahub (✉) • R. Zhang
Department of Chemistry, Institute for Biocomplexity and Informatics, University of Calgary, 2500 University Drive N.W, Calgary, AB, Canada T2N 1N4,
e-mail: dennis.salahub@ucalgary.ca; zhar@ucalgary.ca

A. de la Lande
Laboratoire de Chimie Physique, UMR 8000, CNRS. Université Paris Sud, Bâtiment 349. 15, avenue Jean Perrin, 91405 Orsay, Cedex, France
e-mail: aurelien.de-la-lande@u-psud.fr

A. Goursot
ICGM, UMR 5253 CNRS, Ecole de Chimie de Montpellier, 34296 Montpellier, France
e-mail: annick.goursot@gmail.com

Y. Zhang
Key Laboratory for Macromolecular Science of Shaanxi Province, School of Chemistry & Materials, Shaanxi Normal University, Xi'an, Shaanxi, China 710062,
e-mail: zhangyue_david@hotmail.com

Keywords Biomolecular modeling • Born–Oppenheimer molecular dynamics • Constrained DFT • Density Functional Theory • Dispersion-corrected DFT • Interpretational DFT • QM/MM methodology

Contents

1 Introduction

Density Functional Theory (DFT) has become the workhorse of applied computational chemistry because of its particularly appealing combination of accuracy, speed, and interpretability. Nearly five decades have transpired since the seminal papers of Hohenberg and Kohn [1] and Kohn and Sham [2] and nearly nine since the first use of a density functional in atomic theory by Thomas [3] and by Fermi [4]. Each decade has seen steady, inexorable, progress—more accurate functionals have been developed, better and faster algorithms have been implemented, and new analytical concepts have been devised (see e.g., [5] for a review covering the 1964–2004 period). The new methodologies and techniques have allowed systems of ever growing complexity to be addressed, to the point where, now, DFT is starting to have a real impact on biological questions.

In this review, we will focus on biology and try to capture the state of the art in studies that approach aspects of biological systems and processes from various points of view, all of them involving DFT. We will exclude from our scope the semiempirical Tight-Binding DFT (DFTB) and also Time-Dependent DFT (TDDFT). Although these methods have been used to study systems of biological interest, we choose to discuss here methods and applications that involve "real" ground-state DFT, including hybrid functionals, at the core.

We have chosen a methodological hierarchy that extends from "simple" molecules and cluster models in which the usual tools of quantum chemistry, (geometry optimization, transition state localization, reaction path following), are applicable, to dynamical studies within the Born–Oppenheimer approximation, to systems of greater extent and complexity in which part of the system is treated with a molecular-mechanical force field (the so-called QM/MM approaches) and, finally, to systems involving long-range charge transfer for which the newly developed constrained DFT has considerable merit. Calculations at any level of the hierarchy require an interpretational framework and a section of the chapter is devoted to some of the concepts involved, most notably various population analyses and the Electron Localization Function (ELF).

We have chosen this particular cross section of DFT-centric subjects because of our own contributions to the methodologies and their applications but we will attempt to put our own work in the context of other contributions by reviewing some of the main contributions over the last few years. Our review will not be exhaustive, and we apologize for the omission of any particular works that undoubtedly may be as pertinent as those we have chosen. Our goal is to paint as faithful a picture as possible of the state of the art through inclusion of a representative sampling rather than an exhaustive review. We do, however, aim to give a comprehensive account of the main issues involved in choosing a computational protocol for the types of biological models we have treated, including the strengths and the limitations of the various methodologies.

To the novice (or even expert) biomolecular modeler the choice of a problem and of the appropriate methodology to address it can often be a daunting challenge. In order to put the particular biological processes we have chosen for this review, along with the six types of methodology we describe, within a common organizational framework we offer the following protocol (enzymatic catalysis has been chosen as an exemplar but we think the protocol also applies to other problems (ion channels, lipid conformations, peptide agglomeration, electron transfer, etc., with some changes)).

1. Choose a relevant biological process and system. What are the biological/biochemical/biophysical questions that are to be addressed? This choice will of necessity involve only a small part of biological reality. We are in reductionist mode here; the question of integration into a systems framework will, for the most part, be left to other publications.
2. Choose an appropriate active component of the biological system (in our exemplar, an enzyme active site).
3. Consider a proposed mechanism and build a model, using crystallographic data if available, from the active site outwards. Choose important protein residues, substrates, intermediates and products, water molecules involved in the mechanism, etc., knowing that the validity of the results will depend critically on these choices.
4. In the case of a "simple" cluster model, decide whether to fix any of the peripheral atoms at their crystallographic positions.

5. Decide on the charge state(s) of acidic residues by calculating pK_a's. If more than one charge state (protonated or unprotonated) seems possible, extend the study to examine both possibilities.
6. Also in the case of a cluster model, decide whether to include the effects of the surrounding protein and solvent through the use of a Polarizable Continuum Model (PCM). Choose the value of the dielectric constant (often chosen to be 4.0).
7. Since in this review we are considering only DFT, decide on the variant of DFT to use (mainly GGA or hybrid functionals, with dispersion corrections for highest accuracy) along with the computational parameters (basis sets, auxiliary basis sets, integration grids, SCF and geometry convergence criteria, etc.). Choose a software package (or write one. . .).
8. Design the study to check that the cluster model is large enough either through cluster-convergence tests or, if that is not possible, through well-reasoned choices of the residues to include, ideally calibrated against experimental data.
9. Decide whether the methodology is accurate enough to allow the use of calculated energies for all steps of the mechanism or if the use of some empirical data is preferred/necessary.
10. If a simple cluster model is thought or shown to be inadequate, extend the model, most often by using hybrid QM/MM methods to incorporate the effects of the surrounding protein and solvent.
11. Decide whether dynamical effects are of interest. If so, design a Born–Oppenheimer Molecular Dynamics protocol using either a cluster or QM/MM.
12. Decide whether entropic effects are likely to be important (for example if charged species are released to the solvent) and, if so, decide on whether a quantum chemical approach (calculating the partition function within a harmonic-oscillator approximation) may be used or whether a molecular dynamics-based approach (e.g., free-energy perturbation theory) should be used to properly sample phase space.
13. For an MD approach with QM/MM design a protocol (preequilibration with a classical force field, substrate docking, boundary conditions, number and length of sampling "windows," etc.).
14. In all of the above, one of the most delicate choices is that of a suitable reaction coordinate.
15. For all approaches, analyze the results in terms of structural, energetic, and dynamic aspects and using the tools of interpretational-DFT as appropriate.

Although we do not pretend that the above protocol is unique or complete in all aspects, we will use it as a guide in the following sections, in the hope that it will be helpful to novices as they gain experience in what is, in the end, a complex field of computational chemistry/biology.

2 DFT and ADFT

2.1 Methodology

We take as a starting point the Kohn–Sham equations:

$$\left[-\frac{1}{2}\nabla^2 + v(r) + \int \frac{\rho(r)}{|r-r'|}\mathrm{d}r' + v_{\mathrm{xc}}(r)\right]\varphi_i(r) = \varepsilon_i\varphi_i(r) \tag{1}$$

with the operators in parentheses being, respectively, the kinetic energy of the noninteracting Kohn–Sham reference system, the external potential (nuclear-electron attraction), the classical coulomb potential felt at the point **r** due to the electronic charge distribution (including the self-interaction), and the exchange-correlation potential. φ_i is a Kohn–Sham orbital and ε_i is the corresponding Kohn–Sham orbital energy. The exchange-correlation potential is the functional derivative of the Kohn–Sham exchange-correlation energy with respect to a variation of the density:

$$v_{\mathrm{xc}}(\mathbf{r}) = \frac{\delta E_{\mathrm{xc}}[\rho(\mathbf{r})]}{\delta\rho(\mathbf{r})}. \tag{2}$$

The Kohn–Sham equations are exact but, of course, for practical calculations approximations have to be made and these will determine the accuracy, the speed, and the interpretability of approximate KS-DFT methods.

The most important choice is that of the exchange-correlation functional. Here there is a wide variety of options. Some standard choices are functionals of the Generalized Gradient Approximation (GGA) (which we favor because of their computational speed when used within the Auxiliary DFT approach and their overall good level of accuracy—see below; PBE [6–8] is a common choice) and the hybrid functionals that involve a component of Hartree–Fock exchange, the B3LYP functional [9, 10] providing the prototypical example:

$$E_{\mathrm{xc}}^{\mathrm{B3LYP}} = E_{\mathrm{xc}}^{\mathrm{VWN}} + a_0(E_{\mathrm{x}}^{\mathrm{HF}} - E_{\mathrm{x}}^{\mathrm{VWN}}) + a_{\mathrm{x}}(E_{\mathrm{x}}^{\mathrm{Becke}} - E_{\mathrm{x}}^{\mathrm{VWN}}) + a_{\mathrm{c}}(E_{\mathrm{c}}^{\mathrm{VWN}} - E_{\mathrm{c}}^{\mathrm{LYP}}), \tag{3}$$

where $a_0 = 0.20$, $a_{\mathrm{x}} = 0.72$, and $a_{\mathrm{c}} = 0.81$.

Other functionals have been specially parameterized for various applications, notably by the Truhlar group [11]. GGA functionals of the OPTX type have had considerable success in the area of bio-organic complexes [12–16].

Once the functional has been chosen, one needs to solve the Kohn–Sham equations, typically using some sort of basis set, although numerical approaches have seen some use [17]. Here we will focus on the use of Gaussian basis sets which are used in a number of software packages (Gaussian [18], NWCHEM [19],

Q-Chem [20], etc.) and also in our own code deMon2k [21]. We further outline the use of fitting functions [22] and the Auxiliary-DFT (ADFT) [23] methodology because it provides a real computational advantage provided (for the moment) that pure, nonhybrid, functionals are employed. Of course one has to address the accuracy question, both as concerns the inherent errors of approximate functionals and the numerical errors associated with the choice of basis sets, auxiliary basis sets, numerical integration grids, etc. These issues will be addressed in the discussion of the applications below, but first we complete the overview of the methodology with a discussion of density fitting and the ADFT methodology.

The use of Gaussian functions in DFT was pioneered by Dunlap, Connolly, and Sabin [22] who, in 1982, formulated the LCGTO-Xα method that incorporated a variational fit of the coulomb terms. With the usual LCAO approximation, and using a general form for the exchange-correlation energy rather than Xα (an early density functional that uses a local density approximation for exchange, incorporating a parameter α that, arguably, accounts for correlation to a certain extent), one can write the total energy in terms of the density matrix:

$$E = \sum_{\mu,\nu} P_{\mu\nu} H_{\mu\nu} + \frac{1}{2} \sum_{\mu,\nu} \sum_{\sigma,\tau} P_{\mu\nu} P_{\sigma\tau} \langle \mu\nu \parallel \sigma\tau \rangle + E_{\mathrm{xc}}[\rho(\mathbf{r})]. \tag{4}$$

Here $P_{\mu\nu}$ and $P_{\tau\sigma}$ denote elements of the density matrix and $H_{\mu\nu}$ an element of the core Hamiltonian (kinetic energy and nuclear attraction). The basis functions that are used to expand the molecular orbitals are indicated by μ, ν, σ and τ. The (up to) four-center electron repulsion integral is represented by:

$$\langle \mu\nu \| \sigma\tau \rangle = \int\int \frac{\mu(\mathbf{r}')\nu(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \sigma(\mathbf{r})\tau(\mathbf{r}) \mathrm{d}\mathbf{r}' \mathrm{d}\mathbf{r}. \tag{5}$$

The key development came from the realization that if one could fit the density using an auxiliary basis set then in practice one index could be saved, turning the problem from a basically N^4 scaling, where N is the size of the orbital basis set, to N^2M where M is the size of the auxiliary basis set. Writing

$$\tilde{\rho}(\mathbf{r}) = \sum_k x_k k(\mathbf{r}) \tag{6}$$

along with the expansion of the density using the density matrix:

$$\rho(\mathbf{r}) = \sum_{\mu,\nu} P_{\mu\nu} \mu(\mathbf{r}) \nu(\mathbf{r}) \tag{7}$$

and minimizing the following error

$$\varepsilon_2 = \frac{1}{2} \int\int \frac{[\rho(\mathbf{r}) - \tilde{\rho}(\mathbf{r})][\rho(\mathbf{r}') - \tilde{\rho}(\mathbf{r}')]}{|\mathbf{r} - \mathbf{r}'|} \mathrm{d}\mathbf{r} \mathrm{d}\mathbf{r}' \tag{8}$$

leads to the following equation:

$$\frac{\partial \varepsilon_2}{\partial x_k} = -\sum_{\mu,v} P_{\mu,v}\langle \mu v \| k\rangle + \sum_l x_l \langle l \| k\rangle \equiv 0 \forall k. \tag{9}$$

Solution of this equation is often called the Resolution of the Identity and is now also common in wave function methods. A second auxiliary set was introduced for the exchange-correlation potential; however, this used a least-squares procedure on a numerical grid and it was not done variationally. This level of theory is identified by the keyword BASIS in our deMon2k software since the density matrix is used (within a min–max scheme).

Going one step further in terms of potential speed, Köster and coworkers [23] studied the use of the auxiliary density throughout the variational process. The corresponding auxiliary DFT (ADFT) energy expression is the following:

$$E = \sum_{\mu,v} P_{\mu v} H_{\mu v} + \sum_{\mu,v,k} P_{\mu,v}\langle \mu v \| k\rangle x_k - \frac{1}{2}\sum_{k.l} x_k x_l \langle k \| l\rangle + E_{\mathrm{xc}}[\tilde{\rho}(\mathbf{r})], \tag{10}$$

where, in practice, as implemented in deMon2k, the orbital basis sets are contracted Cartesian Gaussians while the auxiliary sets are primitive Hermite Gaussians. In Eq. (10) the first term represents the core energy, the second the coulomb repulsion energy of the electrons using the density matrix once and the fitted density once, the third term complements the second in the variational fitting procedure and involves only the fitted density and the final exchange-correlation term also involves only the fitted density. The derivatives of this energy expression with respect to the density matrix elements define the ADFT Kohn–Sham matrix elements:

$$K_{\mu v} = \frac{\partial E}{\partial P_{\mu v}} = H_{\mu v} + \sum_k \langle \mu v \| k\rangle + \frac{\partial E_{\mathrm{xc}}[\tilde{\rho}(\mathbf{r})]}{\partial P_{\mu v}}. \tag{11}$$

After some algebra and using the variational property of the density fitting the following equation is derived:

$$K_{\mu v} = \frac{\partial E}{\partial P_{\mu v}} = H_{\mu v} + \sum_k \langle \mu v \| k\rangle (x_k + z_k), \tag{12}$$

where z_k are exchange-correlation fitting functions. The important thing about Eq. (12) is that the Kohn–Sham matrix elements are independent of the density matrix elements. As a result, only the fitted density (and, in the case of GGA functionals, the corresponding density derivatives) have to be numerically calculated on a grid. These quantities scale linearly by construction and so the grid work is reduced considerably. In addition, the use of shared auxiliary function exponents between the two auxiliary sets dramatically decreases the number of expensive exponential function evaluations, resulting in very favorable computing times.

2.2 Applications

2.2.1 Cluster Calculations for Histone Lysine Methyltransferase

We start the discussion of biomodeling using straight DFT with a brief overview of some recent work from Per Siegbahn and coworkers using the recent review of Siegbahn and Himo as the leading reference [24]. Siegbahn has been a champion for (properly converged) cluster models for a number of years and his well-chosen applications have shed considerable light on classes of enzymes for which reliable cluster models can be formulated. With recent increases in computer power and program efficiency, cluster models with upwards of 150 atoms are now feasible; for some, but not all, types of reactions, these are able to capture the essence of the relevant free-energy profiles.

Siegbahn's recent work focuses on the B3LYP hybrid functional for which he claims "In spite of numerous attempts, it has been difficult to improve the accuracy beyond that of this functional." Of course, that does not relieve us from the responsibility of addressing the question of accuracy for a given reaction. Siegbahn cites three sources of inaccuracy in DFT methods (1) the self-interaction error, (2) the inherent limitations of a single-determinant approach, and (3) the lack of van der Waals interactions in the usual functionals, including B3LYP. According to Siegbahn, errors (1) and (2) (using a spin-unrestricted formalism) tend to cancel and this can at least partially explain the relatively good performance of B3LYP. He also discusses the effects of varying the amount of exact exchange in the hybrid functionals, leading to the rule of thumb that if the results do not change very much when the amount of exact exchange is decreased from 20 % to 15 % the methodology seems to be reliable. While this seems less than an ideal procedure from an "ab initio" perspective, such procedures are necessary at the present stage of advancement in the search for more accurate and generally applicable functionals. The body of work using B3LYP indicates that it can provide results of useful accuracy if it is applied with due caution. We will discuss the third source of error in the next section on DFT-D methodologies. We only indicate here that the empirical corrections for dispersion-like interactions can now be readily incorporated and that they lead in many cases to significant improvement. They should become the default option.

Two further aspects of Siegbahn's cluster approach are (1) a coordinate-locking scheme and (2) the use of a polarizable continuum method (PCM) to model the electrostatic effects of the surrounding medium (protein and solvent). The coordinate-locking scheme fixes the coordinates of key atoms on the periphery, hence preventing large artificial movements of the active site groups. For very small cluster models this approach can lead to artifacts, but as the cluster grows it behaves better and better (and ultimately becomes unnecessary). The PCM approach assumes that the surroundings can be represented as a homogeneous polarizable medium with a dielectric constant that has to be chosen (often $\varepsilon = 4$ gives good results but in some studies ε is varied to gauge the sensitivity of the results to this parameter).

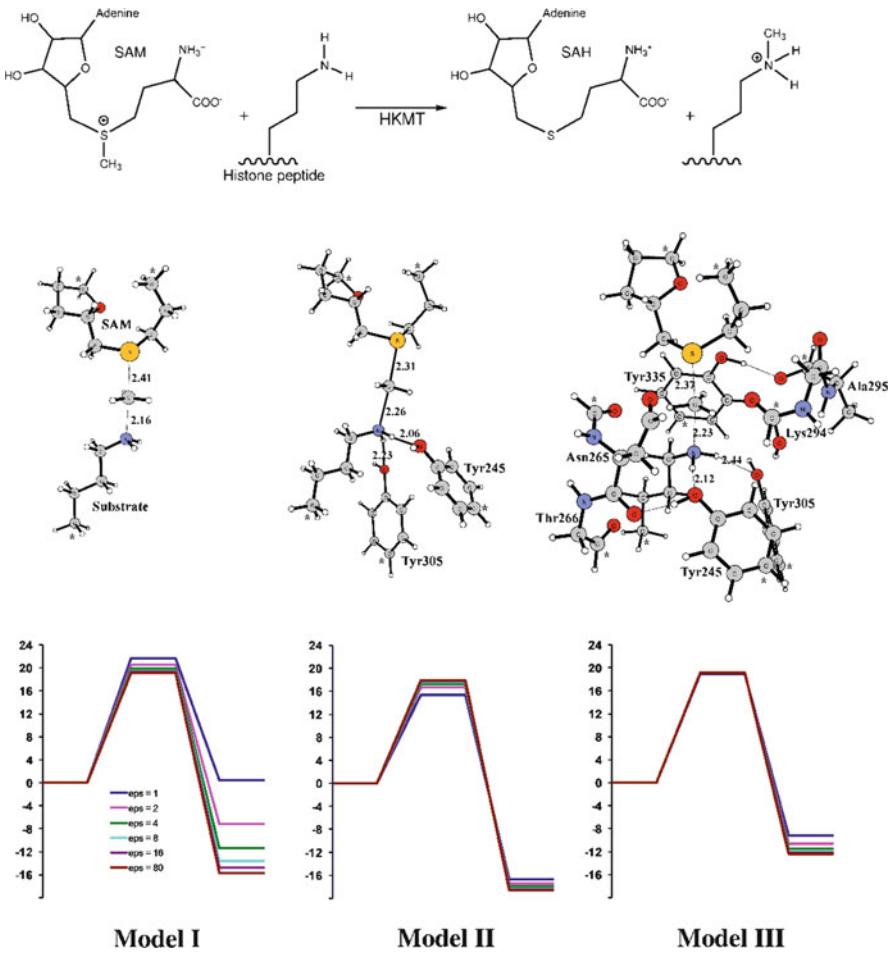


Fig. 1 *Top*: Reaction mechanism for the methylation of the Lysine side chain of histone by *S*-adenosylmethionine using HKMT as the catalyst. *Middle*: The three cluster models used. *Stars* indicate fixed atoms in the coordinate-locking scheme. *Bottom*: Potential energy profiles (kcal/mol). Results for various values of the dielectric constant are shown in *different colors*. Reproduced with permission from [24]

Clearly, the more protein residues and waters of solvation that are included in the cluster model explicitly, the less critical will it be to resort to this somewhat doubtful PCM approximation.

We turn now to histone lysine methyltransferase (HKMT) which catalyzes the methylation of the N-terminal histone tail of chromatin using the *S*-adenosyl-methionine (SAM) cofactor as the methylating agent (see Fig. 1).

Models of increasing size were used; Model I (46 atoms) contained only truncated models of SAM and the substrate; Model II (72 atoms) also contained two important tyrosine residues that form hydrogen bonds with the amino group of

the substrate and Model III (132 atoms, Middle part of Fig. 1) contains additional groups that form a ring around the substrate and interact with the transferred methyl group. The total overall charge on the models is +1. Considering only this particular reaction one has a "quantum-chemistry friendly" situation, where all of the reaction steps (reactants, transition state, and products) take place in a reasonably similar environment (provided by the interacting residues of Model III, for example) and a single positive charge is transferred from the cofactor to the substrate. Geometries of the critical points were calculated at the B3LYP/6-31G(d,p) level with single-point calculations of the energies using a larger 6-311 + G(2d,2p) basis. Solvation effects were calculated as single points at the same level as the geometry optimizations using a range of dielectric constants (2, 4, 8, 16, 80). As expected the larger models showed little dependence on the value of ε chosen (essentially identical values for the entire range of ε for the activation energy from the reactant side and roughly a range of 4 kcal mol^{-1} for the overall exothermicity of the reaction) (see the bottom part of Fig. 1).

It was found that the transition-state geometries for the three models were very similar which is of practical importance, allowing the TS search to be performed for small models and then only having to refine the structures for the larger models. Moreover, the calculated energy differences are not very different in the different models. In fact all of the barriers are close to the experimental barrier of 20.9 kcal mol^{-1}. The solvation effects saturate quickly; 132 atoms in this case seem to be well converged, a fortunate circumstance for this reaction and a good number of other reactions, because such cluster sizes are well within the range of current possibility. It was emphasized that geometry optimization is crucial; taking raw crystallographic geometries usually leads to wrong energy profiles.

Some final comments on the entropy (free energy) are in order. Because of the coordinate-locking scheme, there are a number of small imaginary frequencies ($<30i$ cm^{-1}) which, although they do not affect the energetics significantly, do render the calculation of the harmonic frequencies and their associated entropy contributions inaccurate. So the reported energies correspond to enthalpies and not free energies. For the present HKMT case, Siegbahn refers to work by Hu and Zhang [25] to show that the entropy effects are quite small. Hu and Zhang used a QM/MM methodology (B3LYP with single-point MP2) with free-energy perturbation theory (see below) for the MM contributions to the entropy and harmonic frequencies for the QM part (now possible because there are no frozen atoms). The QM part contained 66 atoms, SAM and the lysine side chain only. The free energy barriers are found to be only about 1.1 kcal mol^{-1} lower than the potential energy barriers, confirming the validity of neglecting entropy for this particular reaction, which is well contained within the QM part of the system. We will see below that this is not a general result. For some reactions, such as those involving DNA or RNA polymerase, entropy effects can be very large; they definitely cannot be neglected.

For other examples of the valid use of the finite cluster model, we refer the reader to [26–31].

2.2.2 Cluster Calculations for RNA Polymerase

In the central dogma of biology ("DNA to RNA to proteins") the first step transcribes the genetic code from DNA to messenger RNA. This task is accomplished by a marvelous nano-molecular machine, the RNA Polymerase (RNAP) enzyme, which, once initiated, processes a DNA template strand, adding successive matching nucleoside triphosphates to a growing chain of m-RNA. RNAP is a complex multidomain protein containing about 3,500 residues and 28,000 nonhydrogen atoms [32]. One of the great triumphs of modern protein crystallography is surely the elucidation of many aspects of the mechanism for transcription for which Roger Kornberg was awarded the 2006 Nobel Prize in Chemistry. In principle, the work of Kornberg and the others who have accomplished this magnificent task is simple—isolate and crystallize proteins that are ligated by various intermediates along the multistate reaction path and derive their structures by X-ray analysis. In this way we get "flash-frozen" snapshots of the reaction mechanism. This has been so successful that a movie of the process has been made [33] which may be viewed at http://www.lmb.uni-muenchen.de/cramer/pr-materials. We show some snapshots from the movie in Fig. 2.

This is clearly a more complex reaction than that of HKMT just described. Of the seven steps shown in the movie all but one involve conformational changes of the enzyme that are more in the domain of Molecular Mechanics than in that of quantum chemistry (we are studying these conformational steps, but they are not the topic of the present chapter). The sole exception is step 4, catalytic incorporation, which actually involves a multistep mechanism of chemical reactions. This is where theory and computation have to step in to help elucidate the mechanism. The first steps of the process involve the construction of cluster models for the calculation of relevant portions of the potential energy surface corresponding to proposed reaction steps. Several key choices have to be made for which reactions to consider.

The basic enzymatic function of RNAP is the transfer of the nucleotidyl motif from the rNTP substrates to the hydroxyl at the 3′-end of the nascent RNA transcript. The nucleotidyl transfer reaction can be simplified as

$$\mathrm{RNA}_i + \mathrm{rNTP} \xrightarrow{\mathrm{RNAP}} \mathrm{RNA}_{i+1} + \mathrm{HPP}_i. \tag{13}$$

The catalytic center of RNAP includes the binding site for the 3′-end of RNA and the insertion site for the incoming rNTP. In the nucleotidyl transfer reaction, the 3′-OH group in the sugar ring of the RNA primer reacts with the α-phosphorous atom of a ribonucleoside triphosphate by nucleophilic attack, then the P_α–$O_{\alpha\beta}$ bond is broken and pyrophosphate (PP_i) is released. Thus, a nucleotidyl addition to the RNA primer is achieved. Structural and biochemical data have shown that the active centers of all polymerases share certain common features: a pair of metal ions (normally divalent magnesium ions Mg^{2+}) and three universally conserved carboxylates. The two-metal-ion mechanism for the nucleotidyl transfer reaction

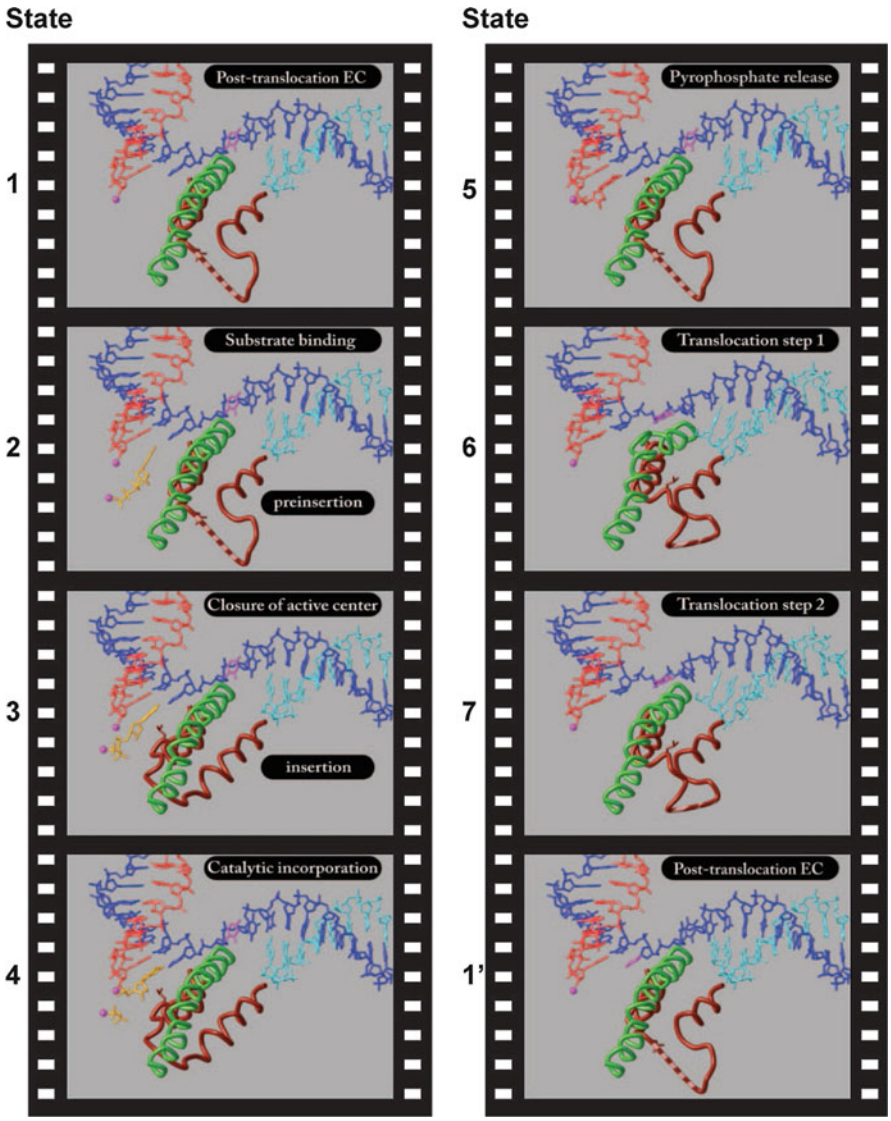


Fig. 2 Snapshots from the NAC (Nucleotide Addition Complex) movie have been depicted that correspond to different functional states of the EC. Reproduced with permission from [33]

was proposed by Steiz [34]. A schematic diagram of the ternary elongation complex for yeast RNA polymerase is shown in Fig. 3.

We have considered several different detailed reaction mechanisms for the nucleotide addition [36]. In all of them, the 3′OH group has to be deprotonated, either by passing its proton to another residue of the enzyme or to a solvent water

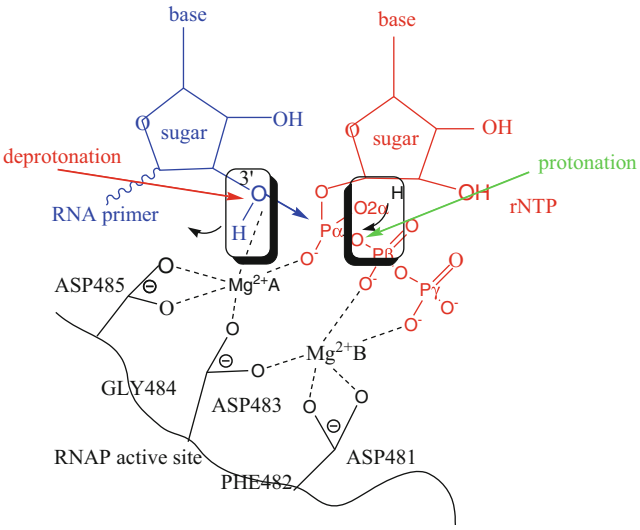


Fig. 3 The two-metal-ion mechanism and two-proton transfer model for the nucleotidyl reaction of yeast RNA polymerase II [34, 35]. The ternary elongation complex consists of three parts: the growing RNA transcript (*blue*), the incoming rNTP (*red*), yeast RNA polymerase active center (*black*) that is mainly composed of two divalent magnesium ions, Mg^{2+} A and B, and three conserved amino acid residues, aspartates. The two protons are directly or indirectly transferred at two sites proposed according to deuterium isotope effect experiments. One site is the primer RNA 3′ terminus (*left square*), where the RNA primer 3′OH must be deprotonated by the incoming rNTP, a nearby residue or a mediated water molecule before nucleophilic attack. The other is the α- and β-phosphate bridging oxygen atom (*right square*), where the pyrophosphate should be protonated by a nearby residue or a mediated water molecule before it leaves. The three aspartate residues ASP481, 483, 485 connected by PHE482 and GLY484 conjugate with two magnesium ions to form the active center of yeast RNA polymerase

molecule or directly to the departing pyrophosphate. One of these mechanisms is illustrated in Fig. 4.

Initial molecular dynamics simulations were performed in order to derive a plausible starting structure for subsequent optimization with DFT. The QM model includes a simplified RNA primer (a sugar ring with 2′ and 3′OH groups), an incoming nucleotide GTP substrate, two divalent metal cations, Mg^{2+}A and B, a solvent water molecule, and three universally conserved Aspartate residues (Asp481, Asp483, Asp485) where Asp481 and Asp483 are connected by PHE482, Asp483 and Asp485 are connected by Gly484. The three conserved aspartate residues form a lotus-like complex so that they are quite robust during geometry optimizations. Our previous work showed that the simplified aspartate residues (formic acid $HCOO^-$ or acetic acid CH_3COO^-) changed a lot during geometry optimizations [37].

We found that the barrier height of direct proton transfer from the 3′OH of the RNA primer to the O2α of rNTP is higher than that for proton transfer from the 3′OH of RNA primer to water. Thus, the latter model will be discussed in the following.

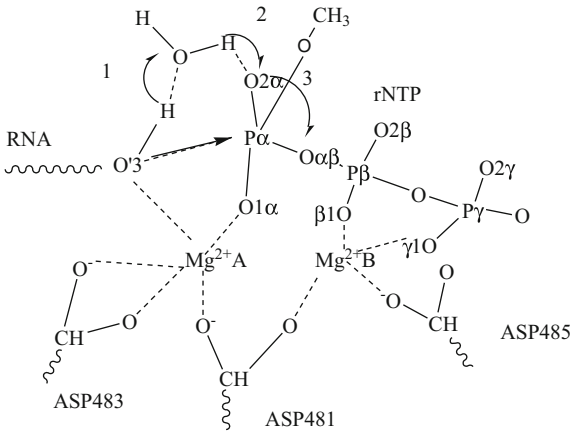


Fig. 4 The proposed nucleotidyl transfer reaction mechanism. Step 1: The proton of the 3′OH of RNA primer transfers to a solvent water and a proton on water transfers to O2α of α-phosphate simultaneously; Step 2: the proton of the O2α atom rotates to the side of β-phosphate; Step 3: the 3′O atom performs a nucleophilic attack at the α-phosphorus atom of the α-phosphate; Step 4: the Pα–O3β bond of the intermediate cleaves to form a phosphodiester bond and the proton on O2α migrates to O3β

The model consists of two Mg^{2+}, three conserved aspartate residues, one ribose, the simplified RNA primer, a simplified rNTP, and a water molecule that is located between the growing RNA primer and the incoming rNTP and is closest to the 3′OH of RNA primer and the α-phosphorus atom of the incoming rNTP. The model has 94 atoms with a total charge of −3. This is due to the three conserved aspartate residues (−3 charge) of the yeast RNA polymerase II active site, the triphosphate (−4 charge) of the incoming rNTP, 2 Mg^{2+}(+4 charge) at physiological pH (7.2–7.5). Considering the physiological pH in the experimental condition and the pK_a of triphosphate, the incoming rNTP is thought to be in a deprotonated state in this study.

All density functional (DFT) calculations were performed using the deMon2k (version 2.4.4) program. We carried out full optimizations on all structures, reactant, intermediates, transition states using the PBE exchange-correlation functional with the basis set DZVP-GGA and auxiliary basis set GEN-A2. No constraints were imposed on any atom of these systems.

We proposed the following detailed nucleotidyl transfer reaction mechanism for yeast RNA polymerase II, shown in Fig. 4. The proton of the 3′OH first transfers to the O2α of α-phosphate via a solvent water molecule, then one of the water molecule's protons transfers to the bridging phosphate O3β atom. Note that the water molecule is located by our CHARMM molecular dynamics (MD) simulations.

The calculated potential energy profile for the reaction is shown in Fig. 5.

If taken at face value the analysis of these four steps could provide detailed insight into the various proton transfer and bond making and -breaking steps and the

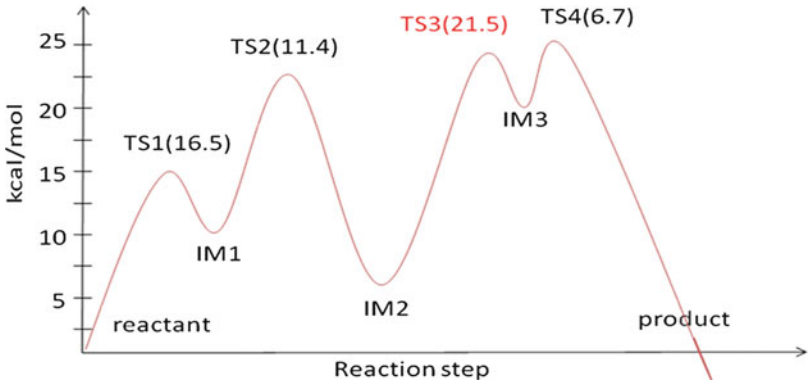


Fig. 5 The calculated potential energy curve for the PBE/DZVP-GGA method

role of water molecules. Indeed, we have performed such an analysis [36] but we have chosen not to publish it yet because of potentially large effects of the protein/solvent environment and, especially, the probably large entropic effects, based on analogous QM/MM calculations for DNA polymerases that are discussed in the QM/MM section below.

3 DFT-D

Biological systems are large, solvated in water and comprise domains with low overlapping density, which are only weakly interacting with each other. They belong to what has generally been called "soft matter." Simultaneous charge fluctuations in the various nonbonded parts of soft matter systems generate attractive van der Waals interactions that are nonlocal correlation effects. These correlation effects are naturally accounted for in correlated wave function methods including a sufficiently high level of electron correlation (the basic dispersion interaction is already accounted for at the MP2 level), but these methods are relevant only for small systems and they are mainly used for benchmark calculations.

Despite the fact that the exact density functional contains van der Waals correlation, the inability of local (LDA) or generalized gradient approximations (GGA) or even more sophisticated meta-GGA functionals to treat properly long-range interactions is now recognized. Significant progress in making DFT more appropriate for weakly interacting systems has been made with long-range density functional theory, applying to nonoverlapping [38–42] and also overlapping electron densities [43, 44]. Molecular polarizabilities, from which the dispersion interaction energy can be calculated, have been estimated from time-dependent calculations [45] or from the instantaneous dipole moment of the exchange hole [46–48]. New combinations and parameterizations of GGA [49, 50] or meta-GGA [51] exchange-correlation functionals have also been proposed to allow the incorporation of the long-range part of electron correlation.

Based on the treatment of asymptotic van der Waals forces [52], a van der Waals correlation density functional has been proposed by Dion et al. [53] and applied to solid and biosystems [54]. The use of this correlation functional for large systems has been limited due to the double spatial integration for calculating the correlation energy. However, a very recent implementation algorithm has been proposed to overcome this bottleneck [55].

Because it has been known for a long time that van der Waals interaction energies are important for large systems, a more empirical approach was first used by Wu and Yang [56], adding an empirical dispersion energy correction (E_{disp}) to the usual DF energy (DFT-D approach) based on atomic C_6 coefficients, derived from molecular C_6 coefficients. The idea of using a correction term originates from Hartree–Fock-based studies [57–59] and from Elstner et al. [60] for correcting the Self-Consistent-Charge DFT-Tight Binding method. Whereas the nonlocal character of the Hartree–Fock exchange provides a correct description of the long-range intermolecular forces, the DF approaches based on the local density expansion are not strictly applicable and their performance depends on the particular exchange-correlation functional. A damping function has to be used in order to set E_{disp} to zero as the electron densities overlap. Different analytic forms were tested [56, 61]. Based on a test set of small van der Waals systems, proper scaling factors of E_{disp} were proposed for different XC density functionals [62]. Instead of scaling the empirical dispersion energy, Jurečka et al. [63] adopted a global scaling factor of the atomic van der Waals radii, optimized for a training set of noncovalent complexes. More recently, inclusion of higher order correction terms, involving C_8 and C_{10} coefficients associated with an adequate damping function, has been proposed by Johnson and Becke [48], aiming at a better description of π stacked systems.

In recent years more sophisticated DFT-D dispersion corrections have been proposed and benchmarked on large test sets, allowing comparison with a very complete set of XC functionals [64]. Improving the use of atomic C_6 coefficients for dispersion corrections, a density-dependent energy correction for long-range dispersion has been proposed, based on the Becke–Johnson exchange hole dipole moment formalism [65] and an extended Tang–Toennies damping function [66] accounting for charge-overlap effects [67]. This formalism has been benchmarked for different functionals and test sets [68].

In our DFT-D applications, the dispersion term, limited to the dipole–dipole contribution to the dispersion interaction energy,

$$E_{\text{disp}} = -\sum_{i=1}^{N}\sum_{j=i+1}^{N-1}\frac{C_6^{ij}}{r_{ij}^6} f_{\text{damp}}(r_{ij}) \tag{14}$$

is expressed as the sum of the i,j atom-pair contributions in an N-atomic system with an interatomic distance r_{ij}. In the DF approach including a damped empirical correction for the van der Waals interactions, E_{disp} is simply added to the usual DF energy and ∇E_{disp} is added to the DF energy gradient. The present E_{disp} expression,

limited to the first C_6/R^6 term, contains implicitly most of the physical intermolecular dispersion via the fitting of the atomic C_6 coefficients to molecular C_6 values, obtained from a training set of 44 pairs of molecules including hydrocarbons and other small organic compounds [56].

The dispersion coefficients

$$C_6^{ij} = \frac{2C^iC^j}{(C^i + C^j)} \tag{15}$$

are computed from the atomic C_6^i, as proposed by Wu and Yang [56], but averaged over the possible hybridization states of the atoms, which are 2.845 and 26.360 a.u. for H and C atoms, respectively. The damping function used is

$$f_{\text{damp}}(r_{ij}) = \frac{1}{1 + e^{-\alpha\left(\frac{r_{ij}}{r_0} - 1\right)}} \tag{16}$$

with $\alpha = 23.0$ [24, 56] and with r_0 being the sum of the atomic van der Waals radii [69, 70].

The dispersion correction energy, its gradient and second derivative were implemented in the deMon2K program [29], allowing geometry optimization, Born–Oppenheimer molecular dynamics, and vibrational frequency analysis. We used a nonscaled empirical dispersion term, unlike previously mentioned DFT-D calculations [62, 63].

In fact, scaling E_{disp} has been used to compensate the erroneous behavior of GGA exchange functionals which either show an unexpected attractive tendency in the van der Waals region (and thus avoids a double counting of the "dispersion") or, in contrast, have a too strong repulsive slope in this region. The latter trend is mainly displayed by the Becke exchange functional [71] whereas PW91 [72] and PBE [6] exchange functionals are responsible for the former.

Exchange-only calculations of alkane dimers illustrate this problem, as well as the related question of choosing a correlation functional which can compensate the exchange repulsion at nonbonding distances that correspond to short-range van der Waals attractive interactions between hydrophobic alkyl chains [73]. Taking as an example two butane molecules, one can analyze the effect of the exchange functional on their interaction energy as a function of distance: the revised version of PBE, i.e., revPBE [74] and TPSS [75] repulsive curves are close to HF for long distances, i.e., between 4.4 and 5 Å where they reach zero, insuring no double counting of any correction for "dispersion" effects. In contrast, Becke exchange leads to a much stronger repulsion, whereas PBE is attractive for distances beyond 4.4 Å. Similar conclusions have been reported for the benzene–benzene interaction [53].

On the other hand, correlation interaction energies also have to be analyzed in combination with the exchange interaction energy term. As expected, correlation interaction becomes more attractive when decreasing the separation between the two butane monomers. In this test example, PBE exchange and correlation

functionals without the empirical dispersion correction lead to a weakly bound butane dimer (-0.8 kcal mol^{-1} at a separation of 4.5 Å), whereas a PBE-D (including E_{disp} correction) calculation leads to a much too large stabilization of 4.1 kcal mol^{-1} compared with the MP2 and CCSD(T) interaction energy of 2.7 kcal mol^{-1}. However, combining the PBE correlation with the revPBE exchange yields a large repulsive interaction in the range of 4–4.4 Å which cannot be compensated by the damped dispersion correction. In contrast, combining the "semiempirical" revPBE exchange, adjusted on atomic exchange energies, with the LYP correlation functional [76], self-interaction (SI) free by construction, allows a good balance of repulsive and attractive effects between the two subsystems. In fact, the revPBE exchange leads to less SI error than PBE (5 times less for the H atom). The PBE exchange SI error is mostly compensated by the SI error of the PBE correlation itself. Therefore, combining the revPBE exchange with the PBE correlation is much less appropriate than combining it with the LYP correlation. This strategy is not based on fundamental arguments but on an empirical analysis of XC functional behavior. Different applications on saturated and aromatic hydrocarbon compounds have shown that the revPBE-LYP exchange-correlation functional augmented with the empirical van der Waals correction gives a good description of soft matter containing alkyl chains, without being considered as the panacea for a general and accurate estimate of long-range interactions [73]. In a similar vein, a recent study of exchange-only interaction energies of small molecules led to the proposal that a re-parameterized PW86 exchange functional can be recommended for its performance in yielding no spurious intermolecular binding when augmented by an empirical dispersion term [77].

The interest of taking dispersion interactions into account for soft matter systems can be illustrated by the thermal properties of lipids in cell membranes, which get more rigid at a given temperature when their alkyl chains get longer. This property is exploited by natural mechanisms for maintaining cell membranes fluid but not leaky, either increasing the number of long chain lipids or making them shorter. This very important property is related to the increasing dispersion-like stabilizing interaction of alkane chains when increasing the number of CH_2 units in the chains. The estimated value, corresponding to the most favorable distance of 4.10 Å between two chains, has been found to be -0.81 kcal mol^{-1} from DFT-D, -0.86 kcal mol^{-1} from MM, and -0.88 kcal mol^{-1} from the G3 (CCSD(T)) results (butane and hexane) [73]. Calculations using MP2 (limit) lead to a slightly overestimated value of 0.97 kcal mol^{-1}, as usually observed [78].

It is worth noting that dispersion energy also contributes to the total energies of the individual *n*-alkane monomers. This contribution varies from -3.7 for butane to -14.8 kcal mol^{-1} for dodecane, with a regular decrease of -1.4 kcal mol^{-1} per CH_2 unit. The internal E_{disp} value is thus about 50 % larger than the dispersion contribution to the dimer binding. This shows that, in the alkane monomers, the damped dispersion at mid-range contributes also to the total energies. However, the amount of stabilization provided by the internal dispersion in an alkane molecule is very small with respect to the other energy contributions (0.55 % of the correlation energy for *n*-butane, 0.57 % for isobutane, 0.70 % for *n*-octane, and 0.81 % for

isooctane). This explains why this E_{disp} term cannot compete with some other effects, such as those governing the branched/linear alkane relative stabilities [79].

Similar intra- and inter-molecular dispersion effects have been estimated in the case of phosphatidyl choline (PC) lipids, with two acyl chains of 12 (dilauroyl, DL), 14 (dimyristoyl, DM), and 16 (dipalmitoyl, DP) CH_2 units. The conformational analysis of these lipids has shown the existence of two relative orientations of the chains with close energies, having their carbon backbones in parallel and in perpendicular planes [80].

It is worth noting that the dispersion-type energies differ for these two chain arrangements with a larger stabilization (about 4.5 kcal mol^{-1}) for the parallel arrangement. Indeed, the optimum distance for dispersion between two alkane monomers in a dimer is about 4.2 Å, which is also the distance found between the PC lipid alkane chains in the "parallel" conformers as compared to the distance of 4.8 Å for "perpendicular" conformers. However, the shorter distance stabilizing effect is compensated by a structural rearrangement in the glycerol backbone and the balance of the above two effects leads to similar total energies for the two types of chain structures. The comparison of the computed empirical dispersion energies for several lipid isomers with 12, 14, and 16 carbons in their alkyl chains and in the two tail arrangements reveals a regular increase of stability of approximately 3.1 kcal mol^{-1} per CH_2 unit [81]. This increase of stabilization from DLPC to DPPC is more regular for parallel chains. These results allow one to anticipate qualitatively the thermodynamic behavior of these lipids as well as, more generally, the fluidity of lipid bilayers (membranes): long chains are more stable than shorter chains and will gain more stability in assemblies of lipids. As a consequence, the evolution from ordered (more stable) chains to disordered (less stable) will be less easy for longer chains, needing higher temperature, than for shorter chains. In fact, the transition from ordered to disordered chains, called the main phase transition for lipid bilayers, is measured at −2°, 27°, and 41 °C for DLPC, DMPC, and DPPC, respectively.

4 BODFT-MD

4.1 Introduction

The concept of motion in science goes back to Newton [82]. But there was a long time gap until 1959 when a general method for the computer simulation of particle dynamics (500 hard spheres) appeared [83]. Molecular dynamics (MD) simulations using atomic parameters started in the 1970s, studying water [84] and very soon protein conformations and folding, with the early mention of "computer simulations" [85]. About 10 years later, dynamics was introduced into the world of quantum chemistry with ab initio molecular dynamics (AIMD) studying the dynamical motion of atoms by solving, as exactly as is required, the entire quantum mechanical electronic structure problem and deriving the forces on the atomic

nuclei using the Hellmann–Feynman theorem. These forces are then used to move the atoms as in classical Newton dynamics. The breakthrough in this domain was the method proposed by Car and Parrinello (CPMD), solving coupled equations of motion for both nuclei and electrons based on DFT potential energy calculations [86]. This work opened perspectives in many areas of science. Converging the electronic structure at every time step, i.e., following the Born–Oppenheimer (BO) potential energy surface was proposed later, bringing the advantage of providing fully converged structures all along the trajectory, which is of great interest for the statistical estimate of dynamical properties [87].

Thanks to the evolution of software and computers, CPMD or BOMD dynamics can be applied now to larger systems than in the 1990s. However, these AIMD calculations are still very time consuming due to the repetition of the self-consistent electronic structure and gradient calculations at every time step, that is generally about 1 fs for BOMD and about ten times less for CPMD. Even if parallelization decreases dramatically the computational effort, long trajectories at the ns scale, which is still very short in biochemistry, are only possible for systems with less than about 100 atoms.

Although questions asked in biology and biochemistry are at a much coarser, macroscopic level, experimental tools and studies have brought accumulated information on molecular structures and mechanisms, reducing the gap with theoretical investigations at the microscopic level. BOMD can thus bring knowledge about the role of relatively fast structural time evolution of properties that are revealed at a macroscopic level. In this perspective, one must stress that ab initio methods are not avoidable, because of their accuracy in taking into account the correlation between all degrees of freedom of the system, whatever are its elements and size, which is not possible with interatomic pair potentials. This particular point may be the major reason to perform AIMD on model systems rather than classical MD on more extended ones.

We believe that it is also why the results presented here on single lipid molecules may bring fundamental information on properties of lipid assemblies.

4.2 Properties of Phosphatidyl Choline Lipids

All living systems are made up of cells and all cells are limited by membranes which are built from lipid bilayers. Membranes are held together in water by hydrophobic intermolecular interactions that favor self-assembly of lipid molecules and tend to close bilayers, avoiding holes. Cell membranes include proteins and carbohydrates and modulate the flow of ions and polar molecules. The fluid nature of the lipid membranes is of critical importance. The fluidity of the cell membranes, normally in the liquid crystalline state, is precisely regulated because lipids undergo phase changes in response to temperature. Phospholipids, according to their fatty acid compositions, have a specific main phase transition temperature, T_m, also called the melting temperature, between the gel (ordered)

and liquid crystalline (disordered) states. The melting transition is accompanied by enthalpy and volume changes.

It is worth noting that lipids in membranes are not covalently bound and are moving individually, diffusing laterally inside their layer (about one time per 100 ns) and, even, undergoing a flip from one leaflet to the other, although much more rarely. These data show that individual lipids behave as separate entities, even if they are subject to a collective physics. Moreover, one can expect that intramolecular bonding prevails over intermolecular interactions, at least, for neutral lipids.

4.2.1 Structural Studies of Phosphatidyl Choline Lipids

Experimental X-ray and NMR studies of lipids have recognized that the knowledge of the structure and dynamics of a phospholipid monomer within an assembly is essential for the understanding of the functional role of the bilayer in biomembranes [88–94]. The effects of the intra and intermolecular electronic forces on the structure and energetics of these systems make their study nontrivial both experimentally and computationally. However, there is general agreement concerning structural differences between the fluid-like and the crystal conformers. This fact questioned the relevance of using the single crystal atomic positions for fluid phase simulations wherein the internal monomer properties appear to be better preserved, due to the relatively small intermolecular interactions [91, 92]. The use of isotopically labeled atoms in Fourier transform infrared (FTIR) spectroscopy allowed the assignment of lipid functional group vibrations, the estimate of their shift in water, and the presence of blue-shifted CH_2 stretching vibrations for temperatures above T_m which have been associated with the appearance of gauche conformations in the alkyl chains [95–99].

Among the large number of MD studies devoted to phospholipid structural properties, dihedral angle values [100, 101], tail orientations [102], head group flexibility [103, 104], phase changes [105, 106], hydration effects [107, 108], and evaluation of local order parameters [109, 110] have been explored. A thorough exploration of the conformational spaces of the dilauroyl (DLPC), dimyristoyl (DMPC), and dipalmitoyl (DPPC) phosphatidylcholine molecules was performed using DFT-D, showing, for the three lipids, the existence of a large number of quasi isoenergetic conformers [80, 81, 111, 112]. These conformers differ by the relative position of the two alkyl chains and by different combinations of the backbone torsion angles (Fig. 6). For the three lipids, the backbone rotational conformers share a common geometric profile which includes a balance of attractive, repulsive, and constraint forces between and within specific groups of atoms. The definition of this profile fits with most of the structural characteristics deduced from measured NMR properties of DMPC and DPPC solutions. The calculated vibrational spectra are similar and in very good agreement with experimental data obtained for these PC bilayers. These results support the idea that these molecules preserve their individual molecular structures in their various assemblies. In fact, this conclusion is confirmed by our studies of the dynamical behaviors of DMPC and DLPC [111, 112].

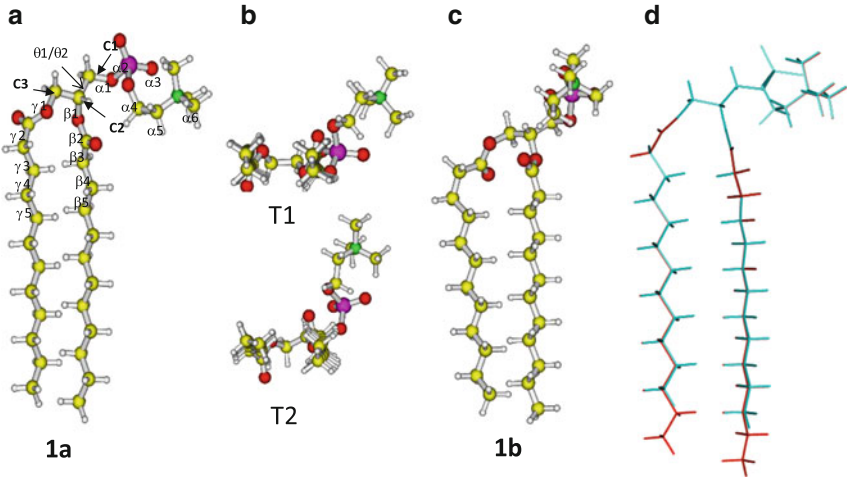


Fig. 6 Structures of phosphatidyl choline lipids: (**a**) isomer 1a; (**b**) parallel (T1) and perpendicular (T2) chain conformations; (**c**) isomer 1b; (**d**) superposed DLPC (*blue*) and DMPC (*red*) isomer 1a optimized structures

4.2.2 Experimental Phase Transitions

Most of the reported structural transformations of PC lipid bilayers with respect to environmental conditions such as temperature, pressure, pH, etc., are associated with isomerizations of the constituent lipid molecules (mostly 14–24 carbon acyl chains). For instance, the transition from the ordered gel to fluid liquid crystalline phase, called the main phase transition, has been related to the melting of the hydrocarbon chains: in the gel phase, phospholipids with all trans alkyl chains are present, whereas in the disordered liquid crystalline phase the most populated conformational states correspond to gauche forms in the alkyl chains.

In the last decade, accurate microcalorimetric and atomic force microscopy experiments have provided detailed information on the lipid melting processes [113–118]. Interestingly, a linear relationship between the isobaric heat capacity and the volume expansion with temperature has been observed for a variety of lipids [115–118] leading to the interpretation that proportional enthalpy and volume changes at the melting transition would be driven by intrinsic structural changes within the lipid molecules, whereas the changes of free volumes and intermolecular interactions could be considered as perturbations [116].

The thermograms of DMPC and longer chain lipid bilayers show a small peak at temperatures well below that of the sharp peak of the main phase transition. The general interpretation is that, when the bilayers are fully hydrated and sufficiently incubated at low temperature, this small endothermic peak (about 1 kcal mol^{-1}) is related to a pretransition between a L_c gel state and a rippled gel state ($P_{\beta'}$), both characterized by all trans chain conformers. This pretransition state is thus not related to molecular structural changes, but to a different packing arrangement of

the molecules in the gel. The pretransition peak is followed by the main phase transition to a liquid crystalline L_α state with predominantly gauche chain conformers of the constituent lipid molecules.

In contrast, the phase pattern of the DLPC thermograms, characterized by one sharp peak followed by a broad band spreading over about 10 K is more complex. This complexity is enhanced by the overlapping of lipids and water solid–liquid phase transitions. Various interpretations of the unusual DLPC thermograms were proposed. From most experiments using various techniques, two phase transitions were assigned to the two peaks [119–123]. These latter studies agreed also in reporting a strong dependence of the position of the first peak on the incubated temperature and the heating rate. Finally, the most recent experiments, using a mixture of water and ethylene glycol and thus avoiding water freezing at the temperature close to the first peak, showed the existence of three peaks [124].

MD simulations of lipid bilayers as a function of temperature have not been extensive. The reliability of force fields is dependent on the accuracy of describing the torsional potential energies of alkanes which prompted improvements of some of the most commonly applied force fields for biomolecules [125] using recent ab initio computations of torsional potential in various trans/gauche *n*-alkanes (up to *n*-decane) [126]. A few recent MD simulations have analyzed the temperature effects on models of PC lipid bilayers, mainly for long chains (16–18 C), with a reasonably successful prediction of the phase transition temperature (error of 12–50 °C) [106, 127–130]. These studies relate the melting temperature to the fast increase of gauche conformers in the alkyl chains.

4.2.3 BODFT-MD of DLPC and DMPC Molecules

In agreement, with the experimental results, the simulated dynamical behaviors of the DMPC [112] and DLPC [111] molecules were found to be quite different. In both cases, BOMD simulations using the DFT-D methodology were performed for a set of temperatures ranging from below to above the experimental T_m for lipid bilayers, i.e., about 295 and 270 K for DMPC and DLPC, respectively.

Quantification of the structural deformations occurring along the molecular dynamics is obtained using the distance-fluctuation criterion initially introduced by Berry et al. [131]. The Berry parameter Δ_B is expressed for a system of N atoms and r_{ij} inter atomic distances between atoms i and j as

$$\Delta_B = \frac{2}{N(N-1)} \sum_{i<j} \frac{\sqrt{\left\langle \Delta r_{ij}^2 \right\rangle - \left\langle r_{ij} \right\rangle^2}}{\left\langle r_{ij} \right\rangle}. \tag{17}$$

The critical value of Δ_B for the solid-to-liquid transition of finite systems has been suggested to be close to 0.1 [132–134].

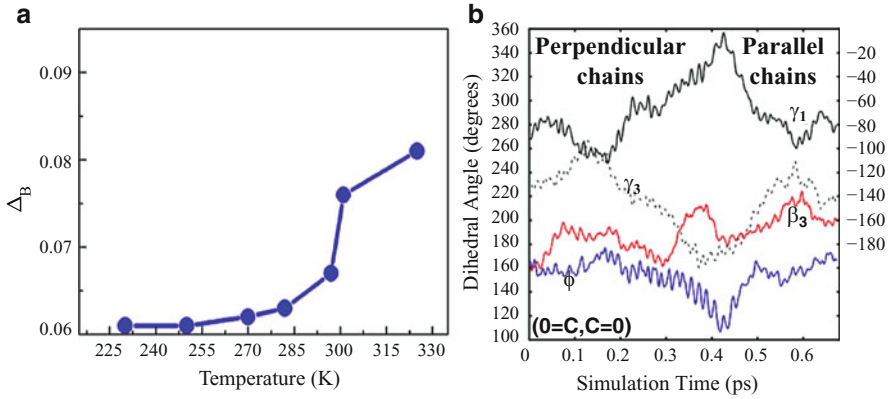


Fig. 7 Dynamics of DMPC: (**a**) Berry parameter Δ_B evolution with temperature; (**b**) Short period chain fluctuations at 250 K, showing variation of the four involved dihedral angles. Reproduced with permission from [112]

Focusing on the microscopic structural changes of these molecules as a function of temperature, one can indeed relate them to the existence of only one peak for DMPC and two peaks for DLPC.

The DMPC atomic motions for $T \leq 282$ K correspond essentially to reorientations of the chain CH_2 groups around the chain axes. The planes of the alkyl carbon skeletons are thus fluctuating between parallel and perpendicular (see Fig. 7b). At about 297 K, mostly perpendicular chain orientations remain, fluctuations keeping both chains at a distance comparable to that found at low temperature, but a trans to gauche conformer is formed (kink) at one chain end. Such a transformation also occurs at 301 K in the middle of the chain, increasing the interchain distances and the Δ_B value. Raising the temperature leads to more configurational states introducing more kinks in the chains. In the temperature range explored (230–325 K), no other conformational change was observed than the kink occurrences in the alkyl chains. Figure 7a illustrates the evolution of the DMPC Δ_B parameter with temperature.

The situation is very different in the case of DLPC. It must be noted that DLPC, with 12 C chains is the shortest possible lipid that can assemble in layers. As a consequence of their shortness, DLPC alkyl chains appear much more mobile than those of DMPC, even at low temperature. Moreover, several isomer interchanges occur along the DLPC BOMD trajectories generated at different temperatures, starting from the two isoenergetic isomers illustrated in Fig. 6a, c (**1a** and **1b**). These isomers differ only by the two dihedral angles β3 and γ4. Figure 8a illustrates the concomitant changes of these dihedral angles along the **1a** dynamics at $T = 240$ K. When starting from **1b**, the first structural change (kink in one chain) occurs at $T = 261$ K. For both **1a** and **1b** isomers, gauche alkane chain conformers appear at $T \geq 261$ K. It is interesting to note that much larger chain and backbone fluctuations occur at 261 K than at 273 and 285 K, despite the occurrence of a glycerol conformational change in **1**b and **1a**, at 273

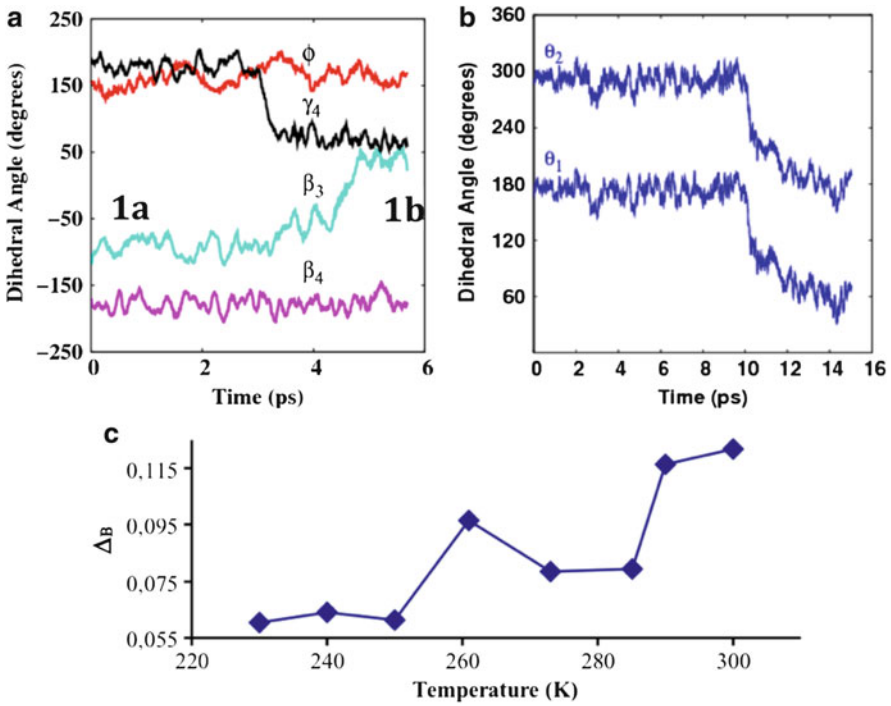


Fig. 8 Dynamics of DLPC: (**a**) Interchange between isomers 1a and 1b at 240 K; (**b**) interchange between two glycerol backbone isomers at $T = 273$ K ($\theta 1/\theta 2 = 180°/-60°$) to ($\theta 1/\theta 2 = 60°/180°$); (**c**) Berry parameter Δ_B evolution with temperature. Reproduced with permission from [111]

(illustrated in Fig. 8b) and 285 K, respectively. At higher temperatures, the number of kinks increases in both chains.

DLPC dynamics thus shows conformational changes associated with the ester groups and the glycerol backbone at various temperatures. This contrast with DMPC is also revealed by the evolution of Δ_B with temperature. The Δ_B value for DLPC (Fig. 8c) is weighted by the populations of isomers **1a** and **1b**. The Δ_B value jumps to 0.097 at $T = 261$ K, the temperature at which a trans to gauche transformation occurs in the alkyl chains. Increasing the temperature to $T = 273$ and 285 K, an unexpected reduction of Δ_B is observed. Despite the fact that the alkyl chains keep one gauche form as is the case at $T = 261$ K, it appears that the fluctuations of the intra-molecular atomic distances decrease. This peculiar behavior of an isolated DLPC molecule corresponds to the conclusions drawn from both DSC [124] and neutron diffraction measurements [123] attributing the broad area between 272.6 and 281 K to an unusual transition from a gel state to an "intermediate" liquid state (L_x), which shows "a substantial slowing of molecular motions" with respect to the first transition [123].

Using the critical Δ_B value of about 0.1 suggested for finite system order–disorder phase transitions, 261 K can be considered as a DLPC intra-molecular "melting"

temperature. The second intra-molecular transition to a more disordered state occurs at $T = 290$ K as follows from the Δ_B evolution. It is interesting to relate these two peaks of the Berry parameter function to the thermograms of the DLPC bilayers, which also display two phase transitions. The calculated 261 K is about 10 K lower than the experimental temperatures, indicated for the first gel to liquid ($P_{\beta'} \rightarrow L_\alpha$ or L_x) phase transition. The second peak of the Δ_B function is shifted upward by about 10–15 K when compared to the experimental temperatures of 275–280 K [124].

It is also worth noting that for the DMPC molecule, the distance-fluctuation criterion shows only one abrupt increase at T between 297 K and 301 K related to the main phase transition at $T = 295 \pm 1.5$ K [114, 115]. This confirms that C12 and C14 PC lipids have different dynamics at the molecular level. It is thus most likely to attribute the origins for the "intrinsically different" [123] gel to liquid phase transitions of PC bilayers with chains shorter than C13 to the substantially different dynamical behavior of the constituent lipid molecules.

From these molecular BOMD results, we conclude that the dynamical behaviors of the individual DMPC and DLPC molecules are largely preserved in their assembly properties that are not related to changes of the ensemble topology. This conclusion is confirmed by results from short BOMD trajectories (14 ps) of DLPC in 50 waters that show isomer interchanges occurring at 240 K and 273 K similar to those illustrated in Fig. 8 for the isolated molecule.

4.3 Activation of Triplet Dioxygen by Bio-inspired Cuprous Complexes

We now present a second example of application of the BODFT-MD methodology for the case of dioxygen activation by bio-inspired copper complexes. Despite the fact that dioxygen can lead to many radical oxygen species such as $HO^\bullet$, $O2^{\bullet -}$, or $HOO^\bullet$, that are harmful for the cells, aerobic organisms massively exploit dioxygen and direct its oxidative power toward many metabolites. Free dioxygen is, however, a triplet molecule while most organic substrates are singlet molecules. This so-called O_2 spin-mismatch implies that (1) reaction of $^3O_2$ with organic molecules is slow at ambient temperature, and (2) that catalysts must be used to overcome this kinetic barrier and to limit the possible formation of ROS in the cells. Biochemists have identified numerous enzymes allowing O_2 to be activated, most of them containing metal ions (primarily Fe and Cu) at their active sites [135]. In fact biochemical studies have revealed an impressively large panel of strategies employed by different enzymes to efficiently activate O_2 depending on the chemical nature of their substrate. In that regard the living realm provides a formidable source of inspiration for the development of innovative models [136].

We have been interested in the noncoupled copper-monooxygenases comprising Peptidylglycine α-Monooxygenase (PHM), Dopamine β-Monooxygenase (DβM), and Tyramine β-Monooxygenase (TβM) that catalyze the hydroxylation of C–H

bonds of C-terminal glycine-extended peptides of dopamine and of tyramine, respectively. Biochemical studies suggest that activation of the substrate C–H bond is promoted by a mononuclear cupric-superoxo species ($Cu(II)/O_2^{\bullet -}$) in the singlet state. The understanding of the process by which such a singlet species is formed from the interaction between a triplet O_2 molecule and the singlet cuprous complex is far from being satisfactory.

The kinetics of spin-forbidden reactions can be described by various theoretical models like those [137] derived from Transition State Theory that include nonadiabatic effects [138]. Basically these approaches follow the key concepts of TST (equilibrium between the reactant and the activated complex, semiclassical description of the molecular system) but take into account the probability the system has to hop from the initial quantum state (the triplet state here) to the final quantum state (the singlet) when reaching degeneracy between the two spin states. We have developed an alternative mixed quantum–classical (MQC) expression building on the work of Prezhdo and Rossky [139] and of Jasper and Truhlar [140] who explored the manifestations of decoherence effects in physico-chemical processes. In theoretical physics the term decoherence denotes processes by which a quantum system comes to behave classically upon interactions with its environment [141]. Applying these ideas to the case of chemical reactions involving two quantum states we obtained [142]:

$$k_{\mathrm{SC}}^{\mathrm{MQC}} = \nu \cdot \frac{8\mu^2}{\hbar^2 + 8\mu^2} \exp\left(-\frac{\Delta G^{\neq}}{k_{\mathrm{B}}T}\right). \tag{18}$$

In this expression μ is the product of the electronic coupling between the two quantum states (including the spin–orbit coupling terms) and of a characteristic decoherence time τ_{dec}. The latter reflects the mixed quantum–classical character of the molecular system: "*although intrinsically obeying quantum mechanical laws, behave semiclassically after a finite but nonzero amount of time* (τ_{dec})" and as such Eq. (18) may be "*regarded as a mixed quantum classical rate constant expression*." We recall here that this rate constant expression is expected to be valid for fast decoherence times (say <100 fs) that are shorter than the time characterizing the fluctuations of the diabatic energy gap.

Now the question is to devise algorithms for estimating τ_{dec} in molecular systems of significant size like those found in biology. A fully ab initio estimation would require performing simulations of the quantum density matrix of the system [143], which is out of the question for molecular systems composed of even tens of atoms. Coming back to the question of dioxygen activation by the PHM enzyme, we have considered the complexes depicted in Fig. 9 that mimic some of the main geometrical features of the parent enzymatic active site [144]. We followed the methodology proposed by Prezhdo et al. where decoherence is modeled by the decaying overlap between nuclear wave packets evolving on different electronic states. Contrary to these authors we have employed a BODFT-MD approach to perform our simulations. The computational protocol is illustrated in Fig. 10. In a

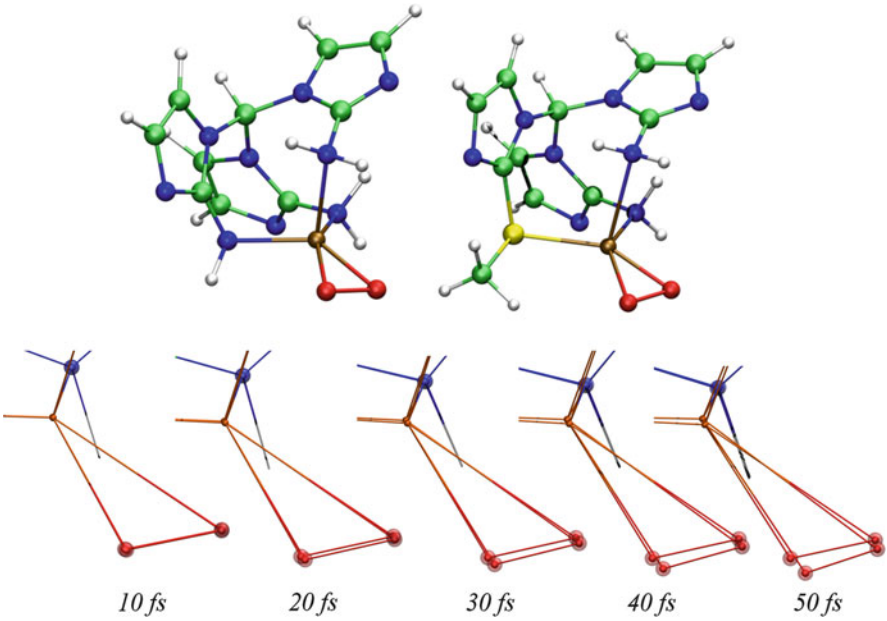


Fig. 9 DFT-based modeling of decoherence within a bio-inspired model of the active site on mononuclear copper enzymes. *Top*: two complexes investigated presenting an N_3 or an N_2S coordination sphere. *Bottom*: diverging motion of the copper and oxygen atoms on the fs timescale for one set of diverging trajectories. The transparent spheres represent the mass-dependent wave packet of each nucleus. *Color code*: Cu in *brown*, O in *red*, N in *blue*, C in *green*, S in *yellow*, and H in *white*

first step we performed a few tens of ps of BODFT-MD simulation in the NVT ensemble (at 300 K) on the singlet or on the triplet PES. In addition since the hopping probabilities between the two states (the quantities we wish to evaluate) reach a maximum in the region of degeneracy of two spin states, the BOMD simulations were biased by application of a harmonic constraint on the energy gap ($V_{\text{bias}} = k_{\text{bias}} (\Delta E_{\text{ST}} - 0)^2$). From this initial trajectory we extracted snapshots providing the starting conditions (positions and momentum of the nuclei and KS determinants) of so-called diverging trajectories which were run independently on the singlet and on the triplet surfaces (Step 2). A final step consists in the postprocessing of the data, that is, the estimation of the characteristic decoherence times and of the hopping probabilities obtained by integration of the Time-Dependant Schrödinger equation along the diverging BOMD [145].

The characteristic decoherence time for this copper complex has been found to be slightly below 10 fs for both complexes, whatever the direction of the reaction (S→T or T→S) Table 1. The computed hopping probabilities entering the transmission coefficient of the rate constant show an interesting feature: the introduction of a sulfur atom within the copper coordination sphere induces an increase of the hopping probability by a factor of more than 3. These effects are related to the

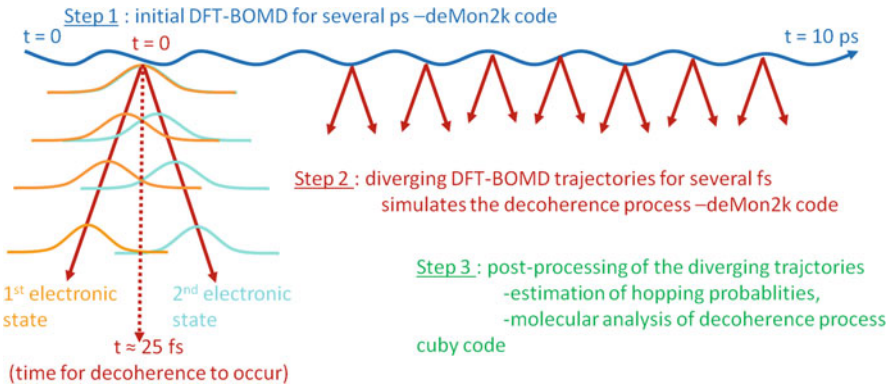


Fig. 10 Three-step computational protocol employed to estimate decoherence times with deMon2k. The *orange* and *cyan* Gaussian functions along the *left-hand side* set of diverging trajectory represent one nuclear wave packet on the singlet and triplet PES

Table 1 Decoherence time and average hopping probabilities for the N_2N- and N_2S-based $[CuO_2]^+$ adducts following the computational protocol described in [142] that is based on DFT MD simulations

		$[(N_2N)CuO_2]^+$	$[(N_2S)CuO_2]^+$
τ_{dec} (fs)	T → S	9.7	9.6
	S → T	9.2	9.4
P_h	T → S	0.035 (0.004)	0.115 (0.010)
	S → T	0.036 (0.004)	0.106 (0.018)

The numbers in parentheses are the standard deviations

higher electronic coupling computed for the $[(N_2S)CuO_2]^+$ adduct (ca. 160 cm^{-1}) compared to $[(N_2N)CuO_2]^+$ (ca. 60 cm^{-1}). It should be remarked, however, that these computations have been carried out in the gas phase for a bio-mimicking model of the real biological enzymatic active sites. New estimates of the above quantities with inclusion of the copper complex environment composed of protein residues and of water molecules are underway in our laboratories with the QM/MM computational approaches described in the next section.

5 DFT/MM-MD

Atomistic simulations of biomolecules remain one of the challenges of present day computational chemistry and biophysics. In such simulations one of the main difficulties is to give an accurate description of the structural and energetic aspects of the biochemical processes, while simulating very large systems for the lengths of time relevant to the studied problems. Particularly if one is concerned with properties that require high accuracy electronic structure, such as bond breaking/making, transition metal sites, magnetic properties, and sufficiently accurate

energetics (for example, relative energies of different conformers or transition states), the choice of methods is restricted to methods based on the calculation of the full electronic structure of the system. These electronic structure, or quantum mechanical (QM), methods are usually based on Density Functional Theory (DFT).

Despite the great development of computer and code performances, the size of the biosystem is still limited to about 1,000 atoms for geometry optimization and 150 atoms for reasonably long dynamics. Moreover, taking into account the solvent, when a global treatment is not sufficient, increases the computational demand even more.

As mentioned in the protocol for studying biomolecular systems in Sect. 1 of this review, if a simple cluster model is thought or shown to be inadequate, the model is most often extended by using hybrid QM/MM methods to incorporate the effects of the surrounding protein and solvent. Hybrid methods where the system is partitioned into a chemically active part, treated by first-principles QM methods and a larger environment, assumed to be chemically inert, which is modeled by classical molecular mechanics (MM), have been introduced in the late 1970s and have incorporated technological improvements over the years [146–152].

QM/MM has proven to be extremely successful in the study of biochemical, especially enzymatic, reactions and therefore has been widely applied in this field [153–165]. Numerous reviews have been devoted to overviews of QM/MM studies of biochemical reactions over the last 20 years [152, 166–175]. In this section, to introduce the methodology, we will first use the implementation of a QM/MM interface [176] between our computational chemistry program, deMon2k [70] and the widely used CHARMM [177] MM/MD package, as an example. For a more comprehensive treatment, please refer to our previous review on this topic [178].

In a typical QM/MM scheme, a system is usually divided into two subsystems: the QM subsystem treated by "high-level" QM methods and the MM subsystem treated by "low-level" force field-based methods. The boundary between these two subsystems distinguishes the QM region from the MM region. Ideally, partition of the system should not cut any covalent bonds to ensure the completeness of the QM subsystem. However, crossing covalent bonds is often unavoidable for large molecules, such as polymers and proteins. In order for a QM/MM calculation to mimic the real system the QM chemical structure has to be complete, i.e., no dangling bonds are permitted. An intuitive way to remedy a dangling bond is to cap it with an artificial atom, which gives rise to the so-called Link Atom approach [148]. In this approach, an additional atom, as a link between the QM and MM regions, is added to saturate the QM frontier atom at one end of the cut covalent bond. The link atom scheme is illustrated in Fig. 11. This link atom in most cases is a hydrogen atom because of its simplicity and practicality.

To explicitly calculate the QM subsystem (QMS)-MM subsystem (MMS) interaction on the QM level, an additive energy scheme has been widely used. It is formulated as

$$E(S) = E_{\mathrm{MM}}(\mathrm{MMS}) + E_{\mathrm{QM}}(\mathrm{QMS} + L) + E_{\mathrm{QM}}[(\mathrm{QMS} + L) - \mathrm{MMS}], \qquad (19)$$

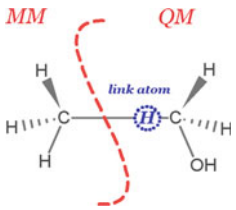


Fig. 11 Illustration of the link atom scheme using ethanol as an example, in which the methyl group is treated as the MM subsystem and the rest is QM. Reproduced with permission from [178]

where L denotes the link atoms. The last term in Eq. (19) accounts for the coupling between QM and MM subsystems. It consists of electrostatic and van der Waals interactions between QM and MM atoms, all of which will be detailed below.

The electrostatic interactions between QM and MM subsystems are often calculated based on the following formula:

$$\hat{H}_{\mathrm{QM-MM}} = -\sum_{i,m} \frac{q_m}{r_{im}} + \sum_{A,m} \frac{Z_A q_m}{R_{Am}}, \tag{20}$$

where q_m are the charges of MM atoms, Z_A the atomic number of QM atoms, i runs over all QM electrons, A over all QM atoms including link atoms, and m over all MM atoms. The first term is a one-electron operator and the second accounts for the nuclei–MM charge interaction. When acting on the QM wave function, Eq. (20) results in the electrostatic interaction between QM and MM subsystems as a portion of E_{QM} [(QMS + L)-MMS].

As one can see in Eq. (20), calculation of the first term on the DFT level can significantly slow down the entire QM/MM calculation, especially when there are a large number of MM charges. However, these charges cannot be simply ignored because long-range electrostatic effects can often be of considerable importance. To speed up the calculation of long-range electrostatics, we have introduced a multipole expansion-based approach [179] in the DFT framework. This has resulted in significant savings in time and plausible accuracy compared to full DFT calculation as shown in Figs. 12 and 13 and Table 2.

Besides electrostatic interactions, the QMS–MMS coupling also includes van der Waals (vdW) interactions. The vdW interaction is usually described by a Lennard–Jones 12-6 potential:

$$E_{ij} = \sum_{i,j} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right), \tag{21}$$

where i runs over QM atoms and j over MM atoms, and A and B are constants pertaining to atom types. In QM/MM calculations, when electrical embedding is used, the vdW interaction of QMS–MMS could be incorrect as the corresponding electrostatic interaction is not the parameterized point charge–point charge interaction any more. However, Cui and co-workers have tested three sets of vdW

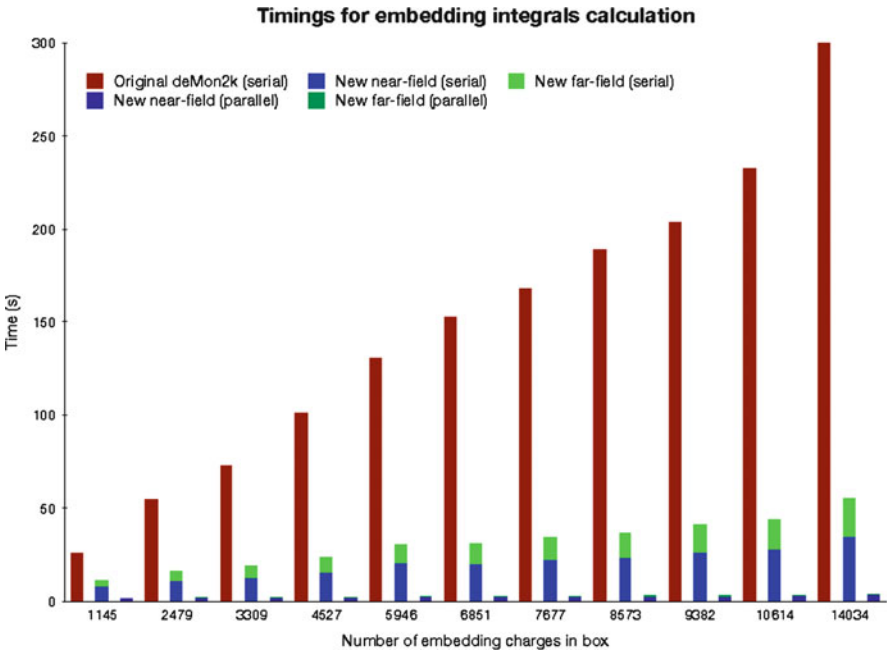


Fig. 12 Comparison of computational timings for the embedding integral calculation in small QM/MM systems

parameters and concluded that the QM/MM energetics were not sensitive to the vdW parameters and efforts to improve QM/MM accuracy should focus elsewhere [180].

