

ATOM-CENTERED POTENTIALS FOR DESCRIBING LONDON DISPERSION FORCES IN DENSITY FUNCTIONAL THEORY

THÈSE N° 4058 (2008)

PRÉSENTÉE LE 4 AVRIL 2008

À LA FACULTÉ DES SCIENCES DE BASE
LABORATOIRE DE CHIMIE ET BIOCHIMIE COMPUTATIONNELLES
PROGRAMME DOCTORAL EN CHIMIE ET GÉNIE CHIMIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

I-Chun LIN

Master of Natural Sciences, University of Cambridge, Royaume-Uni
et de nationalité chinoise

acceptée sur proposition du jury:

Prof. P. J. Dyson, président du jury
Prof. U. Röthlisberger, directrice de thèse
Prof. M. Dal Peraro, rapporteur
Prof. J. Hutter, rapporteur
Prof. M. Sprik, rapporteur



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2008

Acknowledgment

I would like to begin by expressing my gratitude to my supervisor, Ursula. I truly enjoy the freedom she has allowed in my research and have benefited substantially from the breadth and depth of her knowledge. Her wonderful support made attending conferences and summer schools possible, giving me opportunities to meet scientists in/outside the field I am working on.

I truly appreciate the guidance and fascinating insights on how things “work” or “don’t work” in academia that I have received from my post-docs/collaborators, Anatole, Ari, Ivano, Leo and Mauricio. Anatole is “double”-acknowledged for being the punch-bag and movie buddy whenever he is around.

I cannot thank Karin, our group secretary, enough for helping me through endless visa applications and trip organisations.

I am indebted to Profs Dal Peraro, Dyson, Hutter and Sprik for agreeing to be my PhD examiners. Without the help of Marc-Etienne and Pascal, my French résumé will be hardly understandable to anyone but me.

I will always think back fondly the time I have spent with the LCBC gang: Roberto and Thereza, Leo, Ute, Mauricio and Fabiola, Samuele, Marc-Etienne, Camille, Anatole, Denis, Michel, Christian, Patrick, Maria-Carola, Pascal, Enrico, Sam, Geoff, Marilisa, Ivano and Paola, Pablo, Fanny, Stefano and the two LCBC chefs, Michele and Matteo. Special credits go to Topaz, Ruby, Phaedra, Aries and Drexel the cat. It is most entertaining to share the office with Ivano. I am not sure how many times he tried to catch my attentions with no avail but thanks for the patience nonetheless, it is a tough job to share an office with an iPod-nano addict like me.

The acknowledgement will be incomplete without mentioning my friends dating back to the good-old school days: Amina, Arno, Dude, Ellen, Ezzeri, Gabor, Ingrid, JJ, Kamil, Katharina, Stephen, Tomek and my two tata: the very-hard-to-track-down tata Kerim and the most fabulous tata Wookie.

Thanks to everyone who shares his/her space so I could avoid the fate of sleeping under some bridge during my rather frequent get-away; the list is long, just like my gratitude: Amina, Anatole, Ari, Ezzeri, Ingrid and her family, Kamil, Katharina and her parents, Jan, Jochen, Pozzi, Samuele, Tomek and Wookie. I have to specially acknowledge Kamil for getting the city girl to embrace the US wild life, that itself is a true accomplishment, not to mention the long drive across the Golden State.

I am most grateful to my parents for their never-ending trust and love; the complete support they

show in every choices I made is greatly appreciated. I am lucky enough to have aunts and uncles who treat me as their own, and the most wonderful grandma in the world who spoils me to bits. Last but not least, thanks to my elder brother, my sister-in-law and my chatty little brother who make the trips to London always homey and relaxing.

Abstract

The Kohn–Sham formulation of density functional theory (DFT) has posed itself as one of the most popular and versatile methods for condensed phase studies owing to its reasonable accuracy and affordable computational cost. DFT, in principle, yields exact ground state energy, including dispersion forces that are of primordial importance in chemical and biological systems. Yet with many exchange–correlation functionals in practical use such as the local density approximation or generalized gradient approximations, DFT either provides sporadic results or fails completely to account for these forces. In consequence, various methods offering remedy for this shortcoming have been proposed in this active field of research. In particular, dispersion-corrected atom-centered potentials (DCACPs) serve as a robust and efficient way to include these weak forces in a fully self-consistent manner within current DFT frameworks.

The aim of this thesis is twofold: first, to improve the predictive power and the understanding of the DCACP concept; second, applying DCACPs to systems of increasing complexity starting with dimers, continuing through larger clusters and ending with the condensed phase. The success of the second aim not only justifies the use of DCACPs but more importantly, provides insights to the role dispersion forces play in the systems investigated.

We first draw on the atoms-in-molecules theory and a multi-center density expansion to justify the form and universality of DCACPs. A library of DCACPs calibrated with an improved penalty functional against high-level *ab initio* references is presented. With the library in hand, we extend our studies to systems of biological significance, mainly constituents of proteins and DNA; polycyclic aromatic molecules intercalated in between segments of DNA are the center of focus. The application of DCACPs is then furthered to the condensed phase and the importance of van der Waals interactions in liquid water is investigated.

Keywords: amino acids, density functional theory, dispersion-corrected atom-centered potentials, dispersion forces, interaction energy, liquid water, nucleic acids.

Résumé

La formulation de Kohn–Sham de la théorie de la fonctionnelle de la densité (DFT) s’est imposée comme l’une des méthodes les plus populaires et les plus souples pour les études en phase condensée, grâce à sa précision raisonnable pour un coût informatique abordable. La DFT, en principe, permet d’obtenir l’énergie exacte de l’état fondamental d’un système, y compris la contribution des forces de dispersion qui sont de première importance dans les systèmes chimiques et biologiques. Pourtant, avec de nombreuses fonctionnelles d’échange-corrélation couramment utilisées – telles l’approximation de la densité locale ou les approximations du gradient généralisé – la DFT fournit des résultats sporadiques, voire échoue complètement à reproduire ces forces. Par conséquent, diverses méthodes permettant de corriger cette imperfection ont été développées. Les potentiels centrés sur les atomes et corrigés pour la dispersion (DCACPs) constituent une approche prometteuse pour inclure ces forces faibles de façon cohérente dans le cadre de la DFT.

Le but de cette thèse est double: premièrement, l’amélioration de la performance prédictive et de la compréhension du concept de DCACP; deuxièmement, l’application des DCACPs à des systèmes de complexité croissante: des dimères, de plus grands clusters, et finalement la phase condensée. La seconde partie justifie l’utilisation des DCACPs, mais aussi, plus important, elle fournit des informations sur le rôle des forces de dispersion dans les systèmes étudiés.

Nous nous basons sur la théorie de la mécanique moléculaire des atomes et sur un développement multi centré de la densité pour justifier la forme et l’universalité des DCACPs. Une librairie de DCACPs calibrés avec une fonctionnelle de pénalité améliorée et des valeurs de référence calculées *ab initio* à niveau élevé a d’abord été développée. A l’aide de cette librairie, nous avons élargi notre étude à des systèmes d’importance biologique, principalement aux constituants des protéines et aux acides désoxyribonucléique (ADN) pour lesquels; nous nous sommes focalisés sur l’intercalation de molécules aromatiques polycycliques entre des segments d’ADN. L’application des DCACPs a ensuite été étendue à la phase condensée, avec l’étude de l’importance des interactions de van der Waals dans l’eau liquide.

Mots-clé: acides aminés, acides nucléiques, eau liquide, énergie d’interaction, forces de dispersion, théorie de la fonctionnelle de la densité, potentiels centrés sur les atomes et corrigés pour la dispersion.

Contents

Acknowledgment	i
Abstract	iii
Résumé	iv
List of Abbreviations	ix
List of Tables	x
List of Figures	xii
1 Introduction	1
2 Theory	7
2.1 Density Functional Theory	8
2.2 <i>Ab Initio</i> Molecular Dynamics	13
2.3 London Dispersion Forces	15
2.4 Dispersion-Corrected Atom-Centered Potentials	18
3 Multi-Center Density Functionals	21
4 Library of Dispersion-Corrected Atom-Centered Potentials	31
4.1 Introduction	32
4.2 Computational Details	33
4.3 Results and Discussion	33

4.4	Conclusions	43
5	Noncovalent Interactions in Nucleobase–Intercalator Complexes	45
5.1	Introduction	46
5.2	Computational Details	49
5.3	Results and Discussion	52
5.4	Conclusions	60
6	Weak Interactions in Biomacromolecules	63
6.1	Introduction	64
6.2	Computational Details	65
6.3	Results and Discussion	66
6.4	Conclusions	71
7	Importance of van der Waals Interactions in Liquid Water	73
7.1	Introduction	74
7.2	Computational Details	76
7.3	Results and Discussion	77
7.4	Conclusions	83
8	Comparative Study of Dispersion-Corrected BLYP and PBE Water	85
8.1	Introduction	86
8.2	Computational Details	87
8.3	Results and Discussion	88
8.4	Conclusions	96
9	Conclusions and Outlook	97
	Bibliography	101
	Appendix	113
	Publications and Presentations	115

Curriculum Vitae

119

List of Abbreviations

A	adenine
ACP	atom-centered potential
AIMD	<i>ab initio</i> molecular dynamics
a.u.	atomic units
BLYP	Becke-Lee-Yang-Parr (exchange-correlation functional)
B3LYP	Becke's 3-parameter hybrid (exchange-correlation functional)
BP	Becke-Perdew (exchange-correlation functional)
BOMD	Born–Oppenheimer molecular dynamics
BSSE	basis set superposition error
C	cytosine
CBS	complete basis set
CCSD(T)	coupled-cluster with single and double and perturbative triple excitations
CI	configuration interaction
CPMD	Car-Parrinello molecular dynamics
DCACP	dispersion-corrected atom-centered potential
DFT	density functional theory
DNA	deoxyribonucleic acid
G	guanine
GGA	generalized gradient approximation
GTH	Goedecker-Teter-Hutter (pseudopotentials)
H-bond	hydrogen bond

HF	Hartree–Fock
HOMO	highest occupied molecular orbital
KS	Kohn–Sham
LDA	local density approximation
LUMO	lowest unoccupied molecular orbital
MAE	mean absolute error
MD	molecular dynamics
MP2	second order Møller–Plesset perturbation theory
MSD	mean square displacement
NMR	nuclear magnetic resonance
NVE	microcanonical ensemble
NVT	canonical ensemble
PBC	periodic boundary condition
PBE	Perdew-Burke-Ernzerhof (exchange-correlation functional)
RMSD	root mean square deviation
RNA	ribonucleic acid
Ry	Rydberg
SAPT	symmetry-adapted perturbation theory
T	thymine
TM	Troullier-Martins (pseudopotentials)
U	uracil
vdW	van der Waals
WC	Watson–Crick

List of Tables

4.1	Additional CPU cost incurred by using DCACPs at different ℓ channels.	34
4.2	Transferability of the carbon and hydrogen DCACPs calibrated in various schemes.	35
4.3	DCACP parameters for H, C, N, and O.	36
4.4	DCACP parameters for the rare-gas atoms.	36
4.5	Equilibrium interaction energies and distances of the rare-gas dimers.	39
4.6	Polarisabilities and multipole moments of formaldehyde, Ar, Ar-N ₂ , CO ₂ , N ₂ , and benzene.	39
4.7	Equilibrium distances and interaction energies of some weakly bound complexes.	42
4.8	Comparative study on DCACPs applied with different pseudopotential types.	43
5.1	Structure of ellipticine and its derivatives.	46
5.2	Interaction energies of heterocycle and benzene dimers.	52
5.3	Interaction energies of the stacked nucleobase pairs.	53
5.4	Overall dipole moment of the nucleobase in its optimized geometry and the geometry observed in the stacked complex.	54
5.5	Interaction energies and H-bond lengths of the WC base pairs	55
5.6	Equilibrium distances and interaction energies of the ellip–WC complexes.	56
5.7	Interaction energy of the fully relaxed ellip–WC complexes and the H-bonding energy of the WC base pair in the geometry-optimized ellip–WC complexes.	57
6.1	Interaction energies of the H-bonded base pairs.	67
6.2	Interaction energies of the interstrand base pairs.	68
6.3	Interaction energies of the stacked base pairs.	69

6.4	Interaction energies of the amino acid complexes.	70
7.1	Characteristics of the oxygen-oxygen radial distribution function of liquid water	77
7.2	Structural parameters and interaction energies of the water dimer and cyclic trimer.	82
8.1	Structural properties of liquid water.	89
8.2	Dynamical properties of liquid water.	92
8.3	Interaction energies of water clusters from dimer to hexamer.	93

List of Figures

3.1	Value of single-particle density at nucleus as a function of the atomic number.	26
3.2	Interaction energy of $(\text{H}_2)_2$ in parallel computed with DCACPs having multiple projectors.	28
4.1	Interaction energies of H_2 dimer in parallel and cross configurations evaluated with DCACPs at different ℓ channels.	34
4.2	The radial term of DCACPs.	37
4.3	Interaction energy curves of the parallel H_2 dimer.	38
4.4	Interaction energy curves of the sandwich benzene dimer.	38
4.5	Difference between the BLYP- and DCACP-predicted electron densities.	40
4.6	Interaction energy surface of the parallel-displaced benzene dimer.	41
5.1	Graphical representation of the the AT- E^+ and ellipticine- $\text{d}(\text{CG})_2$ complexes.	51
5.2	Detail of the optimized geometry of the G \cdots C stacked base pair.	54
5.3	Interaction energy profile of the intercalation process in the ellipticine- $\text{d}(\text{CG})_2$ complex.	58
5.4	Graphical representation of the orbitals in the ellipticine- $\text{d}(\text{CG})_2$ complex.	59
6.1	Superimposition of the DCACP-optimized stacked base pairs on the MP2-optimized geometries.	70
7.1	Oxygen-oxygen and oxygen-hydrogen radial distribution functions of liquid water.	78
7.2	Angular distributions of liquid water.	79
7.3	Mean square displacements and self-diffusion coefficients of liquid water.	80
7.4	Orientational autocorrelation functions of liquid water.	81
7.5	Interaction energy curves of two configurations of the water dimer at various O-O distance.	83

8.1	Graphical representation of the water dimer.	87
8.2	Variation of g_{OO}^{\max} observed during the 30 ps NVE liquid water simulation.	88
8.3	Oxygen-oxygen radial distribution functions of liquid water.	89
8.4	Angular distributions of liquid water.	91
8.5	Boltzmann-weighted difference between interaction energies of the water dimer calculated with DFT methods and CCSD(T) as a function of r_{OO} and \angle_{OHO}	94

Chapter 1

Introduction

Computer Simulations

The expanding role of computer simulations in basic sciences has been fueled by the steady progress in computer technology and in the development of numerical algorithm. The performance to price ratio has increased significantly over the years and shows no signs of abating. Moreover, the introduction of massive parallelization in computer architecture will certainly maintain this present trend. With this rapid advance in both hardware and software, simulating highly complex systems over a longer time scale with theories or models based on less assumptions and approximations has become feasible.

Predicting properties of intricate molecular systems via computer simulations – *in silico* experiments – is certainly not yet accurate enough to justify a total replacement of experimental measurements; instead, the strength of simulations lies in their ability to give insights into the planning and analysis of their conventional counterparts. At the most basic level, computer simulations provide a standard tool in determining the spatial structures obtained from experiments such as X-ray crystallography and neutron diffraction; electronic structure calculations can also be directly compared with a variety of spectroscopic data (Infrared, Raman, and NMR) without additional assumptions. Simulations, in addition, allow one to probe the relation between microscopic properties and macroscopic behavior. A microscopic model can be modified with a high degree of confidence, and the consequences on the macroscopic behavior of the molecular system can then be evaluated. The level of control that can be exerted *in silico* is nearly impossible to achieve in conventional experiments since modifications in the latter tend to have unintentional secondary effects that can be difficult to differentiate from the primary effect being studied. *In silico* experiments thus offer a clean and clear-cut way to test hypotheses regarding, for example, biological phenomena at the molecular level, including structures of biomolecules or mechanisms of enzyme catalysis. Furthermore, qualitative, sometimes even quantitative, theoretical estimates for quantities like binding constants of ligands to receptors can be obtained even when the production of the specific ligand is too costly or the measurement too time-consuming to carry out. Last but not least, simulations can be performed under unobservable or extreme conditions of temperature and pressure inaccessible by conventional experiments, freeing scientific investigation to what cannot be seen, touched, nor made.

The reliability of simulations depends on whether the relevant phase space is sufficiently sampled, the degree to which the microscopic system simulated reflects the typically macroscopic system in na-

ture, and the accuracy of the theory/model adopted. In order to enhance the accessible time scale, an appropriate model combined with a clever choice of rare-event sampling techniques is necessary. Classical force fields are able to access either very long runs (up to about one microsecond) for a medium number of particles ($< 10^4$) or shorter period of time for systems containing 10^6 particles or more; in coarse-grained methods, an atomistic description is replaced by a lower-resolution model that averages away fine details so even longer time- and length-scale dynamics become tractable. The limitation of length-scale can also be circumvented by multi-scale simulations [1–3] where the system is divided into various subsystems that are treated at different level of details and bridged by transition regions defined in a coherent fashion, *e.g.*, quantum mechanics/molecular mechanics [4–6]. Nevertheless, one needs to strike a fine balance between the level of accuracy targeted in different subsystems so the computing effort for the subsystem treated with a highly accurate method is not wasted because of some distorting effects from the cruder parts of the model. When the degrees of freedom are infinitely dense or sufficiently sampled, the accuracy of simulations will depend solely on the quality of the assumptions and approximations chosen. In order to establish a firm foundation for applying *in silico* experiments, the theoretical predictions and experimental data should be compared whenever possible, bearing in mind that a good agreement between calculated and experimental data can sometimes be due to fortuitous error cancellations.

Complementary to experiments, computer simulations have played an equal, and sometimes pivotal, role in quantitative characterizations and in advancing qualitative understandings. Although there are still problems not yet solved and will remain at the forefront for many years to come, the ongoing development and steady advance will only further stress their status as the central and basic methodological approach in the future.

Motivation

Well-established empirical force fields based on two-body interactions have made headway in computer simulations. Many of them, however, are parameterized against experimental data, calling into question their predictive power for situations differing largely from the reference system. Moreover, since many-body effects are represented in an effective way by modifying the two-body interaction terms, a direct physical interpretation of the simulated phenomena at the molecular level is sometimes not

straightforward, and the neglect of cooperative effects can also be problematic. In addition, most standard force fields employ fixed point charges and do not allow the flexibility for an explicit polarization of atoms. The most obvious limitation of empirical force fields lies in their inability to describe reactive events since they are unable to adapt to changes in chemical environment with their predefined fixed parameters.

The use of *ab initio* methods minimizes the risk of violating assumptions implicit in the parameterization of empirical or semi-empirical approaches; features difficult or impossible to describe with force fields such as electronic polarization effects or bond breaking/forming events can be treated self-consistently. Their application, however, is limited to only a few hundred nuclei, and the accessible time scale is in the (tens of) picosecond range.

Even though the available computer resources are usually one of the main considerations when choosing one particular model over the other, the dominant factor still hinges on the physical adequacy of the chosen model for the specific property one is interested in. For reactive events, a quantum mechanical treatment of electronic degrees of freedom with the nuclear coordinates as parameters is mandatory. In addition, a detailed atomistic investigation of a biological system, for example, often requires the knowledge of its electronic structure.

The Kohn–Sham (KS) formalism of density functional theory (DFT) [7, 8] has proven its popularity in solid state physics and has steadily expanded its stronghold to biological systems such as proteins and nucleic acids. It scales favorably with the system size compared with Hartree–Fock (HF) and high-level correlated *ab initio* methods, and it has the further advantage over HF for being able to treat electron correlation effects to a certain extent. From the implementation point of view, DFT is well suited for modern parallel computing and linear-scaling techniques [9].

DFT is, in principle, exact if the true expression for the exchange-correlation potential were known. Unfortunately, the exact form remains elusive and the many reported deficiencies of DFT stem solely from the approximated nature of the exchange-correlation potentials in practical use. Much effort has been devoted to find good approximations to achieve the goal of chemical accuracy, the accuracy needed to predict rates of chemical reactions (energy errors within 1 kcal/mol). DFT has been quite successful in describing a wide variety of strongly interacting systems, isolated molecules, or dense solid state systems. The local density approximation (LDA) treats largely homogeneous systems such as simple

metals and semiconductors with surprising accuracy; for inhomogeneous systems, semi-local density approximations such as members of the generalized gradient approximation (GGA) family work well for cohesion, bonds, structures and other properties. Nevertheless, systems such as soft matter and biomolecules are at least as abundant and they generally have inter-particle separations for which non-local long-range interactions (*e.g.*, dispersion forces) are influential.

A classical picture of dispersion forces entails the energy lowering associated with polarization by instantaneous fluctuations in the charge distributions of two inert or widely separated systems. These weak forces, which act between separated atoms/molecules even in the absence of charges or permanent electric moments, contribute significantly to phenomena such as solvation, physisorption, molecular recognition, as well as the stability and conformational variability of molecular crystals and biomacromolecules. Without taking dispersion forces into account, proteins may be predicted to be unstable and relative energies of various polypeptide conformations may be largely in error [10]. Even for systems in which stronger intermolecular forces (*e.g.*, direct electrostatic interactions) seem to dominate, these very weak forces may still play a pivotal role; for example, the relative energies of two phases in ionic materials are often sufficiently sensitive to dispersion interactions to have a substantial effect on the transition pressure [11].

Many popular density functionals are based on the local electron density (LDA), its gradient (GGA) and sometimes even the local kinetic energy density (meta-GGA). As dispersion interactions contribute even at distances where the electron overlap is negligible, these purely local functionals fail by construction to reproduce these interactions and are unable to describe correctly the leading R^{-6} dispersion interaction term which originates from correlated instantaneous dipole fluctuations. GGA functionals cover only the short-range exponentially decaying contribution [12, 13] of van der Waals (vdW) interactions. This, however, is described very differently by various GGAs in both their exchange and correlation parts [14]. Although the exchange part should be purely repulsive for rare-gas dimers, the magnitude of repulsion varies with the functional, which can be traced back to their different behaviors at the small density and high reduced-gradient region [15]. Some GGAs lead to an artificial attraction even at the exchange-only level whereas others overestimate Pauli repulsion; correlation effects due to the short range overlap term are also covered in varying degrees [13, 16]. Within some GGAs (PBE, for example), the $(\text{He})_2$ interaction energy tends to be severely overestimated, and much too short equi-

librium bond lengths are predicted. The well depths of the heavier rare-gas dimers, on the other hand, are severely underestimated by GGAs. BLYP [17, 18] yields only repulsive potential energy profiles for rare-gas dimers [12, 13, 15, 19–21]. In addition, many GGAs predict repulsive $\pi - \pi$ stacking [22–27], key interactions in nucleic acids as well as aromatic polymers, to name just a few.

One can thus conclude that the general performance of DFT in predicting dispersion forces is rather dismal. In consequence, if these forces are of any importance in the system of interest, they must be recovered in calculations to grant the results any credibility or significance. Methods for efficient calculation of dispersion forces have been the focus of many recent works [27–33]. From the practical point of view, any proposed scheme that is to be generally applicable to a wide range of chemically and biologically interesting systems needs to be system independent (transferable) and computationally tractable. In this work, we concentrate on one recently introduced approach – dispersion-corrected atom-centered potentials (DCACPs) [34]. The concept of DCACPs is to represent the effect of dispersion forces via atomic orbital-dependent potentials whose two adjustable parameters are obtained by calibrating against references of chosen accuracy, offering a robust way to include these forces in a fully self-consistent manner within current DFT frameworks.

Overview

The thesis is organized as follows: the theoretical background is briefly sketched out in Chapter 2. Chapter 3 draws on the atoms-in-molecules theory and a multi-center density expansion to justify the form and universality of DCACPs. Chapter 4 presents a library of DCACPs for rare-gas atoms and some of the most abundant elements in biological/chemical systems. Furthermore, transferability to systems with different geometrical orientations or chemical compositions from the calibration ones is investigated. In Chapters 5 and 6, we extend our studies to both neutral and charged complexes larger than simple dimers; systems of biological significance, including complexes of nucleobases, amino acids, and intercalator–nucleobase adducts, are the center of focus. The importance of vdW interactions in liquid water is addressed in Chapters 7 and 8. Finally, conclusions are drawn, and possible future extension to this work is discussed.

Chapter 2

Theory

2.1 Density Functional Theory

Density functional theory (DFT) presents a promising alternative to high-level correlated *ab initio* theories. It has posed itself as one of the most popular and versatile methods available for investigating the electronic structure of many-body systems, in particular large molecules and condensed phases.

The N -electron wave function $\Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ contains all the information we could ever have, but more than we usually want; often we are interested in no more than the total energy (as well as its changes) and the electron density $\rho(\mathbf{r})$,

$$\rho(\mathbf{r}_1) = N \int d\mathbf{x}_2 \dots d\mathbf{x}_N \Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) \Psi^*(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N).$$

DFT, on the other hand, allows one to replace $\Psi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ which depends on the coordinates of the particles ($3N$ variables, or $4N$ if the spin is taken into account) by the much simpler $\rho(\mathbf{r})$, a function of three spatial coordinates only (and, for spin polarized systems, the spin).

The fundamental of DFT is described only briefly here, applying it to the simplest case of non-relativistic interacting electrons in a closed shell ground-state system with integer occupation number for which the Born-Oppenheimer approximation is valid. For more details, one is referred to Refs. [35–38]. Atomic units are employed throughout unless otherwise specified.

The first Hohenberg–Kohn theorem [7] states that the external potential $v_{\text{ext}}(\mathbf{r})$ is uniquely determined by the ground-state density $\rho(\mathbf{r})$ within a trivial additive constant. Since $\rho(\mathbf{r})$ determines the number of electrons N , $N = \int d\mathbf{r} \rho(\mathbf{r})$, it follows that $\rho(\mathbf{r})$ also determines the full Hamiltonian and implicitly, all properties of the system. The second Hohenberg–Kohn theorem [7] provides the energy variational principle. It states that for a trial density $\rho(\mathbf{r})$ in which $\rho(\mathbf{r}) \geq 0$ and $\int \rho(\mathbf{r}) d\mathbf{r} = N$, we have $E[\rho] \geq E_0$ (E_0 is the true ground-state energy).

$$\begin{aligned} E[\rho] &= T[\rho] + V_{\text{ee}}[\rho] + V_{\text{ne}}[\rho] \\ &= F_{\text{HK}}[\rho] + \int d\mathbf{r} v_{\text{ext}}(\mathbf{r}) \rho(\mathbf{r}), \\ F_{\text{HK}}[\rho] &= T[\rho] + V_{\text{ee}}[\rho]. \end{aligned}$$

$T[\rho]$, $V_{\text{ne}}[\rho]$, and $V_{\text{ee}}[\rho]$ stand for the kinetic energy, the electron-nucleus attraction energy, and the

electron-electron repulsion energy, respectively. $F_{\text{HK}}[\rho]$ is independent of the system, thus known as the universal functional.

Kohn and Sham proposed to replace the kinetic energy of the interacting electrons with that of an equivalent non-interacting system because the latter can be easily (and accurately) evaluated [8]. In the Kohn–Sham (KS) formalism, $F[\rho]$ thus consists of the classical Coulomb repulsion $J[\rho]$ and the kinetic energy $T_s[\rho]$ in terms of the non-interacting KS orbitals $\{\psi_i\}$ whose corresponding density equals the density of the real target system of interacting electrons,

$$\begin{aligned} F[\rho] &= T_s[\rho] + J[\rho] + E_{\text{xc}}[\rho], \\ T_s[\rho] &= -\frac{1}{2} \sum_i^N \langle \psi_i | \nabla^2 | \psi_i \rangle, \quad J[\rho] = \frac{1}{2} \int d\mathbf{r} d\mathbf{r}' \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}, \quad \rho(\mathbf{r}) = \sum_i^N |\psi_i(\mathbf{r})|^2. \end{aligned}$$

The exchange-correlation term $E_{\text{xc}}[\rho]$ contains all the remaining, yet undefined, contributions to the energy. In other words, $E_{\text{xc}}[\rho]$ not only accounts for the non-classical part of electron-electron interactions, but it also includes the difference in kinetic energy between the fictitious non-interacting and the real interacting systems.

In analogy to the HF scheme, the KS orbitals must minimize the energy under the constraint $\langle \psi_i | \psi_j \rangle = \delta_{ij}$. This leads to the KS equation,

$$\left(-\frac{1}{2} \nabla^2 + v(\mathbf{r}) + \frac{\delta J}{\delta \rho} + \frac{\delta E_{\text{xc}}}{\delta \rho} \right) \psi_i = \epsilon_i \psi_i,$$

in which

$$\frac{\delta J}{\delta \rho} = \int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}, \quad \frac{\delta E_{\text{xc}}}{\delta \rho} = v_{\text{xc}}.$$

Apart from the kinetic energy term, all other terms in the bracket constitute the KS effective potential, *i.e.*, $v_{\text{eff}} = v(\mathbf{r}) + \int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + v_{\text{xc}}(\mathbf{r})$.

KS-DFT is a formally rigorous way of approaching any interacting problem by mapping it to a less complicated non-interacting problem. It looks (and, more importantly, scales with the system size) like a mean-field theory with a self-consistent effective one-electron Schrödinger equation for the KS orbitals, but includes, in principle, all correlation effects on the ground-state electron density and the total energy. If the true expression for v_{xc} were known, solving the KS equations would be the equivalent of solving

the exact Schrödinger equation within the Born–Oppenheimer approximation. Unfortunately, the exact v_{xc} remains elusive and much effort has been devoted to find good approximations.

2.1.1 Exchange-correlation functionals

As mentioned previously, DFT, in principle, is able to provide the exact ground-state energy, the approximation only enters when one is to choose an explicit form for the unknown functionals of E_{xc} and the corresponding v_{xc} . The quality of the density functional approach thus hinges solely on the accuracy of the chosen approximation. The local density approximation (LDA), generalized gradient approximations (GGAs) and the hybrid GGA functionals are sketched out in brief here.

LDA. In the LDA, E_{xc}^{LDA} and v_{xc}^{LDA} are defined as

$$E_{xc}^{\text{LDA}} = \int \rho(\mathbf{r}) \varepsilon_{xc}^{\text{LDA}}(\rho) d\mathbf{r},$$

$$v_{xc}^{\text{LDA}}(\mathbf{r}) = \frac{\delta E_{xc}^{\text{LDA}}}{\delta \rho(\mathbf{r})} = \varepsilon_{xc}^{\text{LDA}}(\rho) + \rho(\mathbf{r}) \frac{\partial \varepsilon_{xc}^{\text{LDA}}(\rho)}{\partial \rho}.$$

$\varepsilon_{xc}(\rho)$ is the exchange and correlation energy density of a uniform electron gas of density ρ . $\varepsilon_{xc}(\rho)$ can be divided into exchange and correlation contributions: $\varepsilon_{xc}^{\text{LDA}}(\rho) = \varepsilon_x^{\text{LDA}}(\rho) + \varepsilon_c^{\text{LDA}}(\rho)$. Accurate values of $\varepsilon_c(\rho)$ from quantum Monte Carlo calculations [39] have been interpolated to provide an analytic form [40]. The exchange part is given by the Dirac exchange-energy functional [41],

$$\varepsilon_x^{\text{LDA}}(\rho) = -C_x \rho(\mathbf{r})^{1/3}, \quad C_x = \frac{3}{4} \left(\frac{3}{\pi} \right)^{1/3}.$$

The LDA is applicable to systems with slowly varying density, but it cannot be formally justified for highly inhomogeneous systems such as atoms and molecules. It predicts structural properties in the solid state with surprising accuracy considering its crudeness; however, the only moderate accuracy it delivers for bond energies and other molecular properties is insufficient for most applications in chemistry.

GGA. GGAs are the first step to go beyond the LDA by incorporating explicit density gradient dependence into E_{xc} and they have been quite successful in repairing the over-binding character of the LDA.

$$E_{xc}^{\text{GGA}} = \int \rho(\mathbf{r}) \varepsilon_{xc}^{\text{GGA}}(\rho, \nabla \rho) d\mathbf{r}.$$

As for the LDA, the contributions from exchange and correlation are normally split, and approximations are sought individually, $E_{xc}^{\text{GGA}} = E_x^{\text{GGA}} + E_c^{\text{GGA}}$. E_x^{GGA} is then re-cast in a form that involves $s(\mathbf{r})$, the reduced density gradient:

$$E_x^{\text{GGA}} = E_x^{\text{LDA}} - \int F(s)\rho^{4/3}(\mathbf{r})d\mathbf{r},$$

$$s(\mathbf{r}) = \frac{|\nabla\rho(\mathbf{r})|}{\rho^{4/3}(\mathbf{r})}.$$

Becke proposed F to be of the following form (abbreviated as B),

$$F^{\text{B}} = \frac{\beta s^2}{1 + 6\beta s \sinh^{-1} s},$$

where β is an empirical parameter determined to be 0.0042 by a least-square fit to the exact exchange energies of the rare-gas atoms (helium through radon) [17]. Among the most commonly used correlation functionals in chemistry are functionals due to Lee, Yang, Parr (LYP) [18] and Perdew (P) [42]. The former is derived from an expression for the correlation energy of the helium atom based on an accurate correlated wave function by Colle and Salvetti [43]; the latter is based on the uniform electron gas and employs an empirical parameter fitted to the correlation energy of the neon atom. In principle, each exchange functional can be combined with any of the correlation functionals, but only a few combinations are in popular use. The exchange part is almost exclusively chosen to be Becke's functional which is either combined with P or LYP – abbreviated as BP and BLYP, respectively. Another popular GGA functional is the one proposed by Perdew, Burke, and Ernzerhof (PBE), it is based on exact results and retains the correct features of the LDA.

Despite the success of GGAs, there is no known systematic procedure for improving such 'gradient corrections' to the LDA, and continued progress in this area relies largely on physical intuition, knowledge of various constraining relationships, and simple trial and error.

Hybrid GGA. The exchange contribution is usually significantly larger in absolute numbers than the corresponding correlation effects, an accurate expression for the exchange functional is thus essential for obtaining meaningful results from DFT. The introducing of hybrid functionals which contain a certain amount of exact exchange has been an important step towards higher accuracy in DFT. One of the most

popular hybrid functionals is known as B3LYP [44],

$$E_{xc}^{\text{B3LYP}} = (1 - a)E_x^{\text{LDA}} + aE_x^{\text{exact}} + bE_x^{\text{B}} + cE_c^{\text{LYP}} + (1 - c)E_c^{\text{LDA}},$$

where a , b , and c are 0.20, 0.72, and 0.81, respectively. Nevertheless, the hybrid functionals have yet to find popularity in solid state physics, owing to the high computational cost associated with calculating exact exchange in plane wave codes.

