

AB INITIO COMPUTATIONAL APPLICATIONS TO
COMPLEX BIOMOLECULAR SYSTEMS

A DISSERTATION IN
Physics
and
Math & Statistics

Presented to the Faculty of the University
of Missouri-Kansas City in partial fulfillment of
the requirements for the degree

DOCTOR OF PHILOSOPHY

by
LEI LIANG

B.Sc., University of Science and Technology of China, 2006
M.Sc., University of Missouri-Kansas City, 2009

Kansas City, Missouri
2011

© 2011

LEI LIANG

ALL RIGHTS RESERVED

AB INITIO COMPUTATIONAL APPLICATIONS TO
COMPLEX BIOMOLECULAR SYSTEMS

Lei Liang, Candidate for the Doctor of Philosophy Degree

University of Missouri-Kansas City, 2011

ABSTRACT

A series of biomaterial related systems — including water and DNA molecules — have been studied using *ab initio* (first-principles) methods. By investigating the properties of water as the preliminary step, the hydrogen bond (HB) interactions, which play important roles in biomolecules, were better understood from the quantum mechanical viewpoint. The calculated *K*-edge x-ray absorption near edge structure (XANES) spectra of all 340 oxygen atoms in the model have been accumulated to reproduce the experimental one. The spectra were shown to be very sensitive to the HB configurations of O atoms, which could be used to elucidate the subtle structural variations in complex biomolecules. The simulation of single-molecule DNA overstretching experiments under torsionally constrained condition has been carried out afterwards. The initial DNA models were stretched stepwisely and eventually gained an extension of 1.5-fold (150% × the original length). The variation of total energy, atomic configuration, and the electronic structure during this process were analyzed in details. At the extension of ~1.3-fold, the ring opening reactions occurred in the backbones. The backbone nicks appeared at elongations of ~1.40-fold. The whole process was accompanied by HB breaking and charge transfers. We have proposed an overstretched

structure named O-DNA (Opened-DNA) to clarify the confusion in understanding the behavior of DNA under high force load. With more experiences gained, a comprehensive methodology revealing the underlying principles of bioprocesses from the quantum mechanical viewpoint eventually come up. For the purpose of better computational accuracy, the scheme of implementing the generalized gradient approximation (GGA) exchange-correlation functionals into the Orthogonalized Linear Combination of Atomic Orbitals (OLCAO) program suite has been discussed, and the computational efficiency has been analyzed correspondingly. Moreover, the parallel strategy for performing evaluation on a regular mesh and relevant updates to the file system were also presented. All the fundamental works above paved the way for more sophisticated study on wet DNA model and interfaces between biomolecules and bioceramic materials in the future.

APPROVAL PAGE

The faculty listed below, appointed by the Dean of the College of Arts and Sciences have examined a dissertation titled "*ab initio* Computational Applications to Complex Biomolecular Systems," presented by Lei Liang, candidate for the Doctor of Philosophy degree, and certify that in their opinion it is worthy of acceptance.

Supervisory Committee

Wai-Yim Ching, Ph.D., Committee Chair
Department of Physics

Jerzy M. Wrobel, Ph.D.
Department of Physics

Daming Zhu, Ph.D.
Department of Physics

Noah Rhee, Ph.D.
Department of Math and Statistics

Yong Zeng, Ph.D.
Department of Math and Statistics

TABLE OF CONTENTS

ABSTRACT	II
LIST OF ILLUSTRATIONS	VIII
LIST OF TABLES	XII
ACKNOWLEDGEMENTS	XIII
CHAPTER	
1 INTRODUCTION	1
1.1 Context	1
1.2 Outline and Motivation	4
2 THEORETICAL METHODS	6
2.1 Density Functional Theory	6
2.1.1 Introduction	6
2.1.2 Kohn-Sham Equation	8
2.1.3 Exchange-correlation Functional	10
2.2 The VASP Package and Calculation	13
2.2.1 Introduction	13
2.2.2 Geometry Optimization	15
2.2.3 Parallel VASP Running Experiences	16
2.3 The SIESTA Method and Calculation	17
2.4 The OLCAO Method and Calculation	19
2.4.1 Overview	19
2.4.2 Technical Aspects	26

2.4.3	Scheme for GGA Implementation	33
2.4.4	Computational Capabilities.....	38
2.4.5	Summary.....	41
3	CASE STUDIES.....	42
3.1	Network Structure of Bulk Water	42
3.1.1	Introduction.....	42
3.1.2	Model and Computational Methods.....	44
3.1.3	Results and Analysis.....	46
3.1.4	Discussions	63
3.1.5	Conclusions.....	70
3.2	Mechanism of DNA Overstretching Deformation.....	71
3.2.1	Introduction.....	71
3.2.2	Model and Computational Methods.....	82
3.2.3	Results and Analysis.....	84
3.2.4	Discussions	101
3.2.5	Conclusions.....	104
3.3	Summary	104
4	FUTURE WORK.....	106
4.1	Overview	106
4.2	Prospects of Quantum Mechanical Modeling in Biology	108
4.3	Final Comments	112
APPENDIX		

A. VASP RELAXATION RUNNING TIME COMPARISON	113
B. TYPICAL INITIAL DATASETS OF OXYGEN ATOM POTENTIAL	115
C. MINIMAL BASIS OF OXYGEN ATOMIC ORBITALS	116
D. ABBREVIATIONS	117
REFERENCES	119
VITA	130

LIST OF ILLUSTRATIONS

Figure	Page
1. SCF setup phase HDF5 File from ref. [70].....	29
2. Schematic display of the OLCAO XC mesh: Wigner cell of each atom (yellow and green boxes); Sampling rays: green (yellow) vectors starting from the atom ending at the boundary of the Wigner cells; Sampling points: green (yellow) dots.....	31
3. Schematic display of the density gradient interpolation on (a) uniform mesh (upper panel) and (b) OLCAO mesh (lower panel).....	35
4. Potential alternative file storage system of XC group under GGA.	37
5. Water phase diagram [87].....	43
6. The periodic water model with 340 H ₂ O molecules.....	45
7. Distributions of (a) covalent bond length r_{O-H} and (b) bond angle θ ($\angle HOH$).....	46
8. Radial pair distribution functions for the water model (red solid): (a) O-H; (b) H-H; and (c) O-O. Experimental data (cyan dash) is from ref. [95].	48
9. Calculated DOS (black solid) and PDOS (blue dash dot: O; red dash: H) in water.....	49
10. Distribution of all O and H bonds in the water model: region I: nonbonded; region II: H-bonded; region III: covalently bonded.	51
11. Schematic illustration of donor and acceptor H bonds between three water molecules H ₂ O, H' ₂ O', and H'' ₂ O".	53
12. Distributions of the H bonds for the three groups: two-HB (blue); three-HB (pink); and five-HB (yellow).	54
13. Scattered plot of $r_{O-O'}$ vs α ($\angle O'-OH$) for all H Bonded O...H pairs. Solid Line: HB definition curve from ref. [89].	56
14. Scattered plot of the average O-H covalent bond length r_{O-H} vs covalent bond angle θ . .	57
15. (a) O K-edge XANES spectra: experimental (i) XRS at 25°C [89]; (ii) XAS at ambient conditions [110]; (iii) averaged spectra over 340 molecules from present calculation. (b) Calculated total spectra for the four categorized groups with two, three, four and five HBs.....	60

16. All eleven individual O K-edge spectra in the two-HB coordinated group from calculation.	61
17. A typical representative spectrum form each group with two to five HBs.....	62
18. Comparison of the distributions of the covalent bond length r_{OH} of Model 1-5.....	66
19. Comparison of the distributions of the covalent bond angle θ of Model 1-5.	67
20. Total averaged O k-edge XANES spectra of Model-1 (purple), Model-2 (red), and experiment (blue) [110].....	68
21. Comparison of the distributions of the O..H HB bond order of Model 1-5.....	69
22. Four nucleotides and bonding formation in double-helix structure.....	73
23. Labeled sections: (1) the entropic elasticity plateau, (2) elastic extension of B-DNA, (3) the B-S transition plateau, (4) elastic extension of S-DNA, (5) the DNA melting plateau, and (6) elastic extension of single-stranded (melted) DNA (ssDNA) (from ref. [112, 114], inset from ref. [113] Figure 1(c))......	74
24.(a) 65 pN overstretching plateau under torsionally unconstrained condition, from ref. [113] figure 1(a); (b) 115 pN overstretching plateau under torsionally constrained condition, from ref. [113] Figure 1(b); (c) overstretching curve fitted with WLC and FJC model: green dot clearly indicates the final state to be ssDNA, from ref. [125] Figure 1(a).	75
25. Overstretching at ~ 65 pN in a topologically closed but rotationally free assay using a biotin embedded in a 5-nucleotide ssDNA loop, from ref. [124] Figure 1(d).	79
26. Potential explanations to retraction-hysteresis: (a) Unpeeled single strand forms self-complementary hairpins upon tension release which resist the recovery to original B-DNA (from ref. [114] Figure 3(a)); (b) Large molten bubble is entropy favored, which may cause entropy-driven hysteresis (from ref. [120] Figure 2).	79
27. (a, b, c) from ref. [117] Figure 2. Increased stability of basepairs decreases overstretching hysteresis. Open squares show the increasing-force scan and solid stars show the decreasing-force scan. (a) Effects of NaCl concentration for stretching a λ -DNA: 150 mM (black), 300 mM (green, same DNA as black), and 1 M (red, a different DNA from black and green); (b) Effects of temperature for stretching λ -DNA: 24°C (black), 19°C (red); (c) Effects of sequence composition: GC-rich DNA (red) shows less hysteresis than AT-rich DNA (black). Extensions are scaled with their respective contour lengths. (d) zip-DNA does not require accurate base pairing but π - π electronic coupling among nearest-neighbor nucleobases (from ref. [112]).	80

28. (a) Tentative DNA phase diagram under tension and torque, from ref. [129] Figure 4. The P-DNA model from ref. [131] Figure 5; Backbones: purple, guanine: blue, cytosine: yellow, anionic oxygens of the phosphate groups: red. (b) Predicted critical lines for denaturation at different temperatures, from ref. [133] Figure 2. The region enclosed by each line corresponds to stable B-DNA at that temperature. Red points: experimental data from single molecule manipulation [130], blue square: the well-known overstretching transition ($\Gamma = 0$, $f = 60$ pN). $T_D = 77$ °C, the denaturation temperature, is critical at zero external load: but even at temperatures of T_D or higher (inset), B-DNA can be stable in an interval of applied positive torque. Solid (dotted) straight line indicates $\Gamma_m = a\tilde{v}\tilde{\Omega}$ (its approximation $\Gamma_m = 11 + 0.092f$), the middle point between critical torques, which also corresponds to the highest critical temperature at given tension. 81
29. (a,b) Initial B-DNA model $(dA-dT)_{10}$ (AT10). (ψ , ω) are the tilt/roll angles, and $d = 3.38$ Å is the stacking height. r is the helix radius, $\theta = 36^\circ$ is the twist angle. 20 Na^+ counter ions were added (purple atoms). (c) Detailed structure. 83
30. Electron total energy of each stretched model from VASP relaxation. The whole overstretching process is divided at 0.80 into stage (I) and (II). 85
31. Average distance between C' and C'' in the direction of (a) $C' \rightarrow \text{PO}_4 \rightarrow \text{CH}_2 \rightarrow C''$ (denoted as D_1) and $C'' \rightarrow C'$ (denoted as D_2). 87
32. Displays of the elongated model at relative extension of 0, 0.80, 1.32, and 1.34. At 1.32, $C'-C''$ single bonds in the sugar group of strand T break resulting in the ring opening reactions (RORs). At 1.40, C-O ester bonds in strand T break resulting nicks in the backbone. 88
33. (a) Schematic display of the dihedral angle α between the nucleobases. $\uparrow A$ ($\uparrow T$) is the normal vector of the ring plane of adenine (thymine). $\uparrow T'$ is the translation of $\uparrow T$. (b) Variations of dihedral angle α between the base planes at each 10% elongation. 89
34. Variation of hydrogen bond (HB) bond length of O..H (blue diamond) and N..H (orange dot) between the basepairs. 91
35. Variation of HB bond order of O..H (blue diamond) and N..H (orange dot) pairs. The average values are shown as black triangles. 93
36. Charge transfer properties (ΔQ^*) categorized according to different strands (left panel: strand A; right panel: strand T) and functional groups (nucleobases: green; yellow: $\text{PO}_4 + \text{Na}^+$; magenta: sugar) at relative extension of 0, 0.30, 0.50, 0.70, 0.80, 0.90, 1.10, 1.30, 1.32, 1.34, 1.40 and 1.42 (the bars from left to right). The shadow represent the average ΔQ^* of one nucleotide. 95

37. Calculated TDOS of the initial structure ($L/L_0 = 0$).....	98
38. Calculated TDOS and decomposed PDOS of different functional groups at $L/L_0 =$ (a) 0 and (b)1.30.	98
39. Calculated TDOS and decomposed PDOS of different functional groups at $L/L_0 =$ (a) 1.32, (b)1.34, (c) 1.40, and (d) 1.42.	99
40. The breaking of C'-C'' single bond results in (a) two C· radicals (· represents the unpaired electron) or (b) C+(-) ions.	100
41. Possible cases of electron redistribution along the backbone after RORs: (a) all C·, (b) alternative C-C+C-C+..., (c) mixed C-C+C·C·C+C-..., and (d) clustered C+C-C-C+C+C-C-C+.....	100

LIST OF TABLES

Table	Page
1. Portion of four groups with different HB numbers.....	53
2. Water models constructed and studied with different methods.....	65
3. Information of B-DNA model AT10.....	85

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude and deep appreciation to my advisor, Dr. Wai-Yim Ching, for his excellent guidance, persistent encouragement, and continuous support throughout this work.

I am grateful to Dr. Jerzy M. Wrobel and Dr. Daming Zhu of the physics department, and Dr. Noah Rhee and Dr. Yong Zeng of the Math and Statistics department for their valuable discussions and insightful comments for the dissertation, and kindly serving in my supervisory committee.

Special thanks to our former group members, Prof. Paul Rulis and Prof. Lizhi Ouyang, for their advices and assistance not only on the research but also in the daily life.

Also, great thanks to the other professors, staff members, and fellow students in the physics department for their help and friendship.

My family has always been the foundation of my life. Without their support and love I never would have able to finish.

Finally, I would like to thank the University of Missouri–Kansas City, the Department of Physics at UMKC, and the US Department of Energy (DOE) for their monetary and training support.

To my dear parents and friends

CHAPTER 1

INTRODUCTION

1.1 Context

Over the recent few decades, a young and rapidly evolving field of research named *nanobiotechnology* has grown out of the embryonic state. At the crossroad of biotechnology and nanoscience, nanobiotechnology concerns the utilization of biological systems optimized through the evolution, such as cells, cellular components, nucleic acids, and proteins, to fabricate functional nanostructured and mesoscopic architectures comprised of organic and inorganic materials. Some major examples include specifically designed structures formed using the self-assembling properties of nucleic acids, microcontact printing of proteins, biomolecular motors, and nanoparticles for drug delivery, etc [1, 2]. The new phenomena exhibited and their fascinating nature behind have triggered deep and wide investigations that involve both experimentalists and theorists.

Theoretical modeling could be very promising in this area of study for two main reasons. First, with the state-of-the-art modeling packages and rapid growth in the computing power, the design and simulation of functional molecular machines are feasible with current computational facilities. For example, present nanobiotechnology uses the most rudimentary molecular structures (DNA, etc.) as building blocks to achieve economical fabrication. By using the Nucleic Acid Builder (NAB) [3], constructions of DNA double helix with different conformations are simplified into the inputs of the sequence order and a few structural parameters. Moreover, the development of parallel programming libraries [4, 5] and the

creation of high-performance computing tools [6] enable new implementations of rigorous physical theories such that systems with greater complexities can be addressed within a favorable time range. Second, computational simulations of nano-scale systems have a lot of advantages compared to real experiments. Such modeling is a relatively economical way to explore the truly wide range of possible molecular machines, allowing the rapid elimination of obvious dead ends or the retention, as well as intensive analysis of more promising designs [7]. It is clear that the right computational support will substantially reduce the developing time. Theoretical simulations also allow us to independently control the intrinsically entangled experimental parameters, such as size, shape, chemical composition and degree of agglomeration, etc., in order to identify the underlying mechanisms responsible for instabilities of nanostructures. Also possible is to study the properties in highly non-equilibrium environments, such as extreme temperatures and pressures, or strong electric or magnetic fields, which are very difficult to probe experimentally [8].

Despite the attractive features mentioned above, theoretical modeling is still some distance away from being a ‘black box’ with little discrimination regarding different systems. Present *ab initio* or first principles quantum mechanical methods [9] can study systems with up to a few hundreds of atoms, and the semi-empirical methods [10] can treat as many as several thousands of atoms. Classical mechanical methods can address problems on a scale ranging from the microscopic size to the mesoscopic size but with limited accuracy [11]. The confluence of all those computational progresses is the multiscale modeling [12] which has already shown great impact on various disciplines in both academic researches and industrial

sectors. In particular to nanobiotechnology, an emerging field where both experimental and theoretical exploring tools are far from mature, the quantum mechanical methods based on density functional theory (DFT) have some intrinsic advantages, and are able to contribute in this area. DFT is an exact theory about the ground state of electrons which greatly reduces the number of degrees of freedom in solving the complicated Schrödinger equation compared to the conventional wave function (WF) approaches [13]. It is extremely powerful in obtaining the accurate atomic structure and investigating the electronic interactions. It also offers the greatest sensitivity to chemical reactions that are important in the surface effects of nanohazards [14, 15] and some bio-nano interactions such as the translocation of nanoparticles through biomembranes [16]. However, the deficiencies of present DFT methods also should be addressed. Most DFT calculations are performed at zero temperature with little consideration of the real experimental conditions which may involve solvent or inter-particle interactions rather than the ideal case of ultra-high vacuum. Fortunately, these problems could be solved within current knowledge scope using the *ab initio* molecular dynamics (MD) methods and properly tuned models [8].

The *ab initio* DFT simulations of complex biomolecules and that of regular crystalline systems have certain differences. First, the periodicity of biomolecular models is not so self-evident that special attentions need to be paid to avoid the problems associated with boundary conditions. Second, biomolecular models usually involve a large number of atoms in one unit cell with little symmetries in the structure. Therefore it is often much more computationally demanding hence the conflicts between the desired accuracy and the

computing resources need to be properly balanced. Moreover, the short-range hydrogen bonding and the long-range Van der Waals force play important roles in determining the biomolecular conformation. The way to analyze and interpret those soft and sensitive interactions has not yet been well established. Generally speaking, to carry out a meaningful *ab initio* calculation of biomolecules is not a simple task. It requires a well-designed model, a well-organized scheme together with the techniques to extract useful but inconspicuous information covered by the seeming mess.

1.2 Outline and Motivation

This work will focus on the *ab initio* DFT studies of several important systems closely related to many biological and bioengineering processes. As the first step, water, the most common and intriguing substances on Earth yet the most important component in living organisms, will be studied to clarify the ambiguity in the hydrogen bond (HB) definition using quantum mechanical language. This will prepare us with the fundamental understanding of the intermolecular interaction for better analysis of more complex biomolecular systems. The next step will be the simulation comparable to the single-molecule DNA overstretching experiments [17, 18]. The purpose of this part is to fully resolve the mechanism of structural deformation in the overstretching process by using accurate atomic-scale information such as atomistic movements and bonding strength variations obtained from the calculation.

Beyond the curiosity to understand the systems mentioned above, testing and improving the performance of various first principles packages (the Vienna Ab initio Simulation Package (VASP) [19], the Spanish Initiative for Electronic Simulations with Thousands of Atoms (SIESTA) method [20], and the Orthogonalized Linear Combination of Atomic Orbitals (OLCAO) program suite [21]) in dealing with the complex biomolecular systems are also important purposes of this work. The combination of those packages has been successfully applied to a variety of crystalline and non-crystalline materials, including complex microstructures such as interfaces and surfaces [22]. In the second chapter, the well-known DFT theory and some technical aspects of the packages used in this work will be introduced. Special discussions will be put on the potential new features and the strategy of their implementations for the present OLCAO method to achieve better accuracy and efficiency.

CHAPTER 2

THEORETICAL METHODS

2.1 Density Functional Theory

2.1.1 Introduction

Ever since the emergence of quantum mechanics, it has been well recognized that given suitable computational tools, all properties of materials can be addressed by viewing the systems as interacting electrons and atomic nuclei and solving the Schrödinger equation. However, the equation with many particle interactions is too complicated to allow a direct solution. After many decades of pursuit, the electron density approach successfully overcame the unfavorable scalability problem associated with the previous WF approach [23], thus the dream to deal with large and complex systems became more realistic.

The idea of using electron density as the basic variable can be traced back to the Thomas-Fermi theory about electron gas [24, 25] in the 1920s, but it is not until the two Hohenberg-Kohn (HK) theorems [26] proposed in 1964 that the density functional theory was put on a firm theoretical footing. The two theorems state:

- (1) The external potential $v_{ext}(\vec{r})$, and hence the total energy, is a unique functional of the electron density $\rho(\vec{r})$,
- (2) The density that minimizes the total energy is the exact ground state density.

where \vec{r} represents the positional coordinates.

Thereafter, the Kohn-Sham (KS) method [27] was ingeniously proposed and eventually turned DFT a practical tool for rigorous calculation. By introducing the non-interacting reference system and conglomerating all the many-body related energy into the exchange-correlation (XC) energy, the resulting single-electron KS equation becomes the heart of 90% of current quantum mechanical calculations in molecules and solids. To solve the KS equation, explicit expression of the XC functional is necessary. In the following years, the development of XC potential term has been continuously pushing forward. Nowadays the widely used approximations include the local (spin) density approximation (L(S)DA) [27] and the generalized gradient approximation (GGA) [28]. L(S)DA assumes that the electron distribution within a system is uniform and has been demonstrated to be very successful for homogeneous systems. GGA is a step forward in the sense that it is a functional of both density and its gradient. It is more accurate in dealing with inhomogeneous systems such as heavy fermion system where the electron correlation is very strong. The early works on GGA dated back to 1983 [29], when Langreth and Mehl (LM) proposed their GGA based on the wave-vector analysis of E_{xc} . After that, GGA has been continuously improved. The most popular functionals include the works of Perdew-Wang (1986, PW86) [30], Beche-Perdew (1988, BP) [31] and Lee-Yang-Parr (1988, LYP) [32]. All of them are parameter-dependent and are optimized only for some particular types of systems. The appearance of Perdew-Wang (PW91) GGA [33] in 1991 ended this situation. It was constructed purely under *ab initio* conditions and fulfilled almost all scaling relations except high density limit of uniform scaling. The much more recent improvement over PW91

include an accurate description of the linear response of the uniform electron gas, correct behavior under uniform scaling, and a smoother potential were proposed by Perdew-Burke-Ernzerhof in 1996 (PBE96) [34].