Built on the energy scheme, the main focus of QM/MM calculations is often on geometry. In principle, the whole QM/MM system can be simultaneously optimized with a uniform optimizer using the QM/MM potential and gradient. Convergence, however, could be difficult to reach when the starting geometry is far from the minimum. Considering the different natures of QM and MM methods, optimization is sometimes easier to run separately for each subsystem on their respective levels. To this end, a macro/microiterative scheme has been implemented [181–187]. In this scheme, the optimization is driven by the QM optimizer. The MM subsystem is optimized to convergence with the QM part frozen, and this is termed as micro iteration. Thereafter, the QM region is optimized till convergence with the MM part frozen, and this is termed a macro iteration. These two iterations alternate until the whole system is fully optimized. A diagram of this scheme is shown in Fig. 14.

In addition to optimized geometries, which are minima on the potential energy surface, curvature on the free energy surface is also of interest in many QM/MM studies. Therefore, efficient calculations of free energy are required. One of the most widely used techniques to calculate the free energy in QM/MM methods is free energy perturbation (FEP) theory. Originally developed in MM, this FEP

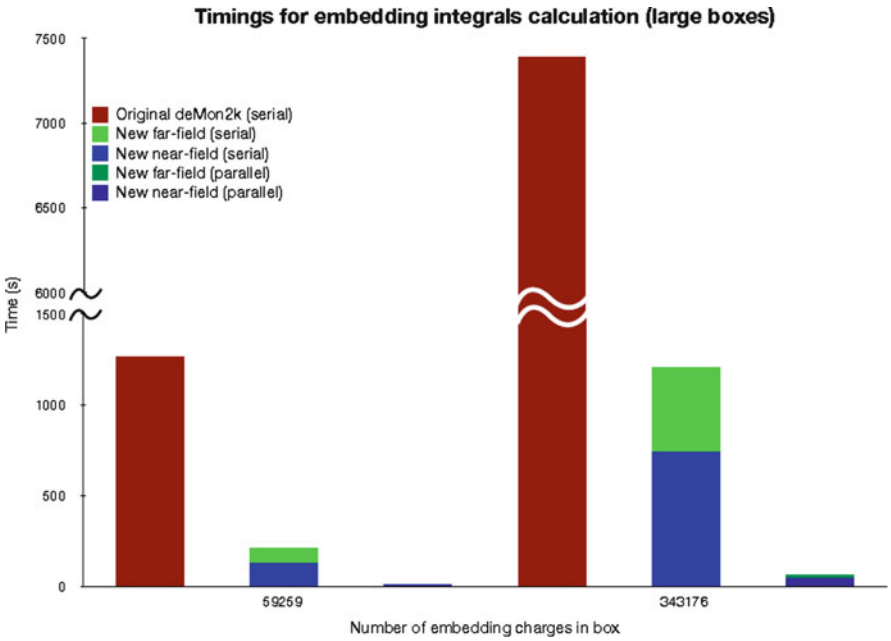


Fig. 13 Comparison of computational timings for the embedding integral calculation in large QM/MM systems

Table 2 Comparison of the total energies (a.u.) of the embedded systems

Embedding charges in box	Original deMon2k (serial)	New code (serial)	New code (parallel 16 processors)
1145	−4880.219852928	−4880.219852922	−4880.219853263
2479	−4880.197609272	−4880.197609273	−4880.197609798
3309	−4880.151039302	−4880.151039303	−4880.151039189
4527	−4880.342973267	−4880.342973268	−4880.342973155
5946	−4880.292634368	−4880.292634369	−4880.292633976
6851	−4880.359550818	−4880.359550818	−4880.359550949
7677	−4880.314753835	−4880.314753835	−4880.314754031
8573	−4880.672278586	−4880.672278586	−4880.672278983
9382	−4880.728606709	−4880.728606709	−4880.728607116
10614	−4880.785583226	−4880.785583226	−4880.785583612
14034	−4880.580437797	−4880.580437798	−4880.580438027
59259	−4880.428695928	−4880.428695929	−4880.428695680
343176	−4880.335242269	−4880.335242272	−4880.335241941

technique relates the free energy difference between state A and state B to the potential energy difference between these two states [188]. If we assume that the perturbation required to transform system A to system B is small (<2 kT), it can be shown that:

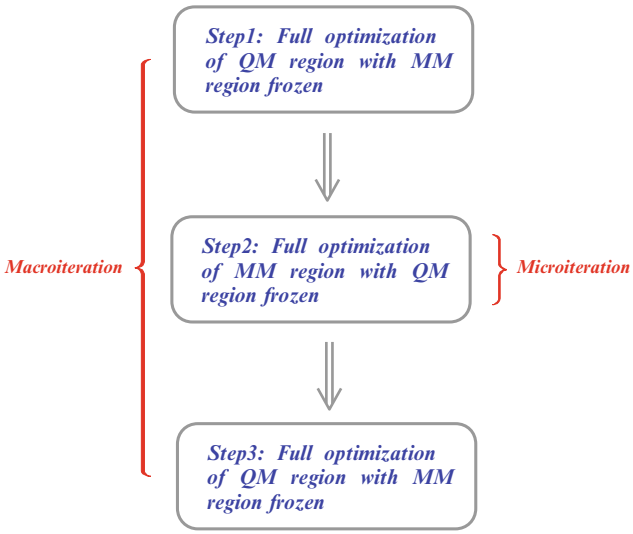


Fig. 14 Schematic of the macro–micro iteration scheme for QM/MM geometry optimization

$$\Delta G_{(A->B)} = -k_{\mathrm{B}}T \ln \left\langle \exp\left(-\frac{E_B - E_A}{k_{\mathrm{B}}T}\right)\right\rangle_A, \tag{22}$$

where the potential energy difference between two states is weighted by the energy of the initial state. In case of larger perturbations, one can always use additional windows to connect starting and ending points of this perturbation. The free energy is a path-independent property of the system and can be evaluated regardless of how "alchemical" the perturbation path may be.

In the formalism of QM/MM, a parameter λ is introduced to relate the two states. Thus, the Hamiltonian describing the system which is changing from state A to state B during FEP calculations can be re-written as

$$H(\lambda) = \lambda E_A + (1 - \lambda)E_B, \tag{23}$$

where E_A and E_B represent distributions of states for the two end points of the perturbation. The dynamics of two replicas of the system each corresponding to end points with $\lambda = 0$ and $\lambda = 1$ is treated explicitly and simultaneously and appropriate weighting is applied to reconstitute the Hamiltonian for intermediate windows.

We have applied this method in our QM/MM implementation [176] to the study of ion solvation and the mechanisms of selectivity in ion channels or ion-coupled transporters. Ion selectivity has been extensively studied at the MM level with both classical and polarizable force fields. However, classical simulations may be compromised by their inability to account for charge transfer and electronic polarization, thought to be critical for ion binding to proteins. We have extended QM/MM FEP to studies of Na^+/K^+ solvation and selectivity by water clusters with

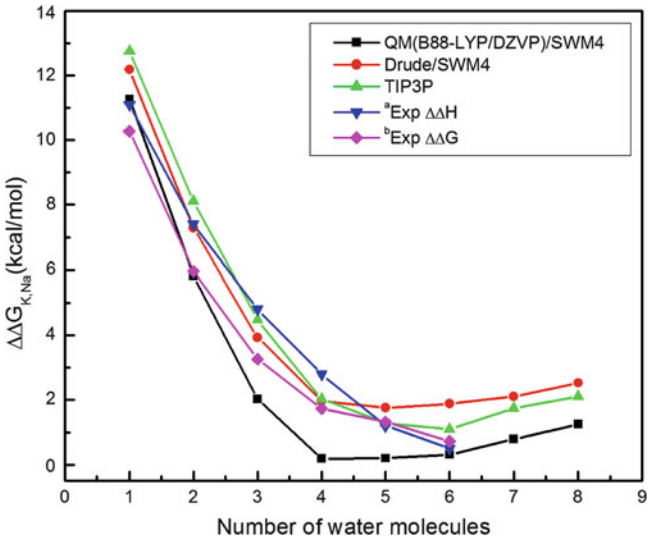


Fig. 15 Relative (to the bulk) free energy of selectivity for Na^+/K^+ in water clusters as a function of cluster size. Reproduced with permission from [176]

variable numbers of ligands, for which results are illustrated in Fig. 15. In [176] QM/MM FEP calculations have been performed for ion-water clusters with different numbers of water molecules, where the ion was treated as the QM region using B88-LYP with the DZVP basis set implemented in deMon2k, and waters were represented by the polarizable Drude force field implemented in CHARMM. The same approach can be used for selectivity calculations in more complex biological systems. Also, the close connection between QM/MM calculations and those with polarizable force fields developed to account for electronic effects is visible. Therefore, the polarizable force fields can be sufficient for some cases.

Hybrid QM/MM methodology looks particularly attractive for reproducing solvent effects when a discrete representation of the solvent molecules is needed, as in the example above, and for many other bio-processes, such as microsolvation at enzymatic sites or at metal binding sites in peptides, proton transfer, free-energy calculations. In fact, comparison of full QM and QM/MM results for sufficiently small peptides in water has been performed recently, including an analysis of the full MM results [189, 190]. These comparative studies led to the following main conclusions (1) the solvation pattern around most groups of the dipeptides is similarly described by the three methods (QM, QM/MM, and MM); (2) solvation patterns around terminal $–CO2^-$ and $–NH_3^+$ groups are poorly reproduced in MM simulations; (3) solvation patterns of the full QM results are well reproduced, at a lower cost, by the QM/MM calculation involving MM waters; (4) DFT/MD computed chiroptical properties of a solvated glycine molecule are comparable to those obtained with the much more expensive coupled cluster CC2 method.

The QM/MM methodology is also powerful in providing accurate structural and free energy changes in solution at a moderate computational cost for relatively large models treated with a QM method in the presence of a solvent calculated explicitly using an MM approach. Several simulation techniques based on ab initio quantum mechanics have been proposed for this purpose. The quantum mechanical free energy (QM-FE) approach mentioned above consists of optimizing the reaction path in the QM system in the gas phase and of applying free-energy perturbation (FEP) along the preoptimized gas-phase reaction path with the inclusion of the QM/MM interactions that are treated classically [191]. This method may have some inaccuracy if the gas-phase reaction path does not yield a good description of the reaction path in solution. The QM(ai)/MM method developed for the study of condensed phase reaction processes approximates the potential-energy surface from ab initio QM with an empirical valence bond (EVB) potential [192]. At the same time, QM/MM-FE was proposed for the simulation of enzymatic reaction processes [193–195]. In this method, the dynamics of the QM and MM systems are assumed to be independent of each other, although the two subsystems interact energetically, and the QM subsystem motions are taken to be harmonic. In order to avoid the dependency of the reaction path on the choice of the initial conformation, the reaction path is thus calculated on the potential of mean force (PMF) surface of the QM/MM system, yielding the QM/MM minimum free-energy path (QM/MM-MFEP). The free-energy gradients are used to perform the path optimization on the free energy surface.

The QM/MM-MFEP method is easily combined with the deMon2k version that includes MM and QM/MM energy and gradient evaluations inside the program core. The QM/MM-MFEP method using this deMon2k version has been tested on glycine in aqueous solutions, namely the path from neutral to zwitterionic glycine in a water droplet (53, 108, and 182 water molecules). The glycine molecule is the QM subset surrounded by SPC MM waters.

Being an amino acid prototype, glycine solvation in water has been studied in the literature using different methods and aiming at different goals. Indeed, neutral glycine only exists in the gas phase, whereas its zwitterionic form is mainly present in water solution [196]. Qualitatively, the larger zwitterion stability is found using a dielectric solvation model [197]. Several studies searching for the minimum number of water molecules that are needed to stabilize the zwitterions concluded to various values ranging from 2 to 7 [197–199]. Statistical averaging from finite-temperature simulations has been performed using MM, QM/MM, full QM methods, and the continuum solvent model [189, 190, 200–204]. In most of these studies, detailed hydration shells and solute–solvent interactions of the zwitterionic form are investigated. The free energy difference, ΔG, between neutral and zwitterionic solvated species has also been studied, despite the fact that neutral aqueous glycine is not observed experimentally. A ΔG value of 8.7 kcal mol^{-1}, close to the experimentally proposed value of about 7.0–7.5 kcal mol^{-1} [196], was obtained using the QM gaseous energies of neutral and zwitterionic forms of glycine and its radical (calculated with the accurate CBS-QB3 procedure [205]) and MM solvation energies (free energy cycle) [200]. Full QM simulations using the Car and

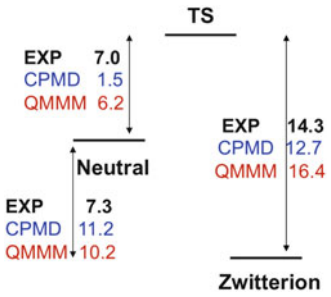


Fig. 16 Schematic representation of the free energy proton transfer path from neutral to zwitterionic glycine or vice versa in 53 waters. The QM/MM-MFEP results are presented in *red*, compared with experimental values in *black* [205, 206] and CPMD calculations in *blue* [201]

Parrinello (CPMD) approach [201] showed variable hydration numbers along the proton transfer path and led to a zwitterion–neutral glycine ΔG of 11.2 kcal mol^{-1} and a proton transfer barrier, ΔG^*, of 12.7 kcal mol^{-1}. The advantage of performing the dynamics of the peptide in a full QM or in a QM/MM calculation is to provide the reaction path and the barrier estimate. Combining the experimental barrier value of 14.3 kcal mol^{-1} for the zwitterion to neutral glycine reaction [206] and the energy difference between the neutral and zwitterionic glycine, one concludes that the experimental neutral glycine ΔG^* is about 7 kcal mol^{-1}. The full QM CPMD calculations [201] thus lead to an underestimated free energy barrier of 1.5 kcal mol^{-1} for the neutral glycine.

Figure 16 illustrates the results obtained for our QM/MM simulation of glycine in 53 MM waters using the QM/MM-MFEP methodology.

First, the neutral glycine was placed into a spherical droplet of 53 waters, cut from a drop of 2500 waters equilibrated at 300 K. The glycine molecule BOMD was performed using the DFT methodology (PBE98-LYP-D, DZVP/GEN-A2 bases), whereas the MD of the water molecules used the SPC force field [207].

The system was left to evolve at 300 K using a Nose–Hoover thermostat in the NVT ensemble. After about 100 ps, the proton of the Ha acidic glycine moved to N. As also mentioned in the CPMD study [201], a decrease of the water coordination to glycine was observed before and during the transfer (2 waters bound to the acidic OH and to the N atom moved farther away). The proton transferred directly to the amine N without any intermediate bonding with water.

In order to calculate the free-energy changes following the PMF description of the proton transfer, we used the QM/MM-MFEP method [195]. The reaction path was divided into six steps and the reaction coordinate along this path was chosen to be the distance between the acidic proton Ha to the amine nitrogen N. Step 0 was the neutral and step 5 the zwitterionic glycine. The Ha$\cdots$N distances of steps 1–4 were taken from snapshots of BOMD-MD dynamics.

The calculated free energy changes from neutral to zwitterionic glycine are obtained from the calculation of the potential of mean force of the proton transfer in water, with the contribution of the water degrees of freedom being ensemble-averaged out.

The optimization of the proton transfer is carried out in a discretized representation, using the free-energy perturbation formula for the chain of conformations (steps)

$$A(r_{\mathrm{QM}}) = A_{\mathrm{ref}} - \frac{1}{\beta} Ln\langle \exp(-\beta[E(r_{\mathrm{QM}}, r_{\mathrm{MM}}) - E_{\mathrm{ref}}(r_{\mathrm{MM}})])\rangle_{E_{\mathrm{ref}}, r_{\mathrm{MM}}}, \tag{24}$$

where $E_{\mathrm{ref}}(r_{\mathrm{MM}})$ is the total energy of the system expressed in terms of the coordinates of the QM and MM subsystems, at the QM reference geometry. The QM geometry and the MM ensemble are mutually dependent. The first QM geometry (step 0) is set as the reference point on the PMF. Its free energy is A_{ref}. The $E_{\mathrm{ref}}(r_{\mathrm{MM}})$ values for each step were obtained by optimizing the full QM/MM system keeping the constraint of the H$\cdots$N distance fixed for every step from 1 to 4. Finally, sampling of the MM space was provided by about 40 ps MM water dynamics keeping frozen the full QM subsystem at the reference geometry of the step. It has been verified that doubling this MM sampling time did not change the $A - A_{\mathrm{ref}}$ values.

In order to take into account the sensitivity of the energetic results with respect to the orbital basis set extension, the $E_{\mathrm{ref}}(r_{\mathrm{MM}})$ values obtained from DZVP/GEN-A2 calculations were followed by one single point energy calculation with the cc-pVTZ basis [208], and the auxiliary GEN-A2* set including up to g functions.

Figure 16 shows that both theoretical methods overestimate the neutral–zwitterion free energy difference, whereas the QM/MM methodology leads to a better estimate of the neutral to transition state free energy difference. However, increasing the number of waters from 53 to 108 brings the results into very close agreement with both experimental values. The effects of glycine dilution are presently under study [209].

We conclude this section with some remarks on the RNA polymerase enzyme for which cluster calculations have been discussed in Sect. 2.2.2. Extensions of that work to include more of the protein and solvent environment through QM/MM calculations and free-energy perturbation theory are in progress, but they are not yet ready for publication. Instead, we will give a glimpse of some related work from the literature on DNA polymerases, which show many of the same essential features as their RNA counterparts.

DNA polymerases and RNA polymerases help replicate DNAs and RNAs with considerably high fidelity, selecting the matched deoxynucleoside triphosphate (dNTP) and NTP over the mismatched dNTPs and NTPs, and catalyzing the nucleotidyl transfer reaction. In the QM/MM studies of DNA pol IV and T7 pol, Zhang and coworkers [210, 211] utilized the pseudobond approach to cap the QM subsystem, the micro/macro iteration scheme to optimize the geometry, the reaction coordinate driven method for the reaction path search, and the free energy perturbation method for free energy calculations along the path. They found that the nucleophilic attack was the rate-limiting step and the initial proton transfer was assisted by water.

Figure 17 shows a comparison of potential energy and free energy profiles for this reaction. The potential energy profile is in fact reminiscent of the results for our cluster models of RNA polymerase shown in Fig. 5. The warning about the very large effects of entropy should be clear from Fig. 17.

Finally, we mention that the initial proton transfer has also been a focus in the study of RNA polymerase II. Ramos and coworkers [212] adopted an ONIOM method in their QM/MM study of RNA polymerase II. They tested several proton transfer possibilities and concluded that none of the adjacent residues or the RNA primer could be the deprotonation agent due to the high activation barriers. As an alternative solution, they introduced a hydroxide ion in the reaction site to be the proton acceptor, which resulted in a reasonable activation barrier. However, the source of the hydroxide ion was not discussed and it is not clear to us whether this hydroxide-ion assisted mechanism is in fact plausible.

6 Constrained-DFT

6.1 Methodological Background

The resolution of the KS equations gives access to the electronic density of the adiabatic ground state of the molecular system of interest. In other words, one finds by the SCF procedure a relaxed electronic density that corresponds to a minimum of the DFT energy. It is, however, possible that one desires to investigate properties of electronic densities that *do not* correspond to the electronic ground states. The purpose of constrained DFT is precisely to enable the user to impose a particular atomic charge or a spin density while solving the KS equations. Although artificial, this procedure can have considerable interest for the modeling of (bio-)chemical systems. We will provide various examples of applications in the second part of this section but we can already mention the question of electron transfer reactions within the context of the Marcus Theory [213]; this theory relies on the definition of phenomenological states that cannot be obtained by standard DFT (which provides adiabatic states). cDFT is a means to define such ad hoc diabatic states. As another example we remark that a strength of computational chemistry is to give access to systems or situations that are not amenable to direct experimental measurements. Chemical trends like the electronic effects of substituents on a reactive energy profile can be investigated by various methods including cDFT or other methods by imposing artificial constraints on the chemical system of interest. As an example of the usefulness of other procedures we refer to [214] where cDFT is not used but the nuclear charge of nitrogen atoms is varied to modify its electronegativity and indirectly the energy cost for activating an alkyl C–H bond.

In the next paragraphs we explain how this objective is achieved with cDFT and how an efficient implementation within the context of ADFT can be devised. Our aim is not to provide here a delineated mathematical derivation of the

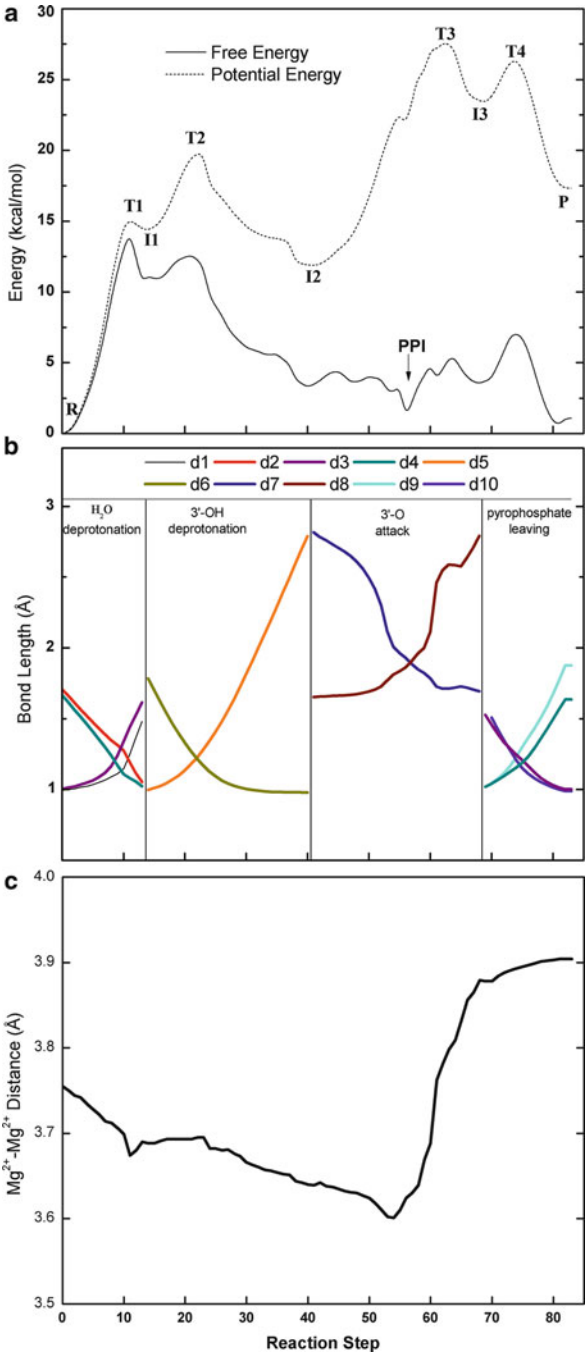


Fig. 17 Potential energy and free energy profiles for the nucleotidyl addition reaction of a DNA polymerase. Reproduced with permission from [211]

formalism, but rather to introduce the key aspects of the method so as to provide the reader with a view of the potential applications of cDFT. For more mathematical details, the reader is referred to recent reviews on the method [215, 216]. The foundations of constrained DFT are to be found in the pioneering works of Gunnarsson and Lundqvist [217] and of Dederich et al. in the early 1980s [218]. Wesolowski, Muller, and Warshel also developed a related, though algorithmically different, constrained DFT approach in the 1990s [219] with remarkable success for the modeling of condensed phase proton transfers, electron transfer reactions [220], and S_N2 reactions [221], in connection with the Empirical Valence Bond approach [222]. We will not, however, cover this methodology in this chapter and the reader is referred to the references for more details. Continuing with the principal works on cDFT formalisms we finally mention that of Wu and Van Voohris in 2005 [223]. In 1976 Gunnarsson and Lundqvist showed that in addition to the true ground state it was possible to obtain relaxed electronic densities of the particular molecular symmetries. A few years later Dederich et al. proposed a generalization of this idea that used a Lagrange Multiplier to constrain the DFT energy:

$$\varepsilon[\rho, \lambda_c] = \min_\rho \max_{\lambda_c} \left(E[\rho] + \lambda_c \left[\int \rho(\mathbf{r}) w(\mathbf{r}) \mathrm{d}\mathbf{r} - N_C \right] \right). \tag{25}$$

In this equation E is the nonconstrained DFT energy whose mathematical expression has been given in Sect. 2.1, $w(r)$ is a weight function that defines the constraining property, and N_C is the set-point supplied by the user. For example to constrain N_C electrons to occupy a volume Ω the weight function would equal 1 inside Ω and zero everywhere else. Both $w(r)$ and N_C are thus user-defined terms. We will come back later to the practical definition of $w(r)$ within the LCAO framework. For the moment we focus on the Lagrange multiplier λ_c that needs to be determined. To emphasize the role of this term in the formalism it is useful to write down the set of modified KS equations. These are obtained by differentiation of the cDFT energy with respect to the MO coefficients under the orthonormalization constraint.

$$\left(-\frac{1}{2}\nabla^2 + v_{\mathrm{ext}}(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \mathrm{d}\mathbf{r}' + v_{\mathrm{xc}}(\mathbf{r}) + \lambda_c w(\mathbf{r}) \right) \psi_i = \varepsilon_i \psi_i(\mathbf{r}) \forall i. \tag{26}$$

Within the LCGTO formalism employed in deMon2k the elements of the constrained KS matrix are given by

$$K_{\mu\nu} \equiv \left(\frac{\partial \varepsilon_{\mathrm{ADFT}}}{\partial P_{\mu\nu}} \right)_x = H_{\mu\nu} + \sum_{\bar{k}} \langle \mu\nu | \bar{k} \rangle (x_{\bar{k}} + z_{\bar{k}}) + \lambda_c W_{\mu\nu}. \tag{27}$$

It is apparent from these equations that the product $\lambda_c W$ acts like a supplementary potential felt by the electrons. Its role is to drive the convergence of the

SCF procedure toward a relaxed density that fulfills the desired constraint. It must be noted, however, that λ_c is not known beforehand (contrary to W or N_C) and must be determined. In the older applications of cDFT, it was customary to scan over possible values of λ_c to find the correct one. This was obviously a cumbersome strategy that probably restricted the application of cDFT to only a few fields of research. In addition, a critical point when solving the cDFT KS equations is related to the fact that the Coulomb and XC potentials on the one hand and the $\lambda_c w$ potential on the other are inter-dependent. This makes the resolution of the constrained KS equations difficult. Wu and Van Voorhis proposed an efficient solution to these difficulties building on Optimized Potential Theory. By examining the stationary conditions of the cDFT energy with respect to λ_c, these authors proved that, for a *given* Coulomb and Exchange Correlation (XC) potential there is a unique λ_c that leads to an electronic density fulfilling the desired constraint. In addition, the correct λ_c corresponds to a maximum of the energy. The first and second derivatives of the energy with respect to λ_c are easily obtained from the KS orbital coefficients so that an automatic algorithm can now be implemented for searching the correct λ_c. In practice this optimization is conducted according to the combination of steepest descents and a Newton–Raphson algorithm for best efficiency [224]. Then to address the difficulty of the coupling between the Coulomb, XC potential, and the constraint, Wu and Van Voorhis proposed a dual-loop strategy whereby the determination of KS orbital energies is decoupled from the determination of λ_c. The algorithm is illustrated in Fig. 18. At every SCF step λ_c is optimized keeping the Coulomb and XC potential constants ("inner loop" in the figure). Once the correct λ_c is determined the constrained density matrix is used to calculate the new Coulomb and XC potentials for the next SCF step. Our experience told us that the convergence of the inner loop is usually not problematic in terms of CPU time, especially if a guess for λ_c from the previous SCF iteration can be provided. At global convergence the SCF procedure provides a converged electronic density fulfilling the desired constraint. The most demanding task in ADFT/cDFT remains the numerical calculation of the XC matrix elements, but since these terms remain constant when converging the inner loop the necessity to optimize λ_c is not lethal for the computational efficiency of the ADFT framework. In other words, cDFT is compatible with ADFT and its implementation with the min–max algorithm [225]. The calculation of the contribution of the constraining term to the energy gradients with respect to the nuclear positions has also been implemented in deMon2k, allowing geometry optimization, frequency analysis, and BOMD with cDFT [226]. However, we identified a computational bottleneck in the calculation that becomes problematic for molecular systems of a few tens of atoms (ca. 70–80 atoms). This is due to the involvement of the orbital density in the constraining term and therefore to the calculation of products of GTOs in the numerical evaluation of the energy derivative terms. We are pursuing our efforts to derive a fully cADFT formalism where the constraint would apply on the auxiliary function density instead of on the orbital density.

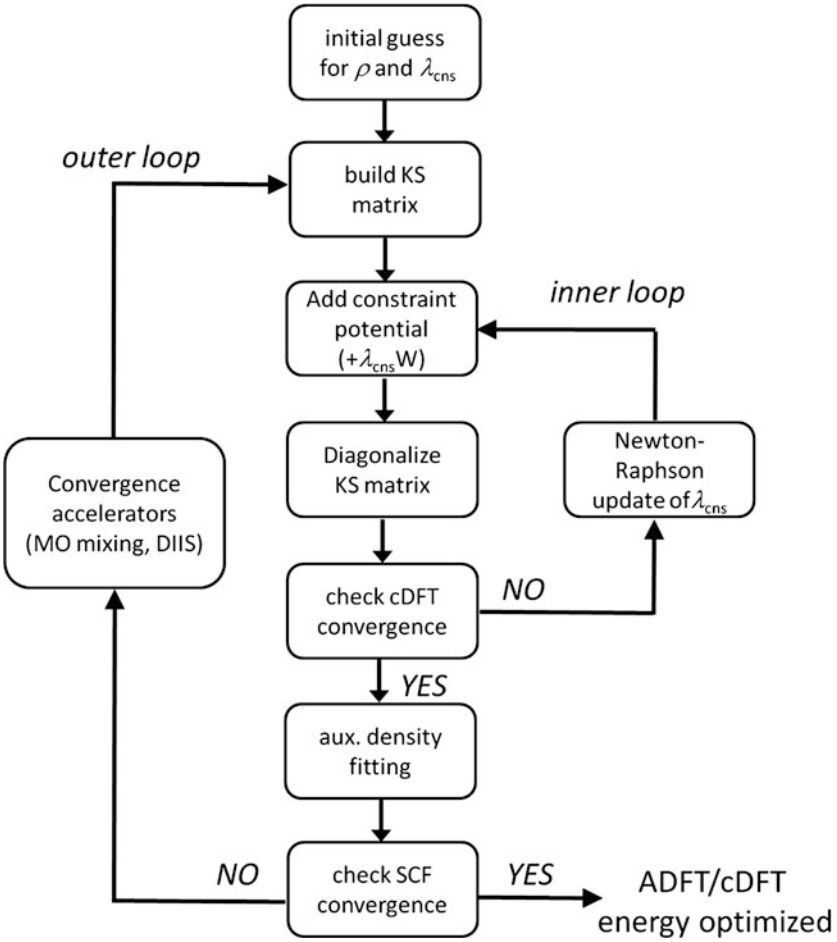


Fig. 18 Dual-loop SCF algorithm for solving the modified ADFT/cDFT KS equations in deMon2k. Reproduced with permission from [225]

6.2 cDFT and Population Analyses

So far we have presented cDFT as a means to impose a charge or a spin density on a group of atoms. However, we have not specified how such charges are calculated in practice. It is well known that atomic charges cannot be strictly defined in quantum mechanics and, as a corollary, that multiple population analysis (PA) approaches can be developed to reach this goal. A central point PAs have to address is to define a criterion for assigning each fraction of the electronic density of every point in space to the individual atoms. Some PAs realize this partition with criteria relying on the KS molecular orbital coefficients. The Mulliken [227], Löwdin [228], and the more sophisticated Natural Bond Orbital [229, 230] approaches are examples of

such methods. Other PAs work on the electronic density itself, a strategy that seems most appropriate to the philosophy of DFT. The Voronoi Deformation Density (VDD) [231], Becke [232], Hirshfeld [233], and the more recent iterative Hirshfeld [234–236] and iterative Stockholder [237] are examples of PAs falling into this second category. For the sake of completeness, we also mention the Atoms-in-Molecules approach developed by Bader that uses a topological analysis of the electron density, but we will come back to this method in Sect. 7. The mathematical expressions for calculating the atomic charges for some of these approaches are given in Table 3.

In this table, P is the density matrix and S is the overlap matrix of the atomic GTOs (μ and ν). Z_A is the nuclear charge of atom A. For the Becke, Hirshfeld, and VDD approaches, the electron density is integrated numerically over a grid of points. The Voronoi cell of atom A is the ensemble of points closer to A than to any other atom. The Becke cell is defined similarly except that a smoothing function is applied at the borders between the cells. ρ^A is the electron density of the neutral atom A.

Coming back to cDFT one needs to make a choice of a particular PA to define the charge of the atoms to be constrained. This choice determines the mathematical definition of the W matrix elements. So far the PAs that have been used for cDFT are the Mulliken, Löwdin, Becke, Hirshfeld, and VDD approaches. Our personal experience has led us to prefer the Hirshfeld scheme that generally provides more reasonable results from a chemical point of view. However, apart from the Mulliken approach that is usually not satisfying, cDFT with the Löwdin, or Becke method is usually satisfactory. A future worthwhile objective would be to devise algorithms for making cDFT compatible with more sophisticated PA approaches such as the iterative Hirshfeld, the iterative Stockholder, or Bader's Atoms-In-Molecules approach. Many of these approaches are available in deMon2k [238], but using them for cDFT is however nontrivial because the mathematical definition of the W matrix would then depend on the density matrix itself (at variance with the cases reported in Table 3). Said otherwise, the W matrix would vary between successive SCF steps like the Coulomb and XC matrices, making the dual-loop procedure of Fig. 18 partially obsolete. We now report on the applications of cDFT to biologically relevant electron-transfer systems. A further example will be given in the section on "interpretational-DFT."

6.3 *Modeling Electron Transfer Reactions*

Along with hydrogen transfers (in the form of protons, hydrogens, or hydrides), electron transfer is one of the most fundamental chemical processes in biological systems. Electron transfers are found in the respiratory chain, in photosynthesis, and in uncountable metabolic pathways. Enzymes such as the family of oxidoreductases have evolved to catalyze chemical reactions such as the functionalization of C–H bonds. Inner-sphere or outer-sphere (Long Range) electron

Table 3 Five population approaches commonly used in cDFT computations to define the weight matrix elements

PA	Definition of the charge in the volume Ω	Weight matrix elements	cDFT weight w_i
Mulliken	$Q_A^{\text{Mulliken}} = Z_A - \sum_{\mu \in A} (PS)_{\mu\mu}$	$W_{\mu\nu} = S_{\mu\nu}$ if μ and $\nu \in A$ $W_{\mu\nu} = \frac{1}{2} S_{\mu\nu}$ if μ or $\nu \in A$ $W_{\mu\nu} = 0$ if μ and $\nu \notin A$	–
Löwdin	$Q_A^{\text{Löwdin}} = Z_A - \sum_{\mu \in A} (S^{1/2} P S^{1/2})_{\mu\mu}$	$W_{\mu\nu} = \sum_{\lambda} S_{\mu\lambda}^{1/2} S_{\lambda\nu}^{1/2}$	–
Becke	$Q_A^{\text{Becke}} = Z_A - \int_{\text{Becke cell of } A} \rho(r) \mathrm{d}r$	$W_{\mu\nu} = \sum_i a_i w_i \mu(r_i) \nu(r_i)$	$w_i^{\text{becke}} = 1.0$ inside the cell $w_i^{\text{becke}} = 0.0$ outside the cell + smoothing function
Hirshfeld	$Q_A^{\text{Hirshfeld}} = Z_A - \int \rho(r) \frac{\rho_A^a(r)}{\sum_{\text{all atoms}} \rho^a(r)} \mathrm{d}r$	$W_{\mu\nu} = \sum_i a_i w_i \mu(r_i) \nu(r_i)$	$w_i = \frac{\sum_{\text{atoms in } A} \rho^a(r_i)}{\sum_{\text{all atoms}} \rho^a(r_i)}$
VDD	$Q_A^{\text{VDD}} = \int_{\text{Voronoi cell of } A} \left[\rho(r) - \sum_{\text{all atoms}} \rho^a(r) \right] \mathrm{d}r$	$W_{\mu\nu} = \sum_i a_i w_i \mu(r_i) \nu(r_i)$	$w_i^{\text{VDD}} = 1.0$ inside the cell $w_i^{\text{VDD}} = 0.0$ outside the cell

transfers are usually at play within the catalytic cycles of these enzymes [239]. However, the precise molecular strategies enabling efficient catalysis are often elusive. Computational chemistry, and in particular DFT, provides a valuable means to complement biochemical studies to reach a detailed understanding of these biological processes. From a theoretical point of view, the most popular conceptual framework for dealing with the kinetics of ET is the Marcus Theory (MT) [213]. The theory assumes the definition of two phenomenological electronic states corresponding to situations where the electron to be transferred is localized on the reductant (initial redox state) or on the oxidant (final redox state). These diabatic electronic states correspond to the empirical description familiar to the chemist. In MT two limiting regimes are considered. First when the quantum coupling between the two diabatic states is strong as in the case of inner-sphere ET, an adiabatic rate constant can be employed:

$$k_{\mathrm{ET}}^{A} = \nu \cdot \exp\left(-\frac{(\Delta G^{\circ} + \lambda)^2}{4\lambda k_{\mathrm{B}} T}\right), \tag{28}$$

where ΔG° is the free energy of the reaction, λ is the reorganization energy, and ν is the effective frequency along the reaction coordinate. Actually the diabatic energy gap is often a good choice of reaction coordinate [240]. The other terms have their usual meaning. When the coupling between the two diabatic states is weak, as in the case of a long-range ET or of symmetry-forbidden ET, a nonadiabatic expression applies

$$\mathrm{k}_{\mathrm{ET}}^{NA} = \frac{2\pi}{\hbar} \frac{1}{\sqrt{4\pi\lambda k_{\mathrm{B}} T}} |H_{DA}|^2 \exp\left(-\frac{(\Delta G^{\circ} + \lambda)^2}{4\lambda k_{\mathrm{B}} T}\right). \tag{29}$$

where H_{DA} is the electronic coupling factor between the initial and final redox states (respectively $|\varphi_D\rangle$ and $|\varphi_A\rangle$):

$$H_{DA} = \langle \varphi_D | H_{\mathrm{el}} | \varphi_A \rangle, \tag{30}$$

H_{el} being the electronic Hamiltonian.

The cDFT formalism of Dederich et al. and Wu et al. has been employed by various groups to model the two diabatic states and to compute the terms entering the adiabatic and nonadiabatic rate constant expressions [241–244]. In particular, cDFT is one of the few DFT-based approaches that enable the evaluation of the electronic coupling terms between the diabatic states with no need for the adiabatic representation [242, 245]. In addition, the possibility of performing BOMD simulations on cDFT potential energy surfaces makes possible the evaluation of free energy terms. In this sense, cDFT holds great promise for the modeling of ET reactions at the DFT level, especially thanks to the computational advantages of the cADFT framework. We already mentioned the cDFT method of Wesolowski et al. that also has found numerous applications in the field [246, 247].

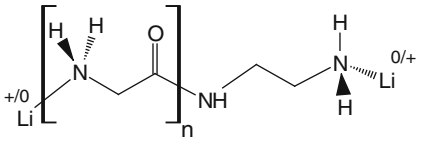


Fig. 19 $Li_2^{+/0}$ redox pairs separated by polypeptide chains of increasing lengths investigated in [226]. Reproduced with permission from [226]

In [226] we investigated the decay of the electronic coupling as a function of the donor to acceptor distance for the electron transfer between two lithiums separated by a polyglycine peptide of increasing length (Fig. 19). The simulations were run in the gas and aqueous phases thanks to the QM/MM scheme described in Sect. 5. Note that the BOMD simulations were biased so as to sample exclusively the molecular configurations corresponding to the degeneracy of the two diabatic states. Note also that the lithium and its first coordination shell were fixed during the simulations to control the *D* to *A* distance. In this work we also reported the first estimates of the characteristic decay times of the Franck–Condon factors by means of cDFT/MM MD simulations. This term reflects the time the molecular system remains in the Franck–Condon region that is active for the ET. It can be computed by the following expression:

$$\tau_{\mathrm{dec}} = \left[\left\langle \sum_n \frac{1}{2a_n \hbar^2} \left(F_{1n} - F_{2n}\right)^2 \right\rangle_T \right]^{-1/2}, \tag{31}$$

where F_{xn} and $a_{n,}$ respectively, denote the forces felt by the *n*th degree of freedom in the electronic state 1 or 2 and the typical width of the wave packet associated with this degree of freedom. Actually Eq. (31) provides a means to evaluate the decoherence time in the context described in Sect. 4.3, but under a short time approximation and within the high temperature limit [139]. Note that this approximate equation enables an estimation of the decoherence time without resorting to computationally expensive diverging trajectories (see Sect. 4.3).

The graphs shown in Fig. 20 are in line with common knowledge of the electronic coupling decaying exponentially with the *D* to *A* distance. Characteristic decays of 0.52 and 0.58 $\mathring{A}^{-1}$ are found for the gas phase and aqueous phase, respectively, in good agreement with the experimental values reported in the literature [248] for LRET rates through β-strands. The characteristic decoherence times are found to be around 4.5 and 1.5 fs in the gas and aqueous phases, but quite interestingly independent of the bridge length. This constant trend can be understood by looking at Eq. (31). The atoms that contribute the most to decoherence are those which feel the different forces in the two redox states. Therefore, the molecular fragments that are far from the donor and from the acceptor do not appreciably contribute to decoherence. And, similarly, when the bridge length is increased the new intervening atoms contribute less and less to decoherence. Recall, however, that in these simulations the most contributing atoms were fixed

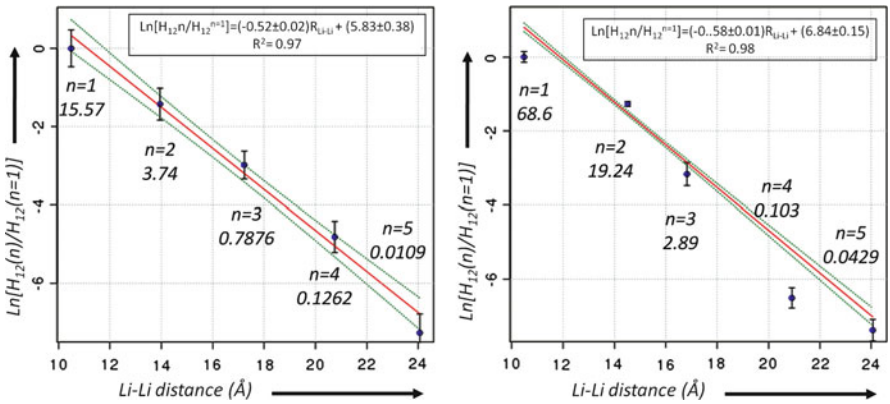


Fig. 20 Linear decay of the electronic coupling logarithm for the inter-lithium ET for different donor-to-acceptor distances in the gas phase (*Left*) and in water (*Right*). Each point represent an average over 1 ps of BOMD simulation in the NVT ensemble ($T = 300$ K). The values of the average H_{DA} on the graphs are given in cm^{-1}. Reproduced with permission from [226]

in space for practical reasons so that the actual characteristic decoherence time for this reaction is probably much below 1 fs. In fact the study reported in [226] is more to be seen as a proof of principles of the use of cDFT/MM MD approaches for the estimation of decoherence times in large molecular systems. Investigations of decoherence times in real enzymatic systems are currently being performed in our laboratories using this methodology and will be reported in due course.

6.4 *Other Applications of cDFT*

Apart from the modeling of electron transfer reactions, the cDFT formalism has been applied to many other chemical problems (see [215]). So far we have considered constraints applied to the atomic charges only; however, the cDFT formalism also allows one to enforce the spin density on specific atoms. Therefore, the method provides a valuable tool to prepare electronic guesses for (nonconstrained) SCF-DFT computation with desired spin-density properties. Such a possibility is certainly of great added value for the modeling of complex electronic structures like those encountered in bio-inorganic metal complexes. In addition, the ability of cDFT to provide trustworthy values of the magnetic coupling constants within polymetallic clusters such as those found in many metalloenzymes holds great promise for future applications in this field [249].

To conclude this section, and although we do not cover the TD-DFT and Tight-Binding DFT methods in this chapter we mention that extensions of the cDFT ideas of Dederich et al. have been developed for these methods. More generally, there is little doubt that further development of cDFT in the coming years will allow its application to numerous biologically relevant systems.

7 Interpretational-DFT

As seen throughout this chapter, DFT represents a powerful computational tool. The primary outputs of DFT computations are the energies of the molecules. Then, when coupled to geometry optimizers or BOMD engines, the stable conformers or Boltzmann ensembles can be determined, giving access, for example, to the free energy balance of a chemical reaction (or its free energy of activation). In addition, the calculations of spectroscopic properties such as, for example, the infra-red or Raman vibrational spectra provide valuable supplements to understand the chemical systems of interest. Properties such as polarizabilities, hyperpolarizabilities [250], NMR (Nuclear Magnetic Resonance) shieldings [251, 252], or NQR (Nuclear Quadrupole Resonance) [253] energy splittings, to mention but a few, are other quantities that can be efficiently computed with DFT. Having recognized the immense merits of DFT it is also worth noting that the outcomes of these first principles computations are not always directly connected to some of the most fruitful concepts of chemistry like the notion of electronegativity, of hardness, or even the notion of chemical bonding. Many efforts have been spent to fill this gap leading for example to the so-called conceptual DFT [254, 255] framework or the topological analysis.

In this section, we have chosen to focus on the interpretational tools that are based on the topological analysis of well-defined functions like the electronic density, ρ and the Electron Localization Function (ELF) η. The former is a direct output of a DFT computation or of experimental measurements while the latter is a function introduced by Becke and Edgecombe as a mean to measure the electron localization in molecular systems [256]

$$\eta = \frac{1}{\left[1 + \left(\frac{D}{D_{\mathrm{h}}}\right)^2\right]}. \tag{32}$$

with

$$D = \frac{1}{2}\sum_i |\nabla\varphi_i|^2 - \frac{1}{8}\frac{(\nabla\rho)^2}{\rho}, \tag{33}$$

$$D_{\mathrm{h}} = \frac{3}{10}(3\pi^2)^{5/3}\rho^{5/3}, \tag{34}$$

where D and D_{h} represent the curvature of the electron pair density for electrons with identical spins for the system under study, and a homogeneous electron gas with the same density, respectively. The ELF function can be interpreted as a measure of the Pauli repulsion in the atomic or molecular space and reflects the probability of finding opposite spin electron pairs. The function η is bounded between 0 and 1.

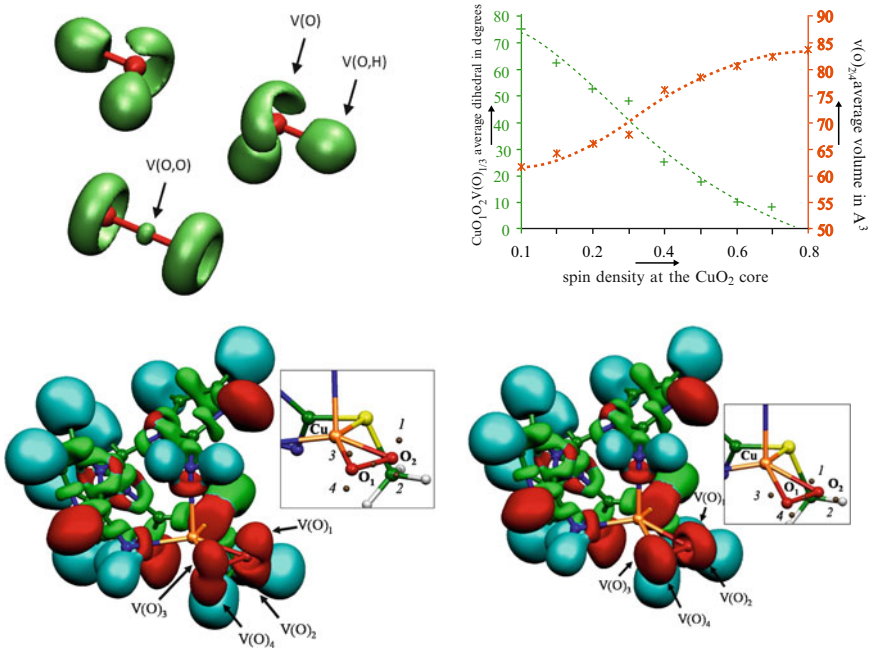


Fig. 21 *Upper left*: ELF isosurface (0.7) for the $O_2(H_2O)_2$ system showing the valence basins within the molecular fragments. *Upper right*: cDFT/ELF analysis showing the rotation of the dioxygen lone pairs and their volumes changes as a function of the spin density imposed at the CuO_2 core. *Bottom*: Orientation of the dioxygen lone pairs revealed by the ELF topological analysis in the singlet (*Left*) and triplet (*Right*) spin states within minimal model of noncoupled monooxygenases. (isosurface for ELF = 0.8). *Insets*: zooms on the CuO_2 core showing the localization of the O_2 valence basin attractors. *Color code*: *red* for nonbonding (lone pair), *green* for bonding, and *light blue* for protonated bond. Reproduced with permission from [259]

Topological analysis, as defined in the context of the theory of dynamical systems, was pioneered by Bader in the 1970s in the case of the electronic density. The approach proposed by Bader performs a partition of the electrons within the molecular space into basins associated with each atom. Therefore, it provides a mathematically grounded way to define an atom within a molecule. Following a similar approach, Savin and Silvi showed that the topological analysis of the ELF leads to the identification of molecular basins that can be connected to the chemical intuitive picture of bonds, lone pairs, π-systems [257]. For example, the topological analysis allows one to distinguish between core basins (labeled $C(X)$, X being a nuclei) encompassing the nuclei and valence regions. These latter then split into bonding basins connected to more than one atomic center (labeled $V(X, Y, \ldots)$ $X, Y \ldots$ being nuclei) and nonbonding basins (lone pairs) connected to only one atomic center (labeled $V(X)$). These valence basins match closely the nonbonding (lone pairs) and bonding domains of Gillespie's VSEPR (Valence Shell Electron Pair Repulsion) model [258]. The ELF isosurface ($\eta = 0.7$) is showed in Fig. 21 for a

triplet water molecule surrounded by two water molecules. The dioxygen lone pairs (V(O)) and the covalent O–O bond are clearly visible.

As a further illustration we take here the example of the $[(N_2S)CuO_2]^+$ complex which was presented in Sect. 4 (Fig. 10). As already mentioned the transition between the triplet and singlet spin state is a prerequisite in the dioxygen activation process. Recently, we reported a systematic study of the ELF topology on a series of metalled dioxygen complexes (including, Fe, Ni, Pd, and Cu atoms) [259]. We observed that upon spin transition the dioxygen lone pairs experienced a systematic rotation around the metal-O_2 plane making them more or less accessible toward exogenous substrate. In addition, an important variation of the lone pair spatial extension was observed upon this transition. This trend could be related to the molecular orbital diagrams of the complexes. We also reported in this communication, the ELF topological analysis on constrained DFT densities where the spin density of the CuO_2 triad was constrained between 0.1 and 0.8 thereby confirming our initial interpretation, leading to the conclusion that metalloenzymes could use spin state control to tune the regioselectivity of substrate oxidations.

More generally the ELF topological analysis has been found to be a powerful interpretational framework to rationalize the regioselectivity of numerous types of chemical reactions [260, 261]. We already mentioned the approach of Bader that is based on the electronic density itself. Other scalar functions such as the electrostatic potential [262] have also been the subject of topological analysis.

8 Conclusions and Perspectives

In this review we have seen examples representing three main types of biomolecules and biomodels: proteins, nucleic acids, and lipids. Although it is not always classified as a "biomolecule," water represents a fourth major type and it has been center stage in many of the examples we have treated. For each of these classes of biomolecules, DFT has played a major role in our work and in that of other workers. But this is not a one-man show; we have also shown how DFT can be combined with molecular dynamics, either in the Born–Oppenheimer-MD approach or in hybrid QM/MM methodologies. And we have shown examples that go beyond strictly Kohn–Sham DFT in the use of constraints that allow connections with other theories and concepts, notably the Marcus theory of electron transfer. Finally, we have given a glimpse of some of the tools that can be used to analyze and interpret the DFT-based computations.

While there has been very remarkable recent progress in DFT towards descriptions of ever more complex models for biological processes, we do not view DFT as a panacea. Much work remains to develop both the functionals necessary for quantitative work and the appropriate suite of modeling tools within which DFT can perform. More and more the lines are blurring between DFT (and quantum chemistry in general) and the more conventional molecular dynamics and statistical mechanics methodologies that use empirical force fields. Much more

work remains to develop powerful and efficient hybrid methodologies that will, in the fullness of time, allow more complete and sophisticated models combining all four classes of biomolecules. We find the prospect of combining proteins, lipids, nucleic acids, and water (not to neglect ions and small molecules...) into a multiscale suite of methodologies to be a very exciting one. Much progress has been made and we hope that our review will help to introduce new workers into this vibrant field.

Acknowledgments We are grateful to the numerous coworkers involved in the deMon Developers consortium for their contributions to the development of the deMon software (http://www.demon-software.com). Operating grants from the Natural Sciences and Engineering Research Council of Canada are gratefully acknowledged as is the provision of ample computational resources from Compute Canada/WestGrid.

References

1. Hohenberg P, Kohn W (1964) Inhomogeneous electron gas. Phys Rev B 136:B864
2. Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. Phys Rev 140:1133
3. Thomas LH (1927) The calculation of atomic fields. Proc Camb Philos Soc 23:542
4. Fermi E (1927) Un metodo statistico per la determinazione di alcune prioprietà dell'atomo. Rend Accad Naz Lincei 6:602
5. Salahub D, Goursot A, Weber J, Köster A, Vela A (2005) Applied density functional theory and the deMon codes 1964-2004. In: Dykstra C, Frenking G, Kim K, Scuseria G (eds) Theory and applications of computational chemistry: the first forty years. Elsevier, Amsterdam, p 1079
6. Perdew J, Burke K, Ernzerhof M (1996) Generalized gradient approximation made simple. Phys Rev Lett 77:3865
7. Perdew J, Burke K, Ernzerhof M (1997) Generalized gradient approximation made simple (vol 77, pg 3865, 1996). Phys Rev Lett 78:1396
8. Perdew J, Burke K, Ernzerhof M (1998) Comment on "Generalized gradient approximation made simple"—Reply. Phys Rev Lett 80:891
9. Becke A (1993) Density-functional thermochemistry. 3: The role of exact exchange. J Chem Phys 98:5648
10. Stephens P, Devlin F, Chabalowski C, Fritsche M (1994) Ab-initio calculation of vibrational absorption and circular dichroism spectra using density-functional force fields. J Phys Chem 98:11623
11. Zhao Y, Truhlar D (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. Theor Chem Acc 120:215
12. Handy NC, Cohen AJ (2001) Left-right correlation energy. Mol Phys 99:403
13. Conradie J, Ghosh A (2006) Iron(III)-nitro porphyrins: theoretical exploration of a unique class of reactive molecules. Inorg Chem 45:4902
14. Conradie MM, Conradie J, Ghosh A (2011) Capturing the spin state diversity of iron(III)-aryl porphyrins OLYP is better than TPSSh. J Inorg Biochem 105:84
15. Swart M (2008) Accurate spin-state energies for iron complexes. J Chem Theory Comput 4:2057

16. Han WG, Noodleman L (2008) Structural model studies for the peroxo intermediate P and the reaction pathway from P -> Q of methane monooxygenase using broken-symmetry density functional calculations. Inorg Chem 47:2975
17. Dickson R, Becke A (1993) Basis-set-free local density-functional calculations of geometries of polyatomic molecules. J Chem Phys 99:3898
18. Frisch MJT, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Mennucci B, Petersson GA, Nakatsuji H, Caricato M, Li X, Hratchian HP, Izmaylov AF, Bloino J, Zheng G, Sonnenberg JL, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Montgomery JA Jr, Peralta JE, Ogliaro F, Bearpark M, Heyd JJ, Brothers E, Kudin KN, Staroverov VN, Kobayashi R, Normand J, Raghavachari K, Rendell A, Burant JC, Iyengar SS, Tomasi J, Cossi M, Rega N, Millam NJ, Klene M, Knox JE, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Martin RL, Morokuma K, Zakrzewski VG, Voth GA, Salvador P, Dannenberg JJ, Dapprich S, Daniels AD, Farkas O, Foresman JB, Ortiz JV, Cioslowski J, Fox DJ (2009) Gaussian 09. Gaussian Inc., Wallingford, CT
19. Valiev M, Bylaska E, Govind N, Kowalski K, Straatsma T, Van Dam H, Wang D, Nieplocha J, Apra E, Windus T, de Jong W (2010) NWChem: a comprehensive and scalable open-source solution for large scale molecular simulations. Comp Phys Commun 181:1477
20. Shao Y, Molnar L, Jung Y, Kussmann J, Ochsenfeld C, Brown S, Gilbert A, Slipchenko L, Levchenko S, O'Neill D, DiStasio R, Lochan R, Wang T, Beran G, Besley N, Herbert J, Lin C, Van Voorhis T, Chien S, Sodt A, Steele R, Rassolov V, Maslen P, Korambath P, Adamson R, Austin B, Baker J, Byrd E, Dachsel H, Doerksen R, Dreuw A, Dunietz B, Dutoi A, Furlani T, Gwaltney S, Heyden A, Hirata S, Hsu C, Kedziora G, Khalliulin R, Klunzinger P, Lee A, Lee M, Liang W, Lotan I, Nair N, Peters B, Proynov E, Pieniazek P, Rhee Y, Ritchie J, Rosta E, Sherrill C, Simmonett A, Subotnik J, Woodcock H, Zhang W, Bell A, Chakraborty A, Chipman D, Keil F, Warshel A, Hehre W, Schaefer H, Kong J, Krylov A, Gill P, Head-Gordon M (2006) Advances in methods and algorithms in a modern quantum chemistry program package. Phys Chem Chem Phys 8:3172
21. Koster A, Geudtner G, Calaminici P, Casida M, Dominguez V, Flores-Moreno R, Gamboa G, Goursot A, Heine T, Ipatov A, Janetzko F, Jd C, Reveles J, Vela A, Zuniga-Gutierrez B, Salahub D (2011) deMon2k, Version 3. The deMon Developers, Cinvestav, Mexico
22. Dunlap B, Connolly J, Sabin J (1979) Some approximations in applications of X-alpha theory. J Chem Phys 71:3396
23. Koster A, Reveles J, del Campo J (2004) Calculation of exchange-correlation potentials with auxiliary function densities. J Chem Phys 121:3417
24. Siegbahn P, Himo F (2011) The quantum chemical cluster approach for modeling enzyme reactions. WIREs Comput Mol Sci 1:323
25. Wang S, Hu P, Zhang Y (2007) Ab initio quantum mechanical/molecular mechanical molecular dynamics simulation of enzyme catalysis: the case of histone lysine methyl transferase SET/79. J Phys Chem B 111:3758
26. Siegbahn P, Blomberg M (2010) Quantum chemical studies of proton-coupled electron transfer in metalloenzymes. Chem Rev 110:7040
27. Blomberg M, Siegbahn P (2010) Quantum chemistry as a tool in bioenergetics. Biochim Biophys Acta Bioenerg 1797:129
28. Siegbahn P (2011) Recent theoretical studies of water oxidation in photosystem II. J Photochem Photobiol B Biol 104:94
29. Siegbahn P, Borowski T (2011) Comparison of QM-only and QM/MM models for the mechanism of tyrosinase. Faraday Disc 148:109
30. Blomberg MA, Siegbahn PM (2012) The mechanism for proton pumping in cytochrome c oxidase from an electrostatic and quantum chemical perspective. Biochim Biophys Acta 1817:495

31. Siegbahn P, Himo F (2009) Recent developments of the quantum chemical cluster approach for modeling enzyme reactions. J Biol Inorg Chem 14:643
32. Bushnell D, Cramer P, Kornberg R (2002) Structural basis of transcription: amanitin–RNA polymerase II cocrystal at 2.8 Å resolution. Proc Natl Acad Sci USA 99:1218
33. Florian Brueckner JO, Cramer P (2009) A movie of the RNA polymerase nucleotide addition cycle. Curr Opin Stuct Biol 19:294
34. Steitz TA (1998) Structural biology—a mechanism for all polymerases. Nature 391:231
35. Castro C, Smidansky E, Maksimchuk KR, Arnold JJ, Korneeva VS, Gotte M, Konigsberg W, Cameron CE (2007) Two proton transfers in the transition state for nucleotidyl transfer catalyzed by RNA- and DNA-dependent RNA and DNA polymerases. Proc Natl Acad Sci USA 104:4267
36. Zhang Y, Zhu R, Zhang R, de la Lande A, Salahub D (2012) On the mechanism of the nucleotidyl transfer reaction catalyzed by yeast RNA polymerase II (Unpublished)
37. Zhu R, Janetzko F, Zhang Y, van Duin ACT, Goddard WA, Salahub DR (2008) Characterization of the active site of yeast RNA polymerase II by DFT and ReaxFF calculations. Theor Chem Acc 120:479
38. Rapcewicz K, Ashcroft NW (1991) Fluctuation attraction in condensed matter—a nonlocal functional approach. Phys Rev B 44:4032
39. Andersson Y, Langreth D, Lundqvist B (1996) Van der Waals interactions in density-functional theory. Phys Rev Lett 76:102
40. Wesolowski T, Tran F (2003) Gradient-free and gradient-dependent approximations in the total energy bifunctional for weaky overlapping electron densities. J Chem Phys 118:2072
41. Hesselmann A, Jansen G (2003) Intermolecular dispersion energies form time-dependent density functional theory. Chem Phys Lett 367:778
42. Misquitta A, Jeziorski B, Szalewicz K (2003) Dispersion energy from density-functional description of monomers. Phys Rev Lett 91:033201
43. Dobson J (1998) Prospects for a van der Waals density functional. Int J Quantum Chem 69:615
44. Rydberg H, Lundqvist B, Langreth D, Dion M (2000) Tractable nonlocal correlation density functionals for flat surfaces and slabs. Phys Rev B 62:6997
45. Osinga V, Ginsbergen SV, Snijders J, Baerends E (1997) Density functional results for isotropic and anisotropic multipole polarizabilities and C-6, C-7 and C-8 van der Waals dispersion coefficients for molecules. J Chem Phys 106:5091
46. Becke A, Johnson E (2006) Exchange-hole dipole moment and the dispersion interaction: high-order dispersion coefficients. J Chem Phys 124:014104
47. Becke A, Johnson E (2005) A density-functional model of the dispersion interaction. J Chem Phys 122:154104
48. Johnson E, Becke A (2006) A post-Hartree-Fock model of intermolecular interactions: inclusion of higher-order corrections. J Chem Phys 124:174104
49. Kurita N, Inoue H, Sekino H (2003) Adjustment of Perdew-Wang exchange functional for describing van der Waals and DNA base-stacking interactions. Chem Phys Lett 379:161
50. Walsh T (2005) Exact exchange and Wilson-Levy correlation: a pragmatic device for studying complex weakly-bonded systems. Phys Chem Chem Phys 7:443
51. Zhang Y, Salahub D (2007) A reparametrization of a meta-GGA exchange-correlation functional with improved descriptions of van der Waals interactions. Chem Phys Lett 436:394
52. Hult E, Rydberg H, Lundqvist B, Langreth D (1999) Unified treatment of asymptotic van der Waals forces. Phys Rev B 59:4708
53. Dion M, Rydberg H, Schroeder E, Langreth D, Lundqvist B (2004) Van der Waals density functional for general geometries. Phys Rev Lett 92:246401
54. Langreth D, Lundqvist B, Chakarova-Kack S, Cooper V, Dion M, Hyldgaard P, Kelkkanen A, Kleis J, Kong L, Li S, Moses P, Murray E, Puzder A, Rydberg H, Schroeder E, Tonhauser T (2009) A density functional for sparse matter. J Phys Condens Matter 21:084203

55. Roman-Perez G, Soler J (2009) Efficient implementation of a van der Waals density functional: application to double-wall carbon nanotubes. Phys Rev Lett 103:096102
56. Wu Q, Yang W (2002) Empirical correction to density functional theory for van der Waals interactions. J Chem Phys 116:515
57. Ahlrichs R, Penco R, Scoles G (1977) Intermolecular forces in simple systems. Chem Phys 19:119
58. Hepburn J, Scoles G (1975) A simple but reliable method for the prediction of intermolecular potentials. Chem Phys Lett 36:451
59. Mooij W, Duijneveldt Fv, Rijdt Jv-D-vd, Eijck Bv (103) Transferable ab initio intermolecuular potentials. 1. Derivation from methanol dimer and trimer calculations. J Phys Chem A 103:9872
60. Elstner M, Hobza P, Frauenheim T, Suhai S, Kaxiras E (2001) Hydrogen bonding and stacking interactions of nucleic acid base pairs: a density-functional-theory-based treatment. J Chem Phys 114:5149
61. Zimmerli U, Parrinello M, Koumotsakos P (2004) Dispersion corrections to density functionals for water-aromatic interactions. J Chem Phys 120:2693
62. Grimme S (2004) Accurate description of van der Waals complexes by density functional theory including empirical corrections. J Comput Chem 25:1463
63. Jurecka P, Cerny J, Hobza P, Salahub D (2007) Density functional theory augmented with an empirical dispersion term. Interaction energies and geometries of 80 noncovalent complexes compared with ab initio quantum mechanics calculations. J Comput Chem 28:555
64. Loerigk L, Grimme S (2011) Efficient and accurate double-hybrid-meta-GGA density functionals—evaluation with the extended GMTKN30 database for general main group thermochemistry, kinetics, and noncovalent interactions. J Chem Theory Comput 7:291
65. Becke A, Johnson E (2007) Exchange-hole dipole moment and the dispersion interaction revisited. J Chem Phys 127:154108
66. Tang K, Toennies P (1984) An improved simple model for the van der Waals potential based on universal damping functions for the dispersion coefficients. J Chem Phys 80:3726
67. Steinmann S, Corminboeuf C (2010) A system-dependent density-based dispersion correction. J Chem Theory Comput 6:1990
68. Steinmann S, Corminboeuf C (2011) Comprehensive benchmarking of a density-dependent dispersion correction. J Chem Theory Comput 7:3567
69. Bondi A (1964) van der Waals volumes and radii. J Phys Chem A 68:441
70. Köster AM, Calaminici P, Casida ME, Flores-Moreno R, Geudtner G, Goursot A, Heine T, Ipatov A, Janetzko F, del Campo JM, Patchkovskii S, Reveles JU, Vela A, Salahub DR (2006) deMon2k. The International deMon Developers Community, Cinvestav-IPN, México
71. Becke A (1988) Density-functional exchange-energy approximation with correct asymptotic behavior. Phys Rev A 38:3098
72. Perdew J, Wang Y (1992) Accurate and simple analytic representation of the electron-gas correlation energy. Phys Rev B 45:13244
73. Goursot A, Mineva T, Kevorkyants R, Talbi D (2007) Interaction between n-alkane chains: applicability of the empirically corrected density functional theory for van der Waals complexes. J Chem Theory Comput 3:755
74. Zhang Y, Yang W (1998) Comment on "generalized gradient approximations made simple". Phys Rev Lett 80:890
75. Tao J, Perdew J, Staroverov V, Scuseria G (2003) Climbing the density functional ladder: nonempirical meta-generalized gradient approximation designed for molecules and solids. Phys Rev Lett 91:146401
76. Lee C, Yang W, Parr R (1988) Development of the cole-Salvetti correlation-energy formula into a functional of the electron density. Phys Rev B 37:785
77. Murray E, Lee K, Langreth D (2009) Investigation of exchange energy density functional accuracy for interacting molecules. J Chem Theory Comput 5:2754

78. Tsuzuki S, Honda K, Uchimaru T, Mikami M (2006) Estimated MP2 and CCSD(T) interaction energies of n-alkane dimers at the basis set limit: comparison of the methods of Helgaker et al. and Feller. J Chem Phys 124:114304
79. Grimme S (2006) Seemingly simple stereoelectronic effects in alkane isomers and the implications for Kohn-Sham density functional theory. Angew Chem Int Ed 45:4460
80. Krishnamurty S, Stefanov M, Mineva T, Begu S, Devoisselle J, Goursot A, Zhu R, Salahub D (2008) Density functional theory-based conformational analysis of a phospholipid molecule (dimyristoyl phosphatidylcholine). J Phys Chem B 112:13433
81. Goursot A, Mineva T, Krishnamurty S, Salahub D (2009) Structural analysis of phosphatidyl choline lipids and glycerol precursors. Can J Chem 87:1261
82. Newton I (1687) Philosophiae Naturalis Principia Mathematica
83. Alder BJ, Wainwright TE (1959) Studies in molecular dynamics. 1 General method. J Chem Phys 31:459
84. Rahman A, Stillinger F (1971) Molecular dynamics study of liquid water. J Chem Phys 55:3336
85. Levitt M, Warshel A (1975) Computer simulation of protein folding. Nature 253:694
86. Car R, Parrinello M (1985) Unified approach for molecular-dynamics and density-functional theory. Phys Rev Lett 55:2471
87. Barnett RN, Landman U (1993) Born-Oppenheimer molecular dynamics simultations of finite systems—structure and dynamics of $(H_2O)_2$. Phys Rev B 48:2081
88. Hauser H, Pascher I, Pearson RH, Sundell S (1981) Preferred conformation and molecular packing of phosphatidylethanolamne and phosphatidylcholine. Biochim Biophys Acta 650:21
89. Hauser H, Guyer W, Pascher I, Skrabal P, Sundell S (1980) Polar group conformation of phosphatidylcholine—effect of solvent and aggregation. Biochemistry 19:366
90. Hauser H, Pascher I, Sundell S (1988) Preferred conformation and dynamics of the glycerol backbone in phospholipids—an NMR and X-ray single-crystal analysis. Biochemistry 27:9166
91. Hong M, Schmidt-Rohr K, Zimmermann H (1996) Conformational constraints on the headgroup and sn-2 chain of bilayer DMPC from NMR dipolar couplings. Biochemistry 35:8335
92. Bruzik KS, Harwood JS (1997) Conformational study of phospholipids in crystalline state and hydrated bilayers by C-13 and P-31 CP-MAS NMR. J Am Chem Soc 119:6629
93. Aussenac F, Laguerre M, Schmitter JM, Dufourc EJ (2003) Detailed structure and dynamics of bicelle phospholipids using selectively deuterated and perdeuterated labels. H-2 NMR and molecular mechanics study. Langmuir 19:10468
94. Pearson RH, Pascher I (1979) Molecular structure of lecithin dihydrate. Nature 281:499
95. Mantsch HH, McElhaney RN (1991) Phospholipid phas-transitions in model and biological-membranes as studied by infrared spectroscopy. Chem Phys Lipids 57:213
96. Lewis R, McElhaney RN (1998) The structure and organization of phospholipid bilayers as revealed by infrared spectroscopy. Chem Phys Lipids 96:9
97. Mendelsohn R, Davies MA, Brauner JW, Schuster HF, Dluhy RA (1989) Quantitative-determination of conformational disorder in the acyl chains of phospholipid-bilayers by infrared spectroscopy. Biochemistry 28:8934
98. Casal HL, McElhaney RN (1990) Quantitative determination of hydrocarbon chain conformational order in bilayers of saturated phosphatidylcholines of various chain lengths by fourier-transform infrared spectroscopy. Biochemistry 29:5423
99. Hubner W, Mantsch HH (1991) Orientation of specifically c-13=O labelled phosphatidylcholine multilayers from polarized attenuated total reflection FT-IR spectroscopy. Biophys J 59:1261
100. Thirumoorthy K, Nandi N, Vollhardt D, Oliveira ON (2006) Semiempirical quantum mechanical calculations of dipolar interaction between dipyridamole and dipalmitoyl phosphatidyl choline in Langmuir monolayers. Langmuir 22:5398

101. Weiner SJ, Kollman PA, Nguyen DT, Case DA (1986) An all atom force-field for simulations of proteins and nucleic acids. J Comput Chem 7:230
102. Feller SE, MacKerell AD (2000) An improved empirical potential energy function for molecular simulations of phospholipids. J Phys Chem B 104:7510
103. Vanderkooi G (1991) Multibilayer structure of dimyristoylphosphatidylcholine dihydrate as determined by energy minimization. Biochemistry 30:10760
104. Stauch T (1993) Lipid membrane structure and dynamics studied by all-atom molecular dynamics simulations of hydrated phospholipid bilayers. Mol Simul 10:335
105. Egberts E, Marrink SJ, Berendsen HJC (1994) Molecular dynamics simulation of a phospholipid membrane. Eur Biophys J Biophys Lett 22:423
106. Marrink SJ, Mark AE (2004) Molecular view of hexagonal phase formation in phospholipid membranes. Biophys J 87:3894
107. Murzyn K, Zhao W, Karttunen M, Kurdziel M, Rog T (2006) Dynamics of water at membrane surfaces: effect of headgroup structure. Biointerphases 1:98
108. Rog T, Murzyn K, Pasenkiewicz-Gierula M (2002) The dynamics of water at the phospholipid bilayer surface: a molecular dynamics simulation study. Chem Phys Lett 352:323
109. Hogberg CJ, Lyubartsev AP (2006) A molecular dynamics investigation of the influence of hydration and temperature on structural and dynamical properties of a dimyristoylphosphatidylcholine bilayer. J Phys Chem B 110:14326
110. Taning J, Hogberg C, Stevensson B, Lyubartsev A, Maliniak A (2007) Molecular conformations in a phospholipid bilayer extracted from dipolar couplings: a computer simulation study. J Phys Chem B 111:13638
111. Mineva T, Krishnamurty S, Goursot A, Salahub D (2012) Temperature dependence of the conformational interchanges of dilauroyl phosphatidylcholine structures: a density functional study. Int J Quantum Chem (in press)
112. Krishnamurty S, Stefanov M, Mineva T, Begu S, Devoisselle J, Goursot A, Zhu R, Salahub D (2008) Lipid thermodynamics: melting is molecular. Chemphyschem 9:2321
113. Heerklotz H (2004) The microcalorimetry of lipid membranes. J Phys Condens Matter 16:R441
114. Enders O, Ngezahayo A, Wiechmann M, Leisten F, Kolb HA (2004) Structural calorimetry of main transition of supported DMPC bilayers by temperature-controlled AFM. Biophys J 87:2522
115. Ebel H, Grabitz P, Heimburg T (2001) Enthalpy and volume changes in lipid membranes. I. The proportionality of heat and volume changes in the lipid melting transition and its implication for the elastic constants. J Phys Chem B 105:7353
116. Heerklotz H, Seelig J (2002) Application of pressure perturbation calorimetry to lipid bilayers. Biophys J 82:1445
117. Heerklotz H (2002) Triton promotes domain formation in lipid raft mixtures. Biophys J 83:2693
118. Kharakoz DP, Panchelyuga MS, Tiktopulo EI, Shlyapnikova EA (2007) Critical temperatures and a critical chain length in saturated diacylphosphatidylcholines: calorimetric, ultrasonic and Monte Carlo simulation study of chain-melting/ordering in aqueous lipid dispersions. Chem Phys Lipids 150:217
119. Morrow MR, Davis JH (1987) Calorimetric and nuclear-magnetic-resonance study of the phase behavior of dilauroylphosphatidylcholine in water. Biochim Biophys Acta 904:61
120. Finegold L, Shaw WA, Singer MA (1990) Unusual phase properties of dilauroylphosphatidylcholine (C12PC). Chem Phys Lipids 53:177
121. Bonev B, Morrow MR (1996) Effects of hydrostatic pressure on bilayer phase behavior and dynamics of dilauroylphosphatidylcholine. Biophys J 70:2727
122. Hatta I, Matuoka S, Singer MA, Finegold L (1994) A new liquid-crystalline phase in phosphatidylcholine bilayers studied by x-ray diffraction. Chem Phys Lipids 69:129

123. Harroun TA, Nieh MP, Watson MJ, Raghunathan VA, Pabst G, Morrow MR, Katsaras J (2004) Relationship between the unbinding and main transition temperatures of phospholipid bilayers under pressure. Phys Rev E 69:031906
124. Tada K, Goto M, Tamai N, Matsuki H, Kaneshina S (2008) Thermotropic and barotropic phase transitions of dilauroylphosphatidylcholine bilayer. Chem Phys Lipids 153:138
125. Klauda JB, Brooks BR, MacKerell AD, Venable RM, Pastor RW (2005) An ab initio study on the torsional surface of alkanes and its effect on molecular simulations of alkanes and a DPPC bilayer. J Phys Chem B 109:5300
126. Klauda JB, Pastor RW, Brooks BR (2005) Adjacent gauche stabilization in linear alkanes: implications for polymer models and conformational analysis. J Phys Chem B 109:15684
127. May ER, Kopelevich DI, Narang A (2008) Coarse-grained molecular dynamics simulations of phase transitions in mixed lipid systems containing LPA, DOPA, and DOPE lipids. Biophys J 94:878
128. Leekumjorn S, Sum AK (2007) Molecular studies of the gel to liquid-crystalline phase transition for fully hydrated DPPC and DPPE bilayers. Biochim Biophys Acta Biomembr 1768:354
129. Leekumjorn S, Sum AK (2007) Molecular characterization of gel and liquid-crystalline structures of fully hydrated POPC and POPE bilayers. J Phys Chem B 111:6026
130. Qin SS, Yu ZW, Yu YX (2009) Structural and kinetic properties of alpha-tocopherol in phospholipid bilayers, a molecular dynamics simulation study. J Phys Chem B 113:16537
131. Berry R, Beck T, Davis H, Jellinek J (1988) Solid-liquid phase behavior in microclusters. Adv Chem Phys 70B:75–138
132. Frantz DD (1995) Magic numbers for classical Lennard-Jones cluster heat capacities. J Chem Phys 102:3747
133. Bartell LS, Dulles FJ (1995) Monte-Carlo study of small benzene clusters. 2. Transition from rigid to fluxional forms. J Phys Chem 99:17107
134. Wales DJ, Ohmine I (1993) Structure, dynamics and thermodynamics of model (H2O)8 and (H2O)20 clusters. J Chem Phys 98:7245
135. Klinman JP (2007) How do enzymes activate oxygen without inactivating themselves? Acc Chem Res 40:325
136. Que L, Tolman WB (2008) Biologically inspired oxidation catalysis. Nature 455:333
137. Harvey JN (2007) Understanding the kinetics of spin-forbidden chemical reactions. Phys Chem Chem Phys 9:331
138. Lorquet JC, Leyhnihant B (1988) Nonadiabatic unmolecular reactions. 1. A statistical formulation for the rate constants. J Phys Chem 92:4778
139. Prezhdo OV, Rossky PJ (1997) Evaluation of quantum transition rates from quantum-classical molecular dynamics simulations. J Chem Phys 107:5863
140. Jasper AW, Truhlar DG (2005) Electronic decoherence time for non-Born-Oppenheimer trajectories. J Chem Phys 123:064103
141. Zurek W (2007) Decoherence and the transition from quantum to classical—revisited. ArXiv: quant-ph:0306072vj
142. de la Lande A, Rezac J, Levy B, Sanders B, Salahub D (2011) Transmission coefficients for chemical reactions with multiple states: role of quantum decoherence. J Am Chem Soc 133:3883
143. Kapral R (2006) Progress in the theory of mixed quantum-classical dynamics. Annu Rev Phys Chem 57:129
144. de la Lande A, Salahub D, Moliner V, Gerard H, Piquemal J, Parisel O (2009) Dioxygen activation by mononuclear copper enzymes: insights from a tripodal ligand mimicking their Cu(M) coordination sphere. Inorg Chem 48:7003
145. de la Lande A, Babcock NS, Rezac J, Levy B, Sanders BC, Salahub DR (2012) Quantum effects in biological electron transfer. Phys Chem Chem Phys 14:5902
146. Warshel A, Levitt M (1976) Theoretical studies of enzymatic reactions—dielectric, electrostatic and steric stabilization of carbonium ions in reaction of lysozyme. J Mol Biol 103:227

147. Singh UC, Kollman PA (1986) A combined abinitio quantum-mechanical and molecular mechanical method for carrying out simulations on complex molecular systems—applications to the CH3Cl + Cl exchange reaction and gas-phase protonation of polyethers. J Comput Chem 7:718
148. Field MJ, Bash PA, Karplus M (1990) A combined quantum-mechanical and molecular mechanical potential for molecular dynamics simulations. J Comput Chem 11:700
149. Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen KA, Su SJ, Windus TL, Dupuis M, Montgomery JA (1993) General atomic and molecular electronic structure system. J Comput Chem 14:1347
150. Frauenheim T, Seifert G, Elstner M, Hajnal Z, Jungnickel G, Porezag D, Suhai S, Scholz R (2000) A self-consistent charge density-functional based tight-binding method for predictive materials simulations in physics, chemistry and biology. Phys Status Solidi B Basic Res 217:41
151. Yarne DA, Tuckerman ME, Martyna GJ (2001) A dual length scale method for plane-wave-based, simulation studies of chemical systems modeled using mixed ab initio/empirical force field descriptions. J Chem Phys 115:3531
152. Hu H, Yang WT (2009) Development and application of ab initio QM/MM methods for mechanistic simulation of reactions in solution and in enzymes. J Mol Struct Theochem 898:17
153. Alberts IL, Wang Y, Schlick T (2007) DNA polymerase beta catalysis: are different mechanisms possible? J Am Chem Soc 129:11100
154. Altarsha M, Benighaus T, Kumar D, Thiel W (2009) How is the reactivity of cytochrome P450cam affected by Thr252X mutation? A QM/MM study for X = serine, valine, alanine, glycine. J Am Chem Soc 131:4755
155. Altun A, Guallar V, Friesner RA, Shaik S, Thiel W (2006) The effect of heme environment on the hydrogen abstraction reaction of camphor in P450(cam) catalysis: a QM/MM study. J Am Chem Soc 128:3924
156. Altun A, Shaik S, Thiel W (2007) What is the active species of cytochrome p450 during camphor hydroxylation? QM/MM studies of different electronic states of compound I and of reduced and oxidized iron-oxo intermediates. J Am Chem Soc 129:8978
157. Banerjee A, Yang W, Karplus M, Verdine GL (2005) Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA. Nature 434:612
158. Bathelt CM, Zurek J, Mulholland AJ, Harvey JN (2005) Electronic structure of compound I in human isoforms of cytochrome P450 from QM/MM modeling. J Am Chem Soc 127:12900
159. Bhattacharyya S, Stankovich MT, Truhlar DG, Gao JL (2007) Combined quantum mechanical and molecular mechanical simulations of one- and two-electron reduction potentials of flavin cofactor in water, medium-chain acyl-CoA dehydrogenase, and cholesterol oxidase. J Phys Chem A 111:5729
160. Blumberger J, Klein ML (2006) Reorganization free energies for long-range electron transfer in a porphyrin-binding four-helix bundle protein. J Am Chem Soc 128:13854
161. Bucher D, Guidoni L, Rothlisberger U (2007) The protonation state of the Glu-71/Asp-80 residues in the KcsA potassium channel: a first-principles QM/MM molecular dynamics study. Biophys J 93(7):2315
162. Bukowski MR, Koehntop KD, Stubna A, Bominaar EL, Halfen JA, Munck E, Nam W, Que L (2005) A thiolate-ligated nonheme oxoiron(IV) complex relevant to cytochrome P450. Science 310:1000
163. Callis PR, Liu TQ (2006) Short range photoinduced electron transfer in proteins: QM-MM simulations of tryptophan and flavin fluorescence quenching in proteins. Chem Phys 326:230
164. Cao Z, Mo Y, Thiel W (2007) Deprotonation mechanism of NH4+ in the Escherichia coli ammonium transporter AmtB: insight from QM and QM/MM calculations. Angew Chem Int Ed 46:6811