2.1.2 Plane waves and pseudopotentials

Plane waves form a complete and orthonormal basis and are defined as

$$f_{\mathbf{G}}^{\text{PW}}(\mathbf{r}) = \frac{1}{\sqrt{\Omega}} \exp[i\mathbf{G} \cdot \mathbf{r}],$$

with the cell volume Ω and the reciprocal space vector \mathbf{G} . The KS orbitals can then be expanded:

$$\psi_i(\mathbf{r}) = \frac{1}{\sqrt{\Omega}} \sum_{\mathbf{G}} c_i(\mathbf{G}) \exp[i\mathbf{G} \cdot \mathbf{r}].$$

This expansion has to be truncated at an energy cutoff $E_{\text{cut}} = \frac{1}{2}\mathbf{G}^2$, which determines the number of plane waves,

$$N_{\text{PW}} = \frac{1}{2\pi^2} \Omega E_{\text{cut}}^{3/2},$$

and, in consequence, the accuracy of the calculations. There are many advantages of plane waves: (1) they are not space-fixed functions, implying that the Pulay forces [45] vanish exactly even within a finite basis, facilitating the computation of forces tremendously; (2) plane waves form an unbiased basis set without favoring certain atoms or regions over others; (3) no basis set superposition errors (BSSEs) occur when evaluating interaction energies; (4) the quality of the basis is controlled simply by the parameter E_{cut} ; and (5) the evaluation of various expressions can be speeded up significantly by using Fast Fourier Transforms [46]. Expanding the core wave functions or the core oscillatory region of the valence wave functions into plane waves, however, is extremely inefficient. In practice, in order to minimize the size of the plane wave basis necessary for calculations, core electrons are always replaced by pseudopotentials.

Pseudopotentials are required to produce pseudo wave functions that approach the full wave func-

tions outside a core radius r_c . Inside this radius, the pseudopotential and the pseudo wave function should be as smooth as possible to allow for a small plane wave cutoff. For the pseudo wave function this requires the nodal structure of the valence wave function to be replaced by a smooth function. In addition, the charge enclosed within r_c for the all-electron and pseudo wave functions must be equal. If a pseudopotential satisfies all the conditions above, it is commonly referred to as a norm-conserving pseudopotential [47, 48], which, by construction, has to be angular-momentum dependent. In its most general form, it is semi-local,

$$V^{\text{PP}}(\mathbf{r}, \mathbf{r}') = \sum_{lm} |Y_{lm}(\hat{\mathbf{r}})\rangle V_l(\mathbf{r}) \delta_{\mathbf{r}, \mathbf{r}'} \langle Y_{lm}(\hat{\mathbf{r}}')|,$$

where Y_{lm} are spherical harmonics. Most applications nowadays adapt a fully separable form in order to achieve substantial savings in computer time and storage [49–52]. More information on pseudopotentials and their construction can be found in Refs. [53, 54].

Plane wave basis sets are not without flaws even when coupled with pseudopotentials. A downside of being an unbiased basis set is that one cannot assign more basis functions to regions where they are more needed, and this is particularly unfavorable for strongly inhomogeneous systems. In addition, to calculate exact exchange with plane waves incurs a large computational overhead, hindering the use of hybrid functionals which in many respects are known to be more accurate than standard GGAs.

2.2 *Ab Initio* Molecular Dynamics

Ab initio molecular dynamics (AIMD) combines first-principles electronic structure methods, most commonly DFT, with molecular dynamics based on Newton’s equations of motion; the basic idea is to generate the dynamical trajectory using forces computed directly from electronic structure calculations which are performed on the fly as simulations proceed. AIMD, unlike its classical counterpart, is thus able to respond to changes in chemical situations during the course of simulation and allows for a proper treatment of electronic polarization and many-body effects. The accessible simulation length, however, is much shorter ($\sim 10^3 - 10^4$ times) than what is affordable with its classical counterpart.

Two most popular AIMD schemes, Born-Oppenheimer molecular dynamics (BOMD) and Car-Parrinello molecular dynamics (CPMD) [55], rely heavily on the Born-Oppenheimer approximation

which states that the ionic and electronic degrees of freedom can be separated adiabatically at all points in the phase space. Yet this approximation breaks down in some cases and more advanced AIMD scheme (e.g., Ehrenfest dynamics) is then required. A brief description on BOMD and CPMD is given here; one is referred to Refs. [56] and [57] for more involved descriptions.

2.2.1 Born-Oppenheimer molecular dynamics

BOMD solves the static electronic structure problem at every molecular dynamics step given the set of fixed nuclear positions at that instant in time. As a result, the time-dependence of the electronic structure is a consequence of the nuclear motion and the electronic structure part is reduced to solving the time-independent Schrödinger equation concurrently to propagate the nuclei via classical mechanics:

$$\begin{aligned} M_I \ddot{\mathbf{R}}_I(t) &= -\nabla_I \min_{\Psi_0} \langle \Psi_0 | \mathcal{H}_e | \Psi_0 \rangle, \\ E_0 \Psi_0 &= \mathcal{H}_e \Psi_0. \end{aligned}$$

Since the force depends exclusively on the minimization of $\langle \mathcal{H}_e \rangle$, a tight convergence of the wave function at every step is essential for a properly conserved classical nuclear Hamiltonian that shows minimal drift.

2.2.2 Car-Parrinello molecular dynamics

CPMD takes advantage of the smooth time-evolution of the dynamically evolving electronic subsystem as much as possible and makes an acceptable compromise on the length of the time step. A set of orbitals optimized initially is given a fictitious time dependence and is propagated along with the nuclear configuration.

The Car-Parrinello Lagrangian,

$$\mathcal{L}_{CP} = \frac{1}{2} \sum_I M_I \dot{\mathbf{R}}_I^2 + \frac{1}{2} \sum_i \mu \langle \dot{\psi}_i | \dot{\psi}_i \rangle - \langle \Psi_0 | \mathcal{H}_e | \Psi_0 \rangle + \sum_{ij} \Lambda_{ij} (\langle \psi_i | \psi_j \rangle - \delta_{ij}),$$

contains the kinetic energies of the nuclei and the electrons, the KS energy, and the constraints Λ_{ij} which ensure that the orbitals remain orthonormal. A fictitious mass μ is assigned to the orbital degrees

of freedom. Newton's equations of motion are obtained from the associated Euler-Lagrange equations:

$$\begin{aligned}\frac{d}{dt} \frac{\partial \mathcal{L}_{\text{CP}}}{\partial \dot{\mathbf{R}}_I} &= \frac{\partial \mathcal{L}_{\text{CP}}}{\partial \mathbf{R}_I}, \\ \frac{d}{dt} \frac{\delta \mathcal{L}_{\text{CP}}}{\delta \dot{\psi}_i^*} &= \frac{\delta \mathcal{L}_{\text{CP}}}{\delta \psi_i^*}.\end{aligned}$$

The corresponding Car-Parrinello equations are:

$$\begin{aligned}M_I \ddot{\mathbf{R}}_I(t) &= -\frac{\partial}{\partial \mathbf{R}_I} \langle \Psi_0 | \mathcal{H}_e | \Psi_0 \rangle + \sum_{ij} \Lambda_{ij} \frac{\partial}{\partial \mathbf{R}_I} \langle \psi_i | \psi_j \rangle, \\ \mu \ddot{\psi}_i(t) &= -\frac{\delta}{\delta \psi_i^*} \langle \Psi_0 | \mathcal{H}_e | \Psi_0 \rangle + \sum_j \Lambda_{ij} \psi_j,\end{aligned}$$

with the conserved energy E_{cons} ,

$$E_{\text{cons}} = \frac{1}{2} \sum_I M_I \dot{\mathbf{R}}_I^2 + \frac{1}{2} \sum_i \mu_i \langle \psi_i | \psi_i \rangle + \langle \Psi_0 | \mathcal{H}_e | \Psi_0 \rangle.$$

The fictitious mass μ must be carefully chosen so to maintain the adiabatic separation [*i.e.*, the lowest electronic frequency ($\omega_e^{\text{min}} \propto E_{\text{gap}}/\mu$, E_{gap} is the energy gap of the system) is much larger than the highest frequency of the nuclei] while still keeping a reasonable time step ($\Delta t^{\text{max}} \propto \mu/E_{\text{cut}}$, E_{cut} is the cutoff energy).

2.3 London Dispersion Forces

2.3.1 Dispersion forces in general

London dispersion forces are also known as dispersion forces, London forces, and sometimes van der Waals (vdW) forces. They are universal attractive forces that arise from the transient dipoles all molecules possess as a result of the fluctuations in the instantaneous positions of electrons. Two instantaneous dipoles are correlated in direction, and because of this correlation, the attraction between the two instantaneous dipoles does not average to zero.

A quantum treatment of dispersion forces is outlined briefly using perturbation theory to calculate the lowering in energy when two closed-shell atoms are brought to a separation R [58]. The perturbation Hamiltonian is the interaction of two electric dipole operators ($\boldsymbol{\mu}_A, \boldsymbol{\mu}_B$) based on the two atoms A and

B. It follows from classical electrostatics,

$$H^{(1)} = \frac{1}{R^3} \left\{ \boldsymbol{\mu}_A \cdot \boldsymbol{\mu}_B - \frac{3(\boldsymbol{\mu}_A \cdot \mathbf{R})(\boldsymbol{\mu}_B \cdot \mathbf{R})}{R^2} \right\}.$$

It is often convenient to choose the z -axis to be parallel with \mathbf{R} , the vector joining A to B and the origin at A . The perturbation is then

$$H^{(1)} = \frac{1}{R^3} \{ \mu_{Ax} \mu_{Bx} + \mu_{Ay} \mu_{By} - 2\mu_{Az} \mu_{Bz} \}.$$

The total Hamiltonian of the system is

$$H = H^{(0)} + H^{(1)}, H^{(0)} = H_A + H_B.$$

The unperturbed states of the pair of atoms are $|n_A n_B\rangle$, with

$$H^{(0)} |n_A n_B\rangle = (E_{n_A} + E_{n_B}) |n_A n_B\rangle.$$

We shall write $E_{n_A n_B} = E_{n_A} + E_{n_B}$ and consider interactions between the two atoms in their ground states, $|00\rangle$. The first-order correction to the energy is zero because every matrix element is the ground-state expectation value of the electric dipole moment operator, which is zero for a non-polar species.

The second-order energy is

$$\begin{aligned} E^{(2)} &= \sum_{n_A, n_B} \frac{\langle 00 | H^{(1)} | n_A n_B \rangle \langle n_A n_B | H^{(1)} | 00 \rangle}{E_{00} - E_{n_A n_B}} \\ &= -\frac{2}{3} \left(\frac{1}{R^3} \right)^2 \sum_{n_A, n_B} \frac{(\boldsymbol{\mu}_{A,0n_A} \cdot \boldsymbol{\mu}_{A,n_A0})(\boldsymbol{\mu}_{B,0n_B} \cdot \boldsymbol{\mu}_{B,n_B0})}{\Delta E_{n_A0} + \Delta E_{n_B0}} \end{aligned}$$

where $\Delta E_{n_A0} + \Delta E_{n_B0} = E_{n_A n_B} - E_{00}$ and $\langle 0 | \mu_{Ax} | n_A \rangle \langle n_A | \mu_{Ax} | 0 \rangle = \boldsymbol{\mu}_{A,0n_A} \cdot \boldsymbol{\mu}_{A,n_A0}$. This expression confirms that there is a nonzero interaction energy which is attractive and is inversely proportional to R^{-6} .

After few approximations, one arrives at the London formula:

$$V = -\frac{C}{r^6}, \quad C = \frac{2}{3} \alpha_1 \alpha_2 \frac{I_1 I_2}{I_1 + I_2}$$

where I and α are the ionization energy and polarizability of the molecule, respectively. The London formula, although only approximate, reveals the essential character of the dispersion interaction and may be used to make rough estimates of its magnitude.

From the quantum chemical point of view, a description of these very weak forces requires an accurate treatment of electron correlation effects. HF is unable to treat these attractive forces at all; high-level correlated *ab initio* methods such as coupled-cluster theory with large basis sets or quantum Monte Carlo are required, but they are too costly to be applicable for all but the smallest systems.

2.3.2 Dispersion forces in density functional theory

DFT has been quite successful in describing a wide variety of strongly interacting systems, but it has not met equal success in describing weak interactions. The precision achievable with the current approximated exchange-correlation functionals on systems dominated by dispersion forces is rather disappointing and is difficult to assess *a priori*. As vdW interactions contribute to the interaction energy even at distances where electron overlap is negligible, current functionals that are based on the local electron density, its local gradient, or the local kinetic energy density fail by construction to reproduce these interactions. The development for possible solutions to this failure has become an active field of research, and numerous approaches have since emerged.

Approaches based on electron density partitioning, involving the assignment of fragments, have been applied to rare-gas dimers and small hydrocarbon complexes [59–61]. Introducing non-local correlations into a vdW functional has shown promising results on rare-gas dimers, aromatic ring complexes and, graphite [29, 62–67]. In addition, a long-range correction scheme combined with a vdW functional has successfully described interactions of small vdW complexes as well as benzene and naphthalene dimers [68, 69]. An alternative approach is to optimize a functional specifically for non-bonded interactions (*e.g.*, M05-2X and PWB6K [70]). Nevertheless, it still gives an exponential decay at long range instead of the asymptotic R^{-6} tail for the interaction energy of two systems with no permanent multipole moments. A method utilizing frequency-dependent density susceptibilities predicted by time-dependent DFT to compute the dispersion energy at finite inter-monomer separations has been proposed [71]. Furthermore, Kohn *et al.* [28] has presented a scheme that is valid at all distances but is computationally very demanding; it is thus conceptually helpful yet not applicable to systems of chemical/biological interest.

The most straightforward solution to recover dispersion forces is to explicitly include empirical pairwise inter-atomic potentials of the $f(R)C_6R^{-6}$ form in the total energy where $f(R)$ is a damping function. This scheme generally leads to reasonable stabilization energies and nuclear forces [16,23,30,72], but it involves parameterization of C_6 coefficients and defining appropriate $f(R)$. Furthermore, the electronic structure is left uncorrected. Recently, a post HF model in which the instantaneous dipole moment of the exchange hole is used to generate the dispersion coefficients (C_6 or higher) [33,73–75] has predicted the geometries and binding energies for a large test set of intermolecular complexes remarkably well, except for the $\pi - \pi$ stacked benzene dimer.

2.4 Dispersion-Corrected Atom-Centered Potentials

In analogy to the concept of atomic pseudopotentials where the parameters are generated by iteratively minimizing a penalty functional that expresses the deviations of pseudo wave function from its all-electron counterpart, a similar procedure can be adapted to design atom-centered potentials (ACPs) for not only atomic but also complex molecular electronic properties. The latter involves having a penalty functional that penalizes deviations of molecular properties (*e.g.*, dispersion forces or electron density) with respect to experimental or high-level *ab initio* references.

Dispersion-corrected atom-centered potentials (DCACPs) are one example of such concept in which ACPs are introduced to recover dispersion forces within DFT-GGA [34]. This approach attempts to model the attractive long-range electron density correlation by an atom–electron interaction instead of an atom–atom interaction of the C_6R^{-6} form.

DCACPs retain the same analytical form as the non-local part of the Goedecker-Teter-Hutter (GTH) pseudopotentials [52] since we believe that the non-local form can inherently cast the non-local character of dispersion forces. In addition, it will be straightforward to apply DCACPs in (plane wave) DFT calculations by simply including the parameters to the corresponding pseudopotential files. The DCACP operator is defined as

$$\hat{v}_I^{\text{DCACP}}(\mathbf{r}, \mathbf{r}') = \sum_{\ell=0}^{\ell_{\max}} \sum_{m=-\ell}^{+\ell} Y_{\ell m}(\hat{\mathbf{r}}) \hat{p}_{\ell}(r) \sigma_1 \hat{p}_{\ell}(r') Y_{\ell m}^*(\hat{\mathbf{r}}'),$$

and the normalized projectors $\hat{p}_\ell(r)$ are Gaussian functions of the form

$$\hat{p}_\ell(r) = \frac{\sqrt{2}r^\ell \exp[-r^2/(2\sigma_2^2)]}{\sigma_2^{\ell+3/2} \sqrt{\Gamma(\ell + 3/2)}}.$$

$r = |\mathbf{r} - \mathbf{R}_I|$ is the distance from the position of nucleus I , $\hat{\mathbf{r}}$ is the unit vector in the direction of $\mathbf{r} - \mathbf{R}_I$, and $Y_{\ell m}$ denotes a spherical harmonic. We have chosen to use only $\ell = 3$ in the above expansion (see Chapter 4 for further discussions). The two parameters σ_1 and σ_2 are generated by minimizing a penalty functional \mathcal{P} composed of two energy- and one nuclear-force-dependent terms:

$$\begin{aligned} \min_{\{\sigma_i\}} \mathcal{P}(r) &= \min_{\{\sigma_i\}} [|E^{\text{ref}}(r_{\text{min}}) - E(r_{\text{min}}, \{\sigma_i\})|^2 + \\ &\quad |E^{\text{ref}}(r_{\text{mid}}) - E(r_{\text{mid}}, \{\sigma_i\})|^2 + \\ &\quad \sum_I w_I |F_I(r_{\text{min}}, \{\sigma_i\})|^2]. \end{aligned}$$

r_{min} and r_{mid} are the equilibrium and midpoint distances¹ of the reference calculations. E^{ref} is the reference interaction energy. Only forces along the intermolecular interaction axis are considered. w_I is a weighting factor chosen so that contributions from the energy and the force (F_I) terms are of the same order of magnitude; the typical value is 10^{-2} . By tuning the amplitude σ_1 and the width σ_2 of DCACPs, \mathcal{P} is minimized in such a way that the reference equilibrium interaction energy/distance and the midpoint interaction energy are optimally reproduced for a chosen calibration system.

DCACPs, with their current form, do not model the asymptotic dipolar nature of dispersion interactions explicitly. As a consequence, this approach can be accurate at the equilibrium distance but not in the long-range limit that would be important for, say, small-angle scattering of rare gases in molecular beams. The introduction of the midpoint energy term in the penalty functional is the first attempt in improving the long-range behavior of the resulting interaction energy curves [76]. Nevertheless, it is possible to achieve the asymptotic r^{-6} behavior by having a more complete (spherical harmonics and/or Gaussian functions) basis, *i.e.*, expanding DCACPs with more than one ℓ (see Chapter 3).

¹the midpoint distance is where the interaction energy is half of the equilibrium energy.

Chapter 3

Multi-Center Density Functionals

Abstract

We propose to express corrections to the approximated universal density functionals as multi-(atom)-centered functionals, drawing on the atoms-in-molecules theory of Bader in combination with a multi-center density expansion. Unlike the conventional density functionals which depend on the origin-centered electron density, variables of these newly formulated multi-center functionals are the positions of the nuclei and their identity which can be unambiguously determined from the topology of the electronic charge density.

We propose a multi-(atom)-centered expansion for the correction of most commonly used (origin-centered) density functionals. Multi-center expansions of classical and quantum fields, such as densities and wave functions, have been exploited in many electronic structure theories [77–80]. Most commonly, the centers of the expansion coincide with the positions of the atoms in a system. Nevertheless, to the best of our knowledge, there is so far no attempt to apply this technique for the derivation of more accurate exchange-correlation functionals to be used in density functional theory (DFT). In this case, the particular choice of multi-(atom)-centered expansion is physically intuitive since deviations from the homogeneous electron gas reference are largest at the atoms. Specifically, the total electronic density of a system can be regarded as a sum of the atomic densities corrected for the inter-atomic interactions [77]. Some of the topological properties of the atomic densities are preserved even in the formation of molecules and solids; cases in point are the locations of the density maxima and the cusps which are uniquely associated with the positions of the atoms [81–83]. Considering all these, it is thus natural to derive a multi-center-expansion-based scheme for the optimization of density functionals that are based on the topology of the total electronic density.

We first introduce the concept of reformulating corrections to the approximated universal density functional (eq 3.2) as a multi-center functional, drawing largely from the idea of multi-center density expansions [80]. Using the atoms-in-molecules theory of Bader [84], we then demonstrate that all quantities required for the functional, *i.e.*, the positions and the charges of the nuclei in a system, can be uniquely derived from the electronic charge density. The approach of atom-centered potentials (ACPs) is briefly recapped before we show its success on treating the lack of dispersion forces in DFT-BLYP, reproducing essentially exact asymptotic r^{-6} behavior when a sufficiently complete (spherical harmonics and/or Gaussian functions) basis is employed.

The concept of multi-center density functionals. In DFT, the basic variable is the electron density $\rho(\mathbf{r})$, a function of three spatial coordinates (and, for spin polarized systems, the spin) defined with respect to an arbitrarily chosen origin. Given a density, the following relation,

$$E[\rho] = F[\rho] + \int d\mathbf{r} v_{\text{ext}}(\mathbf{r})\rho(\mathbf{r}), \quad (3.1)$$

delivers the corresponding energy and upon minimization, the ground state density and energy. $F[\rho]$ is a universal functional which does not depend explicitly on the external potential $v_{\text{ext}}(\mathbf{r})$. Even though

such a functional exist, its exact functional form is still elusive and many works have been devoted to the search of better approximations. One can rewrite the exact universal functional $F[\rho]$ as the sum of an approximate form commonly adapted by the DFT community ($F^{\text{approx}}[\rho]$) and a correction term ($\Delta F[\rho]$) with respect to $F[\rho]$,

$$F[\rho] = F^{\text{approx}}[\rho] + \Delta F[\rho]. \quad (3.2)$$

Instead of an origin-centered $\Delta F[\rho]$, here we propose to utilize a multi-center expansion of the form $\Delta f[\rho](\{\mathbf{r}_j\}_{j=1}^M)$; \mathbf{r}_j is defined with respect to site j , $\mathbf{r}_j = \mathbf{r} - \mathbf{R}_j$. We shall follow closely the argument presented by Averill and Painter [80] who considered the decomposition of the charge density into a sum of atom-centered functions. From now on, the functional form used for $\Delta F[\rho]$ and all its components is $\Delta F[\rho] = \int d\mathbf{r} \Delta F[\rho](\mathbf{r})$.

Using the projection technique pioneered by Boys and Rajagopali [77] and further refined by Becke [78] and Delley [79], $\Delta F[\rho]$ can be written as a sum of site- j -centered functions P_j ,

$$\Delta F[\rho] \approx \Delta f[\rho](\{\mathbf{r}_j\}_{j=1}^M) = \sum_{j=1}^M P_j[\rho](\mathbf{r}_j) = \sum_{j=1}^M \Delta f_j[\rho](\mathbf{r}_j). \quad (3.3)$$

P_j is defined by a set of weight functions ω_j also centered at site j ,

$$P_j[\rho](\mathbf{r}_j) = \omega_j(\mathbf{r}) \Delta F[\rho](\mathbf{r}), \quad (3.4)$$

where for any \mathbf{r} ,

$$\sum_j \omega_j(\mathbf{r}) = 1. \quad (3.5)$$

One can choose ω_j to be the homonuclear fuzzy-cell function of Becke [78]. In this scheme, each ω_j has value unity in the vicinity of its own nucleus, but vanishes in a continuous and well-behaved manner near any other nucleus; the system is thus divided into fuzzy, overlapping analytically continuous cells. Applying ω_j to $\Delta F[\rho](\mathbf{r})$ produces P_j that is large in magnitude near site j and approaches zero away from site j .

It is more convenient at this point to introduce an expansion for the corresponding multi-center

operator $\Delta\hat{f}(\mathbf{r})$ so that

$$\Delta f_j[\rho](\mathbf{r}) = \rho(\mathbf{r})\Delta\hat{f}_j(\mathbf{r}). \quad (3.6)$$

Carrying out partial-wave projections about j ,

$$\hat{p}_{\ell m}^j(r_j) = \int Y_{\ell m}^j(\hat{\mathbf{r}}_j)\hat{P}_j(\mathbf{r}_j) d\Omega, \quad (3.7)$$

determines the approximate representation of \hat{P}_j ,

$$\hat{P}_j(\mathbf{r}_j) \approx \sum_{\ell=0}^{\ell_{\max}} \sum_{m=-\ell}^{+\ell} Y_{\ell m}^j(\hat{\mathbf{r}}_j)\hat{p}_{\ell m}^j(r_j), \quad (3.8)$$

where $d\Omega$ is the angular volume element and the accuracy of the expansion is controlled solely by the cutoff value ℓ_{\max} .

Putting the pieces together and generalizing for the non-local case, we obtain

$$\Delta\hat{f}_j(\mathbf{r}_j, \mathbf{r}'_j) = \sum_{\ell=0}^{\ell_{\max}} \sum_{m=-\ell}^{+\ell} Y_{\ell m}^j(\hat{\mathbf{r}}_j)\hat{p}_{\ell m}^j(r_j)\hat{p}_{\ell m}^j(r'_j)Y_{\ell m}^j(\hat{\mathbf{r}}'_j). \quad (3.9)$$

In the following, however, we would like to consider orbital-dependent corrections of the form,

$$\Delta f_j[\{\phi_i\}] = \sum_i \int \int d\mathbf{r}d\mathbf{r}' \phi_i^*(\mathbf{r}_j)\Delta\hat{f}_j(\mathbf{r}_j, \mathbf{r}'_j)\phi_i(\mathbf{r}'_j), \quad (3.10)$$

where i runs through the Kohn-Sham orbitals. This is the proposed orbital-dependent-functional form we use to expand $\Delta F[\rho]$ in eq 3.2.

Variables of multi-center functionals are uniquely defined by the density ρ only. Sites j in the multi-center scheme are most commonly (and naturally) chosen to be the location of atom I in the system. One thus requires knowledge of the positions (and sometimes the charges) of the nuclei for the evaluation of multi-center functionals. These two pieces of information should be defined by ρ and ρ only; in other words, the following requirements need to be satisfied: (a) the topology of $\rho(\mathbf{r})$ (maxima and cusps in particular) uniquely determines the positions of the multi-center expansion $\{\mathbf{R}_I\}_{I=1}^M$ [83,84]; (b) for any given choice of I -centered weight function ω_I , requirement (a) determines the projection $p_{\ell m}^I(r)$; and (c) a one-to-one mapping between Z_I and $\rho(\mathbf{R}_I)$, I is the index labeling the

atom in the system.

Using the atoms-in-molecules theory of Bader [84], one is able to determine the atomic fragments in molecular systems on the basis of the topology of the total electron density alone. The electronic charge density exhibits a local maximum only at the position of a nucleus in both the ground and excited states of many-electron systems (with relatively few exceptions to be discussed later). The principle justification comes from extensive theoretical studies of charge distributions calculated from wave functions close to the Hartree–Fock (HF) limit. Analyzing the topological properties of the charge distributions of a very large number of polyatomic molecules has shown that maxima found at positions other than the nuclei are trivially small compared with the ones at the nuclear positions [85]. This usual topology of ρ can change when two nuclei with dissimilar charges are in very close contact with each other and the local maximum at the position of the lesser of the two charges will disappear [85]. Fortunately, this behavior is not observed for energies normally encountered in chemical reactions for atoms that possess a core density (hydrogen does not possess a core and may be problematic when in combination with a very electronegative species) [86]. Strictly speaking, the local maxima at the positions of nuclei are not ‘true’ (3,-3) critical points¹ because the gradient vector of the charge density is discontinuous at the nuclear cusp that is present in both the wave function and density. Nevertheless, there always exists a function homeomorphic to $\rho(\mathbf{r}; \mathbf{R}_I)$ which coincides with ρ almost everywhere and for which the nuclear positions are (3,-3) critical points [84]. In this sense, the nuclear positions behave topologically as do (3,-3) critical points in the charge distribution; as a consequence, they can be used to describe the mapping in (b), together with the always-present cusps.

Regarding the requirement (c), the value of the charge density at a nuclear position [$\rho(\mathbf{R}_I)$], except for protons, is much larger than its value at any other of its extrema. $\rho(\mathbf{R}_I)$ of a free atom in the HF approximation is roughly proportional to Z^3 for $Z < 55$ (Z is the atomic number) as depicted in Figure 3.1 [87]. On the other hand, values of ρ at the saddle or bond critical points between certain pairs of nuclei over the range of chemically significant inter-nuclear separations range between $0 \leq \rho \leq 1.0$ a.u [84], of considerably smaller magnitude than $\rho(\mathbf{R}_I)$. This universal mapping still holds in most chemical and physical electronic structure approaches, particularly those in which the frozen-core

¹The critical point is labeled by its rank ω and signature σ in the form of (ω, σ) . The former is equal to the number of non-zero eigenvalues (of the Hessian matrix) of ρ at the critical point and the latter is the algebraic sum of the signs of the eigenvalues, *i.e.*, (3,-3) denotes the critical point where all curvatures are negative and ρ is a local maximum [84].

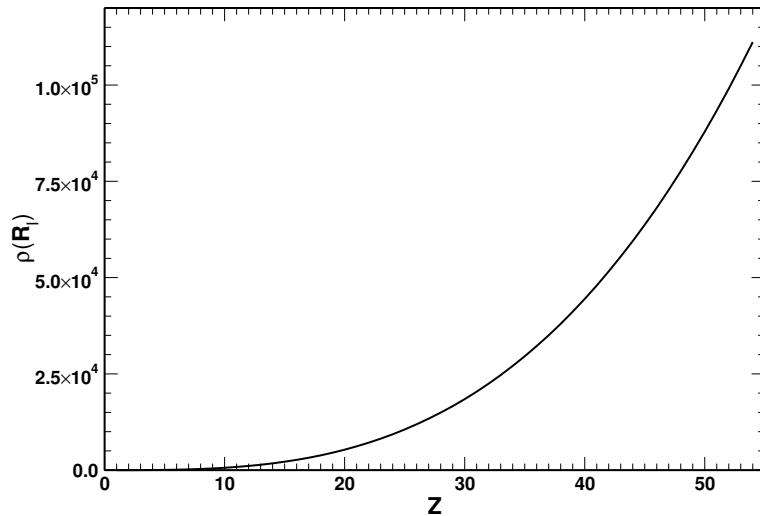


Figure 3.1: Value of spherically averaged single-particle density at nucleus [$\rho(\mathbf{R}_I)$ in a.u.] as a function of the atomic number Z for all neutral atoms with $Z < 55$ in the HF approximation. Data are taken from Ref. [87].

approximation is invoked.

Furthermore, DFT states that, in principle, knowledge of the ground-state electron density is sufficient to determine all molecular properties [82]. Kato's theorem [81] states that

$$Z_I = -\frac{1}{2\rho(\mathbf{r})} \left. \frac{\partial \rho(\mathbf{r})}{\partial r} \right|_{\mathbf{r} \equiv \mathbf{R}_I}, \quad (3.11)$$

where the partial derivatives are taken at the nuclei I . The cusps of the density thus tell us where the nuclei are (\mathbf{R}_I) and what the atomic numbers Z_I are.

In conclusion, given any (even approximate) density of a many-electron system constructed from a sum of atomic density and optimized according to any *ab initio* approach, it is possible to unambiguously locate all atoms (\mathbf{R}_I) in the system and to identify the corresponding physical nature (Z_I) on the basis of the electron density distribution only, and any functionals that involves atomic positions can be re-cast into a density-only (*i.e.*, 'universal') form.

The concept of ACPs. We have demonstrated that the origin-centered $\Delta F[\rho(\mathbf{r})]$ can be projected into a multi-(atom)-centered functional $\Delta f[\rho] \left(\{\mathbf{r}_j\}_{j=1}^M \right)$ (herein referred to as ACP). This correction term can be derived either from theory, but this is almost as difficult as searching for the true universal functional, $F[\rho]$. One can also empirically tune the ACP against some well-defined penalty functionals

to improve chosen atomic or molecular properties [88]. In this case we used an analytic form similar to the one proposed by Goedecker *et al.* [52] in the context of atomic pseudopotentials. These atom-centered analytic forms can be tuned to reach accuracy for specific properties on par with high-level *ab initio* calculations or simply functionals of higher rank. Among others, this approach has been used for (1) generating link atoms to bridge the quantum and classical fragments in quantum mechanics/molecular mechanics simulations, designed in such a way that the link atoms minimally perturb the electronic structure in the quantum mechanical region [88]; for (2) reproducing the electron density and derived molecular properties of hybrid functional quality within BLYP calculations [88]; and for (3) improving the description of dispersion forces in DFT (dispersion-corrected atom-centered potentials, DCACPs) [34].

ACPs employed in the above-mentioned studies are generally defined as

$$\Delta \hat{f}_I(\mathbf{r}_I, \mathbf{r}'_I) = \hat{v}_I^{\text{ACP}}(\mathbf{r}, \mathbf{r}') = \sum_{\ell=0}^{\ell_{\max}} \sum_{m=-\ell}^{+\ell} Y_{\ell m}(\hat{\mathbf{r}}) \hat{p}_\ell^I(r) h_{I\ell}(Z_I) \hat{p}_\ell^I(r') Y_{\ell m}^*(\hat{\mathbf{r}}') \quad (3.12)$$

with the normalized projector,

$$\hat{p}_\ell^I(r) \propto \frac{r^\ell \exp[-r^2/2\beta_{I\ell}(Z_I)^2]}{\beta_{I\ell}(Z_I)^{\ell+3/2}}. \quad (3.13)$$

$h_{I\ell}(Z_I)$ and $\beta_{I\ell}(Z_I)$ are two adjustable atom-dependent parameters per ℓ that, in principle, are uniquely assigned to different atoms in the system according to the mapping in (c).

We will concentrate on one particular application of ACPs in treating the lack of dispersion forces in DFT, but the following argument will be generally applicable.

