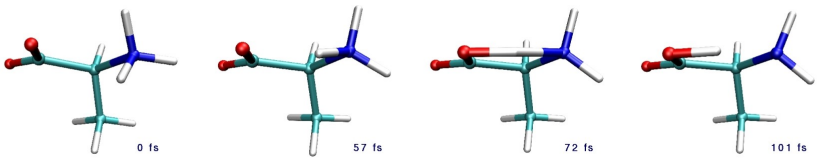
Dissertations Laboratory of Physics, Helsinki University of Technology
Fysiikan laboratorio, Teknillinen korkeakoulu
Espoo 2007 Dissertation 147

FIRST-PRINCIPLES STUDIES OF THE STRUCTURE AND DYNAMICS OF BIOMOLECULES

Ivan Degtyarenko



TEKNILLINEN KORKEAKOULU
TEKNISKA HÖGSKOLAN
HELSINKI UNIVERSITY OF TECHNOLOGY

Dissertations of Laboratory of Physics, Helsinki University of Technology
Fysiikan laboratorio, Teknillinen korkeakoulu
Espoo 2007 Dissertation 147

FIRST-PRINCIPLES STUDIES OF THE STRUCTURE AND DYNAMICS OF BIOMOLECULES

Ivan Degtyarenko

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the Department of Engineering Physics and Mathematics, Helsinki University of Technology for public examination and debate in Auditorium E at Helsinki University of Technology (Espoo, Finland) on the 8th of June, 2007, at 12 o'clock noon.

Helsinki University of Technology
Department of Engineering and Mathematics
Laboratory of Physics

Dissertations of Laboratory of Physics, Helsinki University of Technology
ISSN 1455-1802

Dissertation 147 (2007):
Ivan Degtyarenko: First-principles studies of the structure and dynamics of biomolecules
ISBN 978-951-22-8805-2 (printed version)
ISBN 978-951-22-8806-9 (electronic version)

Available in PDF format at http://lib.tkk.fi/Diss/2007/isbn9789512288069/

Multiprint Oy / Otamedia
Espoo 2007

Abstract

First-principles biosimulations have become an essential tool in the study of atoms and molecules and, increasingly, in modelling complex systems as those arising in biology. With the appearance of density-functional theory, and gradient-corrected exchange-correlation functionals, the ability to obtain an accurate enough solutions to the electronic Schrödinger equation for systems containing hundreds (or even thousands) of atoms has revolutionized biophysics and biochemistry. Biological systems exhibit a far higher degree of complexity than those studied in many other fields of physics. The sizes of the systems, long time scale of processes, the effect of the environment, and the range of intermolecular interactions provide challenging problems for the application of first-principles quantum mechanical simulations to biomolecular studies.

This thesis concentrates on first-principles electronic structure calculations of various biological systems and processes. The dynamics of the active center of myoglobin has been studied by means of Born-Oppenheimer molecular dynamics. Similar methodology has been used to investigate the effect of hydration of the L-alanine amino acid and to predict its actual structure in aqueous solution at finite temperature. The effect of the environment, and the actual structure of several biomolecules in water have been investigated by means of vibrational spectra calculations. Different continuum models have been employed in calculations of the vibrational absorption, vibrational dichroism, Raman and Raman optical activity spectra.

The treatment of large, biological systems, such as proteins in aqueous solution, entirely by *ab initio* methods is extremely expensive. The thesis demonstrates various approaches to overcome size and time scale limits. The work presented here is an example of how quantum mechanical techniques can successfully be applied to biologically relevant problems in rather large and complex systems.

Affiliation

Author

Ivan Degtyarenko

Laboratory of Physics
Helsinki University of Technology
Espoo, Finland
imd@fyslab.hut.fi

Supervisor

Acad. Prof. Risto Nieminen

Laboratory of Physics
Helsinki University of Technology
Espoo, Finland
Risto.Nieminen@hut.fi

Opponent

Prof. Dr. Sándor Suhai

German Cancer Research Center
Division of Molecular Biophysics
Heidelberg, Germany
s.suhai@dkfz-heidelberg.de

Reviewer

Prof. Mikko Karttunen

Dept. of Applied Mathematics
The University of Western Ontario
London (ON), Canada
mkarttu@uwo.ca

Reviewer

Prof. Kari Laasonen

University of Oulu
Department of Chemistry
Oulu, Finland
Kari.Laasonen@oulu.fi

Preface

This thesis has been prepared in the Laboratory of Physics in the Computational Nanoscience center (Helsinki University of Technology, Finland) during the years 2002-2007.

I wish to express my gratitude to the supervisor and the COMP (Computational Nanoscience) leader Academy Professor Risto Nieminen for providing excellent facilities and the financial and other support for the research. I want to pass my sincere gratitude to Dr. Karl Jalkanen (Curtin University, Australia) and Dr. Carme Rovira (University of Barcelona, Spain) for being my co-supervisors and guides in the area of computational physics and chemistry. I thank my coauthor Andrey Gurtovenko (University of Bradford, UK) for his contribution to this thesis. I want to say the words of gratitude to my colleagues in the Laboratory of Physics, present or past, and to all my friends in Helsinki University of Technology, with whom I have been working together these years. Thank you for all the help.

I dedicate this thesis to my family Klara, Aleksandra and Egor, ... (so far), who have been the victims that suffered most from sacrificing our weekends and all my spare time to the progress of this work. Thank you for your endurance and the invaluable support during the preparation of this thesis.

Helsinki, June 2007

Ivan Degtyarenko

Contents

Abstract
Affiliation
Preface
List of publications

1. Molecular modelling

1.1 Introduction....1
1.2 Biological system specificity....3
1.3 *Ab initio* methods in molecular simulations....3
1.4 Overcoming length and time scales....5

2. Density-functional approach

2.1 General formulation: the Hohenberg-Kohn theorems....7
2.2 The Kohn-Sham self-consistent equations....8
2.3 Approximation for the *xc*-energy....11
2.3.1 Local density approximation (LDA)....11
2.3.2 Generalized gradient approximations (GGA)....12
2.3.3 Hybrid functionals....13
2.4 Basis sets....14
2.5 Pseudopotential approximation....16

3. Optimization techniques

3.1 Conjugate-gradients....18
3.2 Second-order methods....18
3.3 Locating the global minimum....20

4. Molecular dynamics

4.1 Newton's equations of motion.....21
4.2 Quantum molecular dynamics.....23
4.2.1 Born-Oppenheimer molecular dynamics.....23
4.2.2 Contributions to the Hellmann-Feynman forces.....23
4.2.3 BOMD in practice.....24

5. Vibrational spectra calculations

5.1 Vibrational normal modes and frequencies.....25
5.2 Potential energy surface calculation.....26
5.2 Energy perturbation.....27
5.3 Intensities.....27

6. First-principles studies of biomolecules

6.1 Structure and vibrational spectra of (R)-phenyloxirane.....29
6.2 Dynamics of the isolated Co- and Fe-heme models.....30
6.3 First-principles studies of the hydrated L-alanine amino acid.....32
6.3.1 Dynamics of the L-alanine in the hydration shell.....32
6.3.2 The structure in water and in crystal.....34
6.3.3 Case study: isolated L-alanine amino acid.....35
6.3.4 Vibrational spectra of L-alanine zwitterion in aqueous solution.....36

7. Summary

Abstracts of the publications.....39

List of abbreviations.....43

List of publications

I. K. J. Jalkanen, V. W. Jurgensen and I. M. Degtyarenko, *Linear response properties required to simulate vibrational spectra of biomolecules in various media: (R)-Phenyloxirane (a comparative theoretical and spectroscopic vibrational study)*, Advances in Quantum Chemistry **50**, 91-124 (2005).

II. I. Degtyarenko, R. M. Nieminen and C. Rovira, *Structure and dynamics of dioxygen bound to cobalt and iron heme*, Biophysical Journal **91(6)**, 2024-2034 (2006).

III. I. M. Degtyarenko, K. J. Jalkanen, A. A. Gurtovenko and R. M. Nieminen, *L-alanine in a droplet of water: a density-functional molecular dynamics study*, Journal of Physical Chemistry B **111(16)**, 4227-4234 (2007).

IV. I. M. Degtyarenko, K. J. Jalkanen, A. A. Gurtovenko and R. M. Nieminen, *The aqueous and crystalline forms of L-alanine zwitterion*, Journal of Computational and Theoretical Nanoscience (in press).

V. K. J. Jalkanen, I. M. Degtyarenko, R. M. Nieminen, X. Cao, L. A. Nafie, F. Zhu and L. D. Barron, *Role of hydration in determining the structure and vibrational spectra of L-alanine and N-acetyl L-alanine N'methylamide in aqueous solution: a combined theoretical and experimental approach*, Theoretical Chemistry Accounts (in press).

The author has had an active role in all the phases of the research reported in this thesis. He has been involved in the planning of the simulations, analysing the results and representing the interpretation. The author has performed the electronic-structure calculations for Publications II, III and IV, and has developed the analysis tools. Publications III and IV were written by the author and he originated the ideas for the studies. The author played an important role in designing the computer experiments and analysing the results for Publication II. The results calculated by the author provided a reference for the Publications I, V. He participated in the interpretation of the results and played an active role in writing the manuscript for Publication V.

1. MOLECULAR MODELLING

1.1 Introduction

The chemical origin of life was recognized almost two hundred years ago. The initial formulation of the concept that life is based on morphological units known as cells is generally attributed to a work by Schleiden and Schwann in 1838.[1] Originally, it was thought that only living beings could produce the molecules of life, so called organic molecules. Ten years before Schleiden and Schwann, in 1828, Wöhler published a paper about the synthesis of urea, demonstrating that organic compounds can be created artificially. About the same time, Payen discovered the first enzyme, diastase (α-, β-, or γ-amylase). Eduard Buchner contributed the first demonstration of a complex biochemical process outside of a cell in 1897: alcoholic fermentation in cell extracts of yeast.[2] It became clear that the living organisms and life itself are the result of a complex combination of individual chemical products and chemical reactions. Discovery of the genetic code and the molecular structure of deoxyribonucleic acid (DNA) in the early 1950s had a tremendous impact on all subsequent biological investigations.[3,4] One of the most fascinating aspects of this discovery was that an understanding of the mechanism by which the genetic code functioned could not be achieved until knowledge of the three-dimensional (3D) structure of DNA was obtained. This established the primary importance of molecular structure for an understanding of the function of molecular systems, and gave rise to revealing the structures of proteins and enzymes. Since the 1950s when the first 3D structures of myoglobin (Figure 1) and hemoglobin appeared,[5,6] the structures of a very large number of proteins and other biological molecules have been determined experimentally.[7,8]

---

1 D. Voet and J. G. Voet, *Biochemistry*, Wiley, Hoboken, 3rd edn., 2004.

2 *Nobel Lectures, Chemistry 1901-1921*, Elsevier, Amsterdam, 1966.

3 R. Olby, *The Path to the Double Helix: Discovery of DNA*, MacMillan, London, 1974.

4 N. Campbell and J. Reece, *Biology*, Benjamin Cummings, San Francisco, 6th edn., 2002.

5 J. C. Kendrew, G. Bodo, H. M. Dunitz, R. G. Parrish, H. Wyckoff and D. C. Phillips, *A three-dimensional model of the myoglobin molecule obtained by X-ray analysis,* Nature **181**, 662-666 (1958).

6 M. F. Perutz, M. G. Rossmann, A. F. Cullis, H. Muirhead, G. Wll, and A. C. T. North, *Structure of hemoglobin: a three-dimensional Fourier synthesis at 5.5 Å resolution obtained by X-ray analysis*, Nature **185**, 416-422 (1960).

7 T. E. Creighton, *Proteins: Structure and Molecular Properties*, Freeman, New York, 2nd edn., 1993.

8 There are currently over 40,000 biological macromolecules structures available, including proteins, nucleic acids, and protein-nucleic acid complexes (http://www.pdb.org/). H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *The Protein Data Bank*, Nucl. Ac. Res. **28(1)**, 235 (2000).

The prediction of the main elements of the secondary structure in proteins, the α-helices and β-sheets, by Pauling and Corey,[9] followed by the prediction of allowed amino acids conformations, by Ramachandran,[10] made theoretical investigation of the proteins possible.[11] Investigation of hemoglobin structure showed that the protein must be flexible in order to perform its biological functions.[6,12] There is no path for the escape of O_2 from the heme-binding pocket in the crystal structure, and the protein must change structure in order for O_2 to be released. This realization and others[13] indicated the need to investigate the dynamical properties of the proteins. A number of experimental techniques have been developed, however, the information content from these studies of dynamics is generally isotropic in nature, affording little insight into the atomic details of these fluctuations.[13,14] Atomic resolution information on the dynamics of biomolecules, in their natural environment rather than in crystalline form, and the relationship of the dynamics to molecular functions and properties, is the field where computational studies can extend our knowledge beyond what is accessible to experimentalists.

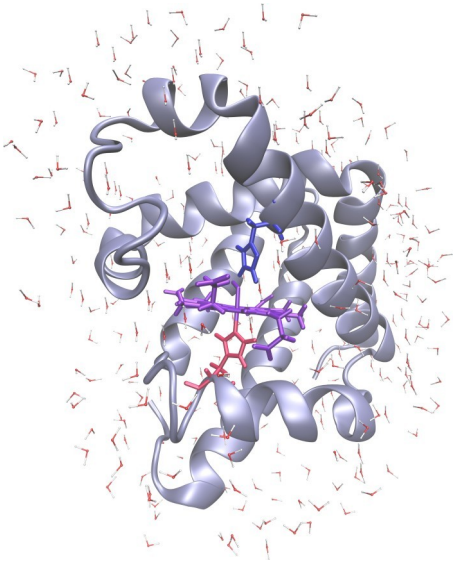


Figure 1. Myoglobin in a water environment.

Molecular modelling is not only the way to mimic the behavior of molecular systems, but also the way to understand experiments which have not been able to be interpreted. Nowadays molecular modelling is associated mostly with computer simulations. However, computers simply allow the higher level complexity problems to be solved.

---

9 (a) L. Pauling, R. B. Corey and H. R. Branson, *The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain*, Proc. Natl. Acad. Sci. U.S.A. **37**, 205-211 (1951); (b) L. Pauling and R. B. Corey, *Configurations of polypeptide chains with favored orientations around single bonds: two new pleated sheets*, Proc. Natl. Acad. Sci. U.S.A. **37**, 729-740 (1951).

10 G. N. Ramachandran and V. Sasisekharan, *Conformation of polypeptides and proteins*, Adv. Prot. Chem. **28**, 283-437 (1968).

11 Proteins have four levels of structure: (i) primary, the amino acid sequence, (ii) secondary, local conformations of short sequences stabilized by hydrogen bonds, like α-helix and β-sheet, (iii) tertiary, how the secondary structural elements fold up, and finally (iv) quatenary, how the tertiary elements assemble. The myoglobin in Figure 1 is a monomer and hence has tertiary structure.

12 M. Perutz, G. Fermi, B. Luisi, B. Shaanan and R. C. Liddington, *Stereochemistry of cooperative mechanisms in hemoglobin*, Acc. Chem. Res. **20(9)**, 309-321 (1987).