165. Cisneros GA, Perera L, Garcia-Diaz M, Bebenek K, Kunkel TA, Pedersen LG (2008) Catalytic mechanism of human DNA polymerase lambda with Mg2+ and Mn2+ from ab initio quantum mechanical/molecular mechanical studies. DNA Repair 7:1824
166. Field MJ (2002) Simulating enzyme reactions: challenges and perspectives. J Comput Chem 23:48
167. Friesner RA, Guallar V (2005) Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis. Annu Rev Phys Chem 56:389
168. Gao JL, Truhlar DG (2002) Quantum mechanical methods for enzyme kinetics. Annu Rev Phys Chem 53:467
169. Hu H, Yang WT (2008) Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods. Annu Rev Phys Chem 59:573
170. Kamerlin SCL, Haranczyk M, Warshel A (2009) Progress in ab initio QM/MM free-energy simulations of electrostatic energies in proteins: accelerated QM/MM studies of pK(a), redox reactions and solvation free energies. J Phys Chem B 113:1253
171. Lin H, Truhlar DG (2007) QM/MM: what have we learned, where are we, and where do we go from here? Theor Chem Acc 117:185
172. Mordasini TZ, Thiel W (1998) Combined quantum mechanical and molecular mechanical approaches. Chimia 52:288
173. Mulholland AJ (2005) Modelling enzyme reaction mechanisms, specificity and catalysis. Drug Discov Today 10:1393
174. Senn HM, Thiel W (2007) QM/MM studies of enzymes. Curr Opin Chem Biol 11:182
175. Senn HM, Thiel W (2009) QM/MM methods for biomolecular systems. Angew Chem Int Ed 48:1198
176. Lev B, Zhang R, de la Lande A, Salahub DR, Noskov SY (2010) The QM-MM interface for CHARMM-deMon. J Comput Chem 31:1015
177. Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoscek M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M (2009) CHARMM: the biomolecular simulation program. J Comput Chem 30:1545
178. Zhang R, Lev B, Cuervo JE, Noskov SY, Salahub DR (2010) A guide to QM/MM methodology and applications. Adv Quantum Chem 59:353
179. Alvarez-Ibarra A, Koester A, Zhang R, Salahub DR (2012), Asymptotic Expansion for Electrostatic Embedding Integrals in QM/MM Calculations. J Chem Theor Comp DOI: 10.1021/ct300609z
180. Riccardi D, Li GH, Cui Q (2004) Importance of van der Waals interactions in QM/MM simulations. J Phys Chem B 108:6467
181. Maseras F, Morokuma K (1995) Imomm—a new integrated ab-initio plus molecular mechanics geometry optimization scheme of equilibrium structures and transition-states. J Comput Chem 16:1170
182. Murphy RB, Philipp DM, Friesner RA (2000) A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments. J Comput Chem 21:1442
183. Vreven T, Morokuma K, Farkas O, Schlegel HB, Frisch MJ (2003) Geometry optimization with QM/MM, ONIOM, and other combined methods. I. Microiterations and constraints. J Comput Chem 24:760
184. Prat-Resina X, Gonzalez-Lafont A, Lluch JM (2003) How important is the refinement of transition state structures in enzymatic reactions? J Mol Struct Theochem 632:297
185. Marti S, Moliner V (2005) Improving the QM/MM description of chemical processes: a dual level strategy to explore the potential energy surface in very large systems. J Chem Theory Comput 1:1008

186. Vreven T, Frisch MJ, Kudin KN, Schlegel HB, Morokuma K (2006) Geometry optimization with QM/MM methods. II: Explicit quadratic coupling. Mol Phys 104:701
187. Kastner J, Thiel S, Senn HM, Sherwood P, Thiel W (2007) Exploiting QM/MM capabilities in geometry optimization: a microiterative approach using electrostatic embedding. J Chem Theory Comput 3:1064
188. Zwanzig RW (1954) High-temperature equation of state by a perturbation method.1. Nonpolar gases. J Chem Phys 22:1420
189. Hugosson HW, Laio A, Maurer P, Rothlisberger U (2006) A comparative theoretical study of dipeptide solvation in water. J Comput Chem 27:672
190. Kundrat MD, Autschbach J (2008) Ab initio and density functional theory modeling of the chiroptical response of glycine and alanine in solution using explicit solvation and molecular dynamics. J Chem Theory Comput 4:1902
191. Jorgensen WL, Tirado-Rives J (2005) Molecular modeling of organic and biomolecular systems using BOSS and MCPRO. J Comput Chem 26:1689
192. Warshel A, Florian J (2004) In: PvR S, Jorgensen WL, Schaeffer-III HF, Schreiner PR, Thiel W, Glen R (eds) Encyclopedia of computational chemistry. Chichester, Wiley
193. Zhang YK, Liu HY, Yang WT (2000) Free energy calculation on enzyme reactions with an efficient iterative procedure to determine minimum energy paths on a combined ab initio QM/MM potential energy surface. J Chem Phys 112:3483
194. Liu HY, Lu ZY, Cisneros GA, Yang WT (2004) Parallel iterative reaction path optimization in ab initio quantum mechanical/molecular mechanical modeling of enzyme reactions. J Chem Phys 121:697
195. Hu H, Lu ZY, Yang WT (2007) QM/MM minimum free-energy path: methodology and application to triosephosphate isomerase. J Chem Theory Comput 3:390
196. Wada G, Tamura E, Okina M, Nakamura M (1982) On the ratio of zwitterion form to uncharged form of glycine at equilibrium in various aqueous media. Bull Chem Soc Jpn 55:3064
197. Jensen JH, Gordon MS (1995) On the number of water molecules necessary to stabilize the glycine zwitterion. J Am Chem Soc 117:8159
198. Ramaekers R, Pajak J, Lambie B, Maes G (2004) Neutral and zwitterionic glycine.H_2O complexes: a theoretical and matrix-isolation Fourier transform infrared study. J Chem Phys 120:4182
199. Bachrach SM (2008) Microsolvation of glycine: a DFT study. J Phys Chem A 112:3722
200. Wood GPF, Gordon MS, Radom L, Smith DM (2008) Nature of glycine and Iis alpha-carbon radical in aqueous solution: a theoretical investigation. J Chem Theory Comput 4:1788
201. Leung K, Rempe SB (2005) Ab initio molecular dynamics study of glycine intramolecular proton transfer in water. J Chem Phys 122:184506
202. Gontrani L, Mennucci B, Tomasi J (2000) Glycine and alanine: a theoretical study of solvent effects upon energetics and molecular response properties. J Mol Struct Theochem 500:113
203. Nair NN, Schreiner E, Marx D (2008) Peptide synthesis in aqueous environments: the role of extreme conditions on amino acid activation. J Am Chem Soc 130:14148
204. Sun J, Bousquet D, Forbert H, Marx D (2010) Glycine in aqueous solution: solvation shells, interfacial water, and vibrational spectroscopy from ab initio molecular dynamics. J Chem Phys 133:114508
205. Montgomery JA, Frisch MJ, Ochterski JW, Petersson GA (2000) A complete basis set model chemistry. VII. Use of the minimum population localization method. J Chem Phys 112:6532
206. Slifkin MA, Ali SM (1984) Thermodynamic parameters of the activation of glycine zwitterion protonation reactions. J Mol Liq 28:215
207. Berendsen HJC, Grigera JR, Straatsma TP (1987) The missing term in effective pair potentials. J Phys Chem 91:6269
208. Dunning TH (1989) Gaussian-basis sets for use in correlated molecular calculations. 1: The atoms boron through neon and hydrogen. J Chem Phys 90:1007
209. Goursot A, Mineva T, Salahub D. unpublished

210. Wang LH, Yu XY, Hu P, Broyde S, Zhang YK (2007) A water-mediated and substrate-assisted catalytic mechanism for Sulfolobus solfataricus DNA polymerase IV. J Am Chem Soc 129:4731
211. Wang LH, Broyde S, Zhang YK (2009) Polymerase-tailored variations in the water-mediated and substrate-assisted mechanism for nucleotidyl transfer: insights from a study of T7 DNA polymerase. J Mol Biol 389:787
212. Carvalho ATP, Fernandes PA, Ramos MJ (2011) The catalytic mechanism of RNA polymerase II. J Chem Theory Comput 7:1177
213. Marcus RA, Sutin N (1985) Electron transfers in chemistry and biology. Biochim Biophys Acta 811:265
214. de la Lande A, Gerard H, Parisel O (2008) How to optimize a C-H cleavage with a mononuclear copper-dioxygen adduct? Int J Quantum Chem 108:1898
215. Kaduk B, Kowalczyk T, Van Voorhis T (2012) Constrained density functional theory. Chem Rev 112:321
216. Van Voorhis T, Kowalczyk T, Kaduk B, Wang LP, Cheng CL, Wu Q (2010) The diabatic picture of electron transfer, reaction barriers, and molecular dynamics. In: Leone SR, Cremer PS, Groves JT, Johnson MA, Richmond G (eds). Annu Rev Phys Chem 61:149–170
217. Gunnarsson O, Lundqvist BI (1976) Exchange and correlation in atoms, molecules and solids by spin-density functional formalism. Phys Rev B 13:4274
218. Dederichs PH, Blugel S, Zeller R, Akai H (1984) Ground states of constrained systems—application to cerium impurities. Phys Rev Lett 53:2512
219. Wesolowski T, Muller RP, Warshel A (1996) Ab initio frozen density functional calculations of proton transfer reactions in solution. J Phys Chem 100:15444
220. Olsson MHM, Hong GY, Warshel A (2003) Frozen density functional free energy simulations of redox proteins: computational studies of the reduction potential of plastocyanin and rusticyanin. J Am Chem Soc 125:5025
221. Hong GY, Rosta E, 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. J Phys Chem B 110:19570
222. Karmelin SCL, Warshel A (2010) The EVB as a quantitative tool for formulating simulations and analyzing biological and chemical reactions. Faraday Discuss (145):71
223. Wu Q, Van Voorhis T (2005) Direct optimization method to study constrained systems within density-functional theory. Phys Rev A 72:024502
224. Press W, Teukolsky S, Vetterling W, Flannery B (1986) Numerical recipes in Fortran. Cambridge University Press, Cambridge
225. de la Lande A, Salahub D (2010) Derivation of interpretative models for long range electron transfer from constrained density functional theory. J Mol Struct Theochem 943:115
226. Rezac J, Levy B, Demachy I, de la Lande A (2012) Robust and efficient constrained DFT molecular dynamics approach for biochemical modeling. J Chem Theory Comput 8(2):418–427. doi:10.1021/ct200570u
227. Mulliken RS (1955) Electronic population analysis on LCAO-MO molecular wave functions. J Chem Phys 23:1833
228. Löwdin (1970) On the nonorthogonality problem. Adv Quantum Chem 5:185
229. Reed AE, Weinstock RB, Weinhold F (1985) Natural-population analysis. J Chem Phys 83:735
230. Reed AE, Curtiss LA, Weinhold F (1988) Intermolecular interactions from a natural bond orbital donor-acceptor viewpoint. Chem Rev 88:899
231. Bickelhaupt FM, Hommes N, Guerra CF, Baerends EJ (1996) The carbon-lithium electron pair bond in (CH3Li)(n) (n = 1, 2, 4). Organometallics 15(13):2923
232. Becke A (1988) A multicenter numerical-integration scheme for polyatomic molecules. J Chem Phys 88:2547
233. Hirshfeld FL (1977) Bonded-atom fragments for describing molecular charge densities. Theor Chim Acta 44:129

234. Mandado M, Krishtal A, Van Alsenoy C, Bultinck P, Hermida-Ramon JM (2007) Bonding study in all-metal clusters containing Al-4 units. J Phys Chem A 111:11885
235. Ghillemijn D, Bultinck P, Van Neck D, Ayers PW (2011) A self-consistent Hirshfeld method for the atom in the molecule based on minimization of information loss. J Comput Chem 32:1561
236. Geldof D, Krishtal A, Blockhuys F, Van Alsenoy C (2011) An extension of the Hirshfeld method to open shell systems using fractional occupations. J Chem Theory Comput 7:1328
237. Lillestolen TC, Wheatley RJ (2009) Atomic charge densities generated using an iterative stockholder procedure. J Chem Phys 131:144101
238. de la Lande A Köster A, Vela A, Lévy B, Demachy I (2012) in preparation
239. Lewis JC, Coelho PS, Arnold FH (2011) Enzymatic functionalization of carbon-hydrogen bonds. Chem Soc Rev 40:2003
240. King G, Warshel A (1990) Investigation of the free-energy functions for electron-transfer reactions. J Chem Phys 93:8682
241. Wu Q, Van Voorhis T (2006) Direct calculation of electron transfer parameters through constrained density functional theory. J Phys Chem A 110:9212
242. Wu Q, Van Voorhis T (2006) Extracting electron transfer coupling elements from constrained density functional theory. J Chem Phys 125:164105
243. Lu YH, Quardokus R, Lent CS, Justaud F, Lapinte C, Kandel SA (2010) Charge localization in isolated mixed-valence complexes: an STM and theoretical study. J Am Chem Soc 132:13519
244. Oberhofer H, Blumberger J (2009) Charge constrained density functional molecular dynamics for simulation of condensed phase electron transfer reactions. J Chem Phys 131:064101
245. Oberhofer H, Blumberger J (2010) Electronic coupling matrix elements from charge constrained density functional theory calculations using a plane wave basis set. J Chem Phys 133:244105
246. Karmelin SCL, Vicatos S, Dryga A, Warshel A (2011) Simulations in studies of biophysical and chemical systems. Annu Rev Phys Chem 62:41
247. Karmelin SCL, Warshel A (2011) The empirical valence bond model: theory and applications. WIREs Comput Mol Sci 1:30
248. Langen R, Chang IJ, Germanas JP, Richards JH, Winkler JR, Gray HB (1995) Electron tunneling in proteins—coupling through a beta-strand. Science 268:1733
249. Rudra I, Wu Q, Van Voorhis T (2007) Predicting exchange coupling constants in frustrated molecular magnets using density functional theory. Inorg Chem 46:10539
250. Flores-Moreno R, Koster A (2008) Auxiliary density perturbation theory. J Chem Phys 128:134105
251. Malkin V, Malkina O, Salahub D (1993) Calculations of NMR shielding constants by uncoupled density functional theory. Chem Phys Lett 204:80
252. Zuniga-Gutierrez B, Geudtner G, Koster A (2011) NMR shielding tensors from auxiliary density functional theory. J Chem Phys 134:124108
253. De Luca G, Russo N, Koster AM, Calaminici P, Jug K (1999) Density functional theory calculations of nuclear quadrupole coupling constants with calibrated O-17 quadrupole moments. Mol Phys 97:347
254. Parr RG, Donnelly RA, Levy M, Palke WE (1978) Electronegativity- density functional viewpoint. J Chem Phys 68:3801
255. Geerlings P, De Proft F (2008) Conceptual DFT: the chemical relevance of higher response functions. Phys Chem Chem Phys 10:3028
256. Becke A, Edgecombe K (1990) A simple measure of electron localization in atomic and molecular systems. J Chem Phys 92:5397
257. Silvi B, Savin A (1994) Classification of chemical bonds based on topological analysis of electron localization functions. Nature 371:683
258. Gillespie RJ, Nyholm RS (1957) Inorganic stereochemistry. Q Rev 11:339

259. de la Lande A, Salahub D, Maddaluno J, Scemama A, Pilme J, Parisel O, Gerard H, Caffarel M, Piquemal J (2011) Rapid communication spin-driven activation of dioxygen in various metalloenzymes and their inspired models. J Comput Chem 32:1178
260. Piquemal JP, Pilme J, Parisel O, Gerard H, Fourre I, Berges J, Gourlaouen C, De La Lande A, Van Severen MC, Silvi B (2008) What can be learnt on biologically relevant systems from the topological analysis of the electron localization function? Int J Quantum Chem 108:1951
261. Polo V, Andres J, Berskit S, Domingo LR, Silvi B (2008) Understanding reaction mechanisms in organic chemistry from catastrophe theory applied to the electron localization function topology. J Phys Chem A 112:7128
262. Leboeuf M, Koster A, Jug K, Salahub D (1999) Topological analysis of the molecular electrostatic potential. J Chem Phys 111:4893

Struct Bond (2013) 150: 65–96
DOI: 10.1007/978-3-642-32750-6_2

Density Functional Theory and Molecular Interactions: Dispersion Interactions

Eugene S. Kryachko

Abstract We are definitely witnessing an ever-increasing need to study dispersion molecular interactions that govern a weakly bound molecular world within the density functional theory. This chapter outlines the basic approaches currently undertaken to resolve this density functional paradigm.

Contents

1 Density Functional Theory: Background

In the last two decades, since Walther Kohn was awarded the Nobel Prize 1991 in Chemistry for the density functional theory (DFT), the latter has become the most popular and useful computational approach to study many-electron systems in the ground states because of its physically transparent underlying concept and lower computational cost [1–8]. The Kohn–Sham version of DFT [1–3, 7] is the most widely used many-body method for electronic structure calculations of atoms, molecules, solids, and solid surfaces.

E.S. Kryachko (✉)
Bogoliubov Institute for Theoretical Physics, Kiev 03680, Ukraine
e-mail: eugene.kryachko@ulg.ac.be

Generally speaking, a molecule **M** is a stable quantum Coulomb system that consists of the following two subsystems:

- The electronic—of N electrons of the mass m_e and the charge $-e$ which positions in the spin-configurational space are determined by the corresponding radii vectors $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N$ where each $\mathbf{r}_i$, $i = 1, 2, \ldots, N$ belongs to the real three-dimensional space $\mathbb{R}^3$ and the spins $\sigma_1, \sigma_2, \ldots, \sigma_N$ where each σ_i, $i = 1, 2, \ldots, N$ takes the value from $\mathbb{Z}_2 = \{\pm 1/2\}$, the discrete two-dimensional spin space
- The nuclear—of M nuclei carrying the nuclear charges $\{Z_\alpha\}_{\alpha=1}^{M}$ and located at $\{\mathbf{R}_\alpha \in \mathbb{R}^3\}_{\alpha=1}^{M}$.

According to Löwdin's definition [9, 10]: "A system of electrons and atomic nuclei is said to form a molecule if the Coulombic Hamiltonian H'—with the center of mass motion removed—has a discrete ground-state energy E_0" (see also [11–13] and references therein) where the total Hamiltonian $H := \widehat{H} = \widehat{H}_{\mathrm{e}} + \widehat{T}_{\mathrm{nn}} + \widehat{U}_{\mathrm{nn}}$ is, respectively, the sum of the electronic Hamiltonian operator, the nuclear kinetic energy operator, and the nuclear–nuclear Coulomb interaction energy operator. Consider, within the Born–Oppenheimer approximation, the electronic Hamiltonian operator (in the atomic units) of **M**:

$$\widehat{H}_{\mathrm{e}} = \widehat{T}_{\mathrm{e}} + \widehat{U}_{\mathrm{ee}} + \widehat{V}_{\mathrm{en}} = -\frac{1}{2}\sum_{i=1}^{N}\nabla_{\mathbf{r}_i}^2 + \sum_{1=i<j}^{N}\frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{i=1}^{N}\hat{v}(\mathbf{r}_i), \tag{1}$$

where $\widehat{T}_{\mathrm{e}}$ is the nuclear kinetic energy operator, $\widehat{U}_{\mathrm{ee}}$ the nuclear–nuclear Coulomb interaction energy operator, and the "external," electron–nuclear potential is defined as

$$\hat{v}(\mathbf{r}_i) := \sum_{\alpha=1}^{M}\frac{Z_\alpha}{|\mathbf{r}_i - \mathbf{R}_\alpha|}. \tag{2}$$

$\widehat{H}_{\mathrm{e}}$ acts on the class $\mathcal{L}_N$ of "admissible" N-electron wavefunctions $\Psi(\mathbf{r}_1, s_1; \ldots; \mathbf{r}_N, s_N)$ obeying the following conditions:

(**Fi**) the wavefunction normalization:

$$\langle\Psi|\Psi\rangle = \sum_{s_1,\ldots,s_N}\int \mathrm{d}^3\mathbf{r}_1 \ldots \int \mathrm{d}^3\mathbf{r}_N |\Psi(\mathbf{r}_1, s_1; \ldots; \mathbf{r}_N, s_N)|^2 < \infty \tag{3}$$

implying that $\mathcal{L}_N \subset \mathcal{L}_\sigma^2\left(\mathbb{R}^{3N} \otimes \mathbb{Z}_2^N\right)$, the Hilbert space of antisymmetric, square-integrable N-electron wavefunctions. Henceforth it is assumed that an arbitrary $\Psi \in \mathcal{L}_N$ is normalized to unity: $\langle\Psi|\Psi\rangle = 1$;

(**Fii**) the boundness from below of the expectation value $\left\langle\Psi\middle|\hat{H}_{\mathrm{e}}\middle|\Psi\right\rangle > -\infty$: In fact, (**Fii**) results from the aforementioned definition of molecule which lowest energy is finite. If $\widehat{U}_{\mathrm{ee}}$ and $\widehat{V}_{\mathrm{en}}$ are of Coulomb type, (**Fii**) is equivalent to

$$T_{\mathrm{e}}[\Psi] = \langle\Psi|T_{\mathrm{e}}|\Psi\rangle < \infty \tag{4}$$

implying that $\Psi \in \mathcal{L}_N$ is a differentiable function of all spatial coordinates, together with each component of $\nabla_{\mathbf{r}_i}\Psi \in \mathcal{L}_N$.

One can prove [4, 14] that the conditions (**Fi**) and (**Fii**) fully determine $\mathcal{L}_N$ of "admissible" N-electron wavefunctions where the energy functional

$$E[\Psi] \equiv \left\langle \Psi \middle| \widehat{H}_{\mathrm{e}} \middle| \Psi \right\rangle \tag{5}$$

is thus well defined. Its lowest energy, the infimum, equal to the ground-state electronic energy E_0 as the lowest eigenenergy of the N-body Schrödinger equation

$$\widehat{H}_{\mathrm{e}}\Psi_0 = E_0\Psi_0, \tag{6}$$

is attained at the ground-state electronic wavefunction Ψ_0, that is

$$E_0 \equiv \inf_{\Phi \in \mathcal{L}_N} \{E[\Phi]\} = E[\Phi]|_{\Phi=\Psi_0 \in \mathcal{L}_N}. \tag{7}$$

The stationary quantum mechanical variational principle then reads as

$$\delta E[\Phi]|_{\Phi=\Psi_0} = 0. \tag{8}$$

The basic postulate of the many-electron density functional theory [1–8] suggests, first, the existence of the so-called functional

$$\mathcal{E}[\rho(x)] = \begin{cases} \mathcal{E}[\rho(\mathbf{r})] & \text{spin} - \text{restricted functional} \\ \mathcal{E}[\rho_\uparrow(\mathbf{r}), \rho_\downarrow(\mathbf{r})] & \text{spin} - \text{polarized functional} \end{cases} \tag{9}$$

that has the meaning of the energy and depends, in some functional manner, on one-electron density $\rho(\mathbf{r})$,

$$\rho_\Psi(\mathbf{r}) := N \sum_{s_1,\ldots,s_N} \int d^3\mathbf{r}_2 \ldots \int d^3\mathbf{r}_N |\Psi(\mathbf{r}, s_1; \mathbf{r}_2, s_2; \ldots; \mathbf{r}_N, s_N)|^2, \Psi \in \mathcal{L}_N \tag{10}$$

or on its both spin components, $\rho_{\Psi\uparrow}(\mathbf{r})$ and $\rho_{\Psi\downarrow}(\mathbf{r})$,

$$\begin{aligned} \rho_{\Psi s}(\mathbf{r}) &:= \rho_\Psi(\mathbf{r}, s) \\ &= N \sum_{s_1,\ldots,s_N} \int d^3\mathbf{r}_2 \ldots \int d^3\mathbf{r}_N |\Psi(\mathbf{r}, s; \mathbf{r}_2, s_2; \ldots; \mathbf{r}_N, s_N)|^2, s = \uparrow, \downarrow . \end{aligned} \tag{11}$$

The latter yield together $\rho_\Psi(\mathbf{r}) = \rho_{\Psi\uparrow}(\mathbf{r}) + \rho_{\Psi\downarrow}(\mathbf{r})$. Each $\rho_{\Psi s}(\mathbf{r})$ is normalized to N_s so that $N_\uparrow + N_\downarrow = N$. The second suggestion is that the infimum of $\mathcal{E}[\rho(\mathbf{r})]$ does exist and

$$E_0 \equiv \inf_{\Phi\in\mathcal{L}_N}\{E[\Phi]\} = E[\Phi]|_{\Phi=\Psi_0} = \inf_{\rho\in\mathcal{P}_N}\{\mathcal{E}[\rho(\mathbf{r})]\} = \mathcal{E}[\rho_\Phi(\mathbf{r})]|_{\Phi=\Psi_0} \tag{12}$$

where $\mathcal{P}_N$ is a set of one-electron densities associated with $\mathcal{L}_N$ (see below). Formally, this postulate looks rather strong, however it is widely accepted that it is guaranteed by the Hohenberg–Kohn theorem [1] (for the new proof of the Hohenberg–Kohn theorem see [15, 16]).

Equation (9) assumes the existence of the "**F**unctional mapping"

$$\mathrm{F} : E[\Psi] \mapsto \mathcal{E}[\rho_\Psi(\mathbf{r})] \tag{13}$$

that implicitly presumes the existence of the "**V**ariable mapping"

$$\Psi \leftrightarrow \rho_\Psi(\mathbf{r}). \tag{14}$$

Obviously, the mapping (14) is valid if, first, there are defined the sets of "variables" on its left- and right-hand sides. Second, the symbol $\leftrightarrow$ does not mean at all that this is precisely a one-to-one correspondence. The sub-mapping of (14), $V{:}\Psi \rightarrow \rho_\Psi(\mathbf{r})$, is given by the reduction mapping, either (10) or (11), that is, $\rho_\Psi(\mathbf{r}) = V(\Psi)$ and $\mathcal{P}_N \equiv V\mathcal{L}_N$. Besides, the reduction mapping has another facet—this is a so-called N-representability: any one-electron density obtained via V possesses its own image in $\mathcal{L}_N$. Generally speaking, the inverse mapping V^{-1} is one-to-many, that is, a given one-electron density has many preimages in $\mathcal{L}_N$. It is trivial to show this. Let us consider any stable two-electron system which ground-state wavefunction and one-electron density are $\Psi_0(\mathbf{r}_1, \mathbf{r}_2)[\alpha(s_1)\beta(s_2) - \beta(s_1)\alpha(s_2)]$ and $\rho_0(\mathbf{r})$, respectively. The two-electron Slater determinant $\sqrt{\rho_0(\mathbf{r}_1)\rho_0(\mathbf{r}_2)}[\alpha(s_1)\alpha(s_2) - \beta(s_1)\alpha(s_2)]/2$ possesses the same one-electron-density $\rho_0(\mathbf{r})$ as well. Q. E. D. The Hohenberg–Kohn theorem [1] (see also [15, 16]) states however that there exists a one-to-one correspondence between the ground-state wavefunctions and ground-state densities.

2 Kohn–Sham Density Functional Theory: The Hohenberg–Kohn Theorem

The Hohenberg–Kohn theorem [1] underlies the foundation of the density functional theory [2–7], the Kohn–Sham density functional theory. On p. B864 of their work [1], Hohenberg and Kohn state that they "... develop an exact formal variational principle for the ground-state energy, in which the density" $\rho(\mathbf{r})$ (in a widely accepted notation) "is the variable function. Into this principle enters a universal functional" $F[\rho(\mathbf{r})]$, "which applies to all electronic systems in their ground state no matter what the external potential is." Following Hohenberg and Kohn [1], let us consider "a collection of an arbitrary number of electrons, enclosed in a large box and moving under the influence of an external potential $v(\mathbf{r})$ and

mutual Coulomb repulsion." The total Hamiltonian H of a given N-electron system is written as Eq. (1), viz.

$$H = T_e + V_{ee} + V, \tag{15}$$

where T_e is the kinetic energy operator of N electrons, V_{ee} is the interelectronic Coulomb operator, and

$$V = \sum_{i=1}^{N} v(\mathbf{r}_i) \tag{16}$$

is the total external potential. Hohenberg and Kohn [1] further assume (p. B865) that H possesses the least bound-state (ground-state) wavefunction $\Psi_0(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) \in \mathcal{H}_N$ (spins are omitted for simplicity) and the latter is nondegenerate. $\mathcal{H}_N$ is the Hilbert space of square integrable N-electron wavefunctions and $\mathbf{r}_i \in \Re^3$, $i = 1, 2, \ldots, N$. Let us define the corresponding ground-state one-electron density [3]

$$\rho_0(\mathbf{r}) \equiv \int \prod_{i=2}^{N} d^3\mathbf{r}_i |\Psi_o(\mathbf{r}, \mathbf{r}_2, \ldots, \mathbf{r}_N)|^2, \tag{17}$$

"which is clearly a functional of $v(\mathbf{r})$" ([1], p. B865), that is, there exist such mappings

$$v(\mathbf{r}) \Rightarrow \Psi_0(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) \Rightarrow \rho_0(\mathbf{r}). \tag{18}$$

Hohenberg–Kohn Theorem [1]. "$v(\mathbf{r})$ is a unique functional" of $\rho(\mathbf{r})$, "apart from a trivial additive constant."

Proof ([1], p. B865). "The proof proceeds by *reductio ad absurdum*." We *assume* the existence of two "external" potentials $v_1(\mathbf{r})$ and $v_2(\mathbf{r})$ such that

$$v_1(\mathbf{r}) \neq v_2(\mathbf{r}) + \text{constant}. \tag{19}$$

Via Eqs. (16) and (15), $v_1(\mathbf{r})$ and $v_2(\mathbf{r})$ define the Hamiltonians H_1 and H_2 associated with two different N-electron systems. Let us further *assume* the existence of the ground-state normalized wavefunctions $\Psi_0^{(1)} \in \mathcal{H}_N$ and $\Psi_0^{(2)} \in \mathcal{H}_N$ of H_1 and H_2, respectively. By virtue of Eq. (17), $\Psi_0^{(1)}$ and $\Psi_0^{(2)}$ yield the corresponding ground-state one-electron densities $\rho_0^{(1)}$ and $\rho_0^{(2)}$. Hohenberg and Kohn [1] finally *assume* that

(i) $\Psi_0^{(1)} \neq \Psi_0^{(2)}$

(ii) $\rho_0^{(1)}(\mathbf{r}) = \rho_0^{(2)}(\mathbf{r}) = \rho_0(\mathbf{r})$.

Applying the Rayleigh–Ritz variational principle, one obtains

$$\begin{aligned} E_0^{(1)} =& \left\langle \Psi_0^{(1)} | H_1 | \Psi_0^{(1)} \right\rangle \overset{\text{(i)}}{<} \left\langle \Psi_0^{(2)} | H_1 | \Psi_0^{(2)} \right\rangle \\ \overset{Eq.(5)}{=}& \left\langle \Psi_0^{(2)} | H_2 | \Psi_0^{(2)} \right\rangle + \left\langle \Psi_0^{(2)} | V_1 - V_2 | \Psi_0^{(2)} \right\rangle \\ =& E_0^2 + \int d^3\mathrm{r} [v_1(\mathbf{r}) - v_2(\mathbf{r})] \rho_0(\mathbf{r}) \end{aligned} \tag{20}$$

and

$$\begin{aligned} E_0^{(2)} =& \left\langle \Psi_0^{(2)} | H_2 | \Psi_0^{(2)} \right\rangle \overset{\text{(i)}}{<} \left\langle \Psi_0^{(1)} | H_2 | \Psi_0^{(1)} \right\rangle \\ \overset{Eq.(5)}{=}& \left\langle \Psi_0^{(1)} | H_1 | \Psi_0^{(1)} \right\rangle + \left\langle \Psi_0^{(1)} | V_2 - V_1 | \Psi_0^{(1)} \right\rangle \\ =& E_0^1 + \int d^3\mathrm{r} [v_1(\mathbf{r}) - v_2(\mathbf{r})] \rho_0(\mathbf{r}) \end{aligned} \tag{21}$$

where the used formulas are indicated above the signs.

Hohenberg and Kohn then conclude ([1], p. B865) that adding (20) to (7) "leads to the inconsistency"

$$E_0^{(1)} + E_0^{(2)} < E_0^{(1)} + E_0^{(2)}, \tag{22}$$

and therefore, (22) implies that the assumption (ii) fails. "Thus $v(\mathbf{r})$ is (to within a constant) a unique functional of" $\rho(\mathbf{r})$, "since, in turn, $v(\mathbf{r})$ fixes H we see that the full many-particle ground state is unique functional of" $\rho(\mathbf{r})$. Q. E. D.

Examine Eq. (22). It is obviously self-contradictory (see also [15–17]). Equation (22) is deduced under the assumption that (19) is true together with the to-be-refuted assumptions (i) and (ii) both composing the negation of the Hohenberg–Kohn theorem. Equation (22) then appears to be absurd in a sense of being obviously false and therefore the statement of the Hohenberg–Kohn theorem is true. This might, logically speaking, imply that one of the to-be-refuted assumptions, (i) or (ii), or simultaneously both, (i) and (ii), lead to the contradiction with (19) *or* they are a priori invalid in a sense that one of them *or* both are incompatible with (5) and therefore, the statement of the Hohenberg–Kohn theorem is not true unless it is proved in the other way (see below). Explicitly, all these cases are the following:

(I) $\Psi_0^{(1)} = \Psi_0^{(2)} = \Psi_0$.

This directly gives $\rho_0^{(1)} = \rho_0^{(2)} = \rho_0$, that is, (ii) does hold. This also yields that

$$V_1 \equiv V_2 \equiv E_0 - \frac{(T_\mathrm{e} + V_\mathrm{ee})\Psi_0}{\Psi_0} \tag{23}$$

if V_1 and V_2 are multiplicative operators, as suggested by Eq. (16). Equation (23) clearly contradicts (19). However, there is no "inconsistency" because the last terms in the last lines of Eqs. (20) and (21) simply vanish.

(II) $\Psi_0^{(1)} \neq \Psi_0^{(2)}$ and $\rho_0^{(1)} \neq \rho_0^{(2)}$
This is precisely in the line of the original reasoning by Hohenberg and Kohn [1] proving that different external potentials determine different ground-state one-electron densities.

(III) $\Psi_0^{(1)} = \Psi_0^{(2)}$ and $\rho_0^{(1)} \neq \rho_0^{(2)}$
These two relations contradict to each other due to (17).

(IV) A self-contradiction (*ad absurdum*) of Eq. (22) might also mean that the to-be-refuted assumptions (i) or/and (ii) of the Hohenberg–Kohn theorem are self-contradictory with Eq. (19) and this is precisely the case of many-electron Coulomb systems with Coulomb-type class of external potentials. In other words, the original *reductio ad absurdum* proof of the Hohenberg–Kohn theorem based on the assumption (19) is incompatible with the *ad absurdum* assumption (ii) since the Kato theorem is valid for such systems [18].

The Kato theorem [18] (see also [19–29]) determines the character of the singularity of the exact N-electron wavefunction at the electron–nucleus coalescence where the external potential $v(\mathbf{r})$ of the Coulomb form (see Eq. (2.2) and the conditions (i) and (ii) on p. 154 and Theorem I on p. 156 of [19]; in atomic units)

$$v(\mathbf{r}) = -\sum_{\alpha=1}^{M} \frac{Z_\alpha}{|\mathbf{r} - \mathbf{R}_\alpha|} \tag{24}$$

is singular. In Eq. (24), the αth nucleus with the nuclear charge Z_α is placed at $\mathbf{R}_\alpha \in \Re^3$. Any N-electron eigenwavefunction Ψ of H with $v(\mathbf{r})$ of the form (24) and its one-electron density ρ_Ψ then satisfy the electron–nucleus cusp condition, that is, their corresponding nonvanishing radial logarithmic derivatives obey at $\mathbf{R}_\alpha$ the following relationships:

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}r_i} \log \Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_i, \ldots, \mathbf{r}_N)|_{\mathbf{r}_i = \mathbf{R}_\alpha} = -Z_\alpha, \quad i = 1, 2, \ldots, N \\ &\frac{\mathrm{d}}{\mathrm{d}r} \log \rho_\Psi(\mathbf{r})|_{\mathbf{r}_i = \mathbf{R}_\alpha} = -2Z_\alpha. \end{aligned} \tag{25}$$

Therefore, the true one-electron density of the given N-electron system exhibits cusps (local maxima) at the positions of the nuclei. Analyzing the topology of the given ground-state one-electron density $\rho_0(\mathbf{r})$ over the whole coordinate space $\Re^3$, one determines the positions of its cusps and evaluates the lhs of Eq. (25) (the last one) at these points. Altogether, the positions of the electron–nucleus cusps [as being always negative, see Eq. (25)] and the halves of the radial logarithmic derivatives of $\rho_0(\mathbf{r})$, taken with the opposite sign at these points, fully determine the external potential $v(\mathbf{r})$, Eq. (24), of the given system. This constitutes a naïve interpretation of the Hohenberg–Kohn theorem originally proposed by Coleman [30], Bamzai and

Deb [31], Smith [32], and E. Bright Wilson (quoted by Löwdin [33]; for the recent applications of the Kato theorem to the Hohenberg–Kohn theorem see also [34–36]). Therefore, if a given pair of N-electron systems with the Hamiltonians H_1 and H_2 of the type (1) are characterized by the same ground-state one-electron densities ($\equiv$ to-be-refuted assumption (ii)), their external potentials $v_1(\mathbf{r})$ and $v_2(\mathbf{r})$ of the form (24) are identical. The latter contradicts (19) and hence, the assumption (ii) cannot be used in the proof via *reductio ad absurdum* of the Hohenberg–Kohn theorem together with the assumption (19). In other words, they are Kato-type incompatible with each other.

Vice versa, the nuclei of the given N-electron system are isolated 3D point attractors behaving topologically as critical points of rank three and signature minus three [37]. However, there exist some "particular many-electron systems" which show local maxima of their ground-state one-electron density at non-nuclear positions ([38–48] and references therein). These local non-nuclear maxima might be the true ones or might appear as a consequence of an incomplete, inadequate quantum mechanical treatment. Therefore, despite the present conclusion that the original proof of the Hohenberg–Kohn theorem by *reductio ad absurdum* is flawed in a sense that its to-be-refuted assumption (ii) is incompatible, by virtue of the Kato theorem, with the assumption (19) (for a similar proof of the ensemble generalization of the Hohenberg–Kohn theorem see Sect. II of [16]), the Kato theorem itself corroborates the existence of the one-to-one correspondence between the Coulomb-type class of external potentials (10) and the ground-state one-electron densities for nearly all many-electron except probably the aforementioned "particular" ones.

According to the work [1] by Hohenberg and Kohn, the Hohenberg–Kohn theorem implies the existence of the universal energy density functional for any isolated many-electron Coulomb system. This statement has been usually interpreted as the second Hohenberg–Kohn theorem [3]. In the density functional theory, there exist two rigorous constructions of the universal energy density functionals based on their own rigorous proofs of the Hohenberg–Kohn theorem. This is the Levy–Lieb energy density functional [49–54] and the universal energy density functional based on the group of the local-scaling deformations in $\Re^3$ consisting of topological deformations mapping or topologically deforming any pair of one-electron densities to each other [4] (see also Sect. 5). The related Jacobian of such deformation gives rise to the first-order nonlinear differential equation, a so-called "Jacobian equation" [55–63] whose solution, namely, the corresponding deformation, does exist within the present approach and it is unique [4]. Solving the "Jacobian equation" enables to determine the deformation for any pair of well-behaved one-electron densities and to consistently extend the action of the local-scaling deformation group onto $\mathcal{H}_N$ [49]. This larger local-scaling transformation group partitions $\mathcal{H}_N$ into disjoint classes, orbits. All orbits exhaust $\mathcal{H}_N$ and within each orbit, there establishes the one-to-one correspondence between its wavefunctions and the one-electron densities. That is, these orbits are endowed with the characteristic that there are no two wavefunctions belonging to the same orbit that have the same density. Each orbit, say $\mathcal{O}^{[\alpha]}$, is determined by its

generator wavefunction $\Psi_g^{[\alpha]}$. Therefore, for a given orbit $\mathcal{O}^{[\alpha]}$, one defines the energy density functional $E_\alpha[\rho(\mathbf{r})]$ as merely the restriction of the energy functional $E[\Psi]$ on those wavefunctions that belong to the αth orbit, $\Psi \in \mathcal{O}^{[\alpha]}$. It is trivial to prove that, first, there are as many different energy density functionals as the orbits in $\mathcal{H}_N$, and second, that each density functional $E_\alpha[\rho(\mathbf{r})]$ implicitly depends on the generator wavefunction $\Psi_g^{[\alpha]}$. Evidently, the true ground-state N-electron wavefunction of the given Hamiltonian operator H belongs to a certain orbit, called the Hohenberg–Kohn one, $\mathcal{O}^{[\mathrm{HK}]}$ [4]. Within $\mathcal{O}^{[\mathrm{HK}]}$, the Levy–Lieb energy density functional [49] exactly coincides with $E_{HK}[\rho(\mathbf{r})]$ that is defined within the local-scaling deformation approach [64–75]. The explicit form of any energy density functional $E_\alpha[\rho(\mathbf{r})]$ for any α has been obtained and the corresponding variational Euler–Lagrange equation has been also derived in [64–75] (see also Chaps. 7 and 8 of [4]). The rigorous mathematical framework of the local-scaling deformation approach to the density functional theory based on the "Jacobian equation" has recently been elaborated in [55, 57–63]. The local-scaling deformation approach to the density functional theory has been also generalized to include spin densities, for the momentum space representation, for excited states, for the fractional occupation numbers, and finally to study nonadiabatic effects. The corresponding analogue of the Kohn–Sham approach has been also formulated in terms of the orbit generators (see Chap. 8 in [4]). A variety of theoretical and computational applications of the local-scaling deformation density functional theory to atoms and molecules has recently been elaborated as well (see [76–87] and references therein).

However, in the computational perspective, the Kohn–Sham DFT remains the most widely used many-body approach to conduct the electronic structure calculations of atoms, molecules, solids, and solid surfaces. In the context of this DFT, only the exchange-correlation energy must be approximated as a functional of the electron density [8]. The simplest approximation is the local spin density (LSD) approximation [88] which uses the local electron spin densities $\rho_\uparrow(\mathbf{r})$ and $\rho_\downarrow(\mathbf{r})$, as in (9), as the only ingredients. In the development of DFT, a significant step was the introduction of the density gradients $\nabla_{\rho\uparrow}(\mathbf{r})$ and $\nabla_{\rho\downarrow}(\mathbf{r})$ as additional local ingredients in the generalized gradient approximation or shortly GGA [89–91].

3 Dispersion Molecular Forces: Introduction

Some people think that there is an air between molecules.
A. S. Kompaneets [92]

According to our current knowledge, there exist four basic types of forces: gravitational, electromagnetic, strong, and weak. The electromagnetic force is the force that binds the electrons and the nuclei inside the atom, and the atoms inside a molecule, and that governs the interaction between atoms and molecules. These are referred to as the so-called intermolecular interactions [92–94, 186, 187]. Indeed,

Table 1 Van der Waals radii R_{vdW} (in Å) of atoms

Atom	R_{vdW}	Atom	R_{vdW}	Atom	R_{vdW}	Atom	R_{vdW}	Atom	R_{vdW}	Atom	R_{vdW}
H	1.20	He	1.30	N	1.50	O	1.40	F	1.35	Ne	1.40
P	1.90	S	1.85	Cl	1.80	Ar	1.70	As	2.00	Se	2.00
Br	1.95	Kr	1.80	Sb	2.20	Te	2.20	I	2.15	Xe	2.05

intermolecular interactions play the important role in the world: it is true that we "touch" them everywhere, at a macroscopic scale, in our everyday life. It is rather hard to imagine what would be the world without them—obviously, the situation would be significantly more dramatic than to have the ideal gases that surround us! Simply imagine that would be no such molecules as the DNA and RNA. Many chemical and biochemical processes involve van der Waals interactions [93].

The intuitively clear idea of that atoms and molecules do interact with each other is a very old indeed and dates back to the ancient times. Recall in particular Democritus, Leukippos, and Lucretius whose philosophical thoughts on the interaction through a direct contact were developed and rationalized by R. Boscovich in "Theory of Natural Science Reduced to the Single Law of Forces Existing in Nature" (1758) and independently by A. C. Clairault in "Théorie de la Figure de la Terre" (1743).

In Nature, atoms are located at different interatomic distances depending on a kind of the forces between them: either by cohesion forces or chemical bonds. The latter prevail at the distances which are smaller or equal to the sum of van der Waals radii of atoms. At such distances atoms form a molecule. By definition, the van der Waals (vdW) radii of a given atom is the half of the shortest distance that is observed in crystals between the nuclei of the same atoms. The vdW radii of atoms are listed in Table 1. At the distances beyond the sum of van der Waals radii of atoms, there exists a specific van der Waals interaction often referred to as the dispersion interaction between atoms, after Johannes Diderik van der Waals who first postulated its existence in his well-known equation of state derived in his PhD thesis in 1837 and which won him the 1910 Nobel Prize in Physics.[1] For the first time van der Waals explained the deviations of gases from the ideal behavior. Let us consider a vessel filled by a gas of atoms. Within this vessel, the pressure exerted by a gas of atoms on its wall is lower compared to that predicted by the ideal gas law since the atoms may collide with the wall and are thus retained by the attraction they undergo from the other atoms in the bulk of the gas that results in the pressure P obeying the equation [94],

$$P = \left(\frac{nRT}{V - nb}\right) - a\left(\frac{n^2}{V^2}\right). \tag{26}$$

Here a is a so-called vdW factor and $b = (16\pi N/3)R_{vdW}^3$.

[1] http://www.nobelprize.org/nobelprizes/physics/laureates/1910/waals-bio.html.

Table 2 Dipole moments d^o of some molecules in D

$\mathcal{M}$	d^o	$\mathcal{M}$	d^o	$\mathcal{M}$	d^o
n-Butane	0.00	Pyridine	2.23	*n*-Pentane	0.00
n-Hexane	0.00	Cyclohexanone	2.90	Acetone	2.900
Benzene	0.00	Propionitrile	3.20	Dimethylacetamide	3.70
Cyclohexane	0.00	Nitroethane	3.2	Lithium fluoride	6.40
Toluene	0.36	Dimethylsulfoxide	3.92	Lithium chloride	7.09
Triethylamine	0.78	Diethylether	1.21	Tetrahydrofurane	1.76
Hydrogen chloride	1.10	Methylacetate	1.80	Water	1,84
Propylenecarbonate	4.98	Natrium chloride	9.06	Potassium chloride	10.70

It merits to notice that the dispersion interaction energy between the ground-state molecules is always negative. Therefore, such molecules always attract each other—in this regard it is appropriate to notice the celebrated work by E. H. Lieb and W. E. Thirring [95] that proves that the attractive van der Waals force between an atom and a molecule exists only at some mutual orientations and was generalized to any orientation in 2006 by M. Lewin [96]. The leading term of the dispersion energy is the dipole–dipole term which is proportional to $1/R^6$ and determined by a change in the zero-point vibrational energy of electric field created by zero-point vibrations of fluctuating dipole moments of interacting species. Since the zero-point fluctuations are the quantum phenomenon, the dispersion interactions have the quantum origin, as though in the beginning of the 1970s, T. H. Boyer [97] derived the London formula for the dispersion interactions within the classical electrodynamics, additionally assuming the existence of the classical electromagnetic zero-point radiation (Table 2). In fact, the van der Waals forces are cohesive attractions between molecules which operate at long intermolecular separations. From a quantitative viewpoint, van der Waals forces between molecules correspond to interactions between electric dipoles. Generally speaking, there exist three types of electric dipoles in molecules. These are permanent, induced, and temporary dipoles. If a molecule $\mathcal{M}$ under study consists of the positive nuclear charges $q_1 = Z_1, \ldots, q_M = Z_M$ and negative electron charges $e_1, \ldots, e_N$, its total permanent dipole moment is defined as

$$\mathbf{d}^o = \sum_{\alpha=1}^{M} q_\alpha \mathbf{R}_\alpha + \sum_{i=1}^{N} e_i \mathbf{r}_i \tag{27}$$

d^o is distinct from zero in some state if the center of charge of the nuclei, $\mathbf{R}_n \equiv \sum_{\alpha=1}^{M} q_\alpha \mathbf{R}_\alpha \Big/ \sum_{\alpha=1}^{M} q_\alpha$, does not coincide with that of the electron subsystem, $\mathbf{r}_n \equiv \sum_{i=1}^{N} e_i \mathbf{r}_i \Big/ \sum_{i=1}^{N} e_i$. If $d^o \neq 0$, a molecule $\mathcal{M}$ is called polar. Permanent dipole moments of neutral molecules usually vary from 0 to 15 Debye (D) that is reflected in Table 3.[2]

[2] Al. McClellan, Tables of Experimental Dipole Moments. Rahara Enterprises, E1 Cerrito, 1974.

Table 3 Intramonomer distances (in Å) and binding energies (in kcal/mol), below, of representative van der Waals complexes, Ne_2, Ar_2, $(CH_4)_2(D_{3d})$, $(C_2H_4)_2(D_{2d})$, T-$(C_6H_6)_2$, SP-$(C_6H_6)_2$, P-$(C_6H_6)_2$, taken from Tables 1 and 2 of [98]

DFT	Ne_2	Ar_2	$(CH_4)_2$	$(C_2H_4)_2$	T-$(C_6H_6)_2$	SP-$(C_6H_6)_2$	P-$(C_6H_6)_2$
VSXC	2.47	3.59	3.21	3.35	4.65	3.47	3.41
	0.49	1.15	4.33	9.96	8.28	16.95	11.23
PW91	3.00	4.0	4.0	4.0	5.15	5.76	4.22
	0.97	0.36	0.52	1.56	1.65	1.55	0.24
HCTH407	2.58	3.84	3.89	4.05	5.45	5.84	4.22
	0.76	0.48	0.71	1.53	1.41	1.63	0.67
PBE	2.52	3.88	3.86	3.76	5.19	5.76	4.31
	0.69	0.19	0.29	1.30	1.33	1.27	−0.22
PBE1PBE	2.54	3.93	3.90	3.75	5.14	5.76	4.40
	0.53	0.13	0.20	1.16	1.32	1.15	−0.42
BHANDHLYP	2.48	3.93	3.99	3.76	5.20	5.77	U
	0.81	0.03	0.02	0.91	1.00	0.90	
MPW1K	2.62	4.31	4.35	3.82	5.19	5.81	U
	0.34	0.06	0.08	0.69	1.00	0.80	
MPW1PW91	2.61	4.30	4.37	3.88	5.27	5.83	U
	0.40	0.07	0.10	0.62	0.89	0.75	
B3LYP	2.51	U	U	3.85	5.37	5.81	U
	0.63			0.55	0.63	0.60	
B3P86	2.64	U	U	3.76	5.21	5.78	U
	0.04			0.47	0.64	0.48	
CCSD(T)	3.2[a]	3.8[a]	3.8[a]	3.8[a]	4.89[b]	3.8[b]	3.7[b]
	0.07	0.26	0.495	1.15	2.74	2.78	1.81

U implies that at this computational level, the studied complex is unbound. The notations [a][99] where the computational level CCSD(T) and the aug-cc-pV5Z basis set are used, and [b][100], and U for unbounded. The experimental value of the bound distance in Ne_2 dimer is equal to 3.09 Å. The benzene dimer exists in three isomeric structures *T*—T-shaped, *P*—parallel, and *SP*—slipped parallel

Any pair of polar molecules, $\mathcal{M}_1$ and $\mathcal{M}_2$, separated by a distance R_{12}, in the states n and m, respectively, interact with each other by their dipoles, $\mathbf{d}_n^{(1)}$ and $\mathbf{d}_m^{(2)}$, via the dipole–dipole interaction read as

$$E_{\mathrm{dd}}^{(1)} = \frac{\mathbf{d}_n^{(1)}\mathbf{d}_m^{(2)}}{R_{12}^3} - \frac{3(\mathbf{d}_n^{(1)}\mathbf{R}_{12})(\mathbf{d}_m^{(2)}\mathbf{R}_{12})}{R_{12}^5}. \tag{28}$$

Structurally, $E_{\mathrm{dd}}^{(1)}$ consists of two terms. A polar molecule $\mathcal{M}_1$ interacts with the electric field $\mathbf{E}^{(2)}(\mathbf{R}_{12})$ created by another molecule $\mathcal{M}_2$ at the position of the first molecule. As known from electrostatics [101–103], molecule $\mathcal{M}_1$ gains the energy $-\mathbf{d}_n^{(1)}\mathbf{E}^{(2)}(\mathbf{R}_{12})$. Expressing the electric field

$$\mathbf{E}^{(2)}(\mathbf{R}_{12}) = \nabla[(\mathbf{d}_m^{(2)}\mathbf{R}_{12})/R_{12}^3] \tag{29}$$

results then in Eq. (28).

Let us consider another, so-called, second-order effect of an external electric field $\mathbf{E}$ on a given molecule $\mathcal{M}_2$. This field influences the molecular charges, electrons, and nuclei, causing their displacements, and as a result, there appears the induced dipole moment $\mathbf{d}^{\text{ind}}$,

$$\mathbf{d}^{\text{ind}} = \alpha \mathrm{E}, \tag{30}$$

where the proportionality coefficient is merely a polarizability α of a given molecule. Assuming that this electric field is generated by the presence of the second molecule $\mathcal{M}_1$, one obtains

$$\mathbf{d}^{\text{ind}}_{2,m} = \alpha \nabla[(\mathbf{d}^{(1)}_m \mathbf{R}_{12})/R^3_{12}]. \tag{31}$$

Therefore, the interaction of the permanent dipole $\mathbf{d}^{(1)}_n$ of $\mathcal{M}_1$ with the dipole $\mathbf{d}^{\text{ind}}_{2,n}$ that is induced on $\mathcal{M}_2$ by $\mathcal{M}_1$ takes the following expression:

$$E^{(2)}_{\text{dd}} = \frac{\mathbf{d}^{(1)}_n \mathbf{d}^{\text{ind}}_{2,n}}{R^3_{12}} - \frac{3(\mathbf{d}^{(1)}_n \mathbf{R}_{12})(\mathbf{d}^{\text{ind}}_{2,n} \mathbf{R}_{12})}{R^5_{12}} \tag{32}$$

that is known as the Keesom dipole–dipole interaction [104].

By a straightforward analogy, the dipole–dipole interaction of two mutually induced dipoles on $\mathcal{M}_1$ and $\mathcal{M}_2$ is described by the expression

$$E^{(2)}_{\text{dd}} = \frac{\mathbf{d}^{\text{ind}}_{1,n} \mathbf{d}^{\text{ind}}_{2,m}}{R^3_{12}} - \frac{3(\mathbf{d}^{\text{ind}}_{1,n} \mathbf{R}_{12})(\mathbf{d}^{\text{ind}}_{2,n} \mathbf{R}_{12})}{R^5_{12}}. \tag{33}$$

If the distance R_{12} between dipoles is small enough compared to the wavelength λ, corresponding to transitions between the ground and excited states, within the second- and higher-order Rayleigh–Schrödinger perturbation theory as [92, 105, 106] there appear, as first shown by Fritz London [105, 106], the dispersion interaction [105, 106]

$$E^{(2)}_{\text{disp}} = -\sum_{m,n\neq 0} \frac{\left|\left\langle \Psi^{(1)}_n \Psi^{(2)}_m | V_{12} | \Psi^{(1)}_0 \Psi^{(2)}_0 \right\rangle\right|^2}{(E^{(1)}_n - E^{(1)}_0) + (E^{(2)}_m - E^{(2)}_0)} \tag{34}$$

where $\Psi^{(i)}_n$ is the n-state eigenfunction of $\mathcal{M}_i, i = 1, 2$ and V_{12} is the electrostatic interaction between molecules $\mathcal{M}_1$ and $\mathcal{M}_2$.

The dispersion energy is traditionally represented by means of the multipole expansion [94]

$$E^{(2)}_{\text{disp}} = -\sum_{n=6}^{\infty} \frac{C_n}{R^n}, \tag{35}$$

where C_n are dispersion constants among which C_6 corresponds to dipole–dipole interaction, C_8 to dipole–quadrupole and C_{10} to dipole–octupole and dipole–quadrupole interactions.

Dispersion interactions play a role of the attractive interaction between rare gas atoms and are also one of the important intermolecular interactions that govern the molecular organic world [93, 107]. Dispersion interactions are mostly responsible for the heats of sublimation of hydrocarbon molecules, make significant contributions to the solvent properties of polar and apolar neutral compounds [108, 109] and are also important for crystal packing of organic molecules [110] as well as for the stacking of nucleic acids in DNA [93, 111]. The world of dispersion interaction is rich (see [112] and references therein), despite the fact that it is a weaker form of intermolecular attractions. Dispersion forces as one of the two types of van der Waals force are also known as "London forces," named after Fritz London [105, 106].

Density functional theory [113–116] as one of the approaches to evaluate electron correlation is considerably less demanding on computational resources than the MP2 or CCSD(T) methods. DFT might therefore be considered as a powerful computational tool, if it can adequately describe and accurately evaluate intermolecular interactions. The suitability of DFT for the evaluation of dispersion interaction has been an important issue in the recent literature [117–125] that is mirrored in Table 3. While DFT calculations with local exchange-correlation functionals lead to overestimate binding energies of weakly bound systems, it was reported that non-local exchange-correlation functionals very often underestimate the attraction [118–120]. The DFT calculations with Becke's exchange and Lee, Yang, and Parr's correlation functionals, BLYP [113, 114], and Becke's three-parameter functional combined with Lee, Yang, and Parr's correlation functional, B3LYP [114, 116], also fail to evaluate the attractive dispersion interaction between hydrocarbon molecules [120–122]. Recently Zhang et al. and Wesolowski et al. reported that the Becke exchange functional due to its erroneous asymptotic behavior at low density is responsible for the failure in the evaluation of the attraction between weakly bound systems [123–125]. It was shown, however, that other non-local exchange-correlation functionals such as Perdew and Wang's exchange and correlation functionals, PW91 [115], are possible alternatives to describe the binding between rare gas dimers or other systems. The performance of some exchange-correlation functionals and the PW91 one in particular for the representative van der Waals systems is demonstrated in Table 3. Notice that the PW91 functional is a general functional, i.e., it is not biased towards the description of intermolecular interaction. In this investigation we will examine the basis set dependence of the interaction energies, and benchmark those against the results obtained from MP2 [188] and CCSD(T) theory. For comparison, the results with the BLYP and B3LYP are also presented. Interestingly, DFT adequately describes, on the one hand, atoms and molecules as stable many-electron systems and on the other hand, the molecules formed under interaction of its composing molecules. However, its description of those molecular interactions is not always perfect. The simplest DFT approximation widely used in computational practice is the local

density approximation (LDA) [2], based on the properties of the uniform electron gas. In principle, DFT yields the exact ground-state energy, including long-range van der Waals energies, very important in organic chemistry and elsewhere. However, the commonly used LDA and GGA, designed for nonuniform electron gases, fail to capture the essence of vdW energies. The latter reflect correlated motions of electrons due to the Coulomb interactions between distant, even non-overlapping atoms, molecules, and solids. In [4] Kohn and coauthors propose a first-principles approach, which contains the following ingredients (i) The density distribution, $\rho(\mathbf{r})$, is approximated by the LDA or GGA. (ii) The Coulomb interaction is divided into short- and long-range parts, of which only the latter contributes to vdW energies. (iii) The contribution of the long-range interactions to the energy is expressed by the so-called adiabatic connection formula. (iv) This expression is transformed into the time domain, avoiding the need to solve a self-consistent equation for the density–density response function.

4 Dispersion-Corrected DFT Approaches

DFT methods with currently available functionals failed completely for London-type clusters for which no minimum was found at the potential energy surfaces.
P. Hobza, J. Reschel and J. Sponer (1995) J Comput Chem 16:1315

Density functional theory is often the preferred electronic structure method to study moderate and large systems. This preference reflects the efficiency of DFT compared to correlated wavefunction theories such as coupled cluster theory even though accuracy, and more importantly, predictability (as in systematic convergence to the right answer) are sacrificed. DFT which incorporates currently accepted exchange-correlation functionals can be used with reasonable reliability on chemically bound systems around the equilibrium geometry but inevitably fail when applied to systems which are bound by weak van der Waals forces [126–130] and to a lesser degree for chemically bound systems away from equilibrium, like transition states [131]. These failings of density functional theory are well known [132]. Here, we address weak interactions. Attempts to compute weak intermolecular forces using DFT fall into two categories. Some would simply modify functionals until reasonable results are obtained (see [133] and references within). Others would focus on an add-on correction that explicitly introduces the van der Waals C_6 coefficient. This can be made to work, but it is unsatisfactory that the weak interactions do not occur naturally as they would in wavefunction methods. This is the experimental or computational fact which has not been still proved. Though, the problem of description of London dispersion in density functional theory (DFT) using (semi) local exchange-correlation functionals is a well-known problem [134, 135] since the first diagnostic in 1994 [134].

A step in the right direction was made by Engel et al. [136] who obtained reasonable results for the helium and neon dimers. In [137], Bartlett and

co-workers proposed ab initio density functional theory has been applied for the weakly interacting, He_2, $[He–Be]^{2+}$, Ne_2 and Be_2 that results in fair agreement with the highly accurate coupled-cluster method. Generally, one assumes that the cause lies in the local character of the widely used correlation functionals, which, in contrast to the correlation contribution in post-Hartree–Fock methods such as Møller–Plesset or coupled cluster, only utilize information on the density of the system at one point and are therefore unsuitable for the description of a non-local phenomenon such as dispersion. Attempts to introduce non-local correlation to DFT, such as the random phase approximation (RPA) [138, 139] or the non-local van der Waals functionals [140–143], are being investigated, but unfortunately the improvement comes with a significant increase in the computational cost. Since the relatively low computational cost of DFT is one of the major factors responsible for its status as the most widely used quantum chemical method today, a range of more pragmatic approaches has been developed to correct the performance of DFT for dispersion interactions. Part of these methods rely on reparametrization of existing local correlation functionals [144–146], motivated by the fact that dispersion is partially included in many functionals and that a suitable reparametrization will allow one to achieve the aspired results more consistently. The drawback of such an approach is that the strong empirical character decreases the reliability. For instance, the performance of the reparametrized functionals often decreases for properties other than the electronic energy. Other attempts are based on adding a correction term, representing the dispersion energy, to the energy calculated using standard DFT methods. Also in this category, one can find highly empirical but computationally attractive methods [147] based on parameters fitted to reproduce high-level results, as well as the methods with deeper theoretical foundation but computationally more expensive, where ab initio information of the systems is used to evaluate the dispersion energy, such as the static or frequency dependent polarizabilities [148–154] or the dipole moment of the exchange-correlation hole (XDM) [156–159]. Another noteworthy approach is the adaptation of the symmetry adapted perturbation theory [160] to the framework of DFT, i.e., SAPT(DFT) [161–164]. SAPT(DFT) has a significant computational advantage against the highly scaling SAPT as the contribution of intramonomer correlation, already embedded within the Kohn–Sham orbitals, does not need to be evaluated. Although possible to use for the correction of DFT dispersion energies [165], SAPT(DFT) is mostly meant for an evaluation of the total interaction energy. The explicit expression for the repulsive contribution of electron-exchange to the dispersion energy within SAPT(DFT), though rarely calculated fully due to the computational expense, offers a more theoretically attractive alternative to the empirical damping functions used in other methods. SAPT(DFT) does have the disadvantage of requiring explicit separation of the system in two parts, which makes it impossible for application on intramolecular dispersion interactions, such as those occurring, for example, in biomolecules. Table 4 reports some developed dispersion-corrected DFT functionals and their performance for the representatives of the van der Waals dimers.

Table 4 Dispersion-corrected DFT functionals

Method	He_2	He–Ne	Ne_2	He–Ar	Ar_2	$(CH_4)_2$	$(C_2H_4)_2$	T-$(C_6H_6)_2$
CCSD(T)	−0.02	−0.04	−0.07	−0.06	−0.27	−0.53	−1.51	−2.74
D-B3LYP/A[a]	0.04	0.04	0.05	0.07	0.17	0.38	0.49	0.98
D-BH-B3LYP/A[a]	−0.00	−0.02	−0.05	−0.02	−0.19	−0.49	−1.68	−2.64
D-B3LYP/B[a]	0.05	0.04	0.03	0.08	0.07	0.76	0.40	0.96
D-BH-B3LYP/B[a]	0.03	0.02	−0.03	0.03	−0.08	−1.12	−0.87	−1.89
DFT-B97-D/C[b]			−0.17		−0.26			−2.99

Basis sets A ≡ aug-cc-pVTZ, B ≡ 6-311++G(2d,f,p), and C ≡ TZV2P
[a][166].
[b][167].

One of the most promising new DFT methods is the non-local van der Waals correlation functional vdW-DF-04 in [168, 169], which was derived from first principles, describes dispersion interactions in a seamless fashion, and yields the correct asymptotic form of intermolecular van der Waals forces. Recently we reported a self-consistent implementation of vdWDF-04 with Gaussian basis functions [168, 169]. The code includes analytic gradients of the energy with respect to nuclear displacements, enabling efficient geometry optimizations. The alternative and consistent approach to the dispersion-corrected DFT can be formulated, as we believe, within the so-called local-scaling transformations' (LST') DFT whose key features are given in the next section.

5 Local-Scaling Transformations' DFT

In order to properly assess the local-scaling transformation formulation of the density functional theory, we first consider the concept of local-scaling transformation and second, apply it to the topological features of atomic and molecular one-electron densities.

5.1 Mathematical Preliminaries: Local-Scaling Transformations

Define on the Euclidean $\mathbb{R}^3$ the following mapping: $\mathbb{R}^3 \xrightarrow{f} \mathbb{R}^3$ such that $r \in \mathbb{R}^3$ is mapped into

$$\mathbf{f}(\mathbf{r}) := f(\mathbf{r})\mathbf{e_r} = f(r;\mathbf{e_r})\mathbf{e_r}, \tag{36}$$

where $\mathbf{e_r} \equiv \mathbf{r}/r \equiv \mathbf{e}(\Omega)$ is a unit vector, specified in $\mathbb{R}^3$ and defined by the spherical angles $\Omega = (\theta, \phi)$ and $r = |\mathbf{r}|$. For a given $\mathbf{e_r}(\Omega)$, the transformation (36) that deforms $\mathbb{R}^3$ onto itself, nonuniformly in general, is referred to as a local-scaling transformation or LST for short [4] and is the special class of point transformation

[170, 171]. LSTs satisfy all axioms of group and hence form the group $\mathcal{F}$ of local-scaling transformations. A scalar function $f(\mathbf{r})$ in Eq. (36) can be arbitrary, though often it belongs to $\mathcal{C}^1$ or higher. In the former, $\mathbf{f}$ is a $\mathcal{C}^1$-diffeomorphism on $\mathbb{R}^3$.

Equation (36) nontrivially generalizes the well-known scaling: $\mathbf{f}_\lambda(\mathbf{r}) := \lambda\mathbf{r}$ which Fock [172] used in 1930 to prove the virial theorem. $\lambda \neq 0$ is a constant that means that all vectors $\mathbf{r} \in \mathbb{R}^3$ are scaled uniformly. Bearing in mind that an arbitrary vector $\mathbf{r}$ is uniquely determined by its Cartesian coordinates $\mathbf{r} = (x, y, z)$, the equivalent representation of (36) is the following:

$$\mathbf{r} \equiv \begin{pmatrix} x \\ y \\ z \end{pmatrix} \xrightarrow{f} \mathbf{f}(\mathbf{r}) \equiv \begin{pmatrix} \frac{x}{r} f(x,y,z) \\ \frac{y}{r} f(x,y,z) \\ \frac{z}{r} f(x,y,z) \end{pmatrix} \equiv \begin{pmatrix} f_x(\mathbf{r}) \\ f_y(\mathbf{r}) \\ f_z(\mathbf{r}) \end{pmatrix} \equiv \begin{pmatrix} x\sigma(\mathbf{r}) \\ y\sigma(\mathbf{r}) \\ z\sigma(\mathbf{r}) \end{pmatrix}, \tag{37}$$

where $\mathbf{f}(\mathbf{r}) = \sigma(\mathbf{r})\mathbf{r}$.

The Jacobian of (1) is defined as

$$\begin{aligned} J\{\mathbf{f}(\mathbf{r});\mathbf{r}\} \equiv J\{\mathbf{f};\mathbf{r}\} &= \begin{vmatrix} \frac{1}{r}f - \frac{x^2}{r^3}f + \frac{x}{r}\frac{\partial f}{\partial x} & -\frac{xy}{r^3}f + \frac{x}{r}\frac{\partial f}{\partial y} & -\frac{xz}{r^3}f + \frac{x}{r}\frac{\partial f}{\partial z} \\ -\frac{xy}{r^3}f + \frac{y}{r}\frac{\partial f}{\partial x} & \frac{1}{r}f - \frac{y^2}{r^3}f + \frac{y}{r}\frac{\partial f}{\partial y} & -\frac{yz}{r^3}f + \frac{y}{r}\frac{\partial f}{\partial z} \\ -\frac{xz}{r^3}f + \frac{z}{r}\frac{\partial f}{\partial x} & -\frac{yz}{r^3}f + \frac{z}{r}\frac{\partial f}{\partial y} & \frac{1}{r}f - \frac{z^2}{r^3}f + \frac{z}{r}\frac{\partial f}{\partial z} \end{vmatrix} \\ &= \frac{f^2}{r^3}\left(x\frac{\partial f}{\partial x} + y\frac{\partial f}{\partial y} + z\frac{\partial f}{\partial z}\right) = \frac{1}{3r^3}\mathbf{r}\cdot\nabla f^3. \end{aligned} \tag{38}$$

In terms of $\sigma(\mathbf{r})$, the Jacobian (38) has the form $J\{\mathbf{f}(\mathbf{r});\ \mathbf{r}\} = \sigma(\mathbf{r})[1 + \mathbf{r}\cdots\ln \sigma(\mathbf{r})]$ [171]. For the uniform scaling $\mathbf{f}_\lambda := =$, the corresponding Jacobian is equal to

$$J\{\mathbf{f}_\lambda;\mathbf{r}\} = \begin{vmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{vmatrix} = \lambda^3. \tag{39}$$

It is trivial to generalize a three-dimensional local-scaling transformation (1) on other dimensions, say $\mathbb{R}^D$, simply by considering a vector $\mathbf{r}$ as a D-dimensional one. If $D = 1$, $f(r)$ is a function of a single variable r. The corresponding Jacobian $J\{f(r);\ r\} = \mathrm{d}f(r)/\mathrm{d}r$. Let us consider some examples of local-scaling transformations $\mathbf{f}_{\mathrm{LST}}$:

(1) $f^{[1]} = \left[\left(\frac{1}{r}\right)^m + \left(\frac{\delta}{\sqrt{r}}\right)^m\right]^{-1/m}$, where δ is the constant [173].

(2) $f^{[2]} = \begin{cases} r(1+ar^2)^{1/3} & \text{if } r \leq R \\ \sqrt{\frac{d_{-2}}{r^2} + \frac{d_{-1}}{r} + d_o + d_1 r + d_L \ln r} & \text{otherwise.} \end{cases}$

This form results from the asymptotes at small and large r [174].

(3) $f^{[3]} = \begin{cases} r & \text{if } r \leq R \\ a\sqrt{\frac{8r}{a} - \frac{8a}{r} + \frac{a^2}{r^2} - 12\ln\left(\frac{r}{a}\right)} & \text{otherwise} \end{cases}$ [8].

(4) Let $\Omega :=]-L/2, L/2[^3 \subseteq \mathbb{R}^3$ be a cube with volume $|\Omega| = L^3$. $\mathbf{f}^{[4]}$ is defined as a periodic deformation on the cube Ω if it is a $\mathcal{C}^1$ -diffeomorphism on $\mathbb{R}^3$ that leaves Ω invariant: $\mathbf{f}^{[4]}(\Omega) = \Omega$ and if $\mathbf{f}^{[4]}(\mathbf{r} + L\mathbf{m}) = \mathbf{f}^{[4]}(\mathbf{r}) + L\mathbf{m}$ for any $\mathbf{m} \in \mathbb{Z}^3$ [175].

(5) $f^{[5]}_{p,q,r}$ is defined by the inverse function $r\left(f^{[5]}_{p,q,r}\right) = \left[f^{[5]}_{p,q,r}\right]^p \left(1 + \alpha\left[f^{[5]}_{p,q,r}\right]^q\right)^r$ where α, p, q, and r are variational parameters. If $p = q = r = 1$, $r\left(f^{[5]}_{p,q,r}\right)$ refers to Hall's transformation [176]. The other $r\left(f^{[5]}_{p,q,r}\right)$ with $q = r = 1$, $p = r = 1$ and $p = q = 1$ were defined in [177, 178]. Hall's local-scaling transformation is then $f^{[5]}_{1,1,1} = [(1 + 4\alpha r)^{1/2} - 1]/(2\alpha)$.

Let $\phi(\mathbf{r})$ be an arbitrary function given on domain $\Sigma \subset \mathbb{R}^3$. A local-scaling transformation (16) transforms $\phi(\mathbf{r})$, generally speaking, into another function

$$\psi(\mathbf{r}) := \phi(\mathbf{f}(\mathbf{r})) \tag{40}$$

within the Jacobian (3), depending on the normalization of $\phi(\mathbf{r})$ on Σ if any. If $\phi(\mathbf{r}) = \exp(-r)$ is a simple exponential orbital, under Hall's local-scaling transformation it converts to

$$\psi(\mathbf{r}) = \frac{(1 + 4\alpha r)^{1/2} - 1}{2\alpha r(1 + 4\alpha r)^{1/4}} \exp\left(-[(1 + 4\alpha r)^{1/2} - 1]/(2\alpha)\right) \tag{41}$$

that was used to approximate the $1s$ orbital.

5.2 *One-Electron Densities: Definition*

A function $\rho(\mathbf{r}) : \mathbb{R}^3 \to \mathbb{R}^1_+$ is defined as a one-electron density associated with some system of N electrons iff:

(**Di**) $\rho(\mathbf{r})$ is non-negative everywhere in $\mathbb{R}^3$;
(**Dii**) $\rho(\mathbf{r})$ is normalized to the total number N of electrons,

$$\int_{\mathbb{R}^3} \mathrm{d}^3\mathbf{r}\rho(\mathbf{r}) = N. \tag{42}$$

Here $\mathbb{R}^1_+$ stands for the non-negative semi-axis of $\mathbb{R}^1$. Equation (42) merely implies that the square root of $\rho(\mathbf{r})$ is a square-integrable function, i.e., $[\rho(\mathbf{r})]^{1/2} \in L^2(\mathbb{R}^3)$;
(**Diii**) $\rho(\mathbf{r})$ is a continuously differentiable function of $\mathbf{r}$ almost everywhere in $\mathbb{R}^3$. It is a well-behaveness of densities.

Let $\mathcal{D}_N$ be the class of the one-electron densities associated with a Coulomb system of N electrons and obeying the conditions (**Di**)–(**Diii**). Obviously, $V\mathcal{L}_N \subset \mathcal{D}_N$. The fact that the condition (**Diii**) is valid for $V\mathcal{L}_N$ is the consequence of the following

Proposal 1. For any $\rho_\Psi(\mathbf{r}) = V\Psi$ where $\Psi \in \mathcal{L}_N, \nabla_{\mathbf{r}}\rho_\Psi(\mathbf{r}) \in \mathcal{L}^2(\mathbb{R}^3)$.

Proof. [88]: According to the Schwarz inequality, it follows from Eq. (11) that

$$[\nabla_{\mathbf{r}}\rho_\Psi(\mathbf{r})]^2 \leq 4N\rho_\Psi(\mathbf{r}) \sum_{s_1,\ldots,s_N} \int \mathrm{d}^3\mathbf{r}_2 \cdots \int \mathrm{d}^3\mathbf{r}_N |\Psi(\mathbf{r}, s_1; \mathbf{r}_2, s_2; \ldots; \mathbf{r}_N, s_N)|^2. \quad (43)$$

□.

Corollary 1.1. $\nabla_{\mathbf{r}}[\rho_\Psi(\mathbf{r})]^{1/2} \in \mathcal{L}^2(\mathbb{R}^3)$.

Proof. [10]:

$$\int \mathrm{d}^3\mathbf{r}(\nabla_{\mathbf{r}}[\rho_\Psi(\mathbf{r})]^{1/2})^2 = \frac{1}{4}\int \mathrm{d}^3\mathbf{r}\frac{[\nabla_{\mathbf{r}}\rho_\Psi(\mathbf{r})]^2}{\rho_\Psi(\mathbf{r})} \leq T_{\mathrm{e}}[\Psi]. \quad (44)$$

□.

The term $[\nabla_{\mathbf{r}}\rho_\Psi(\mathbf{r})]^2/\rho_\Psi(\mathbf{r})$ is known as the von Weizsäcker kinetic energy $t_{\mathrm{W}}[\rho_\Psi(\mathbf{r})]$ [4]. Hence, in the other words, **Corollary 1.1** tells that $t_{\mathrm{W}}^{1/2}[\rho_\Psi(\mathbf{r})]$ is square-integrable. Usually, the von Weizsäcker term is only a part of the total many-electron kinetic energy [4]. The exception is the Hartree–Fock 2-electron model systems for which $t_{\mathrm{W}}[\rho_\Psi(\mathbf{r})]$ is the exact kinetic energy.[3] We further have

Corollary 1.2. Thomas–Fermi [172, 179] one-electron density $\rho_{\mathrm{TF}}(\mathbf{r})$ is not N-representable.

Proof. According to [180, 181], $t_{\mathrm{W}}^{1/2}[\rho_{\mathrm{TF}}(\mathbf{r})]$ is not square-integrable. □.

Furthermore, the Thomas–Fermi energy density functional cannot be inserted in the density functional philosophy presented by the mappings (13) and (14) for all $\rho(\mathbf{r}) \in \mathcal{D}_{\mathrm{N}}$ since the ground-state energies of many Thomas–Fermi atoms and ions[4] lie below the exact ones.[5]

[3] Except H^- which is unstable within the Hartree–Fock picture since its Hartree–Fock ground-state energy is equal to -0.488 hartree and placed above $E_0[H] = -0.5$ hartree. Note that the exact ground-state energy of H^- is -0.5278 hartree.

[4] Thomas–Fermi molecules are unstable (see [189] and e.g., [4] and references therein).

[5] Some of the widely used density functionals predict the ground-state energies below the experimental ones. For example, the B3LYP density functional in conjunction with the 6-31+G(d, p) basis set yields the energy -0.500273 hartree $< E_0[H]$ [182]. The B3LYP and B3PW91 show a similar trend for the atoms Li, C, O, F, Na, and Mg, and diatomics O_2, F_2, and LiF [183]. This implies that the corresponding ground-state wavefunctions, if do exist, are not square-integrable.

Consider an N-electron atom or ion with the nucleus centered at the origin of the Cartesian coordinate system. Let $\rho(\mathbf{r}) \in \mathcal{D}_N$ be one-electron density associated with a given atom and $\rho(\mathbf{r}) = \{\rho(r, \mathbf{e_r}) | r \in \mathbb{R}^1_+, \Omega \equiv (\theta, \phi), 0 \leq \theta \leq \pi, 0 \leq \phi \leq \pi\}$ is merely a bundle of one-dimensional curves. $\tilde{\rho}_2(r)$

Let two densities $\rho_1(\mathbf{r})$ and $\rho_2(\mathbf{r})$ from $\mathcal{D}_N$ be given. Both of them are represented by the corresponding bundles of curves. Let us choose the unit vector $\mathbf{e_r}$ and in these bundles, the projections of $\rho_1(\mathbf{r})$ and $\rho_2(\mathbf{r})$ onto $\mathbf{e_r}$; the curves $\tilde{\rho}_1(r)$ and $\tilde{\rho}_2(r)$ which are, according to (**Diii**), continuously differentiable functions of $r = |\mathbf{r}|$. Hence, they are homotopically equivalent, or equivalently, there does exist such topological deformation that maps or deforms $\tilde{\rho}_1(r)$ into $\tilde{\rho}_2(r)$. Formally,

$$\tilde{\rho}_2(r) = J\{\mathbf{f}(\mathbf{r}); \mathbf{r}\}\tilde{\rho}_1(f(r, \mathbf{e_r})). \tag{45}$$

The Jacobian in (38) ensures the normalization (**Dii**) for both densities ρ_1 and ρ_2. Generalizing Eq. (24) over all directions in $\mathbb{R}^3$ results in that [7]

$$\rho_2(\mathbf{r}) = \frac{1}{3r^3}\mathbf{r} \cdot \nabla f^3 \rho_1(\mathbf{f}(\mathbf{r})). \tag{46}$$

To hold the electron–nuclear Kato cusp, the nuclear position is invariant of f. If f is a uniform scaling f_λ, the latter equation takes the form

$$\rho_\lambda(\mathbf{r}) = \lambda^3 \rho_1(\lambda \mathbf{r}). \tag{47}$$

Given $\mathbf{e_r}$, combining Eqs. (36) and (38) yields

$$\frac{\mathrm{d}f(r, \mathbf{e_r})}{\mathrm{d}r} = \frac{r^2 \rho_2(r, \mathbf{e_r})}{f^2 \rho_1(f(r, \mathbf{e_r}), \mathbf{e_r})} \tag{48}$$

or in spherical coordinates, along a chosen unit vector $\mathbf{e_r}$ determined by $\Omega = (\theta_0, \phi_0)$,

$$\frac{\mathrm{d}f(r, \theta_0, \phi_0)}{\mathrm{d}r} = \frac{r^2 \rho_2(r, \theta_0, \phi_0)}{f^2 \rho_1(f(r, \theta_0, \phi_0), \theta_0, \phi_0)}. \tag{49}$$

Equation (49), or (48), is the first-order nonlinear differential equation for deformation $f(\mathbf{r})$ for given densities ρ_1 and ρ_2. Due to (**Diii**), its solution exists and it is unique (see e.g., [4] and references therein). Therefore, for any pair of well-behaved densities, one enables to determine the deformation that transforms one of them into another. This means that F acts on $\mathcal{D}_N$ transitively, that is, in algebraic terminology, $\mathcal{D}_N$ is a single orbit with respect to $\mathcal{F}$. For a given and fixed density $\rho_1(\mathbf{r})$, defined hereafter as the generator density $\rho_g(\mathbf{r})$, Eq. (46) then implies:

Proposal 2. There exists the one-to-one correspondence between $\mathcal{F}$ and $\mathcal{D}_N$ that is explicitly expressed as $f \in \mathcal{F} \Leftrightarrow \rho_f^{[g]}(\mathbf{r}) := J\ \{\mathbf{f}, \mathbf{r}\}\rho_g(\mathbf{f}(\mathbf{r}))$.

In the integral form, Eq. (28) is as follows:

$$f(r,\theta_0,\phi_0)=\left[3\int_{r_0}^{r}\mathrm{d}r\,\frac{r^2\rho_2(r,\theta_0,\phi_0)}{\rho_1(f(r,\theta_0,\phi_0),\theta_0,\phi_0)}\right]^{1/3}. \tag{50}$$

Note that the rhs of (50) contains a cubic root that reflects that the group $\mathcal{F}$ of local-scaling transformations acts on $\mathbb{R}^3$. It is shown in [184] that the dimensionality D of $\mathbb{R}^D$ enters the corresponding Jacobian in the power D and, respectively, the corresponding integral form as $1/D$, this is on the one hand. On the other, there exists another remarkable facet of Eq. (28). This equation is well known in mathematics as the "Jacobian problem" ([55, 57–63, 190], see also [56]).

5.3 Many-Electron Wavefunctions and Concept of Orbit

To build the "variable mapping" (14), let us generalize the concept of the local-scaling transformations on $\mathcal{L}_N$. This is rather simple and straightforward. For this purpose, let us choose an arbitrary "reference" or generator wavefunction $\Phi_{\mathrm{g}}\left(\{\mathbf{r}_i,\sigma_i\}_{i=1}^{i=N}\right)$ where σ_i is the spin of the i-th electron and $\rho_{\mathrm{g}}(\mathbf{r})\in\mathcal{D}_N$ is the associated one-electron density. Then define a new wavefunction

$$\Phi_f(\{\mathbf{r}_i,\sigma_i\})=\Phi_\rho(\{\mathbf{r}_i,\sigma_i\})\equiv\left[\prod_{i=1}^{N}J(\mathbf{f}(\mathbf{r}_i);\mathbf{r}_i)\right]^{1/2}\Phi_{\mathrm{g}}(\{\mathbf{f}(\mathbf{r_i}),\sigma_{\mathrm{i}}\}) \tag{51}$$

with the density $\rho(\mathbf{r})\equiv\rho_{\mathrm{g}}(\mathbf{f}(\mathbf{r}))$ casting in **Proposal 2**. Φ_f is a locally scaled image of the "reference" wavefunction. Formally, $\Phi_f\equiv F\Phi_g$ where $F\in\mathcal{F}^{\times N}:=[\times]^N\mathcal{F}$ and $F=(f,f,\ldots,f):=f^{\times N}$ and (51) is nothing then else as the definition of the action of the group $\mathcal{F}^{\times N}$ on $\mathcal{L}_N$. Arbitrariness in choosing Φ_{g} ensures the validity of the definition (30) on the entire $\mathcal{L}_N$. Due to the isomorphism of the groups $\mathcal{F}$ and $\mathcal{F}^{\times N}$, it is obvious that a local-scaling transformation that maps a given pair of N-electron wavefunctions into each other matches unambiguously the local scaling that transforms the corresponding one-electron densities into each other. However, although any pair of densities are locally scaled, this property no longer holds for an arbitrary pair of N-electron wavefunctions. Hence, $\mathcal{L}_N$ is nontrivially partitioned, with respect to the group $\mathcal{F}^{\times N}$ of local-scaling transformations, into the orbits

$$\mathcal{L}_N=\bigcup_i\mathcal{O}^{[i]}. \tag{52}$$

In this sense, the group $\mathcal{F}$ entangles $\mathcal{D}_N$ and $\mathcal{L}_N$. By construction, an arbitrary orbit $\mathcal{O}^{[i]}$ is closed with respect to $\mathcal{F}^{\times N}$, that is, for any pair Φ_1 and Φ_2 in $\mathcal{O}^{[i]}$, there exists such local-scaling transformation $F_{1\Rightarrow 2}$ that $\Phi_2=F_{1\Rightarrow 2}\Phi_1$. In the other words, if Φ_1 is the generator wavefunction of $\mathcal{O}^{[i]}$, for all $F\in\mathcal{F}^{\times N}$, $F_{1\Rightarrow 2}\Phi_1\in\mathcal{O}^{[i]}$. We thus prove:

Proposal 3. There exists a one-to-one map of variables on any orbit in $\mathcal{L}_N$.