For the spherical averaged part of the operator $\hat{P}_j[\rho](\mathbf{r}_j)$ in eq 3.7, namely the DCACPs projector \hat{p}_ℓ^I , we can easily identify

$$\omega_j^\ell(r) = \frac{\exp[-r^2/2\beta_{I\ell}(Z_I)^2]}{\beta_{I\ell}(Z_I)^{\ell+3/2}}, \quad (3.14)$$

$$\Delta F^\ell = r^\ell, \quad (3.15)$$

where $\Delta F^\ell[\phi_\alpha] = \int dr \Delta \hat{F}^\ell \phi_\alpha(r)$. Therefore, the atom-centered DCACP corrections in the origin-

centered representation are generated by a weighted sum of ‘multiple moments’,

$$q_{\ell m}^{\alpha} = \int d\mathbf{r} Y_{\ell m}(\hat{\mathbf{r}}) r^{\ell} \phi_{\alpha}(\mathbf{r}). \quad (3.16)$$

With DCACPs, we have opted for a one-channel ($\ell = 3$) expansion [76] since we are mostly interested in the region of the van der Waals (vdW) minimum and less so in the long-range limit. Indeed, despite its general excellent performance and transferability for dispersion interactions within distances up to ~ 5 Å, some deviations are observed at the long-range limit of the asymptotic dipolar nature [89–92].

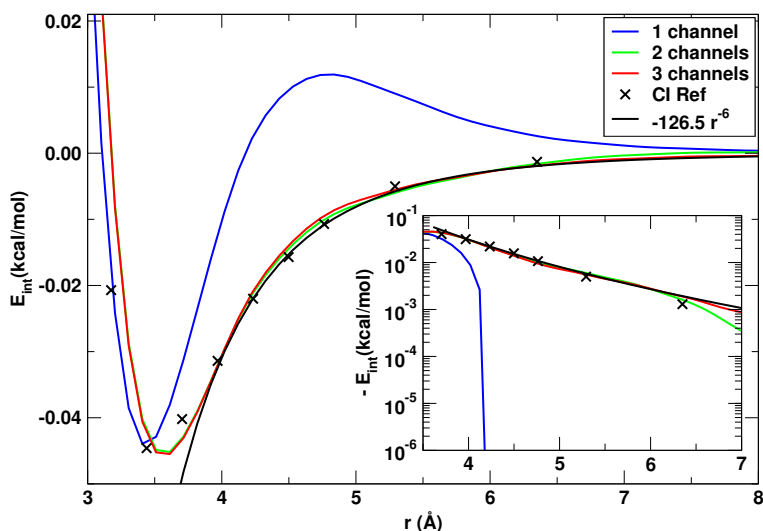


Figure 3.2: Interaction energy of $(\text{H}_2)_2$ aligned in parallel computed with DCACPs having multiple projectors.

Since both the spherical harmonics (angular) and Gaussian functions (radial) expansions adopted here form a complete set of orthonormal functions, they can in principle reproduce any functions with arbitrary accuracy. It should thus be straightforward to achieve the r^{-6} behavior by including more projectors in the expansion. To demonstrate this, H_2 dimer is chosen as an example for hydrogen is devoid of any non-local components in its atomic pseudopotential, and we are allowed to assign l as low as 0 to the DCACP without any interference with the underlying atomic pseudopotential². The results

²All DFT calculations have been carried out using the CPMD code [93], the BLYP functional [17, 18], Goedecker-Teter-Hutter pseudopotentials [52], and a plane wave cutoff of 100 Ry in an isolated cell with dimensions $10 \times 10 \times 20$ Å³. The hydrogen DCACP has been calibrated against full CI reference of H_2 dimer aligned in parallel [94], using the scheme proposed in Ref. [34].

are presented in Figure 3.2. The ‘tail’ of the CI reference is fitted to a function of the form $a \cdot r^{-6}$, and the two deviate slightly from each other, especially in the long-range limit, indicating influence of higher order terms such as r^{-8} . In addition, a drastic improvement in the r^{-6} description is observed when the expansion of DCACPs includes more than one ℓ . Essentially, exact r^{-6} behavior can be reached by as few as two projectors. This clearly demonstrates that the DCACP approach is capable of reproducing the physically correct r^{-6} asymptotic limit in spite of the fact that this functional form is not explicitly imposed.

To summarize, we have drawn on the atoms-in-molecules theory and the multi-center potential approach to show that it is possible to expand corrections to the approximated universal density functional in terms of multi-(atom)-centered contributions. This expansion is unambiguously determined by the topology of the electron density: the cusp condition determines the centers of the expansion (the nuclear positions) and $\rho(R_I)$ the nature of the atom at position R_I , Z_I . In this sense, this expansion can be considered ‘universal’, *i.e.*, it depends solely on the electronic density. The final assessment of the multi-center functionals is obtained through a fitting procedure of the Z_I -dependent parameters used in $\Delta \hat{f}_I(\mathbf{R}_I, \mathbf{r}'_I)$. This procedure, in principle, is carried out only once for each element, and it involves tuning the parameters so that the desired accuracy on any chosen molecular properties can be optimally reproduced. For the multi-center correction aimed to cure the lack of dispersion forces in DFT-GGA, we have shown, with the example of $(\text{H}_2)_2$, that it is possible to achieve the desired level of accuracy (*i.e.*, the correct r^{-6} asymptotic tail in the interaction energy curve) by including a sufficient number of components in the expansion in eq 3.12.

Chapter 4

Library of Dispersion-Corrected Atom-Centered Potentials

Abstract

Parameters for analytical dispersion-corrected atom-centered potentials (DCACPs) are presented to improve the description of London dispersion forces within the generalized gradient approximation functionals BLYP, BP, and PBE. A library of DCACPs for hydrogen, carbon, nitrogen, oxygen, helium, neon, argon, and krypton is generated via calibration against CCSD(T) or full CI references. The performance as well as the transferability of DCACPs are tested on weakly bound complexes, and excellent results are obtained in all investigated systems.

4.1 Introduction

London dispersion forces arise from the transient dipoles all molecules possess as a result of the fluctuations in the instantaneous positions of electrons. They are of primordial importance in chemical and biological systems; they play a crucial role in colloidal systems, in noble-gas chemistry, and in soft matter generally, contributing significantly to phenomena such as solvation, physisorption, molecular recognition, as well as the stability and conformational variability of molecular crystal and biomacromolecules. Kohn–Sham density functional theory (KS-DFT) [7, 8] combined with the local density approximation (LDA) or present-day generalized gradient approximation (GGA) exchange-correlation functionals either provides sporadic results or fails completely to account for these weak forces [12, 13, 15, 19–21]. Considerable efforts have been made to remedy this shortcoming [27–33].

Dispersion-corrected atom-centered potentials (DCACPs) represent the effect of dispersion forces via atomic orbital-dependent potentials whose two parameters are obtained by calibrating against references of chosen accuracy [34]. The first generation of DCACPs calibrated against MP2 references has already shown promising results [89, 95, 96].

Herein, a library of DCACPs, including parameters for hydrogen, carbon, nitrogen, oxygen, helium, neon, argon, and krypton, to be used in conjunction with BLYP [17, 18], BP [17, 42], and PBE [97] functionals, is presented. DCACPs can be tuned to reach any desired accuracy given by the chosen calibration reference. In this work, DCACPs are calibrated against high-level correlated *ab initio* CCSD(T) or full CI references of (H₂)₂ in a parallel configuration [94], sandwich benzene dimer [98], (N₂)₂ (parallel), (CO₂)₂ (cross-shaped), and rare-gas dimers [99]. The calibration systems are chosen in order to fulfill the following criteria: (a) the system is small enough so that high-level reference calculations are tractable, (b) the interaction energy is dominated by the balance between Pauli repulsion and dispersion forces, and (c) the electronic structures of the monomers in the complex are well described by the GGA employed. Testing is restricted to simple van der Waals (vdW) complexes where high-level reference data are available. Furthermore, we investigate the influence of DCACPs on intramolecular geometries and electronic structures by analyzing bond lengths, vibrational frequencies, electron densities, multipole moments, and polarisabilities. The compatibility of DCACPs with various atomic pseudopotential types is also addressed.

4.2 Computational Details

All DFT calculations have been carried out using the program CPMD [93]. Goedecker-Teter-Hutter (GTH) pseudopotentials [52] have been used throughout the calibrations; for the test calculations, Troullier-Martins (TM) [51] and Vanderbilt [50] types have also been employed as specified. For the calibrations, we have used plane wave cutoffs (referred to as cutoff from now on) of 100 Ry (C and Ar), 120 Ry (He), 125 Ry (N), 150 Ry (H, O, Ar, and Kr), 550 Ry (Ne, for PBE and BP), and 600 Ry (Ne, BLYP). The calibrations of H, He, Ar, and Kr have been carried out in an isolated $10 \times 10 \times 20 \text{ \AA}^3$ cell using the Poisson solver implemented in CPMD according to Ref. [100]. The calibration of neon has been carried out in an isolated cell of $10 \times 10 \times 15 \text{ \AA}^3$ (BLYP) and $10 \times 10 \times 18 \text{ \AA}^3$ (BP and PBE). N ($10 \times 10 \times 20 \text{ \AA}^3$ cell) as well as C and O ($15 \times 15 \times 20 \text{ \AA}^3$ cell) have been calibrated with periodic boundary conditions (PBCs). Unless otherwise stated, calibration references have been calculated at CCSD(T)/aug-cc-pVTZ level of accuracy (counterpoise-corrected [101] for basis set superposition errors) using the GAUSSIAN 03 package [102]. The test calculations have been set up as follows: 150 Ry cutoff, $10 \times 10 \times 20 \text{ \AA}^3$ cell with PBCs for Ar-N₂ complexes; 100 Ry cutoff, isolated $12 \times 12 \times 30 \text{ \AA}^3$ cell for formaldehyde dimer; 100 Ry cutoff, $15 \times 15 \times 15 \text{ \AA}^3$ cell with PBCs for H₂-benzene complex.

For the minimization of the penalty functional, a simple but robust simplex-downhill algorithm [46] has been employed.

4.3 Results and Discussion

4.3.1 Preliminary

Choice of ℓ . We found that one projector is sufficient to obtain reasonable accuracy (deviations within 0.05 kcal/mol and 0.1 Å) for the reference and test systems. To ensure that there is no interference with the atomic pseudopotentials centered at \mathbf{R}_I , any angular momentum component ℓ not occupied in the atomic pseudopotentials should, in principle, be used. The very different length scale between the atomic pseudopotentials and DCACPs should also ensure negligible interference between the two.

To illustrate this, interaction energies of H₂ dimer in parallel (calibration system) and cross configurations calculated with DCACPs located at different ℓ channels are plotted in Figure 4.1. The atomic pseudopotential for hydrogen consists of only the local part; DCACP can thus occupy angular momenta

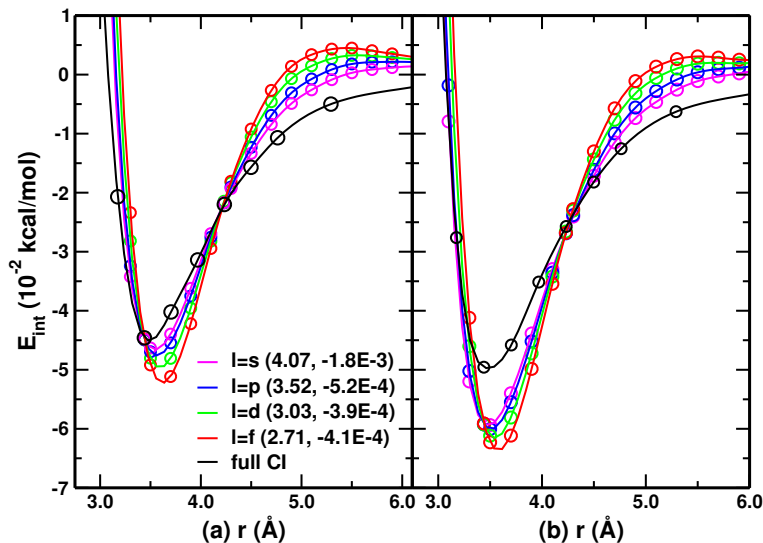


Figure 4.1: Interaction energies of H_2 dimer in (a) parallel and (b) cross configurations evaluated with DCACP at different ℓ channels. The two numbers in the legend correspond to σ_2 and σ_1 (in a.u.), respectively.

	#	Cutoff [Ry]	local only Time	$\ell = 0$		$\ell = 1$		$\ell = 2$		$\ell = 3$	
				Time	%	Time	%	Time	%	Time	%
He	200	100	22.45	23.00	2.4	24.04	7.1	25.09	11.8	26.18	16.6
He	240	100	26.72	27.40	2.5	28.76	7.6	30.24	13.2	31.52	18.0

Table 4.1: CPU time (Time, seconds) and the associated % increase with respect to the pure BLYP calculation (marked as ‘local only’) incurred by using different ℓ as the DCACP channel. CPU time quoted is for one wave function optimization step on 128 nodes of Blue Gene/L (equivalent to 356 CPUs, each has 0.5 gigabytes of memory). # is the number of helium atoms.

as low as $\ell = 0$. Having the DCACP at a lower angular momentum gives slightly better long-range behavior and is more computationally efficient [the scaling for calculating the non-local part of pseudopotentials goes up as $(2\ell + 1)N^3$, N is the number of atoms]. As an example, an increase of almost 20% in CPU time is decreased to mere 3% when ℓ is changed from 3 to 0 (Table 4.1, CPU time per wave function optimization step for a box of 200 or 240 helium atoms on 128 nodes of Blue Gene/L with 0.5 gigabytes of memory per CPU). Despite these, considered that most elements have up to $\ell = 2$, sometimes even $\ell = 3$, occupied in their respective atomic pseudopotentials and the almost negligible improvement when going from $\ell = 3$ to 2, we have chosen to standardize on the rarely used component $\ell = 3$.

Atom-, united-atom-, or atom-type-based. It would be ideal if the element of interest were the sole component of the calibration system. This, however, will not be practical for elements such as carbon

or boron where a ‘pure’ homonuclear dispersion-bound complex tends to have more than tens of atoms (*e.g.*, fullerene, graphene, or boron clusters) so a high-level reference is not readily available. This also raises the question whether DCACP should be atom-, atom-type-, or united-atom-based. The first choice, atom-dependent, is the most general; each element will have one and only one DCACP for all chemical environments without further tuning. The latter two are inspired by classical force fields: atom-type is similar to the all-atom force fields that provide parameters for every chemical type of an atom in a system, including hydrogen (*e.g.*, sp^3 and sp^2 carbon will have different parameters); with united-atom, the approach adopted by the first generation of DCACPs [34], the hydrogen and heavy (for example, carbon) atoms of different hybridization states are treated as a single interaction center.

Dependency & calibration system for C DCACP		BLYP			PBE		
		(C ₆ H ₆) ₂	(CH ₄) ₂	H ₂ -C ₆ H ₆	(C ₆ H ₆) ₂	(CH ₄) ₂	H ₂ -C ₆ H ₆
		$E_{\min}(r_{\min})$	$E_{\min}(r_{\min})$	$E_{\min}(r_{\min})$	$E_{\min}(r_{\min})$	$E_{\min}(r_{\min})$	$E_{\min}(r_{\min})$
United-atom	1. C ₆ H ₆	-1.72 (3.9)	-0.10 (4.1)	-0.42 (3.14)	-1.70 (3.9)	-0.49 (3.8)	-0.93 (2.70)
C only	2. CH ₄	-5.55 (3.7)	-0.44 (3.8)	-0.99 (2.74)	-1.17 (3.8)	-0.51 (3.7)	-1.06 (2.70)
Atom	3. C ₆ H ₆	-1.72 (3.9)	-0.83 (3.7)	-0.98 (2.70)	-1.75 (3.9)	-0.43 (3.8)	-0.90 (2.70)
H: CI	4. CH ₄	+1.60 (3.9)	-0.45 (3.8)	-0.54 (2.95)	-1.83 (3.8)	-0.51 (3.7)	-1.12 (2.70)
Atom	5. C ₆ H ₆	-1.72 (3.9)	-0.69 (3.8)	-0.90 (2.74)	-1.70 (3.9)	-0.34 (4.0)	-0.84 (2.95)
H: CCSD(T)	6. CH ₄	+0.68 (3.9)	-0.45 (3.8)	-0.53 (2.71)	-2.70 (3.7)	-0.50 (3.7)	-1.20 (2.70)
CCSD(T) ^a		-1.70 (3.9)	-0.50 (3.7)	-0.95 (2.75)	-1.70 (3.9)	-0.50 (3.7)	-0.95 (2.75)

^a (C₆H₆)₂: aug-cc-pVQZ* [98], (CH₄)₂: aug-cc-pVTZ (this work) and H₂-C₆H₆: aug-cc-pVTZ [103].

Table 4.2: Transferability of the carbon and hydrogen DCACPs calibrated in various schemes. In the united-atom-based scheme (united-atom), the carbon DCACP is calibrated and contributions from hydrogen are implicitly included; in the atom-based scheme (atom), the hydrogen DCACP is calibrated against either CI or CCSD(T) reference of (H₂)₂ (as labeled) first and is later used in the carbon DCACP calibration against CCSD(T) reference of either methane or benzene dimer as denoted in the first column. E_{\min} and r_{\min} are in kcal/mol and Å, respectively.

To probe the relative performance of these different schemes, we have carried out six sets of calibrations involving hydrogen and carbon: sets 1 and 2 are united-atom-based models in which the DCACP for carbon is calibrated against CCSD(T) reference of either benzene or methane dimer whose hydrogen atoms are DCACP-free; sets 3 and 4 are atom-based sets in which the DCACP for hydrogen is first calibrated against a full CI reference of (H₂)₂ and is then employed in the calibration of the DCACP for carbon against a CCSD(T) reference of either benzene or methane dimer; lastly, sets 5 and 6 use the same procedure as sets 3 and 4 but a CCSD(T) reference of H₂ dimer is used instead. The resulting DCACPs are subject to a small set of transferability tests including benzene as well as methane dimers and hydrogen adsorbed on benzene. The results are tabulated in Table 4.2. The performance of DCACPs

complementing PBE does not seem to depend much on the particular approach or calibration system one takes. On the other hand, DCACPs complementing BLYP show a large variation; using methane dimer as the calibration system for carbon generally leads to worse transferability, probably due to the over-repulsive character of Becke’s exchange. With the united-atom-based approach, the carbon DCACP fails badly regardless of the calibration system. As a compromise, we choose to take the atom-based approach and the highest level of reference possible whenever a decision needs to be made (full CI in the case of hydrogen). With regard to the atom-type-dependency, a separate study [90] on vdW complexes of aliphatic hydrocarbons molecules and crystals of aromatic hydrocarbon compounds illustrates the transferability of the DCACP scheme across different hybridization states of carbon: the typical error of binding energies for gas-phase dimers amounts to 0.3 kcal/mol. This further demonstrates that only one DCACP per element is sufficient, irrespective of the hybridization state, and the atom-based approach is more than adequate.

4.3.2 Parameters

	BLYP		BP		PBE	
	$\sigma_1[10^{-4}]$	σ_2	$\sigma_1[10^{-4}]$	σ_2	$\sigma_1[10^{-4}]$	σ_2
H	-4.06	2.71	-5.55	2.66	0.50	2.47
C	-5.49	3.11	-3.71	3.50	-5.79	2.84
N	-6.05	2.91	-8.06	2.82	-1.77	2.83
O	-7.92	2.57	-10.65	2.64	-6.47	1.73

Table 4.3: DCACP parameters for H, C, N, and O in a.u.

	BLYP		BP		PBE	
	$\sigma_1[10^{-4}]$	σ_2	$\sigma_1[10^{-4}]$	σ_2	$\sigma_1[10^{-4}]$	σ_2
He	-3.92	2.40	-9.91	2.16	3.31	1.98
Ne	-6.41	2.48	-12.92	2.42	3.00	2.07
Ar	-12.96	2.77	-16.42	2.79	-7.44	2.15
Kr	-12.95	3.18	-14.68	3.22	-3.48	3.20

Table 4.4: DCACP parameters for the rare-gas atoms in a.u.

Tables 4.3 and 4.4 list the DCACP parameters for H, C, N, O, and the rare-gas atoms He, Ar, Ne, and Kr to be used in combination with the GGA functionals BLYP, BP, and PBE. The radial term of DCACPs is plotted in the form of $-|\sigma_1|^{1/2}p_{\ell=3}(r)$ in Figure 4.2.

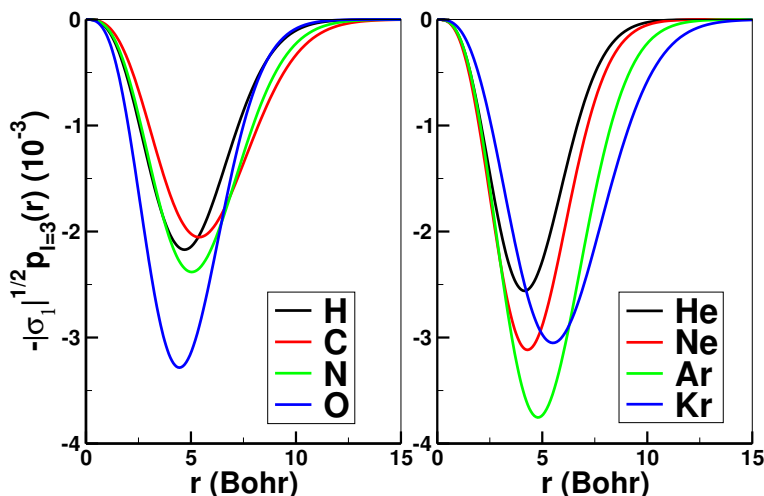


Figure 4.2: The radial term of DCACPs for the BLYP functional plotted in the form of $-|\sigma_1|^{1/2} p_{\ell=3}(r)$ (a.u.).

A positive σ_1 indicates that the original exchange-correlation functional produces over-binding curves whereas a negative σ_1 shows that the uncorrected interaction energy curve is under-binding. While the BLYP and BP intermolecular interaction curves are always repulsive for the calibration systems, PBE shows partial binding in some cases. As a result, DCACPs for BLYP and BP have consistently negative σ_1 . No obvious trend, however, can be observed for DCACPs complementing PBE, and the sign of σ_1 varies according to the performance of uncorrected PBE. Even though we choose to concentrate only on three popular GGA functionals – BLYP, BP, and PBE – the DCACP approach is general, and it can be applied to any functionals, including the LDA. The LDA is accurate for solids and is still widely used in condensed matter physics; however, GGAs show a more consistent performance across various disciplines, including chemistry and biology, and they offer a better description on intramolecular properties over the LDA [104, 105].

4.3.3 Calibrations

Figure 4.3 shows the interaction energy curves of $(\text{H}_2)_2$ calculated using DCACPs calibrated with and without the midpoint term in the penalty functional. The inclusion of the midpoint term significantly improves the midrange to long-range behavior; however, in order to satisfy this additional criterion, compromises are made: r_{\min}^{DCACP} for DCACP-PBE is shifted outwards by 0.06 Å with respect to r_{\min}^{CI} , and the interaction energy is more attractive by 0.002 kcal/mol. DCACP-BLYP behaves slightly worse:

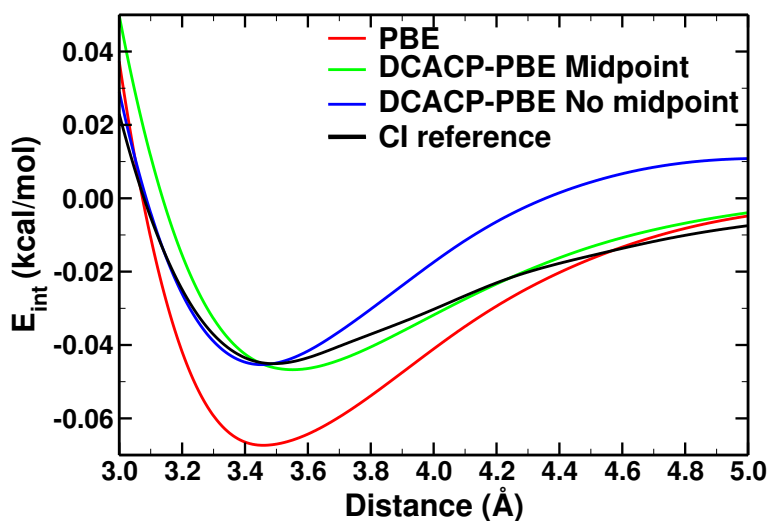


Figure 4.3: Interaction energy curves of the parallel H_2 dimer as a function of the intermolecular distance showing the effect of including the midpoint term in the penalty functional.

r_{\min}^{DCACP} is shifted out by 0.12 \AA compared with r_{\min}^{CI} , and the complex is over-stabilized by 0.007 kcal/mol . Bearing in mind that the interaction energy curve is shallow around the minimum, this shift is negligible for simulations at finite temperatures. Better fitting results are obtained for all other calibration systems using the midpoint penalty functional.

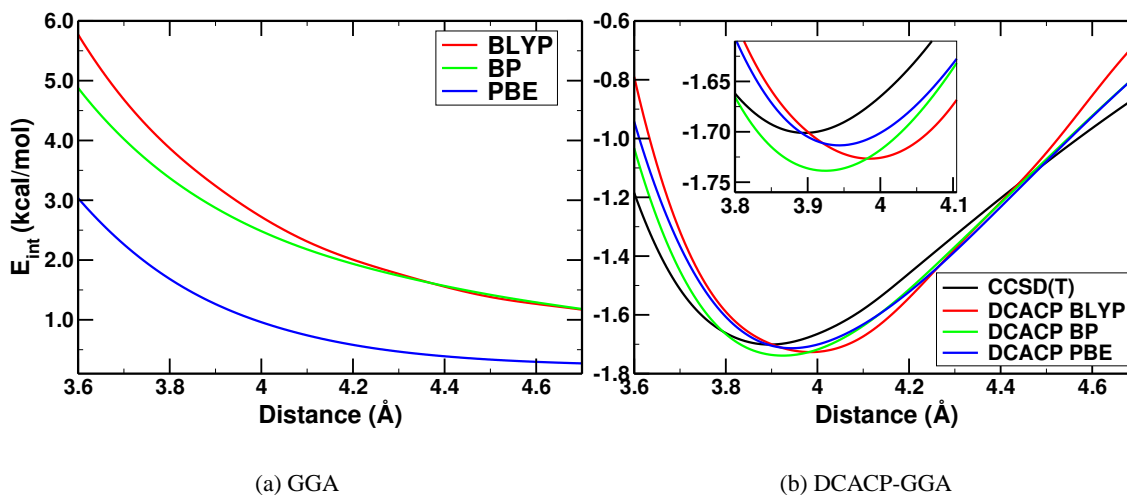


Figure 4.4: Interaction energy curves of the sandwich benzene dimer as a function of intermolecular distance with (a) pure and (b) DCACP-augmented GGA functionals.

Figure 4.4 shows the interaction energy curves of the sandwich benzene dimer as an example of the DCACP's performance on a calibration system. The overall shape of the interaction energy curves cal-

	He ₂		Ne ₂		Ar ₂		Kr ₂	
	E_{\min}	r_{\min}	E_{\min}	r_{\min}	E_{\min}	r_{\min}	E_{\min}	r_{\min}
DCACP-BLYP	0.020	3.0	0.094	3.0	0.247	3.8	0.322	4.2
DCACP-BP	0.022	3.1	0.105	3.2	0.250	3.8	0.322	4.2
DCACP-PBE	0.021	3.0	0.077	3.2	0.248	3.8	0.317	4.2
Ref. [99]	0.020	3.0	0.082	3.2	0.249	3.8	0.322	4.2

Table 4.5: Equilibrium interaction energies (E_{\min} , kcal/mol) and distances (r_{\min} , Å) of the rare-gas dimers.

	$\mu_{\text{H}_2\text{CO}}$	α_{Ar}	$\alpha_{\text{Ar-N}_2}$	Q_{CO_2}	Q_{N_2}	$Q_{\text{C}_6\text{H}_6}$
BLYP	2.295	12.24	24.43	-3.21	-1.13	-5.38
DCACP-BLYP	2.294	12.24	24.43	-3.21	-1.13	-5.39

$$\alpha_{\text{Ar-N}_2} = \alpha_{zz}^{\text{Ar-N}_2\text{T}}, Q = \langle z^2 - \frac{1}{2}x^2 - \frac{1}{2}y^2 \rangle$$

Table 4.6: Polarisabilities (α), dipole (μ), and quadrupole (Q) moments of formaldehyde, Ar, Ar-N₂, CO₂, N₂, and benzene. All values are in a.u., except for dipole moments which are quoted in Debye.

culated at the same level of accuracy as the reference points [CCSD(T)] and the ones evaluated with the DCACP-augmented functionals (DCACP-GGA) agree very well, considering that there are only three penalty terms (two energy- and one nuclear-force-dependent terms) and two adjustable parameters per element. For the rare-gas dimers, r_{\min} and E_{\min} are also in very good agreement with the reference values (Table 4.5), in great contrast to the spurious attractive or repulsive behavior of the pure functionals [12, 13, 20].

4.3.4 Effects on the electronic structure

The facts that DCACPs occupy a polarization channel and are much weaker as well as longer ranged than the atomic pseudopotentials should ensure that there is no interference between the two. To verify this, we have checked the effect of DCACPs on intramolecular geometries; specifically, the bond lengths of the isolated monomers of the calibration systems, including single, double, and triple bonds, have been computed. The values obtained with DCACP-GGA deviate negligibly from those of the corresponding uncorrected functionals (one thousandth of an Å). Furthermore, dynamical properties have been tested via vibrational frequency analysis on geometry-optimized (ionic gradient tolerance of 10^{-5} a.u.) molecular hydrogen, water, and carbon dioxide molecules, and only small deviations of $\leq 5 \text{ cm}^{-1}$ are observed, within numerical accuracy.

Unlike DFT-D [30] in which an explicit force-field-like r^{-6} dependency is included, DCACPs are

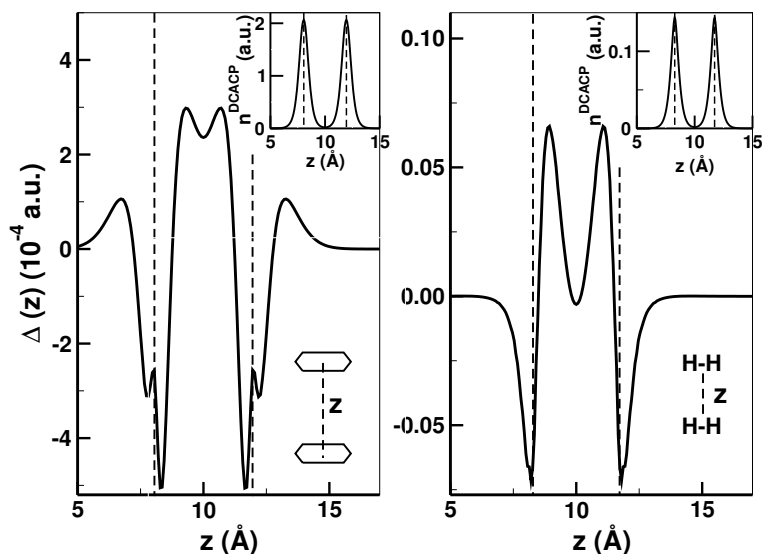


Figure 4.5: Graphs showing $\Delta(z) = \int dx dy [n^{\text{DCACP}}(\mathbf{r}) - n^{\text{BLYP}}(\mathbf{r})]$ (a.u.) against the intermolecular axis z (Å) for the benzene dimer (left) and the H_2 dimer (right). The dotted lines show the position of the moieties. The inset shows the DCACP-BLYP electron density for the respective system.

effective potentials whose contributions to the total energy depend explicitly on the electronic wave function. As a consequence, DCACPs adapt to the chemical environment where the atom is and exhibit remarkable generality. On the other hand, it is not desirable to have DCACPs strongly influencing the already reasonably well-described electronic properties. In this regard, the very small amplitude of DCACPs ($10^{-4} - 10^{-5}$ times smaller than the corresponding atomic pseudopotential) induces a negligible effect on characteristics that depend on the electronic structure; properties such as polarisabilities and multipole moments remain basically unchanged (Table 4.6). Figure 4.5 illustrates that the effect of DCACPs on the electronic structure of a vdW complex is small (the largest electron density difference is 10^{-4} times smaller than the maximum electron density of the complex) but relevant. The electron density in the inter-moiety region is increased at the expense of the density on the molecular plane.

4.3.5 Transferability

To be transferable entails that one and the same potential can be used in calculations of different chemical environments with comparable accuracy. It is one of the prerequisites for a newly developed method to be of any practical use. The assessment of the newly generated DCACPs is limited to a simple set of weakly bound complexes. Even though the size of the testing set is relatively small and is restricted to gas-phase systems only, the transferability of the parameters is expected to be at least on par with

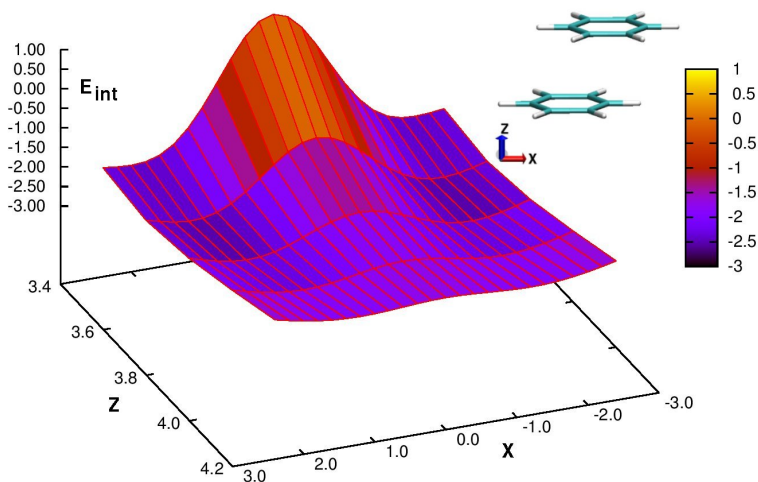
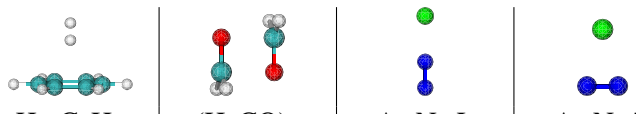


Figure 4.6: Interaction energy (E_{int} , kcal/mol) surface of the parallel-displaced benzene dimer at various x and z distances (both in Å) predicted by DCACP-BLYP. The global minimum is at $(x, z, E_{\text{int}}) = (\pm 1.76, 3.6, -2.60)$; the estimated CCSD(T)/aug-cc-pVQZ* result is $(\pm 1.67, 3.6, -2.63)$ [98].

the ones calibrated against MP2 references where encouraging results have been obtained for rare gases, graphene, and hydrocarbon complexes [34,89]. Furthermore, one recent study [90] has shown successful applications of CCSD(T)-calibrated DCACPs with BLYP on describing different hybridization states of carbon and condensed phase phenomena such as the binding energy of graphene sheets and the cohesive energy of benzene crystal.