13 M. Karplus and G. A. Petsko, *Molecular dynamics simulations in biology*, Nature **347(18)**, 631-639 (1990).

14 I. N. Serdyuk, N. R. Zaccai and J. Zaccai, *Methods in Molecular Biophysics*, Cambridge, New York, 2006.

1.2 Biological system specificity

Biological systems are characterized by specificity, and complexity, both in structure and in chemical reactivity.[15] The origins of complexity are numerous: the size of the systems, the relatively long time of biological reactions and processes, the lack of systematic periodicity, the variety of intermolecular interactions which must be considered, and the need for careful treatment of the effects of the environment.[13] The length scale varies from a few ångströms of chemical bond length to the size of the earth ecosystem. The time scale varies from the femtosecond of electronic rearrangements to the 10^9 years of evolution time (see Figure 2).

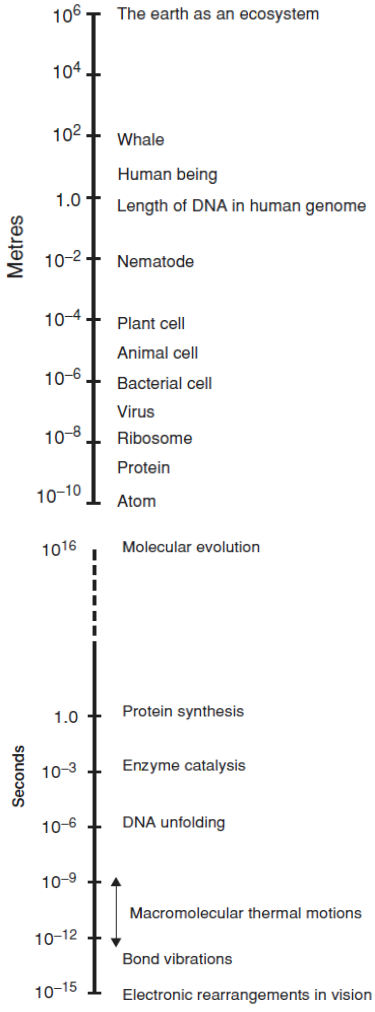


Figure 2. Length and time scales in biology.

Myoglobin is a representative example of a molecular structure which biophysics is dealing with everyday (Figure 1). Myoglobin is a medium size hemoprotein consisting of a polypeptide chain of 151 amino acids,[16] associated with a single heme group, expressed solely in cardiac myocytes and oxidative skeletal muscle fibres. It was called the hydrogen atom of biology due to its molecular properties. This simple example shows the level of complexity of the biological problems. Experimental techniques allow these problems to be investigated to some degree.[14] However, nowadays simulation tools are the key in understanding biology at the molecular scale and in interpreting experimental results. There are various approaches in computational molecular modelling, and the most important ones are outlined in *Sections 1.3* and *1.4*.

1.3 *Ab initio* methods in molecular simulations

First-principles calculations (or *ab initio*, from Latin "from the beginning") are methods aimed at solving the electronic Schrödinger equation without reference to experimental data. These methods do not contain empirical parameters and are mathematically rigorous and thus accurate, but computationally expensive. The key

---

15 H. Frauenfelder, P. G. Wolynes and R. H. Austin, *Biological physics*, Rev. Mod. Phys **71(2)**, S419-430 (1999).

16 The myoglobon crystalline structure investigated by John Kendrew *et al.*[5] contains in all 1,260 atoms excluding hydrogen; in addition there are some 400 atoms of liquid and salt solution, a proportion of which are bound to fixed sites on the surface of the molecule. Nobel Lecture in Chemistry, *Myoglobin and the structure of proteins*, 1962.

size range at which realistic molecular structural models can be constructed is up to a hundred atoms.[17] Systems of this size can be treated at the Hartree-Fock (HF) level;[18] however, for most of the biologically relevant problems, inclusion of electron correlation is critical if results of sufficient accuracy are to be achieved. Because of the steep scaling of computational effort with system size N, traditional high-accuracy quantum chemical methods such as coupled cluster method (CCSD (T) scaling as N^7) or even second-order Møller-Plesset perturbation theory (MP2 scaling as N^5),[19] cannot readily be applied to systems much larger than tens of atoms in a cost-effective fashion. The accurate results obtained by these methods are often used as the "best values" in benchmark tests.[20]

Electron correlation is accurately addressed in another *ab initio* approach: density functional theory (DFT).[21] DFT deals with the electronic total density as opposed to electronic wavefunctions in HF. DFT normally scales as N^3 and potentially allows the calculation of systems with thousands atoms.[22,23] DFT has become the method of choice when high accuracy in the treatment of conformational energetics and nonbonded interactions is required.[24] While the method does not yet uniformly achieve chemical accuracy (1 kcal/mol maximum error for heats of atomization), the performance of the best functionals with hybrid methods achieve an average error of 3.1 kcal/mol for several hundred heats of atomization, ionization potentials, electron affinities, and proton affinities over a wide range of small molecules.[20]

The standard wavefunction methods and density functional approaches can be jointly classified as the electronic structure calculation methods,[25] as opposed to molecular mechanics where electronic degrees of freedom are not taken into account.

1.4 Overcoming length and time scales

It is a fact that the size range of hundreds and even thousands of atoms achievable

---

17 R. A. Friesner, *Ab initio quantum chemistry: Methodology and applications*, Proc. Natl. Acad. Sci. U.S.A. **102(19)**, 6648-6653 (2005).

18 A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry. Introduction to Advanced Electronic Structure Theory*, Dover, New York, 1996.

19 The localized MP2 methods have reduced the scaling to less than N^3, see Ref. 17.

20 L. Curtis, P. C. Redfern and K. Raghavachari, *Assessment of Gaussian-3 and density functional theories on the G3/05 test set of experimental geometries*, J. Chem. Phys. **123**, 124107 (2005) and references therein. See also URL http://chemistry.anl.gov/compmat/comptherm.htm.

21 In practice, the level of theory depends on the number of approximations used, and mainly on the exchange-correlation functional.

22 W. Kohn, *Nobel Lecture: Electronic structure of matter-wave functions and density functionals*, Rev. Mod. Phys. **71**, 1253–1266 (1999), and references therein.

23 Static calculations are meant, nowadays the realistic molecular system size range for studying dynamics is up to several hundreds of atoms and time scale is tens of picoseconds. Examples of such large first-principles molecular dynamics simulations are presented in Publications II, III and IV of this thesis.

24 However, one notable exception being the dispersion forces, J. F. Dobson, K. McLennan, A. Rubio, J. Wang, T. Gould, H. M. Le and B. P. Dinte, *Prediction of dispersion forces: is there a problem?*, Aust. J. Chem. **54**, 513-527 (2001).

25 We skip here the discussion on semi-empirical methods. The subject is outlined in the Ref. 53, ch. 3, pp. 81-97.

by DFT is not enough for simulations of macromolecules. To overcome the system size limitations one should improve underlying theory or employ a number of approximations.[26,27]

Linear-scaling techniques, or so-called *O(N)* methods, are essential tools for large system calculations.[28,29,30] The new algorithms enable the calculations to be less CPU time and memory size consuming while still keeping the underlying theory close to the *ab initio* level.

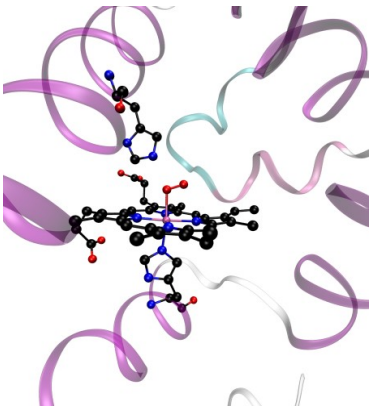


Figure 3. QM/MM approach applied to myoglobin. The active site, heme, is treated quantum mechanically. The rest of the protein is treated at the molecular mechanics level.

The simplest approximation is to make the studied system smaller by isolating the parts of the system which are believed to be important to the process under investigation. Such a mimetic model can be treated fully quantum-mechanically. However, one must ensure that when enlarging the system the results of the calculation do not change significantly. A better approximation is to embed the fragment treated quantum mechanically in a larger system which is described by a classical molecular mechanics approach. These quantum mechanics / molecular mechanics (QM/MM) approaches allow the simulation of entire proteins including solvating water molecules.[31,32] The difficulty lies in achieving a good description of the interface between two parts, where covalent bonds may be cut. In solutions the effect of solvation can be approximated by a dielectric continuum model, using self-consistent reaction field methods (SCRF).[33]

---

26 M. D. Segall, *Applications of ab initio atomistic simulations to biology*, J. Phys.: Condens. Matter **14**, 2957-2973 (2002).

27 Another way to make larger simulations possible is technical, along with the growing computer power, efficient program's code parallelization brings great possibilities.

28 S. Goedecker, *Linear scaling electronic structure methods*, Rev. Mod. Phys. **71(4)**, 1085-1123 (1999).

29 SIESTA (Spanish Initiative for Electronic Simulations with Thousands of Atoms) is both a method and its computer program implementation, to perform electronic structure calculations and *ab initio* molecular dynamics simulations of molecules and solids (http://www.uam.es/siesta/). J. M. Soler, E. Artacho, J. D. Gale, A. Gárcia, J. Junquera, P. Ordejón and D. Sánchez-Portal, *The SIESTA method for ab initio order-N materials simulation*, J. Phys.: Condens. Matter **14**, 2745-2779 (2002).

30 D. R. Bowler, R. Choudhury, M. J. Gillan and T. Miyazaki, *Recent progress with large-scale ab initio calculations: the CONQUEST code*, Phys. Stat. Sol. B **243**, 989 (2006).

31 P. D. Lyne and O. A. Walsh, *Computer simulation of biochemical reactions with QM–MM methods*, in *Computational Biochemistry and Biophysics*, eds. O. M. Becker, A. D. MacKerell Jr., B. Roux and M. Watanabe, Dekker, New York, ch. 11, pp. 221-236, 2001.

32 J. Norberg and L. Nilsson, *Advances in biomolecular simulations: methodology and recent applications*, Quart. Rev. Biophys. **36(3)**, 257-306 (2003).

33 Besides including the solvent molecules explicitly, one can also employ a reaction field continuum model, or use both approaches and model the solute molecule in its first hydration shell surrounding this "supermolecule" with

A step further from *ab initio* calculations are the molecular mechanics (MM) (or force field) methods.[34] The parameters that enter the potential energy function are fitted to experimental or higher level computational data. The molecules are modelled as atoms held together by bonds and are basically described by a "ball and spring" model, and the quantum aspects are neglected. The methods are computationally "cheap" (compared to *ab initio*) and used for molecular dynamics simulations of large systems (~ 10^6 atoms),[35] at long time scales (~ 500 μs).[36] However the methods suffer from several limitations: they require extensive parameterization and thus depend on the available experimental and *ab initio* calculation data; energy estimates obtained are not accurate; they cannot be used to simulate reactions where covalent bonds are broken/formed; and are limited in their abilities for providing accurate details regarding the chemical environment.

The method of choice for solving a particular problem is always a compromise between efficiency of calculations and required accuracy. Nowadays, the force-field methods and DFT are the main workhorses of the computational biophysics and biochemistry simulations.

The computational methods used throughout this thesis are based on density functional theory. In the *Chapter 2*, an introduction to density functional theory and related methods will be given. After the description of the Kohn-Sham equations it will be demonstrated how DFT can be used efficiently in computer simulations. The discussion will address the basis set choice, exchange-correlation functionals and the pseudopotential scheme. In *Chapter 3*, an overview will be given on optimization techniques and finally the molecular dynamics Born-Oppenheimer scheme and the vibrational spectra simulation methods will be introduced in *Chapters 4* and *5*.

---

continuum solvent. J. Tomasi, B. Mennucci and R. Cammi, *Quantum mechanical continuum models*, Chem. Rev. **105**, 2999-3093 (2005), and references therein.

34 Coarse-grained methods, not discussed in this thesis, are a step beyond the atomistic simulations. Instead of explicitly representing every atom of the system, one uses pseudoatoms to represent groups of atoms. *Novel Methods in Soft Matter Simulations*, eds. M. Karttunen, I. Vattulainen, and A. Lukkarinen, Springer-Verlag, Berlin, 2004.

35 (a) P. L. Freddolino, A. S. Arkhipov, S. B. Larson, A. McPherson and K. Schulten, *Molecular dynamics simulations of the complete satellite tobacco mosaic virus*, Structure **14**, 437-449 (2006); (b) K. Y. Sanbonmatsu, S. Joseph and C.-S. Tung, *Simulating movement of tRNA into the ribosome during decoding*, Proc. Natl. Acad. Sci. U.S.A. **102**, 15854-15859 (2005).

36 G. Jayachandran, V. Vishal, A. E. Garcia and V. S. Pande, *Local structure formation in simulations of two small proteins*, J. Struct. Biol. **157(3)**, 491-499 (2007).

2. DENSITY-FUNCTIONAL APPROACH

John Perdew has characterized Density Functional Theory (DFT), as "a simple story about everything".[37] DFT is an alternative approach to the theory of electronic structure, in which the electron density distribution $n(r)$, rather than the many electron wavefunction $\Psi(r_1, r_2, \ldots, r_N)$ plays a central role.[22] The initial work on a rigorous theory (DFT) has been reported by Kohn, Hohenberg and Sham in two publications in 1960s.[38,39] In comparison to other quantum methods, DFT is unique due to its appealing combination of computational efficiency and accuracy. For this reason, DFT has become the standard approach for the study of systems in which a large number of atoms must be considered, for instance biological systems. In this chapter, an introduction to density-functional theory and related methods will be given, based on a number of sources.[37,40,41] The discussion presented here relies upon the Born-Oppenheimer adiabatic approximation.

2.1 General formulation: the Hohenberg-Kohn theorems

The first theorem provides the proof that the electron density $n(r)$ uniquely determines the Hamilton operator, H. The theorem states that *"the external potential $V(r)$ (to within a constant) is a unique functional of $n(r)$; since, in turn, $V(r)$ fixes H, we see that the full many-particle ground state is a unique functional of $n(r)$ ".*[38] From this theorem and the fact that $n(r)$ determines the number of electrons, it follows that $n(r)$ also determines the ground state wave function and all other electronic properties of the system. In particular, it determines the kinetic energy $T[n(r)]$, the potential energy $U[n(r)]$ and the total energy $E_v[n(r)]$. The functional of the total energy is:

$$E_V[n(r)] = \int V(r)n(r)\,dr + F[n(r)] \quad , \tag{1}$$

where the functional $F[n(r)]$ is

---

37 J. Perdew and S. Kurth, *Density functionals for non-relativistic coulomb systems*, in *Primer in Density-Functional Theory*, eds. C. Fiolhais, F. Nogueira and M. Marques, Springer, Heidelberg, ch. 1, pp. 1-55, 2003.

38 P. Hohenberg and W. Kohn, *Inhomogeneous electron gas.* Phys. Rev. **136**, 864-871 (1964).

39 W. Kohn and L. J. Sham, *Self-consistent equations including exchange and correlation effects*. Phys. Rev. **140**, 1133-1138 (1965).

40 W. Koch and M. C. Holthausen, *A Chemist's Guide to Density Functional Theory*, Wiley, Weinheim, 2nd edn., 2000.

41 R. M. Martin, *Electronic Structure: Basic Theory and Practical Methods*, Cambridge University Press, Cambridge, 2004.

$$F[n(r)] = T[n(r)] + J[n(r)] + E_{ncl}[n(r)] \quad . \tag{2}$$

Here, besides the exact kinetic energy, T, the functional includes classical Coulomb repulsion, J:

$$J[n(r)] = \frac{1}{2}\int\int \frac{n(r)n(r')}{|r-r'|} dr dr' \quad , \tag{3}$$

and set of complex quantum effects in $E_{ncl}[n(r)]$. Whereas the nuclei-electron interactions are incorporated in the external potential V of (1). It is notable that $F[n(r)]$ is defined independently of the V and thus it is an universal functional of the electronic density. The $F[n(r)]$ functional plays a central role in DFT, if it were known exactly we would have solved the Schrödinger equation exactly. The difficulties appear due to the non-classical term $E_{ncl}[n(r)]$ in (2), which is the important quantity in practice and basically the major part of the exchange-correlation energy, discussed in the following section.

