

Development of Accurate and Efficient Electronic Structure Approaches for the Description of Organic Systems

THÈSE N° 7042 (2016)

PRÉSENTÉE LE 10 JUIN 2016

À LA FACULTÉ DES SCIENCES DE BASE

LABORATOIRE DE DESIGN MOLÉCULAIRE COMPUTATIONNEL

PROGRAMME DOCTORAL EN CHIMIE ET GÉNIE CHIMIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Riccardo PETRAGLIA

acceptée sur proposition du jury:

Prof. J. Vanicek, président du jury
Prof. A.-C. Corminboeuf, directrice de thèse
Dr B. Hourahine, rapporteur
Dr T. Laino, rapporteur
Prof. U. Roethlisberger, rapporteuse



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2016

Leave this world a little better
than you found it.
— Robert Baden-Powell

To my family...

Abstract

The overarching objective of this thesis is extending and adapting the set of computational tools available for describing molecular precursors of organic semiconductors. The research presented within develops adhering to three principle goals: (1) provide accurate energies and geometries for large-scale assemblies at an affordable computational cost; (2) provide tools that thoroughly explore and map free energy landscapes and emphasize the importance of this comprehensive mapping; (3) improve density functional theory (DFT) descriptions of precursors to charge carrier molecules which are currently quite challenging.

The Self Consistent-Charge Density Functional Tight Binding (SCC-DFTB), particularly the most recent variant (DFTB3), provides an excellent balance between accuracy and efficiency needed to study large molecules and to perform molecular dynamic simulations. As DFTB approximates the DFT Hamiltonian it suffers from the same principle shortcomings, including neglecting dispersion interactions. Such interactions are crucial for establishing the proper relative orientations of organic molecules in aggregate, a property of fundamental importance for elucidating conduction properties. Building upon our laboratories experience with the dDsC dispersion correction, an *a posteriori* pairwise dispersion correction that depends upon Mulliken charges, dDMC, was developed specifically for DFTB. During the course of this work, a caveat in the DFTB parameterization of sulfur was identified that caused sulfur containing molecules to exhibit strong non-covalent binding, even in the absence of a dispersion correction.

From a computational perspective, detecting all conformations present in a chemical space and recognizing those structures that are the most chemically relevant represents a significant challenge. To address this problem, we have borrowed a technique usually used for molecular mechanics simulation and combined it with DFTB (REMD@DFTB3). Using this technique allows through exploration of various chemical systems at a quantum level in which bond breaking and formation is possible, thereby allowing description that surpass the typically employed static picture.

Although the DFTB approach is unquestionable practical, certain situations require more sophisticated and accurate treatments. Radical cation dimers are, for instance, illustrative as models for organic charge carrier molecules, yet their computational description possesses inherent challenges. Although accurate, post-HF methods are too expensive for routine use, while DFT approaches suffer from their typical shortcomings (self-interaction error, lack of dispersion interactions, missing static correlation). To overcome these problems, a new density functional, ω B97X-dDsC, was developed in which the parameters of the dDsC correction and those of ω B97X were fitted together. Preliminary examination of the overall performance of ω B97X-dDsC is very promising.

Key words: density functional tight binding; London dispersion; dispersion correction; radical cation dimer; replica exchange molecular dynamics; density functional theory

Compendio

L'obiettivo di fondo di questa tesi è estendere ed adattare l'insieme degli strumenti computazionali disponibili per descrivere precursori molecolari di semiconduttori organici. La ricerca si sviluppa perseguendo tre obiettivi principali: (1) fornire energie e geometrie accurate per estesi complessi supramolecolari a costi computazionali ragionevoli; (2) fornire strumenti per l'esplorazione di estesi panorami configurazionali; (3) migliorare la descrizione fornita dalla Teoria del Funzionale Densità (DFT) riguardo molecole conduttrici di carica.

La Self Consistent-Charge Density Functional Tight Binding (SCC-DFTB), ed in particolare la variante più recente (DFTB3), fornisce un eccellente bilancio tra l'accuratezza e l'efficienza necessarie per studiare grandi molecole ed eseguire simulazioni di dinamica molecolare. Poiché la DFTB approssima l'Hamiltoniano TFD, essa stessa soffre delle medesime mancanze, tra cui il trascurare le interazioni di dispersione. Tali interazioni sono cruciali per stabilire l'orientazione relativa delle molecole organiche all'interno di aggregati: una proprietà fondamentale per lo studio della conducibilità tra molecole. Grazie all'esperienza già acquisita all'interno del nostro laboratorio con la correzione di dispersione dDsC, abbiamo sviluppato una nuova correzione *a posteriori*, dipendente dalle cariche di Mulliken, specifica per DFTB: la dDMC. Approfondendo la conoscenza dello schema DFTB, abbiamo trovato un problema nella parametrizzazione dello zolfo che causava attrazioni spurie tra lo zolfo e tutti gli altri atomi legati non-covalentemente, anche in assenza di correzioni di dispersione.

Da un punto di vista computazionale, la determinazione di tutte le configurazioni presenti in un determinato spazio chimico e il riconoscimento delle strutture più rilevanti rappresenta una importante sfida. Abbiamo intrapreso tale sfida prendendo in prestito una tecnica solitamente utilizzata per simulazioni nell'ambito della meccanica molecolare e l'abbiamo combinata con la DFTB (REMD@DFTB3). Questa tecnica permette l'esplorazione di vari sistemi chimici a livello quantistico, cosicché la rottura e la formazione dei legami sia possibile. Tale descrizione permette di superare le tecniche statiche impiegate tradizionalmente nello studio dei sistemi organici.

Benché l'approccio DFTB sia senza dubbio pratico, alcune situazioni richiedono l'uso di strumenti più sofisticati ed accurati. I dimeri cationici radicalici sono modelli illustrativi di molecole trasportatrici di cariche. Nonostante la loro importanza, la descrizione computazionale di tali sistemi risulta molto complessa: i metodi *post-HF*, sebbene molto accurati, sono troppo dispendiosi per l'uso quotidiano mentre la TFD soffre dei suoi tipici problemi (errore di auto-interazione, mancanza della dispersione, mancanza di correlazione statica). Al fine di superare questi problemi, abbiamo sviluppato un nuovo funzionale densità, l' ω B97X-dDsC, nel quale i parametri che regolano la correzione di dispersione sono calcolati insieme a quelli dell' ω B97X. I risultati preliminari dell' ω B97X-dDsC sono molto promettenti.

Parole chiave: Density functional tight binding; interazioni di London; correzioni di dispersione; dimeri cationici radicalici; dinamica molecolare con scambio di repliche; Teoria del funzionale densità

Contents

Abstract (English/Italiano)	i
1 Introduction	1
2 Theory	5
2.1 Density Functional Tight-Binding	5
2.2 <i>A Posteriori</i> Atomic-Pairwise Dispersion Corrections	8
2.3 Replica Exchange Molecular Dynamics	10
2.3.1 Parallel Tempering	10
2.3.2 Hamiltonian Replica Exchange Molecular Dynamics	13
3 A Caveat on SCC-DFTB and Non-covalent Interactions Involving Sulfur Atoms	15
3.1 Introduction	15
3.2 Computational Details	16
3.3 Critical Failure of Sulfur at DFTB level	16
3.4 Conclusion	21
4 A fast charge-dependent atom-pairwise dispersion correction for DFTB3	23
4.1 Introduction	23
4.2 Theory	25
4.2.1 Gradient	28
4.2.2 Adjustable parameters and training set	28
4.3 Computational Details	29
4.4 Results	31
4.5 Conclusions	35
5 Putting forth REMD as a tool to solve problems in organic chemistry	37
5.1 Introduction	37
5.2 Computational Details	40
5.3 Results and Discussions	43
5.3.1 Dithiacyclophane	43
5.3.2 Cope rearrangement of semibullvalene	44

Contents

5.3.3	Molecular Rotor	47
5.3.4	Cinchona alkaloid	49
5.4	Scope and Limitations	52
5.5	Toward a More General REMD@DFTB3	52
6	Overcoming the drawbacks of standard density functional approximations	55
6.1	Introduction	55
6.2	Theory	57
6.3	The Algorithm	59
6.4	Computational Details	61
6.5	Results and Discussions	61
6.5.1	Functional development and training	62
6.5.2	Validation	64
6.6	Conclusion	72
7	General Conclusions and Outlook	75
A	Density Functional Theory	79
A.1	The Hohenberg-Kohn theorems	80
A.2	Exchange-correlation functionals	81
B	REMD Error Analysis	83
	Bibliography	101
	Curriculum Vitae	103

1 Introduction

The description and design of molecules through modern quantum chemical approaches is attractive for a variety of reasons: novel insights at the subatomic level can be provided, environmental impact are negligible, man power can be saved and the exploration of larger chemical space is facilitated. Still, the *in-silico* exploration of novel molecular and materials properties is hampered by several limitations involving, for instance, flaws in the electronic structure methods, poor statistical sampling or simply inaccessible computational costs.¹⁻⁶ In this work, we aim at delivering electronic structure methods that accurately and efficiently describe molecules of relevance to the field of organic electronics.

The importance of electronic devices has driven technological breakthroughs affecting all aspects of our everyday life. Aside from silicon and inorganic congeners, growing interest exists for developing a new generation of devices based on π -conjugated polymers and oligomers.⁷⁻⁹ The field was pioneered by the work of MacDiarmid, Heeger and Shirakawa “Synthesis of electrically conducting organic polymer: halogen derivatives of polyacetylene” who were awarded the 2000 Nobel Prize.¹⁰

The benefits of these species potentially include reduced fabrication costs along with novel functionalities (e.g., mechanical flexibility, transparency, impact resistance).¹¹⁻¹³ The performance of organic semiconductors (i.e., charge-carrier mobility in field-effect transistors) depends heavily upon the organization and on the electronic structure of the π -conjugated molecules (or chains) at the molecular level. To achieve the full potential of such materials, technological developments require fine-tuning of the specific intermolecular interactions spanning small ranges of distances, lateral displacements, π -conjugated moiety orientation but also long-range organization (i.e., morphology).^{9,14,15} In this respect, computational methods can significantly speed up the discovery pace of these materials although various challenges need to be overcome to achieve an accurate and reliable description of their structures, energies and charge transport properties.

Chapter 1. Introduction

Our group has already engaged in the development of state-of-the-art theoretical methods for analyzing and describing phenomena underpinned by π -conjugated molecules.^{16–20} In this thesis, we extend and adapt the computational toolbox to improve the quantum chemical description of precursors to organic semiconductors with special attention placed on motifs made of thiophene ring units. The specific objectives of this thesis are (1) providing accurate energies and geometries for large-scale assemblies using low-cost electronic structure levels; (2) highlighting the importance of mapping the free energy landscapes and combining enhanced sampling techniques with these same low-cost levels and (3) improving the density functional theory (DFT) description of the challenging π -dimer radical cations, which constitute a model for charge carriers.

The upcoming chapter, **Chapter 2**, summarizes the main theoretical frameworks and methods at the center of the proposed work. We first introduce Self Consistent-Charge Density Functional Tight Binding (SCC-DFTB)^{21,22} and especially the most recent variant (DFTB3),^{23–25} which provides the balance between accuracy and efficiency desirable herein. Since the DFTB formalism approximates the density functional theory Hamiltonian it suffers from the same shortcomings such as the neglect of long-range correlation (i.e., London dispersion).²⁶ The inability of DFTB to accurately describe London dispersion is one of our focuses and some of the existing pragmatic solutions are already presented in this first technical chapter. The chapter also provides an overview of Replica Exchange Molecular Dynamics (REMD)^{27–29} techniques, which come from biological simulations, and which we combine with DFTB3 in the subsequent chapters.

The research core of this thesis starts with **Chapter 3**, which illustrates the dramatic consequences of a caveat associated with the DFTB parameters of sulfur.³⁰ This spurious parameterization causes the sulfur-containing molecules to strongly bind non-covalently even in the absence of a dispersion correction. This realization was important since the problem hampered the immediate development and application of a dispersion-correction scheme we designed for DFTB. In 2011, LCMD formulated a DFT-based dispersion correction (dDsC)³¹ that is dependent upon the electron density. A targeted advantage of dDsC was its ability to simultaneously improve the performance of a variety of standard density functionals for typical intra- and intermolecular interactions and the situations involving highly polarized systems. The objective of **Chapter 4** is to take advantage of the dDsC philosophy and devise an analogue correction based on Mulliken charge³² that can be combined with DFTB. **Chapters 5** exploits the efficiency of DFTB to go beyond the static picture and provide a tool that can reveal useful chemical insights.³³ In particular, we borrow a computational technique originally conceived to be used in the context of biological simulations, together with empirical force fields, and apply it to organic chemical problems. This technique, Replica-Exchange Molecular Dynamics (REMD), permits thorough exploration of the potential energy surface.

We combined REMD with density functional tight binding, the level of accuracy and efficiency that is necessary for enhancing the statistical sampling in quantum chemistry problems.

Chapter 6 tackles the highly challenging electronic description of radical π -dimer cationic systems that are the prototype model of organic conductors. Within the context of DFT, the energy and geometries of these systems are very poorly described owing to their sensitivity to both the delocalization error and the missing dispersion. This chapter tries to overcome these challenges through the parameterization of a novel generation of ω B97X-based functional. In this so-called ω B97X-dDsC variant, 16 functional parameters are trained jointly with the dDsC dispersion correction parameters to ensure a subtle interplay between the two sources of error.³⁴ In conjunction with proposing a robust state-of-the-art DFT approximation, **Chapter 6** also discusses the cornerstone of the work that is an efficient optimization algorithm that provides a clever way to optimize a large set of parameters through proposing tricks to improve its computational efficiency.

Finally, **Chapter 7** concludes this thesis and summarizes the main findings and future prospects.

2 Theory

This chapter provides an overview of the theoretical methods relevant to the rest of the thesis. In particular we introduce the basics of density functional tight binding that is the most used electronic approach here. We also present the fundamentals of replica exchange molecular dynamics that are at the center of the last two chapters. Finally, a brief introduction on the combination of atom-pairwise dispersion corrections with electronic structure approaches terminates this chapter.

2.1 Density Functional Tight-Binding

As will be clear in the following chapters, our work necessitates the use of an efficient electronic structure approach enabling the fast computations of energy and forces of about thousands structures or of large molecules. In light of these considerations, we decided to rely upon the density functional tight-binding (DFTB) framework, as an approximation to the density functional theory Hamiltonian. The DFTB scheme exists under three different flavors^{22–24,35,36} In its original and simplest form the total energy in DFTB is expressed as in tight-binding models³⁶

$$E = E_{el} + E_{rep} = \sum_{i=1}^M \varepsilon_i + \frac{1}{2} \sum_{\alpha\beta}^N V(R_{\alpha\beta}) \quad (2.1)$$

where ε_i are eigenvalues of a Schrödinger-like equation and $V(R_{\alpha\beta})$ is a short-range pairwise repulsion between atoms that depends exclusively on the inter-atomic distance.

The electronic terms, given as E_{el} , differ among the DFTB variants, in the description of hydrogen bond properties and the charge transfer between bonding atoms. The self-consistent charge density functional tight-binding scheme, developed by Elstner and co-workers,²² accounts for the charge transfer between bonded atoms, which is absent in the

Chapter 2. Theory

Not-Charge-Consistent-DFTB (NCC-DFTB) approach. The Self-Consistent-Charge DFTB approach (SCC-DFTB)²² enables the description not only of solid state materials but also of more complex systems such as biological macromolecules and nanosystems.³⁷ The most recent DFTB3^{23,24} formalism shows improved performance for systems in which atoms carry significant partial charges.³⁸

In the most recent DFTB3 variant, the electronic term is defined as follows

$$E_{el} = E_{NCC} + E_{SCC} + E_{3rd} \\ = \sum_i^M \sum_{\alpha\beta} \sum_{\mu\in\alpha} \sum_{\nu\in\beta} c_{\mu i} c_{\nu i} \mathcal{H}_{\mu\nu}^0 + \frac{1}{2} \sum_{\alpha\beta}^N \Delta q_{\alpha} \Delta q_{\beta} \gamma_{\alpha\beta} + \frac{1}{3} \sum_{\alpha\beta}^N \Delta q_{\alpha}^2 \Delta q_{\beta} \Gamma_{\alpha\beta} \quad (2.2)$$

where N is the number of atoms in the system, α and β are atom indexes, and M is the number of occupied molecular one-electron orbitals i that are expanded within the LCAO ansatz using a suitable set of constrained atomic orbitals

$$\psi(\mathbf{r}) = \sum_{\nu} c_{\nu i} \phi_{\nu}(\mathbf{r} - \mathbf{R}_{\alpha}) \quad (2.3)$$

These constrained atomic orbitals are determined by solving a modified set of Schrödinger equations for free neutral pseudoatoms

$$\left[\mathcal{T} + V_{nuc l} + V_H + V_{XC} + \left(\frac{r}{r^{wf}} \right)^2 \right] \phi_{\nu} = \varepsilon_i \phi_{\nu} \quad (2.4)$$

The first three terms of the equation are the kinetic energy, the nuclear repulsive potential and the Hartree potential. V_{XC} is the exchange-correlation potential that, generally, is expressed as PBE XC potential.^{39,40} r^{wf} is the wave function compression radius that is often taken as twice the covalent radius (e.g. for sulfur: $r^{wf} = 2\text{\AA} = 2r^{cov}$). The Hamiltonian matrix elements are written in a two-center approximation as

$$\mathcal{H}_{\mu\nu}^0 = \begin{cases} \varepsilon^{\text{free atom}} & \text{if } \alpha = \beta, \mu = \nu \\ \langle \phi_{\mu}^{\alpha} | \mathcal{T} + V[\rho_{\alpha}^0 + \rho_{\beta}^0] | \phi_{\nu}^{\beta} \rangle & \text{if } \alpha \neq \beta, \mu \neq \nu \\ 0 & \text{if } \alpha = \beta, \mu \neq \nu \end{cases} \quad (2.5)$$

The $\varepsilon^{\text{free atom}}$'s are computed neglecting the extra term (r/r^{textwf}) in eq. 2.4 so that this formulation of $\mathcal{H}_{\mu\nu}^0$ ensures the correct limit for free atoms. ρ_i^0 represents free atom compressed densities computed using eq. 2.4 by substituting r^{wf} with r^{dens} whereas the densities were originally determined directly from the compressed wave functions, the use of alternative constrained radii (generally larger, e.g., 4.76Å for sulfur) has been shown to improve results (see ref.²³ and reference cited therein). The choice of r^{dens} is not trivial but is known to be critical for obtaining good energies.³⁶ In order to reduce computational demand, the Hamilto-

nian and overlap matrix elements are precomputed using a minimal basis set and tabulated for different distances in a Slater-Koster approach.⁴¹ Using compressed wave functions and densities improves the portability of these parameters.²²

The DFTB approach is made self-consistent by improving the NCC-DFTB scheme with

$$E_{\text{SCC}} = \frac{1}{2} \sum_{\alpha\beta}^N \Delta q_{\alpha} \Delta q_{\beta} \gamma_{\alpha\beta} \quad (2.6)$$

where Δq_i is the charge difference with respect to the neutral state with density ρ_i^0 and $\gamma_{\alpha\beta}$ is a function describing the interactions between the non-neutral atoms α and β . γ is chosen such that at large inter-atomic distances E_{SCC} tends to a pure Coulombic interaction energy between charges Δq_i .²² $\gamma_{\alpha\alpha}$ is approximated as the difference between the ionization potential and the electron affinity of the free atom. This difference is related to the chemical hardness or to the Hubbard parameters (U); hence $\gamma_{\alpha\beta}$ is approximated as a function of U_{α} , U_{β} and $R_{\alpha\beta}$ even at medium distances.²² To improve the description of hydrogen bonds, a suitable scaling term is added to $\gamma_{\alpha\beta}$ at covalent distance when one of the atoms involved is hydrogen.²³ Since Δq_i are computed with the Mulliken scheme, where the charge depends on the coefficients c_{μ} and c_{ν} , the SCC-DFTB scheme is self-consistent.²²

Choosing the Hubbard parameters independent from the atomic charge could be a severe limitation.^{42,43} To improve the description of systems with large inter-atomic charge transfer, the following “third order” term is introduced^{23,24}

$$E_{\text{3rd}} = \frac{1}{3} \sum_{\alpha\beta}^N \Delta q_{\alpha}^2 \Delta q_{\beta} \Gamma_{\alpha\beta} \quad (2.7)$$

where the Hubbard parameters depend on Δq_i through the function $\Gamma_{\alpha\beta}$. Hubbard parameters needed in SCC-DFTB, here, are replaced by their derivative with respect to the charge. In practice, this term is a correction on the energy that accounts for the dependence on atomic charges of the Hubbard parameters.

The repulsive term accounts for all approximations made in the electronic part. In practice, $V(R_{\alpha\beta})$ is a spline^{24,25,44} (or polynomial²²) function fit to the difference in energies between DFT (usually the B3LYP functional) and the electronic term of DFTB at various bond lengths²² for an appropriate, usually small, set of molecules.

None of the DFTB variants are considered as “semi-empirical” in sense that the parameters are not fitted to experimental data: some (such as the Hubbard parameters and derivatives) are directly obtained from the DFT computations while others (such as the repulsive potential) are fitted to DFT results.^{22,42–44} This makes DFTB a DFT-based approximation, affected by the same deficiencies as DFT as it will be discussed in Chapter 3.

2.2 *A Posteriori* Atomic-Pairwise Dispersion Corrections

When performing computation on organic molecule, a reliable description of intermolecular interactions requires an accurate account of dispersion forces. Unfortunately, only computationally very expensive methods (e.g. *post*-HF) are able to account for dispersion interactions. To avoid the limitations implicit with so expensive approaches, a series of techniques have been developed in the last decades to include weak forces in Hartree-Fock (HF)⁴⁵ and Density Functional Theory (DFT).^{46,47}

On the DFT side, the one we will focus on, the main responsible for the lack of dispersion interactions are the inevitable approximations in the exchange-correlation functional, often separated in an exchange functional,^{48–51} which cannot be expressed analytically for systems beyond an homogeneous electron gas,⁵² and a correlation functional, modeled on accurate Quantum Monte Carlo simulations and on low and high-density asymptotic limits.⁵³ In particular, the missing of non-locality in the correlation approximation⁵⁴ induce the lack of dispersion interactions²⁶ which can be overcome with three main strategies: (1) improving the non-locality of the correlation functional to eliminate the root of the problem;^{55–60} (2) fitting very flexible functional forms^{61–65} (M06-2X);⁶⁶ or (3) incorporating a posteriori energy corrections.^{31,67–70} Another category should be devoted to the Dispersion Corrected Atom Centered Potentials (DCACP), initially developed for the plane-waves approach^{71–73} and later generalized to Gaussian basis sets.^{74,75} This approach consists in adding a non-local atom-centered potential that emulates dispersion interactions just like pseudo-potentials account for core electrons in plane-waves^{76,77} or effective core potentials account for relativistic effects in Gaussian basis sets.^{78,79} DCACP tends to correct the symptoms of the non-locality of the correlation hole, similarly to the *a posteriori* corrections. However, the modified basis set will influence the self-consistent computation, thus, the final electron density. For this reason DCACP should not be included in the *a posteriori* dispersion correction.

Dispersion interactions are incorporated in all post-HF methods (from MP2). A possible approach to eliminate the root problem in DFT is to take advantages from wave function theory to correct the locality of the correlation functional. An example are the double hybrid functionals including a percentage of many-body second order perturbation theory correlation energy.^{59,60} However, despite the higher computational cost, an additional dispersion correction is recommended.^{80–82} Another class of functionals, alternatives to the double hybrids, exist.^{55–58} These functionals model the dispersion interactions as coupled local oscillators having a frequency determined by the local density and its derivatives. Despite the approach is very expensive in principle, some approximations^{83,84} make the approach routinely applicable (e.g. vdW-DF family^{57,58} and VV family^{55,55}). The Local Response for Dispersion (LRD)^{85,86} formalism is a further approximation on the non-local Density Functional Approximations (DFAs): the oscillators are considered independent so that their interaction can be simplified

2.2. *A Posteriori* Atomic-Pairwise Dispersion Corrections

in an atom pairwise sum, leading to the same scheme found in the *a posteriori* dispersion corrections.

The *a posteriori* dispersion corrections are a more general approach developed to account for dispersion interactions at HF level.^{87–90} The idea is to correct the electronic structure energy with a post-SCF contribution. Most of the *a posteriori* dispersion corrections can be written as a sum of atomic pairwise contributions^{31,67–70,91–93}

$$E_{disp} = - \sum_{i=1}^{N_{at}} \sum_{j>i}^{N_{at}} f_d(R_{ij}) \frac{C_6^{i,j}}{R_{ij}^6} \quad (2.8)$$

where the indexes i and j run over all the atoms in the system, R_{ij} is the internuclear distance and $C_6^{i,j}$ is the dispersion coefficient associated with the atom couple. $f_d(R_{ij})$ is a damping function whose task is turning the dispersion correction off at short internuclear distances, where the electronic structure method can describe more accurately the interaction. Moreover, without a damping function, the dispersion correction would diverge when $R_{ij} \rightarrow 0$.

Dispersion coefficients and damping functions can be defined in many different way. One of the most popular dispersion correction DFT-D3,⁶⁹ developed by Grimme and providing parameters for most of the elements in the periodic table, interpolates the value of the C_6 from a defined set of data and features, in its original version, the Head-Gordon's power law damping function³⁴

$$f_{HG}(R) = - \frac{1}{1 + a(R/R_0)^{-q}} \quad (2.9)$$

with a and q adjustable parameters fitted to reproduce the results of a representative ensemble of more accurate data. The exchange-correlation functional adopted in the DFT computation strongly influences the two parameters, requiring a different parameterization for each functionals. Also the Fermi damping has been successfully applied in this field^{67,70,94,95}

$$f_F(R) = - \frac{1}{1 + e^{-d(R/R_0-1)}} \quad (2.10)$$

where d determines the steepness of the switching function and R_0 is the vdW distance: at which the function goes to zero. Tang and Toennies' damping function (TT)^{96,97} have been successfully used for dispersion interaction of several noble gas and metal atom pairs

$$F_{TT} = - \left(1 - e^{-b \cdot R} \sum_{k=0}^6 \frac{(b \cdot R)^k}{k!} \right) \quad (2.11)$$

b is a fitted parameter to define the distance at which the damping function goes to zero.

Chapter 4 presents a dispersion correction developed within the SCC-DFTB framework (dDMC)⁹⁸ while in Chapter 6 the Density Dependent Dispersion Correction (dDsC), developed few years ago by Corminboeuf *et al.*,³¹ is parameterized jointly with the ω B97X³⁴ exchange-correlation functional to increase the performance toward very interesting despite very challenging systems at DFT level: ionic dimers and charge-transfer complexes.²⁰ Both the dispersion corrections take advantage from the TT damping function further damped with a Fermi function to increase the flexibility at medium range. dDsC presents electron density dependence on both the dispersion coefficients (through a simplified version of the XDM formalism)^{99–104} and the damping function using the Hirshfeld¹⁰⁵ (overlap) population to distinguish non-bonded regions from bonded atom pairs, eliminating the correction at covalent distances. More details on dDsC can be found in refs.^{31,106–108}

2.3 Replica Exchange Molecular Dynamics

The quantum chemistry community often relies on high-level static computations, ignoring the entropic contributions. In this thesis we target the modeling of organic functional units and molecular architectures (e.g., shuttles, rotors, switches^{33,109–111}) that would profit from both an accurate quantum description and the consideration of the full entropic effects. As a prelude to Chapters 5, we here summarize the theory behind replica exchange molecular dynamics that is combined with DFTB. Replica Exchange Molecular Dynamics (REMD)^{27–29,112} aims at improving the statistical mapping of the phase space for systems presenting a rough potential energy surface: allowing the exploration of basins separated by large barriers and retrieving the right relative free energies among all the basins.

A REMD simulation consists in performing a series of energetically independent simulations (namely *replicas*) of the same system in different equilibrium conditions and allowing them to occasionally exchange their configurations in a way that still ensures a canonical sampling within each simulation.

2.3.1 Parallel Tempering

The most common version of REMD is the Parallel Tempering (PT)^{113,114} where the series of replicas differ for the ensemble temperatures: from the lowest T_0 , the temperature of interest, to a higher temperature T_M , large enough to enhance energy-barriers crossing. Several replicas use temperature in between T_0 and T_M (*vide infra*).

The overall equilibrium distribution is

$$P(X_0, T_0, \dots, X_M, T_M) = \prod_{r=0}^M P(X_r, T_r) \quad (2.12)$$

2.3. Replica Exchange Molecular Dynamics

where $P(X_r, T_r)$ is the canonical distribution of a single replica r for a given state X_r and a given temperature T_r

$$P(X_r, T_r) = \frac{e^{-\beta_r V(X_r)}}{Z_r} \quad (2.13)$$

with $\beta_r^{-1} = k_B T_r$, $V(X_r)$ the potential energy and $Z_r = \int e^{-\beta_r V(X_r)} dX_r$ is the configurational partition function for the r -th replica. To ensure that each replica will maintain the canonical distribution during the REMD, the detailed balance must hold. Since the replica are energetically independent the detailed balance requires

$$P(X_n, \beta_n, X_m, \beta_m) W(X_n, \beta_n; X_m, \beta_m) = P(X_n, \beta_m, X_m, \beta_n) W(X_n, \beta_m; X_m, \beta_n) \quad (2.14)$$

$P(X_n, \beta_n, X_m, \beta_m)$ is the probability to get the two replica at given temperatures and configurations, $W(X_n, \beta_n; X_m, \beta_m)$ is the probability of exchanging the two replicas. The right side of the equation is consistently the opposite of the left side. The detailed balance is satisfied by accepting exchanges with the Metropolis acceptance criteria

$$P(acc) = \min(1, e^{\Delta\beta\Delta V}) \quad (2.15)$$

with $\Delta V = V(X_m) - V(X_n)$ and $\Delta\beta = \beta_m - \beta_n$.

Assuming that $T' > T$ ($\Delta\beta < 0$),¹¹⁵ the total acceptance probability can be factorized as sum of the joint probability to observe a $\Delta V > 0$ and accept the exchange ($P(acc, \Delta V > 0)$) and the analogous probability when $\Delta V < 0$ ($P(acc, \Delta V < 0)$)

$$P(acc) = P(acc, \Delta V > 0) + P(acc, \Delta V < 0) \quad (2.16)$$

As a consequence of the detailed balance, at equilibrium the condition that $P(acc, \Delta V > 0) = P(acc, \Delta V < 0)$ must hold. Equation 2.15 and the assumption that $\Delta\beta < 0$ make the probability of accepting any exchange with $\Delta V < 0$ unitary. As a result, the probability of accepting an exchange is two times the probability that the higher temperature replica has a lower energy than the lower temperature replica

$$P(acc) = 2P(acc, \Delta V < 0) = 2P(\Delta V < 0) \quad (2.17)$$

Assuming a Gaussian distribution for the potential energy (see fig. 2.1) the acceptance probability is

$$P(acc) = \text{erfc}\left(\frac{\langle V \rangle'_T - V^*}{\sqrt{2}\sigma'_T}\right) \quad (2.18)$$

where $\langle V \rangle_{T'}$ is the mean potential energy at the higher temperature T' , V^* is the value of the potential energy such that $P_T(V^*) = P_{T'}(V^*)$ and $\sigma_{T'}$ is the variance of the potential energy distribution at T' . Equation 2.18 displays how the exchange probability depends on

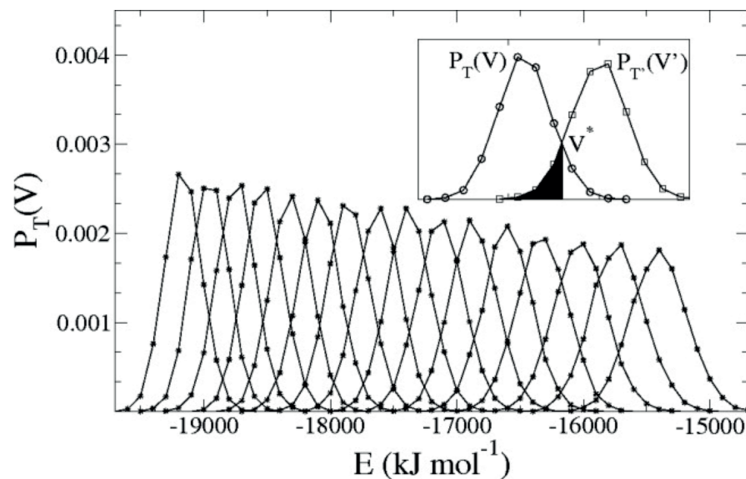


Figure 2.1: Potential energy distribution to determine the acceptance probability for an exchange of replicas. Taken from ref. ¹¹⁵

the temperature difference of the replicas: increasing the temperature increase the mean potential energy therefore increase the distance between the two distribution and decrease the overlap between them. ^{116,117} Also the number of particles in the simulated system (N) plays a role in determining the exchange probability since $\sigma_T \propto \sqrt{N}$ while $\langle V \rangle_T \propto N$: the higher the number of particles, the lower is the acceptance probability so more replicas are needed to span a given temperature range. Since the acceptance probability depends on the difference in temperatures and in potential energies between the two replicas, several replicas at intermediate temperatures allow for an indirect exchange between replicas at very different temperatures. Since the temperature changes among replicas, a rescaling of the atomic velocities is needed each time an exchange is accepted.