The 2-dimensional interaction energy surface of the parallel-displaced benzene dimer is plotted in Figure 4.6. The global minimum predicted by DCACP-BLYP is only off by 0.1 Å (in x) and -0.03 kcal/mol compared with the estimated CCSD(T)/aug-cc-pVQZ* reference [98], showing the excellent transferability of DCACPs to complexes of different orientations from the calibration one.

The interaction energy of H_2 -benzene (bound by dispersion forces), the Ar- N_2 complex in two different configurations (bound by dispersion forces), and the anti-parallel formaldehyde (H_2CO) dimer (bound by dipole-dipole interaction and dispersion forces) have been evaluated with fixed monomer geometries. The results are tabulated in Table 4.7. DFT-GGAs fare badly; the minimum interaction energy E_{min} is overly underestimated [mean absolute error (MAE) 0.80 kcal/mol], and the equilibrium distance r_{min} , if predicted at all, is shifted out, in one case by as much as 0.6 Å. BLYP (MAE 0.95 kcal/mol) and BP (MAE 1.10 kcal/mol) fail to predict minima in all cases apart from the dipole-dipole-interaction-dominated formaldehyde dimer; PBE shows qualitative binding, albeit at the wrong r_{min} (MAE: 0.35 kcal/mol, 0.22 Å).



	H ₂ -C ₆ H ₆		(H ₂ CO) ₂		Ar-N ₂ L		Ar-N ₂ T	
	E_{\min}	r_{\min}	E_{\min}	r_{\min}	E_{\min}	r_{\min}	E_{\min}	r_{\min}
BLYP	0.47	2.90	-0.91	3.8	0.23	4.40	0.34	3.80
DCACP-BLYP	-1.06	2.90	-2.35	3.3	-0.15	4.40	-0.25	3.80
BP	0.56	2.75	-0.71	3.6	0.35	4.41	0.52	3.76
DCACP-BP	-1.24	2.75	-2.39	3.2	-0.11	4.41	-0.26	3.76
PBE	-0.48	3.05	-1.55	3.4	-0.12	4.45	-0.12	3.98
DCACP-PBE	-0.96	2.85	-2.06	3.3	-0.20	4.31	-0.24	3.80
CCSD(T) ^a	-0.95	2.75	-2.18	3.3	-0.23	4.26	-0.30	3.70

^a Reference: Ar-N₂ [106], H₂-C₆H₆ [103], H₂CO dimer calculated in this work with aug-cc-pVDZ basis set.

Table 4.7: Equilibrium distances (r_{\min} , Å) and interaction energies (E_{\min} , kcal/mol) of some weakly bound complexes. For repulsive interaction curves, the interaction energy is calculated at the equilibrium distance predicted by DCACP-GGAs.

The inclusion of DCACPs greatly improves the performance of DFT-GGAs on these complexes with only an additional CPU cost of no more than 10%. r_{\min} is predicted extremely well throughout the set, with deviations of less than 0.2 Å from the CCSD(T) values. Overall MAEs are reduced down to 0.08 Å and 0.11 kcal/mol for r_{\min} and E_{\min} , respectively. DCACP-PBE with a MAE of 0.05 kcal/mol stands out as the best among the three functionals tested (MAE: 0.10 and 0.16 kcal/mol for DCACP-BLYP and DCACP-BP, respectively).

As mentioned previously, the DCACP parameters depend heavily on the performance of the underlying GGA functionals. PBE provides some spurious interactions for dispersion-bound systems. DCACPs complementing PBE can thus be either attractive or repulsive depending on the element in question, making it possibly more system dependent and less transferable. On the other hand, for BP and BLYP in which spurious dispersion interactions are entirely absent, DCACPs consistently provide an attractive correction to the underlying repulsive functional, in line with the idea of a dispersion-motivated correction. Thus, we believe that the latter combinations are preferable over the occasional superior performance of DCACP-PBE due to the more clean-cut interpretations they offer. As a side note, the linear Ar-N₂ complex (labeled L in Table 4.7) shows the largest deviation of all test cases. Since r_{\min} for this complex lies at a relatively distant 4.3 Å, this might be attributed to the slight discrepancy between the midrange to long-range description of DCACP-GGA and the actual r^{-6} asymptotic behavior.

	cutoff [Ry]	E_{\min} [kcal/mol]	r_{\min} [Å]
GTH	150	-0.20	3.85
TM	70	-0.20	3.80
TM*	70	-0.20	3.80
Vanderbilt	40	-0.20	3.80

Table 4.8: A comparative study (N_2 dimer, DCACP-PBE) on DCACPs applied with GTH, TM, and Vanderbilt pseudopotentials. TM* denotes the numerical version of DCACP which is included as the f channel in the TM pseudopotential.

Besides being applicable to different exchange-correlation functionals, DCACPs are not restricted to be employed with the analytical format of GTH pseudopotentials [52] only. The constraint to the analytical format is only enforced during calibrations. Numerical pseudopotentials such as TM [51] or Vanderbilt [50] types can be (and have been) used alongside. The DCACP-PBE N_2 dimer is used as an illustration for this “pseudopotential transferability” and shows that the improvement brought about by DCACPs is equally good in all three cases with the usual wave function cutoff for each pseudopotential type (Table 4.8). In addition, a numerical version of DCACPs can easily be included as an extra channel (f channel in this case) in TM pseudopotentials (the procedure is described in the Appendix).

4.4 Conclusions

We present a library of DCACPs that are calibrated against CCSD(T) or full CI (H only) references and can be used in combination with the GGA functionals BLYP, BP, and PBE. The results indicate that without introducing any significant distortions on intramolecular geometries and electronic structures, the effects of London dispersion forces can be well described within DFT-GGAs with the DCACP approach. Furthermore, DCACPs display a strong transferability to systems other than the calibration ones, *i.e.*, once calibrated, DCACPs can be applied in various chemical environments without further tuning the parameters. In brief, the DCACP approach shows promising outcomes despite its empiricism, suggesting a more physical interpretation underlying this remarkable performance.

Chapter 5

Noncovalent Interactions in Nucleobase–Intercalator Complexes

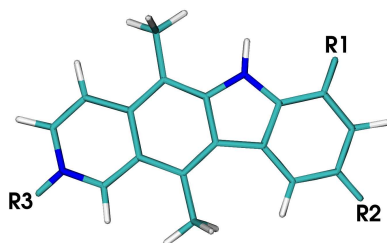
Abstract

Within the framework of Kohn–Sham density functional theory, interaction energies of hydrogen bonded and π – π stacked supramolecular complexes of aromatic heterocycles, nucleobase pairs, and nucleobases with the anti-cancer agent ellipticine as well as its derivatives are evaluated. Dispersion-corrected atom-centered potentials (DCACPs) are employed together with the BLYP functional. For all systems presented, the DCACP results are in very good agreement with available post Hartree–Fock quantum chemical results. Estimates of 3-body contributions (<15% of the respective interaction energy) and deformation energies (5-15% of the interaction energy) are given. Based on our results, we predict a strongly bound interaction energy profile for the ellipticine intercalation process with a stabilization of nearly 40 kcal/mol when fully intercalated (deformation energy not taken into account). The frontier orbitals of the intercalator–nucleobase complex and the corresponding non-intercalated nucleobases are investigated and show significant changes upon intercalation. The results not only offer chemical insights but also suggest that DCACPs can serve as an effective way to achieve higher accuracy in density functional theory without incurring a prohibitive computational cost, paving ways for realistic studies on biomolecular complexes in the condensed phase.

5.1 Introduction

One of the goals in modern biology is to understand biological phenomena at the molecular level, which involves studying structures of biomolecules and their functions. Noncovalent interactions between aromatic molecules are believed to contribute significantly to the stability and conformational variability of many biomacromolecules. In particular, π - π interactions play a key role in assembling many biologically important architectures such as DNA and RNA. These interactions not only influence the structure and dynamics of nucleic acids but also their interactions with polycyclic aromatic molecules whose ability to intercalate between base pairs of DNA has attracted much attention owing to the clinical success of many intercalators in anti-tumor chemotherapy [107–109]. Detailed knowledge of π - π interactions may prove invaluable in designing novel DNA–intercalation drugs.

The macromolecular effects of intercalator–DNA interactions, such as the unwinding and the lengthening of DNA, have been studied extensively by experiments [110–114]. These provide only limited information on the nature of this association at the atomistic level [115]. On the other hand, computer simulations can give atomistic insights into DNA-sequence specific interactions, binding selectivity, and the role of solvents as well as counter-ions in the intercalation process without additional assumptions.



Intercalator	Charge	R1	R2	R3
E	0	H	H	-
9HE	0	OH	H	CH ₃
E ⁺	+1	H	H	H
9AE ⁺	+1	NH ₂	H	H

Table 5.1: Structure of ellipticine and its derivatives.

For this study, the alkaloid ellipticine and its derivatives (Table 5.1) are chosen. Ellipticine is isolated from *Ochrosia elliptica* [116], and many of its more soluble derivatives yield promising results for cancer treatments [117]. NMR studies have shown intercalation to be a DNA binding mode for ellipticine [118]. Further searches for ellipticine-derived drugs are likely to profit from a detailed atomistic knowledge of their binding mechanism.

For simple heterocycles or large polarizable aromatic polycyclic compounds such as ellipticine and nucleobases, London dispersion forces constitute one of the major stabilizing components for their supramolecular complexation [119–122]. Unfortunately, a description of these forces requires an accurate treatment of electron correlation effects; high-level correlated *ab initio* methods such as coupled-cluster theory with large basis sets or quantum Monte Carlo allow for an accurate treatment, but they are not applicable for all but the smallest systems. The tractable size of aromatic heterocycle complexes has prompted studies using MP2 and CCSD(T) methods [24, 72, 123–127]. For larger systems such as stacked DNA base pairs, interaction energies have been computed with *ab initio* methods, albeit to our knowledge, CCSD(T) calculations with large basis set have yet to be attempted. Instead, MP2 calculations extrapolated to the complete basis set (CBS) limit are augmented with a Δ CCSD(T) correction computed in a smaller basis, which systematically corrects the over-binding nature of MP2 for π - π interactions, to estimate the CBS CCSD(T) results [25, 122, 123, 125, 128–135]. Still, this approach is computationally infeasible for biologically more relevant studies such as the intercalator–nucleobase complex investigated here.

The Kohn–Sham (KS) formalism of density functional theory (DFT) [7, 8], in principle, is exact and should correctly describe dispersion forces if the true exchange–correlation functional were known. Many popular exchange–correlation functionals, however, are unsuitable for a proper treatment of these forces [12, 13, 22–26, 28, 136, 137]. New methods for efficient calculations of dispersion forces in DFT have thus been the focus of many recent works. Approaches based on electron density partitioning, involving the assignment of fragments, have been used to describe rare gas and small hydrocarbon complexes [59–61]. The introduction of non-local correlations into a van der Waals (vdW) functional has shown promising results on systems such as rare-gas dimers, aromatic ring complexes, and graphite [29, 62–67]. In addition, a long-range correction scheme combined with a vdW functional has been successfully applied to small vdW complexes, benzene dimers, and naphthalene dimers [68, 69]. Recently, a post Hartree–Fock model has been proposed in which the instantaneous dipole moment of the exchange hole is used to generate the dispersion coefficients [33, 73, 74]. This model has predicted the geometries and binding energies for a large test set of intermolecular complexes remarkably well with the notable exception of the π -stacked benzene dimer. An alternative approach involves optimizing a functional for better descriptions of non-bonded interactions, for example, PWB6K [138] and M05-

2X [70]. A combination of a DFT description of the interacting monomers with a symmetry-adapted perturbation theory (SAPT) treatment of inter-monomer interactions (DFT-SAPT) [71, 139] and the introduction of the density-fitting approximation has made SAPT calculations with extended basis sets on medium-sized systems possible [140]. A more pragmatic solution is to explicitly include empirical pairwise inter-atomic dispersion terms. This scheme generally leads to reasonable stabilization energies and nuclear forces [16, 23, 30, 72], but it involves defining the appropriate damping functions and the parameterization of the C_6 coefficients. Furthermore, the electronic structure is left uncorrected.

The use of dispersion-corrected atom-centered potentials (DCACPs) provides an alternative approach to include dispersion forces in a fully self-consistent manner within the framework of KS-DFT. DCACPs represent corrections to standard exchange-correlation functionals as non-local angular-momentum-dependent atom-centered electronic potentials [34]. So far, DCACPs have been successfully applied to small vdW clusters [76, 89, 90], crystals of graphite and benzene [90], the adsorption of argon on graphite, liquid crystals [96], and a large set of biomolecules including more than 100 nucleobase and amino acid complexes [92].

Here, we study complexes of weakly bound heterocycles, nucleobase pairs, and intercalator–nucleobase complexes using DCACPs which have been calibrated against CCSD(T) (with large basis sets) or full CI results [76]. By including systems ranging from relatively simple molecules to intercalator–nucleobase complexes, we aim to both validate and challenge the DCACP concept in a systematic way using realistic applications. In particular, much attention is paid to the electronic and structural modifications of the monomers upon binding. We begin by studying the interaction energy of three model heterocyclic complexes: (furan)₂, (pyridine)₂, and (pyrimidine)₂. We then evaluate the stabilization energy of few selected stacked nucleobase pairs: guanine···cytosine (G···C), adenine···thymine (A···T), uracil···uracil (U···U), and cytosine···cytosine (C···C) for which high-level reference data is available. The performance of DCACPs on H-bonding is assessed by considering two Watson–Crick (WC) base pairs [141]: adenine···thymine (A···T WC) and guanine···cytosine (G···C WC). Interaction energies of the neutral and charged complexes formed between the ellipticine derivative and the WC base pair (abbreviated as ellip–WC from now on): GC–9HE, GC–E⁺, GC–9AE⁺, GC–E, AT–E⁺, and AT–E [defined in Table 5.1 and Figure 5.1(a)] are evaluated. Lastly, an *in vacuo* model for the DNA intercalation process is investigated. The model is based on an available crystal structure [142] and con-

sists of an ellipticine molecule intercalated between two WC base pairs that are connected by phosphate backbone groups [ellipticine-d(CG)₂ complex].

5.2 Computational Details

All DFT calculations have been carried out using the program CPMD [93], the BLYP functional [17, 18], pseudopotentials of Troullier-Martins type [51], and the Poisson solver implemented in CPMD according to Ref. [100]. DCACPs calibrated against references at CCSD(T) or full CI level have been taken from Ref. [76] in which the fitting procedure and the dependency of the correction on the employed functional are discussed. For the larger complex, ellipticine-d(CG)₂, we have employed a plane wave cutoff of 70 Ry and a gradient tolerance of 1.5×10^{-3} a.u. per nuclear degree of freedom for geometry optimizations. For all other systems we have used a plane wave cutoff of 75 Ry and a tolerance of 5×10^{-4} a.u. for geometry optimizations.

The interaction energy between moieties X and Y is most commonly defined as the difference between the total energy of the fully relaxed complex (E_{XY}^{opt}) and the fully optimized isolated moieties ($E_X^{\text{opt}}, E_Y^{\text{opt}}$),

$$\Delta E_{X-Y}^{\text{opt}} = E_{X-Y}^{\text{opt}} - E_X^{\text{opt}} - E_Y^{\text{opt}}. \quad (5.1)$$

E^{opt} thus takes the intramolecular deformation energy into account. Nevertheless, to be consistent with previously published results, two other definitions for the interaction energy, when explicitly mentioned, are also used:

$$\Delta E_{X-Y}^{\text{fix}} = E_{X-Y}^{\text{fix}} - E_X^{\text{opt}} - E_Y^{\text{opt}}, \quad (5.2)$$

$$\Delta E_{X-Y}^{\text{exp}} = E_{X-Y}^{\text{exp}} - E_X^{\text{fix}} - E_Y^{\text{fix}}. \quad (5.3)$$

The superscript ‘opt’ is to stress that the energy is calculated with the moiety fully relaxed, ‘exp’ denotes the structure is taken from some known experimental structures, and ‘fix’ means that the complex/monomer is constructed/taken from some defined geometry and not relaxed [in eq 5.2, this means the complex X-Y is constructed from the optimized monomer (X, Y) geometry and the only variable is the intermolecular distance; in eq 5.3, the monomer (X, Y) is fixed at the geometry assumed in the experimental complex].

Another quantity that can shed light on the physical nature of the stacking interaction is the deformation energy of the complex components. The deformation energy for a moiety (E_{def}) is defined as the energy difference between two conformations: one corresponding to the isolated moiety's optimized conformation and the other corresponding to the conformation assumed in the relaxed complex. Large E_{def} is usually an indication of interactions other than the London type such as dipole-dipole or H-bonding interactions. In addition, for complexes consisting of an ellipticine derivative and one WC base pair (ellip–WC), the 3-body term $\Delta^{(3)}$ gives an indication on the importance of the cooperative effects as a result of, say, charge transfer or polarization upon complexation.

Applying a new method on new ground requires validation. We rely, whenever possible, on previously published high-level *ab initio* results as our benchmark. For the relatively small aromatic heterocycle complexes, this does not pose a problem. Yet there are only few high-level calculations on stacked DNA base pairs and none involves intercalators. In general, MP2 calculations overestimate π – π interactions with respect to the CCSD(T) values. Owing to error cancellation, results from MP2 calculations using a modified medium-sized basis set [6-31G*(0.25)] where the standard *d*-polarization functions were replaced by more diffuse ones are found to be much closer to the CCSD(T) values than the ones with a large basis set. For example, MP2/6-31G*(0.25) calculations overestimate the equilibrium interaction energy of the sandwich benzene dimer by only 8% compared with the CCSD(T)/aug-cc-pVQZ result whereas MP2/aug-cc-pVQZ overestimates by 97% [24, 98]. In light of these observations, for systems where no higher reference value is available from literature, MP2/6-31G*(0.25) calculations have been carried out using the GAUSSIAN 03 [102] package. These values, however, should not be considered as high quality benchmark but more as guiding values for qualitative trends.

Interaction energies ΔE^{fix} for dimers of benzene, pyrimidine, pyridine, and furan have been computed at various intermolecular distances r with fixed DCACP-BLYP optimized monomer geometries, as shown in Table 5.2.

Geometry optimizations of the stacked and H-bonded A···T and G···C base pairs and their isolated moieties have been carried out to evaluate the interaction energy ΔE^{opt} . Energies of the isolated moieties at the geometry assumed in the complex have also been evaluated for the calculations of E_{def} . For U···U and C···C stacked base pairs, interaction energies ΔE^{exp} have been computed using the geometry from Ref. [135] (UUst, CCst).

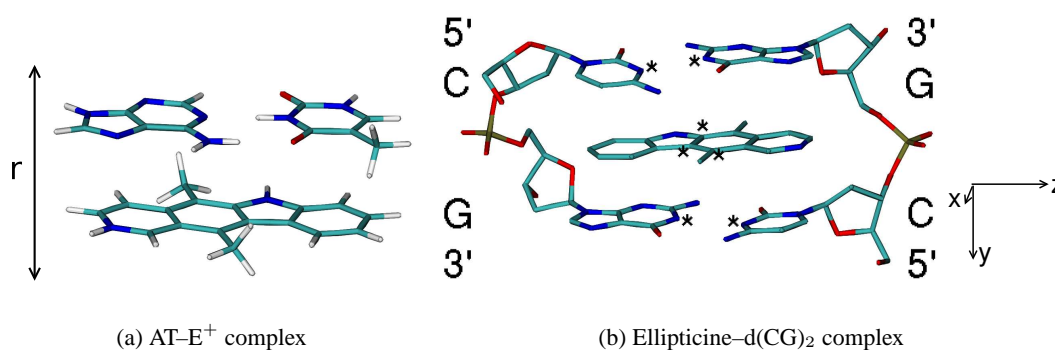


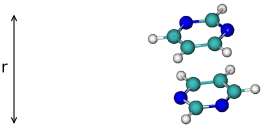
Figure 5.1: Graphical representation of the simulation set-up. (a) Configuration of the AT-E⁺ complex. (b) Structure of the ellipticine-d(CG)₂ complex; the x -coordinate of the atoms marked with an asterisk * is constrained during the simulations. H atoms are omitted for clarity in (b).

The ellipticine and its derivatives are described in Table 5.1. The intercalator and the WC base pair are arranged in a coplanar fashion with a vertical separation r [the AT-E⁺ complex is shown in Figure 5.1(a) as an example]. Single point energy calculations have been carried out for $r = 3.2, 3.4, 3.5,$ and 3.6 \AA ; the bases and the intercalator moieties are kept rigid at the optimized B3LYP/6-31G* geometries to obtain the interaction energy ΔE^{fix} .

The interaction energies E_{int} of the ellipticine-d(CG)₂ complex [Figure 5.1(b)] have been evaluated by defining the intercalator (ellipticine) as one moiety and the 4-nucleobase complex as the other. As we are mostly interested in computing energies arising from the stacking interaction, we define, for the present system, $E_{\text{int}} = E_{\text{whole complex}}^{\text{opt}} - E_{\text{ellipticine}}^{\text{fix}} - E_{\text{4-nucleobase complex}}^{\text{fix}}$. An inter-moiety displacement coordinate Δx defines the intercalation coordinate according to the axis frame shown in Figure 5.1(b). Constraints on the key atoms involved in the Δx displacement are imposed to preserve the parallel displacement throughout the simulations. Configurations with Δx ranging from 0 up to 5 Å have been prepared in 1 Å step increments using the bis-intercalated hexanucleotide crystal structure from Ref. [142] as the initial structure for $\Delta x = 0 \text{ \AA}$. Since the ellipticine-d(CG)₂ complex we adopted is situated at the extremity of the crystal, to closely mimic its non-intercalated state, we have chosen the 5'-d(CpG)-3' segment of a B-DNA decamer, 5'-d(CpGpApTpTpApApTpCpG)-3' [143], as the initial structure of the intercalator-free 4-nucleobase complex. Simulated annealing has been employed in order to obtain the optimized structure for all configurations.

5.3 Results and Discussion

5.3.1 Heterocycles



	(Pyrimidine) ₂			(Pyridine) ₂			(Furan) ₂			(Benzene) ₂		
r [Å]	3.4	3.6	3.7	3.7	3.8	3.9	3.8	4.0	4.2	3.8	3.9	4.0
BLYP	4.03	2.22	1.66	4.23	3.52	2.97	2.77	1.99	1.51	3.84	3.23	2.74
DCACP	-2.56	-3.35	-3.42*	-1.33	-1.53	-1.58*	-0.98	-1.03*	-0.87	-1.61	-1.72*	-1.71
MP2	-3.66	-3.92*	-3.83	-1.81	-1.84*	-1.78	-0.76	-0.80*	-0.71	-1.83*	-1.82	-1.74
CCSD(T)	-2.03	-2.64*	—	—	—	—	—	—	—	-1.66	-1.70*	-1.67

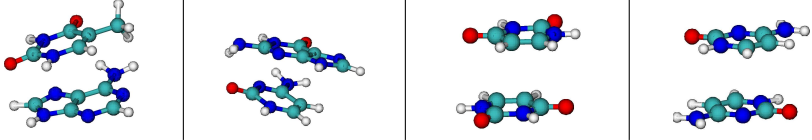
Table 5.2: Interaction energies (ΔE^{fix} , kcal/mol) of heterocycle and benzene dimers at various vertical separations r . MP2 are MP2/6-31G*(0.25) results calculated in this work, apart from the pyrimidine dimer whose values have been taken from Ref. [123]. CCSD(T) are CCSD(T) calculations with either diffuse cc-pVDZ (pyrimidine dimer [123]) or estimated aug-cc-pVQZ (benzene dimer [98]) basis sets. The asterisk * denotes the equilibrium ΔE^{fix} .

Table 5.2 summarizes the interaction energies of heterocycle dimers at various vertical separations. Our results show that the uncorrected BLYP functional predicts no interaction minimum for any of the investigated pairs, failing to describe π - π interactions altogether. In contrast, results from the DCACP-augmented BLYP functional (DCACP-BLYP) not only show attractive interactions between all stacked dimers but also reproduce the trend in the MP2 binding energies when going from pyrimidine to pyridine to furan.

The DCACP-BLYP results agree to within 0.5 kcal/mol with the MP2/6-31G*(0.25) values for the equilibrium ΔE^{fix} . For N-heterocycles, the DCACP-BLYP equilibrium ΔE^{fix} are underestimated by no more than 15% and the equilibrium distances are shifted outwards by 0.1 Å with respect to the MP2/6-31G*(0.25) results. Since the potential energy surface around the minimum is very shallow, a discrepancy of only 0.1 Å is doubtfully relevant. On the other hand, DCACP-BLYP overestimates the equilibrium ΔE^{fix} in furan, O-heterocycle, by 0.2 kcal/mol (approximately 30%) but reproduces the MP2/6-31G*(0.25) equilibrium distance perfectly.

5.3.2 Base pairs

For the stacked and H-bonded base pairs, the DCACP-BLYP results are in good agreement with the reference data [MP2 calculations in CBS limit augmented with $\Delta\text{CCSD(T)}$ corrections]. Results from three



	A...T		G...C		U...U st		C...C st	
	ΔE^{opt}	Δ_{atom}	ΔE^{opt}	Δ_{atom}	ΔE^{exp}	Δ_{atom}	ΔE^{exp}	Δ_{atom}
BLYP	1.2	-0.43	-7.4	-0.33	1.8	-0.39	1.0	-0.42
DCACP	-9.5	-0.07	-14.6	-0.08	-6.3	-0.05	-7.9	-0.08
BLYP-D [144]	-11.2	-0.01	-15.6	-0.04	-7.2	-0.01	-9.1	-0.03
PWB6K [145]	-9.5	-0.07	-14.9	-0.07	-	-	-	-
SAPT [140]	-10.9	-0.02	-17.8	+0.03	-	-	-	-
Ref. [135]	-11.6	0.00	-16.9	0.00	-7.5	0.00	-10.0	0.00

$$\Delta_{\text{atom}} = (\Delta E_{\text{Ref.}}^{\text{opt}} - \Delta E_{\text{DFT}}^{\text{opt}}) / N_{\text{atom}}$$

Table 5.3: Interaction energies (ΔE , kcal/mol) of the stacked nucleobase pairs. Two ΔE definitions are used: optimized (ΔE^{opt}) and experimental (ΔE^{exp}) complex geometry; the latter is not corrected for the deformation energy. BSSEs are not considered in the cited BLYP-D and PWB6K values while deformation energy is not taken into account in the SAPT results. References are MP2 calculations in CBS limit with a $\Delta\text{CCSD(T)}$ correction.

other approaches are also included for comparison: BLYP augmented with the empirical dispersion corrections [BLYP-D, with TZV(2d,2p) basis] [144], hybrid meta GGA functional PWB6K [6-31+G(d,p) basis] [145], and DFT-SAPT (CBS limit, abbreviated as SAPT from now) [140]. One should bear in mind that PWB6K and BLYP-D results are not counterpoise-corrected [101] for the basis set superposition errors (BSSEs) and once corrected should lead to smaller binding energies. As a reference, the BSSEs with PWB6K/6-31+G(d,p) for the stacked and the H-bonded G...C base pairs are estimated to be 1.6 and 1.1 kcal/mol, respectively [145]. On the other hand, the CBS SAPT values do not take deformation energies into account.

The interaction energies ΔE of the stacked base pairs are presented in Table 6.3. As expected, the BLYP functional leads to grossly underestimated, sometimes even repulsive, ΔE for all stacked base pairs investigated; the mean deviation per atom (Δ_{atom}) is -0.39 kcal/mol. Compared with the reference data, DCACP-BLYP systematically underestimates ΔE with Δ_{atom} of roughly -0.07 kcal/mol and a deviation of approximately 20% or lower from the benchmark. These results demonstrate the notable improvement obtained with the DCACP approach. The DCACP-BLYP performance is comparable to BLYP-D (where electronic structure remains uncorrected) and the two more complex but computationally more demanding methods: PWB6K and SAPT.

The DCACP-BLYP ΔE per atom ΔE_{atom} is approximately -0.3 kcal/mol for all stacked base pairs

	A	T	G	C
μ^a	2.36	4.37	6.55	6.43
μ^b	2.49	4.39	6.68	6.50

Table 5.4: Overall dipole moment (Debye) of the nucleobase in its optimized geometry (μ^a) and the geometry observed in the stacked complex (μ^b). The values obtained with uncorrected and DCACP-augmented BLYP functionals are the same.

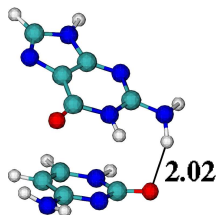
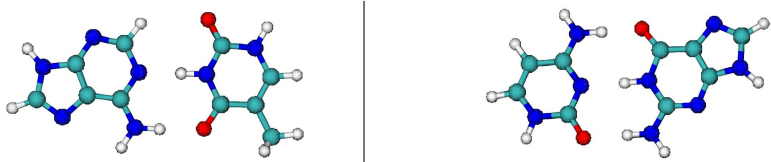


Figure 5.2: Detail of the optimized geometry of the G···C stacked base pair. The distance (Å) between the acceptor and the donor of the H-bond is indicated.

here apart from the G···C pair (-0.5 kcal/mol). Furthermore, the G···C base pair is the only one which shows non-negligible binding with the uncorrected BLYP functional, suggesting that the larger ΔE_{atom} is due to interactions other than dispersion forces. The stacking interaction of nucleobases involves a combination of dispersion forces and permanent multipole–multipole interactions. The G···C pair has the largest dipole–dipole interaction among all studied pairs due to the large dipole moment of the respective bases (Table 5.4). Two points are worth noting: first, the formation of stacked base pairs increases the dipole moment of the individual moiety only by a very small amount, the largest increase is 5%; second, the uncorrected and the DCACP-augmented BLYP functionals produce the same dipole moments. The latter has been observed previously for a benzene–argon complex, indicating that static electronic quantities change only slightly upon including DCACPs [34]. This observation is not surprising considering that DCACPs constitute very weak potentials compared with the atomic pseudopotentials. In addition to the strong dipole–dipole interactions, H-bonding (Figure 5.2, N-H···O) originating from a tilting of the two monomers, G and C, is also likely to contribute. This H-bond has a H···O separation of 2.0 Å, close to the typical H-bond length in water (1.97 Å) and an acceptable directionality ($\angle_{\text{NH}\cdots\text{O}} = 150^\circ$). The other complexes also have potential H-bond sites; however, the distances between the corresponding hydrogen and the H-acceptor atoms are all larger than 2.7 Å, clearly excluding H-bonding.

The deformation of monomers upon binding depends on the subtle balance between intra- and

inter-molecular forces, and it is intimately related to the cluster’s vibrational spectra, polarizability, and charge-transfer properties [131]. The DCACP-BLYP deformation energy E_{def} is small: 0.6 and 1.5 kcal/mol for the stacked $A \cdots T$ and $G \cdots C$ complexes, respectively (*i.e.*, less than 10% of the corresponding interaction energy). For the DCACP-BLYP geometry-optimized stacked base pairs, the NH_2 group enhances its pyramidal shape compared with the isolated monomer, owing to the polar attraction between one of the hydrogen atoms of the amino group and the closest electronegative atom of the adjacent base. In particular, the $G \cdots C$ complex exhibits the largest deformation energy among all stacked base pairs due to the formation of the H-bond previously mentioned (Figure 5.2).



	A \cdots T WC				G \cdots C WC				
	ΔE^{opt}	Δ_{atom}	$r_{\text{N-N}}$	$r_{\text{N-O}}$	ΔE^{opt}	Δ_{atom}	$r_{\text{N-O}}$	$r_{\text{N-N}}$	$r_{\text{O-N}}$
BLYP	-10.9	-0.15	2.86	2.90	-22.6	-0.21	2.95	2.96	2.81
DCACP	-15.4	+0.00	2.86	2.90	-27.7	-0.04	2.93	2.93	2.79
BLYP-D [144]	-15.5	+0.00	2.85	2.94	-28.0	-0.03	2.95	2.94	2.79
PWB6K [145]	-14.2	-0.04	-	-	-28.4	-0.01	2.91	2.94	2.80
SAPT [140]	-15.7	+0.01	-	-	-30.5	+0.06	-	-	-
Ref. [135]	-15.4	0.00	2.86	2.94	-28.8	0.00	2.94	2.93	2.79

$\Delta_{\text{atom}} = (\Delta E_{\text{Ref.}}^{\text{opt}} - \Delta E_{\text{DFT}}^{\text{opt}}) / N_{\text{atom}}$.

Table 5.5: Interaction energies (ΔE^{opt} , kcal/mol) and H-bond lengths (\AA) of the WC base pairs. BSSEs are not taken into account in the quoted BLYP-D and PWB6K results while the SAPT result is not corrected for E_{def} . References are MP2 calculations in CBS limit with $\Delta\text{CCSD(T)}$ corrections.

The results for the WC base pairs are presented in Table 5.5. BLYP describes H-bonding interactions reasonably well, in great contrast to its complete failure in treating π - π interactions. The DCACP approach further improves these estimates: Δ_{atom} , the mean deviation from the reference value per atom, is decreased to 0.02 kcal/mol from BLYP’s 0.18 kcal/mol, making it one of the most accurate methods out of the five discussed here. In addition, the DCACP-BLYP-estimated H-bond lengths are closer to the references than the corresponding BLYP values, providing evidence that the inclusion of DCACPs also improves the description of H-bonding.

The DCACP-BLYP deformation energy E_{def} for the WC base pairs is much larger than in the stacked cases: 1.6 and 3.5 kcal/mol for the $A \cdots T$ WC and the $G \cdots C$ WC base pairs, respectively. These compare well with the respective estimated CCSD(T) values [MP2 augmented with $\Delta\text{CCSD(T)}$ corrections]

of 1.5 and 3.6 kcal/mol [131] and are slightly improved from the already reasonable BLYP values of 1.7 and 3.3 kcal/mol. Other functionals predict E_{def} much worse: for the G···C pair, PWB6K and PW91 yield 2.9 [138] and 5.4 kcal/mol, respectively.

5.3.3 Intercalator–DNA complexes

As mentioned before, high-level correlated calculations for systems larger than base pairs are not available and the quoted MP2/6-31G*(0.25) values should be considered as guiding values for qualitative trend only. Direct comparisons between the MP2/6-31G*(0.25) and the DCACP-BLYP results have to be made with care as DCACPs have been calibrated against CCSD(T) (full CI for hydrogen) references. The study on complexes of one ellipticine derivative stacked with one WC base pair (ellip–WC) involves not only larger systems but also complexes containing both charged and neutral species, showing DCACPs’ transferability across not only hybridization but also charge states.