To show that a certain density is really the ground state density, $n_0(r)$, Hohenberg and Kohn employed the variational principle. The second theorem states, that *"* $E[n(r)]$ *assumes its minimum value for the correct* $n(r)$*, if the admissible functions are restricted by the condition*

$$N[n(r)] \equiv \int n(r)\,dr = N \quad .\text{"} \tag{4}$$

That means, as soon as we uniquely determine the energy by the electron density $n(r) \equiv n_0(r)$, any other density will give rise to a higher energy.

2.2 The Kohn-Sham self-consistent equations

Density-functional theory was turned into a practical tool for computer simulations by Kohn and Sham[39] who invented an elaborate indirect approach to the kinetic energy functional $T[n(r)]$.[42] The exact formula for ground state kinetic energy of a system is

$$T = \left\langle \Psi \left| -\frac{1}{2}\nabla^2 \right| \Psi \right\rangle \quad , \tag{5}$$

where Ψ is the N-electron wave function. According to the first Hohenberg-Kohn theorem the kinetic energy is a functional of the total electron density (*s* is a spin variable)

$$n(r) = |\Psi(r,s)|^2 \quad . \tag{6}$$

Kohn and Sham showed that one can build a theory using simpler formulas, namely

$$T_s[n(r)] = \sum_i^N \left\langle \psi_i \left| -\frac{1}{2}\nabla^2 \right| \psi_i \right\rangle \quad , \tag{7}$$

[42] The simple functional form for the kinetic energy was the reason for a bad perfomance of the original Thomas-Fermi model, see Ref. 40 for the details.

and

$$n(r)=\sum_{i}^{N}\sum_{s}|\psi_i(r,s)|^2 \quad . \tag{8}$$

This representation of the kinetic energy holds true for the determinantal wave function that exactly describes N *non-interacting* electrons. The idea is now to introduce a non-interacting reference system, with the Hamiltonian:

$$\hat{H}_s=\sum_{i}^{N}\left(-\frac{1}{2}\nabla_1^2\right)+\sum_{i}^{N}V_s(r_i) \quad , \tag{9}$$

in which there are no electron-electron repulsion terms, but its density equals the exact ground-state density of the interacting system. For this system there will be an exact determinantal ground state wave function:

$$\Psi_s=\frac{1}{\sqrt{N!}}det[\psi_1\psi_2...\psi_N] \quad , \tag{10}$$

where the ψ_i are the N lowest eigenstates of the one-electron Hamiltonian $\hat{h}_s$:

$$\hat{h}_s\psi_i=[-\frac{1}{2}\nabla_i^2+V_s(r)]\psi_i=\epsilon_i\psi_i \quad .$$

The kinetic energy $T_s[n(r)]$ of this system is given by equation (7). In order to use the expression for the kinetic energy of the non-interacting system, equation (2) is rewritten as:

$$F[n(r)]=T_s[n(r)]+J[n(r)]+E_{xc}[n(r)] \quad , \tag{11}$$

where

$$E_{xc}[n(r)]=(T[n(r)]-T_s[n(r)])+E_{ncl}[n(r)] \quad , \tag{12}$$

The defined quantity $E_{xc}[n(r)]$ is called the *exchange-correlation* energy functional, which contains everything that is unknown: the part of the true kinetic energy, T, not covered by T_s, plus the contributions from exchange and correlation (*xc*) energies.

The energy functional $E[n(r)]$ in terms of the Kohn-Sham orbitals, as defined in equation (10), has the following form:

$$E[n(r)]=T_s[n(r)]+J[n(r)]+E_{xc}[n(r)]+\int n(r)V(r)\,dr \quad . \tag{13}$$

The total energy has been split into four contributions: (i) the kinetic energy of the corresponding non-interacting system, (ii) the electrostatic energy, (iii) an exchange and correlation term and (iv) the interaction energy with the external potential.

For the minimization of the functional (13), one has to take into account the constraint $\int n(r)\,dr=N$ using a Lagrange parameter μ,

$$\frac{\delta[E[n(r)]+\mu(N-\int n(r)\,dr)]}{\delta n(r)}=0 \quad .$$

Instead of varying the density, it is also possible to vary the one-particle wave functions ψ_i, whereby the constraint $\int n(r)\,dr=N$ is replaced by $\int|\psi_i(r)|^2\,dr=1$:

$$\frac{\delta[E[n(r)]+\sum_{i=1}^{N}\epsilon_i(1-\int|\psi_i(r)|^2 dr)]}{\delta\psi_i^*(r)}=0 \quad ,$$

where ϵ_i are the corresponding Lagrange parameters (eigenvalues). The minimisation of the energy functional leads to the Kohn-Sham equations

$$\left[-\frac{1}{2}\nabla_i^2+V_{eff}(r)\right]\psi_i=\epsilon\,\psi_i \quad , \tag{14}$$

where the effective potential V_{eff} is defined by,

$$V_{eff}(r)=V(r)+\int\frac{n(r')}{|r-r'|}dr'+V_{xc}(r) \quad , \tag{15}$$

with the *xc*-potential:

$$V_{xc}(r)=\frac{\delta E_{xc}[n(r)]}{\delta n(r)} \quad . \tag{16}$$

These equations correspond to a system of non-interacting electrons moving in an external potential $V_s(r)=V_{eff}(r)$. For a given $V_{eff}(r)$, one obtains the ground state density $n(r)$, by solving the N one-electron equations (14) and setting

$$n(r)=\sum_{i}^{N}\sum_{s}|\psi_i(r,s)|^2 \quad . \tag{17}$$

The effective potential V_{eff} depends on $n(r)$ through equation (15). Hence, the equations (14), (15) and (17) have to be solved self-consistently. One begins with a guessed $n(r)$, constructs $V_{eff}(r)$ from (15) and then finds a new $n(r)$ from (14).

In conclusion, we have rewritten the variational principle stated in the Hohenberg-Kohn theorem as the solution of single-particle Schrödinger equations with an effective potential V_{eff}. By splitting the energy functional in equation (13), the system of interacting electrons in an external potential V was mapped onto a system of noninteracting electrons in an effective potential V_{eff}, in such a way that the same electron density is obtained. This construction leads to the correct electron density but the wave function is not real. However, according to the first Hohenberg-Kohn theorem, the electron density is sufficient to have access to the ground state properties. Note, that the Lagrange parameters ϵ_i in the equation (14) are not one-particle energies in DFT. However, it is a common procedure to compare theses values to experimental ionization energies. In practice, this approach has turned out to be

successful. The Kohn-Sham equations can (like single-particle Schrödinger equations) in principle be solved exactly and the presented scheme is indeed exact. However, for the exchange-correlation functional approximations have to be used to calculate this contribution to the total energy. In the following section different exchange-correlation functionals are presented.

2.3 Approximation for the *xc*-energy

2.3.1 Local density approximation (LDA)

An approximation to the *xc*-energy, $E_{xc}[n(r)]$, must be made to make the Kohn-Sham approach of any practical use. The exchange-correlation functional which has been widely used after the formulation of the Kohn-Sham theory is the local density approximation (LDA).[39] It is based on the model of the homogenous gas, in which the electron density is constant throughout all space. The idea in LDA is, to approximate the *xc*-energy density at point r of the actual inhomogeneous system with the corresponding expression of the homogenous electron gas $e_{xc}^{hom}[n(r)]$:

$$E_{xc}^{LDA} = \int n(r)\, e_{xc}^{hom}(n(r))dr \quad . \tag{18}$$

The *xc*-functional is obtained by differentiation of this expression:

$$V_{xc}^{LDA} = n(r)\frac{d\, e_{xc}^{hom}(n(r))}{d\, n(r)} + e_{xc}^{hom}(n(r)) \quad . \tag{19}$$

In most applications of LDA, the correlation energy is expressed by analytical parametrization of the results for the homogenous electron gas,[43] as derived by Ceperley and Alder from quantum Monte Carlo simulations:[44]

$$\begin{aligned} e_c(n(r)) &= -0.1423/(1+1.9529\, r_s^1/2+0.03334\, r_s) \quad , \quad for \quad r_s, \geqslant 1 \quad , \quad and \\ e_c(n(r)) &= -0.0480+0.0311\ln r_s - 0.0116\, r_s + 0.002\, r_s \ln r_s \quad , \quad for \quad r_s < 1 \quad , \end{aligned} \tag{20}$$

where

$$r_s = \left(\frac{3}{4\pi n(r)}\right)^{1/3} \tag{21}$$

is the *Wigner-Seitz radius*. For the exchange energy the following expression is commonly used, which is obtained by evaluating the Fock integral for a Slater determinant of orbitals in the case of the uniform gas:

$$E_x^{LDA}[n(r)] = -\frac{3}{2}\left(\frac{3}{4\pi}\right)^{1/3} \int n(r)^{4/3}(r)\, dr \quad . \tag{22}$$

---

43 J. P. Perdew and A. Z. Zunger, *Self-interaction correction to density-functional approximations for many-electron systems*, Phys. Rev. B **23**, 5048 (1981).

44 D. M. Ceperly and B. J. Alder, *Ground state of the electron gas by a stochastic method*, Phys. Rev. Lett. **45**, 566-569 (1980).

LDA should in principle only be used for systems with a slowly varying density. However, LDA works well for many systems that go beyond this limit. This surprising success can most often be explained by cancellations of errors. Other deficiencies are the inadequate cancellation of self-interaction contributions. As a consequence, the exchange-correlation potential, V_{xc}^{LDA}, does not exhibit the correct asymptotic behavior proportional to $1/r$ for localized systems (atoms, molecules, etc.) but instead it declines exponentially.

2.3.2 Generalized gradient approximations (GGA)

The large improvement over the LDA approach was the introduction of the gradient of the density into the expression of the *xc*-energy:

$$E_{xc}^{GGA} = \int n(r)\, e_{xc}(n(r); \nabla n(r))\, dr \quad . \tag{23}$$

The so-called generalized gradient approximations (GGA) extended the applicability of DFT calculations significantly.[45] Two of the most prominent examples of GGAs are the Perdew-Burke-Ernzerhof (PBE)[46] and the Becke-Lee-Yang-Parr (BLYP)[47,48] functionals. The analytic expressions for the GGAs are rather complicated. The expression for PBE, which is the most extensively used functional in this thesis, has the following form in the non-spin-polarized case:

The correlation energy is given by

$$E_c^{PBE} = \int n(r)[e_c^{hom}(r_s) + G(r_s, t)]\, dr \quad ,$$

where $t = |\nabla n(r)| / 2 k_s n(r)$ is a dimensionless density gradient with $k_s = \sqrt{4 k_F / \pi}$, as the Thomas-Fermi screening wave number $k_F = (3\pi^2 n(r))^{1/3}$. The expression for the gradient contribution G is:

$$G = \gamma \ln\left\{1 + \frac{\beta}{\gamma} t^2 \left[\frac{1 + At^2}{1 + At^2 + A^2 t^4}\right]\right\} \quad ,$$

where

$$A = \frac{\beta}{\gamma}[\exp\{-e_c^{hom}/\gamma\} - 1]^{-1} \quad ,$$

and $\beta \simeq 0.066725$ and $\gamma \simeq 0.031091$.

---

45 J. P. Perdew, J. A. Chvary and S. H. Vosko, *Atoms, molecules, solids, and surfaces: Applications of the generalized approximation for exchange and correlation*, Phys. Rev. B **46**, 6671-6687 (1992).

46 (a) J. P. Perdew, K. Burke and M. Ernzerhof, *Generalized gradient approximation made simple*, Phys. Rev. Lett. **77**, 3865 (1996); (b) J. P. Perdew, K. Burke and M. Ernzerhof, *Erratum: Generalized gradient approximation made simple [Phys. Rev. Lett. 77, 3865 (1996)]*, Phys. Rev. Lett. **78**, 1396-1396 (1997).

47 A. D. Becke, *Density-functional exchange-energy approximation with correct asymptotic behavior*, Phys. Rev. A **38**, 3098 (1988).

48 C. Lee, W. Yang and R. G. Parr, *Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density*, Phys. Rev. B **37**, 785 (1988).

The gradient-corrected exchange energy is given by

$$E_x^{PBE} = \int n(r)\, e_x^{hom}(n(r))\, F_x(s)\, dr \quad ,$$

where F_x is

$$F_x(s) = 1 + \kappa - \frac{\kappa}{1 + \mu s^2/\kappa} \quad ,$$

with $\kappa = 0.804$ and $\mu = 0.219$. The enhancement factor F_x is a function of $s = |\nabla n| / 2 k_F n(r)$, which is another dimensionless density gradient.

Mainly, the choice of the PBE functional for molecular dynamics simulations in Publications II and III is based on its reliability in the description of hydrogen bonds.[49] In our simulations the hydrogen bonding, i.e. D-H..A type of interaction (donor-H..acceptor), play the key role in stabilizing L-alanine amino acid in aqueous solution and significantly influence on the dynamics of O_2 ligand in the myoglobin active center simulations.

2.3.3 Hybrid functionals

Hybrid functionals are combinations of orbital-dependent HF and an explicit density functional, $E_{xc} = \frac{1}{2}(E_x^{HF} + E_{xc}^{DFA})$, where DFA denotes an LDA or GGA functional. The hybrid DFT calculations we have performed have been with the B3LYP[50] and B3PW91[51] as implemented in Gaussian 03.[52] They are widely used in calculations with localized atom-centered Gaussian basis sets, because the matrix elements of the exact exchange operator are relatively easy to evaluate (as opposed to plane wave methods). In particular the B3LYP functional was used in our vibrational spectra simulations. The benchmark tests indicate that nowadays these functionals are the most accurate.[16] That is the reason why they have been so much in use in computational chemistry for small molecules simulations. The definition of the *xc*-energy for B3LYP is:

$$E_{xc}^{B3LYP} = E_{xc}^{LDA} + a_0 (E_x^{HF} - E_x^{DFA}) + a_x E_x^{Becke} + a_c E_c \quad , \tag{24}$$

with a coefficients that are empirically adjusted to fit atomic and molecular data.

---

[49] (a) J. Ireta, J. Neugebauer and M. Scheffler, *On the accuracy of DFT for describing hydrogen bonds: dependence on the bond directionality*, J. Phys. Chem. A **108**, 5692-5698 (2004); (b) X. Xu and W. Goddard, *Bonding Properties of the Water Dimer: A Comparative Study of Density Functional Theories*, J. Phys. Chem. A **108**, 2305-2313 (2004).

[50] Becke three parameter hybrid functionals as defined by A. D. Becke, *Density-functional thermochemistry. III. The role of exact exchange*, J. Chem. Phys. **98**, 5648 (1993). B3LYP uses the non-local correlation provided by the LYP expression,[48] and VWN functional (III) for local correlation. Vosko, Wilk, and Nusair correlation functional (VWN) defined by S. H. Vosko, L. Wilk, and M. Nusair, Can. J. Phys. **58**, 1200 (1980).