The other important parameter in the PT (as well as in other REMD schemes) is the frequency of attempted exchanges. The computation of the acceptance probability is very cheap since it depends only on the potential energies which are computed together with the forces, thus, at each step of the simulation. A low frequency of the attempted exchanges could deteriorate the statistical sampling ¹¹⁸ while, in principle, there are not negative effects in using an high frequency other than an increase in the computational cost. A typical method is to increase the frequency until the overhead due to the acceptance probability computation is negligible. ¹¹⁹

2.3.2 Hamiltonian Replica Exchange Molecular Dynamics

A more flexible alternative to the PT is the Hamiltonian Replica Exchange Molecular Dynamics (H-REMD).^{117,120} In H-REMD each replica is characterized by a different Hamiltonian rather than by a different temperature. Many different scheme have been proposed.^{117,120–124} In the most basic version this is obtained by scaling the potential energy (and therefore the forces) by a factor c_r (normally defined in the interval (0, 1]) different in each replica.

$$V_r(X) = c_r V(X) \quad (2.19)$$

where $V_r(X)$ is the potential energy of the replica r -th and $V(X)$ the unscaled (original) potential energy of the same phase point. The probability of a configuration X in the replica r is

$$P_r(X) = \frac{e^{-\beta c_r V(X)}}{Z_r} \quad (2.20)$$

with $Z_r = \int e^{-\beta c_r V(X)} dX$ being the configurational partition function. As thermodynamics is concerned, scaling the potential energy of a canonical system is equivalent to scaling the temperature since

$$e^{-\beta c V} = e^{-\beta' V} \quad (2.21)$$

with $\beta' = c\beta$ then, being $\beta^{-1} = k_B T$ and $(\beta')^{-1} = k_B T'$, $T' = T/c$ must hold.

In the case the Hamiltonian is a sum of terms, the c_r can be different for each term

$$V_r(X) = \sum_{i=1}^K c_{ri} v_i(X) = \mathbf{c}_r \mathbf{v}(X) \quad (2.22)$$

such that each replica differs from the others by a k -dimensional scaling vector \mathbf{c}_r . This scheme is at the origin of the solute tempering REM.^{120,125}

Rather than exchanging temperatures, in H-REMD replicas exchange their scaling vectors. The acceptance probability of an exchange between two replica n and m at temperature β

$$P(acc) = \min \left\{ 1, e^{\beta \Delta \mathbf{c} \Delta \mathbf{v}} \right\} \quad (2.23)$$

with $\Delta \mathbf{c} = \mathbf{c}_n - \mathbf{c}_m$ and $\Delta \mathbf{v} = \mathbf{v}(X_n) - \mathbf{v}(X_m)$ satisfy the detailed balance condition.

The main advantages in using H-REMD instead of PT are (1) it offers the possibility to avoid atom velocities rescaling following a replicas exchange, (2) since the atomic velocities distribution is the same in all the replicas, there is no need to adapt the time step to the temperature,

(3) the different contributions summed in the Hamiltonian can be scaled to a different magnitude allowing to enhance the sampling of only parts/interactions of the system reducing the necessary number of replica to span from the lowest temperature to the highest, (4) if entropic barriers are found, they would increase with the temperature making PT useless while H-REMD can easily overpass them.¹²⁶

In the present work, we will demonstrate how to couple REMD techniques with electronic structure approaches, so that to enhance the sampling of conformations and reaction paths of organic systems.

3 A Caveat on SCC-DFTB and Non-covalent Interactions Involving Sulfur Atoms

This chapter is based on following publication:

Petraglia, Corminboeuf, C.

A Caveat on SCC-DFTB and Non-covalent Interactions Involving Sulfur Atoms.

J. Chem. Theory Comput. **2013**, 9, 3020

3.1 Introduction

Non-covalent interactions play key roles in many areas of chemistry, including being of crucial importance for self-assembly and molecular recognition processes. More specifically, interactions involving sulfur-containing heterocycles currently receive significant interest due to their primary role in the field of organic electronics. For example, aggregation between oligothiophene chains provides the core functionality that makes plastic electronics an attractive and viable alternative to traditional silicon analogs.^{8,127–132} In petroleum chemistry, the situation is reversed as these same dispersion forces are detrimental: thiophenes are a key component of asphaltenes, where their interaction leads to increased oil viscosity that alters oil upgrading,¹³³ rather than providing functionality. Non-covalent interactions involving sulfur atoms are also significant in the biochemical world,^{127,134} where, for instance, sulfur- π interactions assist in forming protein tertiary structures.^{134–138}

The computational analysis of organic electronics suffers from two main challenges; the first involves modeling of nanoscale oligothiophene assemblies themselves (e.g., what are the structural characteristics? Figure 3.1¹⁷). The second challenge is providing a reliable characterization of electronic properties and functions.^{14,139–141} In principle, these structural and electronic aspects could be addressed conjointly using the density functional theory

framework,⁴⁷ however, quantum descriptions of systems of this size remain prohibitive due to computational cost. Thus, the development and application of more approximate schemes that enable the treatment of thousands of atoms are indispensable.^{92,142} The density functional tight binding³⁶ (DFTB) approach (especially in its SCC-DFTB²² and DFTB3^{23–25} variants – see Section 2.1), based on DFT, can simulate large systems¹⁴² with reasonable accuracy^{143–145} and provide a majority of electronic structure related properties. Like DFT, DFTB suffers from several shortcomings, including unreliable descriptions of long-range intermolecular interactions. However, it is now quite clear that this particular shortfall can be overcome using a posteriori dispersion corrections.^{91–93,146,147}

In principle, the DFTB scheme can be applied to investigate organic electronic materials featuring non-covalent interactions between sulfur atoms,^{37,145,148–151} when combined with a dispersion correction.^{91–93,146,147} Unfortunately, as demonstrated here, DFTB suffers from a major qualitative drawback that affects both geometries and energies of non-covalent interactions and ultimately leads to completely erroneous data any time sulfur atoms are present. Here, we illustrate a few alarming examples of this DFTB flaw. For a brief overview of the DFTB scheme see Section 2.1

3.2 Computational Details

All SCC-DFTB²² and DFTB3^{23,24} computations were performed using the MIO/NHorg^{22,152} parameter set in the DFTB+¹⁴² release 1.2.26 DFTB3 computations used the calc parameters from Gaus et al.²⁴ The DFTB geometry optimizations were performed at the DFTB3 level using default settings. DFT computations were performed in MOLPRO at the PBE/def2-TZVP^{39,40,153,154} or B3LYP/def2-TZVP^{153–156} levels using default thresholds and grids.

3.3 Critical Failure of Sulfur at DFTB level

The DFTB description of sulfur-containing compounds relies upon a few critical aspects within the parameterization: the electronic terms in the existing MIO^{22,152} parameters are based on DFT computations using the PBE exchange-correlation functional. The proper description of hypervalent species is achieved by inclusion of d-orbitals in the minimal basis set. The repulsive potential is fitted to various bond lengths for a small set of seven sulfur-containing molecules (see Niehaus et al.¹⁵² for further details). Given the parametrization is based on PBE and B3LYP DFT computations, DFTB results will be compared with values from B3LYP/def2-TZVP and PBE/def2-TZVP, referred herein as “DFT level”. Note that since neither standard DFT approximations nor DFTB accounts for dispersion interactions, the energy profiles of van der Waals complexes are generally expected to be repulsive even though some functional

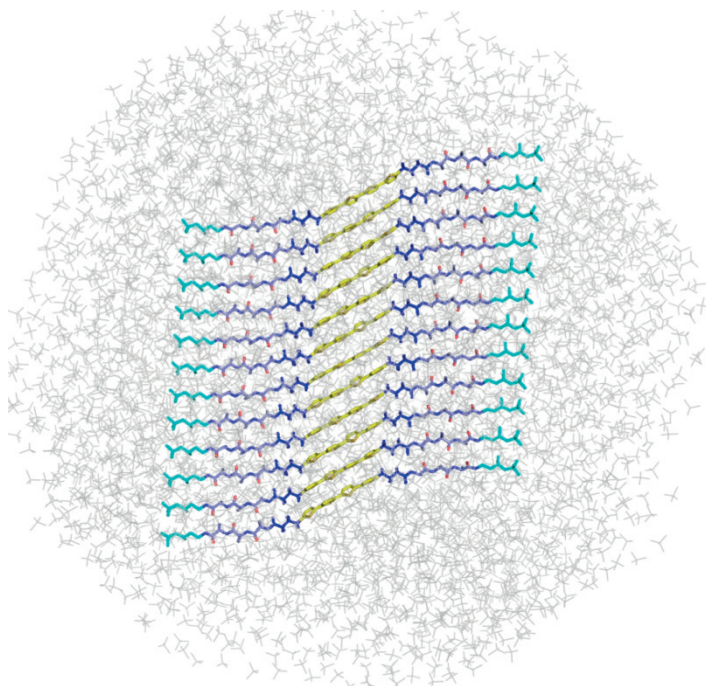


Figure 3.1: A recent example of hierarchically structured microfibers of a “single stack” quaterthiophene (yellow) nanowire.¹⁷ The oligopeptide-substituted quaterthiophenes benefit from a synergistic enhancement of hydrogen-bonding and π - π interactions that are important for electronic applications. The computational description of such assemblies relies upon the proper evaluation of non-covalent interactions.

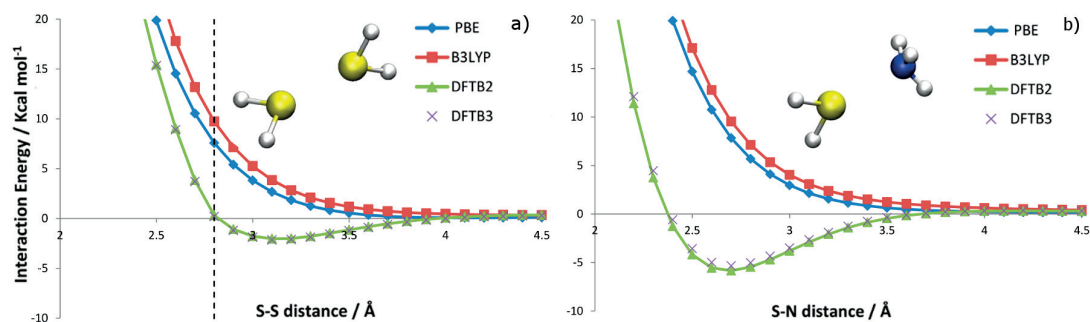


Figure 3.2: Interaction energy profiles of the H_2S dimer (left) and of H_2S with NH_3 (right) at the DFT and DFTB levels. The SCC-DFTB and DFTB3 computations were performed using the MIO parameters. DFT energies are computed at the PBE/6-31G*^{39,40,157–164} and B3LYP/6-31G*^{155–164} levels. The vertical dashed line indicates the cut-off of the spline in the sulfur-sulfur interaction repulsive term. Note that the H_2S dimer does not bind in the absence of a dispersion correction to DFT approximations.

approximations (e.g. PBE^{39,40} TPSS⁵¹) can bind rare-gas dimers and other non-covalently bound diatomic without a dispersion correction.^{165–168}

Figure 3.2a shows DFT and DFTB energy profiles along the S-S intermolecular distance for two hydrosulfuric acid molecules, where the sulfur atoms point toward one another. To our surprise, the SCC-DFTB profile is attractive and thus qualitatively wrong; it does not parallel the repulsive DFT curves. Instead, a binding region with a maximum value of about 2 kcal/mol is found between 2.8–4.0 Å, which potentially leads to serious consequences (vide infra) in, for instance, molecular dynamic simulations. The shortcomings illustrated by Figure 3.2 persists in more complex systems, including the antiparallel thiophene dimer and a T-shape dimer of annelated β -trithiophenes¹⁶⁹ (Figure 3.3). Comparisons of the antiparallel thiophene dimer reference geometries computed by Steinmann et al.,¹⁷⁰ Figure 3.3A, and by the non-dispersion corrected DFTB3 level (Figure 3.3B) show a decrease in the S-S distance from 4.66 Å to 3.27 Å. The latter distance is significantly shorter than the accurate S-S equilibrium distance and the final geometry is spurious. At the DFT levels, the overall intermolecular interaction is repulsive and, as one expects, the two molecules move away from one another if no dispersion correction is utilized. Invoking a dispersion correction in the DFT optimization of the thiophene dimer maintains the antiparallel configuration (that is a local minimum¹⁷⁰), which can be attributed to favorable π -interactions. Figure 3.3C displays the optimized geometry at PBE0-dDsC/def2-SVP level of a T-shape dimer of annelated β -trithiophenes. Geometry optimization at the DFTB3 level (without dispersion correction) causes the S-S distance to decrease from 3.84 Å (Figure 3.3C) to 3.02 Å (Figure 3.3D). At the non-corrected DFT levels, the overall intermolecular interaction is again repulsive. Clearly, the DFTB3 optimization is dominated by the “wrong” S-S attractive potential, which negatively impacts the description of any stacked oligothiophenes or more complex systems, such as those given in Figure 3.3. This shortcoming is specific to sulfur atoms as the non-covalent interaction involving other heavy atoms is repulsive at the DFTB level.

These same drawbacks affect energy profiles of systems that contain non-covalent interactions between sulfur and any other atoms, e.g., H, N, C, O. As an example, Figure 3.2b shows the contrasting DFT and DFTB interaction energy profile along the S-N distance for an ammonia/hydrosulfuric acid dimer where the nitrogen and sulfur atoms point towards one another. The maximum binding energy in the DFTB profile is located at 2.7 Å and corresponds to a stabilization of 5.9 kcal/mol, while the expected profile, given by the two density functionals is repulsive at all distances! Note that DFTB3 exhibits the same behavior as SCC-DFTB and that the problem is limited only to non-covalent interactions.¹⁵² These findings are preliminary indications that “the sulfur problem” arises from the parameterization (only one set of parameters is currently available for sulfur)¹⁵² and not from a defect of the SCC-DFTB theory.

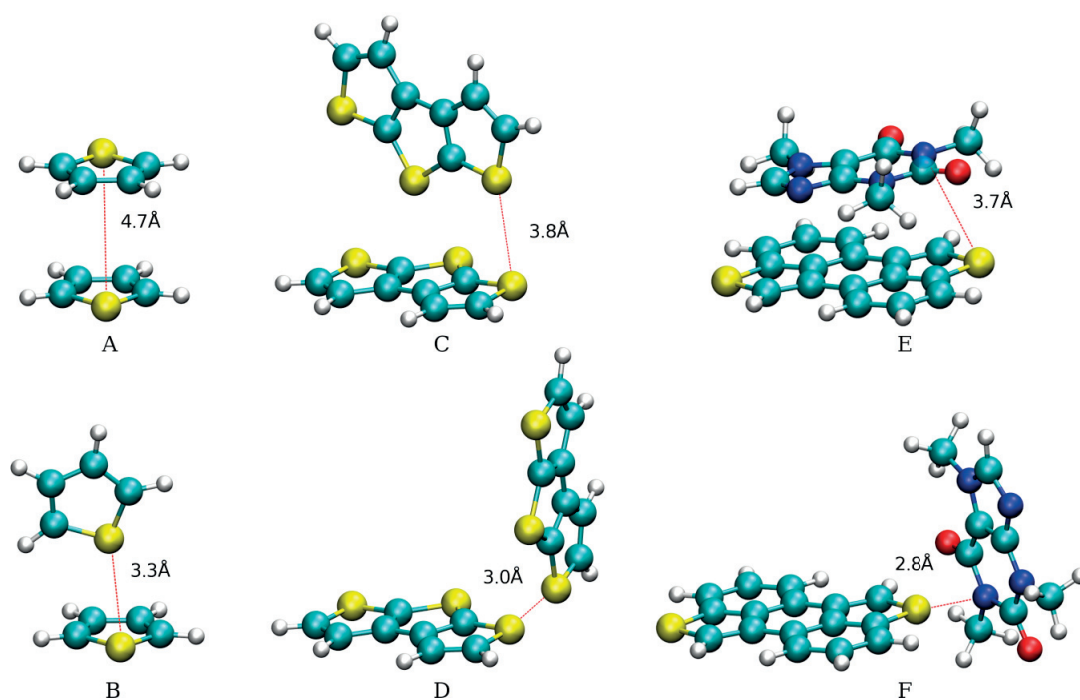


Figure 3.3: The reference geometry of the antiparallel thiophene dimer (A) ¹⁷⁰ and the DFTB3 optimized geometry (B) using the MIO set of parameters. The T-shape annelated β -trithiophene dimer optimized at PBE0-dDsC/Def2-SVP level (C) and at the DFTB3 level (D). A recent example of caffeine sensing ^{171,172} optimized at PBE0-dDsC/Def2-SVP level (E) and at DFTB3 level (F).

A related example includes the improper description of the π - π -stacking interactions between caffeine and a novel chemosensor recently designed by Corminboeuf et al.¹⁷¹ and synthesized by Severin et al.¹⁷² Figure 3.3 shows the resulting DFTB3 geometry optimization of the caffeine-dye complex (3.3F) starting from the PBE0-dDsC/def2-SVP65 geometry (3.3E): the S-N distance in the reference structure is significantly longer (3.67 Å) than that given by DFTB3 (2.83 Å). As mentioned before, the expected behavior is for the two molecules to move apart from each other. Note that the shortest S-N distances in the caffeine-dye complex optimized at DFTB3 level closely match the spurious 2.7 Å equilibrium energy distance of the S-N DFTB profiles (Figure 3.2). Obviously, the spurious S-N overbinding drives the geometry optimization of the caffeine-dye complex towards the incorrect structure.

Providing a quick fix to the problem is not straightforward. In addition, the inclusion or parametrization of a proper a posteriori dispersion correction to DFTB is irrelevant, since this would only increase the observed overbinding. One possible solution considered was modification of the repulsive term $V(R_{\alpha,\beta})$ (see Section 2.1) to increase the repulsion between the sulfur and the other atoms in order to prevent the non-covalent overbinding. Similar modifications of the repulsive terms have been utilized in the past to correct for the N-H binding energy in specific environment, but only small compensations were needed in that case.^{44,173} To correct for the error of the sulfur non-covalent interaction, $V(R_{\alpha,\beta})$ must be radically modified: the cut-off of the repulsive term is never larger than 2.8 Å (see Figure 3.2) in the original parameters, while it must be extended to about 4.0 Å, depending on the atom pair, to overcome the error in the electronic term. This type of modification contrasts with the principle on which the repulsive term is based: $V(R_{\alpha,\beta})$ must be strictly pairwise, short range, and repulsive.^{22,174,175} Short range implies that the cut-off of the repulsive term should be only slightly longer than a covalent bond length, to make it strictly pairwise. Despite this conceptual challenge, we attempted reparameterization by fitting the repulsive term to the difference between the DFT-B3LYP/def2-TZVP total energy and SCC-DFTB electronic energy at various distances for one representative molecular system for each parameter (e.g., two H₂S molecules arranged in the Figure 3.2 conformation were used for the S-S repulsive parameter). Note that we fitted the intermolecular energies instead of the covalent bond energies as done by Niehaus et al.¹⁵² Using this approach, the intramolecular properties were not significantly affected.

An alternative approach to circumvent the problem was using a brute force simplex algorithm that modifies the cubic splines associated with $V(R_{ss})$ and $V(R_{sc})$ to fit an ensemble of DFTB to DFT profiles. Even in this case, the results were not satisfactory, as the final DFTB values remained qualitatively different from DFT.

Our efforts indicate that modifying the repulsive term is insufficient to fix the “sulfur problem” discussed here. Recently, a new set of parameters for hydrogen and second row non-metal

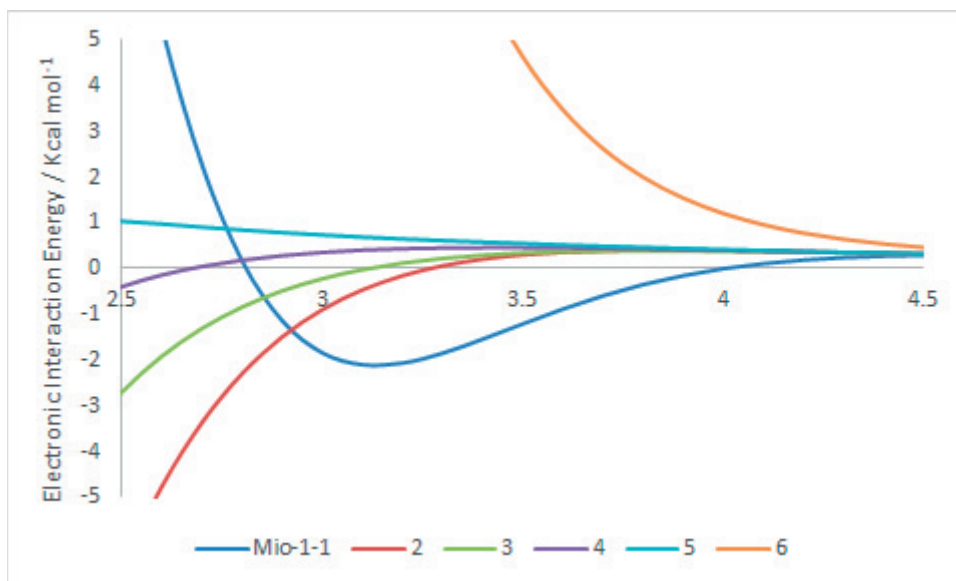


Figure 3.4: Influence of the wave function compression radius of sulfur on the S-S interaction energy profile between 2.5 Å and 4.5 Å (the distances around which the overbinding problem occurs). The profiles correspond to the interaction energy (minus the repulsive contribution between the sulfur atoms, $V(R_{SS})$), of a hydrogen sulfide dimer (see Figure 3.2). The Mio-1-1 curve is obtained with the default Mio-1-1 electronic parameters ($r_{s,p}^{wf} = 2.0\text{Å}$, $r_d^{wf} = 2.3\text{Å}$). The curve “2” uses $r^{wf} = 2\text{Å}$, the curve “3” $r^{wf} = 2\text{Å}$, etc. and were obtained using a homemade program. They demonstrate that the profiles get less attractive when increasing r^{wf} . Note that the difference between “2” and “Mio-1-1” is due to the missing d orbital contribution in the former profile. In 2-6, the same compression radii were use for both s and p sulfur orbitals without the inclusion of d orbitals.

elements was developed,⁴⁴ herein it was found that the wave-function compression radii (r^{wf}) affect non-covalent bond distances. In particular a large r^{wf} causes an increase in Pauli repulsion, and therefore a larger non-covalent bond distance. Our examination of the influence of r^{wf} values on the interaction energy (see Figure 3.4) seems to confirm that a larger r^{wf} could lead to less attractive profiles and that reparameterization of the electronic term is the only way toward fixing the spurious interactions involving sulfur atoms.

3.4 Conclusion

SCC-DFTB and DFTB3 approaches can lead to highly valuable insights and unravel complex problems involving large molecular systems. Further development and validation of such methods could be strongly beneficial for the design of organic based functional nanostructures. Unfortunately, we demonstrate here that the DFTB geometries of sulfur-containing

Chapter 3. A Caveat on SCC-DFTB and Non-covalent Interactions Involving Sulfur Atoms

compounds featuring non-covalent interactions are unreliable, and often result in dramatic qualitative failures preventing the use of all DFTB variants in geometry optimization and more importantly in molecular dynamic simulations. Thus, to successfully apply DFTB approaches to supramolecular systems containing sulfur, a revision of the DFTB parameters is urgent. Attempts to correct this “sulfur problem” by adding a correction within the repulsive energy were unsuccessful. Our analysis indicates that only a complete revision of the electronic parameterization term will render the DFTB framework reliable for simulations of sulfur containing compounds. We have also suggested that larger wave function compression radii could solve the problem. The DFTB community should be cautioned from using the existing sulfur parameters, especially if the sulfur non-covalent interactions are the key factors in determining a complex geometry, and hope that this work will stimulate the development of new parameters.

4 A fast charge-dependent atom-pairwise dispersion correction for DFTB3

This chapter is based on following publication:

Petraglia, R., Steinmann, N. S., & Corminboeuf, C.

A fast charge-Dependent atom-pairwise dispersion correction for DFTB3.

Int. J. Quantum Chem. **2015**, 115, 1265

4.1 Introduction

The development of electronic devices based on π -conjugated polymers and oligomers^{7–9} is driven by the opportunity to achieve novel functionalities (e.g. mechanical flexibility, transparency, impact resistance)^{11–13} at reduced fabrication costs. The performance of such organic devices depends heavily upon the organization of π -conjugated molecules or chains at the molecular level^{9,15} and upon the electronic structure mirrored by the wavefunction. In this context, insightful structure-property relationships can be exploited if quantum chemistry is used concurrently with experiment. The main attractive interactions occurring between π -conjugated moieties arise from van der Waals forces that decay at large intermolecular distances. The central role of computational approaches is hence to achieve an accurate description of London dispersion and establish how to fine-tune the relative displacements or orientations between π -conjugated cores. Despite their omnipresence, van der Waals interactions are not accounted for by standard semi-local and hybrid density functionals^{91–93} that provide a practical balance of accuracy and computational cost unmatched by other methods. Over the last decade, tremendous efforts have been devoted to improving the description of dispersion forces within the DFT framework.^{176,177} The most extensively used approaches consist in adding *a posteriori* an atom pairwise energy correction term (*vide infra*).^{31,69,70,85,101} The various available pairwise schemes differ in the way the dispersion co-

efficients are obtained. For instance, Grimme's popular DFT-D is based only on pre-tabulated values,^{60,68,69} the XDM model from Becke and Johnson computes the dispersion coefficients from the exchange-hole dipole moment,^{104,178} Tkatchenko and Scheffler's vdW-TS connects the dispersion coefficients to the size of the atom in the molecule,⁷⁰ while their latest variant also accounts for the many-body physics.^{179,180} Closer to the present context, Steinmann *et al.* formulated a classical¹⁰⁸ and density-dependent dispersion correction (dDsC),^{31,107,181,182} which simplifies the computation of the XDM and exploits Hirshfeld (overlap) populations¹⁸³ to distinguish non-bonded regions from bonded atom pairs, eliminating the correction at covalent distances.³¹

The DFT framework as used in practice is convenient and efficient albeit restricted to systems made of few hundred atoms only. This limitation prevents the modeling of large-scale organic molecular materials. In comparison, tight binding and other semi-empirical approaches are capable of producing molecular geometries and energetics at dramatically reduced computational costs.¹⁴² In particular, the Self-Consistent Charge Density Functional Tight-Binding (SCC-DFTB)¹⁸⁴ scheme, rooted within the DFTB method developed by Seifert *et al.*^{21,35} as well as its most recent DFTB3 variant,^{22–24} provide valuable insights and have already unraveled complex problems.^{38,148,149,151} In the context of organic electronic materials, the novel parameterization for (bio)organic molecules (so called 3OB) is especially relevant as it restores the proper qualitative behavior^{25,44} for molecules involving non-covalently bound sulfur atoms that were poorly described by the previous MIO11 parameter set.^{30,152,185} Yet, SCC-DFTB suffers from the same deficiency as DFT functionals and does not account for dispersion interactions.^{92,147} In this work, we propose a dispersion correction tailored for SCC-DFTB but inspired from the density-dependent dDsC correction. The proposed model is called dDMC due to its dependence on Mulliken charges³² that are readily available from a SCC-DFTB computation. As such, dDMC does not require any additional information and is computationally very cheap. Alternative dispersion energy corrections adapted to SCC-DFTB/DFTB3 approaches exist. Elstner *et al.* proposed a method suitable for biological system that is based on the Slater-Kirkwood effective number of electrons.⁹² Zhechkov *et al.* used the Universal Force Field London coefficients to correct the SCC-DFTB energy.¹⁴⁷ Řezáč *et al.*, introduced the more sophisticated D3H4¹⁴⁶ method, which corrects for dispersion interaction using Grimme's D3⁶⁹ correction and improve the description of interactions involving hydrogen atoms. More recently, Grimme proposed a new parameterization of D3 specific for the DFTB3/3OB.¹⁸⁶ However, all these schemes have been parameterized and validated on biological systems with no specific consideration of typical π - π stacked molecules characteristics of organic electronics. The challenges associated with the modeling of these systems involve overcoming the interplay arising from the poor description of both the sulfur-containing moieties^{25,30} inherent to the DFTB parameters and the vdW interactions. The pragmatic dispersion correction proposed herein aims at providing efficiently reliable energies, geometries and molecular

dynamic trajectories for sulfur-containing organic complexes. The next section describes the theoretical aspects of the dDMC that is followed by its validation.

4.2 Theory

dDMC is an *a posteriori* pairwise dispersion correction that adjusts the idea behind dDsC^{31,107,108} to the simpler DFTB scheme. The general approach to compute the dispersion energy is

$$E_{disp} = \sum_{i=1}^N \sum_{j>i}^N f_d(R_{ij}) \frac{C_6^{i,j}}{R_{ij}^6} \quad (4.1)$$

The indexes i and j run over all the nuclei, R_{ij}^6 is the internuclear distance, $C_6^{i,j}$ is the dispersion coefficient associated with the interaction between the atom i and the atom j , f_d is a function that damps the correction at short internuclear distances that are better described by the DFTB Hamiltonian. The commonly used density-dependent schemes (e.g., XDM, dDsC) compute atomic dispersion coefficients from partitioning functions such as the Hirshfeld scheme. The same Hirshfeld partitioning is also used in the sophisticated damping function of dDsC. The simplification in dDMC aims at avoiding the computation of: i) integrals inherent to the Hirshfeld partitioning and ii) local density derivatives (i.e., the XDM) that are more demanding than SCC-DFTB itself. The Hirshfeld partitioning defines a weighting function:

$$w_i = \frac{\rho_i^{\text{free}}(\mathbf{r})}{\sum_j \rho_j^{\text{free}}(\mathbf{r})} \quad (4.2)$$

where $\rho_i^{\text{free}}(\mathbf{r})$ represents the electron density associated with the free atom i while $\rho(\mathbf{r})$ is the molecular electron density. The j index runs over all the atoms in the molecule.

One of the central quantities, on which density dependent dispersion corrections are based, is the estimate of the size of the atom in a molecule. In particular, the ratio between the volume of the Atom In Molecule (AIM) and the free atom:

$$\frac{V_i^{\text{aim}}}{V_i^{\text{free}}} = \left(\frac{\int r^3 w_i(\mathbf{r}) \rho(\mathbf{r}) d^3\mathbf{r}}{\int r^3 \rho_i^{\text{free}}(\mathbf{r}) d^3\mathbf{r}} \right) \quad (4.3)$$

can be conveniently approximated by Eq. 4.4.

$$\frac{V_i^{\text{aim}}}{V_i^{\text{free}}} \approx \frac{N_i}{Z_i} \quad (4.4)$$

where N_i and Z_i are the Mulliken electronic population for the atom in the molecule and the number of electrons for the free atom i . This seemingly very crude approximation is motivated

by a model of atoms with a uniform density inside the volume of the atoms. Our approach is correct in two limiting cases: (i) the neutral, “free” atom and (ii) when the atom has no electrons.