		GC–9HE	AT–9HE	GC–E ⁺	AT–E ⁺	GC–9AE ⁺
r_{min}	DCACP	3.5	3.5	3.5	3.5	3.5
	MP2	3.4	3.4	3.4	3.3	3.3
ΔE^{fix}	DCACP	–14.89	–14.62	–16.67	–19.13	–21.18
	BLYP	8.31	8.67	6.35	3.82	2.37
	MP2	–18.63	–18.38	–21.82	–24.52	–26.82
$\Delta^{(3)}$	DCACP	0.88	0.24	1.77	0.49	0.35
Δ_{atom}		–0.06	–0.06	–0.08	–0.08	–0.09
$\Delta_{\text{atom}} = (\Delta E_{\text{MP2}}^{\text{fix}} - \Delta E_{\text{DCACP}}^{\text{fix}}) / N_{\text{atom}}$						

Table 5.6: Equilibrium distances (r_{min} , Å) and interaction energies (ΔE^{fix} , kcal/mol) of the ellip–WC complexes. MP2 values are calculated with 6-31G*(0.25) basis set [146]. The 3-body term ($\Delta^{(3)}$, kcal/mol) is calculated at the equilibrium distance with the DCACP-augmented BLYP functional.

Table 5.6 lists the DCACP-BLYP equilibrium distances r_{min} and interaction energies ΔE^{fix} of the ellip–WC complexes. Results for the neutral and the positively charged complexes are both consistently underestimated by roughly 20% with respect to MP2/6-31G*(0.25). The equilibrium distances, on the other hand, are in good agreement with MP2/6-31G*(0.25), slightly larger by ~ 0.1 Å on average. The 3-body term $\Delta^{(3)}$ gives information on the non-additivity of pair interactions. In all cases investigated here, $\Delta^{(3)}$ is less than 15% of the corresponding ΔE^{fix} , the largest 3-body contribution is found in the GC–E⁺ complex whose $\Delta^{(3)}$ is roughly 11% of ΔE^{fix} .

A detailed atomistic investigation of a biological system often requires the knowledge of the electronic structure. One attractive feature of *ab initio* calculations is that one can investigate the effects of complexation on the electronic structure, a phenomenon which is inaccessible to classical force field simulations. The electron density of the AT-E⁺ and AT-E complexes is analyzed further. Upon protonation, a new σ bond is formed resulting in an increase in the in-plane electron density. To compensate for this change, the electron density in the π system of protonated ellipticine is slightly depleted.

Complex	GC-9AE ⁺	AT-E ⁺	GC-E ⁺	AT-E	GC-E
$\Delta E_{\text{complex}}^{\text{opt}}$	-25.02	-21.47	-20.47	-15.48	-16.56
$\Delta E_{\text{base}}^{\text{fix}}$	-27.23	-14.37	-27.18	-15.01	-27.97
μ_{base}	6.76	2.31	6.15	1.70	5.92

Table 5.7: Interaction energy of the fully relaxed ellip-WC complexes ($\Delta E_{\text{complex}}^{\text{opt}}$, kcal/mol) and the H-bonding energy of the WC base pair in the geometry-optimized ellip-WC complexes ($\Delta E_{\text{base}}^{\text{fix}}$, kcal/mol). The latter should be compared with the corresponding values in Table 5.5. The overall dipole moment (μ_{base} , Debye) of the WC base pair is also included. ‘In’ denotes the corresponding intercalator.

Table 5.7 summarize the results of DCACP-BLYP geometry optimizations on the five selected ellip-WC complexes. The interaction energy $\Delta E_{\text{complex}}^{\text{opt}}$ follows the same trend as observed in calculations using rigid monomer geometries (ΔE^{fix} , Table 5.6), but the values are 10–20% larger. The contributions to the interaction energy can be roughly separated into dispersion, multipole-multipole interactions, H-bonding, and for the charged complexes, multipole-charge interactions. The deformation of the planar WC base pair leads to slightly weaker H-bonds: the H-bond strength $\Delta E_{\text{base}}^{\text{fix}}$, on average, is roughly 1.0 kcal/mol weaker in the significantly deformed charged ellip-WC complexes than the corresponding isolated WC base pairs. In the neutral complexes it is only weaker by roughly 0.3 kcal/mol due to the largely preserved planar structures. The loss of H-bonding, however, is more than compensated for by other favorable interactions introduced upon deformation. An increase in the z -component (z -axis as shown in Table 5.7) of the dipole moment results in a stronger dipole-charge as well as dipole-dipole interactions in the charged complexes (Table 5.7). For comparison, the dipole moments of the planar A···T WC and G···C WC base pairs are 1.42 and 5.87 Debye, respectively. It is worth noting that the GC-E complex is more stable than the AT-E complex by a small margin (0.8 kcal/mol), supporting

the observation that ellipticine has a preference to intercalate between the d(GpC)₂ over the d(ApT)₂ stacked pairs [142].

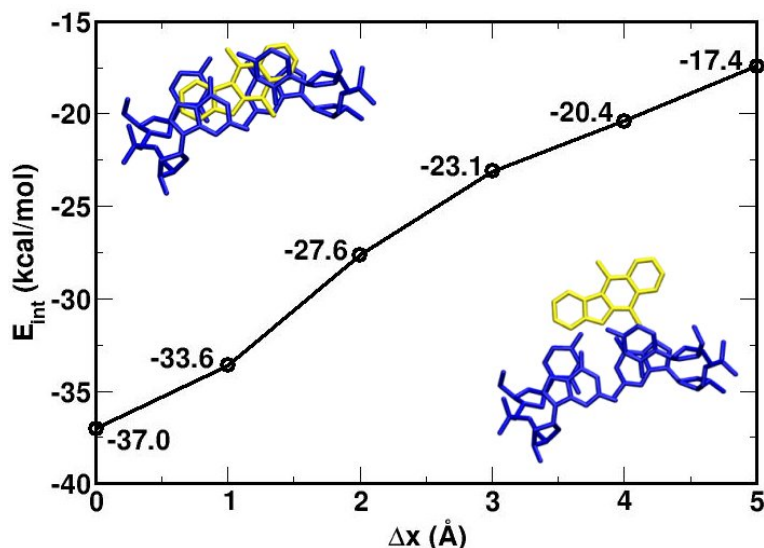


Figure 5.3: Interaction energy profile of the intercalation process in the ellipticine–d(CG)₂ complex as a function of the inter-moiety displacement Δx . The interaction energy (E_{int} , kcal/mol) is quoted in the plot. The upper left and lower right insets correspond to the $\Delta x = 0 \text{ \AA}$ and $\Delta x = 5 \text{ \AA}$ configurations, respectively; H atoms are omitted for clarity.

Figure 5.3 summarizes the interaction energy of the intercalation process evaluated at different inter-moiety displacements Δx in increments of 1 \AA . The maximal interaction is, as expected, found at the fully inserted configuration $\Delta x = 0 \text{ \AA}$, and a large well depth, defined as $E_{\text{int}}(\Delta x = 5) - E_{\text{int}}(\Delta x = 0)$, of roughly 20 kcal/mol is observed. At $\Delta x = 5 \text{ \AA}$, E_{int} is still very attractive. The DCACP-BLYP E_{int} has been calculated with the isolated moieties assuming the same geometries as found in the respective optimized ellipticine–d(CG)₂ complexes. In accommodating the ellipticine molecule, DNA is known to unwind and lengthen [142]; therefore, one can expect a large deformation energy upon intercalation. Geometry optimization on the corresponding intercalator-free 4-nucleobase complex, however, shows little change in energy compared with the ones calculated using geometries taken from the optimized ellipticine–d(CG)₂ complexes, indicating the presence of a local minimum with large inter-base separation. This data is supported by the fact that base pairs found at the ends of a DNA segment tend to be more distorted with little π – π interaction. Our results show that the intercalation process is energetically favorable and that intercalating ellipticine at such position should not cause a large initial loss of π – π stacking. This may serve as an alternative explanation as to why ellipticine prefers the d(CpG)₂ inter-

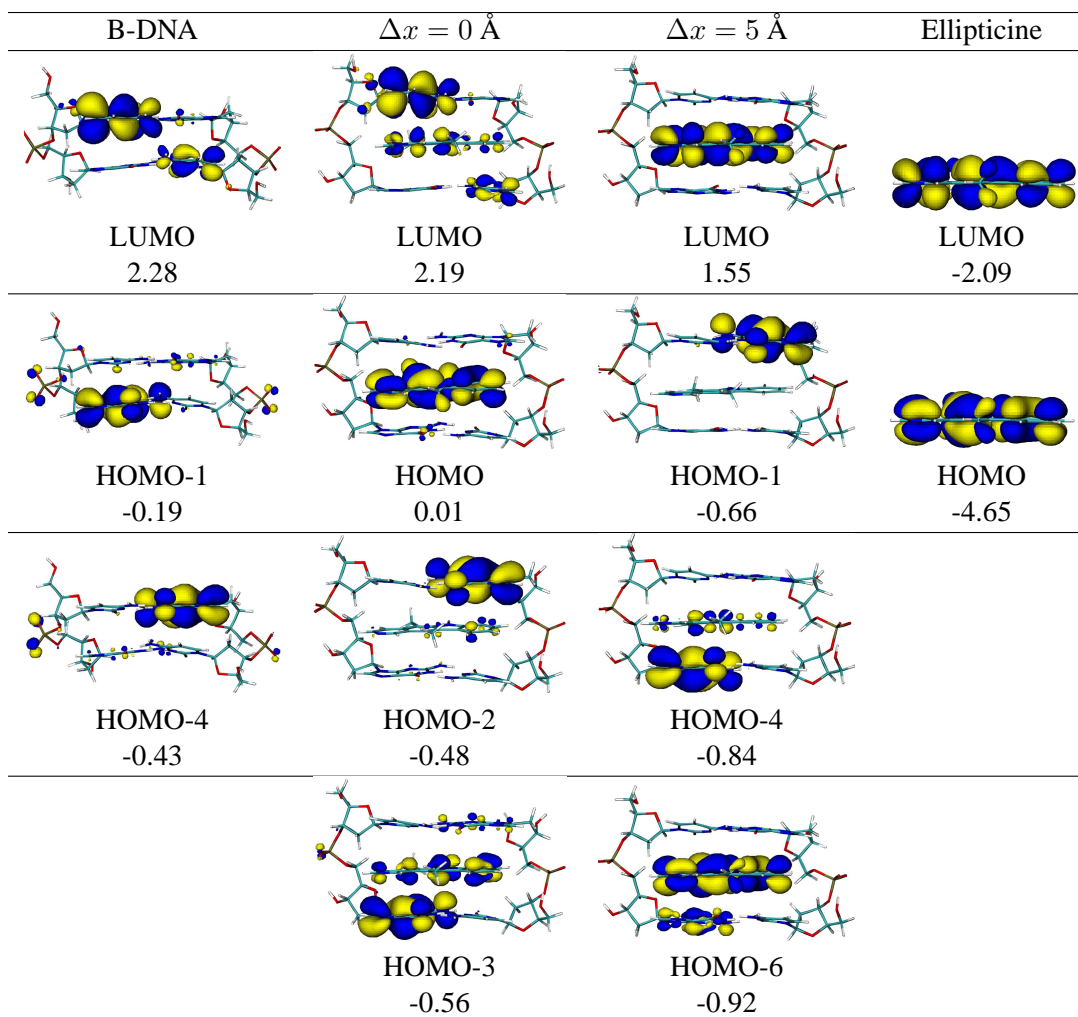


Figure 5.4: Graphical representation of the KS orbitals in the ellipticine–d(CG)₂ complex. Only orbitals localized on the base or ellipticine are depicted. The corresponding energy (eV) is listed beneath.

calation site as reported in Ref. [142]. The binding site preference of ellipticine observed in Ref. [142] could be more related to the sequence position than to the nature of the base. The use of larger DNA fragments or a fragment where d(CpG)₂ is not located at the extremities may be necessary for a proper evaluation of the sequence-dependent binding preference for intercalators. To evaluate the effect of DCACPs, single-point energy calculations with the DCACP-BLYP optimized geometries for $\Delta x = 0 \text{ \AA}$ and $\Delta x = 5 \text{ \AA}$ have been carried out using the BLYP functional alone. For $\Delta x = 0 \text{ \AA}$, the resulting interaction energy is repulsive (18.4 kcal/mol) whereas for $\Delta x = 5 \text{ \AA}$ the complex is slightly stable (-3.3 kcal/mol). The use of DCACPs is thus fundamental to obtain even a qualitatively correct picture of ellipticine's intercalation properties.

To complement the study on the effects of ellipticine binding on DNA's electronic structure, few

selected KS orbitals of interest and their energy are presented in Figure 5.4. Since the orbitals on the phosphate groups do not mix with the bases nor the intercalator, we choose to present orbitals localized on the bases or the intercalator only. Upon intercalation, the characteristics of the frontier orbitals change significantly: the highest occupied molecular orbital (HOMO) of the fully intercalated complex is now localized on the intercalator, ellipticine, and the lowest unoccupied molecular orbital (LUMO) has the same characteristics as in the non-intercalated B-DNA but now with a small mixing (binding in nature) to the LUMO of ellipticine. Furthermore, the energy of the ellipticine HOMO has risen by almost 5 eV upon intercalation. These changes in the electronic structure should have a pronounced influence on the electron/charge transfer properties of DNA, which is of great interest for DNA wire and DNA repair studies [147, 148], to name just a few. Nevertheless, it should be emphasized that the current model has been simulated at anhydrous condition with no counter-ions for the phosphate backbone. The results should thus be interpreted with caution and future studies based on more realistic models should be carried out to verify its validity.

5.4 Conclusions

We have carried out DFT DCACP-BLYP calculations to study interactions of aromatic heterocycle, nucleobase, and intercalator–nucleobase complexes. We show that the DCACP approach is able to describe π – π interactions remarkably well in all systems studied, including neutral as well as positively charged complexes. H-bonding is also well described; the binding geometries as well as interactions and the deformation energies are both clearly improved compared with DFT BLYP. In addition, even with the current implementation in which further optimization for computational efficiency is still possible, for systems studied in this work, the computational overhead is increased by no more than 30%.

The 3-body energy in all investigated cases is <15% of the respective interaction energy. Furthermore, H-bonds between the WC base pairs are deformed noticeably to achieve more favorable multipole–charge interactions in all charged ellip–WC complexes. The interaction energy of ellipticine–d(CG)₂ complex when fully intercalated is estimated to be approximately 40 kcal/mol. Whenever CCSD(T) references are available, the DCACP-BLYP estimates of interaction energies are closer to the CCSD(T) values than MP2 methods, which is consistent with the fact that all DCACPs used in this study apart from the hydrogen DCACP have been calibrated against CCSD(T) references. In short, the

availability of a library for most commonly encountered elements in biological systems [76] render the DCACP approach a practical way to achieve higher accuracy without incurring a prohibitive computational penalty. More realistic DFT-based quantum mechanical/molecular mechanical calculations [149] of intercalator–DNA complexes solvated in water or a protein environment can now be envisaged.

Chapter 6

Weak Interactions in Biomacromolecules

Abstract

Interaction energies of the biomolecules in the JSCH-2005 database are calculated with density functional theory using the exchange-correlation functional BLYP augmented with dispersion-corrected atom-centered potentials. The results are in excellent agreement with extrapolated CCSD(T) complete basis set limit references with unsigned mean errors of less than 1.6 kcal/mol. Geometry optimizations all reach stable configurations that are close to the MP2-optimized geometries.

6.1 Introduction

Weak interactions such as hydrogen bonding (H-bonding) and London dispersion forces are key interactions for biological systems; they are believed to contribute significantly to the stability as well as conformational variability of many biomacromolecules and their interactions with each other or the environment.

Unfortunately, a proper description of dispersion forces requires an accurate treatment of electron correlation effects; high-level *ab initio* methods such as coupled-cluster theory with large basis sets or quantum Monte Carlo provide proper descriptions, but they are computationally too demanding for systems larger than tens of atoms.

In principle, Kohn–Sham density functional theory (KS-DFT) [7, 8], whose computational cost is much lower than that of conventional correlated methods, allows for accurate calculations of dispersion forces if the exact exchange-correlation functional were known. Yet with many of the most commonly used exchange-correlation functionals such as the local density approximation or generalized gradient approximations, DFT is unable to properly account for these forces [12, 13, 23]. Developing possible cures for this deficiency of DFT has been the focus of many recent works [27–33]. Nevertheless, to be generally applicable to a wide range of chemically and biologically interesting systems such as nucleic acids and proteins, the scheme needs to be system independent (*i.e.*, highly transferable) and computationally tractable.

Dispersion-corrected atom-centered potentials (DCACPs) [34], which represent the effect of dispersion forces via electronic potentials centered at the nuclear positions, are characterized by a particularly high transferability. They have been successfully applied to various systems in both gas and condensed phases without increasing the computational cost significantly [76, 90, 91, 150]. Nevertheless, a systematic evaluation of their performance against some well-established benchmark of biological importance is still missing.

Recently, high-level *ab initio* interaction energies of more than 100 nucleobase and amino acid complexes have been collected and published under the acronym JSCH-2005 [135]. Both MP2-geometry-optimized and experimental (crystal and NMR solvent) geometries are considered and the interaction energies are obtained by extrapolating the MP2 results to complete basis set (CBS) limit which are then augmented with CCSD(T) correction terms evaluated with smaller basis sets (6-31G** and cc-pVDZ).

This collection forms an extensive benchmark set for testing the performance of less rigorous but more computationally efficient methods on complexes of biological relevance.

Here, we present DFT calculations for all the molecules in the JSCH-2005 database using the BLYP functional [17, 18] augmented with DCACPs in order to provide a more systematic validation of this recently introduced approach.

6.2 Computational Details

All DFT calculations have been carried out using the CPMD code [93], pseudopotentials of Troullier and Martins [51], and a plane wave cutoff of 75 Ry in a 25^3 \AA^3 cubic cell under periodic boundary conditions (isolated cell [100] for the amino acid pairs). DCACPs for hydrogen, carbon, nitrogen, oxygen, and sulfur have been taken from Refs. [76] and [151]; no DCACP has been applied on fluorine. We choose to use the BLYP functional [17, 18] in this study; however, any other exchange-correlation functional can equally well be employed in combination with the appropriately parameterized DCACPs with a (slightly) varying degree of transferability [76].

To be on equal ground with the JSCH-2005 database, we have, in general, followed the same procedure as Ref. [135] in considering geometry optimizations for some complexes, while for others, experimental geometries have been used. For complexes labeled OG in Ref. [135], we have evaluated the interaction energies of the optimized complexes with respect to the optimized monomers (the intramolecular deformation energy is thus taken into account). The only exceptions are complexes labeled CC1–CC14 whose interaction energies have been calculated at fixed monomer separations to scan the interaction energy curves. In particular, complexes CC1–CC4 map the twist dependence of the stacking interactions for undisplaced dimers whereas complexes CC5–CC7 and CC8–CC13 show the mutual displacement of parallel and anti-parallel dimers. Interaction energies for these complexes (CC1–CC14) and the complexes labeled EG have been computed from single-point energy calculations based on the reference structures. In addition, pure BLYP calculations have been carried out on all EG and CC1–CC14 complexes to evaluate the contribution of DCACPs to the interaction energy ($D = E_{\text{int}}^{\text{DCACP}} - E_{\text{int}}^{\text{BLYP}}$).

An alternative solution to remedy the deficiency of current density functionals in describing dispersion forces is to explicitly include empirical pairwise inter-atomic potentials of the C_6R^{-6} form in the total energy (DFT-D) [16, 23, 30, 72]. This involves parameterization of the C_6 coefficients and defining

appropriate damping functions; furthermore, the electronic structure is left uncorrected. For the sake of comparison, results calculated using DFT-D with the BLYP functional (BLYP-D) [144] are included whenever deemed necessary.

6.3 Results and Discussion

Calculations using BLYP augmented with DCACPs (abbreviated as DCACP from now on) are in excellent agreement with the reference data for all ranges of noncovalently bound complexes in the database. These results demonstrate that DCACPs not only quantitatively correct the BLYP description of dispersion-bound complexes but also improve the one of H-bonded complexes (Tables 6.1 – 6.4, abbreviation used are the same as in Ref. [135]).

Before making further comparison, we would like to stress that both the benchmark and the DCACP calculations do not suffer from the basis set superposition error (BSSE) since the former has been extrapolated to the CBS limit and the latter has been carried out using plane wave instead of localized basis sets. On the other hand, the BLYP-D results from Ref. [144] have been computed with TZV(2d,2p) basis sets without correcting for BSSEs. Previous studies show that BSSEs for TZV(2d,2p) or TZV(2df,2pd) basis sets give a non-negligible positive contribution within 10-20% of the binding energy [30, 152]. Furthermore, the π - π stacked complexes tend to have slightly higher BSSE contribution than the H-bonded complexes. With this in mind, comparisons to the BSSE-uncorrected BLYP-D results should be made with caution.

For instance, on the first sight, the average unsigned mean errors show that BLYP-D gives a slightly better agreement with the reference apart from the H-bonded complexes [H-bonded: 0.69 (DCACP), 0.76 (BLYP-D); interstrand: 0.44 (DCACP), 0.29 (BLYP-D); stacked: 1.51 (DCACP), 0.53 (BLYP-D); amino acids: 1.44 (DCACP), 1.52 (BLYP-D); all in kcal/mol]. Yet D computed from the DCACP results agrees well with the ones estimated by BLYP-D [144]; the average D evaluated from the DCACP and BLYP-D calculations differ only by 0.47, 0.07, -0.16, and 0.66 kcal/mol in the respective class of H-bonded bases, interstrand bases, stacked bases, and amino acids complexes. In addition, we do not expect the plane wave cutoff to influence D to any significant extent; as a test, calculations with both 75 and 150 Ry for the UUst complex give identical values ($D = 8.07$ kcal/mol). Considered that for most complexes, the differences between the dispersion-corrected DFT results (be it BLYP-D or DCACP) and

Table 6.1: Interaction energies (kcal/mol) of the H-bonded base pairs.

Complex	BLYP	DCACP	Benchmark	Error	$-D$
Optimized geometry					
G···C WC	–	–28.15	–28.80	0.65	–
mG···mC WC	–	–27.94	–28.50	0.56	–
A···T WC	–	–15.28	–15.43	0.15	–
mA···mT H	–	–16.33	–16.27	–0.06	–
8oG···C WC pl	–	–29.26	–29.40	0.14	–
I···C WC pl	–	–21.60	–22.70	1.10	–
G···U wobble	–	–15.99	–16.10	0.11	–
CCH+	–	–46.58	–46.50	–0.08	–
U···U Calcutta pl	–	–9.46	–9.80	0.34	–
U···U pl	–	–12.33	–12.60	0.27	–
6tG···C WC pl	–	–26.53	–25.50	–1.03	–
A···4tU WC	–	–14.05	–13.20	–0.85	–
2-aminoA···T	–	–18.17	–17.60	–0.57	–
2-aminoA···T pl	–	–18.29	–17.30	–0.99	–
A···F	–	–4.46	–4.90	0.44	–
G···4tU	–	–15.44	–15.90	0.46	–
G···2tU	–	–13.96	–14.60	0.64	–
A···C pl	–	–16.28	–15.90	–0.38	–
G···G pl	–	–18.91	–18.40	–0.51	–
G···6tG pl	–	–19.71	–19.00	–0.71	–
6tG···G pl	–	–20.22	–19.60	–0.62	–
G···A 1	–	–17.36	–17.50	0.14	–
G···A 1 pl	–	–16.76	–16.10	–0.66	–
G···A 2	–	–12.33	–10.90	–1.43	–
G···A 2 pl	–	–11.78	–10.50	–1.28	–
G···A 3	–	–16.77	–16.80	0.03	–
G···A 4	–	–13.04	–12.10	–0.94	–
A···A 1 pl	–	–13.64	–13.10	–0.54	–
A···A 2 pl	–	–13.04	–12.30	–0.74	–
A···A 3 pl	–	–11.50	–10.90	–0.60	–
8oG···G	–	–18.39	–19.60	1.21	–
2tU···2tU pl	–	–11.03	–11.60	0.57	–
Experimental geometry					
A···T WC	–11.35	–15.82	–16.40	0.58	4.47
G···C WC*	–27.30	–32.82	–35.80	2.98	5.52
A···T WC	–12.93	–17.39	–18.40	1.01	4.46
G···A HB	–7.94	–12.89	–11.30	–1.59	4.95
C···G WC	–25.04	–30.50	–30.70	0.20	5.46
G···C WC	–24.93	–30.34	–31.40	1.06	5.41

$$\text{Error} = E_{\text{int}}^{\text{DCACP}} - E_{\text{int}}^{\text{Benchmark}}, D = E_{\text{int}}^{\text{DCACP}} - E_{\text{int}}^{\text{BLYP}}.$$

Table 6.2: Interaction energies (kcal/mol) of the interstrand base pairs.

Complex	BLYP	DCACP	Benchmark	Error	$-D$
GG0/3.36 CGis036	-1.37	-3.47	-3.68	0.21	2.10
GG0/3.36 GCis036	7.94	-2.50	-4.82	2.32	10.44
AA20/3.05 ATis2005	-0.15	-2.06	-2.34	0.37	1.91
AA20/3.05 TAis2005	5.91	-3.76	-2.16	-1.60	9.67
GC0/3.25 C//Cis	3.55	3.33	3.09	0.24	0.22
GC0/3.25 G//Gis	5.17	2.06	1.93	0.13	3.11
CG0/3.19 G//Gis	2.03	1.19	1.24	0.05	0.84
CG0/3.19 C//Cis	3.96	-3.24	-3.91	0.67	7.20
GA10/3.15 A//Cis	1.93	0.19	-0.31	0.50	1.74
GA10/3.15 T//Gis	2.14	1.07	0.58	0.49	1.07
AG08/3.19 T//Gis	2.52	-0.36	-0.47	0.11	2.88
AG08/3.19 A//Gis	1.61	-0.02	-0.18	0.16	1.63
TG03.19 A//Gis	2.04	-3.67	-4.22	0.55	5.71
TG03.19 T//Cis	-0.47	-0.98	-1.15	0.17	0.51
GT10/3.15 T//Cis	0.98	0.69	0.30	0.39	0.29
GT10/3.15 A//Gis	-0.57	-3.80	-4.06	0.26	3.23
AT10/3.26 T//Tis	1.86	1.16	0.88	0.28	0.70
AT10/3.26 A//Ais	1.59	-0.67	-0.92	0.25	2.26
TA08/3.16 A//Ais	4.20	-0.52	-1.55	1.03	4.72
TA08/3.16 T//Tis	1.48	1.11	0.70	0.41	0.37
AA0/3.24 A//Tis	0.00	-1.44	-1.71	0.27	1.44
AA0/3.24 T//Ais	0.09	-1.16	-1.30	0.14	1.25
A...A IS	2.41	-0.37	-0.70	0.33	2.78
T...T IS	1.99	1.25	1.00	0.25	0.74
G...G IS	4.03	-3.93	-4.50	0.57	7.96
C...C IS	2.71	1.30	1.40	0.10	1.41
A...G IS	-1.44	-4.51	-4.80	0.29	3.07
T...C IS	0.67	0.30	-0.10	0.40	0.37
C...A IS	0.48	-2.00	-3.00	1.00	2.48
G...G IS	-3.06	-5.06	-5.20	0.14	2.00
G...G IS	5.48	0.94	0.80	0.14	4.54
C...C IS	3.45	3.23	3.10	0.13	0.22

$$\text{Error} = E_{\text{int}}^{\text{DCACP}} - E_{\text{int}}^{\text{Benchmark}}, D = E_{\text{int}}^{\text{DCACP}} - E_{\text{int}}^{\text{BLYP}}.$$

Table 6.3: Interaction energies (kcal/mol) of the stacked base pairs.

Complex	BLYP	DCACP	Benchmark	Error	$-D$
Optimized geometry					
G...C S	–	–15.27	–16.90	1.63	–
mG...mC S	–	–14.65	–18.00	3.35	–
A...T S	–	–9.48	–11.64	2.16	–
mA...mT S	–	–11.81	–13.10	1.29	–
CC1–CC14 and experimental geometry					
CC1	11.82	3.04	2.45	0.59	8.78
CC2	6.66	–2.60	–3.85	1.25	9.26
CC3	2.14	–7.19	–8.88	1.69	9.33
CC4	1.31	–7.98	–9.92	1.94	9.29
CC5	10.36	1.39	0.32	1.07	8.97
CC6	10.88	1.79	0.64	1.15	9.09
CC7	5.41	–0.76	–0.98	0.22	6.17
CC8	0.74	–7.51	–9.10	1.59	8.25
CC9	2.04	–7.18	–9.11	1.93	9.22
CC10	2.30	–6.73	–8.27	1.54	9.03
CC11	1.10	–7.56	–9.43	1.87	8.66
CC12	–1.18	–6.51	–7.43	0.92	5.33
CC13	0.75	–7.29	–8.80	1.51	8.04
CC14	2.05	–7.32	–9.11	1.79	9.37
AAst	3.97	–6.53	–8.58	2.05	10.50
GGst	1.24	–10.41	–12.67	2.26	11.65
ACst	1.78	–8.07	–10.22	2.15	9.85
GAst	2.08	–8.96	–11.38	2.42	11.04
CCst	0.82	–8.09	–10.02	1.93	8.91
AUst	2.14	–7.99	–9.79	1.80	10.13
GCst	0.73	–9.08	–10.60	1.52	9.81
CUst	–0.02	–8.91	–10.42	1.51	8.89
UUst	1.63	–6.45	–7.46	1.01	8.08
GUst	0.11	–10.32	–12.09	1.77	10.43
GG0/3.36 GGs036	5.94	–1.33	–1.62	0.29	7.27
GG0/3.36 CCs036	–3.30	–4.27	–3.54	–0.73	7.57
AA20/3.05 AAs2005	6.75	–4.02	–6.06	2.04	10.77
AA20/3.05 TTs2005	0.50	–1.80	–4.18	2.38	2.30
GC0/3.25 G//Cs	1.87	–8.95	–10.80	1.85	10.82
CG0/3.19 G//Cs	0.37	–6.71	–7.88	1.17	7.08
GA10/3.15 A//Gs	4.79	–7.18	–9.14	1.96	11.97
GA10/3.15 T//Cs	3.95	–4.22	–4.69	0.47	8.17
AG08/3.19 A//Gs	2.86	–6.48	–7.58	1.10	9.34
AG08/3.19 T//Cs	2.07	–5.80	–6.07	0.27	7.87
TG03.19 T//Gs	2.53	–4.64	–5.67	1.03	7.17
TG03.19 A//Cs	4.16	–3.96	–4.96	1.00	8.12
GT10/3.15 T//Gs	7.50	–3.65	–4.96	1.31	11.15
GT10/3.15 A//Cs	5.54	–3.99	–5.44	1.45	9.53
AT10/3.26 A//Ts	4.31	–5.75	–6.64	0.89	10.06
TA08/3.16 A//Ts	10.52	–1.91	–6.07	4.16	12.43
AA0/3.24 A//As	5.88	–4.40	–6.25	1.85	10.28
AA0/3.24 T//Ts	6.50	–3.12	–3.86	0.74	9.62
A...T S	5.11	–7.12	–8.10	0.98	12.23
G...C S	–1.22	–7.43	–7.90	0.47	6.21
A...C S	5.27	–5.56	–6.70	1.14	10.83
T...G S	5.44	–6.07	–6.20	0.13	11.51
G...C S	3.64	–5.84	–7.70	1.86	9.48
A...G S	6.05	–4.84	–6.50	1.66	10.89
C...G S	0.36	–9.45	–12.40	2.95	9.81
G...C S	–0.11	–9.60	–11.60	2.00	9.49

$$\text{Error} = E_{\text{int}}^{\text{DCACP}} - E_{\text{int}}^{\text{Benchmark}}, D = E_{\text{int}}^{\text{DCACP}} - E_{\text{int}}^{\text{BLYP}}.$$

Table 6.4: Interaction energies (kcal/mol) of the amino acid complexes.

Complex	BLYP	DCACP	Benchmark	Error	$-D$
F30-K46	0.67	-2.95	-3.10	0.15	3.62
F30-L33	2.87	-4.85	-5.00	0.15	7.72
F30-Y13	1.17	-3.82	-3.90	0.08	4.99
F30-F49	1.48	-2.45	-3.30	0.85	3.93
F30-Y4	4.00	-4.65	-7.00	2.35	8.65
F49-C39	1.99	-1.90	-2.10	0.20	3.89
F49-C6	3.82	-4.42	-5.00	0.58	8.24
F49-K46	1.28	-4.31	-4.80	0.49	5.59
F49-V5	3.20	-5.97	-6.70	0.73	9.17
F49-Y37	1.31	-1.51	-2.50	0.99	2.82
F49-Y4	4.09	-3.24	-3.10	-0.14	7.33
F49-PB (Y4-V5)	1.17	-3.11	-2.80	-0.31	4.28
F49-PB (V5-C6)	1.04	-7.55	-8.20	0.65	8.59
E47-K6 (1IU5)	-71.65	-78.52	-80.73	2.21	6.87
E49-K6 (1BQ9)	-102.27	-108.84	-113.35	4.51	6.57
E54-K2 (1SMM)	-89.83	-92.48	-88.29	-4.19	2.65
E50-K30 (1BRF)	-61.27	-61.44	-60.36	-1.08	0.17
E50-K52 (1BRF)	-86.35	-93.87	-97.14	3.27	7.52
E49-K6 (1BRF)	-63.40	-69.78	-74.24	4.46	6.38

$$\text{Error} = E_{\text{int}}^{\text{DCACP}} - E_{\text{int}}^{\text{Benchmark}}, D = E_{\text{int}}^{\text{DCACP}} - E_{\text{int}}^{\text{BLYP}}.$$

the benchmark values are positive (*i.e.*, both approaches tend to under-bind), DCACP should provide a better description in, for example, the interstrand bases and amino acid complexes, after taking the positive contribution of BSSEs in the BLYP-D results into account.