[51] B3PW91 specifies the functional with the non-local correlation provided by Perdew and Wang 1991. (a) J. P. Perdew, K. Burke and Y. Wang, *Generalized gradient approximation for the exchange-correlation hole of a many-electron system*, Phys. Rev. B **54**, 16533-16539 (1996); (b) J. P. Perdew, K. Burke and Y. Wang, *Erratum: Generalized gradient approximation for the exchange-correlation hole of a many-electron system [Phys. Rev. B 54, 16 533 (1996)]*, Phys. Rev. B **57**, 14999-14999 (1998).

[52] M. J. Frisch *et al.*, *Gaussian 03, Revision C.03,* Gaussian Inc., Wallingford CT, 2004.

2.4 Basis sets

For the application of DFT in a computer simulation, an appropriate choice of the *basis set* has to be considered.[53,54] In this thesis a numerical atomic orbitals (NAO) basis set has been used in molecular dynamics simulations with the SIESTA program[29,110] and a contracted Gaussian type orbitals (CGF) basis set has been used in vibrational spectra calculations with the Gaussian program.[52] Both approaches employ the popular, in molecular simulations, linear combination of atomic orbitals (LCAO) ansatz.[55] The LCAO expansion allows to translate the non-linear optimization problem, which requires a set of coupled integro-differential equations (14 and 15) to be solved, to a linear one. The LCAO expansion of the Kohn-Sham molecular orbitals:

$$\psi_i = \sum_{j=1}^{L} c_{ji} \eta_j \quad . \tag{25}$$

For the complete basis set $\{\eta_j\}$, the number of basis set functions L must be infinite. However, in practical calculations, L is finite and it is of crucial importance to choose the $\{\eta_j\}$ such that the linear combination in Eq. (25) provides an accurate approximation of the exact Kohn-Sham orbitals.

The CGF basis sets combines several primitive Gaussian functions (GTO):

$$\eta^{CGF} = \sum_i d_i \eta_i^{GTO} \quad ,$$

$$\eta^{GTO} = N x^l y^m z^n \mathrm{e}^{-\alpha r^2} \quad ,$$

where N is a normalization factor, x, y, z are Cartesian coordinates and α is the orbital exponent, which determines how compact (large α) or diffusive (small α) the resulting function is. The contraction coefficient d is chosen in a way to resemble a single Slater type orbital (STO):

$$\eta^{STO} = N r^{n-1} \mathrm{e}^{-\zeta r} Y_{lm}(\theta, \phi) \quad ,$$

where the orbital exponent is termed ζ and Y_{lm} are the usual spherical harmonics that describe the angular part of the function. STO is more natural choice for basis functions. In contrast to GTO, Slater orbitals exhibit the correct cusp behavior at $r \to 0$ with a discontinuous derivative and desired exponential decay in the tail regions as $r \to \infty$. Whereas Gaussians have the property that the product of any two Gaussians is a Gaussian. Thus the reason for the popularity of the atom-centered Gaussian basis sets among chemists is that for a basis set consisting of Gaussians times polynomials centered at any site, all multi-centered integrals can be evaluated analytically.

The numerical basis functions $\{\eta_j^{num}\}$, not analytical like GTO and STO, are

---

53 F. Jensen, *Introduction to Computational Chemistry*, Wiley, Chichester, ch. 5, pp. 150-175, 2001.

54 See Ref. 41, chs. 14-15, pp. 273-311, and Ref. 40, ch. 7, pp. 93-116.

55 The scheme is introduced in the framework of the Hartree-Fock method. C. C. J. Roothaan, *New developments in molecular orbitals theory*, Rev. Mod. Phys. **23**, 69 (1951).

represented numerically on atomic centered grids, with cubic spline interpolations between mesh points. These basis functions are generated numerically solving the atomic Kohn-Sham equations with the corresponding approximate *xc*-functional. Thus, numerical basis sets provide exact energies for atomic fragments, but necessitate the use of purely numerical techniques for solving the integrals. In particular, in the SIESTA method, in order to achieve sparsity of the Hamiltonian and overlap matrices, confined basis orbitals are used, which go to zero beyond a certain radius from their center.[56,57]

Irrespective of whether one uses Gaussian functions or numerical basis set a certain classification is used to characterize the quality of basis functions. The minimal basis set (single-ζ or SZ) utilizes one basis function per each atomic orbital. Double-ζ (DZ) has two functions per each orbital, in case of CGF sets, two contracted functions per atomic orbital. Doubling of the numerical functions is achieved by adding numerically generated atomic orbitals from calculations on doubly or even higher positively charged ions. Taking into account that chemical reactivity induces changes in the valence region mainly and that the core electrons are chemically inert, one can limit the use of DZ to the valence orbitals only. This defines the *split-valence* type sets.

Polarization functions, i.e. functions of higher angular momentum than the occupied in the atom, can also be used. The polarization functions have more angular nodal planes than the occupied atomic orbitals and thus ensure that the orbitals can distort from their original atomic symmetry and better adapt to the molecular environment. The polarized double-ζ basis set (DZP or 6-31G(d,p) in terms of CGF sets), also used in several calculations in this work, provides a good compromise between accuracy and efficiency.

The quality of the numerical basis set, used in the Publication II, III and IV, is, in general, comparable to that of plane waves.[58] In Publication II, we performed our own comparison of the numerical atomic orbitals calculation, against Car-Parrinello plane wave calculations.[59] The results were in excellent agreement. The major advantage of the numerical basis set is the efficiency in computer simulations. One of the disadvantages is the basis set superposition error (BSSE), which appeared to be crucial for the interaction energy calculations, as has been shown in Publications II and III.

---

56 E. Artacho, D. Sánchez-Portal, P. Ordejón, A. García, and J. L. Soler, *Linear-scaling ab-initio calculations for large and complex systems*, Phys. Stat. Sol. B **215**, 809-817 (1999).

57 J .Junquera, O. Paz, D. Sánchez-Portal and E. Artacho, *Numerical atomic orbitals for linear-scaling calculations*, Phys. Rev. B **64**, 235111-235120 (2001).

58 The plane waves basis sets do not comply with the LCAO scheme. These are the solutions of the Schrödinger equation of a free particle and exponential function of the general form $\eta^{PW}=\exp(i\vec{k}\vec{r})$, where the vector $\vec{k}$ is related to momentum $\vec{p}$ of the wave through $\vec{p}=\hbar\vec{k}$. Plane waves are not centered at the nuclei but extended throughout the complete space. See Ref. 41, chs. 12-13, pp. 236-271.

59 R. Car and M. Parrinello, *Unified approach for molecular dynamics and density functional theory*, Phys. Rev. Lett. **55**, 15730-15741 (1985).

2.5 Pseudopotential approximation

The pseudopotential approximation (PS) is the third major approximation in practical calculations along with the *xc*-functional and basis sets.[60] In this thesis, vibrational spectra calculations were performed with an all-electron method as implemented in Gaussian code.[52] The quantum molecular dynamics simulations were performed with the pseudopotential method using the SIESTA approach.[29]

The method is based on the idea that the chemical properties of atoms are defined mainly by the valence electrons. The pseudopotential approximation uses this fact to remove the core electrons and the strong nuclear potential and replace them with a pseudo ionic core which interacts with the valence electrons. The pseudopotential is the effective interaction that approximates the potential felt by the valence electrons. Practically, for calculations that means that Hartree and *xc* terms of (15) are evaluated for valence electron density only. Besides, the core electrons are accounted for, by replacing the external potential V with a pseudopotential V^{PS}. In practice, there are various approaches for PS generation,[61,62,63] of which the SIESTA method uses the norm-conserving pseudopotentials proposed by Troullier and Martins[64] with Kleinman-Bylander separation (KB),[65] as they bring the desired smoothness. The pseudopotentials in semilocal form (a different radial potential $V_l(r)$ for each angular momentum l)[66] are transformed into the fully non-local form:

$$V^{PS} = V_{local}(r) + V^{KB} \quad ,$$

$$V^{KB} = \sum_{l=0}^{l_{max}^{KB}} \sum_{m=-l}^{l} \sum_{n=1}^{N_l^{KB}} |\chi_{lmn}^{KB}\rangle\langle\phi_{\ln}|\delta V_l(r)|\phi_{\ln}\rangle\langle\chi_{lmn}^{KB}| \quad ,$$

where $\delta V_l(r) = V_l(r) - V_{local}(r)$ and n, m, l are quantum numbers. $\chi_{lmn}^{KB}(\mathbf{r})$ are the KB projection functions:

$$\chi_{lmn}^{KB}(\mathbf{r}) = \chi_{\ln}^{KB} Y_{lm}(\hat{\mathbf{r}}) \quad ,$$

$$\chi_{\ln}^{KB}(r) = \delta V_l(r)\,\phi_{\ln}(r) \quad .$$

where $r = |\mathbf{r}|$, $\hat{\mathbf{r}} = \mathbf{r}/r$. The functions $\phi_{\ln}$ are thus obtained from the eigenstates $\psi_{\ln}$

---

60 Although the use of pseudopotentials is not strictly necessary with atomic basis sets.

61 W. E. Pickett, *Pseudopotential methods in condensed matter applications*, Comp. Phys. Rep. **9(3)**, 115-197 (1989).

62 F. Nogueira, A. Castro and M. A. L. Marques, *Pseudo-potentials*, in Ref. 37, ch. 6, pp. 230-239, 2003.

63 D. Porezag, M. R. Pederson and A. Y. Liu, *Accuracy of the pseudopotential approximation within density-functional theory*, Phys. Stat. Sol. B **217**, 219 (2000).

64 N. Troullier and J. L. Martins, *Efficient pseudopotentials for plane-wave calculations*, Phys. Rev. B **43**, 1993-2006 (1991).

65 L. Kleinman and D. M. Bylander, *Efficacious form for model pseudopotentials*, Phys. Rev. Lett. **48**, 1425-1428 (1982).

66 A pseudopotential that uses the same potential for all the angular components of the wave function is called a local pseudopotential. The local potential is a function only of the distance from the nucleus.

of the semilocal PS at energy $\epsilon_{\ln}$, using the orthogonalization scheme.[67]

The local part of the potential $V_{local}(r)$ is in principle arbitrary, but it must join the semilocal potentials $V_l(r)$ which, by construction, all become equal to the unscreened all-electron potential beyond the PS core radius r_c, i.e. $V_l(r)=V_{local}(r)$ for $r>r_c$. Thus $\chi^{KB}_{\ln}(r)$ is strictly zero beyond r_c, irrespective of the value of $\epsilon_{\ln}$. For a sake of transferability, which means that the same PS can be used in calculations in different chemical environment, it is sufficient to have a single projector $\chi^{KB}_{\ln}(r)$ per each angular momentum. Introducing the "norm-conservation", i.e. the charge within the r_c is the same for pseudo and true wave functions, $\epsilon_{\ln}$ must equal to the valence atomic eigenvalues ϵ_l, and the function $\phi_{\ln}$ is identical to the corresponding $\psi_{\ln}$.

In our simulations, in order to check the transferability of pseudopotentials, we calculated the same structure in different conformations with and without the pseudopotentials usage (all-electron simulations (AE) in the later case). The total energy differences should be equal, $(\Delta E^{AE}=\Delta E^{PS})$, for the pseudopotential to be transferable. The balance between softness of the pseudopotential and its transferability is controlled by cutoff radii, r_c. Larger r_c corresponds to a softer pseudopotential.[68]

---

67 P. E. Blöchl, *Generalized separable potentials for electronic-structure calculations*, Phys. Rev. B **41**, 5414-5417 (1990). See also Ref. 29.

68 In particular, for simulations of heme model and L-alanine molecule the pseudopotentials were generated using the following reference configurations: $2s^22p^2$ for carbon, $3s^23p^3$ for nitrogen, $3s^23p^4$ for oxygen, and $1s^2$ for hydrogen. The cutoff radii, r_c, (in Å) were 1.25 for both the s and p components in C, 1.24 for the s and p components in N, 1.14 for the s and p components in O, and 1.25 for the s component in H, respectively.

3. OPTIMIZATION TECHNIQUES

Atomic configuration optimization to the local potential energy minimum is a basic task in computer simulations performed in this thesis. As will be seen later, the minimum energy structure is needed for vibrational spectra calculations as well as for interaction energy calculations and structure comparisons. The potential energy surface is generally a complicated, multi-dimensional function of the coordinates. A number of efficient methods exist to determine a minimum on the potential energy surface. Those used in this thesis are outlined in the following *Section*.

3.1 Conjugate-gradients

Minimization techniques can be classified in two categories: first-order methods that use the first derivative and second order methods that use both first and second derivatives. The conjugate-gradients technique, a first-order method, is the one which has been used in the SIESTA simulations. Stationary points of a function $F(x)$ are defined by

$$\left(\frac{\partial F(x)}{\partial x}\right)_{x_i} = g_i = 0 \quad . \tag{26}$$

In the absence of any information about the function $F(x)$, the first step to the optimum direction from the point x_i is done in the direction of steepest descent g_i, $x_{i+1} = x_i - \alpha g_i$, but subsequent searches are preformed along a line which is a mixture of the current negative gradient and the previous search direction, $d_i = -g_i + \beta_i d_{i-1}$, where β is a scalar constant usually given by $\beta_i = \frac{g_i^t (g_i - g_{i-1})}{g_{i-1}^t g_{i-1}}$.[69]

3.2 Second-order methods

The second-order methods use in addition the second derivatives to locate a minimum. The matrix of second derivatives, the Hessian matrix, is

$$\left(\frac{\partial^2 F(x)}{\partial x^2}\right)_{x_i} = H_i \quad . \tag{27}$$

The major advantage is that these methods provide valuable information about the curvature of the function, which can be exploited to find stationary points more

[69] There are several ways of choosing β value, the one used is the Polak-Ribiere method. See for general discussion of the optimization techniques Ref. 25, ch. 14, pp. 316-346.

efficiently. If all eigenvalues of H are negative (positive) the stationary point corresponds to a maximum (minimum). In the case of negative and positive eigenvalues a saddle point was found. The expansion of the function $F(x)$ around x_i into a Taylor series is:

$$F(x)=F(x_i)+g_i^{\dagger}(x-x_i)+\frac{1}{2}(x-x_i)^{\dagger}H_i(x-x_i) \ .$$

Requiring the gradient of the second-order approximation to be zero produces the following step: $(x-x_i)=-H^{-1}g$. This scheme is called the *Newton-Raphson* method. For a purely quadratic function it finds the minimum in one step from any point on the surface. However, in practice the surface is only quadratic to a first approximation. This means that a number of steps will be required, at each of which the Hessian matrix must be calculated and inverted. The direct calculation of the second derivative can be a factor of 10 more demanding than calculating the gradient. The alternative *updating scheme* is used instead, so-called *quasi-Newton* methods in which the inverse Hessian matrix is gradually built up in successive iterations instead of considering the exact H . The initial H matrix can be just a unit matrix, the first step will thus resemble a steepest descent step. As the optimization proceeds, the gradients at the previous points are used to make the Hessian a better approximation for the actual system. After two steps, the updated Hessian is a rather good approximation to the exact Hessian, in the direction defined by these two points. There are many updating schemes, one of which was also used in the calculations for this thesis, is the Broyden-Fletcher-Goldfarb-Shano (BFGS) scheme,[70] its update formula is given by:

$$H_{i+1}=H_i+\frac{(x_{i+1}-x_i)\otimes(x_{i-1}-x_i)}{(x_{i-1}-x_i)\cdot(g_{i+1}-g_i)}-\frac{[H_i\cdot(g_{i+1}-g_i)]\otimes[H_i\cdot(g_{k+1}-g_i)]}{(g_{i+1}-g_i)\cdot H_i\cdot(g_{i+1}-g_i)}$$
$$+[(g_{i+1}-g_i)\cdot H_i\cdot(g_{i+1}-g_i)]u\otimes u$$

$$\text{where} \quad u=\frac{(x_{i+1}-x_i)}{(x_{i+1}-x_i)\cdot(g_{i+1}-g_i)}-\frac{[H_i\cdot(g_{i+1}-g_i)]}{(g_{i+1}-g_i)\cdot H_i\cdot(g_{i+1}-g_i)} \ .$$

The symbol $\otimes$ when interposed between two vectors means that a matrix is to be formed. It is clear that the quasi-Newton method does not converge as fast as the Newton-Raphson method, but since each step of the former one takes significantly less time than the true Hessian matrix calculation, the overall computational effort is smaller.