In 2009, Tkatchenko *et al.*⁷⁰ linked the dispersion coefficient for an atom in a molecule ($C_6^{\text{aim},j}$) to the dispersion coefficient of the free atoms ($C_6^{\text{free},j}$) through the ratio displayed in Eq. 4.5:

$$C_6^{\text{aim},j} = \left(\frac{V_j^{\text{aim}}}{V_j^{\text{free}}} \right)^2 C_6^{\text{free},j} \quad (4.5)$$

Directly exploiting our assumption we can define a new relation for the dispersion coefficient:

$$C_6^{\text{aim},j} = \left(\frac{N_i}{Z_i} \right)^2 C_6^{\text{free},j} \quad (4.6)$$

We here discuss the results obtained with Eq. 4.6 with the $C_6^{\text{free},j}$ available for most of the elements in the periodic table as provided by Grimme.⁶⁹

We apply the same combination rule as in our previous work¹⁰⁸ for the dispersion coefficients between atoms i and j .

$$C_6^{i,j} = \frac{2C_6^{\text{aim},j} C_6^{\text{aim},i}}{C_6^{\text{aim},j} + C_6^{\text{aim},i}} \quad (4.7)$$

It is important to stress that a dispersion correction based on the Mulliken scheme is ideally suited for SCC-DFTB. In contrast to large basis sets, small or minimal basis sets provide robust Mulliken charges.¹⁸⁷ Since the DFTB Hamiltonian depends on the Mulliken charges using a minimal basis set, these charges are robust and much more convenient than Hirshfeld charges (using minimal basis sets).

The dDsC damping is based on the universal Tang and Toennies function^{96,97} plus a second damping function with both depending on the information extracted from the electron density. dDMC preserves the double damping and the flexibility of dDsC but without density-dependency and without adding extra cost to the electronic structure computation. In line with the “density dependent” approach, the damping function uses an electronic parameter to switch the correction on and off. The Fermi function¹⁸⁸ $F(a, s, R_{ij})$ damps a Tang and Toennies function $TT(b, R_{ij})$ to ensure enough flexibility

$$f_d(b, R_{ij}) = F(a, s, R_{ij}) TT(b_{ij} R_{ij}) \quad (4.8)$$

The Fermi damping function

$$F(a, s, R_{ij}) = \frac{1}{1 + \exp\left(-s\left(\frac{R_{ij}}{aR_{ij}^0} - 1\right)\right)} \quad (4.9)$$

contains an empirical parameter a that scales the van der Waals radii^{189,190} ($R_{ij}^0 = R_i^0 + R_j^0$) and a steepness parameter s that minimizes the effect of the Fermi function at larger internuclear distances. The Tang and Toennies function is

$$TT(b_{ij}R_{ij}) = 1 - e^{-b_{ij}R_{ij}} \sum_{k=0}^6 \frac{(b_{ij}R_{ij})^k}{k!} \quad (4.10)$$

in which the TT-damping factor (b_{ij}) regulates the medium range of the correction. b_{ij} is computed according to the combination rule

$$b_{ij} = 2 \frac{b_i b_j}{b_i + b_j} \quad (4.11)$$

The dDsC b_{ii} factor is defined as

$$b_i = b_0 \sqrt[3]{\alpha_i^{-1}} \quad (4.12)$$

where b_0 is a fitted parameter and α_i is the polarizability of the interacting atoms in the molecule defined as $\alpha_i = \alpha_i^{\text{free}} \frac{V_i^{\text{aim}}}{V_i^{\text{free}}}$. Adopting the same idea to dDMC we get

$$b_i = b_0 \sqrt[3]{\alpha_i^{-1}} = b_0 \sqrt[3]{\frac{V_i^{\text{free}}}{\alpha_i^{\text{free}} V_i^{\text{aim}}}} \approx b_0 \sqrt[3]{\frac{Z_i}{\alpha_i^{\text{free}} N_i}} \quad (4.13)$$

where the free atom polarizabilities (α_i^{free}) are taken from the CRC Handbook.¹⁹¹

This scheme and in particular the usage of a “double-damped” function ensures the right behavior at both the short and medium/long-range.

Note that akin to other dispersion corrections, dDMC cannot solve the issues related to the poor description of H-bonded interactions by semi-empirical approaches.^{192,193} Instead an additional empirical correction would be needed¹⁴⁶ for this purpose. However, our present objective is not to improve the description of H-bonds but rather to provide a fast electronic structure scheme that accurately describes sulfur-containing compounds involved in π -stacking interactions. Nevertheless, the proposed dDMC scheme could be further combined with a “H-bond corrections” to provide a more generally applicable scheme.

4.2.1 Gradient

The validation of the quality of the approximated gradients is essential in the context of both optimizations and molecular dynamic trajectories. As the dispersion correction depends, through the Mulliken charges, on the electronic structure, the gradient has an electronic and a geometric contribution. To improve the computational efficiency, we neglect the electronic contribution, i.e., the Mulliken charges are fixed with respect to the atom displacements. In addition, given that the dispersion correction depends only on the interatomic distances, the gradients are computed directly in function of the distance instead of the coordinate's displacements:

$$F_{i,x} = - \sum_{j \neq i}^N \frac{\partial \left(f_d(R_{ij}) C_6^{i,j} R_{ij}^{-6} \right)}{\partial x} \approx - \sum_{j \neq i}^N C_6^{i,j} \frac{d \left(f_d(R_{ij}) R_{ij}^{-6} \right)}{dR_{ij}} \frac{\partial R_{ij}}{\partial x_i} \quad (4.14)$$

where F_{ix} is the force acting on the atom i along the direction x . The validity of this approximation is verified through a comparison between the dispersion numerical gradients (computed using a displacement of 0.001 Å) and the approximated analytical ones on all the structures available in the S66 dataset.¹⁴⁶ To enforce that gradient values are different from zero, a noise corresponding to a uniformly distributed random displacement comprised between -0.2 and 0.2 Å was added to the coordinates of all atoms. The mean absolute deviation (MAD), computed as

$$\text{MAD} = \sum_i^N \frac{|f_i - r_i|}{N} \quad (4.15)$$

using the numerical derivatives as reference values (r_i) and the approximated analytical gradients as forecast values (f_i), is 6.813E-8 eV/Å on an absolute average dispersion force of 6.443E-3 eV/Å. This implies that the error arising from the approximated analytical gradient is $\sim 1\text{E}-5$ eV/Å times smaller than the average dispersion force arising from a deviation of 0.2 Å from the most stable configuration.

4.2.2 Adjustable parameters and training set

The dDMC correction depends on two adjustable parameters, a and b_0 as well as on the steepness factor, s . In line with our former work,¹⁰⁸ the steepness factor, $s = 46$, was chosen such as to minimize the effect of the Fermi function on the overall damping at large internuclear separations by imposing the constraint $F(a, s, 1.1 \cdot a \cdot R_{i,j}^0) \geq 0.99$. Such a limitation turns the Fermi function off when the distance between the atoms (R_{ij}) is larger then $1.1 \cdot a \cdot R_{i,j}^0$ so that only the $TT(bR_{ij})$ damping is active in this region.

The two parameters, a and b_0 , are trained for each electronic structure approach

	DFTB/3OB	PM6
a	1.857	1.530
b_0	1.018	1.042
s	23.000	23.000
Sc	-	1.842

with the Nelder–Mead optimization method to reproduce a set of interaction energies. The training set includes a subset of the S66x8 data set¹⁴⁶ (namely the “dispersion” and “mixed” structures). To ensure a good performance on sulfur-containing compounds, the training set is completed with an expanded version of the Pi29n (i.e. Pi29nx8) that mimics the S66x8 data set adding seven displaced structures for each dimer included in the original Pi29n.²⁰ For each structure, one scales the equilibrium distance between the monomers center of mass (by 0.9, 0.95, 1.05, 1.10, 1.25, 1.50, 2.00). The interaction energies for each dimer (232 structures) are estimated at the CCSD(T)/CBS level^{194,195} and corrected for basis set superposition error using the counterpoise (CP) correction¹⁹⁶ following the same scheme as used in the Pi29n data set. The inclusion of energy profiles in the training phase serves to improve the response of the damping function at small intermolecular distance. To avoid issues arising from the self-interaction error, the charge transfer complex TTF-TCNQ originally presents in the Pi29n data set was not considered herein. In addition, as the emphasis is placed on improving the description of π -stacked motifs that are prevalent in organic electronic materials, we did not consider the hydrogen-bond complexes of the S66x8 dataset. Note finally that the PM6 Hamiltonian¹⁹⁷ that was used for comparisons required the training of one additional global scaling factor for the total dispersion energy.

4.3 Computational Details

All computations are performed using a modified version of the ASE package¹⁹⁸ that applies the dDMC and D3 dispersion corrections to PM6 and DFTB3. The DFTB3 computations are performed using the DFTB+ 1.2.1 software^{142,199} with the 3OB Slater–Koster files.^{25,41,44} The Hubbard parameters, their derivatives as well as the γ factors are chosen according to Cui *et al.*^{25,44} PM6 computations were performed in MOPAC2012¹⁹⁷ with default options. The D3 parameters are taken from Řezáč *et al.*¹⁴⁶ and used with the software available from Grimme’s website. The dDMC correction terms were computed using a standalone code that is distributed freely. The ASE package delivers the QuasiNewton method to perform the geometry optimizations using homemade interfaces with the previous cited software. Optimizations are considered as converged if the forces on all individual atoms are below $5\text{E} - 3\text{ eV/\AA}$.

The Born–Oppenheimer molecular dynamics²⁰⁰ uses the implementation of the velocity Verlet algorithm²⁰¹ in ASE to integrate the DFTB3/3OB and the corrected trajectories. The PBE-

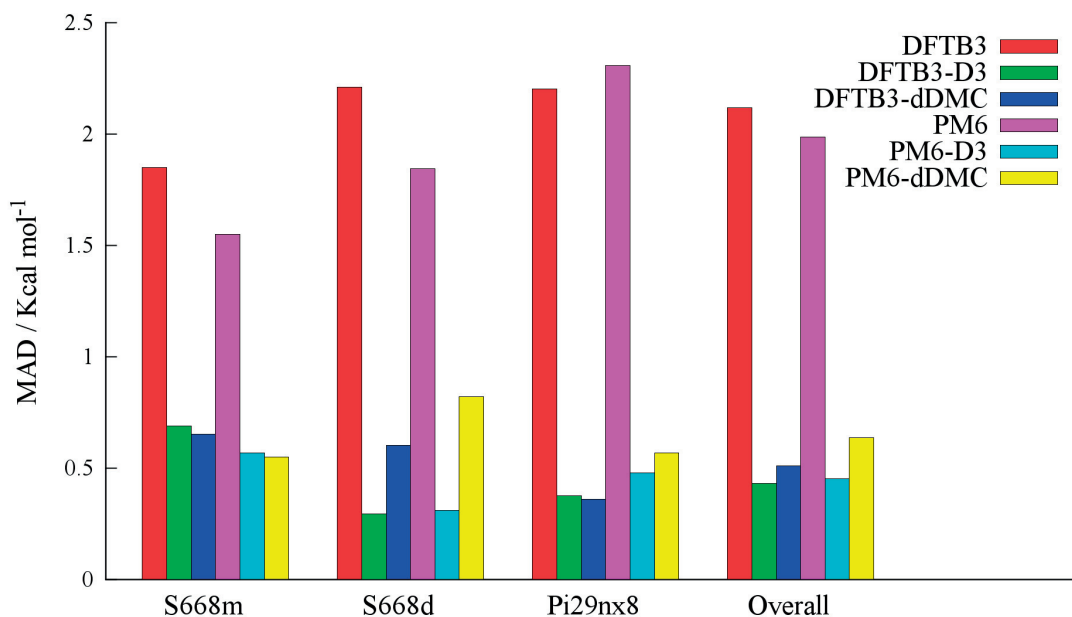


Figure 4.1: Mean absolute deviation of DFTB3/3OB and PM6 and the D3 and dDMC dispersion corrected variants. The “overall” data set displays the MAD of all the dataset together.

dDsC^{31,39,40} simulations are performed with a modified version of the QCHEM4.0 software package.²⁰²

The DFTB3/3OB simulations involving the dithiacyclophane molecule are performed in the microcanonical (NVE)²⁰³ ensemble using a 0.5 fs time step in line with the PBE and PBE-dDsC trajectories obtained previously by Corminboeuf and coworkers.²⁰⁴ The starting structures were the same as in Ref.²⁰⁴ The initial temperature was set to 300 K. With this condition no energy drift has been observed. To ensure that the approximate gradients do not introduce instability in the MD trajectory, we used dithiacyclophane to verify what is the maximum time step that does not introduce an energy drift. The drift was defined as the angular coefficient of the trend line that best fits the energy profile as a function of time. We computed the drift on the NVE simulations starting at 300 K from the same initial structure. Each simulation lasts 10 ps for both DFTB3/3OB and DFTB3/3OB-dDMC. The result shows that the approximate gradient does not influence the energy conservation as both methods show a critical drift for a time step of 2.4 fs.

The simulations on the caffeine–receptor complex^{16,171,172} are conducted in the canonical (NVT) ensemble²⁰³ using the Langevin thermostat²⁰⁵ with a 2 ps⁻¹ friction. We found that a 1 fs time step is small enough to avoid energy drifts. The so-called “reference structure”, optimized at the PBE0-dDsC/def2-SVP^{16,153,154,206} level, was taken as the initial structure for

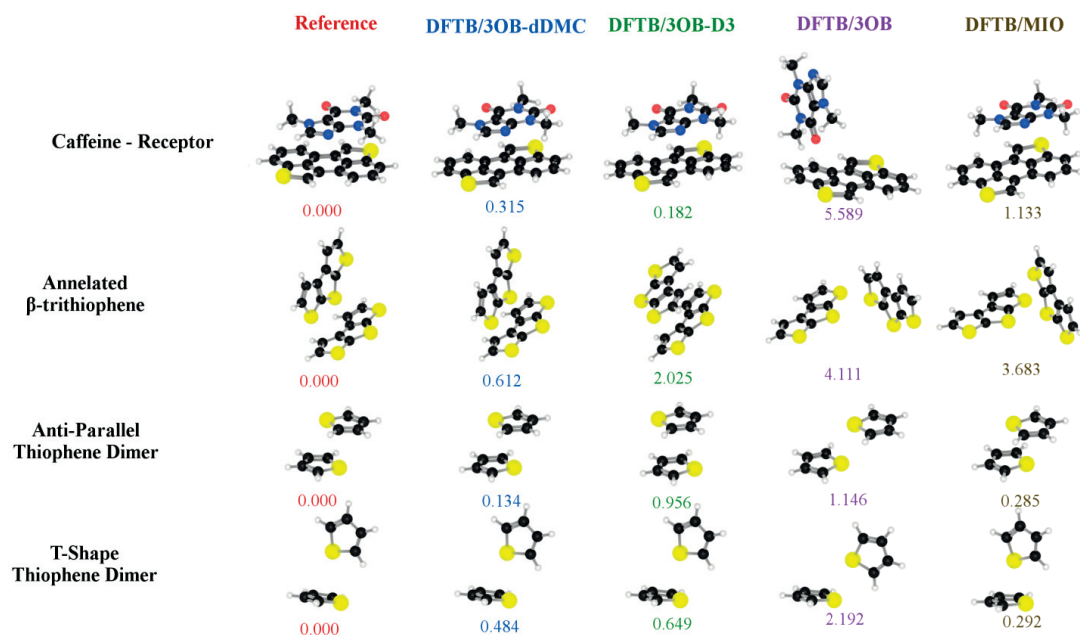


Figure 4.2: Comparisons of the DFTB geometry of sulfur-containing compounds at both dispersion-corrected (dDMC and D3) and non-corrected levels (with the 3OB and MIO parameters). RMSD (\AA) with respect to the reference geometry are reported for each complex. The reference geometries of the caffeine-receptor and the annelated β -trithiophene complexes are computed at the PBE0-dDsC/def2-SVP level. The antiparallel thiophene dimer system is optimized at the RI-MP2/TZ level with counterpoise correction in Turbomole5.1.

the DFTB3/3OB simulations. A snapshot of the DFTB3/3OB trajectory after 5 ps was taken as the starting structure for the PBE-dDsC simulation. All trajectories were thermalized for 5 ps at their respective level.

4.4 Results

The following illustrates the performance of dDMC not only on interaction energies of static dimer structures but especially on practical examples featuring geometry optimizations and molecular dynamics simulations.

Figure 4.1 displays the mean absolute error of the dDMC correction applied to DFTB3/3OB and PM6 compared to the uncorrected variants and the D3 corrected energies.

Overall, the mean absolute errors for the DFTB3 corrected energies are below 0.7 kcal/mol. Despite its simplicity, the performance of dDMC is very similar to D3 except for the S66x8d subset that is notably better described at the DFTB3-D3 level (MAD = 0.29 kcal/mol). The

poorer performance of dDMC for this specific subset essentially arises from an overbinding of the hydrogen-rich dimers such as those made of aliphatic chains (e.g., neopentane, pentane). The present focus is essentially placed on π - π stacking but the combination of dDMC with a H4¹⁴⁶ type correction that contains a specific repulsive term to correct the interaction between hydrogen atoms would surely improve the results for these complexes. The superior performance of DFTB3-dDMC as compared to PM6-dDMC is rooted in the less reliable Mulliken charges associated with the PM6 Hamiltonian.¹⁹⁷ Besides the reasonable performance, a clear benefit of using the dDMC scheme is certainly the gain in computational speed, which is especially visible when computing the gradients on large-scale systems (80% faster than the DFTD3 program version 2.1 rev 3 in calculating gradients on around 1000 atoms).

Chapter 3 demonstrated that the original DFTB3/MIO11¹⁵² parameters lead to spurious energies and geometries for any systems that features a sulfur atom involved in a non-covalent interaction.³⁰ As illustrated by the examples provided in Figure 4.2, the latest 3OB parameterization by Cui *et al.*^{25,44} offers a dramatic improvement over MIO11 for the dispersion corrected gas phase geometries. DFTB/3OB-dDMC leads to four geometries that are in close agreement with the reference PBE0-dDsC/def2-SVP¹⁷² or RI-MP2/TZ^{153,207,208} data. In particular, the T-shape thiophene dimer and the illustrative caffeine-receptor complex^{16,172} are well reproduced with both DFTB/3OB-dDMC and DFTB/3OB-D3 (root mean square deviation (RMSD) < 0.4 Å). The D3 description of the annelated β -trithiophene dimer^{30,169} converges toward another minimum (RMSD = 2.025 Å), whereas dDMC remains in agreement with the reference data (RMSD = 0.6 Å). Similar discrepancies are observed for the antiparallel thiophene dimer. Note that our training set, placing more emphasis on improving the treatment of sulfur interactions, could be at the origin of this difference. Although the spurious overbinding characteristic of the non-dispersion corrected DFTB/MIO11 geometries is recurrent and relatively large in magnitude (sometime even larger than the reference interaction energy), the 3OB parameters offer a significant improvement: the DFTB3/3OB optimized geometries still bind but the interaction energies involved are nevertheless much smaller than the reference and dispersion corrected values. The performance of DFTB3/3OB-dDMC is further validated on the Born-Oppenheimer molecular dynamic simulations of two examples dominated by intramolecular and intermolecular interactions, respectively. Figure 4.3 shows MD trajectories of an illustrative dithiacyclophane incorporating a thieno-[2,3-b]-thiophen that was originally chosen to evaluate the importance of self-consistency in dDsC.²⁰⁴ This compound is rather challenging due to its large flexibility inherent to the existence of several low energy conformers featuring both π - π stacked and open conformations. We here present the molecular dynamic trajectories starting from two closed (i.e. π - π stacked) conformers (**A** and **B**) and monitor the distance between the barycenter of the benzene ring and the middle of the C-C bond of the thienothiophene ring over 8 ps trajectories that are directly compared to our former PBE-dDsC simulations.²⁰⁴ On average, the three tested dispersion-corrected schemes lead to very similar distances with no systematic over/underestimation: the average distance at the

DFTB/3OB-dDMC level is the longest for the first trajectory (i.e. starting from conformer **A**) but the shortest for the second trajectory. The simulation of the closed conformer **B** performed at both non-corrected levels, PBE and DFTB/3OB, readily open. In contrast, the opening of conformer **A** differs significantly between PBE and DFTB/3OB. The PBE opening process is relatively sudden, whereas the DFTB/3OB structure opens more gradually. Other deviations

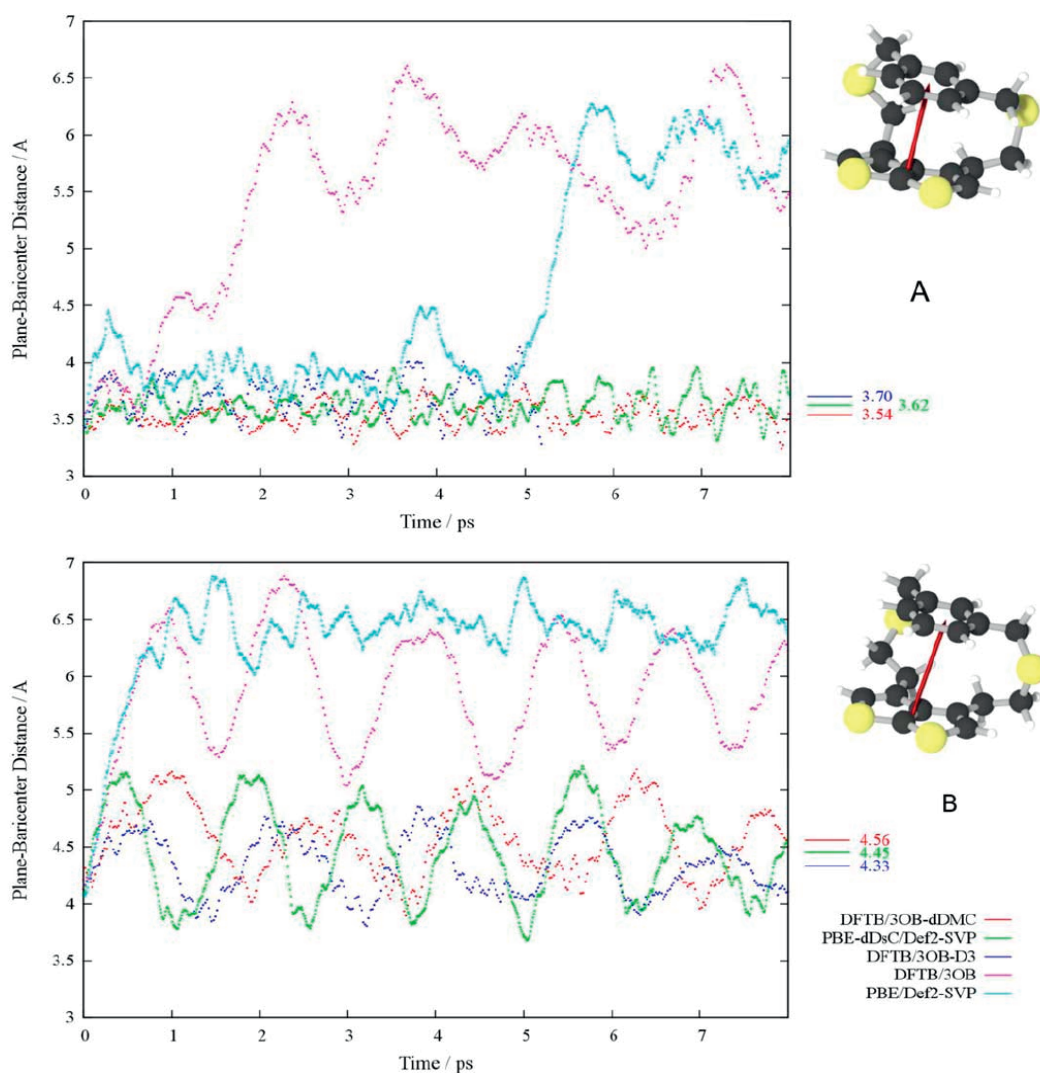


Figure 4.3: Profiles of the distance (see red arrow) between the barycenter of the benzene ring and the middle of the C-C bond of the thienothiophene ring over a NVE Born-Oppenheimer molecular dynamic trajectory. Computations are performed at room temperature with non-corrected and dispersion-corrected PBE and DFTB/3OB. The starting structures are optimized at ω B97X-D/6-31G* level. The methods used to perform the simulations are distinguished by color.

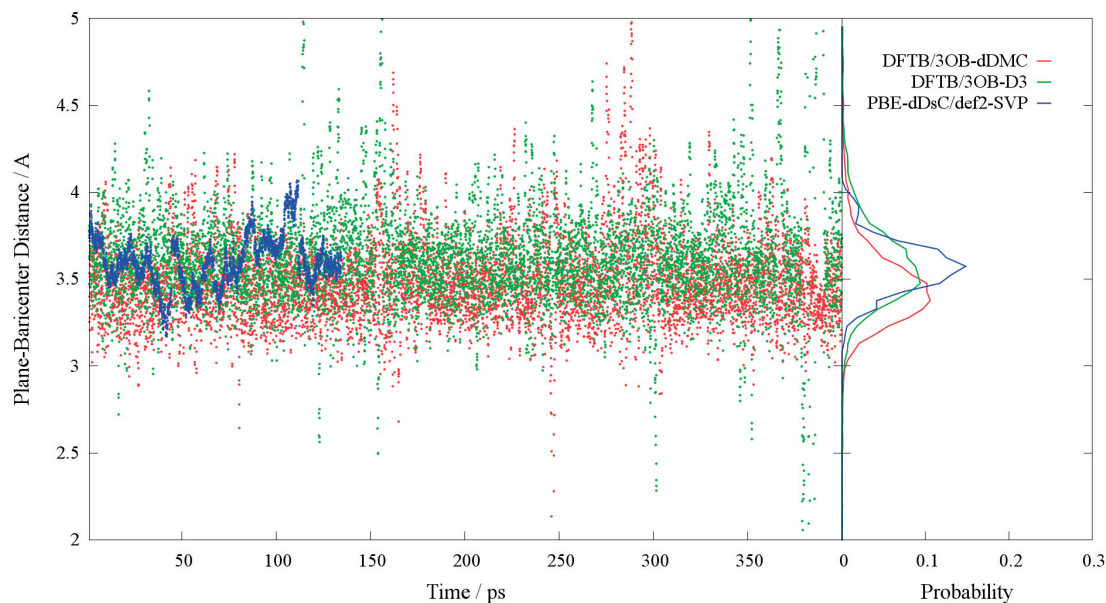


Figure 4.4: Profiles and histograms of the distance between the average plane of the receptor and the the barycenter of the caffeine over a NVT Born-Oppenheimer molecular dynamics trajectory. Computations are performed at room temperature with DFTB3/3OB-dDMC, DFTB3/3OB-D3, and PBE-dDsC/def2-SVP.

observed between the non-corrected approaches include the larger flexibility of DFTB as compared to PBE.

The last molecular dynamics example (Figure 4.4) inspects the longer range intermolecular interaction of an illustrative caffeine–receptor dimer already studied in Refs.^{171,172} that is also included in our comparisons of geometry optimizations (Figure 4.2). The distance monitored is taken between a plane that best incorporates the atoms of the receptor and the barycenter of the caffeine molecule. The 350 ps DFTB3/3OB-dDMC trajectory is compared to that of DFTB3/3OB-D3 and to a shorter 134.5 ps PBE-dDsC/def2-SVP trajectory. The histogram and overall trajectories show a nice correlation between the two DFTB/3OB corrected approaches. The overall observation is that accounting for dispersion is mandatory when performing *ab-initio* molecular dynamics trajectories. Our simulations also demonstrate that the residual error related to the 3OB sulfur parameters is counter balance if combining DFTB/3OB with a dispersion correction. Within the framework of DFTB, dDMC represents a very simple and efficient alternative to the existing schemes specifically adapted to biomolecules for addressing problematic relevant to the field of organic electronics.

4.5 Conclusions

This work introduces a fast atom-pairwise dispersion correction based on Mulliken charges that is specifically tailored for DFTB3. Unlike previous dispersion corrected DFTB computations focusing on biological systems, we here place a special emphasis on improving the description of compounds prevalent in the field of organic electronics. In this respect, the dDMC scheme performs especially well for the energies (MAD = 0.7 kcal/mol for the test set of 94 compounds with a total of 752 different systems), geometries and molecular dynamics of sulfur-containing moieties involved in π - π stacking interactions that are known to be especially challenging for DFTB. We have thus provided both, a valuable extension to DFTB3 by providing a charge-dependent dispersion correction and a careful validation of the provided scheme on test sets for typical weak interactions (S66) and motives typical for organic electronics (Pi29n).

The rising interest in organic electronic materials along with the simplicity of the proposed correction suggests that this approach has great potential. Future developments should enable the treatment of explicit solvent and the consideration of many body contributions that potentially play a role in determining the geometries and thermodynamics of nanoscale assemblies of organic molecules.

5 Beyond static structures: Putting forth REMD as a tool to solve problems in organic chemistry

This chapter is based on following publication:

Petraglia, R., Nicolai, A., Wodrich, M. D., Ceriotti, M., & Corminboeuf, C.

Beyond static structures: Putting forth REMD as a tool to solve problems in computational organic chemistry.

J.Comput. Chem. **2015**, 37, 83

5.1 Introduction

Studies of organic systems frequently utilize computational results as an essential tool to elucidate mechanistic reaction details that are difficult or impossible to access experimentally.^{109,110} The literature is rife with examples of reaction pathways in which reactants, intermediates, and products, as well as their associated transition states, are cartooned as a series of static geometries each possessing a specific reaction enthalpy. Often, a picture of this type successfully captures the key aspects of a system allowing, for example, identification of the primary mechanistic pathway^{209–214} or rationalization of the presence of a specific intermediate.^{214–218} Such descriptions, however, are occasionally insufficient to chemistry occurring in an experimental setting,^{219–236} in the most extreme cases leading to disastrous failures.²³⁷ Inside the computer, a host of factors that govern “real world” chemical reactions must necessarily be approximated or ignored altogether in a static picture. Chiefly among these is a precise description of a reaction’s free energy, as opposed the frequently reported enthalpy. While estimates of the free energy within a harmonic approximation are provided by most quantum chemistry codes, techniques based on molecular dynamics (MD) represent a more appealing option to fully access the entropic contribution to the stability of different

molecular states, that can also be extended to explicitly include environmental effects (e.g., temperature, pressure, etc.).

From the perspective of the computational organic community, an ideal methodology would directly provide *ab-initio* free energies at accessible computational costs, for example, by combining the physical quantities associated with the free energy obtained from MD simulations with the accuracy of static quantum chemical computations. Roughly speaking, the current state of MD simulations to estimate these quantities can be subdivided into two classes: classical or molecular mechanics (MM) and *ab-initio* (AIMD) methods. On one hand, conventional MM force fields, such as those successfully used in molecular biology and pharmaceutical chemistry,^{238,239} rely only on the nuclear coordinates of a system, making them very fast. Unfortunately, in most cases they are incapable of describing chemical processes involving the breaking and formation of bonds (Reactive force fields capable of simulating chemical reactions do exist, but are relatively few in number).²⁴⁰ This makes MD simulations based on MM force fields very useful to examine biological phenomena such as protein folding.²⁴¹ On the other hand, AIMD is capable of describing chemical reactions involving bond breaking and formation. Indeed, the combination of density functional theory (DFT) with AIMD methods has already been used within the framework of Car-Parrinello²⁴² and Born–Oppenheimer^{243,244} MD to resolve problems associated with reaction pathways,^{245–247} phase transitions,^{248,249} and solute/solvent interactions.^{250,251}

The above mentioned examples utilize GGA functionals and plane-waves, which benefit from being both relatively fast^{242,252} and highly scalable,^{253–256} but are, nonetheless, not ideal for the organic community as chemically intuitive concepts and properties are lost owing to the delocalized nature of the plane waves and the inevitable use of effective core potentials. In contrast, Born–Oppenheimer MD simulations, using typical localized basis set, quantum chemistry codes featuring the more reliable global hybrid functionals and post-Hartree Fock methods, are rather sparse (see e.g., Refs.^{257–259}) (For a seminal implementation on GPU see Ref.²⁵⁹). A serious limitation involves the duration for which a simulation can be propagated in time. Naturally, the increased computational expense of AIMD simulations, which require computing potentials from a first principle method such as DFT, limits their applicability to short $\sim 10^2 - 10^3$ ps intervals. Depending on the complexity of the potential energy surfaces (PES), chemically relevant interconversions between different possible states may not appear on such short time scales. One feasible solution is to lengthen the simulation time by reducing the computational expense of determining the first principles potential by replacement with a potential derived from semi-empirical methods, such as density functional tight binding^{22,260} (DFTB). For instance, replacing the DFT by a semi-empirical DFTB potential permits access to simulations that are three orders of magnitude longer (on the 1–10 ns time scale) while also accommodating a tenfold increase in the number of atoms.²⁶¹ Despite these improvements, the complexity of the PES of many organic reactions remains sufficiently large

that visiting all chemically relevant regions remains essentially impossible. To ensure that all of the PES is explored, additionally computational tricks must be used.