Corollary 3.1. Orbit $\mathcal{O}^{[i]}$ is invariant relative to generator wavefunction.

Corollary 3.2. On each orbit $\mathcal{O}^{[i]} \subset \mathcal{L}_N$, there exists one and only one N-electron wavefunction which one-electron density is $\rho(\mathbf{r}) \in \mathcal{D}_N$.

Corollary 3.3. For any given orbit $\mathcal{O}^{[k]} \subset \mathcal{L}_N$ generated by $\Phi_{\mathrm{g}}^{[k]}$ and the latter one-electron density $\rho_{\mathrm{g}}^{[k]}, \mathcal{F}_{\rho\mathrm{g}}^{[k]}$ exhausts the whole $\mathcal{D}_N$.

Remark 1: Corollary 3.3 implies that any density $\rho(\mathbf{r}) \in \mathcal{D}_N$ is N-representable. In other words, the group $\mathcal{F}$ of local-scaling transformations and its actions on $\mathcal{D}_N$ and $\mathcal{L}_N$ defined above ensures the N-representability of $\mathcal{D}_N$.

The uniqueness of the local-scaling transformation as the solution of Eq. (48) guarantees that the transformed wavefunction $\Phi_{\rho}^{[i]}$ is also unique. Thus, for any $\rho(\mathbf{r}) \in \mathcal{D}_N$ there exists a unique wavefunction $\Phi_{\rho}^{[i]}$ generated by means of local-scaling transformation from the arbitrary generator wavefunction $\Phi_{\mathrm{g}}^{[k]}$. The orbit in $\mathcal{L}_N$ is actually the set of all the wavefunctions thus generated which yield one-electron densities $\rho(\mathbf{r})$ in $\mathcal{D}_N$:

$$\mathcal{O}^{[i]} \equiv \left\{ \Phi_{\rho}^{[i]} | \Phi_{\rho}^{[i]} \to \rho(\vec{r}); \Phi_{\rho}^{[i]} \in \mathcal{L}_N; \rho(\vec{r}) \in \mathcal{D}_N \right\}. \tag{53}$$

Therefore, the orbit patterns in $\mathcal{L}_N$ predetermine the inverse "variable mapping" V that was the premise in (13) and (14) and that naturally generalizes the Hohenberg–Kohn theorem on the entire set $\mathcal{D}_N$.

Note that $\mathcal{L}_N$ includes N-electron Slater determinants which are structurally invariant with respect to $\mathcal{F}^{\times N}$. Define $\mathcal{S}_N$ as the proper subset of $\mathcal{L}_N$ consisting of Slater determinants. Since $\mathcal{F}^{\times N}\mathcal{S}_N \subseteq \mathcal{S}_N$, then $\mathcal{S}_N = \cup_i \mathcal{O}_{\mathcal{S}}^{[i]}$ over all Slater orbits. Thus, we have:

Corollary 3.4. An arbitrary one-electron density $\rho(\mathbf{r}) \in \mathcal{D}_N$ is N-representable in $\mathcal{S}_N$.

5.4 Energy Density Functional and Variational Principle

5.4.1 Energy Density Functional: Definition

Proposal 3 definitely allows to propose the following rigorous definition of the energy density functional

$$\mathcal{E}_i[\rho(\mathbf{r})] \equiv \mathcal{E}_{\mathbf{i}}[\rho(\mathbf{r}); \Phi_{\mathrm{g}}^{[\mathbf{i}]}] := \mathbf{E}[\Phi]|_{\Phi \in \mathcal{O}^{[\mathbf{i}]} \subset \mathcal{L}_{\mathbf{N}}} \tag{54}$$

and hence express the "functional mapping" (13) in the explicit way. It is evident that this mapping is one-to-many and there exist as many density functionals as there are orbits in $\mathcal{L}_N$. To derive $\mathcal{E}_i[\rho(\mathbf{r})]$ that is defined on the orbit $\mathcal{O}^{[i]} \subset \mathcal{L}_N$

explicitly, let us first write down the explicit expression for the energy functional $E[\Phi_{\rm g}^{[i]}]$ of the orbit-generating wavefunction $\Phi_{\rm g}^{[i]}$, in terms of its 1- and 2-matrices, $D_{\rm g}^{1[i]}(x_1, x_1')$ and $D_{\rm g}^{2[i]}(x_1, x_2; x_1, x_2)$, respectively, and its one-electron density $\rho_{\rm g}(x)$:

$$E[\Phi_{\rm g}^{[i]}] = \frac{1}{2}\int {\rm d}^4 x_1 \nabla_{x_1}\nabla_{x'_1} D_{\rm g}^{1[i]}(x_1; x'_1)|_{x'_1=x_1} + \int {\rm d}^4 x \rho(x)\hat{v}(\mathbf{r}) + \int {\rm d}^4 x_1 \int {\rm d}^4 x_2 \frac{D_{\rm g}^{2[i]}(x_1, x_2; x_1, x_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}, \tag{55}$$

where $\int {\rm d}^4 x \equiv \sum_s \int {\rm d}^3 \mathbf{r}$. Let us apply the local-scaling transformation that casts in **Proposal 3** to the wavefunction $\Phi_{\rm g}^{[i]}$, precisely to its 1- and 2-matrices and its density. This yields

$$D_{\rho}^{1[i]}(\mathbf{r}'_{\mathbf{1}}, s_1; \mathbf{r}'_{\mathbf{1}}, s'_1) = \left[J(\mathbf{f}(\mathbf{r}_1); \mathbf{r}_1) J(\mathbf{f}(\mathbf{r}'_1); \mathbf{r}'_1)\right]^{1/2} D_{\rm g}^{1[i]}(\mathbf{f}(\mathbf{r}_1), s_1; \mathbf{f}(\mathbf{r}'_1), s'_1), \tag{56}$$

$$D_{\rho}^{2[i]}(\mathbf{r}_1, s_1, \mathbf{r}_2, s_2; \mathbf{r}_1, s_1, \mathbf{r}_2, s_2) = J(\mathbf{f}(\mathbf{r}_1); \mathbf{r}_1) J(\mathbf{f}(\mathbf{r}_2); \mathbf{r}_2) D_{\rm g}^{2[i]}(\mathbf{f}(\mathbf{r}_1), s_1, \mathbf{f}(\mathbf{r}_2), s_2; \mathbf{f}(\mathbf{r}_1), s_1, \mathbf{f}(\mathbf{r}_2), s_2), \tag{57}$$

and

$$\rho(\mathbf{r}, s) = J(\mathbf{f}(\mathbf{r}); \mathbf{r}) \rho_{\rm g}(\mathbf{f}(\mathbf{r}), s). \tag{58}$$

Partitioning the 1- and 2-matrices, appearing in the rhs of Eqs. (56) and (57), into their local and non-local components:

$$D_{\rm g}^{1[i]}(\mathbf{f}(\mathbf{r}_1), s_1; \mathbf{f}(\mathbf{r}'_1), s'_1) = \left[\rho_{\rm g}(\mathbf{f}(\mathbf{r}_1), s_1)\rho_{\rm g}(\mathbf{f}(\mathbf{r}'_1), s'_1\right]^{1/2} \tilde{D}_{\rm g}^{1[i]}(\mathbf{f}(\mathbf{r}_1), s_1; \mathbf{f}(\mathbf{r}'_1), s'_1), \tag{59}$$

$$D_{\rm g}^{2[i]}(\mathbf{f}(\mathbf{r}_1), s_1; \mathbf{f}(\mathbf{r}_2), s_2; \mathbf{f}(\mathbf{r}_1), s_1, \mathbf{f}(\mathbf{r}_2), s_2) = \frac{1}{2}\rho_{\rm g}(\mathbf{f}(\mathbf{r}_1), s_1)\rho_{\rm g}(\mathbf{f}(\mathbf{r}_2), s_2)\left(1 + \mathcal{F}_{{\rm xc,\,g}}^{[i]}(\mathbf{f}(\mathbf{r}_1), s_1, \mathbf{f}(\mathbf{r}_2), s_2)\right) \tag{60}$$

where $\tilde{D}_{\rm g}^{1[i]}$ is the non-local part of the 1-matrix and $\mathcal{F}_{{\rm xc,\,g}}^{[i]}$ is the non-local exchange-correlation factor. Therefore, the 1- and 2-matrices of (51) take the appearance:

$$D_{\rho}^{1[i]}(\mathbf{r}_1, s_1; \mathbf{r}'_1, s'_1) = \left[\rho(\mathbf{r}_1, s_1)\rho(\mathbf{r}'_1, s'_1)\right]^{1/2} \tilde{D}_{\rm g}^{1[i]}(\mathbf{f}(\mathbf{r}_1), s_1; \mathbf{f}(\mathbf{r}'_1), s'_1), \tag{61}$$

and

$$D_{\rho}^{2[i]}(\mathbf{r}_1, s_1, \mathbf{r}_2, s_2; \mathbf{r}_1, s_1, \mathbf{r}_2, s_2) = \frac{1}{2}\rho(\mathbf{r}_1, s_1)\rho(\mathbf{r}_2, s_2) \times \left(1 + \mathcal{F}_{\mathrm{xc,g}}^{[i]}(\mathbf{f}(\mathbf{r}_1), s_1, \mathbf{f}(\mathbf{r}_2), s_2)\right). \tag{62}$$

Finally, we obtain [4]

$$\begin{aligned} E[\Phi_{\rho}^{[i]}] &\equiv \mathcal{E}\left[\rho(x); \Phi_{\mathrm{g}}^{[i]}\right] \\ &= \frac{1}{8}\int \mathrm{d}^4x \frac{[\nabla_{\mathrm{r}}\rho(x)]^2}{\rho(x)} + \frac{1}{2}\int \mathrm{d}^4x \rho(x) \nabla_{\mathbf{r}}\nabla_{\mathbf{r}'}\tilde{D}_{\mathrm{g}}^{1[i]}(\mathbf{f}(\mathbf{r}), s; \mathbf{f}(\mathbf{r}'), s')|_{x'=x} \\ &+ \int \mathrm{d}^4x \rho(x) v(\mathbf{r}) \\ &+ \frac{1}{2}\int \mathrm{d}^4x_1 \mathrm{d}^4x_2 \frac{\rho(x_1)\rho(x_2)\left(1 + \mathcal{F}_{\mathrm{xc,g}}^{[i]}(\mathbf{f}(\mathbf{r}_1), s_1, \mathbf{f}(\mathbf{r}_2), s_2)\right)}{|\mathbf{r}_1 - \mathbf{r}_2|}. \end{aligned} \tag{63}$$

Few statements can be drawn from Eq. (63):

(i) The kinetic energy density functional is composed of two components. The first, the von Weizsäcker term, is local and orbit-invariant. The second is non-local, orbit-dependent, and due to the one-third power in Eq. (50), transformed to the modified Thomas–Fermi term within the local density approximation;

(ii) The exchange-correlation energy density functional is explicitly expressed as

$$\begin{aligned} E_{\mathrm{xc}}[\Phi_{\rho}^{[i]}] &\equiv \mathcal{E}_{\mathrm{xc}}\left[\rho(x); \Phi_{\mathrm{g}}^{[i]}\right] : \\ &= \frac{1}{2}\int \mathrm{d}^4x_1 \mathrm{d}^4x_2 \frac{\rho(x_1)\rho(x_2)\mathcal{F}_{\mathrm{xc,g}}^{[i]}(\mathbf{f}(\mathbf{r}_1), s_1, \mathbf{f}(\mathbf{r}_2), s_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}; \end{aligned} \tag{64}$$

(iii) In fact, each density functional $\mathcal{E}\left[\rho(x); \Phi_{\mathrm{g}}^{[i]}\right]$ depends on two basic variables: the one-electron density $\rho(x)$ and the generator wavefunction $\Phi_{\mathrm{g}}^{[i]}$. Equation (42) expresses the energy as a functional of the one-electron density $\rho(x)$ within the orbit $\mathcal{O}^{[i]}$. True, Eq. (63) satisfies the condition of N-representability; (iv) One of the orbits in the decomposition (52) of $\mathcal{L}_N$ is actually the orbit that contains the exact ground-state wavefunction. Refer it as the Hohenberg–Kohn orbit $\mathcal{O}^{[\mathrm{HK}]} \subset \mathcal{L}_N$. If a generator wavefunction is chosen to belong to $\mathcal{O}^{[\mathrm{HK}]}$, Eq. (63) then determines the Hohenberg–Kohn energy density functional in the explicit manner.

5.4.2 Orbit Variational Principle and Euler–Lagrange Equation

The variational principle of the energy density functional theory based on the definition (33) is a straightforward consequence of the quantum mechanical variational principle (8) and the "functional mapping" (13). It is clearly orbit-dependent

or, equivalently, it is of the intra-orbit type. Let us consider the energy density functional $\mathcal{E}\left[\rho(\mathbf{r},s);\Phi_{\mathrm{g}}^{[i]}\right]$ given by Eq. (63) and defined within the $\mathcal{O}^{[i]}$ only. In this functional, $\rho(\mathbf{r}, s)$ stands for the density variable resulted from the initial density $\rho_g(\mathbf{r}, s)$ associated with the generator wavefunction $\Phi_{\mathrm{g}}^{[i]}$. The extremum of $\mathcal{E}\left[\rho(\mathbf{r},s);\Phi_{\mathrm{g}}^{[i]}\right]$ on $\mathcal{D}_N$ is attained at the i-th-optimal density $\rho_{\mathrm{opt}}^{[i]}(\mathbf{r},s)$ which is obtained by varying the following auxiliary functional:

$$\mathcal{E}\left[\rho(\mathbf{r},s);\Phi_{\mathrm{g}}^{[i]}\right]-\mu^{[i]}\left(\int \mathrm{d}^4x\rho(\mathbf{r},s)-N\right), \tag{65}$$

where $\mu^{[i]}$ is the Lagrange multiplier that accounts for the normalization of the density and that actually plays the role of a chemical potential on the orbit $\mathcal{O}^{[i]}$. Therefore, the stationary ground-state variational principle for the energy density functional $\mathcal{E}\left[\rho(\mathbf{r},s);\Phi_{\mathrm{g}}^{[i]}\right]$ is given by

$$\frac{\delta}{\delta\rho(\mathbf{r},s)}\left\{\mathcal{E}\left[\rho(\mathbf{r},s);\Phi_{\mathrm{g}}^{[i]}\right]-\mu^{[i]}\left(\int \mathrm{d}^4x\rho(\mathbf{r},s)-N\right)\right\}=0,\rho(\mathbf{r},s)\in\mathcal{D}_N, \tag{66}$$

we obtain the following integro-differential equation for the one-electron density [4]:

$$\begin{aligned}&\frac{1}{8}\left[\frac{\nabla\rho(\mathbf{r},s)}{\rho(\mathbf{r},s)}\right]^2-\frac{1}{r}\frac{\nabla^2\rho(\mathbf{r},s)}{\rho(\mathbf{r},s)}+v_{\mathrm{T,g}}^{[i]}([\rho(\mathbf{r},s)];\mathbf{r},s)+v(\mathbf{r})+v_{\mathrm{H}}([\rho(\mathbf{r},s)];\mathbf{r})+\\&v_{\mathrm{xc,g}}^{[i]}([\rho(\mathbf{r},s)];\mathbf{r},s)=\mu^{[i]},\end{aligned} \tag{67}$$

where $v_{\mathrm{H}}([\rho(\mathbf{r},s)];\mathbf{r})=\int \mathrm{d}^4x\rho(\mathbf{r},s)|\mathbf{r}-\mathbf{r}'|^{-1}$ is the Hartree potential,

$$\begin{aligned}v_{\mathrm{T,g}}^{[i]}([\rho(\mathbf{r},s)];\mathbf{r},s)=\frac{1}{2}\Big\{&[\nabla_{\mathbf{r}}\nabla_{\mathbf{r}'},\tilde{D}_{\mathrm{g}}^{1[i]}(\mathbf{f}(\mathbf{r}),s;\mathbf{f}(\mathbf{r}'),s')]_{\mathbf{r}'=\mathbf{r},s'=s}\\+\rho(\mathbf{r},s)&\tfrac{\delta}{\delta\rho(\mathbf{r},s)}([\nabla_{\mathbf{r}}\nabla_{\mathbf{r}'},\tilde{D}_{\mathbf{g}}^{1[i]}(\mathbf{f}(\mathbf{r}),s;\mathbf{f}(\mathbf{r}'),s')]_{\mathbf{r}'=\mathbf{r},s'=s})\Big\}\end{aligned} \tag{68}$$

is the potential originated from the non-local component of the kinetic energy in (63), and,

$$v_{\mathrm{xc,g}}^{[i]}([\rho(\mathbf{r},s)];\mathbf{r},s)=\mathcal{E}_{\mathrm{xc,g}}^{[i]}([\rho(\mathbf{r},s);\Phi_{\mathrm{g}}^{[i]}];\mathbf{r},s)+\rho(\mathbf{r},s)\frac{\mathrm{d}\mathcal{E}_{\mathrm{xc,g}}^{[i]}([\rho(\mathbf{r},s);\Phi_{\mathrm{g}}^{[i]}];\mathbf{r},s)}{\delta\rho(\mathbf{r},s)} \tag{69}$$

the exchange-correlation potential resulted from the non-local part of the electron–electron interaction where

$$\mathcal{E}^{[i]}_{\mathrm{xc,g}}([\rho(\mathbf{r}_1,s_1);\Phi^{[i]}_{\mathrm{g}}];\mathbf{r}_1,s_1)=\frac{1}{2}\int \mathrm{d}^4x_2\,\frac{\rho(\mathbf{r}_2,s_2)\mathcal{F}^{[i]}_{\mathrm{xc,g}}(\mathbf{f}(\mathbf{r}_1),s_1;\mathbf{f}(\mathbf{r}_2),s_2)}{|\mathbf{r}_1-\mathbf{r}|}. \quad (70)$$

Solving Eq. (67) for the given generator wavefunction $\Phi^{[i]}_{\mathrm{g}}$, we obtain the i-th optimal or i-th approximate ground-state density $\rho^{[i]}_0(\mathbf{r},s)\in\mathcal{D}_N$ and the i-th optimal or i-th ground-state energy

$$E^{[i]}_0\equiv\mathcal{E}_i[\rho_0(\mathbf{r})] \quad (71)$$

that simply casts as the i-th orbit variational principle:

$$E^{[i]}_0\equiv\inf_{\Phi\in\mathcal{O}^{[i]}\subset\mathcal{L}_N}\{E[\Phi]\}=E[\Phi]\big|_{\Phi=\Psi^{[i]}_0\in\mathcal{O}^{[i]}\subset\mathcal{L}_N}=\inf_{\rho_\Phi\to\Phi\in\mathcal{O}^{[i]}}\{\mathcal{E}_i[\rho_\Phi]\}. \quad (72)$$

The next step is to substitute the densities $\rho_1(\mathbf{r})$ and $\rho_2(\mathbf{r})$ by $\rho^{[i]}_{\mathrm{g}}(\mathbf{r})$ and $\rho^{[i]}_0(\mathbf{r})$ in Eq. (46) correspondingly and to solve the latter. The solution is the i-th optimal local-scaling transformation $\mathbf{f}^{[i]}_0(\mathbf{r})\in\mathcal{F}$ which is further applied to $\Phi^{[i]}_{\mathrm{g}}$ to get, via Eq. (51), the i-th optimal, ground-state wavefunction $\Phi^{[i]}_0\in\mathcal{L}_N$. True, generally speaking, the latter is the approximate ground-state wavefunction that yields an upper bound to the exact ground-state energy E_0 which is attained, by definition, only at the Hohenberg–Kohn orbit $\mathcal{O}^{[\mathrm{HK}]}$, that is, $E_0{}^{[\mathrm{HK}]}=E_0$.

5.4.3 Global Variational Principle: The Concept of Local-Scaling Self-Consistent Field

The orbit variational principle (65) deduced in Sect. 5.4.2 is solely defined on a particular orbit. The reason is trivial: this is precisely that orbit where the energy density functional is defined according to Eq. (54). In contrast, the global ground-state quantum mechanical variational principle (8) is carried out over the whole Hilbert space $\mathcal{L}_N$. Within the local-scaling formulation of the density functional approach is achieved due to the fact that the energy density functional in fact depends on two basic variables of theory: one—the one-electron density—is the key variable of the density functional theory and the other is the generator wavefunction that determines an orbit. Hence, the orbit partitioning (52) of $\mathcal{L}_N$ is governed by the orbit generators. Therefore,

$$E_0=\inf_{\substack{\text{over all orbits}\\ \mathcal{O}^{[i]}\subset\mathcal{L}_N}}\left\{\inf_{\rho(\mathbf{r})\in\mathcal{D}_N}\left\{\mathcal{E}\left[\rho(\mathbf{r});\Phi^{[i]}_{\mathrm{g}}\right]\right\}\right\}. \quad (73)$$

Equation (73) implies that the search for the exact ground-state wavefunction must be carried out by a combined intra-orbit and inter-orbit minimization [7]. The former reflects the charge consistency variational principle, whereas the latter the

inter-orbit one, the orbit consistency. The latter is actually the variational principle of the "inter-orbit" self-consistent field that resembles the Kohn–Sham self-consistent field approach and results in inter-orbit "jumps" that finally leads to the exact, Hohenberg–Kohn orbit.

6 Overlook

All of what is scientific in chemistry is physics—the rest is cooking.
L. D. Landau [187]

I have a feeling that a success in the adequate description of dispersion interactions within the DFT has been achieved. Again what we have at our current disposal is a number of D-DFT functionals among which choosing the best still remains a challenge that leaves a feeling of certain yet-imperfectness in this DFT area and, on the other hand, shows the ways to improve it. This was in fact the key goal undertaken in the present review which has definitely been achieved, I believe.

Acknowledgments Thanks are due first to all friends and colleagues with whom I shared the ideas of the density functional theory during the last three decades, in particular Ivan Zh. Petkov, Mario V. Stoitsov, Eduardo V. Ludeña, Toshi Koga, Jean-Lois Calais, Per-Olov Löwdin, Julian Schwinger, B. M. Deb, Enrico Clementi, Bob Parr, and Erkki Brändas, and Stefan Grimme. Mihai V. Putz and D. Michael P. Mingos are gratefully acknowledged for their kind invitation.

References

1. Hohenberg P, Kohn W (1964) Phys Rev 136:B864
2. Kohn W, Sham LJ (1965) Phys Rev 140:A1133
3. Parr RG, Yang W (1989) Density functional theory of atoms and molecules. Oxford University Press, Oxford
4. Kryachko ES, Ludeña EV (1990) Energy density functional theory of many-electron systems. Kluwer Academic, Dordrecht
5. Dreizler RM, Gross EKU (1990) Density functional theory. Springer, Berlin
6. March NH (1992) Electron density theory of atoms and molecules. Academic, New York
7. Perdew JP, Kurth S (2003) A primer in density functional theory. In: Fiolhais C, Nogueira F, Marques M (eds) Lecture notes in physics, vol 620. Springer, Berlin
8. Kryachko ES (2011) In: Roy AK (ed) Theoretical and computational developments in modern density functional theory. Nova Science, New York
9. Löwdin P-O (1988) In: Maruani J (ed) Molecules in physics, chemistry, and biology, vol II. Kluwer, Dordrecht, pp 3–60
10. Löwdin P-OJ (1991) Mol Struct Theochem 230(13)
11. Woolley RG, Sutcliffe BTP-O (2003) Löwdin and the quantum mechanics of molecules. In: Brändas EJ, Kryachko ES (eds) Fundamental world of quantum chemistry. A tribute to the memory of Per-Olov Löwdin, vol 1. Kluwer, Dordrecht, pp 21–65
12. Sutcliffe B (2002) Int J Quantum Chem 90:66
13. Kryachko ES (1930) Int J Quantum Chem 2008:108

14. Lieb EH (1981) Phys Rev Lett 46:457; 47:69
15. Kryachko ES (2005) Int J Quantum Chem 103:818
16. Kryachko ES (2006) Int J Quantum Chem 106:1795
17. Szczepanik W, Dulak M, Wesolowski TA (2007) Int J Quantum Chem 107:762–763
18. Kato T (1957) Commun Pure Appl Math 10:151
19. Steiner E (1963) J Chem Phys 39:2365
20. McWeeny R, Sutcliffe BT (1969) Methods of molecular quantum mechanics. Academic, London
21. Davidson ER (1976) Reduced density matrices in quantum chemistry. Academic, New York
22. Bingel W (1966) Theor Chim Acta 5:341
23. Pack RT, Brown WB (1966) J Chem Phys 45:556
24. Kimball JC (1973) Phys Rev A 7:1648
25. Kimball JC (1975) J Phys A Math Gen 8:1513
26. Rajagopal AK, Kimball JC, Banerjee M (1978) Phys Rev B 18:2339
27. Hoffmann-Ostenhof M, Hoffmann-Ostenhof T, Thirring W (1978) J Phys B At Mol Opt 11 L571
28. Hoffmann-Ostenhof M, Seiler R (1981) Phys Rev A 23:21
29. Carlsson AE, Ashcroft NW (1982) Phys Rev B 25:3474
30. Coleman AJ (1981) In: Deb BM (ed) The force concept in chemistry. Van Nostrand, New York
31. Bamzai AS, Deb BM (1981) Rev Mod Phys 53:95; 53:593(E)
32. Smith VH Jr (1982) In: Coppens P, Hall MB (eds) Electron distribution and the chemical bond. Plenum, New York, p 3
33. Löwdin P-O (1986) Int J Quantum Chem S19:19. See also Löwdin P-O (1997) Book of Abstracts. In: 213th ACS national meeting, San Francisco, 13–17 Apr 1997. American Chemical Society, Washington, DC and the discussion on pp. 411*ff* of Ref. [2c]
34. Nagy Á (1998) Int J Quantum Chem 70:681
35. Moiseyev N (2000) Chem Phys Lett 321:469
36. March NH, Howard IA, Holas A, Senet P, Van Doren VE (2000) Phys Rev A 63:012520
37. Bader RFW (1990) Atoms in molecules. A quantum theory. Clarendon, Oxford
38. Gatti C, Fantucci P, Pacchioni G (1987) Theor Chim Acta 72:433
39. Cao WL, Gatti C, MacDougall PJ, Bader RFW (1987) Chem Phys Lett 141:380
40. Cioslowski J (1990) J Phys Chem 94:5496
41. Cooper D (1990) Nature 346:796
42. Edgecombe KE, Esquivel RO, Smith VH Jr, Müller-Plathe F (1992) J Chem Phys 97:2593
43. Bersuker GI, Peng C, Boggs JE (1993) J Phys Chem 97:9323
44. Esquivel RO, Chen J, Stott MJ, Sagar RP, Smith VH Jr (1993) Phys Rev A 47:936
45. Esquivel RO, Sagar RP, Smith VH Jr, Chen J, Stott MJ (1993) Phys Rev A 47:47335
46. Mei CJ, Edgecombe KE, Smith VH Jr, Heilingbrunner A (1993) Int J Quantum Chem 48:287
47. Pendas AM, Blanco MA, Costales A, Sánchez PM, Luaña V (1999) Phys Rev Lett 83:1930
48. Zhikol OA, Oshkalo AF, Shishkin OV, Prezhdo OV (2003) Chem Phys 288:159
49. Levy M (1979) Proc Natl Acad Sci USA 76:6062
50. Levy M (1982) Phys Rev A 26:1200
51. Lieb EH (1983) Int J Quantum Chem 24:243
52. Lieb EH (1982) In: Feshback H, Shimony A (eds) Physics as natural philosophy: essays in honor of Laszlo Tisza on his 75th birthday. MIT Press, Cambridge, MA, p 111
53. Lieb EH (1985) In: Dreizler RM, da Providencia J (eds) Density functional methods in physics. Plenum, New York, p 31
54. M. Levy (1990) Adv Quantum Chem 21:69. See also Sect 6.5 in Ref [2c]
55. Bokanowski O, Grebert B, Mauser NJ (1999) C R Acad Sci Paris Serie I-Math 329:85
56. Dacorogna B, Moser J (1990) Ann Inst Henri Poincaré Analyse non Linéaire 7:1 and references therein
57. Bokanowski O, Mauser NJ (1999) Math Models Methods Appl Sci 8:941

58. Bokanowski O, Grebert B, Mauser NJ (2000) J Mol Struct Theochem 501–502:47
59. Bokanowski O, Grebert B (1996) J Math Phys 37:1553
60. Bokanowski O, Grebert B (1998) Int J Quantum Chem 68:221
61. Bokanowski O (1999) J Math Chem 26:271
62. Bokanowski O (2000) J Math Phys 41:2568
63. Bokanowski O, Grebert B, Mauser NJ (2003) Math Models Methods Appl Sci 13:1185
64. I. Zh. Petkov, M. Stoitsov, and E. S. Kryachko, Int. J. Quantum Chem. 29, 149 (1986)
65. Kryachko ES, Petkov IZ, Stoitsov MV (1987) Int J Quantum Chem 32:467
66. Kryachko ES, Petkov IZh, Stoitsov MV (1987) Int J Quantum Chem 32:473; (1988) 34:305(E)
67. Kryachko ES, Ludeña EV (1987) Phys Rev A 35:957
68. Kryachko ES, Koga T (1989) J Chem Phys 91:1108
69. Kryachko ES, Ludeña EV (1991) Phys Rev A 43:2179
70. Kryachko ES, Ludeña EV (1991) Phys Rev A 43:2194
71. Koga T, Kryachko ES (1991) J Chem Phys 94:2910
72. Kryachko ES, Ludeña EV (1992) Int J Quantum Chem 43:769
73. Kryachko ES, Koga T (1992) Int J Quantum Chem 42:591
74. Kryachko ES, Ludeña EV, Mujica V (1991) Int J Quantum Chem 40:589
75. Kryachko ES, Ludeña EV (1991) J Chem Phys 95:9054
76. Kosaka K (1994) J Phys Soc Jpn 63:1325
77. Valderrama E, Ludeña EV, Hinze J (1999) J Chem Phys 110:2343
78. Koga T (1990) Phys Rev A 42:3763
79. Koga T (1990) Phys Rev A 41:1274
80. Pavlov RL, Maruani J, Delchev YI, McWeeny R (1997) Int J Quantum Chem 65:241
81. Ludeña EV, López-Boada R (1996) Top Curr Chem 180:169
82. López-Boada R, Ludeña EV, Karasiev V, Pino R (1997) J Chem Phys 107:6722
83. López-Boada R, Karasiev V, Ludeña EV, Colle R (1998) Int J Quantum Chem 69:503
84. Kanhere DG, Dhavale A, Ludeña EV, Karasiev V (2000) Phys Rev A 62:065201
85. Pino R (2003) Europhys Lett 63:200
86. Seidl M, Perdew JP, Levy M (1999) Phys Rev A 59:51
87. Seidl M (1999) Phys Rev A 60:4387
88. Perdew JP, Wang Y (1992) Phys Rev B 45:13244
89. Langreth DC, Mehl MJ (1983) Phys Rev B 28:1809
90. Perdew JP, Wang Y (1986) Phys Rev B 33:8800; (1989) 40:3399(E); Perdew JP (1986) 33:8822; (1986) 34:7406(E)
91. Perdew JP, Burke K, Ernzerhof M (1996) Phys Rev Lett 77:3865; (1997) 78:1396 (E)
92. Kaplan IG (1982) Introduction to theory of intermolecular interactions (in Russian). Nauka, Moscow, p 20
93. Hobza P, Zahradnik R (1988) Intermolecular complexes. Academia, Prague
94. Israelachvili JN (1985) Intermolecular and surface forces, Sect 2.4. Academic, London
95. Lieb EH, Thirring WE (1986) Phys Rev A 34:40
96. Lewin M (2006) Ann Henri Poincaré 7:365
97. Boyer TH (1972) Phys Rev A 5:1799; 6:314; (1973) 7:1832; (1974) 9:2078
98. Johnson ER, Wolkow RA, DiLabio GA (2004) Chem Phys Lett 394:334–338
99. Tsuzuki S, Lüthi HP (2001) J Chem Phys 114:3949
100. Sinnokrot MO, Valeev EF, Sherrill CD (2002) J Am Chem Soc 124:10887
101. Debye P (1929) Polar molecules. Chemical Catalogue, New York
102. Sack H (1934) In: Marx E (ed) Handbuch der Radiologie Die Theorien der Radiologie. Akademische Verlaggesellschaft, Leipzig
103. See also Kryachko ES, Yanovitskii OE (1989) Phys Rev A 40:5533
104. Israelachvili J (1992) Intermolecular and Surface Forces, 2nd ed., Academic Press, Chapter 5, pp. 67–10
105. London F (1930) Z Phys 60:491
106. Dzyaloshinskii IE, Lifshits EM, Pitaevskii LP (1961) Uspekhi Fiz Nauk 73(381) (in Russian)

107. Buckingham AD, Fowler PW, Hutson JM (1988) Chem Rev 88:963
108. Jorgensen W (1984) J Am Chem Soc 106:6638
109. MacKerell AD Jr, Karplus MJ (1991) J Phys Chem 95:10559
110. Kitaigorodsky AI (1973) Molecular crystals and molecules. Academic, New York
111. Sponer J, Hobza P (1997) Chem Phys Lett 267:263
112. Buckingham AD (1978) In: Pullman B (ed) Intermolecular interactions: from diatomics to biopolymers. Chichester, Wiley
113. Becke AD (1988) Phys Rev A 38:3098
114. Lee C, Yang W, Parr RG (1988) Phys Rev B 37:785
115. Perdew JP, Wang Y (1992) Phys Rev B 45:13244
116. Becke AD (1993) J Chem Phys 98:5648
117. Lackes DJ, Gordon RG (1993) Phys Rev A 47:4681
118. Kristyan S, Pulay P (1994) Chem Phys Lett 229:175
119. Perez-Jorda JM, Becke AD (1995) Chem Phys Lett 233:134
120. Hobza P, Spooner J, Reschel T (1995) J Comput Chem 16:1315
121. Meijer EJ, Sprik M (1996) J Chem Phys 105:8684
122. Tsuzuki S, Uchimaru T, Tanabe K (1998) Chem Phys Lett 287:202
123. Zhang Y, Pan W, Yang W (1997) J Chem Phys 107:7921
124. Wesolowski TA, Parisel O, Ellinger Y, Weber J (1997) J Phys Chem A 101:7818
125. Zhang RB, Somers KRF, Kryachko ES, Nguyen MT, Zeegers-Huyskens T, Ceulemans A (2005) J Phys Chem A 109:8028
126. Kristyan S, Pulay P (1994) Chem Phys Lett 220:175
127. Hobza P, Sponer J, Reschel T (1995) J Comput Chem 16:1315
128. Meijer EJ, Sprik M (1996) J Chem Phys 105:8684
129. Tuma C, Boese AD, Handy NC (1999) Chem Phys 1:3939
130. Rappe AK, Bernstein ER (2000) J Phys Chem A 104:6117
131. Jones GA, Paddon-Row MN, Sherburn MS, Turner CI (2002) Org Lett 4:3789
132. Koch W, Holthausen MC (2001) A chemist's guide to density functional theory. Wiley-VCH, Weinheim
133. van Mourik T, Gdanitz RJ (2002) J Chem Phys 116:9620
134. Kristyán S, Pulay P (1994) Chem Phys Lett 229:175
135. Pérez-Jordá JM, Becke AD (1995) Chem Phys Lett 233:134
136. Engel E, Hock A, Dreizler RM (2000) Phys Rev A 61:032502
137. Lotrich VF, Bartlett RJ, Grabowski I (2005) Chem Phys Lett 405:43–48
138. Lotrich V, Bartlett RJ (2011) J Chem Phys 134:184108
139. Eshuis H, Furche FJ (2011) Phys Chem Lett 2:983–989
140. Dobson JF, Dinte BP (1996) Phys Rev Lett 76:1780–1783
141. Dion M, Rydberg H, Schröder E, Langreth DC, Lundqvist BI (2004) Phys Rev Lett 92:246401
142. Vydrov OA, Voorhis TV (2009) Phys Rev Lett 103:063004
143. Lee K, Murray ÉD, Kong L, Lundqvist BI, Langreth DC (2010) Phys Rev B 82:092202
144. Zhao Y, Truhlar DG (2008) Theor Chem Acc 120:215–241
145. Xu X, Goddard WA III (2004) Proc Natl Acad Sci USA 101:2673–2677
146. Zhang Y, Vela A, Salahub DR (2007) Theor Chem Acc 118:693
147. Grimme SJ (2004) Comput Chem 25:1463–1473
148. Sato T, Nakai H (2009) J Chem Phys 131:224104
149. Sato T, Nakai H (2010) J Chem Phys 133:194101
150. Tkatchenko A, Scheffler M (2009) Phys Rev Lett 102:073005
151. Grimme S, Antony J, Ehrlich S, Krieg H (2010) J Chem Phys 132:154104
152. Krishtal A, Vannomeslaeghe K, Geldof D, Van Alsenoy C, Geerlings P (2011) Phys Rev A 83:024501
153. Krishtal A, Vanommeslaeghe K, Olasz A, Veszprémi T, Van Alsenoy C, Geerlings P (2009) J Chem Phys 130:174101

154. Neese F, Wennmohs F ORCA, an ab initio, DFT and semiempirical SCFMO package. Version 2.9
155. Becke AD, Johnson ER (2007) J Chem Phys 127:154108
156. Kannemann FO, Becke AD (2010) J Chem Theory Comput 6:1081–1088
157. Steinmann SN, Corminboeuf C (2010) J Chem Theory Comput 6:1990–2001
158. Steinmann SN, Corminboeuf C (2011) J Chem Phys 134:044117
159. Olasz A, Vanommeslaeghe K, Krishtal A, Veszprémi T, Van Alsenoy C, Geerlings P (2007) J Chem Phys 127:224105
160. Jeziorski B, Moszynski R, Szalewic K (1994) Chem Rev 94:1887–1930
161. Heelmann A, Jansen G (2003) Chem Phys Lett 367:778–784
162. Heelmann A, Jansen G, Schtz M (2005) J Chem Phys 122:014103
163. Misquitta AJ, Szalewicz K (2005) J Chem Phys 122:214109
164. Misquitta AJ, Podeszwa R, Jeziorski B, Szalewicz K (2005) J Chem Phys 123:214103
165. Rajchel L, Zuchowski PS, Szczesniak MM, Chalasinski G (2010) Phys Rev Lett 104:163001
166. Krishtal A, Geldof D, Vanommeslaeghe K, Van Alsenoy C, Geerlings P (2012) J Chem Theory Comput 8:125
167. Grimme S (2006) J Comput Chem 27:1787
168. Vydrov OA, Voorhis TV (2008) J Chem Phys 129:014106
169. Vydrov OA, Voorhis TV (2009) J Chem Phys 130:104105
170. Eger FM, Gross EP (1963) Ann Phys 24:63
171. Kryachko ES, Ludeña EV (1987) Phys Rev A 35:957
172. Fock VA (1930) Z Phys 63:855
173. Moro AM, Arias JM, Gómez-Camacho J (2010) Phys Rev C 80:054605
174. Pittel S, Stoitsov MV (2001) Phys At Nucl 64:1055
175. Bokanowski O, Grebert B, Mauser NJ (2000) J Mol Struct Theochem 501–502:47
176. Hall GG (1960) Proc Phys Soc Lond 75:575
177. ten Hoor MJ (1988) Int J Quantum Chem 33:563
178. ten Hoor MJ (1989) J Phys B At Mol Opt Phys 22:L89
179. Fermi E (1927) Rend Accad Naz Lincei 6:602
180. Gombás P (1949) Die Statistische Theorie des Atoms und Ihre Anwedungen. Springer, Wien
181. Katriel J, Nyden MR (1981) J Chem Phys 74:1221
182. Eriksson LA, Kryachko ES, Nguyen MT (2004) Int J Quantum Chem 99:841
183. Paier J, Marsman M, Kresse G (2007) J Chem Phys 127:024103
184. Kryachko ES (1989) Lecture notes in chemistry, vol 50. Springer, Berlin, pp 503–522
185. Quoted from Gorobets B (2006) Landau's circle. Moscow, Summer Garden
186. Chalasinski G, Szczesniak MM (1994) Chem Rev 94:1723
187. Stone AJ (1996) The theory of intermolecular forces. Clarendon, Oxford
188. Møller C, Plesset MS (1934) Phys Rev 46:618
189. Thomas LH (1927) Proc Camb Philos Soc 23:542
190. Moser J (1965) Trans Am Math Soc 120:286

Struct Bond (2013) 150: 97–118
DOI: 10.1007/978-3-642-32750-6_3

Redox Activation of Small Molecules at Biological Metal Centers

Radu Silaghi-Dumitrescu

Abstract Some recent applications of density functional theory (DFT) are described and specifically the review addresses the activation of small molecules such as water, molecular oxygen and its reduced congeners, nitrogen oxides and oxyanions, sulfide, sulfur oxides and oxyanions, carbon dioxide, organic compounds, halogens, molecular hydrogen, and protons. A range of predictions on geometry, electronic structure, and spectroscopic properties are made and binding energies and activation energies are critically reviewed.

Keywords DFT • Metalloenzyme • Small molecule activation

Contents

R. Silaghi-Dumitrescu (✉)
Department of Chemistry, Babes-Bolyai University, Romania, Arany Janos str 11, RO-400028, Cluj-Napoca, Romania
e-mail: rsilaghi@chem.ubbcluj.ro

Abbreviations

CANH	Carbonic anhydrase
CPOI	Chloroperoxidase
DFT	Density functional theory
Fe-SOD	Iron-containing SOD
MM	Molecular mechanics
MPO	Myeloperoxidase
Ni-SOD	Nickel-containing SOD
NOR	Nitric oxide reductase
P450NOR	P450-like nitric oxide reductase
P450	Cytochrome P450
QM	Quantum mechanics
SOD	Superoxide dismutase
SOR	Superoxide reductase
TD-DFT	Time-dependent DFT
TSR	Two-state reactivity
VPO	Vanadium-containing peroxidase

1 Introduction

Activation of small molecules by metal centers may occur via mechanisms involving changes in redox state, spin state, bond breaking and formation, or acid–base properties [1, 2]. Density functional methods (DFT) are particularly suited for providing insights into such aspects [3]. Biological metal centers may be deemed to have evolved so that the biomolecule part would selectively favor one (or at least a small number) of the many reactions and mechanisms of which the metal center is intrinsically capable. On one hand, this selectivity implies a less complicated array of reactions connecting reactants and products (presumably a single shortest path allowed by chemical rules), compared to what would be observed for simpler synthetic metal complexes (multiple uncontrolled parallel reactions with multiple products). Such presumed simplicity is a remarkable advantage for experimentalists seeking to investigate these pathways. On the other hand, precisely the specialization brought about by the biomolecules around the metal, often entails not only particular sensitivity to reaction conditions (e.g., proteins stable only in biological membranes or with very labile metal-binding sites) but also very efficient catalysis—meaning that the lifetimes of the investigated reaction intermediates will often seriously challenge the performance of most cutting-edge experimental techniques. In such cases, DFT methods have the advantage of allowing extensive examination not only of fleeting/unstable reaction intermediates but also of otherwise unobservable transition states.

In the following pages, some recent applications of DFT are discussed for activation of small molecules such as water, molecular oxygen and its reduced congeners, nitrogen oxides and oxyanions, sulfide, sulfur oxides and oxyanions, carbon dioxide, organic compounds, halogens, molecular hydrogen, and protons. A range of predictions on geometry, electronic structure, and spectroscopic properties are made and binding energies and activation energies are critically reviewed.

2 Molecular Oxygen, Superoxide, Peroxide, Water

Oxygen-based small molecules are among the longest known substrates for bioinorganic chemistry and are pervasive throughout processes essential for life—respiration, photosynthesis, defense, or selective biosynthesis of essential building blocks within the human organisms [1, 2]. The constant progress made by theoretical chemists in these areas, with DFT being a key tool, is therefore not surprising.

2.1 Dioxygen Activation

In heme-containing monooxygenases and peroxidases, a general theme is cleavage of the oxygen–oxygen bond, leading to formally high oxidation states at the metal (Fe(IV), Fe(V), also referred to in the literature as "high-valent") [4–7]. Figure 1 illustrates the main reaction pathways accepted to be involved in these processes and the key players. It may be seen that small changes in the first coordination sphere, as well as changes in the second coordination sphere, can significantly affect the choice of reaction pathway. Globins normally oscillate between the "oxy" and the Fe(II) forms, with side reactions leading to Fe(III) ("autooxidation") and then possibly to the ferric-hydroperoxo → Compound I route. Peroxidases normally use the path between the ferric, ferric hydroperoxo, Compound I/II, and ferric-hydroxo states. Cytochrome oxidases close a catalytic cycle that connects the Fe(II) state with the "oxy," then all the way to Compound II and Fe(III)-OH, and then back to Fe(II). Heme oxygenases, similarly to cytochrome oxidases, start out at Fe(II) and to the ferric-hydroperoxo state via the oxy; however, at the ferric-hydroperoxo stage the site self-destructs, as the heme itself is the substrate of this enzyme. Notably, in all these enzymes of such different reactivities, the axial ligand (*X* in Fig. 1) is histidine—illustrating how important the role of the second coordination sphere may be.

Binding of molecular oxygen (often referred to as "dioxygen") to heme centers has been described computationally for several proteins, with emphasis on the electronic structure within the Fe–O_2 moiety, on the reversibility of the process and on the importance of spin state preferences [8–11]. An interesting question to

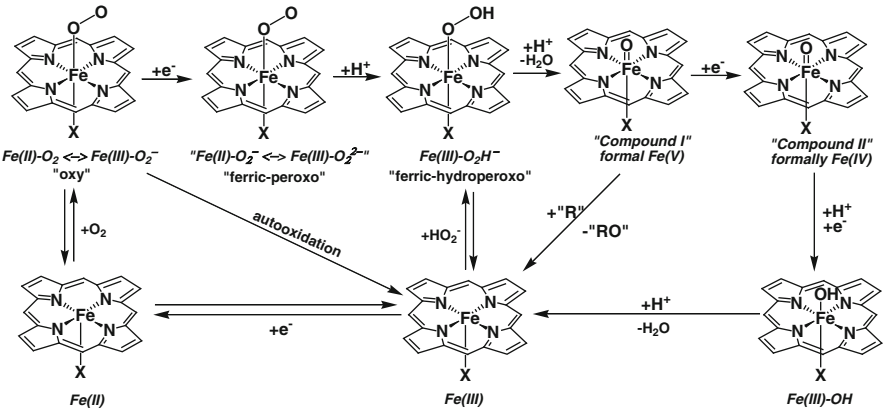


Fig. 1 Reactions involving molecular oxygen and its reduced congeners in heme proteins. Reactions where equilibria are manifest are indicated as such. In globins, peroxidases, cytochrome oxidases, and heme oxygenases, the axial ligand *X* is a protein-derived histidine; in catalases, *X* is a tyrosine phenolate; in chloroperoxidases and cytochromes P450, *X* is a cysteine thiolate

investigate is also the reason why globins and heme oxygenase (or indeed cytochrome oxidases) make such drastically different choices with respect to the stability of the ferrous–dioxygen adduct towards reduction: stable in one case, easily reduced in the other cases. The computed thermodynamics indicate that the one-electron reduction of the ferrous–dioxygen adduct is indeed reasonably allowed—certainly more allowed than the reduction of ferrous–dioxygen adducts of cytochrome P450, where the axial ligand is not a neutral imidazole ligand but rather an anionic thiolate from a cysteine [8, 9]. Paradoxically, precisely this cysteinate has been considered a key element in *facilitating* dioxygen activation—an effect that was dubbed *thiolate push* on the grounds that the extra charge density at a sulfur would, by being transferred onto the dioxygen via the iron, help occupy the antibonding orbitals of the O_2 ligand and thus weaken the O–O bond as well as increase its proton affinity (protonation being, as seen in Fig. 1, a key prerequisite for O–O bond cleavage) [9, 10]. The *thermodynamic* obstacle presented by the thiolate ligand in the early step of dioxygen reduction has in this context been dubbed *thiolate obstruction*, and, as detailed in the next sections, is seen at work in more than one instance in small molecule activation [9, 12, 13].

The nature of the higher oxidation state species in hemoproteins, especially the Compound I and Compound II depicted in Fig. 1, has been thoroughly revisited in recent years. DFT-based predictions, based on proton affinities and electronic structures [14], have argued that many of these high oxidation states would in fact occur in a protonated form (Fe(IV)-hydroxo, as opposed to Fe(IV)=oxo)—and these predictions were soon conformed experimentally [4, 7, 15–20]. A debate is nevertheless still open about the extent of these protonations, with arguments supporting either one of two theories (1) that only thiolate-ligated Compound II species may be anionic (and, possibly, other anion-ligated species) [19] and (2) that all Compound II species are protonated [4]. Essentially, this debate revolves around

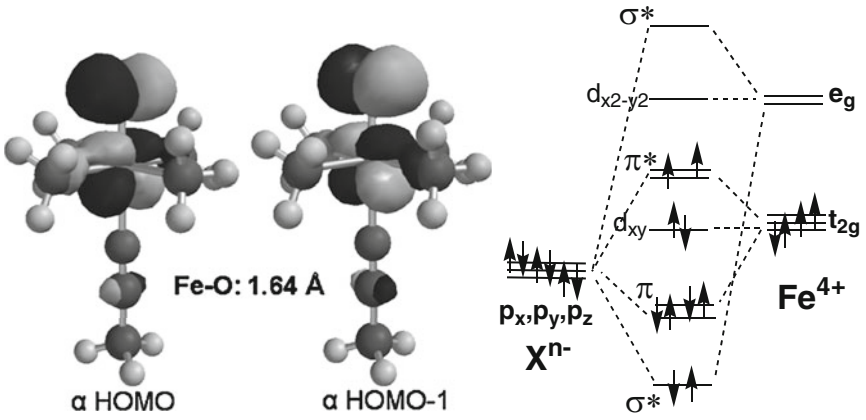


Fig. 2 Molecular orbital diagram and iron–oxygen π* orbitals for an octahedral ferryl model, with four equatorial amine ligands and an axial acetonitrile, illustrating a high degree of covalence and a triplet dioxygen-like electronic structure [22]

the pK_a of the formally Fe(IV)=oxo moiety in the histidine-ligated Compound II. Estimates based on experiment have placed this pK_a at ~4.7 [21], which being close to the pH range where most experiments are conducted, and being subject to temperature dependence in low-temperature crystallography and spectroscopy measurements, may partly explain why different groups looking at the same experiments appear to obtain conflicting results [4, 19, 21]. On the other hand, for Compound I species where the extra oxidizing equivalent compared to Compound II is still located on the heme-oxo moiety, the protonated state is predicted by DFT to be too acidic to be observed under physiological conditions [21].

Bonding within the Fe(IV)=oxo moiety has been described by several groups, highlighting a strongly covalent interaction with a preferred dioxygen-like triplet ground state, as illustrated in Fig. 2 [22–24]. Nevertheless, Fig. 3 illustrates that according to non-DFT computations (Hartree-Fock, MP2 [22]) an alternative description of these centers may be proposed, where the two unpaired electrons reside one on the oxygen and one on the iron—making for an Fe(III)-oxyl unit (formally, a deprotonated hydroxyl radical). Such a description would fit well with the fact that "Fe(IV)=oxo" complexes are capable of hydrogen atom abstraction, much like hydroxyl radical. In models where the oxo ligand was replaced by sulfido or by nitride ligands, DFT has allowed the "covalent deadlock" to be broken: population analyses and molecular orbitals have revealed that with sulfide and nitride the interaction is no longer a "triplet dioxygen-like," and that the mono-atomic ligands are in fact partially oxidized, in the same manner previously predicted by MP2 calculations [24, 25]. Under these circumstances, one may argue that many biological Fe(IV)=oxo complexes are in fact better described as *not* containing iron in a high oxidation state—but rather an oxyl ligand bound to Fe (III) [24]. Evidence for an oxyl-type structure as well as for an oxyl-type reactivity with high-oxidation state centers of other metals, such as manganese, is also available [26, 27].

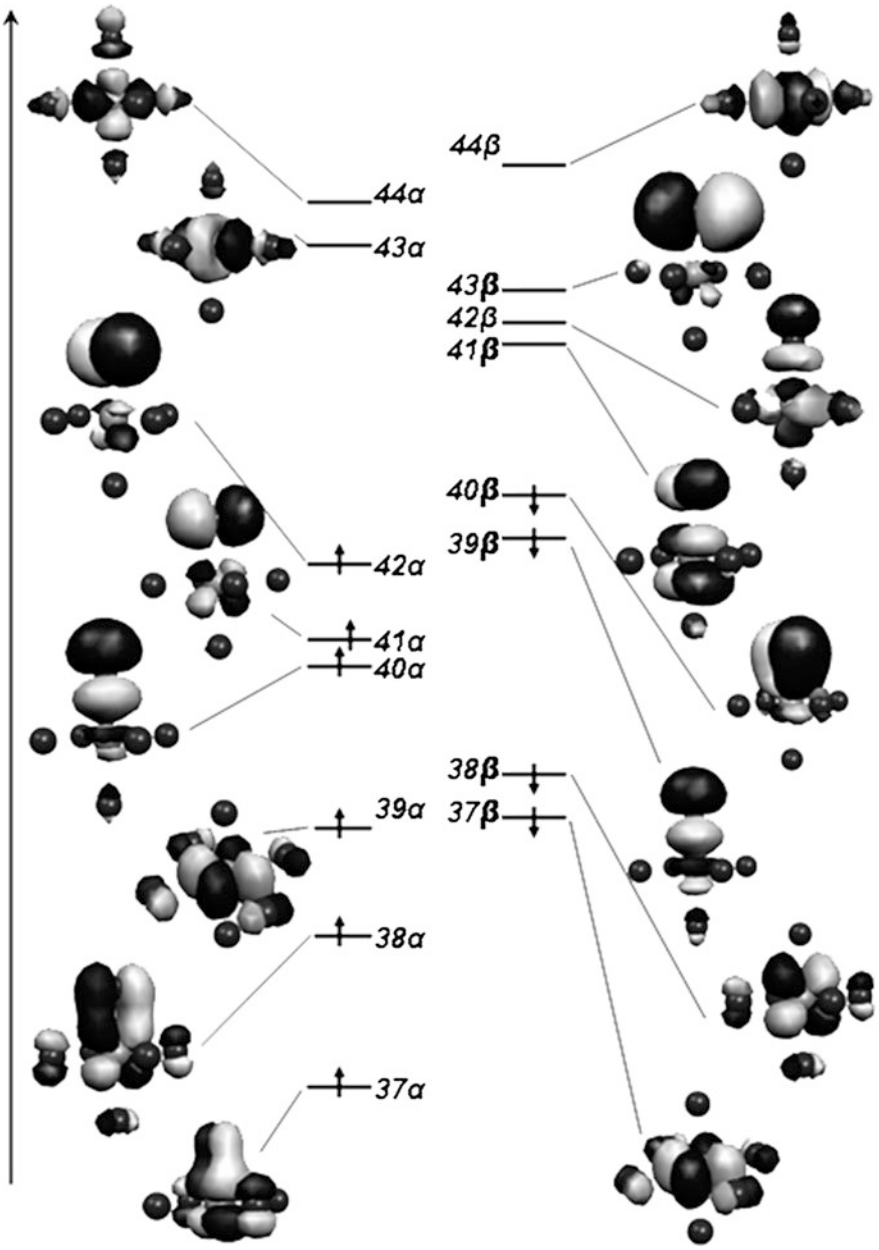


Fig. 3 UMP2/6-311 + G** molecular orbital diagram for a $S = 1$ $[FeO]^{2+}$ complex in an octahedral environment [22]. The π* orbitals are 41 and 42 on the α manifold, and 41 and 43 on the β manifold; 43β illustrates a distinct "hole" in the oxygen p orbitals, and implicitly 38β illustrates one extra electron in the Fe orbitals, compared to the expected Fe(IV) d^4 description—hence, a d^5 Fe(III) iron bound to an O^{-1} oxyl ligand

A complication of the oxo-containing Compound I species is the presence of one extra unpaired electron, which may couple to the triplet iron-oxo moiety either ferromagnetically (to yield a net doublet) or antiferromagnetically (yielding a quartet). According to DFT data, these two spin states display slightly different reactivities in terms of kinetics and of mechanism. Such differences were proposed to explain experimental observations according to which some cytochrome P450-catalyzed reactions would incur two different oxidants. This concept is now known as *two-state reactivity* [23]. On the other hand, with more recent data now suggesting that protonated oxo groups may also be effective oxidants (vide infra), and with other data supporting the involvement of other intermediates in the catalytic cycle as direct oxygenation agents, or the fact that in nonheme enzymes the ferric-hydroperoxo rather than a ferryl species is accepted as the active oxygenating agent [28], the concept of *mechanistic promiscuity* was proposed—describing a situation where a substrate may be converted into a single product via more than one mechanism [27, 29]. Examples have been given of other enzyme mechanisms where this concept may apply—pointing out that mechanistic promiscuity would be an evolutionary advantage for an enzyme [30–32].

Comparisons between the reactivities of the protonated and nonprotonated high-oxidation state Fe(IV)-hydroxo//Fe(IV)=oxo states towards hydrocarbons have been attempted, with hydrogen abstraction processes being taken as representative for enzymatic P450 chemistry. In some of the studies, the oxo and hydroxo ligands were in fact placed as ligands to the same metal in the same model, in order to allow more straightforward comparisons [27, 33, 34]. Although the free-radical character assigned to the oxo ligand (both in the dioxygen-like description and in the MP2-like description) would suggest that the oxo atom would be distinctly more adept than a hydroxo unit at activating hydrocarbons, it was surprisingly found that the barriers for hydrogen atom abstraction from model hydrocarbons by oxo vs. hydroxo ligands are remarkably close to each other [27]. As such, the thermodynamics of the process (and not the kinetics) would be the ones dictating the preferred reaction pathways.

Nonheme iron proteins deal with molecular oxygen on the same principles as illustrated in Fig. 1. Differences arise from the increased mobility of the coordination sphere, which often allows dioxygen, superoxide and peroxide to bind in a bidentate fashion to the metal. Additionally, the absence of the porphyrin "charge reservoir" may disfavor oxygen–oxygen bond cleavage thermodynamically. The case of Rieske dioxygenases, where two different DFT functionals, coupled with two slightly different choices of model, yielded two apparently conflicting results, is particularly instructive. Thus, on one hand B3LYP data suggest that indeed substrate oxygenation occurs reasonably well directly from the ferric-hydroperoxo state, with an activation barrier higher by a few kcal/mol than the barrier for O–O bond cleavage [28]. On the other hand, BP86 results suggest the opposite: O–O bond cleavage favored by a few kcal/mol over the direct oxygenation of substrate by a ferric-hydroperoxo adduct; then, such a high oxidation state would very easily oxygenate the substrate [35, 36]. This dependence of the outcome on the choice of the functional is important to remember whenever one examines conclusions based

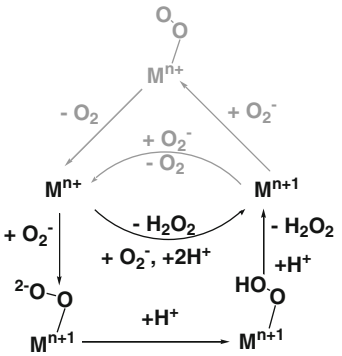


Fig. 4 Reaction pathways for superoxide dismutases and superoxide reductases. Shown in *black* are the two possible pathways leading to oxidation of the active site by superoxide, producing peroxide; this half of the catalytic cycle is common to SOR and SOD. Shown in gray is the re-reductive path specific to SOD, and which in SOR is replaced by a single-electron transfer from the specialized redox protein, rubredoxin. *M* may be Fe ($n = 2$, in SOR, or in SOD), Ni, Mn ($n = 2$, in SOD), or Cu ($n = 1$, in SOD)

on energy differences of less than ~10 kcal/mol: two barriers computed to be different by that much, may, for all we know, be perfectly equal in reality—or even display the opposite trend compared to what our single DFT method has predicted.

2.2 *Superoxide Reductases and Dismutases*

Figure 4 shows proposed catalytic mechanisms for superoxide dismutases (SOD) and superoxide reductases (SOR) [37, 38]. A single redox metal is present at all these enzyme active sites, oscillating between two oxidation states (+2 and +3 for Fe, Ni, or Mn, +1 and +2 for Cu).

A common theme of the SODs and SORs is the question of whether the reduction and/or oxidation of superoxide occurs inner-sphere or outer-sphere. Although there appears to be a consensus, strongly supported computationally, that on both pathways binding of the superoxide to the metal is feasible [38–40], the issue of whether this binding is a prerequisite for completing the electron transfer, or a side effect of it, is still open for discussion. Thus, for nickel-containing superoxide dismutases (NiSOD) the nonprotonated Ni(II)–superoxide complex was predicted by DFT to be unfeasible, featuring the superoxide noncovalently associated with the nickel ligands, rather than coordinated at the Ni itself; at this point, contrary to the formal description from Fig. 4, a Ni(II)-superoxo rather than a Ni(III)-peroxo description was given based on the population analyses and on the O–O bond length [38]. However, protonation of the superoxide led to a state where the OOH moiety was clearly describable as a hydroperoxide, and was shown to be able to coordinate to Ni—hence, a Ni(III)-hydroperoxo state [38]. The one-electron

reduction of the superoxide to (hydro)peroxide appears to be mainly provoked by the protonation event and not by coordination to the metal: indeed, even at Ni–OOH distances excluding coordinative interaction, the OOH moiety retained its hydroperoxo character, suggesting that the reduction of superoxide in NiSOD operates via a protonation-induced outer-sphere electron transfer [38]. Under these conditions, while it is entirely possible for the hydroperoxide to bind to the metal, this event would happen only *after the reduction of superoxide was completed*—and may thus be regarded as a form of substrate inhibition rather than as a key step within the catalytic cycle. Interestingly, for Fe–SOD [38] or for copper-containing SOD [40] such a clear-cut description is not possible: the M^{n+}-superoxo adduct is computed to feature reasonably short Fe–OO bonds (and, incidentally, nonnegligible M^{i+1}-peroxo character). Remarkably, while the NiSOD active site is completely solvent exposed, the iron, the manganese, and the copper SODs have their metal sites further buried inside the protein, as if to preclude superoxide–metal coordination as a form of a nonproductive/inhibitory interaction [38].

For the putative ferrous–superoxide complexes in SOD as well as in SOR, the alternative electromeric state, resulted from a one-electron transfer (ferric-peroxo), has been favored [38], although the O–O bond length, computed at ~1.36 Å with the BP86 functional and slightly shorter with hybrid functionals, was intermediate between those computed to free hydrogen peroxide and free superoxide [37]. Nevertheless, the most recent computational data for SOR suggest a high-spin ferrous-superoxo state, as characterized by ~ one unit of unpaired electron density at the OO ligand, and by O–O bond lengths closer to free superoxide than to free hydrogen peroxide [37, 41]. Furthermore, TD-DFT UV–vis computations suggest that the electronic absorption spectrum of such s species may even feature a band at ~600 nm, in the same region where the resting ferric state features a ligand-to-metal transition (sulfur-to-ferric); spectacularly, in the ferrous-superoxo adduct this band would now arise from *ligand-to-ligand* and *metal-to-ligand* charge transfers [41]. An ongoing debate on whether such a nonprotonated reaction intermediate would be observable is fuelled by reports according to which only one reaction intermediate is observed experimentally for SOR [42]. Based on DFT-derived thermodynamics, correlated with crystallographic information on the O–O bond length [39], this intermediate would most likely be the ferric-hydroperoxo, rather than the nonprotonated ferrous-superoxo—especially as the active site is water exposed and decorated with two potential proton-donating amino acids [37, 39].

The re-reductive pathway, shown in gray in Fig. 4 and specific to SOD but not to SOR, appears more likely to prefer the outer-sphere pathway. Thus, according to DFT data, superoxide coordination to Ni(III)–SOD would be disfavored thermodynamically and unnecessary: the electron transfer from superoxide to Ni, to yield a neutral dioxygen molecule loosely associated with Ni(II)–SOD, occurs even at nonbonding Ni–OO distances [38]. For Fe–SOD, although the diamagnetic ferrous-dioxygen/ferric-superoxo state appeared to feature bond lengths reminiscent of the isoelectronic stable oxy form of globins, the higher spin states, including an essentially nonbonding $S = 3$ situation, were thermodynamically favored, suggesting a drive to avoid proper metal–superoxide adducts on this pathway [38].

2.3 Redox Activation of Water/Hydroxide

As detailed above in the discussion on metal centers in high oxidation states, water may in principle be oxidized upon binding to a metal center, to yield oxyl units or even electrically neutral oxygen atoms. The photosynthetic pathways in plants use this reaction as a convenient source of electrons—with a specialized four-manganese cluster at the active site of the enzyme responsible for the reaction, in an interesting parallel with the Mn-oxyl evidence cited above for mononuclear model compounds. Recent progress in X-ray diffraction determinations has allowed intimate attempts of correlating geometrical and EPR (electron paramagnetic resonance) data with DFT-derived data, attempting to assign oxidation and protonation states in the reactive intermediates en route from water to dioxygen. Paradoxically to some extent, much more appears to remain uncovered on this pathway than on the reverse pathway—dioxygen activation discussed earlier in this chapter [43–46].

3 Nitrogen Oxides and Oxyanions

The nitrogen cycle in nature involves complex transformations, via at least three separate pathways, between nitrate and organic nitrogen/ammonia. Every single step in these processes is catalyzed by metalloenzymes [47]—some of which (e.g., nitrogenase, nitrous oxide reductase) have offered remarkable findings in recent years and continue to pose interesting challenges to computational chemists [48]. Among other enzymes, nitrogenase continues to tantalize computational chemists with new structural features of the iron–sulfur–molybdenum cofactor at active site, especially the mysterious central atom within this cluster; while a few years ago it appeared that computational methods would correctly predict not only structural but even spectroscopic parameters going as far as Mössbauer, recent data suggests that in the center of the cluster is a previously unsuspected carbon atom—hence a new member of the bioorganometallic family [49].

3.1 Nitric Oxide

Most nitric oxide reductases (NOR) fall into two categories: fungal P450-type NOR and bacterial heme/nonheme NOR. P450NOR illustrate a unique mechanism wherein a hydride atom is transferred directly from NAD(P)H to an NO moiety bound at the single heme active site [12]. By contrast, bacterial NOR feature a binuclear active site with a heme and a nonheme iron facing each other; here, the electrons required for reducing nitric oxide are delivered in sequential one-electron steps. Interestingly, while in P450NOR the axial ligand at the heme is a thiolate, in

the bacterial NOR the axial ligand is a histidine. DFT calculations have shown that a sequence of two successive one-electron reductions of an NO molecule bound to an Fe(III) heme would be thermodynamically feasible with a neutral ligand (e.g., histidine), whereas with a thiolate axial ligand the reaction would be blocked at the Fe(II)–NO stage, after the first electron transfer (an example of thiolate obstruction, similar to those discussed above for dioxygen activation) [9, 12]. The thiolate-bound active site of P450NOR manages to bypass the Fe(II)–NO state, by transferring both electrons at once, in the form of a hydride, to the initial Fe(III)–NO adduct—yielding an Fe(II)–HNO state.

Oxidation of nitric oxide by ferrous–dioxygen adducts of hemoproteins occurs accidentally in blood (with hemoglobin) but also as physiological function for certain bacterial globins (e.g., flavohemoglobin); the mechanism of this reaction appears to entail ferric–peroxynitrite adducts at the heme, whose extremely short lifetimes were predicted by DFT calculations [50, 51].

3.2 Nitrite and Nitrate

Nitrite binding to metals is well known in traditional inorganic chemistry as an example of linkage isomerism—occurring either via the nitrogen or via the oxygen atom, cf. Fig. 5. In the past few years it has been recognized that the same linkage isomerism is also at work in metalloproteins [30, 52–56]. An initial DFT prediction was made despite a wealth of experimental data showing nitrite binding to hemoproteins only via its nitrogen atom, binding via the oxygen should also be possible [30]. Soon afterwards, the crystal structures of nitrite-bound myoglobin and hemoglobin were reported, showing indeed the DFT-predicted isomer [52–56]. Subsequent detailed studies, with DFT, ab initio dynamics, and QM/MM approaches [57] have also revealed that nitrite reduction is almost equally feasible via the two isomers, suggesting perhaps one more case of mechanistic promiscuity, as illustrated in Fig. 5 [57]. Also based on DFT data, linkage isomerism was proposed to be feasible in copper-containing nitrite reductases [58], as well as in molybdenum-containing nitrate reductases—as the latter enzymes can reduce nitrite at competition with nitrate [32]. In the case of nitrate reductase, nitrate and nitrite were computed to be reducible at very similar rates [32].

4 Sulfur-Based Compounds

Sulfur, its oxides and its oxyanions are placed on metabolic pathways involved in energy conservation/extraction in bacteria, and are more generally important as basic reagents in biochemical experiments (with dithionite a prime example) [59, 60].

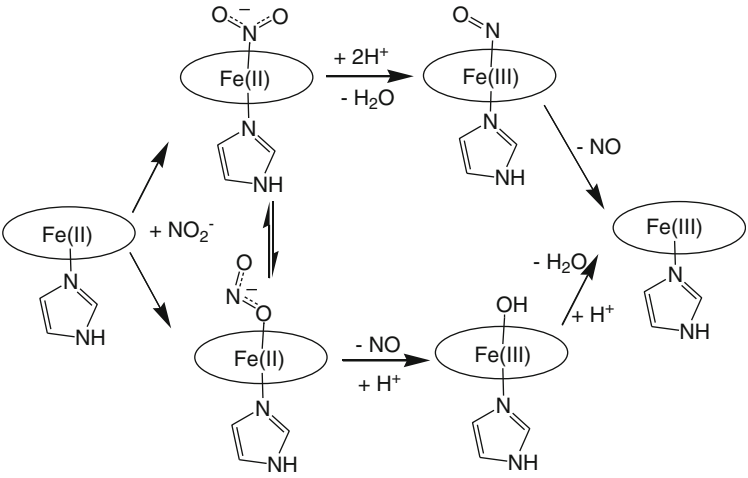


Fig. 5 Nitrite linkage isomerism and reduction at hemoprotein active sites

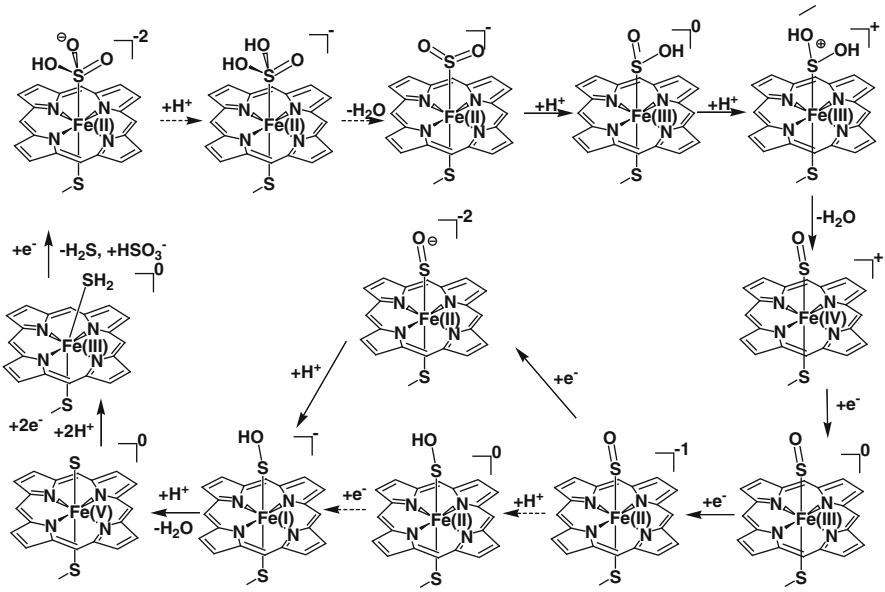


Fig. 6 Proposed catalytic cycle for sulfite reductase

Sulfite reductase features a siroheme-containing active site and catalyzes the six-electron reduction of sulfite to sulfide at the heme iron [31, 61]. A detailed account of each step of the mechanism has recently been provided as detailed in Fig. 6, based on DFT data. Thus, the catalytic cycle would be initiated with a Fe(III)–SO_3H^- adduct, from which protonation and water elimination lead to a

Fe(III)–SO_2^0 species where the ligand binds to iron via an *oxygen* atom, in an example of linkage isomerism reminiscent of the nitrite case and unprecedented for sulfur compounds in enzymes. A facile one-electron reduction to a Fe(III)–SO_2^- adduct ensues. This adduct would then undergo two subsequent protonation events and eliminate a water molecule to yield a Fe(III)–SO^+ species. The latter was computed to easily undergo three successive one-electron reduction events, to Fe(II)–SO^-. From here, two successive protonation events were computed to lead to a Compound I-type species, with a Fe(IV)–sulfide moiety coupled to a cation radical delocalized onto the axial thiolate and onto the porphyrin. A one-electron reduction and a subsequent protonation would lead to an Fe(IV)-SH state akin to the Fe(IV)-OH Compound II species seen in thiolate-ligated hemoproteins from the chloroperoxidase family. This protonated sulfide Compound II would then return to the ferric state by a further one-electron reduction, liberating hydrogen sulfide in a protonation-dependent final step of the catalytic cycle. The sulfide adducts are likely to provide further computational interest, in light of hydrogen sulfide's recently discovered implications in cellular signaling; [62] sulfur oxyanions also appear to offer surprises in coordination chemistry not only in heme proteins but also in systems such as cobalamin [63].

5 Carbon-Based Compounds

Carbon-based compounds such as carbon dioxide or aliphatic hydrocarbons pose challenges of reactivity to metalloenzymes insofar as they are generally reluctant to bind to metal active sites; a general rule is that such substrates would usually be attacked by metal-activated nucleophiles. However, in selected cases, even binding to metal may be feasible, as detailed below.

One class of carbon dioxide-activation enzymes, the carbonic anhydrases (CANH), employ a metal-bound hydroxide as a nucleophile with (bi)carbonate as a final product [64]. As illustrated in Fig. 7, the metal employed in CANH is zinc; recent DFT data shows that the choice of zinc in this enzyme (as opposed to other transition metals, such as cobalt, iron, or copper) is not so much related to differences imparted on the nucleophilicity of the metal-bound hydroxide by the various metals: indeed, hydroxides in other enzyme models (including well-known species such as met-hemoglobin, lipoxygenase, or a superoxide reductase mutant), as well as with other metal centers, were all computed to require reasonably small activation energies when reacting with carbon dioxide [64]. Nevertheless, in most cases the coordination environment around the metal was too crowded, thus hindering an efficient binding of the product to the metal—eventually resulting in an endergonic process. Therefore zinc, with its preferred tetrahedral coordination and reasonable kinetic lability (arguably, eventually linked to other intrinsic electronic properties [65, 66]), is ideally suited for the CANH reaction.

Carbon dioxide may also be activated via direct redox reaction with metals—possibly involving metal–carbon bonds, and necessarily involving low oxidation

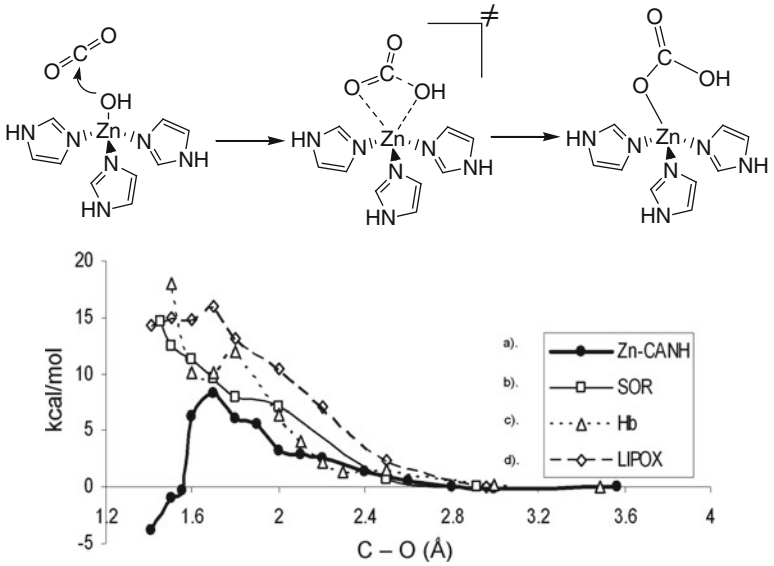


Fig. 7 *Upper panel*: currently accepted mechanism for CO_2 activation by carbonic anhydrases. *Lower panel*: comparison of reaction profiles for CANH versus three other enzyme active sites, not known to catalyze carbon dioxide hydration (*SOR* superoxide reductase, *Hb* hemoglobin, *LIPOX* lipoxygenase)

states at the metal [67]. A recent survey of low-oxidation state heme and related systems, with Fe(0) and Fe(I) states [67, 68], has revealed that although in general the extra oxidizing equivalents in such "super-reduced" states will tend to be localized on the ligands rather than on the metal itself, in one case an Fe–CO_2 bond could be established, with an electronic structure suggestive of a true low oxidation state at the metal. Thus, Fig. 8 illustrates the frontier molecular orbitals for an Fe(0) heme system ligated axially by CO_2 and imidazole; it may be seen that only the d_{x2-i2} orbital at the metal is empty—hence a formally d^8 description, suggestive of an Fe(I) center bound to an activated carbon dioxide radical. This is relevant in the context where such low oxidation states are known experimentally to be involved in catalytic reduction of carbon dioxide [69, 70]. Interestingly, the CO adducts of Fe(0) or Fe(I) heme-like systems were computed to in fact consist of Fe(II) coupled to macrocyclic anion radicals [67].

The oxygenation of organic compounds at metal centers is often accomplished by metal-bound oxygen atoms—be it a hydroxide as in lipoxygenase, or a peroxide-derived atom as in cytochromes P450 or Rieske dioxygenases discussed above [25, 27, 29, 71]. An interesting alternative, not seen in enzymes but seen in model compounds, are the metal-bound nitrites: these are known experimentally to act as reasonable oxygenation agents; their reactivity in this respect has been tested with DFT methods, revealing that despite an apparently drastic difference in electronic structure, the oxyl-like ligand in P450 Compound I (as discussed above) and an

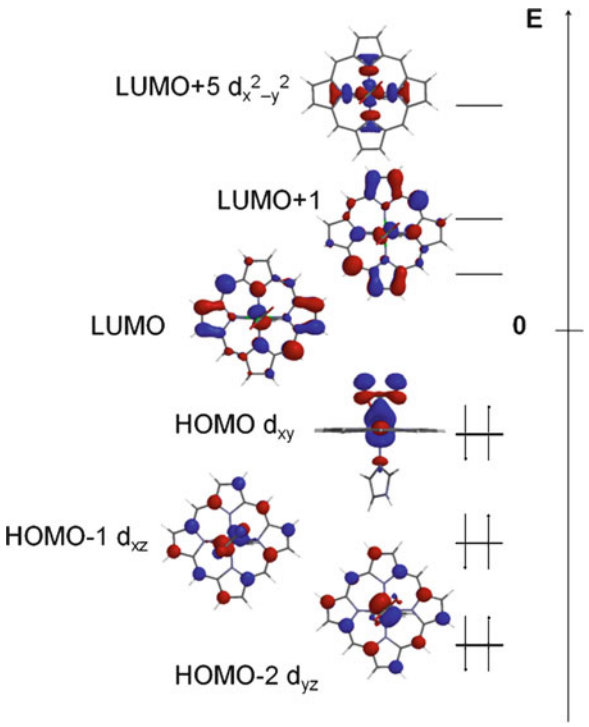


Fig. 8 Metal-localized frontier orbitals for a formally Fe(0) heme bound to carbon dioxide, with an imidazole ligand trans to the CO_2

oxygen atom within a ferric-nitro adduct, feature very similar energy profiles [71]. A key role was therefore inferred the redox potential in dictating oxygenation ability, in good agreement with a previous proposal of Green and coworkers based on P450 models alone [7, 71] (Fig. 9).

6 Halides and Their Derivatives

Heme enzymes such as chloroperoxidase (CPO) and myeloperoxidase (MPO) perform a two-electron oxidation of halides to formally generate hypohalous acids, HOX. These acids can either be released or used for halogenation reactions. Most peroxidases oxidize iodide, some oxidize bromide, and very few (of which MPO and CPO are classical examples) are known to oxidize chloride. Using UV–vis spectroscopy, intermediates arising from the reaction of Compound I with chloride and proposed to consist of Fe(III) adducts with HOCl and OCl^- were detected with model compounds as well with CPO. Recently, the structure of these intermediates was explored with DFT. The oxygen–halogen bonds were computed to be particularly weak (cf. Fig. 10, broken exergonically with a barrier of only ~10 kcal/mol), to the extent where ferric hypohalous acid complexes with intact oxygen–halide bonds should be difficult to observe in heme peroxidases. Instead, a "caged" adduct, featuring a ferric species in close (but not covalent) contact with the oxidized halide, appears

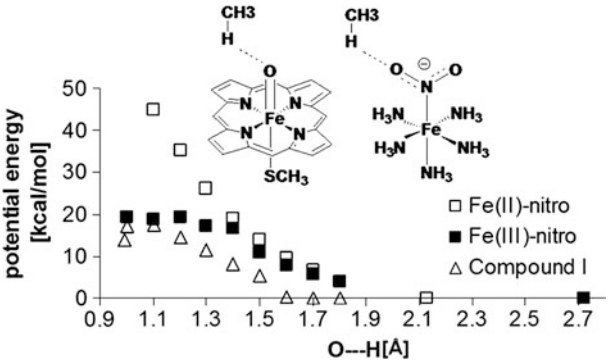


Fig. 9 Potential energy as a function of the O···H distance for hydrogen atom abstraction from methane by three different agents: a thiolate-ligated Compound I, or a ferric or ferrous-nitro complex (cf. structures shown). For each model, the energy of the reactants (*far right side* of the plot) is taken as reference [71]

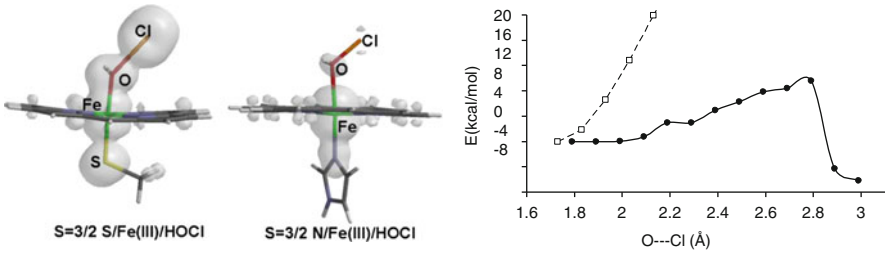


Fig. 10 *Left*: optimized geometries and spin density distributions in models of ferric heme-containing haloperoxidases bound to hypochlorous acid, with axial thiolate or imidazole ligands. *Right*: Potential energy surface for O–Cl bond elongation (*black circles*) in the thiolate-ligated Fe–O–Cl model. *White squares*: equivalent potential surface in an isolated HOCl molecule

as a likely alternative for an observable intermediate in CPO and MPO. The presence of an axial thiolate (cysteinate in CPO) or imidazole (histidine in MPO) ligated to the iron trans to the hypohalous moiety was calculated to have a distinct effect on halide activation, including a spin-state preference with relevance to the TSR (two-state reactivity) concept discussed above [72]. Indeed, Fig. 10 illustrates optimized geometries and computed spin densities for ferric-HOCl models of CPO and MPO. Notably, the O–Cl bond is almost completely broken in the thiolate-ligated model (and consequently significant spin density accumulates to the chlorine), whereas the same bond is essentially intact in the imidazole-ligated model (with the Fe–O bonds following opposite trends—i.e., weaker in the imidazole-ligated model). Furthermore, the thiolate-ligated model, but not the imidazole-ligated one, has a low-spin ground state, where the O–Cl bond is even further activated. Remarkably, in good correlation with these differences, the thiolate-ligated CPO specializes in immediate cleavage of the O–Cl bond so that it can insert the resulted chlorine atom into a hydrocarbon, whereas the imidazole-ligated MPO specializes in liberating intact HOCl.