The actual method, which is used in the Gaussian calculations for this thesis is the *Synchronous Transit-Guided Quasi-Newton* (STQN) method.[71] It is the default algorithm for both minimizations (optimizations to a local minimum) and optimizations to transition states and higher-order saddle points. The method uses a linear synchronous transit or quadratic synchronous transit approach to get closer to the

---

70 R. Fletcher, in *Practical Methods of Optimization*, Wiley, New York, vol.1, 1980.

71 In terms of the Gaussian program it is called Berny algorithm. C. Peng and H. B. Schlegel, *Combining synchronous transit and quasi-Newton methods to find transition states*, Israel J. of Chem. **33**, 449 (1993).

quadratic region around the transition state and then uses a quasi-Newton algorithm with the BFGS update scheme to complete the optimization. The optimization is performed by using all distances, bending and torsional angles between atoms within bonding distance as variables, i.e. using redundant internal coordinates (the number of coordinates is larger than $3N-6$).[72] Internal coordinates appear to be a good choice for efficient quasi-Newton optimization.[73]

3.3 Locating the global minimum

The methods described in the previous sections can only locate the "nearest" minimum, which is normally a local minimum, when starting from a given set of variables. It is notable that the energy minimum search is especially difficult if the systems with the large number of atoms to be studied. As an alternative to "static" optimization techniques, molecular dynamics or simulated annealing might be used in search for global minimum.

Molecular dynamics methods solve Newton equation of motion on an energy surface (see next *Section*). The available energy is distributed between potential and kinetic energy, and molecules are thus able to overcome barriers separating minima if the barrier height is less than the total energy minus the potential energy. Given a high enough energy, which is related to the simulations temperature, the dynamics will sample the whole surface, but in practice it requires long time simulations. However, the current state of *ab initio* molecular dynamics methods and computer facilities allow the simulation times within the pico- or nanosecond range. Combined with the use of finite temperature, this means that only the local area around the starting point is sampled, and that only relatively small barriers can be overcome.

A search for the global minimum structure can be performed by using simulated annealing. Starting from sufficiently high temperature (for the system being investigated) at long run time, all the conformations have to be sampled. However, in theory the system cooling must be done infinitely slowly, implying infinite run time.

---

[72] C. Peng, P. Y. Ayala, H. B. Schlegel and M. J. Frisch, *Using redundant internal coordinates to optimize equilibrium geometries and transition states*, J. Comp. Chem. **17**, 49 (1996).

[73] F. Eckert, P. Pulay and H.-J. Werner, *Ab initio geometry optimization for large molecules*, J. Comp. Chem. **18(12)**, 1473-1483 (1997).

4. MOLECULAR DYNAMICS

Although the term molecular dynamics (MD) sounds very much like a topic specific to molecules, it is a very general method. MD describes the time dependent behavior of a group of particles in space under the influence of forces acting between the particles, and possibly external forces as well.[74] MD methods are now routinely used to investigate the structures, dynamics and thermodynamics of biological molecules and their complexes, although the particles in question can be atoms, either as part of a larger molecular system or isolated, or asteroids and planets or even parts of galaxies.[75] At the molecular scale, molecular dynamics simulations generate information at the microscopic level, including atomic positions and velocities.[76] Statistical mechanics provides the rigorous mathematical expressions that relate macroscopic properties, such as energy, pressure, heat capacity, etc., to the distribution and motion of the atoms of the N-body system.[74,77]

4.1 Newton's equations of motion

The basic formalism of MD simulations is common for all methods (the electronic structure calculations or the force fields) and based on classical *Newton's laws*. For an assembly of N atoms with positions r_i and masses m_i, the trajectory is obtained by solving the differential equations:

$$\frac{d^2 r_i}{dt^2} m_i = F_i \equiv -\frac{dE}{dr_i}, \tag{28}$$

where F_i is the force exerted on the particle i. Quantum and classical approaches differ in the way of treating the potential energy function $E(r_i)$. It is computed either by solving the Schrödinger equation or by substituting a parametrized force field.

The equations of motion are integrated further by using a *finite difference method*, or so-called integrator. The integration is broken down into many small stages separated by a fixed time δt (time step). The total force on each particle in the configuration at a time t is calculated as the vector sum of its interactions with other particles, which are then combined with the positions and velocities at a time t to calculate the positions

---

74 M. Allen and D. Tildesley, *Computer Simulations of Liquids*, Oxford Press, Oxford, 1989.

75 For the latter examples, the method is not called molecular dynamics, but the basic algorithm is the same.

76 By saying atom, in quantum molecular dynamics we assume the atom's nuclear and in coarse-grained methods a pseudoparticle.

77 J. M. Thijssen, *Computational Physics*, Cambridge University Press, Cambridge, ch. 7, pp. 146-174, 1999.

and velocities at time $t+\delta t$. All the integration schemes assume that the positions and dynamic properties, can be approximated as Taylor series expansions. A number of numerical algorithms have been developed for integrating the equations of motion, and of these algorithms the velocity Verlet,[78] based on the original Verlet algorithm,[79] is probably the most popular one.[80] Starting from $r_i(0)$ and $v_i(0)$, the trajectories of the N particles are generated iteratively as:

$$r_i(t+\delta t)=r_i(t)+\delta t\, v_i(t)+\frac{\delta t^2}{2m_i}F_i(t) \quad ,$$

$$v_i(t+\delta t)=v_i(t)+\frac{\delta t}{2m_i}[F_i(t)+F_i(t+\delta t)] \quad .$$

It is notable that the most computationally demanding part is the calculations of the forces at each iteration. The major criteria for the integration algorithm choice are that: (i) the algorithm should conserve energy and momentum, (ii) it should be computationally efficient, and (iii) it should permit a long time step for integration.

The integration procedure shows that to calculate a trajectory, one only needs the initial positions of the atoms, initial distribution of velocities and the potential energy function. The equations of motion are deterministic, e.g., the positions and the velocities at time zero determine the positions and velocities at all other times, t. The initial positions can be obtained from experimental structures and velocities are usually determined from a random distribution with the magnitudes conforming to the required temperature and corrected so that there is no overall momentum, i.e., for a system of N atoms: $P=\sum_{i=1}^{N} m_i v_i=0$. The velocities, v_i, are often chosen randomly from a Maxwell-Boltzmann or Gaussian distribution at a given temperature, which gives the probability that an atom i has a velocity v_x in the x direction at temperature T:

$$p(v_{x_i})=\left(\frac{m_i}{2k_BT}\right)^{1/2}\exp\left[-\frac{m_i v_{x_i}^2}{2k_BT}\right] \quad ,$$

with a temperature, T, given as

$$T=\frac{1}{(3\mathrm{N})}\sum_{i=1}^{N}\frac{|p_i|}{2m_i} \quad .$$

---

78 W. C. Swope, H. C. Anderson, P. H. Berens and K. R. Wilson, *A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: application to small water clusters*, J. Chem. Phys. **76**, 637-649 (1982).

79 L. Verlet, *Computer "Experiments" on classical Fluids I. Thermodynamical properties of Lennard-Jones molecules*, Phys. Rev. **159**, 98-103 (1967).

80 Verlet, leap-frog algorithm and Beeman's algorithm are discussed in A. Leach, *Molecular Modelling. Principles and Applications*, Pearson, Harlow, 2nd edn., ch. 7, pp. 353-406, 2001, and the original references therein. Similar discussion can be found in other textbooks.

4.2 Quantum molecular dynamics

4.2.1 Born-Oppenheimer molecular dynamics

In this thesis we have been using the Born-Oppenheimer molecular dynamics (BOMD).[81] In general, from equation (28) it is evident that MD is not tied to any particular electronic structure method. The forces can be computed at Hartree-Fock level, however, the efficiency in terms of CPU time requirements is the main obstacle for HF and post-HF methods and the main advantage of DFT. It is also clear that the strength and/or weakness of a particular *ab initio* MD implementation is a strength and/or weakness of the chosen electronic structure method. The Newton equation of motion (28) for BOMD in general:

$$\frac{d^2 R_i}{d t^2} m_i = -\nabla_i min_{\Psi_0}\{\langle \Psi_0 | H_e | \Psi_0 \rangle\} \ , \tag{29}$$

and for DFT:

$$\frac{d^2 R_i}{d t^2} m_i = -\nabla_i min_{\{\psi_i\}} E^{KS}[\{\psi_i\}] \ . \tag{30}$$

Thus the static electronic structure problem in each molecular time step is solved for the set of fixed nuclear positions $\{R_i\}$. The nuclei move according to classical mechanics in an effective potential due to the electrons. This potential is a function of only the nuclear positions at time t as a result of averaging H_e over the electronic degrees of freedom. In BOMD, the minimum of $\langle H_e \rangle$ has to be reached at every step, in DFT the minimum of the Kohn-Sham functional E^{KS} is obtained by varying the energy functional (13) with respect to the density $n(r)$.

4.2.2 Contributions to the Hellmann-Feynman forces

An important ingredient of all dynamics methods is the efficient calculation of the forces on the nuclei (29). If the gradients are evaluated analytically, in addition to the derivative of the Hamiltonian itself, there are in general also contributions from variations of the wavefunction $\sim \nabla_I \Psi_0$. The Hellmann-Feynman theorem stands, that if the wavefunction is an exact eigenfunction of the particular Hamiltonian under consideration, then these contributions vanish exactly, which is also valid for many variational wavefunctions provided that complete basis sets is used. However, it is not the case for numerical calculations and thus the additional terms have to evaluated explicitly.

For the LCAO approach (25), where the basis functions depend explicitly on the nuclear position (atom-centered orbitals), two sorts of additional forces are expected in addition to Hellmann-Feynman force:

$$F_I^{HF} = \langle \Psi_0 | \nabla_I H_e | \Psi_0 \rangle \ .$$

[81] D. Marx and J. Hutter, *Ab initio molecular dynamics: theory and implementation*, in *Modern Methods and Algorithms of Quantum Chemistry*, ed. J. Grotendorst, vol. 1, pp. 301-449, 2000.

One is a Pulay force (incomplete basis set correction), which contains the nuclear gradients of the basis functions and the effective one-particle Hamiltonian:

$$F_I^P = -\sum_{i\nu\mu} (\langle \nabla_I \eta_\nu | H_e^{NSC} - \epsilon_i | \eta_\mu \rangle + \langle \eta_\nu | H_e^{NSC} - \epsilon_i | \nabla_I \eta_\mu \rangle) \quad .$$

The second term is the so-called "non-self-consistency-correction" (NSC) and is governed by the difference between the self-consistent potential or field V^{SCF} and its non-self-consistent counterpart V^{NSC} associated to H_e^{NSC}:

$$F_I^{NSC} = -\int dr\, (\nabla_I n(r))(V^{SCF} - V^{NSC}) \quad .$$

The total force needed in QMD simulations:

$$F_I = F_I^{HF} + F_I^P + F_I^{NSC} \quad . \tag{31}$$

4.2.3 BOMD in practice

The time step choice in BOMD depends on the structure and, in particular, on the type of atoms involved. The value of the time step is a compromise between the efficiency and accuracy of simulations. The total energy conservation and small range of the energy fluctuations are the main indicators of the correct time step choice. The most typical time steps for the simulations with biological molecules are in the range of 0.5 fs to 1 fs.

Ab initio MD in its initial setup assumes a microcanonical ensemble, however, for constant pressure or constant temperature simulations this can be fixed by using Nose-Hoover or Parrinello-Rahman thermostats.[82] Another crucial point which must be taken into account is the periodic boundary conditions (PBC). PBC can be useful, for example, in simulations of solutions where large number of solvent molecules must be added explicitly. In the opposite case, when the isolated molecule is being studied, one has to make sure that the simulation shell is large enough, so that interactions with the neighboring species do not affect the process being investigated.

---

[82] Molecular dynamics simulations at constant temperature and pressure are discussed in the Ref. 80, ch. 7, pp. 382-387.

5. VIBRATIONAL SPECTRA CALCULATIONS

Vibrational spectroscopy[83,84] is among the traditional methods of analysis of substances including living material.[85,86] This *Section* introduces the underlying theory of the first principles calculations of the vibrational absorption (VA), vibrational circular dichroism (VCD), Raman scattering and Raman optical activity (ROA). The applications of the vibrational spectroscopy on several biomolecules are presented in Publications I and V. The accurate first-principles calculations are an important means of interpreting the resulting spectra.[87,88,89] Along with the well studied infrared (IR or VA) and Raman vibrational spectroscopical techniques,[90,91] the advanced methods, such as chirality sensitive VCD and ROA, are of special interest.[92] The large biomolecules, being investigated in solution, or in vivo, are likely to yield conventional IR and Raman spectra with many close-lying peaks. Thus spectroscopical techniques, which filter out information selectively, are needed.

5.1 Vibrational normal modes and frequencies

Spectra, the absorption or emission of light at characteristic wavelengths, have been long known to provide accurate fingerprints of molecular species. The detailed information contained in the many transitions, that characterize the spectrum of a molecule, is a reflection of the underlying atoms interactions within that molecule. The

---

83 Recently issued five-volume set handbook of vibrational spectroscopy covers practically all issues and applications to a greater or lesser extent. *Handbook of Vibrational Spectroscopy*, eds. J. M. Chalmers and P. R. Griffiths, Wiley, Hoboken, 2002.

84 J. I. Steinfeld, *Molecules and Radiation: An introduction to Modern Molecular Spectroscopy*, Dover, New-York, 2nd edn., 2005, (originally published by MIT Press, Cambridge, 2nd edn., 1985).

85 See Ref. 14, part E, pp. 573-687.

86 A. Barth and C. Zscherp, *What vibrations tell us about proteins*, Quart. Rev. Biophys. **35(4)**, 369-430 (2002), and references therein.

87 C. W. Bauschlicher Jr., and S. R. Langhoff, *The study of molecular spectroscopy by ab initio methods*, Chem. Rev. **91**, 701-718 (1991).

88 J. Tennyson and S. Miller, *Calculating molecular spectra*, Chem. Soc. Rev. **21**, 91-99 (1992).

89 B. A. Hess Jr., L. J. Schaad, P. Carsky and R. Zahradnik, *Ab initio calculations of vibrational spectra and their use in the identification of unusual molecules*, Chem. Rev. **86**, 709-730 (1986).

90 E. B. Wilson Jr., J. C. Decius and P. C. Cross, *Molecular Vibrations, The Theory of Infrared and Raman Vibrational Spectra*, Dover, New York, 2005, (originally published by McGraw-HBC, New York, 1955).

91 J. Neugebauer, M. Reiher, C. Kind and B.A. Hess, *Quantum chemical calculation of vibrational spectra of large molecules - Raman and IR spectra for Buckminsterfullerene*, J. Comp. Chem. **23(9)**, 895-910 (2002).