In this chapter, the *ab initio* computational packages used in this work (VASP and OLCAO) will be introduced with special discussions on the practical scheme of incorporating GGA XC functionals into the OLCAO method. The updated data storage structure is also considered.

2.1.2 Kohn-Sham Equation

By conglomerating all the many-body related energy into the XC term, the total energy functional can be expressed as

$$\begin{aligned}
 E[\rho] &= T_s[\rho] + J[\rho] + E_{XC}[\rho] + \int \rho v_{ext} d\vec{r} \\
 &= \sum_i^N \sum_s \int \psi_i^* (-\frac{1}{2} \nabla^2) \psi_i d\vec{r} + J[\rho] + E_{XC}[\rho] + \int \rho v_{ext} d\vec{r}
 \end{aligned}
 \tag{2.1},$$

where the electron density is:

$$\rho = \sum_i^N \sum_s |\psi_i(\vec{r}, s)|^2
 \tag{2.2},$$

and $T_s[\rho]$ is the kinetic energy of the non-interacting reference system, $J[\rho]$ is the classical Coulomb potential, E_{XC} represents the XC potential and the last term describes the external potential.

If the N orbitals are continuous and square integrable, then the density defined in (2.2) covers the N-representable ensemble. So the variational search for the minimum $E[\rho]$

can be equivalent in the orbitals space $\{\psi_i\}$. In order to have the electron density defined in (2.2) to be valid, the N orbitals must actually be orthonormal,

$$\int \psi_i^*(x) \psi_j(x) dx = \delta_{ij} \quad (2.3).$$

In the variational search formula, the ground-state total energy can be written as

$$E_g = \underset{\int \psi_i^*(x) \psi_j(x) dx = \delta_{ij}}{\text{Min}} \left\{ \sum_i^N \sum_s^N \int \psi_i^* (-\frac{1}{2} \nabla^2) \psi_i d\vec{r} + J[\rho] + E_{XC}[\rho] + \int \rho v_{ext} d\vec{r} \right\} \quad (2.4).$$

With the Lagrange Multipliers method, the condition for minimum $E[\rho]$ is

$$\delta \left\{ E[\rho] - \sum_i^N \sum_j^N \varepsilon_{ij} \int \psi_i^*(x) \psi_j(x) dx \right\} = 0 \quad (2.5),$$

which leads to the equations

$$\widehat{H} \psi_i = [-\frac{1}{2} \nabla^2 + v_{eff}] \psi_i = \sum_j^N \varepsilon_{ij} \psi_j \quad (2.6),$$

where the effective potential v_{eff} is defined by

$$\begin{aligned} v_{eff} &= v_{ext}(\vec{r}) + \frac{\delta J[\rho]}{\delta \rho} + \frac{\delta E_{XC}[\rho]}{\delta \rho} \\ &= v_{ext}(\vec{r}) + \int \frac{\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r}' + v_{XC}(\vec{r}) \end{aligned} \quad (2.7).$$

Since \widehat{H} is a Hermitian operator, ε_{ij} is a Hermitian matrix which can be diagonalized by a unitary transformation of the orbitals. Such a transformation leaves the density and the Hamiltonian invariant. Thus, the famous **Kohn-Sham orbital equations** [27] are obtained in the canonical form

$$\widehat{H}\psi_i = [-\frac{1}{2}\nabla^2 + v_{eff}] \psi_i = \varepsilon_i \psi_i \quad (2.8).$$

The total energy can be determined from the resultant density calculated via (2.4)

$$E = \sum_i^N \varepsilon_i - J[\rho] + E_{XC}[\rho] - \int v_{XC} \rho d\vec{r} \quad (2.9).$$

Here

$$\sum_i^N \varepsilon_i = \sum_i^N \langle \psi_i | -\frac{1}{2}\nabla^2 + v_{eff} | \psi_i \rangle = T_s[\rho] + \int v_{ext} \rho d\vec{r} + 2J[\rho] + \int v_{XC} \rho d\vec{r} \quad (2.10).$$

So far, the Kohn-Sham equation turns out to be exact for the ground state without any approximation. To solve the KS equation, the explicit expression of the many-body and non-classical XC functional term $E_{XC}[\rho]$ must be given.

2.1.3 Exchange-correlation Functional

The LDA and GGA approximations to $E_{XC}[\rho]$ have the general forms

$$E_{XC}^{LDA}[\rho] = \int f_{LDA}(\rho(\vec{r})) d\vec{r} \quad (2.11),$$

$$E_{XC}^{LDA}[\rho] = \int f_{GGA}(\rho(\vec{r}), \vec{g}(\vec{r})) d\vec{r} \quad (2.12),$$

where $f = \rho \varepsilon_{XC}$ is the local XC energy density, and ε_{XC} is the XC energy per electron. The notations $\vec{g}(\vec{r}) \equiv \nabla \rho(\vec{r})$, $\vec{g}_\alpha(\vec{r}) \equiv \nabla_\alpha \rho(\vec{r})$, and $g(\vec{r}) \equiv |\nabla \rho(\vec{r})|$ are used where α labels the three Cartesian coordinates. In this work, we restrict the discussion to the spin non-polarized case for simplicity. The inclusion of spin polarized electron density is trivial. It is easy to see that in the limit of slow varying density, we should have $f_{GGA}(\rho(\vec{r}), 0) = f_{LDA}(\rho(\vec{r}))$.

There are two different strategies for determining the function f_{GGA} , either adjust f such that it satisfies all (or most) known properties of the XC hole and energy, or fit f to a large dataset of exactly known binding energies of atoms and molecules. The former was preferred, when constructing GGA functionals, one usually tries to incorporate a number of known properties of the exact functional into the restricted functional form of the approximation.

In LDA the XC potential has an explicit analytical form

$$v_{XC}^{LDA} = \frac{df_{LDA}}{d\rho} = \varepsilon_{XC}^{LDA} + \rho \frac{d\varepsilon_{XC}^{LDA}}{d\rho} \quad (2.13),$$

whereas in the GGA it is considerably more complicated [35]

$$\begin{aligned} v_{XC}^{GGA} &= \frac{\partial f_{GGA}}{\partial \rho} - \nabla \frac{\partial f_{GGA}}{\partial \vec{g}} \\ &= \varepsilon_{XC}^{GGA} + \rho \frac{\partial \varepsilon_{XC}^{GGA}}{\partial \rho} - g \frac{\partial \varepsilon_{XC}^{GGA}}{\partial g} - \rho g \frac{\partial^2 \varepsilon_{XC}^{GGA}}{\partial g \partial \rho} - \\ &\quad - \frac{\rho}{g} \frac{\partial \varepsilon_{XC}^{GGA}}{\partial g} \nabla^2 \rho + \frac{\rho}{g^2} \frac{\partial \varepsilon_{XC}^{GGA}}{\partial g} \vec{g} \nabla g - \frac{\rho}{g} \frac{\partial^2 \varepsilon_{XC}^{GGA}}{\partial g^2} \vec{g} \nabla g \end{aligned} \quad (2.14).$$

It is easy to find that v_{XC}^{GGA} depends on both the first and the second gradient of the electron density. Since $\nabla \rho(\vec{r})$ has cusps at extreme of ρ , ∇g is discontinuous at these points, which causes problems for its numerical representation. Furthermore, some parameterizations of ε_{XC}^{GGA} that apparently join seamlessly different density regimes, may have higher derivatives that do not behave well in those transition regions.

The following comparison of two well-known XC functionals — the Ceperley-Alder [36] (LDA) and PBE96 [34] (GGA) — will explicitly show the differences. The analytical expression of the former one is as follows

$$f_{LDA}(\rho) = \rho \varepsilon_{XC}^{LDA} = \rho(\varepsilon_X^{LDA} + \varepsilon_C^{LDA}) \quad (2.15),$$

in which

$$\begin{aligned} & \text{if } r_s \geq 1.0 && \text{else} \\ & \varepsilon_X^{LDA} = -\frac{3}{2\pi\alpha} r_s^{-1} && \\ & \varepsilon_C^{LDA} = \frac{g}{1 + B_1 r_s^{1/2} + B_2 r_s} && , \quad \varepsilon_C^{LDA} = A_F \ln(r_s) + B_F + C_F r_s \ln(r_s) + D_F r_s \\ & v_X^{LDA} = \frac{4}{3} \varepsilon_X^{LDA} && (2.16). \\ & v_C^{LDA} = \frac{(\varepsilon_C^{LDA})^2}{g} \left(1 + \frac{7}{6} B_1 r_s^{1/2} + \frac{4}{3} B_2 r_s \right) && v_C^{LDA} = A_F \ln(r_s) + \left(B_F - \frac{1}{3} A_F \right) + \\ & = \frac{g \left(1 + \frac{7}{6} B_1 r_s^{1/2} + \frac{4}{3} B_2 r_s \right)}{\left(1 + B_1 r_s^{1/2} + B_2 r_s \right)^2} && + \frac{2}{3} C_F r_s \ln(r_s) + \frac{1}{3} (2D_F - C_F) r_s \end{aligned}$$

Here $r_s = \left(\frac{3}{4\pi\rho} \right)^{1/3}$ is the local Seitz radius, and other coefficients are predefined constants.

The expression of XC energy per electron (ε_{XC}^{GGA}) in PBE96 is different such that

$$f_{GGA}(\rho) = \rho \varepsilon_{XC}^{GGA} = \rho \left(\varepsilon_X^{LDA} F_X(s) + \varepsilon_C^{LDA} + H_C(r_s, t) \right) \quad (2.17),$$

where

$$\begin{aligned}
F_X(s) &= 1 + \kappa - \kappa / (1 + \mu s^2 / \kappa) \\
H_C(r_s, t) &= \gamma \times \ln \left(1 + \frac{\beta}{\gamma} t^2 \left[\frac{1 + At^2}{1 + At^2 + A^2 t^4} \right] \right) \\
A &= \frac{\beta}{\gamma} [\exp\{-\varepsilon_C^{LDA} / \gamma\} - 1]^{-1} \\
s &= |\nabla \rho| / 2k_F \rho, \quad t = |\nabla \rho| / 2k_s \rho \\
\rho &= k_F^3 / 3\pi^2, \quad k_s = \sqrt{\frac{4k_F}{\pi}}, \quad \kappa, \mu, \beta \text{ and } \gamma \text{ const}
\end{aligned} \tag{2.18}.$$

It is too complicated to obtain a simple analytical form of the XC potential v_{XC}^{GGA} for direct implementation. For most popular *ab initio* computational packages, GGA is implemented through numerical interpolation. Some practical schemes for evaluating ε_{XC}^{GGA} and v_{XC}^{GGA} will be reviewed later on.

2.2 The VASP Package and Calculation

2.2.1 Introduction

The VASP introduced by Kresse and co-workers in 1993 [19] is a package for performing *ab initio* quantum mechanical calculations, using either Vanderbilt pseudo-potentials or the projector augmented wave (PAW) method, and a plane-wave (PW) basis set. VASP uses a self-consistent cycle with a Pulay mixer and an iterative matrix diagonalisation scheme to calculate the KS ground-state equation. These techniques avoid the problems occurring in the original Car-Parrinello method [37, 38] which is based on the simultaneous integration of electronic and ionic equations of motion.

The use of the PW basis set enables VASP to take advantage of fast-Fourier-transforms (FFT) technique for calculating the operation of the Hamiltonian and avoid the Pulay-term in the Hellmann-Feynman forces [39] calculation. Forces and the full stress tensor can therefore be easily calculated and used to relax atoms into their instantaneous ground-state. It is well recognized that VASP is an accurate and efficient package suitable for geometry optimization of lots of different systems. In order to solve the difficulties associated with the PW basis set, VASP introduced pseudo-potential and pseudo-WF constructed from spherical Bessel functions to describe the core region.

Pseudo-potential is an effective potential constructed to replace the all-electron atomic potential such that core states are eliminated and the valence electrons are described by nodeless pseudo-WFs. In this approach only the chemically active valence electrons are dealt with explicitly, while the core electrons are ‘frozen’, being considered together with the nuclei as rigid non-polarizable ion cores. There are three strategies supported by VASP to describe core region ion-electron interaction: optimized norm-conserving pseudo-potential [40-42], ultra-soft Vanderbilt pseudo-potentials (US-PP) [43-45], and the PAW method [46]. Norm-conserving pseudo-potentials are derived from an atomic reference state, requiring that the pseudo- and all electron valence eigenstates have the same energies and amplitude (and thus density) outside a chosen core cutoff radius r_c . Pseudo-potentials with shorter cutoff radius are said to be soft, which is more rapidly convergent, but less transferable, that is less accurate to reproduce realistic features in different environments. The recently introduced

PAW method performs a linear transformation from the pseudo- to all-electron WFs, and hence greatly reduces the number of parameters needed for the construction of its datasets.

VASP has already been successfully applied to a large number of different systems including liquid, amorphous semiconductors, transition metals, metallic and semiconducting surfaces and interfaces as well as phonons in solids. However, as is well known that no method can be good for everything, there are limitations in VASP. The number of the WFs in the basis set depends on the energy cutoff, which will restrict VASP from efficient calculation of core-level spectral properties for large systems. In addition, using PW basis set increases the difficulty to trace the detailed information in the electronic structure. In order to compensate these disadvantages, other *ab initio* DFT packages including the OLCAO method and SIESTA have been adopted.

2.2.2 Geometry Optimization

Geometry optimization is the process of finding the minima of the potential energy on the mesh surface. The structures with the minimum energy are called equilibrium structures where the force on each atom is close to zero. It is an important step in computational simulation, because the atomic configuration determines many physical and chemical properties. To perform geometry optimization, the forces on each atom need to be calculated first using derivatives of the total energy. Then the atoms are moved to the equilibrium positions where the total force is zero. With the new structure configuration, new potential energy can be obtained followed by new forces exerted on the atoms. With a sufficient

number of iterations, the differences between the new energy value and the previous one will fall into the acceptable convergence range. Then the geometry optimization is considered to be complete. Apparently, the optimization relies on the level of energy convergence level. Without this step, if there is still large force exerted on atoms, the structure is not stable in reality, thus the following analysis might distort our understanding. Possible false results could be gap states between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) in the insulator. Optimization can also be used to locate transition structures, which are represented by saddle points on the total energy surface.

One important property need to be mentioned about the *ab initio* relaxation is that each atom in the molecule is considered as an independent particle such that the short ranged interatomic interaction and the longer ranged intermolecular forces are treated on equal footing. The intermolecular forces determined in this way tend to allow the system to relax into the most favorable instantaneous configuration among many low energy possibilities. This is in contrast with many molecular dynamics (MD) simulations. Taking water for example, MD will treat each H₂O molecule as a fixed entity with only orientational and translational degrees of freedom.

2.2.3 Parallel VASP Running Experiences

Another experience gained from using the parallel VASP code on the supercomputer Hopper Cray XE6 at the National Energy Research Scientific Computing Center (NERSC)

[47] is that with the optimized parameter options, we can speed up the computational process significantly. Hopper has 2 twelve-core AMD 'MagnyCours' 2.1 GHz processors per node with a per core memory of 1.3Gb. While running jobs on the reduced number of cores per node, we found that by using a couple of aprun command line options, -S and -cc, the VASP relaxation on a 660-atom DNA double-helix model (results shown in Chapter 3) ran around 2.5 times faster. The detailed comparison of the time used for an ionic step before and after adding the command options are listed in Appendix A. Both jobs run on 32 Hopper nodes. The LOOP time is the time used for one electronic step and the elapsed time is for one ionic step. We see that one ionic step was shortened from 8631.025s to 3327.090s. The reason for this improvement is that the jobs are better distributed to utilize the parallel programming algorithm implemented in the package.

2.3 The SIESTA Method and Calculation

SIESTA is a DFT method for performing electronic structure calculations and *ab initio* MD simulations of molecules and solids [20, 48, 49]. It was born in 1996 to implement, in self-consistent DFT, the order-N ($O(N)$) algorithms developed for tight-binding model [50, 51]. It allows DFT simulations for more than a thousand atoms in modest PC workstations, and over a hundred thousand atoms in parallel platforms. By far, a large variety of systems involving nano-scale materials have been addressed using this method, including nanoclusters, nanotubes, biomolecules, adsorbates at surfaces, etc [52].

SIESTA uses the standard norm-conserving pseudo-potentials [40-42] to get rid of the core electrons and allow for the expansion of a smooth (pseudo-) electron density on a uniform spatial grid. The use of a linear combination of strictly confined atomic orbitals basis set generates sparse Hamiltonian and overlap matrices required by the $O(N)$ method [53]. The basis set includes multiple- ζ functions to describe the atomic valence orbitals, and the polarization orbitals to account for the deformation induced by bond formation. There are different basis set options available in SIESTA. The minimal single- ζ (SZ) basis is appropriate for the extremely fast semiquantitative simulations, while the standard double- ζ polarization (DZP) basis yields high-quality results for most of the systems under reasonable computational costs. The basis functions and the electron density are projected on a real-space grid, in order to calculate the Hartree and XC potentials and matrix elements, with a number of operations that scales linearly with the size of the system. SIESTA allows both LDA and GGA XC functional, and the density gradients for the latter are evaluated by finite differences in the grid.

Several additional important features of SIESTA are mentioned briefly: (1) The atomic forces and stresses are obtained by direct differentiation of the KS total energy with respect to atomic positions, which allows the simultaneous relaxation of atomic coordinates and cell size using conjugate gradients minimization or several other annealing algorithms; (2) MD simulations are also allowed at constant energy or temperature, and at constant volume or pressure [54]; (3) There exist an auxiliary program VIBRA to compute the Hessian matrix and phonon spectrum as well as an interface between the PHONON program

[55] and SIESTA; and (4) Transport properties and optical responses are also possible to study using difference approaches.

2.4 The OLCAO Method and Calculation

2.4.1 Overview

The concept of using linear combinations of atomic orbitals (LCAOs) in the band theory of solids was originally introduced by Bloch in 1928 [56]. After many decades of systematic refinement [57-59], the current real-space self-consistent field (SCF) [60] orthogonalized-LCAO (OLCAO) method combined many advantages of the developments from various groups and has become computationally very efficient. OLCAO is an all electron DFT method using LDA approximation. It is particularly suitable for the calculation of electronic, optical, magnetic, and spectroscopic properties of complex condensed matter systems. By far, it has been applied to the supercell amorphous solids and liquids [61, 62], grain boundaries [63], surfaces and interfaces [64, 65], biomaterials [66, 67], as well as superconductors [68]. The mathematical rigor and technical details behind the present OLCAO package have been well described previously [68-70]. We will simply go through some of the key terminologies and concepts, selectively discuss the technical aspects, and introduce the computational capabilities that have been used to investigate the systems in this work. In the end, special discussions will be put on the practical GGA implementation scheme.

The Atom-centered Potential Functions

In the OLCAO method, the effective potential v_{eff} is represented by a superposition of the atom-centered potential functions which are themselves linear combinations of atom-centered gaussians, plus corresponding Gaussian dampened pointer potentials to correctly simulate the near nucleus behavior.

$$v_{eff}(\vec{r}) = \sum_A v_A(\vec{r} - \vec{t}_A) \quad (2.19),$$

$$v_A(\vec{r}) = -\frac{Z_A}{r} \exp(-\xi_A r^2) + \sum_{j=1}^N D_{j,A} \exp(-\beta_{j,A} r^2) \quad (2.20),$$

where Z_A is the atomic number of the atom at the site, and $\xi_A, \beta_{j,A}$ are the exponential parameters which are predetermined. The first term in (2.20) comes from the ion potential screened by the electrons which describes the behavior of the nuclear Coulomb effect. The second term consists of the contributions from electronic Coulomb effect and many-body XC interactions. During the SCF iterations, the first term will not change whereas the coefficients $\{D_{j,A}\}$ in the second term are updated each electronic step. Typical datasets of $\{D_{j,A}\}$ and $\{\beta_{j,A}\}$ of oxygen atoms are shown in APPENDIX B.

Here, a very crucial approximation is adopted. For the same atomic site, the same exponential set $\{\beta_{j,A}\}$ in the Gaussian expansion in (2.20) is used. This approximation reduces the number of multicenter integrals between GTOs and enables OLCAO method to be applied to large complex systems. The optimal choices of $\{\beta_{j,A}\}$ are usually attained through the application of OLCAO to a large number of simple crystals. OLCAO has a

relatively comprehensive database of the exponential coefficients for most common elements.

The WF and Atomic Basis Set

The solid state WF solution of the KS equation is expanded in terms of Bloch functions $b_{i\gamma}(\vec{k}, \vec{r})$:

$$\psi_{nk}(\vec{r}) = \sum_{i,\gamma} C_{i\gamma}^n(\vec{k}) b_{i\gamma}(\vec{k}, \vec{r}) \quad (2.21),$$

where γ labels the atoms in the cell and i represents the orbitals' quantum number (l, m).

The Bloch function is constructed from the linear combination of atomic orbitals u_i centered at each atomic site:

$$b_{i\gamma}(\vec{k}, \vec{r}) = \frac{1}{\sqrt{N}} \sum_{\nu} \exp(i\vec{k} \cdot \vec{R}_{\nu}) u_i(\vec{r} - \vec{R}_{\nu} - \vec{t}_{\gamma}) \quad (2.22).$$

Here \vec{R}_{ν} represents the lattice, \vec{t}_{γ} is the position of the γ th atom in the cell, and \vec{k} is the reciprocal space vector in the Brillouin zone. The atomic orbital u_i is represented by product of the linear combination of a suitable number of Gaussian Type Orbitals (GTOs) and a spherical harmonic function

$$u_i(\vec{r}) = \left[\sum_j C_j r^{n-1} \exp(-\alpha r^2) \right] \cdot Y_{lm}(\theta, \varphi) \quad (2.23),$$

where the quantum numbers n, l, m are collectively labeled as i . The exponentials $\{\alpha_j\}$ are predefined and their coefficients $\{C_j\}$ are obtained through separate atomic calculations.