MD simulations of systems containing a large, complex PES necessitate the use of enhanced sampling techniques that facilitate thorough exploration of the free energy landscape. These techniques are required to overcome problems associated with running insufficiently short simulations by reducing the amount of time that a system spends trapped in a local energy minimum. Such enhanced sampling techniques can be roughly divided into two groups. The first is concerned with the identification of pathways between known initial and final states. This category includes, for instance, transition path sampling²⁶² or constrained dynamics.²⁶³ However, these approaches do not facilitate searching for the free energy global minimum or other important configurations of a system within a complex free energy landscape.^{264–266} The second category of enhanced sampling methods is better suited to tackle this problem, as it is aimed at obtaining a thorough exploration of the low-energy portions of the free energy landscape. Some of these techniques, such as metadynamics²⁶⁴ or accelerated MD²⁶⁷ (aMD), rely on modified potentials. While most of these methodologies were originally envisioned for the study of biological systems using classical MD,^{119,239,268} both metadynamics²⁶⁹ and aMD²⁷⁰ have been used to study chemical reactions requiring quantum mechanical treatments. The principal disadvantage, in this case, is that some insight about the system's reactivity is needed to choose an appropriate reaction coordinate, which could in turn affect the outcome of the simulation.

In contrast, modified sampling approaches, such as replica exchange MD¹¹³ (REMD), do not require any prior insight. However, they have mostly been used in the context of biological simulations together with empirical force fields, and have only rarely been used in concert with quantum chemical methods. The few existing studies are limited to molecular clusters containing only a handful of atoms.^{271–273} The basic idea of parallel tempering replica exchange, as discussed in more details in Chapter 2, is to simulate N replicas of a system at a range of different temperatures. Replicas propagated at high temperature freely explore a large amount of the PES in an unencumbered manner by overcoming any barriers present, while low temperature replicas explore local minima regions from which they are unable to escape. The key to the improved sampling seen in REMD involves exchanging complete configurations from replicas at different temperatures via a Metropolis–Teller algorithm,²⁷³ thereby enhancing exploration of the entire free energy landscape.

In principle, the coupling of potentials derived from semi-empirical methods with REMD would allow access to additional information and larger systems than for the first principles potentials obtained in AIMD. Such a tool would be very useful within the realm of organic chemistry, where the PES landscapes of reactions may be quite complex. Moreover, the importance of directly determining free energies can assist in exploring the chemistry of

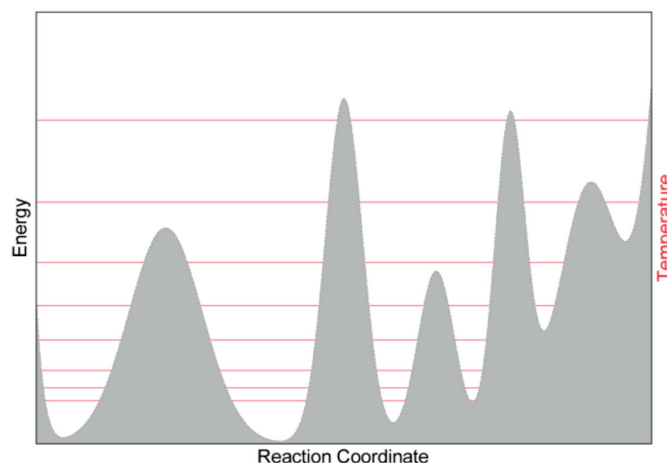


Figure 5.1: Illustration of replica exchange: the potential energy surface of a hypothetical system is represented in gray. The red lines illustrate the replica (i.e., trajectories) at different temperatures. Frequent exchanges are attempted between the replicas based on a Metropolis criterion (with a probability given by Eq. 5.1), which guarantees that at any time the statistical distribution of each replica is consistent with its current temperature. Thanks to the exchanges, each replica can go up and down across the red lines and visit all the minima and maxima on the PES.

systems where entropy is known to play an important role. Here, we present results obtained by coupling the i-PI interface for advanced molecular simulations²⁷⁴ with the DFTB3^{21–23,260} semi-empirical framework to conduct REMD with the objective of exploring the free energy surfaces of organic systems. Such an approach is appealing as it combines the thorough statistical sampling enabled by enhanced MD with the reliability of approximate quantum chemical techniques, which are capable of accurately describing the energetics of structures of organic systems at an affordable cost. Through coupling with i-PI, we highlight the abilities of the REMD@DFTB3 method to address prototypical cases relevant to the computational organic community including: (i) exploration of the conformational space of a dithiacyclophane molecule possessing multiple local minima, (ii) estimation of the minimum energy pathway (MEP) and free energy barrier of the Cope rearrangement (CR) of semibullvalene (SBV), (iii) distinguishing entropically versus enthalpically favored conformational states of a molecular rotor, and (iv) identifying the key conformations of a widely used organocatalyst, cinchona alkaloid.

5.2 Computational Details

All forces are computed at the DFTB3/3OB^{25,44} using the Universal Force Field parameters to account for dispersion forces as implemented in the DFTB+ program.¹⁹⁹ The DFTB+ code was

interfaced with the dynamic driver i-PI after minor modifications. The Hubbard derivatives and the h-damping factor were chosen according to Refs.^{25,44} A finite electron temperature of 300 K was selected to improve the convergence of the geometries arising from the hottest replicas. While a serious drawback for systems sensitive to solvent effects, the current implementation is restricted to gas phase simulations. Regardless, many valuable aspects of a system's behavior can still be extracted from gas phase data. In turn, this newly revealed information may lead to more informed predictions about how the same system would behave in the condensed phase.

The i-PI software drives the REMD, which evolves in the NVT ensemble within a cubic box length of 1000 Å. The large box avoids the spurious interaction between replicas within the periodic boundary conditions. For a given system, the simulations were initiated from the same structure and velocities. Gaussian distributions centered at 900 K provide the atomic velocities for each replica. A time step of 0.25 fs was found to be sufficient to integrate the Newton equation without observing any drift on the conserved quantity. Exchanges between replicas (Figure 5.1) were attempted stochastically every 50 steps²⁷⁵ on average using the following probability:

$$P(acc) = \min \left(1, e^{(E_i - E_j) \left(\frac{1}{k_B T_i} - \frac{1}{k_B T_j} \right)} \right) \quad (5.1)$$

where k_B is the Boltzmann constant and T_i and T_j are the temperature of the exchanged replica. In this way, the mapping of the PES is canonical, meaning that the free energy profile obtained at each temperature corresponds to the free energy profile that would be obtained in a normal MD with much improved sampling efficiency.

A Langevin thermostat, with a time constant of 100 fs, was used to maintain a constant temperature for each replica. The temperature ranges from 300 K for the coldest replica to 1500 K for the hottest. Sixteen replicas were found to be sufficient (i.e. provided enough round-trips among the replicas during the simulation time) for the dithiacyclophane molecule and the CR of SBV. Larger molecules (the cinchona alkaloid and the molecular rotor) needed 48 replicas to effectively exchange among replicas. Snapshots of atomic configurations were saved every 50 steps, and used for further processing.

The initial 10,000 steps were used for equilibration, and discarded from subsequent analysis. Data from different replicas were combined by weighting each frame by $w = e^{\beta_r - \beta_t}$ and computing the observables separately for each parallel tempering temperature. The different replica were then combined with the weighing factors chosen $\frac{\langle w \rangle^2}{\langle w \rangle^2 - \langle w \rangle^2}$ according to the error estimates in reweighed averages given in Ref.²⁷⁶

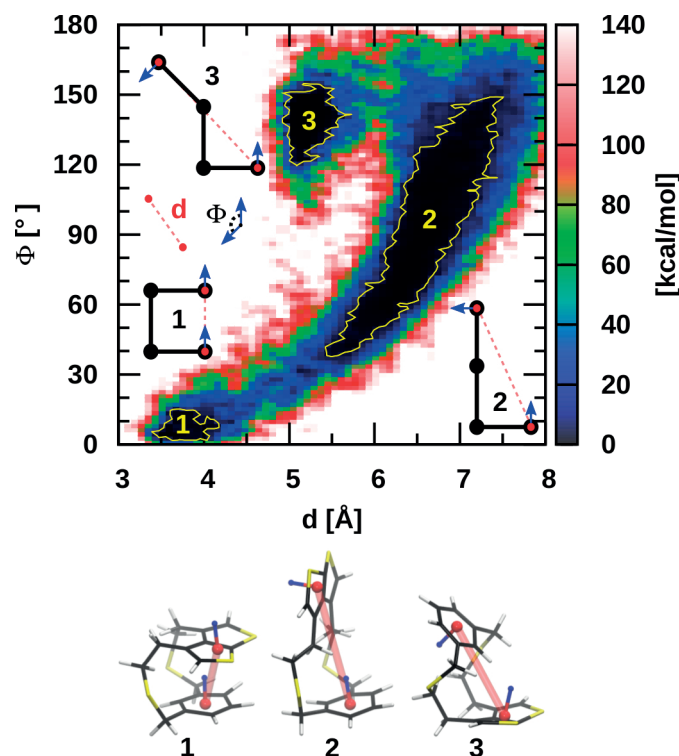


Figure 5.2: Two-dimensional representation of the free energy landscape obtained from the REMD@DFTB3 of the dithiacyclophane molecule. The relevant collective variables are shown in the plot. 1 kcal/mol isocontours are shown in yellow.

Smooth histograms were constructed using kernel binning with a triangular windows function much smaller than the extent of the main features in the free energy landscape. With respect to timing, for each system considered the REMD@DFTB3 computations presented herein could be performed within 2-4 days (real time) depending on the size of the system on Intel-based (Xeon E5-2660) cores. A patch for DFTB+ (version 1.2) is available on demand. The REMD free energies reported in Table 5.2 (cinchona alkaloid) and in Table 5.1 (dithiacyclophane) for each relevant region (i.e. a basin i) are computed through evaluation of the following integral:

$$F_i = k_B T \log \int_{N_i} e^{-F(x)/k_B T} dx \quad (5.2)$$

where each of the non overlapping integration regions (N_i) covers a neighborhood of the local free energy minimum. This equation permits one to account for the thermal fluctuations that distort in a different manner the geometries of the various configurations.

Static electronic structure computations for the dithiacyclophane and CR included optimizations at the M06-2X^{66,277}/def2-SVP level using the “Ultrafine” grid as implemented in Gaus-

sian09.²⁷⁸ Alternative energy assessments were obtained at the PBE0-dDSC/TZ2P^{31,39,107,206,279} (for dithiacyclophane) and CCSD(T)/cc-pVTZ (for the CR) using ADF²⁸⁰ and Molpro,^{281,282} respectively. Reported static free energies include unscaled free energy corrections from M06-2X/def2-SVP computations. Reported static DFTB3/3OB-UFF electronic energies are computed at the DFTB3 optimized geometry.

5.3 Results and Discussions

5.3.1 Dithiacyclophane

Understanding the conformational analysis of molecules represents a cornerstone of organic chemistry. Free energy mapping provides direct relationships between structure and energy and assists in understanding molecular behavior. In this context, a dynamic exploration of the free energy surface may not only alter, but possibly even reverse pictures provided by static relative energy computations. More importantly, it may also reveal unexpected energetically low-lying conformations that were not envisioned owing to preconceived user-based biases about a system. The conformations of dithiacyclophane (Figure 5.2), previously investigated by two of the authors,^{98,204} perfectly illustrates this aspect. This highly flexible molecule was originally found to possess several low-lying conformers featuring both π -stacked (structure 1, meta-stable) and open conformations (structure 2, lower in energy) using accurate electronic structure methods. Not surprisingly, Born–Oppenheimer MD simulations performed at both the DFT (i.e. PBE) and DFTB3 levels were shown to be highly sensitive to the inclusion of a dispersion correction.^{98,204} As might be expected, in the absence of van der Waals corrections, the π -stacked conformation readily converted into the open conformers (on a scale as fast as 250 fs), whereas the stacked meta-stable conformer persisted for a few picoseconds in the dispersion-corrected trajectories. The current REMD@DFTB3 results add a new element to the former picture, revealing a somewhat less intuitive “disarticulated” conformational state (structure 3, Figure 5.2). More importantly, this new conformational region is thermodynamically favored from the MD free energy and static electronic or free energy perspective, and thus would affect any measured properties (e.g. NMR chemical shifts, infrared spectrum, etc.). Note, however, that the conformational entropic contributions are the largest for 2 as indicated by its larger basin (larger number of conformations) and by the small REMD free energy difference between 2 and 3 in comparison to the static picture (see Table 5.1

The population of 3 remains barely dominant at 300 K, when accounting for the full entropic contribution. The two-dimensional (2D) plot of Figure 5.2 provides further insights into the conformational dynamics inaccessible from a static picture. For instance, the direct paths connecting the open conformational regions with the two other areas ($2 \rightarrow 1$, $2 \rightarrow 3$) contrast

	DFTB3/3OB-UFF		M06-2X		PBE0-dDsC	
	Electronic Energy	REMD Free Energy	Electronic Energy	Static Free Energy	Electronic Energy	Static Free Energy
1 Closed	1.96	0.68	2.19	2.77	3.75	4.34
2 Open	1.61	0.19	1.77	1.00	2.38	1.61
3 Disarticulated	0.00	0.00	0.00	0.00	0.00	0.00

Table 5.1: Relative electronic or free energies (in kcal/mol) of dithiacyclophane conformers determined from static computations or REMD simulations. Static values are taken from the DFTB3/3OB-UFF//M06-2X/def2-TZVP, M06-2X/def2-TZVP and PBE0-dDsC/def2-TZVP//M06-2X/def2-TZVP levels respectively. Free energy contributions are given at 300 K.

with the absence of a low-energy pathway directly connecting the closed and “disarticulated” conformer (1 \rightarrow 3). Having access to information of this type could be useful when, for example, aiming to alter dynamic fluctuation through chemical modification. Finally, it is worth noting that “standard” Born–Oppenheimer MD simulations (as opposed to REMD) performed on an even longer timescale (i.e. 1.3 ns) than those in Refs.^{241,242} remained, for most of the time, trapped in its original stacked conformational region without identifying the thermodynamically more important region associated with the new conformation.^{98,204}

5.3.2 Cope rearrangement of semibullvalene

Aside from the identification of chemically important conformers presented in the previous example, REMD can also be used to explore the PES of organic reactions. This second type of case has been explored less often, as most REMD simulations to date have been performed together with a fixed-bonds, nonreactive force field. The CR of SBV represents a prototypical example for which a great deal of mechanistic detail has already been amassed, including information regarding: the (a)synchronous nature and the role of tunneling in the reaction mechanism,^{283,284} the presence and magnitude of homoaromaticity in the transition state,^{285–296} and the link between molecular properties and structure.²⁹⁷ In its simplest form, the CR transitions between two equivalent structures each characterized by three-membered ring on one side of the bridging ethyl sub-unit (Figure 5.3). Utilizing a static picture derived from electronic structure theory [CCSD(T)/cc-pVTZ//M06-2X/def2-TZVP], the transition state connecting these structures possesses equivalent C2–C8 and C4–C6 bond distances (2.00 Å) associated with a concerted reaction mechanism with an overall barrier height of 7.3 kcal/mol.

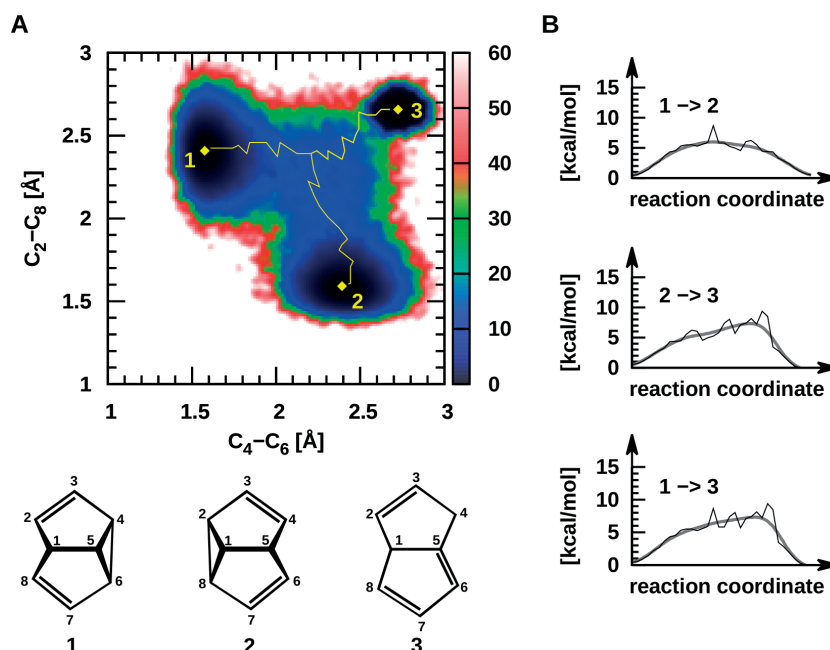


Figure 5.3: Cope rearrangement for semibullvalene. **A:** Free energy map obtained from REMD@DFTB3 simulations indicating the expected Cope rearrangement (minimum energy pathway, $1 \leftrightarrow 2$, given in yellow) as well as an unexpected region corresponding to opening of the semibullvalenes structure to give dihydropentalene ($1 \leftrightarrow 3$, $2 \leftrightarrow 3$). **B:** Minimum energy pathways between the stable conformations presented in Figure 5.3A.

From the organic chemist's perspective, this reaction likely appears far too simple to merit study using MD techniques, predominantly because the essential mechanistic components appear well described from static computations. In contrast, our exploration of the PES using REMD@DFTB3 reveals considerably more information than might have been anticipated (Figure 5.3). More than 600,000 structures were analyzed to generate a 2D free energy plot, that identifies chemically meaningful regions including the expected SBV minimum (1 and 2) and the CR TS (blue central area), as well as the minimum energy reaction pathways connecting these structural regions (yellow). From this data, the relative height of the TS barrier (18.7 kcal/mol is the highest point along the DFTB3 MEP) and any free energy differences between the reactant and products can be estimated (zero in this case) by the upper most reaction coordinate plot of Figure 5.3b, which closely mirror the type of energy plots typically associated with static computations. A quantitative assessment of the reaction rate, however, would require a careful analysis of deviations from transition-state theory that are to be expected whenever the order parameters chosen to describe the transition are not ideal. Perhaps, the most relevant feature of Figure 5.3a is the presence of three minimum, as opposed to the two that we expected. As mentioned earlier, regions 1 and 2 correspond to the two symmetric SBV structures, however, the third black region (3, upper right) represents

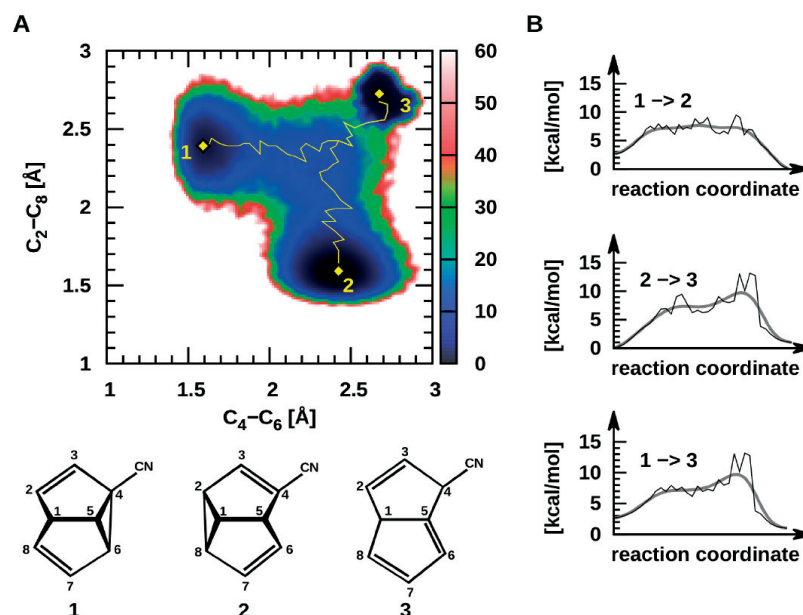


Figure 5.4: Cope rearrangement for CN substituted (C4) semibullvalene. **A:** Free energy map obtained from REMD@DFTB3 simulations indicating the expected Cope rearrangement (minimum energy pathway, $1 \leftrightarrow 2$, given in yellow) as well as an unexpected region corresponding to opening of the semibullvalenes structure to give dihydropentalene ($1 \leftrightarrow 3$, $2 \leftrightarrow 3$). **B:** Minimum energy pathways between the stable conformations presented in Figure 5.4A.

conformations that possess significant lengthening of both the C2-C8 and C4-C6 interatomic distances. Molecular structures of this type are no longer three-dimensional cages, instead adopting quasi-planar conformations, as exemplified by dihydropentalene (3, Figure 5.3). Indeed, such structures are linked to SBV by two distinct, yet directly connected TS structures and a valley-ridge inflection point indicative of a surface bifurcation.^{222,298} Examining the $3 \leftrightarrow 2$ MEPS reveals this feature: the structure must first pass over a relatively high barrier close to region 3 (Figure 5.3a, also visible $1 \rightarrow 3$ and $2 \rightarrow 3$ reaction coordinate plots, Figure 5.3b). Overcoming this first, energetically more costly TS barrier leads to the second TS associated with the CR transition between the two symmetric SBV structures. As we observe just a few transitions toward region 3, we cannot deem simulations to be converged and we cannot be confident of the quantitative accuracy of the free energy landscape we computed. More sophisticated REMD implementations or combinations of REMD and metadynamics²⁹⁹ are probably needed to obtain reversible sampling. Similar in nature to the bare SBV picture, placing an electron withdrawing substituent onto selected positions either enhances or suppresses the CR TS barrier.^{285,286,300,301} For instance, the placement of a CN- onto the C8 carbon results in a decrease in the CR transition state barrier as well as creating asymmetrical products and reactants, yet the substituted dihydropentalene is again present (see Figure 5.4).

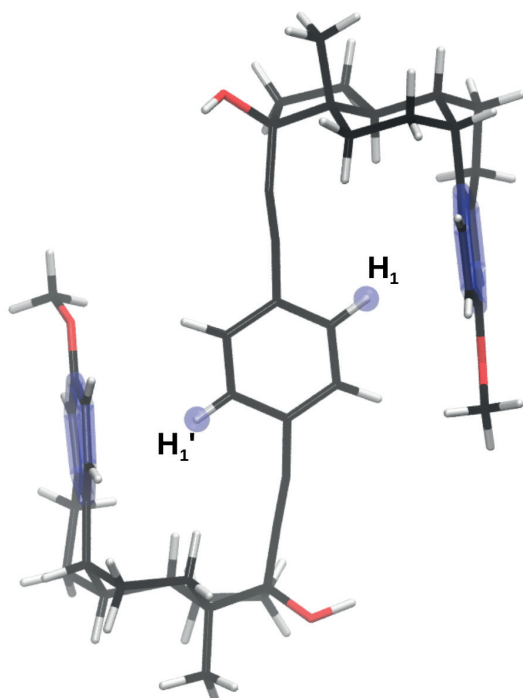


Figure 5.5: Schematic representation of the molecular rotor and the key hydrogen atoms.

Clearly, the inherent nature of sampling the PES using REMD provides an enhanced view of the ways in which organic systems move between different conformers or along competing energetic pathways. For the example provided here, the three energetic basins corresponding to structural minima, as well as the minimum energy pathways connecting them, were obtained with no prior knowledge of the system. Assessments with no *a priori* knowledge that are provided by enhanced sampling techniques are likely to become of increasing importance, particularly for systems featuring a more complex PES. Taken to its extreme, REMD can be used to better understand the behavior of systems consisting of thousands of isomers, such as shape-shifting organic molecules (e.g., bullvalene).²¹⁷

5.3.3 Molecular Rotor

The dynamics of a recently investigated molecular rotor¹¹¹ (Figure 5.5), is a further illustrative case study highlighting the importance of accessing conformational states associated with both entropically and enthalpically favored regions. For this rotor, structural changes occur upon increases in temperature, which are directly visible in the variable temperature ¹H NMR spectra. The measured chemical shifts imply that enthalpy favored conformational states, characterized by CH/ π interactions between the hydrogen of the central rotating phenylene unit and the aromatic ring of the steroid, are at the origin of the pronounced upfield ¹H chem-

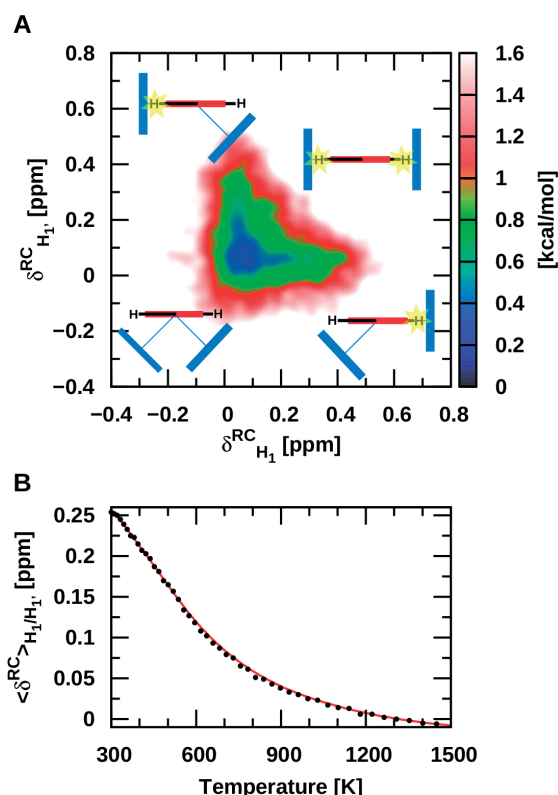


Figure 5.6: Two-dimensional map of the shift induced by the current created by the aromatic ring of the steroids on H₁ and H_{1'}. Pople's ring current model was used to calculate the ring current chemical shifts. B: the averaged ring current chemical shifts with respect to temperature.

ical shift evident at low temperatures. Increasing the temperature results in an increasing population of CH/ π unbound rotational states, leading to the displacement of the ^1H NMR signal of the rotator C-H group toward lower fields. In such a situation, static computation scan help to identify a handful of relevant conformers of both types (bound and unbound), yet the direct one-to-one relationship between the most prevalent conformational states at a given temperature and the observed chemical shifts remains undetectable. REMD, conversely, delivers a more comprehensive picture by disclosing key changes in the relative populations associated with temperature-dependent rotational processes. The REMD@DFTB3 results in Figure 5.6 a for a temperature of 300 K are analyzed in terms of ring current chemical shielding at the position of the hydrogen atoms H₁ and H_{1'} for each structure. The shielding cone created by the aromatic ring of the steroids at proton H₁ and H_{1'} were evaluated using Pople's ring current model.³⁰² The region above 0 ppm is characterized by CH/ π bound states, whereas values around 0 ppm are representative of CH/ π unbound states. Accordingly, the results show

that the enthalpy driven region corresponds to a set of conformers possessing only one CH/ π interaction. In other words, if a CH/ π interaction exists with one of the steroid aromatic rings ($\delta_{H_1 \text{ or } H'_1} > 0$ ppm in Figure 5.6a), the other side is essentially unbound ($\delta_{H_1 \text{ or } H'_1} = 0$ ppm) at 300 K. Increasing the temperature displaces the population toward the CH/ π unbound states (around 0,0) with both hydrogen atoms lying outside the shielding zone (Figure 5.6b). The overall picture is clearly visible in the variable temperature ^1H NMR experiment¹¹¹ but the REMD data provides significantly deeper insight into the dynamic process by revealing the relative conformational population evolution when varying the temperature as well as the direct temperature dependence of the measure properties (Figure 5.6b). For such processes, it is rather unrealistic to neglect the conformational entropic contributions and calculate the “static” population of a few representative conformers randomly extracted from a minimum-energy structures search, as is often done in computational chemistry.