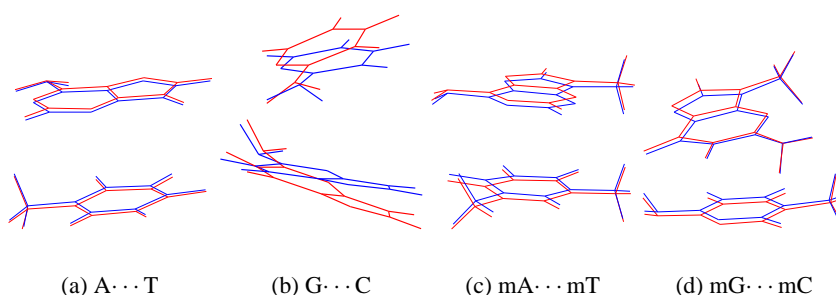


Figure 6.1: The DCACP-optimized stacked base pairs (red) superimposed on the reference geometries (blue, RI-MP2); the corresponding RMSDs per atom are (a) 0.09, (b) 0.50, (c) 0.16, and (d) 0.13 Å.

DCACP also predicts stable configurations for all geometry-optimized complexes, the root mean square deviations (RMSDs) over all atoms (*i.e.*, heavy atoms and hydrogen) between the DCACP-optimized geometries and those of the reference OG complexes do not exceed 0.08 Å in all cases apart

from the stacked base pairs. The G···C base pair has the highest RMSD – 0.50 Å – out of the four stacked base pairs that are optimized; however, the stacked characteristics are still largely maintained, and no motif change to either H-bonded or T-shaped configurations is observed (Fig. 6.1).

6.4 Conclusions

By testing against a well-established benchmark set, we have shown that the BLYP functional augmented with DCACPs gives excellent estimates for interaction energies of important biological components bound by different types of intermolecular forces; the unsigned mean errors in all classes are less than 1.6 kcal/mol compared with the best *ab initio* data available. Furthermore, geometry optimizations lead to stable configurations close to the MP2 predictions. Since the addition of DCACPs does not increase the computational cost dramatically, it serves as a good compromise between efficiency and accuracy to be employed in *ab initio* molecular dynamics.

Chapter 7

Importance of van der Waals Interactions in Liquid Water

Abstract

We present *ab initio* molecular dynamics studies on liquid water using dispersion-corrected density functional theory (DFT) by describing van der Waals interactions via dispersion-corrected atom-centered potentials or empirical van der Waals corrections. Our results show that the inclusion of van der Waals interactions leads to a much softer water structure and higher mobility in contrast to the almost-glassy behavior predicted by DFT-BLYP. The results obtained with dispersion-corrected atom-centered potentials are especially encouraging; properties such as radial distribution functions are in excellent agreement with the experiments, and the self-diffusion coefficients increase threefold as compared with the BLYP prediction. This work thus demonstrates, from an *ab initio* point of view, the clear necessity of properly described van der Waals interactions in liquid water simulations.

7.1 Introduction

Water is the solvent of life; its ubiquity in our environment, its anomalous properties, and its indispensable presence in most biological, chemical, and environmental processes render water the most studied liquid by experiments and computer simulations. Much effort has been devoted to develop a microscopic understanding of the properties of liquid water; however, owing to the complex hydrogen bond (H-bond) network and the intricate interplay of several intermolecular forces, an accurate description remains a challenge [153, 154]. A consensus on the radial structure has been achieved only recently from independent X-ray and neutron scattering experiments [155]. On the other hand, the long-held notion of 4-fold coordinated water molecules [156] has lately been questioned, and the existence of newly proposed configurations with H-bonded chains or large rings of water molecules embedded in a weakly H-bonded disordered network [157] is the subject of a heated debate [158].

Complementary to experiments, computer simulations have played an equal, and sometimes pivotal, role in quantitative characterizations and in advancing qualitative understandings of water at ambient and non-ambient conditions. Well-established empirical force fields based on two-body interactions are able to reproduce properties of water in a wide range of temperatures and pressures [153]. Many of them, however, are parameterized against experimental data, calling into question their predictive power for situations differing largely from the reference system. Moreover, since many-body effects are represented in an effective way by modifying the two-body interaction terms, a direct physical interpretation of the simulated phenomena at the molecular level is sometimes not straightforward. The neglect of cooperative effects can also be problematic [159–162]. The latter can be included to a certain extent by parameterizing many-body potentials against *ab initio* potential energy surfaces of small water clusters [163, 164], but the inability to describe reactions and the transferability to vastly different conditions remain questionable.

Ab initio molecular dynamics simulations (AIMD), on the other hand, are free from these shortcomings and give an unbiased picture of water in different environments. Atomic forces are computed directly from first-principles electronic structure calculations, normally density functional theory (DFT) with generalized gradient approximation (GGA) or hybrid exchange-correlation functionals. This approach is more computationally demanding by comparison but also potentially more predictive. Features difficult to describe with force fields such as electronic polarization effects or bond breaking/forming

events are treated self-consistently; X-ray scattering parameters can be obtained directly from the electron density [165], avoiding the ambiguities and errors of inversion to real-space functions in reciprocal space. Accurate assessments of the parameters used in AIMD simulations of water, however, indicate that the performance is very sensitive to the particular choice of functional [166] and basis sets [167]. GGA water is over-structured and diffuses one magnitude more slowly than experiments; hybrid functionals improve the predictions slightly (Ref. [168] and references therein). Including nuclear quantum effects in AIMD should soften the radial structures further [169–171], despite one contrasting study that actually shows a strengthened radial structure [172].

H-bonding interactions have long been recognized as the major force in determining the spatial patterns and the dynamics of water [173–175]. It has also been argued that the nature of water stems from the competition between H-bond and van der Waals (vdW) interactions [176, 177]. For instance, water can be seen as a dynamic, rapidly changing mixture of tetrahedral ice I_h -like and denser ice II-like structures, with the former thought to be favored by H-bonding while dispersion forces prevail in the latter [177]. Any biased preference from theory on either H-bonding or vdW interactions will thus affect the simulated properties to a large extent. The influence of vdW interactions has been probed by empirical potentials [176, 178], yet no AIMD simulations have been carried out to specifically address this question. Simulations, even in the case when they disagree with experiment data, can provide important clues as to what essential physics has been left out. DFT, in principle, is exact; but because of the approximate nature of exchange-correlation functionals in practical calculations, some fundamental issues remain open. The BLYP functional [17, 18] for which dispersion forces are completely absent [12, 23] can serve as a perfect candidate to reveal effects brought about by the subtle imbalance between these two intermolecular interactions.

In this study, dispersion-corrected atom-centered potentials (DCACPs) [34, 76] and empirical vdW corrections [30, 179] are employed as two pragmatic ways to describe vdW interactions within DFT-BLYP. The concept of DCACPs, which represent these weak forces in the form of atom-centered electronic potentials, has shown encouraging results on their overall performance and transferability across different chemical environments [76, 90–92]; however, in addition to dispersion forces, DCACPs can be correcting other missing intermolecular interactions such as exchange or induction. Therefore, we also apply the empirical vdW corrections in which dispersion forces at long range are recovered by explicitly

including empirical inter-atomic potentials of the C^6R^{-6} form in the total energy. These results are discussed in parallel with the ones obtained from the DCACP approach.

7.2 Computational Details

DFT calculations have been carried out using the CPMD code [93], the BLYP and B3LYP [44] functionals, Goedecker-Teter-Hutter pseudopotentials [52], and a plane wave cutoff of 125 Ry. Parameters for DCACPs and empirical vdW corrections have been taken from Refs. [76] and [179], respectively.

The geometries of the water clusters have been optimized in an isolated cubic 20^3 \AA^3 cell against gradient tolerances of 10^{-5} and 10^{-7} a.u. for nuclear and electronic degrees of freedom. The optimized monomer geometry of each method is used to construct the dimer for the interaction energy curves scanning various O–O distances r_{OO} . Reference at CCSD(T)/aug-cc-pVTZ level (counterpoise-corrected [101] for the basis set superposition errors) has been computed with the GAUSSIAN 03 package [102].

The initial structure for the liquid simulations is from Ref. [180] [64 water molecules in a 12.42^3 \AA^3 periodic box, corresponding to a (light water) density of 1 g/cm^3]. The current standard of 56 – 64 water molecules seems sufficient to eliminate the most obvious problems of finite size effects [155], except for the determination of self-diffusion coefficients D [167]. Hydrogens have been substituted by deuterium to allow for a larger integration time step and to avoid significant nuclear quantum effects [181]. Car-Parrinello molecular dynamics [55] simulations with a fictitious electron mass of 600 a.u. and a 4-a.u. time step has been carried out in the NVE ensemble for 30 ps (velocity rescaling has been applied at 335 K for the first 4 ps) with no thermostat on the electronic degree of freedom. Properties of water have been collected after 10 ps of equilibration time.

Diffusion coefficients D have been calculated from the slopes of the mean square displacement curves using the Einstein relation $2tD = \frac{1}{3} \langle |\mathbf{r}(t) - \mathbf{r}(0)|^2 \rangle$. The motion of the center of mass has been decoupled. The dipole moments have been evaluated from maximally localized Wannier functions [182, 183] every 96 fs within the 20-ps-long trajectory. The orientational autocorrelation functions are defined as $C_{l=1,2}^\mu(t) = 1/N \sum_i \langle P_l[\cos\theta_i(t)] \rangle$ where P_l is the Legendre polynomial of order l and $\theta_i(t)$ is the angle made by the dipoles of water molecule i at time t and at time 0.

7.3 Results and Discussion

The effects of including vdW interactions are monitored by comparing the structural and dynamical properties simulated with DCACP-augmented- (DCACP), the empirical-van-der-Waals-correction-augmented- (BLYP-D), and the pure BLYP (BLYP) functionals. Since the simulations have been carried out in the NVE ensemble, the average temperatures during the data collection period are slightly different: 316 K (BLYP), 325 K (DCACP), and 343 K (BLYP-D).

7.3.1 Radial and angular distribution functions

Method	T_{avg}	$T = T_{\text{avg}}$					$T = 300 \text{ K}$	
		r_{max}	$g_{\text{OO}}^{\text{max}}$	r_{min}	$g_{\text{OO}}^{\text{min}}$	n	$g_{\text{OO}}^{\text{max}}$	$g_{\text{OO}}^{\text{min}}$
BLYP* [184]	300	2.77	2.90	3.3	0.6	4.1	2.90	0.6
BLYP	316	2.77	2.94	3.30	0.60	4.2	3.08	0.56
DCACP	325	2.79	2.67	3.38	0.85	4.6	2.90	0.78
BLYP-D	343	2.82	2.57	3.80	0.94	7.1	2.96	0.82
Neutron [156]	298	2.74	2.76	3.39	0.79	4.6	2.72	0.80
X-ray [185]	300	2.73	2.81	3.40	0.79	-	2.81	0.79

* simulations with a converged discrete variable representation basis set.

Table 7.1: Positions (\AA) and heights of the first maximum and minimum of the oxygen-oxygen radial distribution function and the calculated coordination number n . The average temperature during the simulation T_{avg} (K) is also tabulated. The rescaled $g_{\text{OO}}^{\text{max}}$ and $g_{\text{OO}}^{\text{min}}$ are tabulated in the last two columns.

The oxygen-oxygen radial distribution function (g_{OO}) calculated with the BLYP functional is over-structured compared with the neutron scattering data. DCACP and BLYP-D, on the other hand, yield much softer g_{OO} (see Figure 7.1 and Table 7.1). Nevertheless, part of these discrepancies originates from the slightly different simulation temperatures. For a more systematic comparison, we have disentangled the influence of temperature by rescaling the first maximum ($g_{\text{OO}}^{\text{max}}$) and minimum ($g_{\text{OO}}^{\text{min}}$) to their corresponding heights at 300 K, assuming a linear dependence in this rather small temperature range investigated¹ [186]. The BLYP g_{OO} retains its over-structuring characteristics even after rescaling, displaying a stronger first peak and a much deeper first minimum with respect to both neutron scattering [156] and X-ray [185] diffraction data; $g_{\text{OO}}^{\text{max}}$ and $g_{\text{OO}}^{\text{min}}$ of both DCACP and BLYP-D imply a much softer overall structure, albeit still slightly more structured than experimental measurements. Including nuclear quan-

¹By fitting the TIP4P-pol2 results [165] to $g_{\text{OO}}/T = \text{constant}$, we have obtained 9.0×10^{-3} and 2.7×10^{-3} as the constant for $g_{\text{OO}}^{\text{max}}$ and $g_{\text{OO}}^{\text{min}}$.

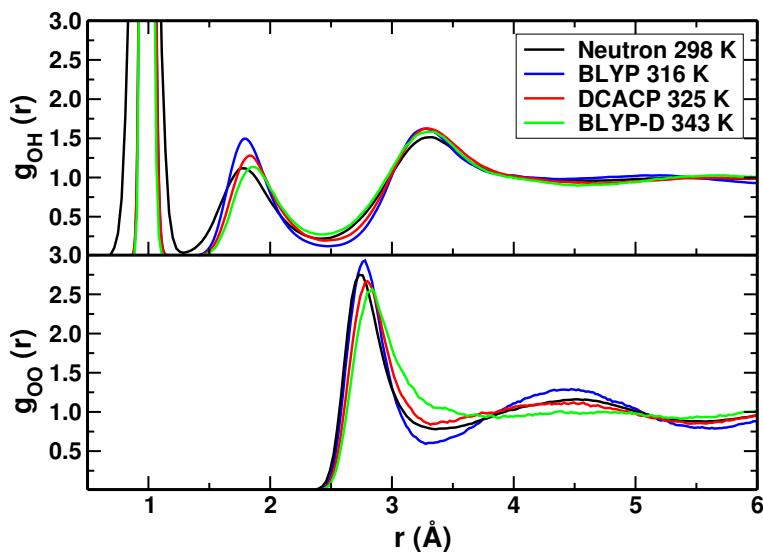


Figure 7.1: Oxygen-oxygen (g_{OO}) and oxygen-hydrogen (g_{OH}) radial distribution functions obtained from neutron diffraction data (Neutron) [156], and DCACP, BLYP-D, and BLYP simulations. The corresponding (average) simulation temperatures are labeled in the legend.

tum effects via *ab initio* path integral molecular dynamics should soften the structure even further, as suggested by previous works [169–171]. As a side note, comparing the normalized BLYP g_{OO} from this work with the one from converged discrete variable representation basis set simulations [167] shows only a variation within the statistical error of the simulations, confirming the adequate choice of the basis set in this work. The coordination number n , obtained by integrating g_{OO} up to the first minimum, is also better described by DCACP. One may suspect this seeming improvement is brought about by the temperature difference; however, the effect of temperature is much smaller than the difference observed here. A wide range of classical force field simulations show that an increase of 53 K alters n by no more than 0.2 [187].

Similar softening is observed in the oxygen-hydrogen radial distribution function g_{OH} (Figure 7.1, lower graph); more importantly, the second intermolecular peaks calculated with DCACP and BLYP-D are much higher than the first, in line with the neutron diffraction data.

Angular distributions of the H-bond-accepting and H-bond-donating water molecules serve as a direct probe of (the activation barrier to) orientational flexibility. One such angle $\beta = \angle_{H-O\dots O}$ has been measured experimentally [188]; the donor angle $\alpha = \angle_{O-H\dots O}$ and the acceptor angle $\theta = \angle_{H\dots O-H}$ are also often used in literature. Distributions of these angles in the first solvation shell (taken as the first intermolecular minimum of g_{OH}) are depicted in Figure 7.2. In general, DCACP and BLYP-D sim-

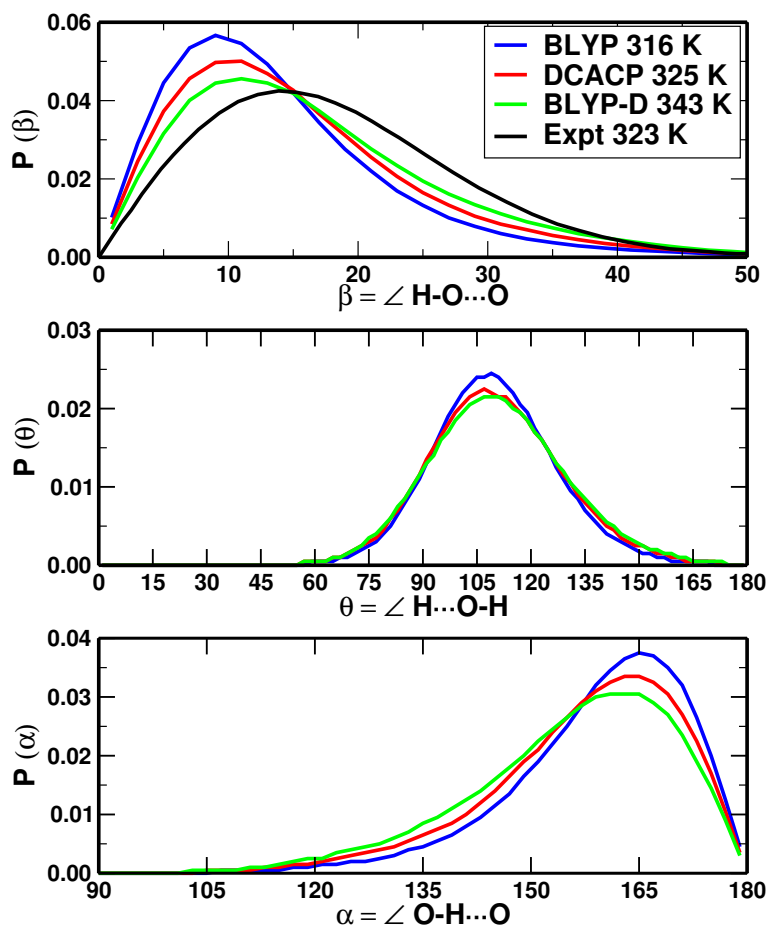


Figure 7.2: Angular distributions (with bin size of 2°) of BLYP, DCACP, and BLYP-D water simulations sampled over 20 ps, together with the experimental data (Expt) for $P(\beta)$ [188].

ulations obtain broader angular distributions with lower maximum than BLYP's ², showing the same trend as TIP4P-pol-2 [187], one of the most accurate force fields available for water. $P(\beta)$ generated with BLYP-D and DCACP agree better with experiment but significant deviations still persist; reaching complete basis set limit [184] and including nuclear quantum effects [187] should further improve the angular structures.

In summary, comparing the pure BLYP results with either the DCACP or the BLYP-D ones shows that the inclusion of vdW interaction softens the structure of liquid water. This softening brings the DCACP and BLYP-D results closer to experiment, and the DCACP results are especially encouraging. Since the electronic configuration of a given water molecule depends on its environment through the

²Although the average simulation temperatures are not identical, previous studies have shown that the position of the maximum and the overall shape of $P(\alpha)$ and $P(\theta)$ are similar within this temperature range whereas $P(\beta)$ is more sensitive to the temperature [187].

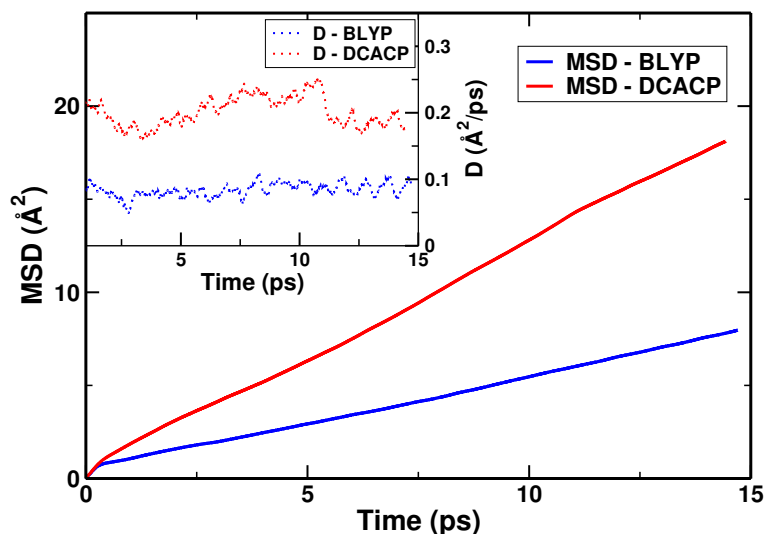


Figure 7.3: Mean square displacements (MSD) and self-diffusion coefficients (D) calculated from the BLYP and DCACP runs.

electronic many-body effects, DCACPs – whose contributions to the total energy depend explicitly on the wave function – take many-body effects implicitly into account and outperform the empirical vdW corrections which are purely pairwise additive.

7.3.2 H-bond analysis and self-diffusion coefficients

A geometric definition, $r_{OO} < 3.5 \text{ \AA}$ and $\angle_{OHO} > 135^\circ$, is used for H-bonding. BLYP data shows a clear preference for 4-fold coordination (64 %). DCACP results, on the other hand, are much more equally distributed, indicating a broader range of coordination numbers. To be precise, DCACP predicts a 1:2 ratio instead of BLYP's 1:3 ratio between 3- and 4-fold coordinations. Nevertheless, all methods support the standard picture where each water is on average almost tetrahedrally coordinated; the average numbers of H-bonds per molecule obtained from BLYP, DCACP, and BLYP-D simulations are 3.76, 3.61, and 3.45, respectively.

As the average number of H-bonds per molecule has a direct impact on D and BLYP predicts, on average, a slightly more coordinated water molecule than DCACP, we expect water simulated with DCACP to be more diffused. The average D calculated from DCACP simulation is indeed much larger, 0.21 (standard deviation of 0.023) $\text{\AA}^2/\text{ps}$, than the BLYP result [0.08 (0.010) $\text{\AA}^2/\text{ps}$] (see Figure 7.3). For comparison, the experimental value for heavy water at 45°C is 0.30 $\text{\AA}^2/\text{ps}$ [189]. D , as mentioned

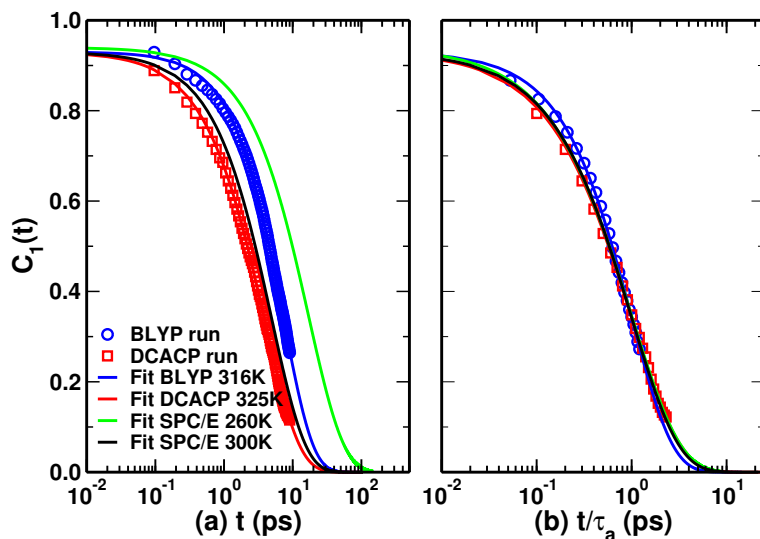


Figure 7.4: (a) The orientational autocorrelation function $C_1(t)$ as a function of time t . The parameters obtained from fitting $C_1(t)$ to the stretched exponential are $\beta = 1.01, \tau_a = 7.30$ and $\beta = 0.85, \tau_a = 3.83$ for the BLYP and DCACP runs, respectively (A is fixed at 0.93); the SPC/E data are taken from Ref. [191]. (b) Test of the time-temperature superposition principle. When the time is rescaled by τ_a , the lines from different temperature should fall on a single curve.

before, is sensitive to the finite size effect. As a rough estimate, assuming a hypothetical experiment were to be carried out on a periodic system of 64 water molecules, the ‘experimental’ D at 45°C would lower down to 0.21 Å²/ps instead (following the same train of thoughts as in Ref. [167]), coinciding perfectly with the D predicted by DCACP.

7.3.3 Orientational correlation times and dipole moments

Typical single molecule relaxation processes have been examined by calculating the orientational autocorrelation functions $C_l^\mu(t)$ for $l = 1, 2$ that have been fitted (between 0.5 and 8 ps) as simple exponential decays with time constants τ_1 and τ_2 , respectively. The experimental values for τ_1 and τ_2 at 300 K are 4.76 and 1.92 ps [190]. Predictions by BLYP fall somewhat outside the experimental values whereas DCACP results show much closer estimates; $\tau_1, \tau_2 = 7.34, 3.97$ (BLYP, 316 K) and 4.39, 2.52 (DCACP, 325 K), respectively.

To separate the temperature effect, we also compared our $C_1^\mu(t)$ with the recent SPC/E simulations at various temperatures [191]; to be in line with Ref. [191], we re-fitted the data to a stretched exponential $A \cdot \exp[-(t/\tau_a)^\beta]$. As shown in Figure 7.4, DCACP predictions compare well with the results of classical simulations. BLYP, on the other hand, predicts $C_1^\mu(t)$ that falls between the SPC/E results

at 260 and 300 K. In addition, the relaxation time τ_α estimated from fitting the DCACP results for $0.1 < t < 9$ ps verifies the time-temperature superposition principle of the mode coupling theory which states that $C_1^\mu(t)$ in the α -relaxation regime at different temperatures follow the same master curve if the time is rescaled by τ_α [191]. The BLYP estimate deviates slightly from the master curve.

The average dipole moments evaluated from the BLYP and DCACP runs are 3.01 (standard deviation 0.28) and 2.91 (0.28) Debye, both close to the experimental value of 2.86 Debye [192]. It has been demonstrated previously that DCACPs only induce a very small change in the electronic structure of monomers, and multipole moments remain basically unaltered [76]. Here, the inclusion of DCACPs influences the liquid structure, resulting in a slight improvement of the average dipole moment over BLYP's.

7.3.4 Cluster studies

Table 7.2 shows that the interaction energies of the optimized water dimer and trimer evaluated with DCACP and BLYP-D have the smallest deviations with respect to high-level reference calculations.

Method	(H ₂ O) ₂			(H ₂ O) ₃	
	r_{OO}	\angle_{OHO}	E_{int}	r_{OO}	E_{int}
BLYP	292	171	-4.09	283	-13.11
DCACP	291	170	-5.19	281	-16.61
BLYP-D	291	170	-5.46	283	-16.97
B3LYP	291	172	-4.43	281	-13.90
Ref. ^a	291	175	-5.02	278	-15.90

^a Dimer: best *ab initio* [193]

^a Trimer: MP2 (extrapolated) [194]

Table 7.2: Structural parameters (r_{OO} in pm, \angle_{OHO} in degree) and the interaction energy (E_{int} , kcal/mol) of the water dimer and the cyclic trimer. Recent quantum Monte Carlo calculations predicted a dimer E_{int} between -5.03 and -5.49 kcal/mol [195].

In liquid water, molecules are constantly in motion and do not have fixed positions and orientations. Hence, any structures of the liquid must be described properly in terms of probabilities and averages; configurations far from optimal H-bond arrangements can also contribute significantly. The interaction energy curves of H-bonding- and vdW-dominant configurations shown in Figure 7.5 demonstrate that both DCACP and BLYP-D improve the energetics, bringing the curves closer to the CCSD(T)/aug-cc-pVTZ values.

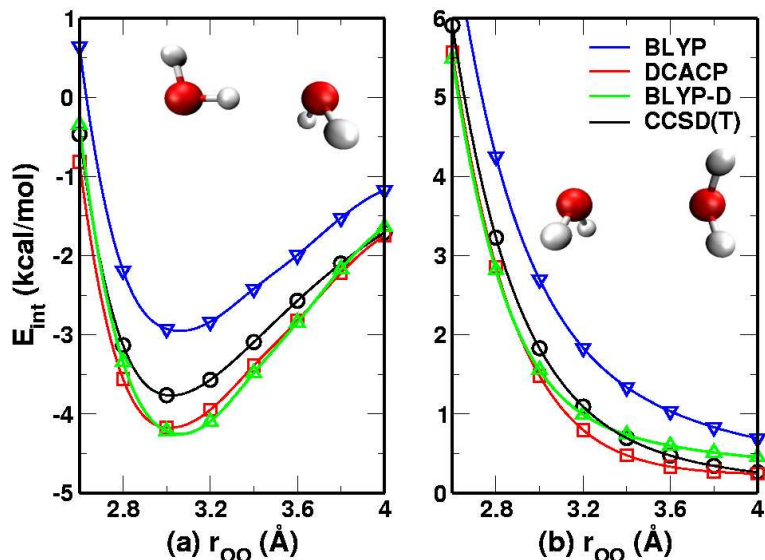


Figure 7.5: Interaction energy curves of two configurations of the water dimer (shown on the upper right hand corner) as a function of r_{OO} . For (a) the H-bonded dimer \angle_{OHO} is fixed at 180° . The interaction energy (E_{int}) is defined as $E_{\text{dimer}}^{\text{fix}} - 2 \cdot E_{\text{mono}}^{\text{opt}}$; $E_{\text{dimer}}^{\text{fix}}$ and $E_{\text{mono}}^{\text{opt}}$ are energies of the fixed dimer at every r_{OO} and of the optimized monomer, respectively.

In short, including vdW interactions not only improves the optimized H-bond geometry, but other configurations are also made more probable. The consequences are shown clearly in the structural and dynamical properties of liquid water such as less structured/wider radial/angular distribution functions.

The approach of empirical vdW corrections is computationally more efficient than DCACPs at its current implementation in which further optimization is still possible³. Nevertheless, DCACPs implicitly include many-body interactions, offering an advantage over the pairwise approach of empirical vdW corrections, allowing one to achieve accuracy similar to the coupled-cluster calculations at much lower computational overhead than the high-level correlated methods.

7.4 Conclusions

AIMD simulations have been instrumental in elucidating microscopic mechanisms in an accurate and unbiased manner in various disciplines. Nevertheless, the predictive power depends heavily on the accuracy of the chosen functional.

In this work, we have demonstrated that by curing the shortcomings of DFT-BLYP in describing

³A rough estimate puts the additional CPU time of DCACP and BLYP-D with respect to the pure BLYP calculation at 97% and 41%, respectively. The CPU time is for one Car-Parrinello step in the liquid water simulation as set up in this study.

vdW interactions, the structure of liquid water becomes much softer, and its dynamical properties are much more liquid-like instead of the almost-glassy behavior predicted by the pure BLYP functional. The self-diffusion coefficient, for example, is increased by almost 3-fold upon the inclusion of DCACPs. A proper description of vdW interactions thus brings the simulated properties into closer agreement to experiments. To best realize the agreement, however, other deficiencies of DFT-BLYP and nuclear quantum effects should also be addressed.

As chemical and biological systems often have hydrophobic groups, a proper treatment of vdW interactions in liquid water simulations have greater implications than simply improving properties of bulk water. Even though the field of research on incorporating vdW interactions in functionals widely employed in AIMD in a robust and efficient way is still very much in its infancy, it is an important prerequisite to gain insights into many important phenomena such as the evaporation and hydration of nanoparticles.

Chapter 8

Comparative Study of Dispersion-Corrected BLYP and PBE Water

Abstract

Ab initio molecular dynamics simulations of liquid water have been carried out using dispersion-corrected density functional theory. The absence of van der Waals interactions in standard DFT is treated by employing either dispersion-corrected atom-centered potentials or the empirical van der Waals corrections. The dispersion-corrected BLYP simulations have previously been shown to give rise to highly mobile liquid water which displays much softer structural properties. With the dispersion-corrected PBE functional, however, the structural and dynamical properties either remain unchanged or become slightly more glassy.

8.1 Introduction

Water is without doubt one of the most important chemical substances known. Yet despite extensive studies, describing its microscopic nature remains a challenge, owing mainly to the strength and the directional nature of hydrogen bonds (H-bonds) which lead to complex cooperative phenomena. Indeed as ever more sophisticated and novel experimental and theoretical techniques are applied to study liquid water, it is becoming increasingly clear that this disparate information is only continuously heating the debate rather than concluding it.

Owing to its favorable performance-to-cost ratio, the Kohn–Sham formalism of density functional theory (DFT) [7, 8] has been one of the most popular *ab initio* methods in the fields of condensed matter physics and material science. Accurate assessments of the parameters used in *ab initio* molecular dynamics (AIMD) simulations of water, however, indicate that the performance of DFT with generalized gradient approximation (GGA) exchange-correlation functionals in describing H-bonding is not yet fully understood. DFT-GGA estimates interaction energies of the water clusters reasonably well [196], but the predicted liquid water is over-structured, diffuses too slowly, and is less dense compared with experimental results [166, 180, 187, 197]. Possible explanations to this discrepancy have been explored in many works. Hybrid functionals offer slightly better predictions [168]; reaching complete basis set (CBS) limit improves both structural and dynamical properties, but the deviations from experiment are still noticeable [167, 184, 198]. In addition, including nuclear quantum effects should further soften the radial structure [169–171].

Nevertheless, it is important to realize that interactions in liquid water are not simply due to H-bonding but rather to a fine balance between H-bonding and the non-directional van der Waals (vdW) interactions [176–178]. The deficiency of many approximated exchange-correlation functionals in describing vdW interactions will certainly contribute to the discrepancy mentioned above, and as demonstrated in the previous chapter, incorporating these weak interactions in DFT-BLYP greatly softens the structure of liquid water [150]. In addition, a proper description of vdW forces are also important prerequisites to gain insights into the influence of hydrophobic effects on the structure and functions of specific amino acids as well as material science problems such as water in zeolites or fluid flow in carbon nanotubes.

Here we present Car-Parrinello molecular dynamics [55] studies on liquid water using the PBE

functional [97] augmented with either dispersion-corrected atom-centered potentials (DCACPs) [34,76] or empirical vdW corrections [30,179]. Developing efficient calculations of vdW interactions in DFT is still in its early stages and the two pragmatic schemes chosen here represent a good compromise between efficiency and accuracy. The radial and angular distribution functions, local dipole moment, diffusion coefficients, and orientational autocorrelation times of the simulated liquid water are reported. Behaviors of water clusters are also investigated, offering clues as to how the improvement, if any, comes about.

8.2 Computational Details

DFT calculations have been carried out using the CPMD code [93], analytic Goedecker-Teter-Hutter pseudopotentials [52], a plane wave cutoff of 125 Ry, and the BLYP [17,18], B3LYP [44], and PBE [97] functionals. DCACPs calibrated against high-level *ab initio* references [CCSD(T) and CI] and parameters for empirical van der Waals corrections have been taken from Refs. [76] and [179], respectively. The geometries of the water monomer and clusters have been optimized in an isolated cubic cell measuring 20 Å on all sides against gradient tolerances of 10^{-5} and 10^{-7} a.u. for nuclear and electronic degrees of freedom. Interaction energy surfaces of the dimer as a function of oxygen-oxygen distances r_{OO} and $O_D-H_D \cdots O_A$ angles \angle_{OHO} (see Figure 8.1) have been evaluated; the water dimer has been constructed from the optimized monomers of respective methods. CCSD(T)/aug-cc-pVTZ calculations have been carried out using the GAUSSIAN 03 package [102].