92 C. Herrmann and M. Reiher, *First-principles approach to vibrational spectroscopy of biomolecules*, Top. Curr. Chem. **268**, 85-132 (2007).

interaction is governed by the electronic potential energy surface (PES) of the system. Unfortunately, there is no general method of extracting potentials from spectroscopic data. This leads to an alternative strategy: the constructed guess potentials are tested by using them to generate spectra which can be compared with the experimental observation. Spectra calculations, in general, are separated into two tasks: (i) the calculations of the vibrational modes and frequencies, and (ii) the calculations of the corresponding intensities.

For a polyatomic molecule, within the Born-Oppenheimer approximation, the nuclear Schrödinger equation, which describe the nuclei as moving on the electronic PES defined by $E_{e,i}$ is

$$[T_n + E_{e,i}(R)]\chi_a(R) = E_{total}\chi_a(R) \quad , \tag{32}$$

where T_n is the nuclei kinetic energy, and χ_a is the nuclear part of the wave function $\Psi_I(R,r) = \chi_a(R)\cdot\phi_{e,i}(r;R)$.[93] The indices a and i label nuclear and electronic energy levels, respectively, and are defined by the composite total index I. The dependence of the PES on mass-weighted Cartesian nuclear coordinates $R_A^{(m)} = m_A^{1/2} R_A$ as a power series expansion about the equilibrium position $(R^{(m)}=0)$:

$$E_{e,i}(R^{(m)}) = E_{e,i}(0) + \sum_{A=1}^{3M}\left(\frac{\partial E_{e,i}}{\partial R_A^{(m)}}\right)_0 R_A^{(m)} + \frac{1}{2}\sum_{A,B=1}^{3M}\left(\frac{\partial^2 E_{e,i}}{\partial R_A^{(m)}\partial R_B^{(m)}}\right)_0 R_A^{(m)} R_B^{(m)} + \ldots \quad , \tag{33}$$

where A and B refers to the Cartesian coordinates of the specific nuclei and their nuclear masses. The first term $E_{e,i}(0)$ can be ignored for the purposes of most vibrational calculations. The second term vanishes since the system is at the equilibrium position. If all terms of order higher than second are neglected, the potential is a quadratic function of $R^{(m)}$:

$$E_e \approx \frac{1}{2}\sum_{A,B=1}^{3M}\left(\frac{\partial^2 E_{e,i}}{\partial R_A^{(m)}\partial R_B^{(m)}}\right)_0 R_A^{(m)} R_B^{(m)} \quad . \tag{34}$$

Equation (34) is known as the *harmonic approximation*.

After transformation of the Hessian matrix of Equation (34) to normal mass-weighted coordinates $Q_A^{(m)}$, the diagonalization of the Hessian yields $3N-6$ vibrational normal modes (for linear molecules $3N-5$), as well as three rotational and three translational modes with the eigenvalues of zero. The diagonal elements of Eq. (34) correspond to the angular frequency ω^2, and thus the frequencies are $\nu = \omega/2\pi$.

5.2 Potential energy surface calculation

Similar to that of molecular dynamics simulations, first principles vibrational spectroscopy calculations are not tied to any particular electronic structure method.

---

[93] The nuclear part χ_a depends on the $3M$ coordinates R of all M nuclei, and the electronic part $\phi_{e,i}(r;R)$ depends explicitly on the $3N$ coordinates r of N electrons, and parametrically on the nuclear coordinates.

DFT is effective and sufficiently accurate method for PES and spectra calculations and in particular applicable to large biomolecules.

In practice, the calculated harmonic vibrational frequencies, depending on the computational method, are underestimated or overestimated as compared to fundamentals observed experimentally. However the error is relatively systematic, and as a result generic frequency scaling factors are often applied.[94] The major source of this disagreement is the neglect of the anharmonicity, finite basis set size, and the errors due to the *xc*-functionals. Several studies reported the quality of the harmonic frequencies predicted by various DFT functionals as compared to other methods.[95,96]

One more practical issue in the vibrational normal modes calculations is the optimized molecular geometry. If the fundamental frequencies for rotations are not "close enough" to zero, it most probably means that the system is not at the equilibrium.

5.2 Energy perturbation

The Hamiltonian of a molecule in the electromagnetic field of a light wave is a sum of the Hamiltonian of the isolated molecule H_{mol} and the Hamiltonian describing its interaction with the radiation H_{rad}, $H = H_{mol} + H_{rad}$. Interactions with the electromagnetic field are described within the first-order time-dependent perturbation theory. The H_{rad} operator in the case of dynamic electromagnetic field with electric and magnetic filed amplitudes F_0 and B_0 can be written as:

$$H_{rad} = -\sum_{i=1}^{3} \mu_i (F_i)_0 - \frac{1}{3} \sum_{i=1,\, j=1}^{3} \theta_{ij} \nabla_i (F_j)_0 - \sum_{i=1}^{3} m_i (B_i)_0 \ \ldots \ , \qquad (35)$$

where μ_i, θ_{ij}, and m_i are the x, y, and z components of the electric dipole, the electric quadrupole, and the magnetic dipole operator, respectively, and $(F_i)_0$, $(\nabla_i F_j)_0$, and $(B_i)_0$ are the x, y, and z components of the electric and magnetic field and its gradient at the gauge origin.[97]

5.3 Intensities

Formulas and the methodology of obtaining the VA, Raman, VCD and ROA are given in the methodology section of Publications I and V. The short summary is presented here.

---

94 (a) A. P. Scott and L. Radom, *Harmonic vibrational frequencies: an evaluation of Hartree-Fock, Moller-Plesset, quadratic configuration interaction, density functional theory and semiempirical scale factors*, J. Phys. Chem. **100**, 16502-16513 (1996), and references therein; (b) M. W. Wong, *Vibrational frequency prediction using density functional theory* , Chem. Phys. Lett. **256(4-5)**, 391-399 (1996).

95 X. Zhou, C. J. M. Wheeless and R. Liu, *Density functional theory study of vibrational spectra. 1. Performance of several density functional methods in predicting vibrational frequencies*, Vib. Spectrosc. **12**, 53 (1996).

96 See Ref. 40, ch. 8, pp.130-136, and references therein.

97 L. D. Barron and C. G. Gray, *The multipole interaction Hamiltonian for time dependent fields*, J. Phys. A: Math., Nucl. Gen. **6**, 59-61 (1973).

The Raman scattering (RS) intensities and infrared (IR) or vibrational absorption (VA) intensities[91] as arisen from interactions with the electromagnetic field. The RS intensities are proportional to the derivatives of the electric dipole electric dipole polarizability (EDEDP) with respect to normal coordinates,

$$RS \propto \left(\frac{\partial \alpha}{\partial Q_i} \right)^2 \propto \left(\frac{\partial^3 E}{\partial Q_i \partial F_\beta \partial F_\gamma} \right)^2 , \tag{36}$$

and the VA intensities are proportional to the derivatives of the electric dipole moment with respect to normal coordinates,

$$VA \propto \left(\frac{\partial \mu}{\partial Q_i} \right)^2 \propto \left(\frac{\partial^2 E}{\partial Q_i \partial F_\alpha} \right)^2 . \tag{37}$$

Where the first derivative of the energy with respect to the electric field is the *dipole moment*,

$$\mu = -\left(\frac{\partial E}{\partial F} \right) , \tag{38}$$

the second derivative is the *dipole polarizability*,

$$\alpha_{\beta\gamma} = \left(\frac{\partial^2 E}{\partial F_\beta \partial F_\gamma} \right) . \tag{39}$$

Vibrational circular dichroism[98,99,100,101] and vibrational Raman optical activity[102,130,103,104] spectra are in principle recorded in the same way as standard IR and Raman spectra, respectively, but instead of total intensities, the intensity differences between right and left circularly polarized light are plotted.

---

98 P. J. Stephens, *Theory of vibrational circular dichroism*, J. Phys. Chem. **89**, 748-752 (1985).

99 R. D. Amos, N. C. Handy, K. J. Jalkanen and P. J. Stephens, *Efficient calculation of vibrational magnetic dipole transition moments and rotational strengths*, Chem. Phys. Lett. **133**, 21-26, (1987).

100 P. J. Stephens, *Gauge dependence of vibrational magnetic dipole transition moments and rotational strengths*, J. Phys. Chem. **91**, 1712-1715 (1987).

101 K. L. Bak, F. J Delvin, C. S. Ashvar, P. R. Taylor, M. J. Frisch and P. J. Stephens, *Ab initio calculation of vibrational circular dichroism spectra using gauge-invariant atomic orbitals*, J. Phys. Chem. **99**, 14918-14922 (1995).

102 L. A. Nafie, *Infrared and raman optical activity: theoretical and experimental aspects*, Annu. Rev. Phys. Chem. **48**, 357-86 (1997).

103 L. A. Nafie, *Vibrational optical activity*, Appl. Spectr. **50(5)**, 14A-26A (1996).

104 L. Ashton, L. D. Barron, B. Czarnik-Matusewicz, L. Hecht, J. Hyde and E. W. Blanch, *Two-dimensional correlation analysis of Raman optical activity data on the alpha-helix-to-beta-sheet transition in poly(L-lysine)*, Mol. Phys. **104**, 1429-1445 (2006).

6. FIRST-PRINCIPLES STUDIES OF BIOMOLECULES

This thesis is based on a number of publications prepared in 2005-2007, in which we have addressed several biological problems.

6.1 Structure and vibrational spectra of (R)-phenyloxirane

In Publication I we have studied the vibrational spectra of (R)-phenyloxirane biomolecule. We have simulated the VA, VCD, Raman and ROA vibrational spectra by utilizing the first-principles calculations. The goal of the theoretical investigation was to interpret the results of the experimental studies. In this work we use B3LYP xc-functional and 6-31++G(d,p) CGF basis set[52] to optimize he geometry of the phenyloxirane and calculate the Hessian matrix, the atomic polar tensors (APT), the atomic axial tensors (AAT), and the electric dipole - electric dipole polarizability derivatives (EDEDPD). APT, AAT, and EDEDDP allow us to simulate VA, VCD and Raman spectra. In addition to these simulations we calculated electric dipole - magnetic dipole polarizability (EDMDP) and electric dipole - electric quadrupole polarizability (EDEQP) and their derivatives, EDMDPD and EDEQPD, respectively. By having this quantities we are able to simulate the ROA spectrum.[105]

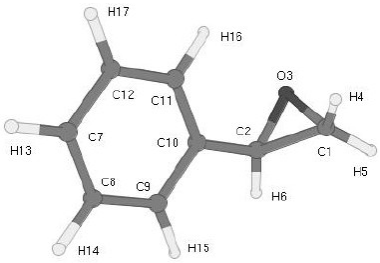


Figure 4. (R)-Pheniloxirane structure.

In addition, in order to investigate the basis set and *xc*-functional dependence, the number of *xc*-functionals and various basis sets have been used. The level of agreement with the experimental spectra being the measures. We have optimized geometry and calculated VA, VCD and Raman spectra with PBE, PW91 and PBE *xc*-functionals. Furthermore, we calculated spectra at B3LYP level of theory by utilizing the split valence double-ζ plus polarization functions (6-31G(d,p)), plus diffusive functions (6-31G++(d,p)), the correlation consistent polarized valence with double-ζ and triple-ζ basis sets, and with diffusive functions (cc-pVDZ, cc-pVTZ, aug-cc-pVDZ, and aug-cc-pVTZ respectively).

[105] For the ROA specta simulation Cambridge Analytical Derivatives Package (CADPAC) has been utilized. The EDMDP and EDEQP were simulated at restricted Hartree-Fock wtih 6-31G(d,p) basis set level of theory and EDMDPD and EDEQPD where calculated with aug-cc-pVDZ basis set. R. D. Amos *et al.*, *CADPAC: The Cambridge Analytical Derivatives Package Issue 6*, 1995.

The overall results of (R)-phenyloxirane structure optimization shows that variations in bond lengths with respect to different *xc*-functional and basis sets, compared against that of B3LYP/6-31G(d,p), are close to the error bar typical for the method used.[106] The largest effect due to basis set appears on dihedral angles $\tau_{3,2,10,11}$ (Figure 4). The values differ by about 7 degrees, reflecting the effect of adding the diffuse functions.

The analysis of the spectra simulation suggests the VA spectra is not that sensitive with respect to xc-functional and basis set. The geometry and Hessian calculated with at B3LYP/6-31G(d,p) level of theory is accurate enough. And in some cases the smaller basis set gives better agreement than the larger one. Similar case with respect to the *xc*-functional changing. However the better agreement for some particular modes can not reflect the overall better accuracy. One has to keep in mind that presented theoretical spectra are calculated within the harmonic approximation, whereas the experimental one is anharmonic. The role of the larger basis set is more pronounced for the VCD intensities than for the VA. The calculations suggest that VCD requires the addition of the diffusive basis functions.

The good agreement between the calculated VA and VCD spectra with the experimental spectra[130] leads us to believe that Raman and ROA spectra are also accurately predicted. We have shown that Raman spectral measurements can be used to validate CH stretch mode not seen by VA. With respect to the calculation level of theory, the polarization orbitals have the largest effect on Raman spectra, and the additional diffuse orbitals are not that important. Changes in peaks heights and frequencies of the Raman spectra due to different basis sets and *xc*-functionals reflect the differences in geometrical parameters.

Overall, results of the simulations suggest that the larger than DZP basis set is required for optimal calculations of VA, VCD, Raman and ROA spectra. In general, this study declares that the above mentioned vibrational spectroscopies of moderate size biomolecules, have now reached the stage where they can be routinely calculated at *ab initio* level.

6.2 Dynamics of the isolated Co- and Fe-heme models

Molecular dynamics simulations of the myoglobin active site, heme, has been reported in Publication II. The idea of the study stems from several reasons. The earlier results of Car-Parrinello molecular dynamics simulations of the myoglobin active centre, performed by one the authors, provide evidence for the rotational motion of the O_2 ligand when bound to the Fe-heme.[107] However, in most of the experiments in which the rotational motion of the ligand has been studied, the prosthetic group (iron

---

[106] A. M. Mebel, K. Morokuma and M. C. Lin, *Modification of the gaussian−2 theoretical model: The use of coupled-cluster energies, density-functional geometries, and frequencies*, J.Chem. Phys. **103**, 7414-7421 (1995). See also Ref. 40, ch. 8, pp.119-136. The reported mean absolute deviations for bond lengths and angles for B3LYP are up to 0.009 Å and 1.7 degrees respectively.

[107] C. Rovira and M. Parrinello, *Harmonic and anharmonic dynamics of Fe-CO and Fe-O_2 in heme models*, Biophys. J. **78**, 93-100 (2000).

protoporphyrin IX) is replaced by cobalt protoporphyrin IX. Due to the S = ½ electronic configuration of Co-heme-O_2, the protein molecule becomes optically and EPR active. The experiments show that rotation of the O_2 about the Co-O bond takes place in the sub-nanosecond time scale.[108] The results of such dynamical studies of the cobalt analogue are typically extrapolated to iron. Nevertheless, a comparative analysis of the dynamics of the metal-ligand bonds in both cases (Fe and Co) has not been available. The first X-ray structure of Co-substituted myoglobin shows that the bound O_2 is at a hydrogen bond distance from the His64 residue.[109] The main objectives of this work are (i) to quantify the differences and similarities between the two prosthetic groups: iron and cobalt protoporphyrin IX, with respect to the dynamics of bound dioxygen, and (ii) to investigate the role of O_2..His64 hydrogen bonding.