5.3.4 Cinchona alkaloid

The final illustrative application of REMD@DFTB focuses on asymmetric catalysis, more specifically on the use of cinchona-based primary amines as chiral phase-transfer catalysts (Figure 5.7). This class of compounds, derived from natural sources, has been identified as a promising alternative to other amino-catalysts such as proline.^{303–305} (**A**) As well as its quinidine diastereoisomer activate various carbonyl compounds with a consistently high level of stereocontrol. Rather than addressing the origin of the stereoselectivity, which depends heavily on: concentration and nature of the acid co-catalyst, formation of hydrogen bond motifs, solvent effects, and so forth, here we revisit the gas phase conformational behavior of (**A**), which is one of the primary elements responsible for the control of stereochemistry. According to our simulations, the rich conformational behavior of (**A**) is not easily rationalized in terms of four relevant (anti/syn and closed/open) conformations.^{298,306} The 2D conformational map with 5, 10, 15, and 20 kcal/mol isocontours given in Figure 5.7 suggests a somewhat richer conformational picture consisting of four easily accessible conformational regions (**1-4**) and one that is less frequently visited (**2'**). The most populated region, **1**, essentially encompasses a 60° range for the $C_8 - C_9 - C'_4 - C_{4a}$ angle with no significant energy barrier. The most illustrative conformation associated with **1** (i.e. lowest free energy) is characterized by an angle around 90°. Previous NMR measurements³⁰⁶ in apolar solvents along with static computations²⁹⁸ established the anti-open (**1**) and syn-open (**4**) conformers as being energetically comparable and ~5 kcal/mol more stable than the “closed” conformers, with the quinuclidine nitrogen lone pair in gauche arrangement with respect to quinoline group (e.g. **2**, see Table 5.2). The information extracted from REMD follows this interpretation while also revealing an additional accessible conformational space (**3**) lying between the free energy of **1** (or **4**) and **2**, and a flatter PES around the lower free energy minimum. The broader conformational space of **1** is particularly illustrative of the importance of capturing the full entropic and anharmonic

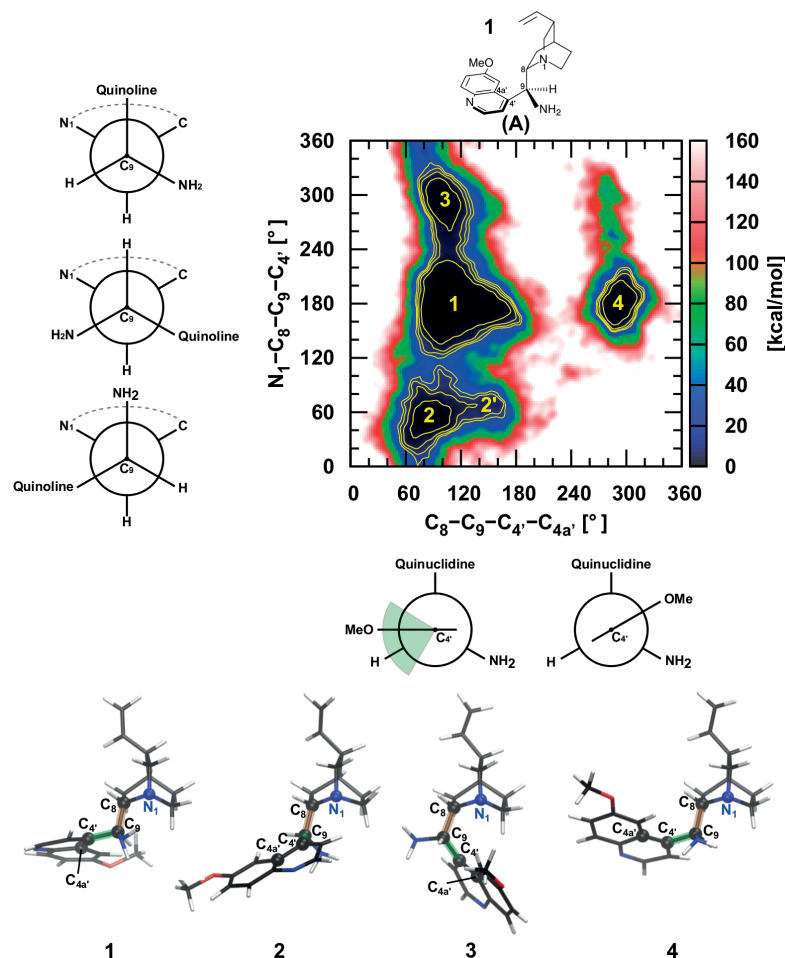


Figure 5.7: Two-dimensional free energy map of the cinchona alkaloid catalyst (A). 5, 10, 15 and 20 kcal/ mol isocontours are shown in yellow. The representative structural figures of the four relevant minima are shown in the bottom.

contributions to the free energy to establish the relative stabilities between **1** and **4**. In fact, while region **4** is slightly favored at both static DFTB3 and DFT-D levels,³⁰⁷ the REMD free energy conformational minimum corresponds to **1** (Table 5.2). Another crucial aspect for catalysis concerns interconversion between the different conformational regions. It is clear from Figure 5.7 that complete rotation around the C₈-C₉ bond is much more favorable in comparison to rotation around the C₉-C_{4'} bond, which is hindered by the contact between the two polycyclic groups. Even though regions **1** and **4** lie close in energy the syn-open conformers (**4**) are trapped and cannot easily convert into other conformational states. Finally, it is interesting to stress that, in line with previous observation, the diastereomers of quinidine, which carry opposite stereochemistries at carbons 8 and 9, essentially behaves like an enantiomer³⁰³ as illustrated by the 2D conformational energy map provided in Figure 5.8

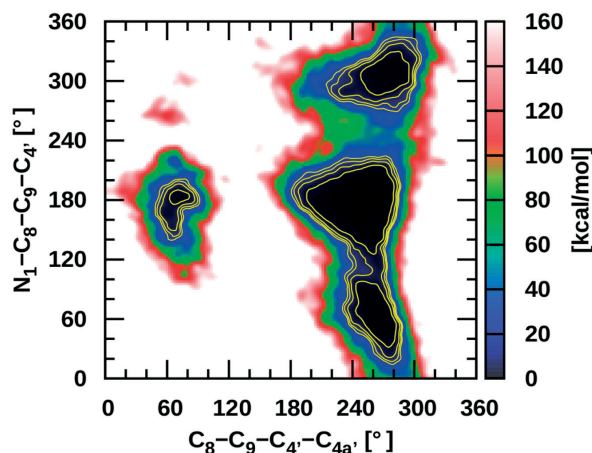


Figure 5.8: Free energy map of the diastereomers of the cinchona alkaloid catalyst. As expected, this map is the mirror of the map presented in Figure 5.7.

	DFTB3/3OB-UFF		BP86-D/TZP
	(Static) Electronic Energy	REMD Free Energy	Static Free Energy
<i>anti-open</i> (1)	1.3	0.00	0.6
<i>syn-open</i> (4)	0.0	1.57	0.0
<i>anti-closed</i> (2)	6.5	2.11	5.4
(3)	1.7	1.94	-

Table 5.2: Relative electronic and free energies (in kcal/mol) of the cinchona alkaloid conformers determined from gas phase static computations or from the REMD simulations (DFTB3/3OB-UFF) at 300K. The geometries for computing the static DFTB3 energies are optimized at the same DFTB3 level. The BP86-D results are taken from Ref.²⁹⁸

Regarding the crucial role played by lowest-energy conformers and by conformational barriers in the enantiodifferentiation processes, clearly having access to the conformational free-energy landscape is an undeniable asset. While we stress that the present analysis was performed at a fairly low electronic structure level (DFTB3) and in the gas phase (a limitation of the current implementation), the key findings related to the conformational landscape and population remain valid. Based on these finding, we believe that computational studies in the field of asymmetric catalysis would benefit greatly from accessing more realistic pictures of the conformational and reaction dynamics for each step of the catalytic cycle.

5.4 Scope and Limitations

Here, the coupling of REMD with DFTB (REMD@DFTB3) via the i-PI dynamic driver has been introduced and its utility demonstrated as a tool to solve problems in computational organic chemistry. Through the study of four illustrative examples, we have highlighted the importance of mapping the free energy of the PES to increase chemical understanding. As opposed to static electronic structure computations, exploring chemistry via REMD often reveals unexpected or unimagined features even for seemingly quite simple systems. This fact is exemplified by the search for low energy conformers of dithiacyclophane and analysis of the CR of SBV. Chemically more complex examples further illustrate this point by showing the important role played by dynamics in reproducing experimentally observable properties (such as proton chemical shifts) and the understanding of underlying chemical structure and its role in determining stereochemistry during organocatalysis. In general, we hope to have shown the utility of using enhanced sampling approaches, such as REMD, to better understand and solve problems of interest to computational organic chemists. Nevertheless, REMD@DFTB3 still presents some flaws: only mono-molecular systems can be treated. The high temperatures of some of the replicas would destroy any non-covalent interaction blowing up the system. Reducing the highest temperature of the REMD is not a valid solution since a lower temperature would correspond at a reduced sampling of the phase space, thus reducing the effectiveness of the approach. The next section illustrate a possible solution of the problem that already provide promising results.

5.5 Toward a More General REMD@DFTB3

Molecular dynamics simulations in the gas phase are usually performed in a periodic boundary box with a size length that is several times the dimension of the investigated molecule. The large dimension of the box avoids spurious interactions between the molecules in the different boxes that would alter the sampling. At the usual temperature of the highest replica, a simulation of more than one molecule in an otherwise empty box, would essentially result in the mapping of configurations of non-interacting molecules, with near zero probability of reaction between the different molecules. For this reason, a containment vessel is needed that will impose the molecules to stay in close proximity. To emulate the presence of a reaction vessel, we apply an artificial harmonic potential (*B* -bias-) with force constant *k* on the distance between the barycenter of the molecules (R_{ij}^{com}), that starts acting when distance R_{ij}^{com} is larger than a threshold (R_{ij}^t), (see Eq. 5.3) emulates the effect of a reaction vessel.

$$B(R_{ij}^{com}) = \begin{cases} k(R_{ij}^{com} - R_{ij}^t)^2 & \text{if } R_{ij}^{com} - R_{ij}^t \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (5.3)$$

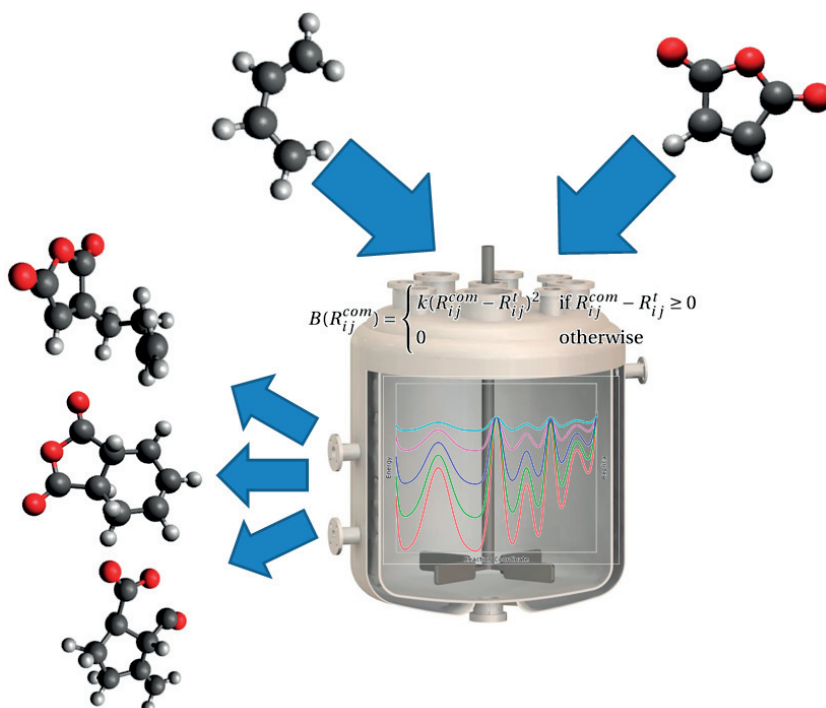


Figure 5.9: A qualitative picture that shows the advanced version of the REMD@DFTB3. Using a 1,3-butadiene and a maleic anhydride, our method is able to provide information on many products other than the Diels-Alder expected one.

The tuning of k is not straightforward: the harmonic potential must counterbalance the kinetic energy of the investigated system, thus, high temperature replicas need an higher k value. On the other hand, the probability to exchange the configuration between two replicas decreases exponentially with the difference between the potential energy B of the simulations. In the depicted scheme, the bias (B) is coupled to the temperature of the replica and, at the same time, acts on the acceptance probability, such that the number of replicas to make the REMD effective increases with the temperature. The Hamiltonian REMD scheme (see Section 2.3.2) allows the system to explore the PES by “alchemically” decreasing the potential energy barriers between two minima rather than through increasing the temperatures. As such, the kinetic energy among the replicas does not change and they can all have exactly the same k .

This generalization of REMD@DFTB3 is still a work in progress, but the first results look promising. For instance, Figure 5.9 illustrates all the products associated with the reaction of 1,3-butadiene with maleic anhydride. In addition to the expected Diels-Alder product (2), alternative products and reaction pathways were also identified (1, 3). The current simulation has not yet reached convergence and some technical adjustments still need to be made, but the current qualitative picture is encouraging. We are excited about the perspective of the

Chapter 5. Putting forth REMD as a tool to solve problems in organic chemistry

proposed technique, which has not previously been exploited by the quantum chemistry community. We believe it will enable the exploration new targets and new territories that could reinforce our understanding of chemical processes and change our way of tackling computational chemistry problems.

6 Overcoming the drawbacks of standard density functional approximations

This work was performed in collaboration with our Master Student, Alberto Fabrizio.

6.1 Introduction

The preceding chapters exploited the fast DFTB electronic scheme to achieve better statistical sampling and prospectively describe large molecules quantum mechanically. In the context of organic semiconductors, the DFTB-dDMC formalism, as introduced earlier, is certainly appealing for the structural and energetics modeling of neutral assemblies of thiophene-based compounds hold together by van der Waals forces. While those neutral systems typically model the resting state of semiconductors, charged radical π -dimers species best represent the typical charge carriers.²⁰ Unfortunately, in contrast to the relatively straightforward computation of neutral units, the charged species are too challenging for DFTB.³⁰⁸

In this context, the logical way beyond DFTB, is to use the parent Kohn-Sham Density Functional Theory,^{46,47} that is the most utilized method in computational chemistry.

While exact in principle, Kohn-Sham DFT suffers in practice from the impossibility to define analytically a functional connecting the electronic density with the exchange and correlation energies.⁴⁷ Thus, the underlying functional dependence must be approximated. It is generally obtained from fits to experimental data or from theoretical arguments. The on-going development of novel density functional approximations (DFAs) is not only of fundamental value but also very useful in the quest of approaching an object that is known to exist.³⁰⁹ As of today, the commonly used approximations still suffers from major shortcomings³¹⁰ and those are directly affecting the description of charged radical π -dimers that we target herein.²⁰ In those systems, as well as in various other chemical applications, the inaccurate treatment of both

the long-range effects in the exchange and correlation holes,⁵⁴ causes crucial failures.^{26,311–322} The exchange-correlation potential (v_{xc}) of standard functionals decays exponentially and lack the correct asymptotic ($-1/r$) behavior.^{323,324} This too rapid decay is at the origin of the delocalization error that causes the overstabilization of fractionally charged fragments^{325–329} or charge transfer complexes.^{330,331} Alternatively, the approximations associated with the correlation hole results in a lack of dispersion interactions,^{26,314–317} which are essential to describe many intra- and intermolecular phenomena such as those present in organic semi-conductor materials.³³²

Numerous efforts have been placed in resolving each of these shortcomings. The most attractive and practical approach to reduce the delocalization error is the introduction of long-range corrected (LC) exchange functionals³³³ originally developed by Savin and coworkers.³³⁴ The separation generally makes use of an Ewald-type partition based on an error function $\text{erfc}(\omega r/r)$ that regulates the Hartree-Fock exchange contribution at different electron-electron distance (r). The contribution of the DFT exchange decreases as $\text{erfc}(\omega r)/r$ with the parameters ω controlling the range partitioning. The LC approach corrects the asymptotic behavior “forcing” the proper convergence of v_{xc} at large distances. Long-range-corrected exchange functionals (e.g., LC-BLYP,³³⁵ LC- ω PBE,^{336,337} CAM-B3LYP,³³⁵ ω B97X³⁴) have improved the results considerably when the delocalization error is important even though the choice of range-separation parameter remains inconvenient and debated.^{338–340} System-dependent alternatives, consisting in tuning the parameter such as to reproduce the HOMO and LUMO energies, lack size-consistency and are thus even more cumbersome. As emphasized later, the present work relies on another philosophy where the range parameter is fitted jointly with other functional parameters on a large training set.^{34,341} Still the class of LC functionals offers no improvements for the treatment of van der Waals interactions. In the computational chemistry community, this issue is usually addressed using various strategies such as (1) improving the non-locality of the correlation functional to eliminate the root of the problem^{55–58} (2) fitting very flexible functional forms^{61–65} (M06-2X);⁶⁶ (3) incorporating a posteriori energy corrections;^{31,67–70} or (4) adding an atom centered nonlocal potential that accounts approximately for dispersion effects, as such as the Dispersion Correcting Atom Centered Potentials (DCACP).^{71–75} The *a posteriori* energy correction approach has been one of the research focuses of our laboratory that is best exemplified by the implementation of the density-dependent, dDSC³¹ correction in multiple computational chemistry software (e.g., QChem,³⁴² GAMESS,^{343,344} ADF,^{280,345} VASP³⁴⁶).

Interestingly, the DFT community generally attempts to improve the description of systems suffering from one of the two shortcomings but never of those affected by a subtle interplay between both the delocalization error and the London dispersion issues. As an example, Steinmann and Corminboeuf showed in 2012 that none of the standard dispersion corrected functionals achieve accurate binding energies and proper dissociative behavior of the charged

radical π -dimers mentioned above.²⁰ More specifically, dispersion-corrected global hybrids strongly overbind those complexes at the equilibrium and exhibit to the wrong dissociative behavior. Alternatively, long-range corrected exchange approximations combined with density dispersion corrections restore the correct asymptotic behavior but strongly underestimate the interaction energies.

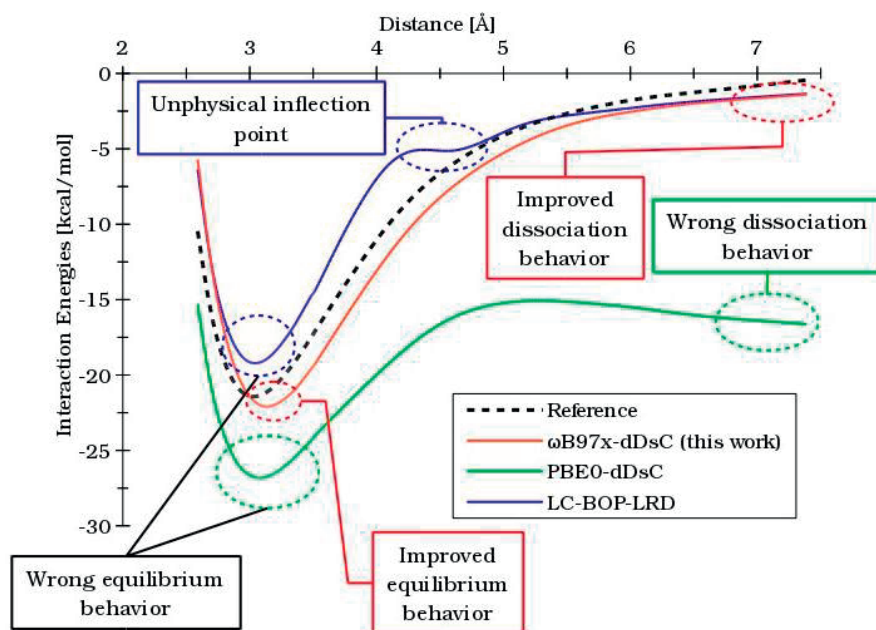


Figure 6.1: This picture illustrates the main problems when applying popular DFAs to radical cation π -dimer and compare the results with the ω B97X-D³⁴ and the functional we developed.

The objective of this chapter is to develop a well-balanced long-range corrected exchange functional rooted in the ω B97X^{34,341} family that is trained jointly with the density-dependent dispersion correction, dDsC. The more general purpose is to improve the description of π -dimer radical cations without deteriorating other general trends for thermochemistry, kinetics and non-covalent interactions.

6.2 Theory

The ω B97X-dDsC functional is built around the B97 density functionals from Becke.⁶² In this functional the Local Density Spin Approximation (LSDA) energies^{52,53} are corrected with a 4th-order polynomial expansion, whose coefficients are indicated with c and fitted on higher

level computations, of the reduced density gradient (s) attenuated by the factor γ

$$u = \frac{\gamma s^2}{1 + \gamma s^2} \quad (6.1)$$

$$E_x^{B97} = \sum_{\sigma} \int e_{x\sigma}^{LSDA}(\rho_{\sigma}) \sum_{i=0}^4 c_{x,i} u^i d\mathbf{r} \quad (6.2)$$

$$E_{c\sigma\sigma}^{B97} = \int e_{c\sigma\sigma}^{LSDA}(\rho_{\sigma\sigma}) \sum_{i=0}^4 c_{c\sigma\sigma,i} u^i d\mathbf{r} \quad (6.3)$$

$$E_{c\alpha\beta}^{B97} = \int e_{c\alpha\beta}^{LSDA}(\rho_{\alpha\beta}) \sum_{i=0}^4 c_{c\alpha\beta,i} u^i d\mathbf{r} \quad (6.4)$$

$$(6.5)$$

$e_{x\sigma}^{LSDA}(\rho_{\sigma})$ is the LSDA exchange potential while $e_{c\sigma\sigma}^{LSDA}(\rho_{\sigma\sigma})$ and $e_{c\alpha\beta}^{LSDA}(\rho_{\alpha\beta})$ are the same-spin and the opposite-spin correlation potentials. The ω B97X³⁴ scheme from Head-Gordon and Chai imposes a Long-Range hybrid correction to the B97 functional and add a small contribution ($c_{x,HF}$) of exact exchange also at short distances, so that the exchange-correlation energy become

$$E_X^{\omega B97X} = E_x^{B97} + c_{x,HF} E_x^{HF-SR(\omega)} + E_x^{HF-LR(\omega)} \quad (6.6)$$

The ω B97X adds two adjustable parameters (ω and $c_{x,HF}$) to the 15 of the B97 DFA. The energy of the ω B97X-dDsC writes

$$E^{\omega B97X-dDsC} = E^{\omega B97X} + E^{dDsC} \quad (6.7)$$

where only the $E^{\omega B97X}$ serves to optimize the electron density since the inclusion into the self consistent field of the *a posteriori* dispersion correction (E^{dDsC}) has been proven to do not improve any results.²⁰⁴

As already mentioned in Section 2.2, dDsC is an atom-pairwise *a posteriori* dispersion correction

$$E_{disp} = - \sum_{i=1}^{N_{at}} \sum_{j>i}^{N_{at}} \sum_{n=3}^{n=5} f_{2n}(R_{ij}) \frac{C_{2n}^{i,j}}{R_{ij}^{2n}} \quad (6.8)$$

The total dispersion energy is the sum of the contributions from each couple of atoms in the system. The atomic-pairwise dispersion energy is computed damping, trough a function of the interatomic distance ($f_{2n}(R_{ij})$), the pure dispersion contribution ($C_{2n}^{i,j}/R_{ij}^{2n}$). The sum over n accounts for high order dispersion terms than C_6 : C_8 and C_{10} . The characteristic of dDsC is that both, the $C_{2n}^{i,j}$ ¹⁰⁶ and the $f_{2n}(R_{ij})$,³¹ depend on electronic structure properties. In particular the $C_{2n}^{i,j}$ are computed accordingly to the Becke-Jonson exchange-hole-dipole-moment formal-

ism^{99–104} but simplifying the computation using the dominant classical Hirshfeld scheme¹⁰⁵ to partition the electron density among the atoms. The damping function is the Tang and Toennies (TT)⁹⁶ one improved with a damping factor b

$$f_{2n}(bR_{ij}) = 1 - \exp(-bR_{ij}) \sum_{k=0}^{2n} \frac{(bR_{ij})^k}{k!} \quad (6.9)$$

to further damp the correction at covalent distances, where the density functional describe better the interactions.¹⁸² The factor b is actually a function itself

$$b(x) = F(x)b_{ij,asym} \quad (6.10)$$

The argument x is a covalent bond index substituting of a “traditional interatomic distance”.³¹

$$F(x) = \frac{2}{e^{a_0 x} + 1} \quad (6.11)$$

damp the TT function when the atoms interact covalently ($x \rightarrow \infty$). $b_{ij,asym}$ is the asymptotic value of $b(x)$ and its value is based on the polarizability of the atoms in the molecule evaluated through the atomic volumes provided by the classical dominant Hirshfeld approach. The parameters a_0 and b_0 (adjusting the short and the medium range behavior of the correction) are empirical and, in the present work, fitted together with all the ω B97X parameters rising the final number of parameters to 19. Fortunately, some constraints decrease the number of the parameters to 16:⁶² the first coefficient of the correlation energy terms must be one and the exchange contributions at short range must sum to one.

6.3 The Algorithm

The optimization algorithm minimizes the Mean Absolute Error (MAE) computed over a large training set (*vide infra*). The optimization of 16 parameters is not an easy task, especially if each MAE evaluation requires several minutes of computation. To reduce the computational costs of the optimization, a single point energy computation with a given ensemble of parameter values provide the optimized electron density for each molecules in the training set. The density is then frozen while optimizing the parameters until reaching, a density specific, minimum MAE. The optimized parameters serve, then, to retrieve a new electron density that will be frozen during a new parameters optimization. Once the parameters do not change among two consecutive optimization the procedure stops since the algorithm found a minimum. The Figure 6.2 displays the entire procedure in a block diagram. The red square identify the inner part of the loop, where the parameters are optimized with the densities frozen while the blue part is responsible for the i/o handling and the decision making. A variant of the Broyden-Fletcher-Goldfarb-Shanno algorithm (L-BFGS-B),^{348,349} as implemented

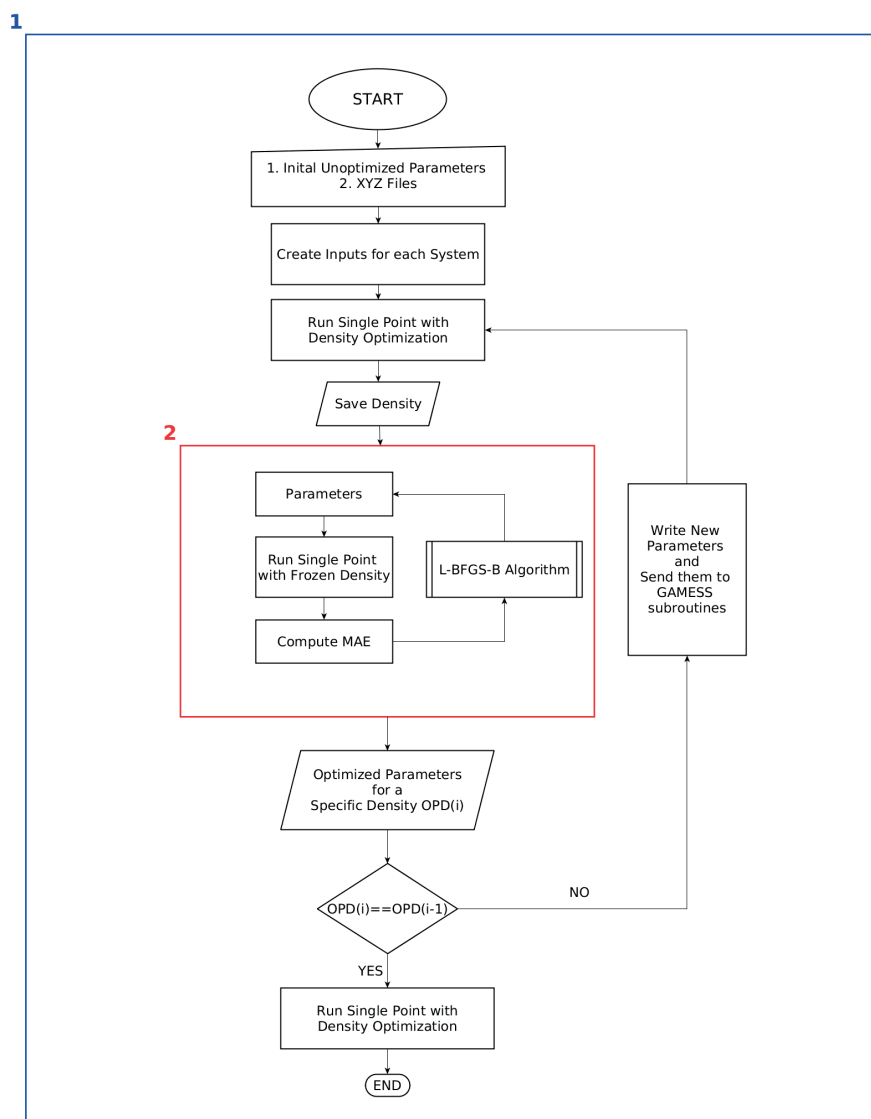


Figure 6.2: Optimization algorithm scheme. Taken from ref.³⁴⁷

in the Scipy packages,³⁵⁰ permits the fitting of the specific-density parameters specifying constraints to avoid non-physical results (a_0 and b_0 must be positive). A small number of subroutine, extracted from the GAMESS-US software,^{343,344} optimize the performance of the frozen density single point computation. Moreover, storing all the useful data produced by the density optimization in the Random-Access Memory (RAM) as binary files improve the speed performance so much that only 52 seconds are necessary to run a frozen-density single point computation over 350 molecules with 16 CPUs.

6.4 Computational Details

The development and the validation of the ω B97X-dDsC are performed with a modified version of the GAMESS-US software.^{343,344} Spin-restricted theory is used for singlet state systems and unrestricted-spin theory for triplet ones. All computations use Karlsruhe def2-TZVP basis set.^{153,154} To speed up the fitting, the integration uses the SG-1 grid.³⁵¹ During the validation and evaluation the Euler-Maclaurin-Lebedev (75/3012) provided more accurate results.

Being still a work in progress, two versions of the functional exist: ω B97X-dDsC and ω B97X-dDsC(Orel) that differ for the training set. Seven data set, including atomization energies (G2),³⁵² ionization potentials (IP),³⁵³ electron affinities (EA),³⁵³ proton affinities (PA),³⁵³ hydrogen and non-hydrogen transfer barrier heights (HTBH and NHTBH)⁸² and non covalent interactions (S22),³⁵³ train the ω B97X-dDsC. Since the influence of dispersion on the global training set is small, the S22 data set is weighted ten times in computing the MAE. The ω B97X-dDsC(Orel) has six systems taken from the Orel26rad added to the same training set of ω B97X-dDsC. In both cases the guessed parameters are the ω B97X-D³⁴ for the core functional and the ω B97X+dDsC for the dDsC part. The ω B97X+dDsC refers to the ω B97X functional plus the dispersion correction fitted *a posteriori* on the functional.

The performance of both versions of the functionals are evaluated computing the MAE over the entire Orel26rad,²⁰ Pi29n²⁰ and S66¹⁴⁶ data sets. The S66 is a popular data set for dispersion interactions based on biological systems while Pi29n provide organic electronics precursor in their resting state, it is similar in spirit to the Orel26rad but contains only neutral systems.

6.5 Results and Discussions

This section presents all the results of the parameterization and validation of the two version of the functional (ω B97X-dDsC and ω B97X-dDsC(Orel)). Often the results will be compared with the ω B97X-D since these functionals share the same core and because ω B97X-D is the best performer on the Orel26rad and one of the best on Pi29n. As a meter of comparison the results

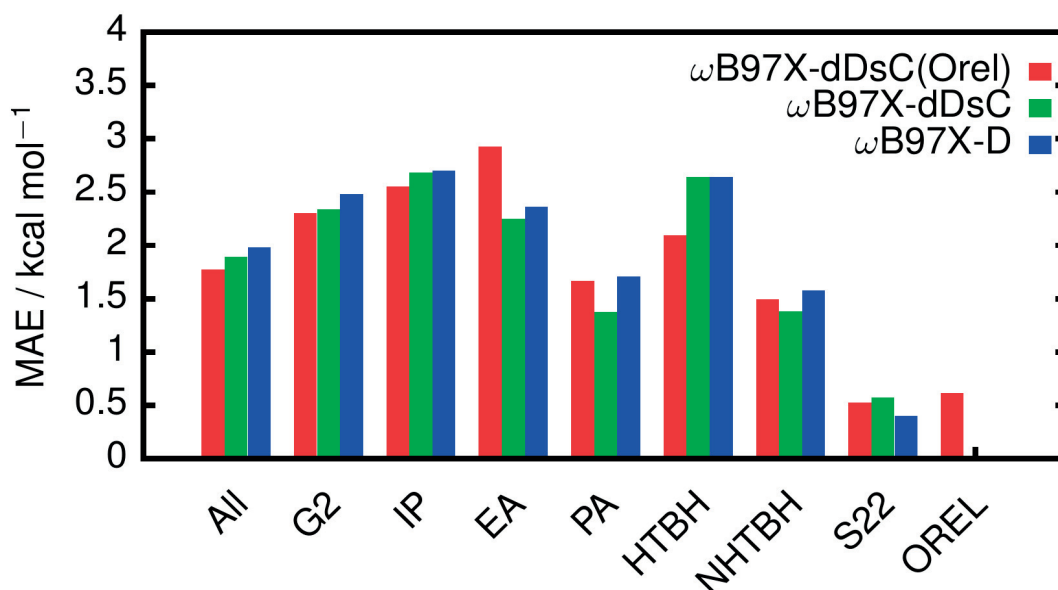


Figure 6.3: Mean Absolute Error on the training set, including atomization energies (G2), ionization potential (IP), electron affinities (EA), proton affinities (PA), hydrogen transfer barrier heights (HTBH), non-hydrogen transfer barrier heights (NHTBH), non-covalent interactions (S22) and radical cation dimers (OREL). For each data set is indicated the number of systems taken in the training set.

of two other functionals with similar characteristics (PBE0-dDsC,^{31,206} LC-BOP-LRD^{85,86} and LC- ω PBE-dDsC^{31,354–356}) will be presented.

6.5.1 Functional development and training

Figure 6.3 compares the performance of the ω B97X-dDsC, ω B97X-dDsC(Orel) and the ω B97X-D against the training set. The first column show that the best performing functional is ω B97X-dDsC(Orel). However this result could be biased because of the additional six systems from Orel26rad which are absent in the training set. Concerning the ability of each functional to describe specific properties, ω B97X-dDsC and ω B97X-dDsC(Orel) are generally better than the ω B97X-D, even though the introduction of the radical systems in the training set deteriorates (0.6 kcal/mol) the performance of the ω B97X-dDsC(Orel) on the electron affinity in respect of both the other two functionals. Interestingly, both our functionals describe the dispersion interactions better than the ω B97X-D, as shown by the smaller MAE associated with the S22 data set (0.3 kcal/mol). This is very impressive since the Head-Gordon and Chai's functional is already one of the best functionals in the treatment of non-covalent complexes.