More specifically, we have

$$\begin{aligned}
s\text{-type} &: C_s r^{n-1} \exp(-\alpha r^2) \\
p\text{-type} &: C_{p,x} r^{n-2} x \exp(-\alpha r^2), \text{ same for } y, z \\
d\text{-type} &: C_{d,xy} r^{n-3} xy \exp(-\alpha r^2), \text{ same for } yz, zx, x^2 - y^2, 2z^2 - x^2 - y^2 \quad (2.24). \\
f\text{-type} &: C_{f,xyz} r^{n-4} xyz \exp(-\alpha r^2), \text{ same for } z(x^2 - y^2), x(x^2 - 3y^2), \\
&\quad y(y^2 - 3x^2), z(5z^2 - 3r^2), x(5z^2 - r^2), y(5z^2 - r^2)
\end{aligned}$$

There are three types of basis set for every element: minimal basis (MB), full basis (FB) and extended basis (EB). The general rules are as follows: the MB contains the minimum number of orbitals which has at least one occupying electron, the FB is obtained by adding one more orbital for each of the present angular quantum numbers, and the EB follows the same manner by adding one more layer to FB. Taking the C atom for example, the three types of basis set are $(1s, 2s, 2p)$, $(\text{MB} + 3s, 3p)$, and $(\text{FB} + 4s, 4p)$. Instead of using the above as fixed rules, OLCAO offers a lot of flexibilities for incorporating more orbitals into the basis which is especially useful for heavy atoms. Taking the transition metal Fe for example, the MB is $(1s, 2s, 2p, 3s, 3p, 4s, 4p, 3d)$, the FB is $(\text{MB} + 5s, 4d)$, and the EB can be $(\text{FB} + 5p, 6s, 5d, 6p)$. With many years of efforts, a complete database of the atomic orbitals has been built up for most of the common atoms.

An example of the standard MB dataset of O atom can be found in APPENDIX C. From this file, we can get the following information: (1) the exponentials $\{\alpha_j\}$ are simply

ranging from a minimum α_{\min} and a maximum α_{\max} distributed in a geometric series; (2) the $1s$ orbital is accounted as the CORE, while the $(2s, 2p)$ are put into the VALENCE orbitals; and (3) all orbitals of one atom are expanded using the same set of Gaussian functions. Something additional features of the OLCAO basis functions deserving comments include: (1) the exponential set can be obtained using the geometry optimization scheme within the OLCAO method; (2) a core orbital is generally defined as the energy level of the orbital is lower than that of oxygen $2s$ orbital; (3) there is plenty of flexibility to define a core state (e.g. Ca $3p$ semi-core orbital is treated as a valence state for better accuracy); and (4) the WF of the same element share the same set of exponential values. Therefore, elements of the Hamiltonian and Overlap matrices belonging to the same atom pair can be calculated using the results of the same set of GTO pairs. For atoms with large number of orbitals, the advantages of using the same sufficiently large set of exponentials becomes apparent since the computational cost can be lowered by up to an order of magnitude.

Secular Equations

By substituting the WF with a linear combination of atomic orbitals, the standard solution of the KS equation (2.8) lead to the secular equation

$$\left| H_{i\gamma, j\delta}(\vec{k}) - S_{i\gamma, j\delta}(\vec{k}) E(\vec{k}) \right| = 0 \quad (2.25),$$

where $H_{i\gamma, j\delta}(\vec{k})$ and $S_{i\gamma, j\delta}(\vec{k})$ are the Hamiltonian and Overlap matrices respectively:

$$\begin{aligned}
H_{i\gamma,j\delta}(\vec{k}) &= \langle b_{i\gamma}(\vec{k}, \vec{r}) | \hat{H} | b_{j\delta}(\vec{k}, \vec{r}) \rangle \\
&= \sum_{\mu} \exp(-i\vec{k} \cdot \vec{R}_{\mu}) \int u_{i\gamma}(\vec{r} - \vec{t}_{\gamma}) \left(-\frac{1}{2}\nabla^2 + v_{eff}\right) u_{j\delta}(\vec{r} - \vec{t}_{\delta} - \vec{R}_{\mu}) d\vec{r} \\
S_{i\gamma,j\delta}(\vec{k}) &= \langle b_{i\gamma}(\vec{k}, \vec{r}) | b_{j\delta}(\vec{k}, \vec{r}) \rangle \\
&= \sum_{\mu} \exp(-i\vec{k} \cdot \vec{R}_{\mu}) \int u_{i\gamma}(\vec{r} - \vec{t}_{\gamma}) u_{j\delta}(\vec{r} - \vec{t}_{\delta} - \vec{R}_{\mu}) d\vec{r}
\end{aligned} \tag{2.26}$$

In the case of s-type GTOs (n=0), the interaction integrals have the following forms:

Overlap integral:

$$I_1 = \langle s_A | s_B \rangle = \int \exp(-\alpha_A r_A^2 - \alpha_B r_B^2) d\vec{r} \tag{2.27},$$

Kinetic energy integral:

$$I_2 = \langle s_A | -\nabla^2 | s_B \rangle = \int \exp(-\alpha_A r_A^2) (-\nabla^2) \exp(-\alpha_B r_B^2) d\vec{r} \tag{2.28},$$

Three center integral:

$$I_3 = \langle s_A | e^{-\alpha_C r_C^2} | s_B \rangle = \int \exp(-\alpha_A r_A^2) \exp(-\alpha_C r_C^2) \exp(-\alpha_B r_B^2) d\vec{r} \tag{2.29},$$

Three center coulomb integral:

$$I_4 = \left\langle s_A \left| \frac{e^{-\alpha_C r_C^2}}{r} \right| s_B \right\rangle = \int \exp(-\alpha_A r_A^2) \frac{\exp(-\alpha_C r_C^2)}{r} \exp(-\alpha_B r_B^2) d\vec{r} \tag{2.30},$$

Momentum integral:

$$I_5 = \langle s_A | \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) | s_B \rangle = \int \exp(-\alpha_A r_A^2) \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z} \right) \exp(-\alpha_B r_B^2) d\vec{r} \tag{2.31}.$$

We can see that the use of Gaussians as the ultimate basis facilitates the analytical integral calculations. Regarding the higher angular momentum orbitals such as p-type, d-type and f-type orbitals, the recursive formula are used to obtain their expression:

$$\begin{aligned}
& \left\langle (x - A_x)^{i_A} (y - A_y)^{j_A} (z - A_z)^{k_A} \hat{s}_A \left| \hat{O} \right| (x - B_x)^{i_B} (y - B_y)^{j_B} (z - B_z)^{k_B} \hat{s}_B \right\rangle \\
& = \frac{1}{2\alpha_B} \frac{\partial}{\partial B_z} \left\langle (x - A_x)^{i_A} (y - A_y)^{j_A} (z - A_z)^{k_A} \hat{s}_A \left| \hat{O} \right| (x - B_x)^{i_B} (y - B_y)^{j_B} (z - B_z)^{k_B} \hat{s}_B \right\rangle
\end{aligned} \tag{2.32}$$

Electron Density Representation

The secular equations solved using the SCF method give us the KS orbitals $\psi_{nk}(\vec{k}, \vec{r})$, from which the electron density can be derived as follows:

$$\rho^v(\vec{r}) = \int \sum_{BZ \text{ occ.}} |\psi_{nk}(\vec{k}, \vec{r})|^2 d\vec{k} \tag{2.33}$$

There is one thing to be mentioned that $\rho^v(\vec{r})$ represents the contribution of valence electron density whereas the core electron density $\rho^c(\vec{r})$ is obtained separately. This is an important feature of the OLCAO method resulting from the core-orthogonalization process which performs a unitary transformation to the Hamiltonian and Overlap matrices to eliminate the non-diagonal blocks of the core states. We can do this because in most cases the core states are highly localized and of little research interest in solids. The purpose of doing so is to improve the computational efficiency while not affecting the accuracy, and it has been proved extremely advantageous in complex systems. The theoretical details of core-orthogonalization has been presented in Ouyang's work [69] and will not be repeated here.

To reduce the dimensionality of the potential energy integrals, $\rho^v(\vec{r})$ and $\rho^c(\vec{r})$ are further cast into the same set of atom centered gaussian functions as those used in the potential representation using least square fitting

$$\rho_A^v(\vec{r}) = \sum_j F_{j,A} \exp(-\beta_{j,A} r^2) \quad (2.34),$$

$$\rho_A^c(\vec{r}) = \sum_j G_{j,A} \exp(-\beta_{j,A} r^2) \quad (2.35).$$

Generally, we can use the following schematic diagram to summarize the SCF process

$$\left. \begin{matrix} \{D_{jA}\} \\ v_{eff} \end{matrix} \right\} \xrightarrow{\text{secular equation}} \left. \begin{matrix} \{C_{jA}\} \\ \Psi_{nk} \end{matrix} \right\} \xrightarrow{\text{Least square fit}} \left. \begin{matrix} \{F_{jA}, G_{jA}\} \\ \rho_A \end{matrix} \right\} \xrightarrow{\text{Least square fit}} \left. \begin{matrix} \{D_{jA}\} \\ v_{eff} \end{matrix} \right\} \quad (2.36),$$

which generates an updated set of potential coefficients $\{D_{jA}\}$ and eventually leads to the required convergence. During this process, the electron density needs to be numerically evaluated on a predefined real-space mesh to calculate the contribution of XC energy and potential. We will discuss the characteristics of this mesh shortly. The OLCAO method contains more complicated details beyond what we have briefly discussed. To fully understand them requires not only the mathematical derivation but also careful reading of the source code and running jobs on actual computer platforms.

2.4.2 Technical Aspects

An OLCAO calculation is usually carried out in two stages: SCF portion and post-SCF portion. In this section, the responsibilities of both stages will be introduced first followed by the file-storage system which organizes the massive intermediate data created by the OLCAO calculation. There has already been a very comprehensive description of these technical details by Rulis [70], but the purpose of repeating them here is to help the readers

get a general picture of the computing procedures in their mind which later on will facilitate understanding of the discussion on the GGA XC functional implementation.

The SCF portion contains two consecutive program executables. The first one is called the setup phase in which the analytic integration of multi-centered Gaussian functions is performed, electrostatic interactions with the atomic nuclei are computed, and the mesh for the evaluation of the XC effect is initialized [70]. The second step is called the main phase. Based on the initial guess of the potential, the KS equations are solved using the Linear Algebra Package library [71]. The eigenvectors, i.e. wave functions, are then used to compute the electron density. Finally, the electron density is used to compute a new potential and thus complete an iteration of the SCF cycle. Once the potential is converged to a desired level of accuracy, it can be used for the next stage post-SCF calculation.

The post-SCF portion consists of five program executables. The first one called integral phase is responsible for evaluating the analytic multi-center Gaussian integrations for specific k-points within the reciprocal-space lattice of the crystal system. But if the system has no symmetry like the biomolecules in the present work, or if the supercell is large enough such that the reciprocal-space lattice is very small, only one Γ -point (0, 0, 0) will be sufficient. The second program is the wave function phase which calculates the wave functions and energy eigenvalues at each selected k-point. The three remaining programs use the results of the wave function phase to compute either the partial density of states (PDOS) and local index (LI), bond order (BO) and effective charge (Q^*), or various spectroscopic properties.

The calculations in both SCF and post-SCF stages will create large amounts of intermediate data that need to be read and rewritten frequently. To reduce the data storage footprint and easily access the intermediate data, the Hierarchical Data Format Version 5 (HDF5) [72] file system introduced by Rulis [70] plays the key role. Taking the setup HDF5 file generated from the SCF portion for examples (Figure 1 [70]), the root of the file is divided into three distinct sections corresponding to the atomic interaction integrals, the XC mesh, and the electrostatic interactions. Within the atomic interaction integral section are four sub-groups. Three of the sub-groups each contain an array of datasets corresponding to the k-points for the atomic overlap, kinetic energy, and nuclear potential integral contributions to the Hamiltonian. The fourth sub-group is itself divided into an array of sub-sub-groups. Each sub-sub-group is created for one term from the atomic potential contribution to the Hamiltonian. These sub-sub-groups then contain an array of datasets corresponding to the k-points for the potential terms. The XC group does not contain any sub-groups but instead has an array of triplets of datasets. The number of triplets is equal to the number of potential sites in the system. Each triplet is composed of a dataset holding a single number indicating the number of points for the XC mesh at this potential site, a dataset for the radial weight of each point for the associated potential site, and a dataset for the exchange density operator for this potential site. In addition to the potential site datasets a single dataset is attached to the XC group that holds the overlap dataset. The electrostatic group contains only single datasets. The local neutral charge, local nuclear charge, non-local neutral charge, non-local nuclear charge, non-local polar charge, core charge density, and

electrostatic overlap datasets are held here. The vast majority of the data are stored in the atomic potential matrices and the XC groups.

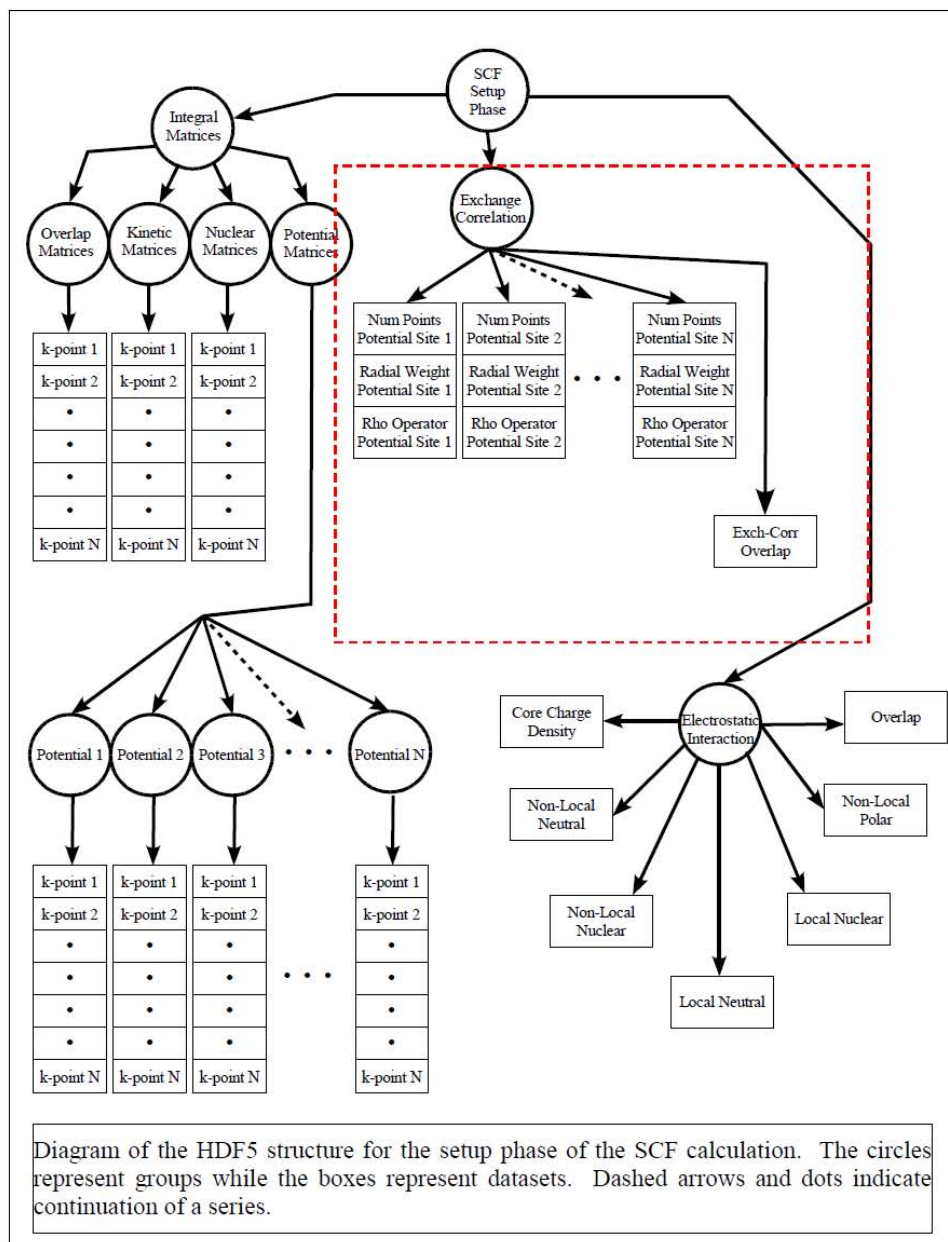


Figure 1. SCF setup phase HDF5 File from ref. [70].

It is necessary to put some detailed description on the XC group (see Figure 1 inside the red box) of the setup HDF5 file because of its tight relation to the GGA implementation scheme. As the first step, a picture of the real-space XC mesh (see Figure 2, for simplicity only a 2D lattice is considered) in mind will help us understand the advantages of organizing data in this style.

The OLCAO mesh is specially designed to fit for the atom-centered atomic orbitals and potentials. The basic unit for mesh construction is the Wigner cell of each atom (yellow and green boxes). In each Wigner cell, many rays originating from the center of the atom are uniformly distributed. Along each ray, many mesh points are assigned according to the weighting scheme such that more points closer to the nuclear sites than far away from the sites. The number of rays can be changed depending on the system size, but the default value is 99 rays. According to the default weighting scheme for point distribution along the rays each atom contributes approximately 4000 points to the system mesh.

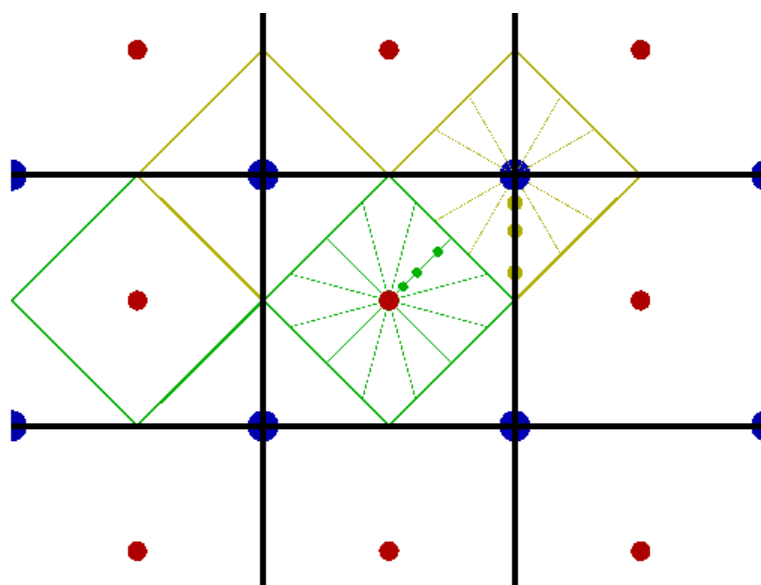


Figure 2. Schematic display of the OLCAO XC mesh: Wigner cell of each atom (yellow and green boxes); Sampling rays: green (yellow) vectors starting from the atom ending at the boundary of the Wigner cells; Sampling points: green (yellow) dots.

Each triplet in the XC group stores the weighting factor and density operator of every mesh point in the Wigner cell of one particular potential site. We have mentioned that the electron density contributed by each potential site has been fitted into the same set of Gaussian functions as the energy potential, therefore the evaluation of charge density at certain mesh point is equivalent to summing up the contribution from every Gaussian function within the cutoff range. Since only the fitting coefficients change during the SCF cycles, we can add up the invariant exponential parts according to their exponential values $\beta_{j,A}$ to construct density operators (Rho Operators). The operators are then saved in the HDF5 file and later on will be recalled once the updated coefficients are ready.

The computing procedures to get density operators can be expressed as follows

loop :

 each mesh point k

 each potential site A

 each replicated cell m

 each exponential gaussian $\beta_{j,A}$ of A

$$\text{RhoOperator}(j, k) = \text{sum}(\exp[-\beta_{j,A}(\vec{r}_k - \vec{r}_{m,l})^2])$$

 ...

 end

(2.37),

in which \vec{r}_k represents the position of mesh point k , $\vec{r}_{m,l} = \vec{r}_A + \vec{r}_{LatticeVec} + \vec{r}_{Cell}$ locates the center of the Gaussian function. Finally, according to (2.34), the numerical value of valence electron density at mesh point k is $\rho_k^v = \sum_{j,A} F_{j,A} \times \text{RhoOperator}(j, k)$.

2.4.3 Scheme for GGA Implementation

Existing GGA Implementation Scheme

As mentioned previously, GGA XC functionals has been well accepted due to the higher accuracy for many systems. Nowadays different implementation schemes have been developed to allow its practical use by various first principles packages. In DFT calculations, the XC energy and potential of the system are usually obtained by evaluating the quantities on a predefined real-space numerical mesh then taking the integral. GGA XC functionals depend on both the electron density ρ and the electron density gradient $\nabla\rho(\vec{r})$. If a well-defined basis set is used, ρ and $\nabla\rho(\vec{r})$ can be calculated exactly at any point in the space, from the electron WFs and their gradients. In practice, this may add a considerable overhead in terms of computer time and memory. Alternatively, the gradient can be expressed as a linear operator of the density at the neighboring grid points (see Figure 3(a), for simplicity only a 2D lattice is considered), using either FFTs or finite differences [35, 73]. The former method usually comes with uniform integration mesh and PW basis set, while the later can be generalized to nonuniform mesh and arbitrary basis using Lagrange interpolation. For example, VASP has adopted the former scheme, and SIESTA uses the later. By the meaning of nonuniform mesh, it indicates the interval between two points can vary. Other than that it is similar to the uniform one. We will simply categorize them as regular mesh in which the crossings of the horizontal and vertical lines are the locations of the mesh points for interpolation. The OLCAO XC mesh does not have this feature so that it will be referred to as irregular mesh in the following text.