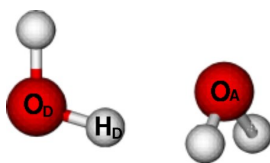


Figure 8.1: Graphical representation of the water dimer.

Simulations of the liquid state have been carried out in a cubic 12.42^3 \AA^3 periodic box containing 64 water molecules. The starting configuration is from Ref. [180]. The mass of hydrogen has been replaced with the one of deuterium to allow for a larger integration time step and to be able to compare with experimental results with reduced nuclear quantum effects. Car-Parrinello molecular dynamics simulations with a fictitious electron mass of 600 a.u. and a 4-a.u. time step have been carried out in the

NVE ensemble (velocity rescaling has been applied at 335 K for the first 2-5 ps) for approximately 30 ps including 10 ps of equilibration. No thermostat has been applied on the fictitious electronic degrees of freedom. The dipole moments have been evaluated from maximally localized Wannier functions [182, 183] averaged over configurations taken at every 96 fs within the 20-ps-long trajectory.

Data from Chapter 7 are included whenever deemed necessary for detailed technical discussions omitted previously and for the sake of easy comparisons.

8.3 Results and Discussion

The small time step and fictitious electron mass have allowed for long and stable simulations without introducing a significant energy exchange between the ionic and electronic degrees of freedom (*i.e.*, the adiabatic separation is maintained throughout). No significant drift in potential energy is seen, and the drifts in the fictitious electron kinetic energy over 30 ps for all runs are no more than 5×10^{-4} a.u. ps⁻¹. The ionic temperatures stabilize after the 10-ps equilibration time, showing no systematic drift during the 20-ps data-collection period; the average ionic temperature for each simulation is tabulated in Table 8.1 (each has a standard deviation of less than 20 K).

8.3.1 Radial and angular distribution functions

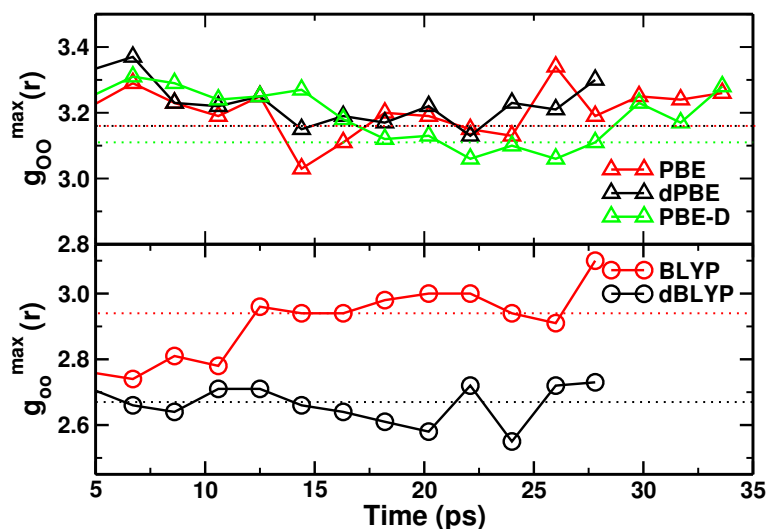


Figure 8.2: Variation of g_{OO}^{\max} observed during the 30-ps NVE simulation. The values are obtained from non-overlapping data blocks of 2 ps. Dashed lines indicate the average values over the whole data-collection periods.

Method	T=T _{avg}								T=298 K	
	T _{avg}	r _{max}	g _{OO} ^{max}	r _{min}	g _{OO} ^{min}	n	h	μ	g _{OO} ^{max}	g _{OO} ^{min}
BLYP	316	2.77	2.94	3.30	0.60	4.2	3.72	3.01	3.10	0.55
dBLYP	325	2.79	2.67	3.38	0.85	4.6	3.59	2.91	2.91	0.77
PBE	317	2.72	3.16	3.26	0.43	3.9	3.80	3.12	3.33	0.38
dPBE	316	2.71	3.16	3.26	0.45	3.9	3.79	3.13	3.32	0.40
PBE-D	327	2.71	3.11	3.29	0.39	4.2	3.88	3.17	3.37	0.31
Neutron [156]	298	2.74	2.76	3.39	0.79	4.6	–	–	2.76	0.79

Table 8.1: The average simulation temperature T_{avg} , characteristics of the oxygen-oxygen radial distribution function (r in Å), the coordination number n , the average number of H-bonds per water molecule h , and the average dipole moment μ from different simulations. The temperature-rescaled $g_{\text{OO}}^{\text{max}}$ and $g_{\text{OO}}^{\text{min}}$ are tabulated in the last two columns. BLYP and dBLYP data are taken from Ref. [150] apart from h where a different definition is used here.

Radial distribution functions. NVE simulations have longer correlation time in radial distribution functions than NVT simulations, and a simulation much longer than 10 ps is necessary to extract the equilibrium properties with reasonable accuracy even if the system has been equilibrated for several tens of ps [167]. Figure 8.2 illustrates that $g_{\text{OO}}^{\text{max}}$ obtained from non-overlapping windows of 2 ps show no significant systematic drift and lie within ± 0.2 units to the average calculated from the last 20 ps of trajectory. Nevertheless, the abrupt increase seen in the end of the BLYP simulation, even though still within the statistical error, warrants a longer simulation time to confirm the stability of this simulation.

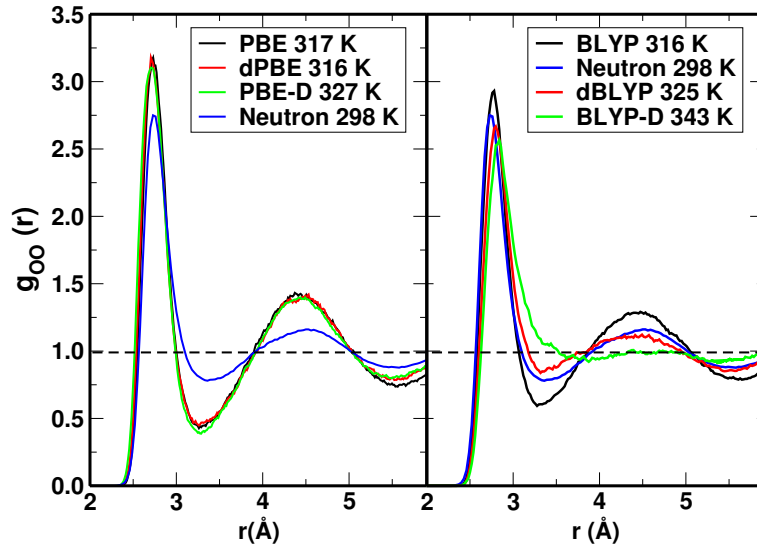


Figure 8.3: Oxygen-oxygen (g_{OO}) radial distribution functions obtained from neutron diffraction data (Neutron) [156] and various methods (data for the BLYP associated simulations are taken from Ref. [150]). The corresponding simulation temperatures are labeled in the legend.

Oxygen-oxygen radial distribution functions (g_{OO}) evaluated from simulations with PBE, DCACP-

augmented-PBE (dPBE), empirical-vdW-corrections-complemented-PBE (PBE-D), BLYP, and DCACP-augmented-BLYP (dBLYP) are shown together with the neutron diffraction data [156] in Table 8.1 and Figure 8.3. Since the simulations have been carried out in the NVE ensemble, the average temperatures are slightly different. For a fair comparison, we rescale the $g_{\text{OO}}^{\text{max}}$ and $g_{\text{OO}}^{\text{min}}$ with the relation $g_{\text{OO}}/T = \text{constant}$ [186] (by fitting to TIP4P-pol2 results [165], we obtain the constant as 9.0×10^{-3} and 2.7×10^{-3} for $g_{\text{OO}}^{\text{max}}$ and $g_{\text{OO}}^{\text{min}}$, respectively).

g_{OO} generated from simulations that incorporate vdW interactions to DFT-GGAs via either DCACPs or empirical vdW terms agree qualitatively: the liquid water turns less structured when BLYP is complemented with either of the two schemes [150]; for dispersion-corrected PBE, however, both g_{OO} show little change from the one predicted by PBE. One can even argue that adding empirical vdW terms to PBE (*i.e.*, PBE-D) actually makes situations slightly worse as seen from the temperature-scaled $g_{\text{OO}}^{\text{min}}$ and $g_{\text{OO}}^{\text{max}}$.

The coordination numbers n of dPBE and PBE (both 3.9) are also slightly lower than the BLYP one (4.2) and deviate significantly from the dBLYP and experimental estimates (both are 4.6).

Angular distributions. Angular distributions of the H-bond accepting and donating water molecules are important quantities for characterizing the flexibility of H-bonds; the donor angle $\alpha = \angle_{\text{O-H}\dots\text{O}}$ and the acceptor angle $\theta = \angle_{\text{H}\dots\text{O-H}}$ serve as a direct probe of the orientational flexibility. Distributions of these two angles [$P(\alpha)$ and $P(\theta)$], together with the only experimentally measured angle $\beta = \angle_{\text{H-O}\dots\text{O}}$ [$P(\beta)$] [188], in the first solvation shell (taken as the first intermolecular minimum of g_{OH}) are depicted in Figure 8.4. All three distributions evaluated from the dPBE, PBE-D, and PBE runs seem identical and are much sharper than distributions extracted from the dBLYP and BLYP data [150], reminiscent of the similar effect afflicting g_{OO} .

Positions of maxima are similar in all water models, differing by no more than 10° , and are consistent with a slightly bent H-bond. Nevertheless, it is noticeable from the two distributions $P(\alpha)$ and $P(\beta)$ that PBE-related methods have much narrower distributions than dBLYP, and the maxima are located at angles close to a linear arrangement of H-bonds, indicating a preference over non-directional vdW interactions for strongly directional H-bonding.

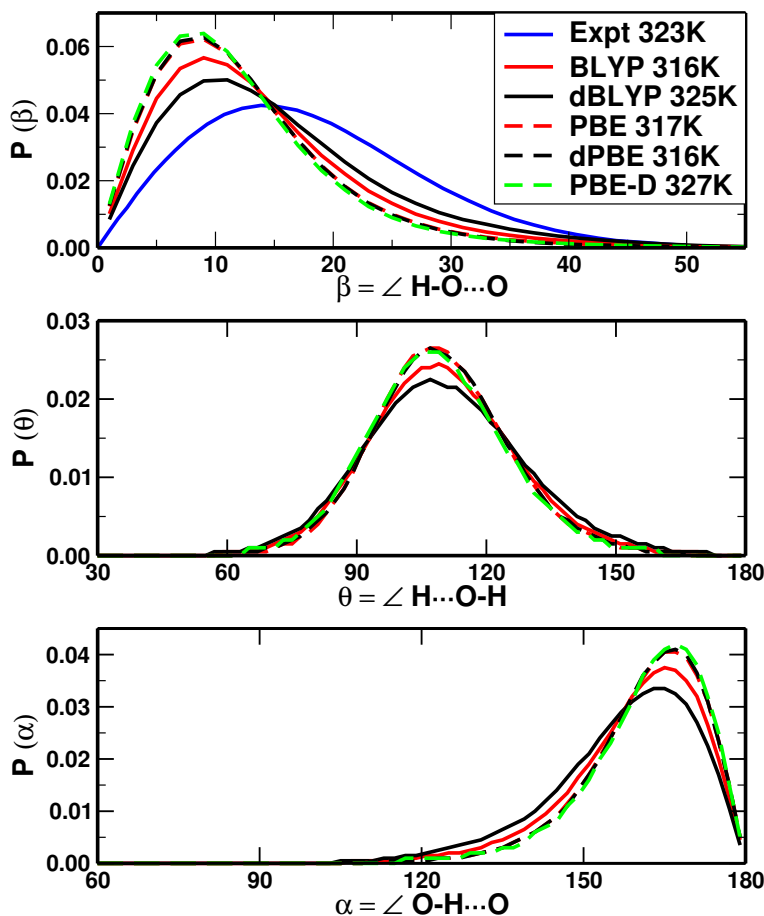


Figure 8.4: Angular distributions of BLYP, dBLYP, PBE, dPBE, and PBE-D water simulations sampled over 20 ps, together with the experimental data (Expt) of $P(\beta)$ [188].

8.3.2 Orientational autocorrelation times and dipole moments

The orientational autocorrelation functions are defined as

$$C_{l=1,2}^{\alpha}(t) = 1/N \sum_i \langle P_l[\cos\theta_i(t)] \rangle$$

where P_l is the Legendre polynomial of order l and $\theta_i(t)$ is the angle made by either the dipoles ($\alpha = \mu$) or the OH vector ($\alpha = \text{OH}$) of water molecule i at time t and at time 0. τ_l^{α} are estimated by fitting $C_l^{\alpha}(t)$ between 0.5 and 9 ps to a mono-exponential decay and the results are tabulated in Table 8.2. All PBE methods (PBE, dPBE, and PBE-D) grossly overestimate the orientational correlation times, opposite to the reasonable behavior displayed by BLYP-related methods (the dBLYP estimate is within the experimental range, and the BLYP result falls only slightly outside). Yet the ratio of $\tau_1^{\mu}/\tau_2^{\mu}$ predicted by dPBE

	T_{avg}	D	τ_1^μ	τ_2^μ	τ_1^{OH}	τ_2^{OH}	τ_1^μ/τ_2^μ	$\tau_1^{\text{OH}}/\tau_2^{\text{OH}}$
BLYP	316	0.08(1)	7.34	3.97	9.78	4.92	1.85	1.99
dBLYP	325	0.21(2)	4.39	2.52	5.39	2.40	1.74	2.24
PBE	317	0.04(1)	16.6	8.02	21.96	10.44	2.07	2.10
dPBE	316	0.03(1)	22.93	9.83	26.81	11.89	2.33	2.25
PBE-D	327	0.03(1)	24.11	10.47	25.58	13.83	2.30	1.85
Expt	*	0.30	4.76	1.92	2-7.5	1.7-2.6	2.48	–

* D at 318K [189], τ_i^α all at 300 K [190, 207–212].

Table 8.2: Dynamical properties of liquid water, including the self-diffusion coefficient D ($\text{\AA}^2/\text{ps}$) and the orientational correlation times τ_i^α (ps), calculated from various simulations.

is the closest to experiment. Incidentally, the average dipole moments evaluated from PBE, dPBE, and PBE-D simulations are all greater than the corresponding values from the BLYP and dBLYP runs, but none deviates significantly from the experimental measurement of 2.86 Debye [192] (Table 8.1).

The ratio $\tau_1^{\text{OH}}/\tau_2^{\text{OH}}$ are reported in Table 8.2. Comparing with values from the fully diffusive model [173] and the extended jump model [199] suggests that the reorientation mechanism predicted here follows more closely the jump mechanism, in agreement with recent classical and quantum force field calculations [200].

8.3.3 Self-diffusion coefficients

The distinct properties of water can be ascribed largely to the strength and the directional nature of H-bonds which has a direct influence on the self-diffusion coefficient D . In Table 8.1, the average number of H-bonds is listed using the same H-bond definition as in Ref. [201]. Even though the number varies from 3.6 to 3.9, none of them challenges the concept that liquid water is tetrahedrally coordinated, one of the disputes concerning liquid water [157, 158, 202, 203].

One should bear in mind that D is an elusive quantity to measure *in silico*; it is sensitive to the protocol of AIMD simulations, such as the length of equilibration as well as the production runs, the pseudopotential type, and the basis set size [167, 168, 180, 187, 204–206]. Furthermore, the system size of 64 water molecules is prone to finite size effects, as discussed in Chapter 7.

D are calculated from the slopes of the mean square displacement (MSD) curves using the Einstein relation, $2tD = \frac{1}{3}\langle|\mathbf{r}(t) - \mathbf{r}(0)|^2\rangle$. For cleaner statistics, the MSD curves from a series of overlapping data blocks (with lengths equal to half of the total simulation time, the starting configuration of each block is separated from those of neighboring ones by 0.001 ps) are averaged. D estimated from the

dPBE, PBE-D, and PBE runs (3×10^{-3} , 3×10^{-3} , and $4 \times 10^{-3} \text{ \AA}^2/\text{ps}$) are only half of the value predicted by BLYP ($8 \times 10^{-3} \text{ \AA}^2/\text{ps}$) [150] and far below the experimental value (see Table 8.2) even after taking the finite size effects into account, in line with the over-structured radial and angular distribution functions.

8.3.4 Clusters

While characteristics of liquid water remain controversial, the gas-phase water cluster community has provided detailed experimental and theoretical pictures of the behavior of small water clusters. From Table 8.3, one can see that all methods, bar BLYP and B3LYP, tend to over-bind compared with MP2 results at CBS limit (MP2/CBS) [213]. Surprisingly, PBE gives the best results even though the liquid water is grossly over-structured. As a side note, the differences between interaction energies predicted in this study and Ref. [196], which employed aug-cc-pV5Z basis set, are less than 0.1 kcal/mol in all available clusters/functionals, validating the choice of wave function cutoff for the liquid water simulations.

	dimer		trimer cyclic		tetramer cyclic		pentamer cyclic		cage		hexamer prism		cyclic	
	E_{int}	e_{int}	E_{int}	e_{int}	E_{int}	e_{int}	E_{int}	e_{int}	E_{int}	e_{int}	E_{int}	e_{int}	E_{int}	e_{int}
BLYP	-4.09	-4.09	-13.11	-4.37 (7)	-24.14	-6.04 (48)	-32.02	-6.40 (56)	-37.34	-4.67 (14)	-37.37	-4.15 (2)	-39.53	-6.59 (61)
dBLYP	-5.19	-5.19	-16.61	-5.54 (7)	-29.71	-7.43 (43)	-38.79	-7.76 (50)	-49.65	-6.21 (20)	-49.95	-5.55 (7)	-47.29	-7.88 (52)
BLYP-D	-5.46	-5.46	-16.99	-5.66 (4)	-30.19	-7.55 (38)	-39.54	-7.91 (45)	-51.15	-6.39 (17)	-51.69	-5.74 (5)	-48.52	-8.09 (48)
B3LYP	-4.41	-4.41	-13.87	-4.62 (5)	-25.15	-6.29 (43)	-33.31	-6.66 (51)	-39.87	-4.98 (13)	-39.57	-4.90 (0)	-41.22	-6.87 (56)
PBE	-5.06	-5.06	-16.09	-5.36 (6)	-29.04	-7.26 (43)	-38.40	-7.68 (52)	-46.56	-5.82 (15)	-46.48	-5.16 (2)	-47.41	-7.90 (56)
dPBE	-5.20	-5.20	-16.70	-5.57 (7)	-29.96	-7.49 (44)	-39.51	-7.90 (52)	-48.03	-6.00 (15)	-47.95	-5.33 (3)	-48.75	-8.13 (56)
MP2/CBS	-4.98	-4.98	-15.83	-5.28 (6)	-27.63	-6.91 (39)	-36.3	-7.26 (46)	-45.8	-5.74 (15)	-45.9	-5.10 (2)	-44.8	-7.5 (50)

Table 8.3: Total interaction energy (E_{int} , kcal/mol) and interaction energy per H-bond (e_{int} , kcal/mol) of optimized water clusters from dimer up to hexamer. The number in the parenthesis is the percentage increase in e_{int} of various clusters with respect to that of the dimer. MP2/CBS reference is taken from [213].

The number in the parenthesis is the percentage increase in the interaction energy per H-bond e_{int} with respect to that in the dimer, *i.e.*, the absolute % enhancement. The absolute % enhancements predicted are correct to within 10% of the MP2/CBS results and follow the same trend: e_{int} increases from dimer to pentamer but decreases when going to the cage and prism configurations of the hexamer. The agreement, however, exhibits a strong variation among the clusters; for examples, all methods bar dBLYP and BLYP-D predict over 35% and 50% decrease in e_{int} going from pentamer to the cage and prism configurations of the hexamer. dBLYP (30% and 42%) agree most closely to the MP2 values (31% and 44%); however, it fares worse when comparing the relative % enhancement from trimer to either of the hexamers.

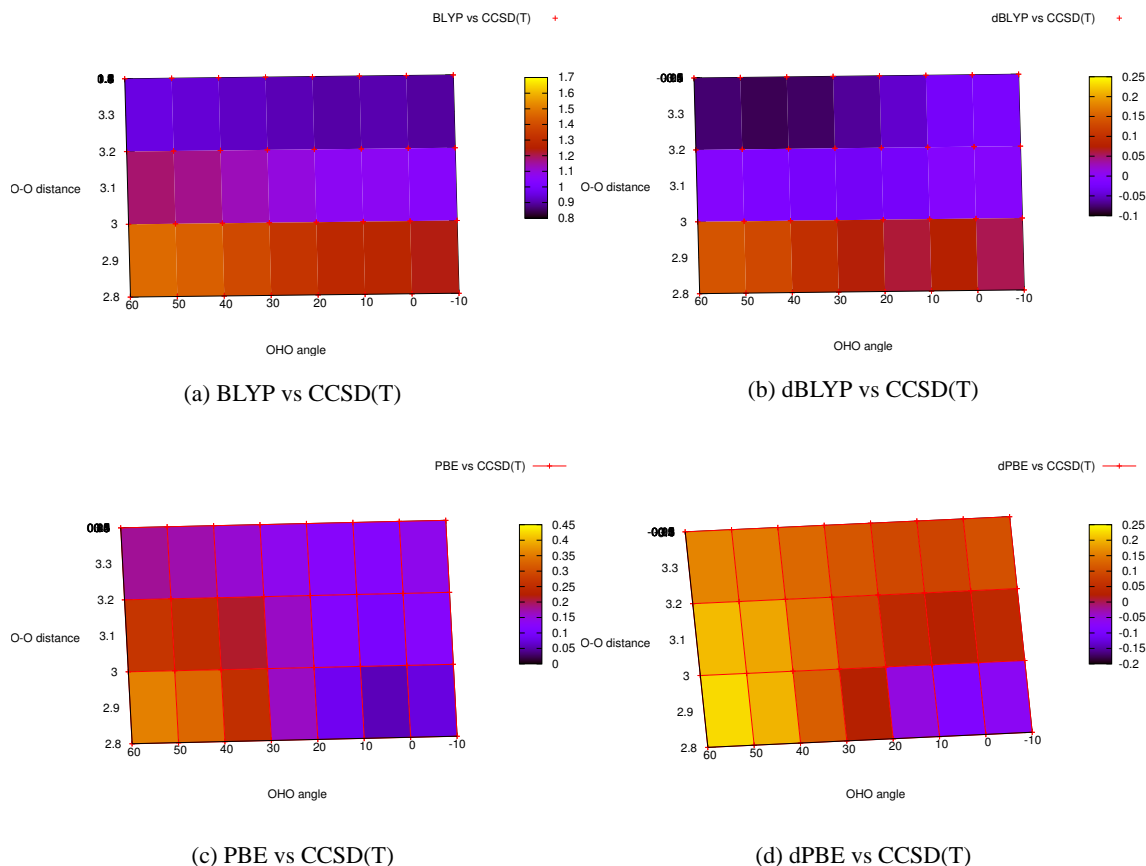


Figure 8.5: Boltzmann-weighted difference between interaction energies of the water dimer calculated with four DFT methods (BLYP, dBLYP, PBE and dPBE) and CCSD(T)/aug-cc-pVTZ (not counterpoise-corrected) method as a function of r_{OO} and \angle_{OHO} .

Although DFT-GGAs, in general, predict H-bonding quite well, their poor descriptions on the angular dependence of the interaction energy surfaces have been reported, probably due to the missing dispersion interaction [214, 215]. Figure 8.5 shows the Boltzmann-weighted difference¹ between the interaction energies calculated with the four DFT approaches (BLYP, dBLYP, PBE, and dPBE) and CCSD(T)/aug-cc-pVTZ (basis set superposition errors not corrected) as a function of r_{OO} and \angle_{OHO} . Similar to the results of the optimized clusters, BLYP has the highest Boltzmann-weighted difference among the four, followed by PBE, dPBE, and dBLYP. Surprisingly though, the performance of BLYP is insensitive to changes in \angle_{OHO} , unlike the two PBE approaches where there is a pronounced angular dependence, especially at $r_{\text{OO}} < 3.2 \text{ \AA}$. dBLYP displays an even better angular dependence which slightly worsens at longer r_{OO} but is still on the same scale, if not better, as other approaches. In addition, if one

¹ $\Delta = \exp[-(E^{\text{CCSD(T)}} - E_{\text{min}}^{\text{CCSD(T)}})/kT] \cdot (E^{\text{DFT}} - E^{\text{CCSD(T)}})$, kT is chosen to be 600.

traces the diagonal path, going from short r_{OO} and small \angle_{OHO} (mainly H-bonding) to large r_{OO} and \angle_{OHO} (vdW-dominant), the Boltzmann-weighted differences of PBE and dPBE increase along the way, in great contrast to the trend observed for BLYP and dBLYP.

The salient angular dependence of PBE and dPBE at r_{OO} close to the first maximum and minimum of the g_{OO} (Table 8.1) is liable to introduce a significant bias towards one particular configuration over the other, in this case, directional H-bonded dimer over non-directional vdW-bound complexes. This may explain why even though the equilibrium interaction energies are better predicted by PBE and dPBE, BLYP with its large deviation in interaction energies still manages to capture properties of liquid water closer to experiments than both of them. With its superior performance on interaction energies, especially around the region of interest, dBLYP is expected to (and does) give the best prediction on liquid water.

Improving the description of vdW interactions in BLYP makes the probability of sampling non H-bonded complexes, which are otherwise too repulsive energetically, closer to reality. This improvement manifest itself in the liquid simulation, leading to much softer water that is highly mobile.

In contrast, the geometry-optimized clusters and the dimer interaction energy surfaces estimated with PBE already show very good agreement with high-level references [MP2/CBS or CCSD(T)/aug-cc-pVTZ level of theory]. When it comes to predict the properties of liquid water, however, the performance is rather dismal: PBE water, as shown by the radial and angular distribution functions, is too structured and all dynamical properties investigated are too slow, even more so than the BLYP water. Augmenting PBE with either DCACPs or empirical vdW corrections change the interaction energy surfaces of the water dimer and the equilibrium energies of the water clusters to a very small extent, similar to the minimum influence it exerts on liquid simulations.

As discussed in Chapter 4, BLYP in which dispersion interactions are entirely absent predicts repulsive interaction energies for all vdW-dominated complexes such as rare-gas and hydrogen dimers [12, 13, 15, 19–21]. PBE, on the other hand, is rather system-dependent and shows spurious interactions for some dispersion-bound systems: for example, the interaction energy of $(\text{H}_2)_2$, the calibration system for hydrogen DCACP, estimated by PBE is over-binding with respect to the full CI reference; for $(\text{CO}_2)_2$, it is weakly bound. In consequence, DCACPs complementing BLYP always provide an attractive cor-

rection to the underlying functional, in line with the idea of a dispersion-motivated correction. The DCACP parameters for PBE, however, can be repulsive (hydrogen and helium) or attractive (carbon, nitrogen, oxygen, *etc.*). In addition, the strong angular dependence in the difference between the interaction energy predicted by PBE and high-level calculations cannot be efficiently corrected for by simply adding the empirical vdW corrections that are purely radial-dependent or DCACPs whose effects are only weakly angular-dependent.

8.4 Conclusions

As a natural continuation of Chapter 7, we have carried out AIMD simulations on liquid water using pure and dispersion-corrected density functional theory combined with the PBE functional. Two empirical approaches, DCACPs and van der Waals corrections, are employed to compensate for the ill description of vdW interactions in DFT-GGAs.

All structural and dynamical results point to a much more glass-like liquid water with PBE, even when compared with the already over-structured BLYP. In addition, unlike with BLYP in which the addition of either DCACPs or empirical vdW corrections leads to superior interaction energies for water clusters and a softer liquid structure, augmenting PBE with either of the two does not seem to cure the discrepancy, if not to enhance the structure even more.

Nevertheless, we are still convinced that a proper treatment of vdW interactions is essential in liquid water simulations, not only for a better description of its properties but more importantly, for studying the many phenomena occurring in wet environment such as hydrophobicity. Yet the ‘spurious’ behavior of PBE in treating dispersion forces renders the corresponding DCACPs less transferable than the ones for BLYP or BP where the contribution of DCACPs is much more clear-cut.

Chapter 9

Conclusions and Outlook

London dispersion forces play a crucial role in many phenomena and processes in chemical and biological systems such as physisorption and molecular recognition. Yet KS-DFT combined with many popular approximated exchange-correlation functionals fails to capture these non-local long-range forces. This thesis aims to contribute to the active field of research on curing this deficiency, in particular, we focus on the recently introduced approach of DCACPs that model the effects of these weak forces via atom-electron potentials instead of the usual atom-atom potentials of the R^{-6} form.

A library of DCACPs for rare-gas atoms, hydrogen, carbon, nitrogen and oxygen to be used with BLYP, BP and PBE are calibrated against CCSD(T) or full CI references of small vdW complexes. The values of DCACP parameters depend heavily on the performance of the underlying exchange-correlation functionals. Since PBE provides spurious interactions for some vdW complexes, DCACPs complementing PBE can be either attractive or repulsive depending on the element in question. On the other hand, for BP and BLYP in which dispersion interactions are entirely absent, DCACPs are always attractive, in line with the dispersion-motivated corrections. As a consequence, the latter combinations, in our view, are preferable over the occasional superior performance of DCACP-PBE due to the more clean-cut interpretations they offer.

By augmenting the BLYP functional with the corresponding DCACPs, $\pi - \pi$ interactions between the intercalator and aromatic biomolecules are predicted remarkably well in all systems studied, including neutral as well as positively charged complexes. When tested against a well-established benchmark suite of noncovalently bound nucleic acids and amino acids, unsigned mean errors of <1.6 kcal/mol compared with the best *ab initio* data available are obtained. The overall descriptions for systems that are bound predominantly by interactions other than dispersion forces are also, in general, further improved with DCACPs. In summary, the results demonstrate that the effect of dispersion forces can be well described with DCACP-augmented DFT-GGA without introducing any significant distortions on intramolecular geometries or electronic structures. Furthermore, they indicate that DCACPs display a strong transferability to systems other than the calibration ones, *i.e.*, once calibrated, the DCACP for a specific element can be applied in different chemical environments with comparable accuracy.

Chemical insights to the role vdW interactions play are also gained. For example, H-bonded base pairs are deformed noticeably to achieve more favorable multipole–charge interactions in all charged nucleobase–intercalator complexes. The intercalation process in the ellipticine–d(CG)₂ complex is en-

ergetically favorable – the stabilization energy when fully intercalated is estimated to be approximately 40 kcal/mol. By curing the shortcomings of DFT-BLYP in describing vdW interactions, the structure of liquid water becomes much softer, and its dynamical properties are much more liquid-like; the self-diffusion coefficient, for example, is increased by almost threefold. A proper description of vdW interactions thus brings the simulated properties in closer agreement to experiments, showing that even interactions as weak as dispersion forces may significantly influence the structural and dynamical properties of systems believed to be held together by much stronger interactions. Surprisingly, the PBE functional whose description of water clusters is one of the closest to the best *ab initio* estimates, fails to reach the same level of accuracy when it comes to the liquid state. Including either DCACPs or empirical vdW terms has negligible effects, probably due to the problem of transferability mentioned above.

Owing to the relative infancy of the DCACP approach, there are still plenty of questions left to be addressed in this field; for instance, the performance on chemical reactions, such as the prediction of reaction barriers, remains to be tested. In addition, DCACPs do not give the asymptotic R^{-6} tail, characteristic of the interaction energy of two systems with no permanent multipole moment. Tests on many different systems have nevertheless shown that the shape of the interaction profile is closely reproduced within distances up to 5 Å. The introduction of the midpoint term in the penalty functional is the first step taken for an improved long-range behavior, but a systematic way to include the correct asymptotic tail should be sought out for applications sensitive to dispersion interactions at the long-range limit, *e.g.*, small-angle scattering of rare gases in molecular beams; one possibility is to include more ℓ in the expansion of DCACPs. Perhaps the most important of all is to develop a rigorous physical interpretation of this approach based on fundamental principles.

AIMD simulations that utilize DFT for solving electronic structures have been instrumental in elucidating microscopic mechanisms in an accurate and unbiased manner in various disciplines. Their predictive power, however, depends strongly on the accuracy of the chosen exchange-correlation functional. We show that DCACPs allow one to achieve accuracy similar to coupled-cluster calculations with much lower computational overhead. Because DCACPs are electronic potentials, their contributions to the total energy depend explicitly on the wave function; many-body interactions are therefore implicitly included, offering an advantage over the pairwise approach of empirical vdW corrections that is less demanding in computer resources. Adopting this approach in AIMD simulations or DFT-based quantum

mechanical/molecular mechanical calculations are relatively straightforward owing to the availability of the easy-to-use library for many common elements in biochemical systems. With its efficiency and proven robustness, we hope that it will be widely used and contribute to the fields of biochemistry and biophysics, allowing for novel applications such as adsorption on surfaces, investigating reactions in environments where $\pi - \pi$ interactions are paramount, or generating new classical force fields for the description of DNA-intercalator interactions.

Bibliography

- [1] F. Müller-Plathe. *Chem. Phys. Chem.*, 3:754, 2002.
- [2] G. Gsányi, T. Albaret, G. Moras, M. C. Payne, and A. De Vita. *J. Phys: Condens. Matter*, 17, 2005.
- [3] C. Y. Wang and X. Zhang. *Curr. Opin. Solid State Mater. Sci.*, 10:2, 2006.
- [4] A. Warshel and M. Levitt. *J. Mol. Biol.*, 103:227, 1976.
- [5] U. C. Singh and P. A. Kollmann. *J. Comput. Chem.*, 7:718, 1986.
- [6] M. J. Field, P. A. Bash, and M. Karplus. *J. Comput. Chem.*, 11:770, 1990.
- [7] P. Hohenberg and W. Kohn. *Phys. Rev.*, 136:B864, 1964.
- [8] W. Kohn and L. J. Sham. *Phys. Rev.*, 140:A1133, 1965.
- [9] G. E. Scuseria. *J. Phys. Chem. A*, 103:4782, 1999.
- [10] M. Elstner, T. Frauenheim, and S. Suhai. *J. Mol. Struct. (THEOCHEM)*, 632:29, 2003.
- [11] P. A. Madden, R. Heaton, A. Aguado, and S. Jahn. *J. Mol. Struct. (THEOCHEM)*, 771:9, 2006.
- [12] S. Kristyán and P. Pulay. *Chem. Phys. Lett.*, 229:175, 1994.
- [13] J. M. Pérez-Jordá and A. D. Becke. *Chem. Phys. Lett.*, 233:134, 1995.
- [14] S. M. Cybulski and C. E. Severesen. *J. Chem. Phys.*, 122:014117, 2005.
- [15] Zhang Y, W. Pan, and W. Yang. *J. Chem. Phys.*, 107:7921, 1997.