The fact that the three functionals perform similarly should not be unexpected since all of them share the same mathematical structures except for the dispersion correction. Moreover, ω B97X-dDsC has been trained on the same training set of ω B97X-D and only few differences are present in the training set of the Orel version.

The Optimized Parameters

Table 6.1 shows the 19 optimized parameters that tune ω B97X-dDsC and ω B97X-dDsC(Orel). The first two columns report the parameters of the ω B97X-D and of ω B97X+dDsC.^{31,34} The horizontal lines separate the numbers by the specific aspect of the exchange-correlation functionals they rule: (from the top to the bottom) (1) exchange, (2) same-spin and (3) opposite-spin correlation, (4) range-separation and (5) dispersion damping function.

Table 6.1: The Adjustable Parameters of the Functionals

	ω B97X-D	ω B97X+dDsC	ω B97X-dDsC	ω B97X-dDsC_Orel
$c_{x,hf}$	0.222036	0.157706	0.227445	0.186914
$c_{x,0}$	0.777964	0.842294	0.772555	0.813086
$c_{x,1}$	0.66116	0.726479	0.596070	0.694966
$c_{x,2}$	0.574541	1.04476	0.534383	0.821741
$c_{x,3}$	-5.25671	-5.70635	-5.383440	-5.563391
$c_{x,4}$	11.6386	13.2794	11.415442	12.398156
$c_{c,\sigma\sigma,0}$	1	1	1	1
$c_{c,\sigma\sigma,1}$	-6.90539	-4.33879	-6.923696	-5.594674
$c_{c,\sigma\sigma,2}$	31.3343	18.2308	31.363830	24.741653
$c_{c,\sigma\sigma,3}$	-51.0533	-31.743	-50.961373	-41.391495
$c_{c,\sigma\sigma,4}$	26.4423	17.2901	26.548992	21.866294
$c_{c,\alpha\beta,0}$	1	1	1	1
$c_{c,\alpha\beta,1}$	1.79413	2.37031	1.835821	2.113316
$c_{c,\alpha\beta,2}$	-12.0477	-11.3995	-12.133922	-11.715107
$c_{c,\alpha\beta,3}$	14.0847	6.58405	14.279191	10.488708
$c_{c,\alpha\beta,4}$	-8.50809	-3.78132	-8.665881	-6.236470
ω	0.2	0.3	0.204365	0.248442
a	6	-	-	-
a_0	-	27.7	3.77	64.31
b_0	-	1.38	1.14	1.42

Even if a conclusive consideration is impossible just from the analysis of the functional parameters, general trends are interesting. The parameters of ω B97X-dDsC are the closest to those in the ω B97X-D, while the parameters that tune the exchange-correlation core of ω B97X-dDsC(Orel) are always halfway between those of ω B97X-dDsC and those of ω B97X+dDsC. Surprisingly, the the dispersion correction damping function parameters are quite differ-

ent: especially in the ω B97X-dDsC(Orel) the b_0 parameter of the damping factor is much higher in respect of any other dDsC parameterization, meaning that the dispersion correction contribution is low. This last point worth some further investigation.

Besides any possible interpretation, these results still reflect the remarkable effects obtained adding the six Orel26rad systems to the training set of ω B97X-dDsC(Orel).

6.5.2 Validation

The purpose of our new dispersion-corrected, range-separated hybrid density functionals is to improve the treatment of systems that are severely affected by the coexistence of the delocalization error and dispersion interactions. This consideration suggests that the Orel26rad data set, containing several examples of such systems, must be included in the test set. The Pi29n is a data set that provide neutral organic electronics precursor complexes. In other words, Pi29n is the neutral counterpart of the Orel26rad. Another important test set to validate our results is the S66: a popular data set concerning weakly bounded dimers of biological interest.

The validation will take place in three different sections separating the three different subject of the benchmark: (1) interaction energies at equilibrium, (2) profiles, (3) geometries, (4) basis set dependence.

Interaction Energies at Equilibrium

Figure 6.4 displays the accuracy of equilibrium interaction energies of the ω B97X-dDsC and the ω B97X-dDsC(Orel) using the Mean Absolute Error in respect of reference data. The performance of ω B97X-dDsC is comparable to the one of ω B97X-D on both the Orel26rad and the Pi29n data set. In contrast, ω B97X-dDsC(Orel) outperforms all the other functionals in the treatment of the radical cation dimers but at the price of a small deterioration of the neutral compounds description. The performance on radical cations of ω B97X+dDsC is still quite satisfying, but this functional clearly fails to describe the correct behavior of the neutral complexes of Pi29n, with a mean absolute deviation around 2.9 kcal/mol. The terrible performances of the PBE0-dDsC and LC- ω PBE-dDsC confirm the challenging nature of the Orel26rad compounds: both functionals fail severely to describe the equilibrium structure of the radical cation complexes, restoring their accuracy only on the neutral data set. Undoubtedly, the particularly poor results of those standardly used functionals gives prominence to the importance of this work.

The ω B97X-dDsC(Orel) is the only functional able to capture the behavior of both radical cation and neutral complexes within 1 kcal/mol of mean absolute error. The importance of a joint fitting of the functional with the dispersion correction emerges clearly from the

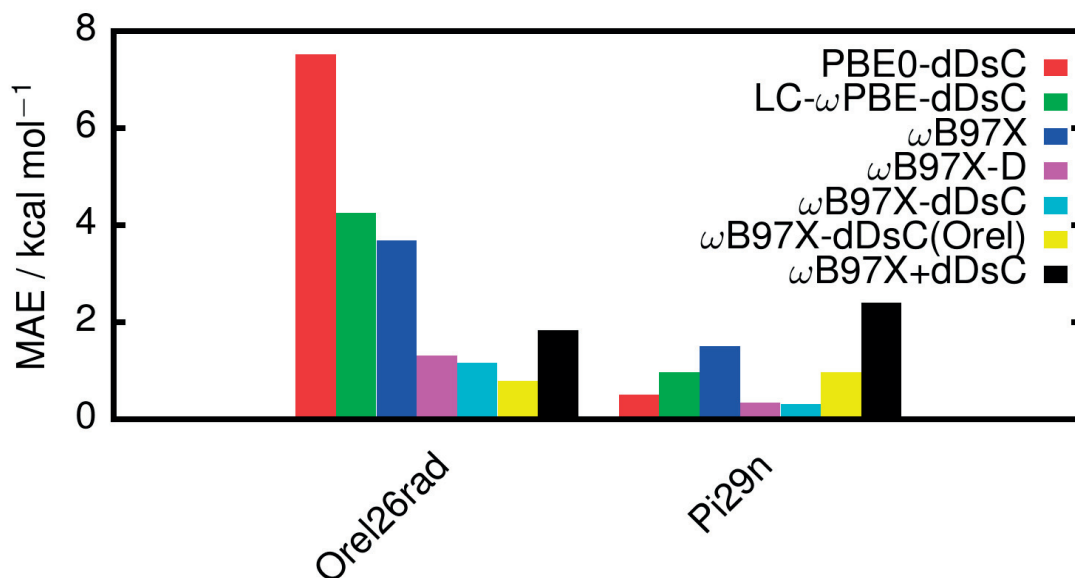


Figure 6.4: Mean Absolute Error of the test set, including the interaction energies of radical cation dimers (Orel26rad) and those of the corresponding neutral compounds (Pi29n).

impressive results (MAE = 0.3 kcal/mol) achieved from the ω B97X-D and ω B97X-dDsC on the neutral Pi29n data set when compared with ω B97X+dDsC (MAE = 2.9 kcal/mol). The ω B97X+dDsC is the worse functional tested on the Pi29n data set. In particular, this functional significantly overbinds the complexes while the ω B97X underestimate the equilibrium interaction energies even though providing more accuracy than the *a posteriori* corrected one. As already claimed by Head-Gordon and Chai,³⁴ the ω B97X functional captures some of the non-local contribution to the correlation even without an explicit dispersion correction. Thus, the more accurate balance between the core and the dispersion correction provided by the joint fitting of the parameters solve the problem.

ω B97X-dDsC(Orel), despite being the best functional up to now, provides inconsistent results on the dispersion focused data sets: the MAE on the S22 is comparable with the other jointly-fitted functionals but when considering the Pi29n, ω B97X-dDsC(Orel) performs sensitively worse than ω B97X-D and ω B97X-dDsC. Figure 6.5 presents the MAE for the S66 dimers divided per dominant interaction energy. The results of the ω B97X-dDsC(Orel) on the S66 are consistent with what observed on the S22 data set. This fact suggests that the deterioration of ω B97X-dDsC(Orel) on the Pi29n is not due to a general problem of the functional concerning neutral systems but is clearly related to some properties of the Pi29n data set itself.

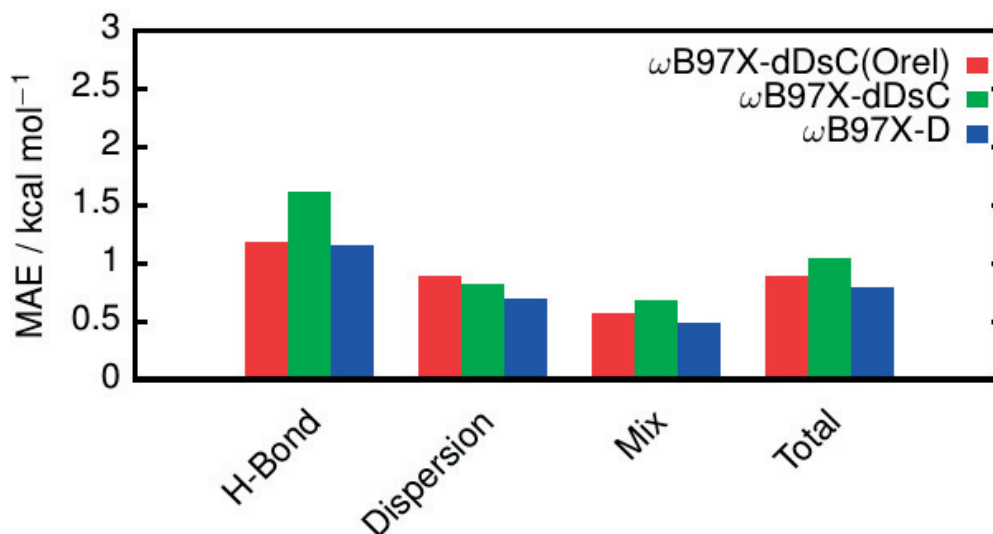


Figure 6.5: Mean Absolute Error on the S66 data set, divided per dominant interactions.

Profiles

The ability to reproduce equilibrium properties is an important feature of any exchange-correlation functional. However, to focus exclusively on equilibrium structures may not be sufficient since the performance of the functional is evaluated just in one point of the entire ground-state potential energy surface. Moreover, the correct description of complex chemical systems out of their equilibrium geometries is fundamental for applications outside the field of electronic structure computations: for example molecular dynamics require accurate performances even far from the equilibrium configuration.

Therefore, the ability of the new functionals to reproduce the interaction energy profiles is monitored using four systems: the antiparallel configuration of the furan and the thiophene dimer in their radical cation state (Figure 6.6 and Figure 6.7) and neutral state (Figures 6.8 and 6.9) provided from the Orel26rad and Pi29n data set.²⁰

PBE0-dDsC and LC-BOP-LRD functionals in the radical cation context show how the accurate description of those systems is challenging. The dispersion corrected global-hybrid PBE0-dDsC displays an unphysical dissociation barrier, in addition to the systematic over-estimation of the binding energy. In contrast, the performance of the dispersion and long-range corrected LC-BOP-LRD is flawed by a sensible under-estimation of the interaction energy of the equilibrium structures and shows an unphysical inflection point near the dissociation limit. ω B97X+dDsC over-binds the equilibrium structures of both furan and thiophene radical

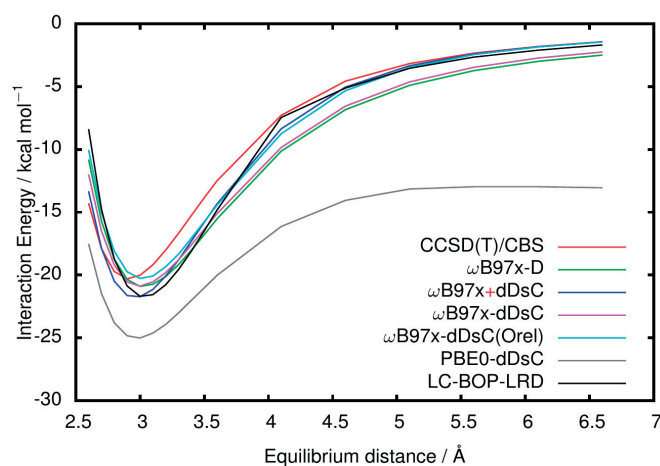


Figure 6.6: Interaction energy profile for the dissociation of the fully optimized furan radical cation dimer.

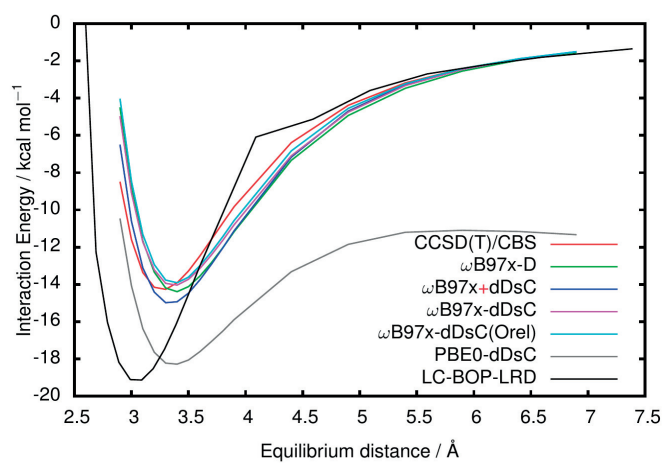


Figure 6.7: Interaction energy profile for the dissociation of the fully optimized thiophene radical cation dimer.

cationic dimers. Nevertheless, this functional describes surprisingly well the dissociation limit of the problematic furan case, indicating that the joint-fitting slightly deteriorates the asymptotic behavior since all the jointly-fitted functionals perform worse in this region of the profile. On the other hand, ω B97X-D and ω B97X-dDsC show again a very similar behavior, providing satisfying results around the equilibrium structure, but missing the correct long-distance trend. ω B97X-dDsC(Orel) is the only fully re-parametrized version of the ω B97X family that describes the long distance behavior almost as satisfyingly as ω B97X+dDsC and, additionally, it does not deviate significantly from ω B97X-D and ω B97X-dDsC around the equilibrium structure.

Higher ω corresponds to a larger portion of the real space where the exact exchange contribution from the long-range correction is present. The solid performance of ω B97X-dDsC(Orel) ($\omega = 0.25$) and ω B97X+dDsC ($\omega = 0.3$) on the description of the dissociation limits, compared to ω B97X-D ($\omega = 0.2$) and ω B97X-dDsC ($\omega = 0.2$) can be explained as an effect of the range-separation parameter (ω): the underestimation of the dissociation energy is probably due to the delocalization error that is deleted by an higher contribution of exact exchange from the long-range correction.

Concerning the equilibrium region of the radical systems, all the jointly-fitted functionals slightly overestimate the interaction energy with the ω B97X+dDsC being the worst of the ω B97X family. A slightly deeper analysis suggests that the dDsC is responsible for this overbind. Unfortunately a more careful fitting of the dDsC parameters would not be sufficient to solve the problem: the values of the C_6 parameters, as well as the damping factor, (see Section 2.2) depend explicitly from the electron density which in turn depends on the ω B97X parameters. Only looking at all the parameters of the functional as an whole could provide more insights. Anyway, both the presented functionals performs excellently especially compared with more popular functionals.

From the preceding statistical analysis, results that ω B97X-dDsC(Orel) deteriorates the description of the Pi29n if compared to the first version of the functional (ω B97X-dDsC). Figures (6.8) and (6.9) show the dissociation curve of the neutral furan and thiophene dimers, treated with all the four functionals of the ω B97X family. These curves confirm that ω B97X-dDsC(Orel) slightly worsen the description of equilibrium structures of both the compounds chosen from the Pi29n data set. The energies obtained subtracting the dispersion contribution from the total energy still show a binding region for the ω B97X+dDsC and the ω B97X-dDsC(Orel). This observation confirms that the ω B97X functional is able to account for a small amount of the dispersion interactions.

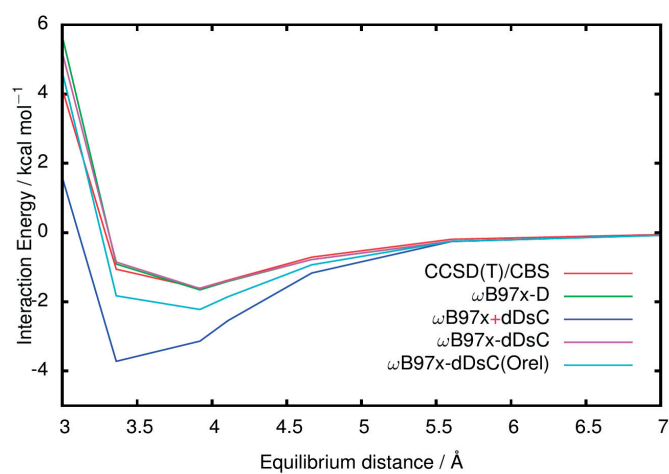


Figure 6.8: Interaction energy profile for the dissociation of the furane neutral dimer.

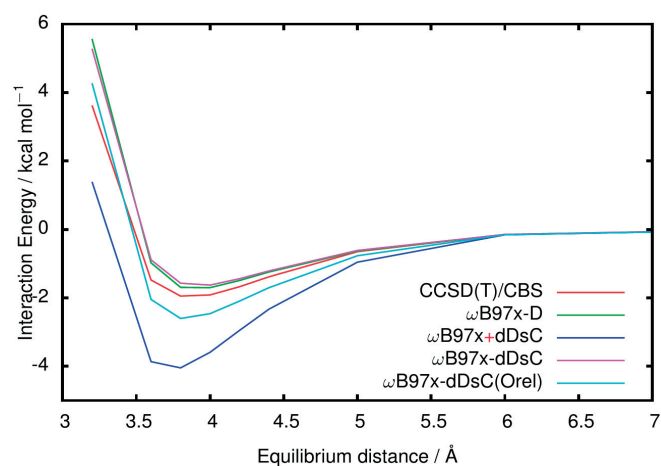


Figure 6.9: Interaction energy profile for the dissociation of the thiophene neutral dimer.

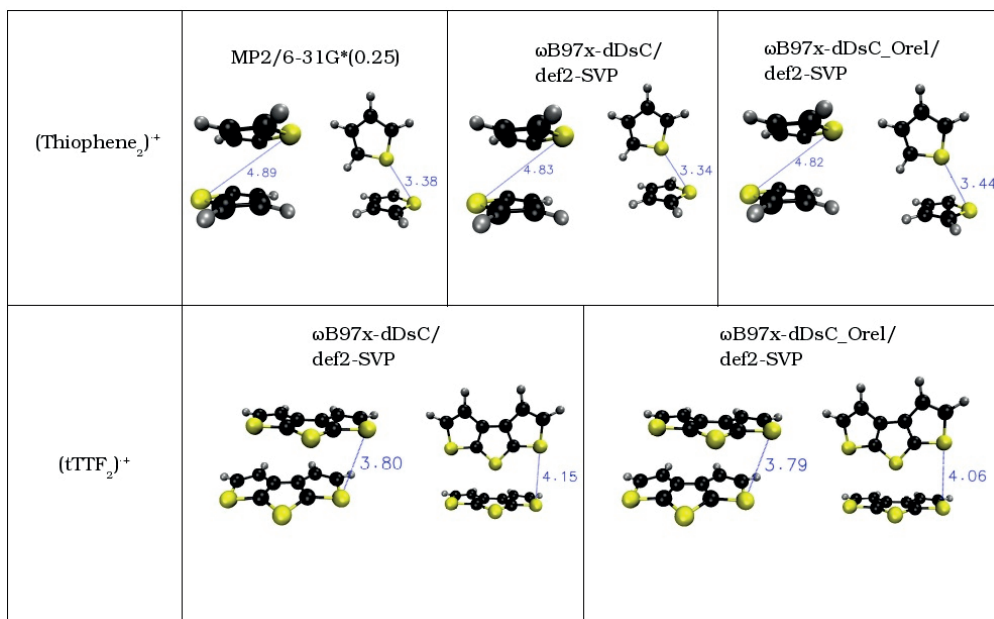


Figure 6.10: Optimized geometries of the radical cationic dimer of thiophene and tri-thienothiophene.

Geometry Optimization

Testing geometry optimization of realistic radical cation complexes serves to demonstrate the broad applicability of our new functionals in routine computations. Note that, in the context of this work, a realistic system is defined as opposed to a model system, such as those of the Orel26rad data set. For this purpose, two different conformations of the radical cation dimer of thiophene and tri-thienothiophene (tTTF₂⁺) are fully optimized. Figure 6.10 shows the results using ω B97X-dDsC and ω B97X-dDsC(Orel). Optimization at MP2/6-31G*(0.25) level^{357,358} provide reference structures. The tests are performed starting with the reference geometries and testing both ω B97X-dDsC and ω B97X-dDsC(Orel). The optimized DFT geometries compare very well with the reference. However, resulting energies differ from what obtained with the reference method: MP2/6-31G*(0.25) predicts the stacked thiophene dimer to be 18.15 kcal/mol more stable than the T-shaped one, while in the case of ω B97X-dDsC and ω B97X-dDsC_Orel the relative stabilization is only about 10 kcal/mol.

The tests over the tri-thienothiophene radical cationic dimers confirm that both the new functionals have the same performance on the optimization of molecular geometries. In this case, both the qualitative and the quantitative agreement between ω B97X-dDsC and ω B97X-dDsC(Orel) is impressive. For instance, the root mean square displacement (RMSD) between

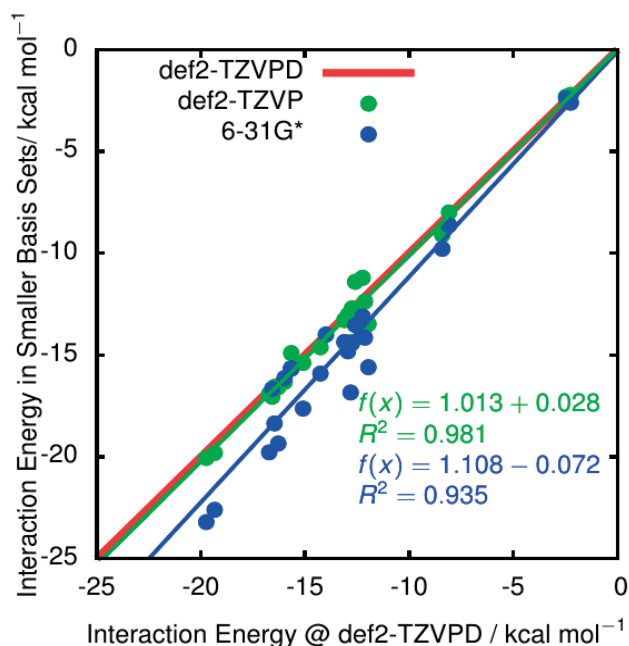


Figure 6.11: Correlation between the interaction energies of Orel26rad, computed with def2-TZVPD and smaller basis sets.

the functionals was as small as 0.071 Å for the parallel-displaced structure and 0.031 Å for the T-shaped one.

Basis Set Dependence

The evaluation of the performance of the functionals uses the Karlsruhe def2-TZVP basis set. However, changing the basis set used in the computations may influence the results. For this reason a comparison of results obtained with different basis sets is mandatory. ω B97X-dDsC(Orel) results has been compared among the def2-TZVPD,^{153,154} def2-TZVP and the very cheap 6-31G*^{157–164} basis sets. Figures (6.11) and Figure (6.12) illustrate the correlation between the interaction energies computed using either def2-TZVP or 6-31G* in function with the energy obtained with def2-TZVPD.

The graphs displays that the ω B97X-dDsC(Orel) is very robust in respect of changes on the basis set. Given that the extremely expensive diffuse functions of def2-TZVPD does not influence much the results of ω B97X-dDsC(Orel) neither on Orel26rad nor on the Pi29n data set. The robustness of the functional is confirmed with the 6-31G*: a basis set with a different conception that, anyway, provide consistent results.

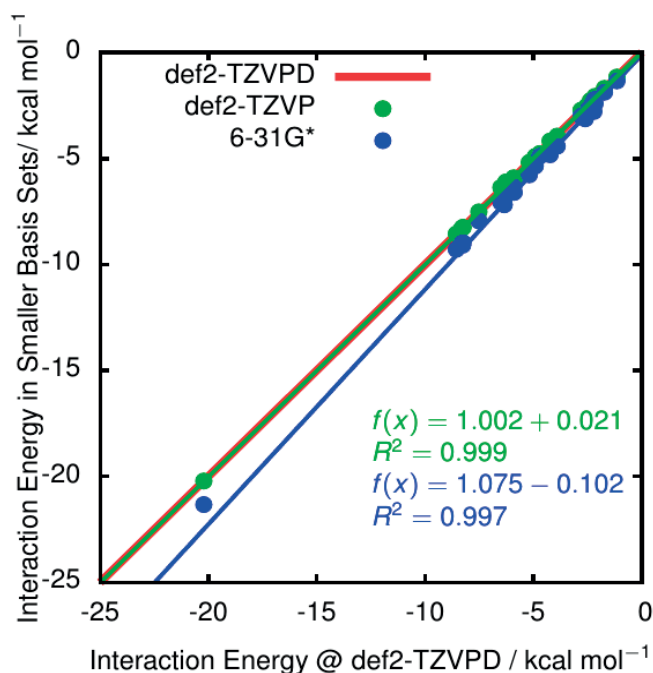


Figure 6.12: Correlation between the interaction energies of Pi29n, computed with def2-TZVPD and smaller basis sets.

To summarize, there is no significant deviation in the performance of ω B97X-dDsC(Orel) using a smaller basis set than def2-TZVPD on both test sets. This robustness of the functional should allow, in principle, the use of a smaller basis set in the future, reducing significantly the computational cost for routine applications.

6.6 Conclusion

This chapter introduced two variants of range-separated exchange functionals fitted jointly with the dDsC density-dependent dispersion correction, ω B97X-dDsC and ω B97X-dDsC(Orel). While the driving force of this work was to improve the description of model systems for organic charge carriers, a relevant byproduct was the elaboration of an efficient optimization algorithm that generates the optimum 16 functional parameters in a reasonable time. The ω B97X-dDsC(Orel) variant was able to greatly improve the description of both the equilibrium and the dissociation limit of π -dimer radical cations but at the cost of deteriorating the treatment of neutral dimer systems. On the other hand, ω B97X-dDsC was found to be more robust than any other ω B97X variant but further improvement on the Orel26rad data set are needed. Now that the optimization machinery is established, our short-term goal is to increase the flexibility of our functional form through incorporating the dispersion contributions up

to C8 and C10. We expect this change to facilitate the achievement of a balance description between dispersion and the delocalization error but also to be strongly beneficial to the modeling of larger molecules than those presented herein.

7 General Conclusions and Outlook

Modern quantum chemistry methods facilitate the description and set the discovery pace of novel functional molecules and materials. Nevertheless, the *in silico* exploration of molecular properties suffers from significant limitations that originate from the trade-off between accuracy and computational cost. In essence, the computational cost of quantum chemistry methods remains too expensive to enable the thorough mapping of extensive chemical spaces or treatments of large chemical systems, yet, at the same time, the inherent approximations associated with these commonly used schemes leads to major qualitative and quantitative failures. In this thesis, some original techniques capable of improving both the accuracy and efficiency of density functional theory (DFT) based methods were presented, with emphasis placed on describing molecular motifs, such as thiophene units, that are relevant to the field of organic electronics.

This thesis began by considering the density functional tight binding (DFTB) formalism and, in particular, its most recent self-consistent variant, DFTB3. The preliminary goal was to improve the accuracy of this efficient scheme and use it as a basis for Born-Oppenheimer molecular dynamics simulations. The DFTB class of methods suffers from the same drawbacks as the parent DFT approximations. In particular, they lack the ability to properly describe the $1/r^6$ attractive term arising from London dispersion. Exploiting the experience of our group in developing an a posteriori density-dependent dispersion correction to account for weak van der Waals forces in DFT (the so-called dDsC), a charge dependent dispersion correction (dDMC), compatible with the DFTB scheme, was introduced. The charge dependency results in a dDMC correction that is influenced by the electronic structure and its performance was demonstrated on the S66 benchmark and on a set of model complexes of typical of organic electronics materials. Given that the self-consistent variants of DFTB readily provide Mulliken charges, dDMC saves up to the 80% of the time compared to alternative corrections, while still providing comparable accuracy to the popular Grimme's D3 correction.

Most of the previous existing DFTB studies are essentially restricted to biological systems while our laboratory focuses on organic semiconductors. For this reason, a preliminary step was validating the use of this fast and efficient scheme for the modeling of molecular precursors to organic electronics such as those represented in the in-house Pi29n training set. Within this context, a caveat in the original DFTB parameterization of sulfur was discovered. In fact, it was shown that any sulfur-containing molecule spuriously binds non-covalently to any other atoms and molecules even in the absence of a dispersion correction. Initially, this artifact prevented both the fitting and the application of our dDMC dispersion correction that could only worsen the already present overbinding behavior. Fortunately, the DFTB community rapidly responded to this concern, and introduced a new set of parameters for third row atoms including sulfur (3OB) where the problem was attenuated (although not suppressed).

DFTB3 offers a considerable improvement in computational cost compared to the parent DFT approaches but the time-scale of DFTB3 molecular dynamic simulations is still very limited, even for small organic molecules. To overcome this problem a technique from classical mechanics was borrowed; we introduced REMD@DFTB3, which maps the DFTB Potential Energy Surface (PES) using energetically independent simulations (replicas) at different temperatures. The highest replicas span all the relevant phase space, whereas the lowest replica extensively samples the energy minima. When applied to standard computational organic chemistry problems, this strategy provided numerous insightful pictures ranging from the identification of unexpected conformers to the demonstration that conformational entropy cannot always be neglected when aiming at rationalizing reaction processes. Notably, this entropy is absent from traditional static quantum chemistry computations. Although very useful, the initial implementation of REMD@DFTB3 was not general in the sense that only intramolecular phenomena could be investigated. Indeed, the use of high temperatures prevents two molecules from meeting one another. Work is now underway on a more general version that couples DFTB3 to the more flexible Hamiltonian-REMD used to build a containment vessel around the molecular system imposing artificial potential barriers. The preliminary results on a Diels-Alder reaction already demonstrate that this more general approach could be of great interest to the computational organic chemistry community.

The final work was focused on improving the challenging electronic description of π -dimer radical cations that are used as prototypical model systems for organic electronics charge carriers. The accurate description of these complexes relies upon achieving a subtle balance between delocalization error and vdW interactions that makes them too challenging for DFTB3 and for most of the existing density functional approximations. (DFAs). In this thesis, an exchange-correlation functional, based on the ω B97X family, was optimized where the core parameters of the functional are trained together with the parameters of the dDsC density-dependent dispersion correction, bringing the total number of parameters to 16. ω B97X-dDsC exploits the long-range exchange correction to reproduce the asymptotic behavior

associated with the dissociation profile of the radical dimers. In addition, the near equilibrium region is improved thanks to an optimal balance between the electronic structure description provided by the functional and the dispersion contributions fitted jointly. The impressive improvement of both the asymptotic and the equilibrium regions in comparison to other density functionals is demonstrated through computing the binding energies and energy profiles of the characteristics Orel26rad data set and the analogue neutral Pi29n data set designed by our group.

Overall, this thesis increases the computational chemistry toolbox to rationalize, design and describe molecular materials. The presented approaches are original and useful especially to the modeling of organic semiconductors, which involve molecular chains made of thiophene units.