Difficulties with the Present OLCAO Mesh Scheme

Although the atom-centered irregular mesh is very convenient for evaluating the electron density and XC potential under LDA, it creates troubles in GGA. The most crucial problem is that the mesh points lack of a concise sequence order within the whole system because they are generated locally according to each potential site. Moreover, the sequence order cannot reflect how close two points are as handy as the regular mesh. To be more specific, the comparison of density gradient interpolation on both meshes surrounding the purple point is presented as follows. Beforehand we should be clear that simply including all existing points for interpolation will be unaffordable so that some cutoff limits are used. On a regular mesh (Figure 3(a)), the interpolating points (yellow ones) for the purple point (i, j) can be picked up immediately at $(i-n, j), \dots, (i-1, j), (i+1, j), \dots, (i+n, j)$ and $(i, j-n), \dots, (i, j-1), (i, j+1), \dots, (i, j+n)$. On the OLCAO mesh (Figure 3(b)), for the purple point (i, j, k) at the center — the k^{th} point on the j^{th} ray of the i^{th} atom site, all points within the cutoff region (black circle) from both the original (green) and the neighboring (yellow) Wigner cell need to be considered. We probably can identify those points by tracking their distances to the center and trace back to the coordinates, but that will require large additional computer memory to store the distance values, and the searching process, which will be repeated for every mesh point in every SCF cycle, will impede the computational efficiency significantly. For normal solid state systems which easily contain tens or hundreds of atoms, the possibility of using the original mesh for GGA interpolation has been excluded. Beyond this major problem, how to decide the cutoff region (sphere? box? how big? etc.) to ensure a valid interpolation is also

an open question. The density of the mesh points in different regions varies due to irregularity of the mesh, which prevents us from using a unified criterion.

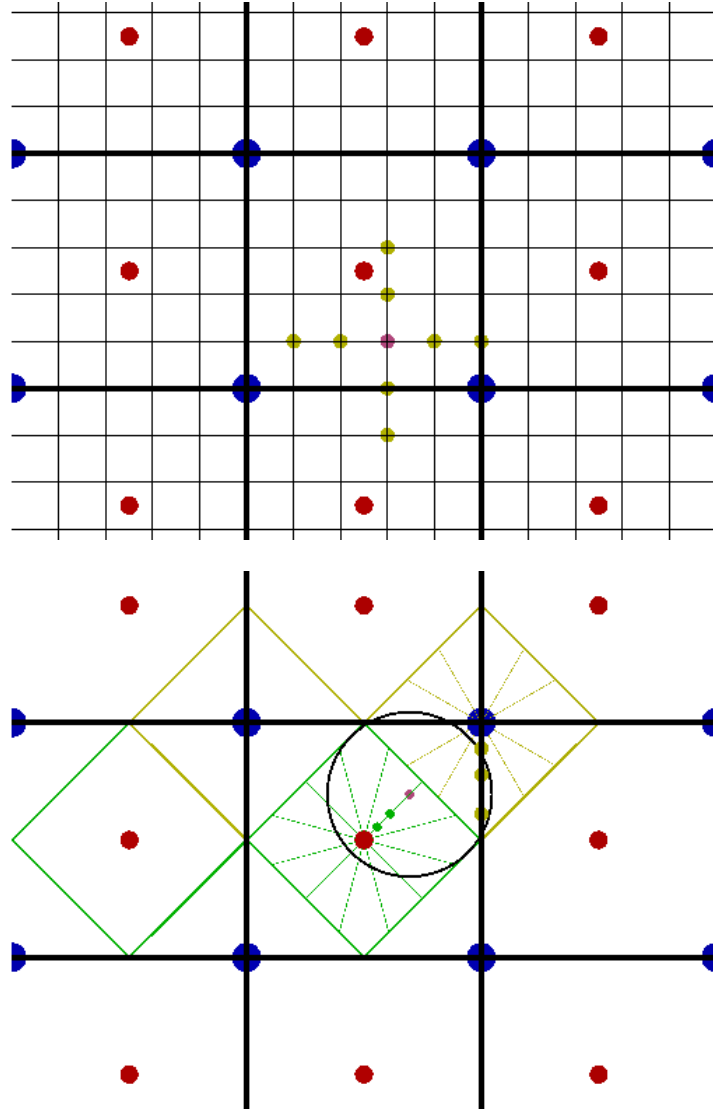


Figure 3. Schematic display of the density gradient interpolation on (a) uniform mesh (upper panel) and (b) OLCAO mesh (lower panel).

Discussions on Potential Solutions for OLCAO

By far, we have been aware of the intrinsic advantages and inadequacies associated with the present OLCAO mesh. To implement GGA, we could learn from SIESTA by introducing a uniform mesh for electron density (ρ) evaluation and using the finite differences for density gradient ($\nabla\rho(\vec{r})$) interpolation. Accordingly, the storage of the XC intermediate data (see Figure 1 red box region) need to be updated because the mesh points are no longer localized to each potential site. One potential file system is displayed in the following schematic diagram (Figure 4): the large data block contains ($N_1 \times N_2 \times N_3$) data cells, and in each cell the information associated with one particular mesh point are categorized into five different sections, including the mesh index, Cartesian coordinates, ρ operator, interpolating matrix, and the density gradient $\nabla\rho(\vec{r})$ value. Among these, the first four sections will not be changed during the SCF iterations. The ρ operators can be calculated following the same procedures in the present OLCAO program suite. The $\nabla\rho(\vec{r})$ values will be obtained after the evaluation of ρ values at all mesh points has been done, and this data section will be updated every iteration.

Although regular mesh is more suitable for electron density gradient evaluation compare to the original one in OLCAO, it also loses some benefits. The original mesh allows a much denser distribution of mesh points when close to the atom, which is easier for sampling the electrons near the core region. Let us consider a simple cubic lattice with two inequivalent atom sites thus two Wigner cells, a total of only 8000 mesh points by default (4000×2) will be sufficient for correctly interpreting the electron distribution. On the other

hand, using a $(20 \times 20 \times 20)$ regular mesh might be too sparse for the core region but too much for the region far away from the atom site. Apparently, simply increasing the mesh density will result in a much more expensive calculation in terms of computer storage and simulation time. To solve this dilemma, data parallelization algorithms will be very helpful. However, partial parallelism does not necessarily mean more efficiency, while detailed discussions on the parallelization of the entire package are beyond the scope of this present work. Rulis has proposed the parallelization schemes with respect to different portions of the OLCAO program suite [70], but to integrate them definitely requires more efforts.

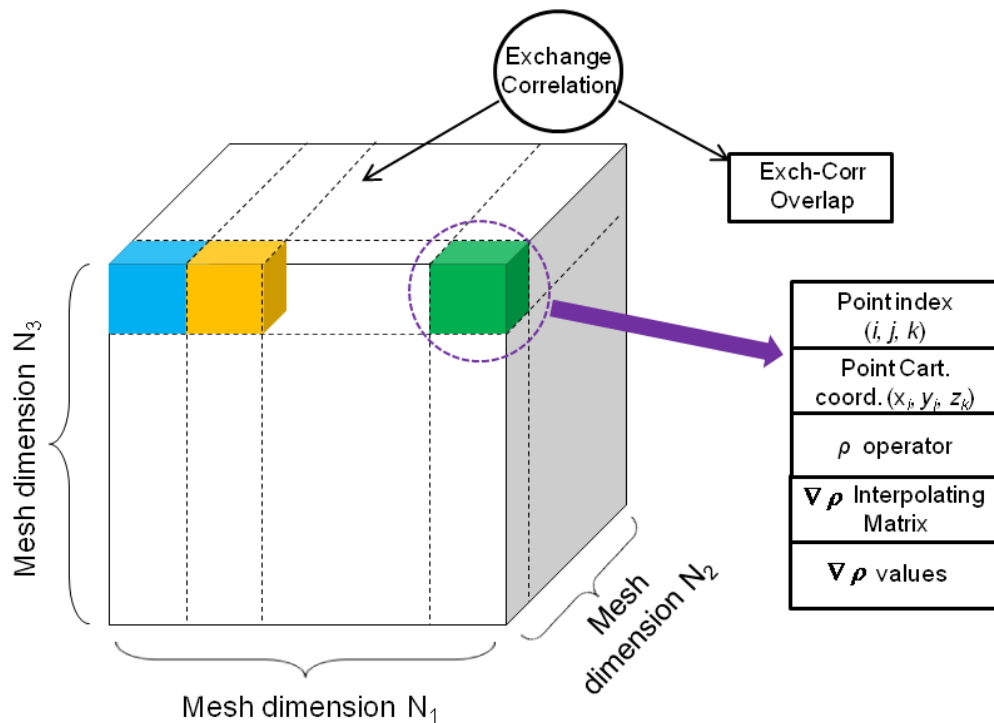


Figure 4. Potential alternative file storage system of XC group under GGA.

2.4.4 Computational Capabilities

Many electronic-structure properties can be constructed from the wave functions of the KS equation. The density of states (DOS), $G(E)$, depicts the electron state distribution as a function of energy:

$$G(E) = \frac{\Omega}{(2\pi)^3} \frac{d}{dE} \int d\vec{k} = \frac{\Omega}{(2\pi)^3} \int \frac{dS}{|\nabla E|} \quad (2.38),$$

where Ω is the volume of the unit cell and the integral is over the constant energy surface in the \vec{k} space. Many methods have been devised to perform the integration over the BZ in the equation (2.38) accurately. The most accurate and widely used method is the linear analytic tetrahedron (LAT) method [74-76]. In OLCAO, the total-DOS (TDOS) can be resolved into partial components called partial-DOS (PDOS), according to different atoms, different atom orbitals, and even different spins in the spin-polarized calculations.

It is also instructive to define the fractional charge $\rho_{iA}^{n\vec{k}}$ for the i th orbitals of the A th atom in the normalized state $\Psi_{n\vec{k}}(\vec{r})$ according to the Mulliken's population analysis scheme [77]

$$1 = \int |\Psi_{n\vec{k}}(\vec{r})|^2 d\vec{r} = \sum_{iA} \rho_{iA}^{n\vec{k}} \quad (2.39),$$

$$\rho_{iA}^{n\vec{k}} = \sum_{jB} C_{iA}^{n\vec{k}*} C_{jB}^{n\vec{k}} S_{iA,jB} \quad (2.40),$$

where S is the overlap matrix between atoms A and B with orbitals labeled as i and j . The C values are the eigenvector coefficients of the n^{th} occupied state. Then the effective charge on each atom can be obtained by summing over the occupied orbitals,

$$Q_A^* = \sum_{n\bar{k}, occ} \sum_i \rho_{iA}^{n\bar{k}} \quad (2.41).$$

We can further derive the charge transfer properties by taking the difference between Q_A^* and the charge of neutral atom, which enable us to study the charge redistribution effect.

The bond order which quantitatively characterizes the strength of the bond between atom pair A and B can be defined as

$$q_{AB}^* = \sum_{n\bar{k}, occ} \sum_{i,j} C_{iA}^{n\bar{k}*} C_{jB}^{n\bar{k}} S_{iA,jB} \quad (2.42).$$

Compared to the purely geometric criteria using bond length and bond angle, it is far more accurate and concise to describe the bonding properties especially for the weak ones like hydrogen bonds which play an important role in biomolecules. A minimal basis set is usually sufficient for BO calculations because the Mulliken scheme is more effective when the basis is more localized.

The physical processes underlying x-ray Absorption Near Edge Structure (XANES) [78] spectroscopy can be described as follows: (1) the absorption of an x-ray photon by a core level of an atom induces the emission of a photoelectron; (2) the resulting core-hole is filled either by capture of an electron from another level accompanied by emission of a fluorescent photon or via an Auger process [79]. In XANES the fluorescent photon, Auger electron and an inelastically scattered photoelectron may be measured together with the

initial photoelectron. This enables XANES to probe the final state of the photoelectron in a bound state such as an exciton [80]. The effect of measuring fluorescent photons, Auger electrons, and directly emitted electrons is equivalent to sum over all possible final states of the photoelectrons, meaning that what XANES measures is the total joint density of states of the initial core level with all final states, consistent with conservation rules. When the x-ray photon energy resonantly connects a core level with a narrow final state in a solid, such as an exciton, readily identifiable characteristic peaks will appear in the spectrum. These narrow characteristic spectral peaks are usually called fingerprints.

To correctly describe the interaction between the core-hole and the excited electron, the supercell-OLCAO method [22, 81] has been developed. The spectra are computed in a two-step process starting with a standard SCF ground state calculation followed by an excited state calculation where one electron is removed from the targeted core orbital of a selected atom and is placed in the bottom of the conduction band. The second step results in changes of the electron density distribution hence a different set of converged wave functions. SCF iterations account for the excited-electron core-hole interaction. Then, dipole transitions that explicitly include the momentum matrix elements are computed according to Fermi's Golden rule between the targeted core orbital from the ground state calculation and the conduction band from the excited state calculation. The sufficiently large supercell ensures no core-hole-core-hole interactions between the adjacent cells. For more accurate determination of higher energy states, the EB basis set is the default choice for XANES calculation. By far, the supercell-OLCAO method has been successfully applied to a large

number of complex crystals, defects, surfaces, interfaces, and complex microstructures [65, 82-86].

2.4.5 Summary

The OLCAO program suite has undergone many evolutionary steps from generation to generation. Beyond implementing GGA XC functionals that has been discussed in the present work, some other enhancements and extensions that are being made or can be made are listed here. The versatility of the OLCAO method can thus be enhanced by defining an improved and more flexible basis set, implementing algorithms for the computation of Van der Waals forces, or by providing a framework for adding more complex XC functional (e.g. LSDA+U, hybrid functional, etc.), introducing practical schemes (e.g. configuration interaction, etc.) to address some of the drawbacks in the local density approximation of density functional theory, etc. Moreover, the improvement in computational efficiency, friendlier user interfaces, and easier communication with Third Party Software are also aims of the OLCAO developers [21].

CHAPTER 3

CASE STUDIES

3.1 Network Structure of Bulk Water

3.1.1 Introduction

Water is one of the most common and intriguing substances on Earth yet it is far from being fully understood. It is also the most important medium in biological systems. In addition to the liquid and vapor phases, there exist many phases of solid H₂O (see Figure 5 [87]) covering a wide range of temperatures and pressures that exhibit many interesting phenomena. All the properties of these phases are closely related to the local configurations of the water molecules as determined by their intermolecular interactions – especially their HBs [88]. The existence of a network structure of H₂O in the liquid phase is a fundamental issue that has attracted a lot of attention both experimentally and theoretically. Due to the variety of existing models of water, the limitations of experimental techniques to verify them, the need for precise criteria to define a HB, and the lack of unambiguous criteria to analyze them, there are conflicting opinions as to whether the molecules in water form a network structure primarily composed of tetrahedrally bonded H₂O or of one-dimensional filamentous structures [89-94]. Several models regarding the network characteristics have been proposed to interpret experimental neutron scattering and XANES spectral data [89, 95]. On the theoretical side, models with non-periodic clusters ranging from several to a few hundred molecules have been studied using different first principles computational methods including *ab initio* MD and DFT [96-98]. There have also been discussions that the ambient-condition

simulations using DFT functionals will rather describe a supercooled state that is overstructured compared to liquid water [99].

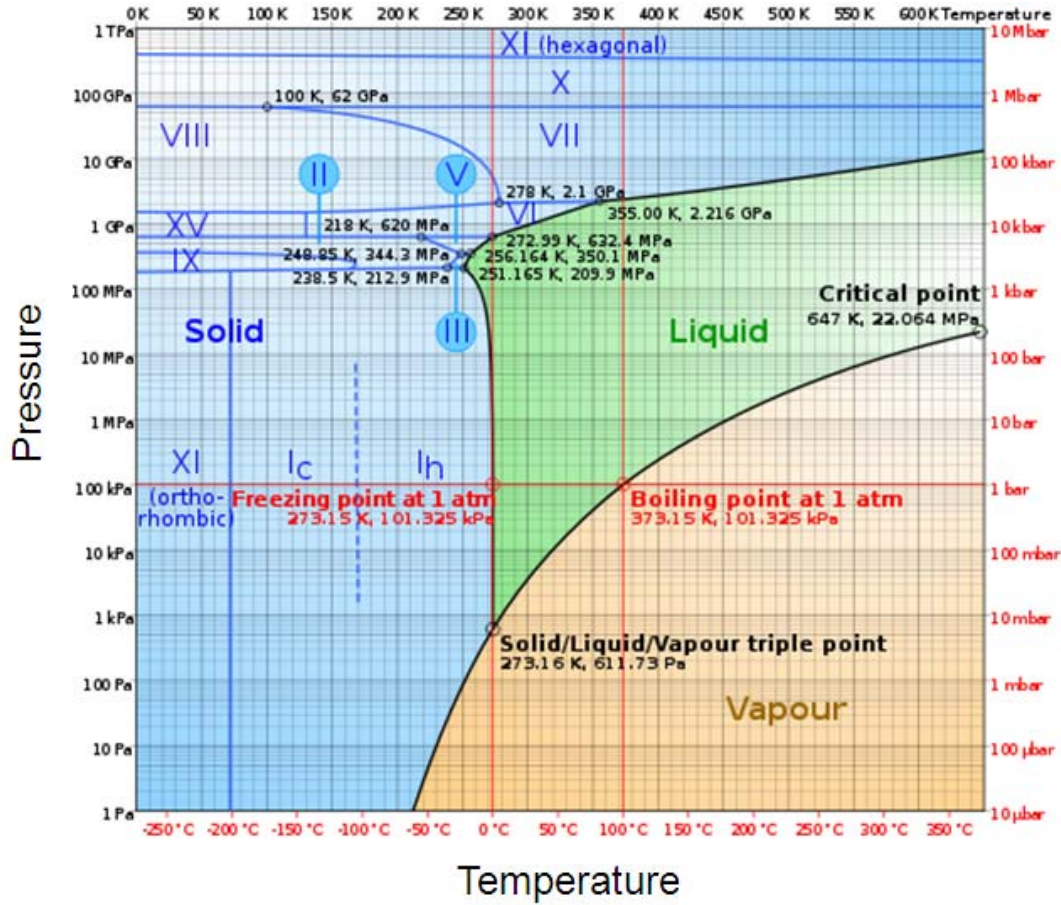


Figure 5. Water phase diagram [87].

3.1.2 Model and Computational Methods

To obtain a sufficiently large and valid model for bulk water, a combination of different methods was applied. The first step was to construct an initial model with periodic boundary conditions and 340 water molecules ($a = 21.6641\text{\AA}$, density: 1g/cc) using Packmol [100]. The use of periodic boundary conditions eliminates the artificial surface effects that would be present when using finite molecular clusters.

Next, we used the SIESTA code for *ab initio* MD optimizations of the molecular system. The model was equilibrated at 300K for 3000 fs. Snapshot configurations from the *ab initio* MD trajectory were selected and analyzed. Of these snapshots, the one with the most ideal configuration (identified by the absence of ionized species such as OH^- and OH_3^+ ions) was chosen and its structure was fully relaxed using the VASP with high precision to minimize the deviation of the O-H covalent bond lengths (BLs) and H-O-H covalent bond angle (BA) θ compared to the ideal H_2O molecule. The following parameters were adopted: (1) the PAW-PBE potential with GGA; (2) a high energy cutoff of 500 eV depending on different models; (3) an electronic convergence criterion of 10^{-5} eV; (4) a force convergence criteria of 10^{-3} eV/ \AA for the ionic relaxation; and (5) one k point at Γ which is fully justified due to the large cell size and the correspondingly small Brillouin zone. A ball and stick picture of the final relaxed model is shown in Figure 6.

The electronic structure calculations were then performed using the OLCAO method. An FB set of H ($1s, 2s, 2p$) and O ($1s, 2s, 2p, 3s, 3p$) was used to investigate the density of

states, while the bond order properties and XANES spectra were calculated using MB (H (*1s*) and O (*1s, 2s, 2p*)) and EB (H (*1s, 2s, 2p*) and O (*1s, 2s, 2p, 3s, 3p, 4s, 4p*)) respectively.

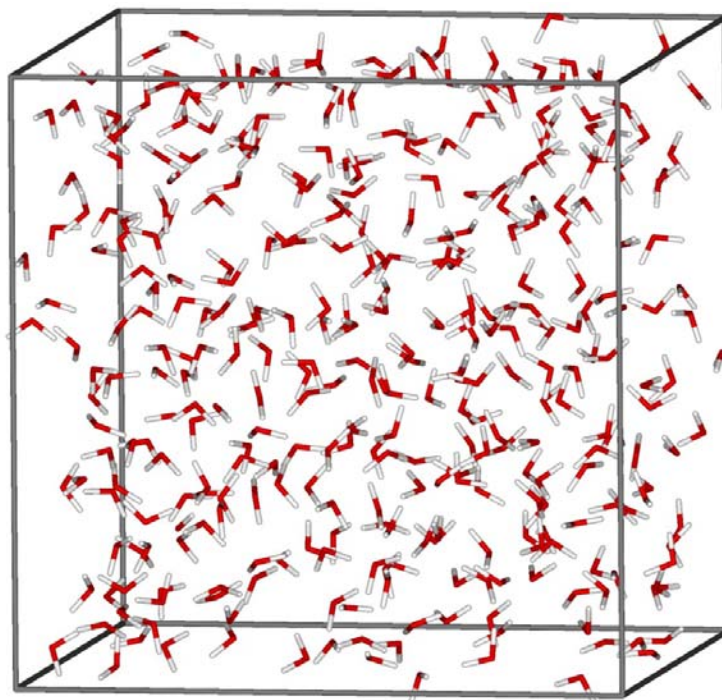


Figure 6. The periodic water model with 340 H₂O molecules.

3.1.3 Results and Analysis

Structural information

The distribution of values for the bond lengths and bond angles is very narrow (Figure 7), with small values for the standard deviations from the mean (1.00 ± 0.006 Å and $106.32^\circ\pm 2.528^\circ$ respectively). These values are very reasonable when compared to experimental data for liquid water [101]. When performing MD simulations where the atoms are treated as independent particles, it is not generally necessary to minimize the BA and BL distortions because they are usually regarded as fluctuations and tend to be inconsequential when averaged over many configurations. However, for *ab initio* electronic structure calculations based on a *single model* this is generally not possible. An exceptionally large or unrealistic distortion in one or a few water molecules can adversely affect the quantum mechanical calculation of the electronic structure by introducing gap states.