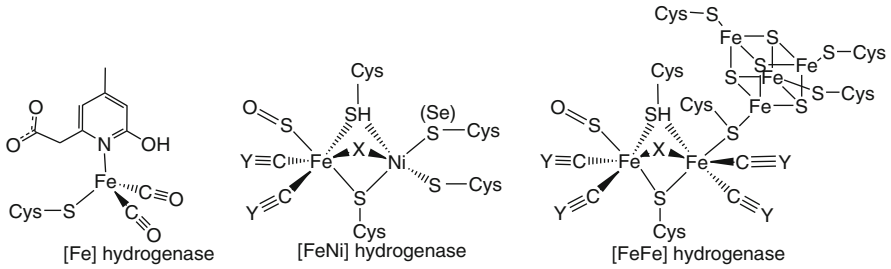


Fig. 11 Hydrogenase active sites. *Y* can be nitrogen, oxygen or sulfur; *X* is most likely sulfur

Vanadium haloperoxidases (VPO) offer a perplexing alternative to heme-containing CPO or MPO. While VPOs also use hydrogen peroxide to activate halides, their active site contain a vanadyl unit already in a high oxidation state—as opposed to CPO and MPO where the role of the hydrogen peroxide was precisely to mount the metal into a higher oxidation state [73–77]. Experimental and DFT data appear to concur on the fact that in VPO the halide is directly activated via a side-on peroxo adduct at vanadium; the protonation states of the ligands, as well as the influence of the protein on the reactivity and spectroscopic properties, have been thoroughly indexed with DFT, TD-DFT, and QM/MM calculations [73–77].

7 Hydrogen/Protons

Hydrogenases catalyze the reversible transformation of molecular hydrogen into protons + electrons; both of the partners in this reaction—but especially the protons—are particularly reluctant to ligate to most biological metal centers. Hydrogenases feature three types of active sites: mononuclear iron, diiron, and binuclear iron–nickel, cf. Fig. 11.

In all cases, unusual ligands are found at the metals—with carbon monoxide or cyanide being the least expected; together with the sulfur-rich ligands, this makes for an unusual coordination environment, where, among others, lower oxidation states at the metal are possible—a clear advantage in the attempt to convince protons to bind to what may otherwise be regarded as an ordinary transition metal in aqueous phase. The small size of the substrate implies serious difficulties in applying some of the most accurate and informative experimental techniques for trapping reaction intermediates—X-ray crystallography and X-ray absorption spectroscopy. The complete details of the mechanisms of hydrogenases are therefore still a matter of investigation. This is thus a field where the reliability of DFT applications is of particular importance; not surprisingly, a significant number of reports have covered the area in recent years [78–91]. The *qualitative* dependence of the results on the size of the model as well as on the functional employed [87], imply a significant need for reconsideration of the performance of the DFT

methods, as different functionals may yield opposing conclusions on the same problem (as also discussed above for e.g., Rieske dioxygenases). In this respect, the progress made by newer parameterized methods, where selective targets are set for performance (e.g., only weak interactions or only (organo)metallic compounds) [3] offers significant hope. Nevertheless, the concept of such target-dependent functionals reignites, in parallel, the interest in true ab initio methods—whose performance would be less dependent on the type of chemical system examined.

8 Summary and Outlook

Recent progress in methodology with DFT methods has allowed tackling of increasingly challenging tasks in terms of understanding mechanisms and making predictions on small molecule activation by biological metal centers. The potential practical applications of such investigations cannot be overstated: fields like proton activation for molecular hydrogen production, water splitting, molecular nitrogen activation, or carbon dioxide fixation still have unexplored facets, with computational methods able to provide key input for generating robust economically feasible catalysts. On a different note, species that may be too short lived to be observed experimentally are facile targets for DFT predictions. Also, recent years have seen a focus on allowing predictions on spectroscopic parameters [92–95]. One important component is, however, the responsibility of the researcher to properly assess the predictive value of the DFT results. The times where a single method would be applied on a small set of models, concluding that "theory confirms experiment," have long passed. We have given here examples of how some of the mainstream functionals can easily lead to opposing conclusions on key aspects of the topic investigated; in our experience these differences between functionals may amount to as high as 30 kcal/mol in terms of spin state preferences, and not much lower for activation barrier differences; these values are much larger than the 1–5 kcal/mol commonly assumed in past years. From this point of view, methodological contributions such as reference [3] will in our opinion prove to be by far among the most useful for the future development of the field.

Acknowledgment Funding from the Romanian Ministry for Education and Research project PCCE 140/2008 is gratefully acknowledged.

References

1. Tolman WB (2006) Activation of small molecules: organometallic and bioinorganic perspectives. Wiley, New York, 382 pp
2. Kraatz HB, Metzler-Nolte N (2006) Concepts and models in bioinorganic chemistry. Wiley-VCH, Weinheim
3. Zhao Y, Truhlar DG (2006) Acc Chem Res 41:157

4. Hersleth HP, Ryde U, Rydberg P, Gorbitz CH, Andersson KK (2006) J Inorg Biochem 100:460
5. Groves JT (2006) J Inorg Biochem 100:434
6. Decker A, Clay MD, Solomon EI (2006) J Inorg Biochem 100:697
7. Green MT, Dawson JH, Gray HB (2004) Science 304:1653
8. Silaghi-Dumitrescu R, Silaghi-Dumitrescu I (2004) Rev Roum Chim 3–4:257
9. Silaghi-Dumitrescu R (2006) Studia Univ Babes-Bolyai Chemia 51:167
10. Rydberg P, Sigfridsson E, Ryde U (2004) J Biol Inorg Chem 9:203
11. Jensen KP, Roos BO, Ryde U (2005) J Inorg Biochem 99:45–54
12. Silaghi-Dumitrescu R (2003) Eur J Inorg Chem 1048
13. Silaghi-Dumitrescu R (2008) Revue Roumaine de Chimie 53:1149
14. Silaghi-Dumitrescu R (2004) J Biol Inorg Chem 9:471
15. Stone KL, Behan RK, Green MT (2006) Proc Natl Acad Sci USA 103:12307
16. Stone KL, Hoffart LM, Behan RK, Krebs C, Green MT (2006) J Am Chem Soc 128:6147
17. Green MT (2006) J Am Chem Soc 128:1902
18. Behan RK, Hoffart LM, Stone KL, Krebs C, Green MT (2006) J Am Chem Soc 128:11471
19. Behan RK, Green MT (2006) J Inorg Biochem 100:448
20. Stone KL, Behan RK, Green MT (2005) Proc Natl Acad Sci USA 102:16563
21. Silaghi-Dumitrescu R, Reeder BJ, Nicholls P, Cooper CE, Wilson MT (2007) Biochem J 403:391
22. Silaghi-Dumitrescu R (2005) Studia Univ Babes-Bolyai Chemia 50:17
23. Shaik S, de Visser SP, Kumar D (2004) J Biol Inorg Chem 9:661
24. Silaghi-Dumitrescu R (2007) Rev Chim 58:461
25. Silaghi-Dumitrescu R, Makarov SV, Uta M-M, Dereven'kov IA, Stuzhin PA (2011) New J Chem 35:1140
26. Lassalle-Kaiser B, Hureau C, Pantazis DA, Pushkar Y, Guillot R, Yachandra VK, Yano J, Neese F, Anxolabéhère-Mallart E (2010) Energy Environ Sci 3:924
27. Silaghi-Dumitrescu R (2010) New J Chem 34:1830
28. Bassan A, Blomberg MR, Siegbahn PE (2004) J Biol Inorg Chem 9:439
29. Silaghi-Dumitrescu R, Cooper CE (2005) Dalton Trans 3477
30. Silaghi-Dumitrescu R (2004) Inorg Chem 43:3715
31. Silaghi-Dumitrescu R, Makarov SV (2012) Int J Quant Chem 112:900
32. Silaghi-Dumitrescu R, Mich M, Matyas C, Cooper CE (2012) Nitric Oxide 26:27
33. Balcells D, Moles P, Blakemore JD, Raynaud C, Brudvig GW, Crabtree RH, Eisenstein O (2009) Dalton Trans 5989
34. Balcells D, Raynaud C, Crabtree RH, Eisenstein O (2009) Chem Commun (Camb) 1772
35. Silaghi-Dumitrescu R (2007) Studia Universitatis Babes-Bolyai Chemia 103
36. Silaghi-Dumitrescu R (2007) Studia Universitatis Babes-Bolyai Chemia 2:127
37. Silaghi-Dumitrescu R, Silaghi-Dumitrescu I, Coulter ED, Kurtz DM Jr (2003) Inorg Chem 42:446
38. Silaghi-Dumitrescu R (2009) J Mol Graph Model 28:156
39. Katona G, Carpentier P, Niviere V, Amara P, Adam V, Ohana J, Tsanov N, Bourgeois D (2007) Science 316:449
40. Pelmenschikov V, Siegbahn PE (2005) Inorg Chem 44:3311
41. Bonnot F, Molle T, Menage S, Moreau Y, Duval S, Favaudon V, Houee-Levin C, Niviere V (2012) J Am Chem Soc 134:5120–5130
42. Kurtz DM Jr (2004) Acc Chem Res 37:902
43. Ames W, Pantazis DA, Krewald V, Cox N, Messinger J, Lubitz W, Neese F (2011) J Am Chem Soc 133:19743
44. Cox N, Rapatskiy L, Su JH, Pantazis DA, Sugiura M, Kulik L, Dorlet P, Rutherford AW, Neese F, Boussac A, Lubitz W, Messinger J (2011) J Am Chem Soc 133:3635
45. Duboc C, Collomb MN, Pecaut J, Deronzier A, Neese F (2008) Chemistry 14:6498
46. Zein S, Kulik LV, Yano J, Kern J, Pushkar Y, Zouni A, Yachandra VK, Lubitz W, Neese F, Messinger J (2008) Philos Trans R Soc Lond B Biol Sci 363:1167

47. Silaghi-Dumitrescu R (2009) Revue Roumaine de Chimie 54:513
48. Silaghi-Dumitrescu R, Silaghi-Dumitrescu I (2005) Chemtracts Inorg Chem 18:595
49. Lancaster KM, Roemelt M, Ettenhuber P, Hu Y, Ribbe MW, Neese F, Bergmann U, DeBeer S (2011) Science 334:974
50. Silaghi-Dumitrescu R (2005) J Mol Struct Theochem 722:233
51. Blomberg LM, Blomberg MRA, Siegbahn PEM (2004) J Biol Inorg Chem 9:923–935
52. Yi J, Orville AM, Skinner JM, Skinner MJ, Richter-Addo GB (2010) Biochemistry 49:5969–5971
53. Yi J, Heinecke J, Tan H, Ford PC, Richter-Addo GB (2009) J Am Chem Soc 131:18119
54. Zahran ZN, Chooback L, Copeland DM, West AH, Richter-Addo GB (2008) J Inorg Biochem 102:216
55. Yi J, Safo MK, Richter-Addo GB (2008) Biochemistry 47:8247
56. Copeland DM, Soares AS, West AH, Richter-Addo GB (2006) J Inorg Biochem 100:1413
57. Perissinotti LL, Marti MA, Doctorovich F, Luque FJ, Estrin DA (2008) Biochemistry 47:9793
58. Silaghi-Dumitrescu R (2006) J Inorg Biochem 100:396
59. Simon J, Kern M, Hermann B, Einsle O, Butt JN (2011) Biochem Soc Trans 39:1864
60. Zhou J, He Q, Hemme CL, Mukhopadhyay A, Hillesland K, Zhou A, He Z, Van Nostrand JD, Hazen TC, Stahl DA, Wall JD, Arkin AP (2011) Nat Rev Microbiol 9:452
61. Crane BR, Siegel LM, Getzoff ED (1995) Science 270:59
62. Pietri R, Roman-Morales E, Lopez-Garriga J (2011) Antioxid Redox Signal 15:393
63. Salnikov DS, Silaghi-Dumitrescu R, Makarov SV, van Eldik R, Boss GR (2012) Dalton Trans 40:9831
64. Silaghi-Dumitrescu R (2010) J Mol Struct Theochem 942:15
65. Putz MV, Russo N, Sicilia E (2004) J Comput Chem 25:994–1003
66. Putz MV (2008) MATCH Commun Math Comput Chem 60:845
67. Kis Z, Silaghi-Dumitrescu R (2010) Int J Quant Chem 110:1848
68. Silaghi-Dumitrescu R, Makarov SV (2010) J Biol Inorg Chem 15:977
69. Grodkowski J, Dhanasekaran T, Neta P, Hambright P, Brunschwig BS, Shinozaki K, Fujita E (2000) J Phys Chem A 104:11332
70. Makarov SV, Salnikov DS, Pogorelova AS, Kis Z, Silaghi-Dumitrescu R (2008) J Porph Phthalocyan 12:1096
71. Silaghi-Dumitrescu R, Makarov SV (2010) Eur J Inorg Chem 34:1830
72. Silaghi-Dumitrescu R (2008) Eur J Inorg Chem 5404
73. Geethalakshmi KR, Waller MP, Thiel W, Buhl M (2009) J Phys Chem B 113:4456
74. Zhang Y, Gascon JA (2008) J Inorg Biochem 102:1684
75. Schweitzer A, Gutmann T, Wachtler M, Breitzke H, Buchholz A, Plass W, Buntkowsky G (2008) Solid State Nucl Magn Reson 34:52
76. Wikete C, Wu P, Zampella G, De Gioia L, Licini G, Rehder D (2007) Inorg Chem 46:196
77. Zampella G, Fantucci P, Pecoraro VL, De Gioia L (2006) Inorg Chem 45:7133
78. Dey A (2011) J Am Chem Soc 132:13892
79. Leidel N, Chernev P, Havelius KG, Ezzaher S, Ott S, Haumann M (2012) Inorg Chem 51:4546
80. Baffert C, Bertini L, Lautier T, Greco C, Sybirna K, Ezanno P, Etienne E, Soucaille P, Bertrand P, Bottin H, Meynial-Salles I, De Gioia L, Leger C (2011) J Am Chem Soc 133:2096
81. Kachmar A, Vetere V, Maldivi P, Franco AA (2010) J Phys Chem A 114:11861
82. Greco C, Fantucci P, De Gioia L, Suarez-Bertoa R, Bruschi M, Talarmin J, Schollhammer P (2010) Dalton Trans 39:7320
83. Ryde U, Greco C, De Gioia L (2010) J Am Chem Soc 132:4512
84. Surawatanawong P, Tye JW, Darensbourg MY, Hall MB (2010) Dalton Trans 39:3093
85. Yang X, Hall MB (2009) J Am Chem Soc 131:10901
86. Bertini L, Greco C, De Gioia L, Fantucci P (2009) J Phys Chem A 113:5657
87. Jayapal P, Sundararajan M, Hillier IH, Burton NA (2008) Phys Chem Chem Phys 10:4249
88. Jayapal P, Robinson D, Sundararajan M, Hillier IH, McDouall JJ (2008) Phys Chem Chem Phys 10:1734

89. Fiedler AT, Brunold TC (2005) Inorg Chem 44:9322
90. Justice AK, De Gioia L, Nilges MJ, Rauchfuss TB, Wilson SR, Zampella G (2008) Inorg Chem 47:7405
91. Zhou T, Mo Y, Zhou Z, Tsai K (2005) Inorg Chem 44:4941
92. Petrenko T, Neese F (2007) J Chem Phys 127:164319
93. Ganyushin D, Neese F (2008) J Chem Phys 128:114117
94. Pantazis DA, Orio M, Petrenko T, Zein S, Bill E, Lubitz W, Messinger J, Neese F (2009) Chemistry 15:5108
95. Ganyushin D, Neese F (2006) J Chem Phys 125:24103

Struct Bond (2013) 150: 119–142
DOI: 10.1007/978-3-642-32750-6_4

The Bond Analysis Techniques (ELF and Maximum Probability Domains) Application to a Family of Models Relevant to Bio-Inorganic Chemistry

Mauro Causà, Maddalena D'Amore, Carmine Garzillo, Francesco Gentile, and Andreas Savin

Abstract Electron Localization Function (ELF) and Maximum Probability Domain (MPD) analyses have been applied to model metal–porphyrins and show compatible and complementary results. ELF basins are quite different from MPDs, but are a necessary starting point for optimizing them. The analyses of the bond between the metal and porphyrin do not show significant differences between non-transition and transition metals. In all the cases considered, we find signatures characteristic of essentially ionic bonds.

Keywords Density functional theory • Electron localization function • Maximum probability domains • Molecular orbitals

M. Causà (✉) • M. D'Amore
Dipartimento di Scienze Chimiche, Università di Napoli "Federico II", via Cintia, 80126 Napoli, Italy
e-mail: mauro.causa@unina.it; mdamore@unina.it

C. Garzillo
Dipartimento di Scienze mediche preventive, Università "Federico II", via S. Pansini 5, 80131 Napoli, Italy
e-mail: carmine.garzillo@unina.it

F. Gentile
Italian Institute of Technology, CRIB, Naples, Italy

Dipartimento di Ingegneria dei Materiali e della Produzione, Universita' di Napoli "Federico II", Piazzale V. Tecchio 80, 80125 Napoli, Italy
e-mail: francesco.gentile@unina.it

A. Savin
Laboratoire de Chimie Theorique, Universite "Pierre et Marie Curie", Place Jussieu, Paris, France
e-mail: andreas.savin@lct.jusieu.fr

Contents

Abbreviations

AIM	Atoms in molecule
AO	Atomic orbital
DFT	Density functional theory
ELF	Electron localization function
ELI	Electron localizability indicator
LOL	Localized orbital locator
MO	Molecular orbital
MPDs	Maximum probability domains

1 Introduction

In the recent years, the traditional discussion of chemical bonding in terms of the canonical orbitals has been challenged [1–3]. In fact, in general SCF theory, the definition of orbitals remains ambiguous [4–6] and several localization procedures have been introduced by suitable unitary transformations of the set of canonical orbitals in order to provide a more rigorous mathematical meaning to chemical concepts such as bonded atoms, reactants, chemical bonds, electron shells, and lone electron pairs [7–9]. Yet, the realization that theoretical partitionings are not unique "casts a shadow of mistrust over certain definitions" [10].

Alternative approaches to the many-electron problem, working in real space rather than in Hilbert space and with the electron density playing the major role, are provided by Bader's "atoms in molecule" [11, 12], which partitions the molecular space into basins associated with each atom and density-functional methods [3, 13]. These latter are based on a modified Kohn–Sham form of the one-electron effective Hamiltonian, differing from the Hartree–Fock operator for the inclusion of a "correlation potential." In these methods, it is possible to mimic correlated natural orbitals, as eigenvectors of the first-order reduced density operator, directly

from one-electron eigenvalue equations, bypassing the calculation of multi-configurational Möller–Plesset or CI wave functions [14]. Such a localized "Lewis-like" description of bonding and reactivity, which uses a natural bond orbital analysis (associating bonds with localized two-center two-electron wave functions), has recently been expanded to bulk materials and/or periodic surface models [15].

As a matter of fact, there are systems that cannot be easily characterized in terms of localized molecular orbitals, and thus not easily understood using localized bonds. An extreme example is provided by the fluxional behavior of CH_5^+, where quantum-mechanical fluctuations induce delocalization phenomena and thus fluctuating bonds [16–20].

Most importantly, these systems are amenable to the Electron Localization Function (ELF) method [21]. This is a local measure based on the reduced second-order density matrix, which as pioneered by Lennard-Jones [22] should retain the chemical significance and at the same time reduce the complexity of the information contained in the square of the wave function Ψ. ELF is defined in terms of the excess of local kinetic energy density due to the Pauli exclusion principle, $T(r)$, and the Thomas–Fermi kinetic energy density, $T_h(r)$:

$$\mathrm{ELF} = \left[1 + \left(\frac{T(r)}{T_h(r)}\right)^2\right]^{-1}. \tag{1}$$

In the seminal paper of Becke and Edgecombe [20] it was proposed that a topological analysis of the ELF [23–25] permitted a more quantitative analysis of the three-dimensional function. The latter is normalized to the interval between 0 and 1, and is large where the Pauli repulsion is small (two electrons with anti-parallel spin are paired), and is small in the regions between electron pairs. To identify a region around a maximum (an attractor), which forms a basin, one can consider all the points in the space with ELF $\geq \zeta$, which defines the ζ localization domain and with ζ being a positive constant smaller than the value of the maximum. This region in the space can be visualized by showing the iso-surface ELF $= \zeta$. All points leading to the same attractor belong to the same basin.

The ELF has been extensively applied to a large number of systems and has also been used to quantify chemical concepts like the strength of the hydrogen bond [26] and aromaticity [27, 28]. Moreover, other interesting ELF-like scalar fields have been developed, such as the localized orbital locator (LOL) [29], based on the comparison of the local noninteracting kinetic energy density with that of the uniform electron gas, and the electron localizability indicator (ELI) [30–32], derived directly from the electron pair density without any reference to the uniform electron gas.

In order to get out of the somewhat confusing definition and interpretation of ELF, a more thorough investigation of the competition of kinetic terms $T(r)$ and $T_h(r)$ of Eq. (1), between regions in which the inhomogeneous or the homogeneous behavior dominates, was carried out by Putz [33, 34]. Employing a path integral

Markovian pair conditional probability density with the basic concepts of the catastrophe theory, he succeeded in introducing new Markovian ELF classes which generalize the previous Becke–Edgecombe definition. Going beyond the actual interpretation of ELF as the error in electron localization, this new approach provides a quantum step-function indicating where the electrons are trapped rather than where they have peaks of spatial density.

From a computational point of view, it should be stressed that the computational tool of Francisco et al. [35] results in obtaining the electron number probability distribution functions of an *n*-electron molecule through an exhaustive partitioning of the real space into arbitrary regions. From the computed probabilities, several magnitudes relevant to chemical bonding theory are obtained, such as average electronic populations and localization/delocalization indices.

Recently, Corminboeuf et al. [36] have compared the ELFs of both the canonical wave functions and electron-localized states with those of ELI and LOL in order to discriminate between enhanced or reduced electron (de)localization within cyclic π-conjugated systems. As a result, the authors say, the simplest LOL function gives a more appealing and intuitive picture of the π-bond, whereas the most popular ELF fails to capture subtle contrasting local electronic properties and suffers from the arbitrariness of the σ/π dissection. Some ELF limitations have been thoroughly discussed in recent years [37] and attempts to search for new tools, capable of providing solutions to still unsolved difficulties in chemical understanding, have been done. A promising methodology has been obtained through a topological analysis by means of Maximum Probability Domains (MPDs) [38–41, 53] and this is discussed in detail below.

In this paper, the chemical bonding of some metal porphyrins is extensively analyzed using both ELF and MPDs, the latter revealing itself to be necessary to gain further chemical information and allowing us to identify and discuss the subtle variations of the electron density with respect to the various macrocycle/metal junctions. Macrocycles such as porphyrins and their analogues [42, 43] have attracted much attention because of their remarkable role in biological photoredox processes and in potentially semiconducting polymers; more recently, they have found applications in the detection and photodynamic therapy of tumors [44]. Therefore, attempts to clearly establish the structure–property relationships in this class of molecules promise to be extremely useful for a better understanding of the mechanisms of those processes. A preferential way to achieve that goal consists in trying to identify the structural features most likely to be responsible for electron donor/acceptor properties and eventually to compare the properties of porphyrin macrocycles to those of other macrocycles with different features, where the features in question are absent. In the present study, for the first time, we explore the performances of the MPDs method when applied to metal porphyrin compounds.

2 Methods

2.1 Algorithms for Managing and Optimizing Maximum Probability Domains

The optimization algorithm of the Domains in accordance with MPDs approach has been described in references [26, 38]. An atomic or ELF basin, defined on a grid, or another domain chosen by the user (a sphere, an ellipsoid, a cube, or a previously obtained MPD) can be a first guess for the MPDs.

2.1.1 Calculation of Surface and Integration

First a regular cubic grid of points $G = \{g_i\}$ is established with a given increment dx. The classical value of dx is 0.05 bohrs. A volume $dV = dx^3$ is associated to the points g_i.

Definition of Domains

A subset $D = \{d_i\}$ of G is defined As constituted by grid points belonging to the domain. The number of the d_i points is n_v. The volume of the domain is: $V_1[D] = n_v\ dV$.

Definition of D for AIM and ELF Basin

The scalar field $s(\mathbf{x})$ (where s is the electronic density for AIM or ELF) is calculated for all the points $\{g_i\} = G$. Afterwards, the set of attractors (maxima) is calculated, first through the use of the numerical gradient of $s(\mathbf{x})$ on the grid G, and next the positions of the attractors are refined using the analytical gradient of s. The g_i points are attributed to each attractor using the numerical gradient. $V_1[D]$ is the most accurate method of integration of the volume for ELF and AIM basins.

2.1.2 Definition of D for MPD

A first guess D can be defined by:

- Using a D set of points taken from AIM or ELF analysis.
- Searching for a set of grid points interior to a guess surface (sphere, ellipse, cube, or a previous MPD surface).

The set D of points d_i is maintained and redefined during the surface optimization for keeping the information about the interior–exterior space of the surface.

2.1.3 Construction of the Surface

Given the set of domain of grid points D, a subset $S = \{s_i\}$ of D is defined as constituting points that have less than six neighbors belonging to D. This is the surface set of grid points. The set S is then made independent of G and D, because the position of s_i may be changed.

2.1.4 Elaboration of the Surface *S*

The set S is then multiply screened and smoothed (the positions of s_i are a little modified to set triangularization).

2.1.5 Triangularization

The set of points S is triangularized. A set of triangles $T = \{t_i\}$ having corners s_i is defined. The process is a non-Delaunay one to allow for concavities. The triangles are defining walking in the set of nearest neighbors s_i for obtaining locally the most regular triangles. The set $B = \{\mathbf{b}_i\}$ of barycenter of the triangles t_i is defined. The set $N = \{\mathbf{n}_i\}$ of normal vectors to each t_i is defined, pointing toward the external space of the domain: the set of domain grid points D is used for defining the versor of $\mathbf{n}_i$. The information about the domain is then represented by D, B, and N.

2.1.6 Variation and Optimization of the Surface

The set B is changed during optimization, $\mathbf{b}_i$ are moved along $\mathbf{n}_i$. The shape derivative is calculated in the points $\mathbf{b}_i$ and helps defining the step and sign of variation of $\mathbf{b}_i$ along $\mathbf{n}_i$.

For optimization, the barycenter of the triangles (of the surface S) is moved along the normals. The displacements are proportional to the shape derivatives computed at the barycenter [40]. They are larger at the start of the optimization, smaller towards the end. During the optimization process, certain regions of domain can collapse to a surface, or even points. These low-coordinated grid points are then eliminated.

2.1.7 Redefinition of the Surface

After a controlled number of variations of B, the set B is no more coherent with D. So the domain and the surface are redefined: a new set of grid points $D = \{d_i\}$ is defined internally to the surface represented by B and N. Then, the sets S, T, B, and N are redefined on the base of the newly defined D.

2.1.8 Integration Method for Probability Evaluation

The overlap matrix $S[D] = \{S_{mn}\}$ between AO m and n is calculated; s_{mn} is the overlap limited to the domain D, whereas S_{mn} is the integral extended to all the space. A Becke atomic partition method is adopted for defining the quadrature set of points $X = \{x_i\}$ and the weights w_i:

$$S_{mn} = \chi_m(\mathbf{x_i})\chi_n(\mathbf{x_i})\mathrm{w}_i$$

$$s_{mn} = \chi_m(\mathbf{x_i})\chi_n(\mathbf{x_i})w_i f(\mathbf{x}_i).$$

$f(\mathbf{x}_i)$ is the domain weight function for limiting the overlap to D, χ_m is the m-th atomic orbital of the molecule model or of the cluster chosen as a representation of the periodic system in terms of a set of local Wannier functions. The volume of the domain D can be calculated as:

$$V_2[D] = \sum\nolimits_i w_i f(\mathbf{x}_i).$$

The average number of electrons in the domain $\langle N\rangle[D]$ can be calculated in two ways:

1. $$\langle N\rangle[D] = \sum_{n=0}^{n_e} n\, p_n$$

   where p_n are the probabilities of finding exactly n electrons in the domain D and $\mathrm{n_e}$ is the number of electrons in the molecule or cluster. The set of probabilities $P = \{p_n\}$ is a functional of the domain overlap matrix $S[D]$.
2. $$\langle N\rangle[D] = \sum_i w_i\, \rho(x_i)$$

   where $\rho(x_i)$ is the electronic density calculated in grid point x_i.This method is used only for checking the accuracy of p_i.

There are two methods for calculating the domain function $f(\mathbf{x}_i)$

1. Given the quadrature point $\mathbf{x}_i$, the closest point of the grid $\mathrm{g_i}$ is found: $f(\mathbf{x}_i)$ is equal to 1 if g_i belongs to the set of domain points D; otherwise $f(\mathbf{x}_i)$ is 0.
2. Given the quadrature point $\mathbf{x}_i$, the closest barycenter b_i of the surface triangles is found. The scalar product ps is calculated: $ps = (x_i - b_i)\cdot n_i$. A "Fermi type" step function is calculated: $f(x_i) = \frac{1}{1-\mathrm{Exp}[sf\cdot ps]}$ where sf has been optimized to a value of 50 atomic units as a compromise between a sharp step and the accuracy of the integration accuracy.

2.1.9 The Adopted Integration Methods

p_n and $\langle N \rangle[D]$ for MP Domains

The probabilities P are calculated on the basis of $S[D]$ evaluated by the method described in Sect. 2.1.8. The $f(\mathbf{x}_i)$ domain function is calculated by method 2 described above. $\langle N \rangle[D]$ is calculated using P. The volume V of MPD is calculated using the method $V_2[D]$, described in Sect. 2.1.8.

AIM end ELF Analysis

The probabilities P are evaluated on the base of $S[D]$ matrix, calculated using the method reported in Sect. 2.1.8; in this case, the $f(\mathbf{x}_i)$ domain function is calculated using method 1 described in Sect. 2.1.8. $\langle N \rangle[D]$ is calculated using P. The volume V of AIM and ELF basins are computed using the method $V_1[D]$ as reported in section "Definition of Domains."

2.2 Computational Details

As starting point for our calculations, we have used the following PDB files for structures: Fe-protoporphyrin PDB Code: HEM, Co-protoporphyrin PDB Code: COH, Mg-protoporphyrin PDB Code: HEG, Ni-protoporphyrin PDB Code: HNI, Zn-protoporphyrin PDB Code: ZNH [45] and Cambridge Crystallographic Data Centre for Ca-protoporphyrin [46].

All the calculations have been performed within the Density Functional Hybrid Approximation, using the B3LYP parametrization [47]. A double zeta plus polarization Gaussian Type Orbital basis set has been adopted (6-31g(dp) [48]).

A 3D extension of MPDs analysis, implemented in a development version of CRYSTAL code [49], able to deal with molecules and crystalline systems in the Hartree–Fock and DFT approximation, has been adopted. The MPD code can study closed and open shell systems.

In MPDs analysis, our first objective has been to identify spatial subsystems $D \in R^3$ in the metal porphyrin compounds by specifying a given electron number n; for example, we considered +2 cationic domains specifying the number of electrons diminished by two (Fe ion with 24 electrons).

We searched two electrons domains in the metal–ligand regions of the macrocycle.

The next choice is the arbitrary definition of a volume D where we are going to find a given number of electrons n with a certain probability p; in fact, according to quantum mechanics, for a given state Ψ, we can find a certain number of electrons n in D with a certain probability. We computed the probability of finding exactly n

electrons in D for one Slater determinant wave functions and define the regions of space for which the probability to find n electrons becomes a maximum; to find out the local maxima for a certain probability with respect to the shape of D, the optimization of D for a given n (indicated as D_n) has been performed by making use of shape derivatives.

As in all distributions, information about the probability distribution $p_{\downarrow}(0)(D)$, $[(p)]_{\downarrow}(1)(D), [(p)]_{\downarrow}(2)(D), \ldots, p_{\downarrow}(N)(D)$ is obtained by means of some parameters, in particular the mean $\langle N \rangle$ has been computed for this system.

For an N electrons system, the average number of electrons in D is given by:

$$\langle N \rangle [D] = \sum_{n=0}^{N} n \, p_n(D). \tag{2}$$

We have obtained it by integrating the electron density $\rho(r)$ over the domain D as indicated in Eq. (3), representing the population of:

$$\int_D \rho(r) d^3(r) = \int_D \langle \psi | \hat{\rho}(r) | \psi \rangle \mathrm{d}^3(r) = \left\langle \psi \middle| \int_D \hat{\rho}(r) \mathrm{d}^3(r) \middle| \psi \right\rangle. \tag{3}$$

3 Results

3.1 *Bader's "Atoms in molecules" Analysis*

The reference analysis of the electronic structure of molecules and crystals, as coming from experiment or calculation, is due to Richard Bader [11], who showed how chemical concepts related to Atoms can be recognized from a pure physical observable, like the electronic density, its derivatives, and its critical points. In the present study, Bader's analysis has been applied to porphyrin and metal porphyrins, mainly for calculating atomic charges. Moreover, we have explored how the electronic density analysis describes the metal–ligand bond (Figs. 1, 2 and 3; Table 1).

3.2 *MPDs Analysis of Pure Covalent and Ionic Systems*

For illustrating the MPDs analyses to the metal–ligand bonding in metal-proto-porphyrins, we report some significant results of MPDs analysis of extreme ionic and covalent situations. We refer to crystalline magnesium oxide as a prototype of pure ionic situation and crystalline silicon, as a pure covalent situation. We select

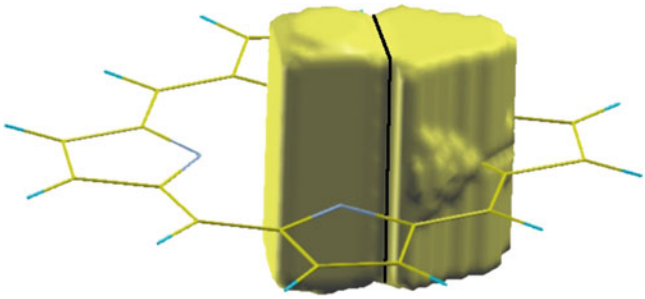


Fig. 1 Atom in molecules (AIM) basins, corresponding to the metal (*center*) and to the ligand (*right*), for the Mg porphyrin. In principle, perpendicular to the molecular plane, the AIM basins extend up to infinite. All the plots of molecules, basins and domains are obtained by XcrysDen computer program [54]

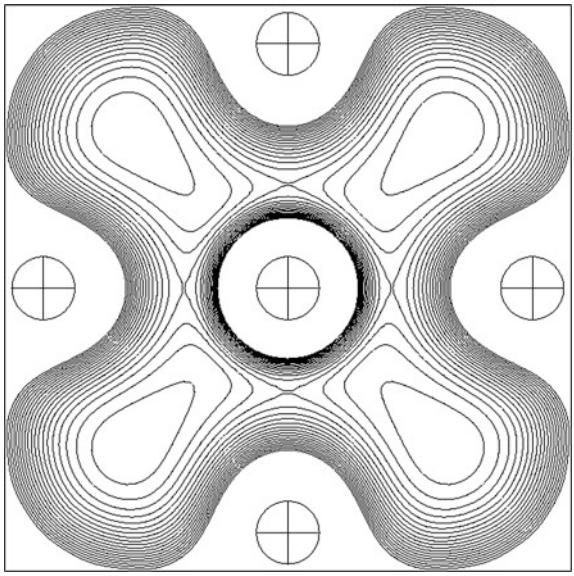


Fig. 2 Electronic charge density map for Mg porphyrin along the molecular plane. The atomic positions are marked by *crossed circles*. The metal is in the center of the map, while the four N ligands are the remaining *circles*. Neighbor iso-lines correspond to values that differ by 0.01 atomic units (electrons per cubic bohr). The lower value iso-lines (0.01 atomic units) lie in the diagonals of the map, out of the metal–ligand directions. The highest reported value iso-lines (0.20 atomic units) lie in the direction of the nuclei, both metal and N. Close to the middle of the metal–ligand segment, a saddle point lies

crystalline situation because the extremely ionic bond exists only in the solid state. Even for the perfect homo-nuclear covalent situation, the crystalline status is a perfect paradigm.

Figures 4, 5 and 6 show that ELF reaches very high values in valence regions; it means that the electronic structure is very localized both in perfect ionic and perfect covalent systems; in terms of ELF, the main difference between ionic and covalent

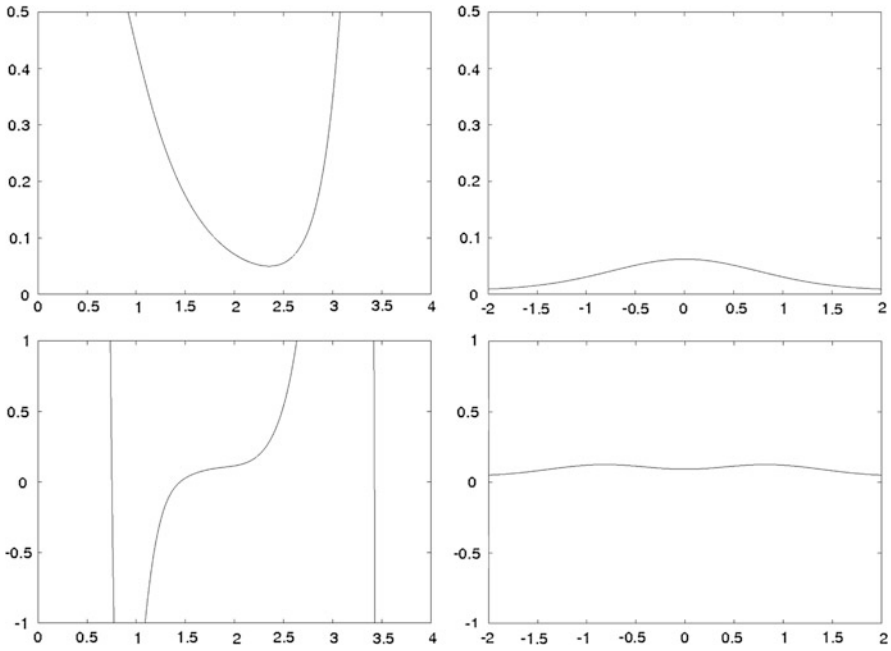


Fig. 3 Electronic charge density (*upper plots*) and Laplacian of the charge density profiles (*lower plots*). *Left plots* refer to the metal–ligand direction: cationic position is at the origin while *N* position is 3.88 bohrs, close to the end of the axis. *Right plots* refer to a direction perpendicular to the "bond" direction, lying on the molecular plane, passing through the minimum of the electronic density (*upper left*). The *upper pictures* show the (3,−1) critical point in the charge density: a saddle point in a scalar field has two negative and one positive hessian eigenvalues, so is defined (3,−1). The *bottom plots* show that the Laplacian of the charge density is positive in the critical point

Table 1 Atomic net charges for metal porphyrins (phy) following Bader's analysis

	Q(M)	V(M)	Q(L)	V(L)
Be phy	+1.32	81.2	−1.25	169.1
Mg phy	+1.45	136.9	−1.24	163.6
Ca phy	+1.47	205.7	−1.17	154.3
Fe phy	+1.39	158.9	−0.94	152.7
Co phy	+1.13	158.9	−0.89	151.6
Ni phy	+0.99	158.1	−1.02	153.7
Zn phy	+1.26	158.0	−0.87	155.6
Fe phy $(CO)_2$	+1.43	46.5	−0.88	92.0
$[Fe(CO)_6]^{2+}$	+1.03	97.5	+1.63	120.7

Basin charges and volumes are reported for the metal (M) and the nitrogen ligand (L): L is Nitrogen in all cases but in the last two lines, where the atomic charge of Carbon in CO is reported. The absolute values of the volumes has no meaning for the square planar coordinated molecules (the first seven systems), because the basins will extend to infinite perpendicularly to the molecular plane. In those cases, the reported volumes refer to basins truncated to a distance of 6 bohrs from the molecular plane

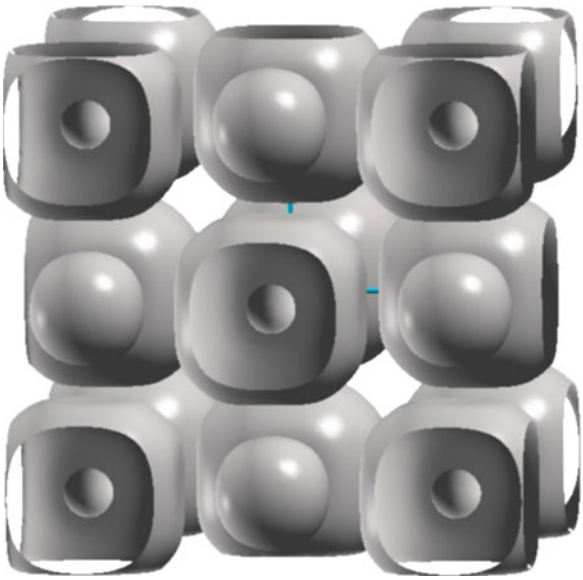


Fig. 4 Electron localization function (ELF) plot for crystalline magnesium oxide. The iso-surface corresponding to ELF = 0.3 is reported. A magnesium ion is at the center of the cube and it appears like a medium-sized spheroid. Oxygen ions appear like large quasi cubes. The *small spheres* represent the oxygen cores

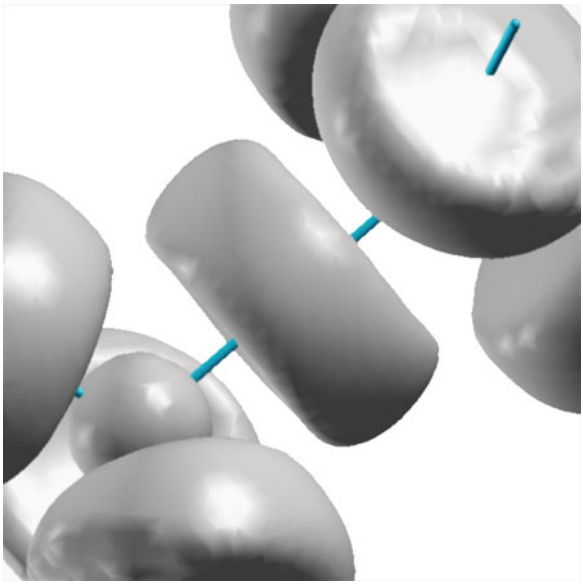


Fig. 5 ELF iso-surface, corresponding to the value of 0.66, for crystalline silicon. The *little spheres* underline the silicon cores; the large cylinders, coaxial with the Si–Si direction, underline the Si–Si covalent bond

lies out of the bond regions: in ionics, ELF drops to very low values, below 0.05, while in covalents, it remains higher, amounting at about 0.1 in bond direction and it is even higher in "out of bond" directions. As a consequence, the ELF basins remain well separated even at low level, like the 0.3 iso surface shown in Fig. 4. In covalent situations, the chemical basins which correspond to cores and covalent bonds easily melt together: the covalent picture is clearly restituted by 0.6 iso-surface reported for silicon in Fig. 5.

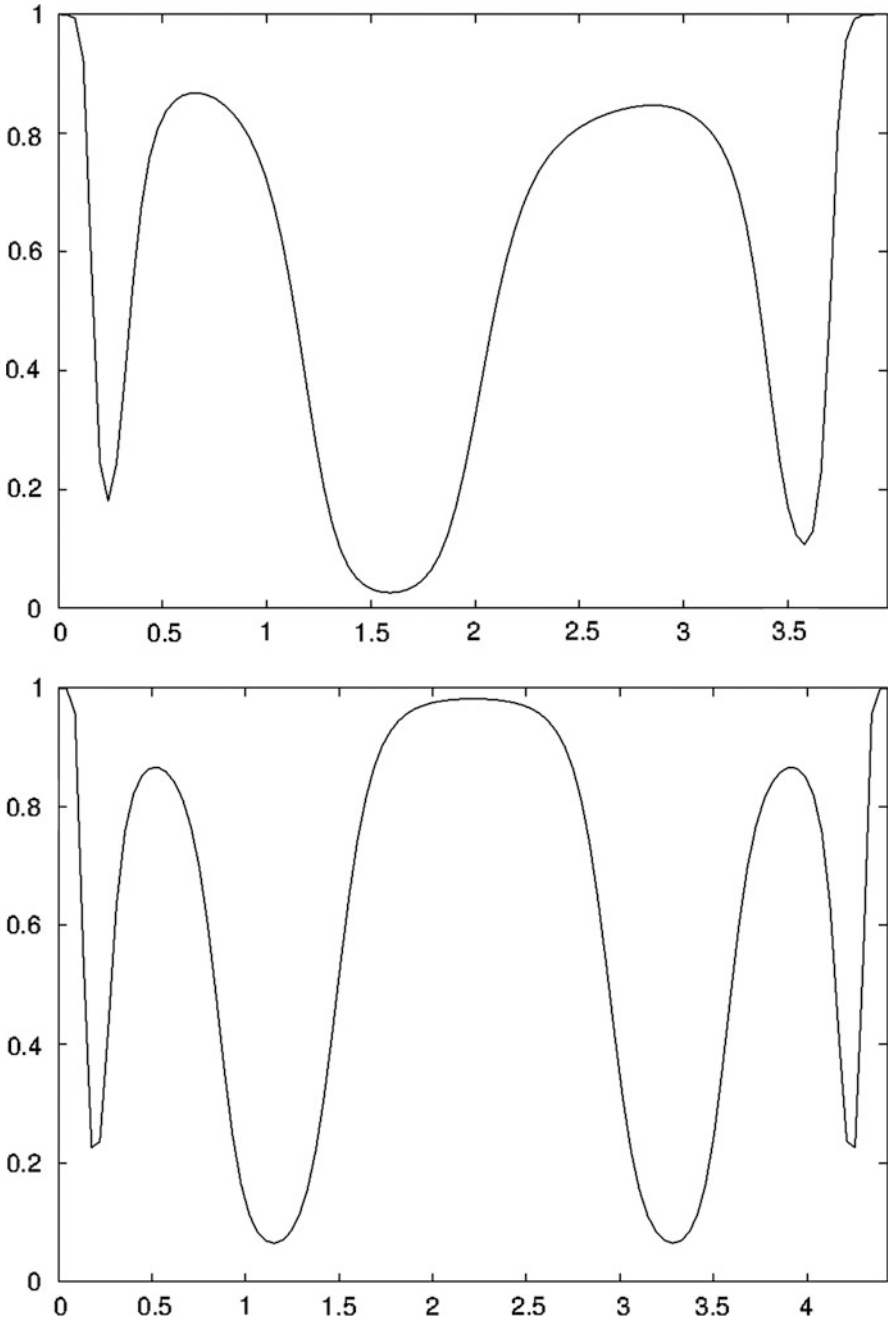


Fig. 6 ELF profile between the closest neighbors of crystalline magnesium oxide (*top*) and silicon (*bottom*). In both cases, the atomic core-valence shell structure is evident. In the MgO, *top plot* Mg position is on the *extreme left* whereas Oxygen is on the *extreme right*. For Si two symmetry equivalent Si nuclei are located at the extremes of *horizontal axis*. ELF reaches maximum values that can be considered as maximum pairing of electrons in the 1 s cores and in the middle of Si–Si segment that we commonly attribute to a Lewis pair covalent bond

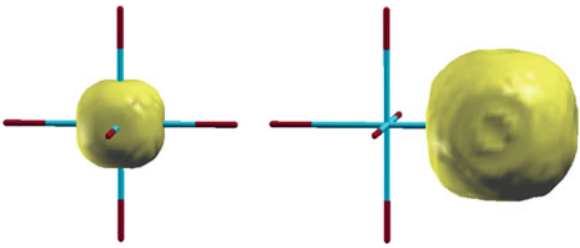


Fig. 7 Cationic (*left*) and anionic (*right*) maximum probability domains for crystalline MgO. Both the domains are optimized by maximizing the probability of containing ten electrons

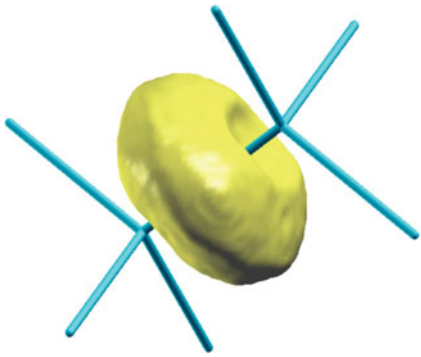


Fig. 8 Two-electron MPD for Crystalline silicon, representing the Si–Si covalent bond

The chemically important Maximum Probability Domains for the same systems are shown in Figs. 7 and 8. For magnesium oxide, the cation and anion 10 electrons MPD have high probability: 0.85 and 0.57 for cation and anion, respectively; the volumes of two ions are very different, 19.4 and 90.5 cubic bohrs; the MPD average net charges are, respectively, +1.85 and −1.85 and these are very close to the formal ones. The two-electron MPD reported in Fig. 8 has a probability equal to 0.42, a volume of 56.0 cubic bohrs, and an average population of 2.04 electrons.

In all our previous MPD studies, the probability lies in the range (0.30–0.45) for covalent bonds, whereas the probability for ionic MPD is higher in fact it amounts to (0.70–0.90) for cations and to (0.5–0.7) for anions.

3.3 ELF and MPDs Analysis of Metal Porphyrins

In the present study, we consider the bond analysis in metal porphyrins. In the following discussion, the porphyrin will be often indicated as "phy." We apply Electron Localization Function (ELF) analysis and Maximum Probability Domains (MPDs) analysis for understanding the eventual differences between transition metal and non-transition metals inserted into the porphyrin ring. As an example of non-transition metal, we consider Be, Mg, and Zn (a d^{10} metal, so d orbitals can be considered core-like). For transition metals, we consider Fe, Co, and Ni. Due to the difficulties in finding experimental examples of pure square planar coordination

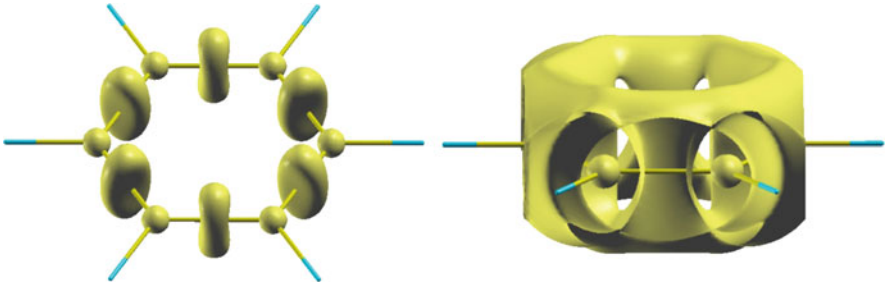


Fig. 9 Electronic localization function (ELF) for benzene. *Left*: ELF = 0.85 iso-surface. *Right*: ELF = 0.66 iso-surfaces. The right iso-surface is cut on the front side for showing the complex topology: the Carbon core basins sharply separate from the scalar ELF field. The aromatic ELF basin is clearly defined. Both in the high or lower ELF value iso-surface, no separation between Sigma and Pi basins can be found, which is a characteristic of delocalized bonds, similar to that shown in Fig. 11. All the molecular plots reporting ELF iso-surfaces or MPD are performed using the Xcrysden program [44]

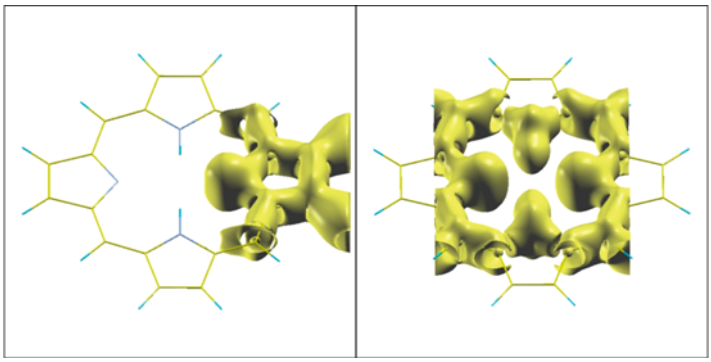


Fig. 10 ELF for porphyrin. In both parts, portions of ELF = 0.66 iso-surface are shown to highlight different features. The ELF shows the same delocalization features as benzene in Fig. 9. The nitrogen lone pairs show a slightly larger localization basin than NH bonds. NH induces a small interruption in the delocalized ring between N and neighbor C(H); this small interruption disappears in deprotonated porphyrin and in metal–porphyrin

for Fe, because the d^6 electronic structure for Fe(II) does not favor this coordination geometry, we also consider the penta-coordinate Fe porphyrin CO and the quasi octahedral Feporphyrin $(CO)_2$ examples. As a classical reference, we also study with the same methods and approximations $[Fe(CO)_6]^{2+}$, as a prototype for the CO–metal bond.

In Figs. 9 and 10, we show the bond delocalization in benzene and porphyrin. In Fig. 9, left, the iso-surface corresponding to ELF equal to 0.85 is presented. At high values, the maxima of ELF relative to atom cores and bonds are shown. Lowering the ELF value to 0.66, the iso-surfaces melt together in a classical hexagonal domain, as shown in Fig. 9, right. Due to the relatively high value of ELF that

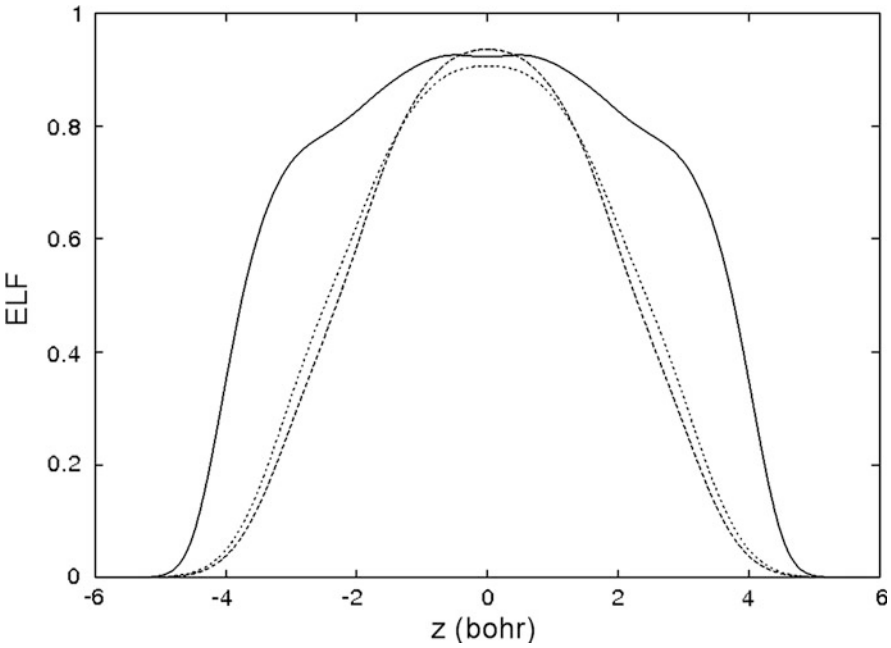


Fig. 11 Vertical ELF profiles for ethylene, benzene, and porhyrin. The z coordinate is perpendicular to the molecular plane. The $z = 0$ value corresponds to the middle of the C–C bond. The *continuous line* refers to ethylene, the *long-dotted one* to benzene and the *short dotted one* refers to porphyrin (the external C–C bond of the ring has been selected)

corresponds to this "aromatic" topology, we interpret as the existence of a wide spatial region where the electrons localize (are paired), without the possibility of distinguishing them in separated pairs. The valence electrons n in the porphyrin ring shows an analogous behavior as the one shown in Fig. 10, where the iso-surface corresponding to ELF equal to 0.66 is shown.

From the comparison with ELF of benzene, we can assume a perfect aromaticity of the porphyrin ring. Further, in Fig. 11 we analyze vertical profiles of ELF.

The ethylene ELF profile is on average higher than the others because the bond is localized; the π bond local maxima are clearly distinguishable. Out of the molecular plane, in the benzene and porphyrin case, ELF is rapidly decaying and only a central maximum is present: we interpret this behavior of benzene and porphyrin as an indication of electron delocalization that corresponds to an objective difficulty in distinguishing between σ and π bonds.

It is noteworthy that Coulson et al. [50], calculating the electron density in benzene in the plane 0.35 above the molecular plane, found that there is only a very small region directly above each of the six carbon nuclei where the charge density of the π electrons is as great as that of the σ electrons, meaning that the π electron charge cloud does overlap considerably with that of σ electrons. Thus "this calculation threw a considerable doubt on the validity of the σ–π separation" [51],

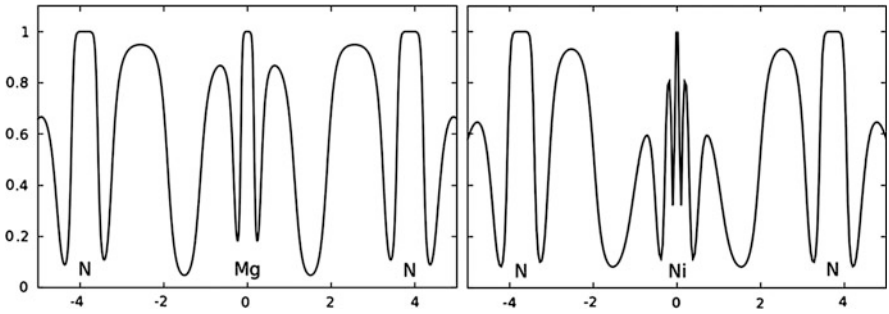


Fig. 12 ELF profiles along the N–M–N segments. *Left plot* refers to Mg phy (*left*); *right plot* refers to Ni phy. *From left to right* of each plot, the following ELF localization domains are distinguishable: N core, N lone pair, Metal outer and inner core, N lone pair and N core. The ELF analysis provides the same picture for the non-transition and the transition metal compound

becoming also more doubtful for larger molecules when σ and π configuration interaction would become more important with the increasing number of atoms.

The ELF analysis of metal porphyrins as reported in Figs. 12, 13, and 14 shows no relevant difference between non-transition and transition metals. There are low values in ELF in the space region between metal and ligands, both nitrogen, or CO. The ELF profiles reported in Fig. 12 show very clearly the localization of the electronic structure in the ligand–metal segment: The higher values of ELF are in the N and Metal 1s cores. Then are clearly distinguishable the 2sp outer core of Mg^{2+} ion and the 2sp and 3spd outer cores of Ni^{2+}. The nitrogen "lone pairs" also show high value of localization. The low value of ELF in d-electron systems has been already remarked upon [29]. At moderately high ELF level (0.66), the porphyrin ring and the central metal regions remain well separated, as shown in Fig. 4. In addition, at the low 0.3 ELF level, the metal and ligand domains are still separated, as shown in Figs. 12 and 13. The ELF level equal to 0.3 is usually adopted for analyzing very ionic systems, like alkaline halides, where the cationic and anionic ELF domains stay distinct also at 0.05 ELF level. In the metal–porphyrin complex, metal and ligand ELF basin remain separated for ELF values larger than 0.12. The same behavior can be noted in $[FeCO_6]^{2+}$, where the metal and ligand (C of CO molecule) ELF basins melt at 0.17 level. From the ELF analysis, as shown in Figs. 12 and 13, there is no difference between non-transition and transition metal porphyrins.

In Fig. 15 two Maximum Probability Domains for metal porphyrins are shown: the central cation MPD and the two-electron domain in the metal–ligand space region. These MPDs are well defined and do not overlap. Following previous applications of MPD analyses [33–36], the non-overlap of MPD is a clue of a good chemical hypothesis. Figure 15 shows that there are no significant topological differences between non-transition and transition metals. This confirms the ELF analysis. Figure 16 shows that in the case of a metal–ligand bond, there is a small difference between the ELF basin and the Maximum Probability Domain. On the other hand, the ELF basin is an essential starting point for optimizing the MPD.

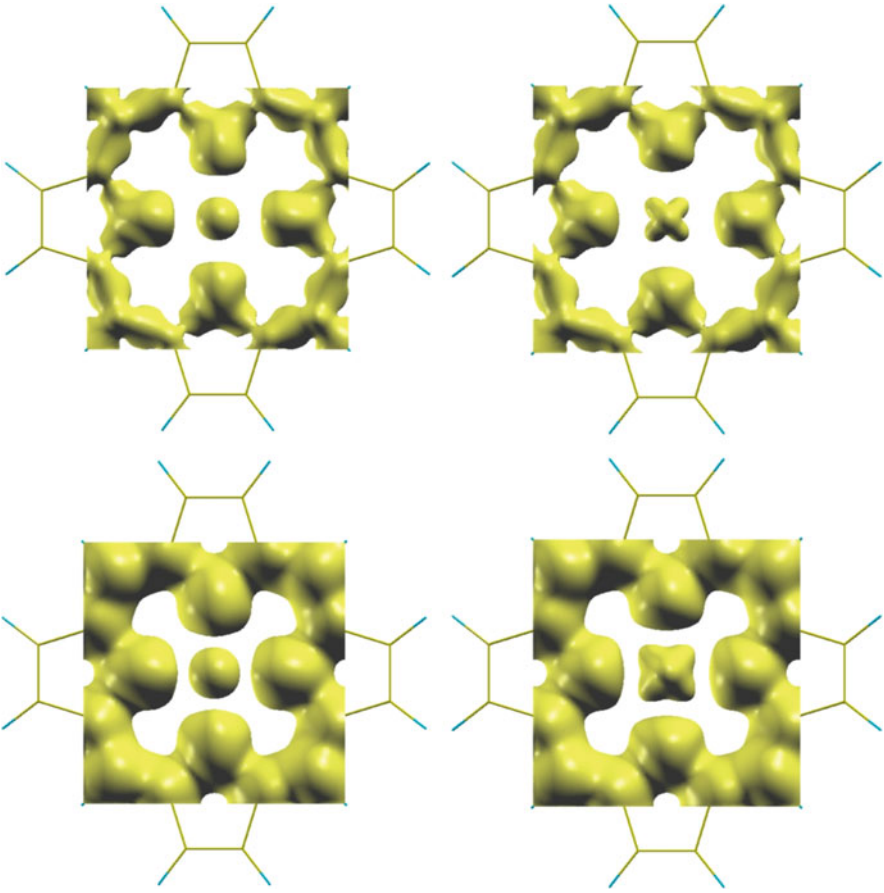


Fig. 13 ELF iso-surfaces: *upper row* corresponds to ELF = 0.66, *lower row* corresponds to ELF = 0.30; *left column* refer to Mg phy, *right column* refers to Ni phy

The numerical data relative to the MPD optimization are reported in Table 2. The trend of $p(M^{2+})$, the maximal probability of the +2 cationic domains, and the trend of their volumes, is chemically understandable. $p(M^{2+})$ monotonically decreases in the alkaline-earth series. Almost iso-electronic Ca phy and Zn phy have the same values. The first transition series Fe phy to Ni phy has almost the same cationic probability $p(M^{2+})$. The cationic probability $p(M^{2+})$ is the only difference between non-transition and transition metal porphyrins. The non-transition cations, including the $3d^{10}$ Zn^{2+} in Zn phy model, are obviously more defined physical systems. The average electronic population $\langle N \rangle$ is always close to the formal number of electrons.

The range of variation of the maximal two-electron probability of metal–ligand domain P(2,M–L) is very small, spanning the interval (0.36–0.41) for porphyrins: a little lower for metal–CO domains where P(2,M–L) is about 0.32. The average

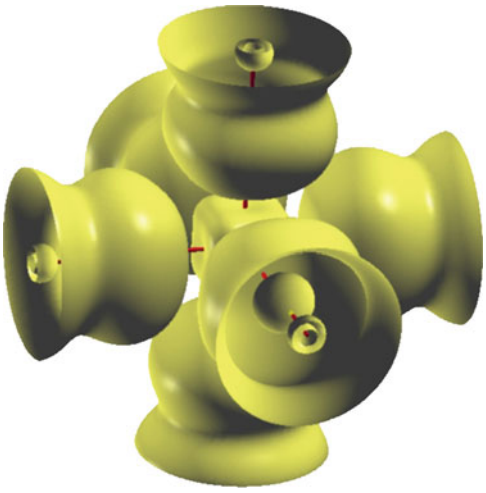


Fig. 14 Iso-surface for ELF = 0.30 of $[Fe(CO)_6]^{2+}$

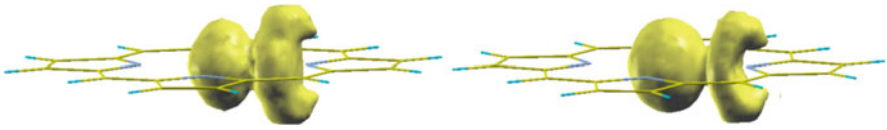


Fig. 15 Metal cation MPD and 2-electron metal–ligand MPDs for Mg phy (*left*) and Ni phy (*right*)

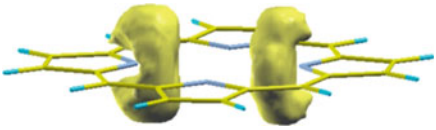


Fig. 16 Difference between ELF Basin (*left*) and MPD domain (*right*) for the metal–ligand bond in Mg-phy. The average number of electrons $\langle N \rangle$ is, respectively, 2.5 for ELF basin and 2.0 for MPD

number of electrons $\langle N \rangle$(M–L) is every time close to 2.0. The volumes are very similar, about 20 cubic bohr, but in the Mg phy model the volume is 40.5 cubic bohr. In Fig. 16, we show the ELF M–L basin and the MPD, obtained optimizing P(2,M–L) starting from the ELF basin; even if topologically equivalent, ELF basin and MPD have a very different dimension in fact: for Mg phy the volume goes from 23.5 to 40.5 cubic bohr upon optimization, whereas P(2,M–L) goes from 0.23 to 0.34.

Table 3 reports some energetic and structural data. The formation energy of the metal porphyrin models is defined as the energy of the reaction:

Table 2 Numerical results about maximum probability domains in metal–porphyrins

	$p(M^{2+})$	$V(M^{2+})$	$\langle N\rangle(M^{2+})$	P(2,M–L)	V(M–L)	$\langle N\rangle$(M–L)
Be phy	0.88	4.7	2.0	0.39	24.1	2.0
Mg phy	0.77	21.6	10.1	0.40	26.3	2.0
Ca phy	0.55	42.6	17.9	0.41	22.6	2.0
Fe phy	0.39	34.8	24.1	0.36	20.3	1.9
Co phy	0.36	34.4	25.2	0.36	19.5	1.9
Ni phy	0.35	32.3	26.2	0.36	19.4	1.9
Zn phy	0.50	33.6	28.1	0.37	20.7	1.9
Fe phy $(CO)_2$	0.29	31.3	24.0	0.36,0.38	19.7,25.6	1.9,1.9
$[Fe(CO)_6]^{2+}$	0.27	30.9	24.0	0.32	17.1	1.9

$p(M^{2+})$ is the maximal probability of a domain centered in the metallic position, relative to a number of electron equal to the atomic number minus two, that is exactly the situation of a +2 ion; V is the volume of the domain in atomic units (cubic bohr); $\langle N\rangle$ is the average number of electrons in the domain, that is the integral value of the electronic density; P(2,M–L) is the maximal probability having two electron in a domain located in the space between the M^{2+} ion and the ligand L core. In metal phorphyrins L is Nitrogen. In $[Fe(CO)_6]^{2+}$ L is Carbon of CO molecule. In Fe phy $(CO)_2$, the two values refer to N and C respectively. V(M–L) and $\langle N\rangle$(M–L) are volumes and average numbers of electrons of M–L domains

Table 3 Relevant geometric parameters and formation energy of metal–porphyrins

	Formation energy (Kcal/mol)	M–L distance (Angstrom)
Q phy	−843.0	
Be phy	−824.7	1.94
Mg phy	−704.1	2.05
Ca phy	−573.8	2.19
Fe phy	−703.9	2.00
Co phy	−805.8	1.98
Ni phy	−803.0	1.96
Zn phy	−726.7	2.05
Fe phy $(CO)_2$	−763.4	2.02, 1.86
$[Fe(CO)_6]^{2+}$	−259.3	1.93

The formation energy refers to the reaction: $M^{2+} + phy^{2-} \rightarrow M\,phy$. The M–L bond distance refers to N ligand in all cases, but for $[Fe(CO)_6]^{2+}$ and for the second value for Fe phy $(CO)_2$, where it refers to M–C. The Q phy compound has the equilibrium geometry of Be phy, but the central ion is substituted by a +2 point charge

$$M^{2+} + phy^{2-} \rightarrow M\,phy$$

where M^{2+} and phy^{2-} are, respectively, the cation and the de-protonated porhyrin. This model reaction is highly exoergonic. As a reference, we calculated the formation energy of the complex Q phy, having a +2 point charge in the cation position. We adopted for Q phy the equilibrium geometry of Be phy, the complex with the smallest cation. The formation energy of Q phy clearly depends on the geometry of the porphyn ring: we calculated it adopting the equilibrium geometry of all the considered complexes, and we obtained a strict correlation between the

formation energy of Q phy and M phy. The extremes of the set of Q phy formation energies are −843 and −731 kcal/mol for the Be phy and Ca phy geometry, respectively, that corresponds to −825 and −574 kcal/mol, that are the extremes of the set of M phy formation energies, for the same compounds. This feature underlines the role of electrostatic in determining M–phy bonds. The formation energy monotonically decreases in absolute value in the series Q–Be–Mg–Ca, as the M–L distance increases, and the dimension of the cation increases (volume of the M^{2+} MPD, reported in Table 2). This trend indicates a strong electrostatic character in the metal porphyrin bond for non-transition metal models. In the model containing first transition row metals, the formation energy is in the same range. Between models containing first transition row metals, the formation energy relative to Fe is lower than in Co phy and Ni phy, perhaps this could be attributed to the disadvantageous square planar coordination of the $3d^6$ Fe^{2+} cation. The quasi octahedral Fe phy $(CO)_2$ shows additivity between Fe–N bond energy and Fe CO bond energy, as calculated for $[Fe(CO)_6]^{2+}$. The metal–ligand equilibrium distances anti-correlate with formation energies.

4 Conclusions and Outlook

We have analyzed the metal–ligand bond in metal porphyrin models, using Electron Localization Function (ELF) and Maximum Probability Domains Analysis (MPDs). We do not find any relevant difference between non-transition and transition metal models. Our MPDs analysis shows that the metal–ligand bond is essentially ionic for all models, similar to the "electride" (charge separated structures) behavior of metal ions within cages [52]. In our MPDs analysis, we do not find clues of stabilization of bonds due to covalent interactions between the central metals and the porphyrin ring both in non-transition and in transition metal porphyrins.

On the basis of results collected hitherto, the Maximum Probability Domains Analysis has been shown as a valuable tool for performing further analysis of bonds in other coordination compounds, without biases coming from historical concepts.

Acknowledgments We, as all the researchers in the field of physical interpretation of chemical bond, are deeply indebted to Richard Bader for his pioneering work [11].

References

1. Kutzelnigg W (2002) Einführung in die Theoretische Chemie. Wiley-VCH, Weinheim
2. Marx D, Hutter J (2009) Ab initio molecular dynamics: basic theory and advanced methods. Cambridge University Press, Cambridge
3. Chattaraj PK (ed) (2009) Chemical reactivity theory: a density functional view. CRC, Taylor & Francis Group, London

4. Feinberg MJ, Ruedenberg K (1971) J Chem Phys 51:1495–1511
5. Kutzelnigg W (1973) Angew Chem Int Ed 12:546–562
6. Hirao K (1974) J Chem Phys 61:3247–3253
7. Boys SF (1960) Rev Mod Phys 32:296–299
8. Foster JM, Boys SF (1960) Canonical configuration interaction procedure. Rev Mod Phys 32:300–302
9. Pipek J, Mezey PG (1989) J Chem Phys 90:4916–4926
10. Del Re G (1974) In: Löwdin PO (ed) Advances in quantum chemistry, vol 8. Academic, New York, pp 95–136
11. Bader R (1994) Atoms in molecules. Oxford University Press, Oxford
12. Matta CF, Boyd RJ (eds) (2007) The quantum theory of atoms in molecules. Wiley, Weinheim
13. Parr RG, Yang W (1989) Density-functional theory of atoms and molecules. Oxford University Press, New York
14. Weinhold F, Landis CR (2005) Valency and bonding: a natural bond orbital donor-acceptor perspective. Cambridge University Press, Cambridge
15. Dunnington BD, Schmidt JR (2012) J Chem Theory Comput. doi:10.1021/ct300002t
16. Marx D, Savin A (1997) Angew Chem Int Ed 36:2077–2080
17. Kumar P, Marx D (2006) Phys Chem Chem Phys 8:573–586
18. Witt A, Ivanov SD, Forbert H, Marx D (2008) J Phys Chem A 112:12510–12517
19. Witt A, Ivanov SD, Mathias G, Marx D (2011) J Phys Chem Lett 2:1377–1381
20. Mathias G, Ivanov SD, Witt A, Baer MD, Marx D (2012) J Chem Theory Comput 8:224–234
21. Becke AD, Edgecombe KE (1990) J Chem Phys 92:5397–5403
22. Lennard-Jones JE (1952) J Chem Phys 20:1024–1030
23. Savin A, Becke AD, Flad J, Nesper R, Preuß H, von Schnering HG (1991) Angew Chem Int Ed 30:409–412 (in English)
24. Savin A, Jepsen O, Flad J, Andersen OK, Preuß H, von Schnering HG (1992) Angew Chem Int Ed 31:187–188 (in English)
25. Silvi B, Savin A (1994) Nature 371:683–686
26. Fuster F, Silvi B (2000) Theor Chem Acc 104:13–21
27. Santos JC, Tiznado W, Contreras R, Fuentealba P (2004) J Chem Phys 120:1670–1673
28. Santos JC, Andres J, Aizman A, Fuentealba PJ (2005) Chem Theory Comput 1:83–86
29. Schmider HL, Becke AD (2000) THEOCHEM 527:51–61
30. Kohout M (2004) Int J Quantum Chem 97:651–658
31. Kohout M, Wagner FR, Grin Y (2006) Int J Quantum Chem 106:1499–1507
32. Wagner FR, Bezugly V, Kohout M, Grin Y (2007) Chem Eur J 13:5724–5741
33. Putz MV (2005) Int J Quantum Chem 105:1–11
34. Putz MV (2009) Int J Mol Sci 10:4816–4940. doi:10.3390/ijms10114816
35. Francisco E, Martín Pendás A, Blanco MA (2008) Comput Phys Commun 178:621–634
36. Steinmann SN, Mo Y, Corminboeuf C (2011) Phys Chem Chem Phys 13:20584–20592
37. Savin A (2005) J Mol Struct Theochem 727:127–131
38. Savin A (2002) Reviews of modern quantum chemistry: celebration of the Contributions of Robert G. Parr. World Scientific, Singapore, p 43
39. Causà M, Savin A (2011) J Phys Chem A 115:13139–13148
40. Lopes Jr OM, Braïda B, Causà M, Savin A (2011) In: Hoggan Ph et al (eds) Advances in the theory of atomic and molecular systems. Prog Theor Chem Phys 22:173. doi 10.1007/978-94-007-3_10
41. Causà M, Savin A (2011) Z Anorg Allg Chem 637:882–884
42. Bossa M, Cervone E, Garzillo C, Del Re G (1995) J Mol Struct Theochem 342:73–86
43. Bossa M, Cervone E, Garzillo C, Peluso A (1997) J Mol Struct Theochem 390:101–107
44. Laisk A, Nedbal L, Govindjee (eds) (2009) Advances in photosynthesis and respiration: photosynthesis in silico, understanding complexity from molecules to ecosystems. Springer, Dordrecht
45. http://www.rcsb.org/pdb/results/results.do?outformat=&qrid=DC002FD5&tabtoshow=Ligand

46. Bonomo L, Lehaire ML, Solari E, Scopelliti R, Floriani C (2001) Angew Chem Int Ed 40:771–774, In English
47. Becke A (1993) J Chem Phys 98:5648–5652
48. Dill JD, Pople JA (1975) J Chem Phys 62:2691–2702
49. Dovesi R, Saunders VR, Orlando R, Zicovich-Wilson CM, Pascale F, Civalleri B, Doll K, Bush IJ, D'Arco P, Lunell M (2009) Crystal 2009 user manual. Turin University, Turin
50. Coulson CA, March NH, Altmann SL (1952) Proc Natl Acad Sci USA 38:372–378
51. Ohno K (1967) Molecular orbital calculation of P electron systems. Adv Quant Chem 3:239–322
52. Constable EC (1991) Nature 353:600–601
53. Cances E, Keriven R, Lodier F, Savin A (2004) Theor Chem Acc 111:373–380
54. Kokalj A (2003) Comput Mater Sci 28:55–63

Struct Bond (2013) 150: 143–180
DOI: 10.1007/978-3-642-32750-6_5

Biological Activity and Toxicity: A Conceptual DFT Approach

Arindam Chakraborty, Sudip Pan, and Pratim K. Chattaraj

Abstract Quantitative structure–activity relationship (QSAR) models are generated for biological activity and toxicity in terms of global and local reactivity descriptors within a conceptual density functional theory framework. Possible anticancer activity of two new metal–borane clusters is analyzed.

Keywords Biological activity • Conceptual DFT • Structure–activity relationships • Toxicity

Contents

A. Chakraborty • S. Pan • P.K. Chattaraj (✉)
Department of Chemistry and Center for Theoretical Studies, Indian Institute of Technology, Kharagpur 721302, India
e-mail: pkc@chem.iitkgp.ernet.in

List of Abbreviations

3D-MEDNEs	Three-dimensional Markovian electron delocalization negentropies
3D-QSAR	Three-dimensional quantitative structure activity relationship
ADME	Allied toxicological physicochemical and absorption Disposition Metabolism and Excretion
AHH	Aryl hydrocarbon hydroxylase
Amax	Maximum acceptor superdelocalizability
ArHs	Aryl-hydrocarbon receptors
CDFT	Conceptual density functional theory
CoMFA	Comparative molecular field analysis
CPs	Chlorophenols
DFT	Density functional theory
DNA	Deoxyribonucleic acid
EPR	Electron paramagnetic resonance
ESIP	Element specific influence parameter
HCV	Hepatitis C Virus
HIV	Human Immunodeficiency Virus
HPA	Hirschfield population analysis
MD	Molecular dynamics
MEP	Minimum electrophilicity principle
MMP	Minimum magnetizability principle
MPA	Mulliken population analysis
MPP	Minimum polarizability principle
MTD	Minimum topological difference
NICS	Nucleus independent chemical shift
NMR	Nuclear magnetic resonance
NPA	Natural population analysis
NQR	Nuclear quadruple resonance
PCA	Principal component analysis
PCBs	Polychlorinated biphenyls
PCDFs	Polychlorinated dibenzofurans
PMH	Principle of maximum hardness
QC	Quantum chemistry
QM/MM	Quantum mechanical/molecular mechanics
QSAR	Quantitative structure–activity relationship
QSPR	Quantitative structure–property relationship
QSRR	Quantitative structure–retention relationships
QSTR	Quantitative structure–toxicity relationship
QTMS	Quantum topological molecular similarity
SAR	Structure–activity relationship
SCC-DFTB	Self-consistent charge density functional tight binding
SCF-MO	Self-consistent field molecular orbital
STR	Structure toxicity relationship
TQSI	Topological quantum similarity index
VSMP	Variable selection and modeling method based on the prediction

1 Introduction

Chemists, the experimentalists, or theoreticians although sharing a different world of knowledge are very much aware of the fact that our mother earth is plundered with millions of molecules that may be segregated into two broad classes, namely organic and inorganic. In order to study the chemical behavior of molecules in such large numbers, an effective correlation technique needs to be built up, which can further categorize them into several classes based on closely related structures or properties. Although both the theorists and experimentalists share the same motivation towards classifying molecules into various categories, leading to a better understanding of reactivity patterns, yet they share a somewhat different platform. While the experimental chemists carry out reactions on a large number of molecules to set up a correlation, the theoreticians at the same time use mathematical algorithms to obtain powerful regression models, which usually help to develop effective structure–activity or structure–property relationships. Nevertheless, the development of fruitful activity or property relationships related to changes in molecular structure is what we need in order to predict the reactivity of an unknown compound. Moreover, hundreds and thousands of reactions are attempted and subsequently published, which eventually enrich the chemical database every day. Such increasing data bank needs to be managed skillfully to bring out a logistic correlation among analogous chemical moieties with a view of acquiring superior prediction towards molecular behavior. In this regard, the ideas developed by both experimental and theoretical chemists must be considered together to offer a multidisciplinary approach in dealing with and to envisage chemical reactivity. Again, the various organic and inorganic molecules upon consumption in any form (diet or medicine) are found to affect the general metabolism in human beings and other bio-organisms, which may even lead to toxicity upon an overdose. The mode of reactivity of the several compounds on the living organisms can be nicely predicted in a quantitative approach through suitable structure–activity relationship (SAR) or structure–toxicity relationship (STR) models that are built from chemical intuition. Once a meaningful rationale is established, the structure and property of an unknown molecule can be predicted from this standardized algorithm. Thus, the extent of the lethal effect of an unknown molecule towards a given living organism can be predicted with respect to the behavior of some structurally analogous molecules chosen as a reference, whose toxic action is already known. So quantitative structure–activity relationship (QSAR) and quantitative structure–property relationship (QSPR) techniques help to build powerful mathematical regression models that efficiently predict the reactivity of an unknown molecule in the backdrop of the behavior of an analogous set of chemically or structurally similar compounds. The possibility of an overdose and hence the development of a toxicity criterion of a particular molecule towards a biological entity can also be analyzed through suitable model equations that relate toxicity with molecular structure. Such mathematical equations are also called quantitative structure–toxicity relationship (QSTR) models and have been widely used to assess bio-toxicity. Consequently, the discovery of suitable antidotes, called drugs, to combat such

toxic effects in bio-organisms has also been made possible by utilizing the fruitful quantitative techniques. So, in addition to providing a reliable rationale towards predicting the chemical reactivity and toxicity of molecules, the QSAR, QSPR, and QSTR techniques also play a useful role in drug discovery and design.

In this article, we first of all outline the various theoretical tools and their role in deciphering the biological activity and toxicity patterns in several molecules. Section 2, in this regard, describes the development of QSAR and various allied quantitative relationship-based models and their applications in predicting the reactivity and toxicity of several classes of molecules. Section 2.1 delineates other theoretical techniques like quantum mechanical density functional theory (DFT), molecular dynamics (MD) simulations, nuclear quadruple resonance (NQR) spectroscopy, their role in building suitable structure–activity relationship (SAR) based models towards analyzing the biological activity of molecules and subsequent applications in medicinal chemistry and drug design. The penultimate section (Sect. 3) outlines a brief overview of conceptual density functional theory (CDFT) and the role of various CDFT-based global and local reactivity descriptors towards developing suitable regression models to correlate and quantify the biological activity and toxicity of molecules. Section 4 finally concludes our discussion with some queries and remarks.

2 Quantitative Structure–Activity/Structure–Property-Based Methods

The trends in the reactivity pattern for a given set of structurally related compounds can be mathematically predicted by building effective structure–activity oriented models. A known set of structurally analogous molecules is chosen and the variation in their reactivity and/or toxicity as a function of the changes in associated molecular framework is studied. The trends obtained as such are then hailed as a benchmark index and further extrapolated to another set of unknown but structurally similar compounds with an aim to correlate their reactivity patterns. Thus, the structure–activity relationship (SAR) based models become useful pathfinders towards predicting chemical reactivity and toxicity of molecules that usually give a nice correlation between theoretical predictions and experimental findings. When such SAR-based models are applied to derive a quantitative relationship between the activity and toxicity trends for a series of chemical compounds, we, hereby, obtain a quantitative structure–activity (or toxicity) relationship, QSAR (or QSTR). QSAR, unlike being conceived merely as an academic tool to predict the reactivity and toxicity of molecules through rigorous data analysis, should be better looked upon as a novel technique that helps to build meaningful models correlating structure, activity, and toxicity with ample judgment and reasoning. The success of such QSAR-based models towards an accurate prediction is sometimes a matter of fortune but, nevertheless a proper judicious rationale and the choice of appropriate

parameters always hold the key. QSAR-based studies, as mentioned above, have got widespread applications in many fields like eco-toxicology, drug discovery, antitumor treatment, molecular modeling, which are nonetheless our direct concern. Schultz et al. [1, 2] in a couple of articles have elaborated the role of QSAR in chemical toxicology. The success beyond any such quantitative prediction lies in the availability of accurate experimental data. An enriched databank renders a given structure–activity relationship (SAR) to be more defined towards an accurate prediction and also helps interpreting the plausible reaction pathway. Some relevant works [3–7] critically highlight the development of the QSAR methodology and its application in predicting bio-toxicity. Insights into the basics of QSAR regarding a nice predictive algorithm can also be obtained from the literature [8–10]. In fact, a hint of the effect of structural changes on reaction mechanisms has been earlier enunciated by Hammett [11, 12] in the well known linear free-energy relationship-based equations applicable to organic reactions involving substituted benzoic acid derivatives. Hammett's idea is also further extended by Taft to understand the varying substituent effects in aliphatic compounds [13]. Based on Hammett's idea, Hansch and his colleagues [14, 15] discussed the role of QSAR, especially the application of biological QSAR containing the terms σ and σ^{-}. Other studies establishing the idea of constructing strong QSAR models towards elucidating chemico-biological interactions [16–27] and drug design [28–37] are well known. Applications of QSAR in designing biomolecules, drug discovery through a quite accurate in silico toxic data analysis and constructing more effective models by introducing new, improved molecular descriptors are also reviewed [38–40]. The roles of QSAR in biochemistry and several allied toxicological, physicochemical, absorption, disposition, metabolism, and excretion (ADME) processes related to the discovery and subsequent development of various bioactive molecules distinguishing drug-like from nondrug-like molecules [41], drug resistance [42], toxicity prediction [43–48], prediction of various physico-chemical parameters [49], gastro-intestinal absorption [50], activity of peptides [51], data mining [52], drug metabolism [53], and prediction of other pharmacokinetic and ADME properties [54, 55] are quite evident in the relevant literature. Some pertinent articles [56–66] and book [67] provide additional insights into the applicability of QSAR in these fields. Now, in order to build an appropriate structure-reactivity model towards assessing toxicity, a strong chemical database is always needed as a backup. Such databases are usually built by experimentally studying several reaction mechanisms for a large number of toxic assays in vitro or in vivo. In this regard, with an aim towards acquiring a more informative database for an even better predictive correlation, various computational approaches are also proposed [68–71]. Some other studies [72, 73] emphasize on the reaction between an electron rich (nucleophile) and electron poor (electrophile) species. Such triggering reactions are considered as a benchmark that provides a detailed in chemico reactivity database towards toxicity prediction.