Despite all the work that has been completed, additional improvements would further extend the applicability of our toolbox and deepen our understanding of various computational and chemical aspects of this project. For example, dDMC should be coupled to a hydrogen bond correction, similar to the H4 from Řezáč *et al.*, which would facilitate its use in more general applications. Likewise, an upcoming implementation of our ω B97X-dDsC functional could include higher-order C8 and C10 dispersion terms which should lead to enhanced robustness for both standard and challenging systems.

REMD@DFTB3 certainly represents an exciting opportunity to revisit standard computational chemistry problems and explore novel perspectives within the quantum chemistry world. However, expertise that accompanies extensive application needs to be gained and a more generalized model that includes implicit solvent effects should be developed. The completion of Hamiltonian-REM and DFTB3 will also open the possibility to run the REMD@DFTB3 in a QM/MM environment and account for explicit solvent effects that are closer to the “reality”. Considering that H-REM allows the decoupling of the solute and the solvent in the replica exchange machinery, the extra cost of the solvent will remain low (see Section 2.3.2). The REMD techniques within the i-PI machinery should also be coupled to other semi-empirical and ab initio methods as implemented in standard quantum chemistry codes. The fast GPU implementation of KS-DFT in Terachem and the HF-3C method of Grimme are first on the list. These methods could provide reliable benchmark results on smaller systems that would serve to validate less accurate methods, such as DFTB. From a practical aspect, the influence of the missing entropy contributions in traditional static computations should be investigated deeply and systematically. In the context of organocatalysis, we have already revealed that the static picture is misleading and that the conformational entropic contribution must be considered. This aspect, which is generally ignored, could be a major influence in pushing quantum chemistry into its next stage. We envision that running a generalized version of

Chapter 7. General Conclusions and Outlook

REMD@DFTB3 on a set of reactants could help in discovering new reaction mechanisms. Within this context, we feel that the field of computational catalysis should be an ideal target.

A Density Functional Theory in a nutshell

The time independent Schrödinger equation (A.1) is the basis for most modern methods aimed at resolving the electronic structure of molecular systems.

$$\hat{\mathcal{H}}_{el}\Psi_{el}(\mathbf{r};\mathbf{R}) = E_{el}\Psi_{el}(\mathbf{r};\mathbf{R}) \quad (\text{A.1})$$

The Hamiltonian operator ($\hat{\mathcal{H}}$) is the quantum-wise sum of the kinetic and potential energies while Ψ_{el} is the many-body wavefunction that depends on the coordinates of all electrons and on the nuclei positions as parameters (in the Born-Oppenheimer approximation). Ψ_{el} describes the actual state of the electronic structure, which in principle contains, all the information regarding the system. Analytic solutions for this equation are possible only for prototypical examples and the computational cost associated with solving the equation scales exponentially with the number of electrons. As a result, accurate numerical solutions are only possible for very small systems.

The first approximated approach to resolve the equation came from Hartree and Fock, who assumed that the electrons are completely uncorrelated. Thus, each electron sees the others as a continuum charge density and a single Slater determinant is used to represent the electron system. An entire class of methods, post-HartreeFock, are aimed at enhancing the Hartree-Fock (HF) solution with different approaches. Yet, they are computationally not feasible for many chemically relevant systems. In 1964, Hohenberg and Kohn published the two foundation theorems that led to a new way of considering electronic structure computations.⁴⁶ The Hohenberg and Kohn work was followed, in 1965, by an operative solution proposed by Kohn and Sham.⁴⁷

A.1 The Hohenberg-Kohn theorems

The formulation of Density Functional Theory proposed by Hohenberg and Kohn⁴⁶ shows that all the properties of many interacting particles systems can be expressed as a functional of the ground state density. In particular, if dealing with electrons and nuclei relying on fixed positions, the Hamiltonian is

$$\hat{\mathcal{H}}_{el} = \hat{T}_e + \hat{V}_{ee} + \hat{V}_{ext}(\mathbf{r}) \quad (\text{A.2})$$

where \hat{T}_e is the kinetic energy of the electrons, \hat{V}_{ee} is the coulomb interactions between electrons and $\hat{V}_{ext}(\mathbf{r})$ is the potential energy due to the Coulombic interactions between electrons and nuclei.

The first theorem states:

For any system of interacting particles, the external potential $V_{ext}(\mathbf{r})$, except for a constant, is a unique functional of the non-degenerate ground state density.

In other words, the external potential uniquely defines, through the Hamiltonian, the ground state density of the system. This implies that all the properties can be written as functional of the ground state density.

The second theorem states that the energy is a functional of the ground state density, which built the operational basis for the future work of Kohn and Sham:

For any external potential $V_{ext}(\mathbf{r})$ can be defined a universal functional of the ground state density for the energy $E[\rho(\mathbf{r})]$:

$$E[\rho(\mathbf{r})] = T[\rho(\mathbf{r})] + V_{ee}[\rho(\mathbf{r})] + \int d\mathbf{r} V_{ext}(\mathbf{r})\rho(\mathbf{r}) \quad (\text{A.3})$$

The minimum of $E[\rho(\mathbf{r})]$ corresponds to the expectation value of the Hamiltonian in the ground state and the density that minimize the functional is the exact ground state density:

$$E_{min}[\rho(\mathbf{r})] = \langle \Psi_0 | \hat{\mathcal{H}} | \Psi_0 \rangle = E[\rho_0(\mathbf{r})] \quad (\text{A.4})$$

This second theorem shows that the energy as well as any other properties, can be written as a functional of the density, without needing the many-body wavefunction. Unfortunately, at this point, the shape of the functional $E[\rho(\mathbf{r})]$ is still unknown.

A.2. Exchange-correlation functionals

The solution, proposed by Kohn and Sham,⁴⁷ was to substitute the complicated many-body system with an auxiliary non-interacting system with the same density as the original one. The non-interacting particles are described by the one-electron Kohn-Sham (KS) orbitals ϕ_i^{KS} :

$$\left[-\frac{1}{2}\nabla^2 + V_{eff}^{KS}(\mathbf{r}) \right] \phi_i^{KS}(\mathbf{r}) = \epsilon_i \phi_i^{KS}(\mathbf{r}) \quad (\text{A.5})$$

where the KS effective potential is

$$V_{eff}^{KS}[\rho(\mathbf{r})] = V_{ext}(\mathbf{r}) + V_H[\rho(\mathbf{r})] + V_{xc}[\rho(\mathbf{r})] \quad (\text{A.6})$$

in which the Hartree potential (V_H) and the exchange-correlation potential (V_{xc}) are added to the external potential V_{ext} . The non-interacting wavefunction is a single Slater determinant of N occupied KS orbitals. Thanks to the previous stated theorems it is known that the total energy is an unique functional of the density

$$E[\rho(\mathbf{r})] = T^{KS}[\rho(\mathbf{r})] + E_H[\rho(\mathbf{r})] + E_{xc}[\rho(\mathbf{r})] + \int d\mathbf{r} V_{ext}\rho(\mathbf{r}) \quad (\text{A.7})$$

Introducing “pseudo-orbitals” allows for an easy way to compute the kinetic energy

$$T^{KS}[\rho(\mathbf{r})] = \sum_i^N \epsilon_i - \int d\mathbf{r} V_{eff}\rho(\mathbf{r}) \quad (\text{A.8})$$

The sole missing term is the $E_{xc}[\rho(\mathbf{r})]$ for what the exact expression is unknown. In practice, this term contains all the differences between the fictitious system of non-interacting electrons and the actual system.

A.2 Exchange-correlation functionals

To make the DFT approach operative it is necessary to define an exchange-correlation functional that approximates the unknown exact one.

One of the first attempt approximated the exchange-correlation energy as a local functional of the density, considering the density uniform in space. This **Local Density Approximation**^{52,53} (LDA) is the only case where the energy can be calculated exactly for a given density. The natural expansion of the LDA consists in considering also a dependence from the local density gradient. The **Generalized Gradient Approximation**³⁵⁹ (GGA) refers to a collection of formulations. This class of functionals contains some of the most popular examples, including BLYP (Becke’s formulation of the exchange³⁶⁰ and Lee-Yang-Parr correlation functional³⁶¹) and the Perdew-Burke-Ernzerhof (PBE) functional.³⁹

Appendix A. Density Functional Theory

GGA and LDA functionals do not consider the long-range contribution to the exchange energy. This failure provokes the so-called self-interaction error, that is completely missing in the non-local Hartree-Fock exchange. The introduction of a small portion (often around 25%) of exact Hartree-Fock exchange gives rise to the **Global Hybrid** (GH) functionals that cure, at least qualitatively, the erroneous results caused by self-interaction. B3LYP^{155,156} is undoubtedly the most famous functional in this class. Moreover, GH functionals largely improve the thermochemical predictions with respect to the predecessor GGAs, this feat prompted the spread of DFT within the chemical sciences as a fast and reliable method for computing many important molecular properties.

As previously mentioned in Section 2.2 there are functionals that cannot be easily placed into of the above mentioned categories. Aside from the self-interaction error, another important issue is the lack of description of weak interactions, for which only a few functionals are conceived to address (M06-2X, vdW-DF).^{57,58,65,66} The self-interaction error itself is not completely canceled by the GH functionals: recently Savin *et al.* introduced the so-called Long-Range corrected exchange functionals³³⁴ (see Chapter 6 for more details) that are able to produce accurate results even when the system presents strong charge-transfer character. In Chapter 6 we present the development of a new functional where the dispersion correction and the electronic structure parameters are fitted together in order to balance the overstabilization due to the self-interaction error and the contribution arising from the dispersion correction.

B Evaluation of the error in the Replica Exchange Molecular Dynamics

The statistical relevance of data analyzed in molecular dynamics need to be validated. In the case of Replica Exchange Molecular Dynamics (REMD), this evaluation is not straightforward owing to the re-weighting to cool the data obtained by the replica at the higher temperatures.

The proper way to assess the accuracy of free energies in the presented context is to run several simulations of the same system starting with different initial conditions. When all the simulations converge to the same results, we can be sure that the system reached the ergodicity. Since this would require an unaffordable time, one can decide to stop the simulations after a reasonable time and use the free energies obtained to create a distribution on top of which conducting statistical analysis. However this second method would be extremely expensive as well. The Block Averaging (BA) method allows to compute free energies distribution using a single trajectory which is divided into blocks of equal length each treated as a different short simulation. A distribution of free energies can be created from the many pieces obtained from partitioning the simulation. This method is very effective to check that a trajectory contains enough transitions among the accessible different states of the Potential Energy Surface. On the other hand, the BA would be not very effective with REMD for which each replica can exchange and the number of transition between different states would be artificially increased. *Bootstrapping* can be seen as a generalization of the BA: instead of creating subsets using consecutive points, the points are taken randomly from the entire trajectory.

We, thus, decided to provide the relative error obtained from bootstrapping the data in the target temperature replica, in order to avoid complications due to the weighting factors. In our analysis, we created 10^6 sets of 2000 points from a trajectory of 10^5 points. With this set-up the free energy average obtained from the bootstrapped sets has been found to converge toward the average values produced by the entire trajectory while the standard deviation obtained by the distribution of the bootstrapped data has been divided by the value of the respective free energy and reported in the Figure B.1 for the dithiacyclophane example of Chapter 5.

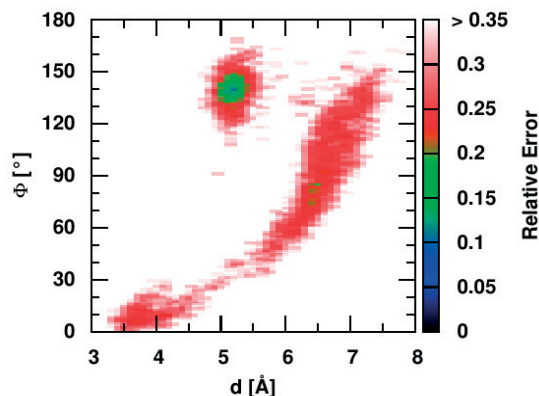


Figure B.1: Calculated relative error of the dithiacyclophane example of Chapter 5. The error is computed bootstrapping the lowest temperature trajectory. This is an overestimation of the true error since none of the warmer replicas has been taken into account. Nonetheless, the minima are described with a reasonable accuracy.

Despite the bootstrapping is performed on the data of a single replica, the minima are well defined (the error is often less than the 20%). We expect the error to be much lower after inclusion of all the replicas in the bootstrapping procedure. Note that transition states errors cannot be computed since they are visited only from higher temperature replicas.

Bibliography

- [1] Siegbahn, P. E.; Blomberg, M. R. *Annu. Rev. Phys. Chem.* **1999**, 50, 221.
- [2] Chermette, H. *Coord. Chem. Rev.* **1998**, 178-180, 699.
- [3] Neese, F. *Coord. Chem. Rev.* **2009**, 253, 526.
- [4] Hu, H.; Yang, W. *Annu. Rev. Phys. Chem.* **2008**, 59, 573.
- [5] Cramer, C. J.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2009**, 11, 10757.
- [6] Nørskov, J. K.; Bligaard, T.; Rossmeisl, J.; Christensen, C. H. *Nat. Chem.* **2009**, 1, 37.
- [7] Allard, S.; Forster, M.; Souharce, B.; Thiem, H.; Scherf, U. *Angew. Chemie - Int. Ed.* **2008**, 47, 4070.
- [8] Gunes, S.; Neugebauer, H.; Sariciftci, N. S. *Chem. Rev.* **2007**, 107, 1324.
- [9] Walzer, K.; Maennig, B.; Pfeiffer, M.; Leo, K. *Chem. Rev.* **2007**, 107, 1233.
- [10] Derivatives, H.; Louis, J.; Macdiarmid, A. G. *Polymer (Guildf)*. **1977**, 36, 578.
- [11] Joachim, C.; Gimzewski, J. K.; Aviram, A. *Nature* **2000**, 408, 541.
- [12] Kay, E. R.; Leigh, D. A.; Zerbetto, F. *Angew. Chemie - Int. Ed.* **2007**, 46, 72.
- [13] Kinbara, K.; Aida, T. *Chem. Rev.* **2005**, 105, 1377.
- [14] Cornil, J.; Beljonne, D.; Calbert, J. P.; Bredas, J.-L. *Adv. Mater.* **2001**, 13, 1053.
- [15] Borsenberger, P. M.; Fitzgerald, J. J. *J. Phys. Chem.* **1993**, 97, 4815.
- [16] Gonthier, J. F.; Steinmann, S. N.; Roch, L.; Ruggi, A.; Luisier, N.; Severin, K.; Corminboeuf, C. *Chem. Commun. (Camb)*. **2012**, 48, 9239.
- [17] Marty, R.; Szilluweit, R.; Sánchez-Ferrer, A.; Bolisetty, S.; Adamcik, J.; Mezzenga, R.; Spitzner, E.-C.; Feifer, M.; Steinmann, S. N.; Corminboeuf, C.; Frauenrath, H. *ACS Nano* **2013**, 7, 8498.

Bibliography

- [18] Liu, H.; Brémond, É.; Prlj, A.; Gonthier, J. E.; Corminboeuf, C. *J. Phys. Chem. Lett.* **2014**, *5*, 2320.
- [19] Nicolai, A.; Liu, H.; Petraglia, R.; Corminboeuf, C. *J. Phys. Chem. Lett.* **2015**, *6*, 4422.
- [20] Steinmann, S. N.; Corminboeuf, C. *J. Chem. Theory Comput.* **2012**, *8*, 4305.
- [21] Seifert, G.; Eschrig, H.; Bieger, W. *Zeitschrift Fur Physikalische Chemie-Leipzig* **1986**, 267, 529.
- [22] Elstner, M.; Porezag, D.; Jungnickel, G.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260.
- [23] Yang, Y.; Yu, H.; York, D.; Cui, Q.; Elstner, M. *J. Phys. Chem. A* **2007**, *111*, 10861.
- [24] Gaus, M.; Cui, Q.; Elstner, M. *J. Chem. Theory Comput.* **2011**, *7*, 931.
- [25] Gaus, M.; Lu, X.; Elstner, M.; Cui, Q. *J. Chem. Theory Comput.* **2014**, *10*, 1518.
- [26] Kristyán, S.; Pulay, P. *Chem. Phys. Lett.* **1994**, 229, 175.
- [27] Swendsen, R. H.; Wang, J.-S. *Phys. Rev. Lett.* **1986**, *57*, 2607.
- [28] Marinari, E.; Parisi, G.; Fisica, D.; Roma, U.; Vergata, T.; Roma Tor Vergata - Roma, S. *Eur. Lett. Eur. Lett* **1992**, *19*, 451.
- [29] Geyer, C. J. Markov Chain Monte Carlo Maximum Likelihood. 1991.
- [30] Petraglia, R.; Corminboeuf, C. *J. Chem. Theory Comput.* **2013**, *9*, 3020.
- [31] Steinmann, S. N.; Corminboeuf, C. *J. Chem. Theory Comput.* **2011**, *7*, 3567.
- [32] Mulliken, R. *J. Chem. Phys.* **1955**, *23*, 1833.
- [33] Petraglia, R.; Nicolai, A.; Wodrich, M. D.; Ceriotti, M.; Corminboeuf, C. *J. Comput. Chem.* **2016**, *37*, 83.
- [34] Chai, J.-D.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615.
- [35] Seifert, G. *J. Phys. Chem. A* **2007**, *111*, 5609.
- [36] Foulkes, W. M. C.; Haydock, R. *Phys. Rev. B* **1989**, *39*, 12520.
- [37] Zhukovskii, Y. F.; Piskunov, S.; Bellucci, S. *Nanosci. Nanotechnol. Lett.* **2012**, *4*, 1074.
- [38] Barone, V.; Carnimeo, I.; Scalmani, G. *J. Chem. Theory Comput.* **2013**, *9*, 2052.
- [39] Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.

-
- [40] Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1997**, *78*, 1396.
- [41] Slater, J. C.; Koster, G. F. *Phys. Rev.* **1954**, *94*, 1498.
- [42] Elstner, M. *Theor. Chem. Acc.* **2006**, *116*, 316.
- [43] Elstner, M. *J. Phys. Chem. A* **2007**, *111*, 5614.
- [44] Gaus, M.; Goez, A.; Elstner, M. *J. Chem. Theory Comput.* **2013**, *9*, 338.
- [45] Szabo, A.; Ostlund, N. S. *Dover Books Chem. Ser.*; 1996; p 480.
- [46] Hohenberg, P.; W., K. *Phys. Rev.* **1964**, *136*, B864.
- [47] Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133.
- [48] Bartlett, R. J.; Lotrich, V. F.; Schweigert, I. V. *J. Chem. Phys.* **2005**, *123*, 62205.
- [49] Bartlett, R. J.; Grabowski, I.; Hirata, S.; Ivanov, S. *J. Chem. Phys.* **2005**, *122*, 34104.
- [50] Grabowski, I.; Teale, A. M.; Śmiga, S.; Bartlett, R. J. *J. Chem. Phys.* **2011**, *135*, 114111.
- [51] Tao, J.; Perdew, J. P.; Staroverov, V. N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.
- [52] Dirac, P. a. M. *Math. Proc. Cambridge Philos. Soc.* **1930**, *26*, 376.
- [53] Slater, J. C. *Phys. Rev.* **1951**, *81*, 385.
- [54] Baerends, E. J.; Gritsenko, O. V. *J. Phys. Chem. A* **1997**, *101*, 5383.
- [55] Vydrov, O. A.; Van Voorhis, T. *Phys. Rev. Lett.* **2009**, *103*, 063004.
- [56] Vydrov, O. A.; Van Voorhis, T. *J. Chem. Phys.* **2010**, *133*, 244103.
- [57] Lee, K.; Murray, É. D.; Kong, L.; Lundqvist, B. I.; Langreth, D. C. *Phys. Rev. B* **2010**, *82*, 081101.
- [58] Dion, M.; Rydberg, H.; Schröder, E.; Langreth, D. C.; Lundqvist, B. I. *Phys. Rev. Lett.* **2004**, *92*, 246401.
- [59] Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 2715.
- [60] Grimme, S. *J. Chem. Phys.* **2006**, *124*.
- [61] Xu, X.; Goddard, W. A. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 2673.
- [62] Becke, A. D. *J. Chem. Phys.* **1997**, *107*, 8554.
- [63] Boese, A. D.; Handy, N. C. *J. Chem. Phys.* **2002**, *116*, 9559.

Bibliography

- [64] Boese, A. D.; Martin, J. M. L. *J. Chem. Phys.* **2004**, *121*, 3405.
- [65] Zhao, Y.; Schultz, N. E.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 364.
- [66] Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2007**, *120*, 215.
- [67] Wu, Q.; Yang, W. *J. Chem. Phys.* **2002**, *116*, 515.
- [68] Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463.
- [69] Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.
- [70] Tkatchenko, A.; Scheffler, M. *Phys. Rev. Lett.* **2009**, *102*, 073005.
- [71] von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. *Phys. Rev. Lett.* **2004**, *93*, 153004.
- [72] von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. *Phys. Rev. B* **2005**, *71*, 195119.
- [73] Lin, I.-C.; Coutinho-Neto, M. D.; Felsenheimer, C.; von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U. *Phys. Rev. B* **2007**, *75*, 205131.
- [74] Torres, E.; DiLabio, G. A. *J. Phys. Chem. Lett.* **2012**, *3*, 1738.
- [75] Mackie, I. D.; DiLabio, G. A. *J. Phys. Chem. A* **2008**, *112*, 10968.
- [76] Troullier, N.; Martins, J. L. *Phys. Rev. B* **1991**, *43*, 1993.
- [77] Goedecker, S.; Teter, M.; Hutter, J. *Phys. Rev. B* **1996**, *54*, 1703.
- [78] Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299.
- [79] Stevens, W. J.; Krauss, M.; Basch, H.; Jasien, P. G. *Can. J. Chem.* **1992**, *70*, 612.
- [80] Schwabe, T.; Grimme, S. *Phys. Chem. Chem. Phys.* **2007**, *9*, 3397.
- [81] Kozuch, S.; Martin, J. M. L. *Phys. Chem. Chem. Phys.* **2011**, *13*, 20104.
- [82] Goerigk, L.; Grimme, S. *J. Chem. Theory Comput.* **2011**, *7*, 291.
- [83] Román-Pérez, G.; Soler, J. M. *Phys. Rev. Lett.* **2009**, *103*, 096102.
- [84] Vydrov, O. A.; Van Voorhis, T. *J. Chem. Phys.* **2010**, *132*, 164113.
- [85] Sato, T.; Nakai, H. *J. Chem. Phys.* **2009**, *131*, 224104.
- [86] Sato, T.; Nakai, H. *J. Chem. Phys.* **2010**, *133*, 194101.

-
- [87] Ahlrichs, R.; Penco, R.; Scoles, G. *Chem. Phys.* **1977**, *19*, 119.
- [88] Conway, A.; Murrell, J. *Mol. Phys.* **1974**, *27*, 873.
- [89] Wagner, A. F. *J. Chem. Phys.* **1974**, *60*, 1885.
- [90] Hepburn, J.; Scoles, G.; Penco, R. *Chem. Phys. Lett.* **1975**, *36*, 451.
- [91] Wu, X.; Vargas, M. C.; Nayak, S.; Lotrich, V.; Scoles, G. *J. Chem. Phys.* **2001**, *114*, 8748.
- [92] Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. *J. Chem. Phys.* **2001**, *114*, 5149.
- [93] Meijer, E. J.; Sprik, M. *J. Chem. Phys.* **1996**, *105*, 8684.
- [94] Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787.
- [95] Jurecka, P.; Cerný, J.; Hobza, P.; Salahub, D. R. *J. Comput. Chem.* **2007**, *28*, 555.
- [96] Tang, K. T.; Toennies, J. P. *J. Chem. Phys.* **1984**, *80*, 3726.
- [97] Tang, K. T.; Toennies, J. P.; Yiu, C. L. *Phys. Rev. Lett.* **1995**, *74*, 1546.
- [98] Petraglia, R.; Steinmann, S. N.; Corminboeuf, C. *Int. J. Quantum Chem.* **2015**, *115*, 1265.
- [99] Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2005**, *123*, 154101.
- [100] Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2007**, *127*, 124108.
- [101] Johnson, E. R.; Becke, A. D. *J. Chem. Phys.* **2005**, *123*, 024101.
- [102] Johnson, E. R.; Becke, A. D. *J. Chem. Phys.* **2006**, *124*, 174104.
- [103] Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2006**, *124*, 014104.
- [104] Johnson, E. R.; Becke, A. D. *J. Chem. Phys.* **2005**, *123*, 24101.
- [105] Bultinck, P.; Cooper, D. L.; Ponec, R. *J. Phys. Chem. A* **2010**, *114*, 8754.
- [106] Steinmann, S. N.; Corminboeuf, C. *J. Chem. Phys.* **2011**, *134*, 044117.
- [107] Steinmann, S. N.; Corminboeuf, C. *J. Chem. Theory Comput.* **2010**, *6*, 1990.
- [108] Steinmann, S. N.; Csonka, G.; Corminboeuf, C. *J. Chem. Theory Comput.* **2009**, *5*, 2950.
- [109] Lin, Z. *Acc. Chem. Res.* **2010**, *43*, 602.
- [110] Cheng, G.-J.; Zhang, X.; Chung, L. W.; Xu, L.; Wu, Y.-D. *J. Am. Chem. Soc.* **2015**, *137*, 1706.

Bibliography

- [111] Pérez-Estrada, S.; Rodríguez-Molina, B.; Xiao, L.; Santillan, R.; Jiménez-Osés, G.; Houk, K. N.; Garcia-Garibay, M. A. *J. Am. Chem. Soc.* **2015**, *137*, 2175.
- [112] Hukushima, K.; Nemoto, K. *J. Phys. Soc. Japan* **1996**, *65*, 1604.
- [113] Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141.
- [114] Hansmann, U. H. *Chem. Phys. Lett.* **1997**, *281*, 140.
- [115] Marsili, S.; Signorini, G. F.; Chelli, R.; Marchi, M.; Procacci, P. *J. Comput. Chem.* **2010**, *31*, 1106.
- [116] Earl, D. J.; Deem, M. W. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3910.
- [117] Fukunishi, H.; Watanabe, O.; Takada, S. *J. Chem. Phys.* **2002**, *116*, 9058.
- [118] Sindhikara, D.; Meng, Y.; Roitberg, A. E. *J. Chem. Phys.* **2008**, *128*, 024103.
- [119] Abrams, C.; Bussi, G. *Entropy* **2014**, *16*, 163.
- [120] Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 13749.
- [121] Affentranger, R.; Tavernelli, I.; Di Iorio, E. E. *J. Chem. Theory Comput.* **2006**, *2*, 217.
- [122] Fajner, M.; Hamelberg, D.; McCammon, J. A. *J. Chem. Theory Comput.* **2008**, *4*, 1565.
- [123] Zacharias, M. *J. Chem. Theory Comput.* **2008**, *4*, 477.
- [124] Meng, Y.; Roitberg, A. E. *J. Chem. Theory Comput.* **2010**, *6*, 1401.
- [125] Wang, L.; Friesner, R. A.; Berne, B. J. *J. Phys. Chem. B* **2011**, *115*, 9431.
- [126] Zuckerman, D. M.; Lyman, E. *J. Chem. Theory Comput.* **2006**, *2*, 1200.
- [127] Facchetti, A. *Chem. Mater.* **2011**, *23*, 733.
- [128] Vondrak, T.; Wang, H.; Winget, P.; Cramer, C. J.; Zhu, X. Y. *J. Am. Chem. Soc.* **2000**, *122*, 4700.
- [129] Shaw, J. M.; Seidler, P. F. *IBM J. Res. Dev.* **2001**, *45*, 3.
- [130] Saito, G. *Phosphorus. Sulfur. Silicon Relat. Elem.* **1992**, *67*, 345.
- [131] Kelley, T. W.; Baude, P. F.; Gerlach, C.; Ender, D. E.; Muyres, D.; Haase, M. A.; Vogel, D. E.; Theiss, S. D. *Chem. Mater.* **2004**, *16*, 4413.
- [132] Hamed, M.; Forchheimer, R.; Inganäs, O. *Nat. Mater.* **2007**, *6*, 357.
- [133] Murgich, J.; Abanero, J. A.; Strausz, O. P. *Energy & Fuels* **1999**, *13*, 278.

-
- [134] Tauer, T. P.; Derrick, M. E.; Sherrill, C. D. *J. Phys. Chem. A* **2005**, *109*, 191.
- [135] Reid, K. S. C.; Lindley, P. F.; Thornton, J. M. *FEBS Lett.* **1985**, *190*, 209.
- [136] Morgan, R. S.; Tatsch, C. E.; Gushard, R. H.; McAdon, J.; Warme, P. K. *Int. J. Pept. Protein Res.* **1978**, *11*, 209.
- [137] Zauhar, R. J.; Colbert, C. L.; Morgan, R. S.; Welsh, W. J. *Biopolymers* **2000**, *53*, 233.
- [138] Morgan, R. S.; McAdon, J. M. *Int. J. Pept. Protein Res.* **1980**, *15*, 177.
- [139] Bredas, J.-L.; Silbey, R.; Boudreaux, D. S.; Chance, R. R. *J. Am. Chem. Soc.* **1983**, *105*, 6555.
- [140] Newman, C. R.; Frisbie, C. D.; da Silva Filho, D. A.; Brédas, J.-L.; Ewbank, P. C.; Mann, K. R.; da Silva, D. A.; Bredas, J.-L. *Chem. Mater.* **2004**, *16*, 4436.
- [141] Grimm, B.; Risko, C.; Azoulay, J. D.; Brédas, J.-L.; Bazan, G. C. *Chem. Sci.* **2013**, *4*, 1807.
- [142] Aradi, B.; Hourahine, B.; Frauenheim, T. *J. Phys. Chem. A* **2007**, *111*, 5678.
- [143] Korth, M.; Thiel, W. *J. Chem. Theory Comput.* **2011**, *7*, 2929.
- [144] Zheng, G.; Irle, S.; Morokuma, K. *Chem. Phys. Lett.* **2005**, *412*, 210.
- [145] Elstner, M.; Frauenheim, T.; Kaxiras, E.; Seifert, G.; Suhai, S. *Phys. status solidi* **2000**, *217*, 357.
- [146] Řezáč, J.; Riley, K. E.; Hobza, P. *J. Chem. Theory Comput.* **2011**, *7*, 2427.
- [147] Zhechkov, L.; Heine, T.; Patchkovskii, S.; Seifert, G.; Duarte, H. A. *J. Chem. Theory Comput.* **2005**, *1*, 841.
- [148] Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. *J. Phys. Chem. B* **2001**, *105*, 569.
- [149] McNamara, J. P.; Hillier, I. H. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2362.
- [150] Page, A. J.; Ohta, Y.; Okamoto, Y.; Irle, S.; Morokuma, K. *J. Phys. Chem. C* **2009**, *113*, 20198.
- [151] Welke, K.; Watanabe, H. C.; Wolter, T.; Gaus, M.; Elstner, M. *Phys. Chem. Chem. Phys.* **2013**, *15*, 6651.
- [152] Niehaus, T. A.; Suhai, S.; Della Sala, F.; Lugli, P.; Elstner, M.; Seifert, G.; Frauenheim, T. *Phys. Rev. B* **2001**, *63*, 85108.
- [153] Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.
- [154] Rappoport, D.; Furche, F. *J. Phys. Chem.* **2010**, *133*, 134105.