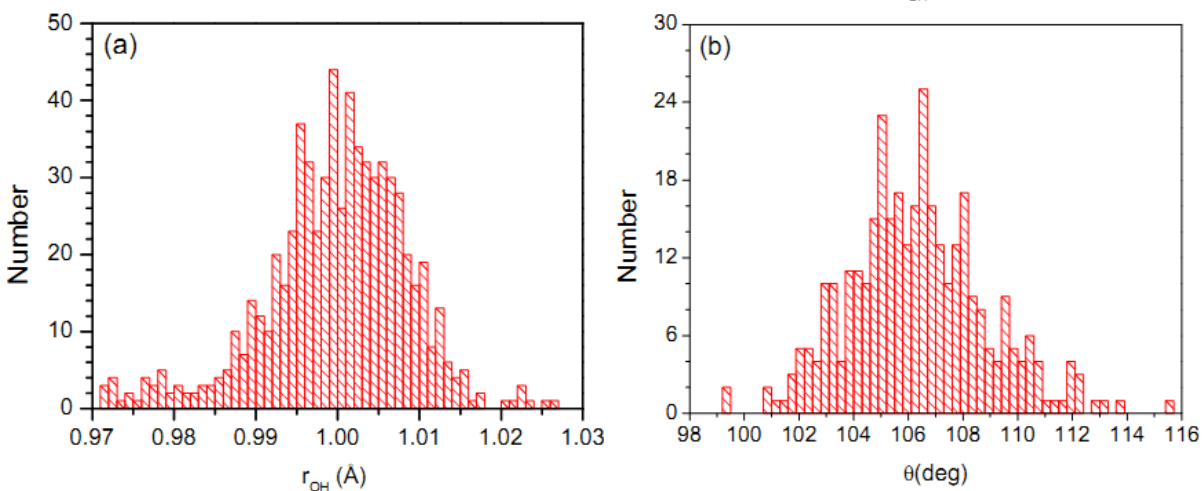


Figure 7. Distributions of (a) covalent bond length r_{O-H} and (b) bond angle θ ($\angle HOH$).

Figure 8 displays the calculated radial distribution functions (RDFs) for the O-H, H-H, and O-O pairs along with neutron diffraction data [95]. The peak positions are in good agreement although, as expected, the peaks in the theoretical curves are much sharper, which has been discussed in the work by Yoo *et al.* [99] as a result of the supercooled state caused by the ambient condition MD DFT simulations. When the temperature rises to a much higher level than the ambient condition, the peak height reduces and becomes comparable to the experimental intensity [99]. Moreover, our single model simulation does not include the kinds of time dependent broadening effects associated with experimental measurements. The RDFs from MD simulations [102] are generally much broader and smoother because they are averaged over many configurations. Investigating the electronic structure of *liquid-water* phase would be desirable in the sense of direct comparison with the experimental results at ambient conditions, but it is beyond the scope of the current study. We will discuss the issues related to the construction of *liquid-water* models, and suggest some potential key points in determining the micro-scale configurations.

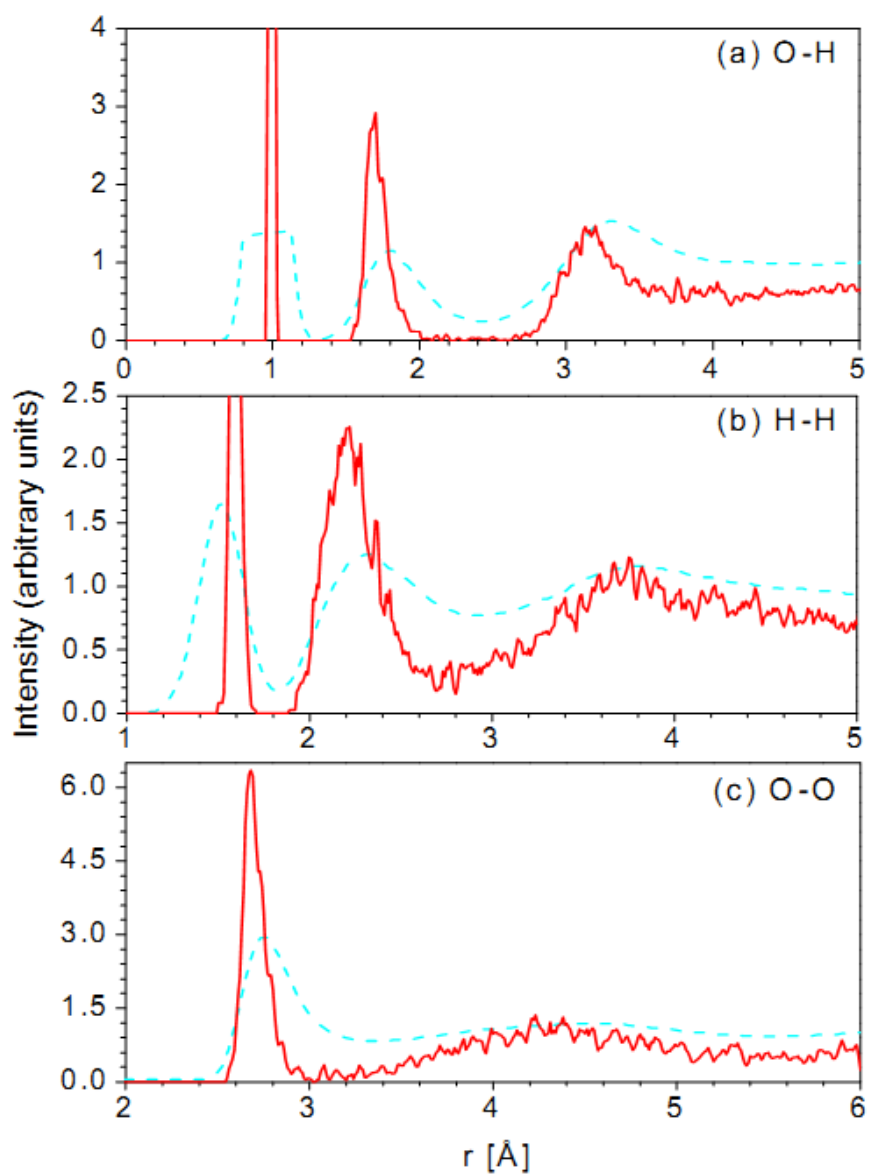


Figure 8. Radial pair distribution functions for the water model (red solid): (a) O-H; (b) H-H; and (c) O-O. Experimental data (cyan dash) is from ref. [95].

Density of states

The calculated DOS and PDOS plots are displayed in Figure 9. The model has a HOMO-LUMO gap of about 4.51 eV which is slightly lower than the values obtained by other DFT calculations using LDA and GGA [103]. The element-resolved PDOS shows the strongly covalent nature of the bonding between O and H in the H₂O molecule. It is characterized by a sharp peak at -1.15 eV from the O lone pair and two other peaks at -6.64 eV and -18.68 eV. These valence electronic structures are in good agreement with both the experimental photoelectron spectroscopy (PES) studies [104, 105], and the theoretical DFT work by Brancato *et al* [97].

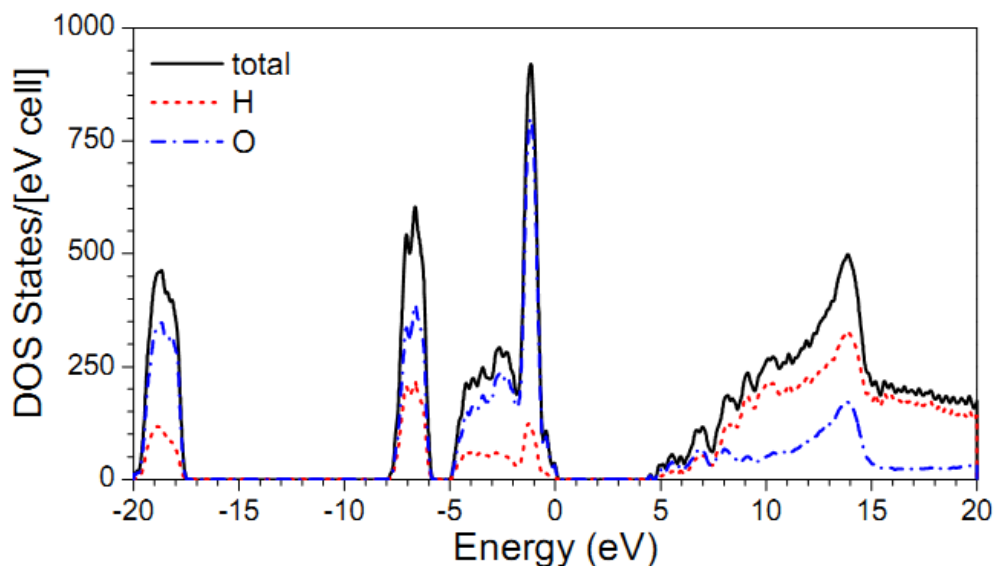


Figure 9. Calculated DOS (black solid) and PDOS (blue dash dot: O; red dash: H) in water.

Hydrogen bond properties

In a water model, HBs (represented as O...H) are much weaker than the O-H covalent bonds, yet they are sufficiently strong to dictate the intermolecular interaction. There have been several criteria proposed for defining a HB. The effective pair potential scheme uses the local minimum in the potential curve as the upper limit of H-bonding energy [106]. The HBs can also be defined by geometric cutoffs using different sets of atomic configuration parameters [89, 107]. The near-neighbor hydrogen and oxygen (NNHO) criterion can be considered as an upgrade of the geometric cutoffs because it is less dependent on rigid arbitrary atomic distances [108]. Finally, the quantum mechanical bond order (BO) values can reflect the stability differences between different kinds of bonds and therefore can help to distinguish HBs [109].

Figure 10 displays our results on the BO distribution of (O, H) pairs with $R_{OH} < 3.5\text{\AA}$. The distribution range can be divided into three regions. Region I is extremely narrow, with many very small BO values, the atoms in this region are considered to be nonbonded. In region III, the distribution has a narrow peak centered at 0.27 originating from covalent O-H bonds within the H₂O molecules. The HB distribution is in region II. It is well separated from the covalent bonding above and it shows a clear minimum at 0.015 (marked with an arrow) separating it from region I. According to this distribution, the HBs are the O...H pairs with BO values in the range of 0.015 to 0.100 centered approximately at 0.05. The lower end of the distribution constitutes the weak HBs, and the high end of the distribution represents the stronger HBs. We have further checked the intermolecular O-H distances of atoms associated with BO values that are less than 0.015, and found that all of them are larger than the well-

accepted HB separation of 1.98\AA in liquid water. In region II, only three HBs have O-H separations that are slightly longer than 1.98\AA . This is because the BO value does not simply scale with the distance of separation but is also affected by the bond angle and presence of other nearby atoms.

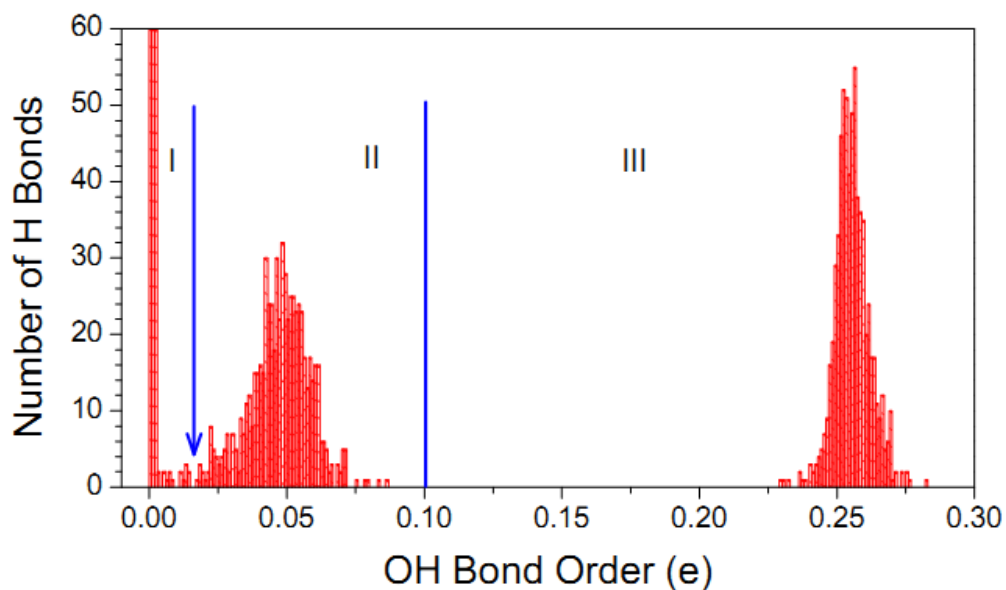


Figure 10. Distribution of all O and H bonds in the water model: region I: nonbonded; region II: H-bonded; region III: covalently bonded.

Figure 11 illustrates the HBs between a H₂O molecule and its two nearest neighbor molecules, (H₂O)' and (H₂O)". A water molecule (H₂O) can play two roles in hydrogen bonding, the donor D (with reference to (H₂O)') and the acceptor A (with reference to (H₂O)"), depending on which atom is involved in the H bonding. According to our calculated BO values for all the atoms in the model, the 340 water molecules can be divided into four groups, with two, three, four, and five HBs, in which both the donors and acceptors are counted. The group with four HBs (two donors and two acceptors or 2D2A) is the fully bonded group, which accounts for 85% of all the HBs. The four HBs in these molecules all tend to have similar strength. Other groups with two (1D1A and 2D0A, 3.2%) or three (2D1A and 2A1D, 10.3%) HBs have a much smaller presence and are usually identified as broken bonded. The smallest group (1.5%) with five HBs (2D3A) is called the overbonded group, and it has relatively weak HBs. The average HB number per molecule is 3.85 and this appears to support the tetrahedral HB network model in bulk water.

The distributions of the groups with two, three, and five HBs are shown in Figure 12. For the two-HB group, the centroid is near the middle with a BO value close to 0.05. For the three-HB group, it shifts to the left indicating a weaker average HB strength for that group. For the five-HB group, they also tend to be at the region of much smaller BO values. We have compared these electronic-structure results with other snapshot configurations obtained from different steps along the equilibrium trajectory of the *ab initio* MD simulation and found that all of them have similar percentages for the four groups. This finding is in contrast with another proposed network model consisting of filamentous H₂O chains that are formed

by H_2O molecules, with each having two strong HBs each. The chains are said to interact with a surrounding disordered cluster network through weak hydrogen bonding [89].

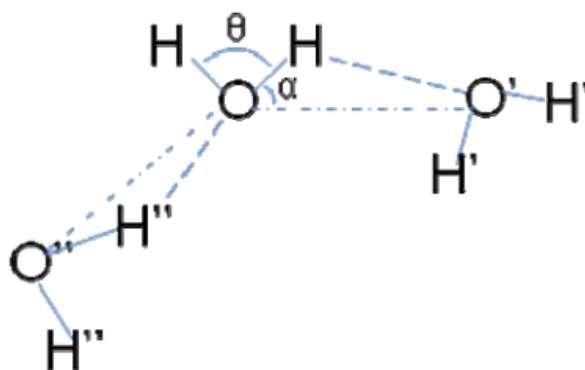


Figure 11. Schematic illustration of donor and acceptor H bonds between three water molecules H_2O , $\text{H}'_2\text{O}'$, and $\text{H}''_2\text{O}''$.

Table 1. Portion of four groups with different HB numbers.

No. of HB	No. of molecule
2	11 (3.2%)
3	35 (10.3%)
4	289 (85%)
5	5 (1.5%)

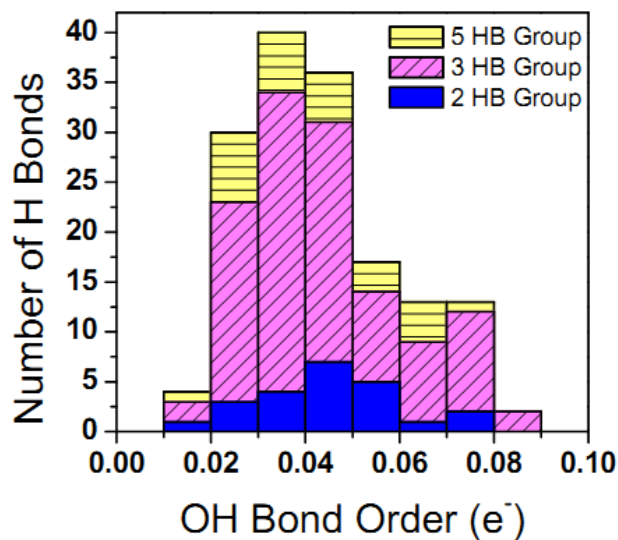


Figure 12. Distributions of the H bonds for the three groups: two-HB (blue); three-HB (pink); and five-HB (yellow).

To have a better understanding of the relationship between HBs and the intermolecular configurations and also to compare with other HB criteria, we plotted the intermolecular parameters $r_{O-O'}$ and α (see Figure 11) of each HB in Figure 13. Wernet et al. [89] suggested a geometrical cutoff of $r_{O-O'} = -0.00044 \alpha^2 + 3.3 \text{ \AA}$ [curve (i) in Figure 13] to show which parametric combinations permit hydrogen bonding. The discrete data from our model stay well under this curve, with the majority in the range of $r_{O-O'} = 2.6\text{-}2.8 \text{ \AA}$ and $\alpha = 1^\circ\text{-}15^\circ$. This indicates that the parabolic curve from Ref. [89] is not stringent enough in describing the condition of HB formation. It is highly unlikely for HBs to form with $r_{O-O'}$ larger than 3.1 \AA and α greater than 28° . The scattered plot in Figure 13 also shows that the most dense population of HB distribution is around $r_{O-O'} = 2.7 \text{ \AA}$ and $\alpha = 8^\circ$ and that there is no explicit relation between the number of HBs for a molecule and its $r_{O-O'}$ or α . From the

discussion above, we can see that BO criterion is a simple and accurate way to define the presence of HBs.

We have also attempted to identify statistical correlations between the number of HBs for a molecule with the intramolecular parameters $r_{\text{O-H}}$ and θ ($r_{\text{O-H}} = (r_{\text{O-H1}} + r_{\text{O-H2}})/2$; $\theta = \angle\text{HOH}$ from Figure 11). In Figure 14, each data point represents the parameters for one water molecule and they are plotted in accordance with the four groups categorized above. The average $r_{\text{O-H}}$ of the broken-bonded HBs (groups i and ii) are mainly distributed in the range of shorter BLs while the full- and over-bonded ones (groups iii and iv) occupy the longer-range region. This is reasonable since the molecules with shorter $r_{\text{O-H}}$ tend to be more charge localized. It is also clear that the fully bonded molecules have a wider distribution in θ , covering both the larger and the smaller ends.

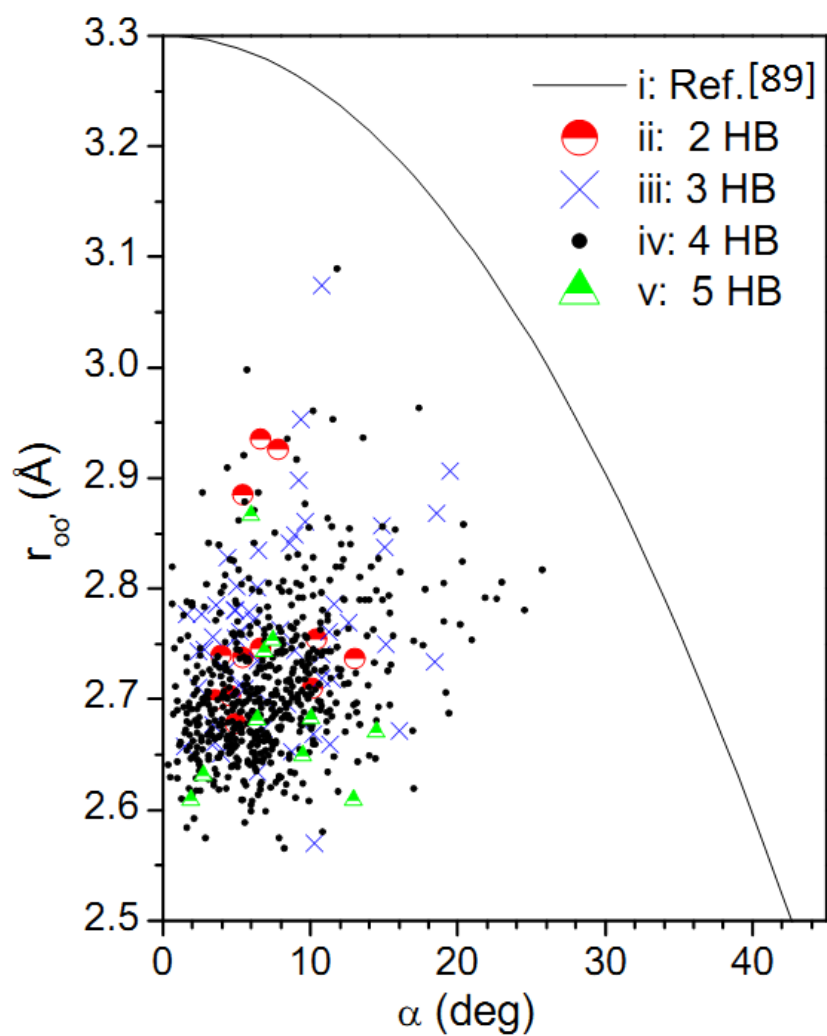


Figure 13. Scattered plot of $r_{O...O'}$ vs α ($\angle O'-OH$) for all H Bonded O...H pairs. Solid Line: HB definition curve from ref. [89].