The models used include the heme active site, the oxygen ligand, and the two closest protein residues. These are the proximal histidine (His93), which is covalently attached to the iron atom) and the distal histidine (His64), which is at hydrogen bond distance with the ligand in X-Ray analyses. Therefore, the actual simulations are performed on four heme models, for both Fe and Co, with and without distal His64, i.e. MeP(Im)O_2 and MeP(Im) O_2···His64, where Me is a metal atom, Fe or Co. Backbone atoms of the distal His64 residue and four hydrogen atoms of the porphyrin plane have been fixed during the simulations, to reproduce the steric restrictions imposed by the protein backbone. The simulations were performed for a total period of 30 ps. Similar model was used in the past to study the short time scale dynamics of FeP(Im)O_2 (up to 14 ps) model.[107]

The overall results show that the structure of the MeO_2 fragments (Me = Fe, Co) are in agreement with the available experimental information.[109] It is found that in CoO_2 both oxygen atoms interact with the distal histidine, forming a hydrogen bond (4.7 kcal/mol), while FeO_2 only uses the terminal oxygen and forms a weaker bond (3.7 kcal/mol). In both cases, the dynamics of the O_2 ligand can be described as oscillations of the O-O axis projection on the porphyrin plane within a porphyrin quadrant, with frequent jumps to the neighboring quadrant. However, the ligand motion is significantly faster for CoO_2 compared to FeO_2. As a result, the iron complex shows localized ligand sites while for cobalt several configurations are possible. This gives support to the highly dynamic motion of these bonds found in several experiments[108] and underlines the higher fluxionality of the CoO_2 fragment compared to FeO_2.

Another task of this study was to establish the level of confidence of the newly

---

[108] J. Bowen, N. Shokirev, A. Raitsimring, D. Buttlaire and F. Walker, *EPR studies of the dynamics of rotation of dioxygen in model cobalt(II) hemes and cobalt-containing hybrid hemoglobins*, J. Phys. Chem.B **101**, 8683-8691 (1997).

[109] E. Brucker, J. Olson, G. Phillips, Y. Dou and I.-S. Masao, *High resolution crystal structure of the deoxy, oxy, and aquomet forms of cobalt myoglobin*, J. Bio. Chem. **271**, 25419-25422 (1996).

applied DFT method SIESTA.[110,111,112,113,114] Our results obtained for the FeP(Im)O_2 model reproduce fairly well those reported earlier for Car-Parrinello MD,[115,116,107] which shows that change from a plane wave basis set to numerical atomic orbitals does not introduce any bias into the calculations.

The simulation setup has room for improvement. The planned extension to this study is to apply the QM/MM approach recently introduced in the SIESTA code. Thus the entire protein and solvent molecules can be treated classically and the active center of the myoglobin at quantum mechanical level. This kind of simulation will give a more realistic picture of the dynamics of the bonded O_2 ligand and will also give rise to simulations of processes which involve releasing and bonding of O_2 ligand and other ligands (CO, NO). Treating the rest of the myoglobin at MM level will allow dispersion forces to be taken into account, which possibly might have an impact on the O_2 dynamics.

6.3 First-principles studies of the hydrated L-alanine amino acid

6.3.1 Dynamics of the L-alanine in the hydration shell

The initial idea of the vibrational spectra calculations for a number of hydrated ionic L-alanine species at different pHs[117] has been expanded firstly to a BOMD study of the L-alanine hydration shell and dynamics (Publication III), and secondly to the

---

110 SIESTA method has been introduce in the number of publications[111] and thoroughly described in details in the recent update.[29] Method has been successfully applied in the area of nanoscale materials, surfaces and interfaces, metallic systems, ferroelectrics and semiconductors.[112] The methodology used in our simulations for isolated heme model, and L-alanine in aqueous solution was previously applied to investigate other biologically relevant systems.[112a] In particular to DNA strand and proteins.[113]

111 (a) D. Sánchez-Portal, P. Ordejón, E. Artacho and J. M. Soler, *Density-functional method for very large systems with LCAO basis sets*, Int. J. Quant. Chem **65**, 453-461 (1997). (b) P. Ordejón, E. Artacho and J. M. Soler, *Self-consistent order-N density-functional calculations for very large systems*, Phys. Rev. B **53**, 10441-10444 (1996).

112 (a) P. Ordejón, *Linear scaling ab initio calculations in nanoscale materials with SIESTA*, Phys. Stat. Sol. B **217**, 335-356 (2000). (b) D. Sánchez-Portal, P. Ordejón and E. Canadell, *Computing the properties of materials from first principles with SIESTA*, Struct. Bond. **113**, 103-170 (2004), and references therein.

113 (a) E. Artacho, M. Machado, D. Sanchez-Portal, P. Ordejon and J. M. Soler, *Electrons in dry DNA in density functional calculations*, Mol. Phys. **101(11)**, 1587-1594 (2003). (b) A. Hubsch, R. G. Endres, D. L. Cox and R. R. P. Singh, *Optical conductivity of wet DNA*, Phys. Rev. Lett. **94**, 178102-178104 (2005). (c) P. J. Pablo, F. Moreno-Herrero, J. Colchero, J. Gómez-Herrero, P. Herrero, A. M. Baró, P. Ordejón, J. M. Soler and E. Artacho. *Absence of dc-Conductivity in λ-DNA*, Phys. Rev. Lett. **85**, 4992-4995 (2000).

114 M. A. Martí, D. A. Scherlis, F. A. Doctorovich, P. Ordejón and D. A. Estrin. *Modulation of the NO trans effect in heme proteins: implications for the activation of soluble guanylate cyclase*, J. Biol. Inorg. Chem. **8**, 595-600 (2003).

115 R. Car and M. Parrinello, *Unified approach for molecular dynamics and density-functional theory*, Phys. Rev. Lett. **55**, 2471-2474 (1985).

116 M. C. Payne, M. P. Teter, D. C. Allan, A. C. Arias and J. D. Joannopoulos, *Iterative minimization techniques for ab initio total-energy calculations: molecular dynamics and conjugate gradients*, Rev. Mod. Phys. **64**, 1045-1097 (1992).

117 Over a wide range of pH the L-alanine amino acid has both a positively charged amine group ($-NH_3^+$) and a negatively charged carboxylate group ($-COO^-$). At low pH the cationic species is dominant with the carboxylate group being protonated (-COOH), while at high pH the anionic species in dominant with the amine group giving up one of its protons to form the neutral amino group ($-NH_2$).

comparative study of L-alanine zwitterion in different media, in water and in crystal (Publication IV). The fact is that the rotational (microwave) spectra in solution are not resolved to give structural information directly, so that the structural parameters of alanine molecules in water come predominantly from computational studies.[118,119] Those computational studies treat the effect of hydration either by including water molecules explicitly or by simulating within the quantum continuum models[33] or both. The number of explicit water molecules was limited due to the computational complexity, four and nine correspondingly.[118,119] The initial positions and orientations of the solvent molecules were chosen intuitively. After an extensive literature search we also found out that a number of the diffraction studies[120,121,122] which focused mainly on the structure of the hydration shell of several amino acids with the hydrophobic side chains, including alanine, are mutually contradictory. Hetch *et al.*[122] detected a strong correlation between hydration shell structure and type of side chain, while Ide *et al.*[120] proposed that the hydration shell is mainly formed by the charged ammonium and carboxylate groups and hydrophobic side chains have little influence on its structure. There was a clear need for a computational experiment in order to resolve two tasks, (i) predicting the actual structure of the L-alanine amino acid in the solution and (ii) detecting the hydration structure.

Ab initio molecular dynamics of the L-alanine molecule hydrated in the relatively large number of water molecules was the essential setup for the simulations. The method allows the structure of the hydration shell and the structure of the amino acid itself in the aqueous solution to be quantitatively and qualitatively determined. The predictive power of the *ab initio* method also allows taking into account the possible proton transfer reaction between the ammonium group and water molecules or the carboxylate group.[123] However we have joined two levels of theory quantum mechanics and molecular mechanics. The simulations were performed in several steps (i) initial system relaxation with a large number of water molecules using classical atomistic molecular dynamics (MD) simulations, (ii) structure optimization of an extracted alanine molecule with a limited number of the water molecules nearest to the alanine

---

118 E. Tajkhorshid, K. J. Jalkanen and S. Suhai, *Structure and vibrational spectra of the zwitterion L-alanine in the presence of explicit water molecules: a density functional analysis*, J. Phys. Chem. B **102**, 5899-5913 (1998).

119 K. Frimand, H. Bohr, K. J. Jalkanen and S. Suhai, *Structures, vibrational absorption and vibrational circular dichroism spectra of L-alanine in aqueous solution: a density functional theory and RHF study*, Chem. Phys. **255**, **255**, 165-194 (2000).

120 M. Ide, Y. Maeda and H. Kitano, *Effect of hydrophobicity of amino acids on the structure of water*, J. Phys. Chem. B **101**, 7022-7026 (1997).

121 (a) Y. Kameda, K. Sugawara, T. Usuki and O. Uemura, *Hydration structure of alanine molecule in concentrated aqueous solutions*, Bull. Chem. Soc. Jpn. **76**, 935-943 (2003). (b) Y. Kameda, M. Sasaki, M. Yaegashi, K. Tsuji, S. Oomori, S. Hino and T. Usuki, *Hydration structure around the carboxyl group studied by neutron diffraction with $^{12}C/^{13}C$ and H/D isotopic substitution methods*, J. Solut. Chem. **33(6-7)**, 733-745 (2004).

122 D. Hetch, L. Tadesse and L. Walters, *Correlating hydration shell structure with amino acid hydrophobicity*, J. Am. Chem. Soc. **115**, 3336-3337 (1993).

123 See the additional Section below were we show how the proton transfer reaction changes the structure of L-alanine zwitterion in the gas-phase. For simulations with the water molecules see S.-W. Park, D.-S. Ahn and S. Lee, *Dynamic path between neutral alanine-water and zwitterionic alanine-water clusters: single, double and triple proton transfer*, Chem. Phys. Lett. **371**, 74-79 (2003).

molecule, and (iii) density functional theory Born-Oppenheimer molecular dynamics simulations starting from the minimum energy structure of step (ii).[124] Five hundred water molecules were used for the atomistic MD and fifty water molecules were adopted for the first principles studies of the dynamics and structure of the L-alanine molecule. Another aspect of the simulations is that the quantum molecular dynamics was performed in a large cell, thus the structure was solved in the "droplet of water". This allowed us to study hydrophobic effects and the role of the side chain on the hydration structure.

The results reported in Publication III suggests the trivial but fundamental fact, which has never been shown in simulations, that L-alanine zwitterion is stable due to hydrogen bonding interactions with the water molecules. Analysis of MD also suggests that the first hydration shell around the amino acid is defined by the charged ammonium and carboxylate groups and practically not affected by the methyl group or α-hydrogen. The hydration structure is rigid compared to bulk water and highly ordered. Water molecules inside the hydration shell are almost always hydrogen bonded to NH_3^+ and COO^- groups. The hydration shell has been defined quantitatively we found that on average there are 7 water molecules inside the shell, and it expands 2.37 Å and 3.04 Å from the ammonium and carboxylate groups, respectively.

6.3.2 The structure in water and in crystal

The simulated data gave rise to a comparative study on L-alanine zwitterion in water and in crystal. In Publication IV we report the averaged structure of the L-alanine in aqueous solution and compare it to amino acid's crystalline form. As mentioned above, the actual structure of L-alanine in the aqueous solution is known from the computational studies, and thus it is generally accepted that the structure of an alanine zwitterion obtained from the crystalline phase can also be applied to alanine in aqueous solution. However, the origin for the zwitterionic form is different in crystal and in water. In crystals all three available protons of the ammonium group (NH_3^+) are used to form single N-H···O hydrogen bonds with oxygen atoms of three carboxylate groups (COO^-) of nearest amino acid molecules, thus linking the molecules together to form a three-dimensional crystal structure. In contrast, the key determinant of the stable zwitterionic structure in aqueous solution is hydrogen bonding interactions with surrounding water molecules (N-H···O_w and O···H_wO_w type of hydrogen bonds).

Additional analysis of the MD trajectory and structure optimization calculations of both aqueous and crystalline structures of L-alanine suggest that the structures can differ significantly. The largest difference between zwitterionic forms of L-alanine in water and in crystal can be attributed to the carboxylate group. Simulations predict strong correlations between the C'-$O^{1,2}$ bond lengths with the number of water molecules which form hydrogen bonds to the COO^- group (C', O^1 and O^2 are atoms of the carboxylate group, see Figure 1 in Publication III). We reported bond lengths and angles for the averaged amino acid structure in water at room and zero temperatures.

[124] Using atomistic MD to relax the structure instead of time consuming *ab initio* approach, can be considered as another method of overcoming time and size scales described in the *Section 1.4*.

However, the relatively short simulation period (40 ps) and the use of the finite temperature means that only the local area (number of conformations) around the starting point is sampled. Especially with the large number of the explicit water molecules in the system, number of local energy minima is large. Thus the different minima generated by selecting and optimizing configurations at suitable trajectory intervals are local minima. And the reported lowest energy structure (i.e. at 0 K) is most probably a local minimum structure. Also taking into account the fact that the geometry parameters at room temperature are averaged over a relatively short simulation time we can conclude that the reported results have to be considered as more qualitative than quantitative.

A search for the global minimum structure can be performed by using simulated annealing. Starting from sufficiently high temperature at long run time, all the conformations have to be sampled. However, in practice the system cooling must be done infinitely slowly, implying infinite run time.

6.3.3 Case study: isolated L-alanine amino acid

Isolated alanine amino acid[125] is one of the representative example of simulations where the molecular system can not be treated classically (Figure on the right). We have reported in our original publication that L-alanine zwitterion is not stable isolated in the gas-phase (Publication III). We performed a number of simulations, starting from different initial conformations, but the result for all of them was the same. The conformations of the isolated L-alanine zwitterion, in the absence of water molecules or neighbor amino acids as in crystal, ended up being neutral non-ionic. For all starting geometries, the L-alanine amino acid adopted conformations in which the acidic carboxylate group received a proton from the basic ammonium group and the molecule became nonionic (NH_3^+ ... COO^- → NH_2 ... COOH). Additional simulations have shown that the hydrated non-ionic alanine takes one of the ionic forms.

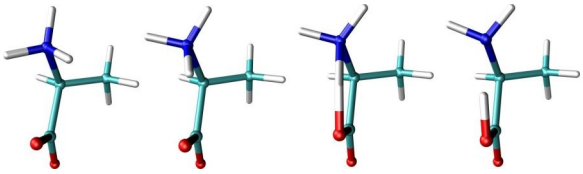


Figure 5. Evolution of the L-alanine zwitterion to nonionic form in the gas phase simulations.

The non-ionic form is detected from the number of experiments[126] and found in the organic content of extraterrestrial meteorites.[127] But the zwitterionic form is the most abundant form in living organisms: they are synthesized in the zwitterionic form and

---

125 Alanine is the smallest chiral α-amino acid, with a non-reactive methyl group ($-CH_3$) as the side chain. Zwitterionic species of amino acids have both a positively charged ammonium group (NH_3^+) and a negatively charged carboxylate group (COO^-). They are the dominant form in aqueous solution over a wide range of pH.

126 S. Blanco, A. Lessari, J. C. Lopez and J. L. Alonso, *The gas-phase structure of alanine*, J. Am. Chem. Soc. **126**, 11675-11683 (2004).