Parallel to the development of the QSAR and several models based on the given phenomenon, the advent of an analogous quantitative structure–property relationship (QSPR) rationale is also in the cards. Several research groups have aimed to strengthen QSPR on a theoretical basis and some novel contributions [74–83] in

this regard are worth-mentioning. Katritzky et al. [84] further describe some significant inroads that establish QSPR as a reliable tool in correlating chemical structure with various molecular (physical, chemical, biological, and technological) properties. Application of QSPR in the field of medicinal chemistry through the generation of whole molecular descriptors is also reported [85]. QSPR equations constructed on an empirical basis have been found to maintain a good regime between structure and property for even a large number of data sets. It is shown [56] that for a large number of molecular structures, data reduction methods, like principal component analysis (PCA) of a matrix formed by the assembly of related properties, provide insight into how these related properties correspond to each other quantitatively. Further applications of this method are also discussed. The development and applications of QSPR in pharmaceutical science engrossing a detailed study of the various pharmacokinetic aspects towards drug modeling is also provided in a two-part review by Grover et al. [86, 87].

In recent years, the development of three-dimensional correlation models for analyzing the structure–property relationships has also become quite popular. The three-dimensional quantitative structure–activity relationships (3D-QSAR) algorithm actually analyzes the quantitative relationship between the biological activity of a set of compounds and their three-dimensional properties using statistical correlation methods. 3D QSAR involves the application of force field calculations requiring three-dimensional structures and is concerned with the overall molecule rather than a single substituent. It examines the steric fields (shape of the molecule) and the electrostatic fields based on the applied energy function [88, 89]. Comparative molecular field analysis (CoMFA) [90], a well known three-dimensional quantitative structure–activity relationship (3D-QSAR) method, has been successfully invoked as a powerful tool towards drug design and other allied applications [91–98]. In the CoMFA approach, the molecules are structurally and pharmacologically oriented on the basis of the assumption that each compound acts via a common binding mode [90, 92, 95, 99, 100]. 3D-QSAR CoMFA in association with QSAR methods have been implemented to build useful structure–activity relationship models towards designing effective enzyme inhibitors [101–104], which are also sometimes guided by molecular docking [105–114]. The above phenomenon is also applied for studying the interactions between different organic polychlorinated derivatives with aryl-hydrocarbon receptors (ArHs) [115–118].

QSAR technique is also found to be effective in the simultaneous designing and study of inhibitors to combat the lethal growth of Human Immunodeficiency Virus (HIV) and Hepatitis C Virus (HCV) within living cells [119–121]. Quantitative structure-relationship-based modelings of different series of compounds acting as anti-HIV enzyme inhibitors are broadly studied [122–132]. QSAR, as an in silico methodology, has also become very relevant towards modeling antitumor enzyme inhibitors [133] and anticancer drugs [134, 135]. Tan et al. [136], in a recent article, have discussed about the use of coinage metals towards modeling antitumor drugs. In this regard, several other studies [137–149] have also reported that the metal-based compounds are indeed effective in tumor therapy.

Applications of other in silico tools like Three-Dimensional Markovian Electron Delocalization Negentropies (3D-MEDNEs) [150, 151] for toxicity prediction are also reported. QSAR technique is also applied to study the toxicity of phenols [152], alkaloids [153], aliphatic esters [154], and modeling anti-tubercular compounds [155–157].

QSAR studies have become extremely useful towards understanding the toxic levels of bio-organisms. In this context, various topological descriptors [158–160] are introduced to quantitatively correlate the hazardous effect of several chemical compounds on the ecosystem. The toxic effects of various compounds on fishes are reported in detail through fruitful SAR-based regression analyses [161–173].

2.1 Role of Different Theoretical Methods Towards Constructing Effective QSAR Models

Several sophisticated theoretical tools have turned out to be quite efficient towards strengthening the basis of a given structure–activity relationship, which, in turn, evokes a more approximate quantitative rationale. Quantum chemistry (QC) based methods are supposed to be quite significant in this aspect for applications in biological systems [174]. QC-based SAR models provide a much better understanding of the electronic correlation effects than empirical methods [175]. The various QC-based molecular descriptors serve as potential determinants to predict the biochemistry of various enzyme-mediated processes, hallucinogenic activity [176–179] as well as the reactivity of several organic molecules, derivation of partition coefficients, and correlation of other physical parameters [180–184]. The influence of quantum chemical methods and the role of the various allied QC-based descriptors in the structure–activity/toxicity studies are also discussed in detail elsewhere [185]. Applications of the SCF-MO [186] and various high level quantum chemical techniques towards modeling biomolecules and correlating their chemical reactivity with molecular structure are a vastly cultivated topic. Sophisticated ab initio methods like localized perturbation theory and density functional theory (DFT) are employed to understand the catalytic activity of enzymatic reactions, conformational energetics, and nonbonded interactions [187]. Biological activity measures of large molecules like proteins, DNA, etc. from first principles quantum chemical approaches sometimes pose problems due to size and complexity in their structures. Still ab initio DFT and DFT-based molecular dynamics (MD) simulation studies [188, 189] are found to be quite fruitful towards understanding the bond cleavage and bond formation processes of various biomolecular interactions. The roles of ab initio DFT and hybrid quantum mechanical/molecular mechanics (QM/MM) methods towards elucidating the biological modeling and reactivity of various complex motifs [190, 191] including the cytochrome P450 family of enzymes [192, 193] are also reported. DFT-based studies in the field of medicinal chemistry [194] as well as the prediction of toxicity measures for molecules acting as anticancer

agents or as drugs are also well documented [195, 196]. Quantum chemistry-based QSAR/QSPR models are also employed towards studying the bio-toxicity of *cis*-platin complexes as effective anticancer drugs [197, 198]. Sarmah and Deka [199] have further made a DFT-based 2D QSAR study on the anticancer activities of various nucleosides.

Superior QSAR-based regression models explaining the predictive nature of enzyme activity and the toxicity of various organic molecules like substituted aromatic compounds are studied vividly using the DFT protocol. Density functional theory as a powerful mathematical algorithm is exploited as a useful tool to provide sets of quantum chemical descriptors that help finding suitable correlations for enzymatic activities and toxicity of various classes of organic compounds and their derivatives [200–217]. Again Elstner and coworkers [218, 219] have elaborated that the self-consistent charge density functional tight binding (SCC-DFTB) formalism, compared to DFT may appear as a better mathematical rationale for large biological systems, as the former method considers the phenomena of nonbonded interactions and charge transfer processes in a better way.

In another approach, with an aim to offer a realistic motive towards handling millions of databases and hundreds of descriptors for a fruitful structure–activity relationship (SAR), Putz and coworkers have proposed a unique QSAR model called spectral-SAR (S-SAR) [220], which considers the spectral norm in quantifying toxicity and reactivity with molecular structure. A handful of applications of the S-SAR algorithm in dealing with ecotoxicity, enzyme activity, and anticancer bioactivity are well established [221–227]. The S-SAR model coupled with Element Specific Influence Parameter (ESIP) formulations [228] are also utilized for predicting ecotoxicity measures. QSAR studies on the anti-HIV-1 activity of HEPT (1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thymine) [229] and further studies involving the minimum topological difference (MTD) method [230, 231] are also reported [232].

A number of other novel methodologies providing good correlations and meaningful structure–activity-based regression models are also available. Latosińska [233, 234] delineates the application of nuclear quadrupole resonance (NQR) spectroscopy in studying drug molecules. Applications of other spectroscopic methods like NMR, EPR, UV, and IR towards investigating the biological activity of molecules are presented earlier [235–237]. The variable selection and modeling method based on the prediction (VSMP) [238], quantitative structure-(chromatographic) retention relationships (QSRR) [239], topological quantum similarity index (TQSI) [240, 241] models, and quantum topological molecular similarity (QTMS) descriptors [242, 243] also have ample contributions towards ensuring refinements in the predictive nature of quantitative structure–activity relationships. Modeling of soft electrophilicity and hydrophobicity [68, 244–246] as descriptors for toxicology-based QSAR predictions has also become a quite effective analytical algorithm.

3 Conceptual Density Functional Theory

Chemistry is all about the formation of new molecules from some parent motifs. All chemical processes follow the basic philosophy of rupture of old inter-atomic linkages (bonds) followed by the formation of newer ones upon electronic interactions between two reacting species. Chemical reactions can be experimentally studied as well as theoretically monitored. The advent of quantum chemistry and its subsequent developments with time have boasted off some exceptionally sophisticated mathematical paradigms, which are proven to be very unique in studying several classes of chemical reactions along with accurately screening the corresponding reaction pathways. Applications of a few such theoretical techniques towards elucidating the toxicity trends of various molecules and building of fruitful SAR-based regression models as a useful benchmark for further prediction are already mentioned in the preceding section. In this section, we discuss a very popular theoretical model, conceptual density functional theory (CDFT) and its utility towards setting up more refined mathematical relationships to quantitatively correlate biological activity and toxicity with molecular structure.

Conceptual density functional theory (CDFT) [247–251], after its inception a few decades ago, has proved itself to be a very powerful theoretical paradigm to understand molecular reactivity. Conceptual DFT basically owes its origin from the research group of Prof. Robert G. Parr, which is later made more coherent and is developed with an aim to offer a logistic idea about electronic interactions. In this effort, several mathematical response functions, also known as reactivity descriptors are introduced, which nicely explain the mode of chemical reactivity between a given set of interacting molecules. Conceptual DFT in association with its various global reactivity descriptors like electronegativity [252, 253] (χ), hardness [254, 255] (η), and electrophilicity [256–259] (ω) originating from the same popular qualitative concepts, respectively, along with the corresponding local variants like atomic charges [260] (Q_k), Fukui functions [261] $f(\mathbf{r})$, and their condensed-to-atom variants [262] (f_k) develops intuitive reactivity trends for a molecule as a whole and for each and every active site of the same as well.

The term electronegativity (χ) was initially coined by Pauling [263, 264], which virtually meant the "power" of an atom in a molecule to attract bonded electrons towards itself. Mulliken [265, 266] further defined the term electronegativity (χ) as a function of the two main experimentally determined parameters of a system, viz., the first ionization energy (IP) and the electron affinity (EA) and showed that χ may be expressed as an arithmetic mean of the respective IP and EA values. Thus, according to Mulliken:

$$\chi = \frac{\mathrm{IP} + \mathrm{EA}}{2}. \tag{1}$$

The chemical potential (μ) of an atom/molecule may be quantitatively assessed by the theory of statistical ensembles. Thus, if an atom or a molecule be considered

as a member of a grand canonical ensemble where the energy (E) and the number of electrons (N) are continuous functions and vary independently, the chemical potential of the ensemble may be expressed as

$$\mu = \frac{\partial \mathrm{E}}{\partial \mathrm{N}}, \text{at constant entropy.} \tag{2}$$

Gyftopoulos and Hatsopoulos [267] proposed a quantum thermodynamic definition of electronegativity (χ) of a system where the electronic chemical potential (μ) is being considered as the negative of the same.

Based on Mulliken's [265, 266] prescription for electronegativity, the Mulliken chemical potential can be similarly expressed as a finite-difference approximation of the derivative of the electronic energy with respect to the number of electrons and can be expressed as

$$\mu_{\text{Mulliken}} = -\chi_{\text{Mulliken}} = -\frac{\text{IP} + \text{EA}}{2} = \left[\frac{\partial E(N)}{\partial N}\right]_{N=N_0}. \tag{3}$$

The concept of "hardness" or "softness" in chemistry was introduced by Ralph G. Pearson [268–272] in connection with the study of generalized Lewis acid–base reactions

$$\mathrm{A}+ : \mathrm{B} \rightarrow \mathrm{A} : \mathrm{B}$$

where A, the electron acceptor is the acid while the electron-pair donor B is the base. Studies on the above donor–acceptor interaction (between B and A) provided a splendid qualitative rationale to conceive global hardness (or softness). Global hardness (η) is simply attributed to the degree of compactness of the electron cloud encompassing the nucleus/nuclei of an atomic/molecular system. Global softness (S) on the other hand being the reciprocal of global hardness describes the extent to which the electronic environment surrounding the nucleus/nuclei of an atomic/molecular species tends to loosen itself. The qualitative concept of molecular hardness was put forward quantitatively in the form of a mathematical descriptor by Parr and Pearson [254, 255], who defined that the chemical hardness (η) for a molecular system can be identified as the first derivative of the chemical potential (μ) or the second derivative of the energy (E) as a function of the number of electrons, N, at a fixed external potential $v(\mathbf{r})$:

$$\eta = \left(\frac{\partial \mu}{\partial N}\right)_{v(\mathbf{r})} = \left(\frac{\partial^2 E}{\partial N^2}\right)_{v(\mathbf{r})}. \tag{4}$$

The convexity in the E vs. N curve renders the value of η to remain positive and from the finite-difference approach the corresponding curvature equals IP–EA, which actually signifies hardness. Therefore:

$$\eta = \left(\frac{\partial \mu}{\partial \mathrm{N}}\right)_{\mathrm{v}(\mathbf{r})} = \left(\frac{\partial^2 \mathrm{E}}{\partial \mathrm{N}^2}\right)_{\mathrm{v}(\mathbf{r})} \approx \mathrm{IP} - \mathrm{EA}. \tag{5}$$

Thus, for a normal charge-transfer interaction between a donor–acceptor couple like a Lewis acid–base pair, the chemical hardness (η) of a given species signifies the measure of its reluctance towards further electron shift to the other species, thereby upholding its qualitative concept of compactness.

Global softness [273] (S), as a reciprocal of the global hardness (η) for a molecular system, is expressed as

$$S = \frac{1}{\eta} = \left(\frac{\partial N}{\partial \mu}\right)_{v(\mathbf{r})}. \tag{6}$$

Among the various reactivity descriptors, electrophilicity [256–259] (ω) and its recently reported mathematical refinement, net electrophilicity ($\Delta\omega^{\pm}$) [274] deserve a special mention from the viewpoint of constructing useful QSAR models.

The term "electrophile" in chemistry is coined to those species, which have got a "special" kind of affinity for electrons. Electrophiles are in general electron-deficient systems, sometimes bearing positive charge(s) or may be a free radical that has got a propensity towards attracting electron-rich motifs called nucleophiles. Electrophiles and nucleophiles together play a vital role in determining the mechanistic courses for several types of organic and inorganic chemical reactions like acid–base, redox, addition, substitution, elimination, molecular rearrangement, etc. An electrophile upon interaction with an electron rich species (nucleophile) will strongly attract the electron density of the latter and get stabilized with the gradual lowering of energy and formation of a stable covalent bond. A clear quantitative estimate of this simple qualitative idea as to which extent an electrophilic system would be energetically stabilized upon transfer of electrons from the adjacent nucleophile was, however, lacking. In this context, Maynard et al. [275], upon experimenting with the human immunodeficiency virus type 1 (HIV-1) nucleocapsid protein p7 (NCp7) with a variety of electrophilic agents, showed that the fluorescence decay rates vary almost linearly with the ratio of the square of electronegativity (χ) to hardness (η), χ^2/η, which can further be designated as the capacity of an electrophile to attract electrons from a nucleophile to create a covalent bond. Such an innovative idea developed by Maynard et al. [275] inspired Parr and coworkers [256] to put forward a more precise mathematical rationale for correlating the electron-attracting power of a species. Subsequently, a new descriptor called electrophilicity index (ω) is introduced. Parr's postulate of electrophilicity index (ω) unlike Maynard et al. (which was developed from a kinetic consideration by studying reaction rates) was based on a thermodynamic background and thus ω became a measure of the favorable change in energy upon saturation of a system with electrons. The electrophilicity index (ω) as prescribed by Parr et al. [256] is

$$\omega = \frac{\mu^2}{2\eta} = \frac{\chi^2}{2\eta}. \tag{7}$$

The above expression of electrophilicity (ω) is comparable with the equation of electrical power (W) in classical physics where $W = V^2/R$, where V and R attribute to the voltage and electrical resistance, respectively. Therefore, ω represents the "electrophilic power" of a species. Comprehensive reviews on electrophilicity index (ω) [256–259] regarding its genesis and rigorous applications towards an understanding of chemical reactivity are available.

In an attempt to further correlate the associated energy changes between the interacting acceptors and donors in a charge-transfer process, Gazquez et al. [276], based on the second-order Taylor series energy expansion formula as a function of the number of electrons (N) in the intervals between $N - 1$ and N, and N and $N + 1$, showed that the electrodonating (ω^-) and the electroaccepting (ω^+) powers may be defined as

$$\omega^- = \frac{(\mu^-)^2}{2\eta^-}; \quad \omega^+ = \frac{(\mu^+)^2}{2\eta^+}, \tag{8}$$

where μ^- and μ^+ denote the chemical potential for electron donation and electron acceptance, respectively. η^- and η^+ signify the hardness for electron donation and electron acceptance, respectively. Further evidences [276] reveal that μ^- and μ^+ can be equalized to μ so that: $\mu^- = \mu^+ = \mu$ and, in the same spirit, $\eta^- = \eta^+ = \eta$. Thus, the electrodonating (ω^-) and the electroaccepting (ω^+) powers become equivalent with the concept of electrophilicity (ω) owing to which ω, ω^-, or ω^+ may be expressed in terms of chemical potential (μ) and hardness (η) as $\omega^- = \omega^+ = \omega = \mu^2/2\eta$. Gazquez et al. [276] proposed two sets of definitions for ω^- or ω^+, which were based on two different approaches; one exploiting the original formula as above, which is expressed as

$$\omega^+ = \frac{\mathrm{EA}^2}{2(\mathrm{IP} - \mathrm{EA})}, \tag{9}$$

$$\omega^- = \frac{\mathrm{IP}^2}{2(\mathrm{IP} - \mathrm{EA})} \tag{10}$$

and the other by using an alternative expression for energy as

$$\omega^+ = \frac{(\mathrm{IP} + 3\mathrm{EA})^2}{16(\mathrm{IP} - \mathrm{EA})}, \tag{11}$$

$$\omega^- = \frac{(3\mathrm{IP} + \mathrm{EA})^2}{16(\mathrm{IP} - \mathrm{EA})}, \tag{12}$$

where *IP* and *EA* correspond to the first ionization energy and electron affinity of the system, respectively. It [276] was also shown that the alternative approach of expressing ω^- or ω^+ yielded better correlations than the other one.

The concept of net electrophilicity ($\Delta\omega^{\pm}$), proposed recently by Chattaraj et al. [274] for a system, is an attempt to assess the resultant electron-accepting power exhibited by a molecule upon chemical response due to the combined attractive and repulsive effects arising out of the presence of both electrons and nuclei. This new dual descriptor ($\Delta\omega^{\pm}$), therefore, seems to be an appraisal of the electrophilicity of a system relative to its own nucleophilicity and serves better to provide a more meaningful physical rationale towards understanding the "electrophilic power" of a system. Energy considerations, however, reveal that a larger value of ω^{+} for a system presupposes its enhanced capability to accept charge whereas a smaller value of ω^{-} necessitates the same to serve as a better donor. The mathematical basis of $\Delta\omega^{\pm}$ thus arises out of a parity between ω^{+} and ω^{-} where the negative (or reciprocal) of ω^{-} is compared with ω^{+}. Thus, net electrophilicity ($\Delta\omega^{\pm}$) in terms of ω^{+} and ω^{-} is formulated as

$$\Delta\omega^{\pm} = \{\omega^{+} - (-\omega^{-})\} = (\omega^{+} + \omega^{-})$$

or

$$\Delta\omega^{\pm} = \left\{\omega^{+} - \left(\frac{1}{\omega^{-}}\right)\right\}. \tag{13}$$

The polarizability (α) of an atom or a molecule means the lowest order response of its electron cloud to an external weak electric field [277, 278]. The static dipole polarizability (α), a linear response property, is defined as the second derivative of the total electronic energy (E) with respect to the external homogeneous electric field as

$$\alpha_{\alpha\beta} = -\left(\frac{\partial^2 E}{\partial F_{\alpha}\partial F_{\beta}}\right)_{F=0}, \tag{14}$$

where F_{α} and F_{β} are the electric field components for a fixed coordinate system with $\alpha, \beta, \gamma = x, y, z$. The polarizability ($\alpha$) is very sensitive to basis set, electron correlation and relativistic effects, and to the vibrational structure in case of a molecule. Qualitatively, polarizability (α) varies inversely with global hardness [279–284] (η) and is proportional to the global softness (S) as expected. Hence, an interesting linear correlation is established between the static dipole polarizability (α) and the third power of global softness [285–287].

The magnetizability (ξ) for a chemical system is a measure of the linear response of its electron cloud to an externally applied magnetic field and is expressed as

$$\left|\xi = -\left(\frac{\partial^2 \varepsilon(B)}{\partial B^2}\right)\right|_{B=0}, \tag{15}$$

where B is the external magnetic field.

The magnetizability (ξ) of a chemical species varies proportionately with its softness (S) and polarizability (α). A softer species is, therefore, more polarizable and hence more magnetizable [288].

In addition to the various conceptual DFT-based global reactivity descriptors, the local reactivity descriptors also provide valuable insights in determining chemical reactivity patterns. The various local reactivity indices like electron density ($\rho(\mathbf{r})$), Fukui function ($f(\mathbf{r})$), local softness ($s(\mathbf{r})$), local hardness ($\eta(\mathbf{r})$), and philicity ($\omega(\mathbf{r})$) help to assess the response of a particular atomic site in a molecule during a chemical attack.

The electron density ($\rho(\mathbf{r})$) provides useful information regarding site-reactivity for a molecule and is assessed in terms of the first-order variation of the energy as a function of the external potential [247, 289] ($v(\mathbf{r})$).

$$\rho(\mathbf{r}) = \left(\frac{\delta E}{\delta v(\mathbf{r})}\right)_N. \tag{16}$$

The Fukui function [261, 290] ($f(\mathbf{r})$) is usually applied to quantitatively evaluate the extent of local response of a particular active site of a chemical system. It is widely applied as a popular local variant and is defined as the differential change in electron density ($\rho(\mathbf{r})$) due to an infinitesimal change in the number of electrons (N).

$$f(\mathbf{r}) = \left(\frac{\partial \rho(\mathbf{r})}{\partial N}\right)_{v(\mathbf{r})} = \left(\frac{\delta \mu}{\delta v(\mathbf{r})}\right)_N. \tag{17}$$

Owing to a discontinuity in the derivative in Eq. (17) for integral values of N, three different types of Fukui functions can be defined by applying the finite difference and frozen core approximations as follows [261, 290]:

For nucleophilic attack

$$f^+(\mathbf{r}) = \left(\frac{\partial \rho(\mathbf{r})}{\partial N}\right)^+_{v(\mathbf{r})} \approx \rho_{N+1}(\mathbf{r}) - \rho_N(\mathbf{r}) \approx \rho_{\mathrm{LUMO}}(\mathbf{r}) \tag{18a}$$

For electrophilic attack

$$f^-(\mathbf{r}) = \left(\frac{\partial \rho(\mathbf{r})}{\partial N}\right)^-_{v(\mathbf{r})} \approx \rho_N(\mathbf{r}) - \rho_{N-1}(\mathbf{r}) \approx \rho_{\mathrm{HOMO}}(\mathbf{r}) \tag{18b}$$

For radical attack

$$f^0(\mathbf{r}) = \frac{1}{2}[f^+(\mathbf{r}) + f^-(\mathbf{r})]. \tag{18c}$$

Thus, $f(\mathbf{r})$, on the basis of the frontier orbital theory proposed by Fukui et al. [291–294], physically correlates the tendency of a local atomic site in a molecule to interact upon chemical response. The Fukui functions are, therefore, referred to as the DFT analogue of the frontier orbitals [295].

A coarse-grained atom by atom representation of the Fukui function, called condensed-to-atom Fukui function, was proposed by Yang and Mortier [262] and based on a finite-difference approach in terms of the Mulliken population analysis (MPA) scheme, they may be represented as

$$f_k^+ = q_k(N+1) - q_k(N) \quad \text{for nucleophilic attack,} \tag{19a}$$

$$f_k^- = q_k(N) - q_k(N-1) \quad \text{for electrophilic attack,} \tag{19b}$$

$$f_k^o = [q_k(N+1) - q_k(N-1)]/2 \quad \text{for radical attack,} \tag{19c}$$

where q_k refers to the electron population at a particular atomic site k in a molecule. The Eqs. (19a)–(19c) can be easily evaluated through population analysis data that helps in further quantum chemical calculations.

The global softness (S) mathematically means an integration of the individual local softness ($s(\mathbf{r})$) values computed for all the atomic sites of an entire molecule. A normalized condition between S and $s(\mathbf{r})$ is thus expressed as

$$\mathrm{S} = \int \mathrm{s}(\mathbf{r})\mathrm{d}\mathbf{r}, \tag{20}$$

which is quite at par with the condition as $\mathbf{f}(r)$ is normalized to unity.

The local softness $s(\mathbf{r})$ is related to the Fukui functions $f(\mathbf{r})$ through a chain rule as

$$s(\mathbf{r}) = \left(\frac{\partial\rho(\mathbf{r})}{\partial\mu}\right)_{v(\mathbf{r})} = \left(\frac{\partial\rho(\mathbf{r})}{\partial N}\right)_{v(\mathbf{r})} \cdot \left(\frac{\partial N}{\partial\mu}\right)_{v(\mathbf{r})} = f(\mathbf{r}) \cdot S. \tag{21}$$

The Fukui function ($f(\mathbf{r})$) makes an assessment of the tendency towards reactivity of the different local sites of the same molecule and thus serves as an intramolecular reactivity descriptor. The local softness ($s(\mathbf{r})$), on the other hand, compares and correlates the propensity of a pair of interacting molecular neighbors during chemical response and unlike $f(\mathbf{r})$ serves as an intermolecular reactivity descriptor.

Local hardness ($\eta(\mathbf{r})$) [296, 297] is introduced to quantify the reluctance of a particular atom or group in a molecule towards chemical attack. Unlike a normalized condition between $s(\mathbf{r})$ and S, however, $\eta(\mathbf{r})$ cannot be normalized to η. $\eta(\mathbf{r})$ is thus differently expressed as

$$\eta(\mathbf{r}) = \frac{1}{N}\int \frac{\delta^2 F[\rho(\mathbf{r})]}{\delta\rho(\mathbf{r})\delta\rho(\mathbf{r}')}\rho(\mathbf{r}')\mathrm{d}\mathbf{r}', \tag{22}$$

where $F[\rho(\mathbf{r})]$ defines the Hohenberg–Kohn universal functional. Local hardness is also expressed [296, 297] differently at par with global hardness (η) by replacing the number of electrons (N) with electron density ($\rho(\mathbf{r})$) as

$$\eta(\mathbf{r}) = \left(\frac{\delta\mu}{\delta\rho(\mathbf{r})}\right)_{v(\mathbf{r})}. \tag{23}$$

But Eq. (23) is ambiguous due to an inter-dependence of $\rho(\mathbf{r})$ and $v(\mathbf{r})$ and thus cannot be considered as a basic definition for $\eta(\mathbf{r})$. Thus, unlike η and S, $\eta(\mathbf{r})$ and $s(\mathbf{r})$ do not even hold a reciprocal relationship and are interconnected as

$$\int \eta(\mathbf{r})\mathbf{s}(\mathbf{r})\mathrm{d}\mathbf{r} = 1. \tag{24}$$

An acceptable definition of local hardness and subsequent identification of the "hard" sites in a molecule is, therefore, in demand [298–304]. The "minimum Fukui function rule" [301] asserts that hard reactions, unlike softer ones, would prefer sites with minimum Fukui function values. Although there are applications [305–307] of this rule, subsequent criticisms [303, 308] are also reported. The "minimum Fukui function rule" is unable to correlate the electrostatic hard–hard interactions that are predominantly charge-controlled with hardly any relevant effect from the associated frontier orbitals. The given rule cannot justify the inadequacy of the frontier orbitals and also misses the role of electrostatic interactions in hard–hard interactions [303, 304, 308].

The local hardness ($\eta(\mathbf{r})$) and the Fukui function ($f(\mathbf{r})$) can be used [309] to obtain the global hardness, which can be expressed as

$$\eta = \int \eta(\mathbf{r})f(\mathbf{r})\mathrm{d}r. \tag{25}$$

The concept of philicity ($\omega^{\alpha}(r)$) introduced by Chattaraj et al. [310, 311] signifies the aptitude towards reactivity (electrophilicity or nucleophilicity) of a local atomic site in a molecule. In other words, an increase (or decrease) in the reactivity of any local site of a molecule does not affect the electrophilicity (ω) of the entire species as a whole—the global reactivity parameter, ω remains conserved. The transformation of the local electrophilicity ($\omega(\mathbf{r})$) into global electrophilicity (ω) is expressed as

$$\omega = \omega \int f(\mathbf{r})\mathrm{d}r$$

or,

$$\omega = \int \omega f(\mathbf{r})\mathrm{d}r = \int \omega(\mathbf{r})\mathrm{d}r$$

thus,

$$\omega(\mathbf{r}) = \omega f(\mathbf{r}). \tag{26}$$

Although $\omega(\boldsymbol{r})$ contains information about both ω and $f(\boldsymbol{r})$ but $f(\boldsymbol{r})$ alone cannot provide any input regarding $\omega(\boldsymbol{r})$ without the knowledge of ω. The concept of philicity is widely applicable to diverse types of chemical reactions, which, therefore, presupposes $\omega(\mathbf{r})$ to be represented as $\omega^{\alpha}(\mathbf{r})$—the symbol of "philicity." The sign of α varies with the different categories of reactions, i.e. $\alpha = +, -$, and 0 represents nucleophilic, electrophilic, and radical attacks, respectively. Thus:

$$\begin{aligned} &\omega^{\alpha}(\mathbf{r}) = \omega \cdot f^{\alpha}(\mathbf{r}) \\ &or, \quad \omega_k^{\alpha} = \omega \cdot f_k^{\alpha} \end{aligned} \tag{27}$$

where $\omega_k{}^{\alpha}$ refers to the condensed-to-atom local philicity variant for the kth atomic site in a molecule. The philicity ($\omega_k{}^{\alpha}$) seems to be a better intermolecular local reactivity index than the Fukui function (f_k^{α}).

The group philicity concept [312] ($\omega_g{}^{\alpha}$) gains importance when we consider the reactivity of an atomic assembly or a group and is usually expressed as a summation of the individual condensed-to-atom philicities ($\omega_k{}^{\alpha}$) over all the relevant atoms. Thus

$$\omega_g^{\alpha} = \sum_{k=1}^{n} \omega_k^{\alpha} \tag{28}$$

where n denotes the number of atoms present in the reacting group and $\alpha = +, -$, and 0 refers to nucleophilic, electrophilic, and radical attacks, respectively.

The CDFT-based global and local reactivity descriptors and their modes of influencing chemical reactivity are further appreciated in terms of the various allied molecular electronic structure principles like the principle of maximum hardness [273, 313–319] (PMH), minimum polarizability principle [279, 320–322] (MPP), minimum electrophilicity principle [323, 324] (MEP), minimum magnetizability principle [288, 325] (MMP), etc. The various reactivity descriptors theoretically justify the observed chemical behavior of a molecule. Widespread applications of conceptual DFT towards modeling a plethora of molecular cluster assemblies accompanied by a rigorous analysis of their bonding, stability, and reactivity trends are adequately reported in the literature. The variation of the conceptual DFT-based reactivity indices as a function of structural changes and/or substituent effects convey invaluable insights into a quantitative correlation between chemical reactivity and toxicity. It [326] is, however, pointed out that the magnitudes of the various reactivity descriptors are quite dependent on the level of theory and the type of the basis set used. Thus, for an effective theoretical benchmarking to provide a structure–activity/toxicity correlation at par with experimental values, proper knowledge regarding the selection of level of theory and/or molecular basis set(s) is necessary. An appropriate choice indeed becomes new pathfinders towards building fruitful regression models for QSAR/QSPR/QSTR analyses.

3.1 Computational Details

The molecular geometries of the different classes of compounds under study are optimized at various levels (semi-empirical and DFT) of theory using the Gaussian 98 [327], Gaussian 03 [328] program packages. The stability of the optimized structures is characterized by their harmonic vibrational frequency values. The number of imaginary frequencies (NIMAG) is zero in every case, which corresponds to the existence of the given molecular geometry at a minimum on the potential energy landscape. The ionization potential (IP) and electron affinity (EA) are computed through the Koopmans' theorem [329] or by exploiting the ΔSCF technique. The atomic charges [260] (Q_k) and Fukui functions [261] ($f(\vec{r})$) are computed from the Mulliken population analysis (MPA) [262], natural population analysis (NPA), or Hirschfield population analysis (HPA) [330] schemes. The Hirschfield charges are computed with the BLYP/DND method using the DMOL^3 package [331].

3.2 Applications of Conceptual DFT in QSAR/QSTR

Various conceptual DFT-based reactivity indices in association with some new parameters are successfully employed in the development of stronger QSAR/QSTR models [332]. Deeper correlations of the toxicity of different classes of organic compounds like chlorinated benzenes [333], polychlorinated biphenyls [312, 334–336], and benzidine [337] at DFT level of theory are reported. The toxicity of the polychlorinated biphenyls as well as benzidine is influenced by its electron affinity and planarity. The interactions of the chlorinated benzo-derivatives and benzidine with other biomolecules like nucleic acid/base pairs or aryl hydrocarbon hydroxylase (AHH) receptors are primarily of charge-transfer type, which can be quantitatively assessed from Parr to Pearson formula [254] and can be given as

$$\Delta N = \frac{\mu_B - \mu_A}{2(\eta_A + \eta_B)}, \tag{29}$$

where, ΔN represents the fractional number of electrons transferred from system A to system B during a charge-transfer process occurring between two interacting systems A and B, which eventually produce an adduct A:B. μ_A, η_A and μ_B, η_B denote the chemical potential and hardness of systems A and B, respectively.

In these cases, the chlorobenzenes act as electron acceptors while the benzidine molecule behaves as an electron donor. The effect of chlorine substitution on the aromaticity of the planar benzene ring of the chlorobenzene derivatives is assessed from the nucleus independent chemical shift (NICS) criterion proposed by Schleyer et al. [338]. Among the various global and local molecular descriptors, electrophilicity (ω) is found to be the most appropriate reactivity parameter regarding toxicity

assessments for the entire molecule as well as for a particular local reactive site. Toxicity-based regression models developed by correlating the experimental biological activities (pIC_{50}) of polychlorinated dibenzofurans (PCDFs) with their corresponding activity (pIC_{50}) values computed by using the electrophilicity and the local electrophilic power illustrate a fine predictive relationship [339]. Padmanabhan et al. [340], in the line of a contemporary article by Toro-Labbé and coworkers [341], utilized the global philicity concept to propose a new multiphilic descriptor ($\Delta\omega_k$), which considers both the nucleophilicity and electrophilicity of a local atomic site in a molecule. The $\Delta\omega_k$ values computed for an entire set of chlorinated benzenes in both the gas and solvent media give a fair potential of its predictive ability regarding toxic effects on aquatic species. The group philicity concept as a possible descriptor in association with electrophilicity is also applied to build structure–toxicity-based relationship models for chlorophenols (CPs) [342] with high internal predictive ability. Effective QSAR models for the toxicity of CPs upon *Daphnia magna* are presented by utilizing the group philicities. Group philicity and electrophilicity together also provide useful structure–activity-based predictions for the CPs against *Brachydanio rerio* and *Bacillus*. Cronin et al. [343], in an earlier study, showed that electrophilicity can be successfully implemented to assess the toxic effects of a large number of aromatic compounds on *Tetrahymena pyriformis* and in that case, the maximum acceptor superdelocalizability (*A*max) proved to be a better descriptor than the energy of the lowest unoccupied molecular orbital (E_{LUMO}) in parametrizing ω. Electrophilicity (global and local) as well as its modified variants are, therefore, quite effective in presenting useful correlation of ecotoxicological trends of various organic molecules with bio-organisms.

Electrophilicity, as a cardinal reactivity index, is quite successfully implemented to study the biological activities of male and female sex hormones like testosterone and estrogen derivatives [344]. Another subsequent article [345] delineates that the number of atoms in a molecule (N_A) can be used as a valid molecular descriptor in the QSAR parlance towards explaining the activity of the sex hormones. A recent QSAR-based modeling [346] on the ecotoxicological effects of a large set of aromatic compounds on micro-organisms like *Tetrahymena pyriformis*, *Daphnia magna*, *Brachydanio rerio,* and *Bacillus* also reflects the usefulness of the atom counting (number of non-hydrogenic atoms) criterion as a simple and effective descriptor in association with ω and the ground state energy (E_{GS}). The electrophilicity protocol is also quite perfect in providing an impeccable prediction of toxicity, which builds a considerable data bank upon interaction of a large number of aliphatic compounds with *Tetrahymena pyriformis* [347]. In most cases, excellent correlations are obtained between toxicity and electrophilicity and/or its local variant, thereby giving rise to a rationale for the toxic action. Quantitative assessments of the logarithm of *n*-octanol/water partition coefficient ($\log K_{ow}$) (an important index for toxicological and pharmacological studies) on a large set of polychlorinated biphenyls (PCBs) [348] using electrophilicity as a descriptor yields reasonably good coefficients and internal predictive ability values. A recent article [349] describes the relative efficacy of electrophilicity, energy of the lowest unoccupied molecular orbital (E_{LUMO}), and $\log P$ (1-octanol/water partition coefficient) as global descriptors in predicting the toxicity ($pIGC_{50}$) of a large number of

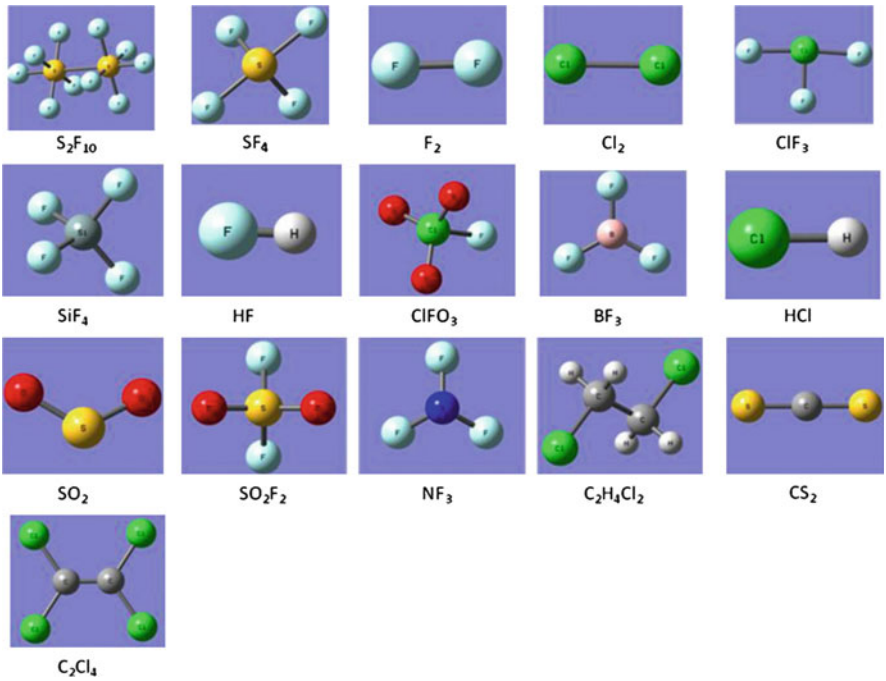


Fig. 1 Optimized ground state geometries (B3LYP/6-311 + G(d)) of some selected sulfur and halogen compounds (Reprinted with permission from [350]. Copyright 2011 IGI Global)

aliphatic compounds towards *Tetrahymena pyriformis*. It further transpires that electrophilicity marginally supersedes over E_{LUMO} to provide better regression analyses in most cases, refinements also being dependent on the units of the IGC_{50} magnitudes.

QSTR-based one-parameter and multi-parameter regression models involving the conceptual DFT-based global reactivity indices and the newly proposed net electrophilicity [274] ($\Delta\omega^{\pm}$) are developed to assess the toxic effects of some halogen, sulfur, and chlorinated aromatic compounds [350]. Two sets of inorganic compounds, containing mainly halogen and sulfur in the first set, displayed in Fig. 1, and chlorinated aromatic compounds in the second set, displayed in Fig. 2, are considered for investigation.

The R^2, R^2_{CV}, and R^2_{adj} values given in Table 1 and the experimental vs. calculated Log (LC_{50}) plots utilizing the one-parameter regression equations for sets 1 and 2 in Figs. 3 and 4, respectively, reveal that in case of the first set, the newly proposed net electrophilicity descriptor [274] ($\Delta\omega^{\pm}$) provides the best result, whereas, for the second set, both electrophilicity index (ω) as well as net electrophilicity index ($\Delta\omega^{\pm}$) put up a comparably decent performance. The newly proposed net electrophilicity index ($\Delta\omega^{\pm}$) as a refinement of electrophilicity (ω) has got a fair potential in constructing effective regression models suitable for QSAR/QSTR studies. Appropriate QSAR-based regression models are quite capable of explaining the toxicity trends of several alkali metal ions and various

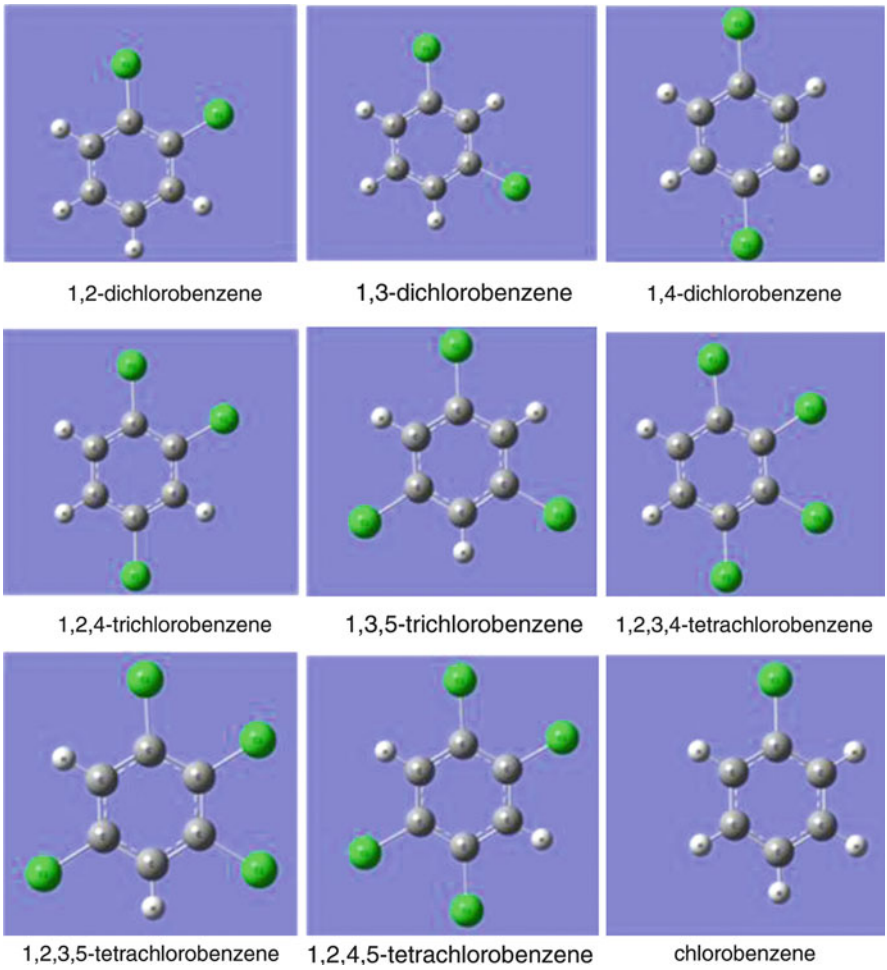


Fig. 2 Optimized ground state geometries (B3LYP/6-311 + G(d)) of chlorinated aromatic compounds (Reprinted with permission from [350]. Copyright 2011 IGI Global)

arsenic compounds [351]. A suitable three-parameter-based regression model consisting of electrophilicity (ω), philicity (ω_{As}^{+}), and atomic charge (Q_{As}) is found to be quite effective in predicting the toxicity of the arsenic compounds.

3.3 Application of Conceptual DFT in Designing Anticancer Drugs

Boron in many cases is now-a-days considered as a "new carbon" in drug modeling [352]. A conceptual DFT-based study is made to assess the anticancer properties of

Table 1 Regression models and various coefficients of determination with the various combinations of μ, ω, $\Delta\omega^{\pm a}$, and $\Delta\omega^{\pm b}$ for the sets 1 and 2

Regression model	R^2	$R_{CV}{}^2$	$R_{adj}{}^2$
Set 1 (Halogen and Sulfur compounds)			
Log LC_{50} = (0.521 × μ) + 6.60	0.694	0.618	0.673
Log LC_{50} = (−0.417 × ω) + 4.64	0.686	0.618	0.662
Log LC_{50} = (−0.2341 × $\Delta\omega^{\pm a}$) + 6.33	0.745	0.713	0.742
Log LC_{50} = (−0.233 × $\Delta\omega^{\pm b}$) + 6.33	0.728	0.696	0.727
Set 2 (Chlorinated aromatic compounds)			
Log LC_{50} = (−2.95 × μ) − 7.89	0.729	0.549	0.691
Log LC_{50} = (2.66 × ω) − 0.4	0.872	0.798	0.854
Log LC_{50} = (1.74 × $\Delta\omega^{\pm a}$) − 10.3	0.686	0.449	0.641
Log LC_{50} = (1.43 × $\Delta\omega^{\pm b}$) − 2.50	0.844	0.751	0.822

[a]Calculated as per Eqs. (9), (10).
[b]Calculated as per Eqs. (11), (12).
Reprinted with permission from [350]. Copyright 2011 IGI Global

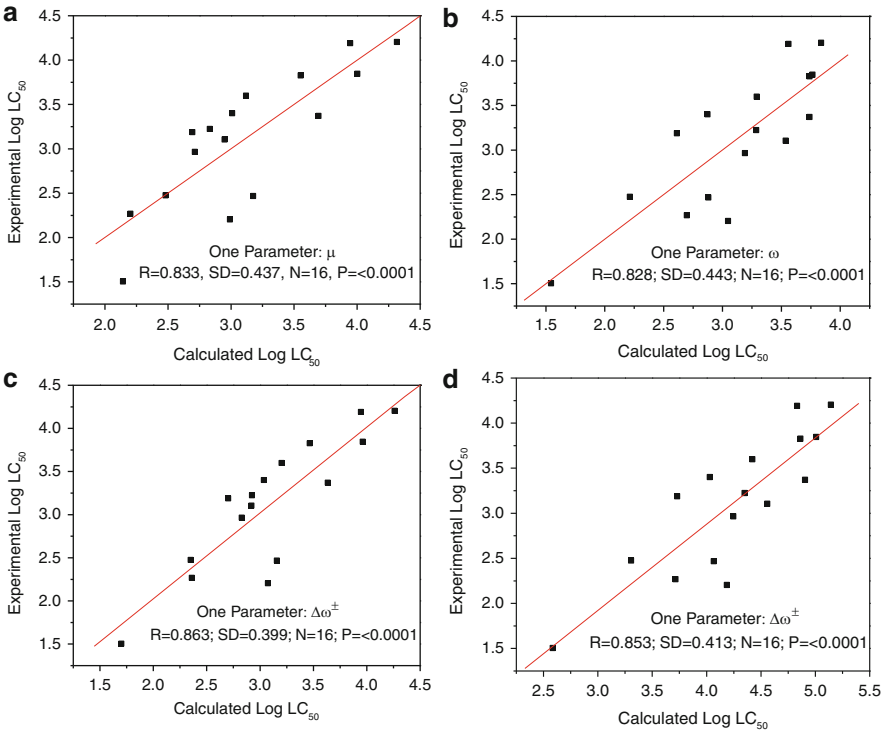


Fig. 3 (**a**) Experimental versus calculated toxicity (Log (LC_{50})) values as obtained using one-parameter chemical potential (μ) regression model for the complete set of Fluorine and Sulfur compounds; (**b**) same for electrophilicity index (ω); (**c**) same for net electrophilicity ($\Delta\omega^{\pm}$) from Eqs. (9), (10) and (13); (**d**) same for net electrophilicity ($\Delta\omega^{\pm}$) from Eqs. (11), (12) and (13) (Reprinted with permission from [350]. Copyright 2011 IGI Global)

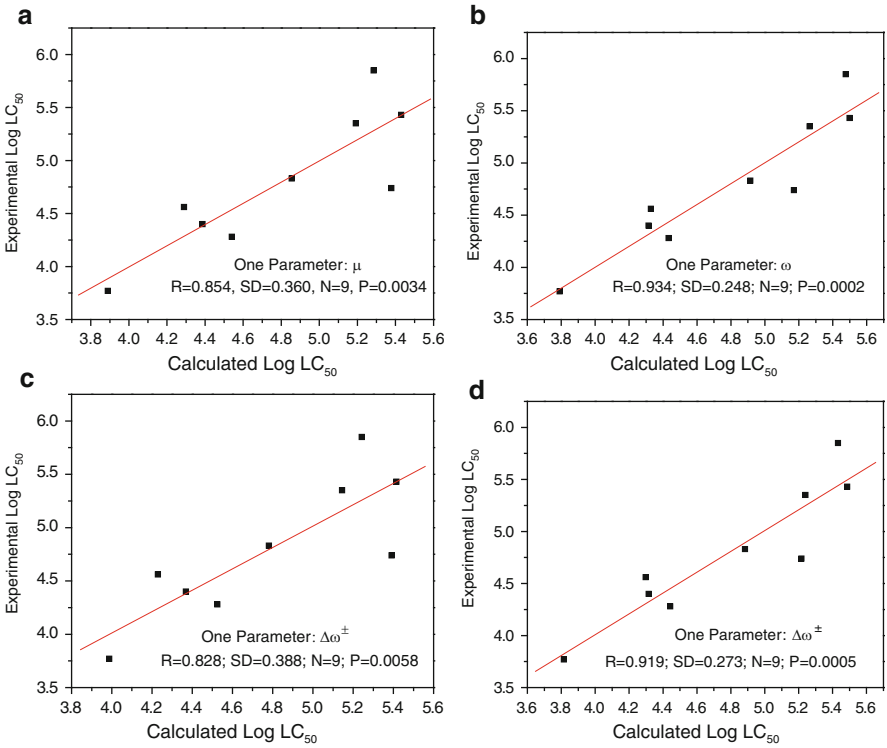
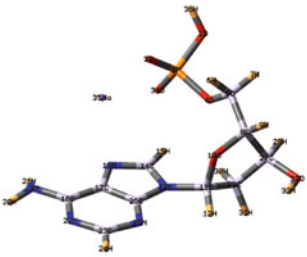


Fig. 4 (**a**) Experimental versus calculated toxicity (Log (LC_{50})) values as obtained using one-parameter chemical potential (μ) regression model for the complete set of chlorinated aromatic compounds; (**b**) same for electrophilicity (ω) index; (**c**) same for net electrophilicity ($\Delta\omega^{\pm}$) from Eqs. (9), (10) and (13); (**d**) same for net electrophilicity ($\Delta\omega^{\pm}$) from Eqs. (11), (12) and (13) (Reprinted with permission from [350]. Copyright 2011 IGI Global)

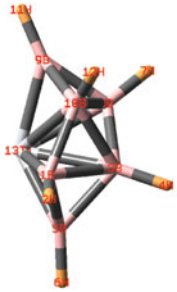
two newly proposed transition metal–borane clusters namely $(B_3H_3)_2Ti$ and $(B_3H_3)_2V$. In this regard the interaction of these metalloboranes with adenine (dAMP), guanine (dGMP), cytosine (dCMP), and thymine (dTMP) nucleotides are explored here. Figure 5 depicts the optimized geometries of four nucleotides, $(B_3H_3)_2Ti$, $(B_3H_3)_2V$, and their corresponding adducts computed at the B3LYP/6-311 + G(d) level of theory.

The interaction energy (ΔE), Gibbs free-energy change (ΔG), reaction enthalpy (ΔH), and reaction electrophilicity ($\Delta\omega$) of all the complexation reactions between the metalloboranes and the nucleotides, provided in Table 2, are negative, which ensures thermodynamic feasibility. The given metalloboranes thus effectively bind with nucleotide moieties and can be conceived as suitable anticancer drugs.

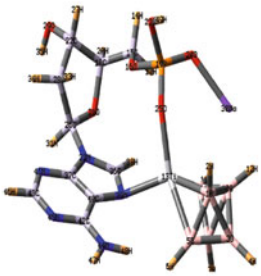
The reaction schemes to derive $(B_3H_3)_2Ti$ and $(B_3H_3)_2$ V from Cp_2Ti^{2+} and Cp_2V^{2+}, respectively, (Table 3) also justify the thermodynamic spontaneity of the given processes. Therefore, $(B_3H_3)_2Ti$ and $(B_3H_3)_2$ V, considered as new boron-based anticancer drugs, are supposed to be a possible alternative of Cp_2Ti^{2+} and Cp_2V^{2+}, respectively.



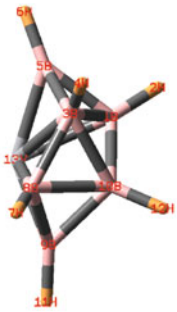
dAMP



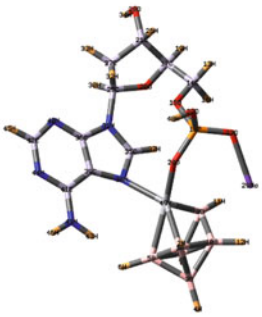
$(B_3H_3)_2$ Ti



$(B_3H_3)_2$Ti(5'- dAMP)

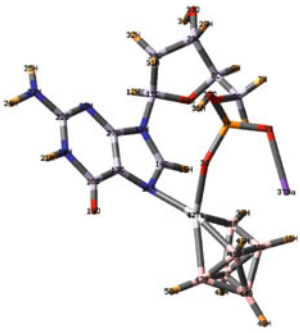


$(B_3H_3)_2$V



$(B_3H_3)_2$V(5'- dAMP)

$(B_3H_3)_2$ Ti



$(B_3H_3)_2$Ti(5'- dGMP)

Fig. 5 (continued)

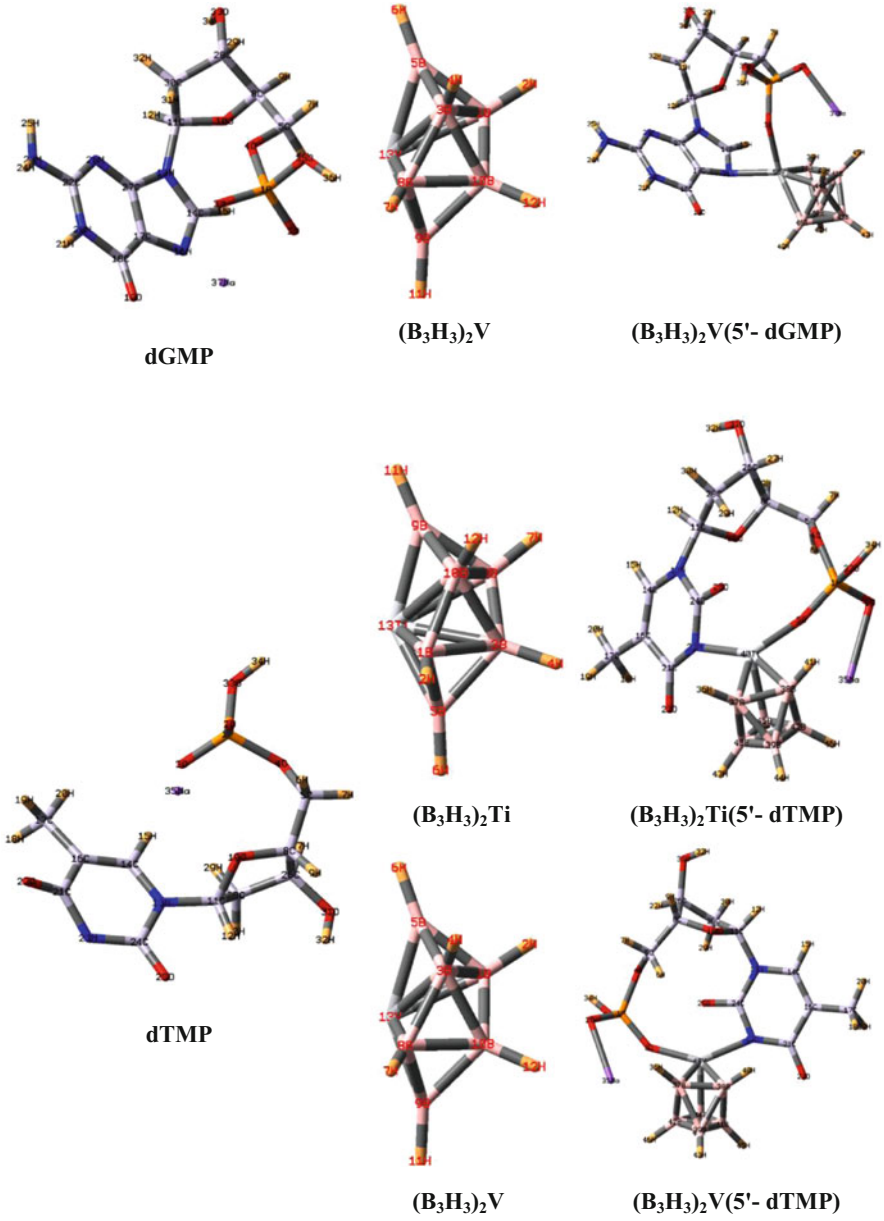


Fig. 5 (continued)

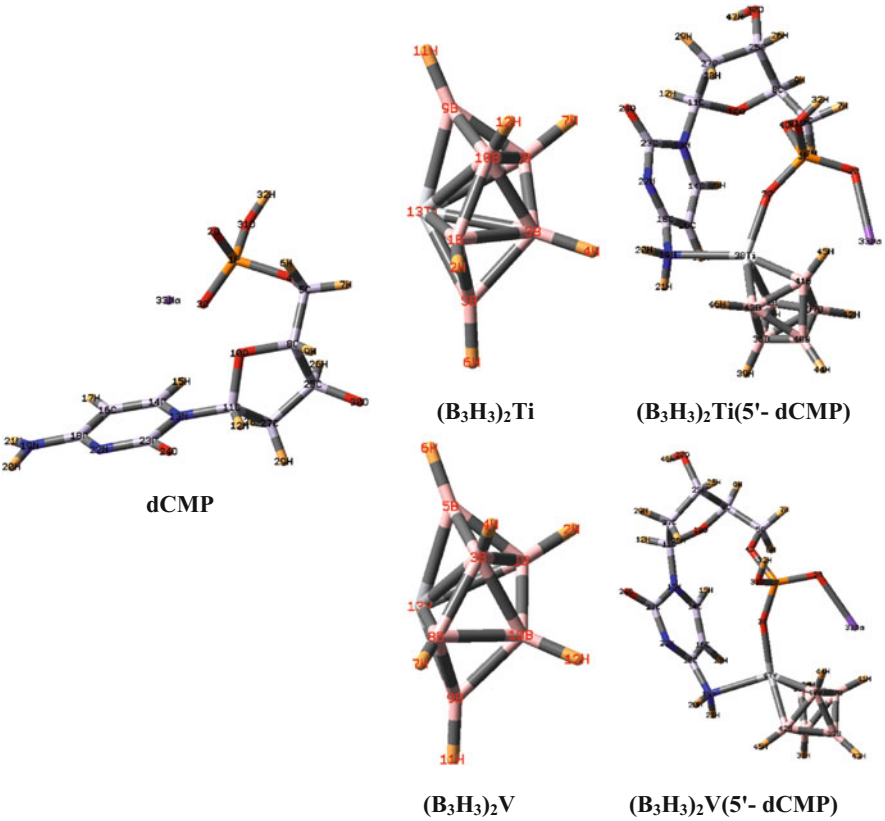


Fig. 5 The optimized geometries of the four nucleotides, $(B_3H_3)_2Ti$, $(B_3H_3)_2V$, and their corresponding adducts at B3LYP/6-311 + G(d) level of theory

Table 2 Interaction energy (ΔE, kcal/mol), Gibbs free-energy change (ΔG, kcal/mol), reaction enthalpy (ΔH, kcal/mol), hardness (η, eV), electrophilicity (ω, eV), and reaction electrophilicity ($\Delta\omega$, eV) of $(B_3H_3)_2Ti$, $(B_3H_3)_2V$, and their corresponding adducts with the four nucleotides at B3LYP/6-311 + G(d) level of theory

Systems	ΔE	ΔG	ΔH	η	ω	$\Delta\omega$
$(B_3H_3)_2Ti$				2.865	3.787	
$(B_3H_3)_2V$				2.810	4.290	
$(B_3H_3)_2Ti(5'\text{-dAMP})$	−89.1	−72.8	−86.5	2.762	2.523	−2.816
$(B_3H_3)_2V(5'\text{-dAMP})$	−89.7	−72.5	−86.9	3.218	2.465	−3.375
$(B_3H_3)_2Ti(5'\text{-dGMP})$	−82.1	−66.2	−79.3	2.776	2.006	−3.055
$(B_3H_3)_2V(5'\text{-dGMP})$	−80.9	−64.5	−78.1	3.211	2.008	−3.554
$(B_3H_3)_2Ti(5'\text{-dTMP})$	−152.3	−133.3	−148.3	3.023	2.851	−5.307
$(B_3H_3)_2V(5'\text{-dTMP})$	−141.7	−119.3	−137.1	2.906	3.754	−4.906
$(B_3H_3)_2Ti(5'\text{-dCMP})$	−85.3	−66.6	−82.3	2.744	2.583	−3.060
$(B_3H_3)_2V(5'\text{-dCMP})$	−86.9	−67.3	−83.8	3.107	2.607	−3.537

Table 3 Interaction energy (ΔE, kcal/mol), Gibbs free energy (ΔG, kcal/mol), reaction enthalpy (ΔH, kcal/mol), and reaction electrophilicity ($\Delta\omega$, eV) of the studied reaction schemes at B3LYP/6-311 + G(d) level of theory

Scheme	ΔE	ΔG	ΔH	$\Delta\omega$
$Cp_2Ti^{2+} + B_3H_3^{2-} = (B_3H_3)CpTi^{+} + Cp^{-}$	−304.3	−306.6	−304.0	−25.960
$(B_3H_3)CpTi^{+} + B_3H_3^{2-} = (B_3H_3)_2Ti + Cp^{-}$	−270.7	−270.6	−270.4	−21.008
$Cp_2V^{2+} + B_3H_3^{2-} = (B_3H_3)CpV^{+} + Cp^{-}$	−311.2	−314.6	−312.1	−25.672
$(B_3H_3)CpV^{+} + B_3H_3^{2-} = (B_3H_3)_2\,V + Cp^{-}$	−271.7	−269.9	−270.8	−18.095

4 Conclusions

A quantitative prediction of molecular behavior based on some known and tested standards is often found to give quite close correlations with allied experimental results. The precise theoretical assessment of a plausible practical thought (molecular property) indeed becomes a useful pathfinder to understand chemical reactivity. The application of this idea to chemistry is abbreviated in many ways, viz. QSAR, QSPR, QSTR, and QSRR. Such methods correlate the changes in molecular structure with chemical behavior like reactivity, toxicity, or any other property to build suitable one or multi-parameter-based regression models for predicting the corresponding parameters of unknown compounds. A comprehensive analysis of the growth of QSAR as a popular predictive algorithm is reported. The roles of various molecular parameters, chemical descriptors in building suitable regression models, and their useful applications towards explaining the biological activity and toxicity of several classes of compounds are mentioned. The crucial roles of QSAR methodology in designing drugs, tumor inhibitors, and in assessment of aquatic toxicity deserve a special mention. A conceptual DFT-based QSAR/QSTR treatment for biomolecules further adds some invaluable insights. Among the various conceptual DFT-based reactivity descriptors, the role of electrophilicity (ω), net electrophilicity ($\Delta\omega^{\pm}$), and philicity (ω_g^{α}) concepts are very useful in predicting molecular toxicity. Atom counting as a simple descriptor gives beautiful regression results in some cases. Novel boron-based transition metal clusters as potential anticancer drugs can be theoretically modeled and their further complexations with nucleotide moieties are studied. Boron can be treated as a "new carbon" in drug design. The search for better variables as well as methodologies for an even better prediction of molecular parameters is on.

Acknowledgments The authors thank Prof. Dr. Mihai V. Putz for inviting them to contribute a chapter in this book. P.K.C. thanks DST, New Delhi, for the J. C. Bose National Fellowship. S.P. thanks CSIR, New Delhi, for financial assistance.

References

1. Schultz TW, Cronin MTD, Walker JD, Aptula AO (2003) J Mol Struct Theochem 622:1–22
2. Schultz TW, Cronin MTD, Netzeva TI (2003) J Mol Struct Theochem 622:23–38
3. Gombar VK, Mattioni BE, Zwickl C, Deahl JT (2006) Computational approaches for assessment of toxicity: a historical perspective and current status. In: Ekins S (ed) Computational toxicology: risk assessment for pharmaceutical and environmental chemicals. Wiley, Hoboken, NJ
4. Nantasenamat C, Isarankura-Na-Ayudhya C, Naenna T, Prachayasittikul V (2009) Excli J 8:74–88
5. Selassie C, Verma RP (2010) History of quantitative structure–activity relationships. In: Abraham D (ed) Burger's medicinal chemistry, drug discovery and development. Wiley, New York, pp 1–95
6. Kar S, Roy K (2010) J Indian Chem Soc 87:1455–1515
7. Cronin MTD (2010) Recent advances in QSAR studies: challenges and advances. In: Puzyn T; Leszczynski J, Cronin MT (eds) Computational chemistry and physics, vol 8. Springer, Berlin, pp 3–11
8. Karcher W, Devillers J (1990) SAR and QSAR in environmental chemistry and toxicology: scientific tool or wishful thinking? In: Karcher W, Devillers J (eds) Practical applications of quantitative structure–activity relationships (QSAR) in environmental chemistry and toxicology. Kluwer Academic, Dordrecht, pp 1–12
9. Selassie CD, Mekapati SB, Verma RP (2002) QSAR: then and now. Curr Top Med Chem 2:1357–1379
10. Roy K, Mitra I (2009) Expert Opin Drug Discov 4:1157–1175
11. Hammett LP (1935) Chem Rev 17:125–136
12. Hammett LP (1937) J Am Chem Soc 59:96–103
13. Taft RW (1956) Separation of polar, steric and resonance effects in reactivity. In: Newman MS (ed) Steric effects in organic chemistry. Wiley, New York, pp 556–675
14. Hansch C (1993) Acc Chem Res 26:147–153
15. Hansch C, Hoekman D, Gao H (1996) Chem Rev 96:1045–1075
16. Lien EJ, Hansch C, Anderson SM (1968) J Med Chem 11:430–441
17. Hansch C, Klein T (1986) Acc Chem Res 19:392–400
18. Hansch C, Kim D, Leo AJ, Norellino E, Silipo C, Vittoria A (1989) CRC Crit Rev Toxicol 19:185–226
19. Langridge R, Klein TE (1990) Molecular graphics and drug design. In: Hansch C, Sammes PG, Taylor JB, Ramsden CA (eds) Comprehensive medicinal chemistry, vol 4. Pergamon, Oxford, pp 413–429
20. Blaney JM, Hansch C (1990) Application of molecular graphics to the analysis of macromolecular structure. In: Hansch C, Sammes PG, Taylor JB, Ramsden CA (eds) Comprehensive medicinal chemistry, vol 4. Pergamon, Oxford, p 459
21. Debnath AK, Shusterman AJ, DeCompadre RLL, Hansch C (1994) Mutat Res 305:63–72
22. Hansch C, Leo A (1995) QSAR in metabolism. In: Heller SR (ed) Exploring QSAR, fundamentals and applications in chemistry and biology. American Chemical Society, Washington, DC, pp 299–343
23. Hansch C, Telzer BR, Zhang L (1995) Crit Rev Toxicol 25:67–89
24. Hansch C, Gao H (1997) Chem Rev 97:2995–3059
25. Hansch C, Gao H, Hoekman D (1998) In: Devillers J (ed) Comparative QSAR. Taylor and Francis, London, pp 285–368
26. Gao H, Katzenellenbogen JA, Garg R, Hansch C (1999) Chem Rev 99:723–744
27. Hansch C, Kurup A, Garg R, Gao H (2001) Chem Rev 101:619–672
28. Hansch C, Fujita T (1995) Status of QSAR at the end of the twentieth century. In: Hansch C, Fujita T (eds) Classical and three-dimensional QSAR in a agrochemistry. American Chemical Society, Washington, DC, pp 1–12

29. Hansch C, Leo A (1995) QSAR of nonspecific toxicity. In: Heller SR (ed) Exploring QSAR, fundamentals and applications in chemistry and biology. American Chemical Society, Washington, DC, pp 169–221
30. Kubinyi H (1997) Drug Discov Today 2:538–546
31. Martin YC (1998) Perspect Drug Discov Des 12:3–23
32. Norinder U (1998) Perspect Drug Discov Des 12:15–39
33. Maddalena DJ (1998) Expert Opin Ther Pat 8:249–258
34. Garg G, Gupta SP, Gao H, Babu MS, Debnath AK, Hansch C (1999) Chem Rev 99:3525–3601
35. Voskresensky ON, Levitsky AP (2002) Curr Med Chem 9:367–1383
36. Yang GF, Huang X (2006) Curr Pharm Des 12:4601–4611
37. Bajot F (2010) The use of QSAR and computational methods in drug design. In: Puzyn T, Leszczynski J, Cronin MTD (eds) Recent advances in QSAR studies—methods and applications. Springer, Dordrecht, pp 261–282
38. Winkler DJ (2002) Brief Bioinform 3(1):73–86
39. Didziapetris R, Reynolds DP, Japertas P, Zmuidinavicius D, Petrauskas A (2006) Curr Comput Aid Drug Design 2:95–103
40. Roy K, Ghosh G (2010) Curr Pharm Des 16:2625–2639
41. Ajay Walters WP, Murcko MA (1998) J Med Chem 41:3314–3324
42. Wiese M, Pajeva IK (2001) Curr Med Chem 8(6):685–713
43. Schultz TW, Seward JR (2000) Sci Total Environ 249(1–3):73–84
44. Benigni R, Giuliani A, Franke R, Gruska A (2000) Chem Rev 100(10):3697–3714
45. Bashir SJ, Maibach HI (2000) Biochem Modulation Skin React 2000:61–64
46. Cronin MTD (2000) Curr Opin Drug Discov Dev 3(3):292–297
47. Garg R, Kurup A, Hansch C (2001) Crit Rev Toxicol 31(2):223–245
48. Freidig AP, Hermens JLM (2001) Quant Struct Activ Relat 19(6):547–553
49. Gombar VK, Enslein K (1996) J Chem Inf Comput Sci 36(6):1127–1134
50. Agatonovic-kustrin S, Beresford R, Yusof A, Pauzi M (2001) J Pharm Biomed Anal 25(2):227–237
51. Brusic V, Bucci K, Schonbach C, Petrovsky N, Zeleznikow J, Kazura JW (2001) J Mol Graph Modell 19(5):405–411
52. Burden FR, Winkler DA (2000) Mol Model Predict Bioact (Proceedings of the 12th European symposium on quantitative structure-activity relationships), pp 175–180
53. Lewis DFV (2000) Toxicology 144(1–3):197–203
54. Vedani A, Dobler M (2000) Prog Drug Res 55:105–13
55. Guba W, Cruciani G (2000) Mol Model Predict Bioact (Proceedings of the 12th European Symposium on Quantitative Structure-Activity Relationships), pp 89–94
56. Leo AJ, Hansch C (1999) Perspect Drug Discov Des 17:1–25
57. Warne MA, Nicholson JK (1999) Prog Environ Sci 1(4):327–344
58. Gupta SP (2000) Prog Drug Res 55:235–282
59. Klebe G (2000) J Mol Med 78(5):269–281
60. Podlogar BL, Ferguson DM (2000) Drug Des Discov 17(1):4–12
61. Lien EJ, Ren S (2000) Phytochem Bioact Agents 2000:21–41
62. Mickle T, Nair V (2000) Drugs Future 25(4):393–400
63. Hadjipavlou-Litina D (2000) Curr Med Chem 7(4):375–388
64. Kurup A, Garg R, Hansch C (2000) Chem Rev 100(3):909–924
65. Livingstone DJ (2000) J Chem Inf Comput Sci 40(2):195–209
66. Katritzky AR, Petrukhin R, Tatham D, Basak S, Benfenati E, Karelson M, Maran U (2001) J Chem Inf Comput Sci 41(3):679–685
67. Winkler DA, Burden FR (2002) Application of neutral networks to large dataset QSAR, virtual screening and library design. In: Bellavance L (ed) Combinatorial chemistry methods and protocols. Humana, Totowa, NJ

68. Schultz TW, Carlson RE, Cronin MTD, Hermens JLM, Johnson R, O'Brien PJ, Roberts DW, Siraki A, Wallace KD, Veith GD (2006) SAR QSAR Environ Res 17:413–428
69. Gerberick F, Aleksic M, Basketter DA, Casati S, Karlberg A-T, Kern PS, Kimber I, Lepoittevin J-P, Natsch A, Ovigne JM, Rovida C, Sakaguchi H, Schultz TW (2008) ATLA Altern Lab Anim 36:215–242
70. Vonk JA, Benigni R, Hewitt M, Nendza M, Segner H, van de Meent D, Cronin MTD (2009) ATLA Altern Lab Anim 37:557–571
71. Cronin MTD, Bajot F, Enoch SJ, Madden JC, Roberts DW, Schwöbel JAH (2009) ATLA Altern Lab Anim 37:513–521
72. McKinney JD (1996) Environ Health Perspect 104:810–816
73. Schwöbel JAH, Koleva YK, Enoch SJ, Bajot F, Hewitt M, Madden JC, Roberts DW, Schultz TW, Cronin MTD (2011) Chem Rev 111:2562–2596
74. Stuper AJ, Brugger WE, Jurs PC (1979) Computer assisted studies of chemical structure and biological function. Wiley-Interscience, New York
75. Kier LB, Hall LH (1986) Molecular connectivity in structure-activity analysis. Wiley, New York
76. Murray JS, Politzer PA (1994) General interaction properties function (GIPF): an approach to understanding and predicting molecular interactions. In: Politzer P, Murray JS (eds) Quantitative treatments of solute/solvent interactions. Elsevier, Amsterdam, pp 243–289
77. Hilal SH, Carreira LA, Karickhoff SW (1994) Estimation of chemical reactivity parameter and physical properties of organic molecules using SPARC. In: Politzer P, Murray JS (eds) Quantitative treatments of solute/solvent interactions. Elsevier, Amsterdam, pp 291–353
78. Abraham MH (1994) New solute descriptors for linear free energy relationships and quantitative structure-activity relationships. In: Politzer P, Murray JS (eds) Quantitative treatments of solute/solvent interactions. Elsevier, Amsterdam, pp 83–133
79. Abraham MH, Chadha HS, Dixon JP, Rafols C, Treiner C (1995) J Chem Soc Perkin Trans 2:887–894
80. Katritzky AR, Lobanov VS, Karelson M (1995) Chem Soc Rev 24:279–287
81. Randic M, Razinger M (1996) On the characterization of three-dimensional molecular structure. In: Balaban AT (ed) From chemical topology to three-dimensional geometry. Plenum, New York, pp 159–236
82. Balaban AT (1997) J Chem Inf Comput Sci 37:645–650
83. Lucic B, Trinajstic N (1999) J Chem Inf Comput Sci 39:121–132
84. Katritzky AR, Maran U, Lobanov VS, Karelson M (2000) J Chem Inf Comput Sci 40:1–18
85. Katritzky AR, Fara DC, Petrukhin RO, Tatham DB, Maran U, Lomaka A, Karelson M (2002) Curr Top Med Chem 2:1333–1356
86. Grover M, Singh B, Bakshi M, Singh S (2000) Pharm Sci Technol Today 3(1):28–35
87. Grover M, Singh B, Bakshi M, Singh S (2000) Pharm Sci Technol Today 3(2):50–57
88. Leach AR (2001) Molecular modelling: principles and applications. Prentice Hall, Englewood Cliffs, NJ
89. Sippl W (2010) Recent advances in QSAR studies: challenges and advances. In: Puzyn T, Leszczynski J, Cronin MTD (eds) Computational chemistry and physics, vol 8. pp 103–125
90. Cramer RDIII, Patterson DE, Bunce JD (1988) J Am Chem Soc 110:5959–5967
91. Jayatilleke PRN, Nair AC, Zauhar R, Welsh WJ (2000) J Med Chem 43:4446–4451
92. Sippl W, Holtje HD (2000) J Mol Struct Theochem 503:31–50
93. Hannongbua S, Nivasanond K, Lawtrakul L, Pungpo P, Wolschann P (2001) J Chem Inf Comput Sci 41:848–855
94. Avery MA, Alvim-Gaston M, Rodrigues CR, Barreiro EJ, Cohen FE, Sabnis YA, Woolfrey JR (2002) J Med Chem 45:292–303
95. Akamatsu M (2002) Curr Top Med Chem 12:1381–1394
96. Aboye TL, Sobhia ME, Bharatam PV (2004) Bioorg Med Chem 12:2709–2715
97. Bhongade BA, Gadad AK (2004) Bioorg Med Chem 12:2797–2805
98. Amin EA, Welsh WJ (2006) J Chem Inf Model 46:1775–1783

99. Kubinyi H (1997) Drug Discov Today 11:457–467
100. Kubinyi H (1997) Drug Discov Today 12:538–546
101. Roy K, Roy PP (2008) Chem Bio Drug Des 72:370–382
102. Roy K, Roy PP (2009) Expert Opin Drug Metabol Toxicol 5:1245–1266
103. jo de Brito MA, Rodrigues CR, Cirino JJV, de Alencastro RB, Castro HC, Albuquerque MG (2008) J Chem Inf Model 48:1706–1715
104. Singh HP, Chaturvedi AP, Sharma CS (2011) Int J Pharm Technol Res 3:231–236
105. Erickson JA, Jalaie M, Robertson DH, Lewis RA, Vieth M (2004) J Med Chem 47:45–55
106. Lei B, Li J, Lu J, Du J, Liu H, Yao X (2009) J Agric Food Chem 57:9593–9598
107. Durdagi S, Mavromoustakos T, Chronakis N, Papadopoulos MG (2008) Bioorg Med Chem 16:9957–9974
108. Liu H, Huang X, Shen J, Luo X, Li M, Xiong B, Chen G, Shen J, Yang Y, Jiang H, Chen K (2002) J Med Chem 45:4816–4827
109. Cheng F, Shen J, Luo X, Zhu W, Gu J, Ji R, Jiang H, Chen K (2002) Bioorg Med Chem 10:2883–2891
110. Zuo Z, Luo X, Zhu W, Shen J, Shen X, Jiang H, Chen K (2005) Bioorg Med Chem 13:2121–2131
111. Kothandan G, Gadhe CG, Madhavan T, Choi CH, Cho SJ (2011) Eur J Med Chem 46:4078–4088
112. Hao M, Li Y, Wang Y, Yan Y, Zhang S (2011) J Chem Inf Model 51:2560–2572
113. Diwan M, Hameed MA, Hamza A, Liu J, Zhan CG (2008) J Chem Inf Model 48:1760–1772
114. Vilar SG, Cozza S, Moro S (2008) Curr Top Med Chem 8:1555–1572
115. Hirokawa S, Imasaka T, Imasaka T (2005) Chem Res Toxicol 18:232–238
116. Jäntschi L, Bolboacă SD, Sestraş RE (2010) J Mol Model 16:377–386
117. Li F, Li X, Liu X, Zhang L, You L, Zhao J, Wu H (2011) Environ Toxicol Pharm 32:478–485
118. Li Y, Wang T, Xin J, Du X (2011) Bioinformatics and biomedical engineering. In: 5th International conference on iCBBE, pp 1–4
119. Gupta SP, Babu MS, Sowmya S (1998) Bioorg Med Chem 6:2185–2192
120. Chen KX, Xie HY, Li ZG, Gao JR (2008) Bioorg Med Chem Lett 18:5381–5386
121. Liu Q, Zhou H, Liu L, Chen X, Zhu R, Cao Z (2011) BMC Bioinformatics 12:294
122. Leonard JT, Roy K (2003) Drug Des Discov 18:165–180
123. Roy K, Leonard JT (2004) Bioorg Med Chem 12:745–754
124. Leonard JT, Roy K (2004) QSAR Comb Sci 23:23–35
125. Leonard JT, Roy K (2004) QSAR Comb Sci 23:387–398
126. Roy K, Leonard JT (2005) QSAR Comb Sci 24:579–592
127. Roy K, Leonard JT (2005) Bioorg Med Chem 13:2967–2973
128. Roy K, Leonard JT (2005) J Chem Inf Model 45:1352–1368
129. Roy K, Leonard JT (2006) Indian J Chem 45A:126–137
130. Leonard JT, Roy K (2006) Bioorg Med Chem 14:1039–1046
131. Leonard JT, Roy K (2006) Bioorg Med Chem Lett 16:4467–4474
132. Mandal AS, Roy K (2009) Eur J Med Chem 44:1509–1524
133. Heravi MJ, Baboli MA, Shahbazikhah P (2008) Eur J Med Chem 43:548–556
134. Shi LM, Fan Y, Myers TG, O'Connor PM, Paull KD, Friend SH, Weinstein JN (1998) J Chem Inf Comput Sci 38:189–199
135. Helguera AM, Rodríguez-Borges JE, García-Mera X, Fernández F, Cordeiro MNDS (2007) J Med Chem 50:1537–1545
136. Tan SJ, Yan YK, Lee PPF, Lim KH (2010) Future Med Chem 2:1591–1608
137. Boyles JR, Baird MC, Campling BG, Jain N (2001) J Inorg Biochem 84:159–162
138. Wang D, Lippard SJ (2005) Nat Rev Drug Discov 4:307–320
139. Centerwall CR, Goodisman J, Kerwood DJ, Dabrowiak JC (2005) J Am Chem Soc 127:12768–12769
140. Kelter G, Sweeney NJ, Strohfeldt K, Fiebig H-H, Tacke M (2005) Anticancer Drugs 16:1091–1098