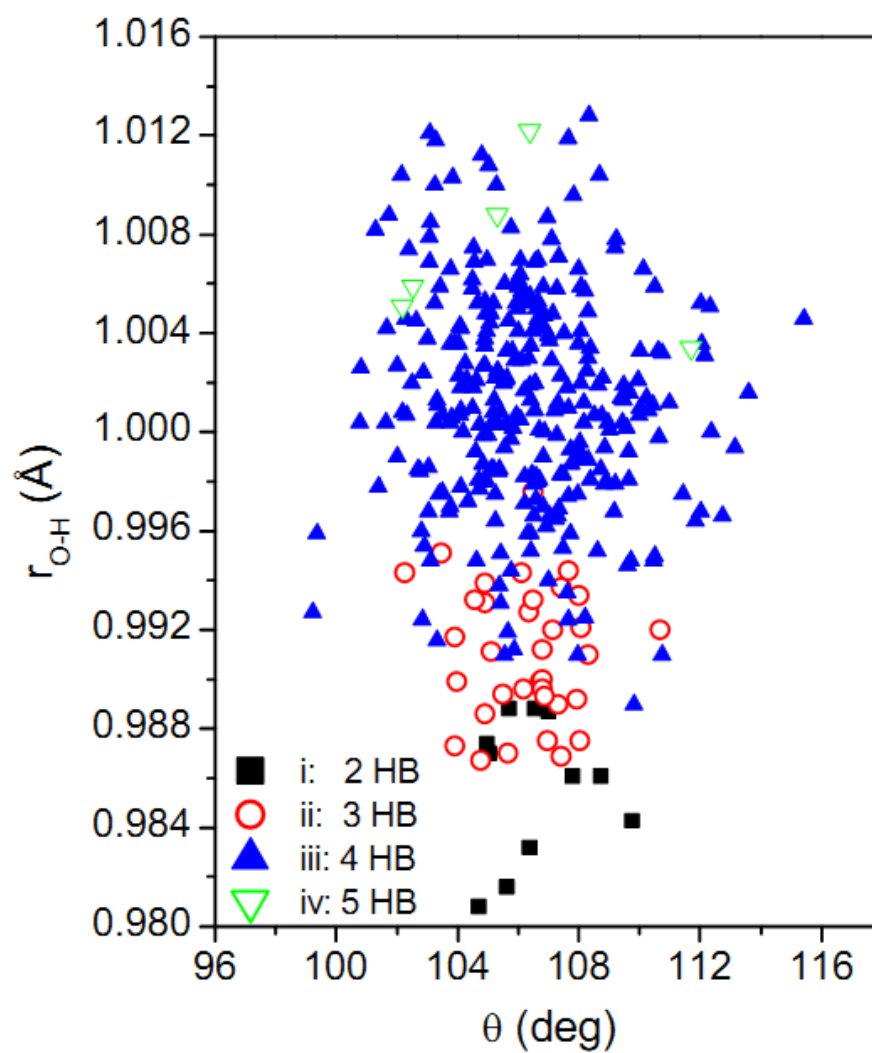


Figure 14. Scattered plot of the average O-H covalent bond length $r_{\text{O-H}}$ vs covalent bond angle θ .

XANES spectroscopic properties

The O *K* edges for *all* 340 oxygen atoms in this large model are obtained using the supercell-OLCAO method. The data provides us with a sufficiently large sample of the XANES spectra to investigate meaningful correlation of spectral features with the HB structure of a model closely related to bulk water. Figure 15(a) shows the experimental x-ray Raman scattering (XRS) [89] and x-ray absorption spectra (XAS) [110] together with our calculated O *K* edge averaged over all single spectra. The calculated total spectrum is shifted by 5.99 eV to the left to align with the main peak in the experimental curves. Clearly, our total spectrum shows a pre-edge peak at ~535 eV, a shoulderlike structure near 537 eV, a main peak at 538 eV, and a post-edge broad peak between 540 and 541 eV. These features are in very good agreement with experimentally measured spectra [89]. The presence and reproduction of the shoulderlike structure at 537 eV is important and this feature has been routinely neglected by others in discussing their research results.

For better interpretation, the 340 O *K*-edge spectra were divided into four groups according to the HB numbers and they are shown in Figure 15(b). It can be seen that the pre-peak originates predominately from the group with two HBs with no contribution from the overbonded group with five HBs. This observation is consistent with other findings on the relation between the pre-peak intensity and broken HBs [111]. On the other hand, the post-edge region is more affected by the fully bonded molecules. Figure 15(a) also shows that the ratio between the pre-peak and main peak is larger for the experimental curve than for the calculated one. This could be attributed to several factors. First, because of the periodic boundary conditions, the model in this work represents the bulk water with no surface

dangling bonds. Therefore it is reasonable to have a smaller portion of broken-bonded water molecules and a higher number of HBs per molecule. Second, the theoretical curve is the superposition of many individual curves with different edge onsets which tends to smooth out the sharp features.

This is illustrated in Figure 16 where we display the eleven O *K*-edge spectra for the group with two HBs, or 3.2% of the total 340 spectra. As can be seen, all have a strong pre-peak but they have different edge onsets, and when added together the resultant is a less prominent pre-peak. The difference in their edge onsets is related to their different local geometries in addition to their HB features. Because the position of the edge onset is determined from the difference in total energy of the ground state and the excited state, it is conceivable that a more accurate determination of those energies would slightly reduce the variation and then increase the strength of the pre-peak in the total spectrum. In Figure 17 we randomly selected O *K* spectra (molecule number: 39, 328, 201, 255) from each of the groups among the 340 spectra calculated. Generally, as the HB numbers increase the observed trend is that the pre-edge peak height decreases and the post-edge height increases.

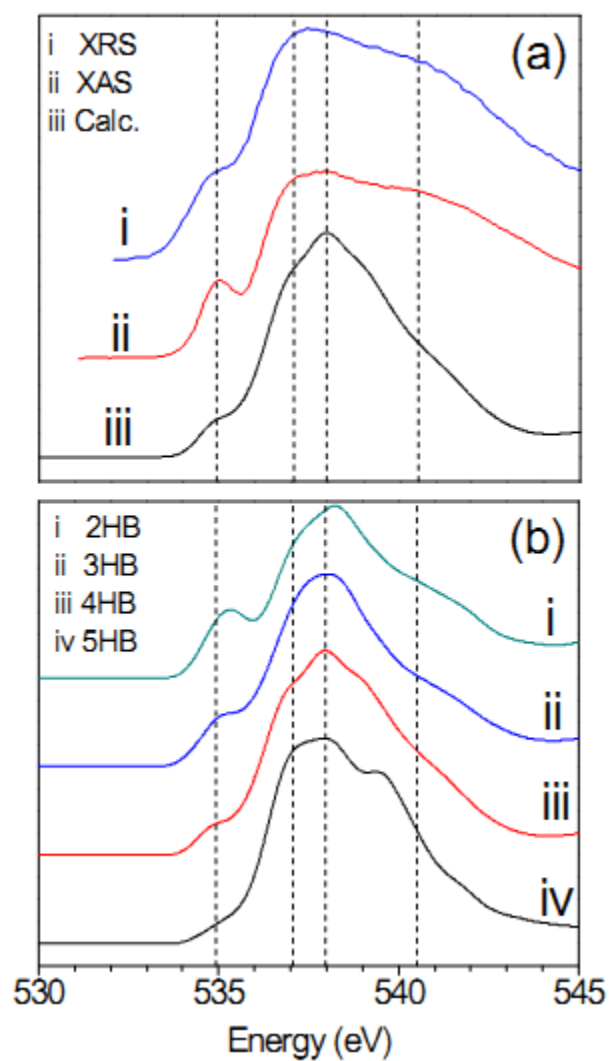


Figure 15. (a) O *K*-edge XANES spectra: experimental (i) XRS at 25°C [89]; (ii) XAS at ambient conditions [110]; (iii) averaged spectra over 340 molecules from present calculation. (b) Calculated total spectra for the four categorized groups with two, three, four and five HBs.

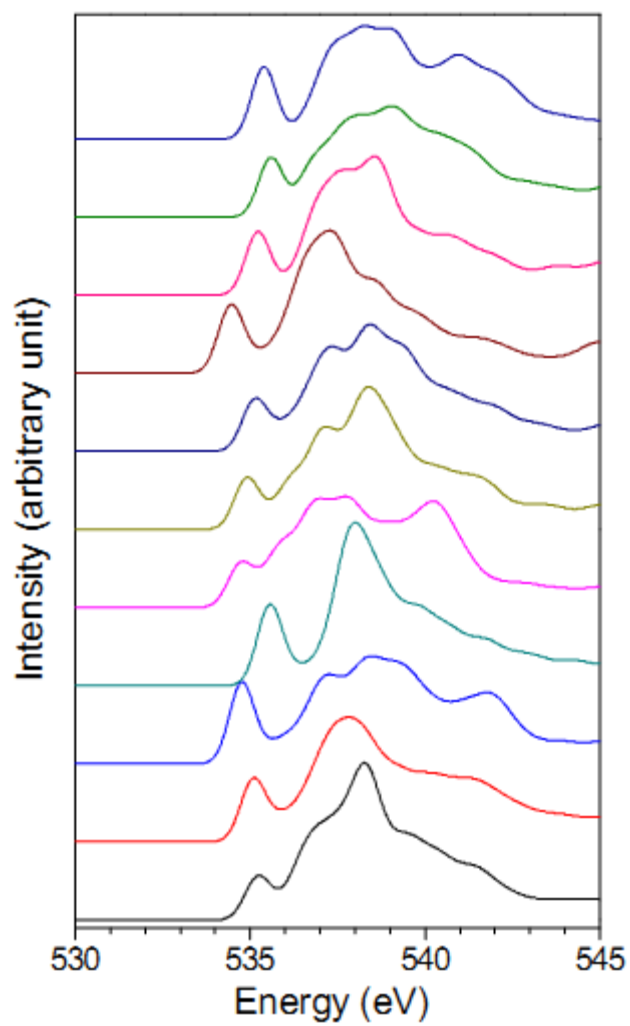


Figure 16. All eleven individual O *K*-edge spectra in the two-HB coordinated group from calculation.

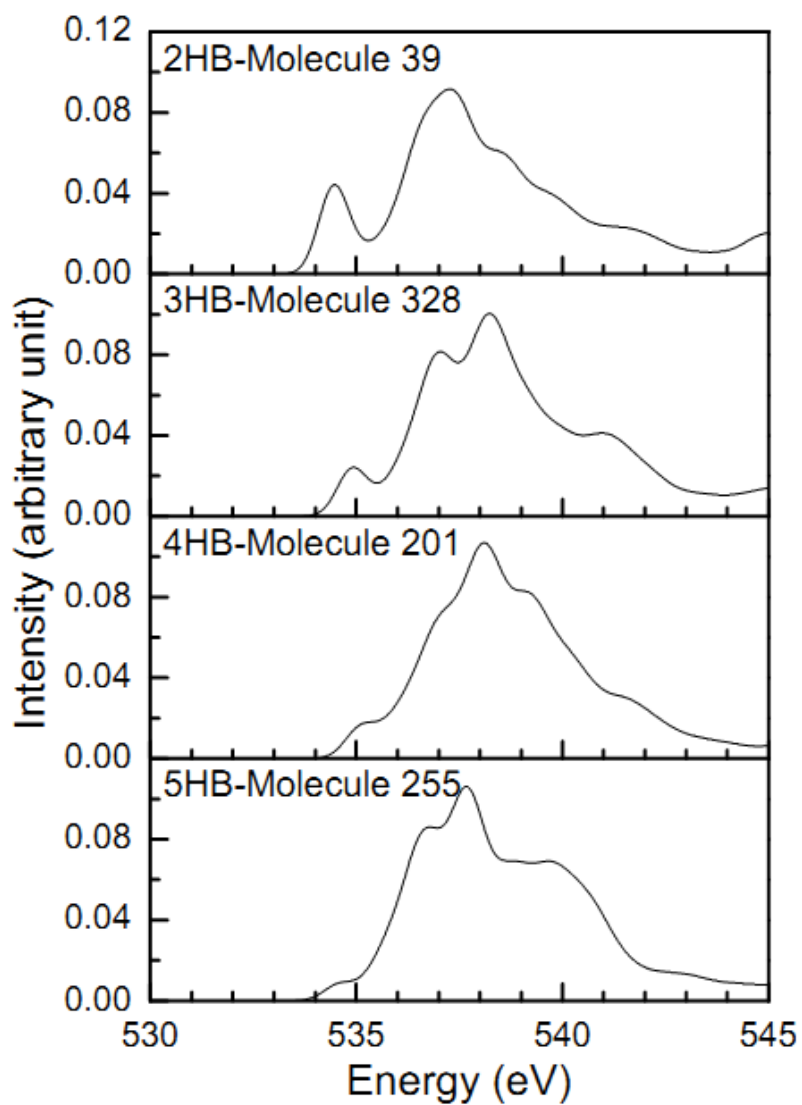


Figure 17. A typical representative spectrum form each group with two to five HBs.

3.1.4 Discussions

No doubt the validation of all statistical analysis depends on the structure. Beyond the 340-H₂O model fully relaxed using SIESTA and VASP (corresponding to Table 2 Model-2), we have built several other periodic water models and studied their electronic structure properties (Model-1, 3-5). The initial structure (IS) of Model-1 obtained from Packmol is equilibrated using SIESTA, and then certain snapshot without dissociated water molecules is picked up from the MD trajectory as the final structure (FS) for OLCAO calculations. Model-1 can be viewed as the preliminary stage of Model-2. By keeping the oxygen atoms at their original positions and the normal vectors of H₂O molecule planes in the same direction, we have adjusted the covalent BAs and BLs of all H₂O molecules according to the experimental values to obtain the IS of model-2, which is referred to as idealized water model. The VASP relaxation gives us the FS of model-2. Model-3 consisting of 800 H₂O molecules is constructed from a totally different way using MD packages by our collaborator. We did the idealization process on Model-3 in the same way as Model-2 to get the IS of Model-4, and then performed the VASP relaxation on it to get the FS. The structure of Model-5 is derived in an even more interesting way. It is based on an amorphous SiO₂ model in which all Si atoms are tetrahedral-bonded to four neighboring O atoms. This network structure has some similarity with water models. By replacing Si with O, and O with H, adjusting the BAs and BLs, and performing the VASP relaxation, we obtained the FS of Model-5. Generally speaking, Model-1 and Model-3 are snapshots directly taken from the MD trajectory, while the others further go through VASP relaxations.

In Figure 18 and 19, the distributions of the covalent BL and BA reveal how the simulation techniques could affect the atomic structures. Conspicuously, the MD snapshots without idealization have much more spread distributions of BL and BA, which turn to be disadvantages in XANES calculations according to our experiences. For example, the averaged total O *K*-edge XANES spectrum of Model-1 does not show any pre-peak or shoulder features compared to Model-2 and the experimental curve (see Figure 20). As has discussed before, the peak features are very sensitive to the intermolecular HB network structure, this may imply that the idealization process reduces the disorder of the system, which facilitates the *ab initio* MD relaxation using VASP to optimize the water model with large quantity of HBs, and enhances the features of those four groups with different number of HBs.

Figure 21 shows the HB BO distributions of the five distinct models. First of all, we see that Model-1 and Model-2 do not differ much on the range or maximum values of BO except that Model-2 has a heavier weight of stronger HBs. Further comparison of the portion of the two to five HB-bonded groups did not show much difference. This indicates the VASP relaxation does not significantly modify the HB connections. For Model-3, the majority of HBs are distributed in the weak-bond range compared to the rest four, and it is not very easy to set the lower bound of HB criterion based on this distribution. Possible interpretation may be that the average covalent BL of Model-3 is relatively shorter such that the electrons are more localized to the water molecules, rendering the intermolecular HBs less strong. Due to the limitation of computing resources, we did not calculate the O *k*-edge XANES spectra of

any 800-H₂O model. In the future, it would be desirable to categorize the O atoms according to their HB numbers for Model-3, since it has a distinct HB BO distribution. With those information, we can estimate what kinds of peak features will exhibit in the final averaged spectrum in order to further verify our conclusion on the relation between HBs and XANES.

Table 2. Water models constructed and studied with different methods.

Model	No. of H ₂ O	a(Å),b(Å),c(Å) α, β, γ	Initial Structure (IS)	Relaxation Method	Final Structure (FS)
1	340	a=b=c=21.66 Å $\alpha=\beta=\gamma=90^\circ$	Built by Packmol	SIESTA	Snapshot from <i>ab initio</i> MD trajectory
2	340	a=b=c=21.66 Å $\alpha=\beta=\gamma=90^\circ$	FS of Model 1 w/ H ₂ O BL & BA idealized	VASP	<i>ab initio</i> MD fully relaxed
3	800	a=b=c=28.815 Å $\alpha=\beta=\gamma=90^\circ$	From collaborator	MD package	MD equilibrated
4	800	a=b=c=28.818 Å $\alpha=\beta=\gamma=90^\circ$	FS of Model 3 w/ H ₂ O BL & BA idealized	VASP	<i>ab initio</i> MD fully relaxed
5	432	a=23.75 Å b=22.17 Å c=24.60 Å $\alpha=90.46^\circ$ $\beta=94.04^\circ$ $\gamma=90.77^\circ$	Amorphous SiO ₂ model	VASP	<i>ab initio</i> MD fully relaxed

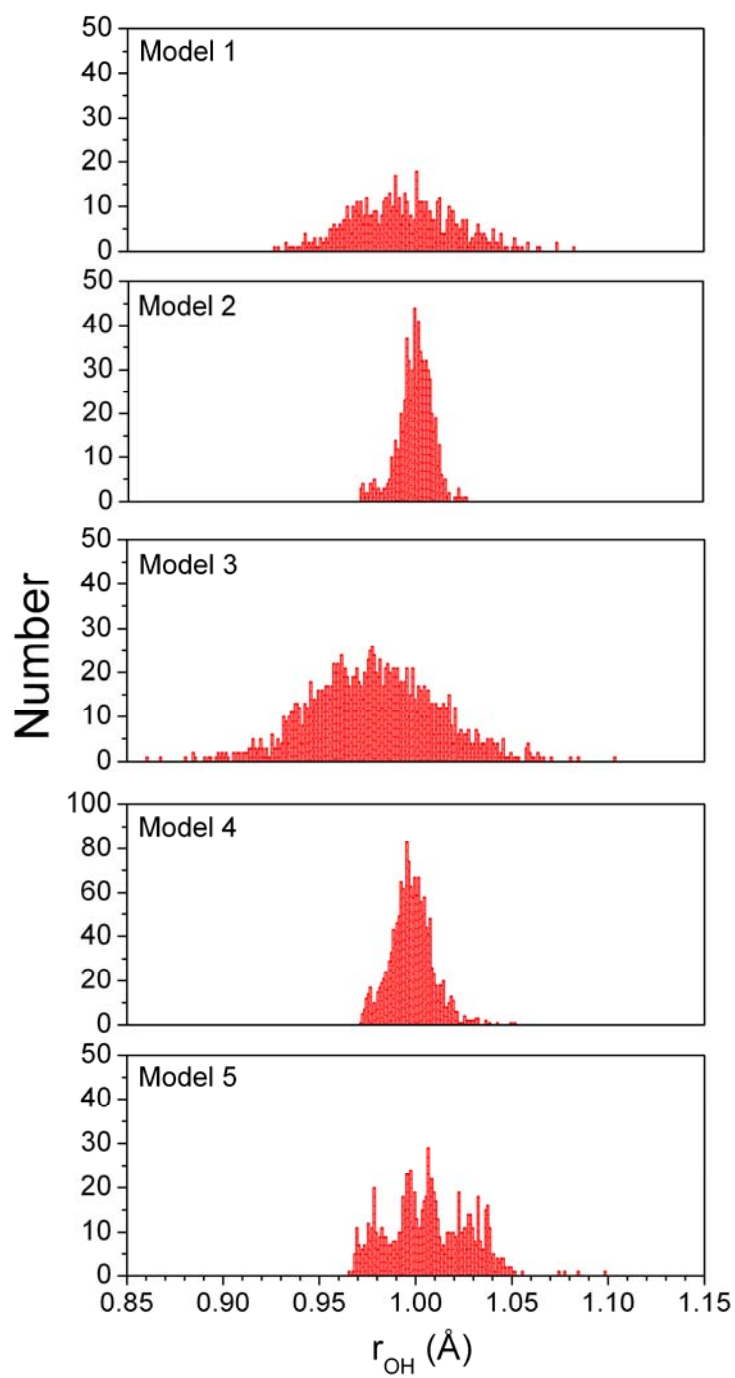


Figure 18. Comparison of the distributions of the covalent bond length r_{OH} of Model 1-5.

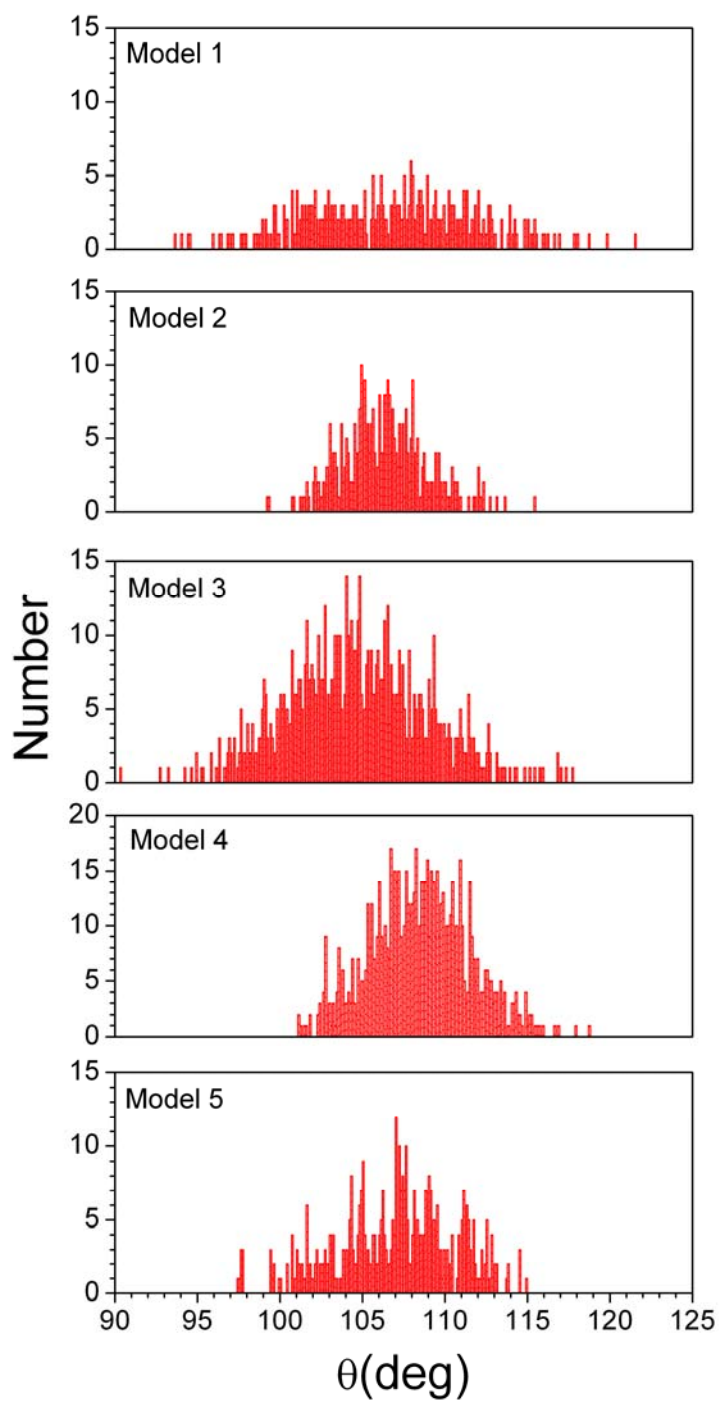


Figure 19. Comparison of the distributions of the covalent bond angle θ of Model 1-5.