127 (a) M. H. Engel and S. S. Macko, *Isotopic evidence for extraterrestrial non-racemic amino acids in the Murchison meteorite*, Nature **389**, 265-268 (1997); (b) S. Pizzarello, *The chemistry of Life's Origin: a carbonaceous meteorite perspective*, Acc. Chem. Res. **39(4)**, 231-237 (2006).

are essentially always in this form at wide range of pH.[1] Our simulations demonstrate the case where a H atom transfer reaction effect can not be omitted.

6.3.4 Vibrational spectra of L-alanine zwitterion in aqueous solution

In parallel to the comparative study on L-alanine zwitterions in water and in crystal we also investigated vibrational absorption, vibrational circular dichroism (VCD), Raman and Raman optical activity spectra (ROA). Experimental and calculated spectra are reported. For calculations, well equilibrated initial structures of L-alanine in water have been extracted from the MD trajectory. Thus we avoid manual positioning of the large water molecules around the amino acid.[128,129] The number of water molecules has been reduced from fifty down to twenty due to the computational complexity of the vibrational spectra calculations (see *Section 5*).

We look for and identify the modes which characterize the zwitterionic structure and also the structural parameters of interest, C-O bond lengths, N-H bond lengths, and the N-C-C-O torsion angles. Here we have for the first time simulated the VA, VCD and Raman spectra for a biomolecule fully encapsulated by an aqueous solvation shell. Note that the zwitterionic form of L-alanine and other amino acids are not stable in the isolated state, see *Publication III* and *Section 6.3.3*. Hence one must treat the environment to even get a stable structure. The already mentioned earlier works with four and nine water molecules in the amino acid hydration shell showed that by increasing the number of water molecules, the agreement with the experimental spectra improved.[128,129] In this work, we not only have fully encapsulated the LAZ and found that 20 explicit water molecules were necessary to fit the experimental spectra, but we have also embedded the LAZ and 20 water molecules within the Onsager, polarized continuum (PCM) and conductor (COSMO) continuum solvent models. The continuum solvent models have been added to treat the effects due to the bulk water molecules, which were found to be significant for the small LAZ clusters. Here we show that the effects due the continuum solvent models is reduced when one has fully encapsulated the LAZ. These type of simulations, especially for the ROA, are computationally expensive. We do not at this time advocate this level of theory for routine work, but only to serve as benchmark calculations.

In this study we also investigated the causes for the stability of the P_π conformer of the alanine dipeptide. There has been some controversy of the causes of the stability of this structure. Our work fully substantiates that this structure is stabilized by the hydrogen bonded network of water molecules. Here we have used both the B3LYP and B3PW91 hybrid exchange correlation functionals and the aug-cc-pVDZ basis set for the N-acetyl L-alanyl N'-methylamide (NALANMA) plus four explicit water molecules. Our spectral simulations at this level of theory, including the PCM and

---

[128] E. Tajkhorshid, K. J. Jalkanen and S. Suhai, *Structure and vibrational spectra of the zwitterion L-alanine in the presence of explicit water molecules: a density functional analysis*, J. Phys. Chem. B **102**, 5899-5913 (1998).

[129] K. Frimand, H. Bohr, K. J. Jalkanen and S. Suhai, *Structures, vibrational absorption and vibrational circular dichroism spectra of L-alanine in aqueous solution: a density functional theory and RHF study*, Chem. Phys. **255**, 165-194 (2000).

COSMO continuum solvent models put an end to this controversy (in our opinion). Hence we advocate the inclusion of the first solvation shell solvent (or ligands, cations, anions and/or combination thereof) which are responsible for stabilizing the structure present in aqueous solution under either native or non-native conditions. Here the combination of experimental and theoretical spectroscopy has been able to answer a question which had escaped X-ray and neutron diffraction and NMR, the structure of the alanine dipeptide in aqueous solution. This structure has been subsequently confirmed by NMR studies and this technique has now been called the standard for structural problems where X-ray and neutron diffraction and NMR techniques have failed. This thesis is the documentation of my work to further extend the applicability of this technique to larger systems. To do so, required one to fully document the interactions responsible for the stability of the P_{π} conformer and to be able to reproduce the newly reported Raman and ROA spectra for this molecule. The goal now is to make this combined technique/method more transparent so that it can be used by the community of researchers in the fields of molecular biology and physical biochemistry, all of whom are not experts in DFT and molecular dynamics.

7. SUMMARY

Computational biophysics is focused on obtaining results relevant to biophysical problems, and on developing new theoretical methods, having a strong interplay with the traditional theoretical physics and chemistry. Developing a new model often enables new problems to be studied, and results from calculations reveal limitations in underlying theory. Nowadays, depending on the accuracy required and the origin of the problem, one can obtain a precise atomization energy for methane molecule or perform long timescale molecular dynamics simulations for systems containing millions of atoms. Undoubtedly, computer simulation methods have predictive power and nowadays play one of the key roles in understanding biological processes. Although many of these processes involve long time and length scales and therefore can be handled only by simulations at the empirical potential level, first principles simulations are very much needed in this area. Despite its obvious advantages, it is evident that a price has to be paid for putting molecular dynamics on quantum mechanical level: the simulation times and system sizes that are accessible are much smaller than what is affordable via molecular mechanics methods. However the fundamental advantages of the first-principle computational methods are the accuracy and the truly predictive power. Empirical methods allow to see only that, what was foreseen, while *ab initio* will show that, what was not foreseen before starting the simulations.

In this thesis density functional theory was applied to various systems taken from the field of biological structures. The overall goal of the work was to answer open questions not answered or emerging form experimental data and to provide information for a better understanding of the selected systems. In order to apply *ab initio* calculations to large systems, such as biological systems it is necessary to strike a balance between accuracy and computational efficiency. The work presented in this thesis is an example of how the quantum mechanical techniques can successfully be applied to biologically relevant problems in rather large and complex systems.

Abstracts of the publications

Publication I K. J. Jalkanen, V. W. Jurgensen and I. Degtyarenko, *Linear response properties required to simulate vibrational spectra of biomolecules in various media: (R)-Phenyloxirane (a comparative theoretical and spectroscopic vibrational study)*, Advances in Quantum Chemistry **50**, 91-124 (2005).

We have simulated the vibrational absorption (VA), also called IR, vibrational circular dichroism (VCD), Raman scattering and Raman optical activity (ROA) intensities utilizing the density functional theory (DFT) B3LYP hybrid exchange correlation functional and other exchange-correlation functionals (PBE, PW91, PBE1) with the 6-31G(d,p), 6-31++G(d,p), cc-pVDZ, aug-cc-pVDZ, cc-pVTZ and augmented correlation consistent polarized valence triple zeta (augcc-pVTZ) basis sets. Previously authors have focused on either the VA and VCD spectra or the Raman and ROA spectra of molecules, since the experimental and theoretical instruments and methods for calculating these quantities are quite distinct. Here we show that the combined analysis gives more information, especially with respect to the electric dipole, magnetic dipole, electric dipole-electric dipole polarizability, electric dipole-electric quadrupole polarizability and electric dipole-magnetic dipole polarizability changes during the various induced transitions. The coupling of vibrational and electronic excitations may be used to aid in understanding the photo induced chemical reactivity observed in many systems. This work is a continuation of our goal to interpret the results of experimental studies on the basis of theoretical results, which can help to understand the structure and function of proteins, other biomolecules and ligands in their native environments. As the physical tools used to observe and study biological processes have evolved, so have the theoretical methods and models to interpret, understand and completely utilize the results of these new measurements. The work on developing methods for modeling amino acids, peptides, proteins and ligands in both the non aqueous (lipid) and aqueous environments has involved, of course, many groups. A review of our contributions to this field has recently been presented. In addition to interpreting existing and new experimental results, we will discuss structural, energetic, conformational, and vibrational studies on a variety of systems that have been used to test and validate levels of theory, and in addition to suggest modications to existing levels of theory, which can make them even more useful than they currently are.

Publication II I. Degtyarenko, R. M. Nieminen and C. Rovira, *Structure and dynamics of dioxygen bound to cobalt and iron heme*, Biophysical Journal **91(6)**, 2024-2034 (2006).

In this study we use *ab initio* molecular dynamics simulations to analyze the structure and dynamics of the oxygen ligand in models of the oxymyoglobin active site and its cobalt-substituted analog. Our calculations are performed for iron-porphyrin

and cobalt-porphyrin complexes with imidazole and oxygen as axial ligands, and we investigate the effect of the distal histidine in the structure and dynamics of the metal-oxygen unit (MeO_2, Me = Fe, Co). We find that the interaction between the distal histidine and the oxygen ligand is stronger for the cobalt complex than for the iron one, consistent with the superoxide ion character of the bound O_2. The dynamics of the O_2 ligand can be described as oscillations of the O-O axis projection on the porphyrin plane within a porphyrin quadrant combined with frequent jumps from one quadrant to another. However, the ligand motion is significantly faster for CoO_2 compared to FeO_2. As a result, the iron complex shows localized ligand sites, whereas for cobalt several configurations are possible. This gives support to the highly dynamic motion of the oxygen ligand found in several experiments on cobalt oxymyoglobin and model complexes and underlines the higher mobility of the CoO_2 fragment compared to FeO_2.

Publication III I. M. Degtyarenko, K. J. Jalkanen, A. A. Gurtovenko and R. M. Nieminen, *L-alanine in a droplet of water: a density-functional molecular dynamics study*, Journal of Physical Chemistry B **111(16)**, 4227-4234 (2007).

We report the results of a Born-Oppenheimer molecular dynamics study on an L-alanine amino acid in neutral aqueous solution. The whole system, the L-alanine zwitterion and 50 water molecules, was treated quantum mechanically. We found that the hydrophobic side chain (R = CH_3) defines the trajectory path of the molecule. Initially fully hydrated in an isolated droplet of water, the amino acid moves to the droplet's surface, exposing its hydrophobic methyl group and α-hydrogen out of the water. The structure of an L-alanine with the methyl group exposed to the water surface was found to be energetically favorable compared to a fully hydrated molecule. The dynamic behavior of the system suggests that the first hydration shell of the amino acid is localized around carboxylate (CO_2^-) and ammonium (NH_3^+) functional groups; it is highly ordered and quite rigid. In contrast, the hydration shell around the side chain is much less structured, suggesting a modest influence of the methyl group on the structure of water. The number of water molecules in the first hydration shell of an alanine molecule is constantly changing; the average number was found to equal 7. The molecular dynamics results show that L-alanine in water does not have a preferred conformation, as all three of the molecule's functional sites (i.e., CH_3, NH_3^+, CO_2^-) perform rotational movements around the C_α-site bond.

Publication IV I. M. Degtyarenko, K. J. Jalkanen, A. A. Gurtovenko and R. M. Nieminen, *The aqueous and crystalline forms of L-alanine zwitterion*, Journal of Computational and Theoretical Nanoscience (in press).

The structural properties of L-alanine amino acid in aqueous solution and in crystalline phase have been studied by means of density-functional electronic-structure and molecular dynamics simulations. The solvated zwitterionic structure of L-alanine ($^+NH_3$-C_2H_4-COO^-) was systematically compared to the structure of its zwitterionic crystalline analogue acquired from both computer simulations and experiments. It turns

out that the structural properties of an alanine molecule in aqueous solution can differ significantly from those in crystalline phase, these differences being mainly attributed to hydrogen bonding interactions. In particular, we found that the largest difference between the two alanine forms can be seen for the orientation and bond lengths of the carboxylate (COO^-) group: in aqueous solution the C-O bond lengths appear to strongly correlate with the number of water molecules which form hydrogen bonds with the COO^- group. Furthermore, the hydrogen bond lengths are shorter and the hydrogen bond angles are larger for L-alanine in water as compared to crystal. Overall, our findings strongly suggest that the generally accepted approach of extending the structural information acquired from crystallographic data to a L-alanine molecule in aqueous solution should be used with caution.

Publication V K. J. Jalkanen, I. M. Degtyarenko, R. M. Nieminen, X. Cao, L. A. Nafie, F. Zhu and L. D. Barron, *Role of hydration in determining the structure and vibrational spectra of L-alanine and N-acetyl L-alanine N'methylamide in aqueous solution: a combined theoretical and experimental approach*, Theoretical Chemistry Accounts (in press).

We here present a combined VA, VCD, Raman and ROA vibrational study of phenyloxirane. In this work we have utilized recent density functional theory Born-Oppenheimer molecular dynamics simulations to determine the first principles locations of the water molecules in the first solvation shell which are responsible for stabilizing the zwitterionic structure. Previous works have used chemical intuition or classical molecular dynamics simulations to position the water molecules. In addition, a complete shell of water molecules was not previously used, only the water molecules which were thought to be strongly interacting (H-bonded) with the zwitterionic species. In a previous work by Tajkhorshid *et al.*[118] the L-alanine zwitterion was stabilized by 4 water molecules, and a subsequent work by Frimand *et al.*[119] the number was increased to 9 water molecules. Here we found that 20 water molecules are necessary to fully encapsulate the zwitterionic species when the molecule is embedded within a droplet of water, while 11 water molecules are necessary to encapsulate the polar region with the methyl group exposed to the surface, where it migrates during the MD simulation. Here we present our vibrational absorption, vibrational circular dichroism and Raman and Raman optical activity simulations, which we compare to the previous simulations and experimental results. In addition, we report new VA, VCD, Raman and ROA measurements for L-alanine in aqueous solution with the latest commercially available FTIR VA/VCD instrument (Biotools) and Raman/ROA instrument (Biotools). The signal to noise of the spectra of L-alanine measured with these new instruments is significantly better than the previously reported spectra. Finally we reinvestigate the causes for the stability of the P_π structure of the alanine dipeptide, also called N-acetyl-L-alanine N'-methylamide, in aqueous solution. Previously we utilized the B3LYP/6-31G* + Onsager continuum level of theory to

investigate the stability of the NALANMA4WC (Han *et al.*)[130]. Here we use the B3PW91 and B3LYP hybrid exchange correlation functionals, the aug-cc-pVDZ basis set and the PCM and CPCM (COSMO) continuum solvent models, in addition to the Onsager and no continuum solvent model. Here by the comparison of the VA, VCD, Raman and ROA spectra we can confirm the stability of the NALANMA4WC due to the strong hydrogen bonding between the four water molecules and the peptide polar groups. Hence we advocate the use of explicit water molecules and continuum solvent treatment for all future spectral simulations of amino acids, peptides and proteins in aqueous solution, as even the structure (conformer) present can not be found without this level of theory.

[130] W. Han, K. J. Jalkanen, M. Elstner and S. Suhai, *Theoretical study of aqueous N-acetyl-L-alanine N'-methylamide: structures and Raman, VCD and ROA spectra*, J. Phys. Chem. B **102**, 2587-2602 (1998).

List of abbreviations

CC or CCSD (T)	Coupled Clusters method
BLYP	Becke and Lee-Yang-Parr exchange-correlation functional
BSSE	Basis Set Superposition Error
DFT	Density Functional Theory
DZP	Double- plus polarization basis set
GGA	Generalized Gradient Approximation
KS	Kohn-Sham
LA	L-alanine
LDA or LSD	Local (Spin) Density Approximation
MD	Molecular Dynamics
MM	Molecular Mechanics
MP2	Møller-Plesset perturbation theory
ROA	Raman Optical Activity
PBC	Periodic Boundary Conditions
PBE	Perdew-Burke-Enrzerhof exchange-correlation functional
PES	Potential Energy Surface
PS	Pseudopotential
QM	Quantum Mechanics
O(*N*)	Order-N
SCRF	Self-consistent Reaction Field
VA	Vibrational Absorption
VCD	Vibrational Circular Dichroism
xc	exchange-correlation

ISBN 978-951-22-8805-2
ISSN 1455-1802