141. Fichtner I, Pampillón C, Sweeney NJ, Strohfeldt K, Tacke M (2006) Anticancer Drugs 17:333–336
142. Dyson PJ, Sava G (2006) Dalton Trans 16:1929–1933 and references therein
143. Jakupec MA, Galanski M, Arion VB, Hartinger CG, Keppler BK (2008) Dalton Trans 2:183–194
144. Hannon MJ (2007) Pure Appl Chem 79:2243–2261
145. Ott I, Gust R (2007) Archiv der Pharmazie 340:117–126
146. Vessieres A Plamont MA, Cabestaing C, Claffey J, Dieckmann S, Hogan M, Mueller-Bunz H, Strohfeldt KA, Tacke M (2009) J Organomet Chem 694(6):874–879
147. Strohfeldt K, Tacke M (2008) Chem Soc Rev 37:1174–1187
148. Gasser G, Ott I, Metzler-Nolte N (2011) J Med Chem 54:3–25
149. Bohari MH, Srivastava HK, Sastry GN (2011) Org Med Chem Lett 1:3
150. Díaz HG, Marrero Y, Hernández I, Bastida I, Tenorio E, Nasco O, Uriarte E, Castañedo N, Cabrera MA, Aguila E, Marrero O, Morales A, Pérez M (2003) Chem Res Toxicol 16:1318–1327
151. Monteagudo MC, Díaz HG, Borges F, Dominguez ER, Cordeiro MNDS (2008) Chem Res Toxicol 21:619–632
152. Shadnia H, Wright JS (2008) Chem Res Toxicol 21:1197–1204
153. Turabekova MA, Rasulev BF (2004) Molecules 9:1194–1207
154. Vlaia V, Olariu T, Vlaia L, Butur M, Ciubotariu C, Medeleanu M, Ciubotariu D (2009) Farmacia 57(4):511–522
155. Saquib M, Gupta MK, Sagar R, Prabhakar YS, Shaw AK, Kumar R, Maulik PR, Gaikwad A, Sinha S, Srivastava AK, Chaturvedi V, Srivastava R, Srivastava BS (2007) J Med Chem 50:2942–2950
156. Ventura C, Martins F (2008) J Med Chem 51:612–624
157. Prathipati P, Ma NL, Keller TH (2008) J Chem Inf Model 48(12):2362–2370
158. Katritzky AR, Gordeevat EV (1993) J Chem Inf Comput Sci 33:835–857
159. Hawkins DM, Basak SC, Kraker JJ, Geiss KT, Witzmann FA (2006) J Chem Inf Model 46:9–16
160. Basak SC, Natarajan R, Mills D, Hawkins DM, Kraker JJ (2006) J Chem Inf Model 46:65–77
161. Könemann H (1981) Toxicology 19:209–221
162. Schüürmann G (1990) Environ Toxicol Chem 9:417–428
163. Cronin MTD, Dearden JC (1995) Quant Struct Activ Relat 14:1–7
164. Veith GD, Mekenyanb OG, Ankleya GT, Call DJ (1995) Chemosphere 30:2129–2142
165. Robert D, Carbo-Dorca R (1999) SAR QSAR Environ Res 10:401–422
166. Katritzky AR, Tatham DB, Maran UJ (2001) Chem Inf Comput Sci 41:1162–1176
167. Yu RL, Hu GR, Zhao YH (2002) J Environ Sci (China) 14:552–557
168. Salvito DT, Senna RJ, Federle TW (2002) Environ Toxicol Chem 21:1301–1308
169. Ren S, Schultz TW (2002) Toxicol Lett 129:151–160
170. Rose K, Hall LH (2003) SAR QSAR Environ Res 14:113–129
171. Roy K, Ghosh GJ (2004) Chem Inf Comput Sci 44:559–567
172. Hoover KR, Acree WE, Abraha MH (2005) Chem Res Toxicol 18:1497–1505
173. Casalegno M, Sello G, Benfenati E (2006) Chem Res Toxicol 19:1533–1539
174. Friesner RA, Beachy MD (1998) Curr Opin Struct Biol 8:257–262
175. Cartier A, Rivail J-L (1987) Chemom Intell Lab Syst 1:335–347
176. Gupta SP, Singh P, Bindal MC (1983) Chem Rev 83:633–649
177. Franke R (1984) Theoretical drug design methods. Elsevier, Amsterdam, pp 115–123
178. Gupta SP (1987) Chem Rev 87:1183–1253
179. Gupta SP (1991) Chem Rev 91:1109–1119
180. Brown RE, Simas AM (1982) Theor Chim Acta 62:1–16
181. Buydens L, Massart D, Geerlings P (1983) Anal Chem 55:738–744
182. Gruber C, Buss V (1989) Chemosphere 19:1595–1609
183. Bodor N, Gabanyi Z, Wong C-K (1989) J Am Chem Soc 111:3783–3786

184. Murugan R, Grendze MP, Toomey JE Jr, Katritzky AR, Karelson M, Lobanov VS, Rachwal P (1994) Chemtech 24:17–23
185. Karelson M, Lobanov V, Katritzky AR (1996) Chem Rev 96:1027–1043
186. Asatryan RS, Mailyan NS, Khachatryan L, Dellinger B (2002) Chemosphere 48:227–236
187. Friesner RA, Dunietz BD (2001) Acc Chem Res 34:351–358
188. Carloni P, Roethlisberger U, Parrinello M (2002) Acc Chem Res 35:455–464
189. Andreoni W, Curioni A, Mordasini T (2001) IBM J Res Dev 45:397–407
190. Raugei S, Gervasio FL, Carloni P (2006) Phys Stat Sol (b) 243:2500–2515
191. Robertazzi A, Magistrato A, Peraro MD, Carloni P (2011) Metallic systems: a quantum chemists's perspective, Chap 1. Taylor and Francis Group, LLC, New York, pp 1–27
192. Segall MD (2002) J Phys Condens Matter 14:2957–2973
193. Shaik S, Cohen S, Wang Y, Chen H, Kumar D, Thiel W (2010) Chem Rev 110:949–1017
194. Sulpizi M, Folkers G, Rothlisberger U, Carloni P, Scapozza L (2002) Quant Struct Activ Relat 21:173–181
195. Cao J, Jin S, Han B, Liu W, Zhao L, Zhong R (2010) 3rd International conference on biomedical engineering and informatics, pp 2374–2376
196. Bayat Z, Vahdani S (2011) J Chem Pharm Res 3(1):93–102
197. Chojnacki H, Kuduk-Jaworska J, Jaroszewicz I, Jański JJ (2009) J Mol Model 15:659–664
198. Sarmah P, Deka RC J (2009) Comput Aid Mol Des 23:343–354
199. Sarmah P, Deka RC (2010) J Mol Model 16:411–418
200. Arulmozhiraja S, Fujii T, Sato G (2002) Mol Phys 100:423–431
201. Arulmozhiraja S, Fujii T, Morita M (2002) J Phys Chem A 106:10590–10595
202. Wan J, Zhang L, Yang G (2004) J Comput Chem 25:1827–1832
203. Arulmozhiraja S, Morita M (2004) J Phys Chem A 108:3499–3508
204. Arulmozhiraja S, Morita M (2004) Chem Res Toxicol 17:348–356
205. Wan J, Zhang L, Yang G, Zhan C-G (2004) J Chem Inf Comput Sci 44:2099–2105
206. Xiu-Fen Y, He-Ming X, Xue-Hai J, Xue-Dong G (2005) Chin J Chem 23947–23952
207. Pasha FA, Srivastava HK, Singh PP (2005) Bioorg Med Chem 13:6823–6829
208. Lameira J, Alves CN, Moliner V, Silla E (2006) Eur J Med Chem 41:616–623
209. Yan X-F, Xiao H-M, Ju X-H, Gong X-D (2006) Huaxue Xuebao 64:375–380
210. Eroglu E, Türkmen H (2007) J Mol Graph Model 26:701–708
211. Gu C, Jiang X, Ju X, Yu G, Bian Y (2007) Chemosphere 67:1325–1334
212. Gu CG, Jiang X, Ju XH, Yang XL, Yu GF (2007) SAR QSAR Environ Res 18:603–619
213. Zhao Y-Y, Tao F-M, Zeng EY (2007) J Phys Chem A 111:11638–11644
214. Zhao YY, Tao F-M, Zeng EY (2008) Chemosphere 73:86–91
215. Mehdipour AR, Safarpour MA, Taghavi F, Jamali M (2009) QSAR Comb Sci 28:568–575
216. Wang Y, Chen J, Li F, Qin H, Qiao X, Hao C (2009) Chemosphere 76:999–1005
217. Manzoni V, Lyra ML, Cavada BS, Neto NS, Freire VN (2011) Int J Quant Chem 111:1270–1279
218. Elstner M, Frauenheim T, Suhai S (2003) J Mol Struct Theochem 632:29–41
219. Elstner M (2006) Theor Chem Acc 116:316–325
220. Putz MV, Lacrămă A-M (2007) Int J Mol Sci 8:363–391
221. Putz MV, Lacrămă A-M, Ostafe V (2007) Res Lett Ecol. doi:10.1155/2007/12813
222. Lacrămă A-M, Putz MV, Ostafe V (2007) Int J Mol Sci 8:842–863
223. Putz MV, Duda-Seiman C, Duda-Seiman DM, Putz AM (2008) Int J Chem Model 1:45–62
224. Putz MV, Putz (Lacrămă) AM (2008)Studia Universitatis Babeş-Bolyai–Seria Chimia 53:73–81
225. Putz MV, Duda-Seiman D, Mancaş S, Duda-Seiman C, Lacrămă A-M (2008) Quantum and topological impact on HMG-CoA reductase inhibitors. In: Putz MV (ed) Advances in quantum chemical bonding structures, chap 15. Transworld Research Network, Kerala, pp 355–387

226. Lacrămă A-M, Putz MV, Ostafe V (2008) Designing a spectral structure–activity ecotoxicologistical battery. In: Putz MV (ed) Advances in quantum chemical bonding structures, chap 16. Transworld Research Network, Kerala, pp 389–419
227. Putz MV, Putz A-M, Lazea Luciana M, Ienciu L, Chiriac A (2009) Int J Mol Sci 10:1193–1214
228. Chicu SA, Putz MV (2009) Int J Mol Sci 10:4474–4497
229. Duda-Seiman C, Duda-Seiman D, Putz MV, Ciubotariu D (2007) Digest J Nanomater Biostruct 2:207–219
230. Ciubotariu D, Deretey E, Oprea TI, Sulea T, Simon Z, Kurunczi L, Chiriac A (1993) Quant Struct Activ Relat 12:367–372
231. Simon Z, Chiriac A, Ciubotariu D, Muresan S, Bologa C, Sulea T, Kurunczi L (1996) Metoda MTD (The MTD method). In: Chiriac A, Ciubotariu D, Simon Z (eds) RelaŃii cantitative structură chimică – activitate biologică (QSAR) Metoda MTD (Quantitative chemical structure – biological activity relationships studies (QSAR). The MTD method), chap 5. Mirton Publishing House, Timisoara
232. Duda-Seiman C, Duda-Seiman D, Dragos D, Medeleanu M, Careja V, Putz MV, Lacrămă A-M, Chiriac A, Nuţiu R, Ciubotariu D (2006) Int J Mol Sci 7:537–555
233. Latosińska JN (2004) Chem Phys Lett 398:324–329
234. Latosińska JN (2005) J Pharm Biomed Anal 38:577–587
235. Holzgrabe U, Diehl BWK, Wawer I (1998) J Pharm Biomed 17:557–616
236. Kalinkova GN (1999) Vib Spectrosc 19:307–320
237. Silverstein RM, Bassler GC, Morrill TC (1991) Spectrometric identification of organic compounds. Wiley, Chichester
238. Liu S-S, Liu H-L, Yin C-S, Wang L-S (2003) J Chem Inf Comput Sci 43:964–969
239. Kaliszan R (2007) Chem Rev 107:3212–3246
240. Lobato M, Amat L, Besalu´ E, Carbo´-Dorca R (1997) Quant Struct Activ Relat 16:465–472
241. Gallegos A, Gironés X (2005) J Chem Inf Model 45:321–326
242. Roy K, Popelier PLA (2008) QSAR Comb Sci 27:1006–1012
243. Roy K, Popelier PLA (2008) Bioorg Med Chem Lett 18:2604–2609
244. Mekenyan OG, Veith GD (1993) SAR QSAR Environ Res 1:335–344
245. Cronin MTD, Dearden JC, Duffy JC, Edwards R, Manga N, Worth AP, Worgan ADP (2002) SAR QSAR Environ Res 13:167–176
246. Li Y, Wang T, Xin J, Du X (2011) 5th International conference on bioinformatics and biomedical engineering (iCBBE). Research Academy of Energy & Environmental Studies, North China Electronics Power University, Beijing, China, pp 1–4
247. Parr RG, Yang W (1989) Density functional theory of atoms and molecules. Oxford University Press, New York
248. Geerlings P, De Proft F, Langenaeker W (2003) Chem Rev 103:1793–1874
249. Chattaraj PK (ed) (2009) Chemical reactivity theory: a density functional view. Taylor and Francis/CRC, Boca Raton, FL
250. Chattaraj PK, Giri S (2009) Ann Rep Prog Chem Sect C Phys Chem 105:13–39
251. Chakraborty A, Duley S, Giri S, Chattaraj PK (2010) An understanding of the origin of chemical reactivity from a conceptual DFT approach. In: Sukumar N (ed) A matter of density: exploring the electron density concept in the chemical, biological, and materials sciences. Wiley, New York
252. Chattaraj PK (1992) J Indian Chem Soc 69:173–183
253. Parr RG, Donnelly RA, Levy M, Palke WE (1978) J Chem Phys 68:3801–3807
254. Parr RG, Pearson RG (1983) J Am Chem Soc 105:7512–7516
255. Pearson RG (1997) Chemical hardness: applications from molecules to solids. Wiley, Weinheim
256. Parr RG, Szentpaly Lv, Liu S (1999) J Am Chem Soc 121:1922–1924
257. Chattaraj PK, Sarkar U, Roy DR (2006) Chem Rev 106:2065–2091
258. Chattaraj PK, Roy DR (2007) Chem Rev 107:PR46–PR74

259. Chattaraj PK, Giri S, Duley S (2011) Chem Rev 111:PR43–PR75
260. Mulliken RS (1955) J Chem Phys 23:1833–1840
261. Parr RG, Yang W (1984) J Am Chem Soc 106:4049–4050
262. Yang W, Mortier WJ (1986) J Am Chem Soc 108:5708–5711
263. Pauling L (1932) J Am Chem Soc 54:3570–3582
264. Pauling L (1960) The nature of the chemical bond. Cornell University Press, Ithaca, NY
265. Mulliken RS (1934) J Chem Phys 2:782–793
266. Mulliken RS (1935) J Chem Phys 3:573–585
267. Gyftopoulos EP, Hatsopoulos GN (1968) Proc Natl Acad Sci USA 60:786–793
268. Pearson RG (1963) J Am Chem Soc 85:3533–3539
269. Pearson RG (1966) Science 151:172–177
270. Pearson RG (1968) J Chem Educ 45:581–587
271. Pearson RG (1968) J Chem Educ 45:643–648
272. Pearson RG, Chattaraj PK (2008) Chemtracts Inorg Chem 21:1–7
273. Pearson RG (1987) J Chem Educ 64:561–567
274. Chattaraj PK, Chakraborty A, Giri S (2009) J Phys Chem A 113:10068–10074
275. Maynard AT, Huang M, Rice WG, Covel DG (1998) Proc Natl Acad Sci USA 95:11578–11583
276. Gázquez JL, Cedillo A, Vela A (2007) J Phys Chem A 111:1966–1970
277. van Vleck J (1932) The theory of electric and magnetic susceptibilities. Oxford University Press, Oxford
278. Dalgarno A (1962) Adv Phys 11:281–315
279. Pearson RG (1986) Proc Natl Acad Sci USA 83:8440–8441
280. Politzer P (1987) J Chem Phys 86:1072–1073
281. Ghanty TK, Ghosh SK (1993) J Phys Chem 97:4951–4953
282. Fuentealba P, Reyes O (1993) J Mol Struct Theochem 282:65–70
283. Pal S, Chandra AK (1995) J Phys Chem 99:13865–13867
284. Ghanty TK, Ghosh SK (1996) J Phys Chem 100:12295–12298
285. Vela A, Gázquez JL (1990) J Am Chem Soc 112:1490–1492
286. Chattaraj PK, Poddar A (1998) J Phys Chem A 102:9944–9948
287. Simon-Manso Y, Fuentealba P (1998) J Phys Chem A 102:2029–2032
288. Chattaraj PK, Arun Murthy TVS, Giri S, Roy DR (2007) J Mol Struct Theochem 813:63–65
289. Parr RG, Yang W (1995) Annu Rev Phys Chem 46:701–728
290. Ayers PW, Levy M (2000) Theor Chem Acc 103:353–360
291. Fukui K, Yonezawa T, Shingu H (1952) J Chem Phys 20:722–725
292. Fukui K, Yonezawa T, Nagata C, Shingu H (1954) J Chem Phys 22:1433–1442
293. Fukui K (1982) Science 218:747–754
294. Fukui K (1975) Theory of orientation and stereoselection. Springer, Berlin
295. Yang W, Parr RG, Pucci R (1984) J Chem Phys 81:2862–2863
296. Berkowitz M, Ghosh SK, Parr RG (1985) J Am Chem Soc 107:6811–6814
297. Ghosh SK, Berkowitz M (1985) J Chem Phys 83:2976–2983
298. Klopman G (1968) J Am Chem Soc 90:223–234
299. Berkowitz M (1987) J Am Chem Soc 109:4823–4825
300. Gázquez JL, Méndez F (1994) J Phys Chem 98:4591–4593
301. Li Y, Evans JNS (1995) J Am Chem Soc 117:7756–7759
302. Damoun S, Van de Woude G, Méndez F, Geerlings P (1997) J Phys Chem A 101:886–893
303. Chattaraj PK (2001) J Phys Chem A 105:511–513
304. Anderson JSM, Melin J, Ayers PW (2007) J Chem Theory Comput 3:358–374
305. Méndez F, Gázquez JL (1994) J Am Chem Soc 116:9298–9301
306. Li Y, Evans JNS (1996) Proc Natl Acad Sci USA 93:4612–4616
307. Pérez P, Simon-Manso Y, Aizman A, Fuentealba P, Contreras R (2000) J Am Chem Soc 122:4756–4762

308. Melin J, Aparicio F, Subramanian V, Galvan M, Chattaraj PK (2004) J Phys Chem A 108:2487–2491
309. Chattaraj PK, Roy DR, Geerlings P, Torrent-Sucarrat M (2007) Theor Chem Acc 118:923–930
310. Chattaraj PK, Maiti B, Sarkar U (2003) J Phys Chem A 107:4973–4975
311. Roy DR, Parthasarathi R, Padmanabhan J, Sarkar U, Subramanian V, Chattaraj PK (2006) J Phys Chem A 110:1084–1093
312. Parthasarathi R, Padmanabhan J, Elango M, Subramanian V, Chattaraj PK (2004) Chem Phys Lett 394:225–230
313. Parr RG, Chattaraj PK (1991) J Am Chem Soc 113:1854–1855
314. Chattaraj PK, Liu GH, Parr RG (1995) Chem Phys Lett 237:171–176
315. Chattaraj PK (1996) Proc Indian Natl Sci Acad A 62:513–531
316. Ayers PW, Parr RG (2000) J Am Chem Soc 122:2010–2018
317. Chattaraj PK, Fuentealba P, Gomez B, Contreras R (2000) J Am Chem Soc 122:348–351
318. Pearson RG (1993) Acc Chem Res 26:250–255
319. Parr RG, Zhou Z (1993) Acc Chem Res 26:256–258
320. Chattaraj PK, Sengupta S (1996) J Phys Chem 100:16126–16130
321. Chattaraj PK, Sengupta S (1997) J Phys Chem A 101:7893–7900
322. Chattaraj PK, Fuentealba P, Jaque P, Toro-Labbé A (1999) J Phys Chem A 103:9307–9312
323. Chamorro E, Chattaraj PK, Fuentealba P (2003) J Phys Chem A 107:7068–7072
324. Parthasarathi R, Elango M, Subramanian V, Chattaraj PK (2005) Theor Chem Acc 113:257–266
325. Tanwar A, Pal S, Roy DR, Chattaraj PK (2006) J Chem Phys 125:056101–056102
326. Vijayaraj R, Subramanian V, Chattaraj PK (2009) J Chem Theory Comput 5:2744–2753
327. Gaussian 98, Revision A.6. Gaussian, Inc., Pittsburgh, PA
328. Gaussian 03, Revision B.03. Gaussian, Inc., Pittsburgh, PA
329. Koopmans TA (1933) Physica 1:104–113
330. Hirschfeld FL (1977) Theor Chim Acta 44:129–138
331. DMOL^3, Accelrys. Inc., San Diego, CA, USA
332. Chattaraj PK, Nath S, Maiti B (2003) Reactivity descriptors. In: Tollenaere J, Bultinck P, Winter HD, Langenaeker W (eds) Computational medicinal chemistry for drug discovery, Chap 11. Marcel Dekker, New York, pp 295–322
333. Padmanabhan J, Parthasarathi R, Subramanian V, Chattaraj PK (2005) J Phys Chem A 109:11043–11049
334. Thanikaivelan P, Subramanian V, Raghava Rao J, Nair BU (2000) Chem Phys Lett 323:59–70
335. Parthasarathi R, Padmanabhan J, Subramanian V, Maiti B, Chattaraj PK (2003) J Phys Chem A 107:10346–10352
336. Parthasarathi R, Padmanabhan J, Subramanian V, Maiti B, Chattaraj PK (2004) Curr Sci 86:535–542
337. Parthasarathi R, Padmanabhan J, Subramanian V, Sarkar U, Maiti B, Chattaraj PK (2003) Internet Electron J Mol Des 2:798–813
338. Schleyer PvR, Maerker C, Dransfeld A, Jiao H, Hommes NJRVE (1996) J Am Chem Soc 118:6317–6318
339. Sarkar U, Padmanabhan J, Parthasarathi R, Subramanian V, Chattaraj PK (2006) J Mol Struct Theochem 758:119–125
340. Padmanabhan J, Parthasarathi R, Subramanian V, Chattaraj PK (2006) J Phys Chem A 110:2739–2745
341. Morell C, Grand A, Toro-Labbé A (2005) J Phys Chem A 109:205–212
342. Padmanabhan J, Parthasarathi R, Subramanian V, Chattaraj PK (2006) Chem Res Toxicol 19:356–364
343. Cronin MTD, Manga N, Seward JR, Sinks GD, Schultz TW (2001) Chem Res Toxicol 14:1498–1505

344. Parthasarathi R, Subramanian V, Roy DR, Chattaraj PK (2004) Bioorg Med Chem 12:5533–5543
345. Roy DR, Pal N, Mitra A, Bultinck P, Parthasarathi R, Subramanian V, Chattaraj PK (2007) Eur J Med Chem 42:1365–1369
346. Giri S, Chakraborty A, Gupta A, Roy DR, Vijayaraj R, Parthasarathi R, Subramanian V, Chattaraj PK (2012) Modeling ecotoxicity as applied to some selected aromatic compounds: a conceptual DFT based quantitative-structure-toxicity- relationship (QSTR) analysis. In: Castro EA, Haghi AK (eds) Advanced methods and applications in chemoinformatics: research progress and new applications, Chap 1. Engineering Science Reference (an imprint of IGI Global), pp 1–24
347. Roy DR, Parthasarathi R, Maiti B, Subramanian V, Chattaraj PK (2005) Bioorg Med Chem 13:3405–3412
348. Padmanabhan J, Parthasarathi R, Subramanian V, Chattaraj PK (2006) Bioorg Med Chem 14:1021–1028
349. Pandith AH, Giri S, Chattaraj PK (2010) Org Chem Int. Doi:. doi:10.1155/2010/545087
350. Gupta A, Chakraborty A, Giri S, Subramanian V, Chattaraj PK (2011) Int J Chemoinform Chem Eng 1:61–74
351. Roy DR, Giri S, Chattaraj PK (2009) Mol Div 13:551–556
352. Hunter P (2009) EMBO Rep 10(2):125–128

Struct Bond (2013) 150: 181–232
DOI: 10.1007/978-3-642-32750-6_6

DFT Chemical Reactivity Driven by Biological Activity: Applications for the Toxicological Fate of Chlorinated PAHs

Mihai V. Putz and Ana-Maria Putz

Abstract The logistic kinetics model of quantitative reactivity–activity relationships is presented as a modern tool for modeling chemical–biological interactions using catalytic progress curves for the chemical ligand species that bind biological cell receptors. Chemical reactivity principles of electronegativity and chemical hardness, their necessity, the associated scenario of chemical bonding, the critical points in ab initio quantum chemical computations of the molecular structures, and the QSAR-OECD principles are reviewed. Illustrative applications are described for chlorinated polycyclic aromatic hydrocarbons (Cl-PAHs) bound to aryl hydrocarbon receptors (AhRs) in human breast cancer MCF-7 cells, *Pimephales promelas* and rats by recognizing key features of *Biological Activity driving Chemical Reactivity* (BioAct-ChemReact) interaction mechanisms. The specific feedback behavior of biological sites against chemical attacks unveil the manner in which chemical reactivity indices and principles of electronegativity and chemical hardness act successively or synergistically (in chemical power and electrophilicity) to regulate chemical reactivity hierarchies in ligand–receptor mechanisms that reflect observed (or *in silico* computed) biological or toxicological effects.

M.V. Putz (✉)
Biology–Chemistry Department, West University of Timişoara, Pestalozzi Str. No. 16, Timişoara 300115, Romania

"Nicolas Georgescu-Roegen" Forming and Research Center of West University of Timişoara, 4th, Oituz Street, Timişoara 300086, Romania
e-mail: mv_putz@yahoo.com; mvputz@cbg.uvt.ro

A.-M. Putz
Institute of Chemistry Timişoara of the Romanian Academy, 24 Mihai Viteazul Bld., Timişoara 300223, Romania

Biology–Chemistry Department, West University of Timişoara, Pestalozzi Str. No. 16, Timişoara 300115, Romania
e-mail: putzanamaria@yahoo.com

Keywords Ab initio computations • AhR-mediated toxicity • Algebraic norm • Carcinogenity • Chemical hardness • Conceptual DFT • Electronegativity • EROL activity • Logistic enzyme kinetics • *Pimephales promelas* • QSAR • Rats • Toxicity

Contents

1 Introduction

In present-day chemistry, and especially in biological chemistry, there is a move toward replacing preindustrial in vivo experiments (expensive in materials and time) and in vitro experiments (expensive in equipment costs and reproducibility efforts across labs) with *in cerebro* and *in silico* experiments (i.e., evaluation and computation/simulation studies) [1]. The latter experiment types are reasonably inexpensive because they [2]

- Are based on physico-mathematical laws and software implementation
- Provide results on a controlled time scale that are dependent only on the researcher and the computation's theoretical and informatics capabilities
- Are more highly reproducible
- Minimize errors, and most importantly
- Are predictable and reproducible (specific to qualitative and quantitative treatments), considerably reducing the cost and time of validation, especially for newly synthesized compounds, by improving toxicity studies and calculating their impact on biological, ecological, medical, and social environments.

In this context, the quantitative structure–activity relationship (*QSAR*) technique has become the standard European Commission-approved computational/*in silico* method for estimating in vivo/in vitro experiments regarding chemical toxicity for

environmental hazards in the alimentary, pharmaceutical, somatic, and agricultural industries [3–6]. For this aim to be feasible, conceptual and computational methods for studying both the chemical structure and biological activity aspects of a given compound should be used. Researchers should employ chemical structural information that furthers reactivity-to-activity conceptual–computational knowledge. Fortunately, the current chemical indices developed within density functional theory (DFT) respond to the *conceptual* need because they describe chemical reactivity in terms of observable quantities such as total (or valence) energy and number of electrons (see [7–12]):

- Electronegativity, viewed as an instantaneous variation of total (or valence) energy for the equilibrium (neutral) charged system [13–27]

$$\chi \equiv -\left(\frac{\partial E_{\mathrm{N}}}{\partial N}\right)_{V(\mathbf{r})} \tag{1}$$

- Chemical hardness, viewed as the instantaneous electronegativity change around atoms in a molecule [28–45]

$$\eta \equiv -\frac{1}{2}\left(\frac{\partial \chi}{\partial N}\right)_{V(\mathbf{r})} \tag{2}$$

- Chemical power, providing the *dynamic charge* of atoms in a molecule

$$\pi = \frac{\chi}{2\eta} \tag{3}$$

- Electrophilicity, relating the *promotion energy* of atoms in a molecule [46–51]

$$\omega = \frac{\chi^2}{2\eta} \tag{4}$$

Electronegativity (EL) and chemical hardness (HD) comprise an orthogonal 2D chemical space in which chemical reactivity can be described analogous to physical phenomena in the kinetic space of velocity and acceleration

$$v = \partial r/\partial t \leftrightarrow \partial E/\partial N = -\chi \tag{5}$$

$$a = \partial^2 r/\partial t^2 \leftrightarrow \partial^2 E/\partial N^2 = 2\eta \tag{6}$$

while recognizing the phenomenological equivalence

$$E(\text{valence energy}) \leftrightarrow r(\text{coordinate}) \tag{7}$$

$$N(\text{no. valence e}^-) \leftrightarrow t(\text{time}) \tag{8}$$

Such conceptual orthogonality is apparent when the chemical principles of EL and HD are discussed in the finite difference (gauge ionic species) framework in Sect. 2. It is worth noting that the numerical evaluation of χ, η, ω, and π actually depends on the computational DFT framework of the various energetic states computed. In this way, the *computational* approach is utilized for the chemical reactivity indices aspect, and it will be surveyed in Sect. 3.

The chemical–biological species interaction, on the other hand, currently requires *further conceptual* tools for modeling. The present approach parallels the substrate–enzyme reaction for the consecrated Michaelis–Menten mechanism [52–62] to provide the ligand-receptor (L–R) logistic kinetics in general and the ligand progress curve $L(t)$ in particular, depending on the 50 %-effective concentration (EC_{50}) dose for the recorded activity (A) for a given species:

$$A = \ln\left(\frac{1}{EC_{50}}\right). \tag{9}$$

Because $L(t)$ usually also depends on kinetic parameters such as the initial chemical concentration L_0 and the maximum biological uptake β_{max}, eventually under the natural exponential form, i.e., the nuclear radioactivity equation–here adapted as

$$L(t) \cong L_0 \exp\left(-\frac{\beta_{max} t}{EC_{50}}\right) \tag{10}$$

its complete unfolding requires *further computational* ligand–receptor activity modeling. This may be performed by using (QSAR) methodology [63–74], which can be achieved by correlating the observed activity with the above reactivity indices within multilinear regressions yielding computed or predicted activity (A^*). However, for reasons that will be immediately revealed, only linear equations will be considered:

$$A^* = A^*(\chi \vee \eta \vee \pi \vee \omega). \tag{11}$$

Following (9), one can immediately make the correspondence between the initial ligand concentration and computed activity

$$L_0 = \exp(-\|A^*\|) \tag{12a}$$

with $\|\bullet\|$ accounting for the algebraic (Banach) norm in the chemical space (with the dimension equal to the cardinal of the set of chemicals considered in the QSAR), while for the EC_{50} parameter, a similar relationship holds at the level of recorded activities, namely

$$EC_{50} = \exp(-\|A\|). \tag{12b}$$

For maximum biological uptake (via interaction with a ligand), one can employ the working definition

$$\beta = -\frac{\mathrm{d}}{\mathrm{d}t}L(t) \to -\frac{L_t - L_0}{t}. \tag{13}$$

With complete consumption of the ligand ($L_t \to 0$), a maximum value is achieved.

$$\beta_{\max} = \frac{L_0}{\Delta t_\infty}. \tag{14}$$

The "infinite" time interval may be shaped by considering the re-scaling first,

$$t = \mathrm{e}^{\frac{1}{1-\tau}} - \mathrm{e} \to \begin{cases} 0 \ldots \tau \to 0 \\ \infty \ldots \tau \to 1 \end{cases} \tag{15}$$

followed by differentiation

$$\mathrm{d}t = \frac{\mathrm{d}\tau}{(1-\tau)^2} \exp\left(\frac{1}{1-\tau}\right). \tag{16}$$

For practical considerations, the interval $(0, \infty)$ projected into $(0, 1)$ can safely use the setting $\tau = 1/2$ as a sufficient condition for maximal biological uptake, so that the associated working time interval (the so-called "*receptor time*") is $\Delta t_\infty = dt(\tau = 1/2, \Delta\tau = 1) = 4e^2$, using Eqs. (12a, 12b) and (14) as the working parameters:

$$\beta_{\max} = \frac{1}{4}\exp(-\|A^*\| - 2) = \frac{L_0}{4e^2}. \tag{17}$$

The working ligand progress curve (10) directly depends on recorded and computed activity, which in turn depend on the chemical reactivity indices considered, employing what can be called *quantitative reactivity–activity relationships* (QRAR). For each DFT framework and each reactivity index, a different progress curve for a given ligand–receptor interaction is obtained. The immediate inference can be made that faster consumption of $L(t)$ is involved for the more reactive index, and therefore, a more preeminent reactivity principle is associated. This allows the researcher

- To formulate chemical–biological interactions with the help of conceptual–computational reactivity indices of electronegativity, chemical hardness, electrophilicity, and chemical power.

- To obtain a hierarchy of the allied chemical principles (electronegativity, chemical hardness, electrophilicity, and chemical power) for a given pool of molecules of certain species and to check whether they are maintained across many species.

This algorithm is schematically presented in Fig. 1. It will be detailed in the next sections by reviewing the basic conceptual and computational concepts invoked here and then illustrated by modeling chlorinated polycyclic aromatic hydrocarbons (Cl-PAHs) on various species (*Pimephales promelas* and *Rattus norvegicus*).

2 Survey of Chemical Reactivity Principles

2.1 Electronegativity and Chemical Hardness Principles

Because the principles of quantum mechanics do not suggest any operator whose eigenvalue is electronegativity, years after Pauling's electronegativity, several definitions and interpretations of electronegativity have been formulated [13–27]. One of most preeminent of these was created by Mulliken in 1934 as the average of the ionization potential (*IP*) and electron affinity (*EA*) for the valence state of an atom [14]. This empirical spectroscopic definition dominated chemistry for almost a half-century until its quantitative counterpart was introduced by Parr [7–10, 18, 19] as the additive inverse of the chemical potential of a multi-electronic system. The link between these two definitions is clear if the finite difference approximation is performed on the ground state energy, E_N, around the reference integer total number of electrons N_0 in Eq. (1) to obtain [7–10]:

$$\chi \equiv -\left(\frac{\partial E_N}{\partial N}\right)_{V(\mathbf{r})} \cong -\frac{E_{N_0+1}-E_{N_0-1}}{2} = \frac{\left(E_{N_0-1}-E_{N_0}\right)+\left(E_{N_0}-E_{N_0+1}\right)}{2}$$
$$= \frac{\mathrm{IP}+\mathrm{EA}}{2} \cong -\frac{\varepsilon_{\mathrm{LUMO}}+\varepsilon_{\mathrm{HOMO}}}{2}. \tag{18}$$

The problem with equivalent EN forms in Eq. (18) is with the mixed potential conditions that they imply. Because the EN definition on the left side of the chain equation (18) involves *ground state* energy in a nonzero constant potential $V(\mathbf{r})$, it assumes almost vertical values for the energies when electrons change with environment. On the other hand, the values at the right extreme of Eq. (18), the so-called finite-difference EN, correspond to the *valence state* and are thus characterized by the almost adiabatic case $V(\mathbf{r}) = 0$ because no further electrons are attached to the system. The intriguing problem of treating chemical reactivity by indices defined with ground state vs. valence state has continuously challenged the conceptual DFT community. There is not always clear insight with respect to the chemical phenomenology covered or the viewpoint that requires such clarification, especially with respect to equalization reactivity principles (starting from electronegativity

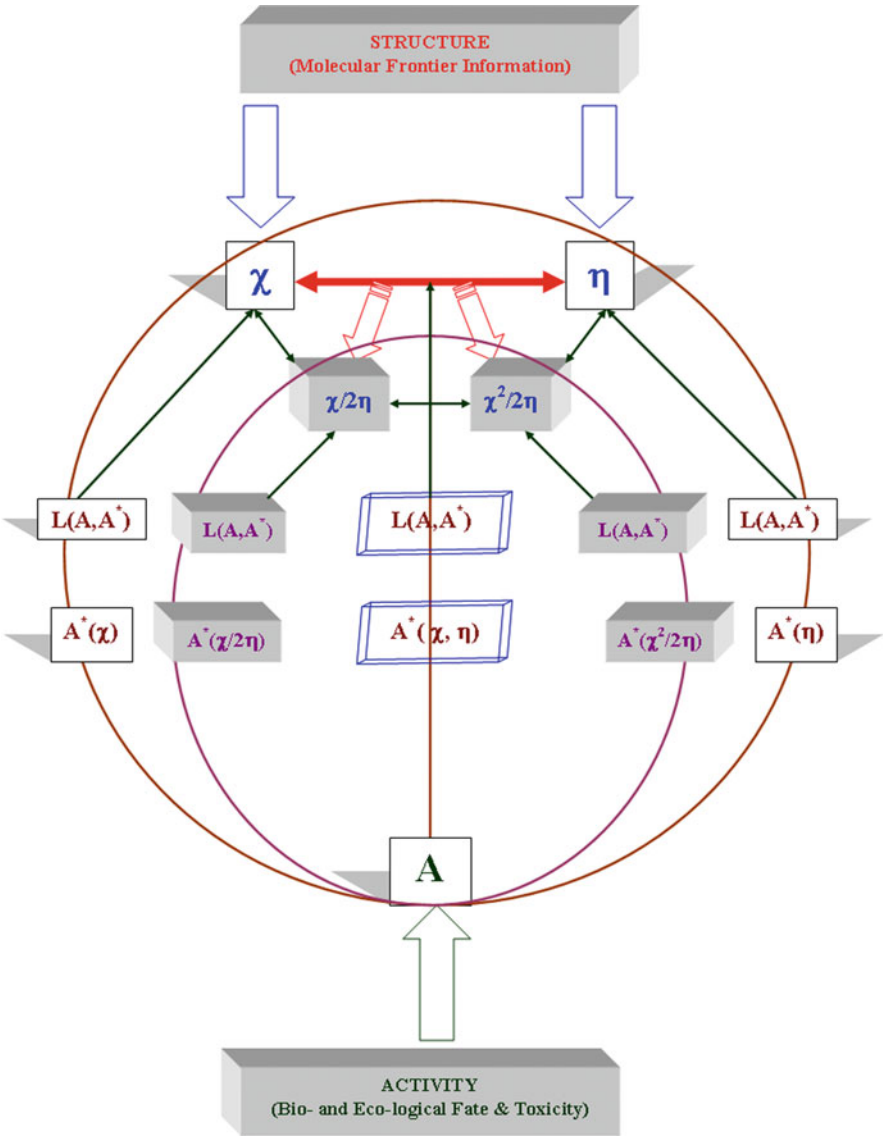


Fig. 1 The workflow of structure–activity relationships involving conceptual and computational density functional theory through molecular frontier information (e.g., HOMO and LUMO), which are primarily employed as electronegativity (χ) and chemical hardness (η) indices and are then combined to form chemical power ($\chi/2\eta$) and electrophilicity ($\chi^2/2\eta$) reactivity measures. These values are correlated with observed biological and ecological activity (A) to provide the QSAR models (A^*) to finally produce ligand progress curves (L) that provide a hierarchy of chemical reactivity principles involved in biological activity within a given DFT computational framework and species of interest

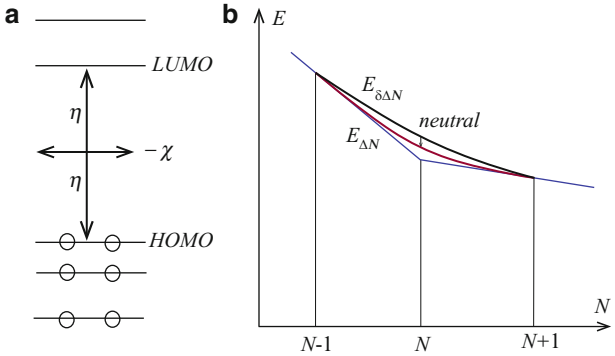


Fig. 2 (**a**) Orbital energy diagram for a molecule [30, 44] showing electronegativity and hardness according to Eqs. (18) and (19), on which basis the equalization of electronegativity and hardness principles, EE and HSAB, follow (see Table 1). (**b**) Plot of the electronic energy vs. electrons for a molecule, for which the electronegativity inequality and maximum hardness principles of chemical reactivity, EL and MH, follow from the parabola of energies of the *valence* states (see Table 1)

equalization principles through to charge transfer between adjacent atoms in forming a bond). However, assuming electronegativity is well defined, its natural companion, chemical hardness, can be immediately introduced and developed in the same finite-difference way, as abstracted from Eq. (2) [28]

$$\eta \equiv -\frac{1}{2}\left(\frac{\partial \chi}{\partial N}\right)_{V(\mathbf{r})} = \frac{1}{2}\left(\frac{\partial^2 E_N}{\partial N^2}\right)_{V(\mathbf{r})} \cong \frac{E_{N_0+1} - 2E_{N_0} + E_{N_0-1}}{2} = \frac{\mathrm{IP} - \mathrm{EA}}{2} \cong \frac{\varepsilon_{\mathrm{LUMO}} - \varepsilon_{\mathrm{HOMO}}}{2}. \tag{19}$$

Like electronegativity [75–80], chemical hardness supports a second-order equalization principle that equalizes the frontier orbitals such that charge transfer is equilibrated between acids and bases in bonding [29, 33, 35, 37]. How this qualitative description of hard-and-soft acids-and-bases (HSAB) [81–87] yields global maximum hardness (MH) is still under debate [88–90], especially when the relationship to electronegativity is invoked. Note that the factor "1/2" in Eq. (19), although considered optional [11, 91], is rooted in parabolic total/valence energy expansion in terms of electronegativity and chemical hardness and describes the fractional quantum distribution of charges in bonding. Only upon averaging do the two adducts with opposite reactivity perspectives compute to unity, as in the classical picture of electronic transfer. Note that the same quantum undulatory information is present (by the appearance of factor "1/2") in the electrophilicity and chemical power definitions (3) and (4).

In order to form a more intuitive (although not complete) idea of how the above-nominated reactivity principles act at the energetic level, Fig. 2 depicts the equality and inequality variants in the left (a) and right (b) parts, respectively. In Fig. 2a, there is a clear difference during a reaction or bonding process in the electronegativities of the partners first encountered: the main effect of reactivity will be the

adjustment of the middle of the HOMO-LUMO gap so that an easier frontier transfer of the electrons is realized. This behavior illustrates the EE principle on a relative energetic scale. Hard molecules are characterized by a large HOMO-LUMO gap, and soft molecules are characterized by a small gap. Therefore, as a second effect of reactivity, hard–hard and soft–soft interactions are favored over hard–soft and soft–hard interactions because the exchange of the charge induced by the ionic character of bonding is much more easily performed through the first two cases due to nearby energies of the frontier orbitals. This justifies the HSAB principle on a relative energetic scale. Instead, from Fig. 2b, one determines the energetic curve of interaction with the paradigmatic parabolic form, whose justification is based on electronegativity/chemical potential equalization. While approaching the optimum reaction path/parabola achieves a minimum for the neutral collection of nuclei, the electronegativity (slope) at that point should be at a minimum and chemical hardness (the curvature) should be at a maximum to assure the stability of the system in the ground state. A summary of these phenomenological considerations is presented in Table 1, which allows for an understanding of chemical reactivity or chemical bonding as the succession of equal and unequal realizations of chemical indices involved in atoms-in-molecules modeling.

Nothing is specified regarding electrophilicity or chemical power equalization for atoms-in-molecules. Because their basic indices in (3) and (4) are defined as a combination of electronegativity and chemical hardness, it is unclear how to a priori combine their associated principles, i.e., equalization of EL with equalization of HD is one option from Table 1 but not the only one, and the same is true for equalization of EL with the maximum of HD, etc. Therefore, this work explores in an a posteriori (applicative) way the degree to which the electrophilicity principle, for instance, derives (via QRAR modeling technique) from electronegativity- and chemical hardness-based reactivity–activity results. In the systematic affirmative cases (for many species, computation frameworks of structures, etc.), electrophilicity cannot be assessed as carrying the new equalization principle because it is a natural consequence of the EL and HD equalizations (most preeminent in driving chemical reactivity/bonding to its equilibrium with atoms-in-molecules; in all other cases, electrophilicity may provide new insight as an equal stabilization energy for atoms in a molecule). It remains to be seen whether there is an intermediate stage either between the EL and HD actions, or before both of them, in which case it will play a unifying role for the allied principles of electronegativity and chemical hardness, as recently challenged [47–51].

2.2 On the Necessity of the Charge-Parabola for Chemical Energy

Here, we explore the reciprocal case usually treated in chemical reactivity, namely, to derive the energetic shape for chemical valence behavior starting from the equalization of chemical potentials of atoms in molecules [92–95],

Table 1 Synopsis of the basic principles of reactivity at the chemical potential (or electronegativity) and chemical force (or hardness) levels, with both equalization and inequality stages on valence states of atomic, pro-molecule, and atoms-in-molecule systems

Reactivity/ bonding index	General principle	Special principle	Principle of chemical reactivity/bonding
$\mu = -\chi$	$\Delta\mu \geq 0$	$\Delta\mu = 0$	*Chemical potential (or electronegativity) equality (EE)*: "the chemical potential of all constituent atoms in a bond or molecule have the same value" [76]
		$\Delta\mu > 0$	*Chemical potential (or electronegativity) inequality (EI)*: "the constancy of the chemical potential is perturbed by the electrons of bonds bringing about a finite difference in regional chemical potential even after chemical equilibrium is attained globally" [77]
η	$\Delta\eta \geq 0$	$\Delta\eta = 0$	*Hard-and-soft acids and bases (HSAB)*: "hard likes hard and soft likes soft" [81–83]
		$\Delta\eta > 0$	*Maximum hardness (MH)*: "molecules arranges themselves as to be as hard as possible" [88, 89]

$$\mu_i = \mu_j = \cdots = \mu, \tag{20}$$

as rooted in a well-established thermodynamic (Gibbs) principle of charge flowing toward equilibrium between two charged regions.

When systems are treated quantum-mechanically, especially within DFT, one may assume that the uni-electronic orbitals $\varphi_i, i = \overline{1,N}$ for the N-electronic system relate to the electronic density via spin–orbital occupancies

$$\rho(\mathrm{r}) = \sum_i^N n_i |\varphi_i(\mathrm{r})|^2, \quad 0 \leq n_i \leq 1, \sum_i n_i = N \tag{21}$$

such that within a continuous limit, one recovers the DFT normalization

$$N = \int \rho(\mathbf{r})\mathrm{d}\mathbf{r} = \int \varphi_{\mathrm{KS}}^*(\mathbf{r})\varphi_{\mathrm{KS}}(\mathbf{r})\mathrm{d}\mathbf{r} \tag{22}$$

involving the Kohn–Sham wave function $\varphi_{\mathrm{KS}}(\mathbf{r})$. It is this wave function that, in equilibrium (dynamic or valence state), corresponds to the global (equalized) chemical potential through the Schrodinger-like equation:

$$\left[-\frac{1}{2}\nabla^2 + V_{\mathrm{eff}}^{\mathrm{DFT}}(\mathbf{r})\right]\varphi_{\mathrm{KS}}(\mathbf{r}) = \mu\varphi_{\mathrm{KS}}(\mathbf{r}). \tag{23}$$

Equation (23) allows for the expression of chemical potential as an eigenvalue of the N-electronic system

$$\mu = \frac{\int \varphi_{KS}^{*}(\mathbf{r})\left[-\frac{1}{2}\nabla^2 + V_{eff}^{DFT}\right]\varphi_{KS}(\mathbf{r})d\mathbf{r}}{\int \varphi_{KS}^{*}(\mathbf{r})\varphi_{KS}(\mathbf{r})d\mathbf{r}} \equiv \frac{\int \varphi_{KS}^{*}(\mathbf{r})\hat{H}\varphi_{KS}(\mathbf{r})d\mathbf{r}}{\int \varphi_{KS}^{*}(\mathbf{r})\varphi_{KS}(\mathbf{r})d\mathbf{r}} = \frac{E}{N} \tag{24}$$

in close agreement with Gibbs (grand canonical) formalization. Until now, we have equated chemical potential compatible with Kohn–Sham DFT with the Gibbs definition. If the energy in Eq. (24) is merely activation or free Gibbs energy, it places this phenomenon in the dynamic (and not really ground-state) framework of chemical reactivity. One may call it chemical energy to differentiate it from physical energy and incorporate the effective potential (the electronic correlation) and activation (valence state) information. This allows for consideration of the chemical force expression, analogous to its classical definition, as the minus of the potential gradient while modeling the driving of charge transfer along the reaction path [92–95]:

$$F_{\mu} = -\frac{d}{dN}(\mu) = \frac{E}{N^2} - \frac{1}{N}\left(\frac{dE}{dN}\right)_{V}. \tag{25}$$

At this point, by recognizing the electronegativity derivative definition of Eq. (1) as the additive inverse of chemical potential

$$\mu = -\chi = \left(\frac{dE}{dN}\right)_{V} \tag{26}$$

and also with chemical force

$$F_{\chi} = -F_{\mu} = -\frac{d}{dN}(\chi) = \eta^{*} \tag{27}$$

one immediately arrives at the working equation ($E^{*}=-E$)

$$E^{*} = -\chi N + \eta^{*}N^2 \tag{28}$$

eventually supporting the energy-charge variation form

$$\Delta E = -\chi\Delta N + \eta^{*}(\Delta N)^2. \tag{29}$$

Note that Eq. (29) differs from conceptual DFT

$$E[N+\Delta N] = E_0[N] + \left(\frac{\partial E}{\partial N}\right)_{V}\Delta N + \frac{1}{2}\left(\frac{\partial^2 E}{\partial N^2}\right)_{V}(\Delta N)^2 + \cdots \tag{30}$$

by a factor of "1/2" that comes from the undulatory nature of the charges themselves, a feature not taken into account when considering the classical-to-quantum analogy in Eq. (27). The correct classical-to-quantum correspondence should

account for the factor "1/2" on the quantum side due to the electronic self-interaction that should be smeared out in analytical considerations.

In this context, the interaction energy $E_{\Delta N}$ of an electronic system that has a change in charge ΔN assumes the paradigmatic parabolic analytical form:

$$E_{\Delta N} = E_0 + \mu_1 \Delta N + \frac{1}{2}\eta_1 (\Delta N)^2. \tag{31}$$

Let us now assume that energy in Eq. (31) is associated with the minimum perturbation to produce the chemical reaction or molecular (trans-)formation. If further perturbation is considered with respect to the change in charge ($\delta\Delta N$ with the small quantity $\delta \in [0, 1]$), another reaction path unfolds through expansion [44]:

$$E_{\delta\Delta N} = E_{0/v} + \mu_1 \delta\Delta N + \frac{1}{2}\eta_1 (\delta\Delta N)^2 \equiv E_{0/v} + \mu_2 \Delta N + \frac{1}{2}\eta_2 (\Delta N)^2, \tag{32}$$

where $\mu_2 = \mu_1 \delta$ and $\eta_2 = \eta_1 \delta^2$ are the new driving chemical potential and force, respectively. Searching for the naturally selected conditions for optimum reactivity, we can immediately observe that as the variational principle is applied to chemical reactivity, it demands a restoring path from interaction energy [Eq. (32)] back to [Eq. (31)] (see also Fig. 2b)

$$E_{\delta\Delta N} \rightarrow E_{\Delta N} \rightarrow \text{minimum} \tag{33}$$

as the differences in their slope and curvature achieve maximum values

$$\Delta\mu = \Delta\mu_{\delta\Delta N \rightarrow \Delta N} = \mu_1 - \mu_2 = \mu_1 (1 - \delta) \geq 0, \tag{34}$$

$$\Delta\eta = \Delta\eta_{\delta\Delta N \rightarrow \Delta N} = \eta_1 - \eta_2 = \eta_1 \left(1 - \delta^2\right) \geq 0 \tag{35}$$

with the specialization and chemical reactivity rules listed in Table 1. The major conclusion is that all principles of chemical reactivity, especially those of chemical hardness, are rooted in the chemical potential equalization principle, though they are not necessarily identical nor do they describe the same stage of bonding. This becomes clear when one translates Table 1 into chemical bonding scenarios through the following four stages [45]:

1. *Encountering stage*: dominated by the difference in electronegativity between reactants and consumed when the electronegativity equalization principle is fulfilled among all product constituents. This stage is associated with charge flow from the more electronegative regions to the lower electronegativity regions in a molecular formation covering the *covalent* binding step.
2. *Promolecule*: even after chemical equilibrium is attained globally, the electrons involved in bonds act as foreign objects between pairs of regions at any level of the molecular partitioning procedure, inducing the appearance of a finite

difference in adjacent electronegativity of neighbor regions in the molecule. This appearance is due to quantum fluctuations associated with the quantum nature of the bonding electrons and corresponds to ionic bond character.

3. *Atoms-in-molecule*: the induced ionic character of bonds is partially compensated by chemical forces through hardness equalization between the pair regions in a molecule. HSAB principles are therefore involved as a second-order effect in charge transfer. See expansion (30) for an example driven by ionic interactions through bonds.
4. *Ω-molecule*: quantum fluctuations provide a further amount of finite difference, this time in attained global hardness that is transposed onto relaxation effects among the nuclear and electronic distributions so that the remaining unsaturated chemical forces are dispersed by stabilization of the molecular structure.

In this way, we have conceptually surveyed how chemical reactivity in general and chemical bonding in particular may be modeled through a minimum set of electronic indices of electronegativity and chemical hardness, which are complementary in their role in promoting and stabilizing electronic structures in chemical combinations.

3 Survey on Computational Quantum Chemistry

3.1 Quantum Chemical Theories and Approximations [45]

Quantum chemistry originates in five levels of quantum approximations imposed on many-electron-many-nuclei systems, either in isolated or interacting states. They are summarized below with mentions of the current limitations, controversies and prospects.

1. *The Born–Oppenheimer approximation* [96], created to simplify the electronic calculus for frozen nuclei approximation, breaks down when computing, for instance, the magnetic dipole moment and its derivative with respect to the nuclear velocities or momenta for assessing the molecular properties of surfaces [97].
2. *The single Slater determinant representation of the ground electronic state* [98] elegantly solves the exchange behavior of electrons by incorporating Pauli repulsion in antisymmetric determinants [99] and was conceptually extended to the configuration interaction by the seminal works of Löwdin [100]. Further generalization was later noted in the valuable concept of the so-called "complete active space" (CAS) [101, 102] that, when combined with other quantum chemical methods such as the self-consistent field (SCF) and density functional theory (DFT), becomes very productive in accounting for *all* electronic states that contribute to the *reactive* space, whether they are electronic states of species

(reactants, intermediates, and products) that are involved in chemical reactions or thermally or photo-induced states [103–105].

3. *Simple Hückel* [106, 107] *and molecular orbital theories* [108–115], which are viewed as the next natural step over the paradigmatic Heitler–London theory of homopolar chemical bonding [116], have been making possible the development of self-consistent field Hartree–Fock–Slater theories [117–122] and associated semi-empirical formulations [123–126] for treating a plethora of chemical systems and phenomena on the basis of their internal symmetry, while remarkably agreeing with (and sometimes predicting) observed spectra and reactivity. Pericyclic reactions [127, 128] and the Woodward–Hoffman rules [129, 130] are eminent examples. This method is met, however, with the so-called quantum correlation problem (i.e., modeling the electronic movement in the dynamical field of the other electrons present in the system) that remains intractable within the Slater (or even with configurational interaction) framework. The solution to obtaining accurate correlations arrives with the advent of Density Functional Theory (DFT) but with the price of modifying the overall wave-function of the system and its spectra (see next).
4. *Thomas–Fermi* [131–135] *and Hohenberg–Kohn–Sham* [136, 137] *theories*, which merged into the celebrated Density Functional Theory [7–9, 11, 12, 138–144], are conceptually exact, i.e., they perform ab initio analysis of the electronic spectra relying only upon universal constants of electronic charge, mass, and Planck constant. Their combinations of bare and effective potential, while providing an approximate set of orbitals (called Kohn-Sham orbitals) correctly resemble the observed electronic density of the system along the measured energies (with correlation effects included) but with less significance than the classical wave-function concept. In fact, the current DFT uses the so-called basis function like a mathematical tool that can be adapted or optimized depending on the accuracy needed in relation to the optimized effective potential [145], adding dispersion effects, etc. The computational implementation of DFT becomes a parameterized procedure that makes it a sort of semi-empirical DFT [146] that can be extended to include time-dependency excited state effects [147–152] or even model Bose–Einstein condensates [12]. The great merit and paradox of DFT is that the theory provides the recipe to compute two- or multi-body interactions as exchange and correlations, respectively, by approximations to single body (density) behavior. From a theoretical physics point of view, the picture is flawed, yet the approximation works very well. This landmark achievement was perhaps awarded the Nobel Prize in Chemistry (in 1998) rather than Physics because it involves no new physics but represents a useful reformulation for Chemistry. Recent DFT work further reduces the limits of DFT, formulating various approximations for exchange-correlation functionals [92] that are more or less in agreement with the fundamental theorems and limits at asymptotic and nuclei ranges [153]; this holds until attempting to formulate expectation values of various physical observables based only on density, similar to those based on the expectation value of quantum mechanical wave functions [154]. True *ab initio DFT* is far

from being fully engaged [155, 156]; *semi-empirical DFT* seems to prevail in computational implementation. Still, conceptual DFT [157] is of upmost importance in formulating chemical reactivity and the indices that help in understanding and modeling the chemical systems [11, 91].

5. *Solvent effects* must almost always be taken into account when using quantum chemical treatment to describe the reactivity of chemical systems. The environment, including any strongly interacting solvent molecules (e.g., water molecules in the case of biomolecules, such as amino acids, peptides, nucleic acids and their complexes) must be considered in any modeling study of open chemical systems in order to fully understand and interpret the experimental results, such as vibrational, NMR and electronic spectra, and the chiral analogues [158]. Overall, the interaction of the system with the environment stands in the foreground of quantum theory when predicting additional quantum fluctuations upon the concerned system due to coupling with the media/observer/solvent [159, 160]. Such interactions can be nevertheless implemented by accounting for additional reactions and the stability of the investigated chemical systems, while also having the quantum statistical tools for treating the macro-canonical samples in a correct physical way. DFT is well equipped from its basic definition of density (associated with the total number of electrons in the system) that can be easily extended to include environmental effects [161–163].

Overall, the quantum many-body theory (i.e., the second quantization of fields) and its "chemical" version as DFT may be regarded as the appropriate quantum tools. Quantum chemistry should be used in formulating and implementing the specific indicators of stability and reactivity, among which electronegativity and chemical hardness are the most preeminent and versatile.

3.2 Self-Consistent Mono-Electronic Orbitals' Equations [164]

Following Dirac's quote, once the Schrodinger equation:

$$H\Psi = E\Psi \tag{36}$$

is established "the underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known" [165].

Unfortunately, the molecular spectra based on the eigen-problem (36) are neither directly nor completely solved without specific atoms-in-molecule and/or symmetry constraints and approximation. As such, at the mono-electronic level of approximation, the Schrodinger equation (36) is rewritten under the so-called "independent-electron problem":

$$H_i^{\text{eff}}\psi_i = E_i\psi_i \tag{37}$$

with the aid of effective electron Hamiltonian partitioning:

$$H = \sum_i H_i^{\text{eff}} \tag{38}$$

and the corresponding molecular monoelectronic wave-functions (orbitals) fulfilling the conservation rule of probability:

$$\int \psi_i^2(\mathbf{r})\mathrm{d}\mathbf{r} = 1. \tag{39}$$

When written as a linear combination over the atomic orbitals, the resulting MO-LCAO wave-function:

$$\psi_i = \sum_v C_{vi}\phi_v \tag{40}$$

is replaced in Eq. (37), and following integration over the electronic space, it allows for matrix version of Eq. (37):

$$\left(H^{\text{eff}}\right)(C) = (S)(C)(E). \tag{41}$$

With the diagonal energy-matrix elements as the eigen-solution

$$(E)_{ij} = E_{ij} = E_i\delta_{ij} = \begin{cases} E_i \ldots i = j \\ 0 \ldots i \neq j \end{cases} \tag{42}$$

to be found in terms of the expansion coefficients matrix (C), the matrix of the Hamiltonian elements:

$$H_{\mu v} = \int \phi_\mu H^{\text{eff}} \phi_v \mathrm{d}\tau \tag{43}$$

and the matrix of the (atomic) overlapping integrals:

$$S_{\mu v} = \int \phi_\mu \phi_v d\tau \tag{44}$$

where all indices in Eqs. (42)–(44) refer to matrix elements because the additional reference to the "i" electron was skipped to avoid confusion.

The solution of the matrix equation (41) may be unfolded through the Löwdin orthogonalization procedure [166, 167], involving the diagonalization of the overlap matrix via a given unitary matrix (U), $(U)^+(U) = (1)$, using the following procedure:

$$(s) = (U)^{+}(S)(U), \tag{45}$$

$$\left(s^{-1/2}\right)_{ii} = \left[(s)_{ii}\right]^{-1/2}, \tag{46}$$

$$\left(S^{-1/2}\right) = (U)\left(s^{-1/2}\right)(U)^{+}, \tag{47}$$

$$\left(\left(S^{1/2}\right)(C)\right)^{+}\left(\left(S^{-1/2}\right)(H^{\text{eff}})\left(S^{-1/2}\right)\right)\left(\left(S^{1/2}\right)(C)\right) = (E). \tag{48}$$

However, the solution given by Eq. (48) is based on the form of effective independent-electron Hamiltonians that can be empirically constructed as in Extended Hückel Theory [168]. Such "arbitrariness" can be nevertheless avoided by the so-called *self-consistent field* (SCF), in which the one-electron effective Hamiltonian is considered to depend on the solution of Eq. (40) itself, i.e., the matrix of coefficients (C). The resulting "Hamiltonian" is called the Fock operator, while the associated eigen-problem is the Hartree–Fock equation:

$$F\psi_i = E_i\psi_i. \tag{49}$$

In matrix representation, Eq. (49) looks like:

$$(F((C)))(C) = (S)(C)(E) \tag{50}$$

and may be solved iteratively through a diagonalization procedure starting from an input (C) matrix or from a starting electronic distribution quantified by the density matrix:

$$P_{\mu\nu} = \sum_{i}^{occ} C_{\mu i}C_{i\nu} \tag{51}$$

with a major influence on the Fock matrix elements:

$$F_{\mu\nu} = H_{\mu\nu} + \sum_{\lambda\sigma} P_{\lambda\sigma}\left[(\mu\nu|\lambda\sigma) - \frac{1}{2}(\mu\lambda|\nu\sigma)\right]. \tag{52}$$

Note that the one-electron Hamiltonian effective matrix components $H_{\mu\upsilon}$ differ from those of Eq. (43) in what they truly represent. In this form, it represents the kinetic energy plus the interaction of a single electron with the core electrons around all nuclei present. The other integrals appearing in Eq. (52) are generally called two-electron-multi-centers integrals and are written as:

$$(\mu\nu|\lambda\sigma) = \int \phi_\mu^A(\mathbf{r}_1)\phi_\nu^B(\mathbf{r}_1)\frac{1}{r_{12}}\phi_\lambda^C(\mathbf{r}_2)\phi_\sigma^D(\mathbf{r}_2)\mathrm{d}\mathbf{r}_1\mathrm{d}\mathbf{r}_2. \tag{53}$$

From definition (53), there is immediate recognition of the special integral $J = (\mu\mu|\nu\nu)$ as the Coulomb integral describing repulsion between two electrons with probabilities ϕ_μ^2 and ϕ_ν^2.

Moreover, the Hartree–Fock Eqs. (50) implemented in Eqs. (51) and (52) are known as Roothaan equations [109] and constitute the basis for closed-shell (or restricted Hartree–Fock, RHF) molecular orbitals calculations. Their extension to spin effects provides the equations for the open shell (or unrestricted Hartree–Fock, UHF), which are also known as the Pople-Nesbet Unrestricted equations [118].

3.3 *Basics of Ab Initio Methods [164]*

The alternative to semi-empirical methods is the full self-consistent calculation, the so-called ab initio approach. It is based on computing all integrals in Eq. (52), with the atomic Slater-type orbitals (STO), $\exp(-\alpha r)$, replaced by Gaussian-type orbitals (GTO) [169]:

$$\phi_A^{\mathrm{GTO}} = x_A^l y_A^m z_A^n \exp\left(-\alpha r_A^2\right) \tag{54}$$

in molecular orbitals expansion, a procedure allowing for much simplification in multi-center integrals computation. Nevertheless, in turn, each GTO may be generalized to a contracted expression constructed upon the primitive expressions of Eq. (54):

$$\phi_\mu^{\mathrm{CGTO}}(r_A) = \sum_p d_{p\mu}\phi_p^{\mathrm{GTO}}\left(\alpha_p, r_A\right) \tag{55}$$

where α_A and $d_{p\mu}$ are the exponents and the contraction coefficients of the primitives, respectively. Note that the primitive Gaussians involved may be chosen as approximate Slater functions [170], Hartree–Fock atomic orbitals [171], or any other set of functions desired to speed up computation. Under these conditions, a minimal basis set may be constructed with one function for H and He, five functions for Li and Ne, nine functions for Na and Ar, 13 functions for K and Ca, 18 functions for Sc to Kr, etc., to describe the core and valence occupancies of atoms [172–174]. Although such a basis does not generally provide accurate results because of its small cardinal, it contains the essential information regarding the chemical bond and may be useful for qualitative studies, as is the present case for aromaticity scales where the comparative trend is studied.

3.4 Hartree–Fock Implementation [164]

When simple ab initio methods are referred to, it means that the Hartree–Fock equation (49) with full Fock matrix elements [117, 175–177] of Eqs. (51) and (52) is solved for a Gaussian-contracted basis [Eq. (55)]. In fact, the method iteratively evaluates the kinetic energy and nuclear–electron attraction energy integrals for the effective Hamiltonian along the overlap and electron–electron repulsion energy integrals (for both the Coulomb and exchange terms), which are written as:

$$T_{\mu\nu} = \langle \mu | \left(-\frac{1}{2} \nabla^2 \right) | \nu \rangle, \tag{56}$$

$$V_{\mu\nu} = \langle \mu | \frac{Z_A}{r_A} | \nu \rangle, \tag{57}$$

$$S_{\mu\nu} = \langle \mu | \nu \rangle, \tag{58}$$

$$(\mu\nu|\lambda\sigma) = \left(\mu\nu | \frac{1}{r_{12}} | \lambda\sigma \right), \tag{59}$$

until consistency in the electronic population of Eq. (51) between two consecutive steps is achieved.

Note that such calculations assume the total wave function as a single Slater determinant, while the resultant molecular orbital is described as a linear combination of the atomic orbital basis functions (MO-LCAO). Multiple Slater determinants in MO description project the configurational and post-HF methods and will not be discussed here.

3.5 Density Functional Theory Implementation [164]

The main weakness of the Hartree–Fock method, namely the lack in correlation energy, is ingeniously restored by the Density Functional method through the introduction of the so-called "effective one-electron exchange-correlation potential." Unfortunately, this means that the analytical form will not be determined. However, the working equations have the simplicity of the HF versions, while replacing the exchange term in Eq. (52) by the exchange-correlation ("XC") contribution; this results in the (general) unrestricted matrix form of the Kohn–Sham equations [137]:

$$F^{\uparrow}_{\mu\nu} = H^{\uparrow}_{\mu\nu} + \sum_{\lambda\sigma} P^{T}_{\lambda\sigma} (\mu\nu|\lambda\sigma) + F^{\mathrm{XC}\uparrow}_{\mu\nu}, \tag{60}$$

$$F_{\mu\nu}^{\downarrow} = H_{\mu\nu}^{\downarrow} + \sum_{\lambda\sigma} P_{\lambda\sigma}^{T}(\mu\nu|\lambda\sigma) + F_{\mu\nu}^{\mathrm{XC}\downarrow}, \tag{61}$$

$$P^{T} \equiv P^{\uparrow+\downarrow} = P^{\uparrow} + P^{\downarrow} \tag{62}$$

in a similar fashion as with the Pople–Nesbet equations of Hartree–Fock theory. The restricted (closed-shell) variant resembles the density constraint:

$$\rho^{\uparrow} = \rho^{\downarrow} \tag{63}$$

in which case the Roothaan analogous equations (for exchange-correlation potential) are obtained.

Either Eq. (60) or (61) fulfills the general matrix equation of type (50) for the energy solution

$$E = \sum_{\mu\nu} P_{\mu\nu} H_{\mu\nu} + \frac{1}{2} \sum_{\mu\nu\lambda\sigma} P_{\mu\nu} P_{\lambda\sigma} (\mu\nu|\lambda\sigma) + E_{\mathrm{XC}}, \tag{64}$$

which can be regarded as the solution of the Kohn–Sham equations. The apparent exchange-correlation energy E_{XC} may be conveniently expressed through energy density (per unit volume) using the integral formulation:

$$E_{\mathrm{XC}} = E_{\mathrm{XC}}\left[\rho^{\uparrow}, \rho^{\downarrow}\right] = \int f\left(\rho^{\uparrow}, \rho^{\downarrow}\right) \mathrm{d}\tau \tag{65}$$

once the Fock elements of exchange-correlation are recognized to be of a density gradient form [178]:

$$F_{\mu\nu}^{\mathrm{XC}\uparrow(\downarrow)} = \int \frac{\partial f}{\partial \rho^{\uparrow(\downarrow)}} \phi_{\mu} \phi_{\nu} \mathrm{d}\tau. \tag{66}$$

The quest for various approximations for the exchange-correlation energy density $f(\rho)$ has spanned decades in quantum chemistry and was recently reviewed [92]. Here, we will present the "red line" of its implementation, as it will be further used for the current applications. The benchmark density functional stands as the Slater exchange approximation, derived within the so-called "Xα theory" [179]:

$$f^{\mathrm{X}\alpha} = -\frac{9}{4}\alpha\left(\frac{3}{4\pi}\right)^{1/3}\left(\rho^{\uparrow 4/3} + \rho^{\downarrow 4/3}\right) \tag{67}$$

with α taking the values:

$$\alpha = \begin{cases} 1 \ldots \text{Slater} \\ 2/3 \ldots \text{uniform electronic gas} \end{cases} \tag{68}$$

With Eq. (67) in Eq. (66), the resulting Kohn–Sham "exchange-correlation" matrix elements (although rooted only in the exchange) yield the integral representation:

$$F_{\mu\nu}^{\mathrm{XC}\uparrow} = F_{\mu\nu}^{\mathrm{X}\uparrow} = -\frac{9}{4\pi}\alpha \int \rho^{\uparrow 1/3}\phi_\mu \phi_\nu \mathrm{d}\tau. \tag{69}$$

Considerable improvement of molecular calculation was given by Becke's density gradient correction of the local spin density (or Slater exchange) approximation of the exchange energy [180]:

$$E_{\mathrm{X}} = \sum_{\sigma=\uparrow,\downarrow} \int e_{\mathrm{X}\sigma}^{\mathrm{LSDA}} g_{\mathrm{X}\sigma} \mathrm{d}\tau \tag{70}$$

where

$$g_{\mathrm{X}\sigma}\left(\frac{|\nabla\rho^\sigma|}{\rho^{\sigma 4/3}}\right) = g_{\mathrm{X}\sigma}(x) = \begin{cases} 1 \ldots \text{Slater} \\ 1 + \dfrac{bx^2}{a\left(1 + 6bx\sinh^{-1}x\right)} \cdots \text{Becke88} \end{cases} \tag{71}$$

with the parameters $a = (3/2)(3/4\pi)^{1/3}$ and $b = 0.0042$ chosen to fit the experiment. Other exchange functionals were developed along the same line, having different realization of the gradient functions [Eq. (71)], most notably those of Perdew and collaborators (e.g., Perdew-Wang-91, PW91) [181].

The correlation contribution was developed using a somewhat different algorithm, namely, employing its definition as the difference between the exact and Hartree–Fock (HF) total energy of a poly-electronic system [182]. Without reproducing the results (more details are provided in the dedicated review of [92]), we mention only the Lee–Yang–Parr (LYP) correlation functional [183–185] along the Vosko–Wilk–Nusair (VWN) local correlation density functional [186].

The exchange and correlation density functionals combine into so-called "hybrid functionals." Those used in the present study refer to

- Becke-97: a hybrid exchange-correlation functional created by extending the $g(x)$ of Equation (71) as a power series containing three terms with an admixture of 19.43 % HF exchange [187].
- Becke88-VWN: combines Becke88 exchange of Eq. (71) with VWN correlation.
- B3-PW91: also developed by Becke with PW91 correlation instead of LYP.
- B3-LYP: advanced by Becke by empirical comparisons against very accurate results; it contains the exchange contribution (20 % HF + 80 % Slater + 72 % Becke88) added to the correlation energy (81 % LYP + 19 % VWN) [188].

These are the main methods, at both conceptual and computational levels, used to assess the structure of molecules involved as ligands in various chemical–biological interactions.

4 Logistic Modeling of Chemical–Biological Interactions

In the post-genomic era, the development of kinetic models that allow simulation of complicated metabolic pathways and protein interactions is becoming increasingly important [189, 190]. Unfortunately, the difference between an in vivo biological system and homogeneous in vitro conditions is large, as shown by Schnell and Turner [191]. Mathematical treatments of biochemical kinetics have been developed from the law of mass action in vitro, but the modifications required to bring them in line with stochastic in vivo situations are still under development [192–194].

The mechanism of the biological activity produced by a substance usually involves the combination of interactions between the molecules of that substance, called the effector or ligand (L), with a receptor (R), a protein, a biologically macromolecule, or a complex of macromolecules within the cell. The intensity of the biological action is illustrated in an ordinary way as the logarithm of the inverse of the concentration C to produce a specific biological answer (see Eq. (9)). Because the C_{50} concentration (the molar concentration that produces 50 % of the maximum biological activity) is used in many situations, it can be shown that the biological activity A is proportional to the affinity of the molecule or ligand L_i for the receptor R, which is at the basis of the explicated biological action. However, the most rational hypothesis concerning the mode of action of bioactive substances presumes that biological activity produced by a ligand L is proportional to the complexation degree of the receptor R with L. In this situation, the presumed biological activity of α% (compared to a maximum of 100 %) is produced by the concentration $[L]$ of ligand, which closely agrees with substrate–enzyme kinetics modeling (see Table 2).

Therefore, we may safely generalize the chemical–biological kinetics in terms of the general rate of biological uptake β while respecting the chemical concentration $[L]$

$$\beta = \frac{\beta_{\max}[L]}{[L] + \mathrm{EC}_{50}} \tag{72}$$

while recognizing that the temporal link in Eq. (72) between the biological activity and chemical concentration is represented as

$$\beta = -\frac{\mathrm{d}}{\mathrm{d}t}[L](t). \tag{73}$$

The full temporal version can be widely formulated as [59–62]:

$$-\frac{\mathrm{d}}{\mathrm{d}t}[L](t) = \frac{\beta_{\max}[L](t)}{[L](t) + \mathrm{EC}_{50}}. \tag{74}$$

The main problem with Eq. (74) is that it accounts only for the velocity of the initial time of the reaction. The information outside the first moments of the inherent progress curve is virtually lost or neglected. Another complication of Eq. (74) is that, even when describing a generalized kinetic, it differs from ordinary chemical curves in its rectangular hyperbola shape instead of the expected exponential form. A further generalized kinetic may instead be assumed, which can be applied to the enzymatic Michaelis-Menten case [196–200] as follows.

We use a probabilistic approach [197, 198], based on the law of mass action, to characterize in vitro ligand-receptor interaction as quoted in Table 2:

$$1 = P_{\mathrm{REACT}}([L]_{\mathrm{bind}}) + P_{\mathrm{UNREACT}}([L]_{\mathrm{bind}}). \tag{75}$$

In Eq. (75), $P_{\mathrm{REACT}}([L]_{\mathrm{bind}})$ is the probability that the ligand–receptor interaction of Table 1 proceeds at a certain concentration of ligand binding to the receptor $[L]_{\mathrm{bind}}$. The limits are:

$$P_{\mathrm{REACT}}([L]_{\mathrm{bind}}) = \begin{cases} 0, [L]_{\mathrm{bind}} \to 0 \\ 1, [L]_{\mathrm{bind}} \gg 0. \end{cases} \tag{76}$$

Note that $P_{\mathrm{REACT}}([L]_{\mathrm{bind}}) = 0$ when the enzymatic reaction does not proceed or when it stops because the ligand fails to bind or is entirely metabolized.

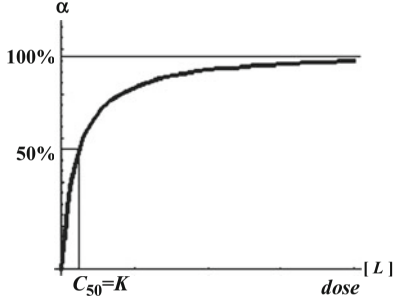
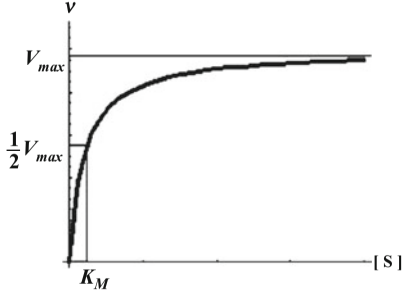
Conversely, $P_{\mathrm{REACT}}([L]_{\mathrm{bind}}) = 1$ when the ligand–receptor interaction proceeds and is related to the standard quasi-steady-states approximation (QSSA). The probability of the occurrence of products in L–R reactions lies between these limits. Similarly, in the case where specific interactions do not take place, $P_{\mathrm{UNREACT}}([L]_{\mathrm{bind}})$, the limits are:

$$P_{\mathrm{UNREACT}}([L]_{\mathrm{bind}}) = \begin{cases} 1, [L]_{\mathrm{bind}} \to 0 \\ 0, [L]_{\mathrm{bind}} \gg 0. \end{cases} \tag{77}$$

This probabilistic treatment of enzyme kinetics is based on the chemical bonding behavior of enzymes that act upon substrate molecules through diverse mechanisms and may offer the key to the quantitative treatment of different types of enzyme catalysis.

To unpack the terms of Eq. (75) to analyze L–R reactions, we first recognize that the bound ligand concentration can be treated as the instantaneous concentration, i.e., $[L]_{\mathrm{bind}} = [L](t)$.

Table 2 The face-to-face ligand–receptor (L–R) and substrate–enzyme (S–E) kinetics [195]

Property	L–R kinetics	S–E kinetics
Reaction	$\mathrm{R}+\mathrm{L}\overset{K}{\leftrightarrow}\mathrm{C}^{*}$	$\mathrm{E}+\mathrm{S}_i\underset{k_{-1}}{\overset{k_1}{\leftrightarrow}}\mathrm{ES}\overset{k_2}{\rightarrow}\mathrm{E}+\mathrm{P}$
Species	$\boldsymbol{R}$	$\boldsymbol{E}$
Chemicals	$\boldsymbol{L}$	$\boldsymbol{S}$
Equation	$\frac{\alpha}{100}=\frac{[L]}{[L]+K}$	$\frac{v}{V_{\max}}=\frac{[S]}{[S]+K_M}$
Constant	$K=\frac{[R][L]}{[C^{*}]}$	$K_M=\frac{k_{-1}+k_2}{k_1}$
Graph		

Maintaining quasi-steady-state conditions for in vitro systems, we may assume constant association–dissociation rates so that probability of interaction/reaction is written as the rate of consumption of the ligand [see Eq. (73)], to saturation:

$$P_{\text{REACT}}([L](t)) = \frac{\beta(t)}{\beta_{\max}} = -\frac{1}{\beta_{\max}} \frac{\mathrm{d}}{\mathrm{d}t}[L](t) \tag{78}$$

after the initial transient of the ligand–receptor adduct-complex interchanges.

We know only that expression (78) behaves like a probability function, with values in the realm [0,1]. Given expressions (75), (78) and the general Michaelis-Menten equation (74), we derive an expression for the unreacted probability term, $P_{\text{UNREACT}}([S](t))$. As such, the expression:

$$P_{\text{UNREACT}}([L](t))^{\text{MM}} = \frac{\text{EC}_{50}}{[L](t) + \text{EC}_{50}} \tag{79}$$

satisfies all probability requirements, including the limits in Eq. (77). When combined with equations (75) and (78), the equation gives the instantaneous version of the classical Michaelis-Menten equation (72). Remarkably, expression (79) can be seen as a generalization of the efficiency of the Michaelis-Menten reaction under steady-state conditions. The efficiency depends on two parameters: EC_{50}, which embodies the toxicological conditions of the ligand–receptor reaction, and the initial ligand concentration $[S_0]$; these determine the ratio of free to total ligand concentration in the L–R interaction. That is, when efficiency is equal to one, we do not expect to find free ligand in the reaction; the L–R reactions are all consumed so that the first branch of the limit (77) is fulfilled and no further binding will occur. It is clear that the Michaelis-Menten term [Eq. (79)] is just a particular choice for a probabilistic ligand kinetic model of the conservation law [Eq. (75)]. However, a more generalized version of equation (79) that preserves all of the above probabilistic features may look like

$$P_{\text{UNREACT}}([L](t))^{*} = \mathrm{e}^{-\frac{[L](t)}{\text{EC}_{50}}} \tag{80}$$

from which the Michaelis-Menten term [Eq. (79)] is returned by performing the $[L](t)$ first-order expansion for the case where the bound ligand approaches zero:

$$P_{\text{UNREACT}}([L](t))^{*} = \frac{1}{\mathrm{e}^{\frac{[L](t)}{\text{EC}_{50}}}} \overset{[L](t)\to 0}{\cong} \frac{1}{1 + \frac{[L](t)}{\text{EC}_{50}}} = P_{\text{UNREACT}}([L](t))^{\text{MM}}. \tag{81}$$

It is worth noting that there is no monotonic form between 0 and 1 other than that of Eq. (80) to reproduce the basic Michaelis-Menten term [Eq. (79)] when approximated for a small $x = [L](t)/K_{\text{M}}$. For instance, if one decides to use $\exp(-x^2)$, the unreactive probability will give $1/(1 + x^2)$ as the approximation for a small x, which is definitely different than expected for basic Michaelis-Menten treatment [Eq. (79)].

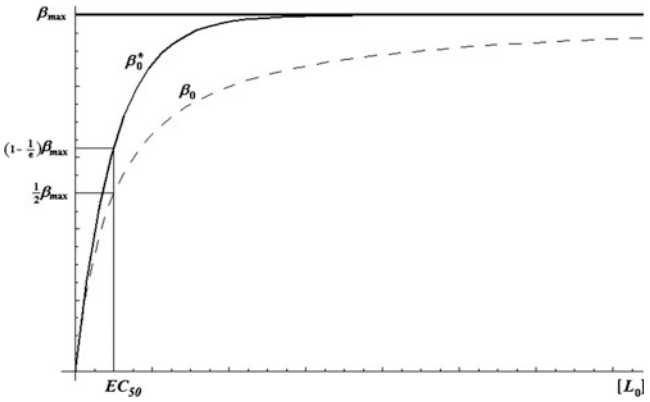


Fig. 3 Initial Michaelis–Menten and logistic velocities plotted against initial ligand concentration for the L–R direct interaction/binding. The *dashed curve* corresponds to the Michaelis–Menten equation (72), while the continuous *thick curve* represents its logistic generalization from (82) (see [198])

The physico-chemical meaning of equation (80) is that the Michaelis-Menten term [Eq. (79)] and its associated kinetics apply to fast ligand–receptor reactions/metabolization, i.e., for fast consumption of $[L](t)$. However, by using Eq. (80) instead of Eq. (79), the range of reaction rates is expanded and provides a new kinetic equation, in the form of the logistic expression

$$-\frac{1}{\beta_{\max}}\frac{\mathrm{d}}{\mathrm{d}t}[L](t) = 1 - \mathrm{e}^{-\frac{[L](t)}{\mathrm{EC}_{50}}} \tag{82}$$

by incorporating Eqs. (75) and (78). Under initial conditions, the logistic equation (82) gives an initial velocity of reaction $\left(\beta_0^*\right)$ that is uniformly higher than that calculated by Michaelis-Menten (79) at all initial concentrations of the ligand, except for the case where $[L_0] \to 0$, when both are zero (see Fig. 3).

To test whether the logistic kinetic equation (82), which is a natural generalization of the Michaelis-Menten equation, may provide a workable analytical solution in an elementary form, we first integrate it under the form

$$\int_{[L_0]}^{[L](t)} \frac{\mathrm{d}[L](t)}{\exp(-[L](t)/\mathrm{EC}_{50}) - 1} = \int_0^t \beta_{\max}\mathrm{d}t \tag{83}$$

generating the new equation to be solved:

$$[L_0] - [L](t) + \mathrm{EC}_{50}\ln\left(\mathrm{e}^{-\frac{[L_0]}{\mathrm{EC}_{50}}} - 1\right) - \mathrm{EC}_{50}\ln\left(\mathrm{e}^{-\frac{[L](t)}{\mathrm{EC}_{50}}} - 1\right) = \beta_{\max}t. \tag{84}$$

This can be solved exactly by substituting

$$\varphi([L](t)) = \frac{[L](t)}{\mathrm{EC}_{50}} \tag{85}$$

into Eq. (84) to obtain the simple equation:

$$-\varphi([L](t)) - \ln\left(\mathrm{e}^{-\varphi([L](t))} - 1\right) = \Psi(t) \tag{86}$$

where we have also introduced the functional notation:

$$\Psi(t) = \frac{1}{\mathrm{EC}_{50}}(\beta_{\max} t - [L_0]) - \ln\left(\mathrm{e}^{-\frac{[L_0]}{\mathrm{EC}_{50}}} - 1\right). \tag{87}$$

Now, the exact solution of Eq. (86) takes the logistic expression:

$$\varphi([L](t)) = \ln\left(1 - \mathrm{e}^{-\Psi(t)}\right). \tag{88}$$

Finally, substituting function (87) into expression (88) gives the logistic progress curve for ligand consumption in an analytically elementary form [198]:

$$[L]_L(t) = \mathrm{EC}_{50} \ln\left(1 + \mathrm{e}^{-\frac{\beta_{\max} t}{\mathrm{EC}_{50}}}\left(\mathrm{e}^{\frac{[L_0]}{\mathrm{EC}_{50}}} - 1\right)\right). \tag{89}$$

This time-dependent solution (89) substitutes an elementary logarithmic dependency for the *W*-Lambert function. It is nevertheless remarkable that the solution of a generalized logistic kinetic version of the Michaelis-Menten instantaneous equation provides an analytically exact solution. It clearly reduces to the above Eq. (74) in the first order expansion of the chemical concentration time evolution with respect to the 50-effect concentration (EC_{50}) observed.

The original chemical–biological-kinetics [Eq. (74)] gave no analytical solution for the actual working kinetics [Eq. (75)] that provides the logistical solution (89) whose reliability was previously tested on various enzyme kinetics mechanisms. This testing produced remarkable results [196, 197, 199, 200] that constituted a trusted background for employing it in the currently envisaged ecotoxicological studies.