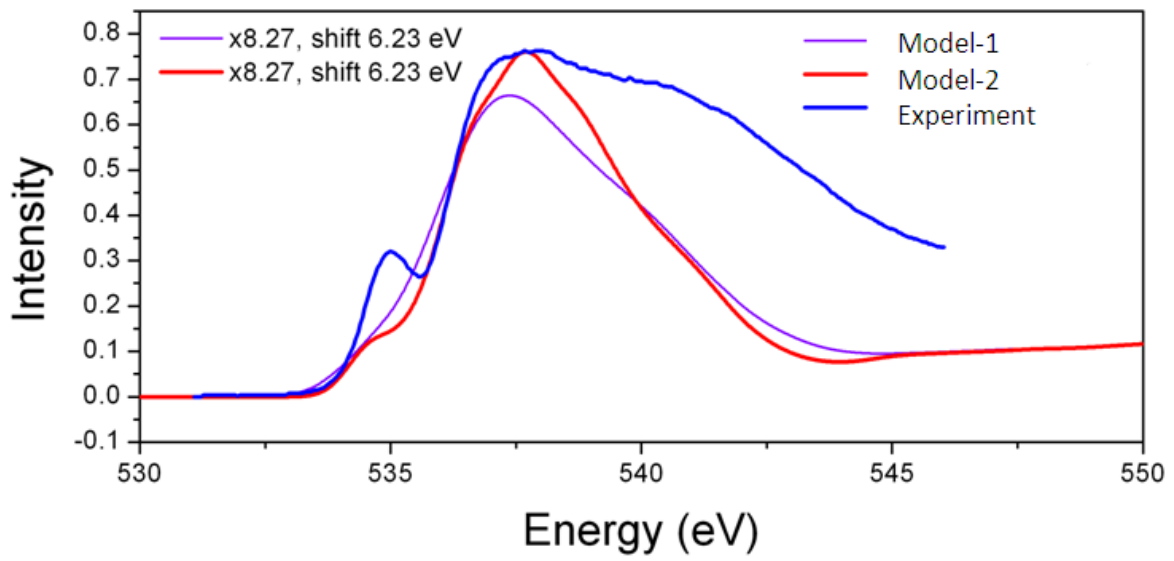


Figure 20. Total averaged O *k*-edge XANES spectra of Model-1 (purple), Model-2 (red), and experiment (blue) [110].

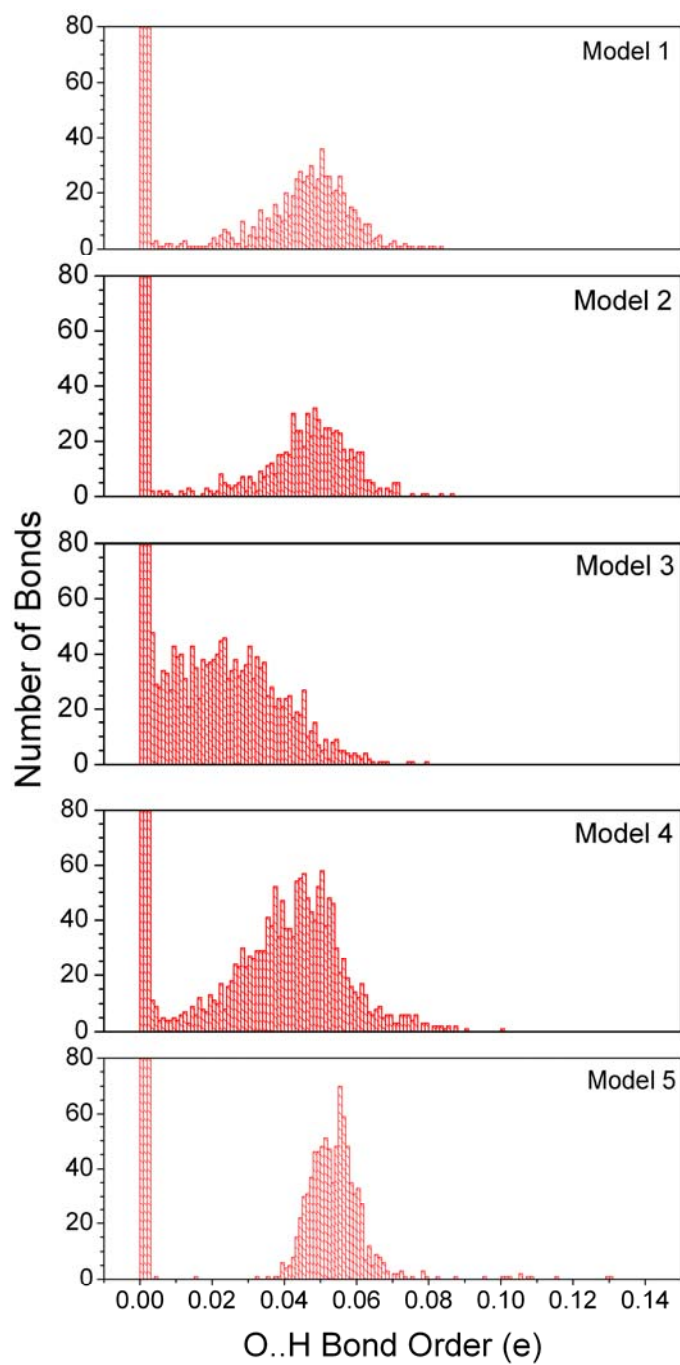


Figure 21. Comparison of the distributions of the O..H HB bond order of Model 1-5.

3.1.5 Conclusions

Based on the detailed *ab initio* calculation of a model to mimic bulk water and the use of quantitative bond-order values to define HBs, we found that tetrahedrally coordinated water molecules dominate with an average of 3.85 HBs per water molecule. The distances (r_{O-O}) and angles ($\angle O'OH$) that characterize the HBs are narrowly confined in the range of 2.6-2.8Å and 1°-15°, respectively. Large-scale non-*ab initio* calculations could be guided by this more stringent criterion when interpreting their results. The calculated O *K*-edge XANES spectra correctly reproduce the peak features observed in the experiments, and the decomposition of the spectra according to HB numbers confirms that the pre-peak is caused by H₂O molecules with fewer than four HBs, while the main- and post-peak structures are determined by the local geometry of the individual molecules and their surroundings. This project clearly demonstrates that the *ab initio* BO criterion used to define weak hydrogen bonding is accurate and efficient, and can be applied to complex liquids, cement hydrates, biomolecular systems, etc. Further, the accurate calculation of many XANES spectra, which are very sensitive to the variations of local atomic environment, can be used to elucidate subtle points of the structural properties of complex materials.

3.2 Mechanism of DNA Overstretching Deformation

3.2.1 Introduction

The structure of DNA molecules has been well known to people. A DNA double-helix strand consists of two long polymers which run in opposite directions to each other. The strands are formed by simple units called nucleotides (see Figure 22). A nucleotide is then formed by one phosphate group, one sugar (deoxyribose), and one of the four types of nucleobases: Adenine (A), Guanine (G), Thymine (T), or Cytosine (C). The sugar and phosphate groups are joined by ester bonds to form the backbone of DNA. The nucleobases on one strand will pair with those on the other through hydrogen bonds. One A-T pair forms two HBs, while one G-C pair forms three HBs.

In nanobiotechnology, DNA has been viewed as a functional material far more than a genetic information carrier. The deformation of the double-helix structure involved in many important biological and bioengineering processes has attracted a lot of research interests. Through direct single-molecule manipulation, the *force–extension* curve shows some typical features of the DNA overstretching process (Figure 23) [112-114]. It usually begins with an “entropic elasticity” plateau where force f of several piconewtons (pN) straightens the knotted DNA. As the DNA extension approaches its full contour length (L_0), the force begins to grow steeply, reflecting the elastic response of the stretched DNA. At $f \sim 65$ pN, the DNA undergoes a significant extension to ~ 1.7 -fold ($1.7 \times L_0$), resulting in a wide plateau that is often attributed to the transition from the classical Watson and Crick’s B-DNA [115] to the so-called S-DNA [116]. The structure of S-DNA — usually depicted as a partially unwound

ladder with base pairing intact (Figure 23 inset) — has not yet been directly observed, but it offers reasonable interpretations to some experimental phenomena [117, 118] that will be discussed shortly, and has also obtained supporting evidence from MD simulations [119]. Beyond the 65 pN plateau, the steep force growth resumes, presumably reflecting the elastic response of S-DNA. Some short-lived intermediate state, such as molten-DNA (M-DNA) [120, 121] in which the base pairing are destroyed, could exist in this stage. Eventually, another plateau at $f \sim 115$ pN appears, corresponding to the force-induced melting which causes the formation of single-stranded DNA (ssDNA) [114].

The number, width, and height of the plateaus in the *force–extension* curve may vary depending on the temperature [122], the DNA sequence [117], the number of nicks in the DNA strands [122-124], solvent pH [122], salt concentration [122], torsional constraints [113], pulling speed, and the stiffness of the AFM cantilever [125]. For example, if the DNA is torsionally constrained (3'5'-5'3' geometry attachment as opposed to 3'-3' (torsionally unconstrained, see Figure 24(a, b)), the S-DNA state may conceivably be skipped, and the overstretched DNA will melt from the nucleation points in the phosphodiester backbone [113]. In the curve, the steeply rising portion can be well fitted using the worm-like chain (WLC) model [126] or freely-joined chain (FJC) model [127] based on the polymer theory (see Figure 24(c)) [125]. However, there still lacks of a clear interpretation of the molecular mechanism underlying the plateau behavior.

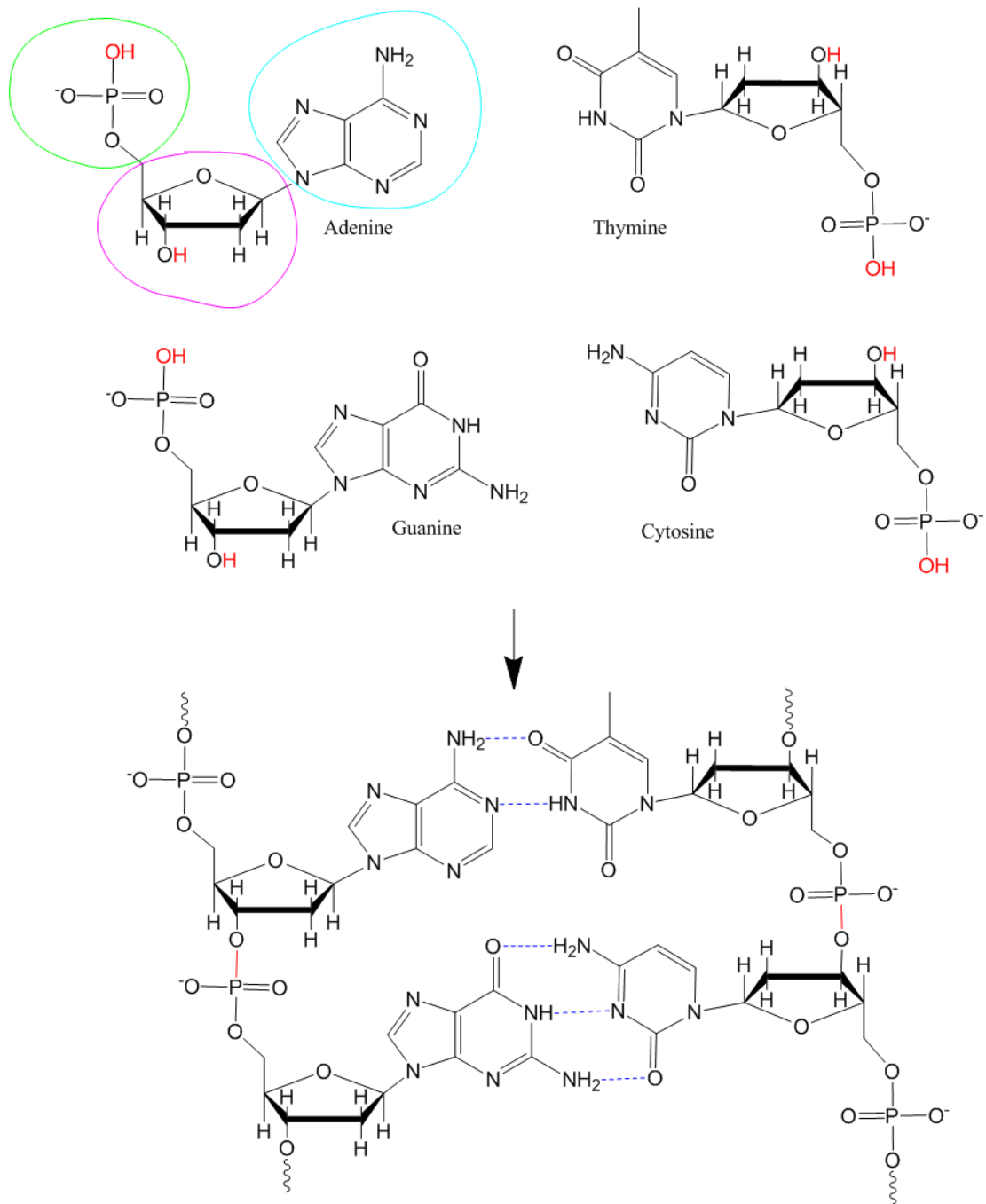


Figure 22. Four nucleotides and bonding formation in double-helix structure.

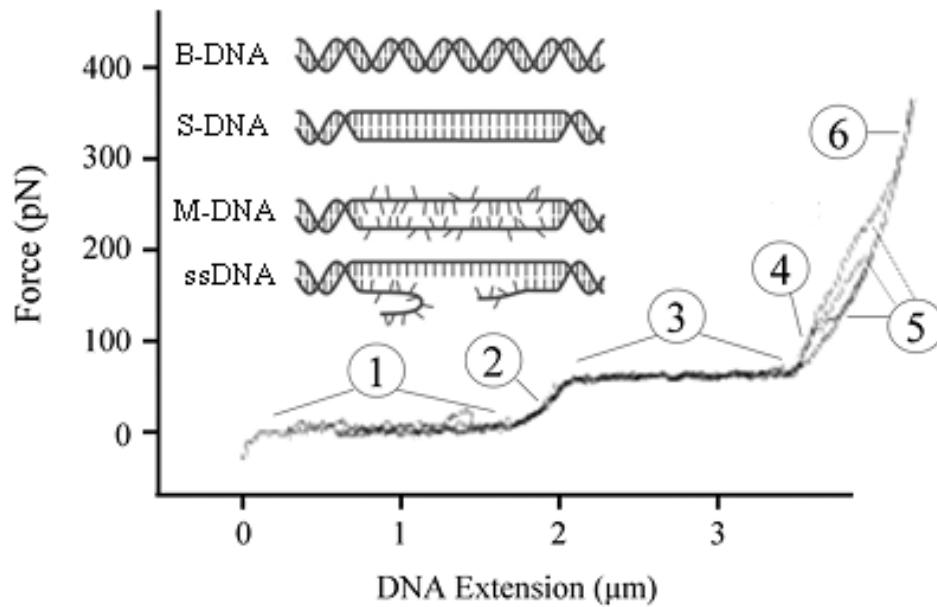


Figure 23. Labeled sections: (1) the entropic elasticity plateau, (2) elastic extension of B-DNA, (3) the B-S transition plateau, (4) elastic extension of S-DNA, (5) the DNA melting plateau, and (6) elastic extension of single-stranded (melted) DNA (ssDNA) (from ref. [112, 114], inset from ref. [113] Figure 1(c)).

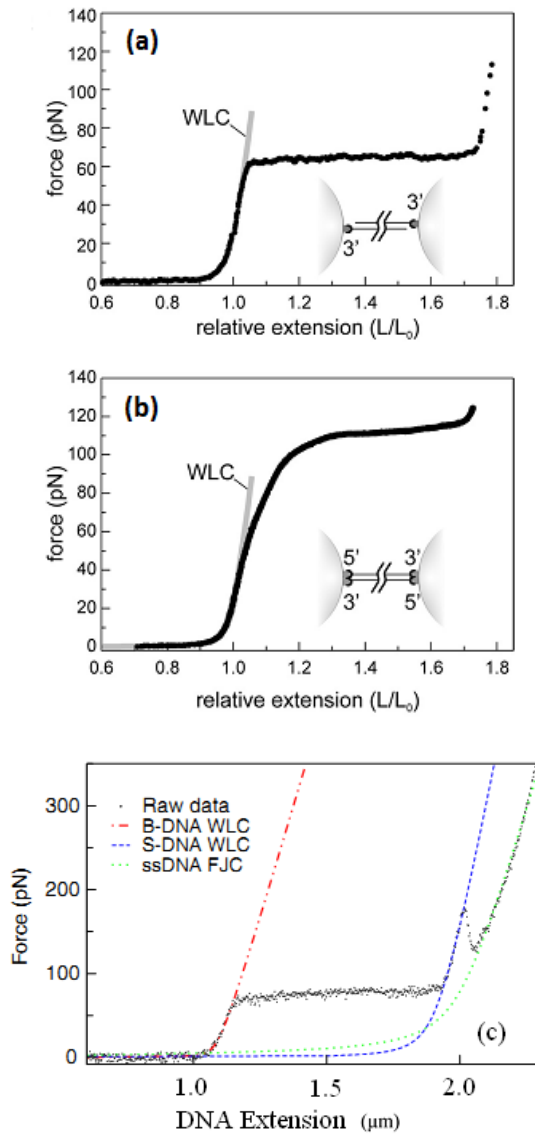


Figure 24.(a) 65 pN overstretching plateau under torsionally unconstrained condition, from ref. [113] figure 1(a); (b) 115 pN overstretching plateau under torsionally constrained condition, from ref. [113] Figure 1(b); (c) overstretching curve fitted with WLC and FJC model: green dot clearly indicates the final state to be ssDNA, from ref. [125] Figure 1(a).

There used to be long heated debates on whether S-DNA is just a hypothetical phase during the overstretching. Recently, in the experiments by van Mameren *et al.* [113], the multicolor fluorescence dyes were used to distinguish the dsDNA and ssDNA. Under the 3'-3' attachment condition, the ssDNA generated from nicks or free ends is directly identified when the molecule underwent an abrupt extension at ~ 65 pN (Figure 24(a)) [113]. Hence, they concluded that peeling is an obligatory step for the ~ 65 pN plateau, which challenged the S-DNA assumption [128].

But shortly after, Paik *et al.* observed the exhibition of the ~ 65 pN plateau using a topologically closed but torsionally free DNA assay with no nicks or free ends (Figure 25) [124]. Moreover, one single overstretched-mode of ssDNA cannot clearly explain the retraction-hysteresis observed through magnetic tweezers measurement [117]. Hysteresis means that if we stretch DNA to some extent, and then gradually decrease the applied force to let it retract, the *force-extension* relation follows a different path from that in the increasing-force stage. It is often associated with irreversible thermodynamic changes. Some potential assumptions of the molecular structural mechanism which may lead to hysteresis have been proposed. Figure 26(a) depicts the so-called hairpin formation within the unpeeled single strands due to the self-complementarity of the nucleobases [114], and Figure 26(b) shows the development of large molten bubbles which possibly lead to nicks under high tension [120]. There could be large free-energy barrier between these structures and the original B-DNA resulting in difficulties in repairing the strands or pinching the bubbles hence a slow recovery [117]. On the contrary, the transition between S-DNA and B-DNA is

much easier. Experiments show that hysteresis decreases significantly under the condition of higher ion concentration, lower temperature, or CG-rich strand (Figure 27(a, b, c)), which indicates there should exist two overstretched modes at ~ 65 pN, presumably S-DNA and ssDNA, such that the aforementioned conditions will resist the formation of ssDNA and stabilize the structure of S-DNA[117].

The structure of overstretched-DNA is so complicated that it is not easy to set an end to the debates on the structures already proposed. Meanwhile, researchers continue to suggest novel predictions. Very recently, another opinion that S-DNA serves as an intermediate state between B-DNA and zip-DNA [112] was suggested based on MD simulations. The newly-termed zip-DNA does not require accurate base pairing but π - π electronic coupling among nearest-neighbor nucleobases (Figure 27(d) [112]).

To summarize the previous studies, some tentative tension-torque phase diagram (Figure 28(a)) has been proposed [129, 130] to illustrate the critical conditions of the phase transitions. We have already been aware of the B- and S-DNA phases. The L-DNA can be obtained by applying the left-hand (negative) torques to unwind the right-hand twisted B-DNA, which results in a large number of basepairs per turn (~ 13 bp/turn). On the contrary, by applying additional right-hand (positive) torques, the overwound P-DNA with just ~ 2.6 bp/turn is obtained. The structure of P-DNA is deduced from MD simulations [131]. It has tightly interwound phosphate backbones and exposed bases in common with Pauling's early DNA structure [131, 132]. The points in the figure are from experiments, and the vertical line indicates the most frequently discussed process of B- to S-DNA transition at ~ 65 pN under

zero torque. More quantitative theory has also been developed to predict the thermal stability of homogeneous DNA under load at various temperatures [133]. Figure 28(b) shows the predicted critical lines for denaturation at different temperatures. The region enclosed by each line corresponds to stable B-DNA at that temperature. Red points are experimental data from single molecule manipulation [130], and the blue square represents the well-known overstretching transition at ~ 65 pN with no twisting torque ($\Gamma = 0$, $f = 60$ pN). The denaturation temperature (T_D) of 77 °C is critical at zero external loads, but even at temperatures of T_D or higher (Figure 28(b) inset), B-DNA can be stable in an interval of applied positive torque. Solid (dotted) straight line indicates $\Gamma_m = a\tilde{v}\tilde{\Omega}$ (its approximation $\Gamma_m = 11 + 0.092f$), in which the middle point between critical torques also corresponds to the highest critical temperature at given tension.