- [16] X. Wu, M. C. Vargas, S. Nayak, V. Lotrich, and G. Scoles. *J. Chem. Phys.*, 115:8748, 2001.
- [17] A. D. Becke. *Phys. Rev. A*, 38:3098, 1988.
- [18] C. Lee, W. Yang, and R. G. Parr. *Phys. Rev. B*, 37:785, 1988.
- [19] D. C. Patton and M. R. Pederson. *Int. J. Quantum Chem.*, 69:619, 1998.
- [20] T. van Mourik and R. J. Gdanitz. *J. Chem. Phys.*, 116:9620, 2002.
- [21] J. Tao and J. P. Perdew. *J. Chem. Phys.*, 122:114102, 2005.
- [22] P. Hobza, J. Šponer, and T. Reschel. *J. Comput. Chem.*, 16:1315, 1995.
- [23] E. J. Meijer and M. Sprik. *J. Chem. Phys.*, 105:8684, 1996.
- [24] P. Hobza and J. Šponer. *Chem. Rev.*, 99:3247, 1999.
- [25] I. Dąbkowska, P. Jurečka, and P. Hobza. *J. Chem. Phys.*, 122:204322, 2005.
- [26] J. Černý and P. Hobza. *Phys. Chem. Chem. Phys.*, 7:1624, 2005.
- [27] Y. Zhao and D. G. Truhlar. *J. Chem. Theory Comput.*, 3:289, 2007.
- [28] W. Kohn, Y. Meir, and D. E. Makarov. *Phys. Rev. Lett.*, 80:4153, 1998.
- [29] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist. *Phys. Rev. Lett.*, 92:246401, 2004.
- [30] S. Grimme. *J. Comput. Chem.*, 25:1463, 2004.
- [31] J. F. Dobson, J. Wang, B. P. Dinte, K. McLennan, and H. M. Le. *Int. J. Quantum Chem.*, 101:579, 2005.
- [32] A. J. Misquitta, R. Podeszwa, B. Jeziorski, and K. Szalewicz. *J. Chem. Phys.*, 123:214103, 2005.
- [33] A. D. Becke and E. R. Johnson. *J. Chem. Phys.*, 124:014104, 2006.
- [34] O. A. von Lilienfeld, I. Tavernelli, U. Rothlisberger, and D. Sebastiani. *Phys. Rev. Lett.*, 93:153004, 2004.

- [35] W. Kohn, A. D. Becke, and R. G. Parr. *J. Phys. Chem.*, 100:12974, 1996.
- [36] R. G. Parr and W. Yang. *Density Functional Theory of Atoms and Molecules*. Oxford Science Publications, 1989.
- [37] R. M. Dreizler and E. K. U. Gross. *Density Functional Theory: An Approach to the Quantum Many-Body Problem*. Springer-Verlag, 1991.
- [38] W. Koch and M. C. Holthausen. *A Chemist's Guide to Density Functional Theory*. Wiley-VCH, 2002.
- [39] D. M. Ceperley and B. J. Alder. *Phys. Rev. Lett.*, 45:566, 1980.
- [40] S. J. Vosko, L. Wilk, and M. Nusair. *Can. J. Phys.*, 58:1200, 1980.
- [41] P. A. M. Dirac. *Cambridge Phil. Soc.*, 26:376, 1930.
- [42] J. P. Perdew. *Phys. Rev. B*, 33:8822, 1986.
- [43] R. Colle and D. Salvetti. *Theor. Chim. Acta*, 37:329, 1975.
- [44] A. D. Becke. *J. Chem. Phys.*, 98:5648, 1993.
- [45] P. Pulay. *Mol. Phys.*, 17:197, 1969.
- [46] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes*. Cambridge University Press, 1986.
- [47] D. R. Hamann, M. Schlüter, and C. Chiang. *Phys. Rev. Lett.*, 43:1494, 1980.
- [48] G. B. Bachelet, D. R. Hamann, and M. Schlüter. *Phys. Rev. B*, 26:4199, 1982.
- [49] L. Kleinman and D. M. Bylander. *Phys. Rev. Lett.*, 48:1425, 1982.
- [50] D. Vanderbilt. *Phys. Rev. B*, 41:7892, 1990.
- [51] N. Troullier and J. L. Martins. *Phys. Rev. B*, 43:1993, 1991.
- [52] S. Goedecker, M. Teter, and J. Hutter. *Phys. Rev. B*, 54:1703, 1996.
- [53] W. E. Pickett. *Comput. Phys. Rep.*, 9:115, 1989.

- [54] M. Fuchs and M. Scheffler. *Comput. Phys. Commun.*, 119:67, 1999.
- [55] R. Car and M. Parrinello. *Phys. Rev. Lett.*, 55:2471, 1985.
- [56] D. Marx and J. Hutter. *Ab Initio molecular dynamics: Theory and implementation*. In *Modern Methods and Algorithms of Quantum Chemistry*, volume 1, page 301, 2000. <http://www.fz-juelich.de/wsqc/proceedings/>.
- [57] R. Iftimie, P. Minary, and M. Tuckerman. *Proc. Natl. Acad. Sci. USA*, 102:6654, 2005.
- [58] P. W. Atkins and R. S. Friedman. *Molecular Quantum Mechanics*. Oxford University Press, Oxford, 1997.
- [59] T. A. Wesolowski and F. Tran. *J. Chem. Phys.*, 118:2072, 2003.
- [60] M. Dulak and T. A. Wesolowski. *Int. J. Quantum Chem.*, 101:543, 2005.
- [61] F. Tran and T. A. Wesolowski. *Int. J. Quantum Chem.*, 101:854, 2005.
- [62] Y. Andersson, D. C. Langreth, and B. I. Lundqvist. *Phys. Rev. Lett.*, 76:102, 1996.
- [63] H. Rydberg, B. I. Lundqvist, D. C. Langreth, and M. Dion. *Phys. Rev. B*, 62:6997, 2000.
- [64] H. Rydberg, M. Dion, N. Jacobson, E. Schröder, P. Hyldgaard, S. I. Simak, D. C. Langreth, and B. I. Lundqvist. *Phys. Rev. Lett.*, 91:126402, 2003.
- [65] D. C. Langreth, M. Dion, H. Rydberg, E. Schröder, P. Hyldgaard, and B. I. Lundqvist. *Int. J. Quantum Chem.*, 101:599, 2005.
- [66] A. Puzder, M. Dion, and D. C. Langreth. *J. Chem. Phys.*, 124:164105, 2006.
- [67] T. Thonhauser, A. Puzder, and D. C. Langreth. *J. Chem. Phys.*, 124:164106, 2006.
- [68] H. Iikura, T. Tsuneda, T. Yanai, and K. Hirao. *J. Chem. Phys.*, 115:3540, 2001.
- [69] T. Sato, T. Tsuneda, and K. Hirao. *J. Chem. Phys.*, 123:104307, 2005.
- [70] Y. Zhao, N. E. Schultz, and D. G. Truhlar. *J. Chem. Theory Comput.*, 2:364, 2006.
- [71] A. J. Misquitta, B. Jeziorski, and K. Szalewicz. *Phys. Rev. Lett.*, 91:033201, 2003.

- [72] M. Piacenza and S. Grimme. *Chem. Phys. Chem.*, 6:1554, 2005.
- [73] E. R. Johnson and A. D. Becke. *J. Chem. Phys.*, 123:024101, 2005.
- [74] A. D. Becke and E. R. Johnson. *J. Chem. Phys.*, 123:154101, 2005.
- [75] J. G. Ángyán. *J. Chem. Phys.*, 127:024108, 2007.
- [76] I-C. Lin, M. C. Coutinho-Neto, C. Felsenheimer, O. A. von Lilienfeld, I. Tavernelli, and U. Rothlisberger. *Phys. Rev. B*, 75:205131, 2007.
- [77] S. F. Boys and P. Rajagopal. *Adv. Quantum Chem.*, 2:1, 1965.
- [78] A. D. Becke. *J. Chem. Phys.*, 88:2547, 1988.
- [79] B. Delley. *J. Chem. Phys.*, 92:508, 1990.
- [80] F. W. Averill and G. S. Painter. *Phys. Rev. B*, 50:7262, 1994.
- [81] T. Kato. *Pure. Appl. Math.*, 10:151, 1957.
- [82] Á. Nagy. *Int. J. Quantum Chem.*, 70:681, 1998.
- [83] S. Liu, R. G. Parr, and A. Nagy. *Phys. Rev. A*, 52:2645, 1995.
- [84] R. F. W. Bader. *Atoms in Molecules: A Quantum Theory*. Oxford University Press, USA, 1994.
- [85] R. F. W. Bader, T. T. Nguyen-Dang, and Y. Tal. *Rep. Prog. Phys.*, 44:893, 1981.
- [86] R. F. W. Bader, Y. Tal, S. G. Anderson, and T. T. Nguyen-Dang. *Israel J. Chem.*, 19:8, 1980.
- [87] J. C. Angulo and J. S. Dehesa. *Phys. Rev. A*, 44:1516, 1991.
- [88] O. A. von Lilienfeld, I. Tavernelli, U. Rothlisberger, and D. Sebastiani. *J. Chem. Phys.*, 122:014113, 2005.
- [89] O. A. von Lilienfeld, I. Tavernelli, U. Rothlisberger, and D. Sebastiani. *Phys. Rev. B*, 71:195119, 2005.
- [90] E. Tapavicza, I-C. Lin, O. A. von Lilienfeld, I. Tavernelli, M. C. Coutinho-Neto, and U. Rothlisberger. *J. Chem. Theory Comput.*, 3:1673, 2007.

- [91] I-C. Lin, O. A. von Lilienfeld, M. D. Coutinho-Neto, I. Tavernelli, and U. Rothlisberger. *J. Phys. Chem. B*, 111:14346, 2007.
- [92] I-C. Lin and U. Rothlisberger. *Phys. Chem. Chem. Phys.*, DOI: 10.1039/b718594d, 2008.
- [93] CPMD, version 3.8. Copyright IBM Corp. 1990-2006, copyright MPI für Festkörperforschung Stuttgart 1997-2001, <http://www.cpmc.org>.
- [94] H. Lavendy, J. M. Robbe, and J. P. Flament. *Chem. Phys. Lett.*, 196:377, 1992.
- [95] A. Tkatchenko and O. A. von Lilienfeld. *Phys. Rev. B*, 73:153406, 2006.
- [96] O. A. von Lilienfeld and D. Andrienko. *J. Chem. Phys.*, 124:054307, 2006.
- [97] J. P. Perdew, K. Burke, and M. Ernzerhof. *Phys. Rev. Lett.*, 77:3865, 1996.
- [98] M. O. Sinnokrot and C. D. Sherrill. *J. Phys. Chem. A*, 108:10200, 2004.
- [99] N. Kurita and H. Sekino. *Int. J. Quantum Chem.*, 91:355, 2003.
- [100] G. J. Martyna and M. E. Tuckerman. *J. Chem. Phys.*, 110:2810, 1999.
- [101] S. F. Boys and F. Bernardi. *Mol. Phys.*, 19:553, 1970.
- [102] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, and *et al.* Gaussian 03, Revision C.02. Gaussian, Inc., Wallingford, CT, 2004.
- [103] A. Ferre-Vilaplana. *J. Chem. Phys.*, 122:104709, 2005.
- [104] A. D. Corso, A. Pasquarello, A. Baldereschi, and R. Car. *Phys. Rev. B*, 53:1180, 1996.
- [105] K. E. Riley, B. T. Op't Holt, and K. M. Merz Jr. *J. Chem. Theory Comput.*, 3:407, 2007.
- [106] K. Patel, P. B. Butler, A. M. Ellis, and M. D. Wheeler. *J. Chem. Phys.*, 119:909, 2003.
- [107] D. Voet and J. Voet. *Biochemistry*. Wiley, 1995.
- [108] M. F. Brana, M. Cacho, A. Gradillas, B. de Pascual-Teresa, and A. Ramos. *Curr. Pharm. Design*, 7:1745, 2001.

- [109] W. A. Denny and B. C. Baguley. *Curr. Top. Med. Chem.*, 3:339, 2003.
- [110] E. C. Long and J. K. Barton. *Acc. Chem. Res.*, 23:271, 1990.
- [111] M. V. Keck and S. J. Lippard. *J. Am. Chem. Soc.*, 114:3386, 1992.
- [112] J. M. Perez, I. Lopez-Solera, E. I. Montero, M. F. Brana, C. Alonso, S. P. Robinson, and C. Navarro-Ranninger. *J. Med. Chem.*, 42:5482, 1999.
- [113] H. Baruah, C. L. Rector, S. M. Monnier, and U. Bierbach. *Biochem. Pharmacol.*, 64:191, 2002.
- [114] E. J. Fechter and P. B. Dervan. *J. Am. Chem. Soc.*, 125:8476, 2003.
- [115] A. H. Elcock, A. Rodger, and W. G. Richards. *Biopolymers*, 39:309, 1996.
- [116] C. Courseille, B. Busetta, and M. Hospital. *Acta Cryst.*, B30:2628, 1974.
- [117] M. Stiborová, C. A. Bieler, M. Wiessler, and E. Frei. *Biochem. Pharmacol.*, 62:1675, 2001.
- [118] N. Patel, J. Bergman, and A. Graslund. *Eur. J. Biochem.*, 197:597, 1991.
- [119] C. A. Hunter and J. K. M. Sanders. *J. Am. Chem. Soc.*, 112:5525, 1990.
- [120] R. A. Friedman and B. Honig. *Biopolymers*, 32:145, 1992.
- [121] P. Hobza, J. Šponer, and M. Polášek. *J. Am. Chem. Soc.*, 117:792, 1995.
- [122] J. Šponer, J. Leszczyński, and P. Hobza. *J. Phys. Chem.*, 100:5590, 1996.
- [123] J. Šponer and P. Hobza. *Chem. Phys. Lett.*, 267:263, 1997.
- [124] S. Tsuzuki, K. Honda, and R. Azumi. *J. Am. Chem. Soc.*, 124:12200, 2002.
- [125] P. Hobza and J. Šponer. *J. Am. Chem. Soc.*, 124:11802, 2002.
- [126] B. W. Hopkins and G. S. Tschumper. *J. Phys. Chem. A*, 108:2941, 2004.
- [127] B. K. Mishra and N. Sathyamurthy. *J. Phys. Chem. A*, 109:6, 2005.
- [128] J. Šponer, J. Leszczyński, and P. Hobza. *J. Comput. Chem.*, 17:841, 1996.
- [129] M. Elstner, P. Hobza, T. Frauenheim, S. Suhai, and E. Kaxiras. *J. Chem. Phys.*, 114:5149, 2001.

- [130] P. Jurečka and P. Hobza. *J. Am. Chem. Soc.*, 125:15608, 2003.
- [131] J. Šponer, P. Jurečka, and P. Hobza. *J. Am. Chem. Soc.*, 126:10142, 2004.
- [132] P. Jurečka, J. Šponer, and P. Hobza. *J. Phys. Chem. B*, 108:5466, 2004.
- [133] I. Dąbkowska, H. V. Gonzalez, P. Jurečka, and P. Hobza. *J. Phys. Chem. A*, 109:1131, 2005.
- [134] M. Kabeláč, L. Zendlova, D. Řeha, and P. Hobza. *J. Phys. Chem. B*, 109:12206, 2005.
- [135] P. Jurečka, J. Šponer, J. Černý, and P. Hobza. *Phys. Chem. Chem. Phys.*, 8:1985, 2006.
- [136] X. Xu and W. A. Goddard III. *Proc. Natl. Acad. Sci. USA*, 101:2673, 2004.
- [137] J. T. Su, X. Xu, and W. A. Goddard III. *J. Phys. Chem. A*, 108:10518, 2004.
- [138] Y. Zhao and D. G. Truhlar. *J. Phys. Chem. A*, 109:5656, 2005.
- [139] A. Hesselmann and G. Jansen. *Phys. Chem. Chem. Phys.*, 5:5010, 2003.
- [140] A. Hesselmann, G. Jansen, and M. Schutz. *J. Am. Chem. Soc.*, 128:11730, 2006.
- [141] J. D. Watson and F. H. C. Crick. *Nature*, 171:737, 1953.
- [142] A. Canals, M. Purciolas, J. Aymani, and M. Coll. *Acta. Cryst.*, D61:1009, 2005.
- [143] K. Grzeskowiak, K. Yanagi, G. G. Prive, and R. E. Dickerson. *J. Biol. Chem.*, 266:8861, 1991.
- [144] C. Morgado, M. A. Vincent, I. H. Hillier, and X. Shan. *Phys. Chem. Chem. Phys.*, 9:448, 2007.
- [145] Y. Zhao and D. G. Truhlar. *Phys. Chem. Chem. Phys.*, 7:2701, 2005.
- [146] M. Dračinský and O. Castano. *Phys. Chem. Chem. Phys.*, 6:1799, 2004.
- [147] R. G. Endres, D. L. Cox, and R. R. P. Singh. *Rev. Mod. Phys.*, 76:195, 2004.
- [148] E. Yavin, E. D. A. Stemp, V. L. O'Shea, S. S. David, and J. K. Barton. *Proc. Natl. Acad. Sci. USA*, 103:3610, 2006.
- [149] A. Laio, J. VandeVondele, and U. Rothlisberger. *J. Chem. Phys.*, 116:6941, 2002.

- [150] I-C. Lin, A. P. Seitsonen, I. Tavernelli, M. D. Coutinho-Neto, and U. Rothlisberger. submitted, 2008.
- [151] P. Aeberhard, J. S. Arey, I-C. Lin, and U. Rothlisberger. in preparation, 2008.
- [152] J. Antony and S. Grimme. *Phys. Chem. Chem. Phys.*, 8:5287, 2006.
- [153] B. J. Guillot. *J. Mol. Liq.*, 101:219, 2002.
- [154] D. Kennedy and C. Norman. *Science*, 309:75, 2005.
- [155] T. Head-Gordon and G. Hura. *Chem. Rev.*, 102:2651, 2002.
- [156] A. K. Soper. *Chem. Phys.*, 258:121, 2000.
- [157] Ph. Wernet, D. Nordlund, U. Bergmann, M. Cavalleri, M. Odelius, H. Ogasawara, L. Å. Näslund, T. K. Hirsch, L. Ojamäe, P. Glatzel, L. G. M. Pettersson, and A. Nilsson. *Science*, 304:995, 2004.
- [158] T. Head-Gordon and M. E. Johnson. *Proc. Natl. Acad. Sci. USA*, 103:7973, 2006.
- [159] J. M. Pedulla, K. Kim, and K. D. Jordan. *Chem. Phys. Lett.*, 291:78, 1998.
- [160] S. S. Xantheas. *Chem. Phys.*, 258:225, 2000.
- [161] B. Kirchner. *J. Chem. Phys.*, 123:204116, 2005.
- [162] E. E. Dahlke and D. G. Truhlar. *J. Phys. Chem. B*, 110:10595, 2006.
- [163] G. Corongiu and E. Clementi. *J. Chem. Phys.*, 97:2030, 1992.
- [164] H. Huber, A. J. Dyson, and B. Kirchner. *Chem. Soc. Rev.*, 28:121, 1999.
- [165] G. Hura, D. Russo, R. M. Glaeser, T. Head-Gordon, M. Krack, and M. Parrinello. *Phys. Chem. Chem. Phys.*, 5:1981, 2003.
- [166] J. VandeVondele, F. Mohamed, M. Krack, J. Hutter, M. Sprik, and M. Parrinello. *J. Chem. Phys.*, 122:014515, 2005.
- [167] H.-S. Lee and M. E. Tuckerman. *J. Chem. Phys.*, 126:164501, 2007.

- [168] T. Todorova, A. P. Seitsonen, J. Hutter, I-F. W. Kuo, and C. J. Mundy. *J. Phys. Chem. B*, 110:3685, 2006.
- [169] M. E. Tuckerman, D. Marx, M. L. Klein, and M. Parrinello. *Science*, 275:817, 1997.
- [170] E. Schwegler, J. C. Grossman, F. Gygi, and G. Galli. *J. Chem. Phys.*, 121:5400, 2004.
- [171] G. S. Fanourgakis, G. K. Schenter, and S. S. Xantheas. *J. Chem. Phys.*, 125:141102, 2006.
- [172] B. Chen, I. Ivanov, M. L. Klein, and M. Parrinello. *Phys. Rev. Lett.*, 91:215503, 2003.
- [173] D. Eisenberg and W. Kauzmann. *The Structure and Properties of Water*. Clarendon Press, Oxford, 1969.
- [174] F. H. Stillinger. *Science*, 209:451, 1980.
- [175] O. Mishima and H. E. Stanley. *Nature*, 396:329, 1998.
- [176] C. H. Cho, S. Singh, and G. W. Robinson. *J. Chem. Phys.*, 107:7979, 1997.
- [177] R. Schmid. *Monatsh. Chem.*, 132:1295, 2001.
- [178] R. M. Lynden-Bell and P. G. Debenedetti. *J. Phys. Chem. B*, 109:6527, 2005.
- [179] R. W. Williams and D. Malhotra. *Chem. Phys.*, 327:54, 2006.
- [180] I-F. W. Kuo, C. J. Mundy, M. J. McGrath, J. I. Siepmann, J. VandeVondele, M. Sprik, J. Hutter, B. Chen, M. L. Klein, F. Mohamed, M. Krack, and M. Parrinello. *J. Phys. Chem. B*, 108:12990, 2004.
- [181] L. Hernández de la Peña and P. G. Kusalik. *J. Chem. Phys.*, 121:5992, 2004.
- [182] N. Marzari and D. Vanderbilt. *Phys. Rev. B*, 56:12847, 1997.
- [183] G. Berghold, C. J. Mundy, A. H. Romero, J. Hutter, and M. Parrinello. *Phys. Rev. B*, 61:10040, 2000.
- [184] H-S Lee and M. E. Tuckerman. *J. Phys. Chem. A*, 110:5549, 2006.
- [185] G. Hura, J. M. Sorenson, R. M. Glaeser, and T. Head-Gordon. *J. Chem. Phys.*, 113:9140, 2000.

- [186] J. Hutter. Private communication.
- [187] Y. A. Mantz, B. Chen, and G. J. Martyna. *J. Phys. Chem. B*, 110:3540, 2006.
- [188] K. Modig, B. G. Pfrommer, and B. Halle. *Phys. Rev. Lett.*, 90:075502, 2003.
- [189] R. Mills. *J. Phys. Chem.*, 77:685, 1973.
- [190] M. S. P. Sansom, I. D. Kerr, J. Breed, and R. Sankararamakrishnan. *Biophys. J*, 70:693, 1996.
- [191] P. Kumar, G. Franzese, S. V. Buldyrev, and H. E. Stanley. *Phys. Rev. E*, 73:041505, 2006.
- [192] Y. S. Badyal, M.-L. Saboungi, D. L. Price, S. D. Shastri, D. R. Haeffner, and A. K. Soper. *J. Chem. Phys.*, 112:9206, 2000.
- [193] W. Klopper, J. G. C. M. van Duijneveldt-van de Rijdt, and F. B. van Duijneveldt. *Phys. Chem. Chem. Phys.*, 2:2227, 2000.
- [194] I. M. B. Nielsen, E. T. Seidl, and C. L. Janssen. *J. Chem. Phys.*, 110:9435, 1999.
- [195] I. G. Gurtubay and R. J. Needs. *J. Chem. Phys.*, 127:124306, 2007.
- [196] B. Santra, A. Michaelides, and M. Scheffler. *J. Chem. Phys.*, 127:184104, 2007.
- [197] M. J. McGrath, J. I. Siepmann, I-F. W. Kuo, C. J. Mundy, J. VandeVondele, J. Hutter, F. Mohamed, and M. Krack. *Chem. Phys. Chem.*, 6:1894, 2005.
- [198] H.-S. Lee and M. E. Tuckerman. *J. Chem. Phys.*, 125:154507, 2006.
- [199] D. Laage and J. T. Hynes. *Science*, 311:832, 2006.
- [200] F. Paesani, S. Iuchi, and G. A. Voth. *J. Chem. Phys.*, 127:074506, 2007.
- [201] P. Raiteri, A. Laio, and M. Parrinello. *Phys. Rev. Lett.*, 93:087801, 2004.
- [202] M. V. Fernández-Serra and E. Artacho. *Phys. Rev. Lett.*, 96:016404, 2006.
- [203] D. Prendergast and G. Galli. *Phys. Rev. Lett.*, 96:215502, 2006.
- [204] J. C. Grossman, E. Schwegler, E. W. Draeger, F. Gygi, and G. Galli. *J. Chem. Phys.*, 120:300, 2004.

-
- [205] M. V. Fernández-Serra and E. Artacho. *Phys. Rev. Lett.*, 96:016404, 2006.
- [206] P. H.-L. Sit and N. Marzari. *J. Chem. Phys.*, 122:204510, 2005.
- [207] P. Madden and D. Kivelson. *Adv. Chem. Phys.*, 56:467, 1984.
- [208] A. Wallqvist and B. J. Berne. *J. Chem. Phys.*, 97:13841, 1993.
- [209] B. Bagchi. *Chem. Rev.*, 105:3197, 2005.
- [210] C. P. Lawrence and J. K. Skinner. *J. Chem. Phys.*, 118:264, 2003.
- [211] H. S. Tan, I. R. Piletic, and M. D. Fayer. *J. Chem. Phys.*, 122:174501, 2005.
- [212] Y. L. A. Rezus and H. J. Bakker. *J. Chem. Phys.*, 123:114502, 2005.
- [213] S. S. Xantheas, C. J. Burnham, and R. G. Harrison. *J. Chem. Phys.*, 116:1493, 2002.
- [214] A. Milet, T. Korona, R. Moszynski, and E. Kochanski. *J. Chem. Phys.*, 111:7727, 1999.
- [215] J. Ireta, J. Neugebauer, and M. Scheffler. *J. Phys. Chem. A*, 108:5692, 2004.

Appendix

Convert DCACPs into the format adapted by Troullier-Martins pseudopotentials

The ionic pseudopotential operator is

$$\hat{V}_{\text{ion}}^{\text{PP}}(r) = V_{\text{ion,local}}^{\text{PP}}(r) + \sum_l V_{\text{semilocal},l}(r) \hat{P}_l$$

where $V_{\text{ion,local}}^{\text{PP}}(r)$ is the local potential;

$$V_{\text{semilocal},l}(r) \hat{P}_l = \Delta V_l(r) = V_{\text{ion},l}^{\text{PP}}(r) - V_{\text{ion,local}}^{\text{PP}}(r)$$

is the semi-local potential for the angular momentum component l , and \hat{P}_l projects out the l th angular momentum component from the wave function. The semi-local form can be transformed into a non-local one using the procedure suggested by Kleinman and Bylander (KB) [49]:

$$V_{\text{nonlocal},l}^{\text{KB}}(\mathbf{r}, \mathbf{r}') = \sum_{m=-l}^l Y_{lm}(\hat{\mathbf{r}}) \frac{|\Delta V_l(r) \phi_l(r)\rangle \langle \Delta V_l(r') \phi_l(r')|}{\langle \phi_l(r'') | \Delta V_l(r'') | \phi_l(r'') \rangle} Y_{lm}(\hat{\mathbf{r}}')$$

where $\phi_l(r)$ is the atomic reference pseudo wave function and Y_{lm} the spherical harmonics.

DCACPs are of the same analytical form as the non-local part of the pseudopotentials developed by Goedecker, Teter, and Hutter [52]:

$$v^{\text{DCACP}}(\mathbf{r}, \mathbf{r}') = \sum_{m=-l}^l Y_{lm}(\hat{\mathbf{r}}) p_{l1}(r) \sigma_1 p_{l1}(r') Y_{lm}(\hat{\mathbf{r}}').$$

To cast DCACPs in the Troullier-Martins format, one can thus write

$$p_{l1}(r) \sigma_1 p_{l1}(r') = \frac{|\Delta V_l(r) \phi_l(r)\rangle \langle \Delta V_l(r') \phi_l(r')|}{\langle \phi_l(r'') | \Delta V_l(r'') | \phi_l(r'') \rangle}.$$

If we chose $\Delta V_l(r)$ to be constant, then

$$\begin{aligned} p_{l1}(r) \sigma_1 p_{l1}(r') &= \phi_l(r) \frac{\Delta V_l}{\langle \phi_l(r'') | \phi_l(r'') \rangle} \phi_l(r'); \\ p_{l1}(r) &= \phi_l(r), \\ \sigma_1 &= \frac{\Delta V_l}{\langle \phi_l(r'') | \phi_l(r'') \rangle}, \\ p_{l1}(r) &= \frac{\sqrt{2} r^l \exp\left(-\frac{r^2}{2r_l^2}\right)}{r_l^{l+3/2} \sqrt{\Gamma(l+3/2)}}, \end{aligned}$$

Γ denotes the gamma function.

Since the projector satisfies the normalisation condition:

$$\int_0^\infty p_{l1}(r) p_{l1}(r) r^2 dr = 1.$$

We have to write in the Troullier-Martins pseudopotential file (as in the CPMD code)

- the “wave function”: $r p_{l1}$
- the “potential”: $\sigma_1 \langle \phi_l | \phi_l \rangle + V_{\text{loc}}(r) \rightarrow \sigma_1 + V_{\text{loc}}(r)$

Publications and Presentations

Publications

• In press or submitted

1. Importance of van der Waals interactions in liquid water
I-C. Lin, A. P. Seitsonen, I. Tavernelli, M. D. Coutinho-Neto, and U. Rothlisberger; submitted (2008)
2. Atom-centered potentials for describing dispersion forces in density functional theory
I-C. Lin and U. Rothlisberger; *Chimia*, accepted (2008)
3. Describing weak interactions of biomolecules with dispersion-corrected density functional theory
I-C. Lin and U. Rothlisberger; *Phys. Chem. Chem. Phys.*, in press, DOI: 10.1039/b718594d (2008)
4. Predicting noncovalent interactions between aromatic biomolecules with London-dispersion-corrected DFT
I-C. Lin, O. A. von Lilienfeld, M. D. Coutinho-Neto, I. Tavernelli, and U. Rothlisberger; *J. Phys. Chem. B*, **111**, 14346, (2007)
5. Weakly bonded complexes of aliphatic and aromatic carbon compounds described with dispersion corrected density functional theory
E. Tapavicza, I-C. Lin, O. A. von Lilienfeld, I. Tavernelli, M. D. Coutinho-Neto, and U. Rothlisberger; *J. Chem. Theory Comput.*, **3**, 1673, (2007)
6. Library of dispersion-corrected atom-centered potentials for generalized gradient approximation functionals: elements H, C, N, O, He, Ne, Ar, and Kr

I-C. Lin, M. D. Coutinho-Neto, C. Felsenheimer, O. A. von Lilienfeld, I. Tavernelli, and U. Rothlisberger; *Phys. Rev. B*, **75**, 205131, (2007)

7. Controlling crystallization and its absence: proteins, colloids and patchy models

J. P. K. Doye, A. A. Louis, I-C. Lin, L. R. Allen, E. G. Noya, A. W. Wilber, H. C. Kok and R. Lyus; *Phys. Chem. Chem. Phys.*, **9**, 2197 (2007)

• **To be submitted or in preparation**

1. Multi-center density functionals

I. Tavernelli, I-C. Lin, and U. Rothlisberger; to be submitted (2008)

2. Ab initio molecular dynamics simulations of liquid water

I-C. Lin, M. D. Coutinho-Neto I. Tavernelli, and U. Rothlisberger; to be submitted (2008)

3. Hydrogen-bonded complexes described with dispersion-corrected density functional theory

J. S. Arey, I-C. Lin, P. Aeberhard, and U. Rothlisberger

4. Dispersion-corrected atom-centered potentials for generalized gradient approximation functionals: element sulfur

P. Aeberhard, J. S. Arey, I-C. Lin, and U. Rothlisberger

5. Hydride transfer in dihydrofolate reductase: Insights from QM/MM simulations

I-C. Lin, L. Guidoni, and U. Rothlisberger

Oral presentations

1. Developments and applications of atom centered potential for the description of weak interactions in DFT, Max Planck Institute for Polymer Research, October 2007, Mainz (Germany)

2. *Ab initio* molecular dynamics simulation of liquid water, Fall meeting of the Swiss Chemical Society, September 2007, Lausanne (Switzerland)

3. Atom centered potentials for describing weak interactions in DFT, Dipartimento di Fisica, Università di Roma 'La Sapienza', March 2007, Rome (Italy)

4. Atom centered potential for the description of weak interactions in DFT, research group of Prof M. E. Tuckerman, New York University, September 2006, New York (USA)
5. Atom centered potential for the description of weak interactions in DFT, 'Progress in *ab initio* modelling of biomolecules: methods and applications', June 2006, Leiden (Netherlands)
6. Atom centered potential for describing London dispersion forces in DFT, CECAM workshop 'New developments for first principles molecular dynamics simulations in condensed matter and molecular physics', May 2006, Lyon (France)
7. From diatomic to DNA Intercalations: atom centered potential for London dispersion forces in DFT, 'Car-Parrinello dyna-mix & co.', April 2006, Paris (France)
8. Atom centered potential for London dispersion forces in DFT, Fall meeting of the Swiss Chemical Society, October 2005, Lausanne (Switzerland)

Poster presentations

1. Developments and applications of atom centered potentials for describing weak interactions in density functional theory, CCP2007 'Conference on computational physics', September 2007, Brussels (Belgium)
2. Atom centered potentials for London dispersion forces in DFT, Research Day EPFL, November 2006, Lausanne (Switzerland)
3. Atom centered potential for London Dispersion forces in DFT, 'Computer Simulations in Condensed Matter: from Materials to Chemical Biology', July 2005, Erice (Italy)
4. Atom centered potential for London dispersion forces in DFT, CECAM workshop 'Van der Waals forces and density functional theory', June 2005, Lyon (France)

Curriculum Vitae

I-Chun Lin,

born October 1982

12.2004 – 04.2008 PhD, Ecole Polytechnique Fédérale de Lausanne

Supervisor: Prof Dr Ursula Röthlisberger

09.2000 – 06.2004 BA and MSci, Trinity College, University of Cambridge