Remarkably, rearranging the logistic solution of the chemical–biological interaction [Eq. (89)] under the equivalent form

$$\mathrm{e}^{\frac{[L]_L(t)}{\mathrm{EC}_{50}}} - 1 = \mathrm{e}^{-\frac{\beta_{\max} t}{\mathrm{EC}_{50}}}\left(\mathrm{e}^{\frac{[L_0]}{\mathrm{EC}_{50}}} - 1\right) \tag{90}$$

provides for a relatively higher concentration $\mathrm{EC}_{50} \gg$ (specific to environmental toxicological fate studies) for the working equation employed in Eq. (10) as the basis for advancing quantitative reactivity–activity relationships.

Table 3 The vectorial descriptors in a Spectral-SAR analysis

Activity	Structural predictor variables					
$\lvert A_{OBS(ERVED)}\rangle$	$\lvert X_0\rangle$	$\lvert X_1\rangle$	$\cdots$	$\lvert X_k\rangle$	$\ldots$	$\lvert X_M\rangle$
$A_{1\text{-OBS}}$	1	x_{11}	$\ldots$	x_{1k}	$\ldots$	x_{1M}
$A_{2\text{-OBS}}$	1	x_{21}	$\ldots$	x_{2k}	$\ldots$	x_{2M}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$A_{N\text{-OBS}}$	1	x_{N1}	$\ldots$	x_{Nk}	$\ldots$	x_{NM}

As such, for a *logistic-spectral analysis*, based on the molecular M-data for N-chemical species and toxicity activities of Table 3, the next steps are considered in producing the chemical–biological progress curves according to which chemical species "dissolve" in biological/environmental receptors.

1. For the activities of Table 3, the Spectral-SAR [74, 201, 202] relationships are formulated

$$|A^*\rangle^{\text{ENDPOINT}} = B_0|X_0\rangle + B_1|X_1\rangle + \cdots + B_M|X_M\rangle \tag{91}$$

for each envisaged molecular-set or model of structural parameters or computational framework.
2. The predicted spectral norm is computed in the N-chemical space

$$\||A^*\rangle\| = \sqrt{\langle A^*|A^*\rangle} = \sqrt{\sum_{k=0}^{M} B_k^2 \langle X_k|X_k\rangle} = \sqrt{\sum_{k=0}^{M} B_k^2 \left(\sum_{j=1}^{N} x_{kj}^2\right)}$$
$$= \sqrt{\sum_{j=1}^{N} \left(A_j^*\right)^2}. \tag{92}$$

3. The initial chemical concentration within the logistical chemical–biological progress picture is related to the predicted S-SAR activity norms (92), according to Eq. (12a, 12b).
4. In the same mechanistic line of the chemical–biological interaction framework, the maximum biological effect is seen as the decrease of the initial chemical concentration in the effector time [see Eq. (17)].
5. Along the EC_{50} parameter computed following the generalization of Eq. (9) toward the algebraic version Eq. (12b), one has all the "ingredients" for progress curves that represent the logistical consumption/metabolization/fate of the chemical species following interaction with the biological species/ecological environment; these are written in the working form, first with the aim of Eqs. (17) and (89), as

$$[L]_L(t) = \text{EC}_{50} \ln\left(1 + \text{e}^{-\frac{1}{4e^2}\frac{[L_0]}{\text{EC}_{50}}t}\left(\text{e}^{\frac{[L_0]}{\text{EC}_{50}}} - 1\right)\right). \tag{93}$$

Finally, by replacing the Eqs. (12a) and (12b) relationships, the (observed) activity–(predicted)activity form of the ligand progress curve

$$[L]_L(t) = \mathrm{e}^{-\||A\rangle\|} \ln\left(1 + \mathrm{e}^{-\frac{\exp(\||A\rangle\|-\||A^*\rangle\|)}{4e^2}t}\left(\mathrm{e}^{\exp(\||A\rangle\|-\||A^*\rangle\|)} - 1\right)\right) \tag{94a}$$

or toxicity–activity expression

$$[L]_L(t) = \mathrm{EC}_{50} \ln\left(1 + \mathrm{e}^{-\frac{\exp(-\||A^*\rangle\|)}{4e^2\mathrm{EC}_{50}}t}\left(\mathrm{e}^{\frac{\exp(-\||A^*\rangle\|)}{\mathrm{EC}_{50}}} - 1\right)\right) \tag{94b}$$

are created for each structural parameter (reactivity indices considered) computed for each DFT framework considered and are correlated with each set of recorded species activities. Through this comparison, one may draw conclusions based on the chemical reactivity principles governing the specific chemical–biological interaction under study.

Next, a pilot application will be carried out to determine the activity of selected Cl-PAHs for various endpoints and biological species.

5 Quantitative Reactivity–Activity Modeling of Chemical–Biological Interactions

5.1 On QSAR Principles

Quantitative structure–activity (or property) relationship (QSA(P)R) methods seem best for unifying the chemical (and biological) interaction into a single model for researchers aiming to quantitatively organize the huge amount of experimental information in comprehensive equations with a predictive value [71, 203]. The QSA(R)R equation is justified by the quantum superposition principle written in the multilinear form of causes ($|X\rangle$ in the Dirac bracket notation of quantum states), resulting in Eq. (91) [204] and providing the appropriate framework in searching for new "natural laws" by various statistical means for computing the coefficients of these expansions (B's), such that the error of the predicted to recorded effects $|A^*\rangle$ is minimized. *Classical QSA(P)R* [67–70] assumes as descriptors the structural indices that directly reflect the electronic structures of the tested chemical compounds, e.g., factors describing the lipophilicity (e.g., LogP, surfaces), electronic effects (e.g., Hammett constants, polarization, localization of charges), and steric effects (e.g., Taft indices, Verloop indices, topological indices, molecular mass, total energy at optimized molecular geometry). Therefore, the optimization is centered on the molecular structure and various descriptors for the system of interest.

The driving QSAR principles have been established for the *in silico* validation of a compound. When completely fulfilled, a QSAR model can give, if not a general natural law [4], a working quantitative model for a given pool of chemical–biological interactions. These principles meet *European normative* regulatory

requirements for a given class of compounds, in accordance with the *Organization for Economic Cooperation and Development* (OECD). These QSAR-OECD principles are [5, 205]

i. **QSAR-OECD-1**: defining the (biological, ecological, or pharmacological) *activity* of a given chemical compound (the so-called *endpoint* A).
ii. **QSAR-OECD-2**: using a non-ambiguous algorithm in the quantitative attribution of activity (endpoint) for a chemical series based on their structure or physico-chemical properties and indices (X_1, X_2, …, X_M in QSAR equation).
iii. **QSAR-OECD-3**: defining the applicability domain relative to compounds and structural diversity considered in correlation with the QSAR model of the envisaged activity with the physico-chemical selected indices for the *in silico* tests (determination of parameters B_0, B_1, B_2, …, B_M in QSAR equation).
iv. **QSAR-OECD-4**: the quality of the QSAR model measured using the regression factor, robustness and predictability.
v. **QSAR-OECD-5**: the possibility of formulation for a mechanistic model of the physico-chemical interactions that yield the assumed activity (endpoint) and QSAR equation in general.

QSAR modeling is crucial for understanding and predicting toxicity-specific mechanisms but is faced with two fundamental problems [6]:

(a) A priori establishment of the physico-chemical factors to be considered in the QSAR equation to produce the most reliable model (i.e., assurance of the QSAR-OECD I, II, and IV principles).
(b) Assuring that activity A^* depends on physico-chemical factors that are *independent* and associated with the *primary causes* that generate the observable effect of the activity (endpoint) (i.e., the assurance of the QSAR-OECD III and V principles).

The present approach, considering structural parameters characterizing the frontier information of the molecular structures (electronegativity χ, chemical hardness η, electrophilicity ω, and chemical power π) that drive the chemical reactivity principles, provides individual associate QSARs for a given biological action/species. This generates the actual quantitative reactivity–activity relationships that have a major role in modeling the chemical–biological interaction along the logistic ligand-receptor progress curves, as previously described.

5.2 *In silico-Based Biological Activities*

Biological activity for a given molecule (available or newly synthesized) interacting with a certain species is not always known or easily produced;

determining these interactions can require extensive laboratory efforts and adherence to many environmental safety conditions.

Instead, at the *in silico* level, the *fill-in-the-data-gaps* technique may be applied (see Fig. 4), featuring the main algorithm [206, 207]:

- Choosing a *target molecule* provides existing analogues whose experimental data are available for a studied "end-point" (activity to be modeled).
- Providing data that are required for the target molecule either by read-across (analogue approach) or by trend analysis (molecular similarity approach).
- All these imply inner QSAR modeling specific for the profiled activity (ligand-receptor) that is hypothesized or a known binding mechanism, involving all available data (analogues and their endpoint measured effects).

The mechanism of toxic action involved in the algorithm loop of Fig. 4 is associated with the critical biological effect of the toxicant at the molecular or cellular level. The main classes of toxic action mechanisms are as follows: nonpolar narcosis, polar narcosis, weak acid respiratory uncoupling, formation of free radicals, electrophilic reactions, and toxic action by specific (receptor-mediated) mechanisms.

However, identification of the mechanism of toxic action is often difficult due to the complex nature of toxic activity [208].

As a general rule, the narcotic mechanism of toxic action is a result of nonspecific noncovalent reversible interactions with cell membranes [208]. Note that nonpolar and polar narcosis can be included in the narcotic mechanism category: nonpolar narcotics are neutral nonreactive compounds (aliphatic alcohols, ketones, and ethers), while polar narcotics are less inert and often possess a hydrogen donor moiety (phenols, anilines) (see [209]). Alternatively, compounds that undergo direct electrophilic interaction may have covalent interactions with biological macromolecules [210]. However, compounds may also undergo metabolic reactions resulting in more toxic forms; other chemicals produce their toxic effects by forming free radicals. For a general guideline, Table 4 presents a summary of structural criteria that can be used to assign mechanisms of toxic action to compounds.

5.3 *Chemical Reactivity Principles Hierarchy According to the Biological Activity of Cl-PAHs*

Polycyclic aromatic hydrocarbons (PAHs) are a class of more than 100 chemicals composed of up to six benzene rings fused together, such that any two adjacent benzene rings share two carbon bonds (e.g., phenanthrenes, naphthalene, and pyrene). They are generally produced during the incomplete burning of organic materials, including coal, oil, gas, wood, garbage, and tobacco. Coal tar ointments containing PAHs are used to treat several inflammatory skin conditions. PAHs are most often generated from motor vehicle exhaust, residential and industrial heating sources, coal, crude oil and natural gas processing, waste incineration, and tobacco smoke.

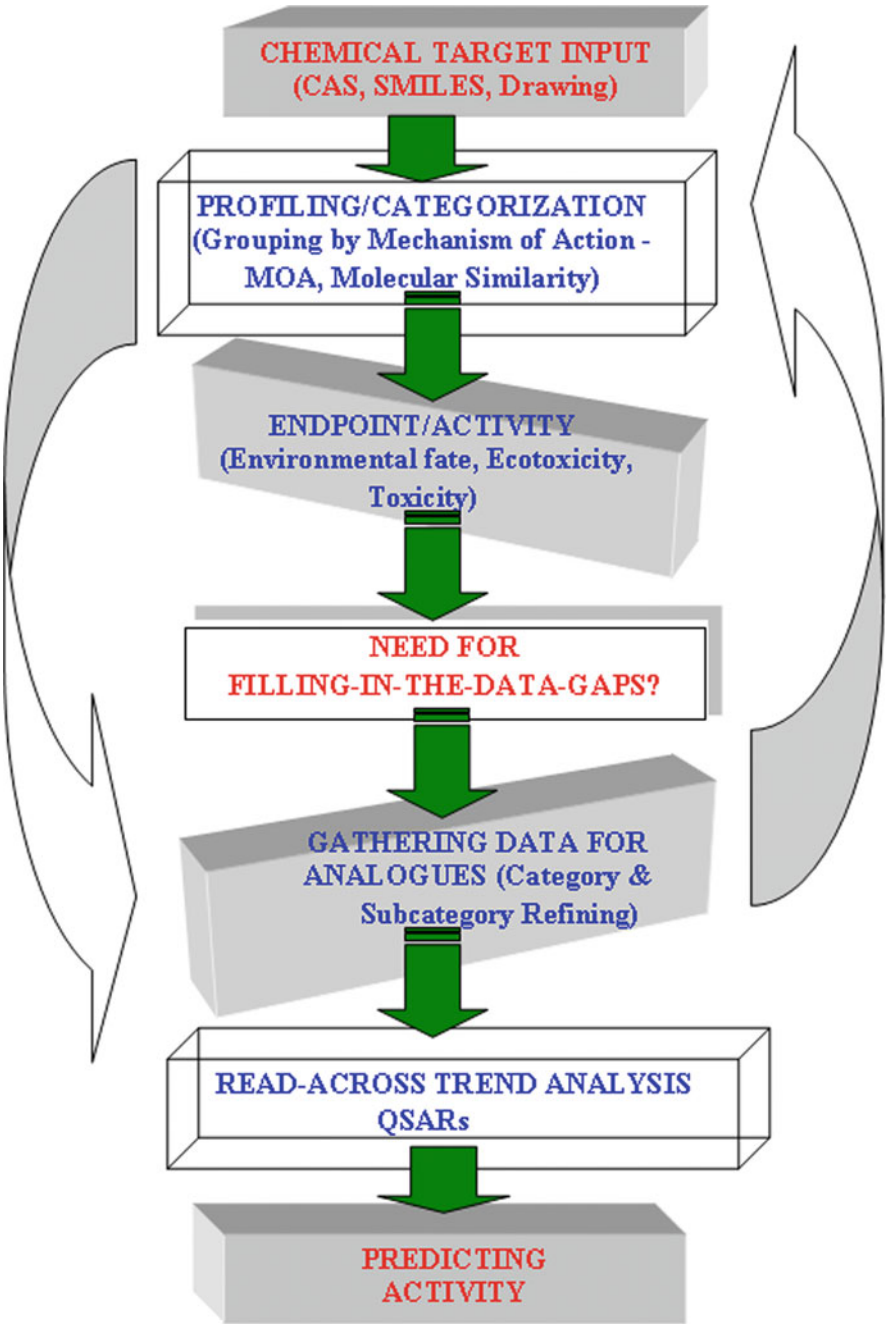


Fig. 4 The *in cerebro* scheme for *in silico* evaluation of environmental or toxicological activities of a given target chemical (of interest, or newly designed or synthesized), following and explicating the Toolbox QSAR computational facility [207]

Table 4 Summary of structural criteria used for classifying compounds according to the mechanism of toxic action [210]

Mechanism of action (MOA)	Structural features
Nonpolar narcosis	Saturated alkanes with, e.g., halogen and/or alkoxy substituents (aliphatic alcohols, ketones, ether, amines); halogens and alkyl substituted benzenes
Polar narcosis	Phenols with a pK_a greater than or equal to 6.0; phenols and anilines with three or fewer halogen atoms and/or alkyl substituents
Weak acid respiratory uncouples	Phenols and anilines with four or more halogen substituents or more than one nitro group or a single nitro group and more than one halogen group
Formation of free radicals	Phenol or aniline substituted with an electron-releasing group (alkoxy, hydroxyl, more than one alkyl group)
Electrophile/ proelectrophile	Activated unsaturated compounds; benzene rings without aniline or phenol substructures that have two nitro groups on one ring; phenols with a single nitro group but not more than one halogen group; aromatic compounds with two or more hydroxy groups in the *ortho* or *para* position and at least one unsubstituted aromatic carbon atom; quinines; aldehydes; compounds with halogens at the α-position of an aromatic bond; ketenes; epoxides
Specific mechanism	Chemicals that interact with specific biological macromolecules. For example, acetylcholinesterase inhibitors with an organophosphate group

Emitted PAHs can bind to particles in the air. Particle size depends in part on the source of the PAHs, while ambient air PAH concentrations show seasonal variation [213, 214]. PAHs are found in meat and in other foods as a result of smoking, grilling, broiling, or other high-temperature processing. Uncooked foods and vegetables also contain low levels of PAHs but can be contaminated by airborne particle deposition or growth in contaminated soil. Humans are usually exposed to PAH mixtures rather than to individual chemicals, and PAH mixture composition varies with the combustion source and temperature [215]. ClPAHs are hybrids of dioxins and PAHs suspected of having similar toxicities [216] and are generally known to be carcinogenic, mutagenic, and teratogenic, with greater mutagenicity, aryl-hydrocarbon receptor activity, and dioxin-like toxicity than the corresponding parent PAHs [217]. Especially at the DNA interaction level, ClPAHs have the ability to bind to and activate the aryl hydrocarbon receptor (AhR), a cytosolic, ligand-activated transcription receptor. The biological pathway involves translocation of the activated AhR to the nucleus. In the nucleus, the AhR binds to the AhR nuclear translator protein to form a heterodimer, leading to transcriptional modulation of genes and causing adverse changes in cellular processes and function [218].

Several ClPAHs have been determined to be AhR-active. AhR-mediated toxicity (Fig. 5) is activated by all embryotoxic HAH and PAH congeners [212, 219, 220] toward nuclear translocation, where the AhR heterodimerizes with the AhR nuclear translocator (ARNT). The resulting ligand-AhR-ARNT complex further combines with various coactivators and promotes their expression through interacting in the

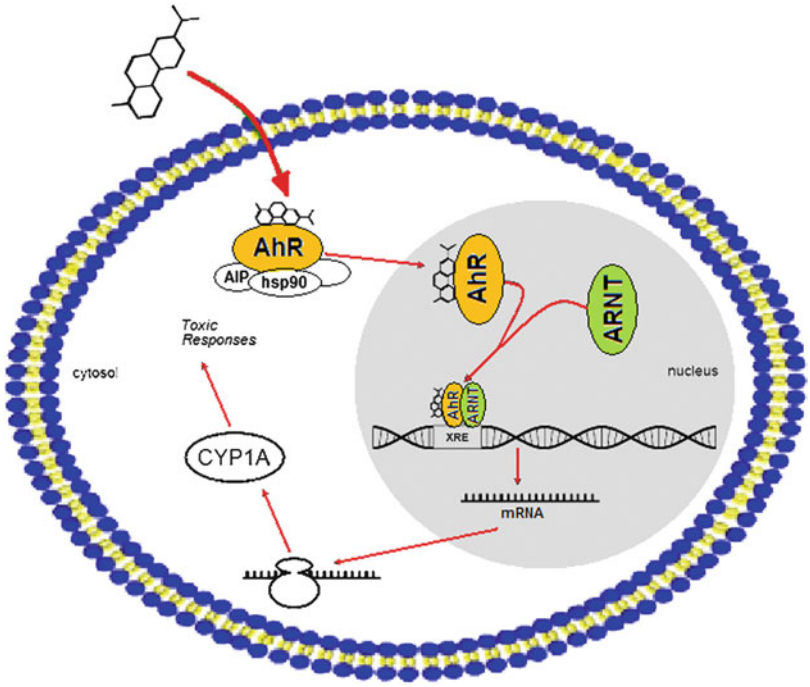


Fig. 5 Generic mechanism of AhR-mediated toxicity: AhR mediates signal transduction by dioxin-like ligands, which form a transcription factor complex with an aryl hydrocarbon nuclear translocator protein (ARNT). This heterodimer recognizes specific DNA sequences, namely dioxin responsive elements (DREs), and leads to induction of several genes forming the so-called Ah gene battery. In this process, the elevated levels of the protein products are assumed to be involved in the toxic action of AhR ligands. *AIP* AhR inhibitory protein, *hsp90* 90-kDa heat shock protein, *ARNT* AhR nuclear translocator, *XRE* xenobiotic response element, *CYP1A* cytochrome P450 1A gene/protein (adapted from [211, 212])

promoter region of AhR-regulated genes with xenobiotic responsive elements (XREs) [221]. This activity plays a role in cell proliferation and differentiation [222] and contributes to the biotransformation of xenobiotics, while also having a functional role in normal development and homeostasis [223–225]. Furthermore, the role of the AhR in TCDD (2,3,7,8-Tetrachlorodibenzo-*p*-Dioxin) and unalkylated PAH toxicity has been assessed [226, 227], but its role in alkylated PAH toxicity has not [211].

ClPAHs may be toxic to humans, and they have an equally important impact on the environment because several ClPAHs have also been found to exhibit mutagenic activity in *Salmonella typhimurium* in the Ames assay [229]. To comprehensively estimate reactivity–activity for representative ClPAHs (see Fig. 6) on human and environmental species, Table 5 presents the ethoxyresorufin-*O*-deethylase (EROD) activities for binding substrates of chlorinated polycyclic aromatic hydrocarbons (ClPAHs) with aryl hydrocarbon receptors (AhRs) in the cytochrome

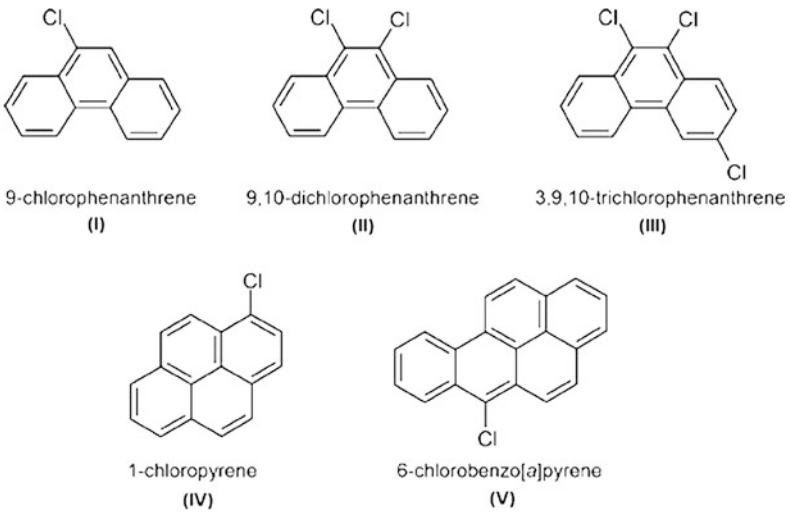


Fig. 6 Chemical structures of ClPAHs of interest [228]

Table 5 Chlorinated polycyclic aromatic hydrocarbons (ClPAH) activity expressed as (a) EROD (ethoxyresorufin-*O*-deethylase) activity as the relative intensity of ClPAH-induced cytochrome P450 (CYP) activity in human breast cancer MCF-7 cells [228]; (b) environmental fate: [Bioaccumulation aquatic] in *Pimephales promelas* over 96 h [10^3 L/kg wet]; (c) ecotoxicological information: [Aquatic Toxicity] LC50 for *Pimephales promelas* after 96 h [10^{-1} mg/L]; and (d) carcinogenity in rats, TD50 [10^{-3} mol/kg]

		Activity/toxicity			
Cl-PAH	CAS	*EROD*[(a)]	Fate/Pp[(b)]	AquaTox/Pp[(c)]	Carcino/Rats[(d)]
(I)	947-72-8	1.2	1.90	3.41	4.72
(II)	17219-94-2	1.4	1.74	1.3	4.05
(III)	800409-57-8	4.4	4.22	0.828	3.84
(IV)	34244-14-9	1.3	4.08	1.3	0.00511
(V)	21248-01-1	9	4.22	0.33	3.46

The values of (b)–(d) are computed using the Filling-in-the Data-Gap Toolbox OECD facility [v.1.1.01/2009] with OASIS baseline surface narcosis through DNA binding of PAHs in Fig. 6 [206, 207]

P450 (CYP) 1 family (CYP1A1 and 1B1) and expression in human breast cancer MCF-7 cells [228]. Also included are ecotoxicities on fish (Pimephales promelas) and rats, as computed by the previously described fill-in-data gaps method. On the molecular-ligand side, each compound in Fig. 6 is a single point in its symmetrical state computed within no-exchange-no-correlation (X0C0) Hartree–Fock (HF), and specific DFT Becke's exchange-correlation forms (Becke97, Becke88-VWN, B3-PW91, B3-LYP) with large Gaussian basis function (6-31G**) schemes within HyperChem software [230] using the HOMO and LUMO frontier information; they were then employed to produce the reactive indices of electronegativity [Eq. (18)],

Table 6 The values of electronegativity (in eV) for the ClPAHs of Fig. 6 computed from the frontier-like formula (18) with the HOMO and LUMO frontier orbital energies that were evaluated in various quantum mechanical frameworks: no-exchange-no-correlation (X0C0) to Hartree–Fock (HF), and specific DFT Becke's exchange-correlation forms (Becke97, Becke88-VWN, B3-PW91, B3-LYP) with large Gaussian basis functions (6-31G**) within HyperChem software [230]

χ	Quantum chemical framework					
Cl-PAH	X0C0	HF	Becke97	Becke88-VWN	B3-PW91	B3-LYP
(I)	251.1668	161.6397	190.1459	191.0285	190.8256	190.7167
(II)	161.12	122.6308	155.8952	129.6772	155.9582	176.9095
(III)	16.58217	156.6598	78.37052	30.02489	233.2496	78.76642
(IV)	115.2805	62.72652	199.2066	220.7729	192.2411	194.9907
(V)	163.9368	256.0867	88.00236	98.01646	215.249	215.0238

Table 7 The values of chemical hardness (in eV) for the ClPAHs of Fig. 6 computed from the frontier-like formula (19) with HOMO and LUMO frontier orbitals' energies evaluated as in Table 6

η	Quantum chemical framework					
Cl-PAH	X0C0	HF	Becke97	Becke88-VWN	B3-PW91	B3-LYP
(I)	1.183525	0.28125	2.195176	2.39061	2.329193	2.317177
(II)	2.44944	0.231994	0.448074	0.690163	0.385002	0.286087
(III)	0.065345	1.273979	1.074822	1.177084	0.214798	0.864036
(IV)	0.065396	0.389016	0.510063	0.298729	0.235435	0.381012
(V)	0.183769	0.43338	0.284748	0.567326	2.196144	1.941712

Table 8 The values of chemical power for the ClPAHs of Fig. 6 computed using definition (3) applied to the electronegativity and chemical hardness values from Tables 6 and 7

π	Quantum chemical framework					
Cl-PAH	X0C0	HF	Becke97	Becke88-VWN	B3-PW91	B3-LYP
(I)	106.1096	287.3595	43.30996	39.95393	40.96389	41.15281
(II)	32.88916	264.2969	173.9615	93.9468	202.542	309.1883
(III)	126.8827	61.48444	36.45746	12.75393	542.9511	45.58049
(IV)	881.4097	80.62214	195.2765	369.5204	408.2671	255.8852
(V)	446.0405	295.4529	154.5267	86.38468	49.00613	55.36966

Table 9 The values of electrophilicity for ClPAHs of Fig. 6 computed using definition (4) applied to the electronegativity and chemical hardness values from Tables 6 and 7

ω	Quantum chemical framework					
Cl-PAH	X0C0	HF	Becke97	Becke88-VWN	B3-PW91	B3-LYP
(I)	26,651.22	46,448.71	8,235.211	7,632.341	7,816.958	7,848.526
(II)	5,299.102	32,410.95	27,119.76	12,182.76	31,588.08	54,698.35
(III)	2,103.989	9,632.138	2,857.19	382.9353	126,643.2	3,590.212
(IV)	101,609.3	5,057.147	38,900.38	81,580.09	78,485.72	49,895.23
(V)	73,122.46	75,661.56	13,598.72	8,467.12	10,548.52	11,905.79

chemical hardness [Eq. (19)], chemical power [Eq. (3)] and electrophilicity [Eq. (4)], with the results reported in Tables 6, 7, 8, and 9.

The connection between the activity data of Table 5 and the molecular structural-frontier information from Tables 6, 7, 8, and 9 is made based on the *biological activity-driven chemical reactivity* algorithm, which is qualitatively presented in the scheme of Fig. 1 and quantitatively represented by the logistic ligand progress curves of Eq. (94b). It is clear that the present chemical–biological interaction (ClPAH molecule-AhR-mediated toxicity, see Fig. 5) is a specific realization of the generic ligand–receptor kinetics, which is modeled quantum mechanically and mostly using DFT methods. The interaction involves the predicted norm of the respective chemical structure–biological activity correlation through the presence of the predicted initial (in time evolution of ligand–receptor kinetics) bound ligand to the receptor site [see Eq. (12a)], as well as the algebraic norm evaluation of the specific EC_{50} [see Eq. (12b)], for each observed or recorded (experimentally or computationally by filling in the data gaps—see Fig. 4) set of activities for the molecules of interest. At this point, one should note that the employed activities for bioaccumulation in *Pimephales promelas* (P.p.), ecotoxicology of P.p., and carcinogenity in rats shown in Table 5 were in fact the correspondent 50 % read-across concentrations for aimed effects obtained using ToolBox in an *in silico* environment. However, it turns out that when considering the EC_{50} and then extracting the activity relationships from the logarithmic forms, in each case, no significant distinction between the influences of the reactivity indices on the bio- and eco-toxicology activity correlation were recorded. Instead, the relationship cancels out all chemical information or mechanisms in producing biological effects. This should be avoided (see QSAR-OECD-5 principle of Sect. 5.1), so we consider the 50 % read-across concentrations as the aimed effects and note this as peculiar *in silico* behavior for ToolBox that should be improved in the future. The biological-driving-chemical interaction results for ClPAHs-AhR-mediated toxicity for human breast cancer MCF-7 cells, aquatic bioaccumulation for *P.p.*, aquatic toxicity for *P.p.* and carcinogenity for rats are summarized in Tables 10, 11, 12 and 13 and in Figs. 7, 8, 9 and 10.

To correctly interpret the results, one can set the following mechanistic rule for hierarchical biological activity-driving chemical reactivity principles: *the higher the QRAR correlating factor is, the closer the predicted initial bound ligand concentration will regulate the toxicity of the in-set EC_{50} concentration*. Thereby, the overall bio- or eco-toxicological effect is modeled. Nevertheless, one notes that in all cases in Tables 10, 11, 12 and 13 and Figs. 7, 8, 9, and 10, the higher Pearson correlation factor associates with the closer L-bound concentrations for the working EC_{50}. In this way, one can systematically identify which reactivity index (and correspondent principle thereof) is dominant in which quantum/DFT computational environment for the biological or ecological system. As such, for the considered systems one finds the following:

- For modeling the interaction of ClPAH ligands that bind human breast cancer MCF-7 cells (Table 10; Fig. 7), it appears that the B3-PW91 DFT exchange-correlation scheme recovers the consecrate chemical reactivity scheme

Table 10 The synopsis of the quantitative (χ, η, π, ω) reactivity indices–EROD activity relationships (EROD-QRAR), including the correlation factor (R) and the initial predicted concentration of Eq. (12a), $L_0 = \exp(-\||A^*\rangle\|)$ [μM], for each index/quantum chemical method considered in Tables 6, 7, 8 and 9 against the EROD activity of Table 5

Quantum-method	Index	$\|A^*\rangle EROD - QSAR - EQUATION$	R	L_0
X0C0	χ	$4.55\ \|1\rangle - 7.7 \times 10^{-3}\ \|\chi\rangle$	0.195	390.622
	η	$4.7\ \|1\rangle - 1.57\ \|\eta\rangle$	0.483	225.796
	π	$3.22\ \|1\rangle + 7.63 \times 10^{-4}\ \|\pi\rangle$	0.08	428.397
	ω	$2.83\ \|1\rangle + 1.51 \times 10^{-5}\ \|\omega\rangle$	0.197	389.811
HF	χ	$-2.78\ \|1\rangle + 0.041\ \|\chi\rangle$	0.856	64.0455
	η	$2.289\ \|1\rangle + 2.243\ \|\eta\rangle$	0.284	345.096
	π	$2.27\ \|1\rangle + 6.02 \times 10^{-3}\ \|\pi\rangle$	0.208	384.824
	ω	$0.96\ \|1\rangle + 7.39 \times 10^{-5}\ \|\omega\rangle$	0.631	146.062
Becke97	χ	$10.38\ \|1\rangle - 0.049\ \|\chi\rangle$	0.813	75.8095
	η	$5.08\ \|1\rangle - 1.80\ \|\eta\rangle$	0.417	265.837
	π	$3.39\ \|1\rangle + 5.44 \times 10^{-4}\ \|\pi\rangle$	0.012	436.279
	ω	$5.15\ \|1\rangle - 9.32 \times 10^{-5}\ \|\omega\rangle$	0.406	272.73
Becke88-VWN	χ	$6.99\ \|1\rangle - 0.026\ \|\chi\rangle$	0.592	165.435
	η	$4.67\ \|1\rangle - 1.18\ \|\eta\rangle$	0.288	342.959
	π	$4.33\ \|1\rangle - 7.21 \times 10^{-3}\ \|\pi\rangle$	0.306	332.855
	ω	$4.299\ \|1\rangle - 3.8 \times 10^{-5}\ \|\omega\rangle$	0.378	289.6
B3-PW91	χ	$-10.75\ \|1\rangle + 0.072\ \|\chi\rangle$	0.62	151.004
	η	$2.15\ \|1\rangle + 1.22\ \|\eta\rangle$	0.394	279.627
	π	$4.23\ \|1\rangle - 3.1 \times 10^{-3}\ \|\pi\rangle$	0.203	386.758
	ω	$3.83\ \|1\rangle - 7.22 \times 10^{-6}\ \|\omega\rangle$	0.109	421.53
B3-LYP	χ	$2.73\ \|1\rangle + 4.25 \times 10^{-3}\ \|\chi\rangle$	0.067	430.681
	η	$1.79\ \|1\rangle + 1.44\ \|\eta\rangle$	0.395	279.377
	π	$5.40\ \|1\rangle - 0.014\ \|\pi\rangle$	0.53	198.381
	ω	$5.24\ \|1\rangle - 6.97 \times 10^{-5}\ \|\omega\rangle$	0.508	211.162

(electronegativity > chemical hardness > chemical power > electrophilicity), which may be called the *I-BioAct-ChemReact* scheme of biological activity-driven chemical reactivity realization. Almost the same situation appears for Becke97 computations, except for an inversion of chemical power with electrophilicity influence; in both cases, the prescribed reactivity hierarchy affirms the electronegativity equalization first, followed by chemical hardness principles (see Table 1), while their combined equalization is naturally released with electrophilicity and then with chemical power equalizations at the level of ligand–receptor site molecular interactions, ***I-BioAct-ChemReact:*** $(\boldsymbol{\chi \wedge \eta}) > (\boldsymbol{\pi \wedge \omega})$. The second scheme of biological–chemical interaction appears when either chemical power or electrophilicity indices and principles dominate those of electronegativity and chemical hardness, as is the case for B3-LYP computation, ***II-BioAct-ChemReact:*** $(\boldsymbol{\pi \vee \omega}) > (\boldsymbol{\chi \vee \eta})$, while the third scheme is recorded when one of the basic chemical reactivity (electronegativity or chemical hardness) indices and principles acts in the first instance and is followed by one of the combined reactivity indices (chemical power or electrophilicity), ***III-BioAct-ChemReact:*** $(\boldsymbol{\chi \vee \eta}) > (\boldsymbol{\pi \vee \omega})$. In this last case, although one of the

Table 11 The same synopsis as in Table 10 but for the FATE/Pp activity of Table 5

Quantum-method	Index	$\|A^*\rangle FATE - QSAR - EQUATION$	R	L_0
X0C0	χ	$4.68\ \|1\rangle - 0.01\ \|\chi\rangle$	0.677	589.933
	η	$4.132\ \|1\rangle - 1.139\ \|\eta\rangle$	0.918	497.636
	π	$2.511\ \|1\rangle + 2.26 \times 10^{-3}\ \|\pi\rangle$	0.617	610.868
	ω	$2.597\ \|1\rangle + 1.52 \times 10^{-5}\ \|\omega\rangle$	0.517	643.122
HF	χ	$2.75\ \|1\rangle + 3.17 \times 10^{-3}\ \|\chi\rangle$	0.173	716.799
	η	$2.31\ \|1\rangle + 1.77\ \|\eta\rangle$	0.586	621.244
	π	$4.53\ \|1\rangle - 6.56 \times 10^{-3}\ \|\pi\rangle$	0.592	619.391
	ω	$3.45\ \|1\rangle - 6.52 \times 10^{-6}\ \|\omega\rangle$	0.145	719.657
Becke97	χ	$4.94\ \|1\rangle - 0.012\ \|\chi\rangle$	0.523	641.171
	η	$3.9\ \|1\rangle - 0.744\ \|\eta\rangle$	0.450	662.262
	π	$3.03\ \|1\rangle + 1.64 \times 10^{-3}\ \|\pi\rangle$	0.095	723.676
	ω	$3.28\ \|1\rangle - 2.53 \times 10^{-6}\ \|\omega\rangle$	0.029	726.44
Becke88-VWN	χ	$4.02\ \|1\rangle - 5.86 \times 10^{-3}\ \|\chi\rangle$	0.343	688.379
	η	$4.07\ \|1\rangle - 0.817\ \|\eta\rangle$	0.524	641.053
	π	$2.91\ \|1\rangle + 2.66 \times 10^{-3}\ \|\pi\rangle$	0.295	698.247
	ω	$2.99\ \|1\rangle + 1.1 \times 10^{-5}\ \|\omega\rangle$	0.286	699.851
B3-PW91	χ	$-3.71\ \|1\rangle + 0.035\ \|\chi\rangle$	0.793	546.71
	η	$3.48\ \|1\rangle - 0.231\ \|\eta\rangle$	0.195	714.148
	π	$2.51\ \|1\rangle + 2.91 \times 10^{-3}\ \|\pi\rangle$	0.501	647.901
	ω	$2.52\ \|1\rangle + 1.4 \times 10^{-5}\ \|\omega\rangle$	0.550	632.856
B3-LYP	χ	$4.165\ \|1\rangle - 5.45 \times 10^{-3}\ \|\chi\rangle$	0.226	709.856
	η	$3.37\ \|1\rangle - 0.123\ \|\eta\rangle$	0.088	724.144
	π	$3.65\ \|1\rangle - 2.94 \times 10^{-3}\ \|\pi\rangle$	0.296	697.942
	ω	$3.596\ \|1\rangle - 1.42 \times 10^{-5}\ \|\omega\rangle$	0.271	702.58

chemical reactivity principles acts first, it is then integrated into the combined reactivity as chemical power or electrophilicity, where implicitly, the other equalization principle for the remaining reactivity index naturally proceeds, as is the case for X0C0 (non exchange and no correlation), HF and Becke-VWN computations. Note that the last two cases, Schemes II and III of the BioAct-ChemReact correlations, indicate a synergism between electronegativity and chemical hardness in the combined reactivity of chemical power and electrophilicity indices that, in this way, promote both indices as global equalization principles, so that electronegativity and chemical hardness are automatically fulfilled as well. For the EROD activity for human breast cancer MCF-7 cells, B3-PW91 and Becke 97 are in accord with an in-chain reactivity hierarchy as prescribed by Table 1, while the other two DFT computational frameworks (Becke88-VWN and B3-LYP) prescribe synergetics for these principles in combined indices and equalization principles of chemical power and electrophilicity.

- For the effect of the bioaccumulation of ClPAHs on the environmental fate of *Pimphales promelas* (Table 11; Fig. 8), a case is made for the ***I-BioAct-ChemReact*** for X0C0 along Becke97 and Becke88-VWN computational schemes, or for ***II-BioAct-ChemReact*** for HF and B3-LYP computations, while the third case ***III-BioAct-ChemReact*** is found only for the B3-PW91 *in silico* approach.

Table 12 The same synopsis as in Table 10 but for the aqua-toxicity/Pp activity of Table 5

Quantum-method	Index	$\|A^*\rangle AQUATOX-QSAR-EQUATION$	R	L_0
X0C0	χ	$0.132\ \|1\rangle + 9.19 \times 10^{-3}\ \|\chi\rangle$	0.668	28,156.8
	η	$1.11\ \|1\rangle + 0.407\ \|\eta\rangle$	0.360	36,315.3
	π	$1.74\ \|1\rangle - 9.7 \times 10^{-4}\ \|\pi\rangle$	0.291	37,716.6
	ω	$1.67\ \|1\rangle - 5.66 \times 10^{-6}\ \|\omega\rangle$	0.211	39,011.7
HF	χ	$2.07\ \|1\rangle - 4.19 \times 10^{-3}\ \|\chi\rangle$	0.250	38,418.2
	η	$1.99\ \|1\rangle - 1.06\ \|\eta\rangle$	0.386	35,736.7
	π	$0.856\ \|1\rangle + 2.92 \times 10^{-3}\ \|\pi\rangle$	0.289	37,739.1
	ω	$1.46\ \|1\rangle - 7.63 \times 10^{-7}\ \|\omega\rangle$	0.019	40,521.2
Becke97	χ	$-0.62\ \|1\rangle + 0.014\ \|\chi\rangle$	0.692	27,454.6
	η	$0.234\ \|1\rangle + 1.329\ \|\eta\rangle$	0.883	21,941.5
	π	$2.268\ \|1\rangle - 6.91 \times 10^{-3}\ \|\pi\rangle$	0.442	34,392.8
	ω	$1.63\ \|1\rangle - 1.07 \times 10^{-5}\ \|\omega\rangle$	0.134	39,909.4
Becke88-VWN	χ	$0.239\ \|1\rangle + 8.92 \times 10^{-3}\ \|\chi\rangle$	0.574	30,855.6
	η	$0.215\ \|1\rangle + 1.189\ \|\eta\rangle$	0.837	23,264.0
	π	$1.56\ \|1\rangle - 1.05 \times 10^{-3}\ \|\pi\rangle$	0.128	39,968.1
	ω	$1.47\ \|1\rangle - 1.56 \times 10^{-6}\ \|\omega\rangle$	0.044	40,464.4
B3-PW91	χ	$4.24\ \|1\rangle - 0.014\ \|\chi\rangle$	0.352	36,491.9
	η	$0.99\ \|1\rangle + 0.414\ \|\eta\rangle$	0.384	35,785.8
	π	$1.92\ \|1\rangle - 1.94 \times 10^{-3}\ \|\pi\rangle$	0.367	36,155.2
	ω	$1.87\ \|1\rangle - 8.46 \times 10^{-6}\ \|\omega\rangle$	0.366	36,182.4
B3-LYP	χ	$0.801\ \|1\rangle + 3.69 \times 10^{-3}\ \|\chi\rangle$	0.168	39,563.0
	η	$0.814\ \|1\rangle + 0.535\ \|\eta\rangle$	0.420	34927.0
	π	$1.606\ \|1\rangle - 1.22 \times 10^{-3}\ \|\pi\rangle$	0.135	39,900.5
	ω	$1.579\ \|1\rangle - 5.683 \times 10^{-6}\ \|\omega\rangle$	0.119	40,042.5

- When the ecotoxicology of ClPAHs on the environmental fate of *Pimphales promelas* (Table 12; Fig. 9) is considered to be the driving scheme, ***I-BioAct-ChemReact*** dominates the computational quantum frameworks because it is specific to Becke97, Becke88-VWN, B3-LYP, and even X0C0, while the schemes HF and B3-PW91 appear specific to the ***III-BioAct-ChemReact*** mechanism of the ligand–receptor interaction. Therefore, this chemical–biological system clearly favors chemical reactivity principles as triggering the chemical biological specific interactions.
- Finally, when ClPAH ligands mediated AhR toxicity in rats, carcinogenic effects are produced (Table 13; Fig. 10). No specific ***I-BioAct-ChemReact*** mechanism of action was recorded because the ***II-BioAct-ChemReact*** hierarchy dominates this system, covering the X0C0, Becke97, Becke88-VWN, and B3-LYP computations, followed by the ***III-BioAct-ChemReact*** hierarchies for the HF and B3-PW91 *in silico* methods. Consequently, the synergistic mechanism of joint electronegativity and chemical hardness equalization in chemical power and electrophilicity is the driving mechanism specific for carcinogenesis.

The present applications illustrate the reliability of the ***BioAct-ChemReact*** logistical-QSAR algorithm in modeling specific biological–chemical interactions for various species, particularly its ability to discriminate between tumor promotion, bio-accumulation, ecotoxicity, and carcinogenity, as presented and analyzed

Table 13 The same synopsis as in Table 10 but for the carcinogenity/rats activity of Table 5

Quantum-method	Index	$\|A^*\rangle CARCINO - QSAR - EQUATION$	R	L_0
X0C0	χ	$2.279\ \|1\rangle + 0.61 \times 10^{-3}\ \|\chi\rangle$	0.305	691.13
	η	$2.539\ \|1\rangle + 0.856\ \|\eta\rangle$	0.481	607.539
	π	$4.79\ \|1\rangle - 4.94 \times 10^{-3}\ \|\pi\rangle$	0.940	339.57
	ω	$4.639\ \|1\rangle - 3.41 \times 10^{-5}\ \|\omega\rangle$	0.808	415.336
HF	χ	$0.872\ \|1\rangle + 0.015\ \|\chi\rangle$	0.585	548.196
	η	$3.022\ \|1\rangle + 0.369\ \|\eta\rangle$	0.085	749.612
	π	$1.37\ \|1\rangle + 9.3 \times 10^{-3}\ \|\pi\rangle$	0.585	548.231
	ω	$2.143\ \|1\rangle + 3.17 \times 10^{-5}\ \|\omega\rangle$	0.493	600.768
Becke97	χ	$4.918\ \|1\rangle - 0.012\ \|\chi\rangle$	0.365	665.551
	η	$2.187\ \|1\rangle + 1.129\ \|\eta\rangle$	0.481	607.539
	π	$5.113\ \|1\rangle - 0.016\ \|\pi\rangle$	0.639	516.73
	ω	$4.994\ \|1\rangle - 9.81 \times 10^{-5}\ \|\omega\rangle$	0.778	432.813
Becke88-VWN	χ	$4.79\ \|1\rangle - 0.012\ \|\chi\rangle$	0.482	606.849
	η	$1.68\ \|1\rangle + 1.49\ \|\eta\rangle$	0.667	499.366
	π	$4.712\ \|1\rangle - 0.012\ \|\pi\rangle$	0.960	328.542
	ω	$4.38\ \|1\rangle - 5.28 \times 10^{-5}\ \|\omega\rangle$	0.957	330.275
B3-PW91	χ	$3.28\ \|1\rangle - 3.42 \times 10^{-4}\ \|\chi\rangle$	0.005	754.817
	η	$2.37\ \|1\rangle + 0.79\ \|\eta\rangle$	0.464	616.591
	π	$4.134\ \|1\rangle - 3.69 \times 10^{-3}\ \|\pi\rangle$	0.443	627.51
	ω	$3.866\ \|1\rangle - 1.28 \times 10^{-5}\ \|\omega\rangle$	0.350	672.034
B3-LYP	χ	$4.55\ \|1\rangle - 7.8 \times 10^{-3}\ \|\chi\rangle$	0.225	719.275
	η	$2.019\ \|1\rangle + 1.03\ \|\eta\rangle$	0.515	588.817
	π	$4.172\ \|1\rangle - 6.76 \times 10^{-3}\ \|\pi\rangle$	0.476	610.099
	ω	$4.244\ \|1\rangle - 4.02 \times 10^{-5}\ \|\omega\rangle$	0.535	577.469

in Tables 10, 11, 12 and 13 and Figs. 7, 8, 9 and 10 for the biological systems of Table 5. As a general output, the chemical reactivity principles of Table 1 act according to the general scenario of chemical bonding (see Sect. 2.2), here extended to chemical–biological interactions (see Sect. 4). Some amendments are required when using the Becke97 exchange-correlation scheme as the main *in silico* DFT modeling paradigm for such a binding mechanism (the hierarchy ***I-BioAct-ChemReact***). Carcinogenity seems to be regulated by the ***II-BioAct-ChemReact*** hierarchy, with Becke97 and Becke88-VWN as the proper computational frameworks. Further studies are necessary to further assess the synergetics of the chemical reactivity principles that work toward biological activity using bi- and multi-dependent quantitative reactivity–activity relationships (QRARs) with various species and toxicants, as well as by considering the biological–chemical hierarchies through the asymptotic times by which the ligand progress curves [Eq. (94b)] first vanish from each computational–conceptual framework (see Figs. 7, 8, 9 and 10). Systematic studies will seek a general formulation of the chemical–biological interactions using quantum catalytic (logistic) activities.

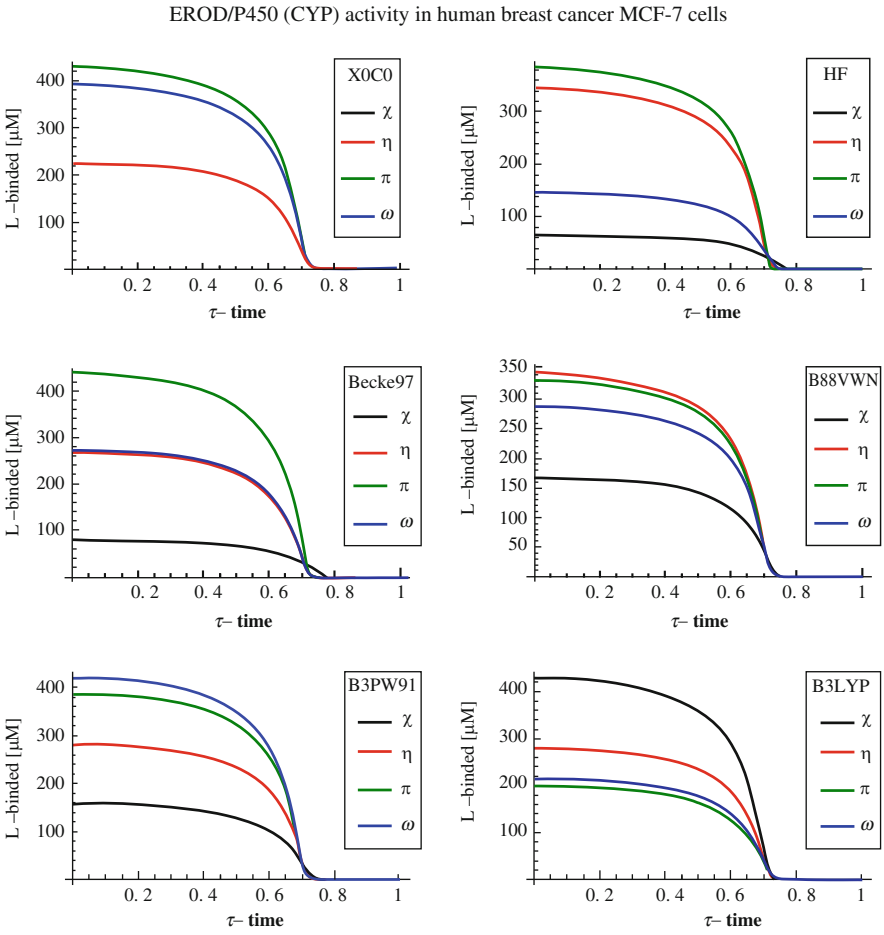


Fig. 7 Comparative representation of the progress curves for the ClPAH ligands of Fig. 6 bound to human breast cancer MCF-7 cells, employing the recorded EROD/human-QRAR reactivity–activity information from Table 10 into logistic chemical–biological interactions modeled by Eq. (89), on the mapped unitary time scale of Eq. (15), for each index/quantum chemical method considered and for an EROD $EC_{50} = 34.696$ μM norm parameter as computed with algebraic definition (12b) and the EROD/human data of Table 5

6 Conclusions

Continuous increases in the synthesis of chemical compounds, accompanied by their disposal in the environment, have made hazard toxicity a global issue, especially at the level of the chemical–biological interaction between toxicants and species. In this context, the marriage between structural molecular computing and catalytic activity should be accomplished in a systematic, yet universal, way. In quantitative terms, structural information is today represented by ab initio quantum

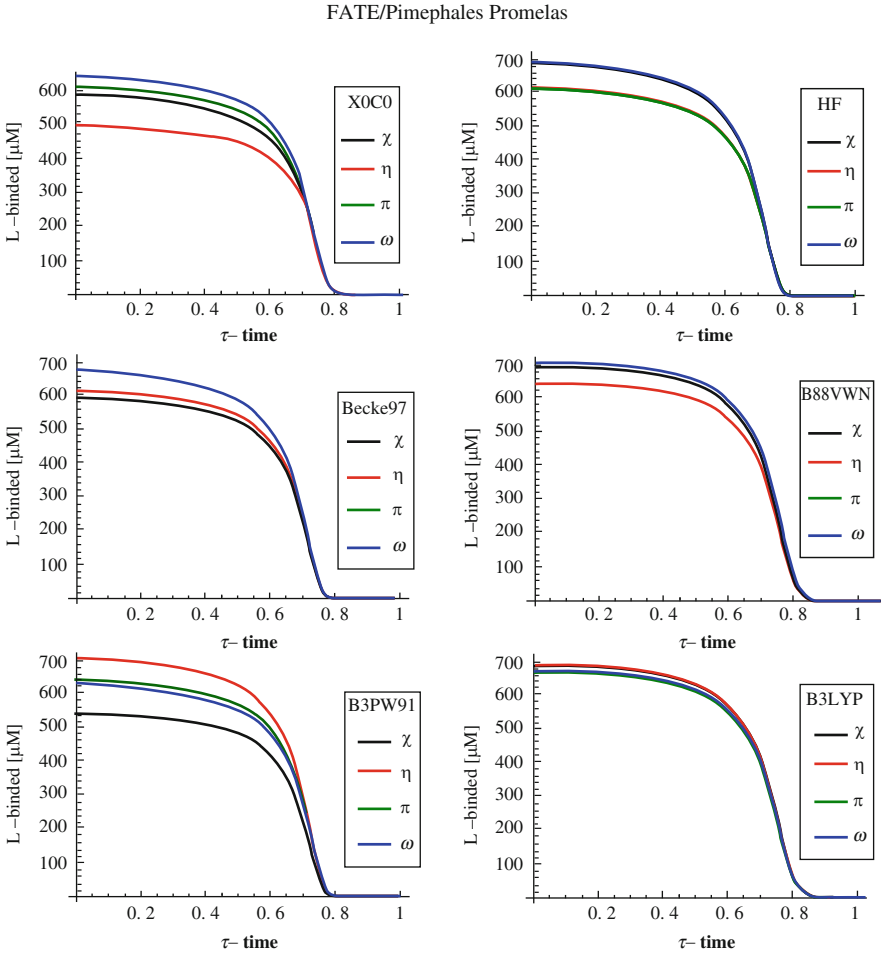


Fig. 8 The same type of representations as in Fig. 7 but for the ClPAH ligand–FATE/Pp QRAR reactivity–activity information from Table 11, with an FATE/Pp $EC_{50} = 464.437$ μM norm parameter as computed using the algebraic definition (12b) with the FATE/Pp data of Table 5

chemical methods including the advanced methods of Density Functional Theory (DFT). For biological uptake, the preferred paradigms are less obvious, and thus, the present work reviews logistic enzyme–substrate kinetics and generalizes it in terms of ligand–receptor interactions. It is worth noting that the resulting logistic progress curves for ligand receptors, themselves a generalization of the consecrated Michaelis-Menten catalytic paradigm, involve the algebraic employment of activity predicted from a quantitative structure–activity relationship (QSAR), Eq. (94). This method is therefore affirmed as the analytical realization of the chemical–biological interaction, which can be employed for various chemical toxicants and biological species. However, the present approach also includes a broader conceptual

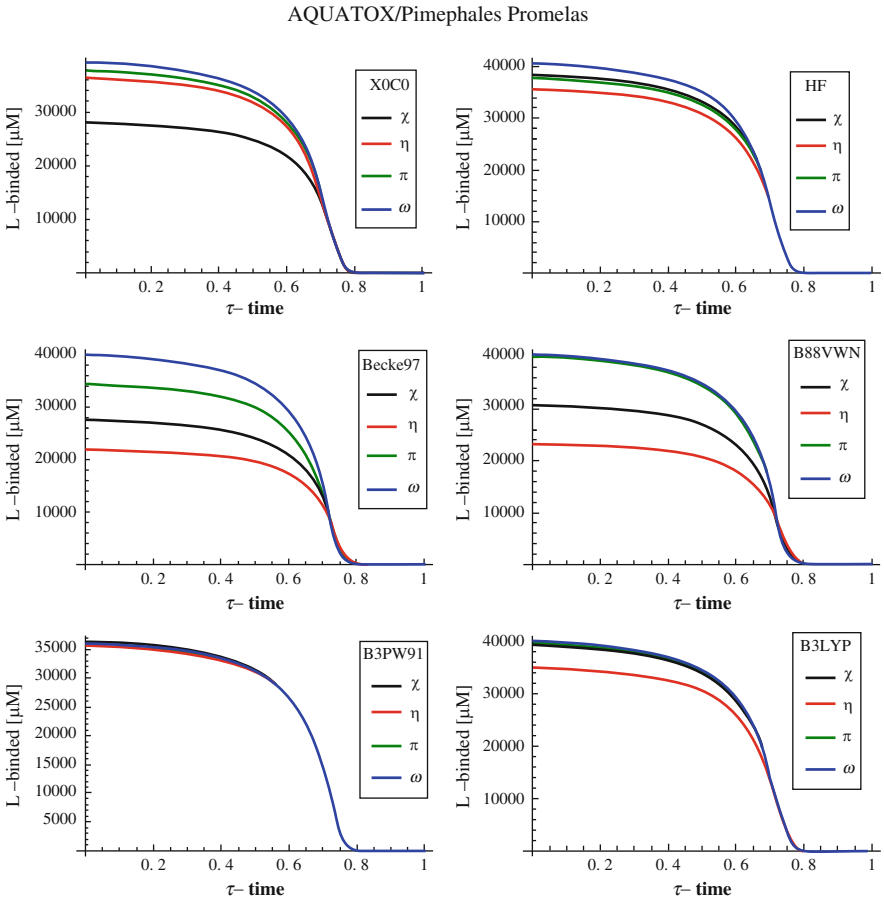


Fig. 9 The same type of representation as in Fig. 7 but for the ClPAH ligand–LC50/Pp QRAR reactivity–activity information from Table 12, with an $EC_{50} = 18774.7$ μM norm parameter as computed using the algebraic definition (12b) with the LC50/Pp data of Table 5

consequence: namely, it allows testing of the hierarchy of chemical reactivity principles of electronegativity and chemical hardness.

Recent advances in conceptual DFT have spurred debate over whether these two concepts are manifestly distinct or act synergistically in unified electrophilicity and chemical power indices. Accordingly, chemical reactivity principles are also reviewed, qualitatively and quantitatively, and tested for their hierarchy or synergistic appearance through the "eyes" of recorded biological or computed *in silico* activity (by, for example, the celebrated Toolbox utility). Calculations were carried out on a pilot set of chlorinated polycyclic aromatic hydrocarbons (ClPAHs) using various species (human breast cancer cells), *Pimephales promelas* and rats (Rattus novegicus) according to the aryl hydrocarbon receptor (AhR)-mediated toxicity, as reflected in EROD (ethoxyresorufin-*O*-deethylase) activity, environmental

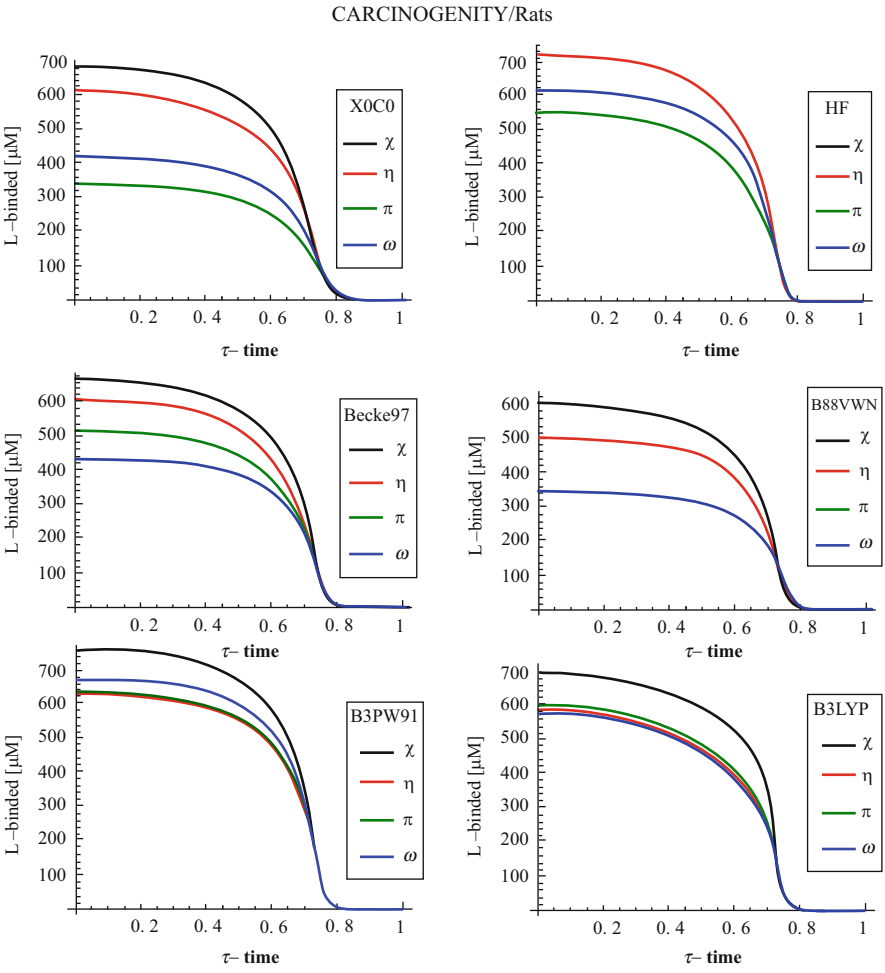


Fig. 10 The same type of representation as in Fig. 7 but for the ClPAH ligand–TD50/rats QRAR reactivity–activity information from Table 13, with an $EC_{50} = 307.538$ μM norm parameter as computed using the algebraic definition (12b) with the TD50/Pp data of Table 5

bioaccumulation and toxicity, and carcinogenity. The results were revelatory because the DFT exchange and correlation methods widely accompanied the chemical bonding scenario (of Table 1) at the level of the biological receptor interactions for most of the environmental hazard toxicity, while modulating the synergistic behavior of the reactivity indices of electronegativity and chemical hardness in the unified chemical bonding measures as the electrophilicity and chemical power, Eqs. (3) and (4), for carcinogenesis modeling. This approach identifies the specific chemical biological interaction for a given toxicant–organism system through chemical reactivity mechanisms, based on logistic ligand–receptor progress-curve-analysis-based QSAR, thus providing the so-called "Biological

Activity driving Chemical Reactivity" realization of the fifth QSAR-OECD principle. The prediction of the initial bound ligand concentration (here in micro-Molar) that triggers the concerned biological uptake is also provided (Figs. 7, 8, 9 and 10). Further studies are envisaged in which quantitative reactivity–activity relationships (QRARs) are revealed for bi- and multi-linear (or even polynomial) correlations between observed or computed activity and frontier chemical information toward systems to determine sequential or synergistic behaviors of chemical reactivity principles and to provide contributions to the biological realm.

Acknowledgments This work was supported by the Romanian National Council of Scientific Research (CNCS-UEFISCDI) through project TE16/2010-2013 within the PN II-RU-TE-2010-1 framework and by Romanian Academy.

References

1. Benfenati E (2007) Chem Cent J 1:32
2. Wang J, Hou T (2010) J Chem Inf Model 50:55–67
3. EC: European Comision, Join Research Centre, Institute for health and Consumer Protection (2010) http://ecb.jrc.ec.europa.eu/qsar/
4. Cherkasov A (2005) Curr Comp Aided Drug Des 1:21–42
5. OECD (2007) OECD (Organization for Economic Co-operation and Development): Guidance document [ENV/JM/MONO(2007)2] on the validation of (Quantitative) structure–activity relationship [(Q)SAR] models. OECD Environment Health and Safety Publications (2007) Series on Testing and Assessment, No. 69, Paris
6. QCS: QSAR and Combinatorial Science (2008) 27:1–132 (Special Issue on Computational Assessment of Toxicity and Environmental Fate)
7. Parr RG (1983) Annu Rev Phys Chem 34:631–656
8. Parr RG, Yang W (1989) Density functional theory of atoms and molecules. Oxford University Press, New York
9. Kohn W, Becke AD, Parr RG (1996) J Phys Chem 100:12974–12980
10. Ayers PW, Parr RG (2001) J Am Chem Soc 123:2007–2017
11. Geerlings P, De Proft F, Langenaeker W (2003) Chem Rev 103:1793–1874
12. Putz MV (2012) Quantum theory: density, condensation, and bonding. Apple Academics and CRC, Taylor & Francis Group, Toronto
13. Pauling L (1932) J Am Chem Soc 54:3570–3582
14. Mulliken RS (1934) J Chem Phys 2:782–793
15. Allred AL, Rochow EG (1958) J Inorg Nucl Chem 5:264–268
16. Iczkowski RP, Margrave JL (1961) J Am Chem Soc 83:3547–3551
17. Klopman G (1965) J Chem Phys 43:S124–S129
18. Parr RG, Donnelly RA, Levy M, Palke WE (1978) J Chem Phys 68:3801–3808
19. Bartolotti LJ, Gadre SR, Parr RG (1980) J Am Chem Soc 102:2945–2948
20. Bergmann D, Hinze J (1987) Struct Bond 66:145–190
21. Allen LC (1989) J Am Chem Soc 111:9003–9014
22. Boyd RJ, Markus GE (1981) J Chem Phys 75:5385–5389
23. Bratsch SG (1985) J Chem Educ 62:101–103
24. Sen KD, Jørgensen CK (eds) (1987) Electronegativity. In: Structure and bonding, vol 66. Springer, Berlin
25. Putz MV (2003) Contributions within density functional theory with applications to chemical reactivity theory and electronegativity. Dissertation.com, Parkland

26. Putz MV, Russo N, Sicilia E (2005) Theor Chem Acc 114:38–45
27. Putz MV (2009) Int J Quantum Chem 109:733–738
28. Parr RG, Pearson RG (1983) J Am Chem Soc 105:7512–7516
29. Pearson RG (1985) J Am Chem Soc 107:6801–6806
30. Pearson RG (1986) Proc Natl Acad Sci USA 83:8440–8441
31. Pearson RG (1988) Inorg Chem 27:734–740
32. Pearson RG (1989) Org Chem 54:1423–1430
33. Gázquez JL, Ortiz E (1984) J Chem Phys 81:2741–2748
34. Berkowitz M, Parr RG (1988) J Chem Phys 88:2554–2557
35. Sen KD, Mingos DMP (eds) (1993) Chemical hardness. In: Structure and bonding, vol 80. Springer, Berlin
36. Parr RG, Gázquez JL (1993) J Phys Chem 97:3939–3940
37. Robles J, Bartolotti LJ (1984) J Am Chem Soc 106:3723–3727
38. Nalewajski RF (1984) J Am Chem Soc 106:944–945
39. Komorowski L (1987) Chem Phys 55:114–130
40. Komorowski L, Boyd SL, Boyd RJ (1996) J Phys Chem 100:3448–3453
41. Ayers PW, Parr RG (2000) J Am Chem Soc 122:2010–2018
42. Putz MV (2006) Int J Quantum Chem 106:361–389
43. Putz MV (2007) J Theor Comput Chem 6:33–47
44. Putz MV (2008) Absolute and chemical electronegativity and hardness. Nova Science, New York
45. Putz MV (2011) Curr Phys Chem 1:111–139
46. Parr RG, Szentpaly LV, Liu S (1999) J Am Chem Soc 121:1922–1924
47. Chattaraj PK, Giri S, Duley S (2011) Chem Rev 111:PR43–PR75
48. Lv S (2011) J Phys Chem A 115:8528–8531
49. Chattaraj PK, Giri S, Duley S (2012) J Phys Chem A 116:790–791
50. Lv S (2012) J Phys Chem A 116:792–795
51. Islam N, Ghosh DC (2012) Int J Mol Sci 13:2160–2175
52. Brown AJ (1892) J Chem Soc Trans 61:369–385
53. Brown AJ (1902) J Chem Soc Trans 81:373–388
54. Henri V (1901) Z Phys Chem 39:194–216
55. Henri V (1902) C R Hebd Acad Sci 135:916–919
56. Michaelis L, Menten ML (1913) Biochem Z 49:333–369
57. Haldane JBS (1930) The enzymes. Longmans-Green, London
58. Pauling L (1946) Chem Eng News 24:1375–1377
59. Voet D, Voet JG (1995) Biochemistry, 2nd edn. Wiley, New York
60. Cornish-Bowden A (1999) Fundamentals of enzyme kinetics. Butterworths, London
61. Copeland RA (2000) Enzymes. Wiley-VCH, New York
62. Schnell S, Maini PK (2003) Commun Theor Biol 8:169–187
63. Topliss JG, Costello JD (1972) J Med Chem 15:1066–1069
64. Topliss JG, Edwards RP (1979) J Med Chem 22:1238–1244
65. Topliss J (1983) Quantitative structure–activity relationships of drugs. Academic, New York
66. Seyfel JK (1985) QSAR and strategies in the design of bioactive compounds. VCH Weinheim, New York
67. Nendza M, Wenzel A (1993) Environ Toxicol Chem Suppl:1459–1470
68. Kubinyi H (1994) Pharmazie in unserer Zeit 23:158–168
69. Klopman G, Zhang Z, Woodgate SD, Rosenkranz HS (1995) Chemosphere 31:2511–2519
70. Lhuguenot JC (1995) Ann Fals Exp Chim 88:293–310
71. Hansch C, Hoekman D, Gao H (1996) Chem Rev 96:1045–1075
72. Klein DJ, Randić M, Babić D, Lučić B, Nikolić S, Trinajstić N (1997) Int J Quantum Chem 63:215–222
73. Schmidli H (1997) Chemomet Intell Lab Syst 37:125–134

74. Putz MV (ed) (2012) QSAR & SPECTRAL-SAR in computational ecotoxicology. Apple Academics and CRC, Taylor & Francis Group, Toronto
75. Parr RG, Bartolotti LJ (1982) J Am Chem Soc 104:3801–3803
76. Sanderson RT (1988) J Chem Educ 65:112–119
77. Tachibana A, Nakamura K, Sakata K, Morisaki T (1999) Int J Quantum Chem 74:669–679
78. Ghosh DC, Islam N (2011) Int J Quantum Chem 111:40–51
79. Ghosh DC, Islam N (2011) Int J Quantum Chem 111:1961–1969
80. Putz MV (2011) Int J Chem Model 3:371–384
81. Pearson RG (1973) Hard and soft acids and bases. Dowden, Hutchinson & Ross, Stroudsberg
82. Pearson RG (1990) Coord Chem Rev 100:403–425
83. Pearson RG (1997) Chemical hardness. Wiley-VCH, Weinheim
84. Chattaraj PK, Sengupta S (1996) J Phys Chem 100:16129–16130
85. Chattaraj PK, Maiti B (2003) J Am Chem Soc 125:2705–2710
86. Chattaraj PK, Sarkar U, Roy DR (2007) J Chem Educ 84:354–358
87. Putz MV, Russo N, Sicilia E (2004) J Comput Chem 25:994–1003
88. Chattaraj PK, Lee H, Parr RG (1991) J Am Chem Soc 113:1854–1855
89. Chattaraj PK, Liu GH, Parr RG (1995) Chem Phys Lett 237:171–176
90. Putz MV (2008) MATCH Commun Math Comput Chem 60:845–868
91. Chermette H (1999) J Comp Chem 20:129–154
92. Putz MV (2008) Int J Mol Sci 9:1050–1095
93. Putz MV (2011) In: Putz MV (ed) Carbon bonding and structures: advances in physics and chemistry. Springer, London, pp 1–32
94. Putz MV (2011) In: Putz MV (ed) Quantum frontiers of atoms and molecules. Nova Science, New York, pp 251–270
95. Putz MV (2012) In: Roy AK (ed) Theoretical and computational developments in modern density functional theory, chap 17. Nova Science, New York
96. Born M, Oppenheimer R (1927) Ann Physik (Leipzig) 84:457–484
97. Buckingham AD, Fowler PW, Galwas PA (1987) Chem Phys 112:1–14
98. Slater JC (1929) Phys Rev 34:1293–1322
99. Pauli W (1940) Phys Rev 58:716–722
100. Löwdin PO (1955) Phys Rev 97:1474–1489
101. Cramer CJ (2002) Essentials of computational chemistry. Wiley, Chichester
102. Jensen F (2007) Introduction to computational chemistry. Wiley, Chichester
103. Roos BO, Taylor PR, Siegbahn PEM (1980) Chem Phys 48:157–173
104. Roos BO, Sadlej AJ, Siegbahn PEM (1982) Phys Rev A 26:1192–1199
105. Roos BO, Malmqvist PÅ (2004) Phys Chem Chem Phys 6:2919–2927
106. Hückel E (1931) Z Phys 71:204–286
107. Hückel E (1931) Z Phys 72:310–337
108. Parr RG, Craig DP, Ross IG (1950) J Chem Phys 18:1561–1563
109. Roothaan CCJ (1951) Rev Mod Phys 23:69–89
110. Roothaan CCJ (1958) J Chem Phys 28:982–983
111. Purvis GD, Bartlett RJ (1982) J Chem Phys 76:1910–1919
112. Roothaan CCJ, Detrich JH (1983) Phys Rev A 27:29–56
113. Pople JA, Head-Gordon M, Raghavachari K (1987) J Chem Phys 87:5968–35975
114. Curtiss LA, Raghavachari K, Redfern PC, Rassolov V, Pople JA (1998) J Chem Phys 109:7764–7776
115. Ohlinger WS, Klunzinger PE, Deppmeier BJ, Hehre WJ (2009) J Phys Chem A 113:2165–2175
116. Heitler W, London F (1927) Z Phys 44:455–472
117. Hartree DR (1957) The calculation of atomic structures. Wiley, New York
118. Pople JA, Nesbet RK (1954) J Chem Phys 22:571–572
119. Roothaan CCJ (1960) Rev Mod Phys 32:179–185

120. Slater JC (1963) Theory of molecules and solids, vol 1, Electronic structure of molecules. McGraw-Hill, New York
121. Corongiu G (2007) J Phys Chem A 111:5333–5342
122. Glaesemann KR, Schmidt MW (2010) J Phys Chem A 114:8772–8777
123. Pariser R, Parr RG (1953) J Chem Phys 21:466–471
124. Pople JA (1953) Trans Faraday Soc 49:1375–1385
125. Streitwieser A (1961) Molecular orbital theory for organic chemists. Wiley, New York
126. Young D (2001) Computational chemistry: a practical guide for applying techniques to real world problems. Wiley, New York
127. Beaudry CM, Malerich JP, Trauner D (2005) Chem Rev 105:4757–4778
128. Hickenboth CR, Moore JS, White SR, Sottos NR, Baudry J, Wilson SR (2007) Nature 446:423–427
129. Woodward RB, Hoffmann R (1965) J Am Chem Soc 87:395–397
130. Hoffmann R, Woodward RB (1968) Acc Chem Res 1:17–22
131. Thomas LH (1927) Proc Camb Philos Soc 23:542–548
132. Fermi E (1927) Rend Accad Naz Lincei 6:602–607
133. Teller E (1962) Rev Mod Phys 34:627–631
134. Balàzs N (1967) Phys Rev 156:42–47
135. Lieb EH, Simon B (1977) Adv Math 23:22–116
136. Hohenberg P, Kohn W (1964) Phys Rev 136:B864–B871
137. Kohn W, Sham LJ (1965) Phys Rev 140:A1133–A1138
138. Dreizler RM, Gross EKU (1990) Density functional theory. Springer, Heidelberg
139. March NH (1991) Electron density theory of many-electron systems. Academic, New York
140. Koch W, Holthausen MC (2002) A chemist's guide to density functional theory, 2nd edn. Wiley-VCH, Weinheim
141. Fiolhais C, Nogueira F, Marques M (eds) (2003) A primer in density functional theory. Springer, Berlin
142. Richard MM (2004) Electronic structure: basic theory and practical methods. Cambridge University Press, New York
143. Kohanoff J (2006) Electronic structure calculations for solids and molecules: theory and computational methods. Cambridge University Press, Cambridge
144. Sholl D, Steckel JA (2009) Density Functional Theory: A Practical introduction. Wiley-Interscience, Hoboken
145. Bokhan D, Bartlett RJ (2006) Phys Rev A 73:022502
146. Derosa PA (2009) J Comput Chem 30:1220–1228
147. Runge E, Gross EKU (1984) Phys Rev Lett 52:997–1000
148. Burke K, Werschnik J, Gross EKU (2005) J Chem Phys 123:062206
149. March NH, Rubio A, Alonso JA (1999) J Phys B 32:2173–2179
150. Besley NA, Peach MJG, Tozer DJ (2009) Phys Chem Chem Phys 11:10350–10358
151. Ploetner J, Tozer DJ, Dreuw A (2010) J Chem Theory Comput 6:2315–2324
152. Nesbet RK (2002) Variational principles and methods in theoretical physics and chemistry. Cambridge University Press, New York
153. Capelle K (2010) A bird's-eye view of density-functional theory. arXiv:cond-mat/0211443v5
154. Bartlett RJ, Musial M (2007) Rev Mod Phys 79:291–352
155. Bartlett RJ, Grabowski I, Hirata S, Ivanov S (2004) J Chem Phys 122:034104
156. Bartlett RJ, Lotrich VF, Schweigert IV (2005) J Chem Phys 123:062205
157. Geerlings P, De Proft F (2001) Chem Rev 101:1451–1464
158. Deng Z, Polavarapu PL, Ford SJ, Hecht L, Barron LD, Ewig CS, Jalkanen KJ (1996) J Phys Chem 100:2025–2034
159. Jalkanen KJ, Suhai S (1996) Chem Phys 208:81–116
160. Han WG, Jalkanen KJ, Elstner M, Suhai S (1998) J Phys Chem B 102:2587–2602
161. Jalkanen KJ, Elstner M, Suhai S (2004) J Mol Struct Theochem 675:61–77
162. Silaghi-Dumitrescu R (2006) J Inorg Biochem 100:396–402

163. Silaghi-Dumitrescu R, Silaghi-Dumitrescu I (2006) J Inorg Biochem 100:161–166
164. Putz MV (2010) Int J Mol Sci 11:1269–1310
165. Dirac PAM (1929) Proc R Soc Lond A123:714–733
166. Löwdin PO (1950) J Chem Phys 18:365–376
167. Löwdin PO (1993) Int J Quantum Chem 48:225–232
168. Hoffmann R (1963) J Chem Phys 39:1397–1412
169. Boys SF (1950) Proc R Soc Lond A200:542–554
170. Szabo A, Ostlund NS (1996) Modern quantum chemistry—introduction to advanced electronic structure theory. Dover, New York
171. Clementi E, Roetti C (1974) At Data Nucl Data Tables 14:177–478
172. Hehre WJ, Stewart RF, Pople JA (1969) J Chem Phys 51:2657–2665
173. Collins JB, Schleyer PvR, Binkley JS, Pople JA (1976) J Chem Phys 64:5142–5152
174. Stewart RF (1970) J Chem Phys 52:431–439
175. Hartree DR (1928) Proc Camb Philos Soc 24:89–111
176. Hartree DR (1928) Proc Camb Philos Soc 24:111–132
177. Fock V (1930) Z Phys 61:126–140
178. Johnson BG, Gill PMW, Pople JA (1993) J Chem Phys 98:5612–5627 (Erratum: Johnson BG (1994) J Chem Phys 101:9202)
179. Slater JC (1974) Quantum theory of molecules and solids, vol 4. McGraw-Hill, New York
180. Becke AD (1988) Phys Rev A 38:3098–3100
181. Perdew JP, Chevary JA, Vosko SH, Jackson KA, Pederson MR, Sing DJ, Fiolhais C (1992) Phys Rev B 46:667–6687
182. Senatore G, March NH (1994) Rev Mod Phys 66:445–479
183. Lee C, Yang W, Parr RG (1988) Phys Rev B 37:785–789
184. Miehlich B, Savin A, Stoll H, Preuss H (1989) Chem Phys Lett 157:200–206
185. Gill PMW, Johnson BG, Pople JA, Frisch MJ (1992) Chem Phys Lett 197:499–505
186. Vosko SH, Wilk L, Nusair M (1980) Can J Phys 58:1200–1211
187. Becke AD (1997) J Chem Phys 107:8554–8561
188. Becke AD (1993) J Chem Phys 98:5648–5653
189. Noble D (2002) Nat Rev Mol Cell Biol 3:459–463
190. Crampin EJ, Schnell S, McSharry PE (2004) Prog Biophys Mol Biol 86:99–112
191. Schnell S, Turner TE (2004) Prog Biophys Mol Biol 85:235–260
192. Savageau MA (1999) Biochemical system analysis: a study of function and design in molecular biology. Addison-Wesley, Reading
193. Savageau MA (1969) J Theor Biol 25:365–369
194. Turner TE, Schnell S, Burrage K (2004) Comput Biol Chem 28:165–198
195. Lacrămă AM, Putz MV, Ostafe V (2008) In: Putz MV (ed) Advances in quantum chemical bonding structures. Transworld Research Network, Kerala, India, pp 389–419
196. Putz MV, Lacrămă AM, Ostafe V (2006) Int J Mol Sci 7:469–484
197. Putz MV, Lacrămă AM (2007) J Optoelectron Adv Mater 9:2529–2534
198. Putz MV, Lacrămă AM, Ostafe V (2007) J Optoelectron Adv Mater 9:2910–2916
199. Putz MV, Putz AM (2011) Int J Chemoinf Chem Eng 1:42–60
200. Putz MV (2011) Molecules 16:3128–3145
201. Putz MV, Putz AM (2011) In: Putz MV (ed) Quantum frontiers of atoms and molecules. Nova Science, New York, pp 539–580
202. Putz MV, Lacrămă AM (2007) Int J Mol Sci 8:363–391
203. Ogihara N (2003) Mod Drug Discov 6:28–32
204. Dirac PAM (1944) The principles of quantum mechanics. Oxford University Press, Oxford
205. Putz MV, Putz AM, Barou R (2011) Int J Chem Model 3:173–190
206. European Chemicals Agency (2011) http://www.qsartoolbox.org
207. Putz MV, Tudoran MA, Putz AM (2012) Int J Chem Model 4 (in press)
208. Schluz TW, Cronin MTD, Netzeva TI, Aptula AO (2003) J Mol Struct Theochem 622:1–22
209. Veith GD, Broderius SJ (1990) Environ Health Perspect 87:207–211

210. Hermes JLM (1990) Environ Health Perspect 87:219–255
211. Scott JA (2009) The mechanism of retene toxicity in the early life stages of fish. PhD Thesis, Queen's University, Kingston, ON
212. Safe S (1990) Crit Rev Toxicol 21:51–88
213. Bostrom CE, Gerde P, Hanberg A, Jernstrom B, Johansson C, Kyrklund T, Rannug A, Tornqvist M, Victorin K, Westerholm R (2002) Environ Health Perspect 110:451–488
214. IPCS: International Programme on Chemical Safety (1998) Selected non-heterocyclic policyclic aromatic hydrocarbons. Environmental Health Criteria 202. http://www.inchem.org/documents/ehc/ehc/ehc202.htm
215. ATSDR: Agency for Toxic Substances and Disease Registry (1995) Toxicological profile for polycyclic aromatic hydrocarbons. http://www.atsdr.cdc.gov/toxprofiles/tp.asp?id=122&tid=25
216. Ohura T (2007) Sci World 7:372–380
217. Kitazawa A, Amagai T, Ohura T (2006) Environ Sci Technol 40:4592–4598
218. Blackenship AL, Kannan K, Villalobos SA, Villeneuve DL, Falandysz J, Imagawa T, Jakobsson E, Giesy JP (2000) Environ Sci Technol 34:3153–3158
219. Denison MS, Heath-Pagliuso S (1998) Bull Environ Contam Toxicol 61:557–568
220. Hahn M (1998) Comp Biochem Physiol C Pharmacol Toxicol Endocrinol 121:23–53
221. Bock KW, Köhle C (2006) Biochem Pharmacol 72:393–404
222. Nebert DW, Roe AL, Dieter MZ, Solis WA, Yang Y, Dalton TP (2000) Biochem Pharmacol 59:65–85
223. Fernandez-Salguero PM, Ward JM, Sundberg JP, Gonzales FJ (1997) Vet Pathol 34:605–614
224. Lund AK, Goens MB, Kanagy NL, Walker MK (2003) Toxicol Appl Pharmacol 193:177–187
225. Walisser JA, Bunder MK, Glover E, Bradfield CA (2004) Proc Natl Acad Sci USA 101:16677–16682
226. Incardona JP, Collier TK, Scholz NL (2004) Toxicol Appl Pharmacol 196:191–205
227. Incardona JP, Day HL, Collier TK, Scholz NL (2006) Toxicol Appl Pharmacol 217:308–321
228. Ohura T, Morita M, Kuruto-Niwa R, Amagai T, Sakakibara H, Shimoi K (2010) Environ Toxicol 25:180–187
229. Nilsson UL, Oestman CE (1993) Environ Sci Technol 27:1826–1831
230. HyperChem 7.01 (2002) Program package. Hypercube, Gainesville

Index