Bibliography

- [155] Becke, A. D. *J. Chem. Phys.* **1993**, 98, 5648.
- [156] Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, 98, 11623.
- [157] Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, 56, 2257.
- [158] Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, 28, 213.
- [159] Hariharan, P. C.; Pople, J. A. *Mol. Phys.* **1974**, 27, 209.
- [160] Gordon, M. S. *Chem. Phys. Lett.* **1980**, 76, 163.
- [161] Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. *J. Chem. Phys.* **1982**, 77, 3654.
- [162] Binning, R. C.; Curtiss, L. A. *J. Comput. Chem.* **1990**, 11, 1206.
- [163] Blaudeau, J.; McGrath, M. P.; Curtiss, L. A.; Radom, L. *J. Chem. Phys.* **1997**, 107, 5016.
- [164] Ditchfield, R. *J. Chem. Phys.* **1971**, 54, 724.
- [165] Ruzsinszky, A.; Perdew, J. P.; Csonka, G. I. *J. Phys. Chem. A* **2005**, 109, 11015.
- [166] Patton, D. C.; Pederson, M. R. *Phys. Rev. A* **1997**, 56, R2495.
- [167] Tao, J.; Perdew, J. P. *J. Chem. Phys.* **2005**, 122, 114102.
- [168] Zhang, Y.; Pan, W.; Yang, W. *J. Chem. Phys.* **1997**, 107, 7921.
- [169] Liu, H.; Kang, S.; Lee, J. Y. *J. Phys. Chem. B* **2011**, 115, 5113.
- [170] Steinmann, S. N.; Piemontesi, C.; Delachat, A.; Corminboeuf, C. *J. Chem. Theory Comput.* **2012**, 8, 1629.
- [171] Rochat, S.; Steinmann, S. N.; Corminboeuf, C.; Severin, K. *Chem. Commun.* **2011**, 47, 10584.
- [172] Luisier, N.; Ruggi, A.; Steinmann, S. N. *Org. Biomol. Chem.* **2012**, 10, 7487.
- [173] Bondar, A. N.; Fischer, S.; Smith, J. C.; Elstner, M.; Suhai, S. *J. Am. Chem. Soc.* **2004**, 126, 14668.
- [174] Gaus, M.; Chou, C. P.; Witek, H.; Elstner, M. *J. Phys. Chem. A* **2009**, 113, 11866.
- [175] Kaminski, S.; Giese, T. J.; Gaus, M.; York, D. M.; Elstner, M. *J. Phys. Chem. A* **2012**, 116, 9131.

-
- [176] Klimeš, J.; Michaelides, A. *J. Chem. Phys.* **2012**, *137*, 120901.
- [177] Grimme, S. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 211.
- [178] Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2005**, *122*, 154104.
- [179] Distasio, R. A.; von Lilienfeld, O. A.; Tkatchenko, A. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 14791.
- [180] Al-Saidi, W. A.; Voora, V. K.; Jordan, K. D. *J. Chem. Theory Comput.* **2012**, *8*, 1503.
- [181] Steinmann, S. N.; Corminboeuf, C. *Chimia (Aarau)*. **2010**, *65*, 240.
- [182] Steinmann, S. N.; Wodrich, M. D.; Corminboeuf, C. *Theor. Chem. Acc.* **2010**, *127*, 429.
- [183] Hirshfeld, F. L. *Theor. Chim. Acta* **1977**, *44*, 129.
- [184] Seifert, G.; Porezag, D.; Frauenheim, T. *Int. J. Quant. Chem.* **1996**, *58*, 185.
- [185] Kruger, T.; Elstner, M.; Schiffels, P.; Frauenheim, T. *J. Phys. Chem.* **2005**, *122*, 114110.
- [186] Brandenburg, J. G.; Grimme, S. *J. Phys. Chem. Lett.* **2014**, *5*, 1785.
- [187] Gonthier, J. F.; Steinmann, S. N.; Wodrich, M. D.; Corminboeuf, C. *Chem. Soc. Rev.* **2012**, *41*, 4671.
- [188] Liu, Y.; Goddard III, W. A. *Mater. Trans.* **2009**, *50*, 1664.
- [189] Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441.
- [190] Rowland, R. S.; Taylor, R. *J. Phys. Chem.* **1996**, *100*, 7384.
- [191] Haynes, W. *Handb. Chem. Phys.*; 2014; Vol. 54; p 2704.
- [192] Goyal, P.; Elstner, M.; Cui, Q. *J. Phys. Chem. B* **2011**, *115*, 6790.
- [193] Goyal, P.; Qian, H.-J.; Irle, S.; Lu, X.; Roston, D.; Mori, T.; Elstner, M.; Cui, Q. *J. Phys. Chem. B* **2014**, *118*, 11007.
- [194] Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243.
- [195] Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. *J. Chem. Phys.* **1997**, *106*, 9639.
- [196] Boys, S.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
- [197] Stewart, J. J. P. *J. Mol. Model.* **2007**, *13*, 1173.
- [198] Bahn, S.; Jacobsen, K. *Comput. Sci. Eng.* **2002**, *4*, 56.

Bibliography

- [199] Dolgonos, G.; Aradi, B.; Moreira, N. H.; Frauenheim, T. *J. Chem. Theory Comput.* **2010**, *6*, 266.
- [200] Barnett, R. N.; Landman, U. *Phys. Rev. B* **1993**, *48*, 2081.
- [201] Verlet, L. *Phys. Rev.* **1967**, *159*, 98.
- [202] Shao, Y. et al. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172.
- [203] Chandler, D.; Percus, J. K. *Phys. Today* **1988**, *41*, 114.
- [204] Brémond, É.; Golubev, N.; Steinmann, S. N.; Corminboeuf, C. *J. Chem. Phys.* **2014**, *140*, 18A516.
- [205] Mehlig, B.; Heermann, D.; Forrest, B. *Mol. Phys.* **1992**, *76*, 1347.
- [206] Adamo, C.; Barone, V. *J. Chem. Phys.* **1999**, *110*, 6158.
- [207] Weigend, F.; Häser, M. *Theor. Chem. Accounts Theory, Comput. Model. (Theoretica Chim. Acta)* **1997**, *97*, 331.
- [208] Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829.
- [209] Bahmanyar, S.; Houk, K. N. *J. Am. Chem. Soc.* **2001**, *123*, 11273.
- [210] Hoang, L.; Bahmanyar, S.; Houk, K. N.; List, B. *J. Am. Chem. Soc.* **2003**, *125*, 16.
- [211] Roy, D.; Najafian, K.; von Ragué Schleyer, P. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 17272.
- [212] Boren, B. C.; Narayan, S.; Rasmussen, L. K.; Zhang, L.; Zhao, H.; Lin, Z.; Jia, G.; Fokin, V. V. *J. Am. Chem. Soc.* **2008**, *130*, 8923.
- [213] Mackie, I. D.; Johnson, R. P. *J. Org. Chem.* **2009**, *74*, 499.
- [214] Zhuo, L.-G.; Zhang, J.-J.; Yu, Z.-X. *J. Org. Chem.* **2012**, *77*, 8527.
- [215] Himo, F.; Lovell, T.; Hilgraf, R.; Rostovtsev, V. V.; Noodleman, L.; Sharpless, K. B.; Fokin, V. V. *J. Am. Chem. Soc.* **2005**, *127*, 210.
- [216] Fristrup, P.; Ahlquist, M.; Tanner, D.; Norrby, P.-O. *J. Phys. Chem. A* **2008**, *112*, 12862.
- [217] Frei, R.; Wodrich, M. D.; Hari, D. P.; Borin, P.-A.; Chauvier, C.; Waser, J. *J. Am. Chem. Soc.* **2014**, *136*, 16563.
- [218] Duarte, F.; Gronert, S.; Kamerlin, S. C. L. *J. Org. Chem.* **2014**, *79*, 1280.
- [219] Li, Y.; Lin, S.-T.; Goddard, W. A. *J. Am. Chem. Soc.* **2004**, *126*, 1872.

- [220] Kaczor, A.; Reva, I. D.; Proniewicz, L. M.; Fausto, R. *J. Phys. Chem. A* **2006**, *110*, 2360.
- [221] Carpenter, B. K. *Annu. Rev. Phys. Chem.* **2005**, *56*, 57.
- [222] Ess, D. H.; Wheeler, S. E.; Iafe, R. G.; Xu, L.; Çelebi-Ölçüm, N.; Houk, K. N.; Celebi-Ölçüm, N.; Houk, K. N. *Angew. Chemie Int. Ed.* **2008**, *47*, 7592.
- [223] Schreiner, P. R.; Reisenauer, H. P.; IV, F. C. P.; Simmonett, A. C.; Allen, W. D.; Mátyus, E.; Császár, A. G. *Nature* **2008**, *453*, 906.
- [224] Xu, L.; Doubleday, C. E.; Houk, K. N. *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 2746.
- [225] Rehbein, J.; Carpenter, B. K. *Phys. Chem. Chem. Phys.* **2011**, *13*, 20906.
- [226] Schreiner, P. R.; Reisenauer, H. P.; Ley, D.; Gerbig, D.; Wu, C.-H.; Allen, W. D. *Science* **2011**, *332*, 1300.
- [227] Black, K.; Liu, P.; Xu, L.; Doubleday, C.; Houk, K. N. *Proc. Natl. Acad. Sci.* **2012**, *109*, 12860.
- [228] Ley, D.; Gerbig, D.; Schreiner, P. R. *Org. Biomol. Chem.* **2012**, *10*, 3781.
- [229] Rossi, M.; Scheffler, M.; Blum, V. *J. Phys. Chem. B* **2013**, *117*, 5574.
- [230] Sandbeck, D. J. S.; Kuntz, C. M.; Luu, C.; Mondor, R. A.; Ottaviano, J. G.; Rayer, A. V.; Sumon, K. Z.; East, A. L. *J. Phys. Chem. A* **2014**, *118*, 11768.
- [231] de Souza, M. A. F.; Ventura, E.; do Monte, S. A.; Riveros, J. M.; Longo, R. L. *Chem. - A Eur. J.* **2014**, *20*, 13742.
- [232] Jiménez-Osés, G.; Liu, P.; Matute, R. A.; Houk, K. N. *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 8664.
- [233] Karmakar, S.; Datta, A. *Angew. Chem. Int. Ed. Engl.* **2014**, *53*, 9587.
- [234] Samanta, D.; Rana, A.; Schmitt, M. *J. Org. Chem.* **2015**, *80*, 2174.
- [235] Goethe, M.; Fita, I.; Rubi, J. M. *J. Chem. Theory Comput.* **2015**, *11*, 351.
- [236] Törk, L.; Jiménez-Osés, G.; Doubleday, C.; Liu, F.; Houk, K. N. *J. Am. Chem. Soc.* **2015**, *137*, 4749.
- [237] Plata, R. E.; Singleton, D. A. *J. Am. Chem. Soc.* **2015**, *137*, 150225164014000.
- [238] Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646.
- [239] Adcock, S. A.; McCammon, J. A. *Chem. Rev.* **2006**, *106*, 1589.

Bibliography

- [240] van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. *J. Phys. Chem. A* **2001**, *105*, 9396.
- [241] Freddolino, P. L.; Harrison, C. B.; Liu, Y.; Schulten, K. *Nat Phys* **2010**, *6*, 751.
- [242] Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471.
- [243] Barnett, R. N.; Landman, U.; Nitzan, A.; Rajagopal, G. *J. Chem. Phys.* **1991**, *94*, 608.
- [244] Wentzcovitch, R. M.; Martins, J. *Solid State Commun.* **1991**, *78*, 831.
- [245] Hytha, M.; Štich, I.; Gale, J. D.; Terakura, K.; Payne, M. C. *Chem. - A Eur. J.* **2001**, *7*, 2521.
- [246] Ivanov, I.; Klein, M. L. *J. Am. Chem. Soc.* **2005**, *127*, 4010.
- [247] Bucko, T.; Hafner, J. *J. Phys. Condens. Matter* **2010**, *22*, 384201.
- [248] Martoňák, R.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2003**, *90*, 075503.
- [249] Martoňák, R.; Donadio, D.; Oganov, A. R.; Parrinello, M. *Nat. Mater.* **2006**, *5*, 623.
- [250] Fois, E.; Fois, E.; Gamba, A.; Gamba, A.; Spano, E.; Spano, E. *J. Phys. Chem. B* **2004**, *108*, 2.
- [251] Trinh, T. T.; Jansen, A. P. J.; van Santen, R. a.; Jan Meijer, E. *Phys. Chem. Chem. Phys.* **2009**, *11*, 5092.
- [252] Kühne, T. D.; Krack, M.; Mohamed, F. R.; Parrinello, M. *Phys. Rev. Lett.* **2007**, *98*.
- [253] Lippert, G.; Hutter, J.; Parrinello, M. *Theor. Chem. Acc.* **1999**, *103*, 124.
- [254] Lippert, G.; Parrinello, M.; Hutter, J. **1997**, *92*, 477.
- [255] Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M. *J. Chem. Theory Comput.* **2005**, *1*, 1176.
- [256] Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M. *J. Chem. Theory Comput.* **2006**, *2*, 1370.
- [257] Todorova, T.; Seitsonen, A. P.; Hutter, J.; Kuo, I.-F. W.; Mundy, C. J. *J. Phys. Chem. B* **2006**, *110*, 3685.
- [258] Wu, R.; Hu, P.; Wang, S.; Cao, Z.; Zhang, Y. *J. Chem. Theory Comput.* **2010**, *6*, 337.
- [259] Ufimtsev, I. S.; Martinez, T. J. *J. Chem. Theory Comput.* **2009**, *5*, 2619.
- [260] Porezag, D.; Frauenheim, T.; Köhler, T.; Seifert, G.; Kaschner, R. *Phys. Rev. B* **1995**, *51*, 12947.
- [261] Cui, Q.; Elstner, M. *Phys. Chem. Chem. Phys.* **2014**, *16*, 14368.

- [262] Dellago, C.; Bolhuis, P. G.; Csajka, F. S.; Chandler, D. *J. Chem. Phys.* **1998**, *108*, 1964.
- [263] Ciccotti, G.; Ferrario, M.; Hynes, J. T.; Kapral, R. *Chem. Phys.* **1989**, *129*, 241.
- [264] Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562.
- [265] Ciccotti, G.; Ferrario, M. *Mol. Simul.* **2004**, *30*, 787.
- [266] Barducci, A.; Bonomi, M.; Parrinello, M. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 826.
- [267] Hamelberg, D.; Mongan, J.; McCammon, J. A. *J. Chem. Phys.* **2004**, *120*, 11919.
- [268] Bernardi, R. C.; Melo, M. C. R.; Schulten, K. *Biochim. Biophys. Acta - Gen. Subj.* **2015**, *1850*, 872.
- [269] Ensing, B.; De Vivo, M.; Liu, Z.; Moore, P.; Klein, M. L. *Acc. Chem. Res.* **2006**, *39*, 73.
- [270] Bucher, D.; Pierce, L. C. T.; McCammon, J. A.; Markwick, P. R. L. *J. Chem. Theory Comput.* **2011**, *7*, 890.
- [271] Ishikawa, Y.; Sugita, Y.; Nishikawa, T.; Okamoto, Y. *Chem. Phys. Lett.* **2001**, *333*, 199.
- [272] Choi, T. H. *Chem. Phys. Lett.* **2012**, *543*, 45.
- [273] Fedorov, D. G.; Sugita, Y.; Choi, C. H. *J. Phys. Chem. B* **2013**, *117*, 7996.
- [274] Ceriotti, M.; More, J.; Manolopoulos, D. E. *Comput. Phys. Commun.* **2014**, *185*, 1019.
- [275] Sindhikara, D. J.; Emerson, D. J.; Roitberg, A. E. *J. Chem. Theory Comput.* **2010**, *6*, 2804.
- [276] Ceriotti, M.; Brain, G. A. R.; Riordan, O.; Manolopoulos, D. E. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **2011**, *468*, 2.
- [277] Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157.
- [278] Frisch, M. J. et al. Gaussian09 Revision E.01.
- [279] Steinmann, S. N.; Corminboeuf, C.; Wu, W.; Mo, Y. *J. Phys. Chem. A* **2011**, *115*, 5467.
- [280] Fonseca Guerra, C.; Snijders, J. G.; te Velde, G.; Baerends, E. J. *Theor. Chem. Accounts Theory, Comput. Model. (Theoretica Chim. Acta)* **1998**, *99*, 391.
- [281] Werner, H.-J.; Manby, F. R.; Knowles, P. J.; Werner, H.-J. *J. Chem. Phys.* **2013**, *8149*, 8149.
- [282] Werner, H.-J.; Knowles, P. J.; Knizia, G.; Manby, F. R.; Schütz, M. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *2*, 242.

Bibliography

- [283] Zhou, C.; Birney, D. M. *Org. Lett.* **2002**, 4, 3279.
- [284] Zhang, X.; Hrovat, D. A.; Borden, W. T. *Org. Lett.* **2010**, 12, 2798.
- [285] Dewar, M. J. S.; Lo, D. H. *J. Am. Chem. Soc.* **1971**, 93, 7201.
- [286] Hoffmann, R.; Stohrer, W. D. *J. Am. Chem. Soc.* **1971**, 93, 6941.
- [287] Williams, R. V.; Kurtz, H. A. *J. Org. Chem.* **1988**, 53, 3626.
- [288] Williams, R. V.; Gadgil, V. R.; Chauhan, K.; Jackman, L. M.; Fernandes, E. *J. Org. Chem.* **1998**, 63, 3302.
- [289] Williams, R. V. *Chem. Rev.* **2001**, 101, 1185.
- [290] Williams, R. *European J. Org. Chem.* **2001**, 2001, 227.
- [291] Brown, E. C.; Henze, D. K.; Borden, W. T. *J. Am. Chem. Soc.* **2002**, 124, 14977.
- [292] Goren, A. C.; Hrovat, D. A.; Seefelder, M.; Quast, H.; Borden, W. T. *J. Am. Chem. Soc.* **2002**, 124, 3469.
- [293] Hrovat, D. A.; Brown, E. C.; Williams, R. V.; Quast, H.; Borden, W. T. *J. Org. Chem.* **2005**, 70, 2627.
- [294] Greve, D. R. *J. Phys. Org. Chem.* **2010**, 24, 222.
- [295] Griffiths, P. R.; Pivonka, D. E.; Williams, R. V. *Chem. - A Eur. J.* **2011**, 17, 9193.
- [296] Ichikawa, Y.; Sakai, S. *J. Phys. Org. Chem.* **2011**, 25, 409.
- [297] Jana, D. F.; Wodrich, M. D.; Corminboeuf, C. *J. Org. Chem.* **2012**, 77, 2548.
- [298] Moran, A.; Hamilton, A.; Bo, C.; Melchiorre, P. *J. Am. Chem. Soc.* **2013**, 135, 9091.
- [299] Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2006**, 128, 13435.
- [300] J.S. Dewar, M.; Jie, C. *Tetrahedron* **1988**, 44, 351.
- [301] Miller, L. S.; Grohmann, K.; Dannenberg, J. J. *J. Am. Chem. Soc.* **1983**, 105, 6862.
- [302] Pople, J. A. *J. Chem. Phys.* **1956**, 24, 1111.
- [303] Dijkstra, G. D. H.; Kellogg, R. M.; Wynberg, H.; Svendsen, J. S.; Marko, I.; Sharpless, K. B. *J. Am. Chem. Soc.* **1989**, 111, 8069.
- [304] Zhou, J.; Wakchaure, V.; Kraft, P.; List, B. *Angew. Chem. Int. Ed. Engl.* **2008**, 47, 7656.

-
- [305] Prakash, G. K. S.; Wang, F.; Ni, C.; Shen, J.; Haiges, R.; Yudin, A. K.; Mathew, T.; Olah, G. A. *J. Am. Chem. Soc.* **2011**, *133*, 9992.
- [306] Bürgi, T.; Baiker, A. *J. Am. Chem. Soc.* **1998**, *120*, 12920.
- [307] Lam, Y.-H.; Houk, K. N. *J. Am. Chem. Soc.* **2015**, *137*, 2116.
- [308] Lundberg, M.; Nishimoto, Y.; Irle, S. *Int. J. Quantum Chem.* **2012**, *112*, 1701.
- [309] Perdew, J. P. *AIP Conf. Proc.* **2001**, *577*, 1.
- [310] Cohen, A. J.; Mori-Sánchez, P.; Yang, W. **2012**, *112*, 289.
- [311] Perdew, J. P.; Parr, R. G.; Levy, M.; Balduz, J. L. *Phys. Rev. Lett.* **1982**, *49*, 1691.
- [312] Dutoi, A. D.; Head-Gordon, M. *Chem. Phys. Lett.* **2006**, *422*, 230.
- [313] Perdew, J. P.; Smith, J. R. *Surf. Sci.* **1984**, *141*, L295.
- [314] Jones, R. O.; Gunnarsson, O. *Rev. Mod. Phys.* **1989**, *61*, 689.
- [315] Milet, A.; Korona, T.; Moszynski, R.; Kochanski, E. *J. Chem. Phys.* **1999**, *111*, 7727.
- [316] Allen, M. J.; Tozer, D. J. *J. Chem. Phys.* **2002**, *117*, 11113.
- [317] Dobson, J. F.; Wang, J.; Dinte, B. P.; McLennan, K.; Le, H. M. *Int. J. Quantum Chem.* **2005**, *101*, 579.
- [318] Swart, M. *J. Chem. Theory Comput.* **2008**, *4*, 2057.
- [319] Reiher, M.; Salomon, O.; Sellmann, D.; Hess, B. A. *Chemistry* **2001**, *7*, 5195.
- [320] Ess, D. H.; Johnson, E. R.; Hu, X.; Yang, W. *J. Phys. Chem. A* **2011**, *115*, 76.
- [321] Cramer, C. J.; Smith, B. A. *J. Phys. Chem.* **1996**, *100*, 9664.
- [322] Kurlancheek, W.; Lochan, R.; Lawler, K.; Head-Gordon, M. *J. Chem. Phys.* **2012**, *136*, 054113.
- [323] Almladh, C.-O.; von Barth, U. *Phys. Rev. B* **1985**, *31*, 3231.
- [324] van Leeuwen, R.; Baerends, E. J. *Phys. Rev. A* **1994**, *49*, 2421.
- [325] Cohen, A. J.; Mori-Sánchez, P.; Yang, W. *Science* **2008**, *321*, 792.
- [326] Mori-Sánchez, P.; Cohen, A. J.; Yang, W. *Phys. Rev. Lett.* **2008**, *100*, 146401.
- [327] Kim, M.-C.; Sim, E.; Burke, K. *J. Chem. Phys.* **2011**, *134*, 171103.

Bibliography

- [328] Jensen, F. J. *Chem. Theory Comput.* **2010**, 6, 2726.
- [329] Heaton-Burgess, T.; Yang, W. J. *Chem. Phys.* **2010**, 132, 234113.
- [330] Ruiz, E.; Salahub, D. R.; Vela, A. J. *Am. Chem. Soc.* **1995**, 117, 1141.
- [331] Ruiz, E.; Salahub, D. R.; Vela, A. J. *Phys. Chem.* **1996**, 100, 12265.
- [332] Tsuzuki, S.; Honda, K.; Azumi, R. J. *Am. Chem. Soc.* **2002**, 124, 12200.
- [333] Cohen, A. J.; Mori-Sánchez, P.; Yang, W. J. *Chem. Phys.* **2007**, 126, 191109.
- [334] Savin, A.; Flad, H.-J. *Int. J. Quantum Chem.* **1995**, 56, 327.
- [335] Yanai, T.; Tew, D. P.; Handy, N. C. *Chem. Phys. Lett.* **2004**, 393, 51.
- [336] Henderson, T. M.; Janesko, B. G.; Scuseria, G. E. *J. Chem. Phys.* **2008**, 128, 194105.
- [337] Rohrdanz, M. A.; Martins, K. M.; Herbert, J. M. *J. Chem. Phys.* **2009**, 130, 054112.
- [338] Livshits, E.; Baer, R. *Phys. Chem. Chem. Phys.* **2007**, 9, 2932.
- [339] Livshits, E.; Baer, R. *J. Phys. Chem. A* **2008**, 112, 12789.
- [340] Stein, T.; Kronik, L.; Baer, R. *J. Am. Chem. Soc.* **2009**, 131, 2818.
- [341] Lin, Y. S.; Li, G. D.; Mao, S. P.; Chai, J.-D. *J. Chem. Theory Comput.* **2013**, 9, 263.
- [342] Shao, Y. et al. *Mol. Phys.* **2015**, 113, 184.
- [343] Gordon, M. S.; Schmidt, M. W. *Theory Appl. Comput. Chem.* **2005**, 1167.
- [344] Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, 14, 1347.
- [345] te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Fonseca Guerra, C.; van Gisbergen, S. J. A.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* **2001**, 22, 931.
- [346] Hafner, J. *J. Comput. Chem.* **2008**, 29, 2044.
- [347] Fabrizio, A. Overcoming the drawbacks of electronic structure methods , in the theoretical investigation of organic electronic materials . Ph.D. thesis, EPFL, 2015.
- [348] Zhu, C.; Byrd, R. H.; Lu, P.; Nocedal, J. *ACM Trans. Math. Softw.* **1997**, 23, 550.
- [349] Byrd, R. H.; Lu, P.; Nocedal, J.; Zhu, C. *SIAM J. Sci. Comput.* **1995**, 16, 1190.
- [350] Millman, K. J.; Aivazis, M. *Comput. Sci. Eng.* **2011**, 13, 9.

- [351] Gill, P. M.; Johnson, B. G.; Pople, J. A. *Chem. Phys. Lett.* **1993**, 209, 506.
- [352] Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. *J. Chem. Phys.* **1997**, 106, 1063.
- [353] Goerigk, L.; Grimme, S. *Phys. Chem. Chem. Phys.* **2011**, 13, 6670.
- [354] Vydrov, O. A.; Scuseria, G. E. *J. Chem. Phys.* **2006**, 125.
- [355] Vydrov, O. A.; Heyd, J.; Krukau, A. V.; Scuseria, G. E. *J. Chem. Phys.* **2006**, 125.
- [356] Vydrov, O. A.; Scuseria, G. E.; Perdew, J. P. *J. Chem. Phys.* **2007**, 126.
- [357] Kroon-Batenburg, L.; van Duijneveldt, F. J. *Mol. Struct. THEOCHEM* **1985**, 121, 185.
- [358] Hobza, P.; Šponer, J. *Chem. Rev.* **1999**, 99, 3247.
- [359] Perdew, J. P.; Burke, K. *Int. J. Quantum Chem.* **1996**, 57, 309.
- [360] Becke, A. D. *Phys. Rev. A* **1988**, 38, 3098.
- [361] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, 37, 785.

Riccardo Petraglia

Curriculum Vitæ

École Polytechnique Fédérale de Lausanne
Institute of Chemical Science and Engineering
Laboratory for Computational Molecular Design

+41 (0)78 795 611

✉ riccardo.petraglia@gmail.com

32 years old. Italian. Married



Last Updated: May 25, 2016

Research Experience

2012 – **Graduate Research Assistant**, *Laboratory for Computational Molecular Design*, EPFL, Lausanne, Switzerland.

Main research topics:

- DFTB parameterization.
- Development of an atom pairwise dispersion correction for DFTB.
- Development of a novel method to study organic molecules and reactions by joining enhanced sampling MD approach to DFTB (REM@DFTB).
- Parametrization of a new XC-functional: ω -B97X-dDsC.
- In silico design of molecular organic electronics.

2011 **Master student in Chemistry**, *Università degli studi di Firenze*, Florence, Italy.

Main research topics:

- Parallel tempering enhance sampling method applications.
- Improving parametrization of force fields for particular saccharides.

2008 **Bachelor student in Chemistry**, *Università degli studi di Firenze*, Florence, Italy.

Main research topics:

- Developing and implementation of an enhanced sampling method acting on the atom's masses.
- Application of the developed method within the steered-MD approach.

Education

Expected June 2016 **Ph.D. in Computational Chemistry**, *Laboratory for Computational Molecular Design*, EPFL, Lausanne, Switzerland.

Thesis title: *Development of Accurate and Efficient Electronic Structure Approaches for the Description of Organic Systems.*

Director: Prof. Clemence Corminboeuf

Sept. 2011 **M.Sc. in Chemistry**, *Università degli studi di Firenze*, Italy, **110/110**.

Orientation: Computational and physical chemistry

Thesis title: Replica exchange simulations of heparin related mono and disaccharides

Directors: P. Procacci and S. Marsili

Feb. 2008 **B.Sc. in Chemistry**, *Università degli studi di Firenze*, Italy, **109/110**.

Thesis title: Atomic mass effect in non-equilibrium processes

Directors: P. Procacci and R. Chelli

Awards

- Best oral presentation Award@Swiss Chemical Society meetings 2015
- Teaching Award@EPFL 2014

Skills

Programming skills

Advanced: Python2/3, Fortran90, UNIX shell scripting, Git

Other: MPI, OpenMP, PERL, PHP, JavaScript, HTML4, C

Linux: Debian and Fedora, set-up of desktop and server stations.

Chemistry software

- Molecular Dynamics:
- **ORAC**: PT-REM, Hamiltonian-REM, Steered-MD.
 - **i-PI**: Biased-MD, PT-REM, implementation of a biased-PT approach.
 - **PLUMED2**: Biased-MD, implementation of the i-PI interface.
- Ab-initio:
- **GAMESS**: Implementation and parametrization of a new XC-functional: ω B97X-dDsC. Post-HF methods(CAS, CASPT, MCSCF, ORMAS), Transition state geometry optimization and reaction paths.
 - **QuantumEspresso**: Geometry optimization of molecular crystals.
 - **DFTB+**: Development of parameters for DFTB and an atom pairwise dispersion correction. Implementation of a socket interface to communicate with the i-PI MD-driver for the REM@DFTB3 approach.
 - **ADF**: Computation of transfer integrals in organic electronics related molecules.

Languages

Italian Native

English Business fluent

French Basic knowledge

Personal Experience

1992 – 2011 **Scout - AGESCI**, *Poppi*, Italy.

104 My Experience as scout has taught me to work and cooperate within a team.

Journal Publications

- Petraglia, R.; Nicolai, A.; Wodrich, M. D.; Ceriotti, M.; Corminboeuf, C. "Beyond static structures: Putting forth REMD as a tool to solve problems in computational organic chemistry." *J. Comput. Chem.* **Jan. 2016**, 37, 83–92.
- Gebers, J.; Hartmann, L.; Özen, B.; Schaer, M.; Suárez, S.; Bugnon, P.; Scopelliti, R.; Steinrück, H.-G.; Konovalov, O.; Magerl, A.; Brinkmann, M.; Petraglia, R.; de Silva, P.; Corminboeuf, C.; Frauenrath, H. "The Effect of Hydrogen Bonding for the Crystallization and OFET Performance of a Quaterthiophene Diacetamide" *Chemical Science*, submitted **2015**, Computation of transfer integral, optimization of structures, definition of couples within the crystals.
- Nicolai, A.; Liu, H.; Petraglia, R.; Corminboeuf, C. "Exploiting Dispersion-Driven Aggregators as a Route to New One-Dimensional Organic Nanowires." *EN J. Phys. Chem. Lett.* **Nov. 2015**, 6, 4422–8.
- Petraglia, R.; Steinmann, S. N.; Corminboeuf, C. "A fast charge-Dependent atom-pairwise dispersion correction for DFTB3" *International Journal of Quantum Chemistry* **2015**, 115, 1265–1272.
- Schrettl, S.; Stefaniu, C.; Schwieger, C.; Pasche, G.; Oveisi, E.; Fontana, Y.; Fontcuberta i Morral, A.; Reguera, J.; Petraglia, R.; Corminboeuf, C.; Brezesinski, G.; Frauenrath, H. "Functional carbon nanosheets prepared from hexayne amphiphile monolayers at room temperature." *Nature chemistry* **2014**, 6, 468–76, Crystal geometry optimization.
- Petraglia, R.; Corminboeuf, C. "A Caveat on SCC-DFTB and Noncovalent Interactions Involving Sulfur Atoms" *Journal of Chemical Theory and Computation* **2013**, 9, 3020–3025.

Scientific talks and posters

- 2015 **"Beyond Static Structures: Putting Forth REMD as a Tool to Solve Problems in Computational Organic Chemistry"**, *Swiss Chemical Society - Fall Meeting 2015*, EPFL, Lausanne, Switzerland.
Talks
- 2012 **"Density Dependent Dispersion Correction Scheme for Large Scale Systems"**, *Swiss Chemical Society - Fall Meeting 2012*, ETH, Zurich, Switzerland.
Poster
- 2013 **"A Caveat on SCC-DFTB and Non-Covalent Interactions Involving Sulfur Atoms"**, *7th Molecular Quantum Mechanics*, ETH, Lugano, Switzerland.
Poster

- 2013 **" ω B97X-dDsC: A density functional approximation designed to describe π -radical dimer cations"**, *15th International Conference on Density Functional Theory and its Applications*, Durham, United Kingdom.
Poster