Apparently, we have not seen a converged model to describe the molecular mechanism underlying the phase transitions. In these processes, formation and breaking of bonds as well as electron redistribution are inevitable, which will significantly affect the atomic and electronic structure of the whole molecule. Fully understanding these changes and their influences require careful and detailed *ab initio* simulations. Therefore, we have performed a step-by-step stretching simulation on a 10-basepair periodic dsDNA strand using *ab initio* methods to illustrate the details. The procedure is comparable to a torsionally constrained case in which the ends of both strands are fixed.

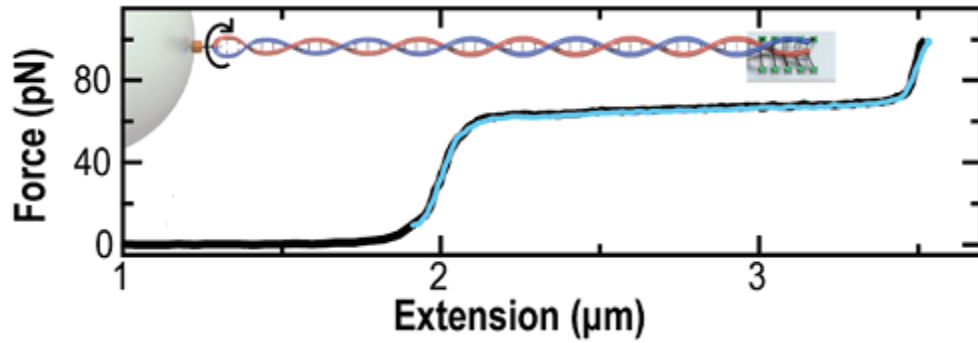


Figure 25. Overstretching at ~ 65 pN in a topologically closed but rotationally free assay using a biotin embedded in a 5-nucleotide ssDNA loop, from ref. [124] Figure 1(d).

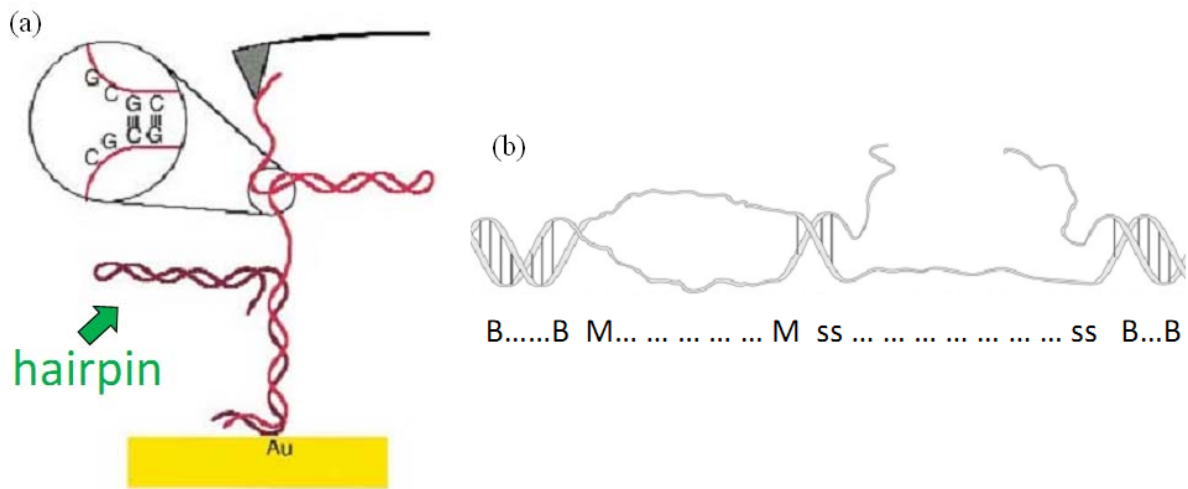


Figure 26. Potential explanations to retraction-hysteresis: (a) Unpeeled single strand forms self-complementary hairpins upon tension release which resist the recovery to original B-DNA (from ref. [114] Figure 3(a)); (b) Large molten bubble is entropy favored, which may cause entropy-driven hysteresis (from ref. [120] Figure 2).

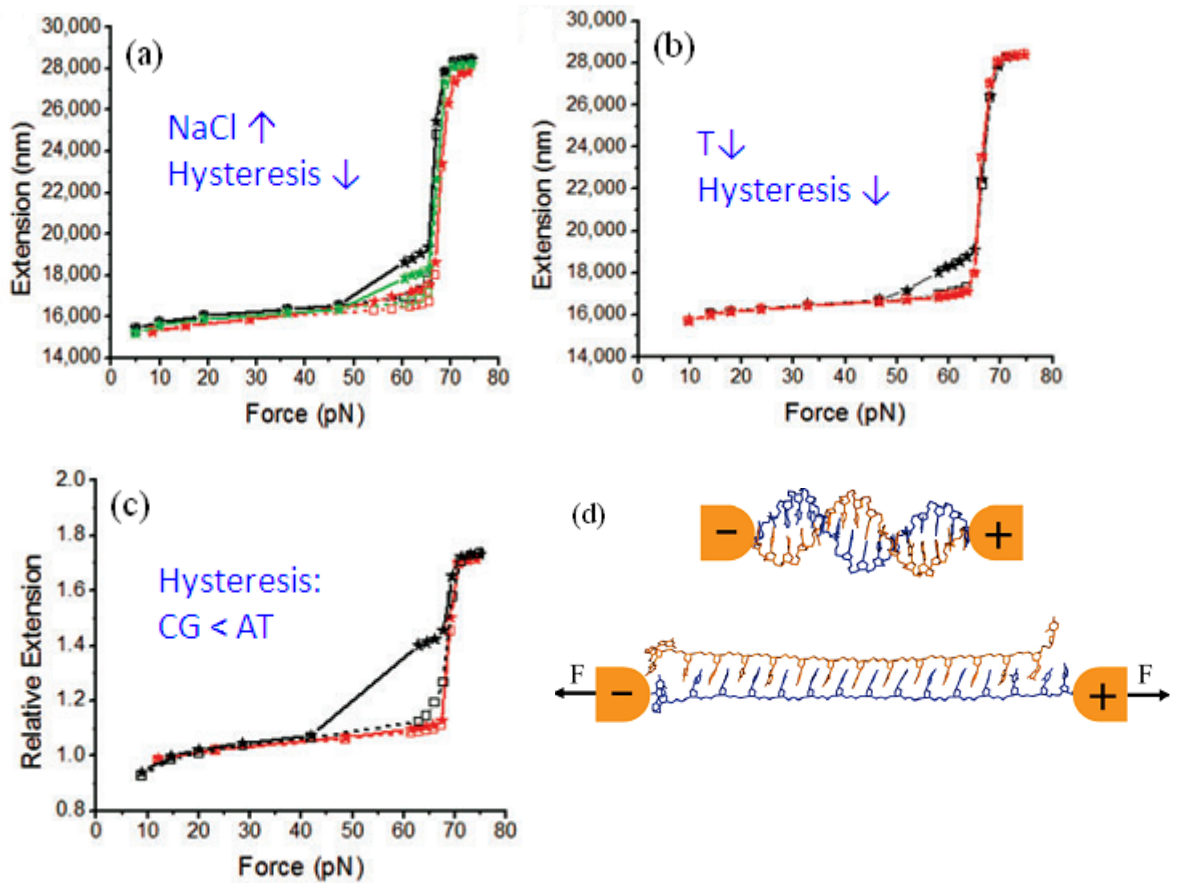


Figure 27. (a, b, c) from ref. [117] Figure 2. Increased stability of basepairs decreases overstretching hysteresis. Open squares show the increasing-force scan and solid stars show the decreasing-force scan. (a) Effects of NaCl concentration for stretching a λ -DNA: 150 mM (black), 300 mM (green, same DNA as black), and 1 M (red, a different DNA from black and green); (b) Effects of temperature for stretching λ -DNA: 24°C (black), 19°C (red); (c) Effects of sequence composition: GC-rich DNA (red) shows less hysteresis than AT-rich DNA (black). Extensions are scaled with their respective contour lengths. (d) zip-DNA does not require accurate base pairing but π - π electronic coupling among nearest-neighbor nucleobases (from ref. [112]).

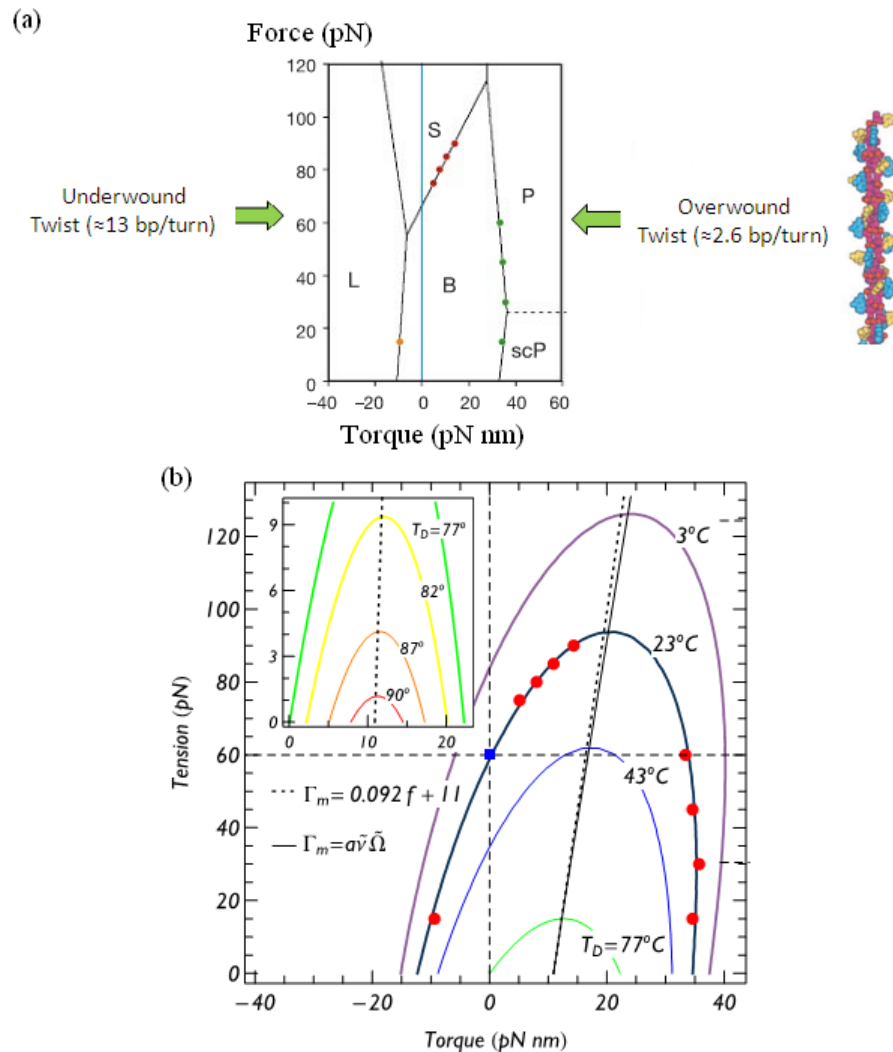


Figure 28. (a) Tentative DNA phase diagram under tension and torque, from ref. [129] Figure 4. The P-DNA model from ref. [131] Figure 5; Backbones: purple, guanine: blue, cytosine: yellow, anionic oxygens of the phosphate groups: red. (b) Predicted critical lines for denaturation at different temperatures, from ref. [133] Figure 2. The region enclosed by each line corresponds to stable B-DNA at that temperature. Red points: experimental data from single molecule manipulation [130], blue square: the well-known overstretching transition ($\Gamma = 0, f = 60$ pN). $T_D = 77^\circ\text{C}$, the denaturation temperature, is critical at zero external load: but even at temperatures of T_D or higher (inset), B-DNA can be stable in an interval of applied positive torque. Solid (dotted) straight line indicates $\Gamma_m = a\tilde{\nu}\tilde{\Omega}$ (its approximation $\Gamma_m = 11 + 0.092f$), the middle point between critical torques, which also corresponds to the highest critical temperature at given tension.

3.2.2 Model and Computational Methods

The initial B-DNA model (dA-dT)₁₀ (denoted as AT10) (Figure 29) was created using NAB. The model has nucleobases A on one strand and T on the other (see Figure 29(c)). To compensate for the negatively charged PO₄⁻¹ group, twenty Na⁺ counter ions were added (purple atoms). The double-helix structure can be described by six parameters: shifts in the *a-b* plane perpendicular to the helix's axis, stacking height *d*, twist angle θ , and tilt/roll angles (ψ , ω). The periodicity imposed in the axial direction (*c*-) reduces the number of free parameters. For this AT10 model, the twist angle is 36° and the *a-b* plane shift reduces to one helix radius $r \sim 10 \text{ \AA}$. The dimension of the unit cell was set to be sufficiently large ($30 \text{ \AA} \times 30 \text{ \AA} \times L_0$, $L_0 = 10d$, $d = 3.38 \text{ \AA}$) such that the interaction between mirror-DNA chains is negligible. There are a total of 660 atoms in this model (C₂₀₀H₂₃₀N₇₀Na₂₀O₁₂₀P₂₀) which are organized into five functional groups: sugar (C₅H₇O)₂₀, phosphate (PO₄)₂₀, Na₂₀, and the nucleobases (A: (C₅H₄N₅)₁₀ and T: (C₅H₅N₂O₂)₁₀).

The parameters for the initial VASP relaxation were similar to those in the water study except a relatively higher energy cutoff of 600 eV. Then the model was gradually stretched at a very moderate rate of 0.02-fold per step along the *c*-axis, and finally acquired an extension of 1.5-fold. To achieve high accuracy and speed up the convergence, the relaxed atomic structure from the previous step was used as the starting structure for the subsequent step. This strategy will help the structures to overcome the local energy minima to find the equilibrium states. The homogeneous stretching exerts similar tension along the backbone. Although extremely demanding in computational time, *ab initio* relaxations are absolutely

necessary. Next, we used the OLCAO method to study the electronic structure and bonding of each optimized model.

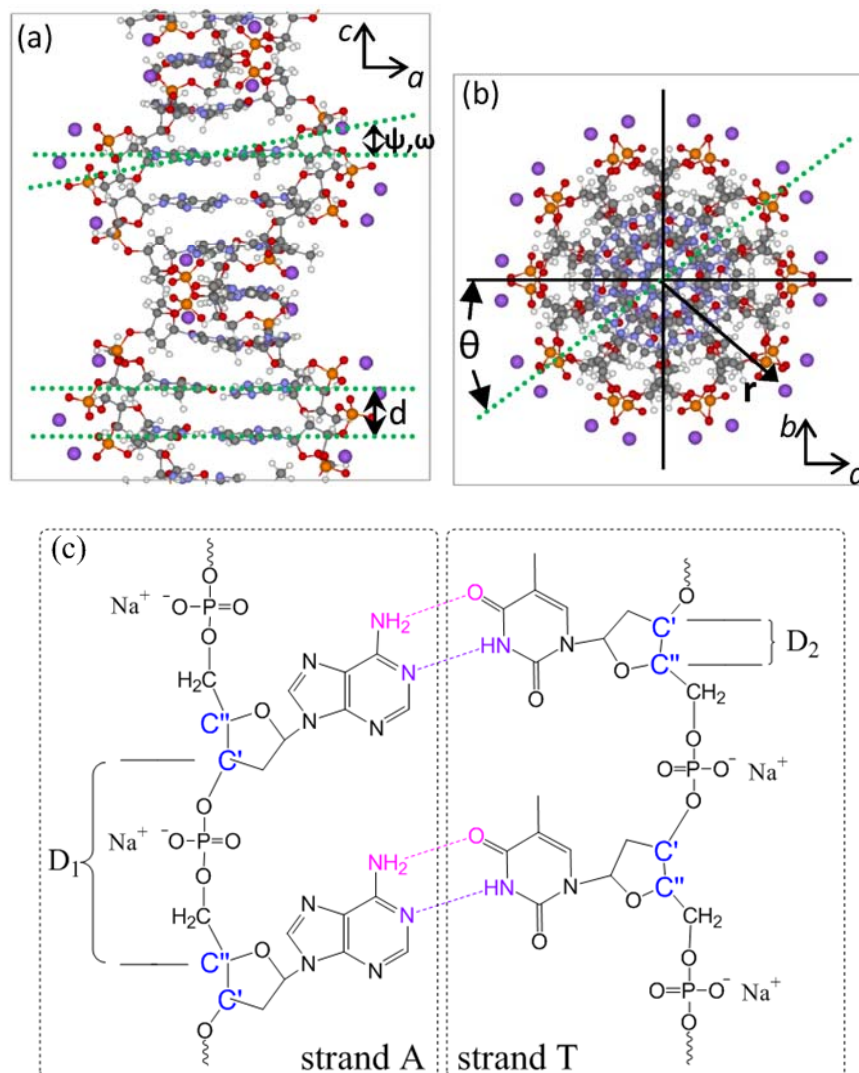


Figure 29. (a,b) Initial B-DNA model (dA-dT)₁₀ (AT10). (ψ, ω) are the tilt/roll angles, and $d = 3.38 \text{ \AA}$ is the stacking height. r is the helix radius, $\theta = 36^\circ$ is the twist angle. 20 Na^+ counterions were added (purple atoms). (c) Detailed structure.

3.2.3 Results and Analysis

Total Energy from VASP Relaxation

Figure 30 shows the calculated electron total energy (TE) of the dsDNA model as a function of the relative extension L/L_0 . We have added a dividing line (black) at 0.80 to separate the stretching process into stages (I) and (II) based on two main considerations: (1) the first overstretching plateau in the *force-extension* curve usually initiates at ~ 0.8 ; and (2) the slopes of the energy curve before 0.80 and after 0.80 are distinct. In region (I), the TE of the model undergoes a smooth and moderate change with a minimum at 0.16, and then slowly increases by about 10 eV compared to the starting model. In region II, TE first drops ~ 2.6 eV unexpectedly, indicating some structural modification which renders the model before and after such an extension in different equilibrium states. Then the TE curve becomes much steeper until it reaches a maximum at 1.30. We can expect some even more significant structural changes to happen afterwards, because a conspicuous V-type change between 1.30 and 1.40 in TE is observed. After 1.40, an energy plateau appears. It has the same energy level as that before the valley occurs. The TE change is due to the rearrangements of the atom positions and electron distributions within the DNA molecule. We should connect these features with the following discussions on the atomic and electronic structural deformation.

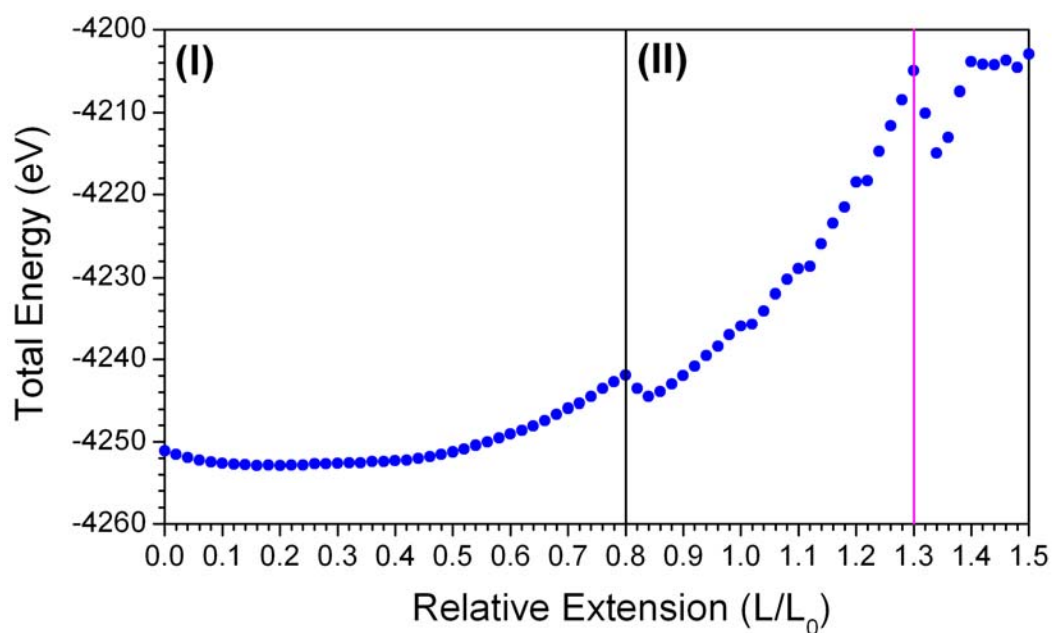


Figure 30. Electron total energy of each stretched model from VASP relaxation. The whole overstretching process is divided at 0.80 into stage (I) and (II).

Table 3. Information of B-DNA model AT10.

Model	Cell Size	Parameters	Chemical Formula	Functional Group
AT10 (660 atoms)	$30 \text{ \AA} \times 30 \text{ \AA} \times L_0$ $L_0 = 10d$	$d = 3.38 \text{ \AA}$ twist angle $\theta = 36^\circ$ helix radius $r \sim 10 \text{ \AA}$	$C_{200} H_{230} N_{70} Na_{20} O_{120} P_{20}$	Sugar: $(C_5H_7O)_20$ Phosphate: $(PO_4)_{20}$ Counter ions: Na_{20} A: $(C_5H_4N_5)_{10}$ T: $(C_5H_5N_2O_2)_{10}$

Structural Deformation

1. Backbone

As labeled in Figure 29(c), the DNA backbone can be viewed as being composed of two types of segments: the C'-PO₄-CH₂-C'' and the C''-C' single bond from the sugar ring. The point-to-point distances between two adjacent C' and C'' atoms in both the C'→PO₄→CH₂→C'' (marked as D₁) and C''→C' directions (D₂) are tracked throughout the entire stretching simulation. Figure 31 displays the average values of D₁ and D₂ in two single strands. The dividing line is also set at 0.82 corresponding to the aforementioned two stages.

In stage (I), D₁ (Figure 31(a)) increases continuously from the beginning until the relative extension reaches 0.80. D₁ in strand T (D₁^T, red circle) extends faster than D₁^A (black square) such that although D₁^T is slightly shorter than D₁^A at the beginning, they are almost identical at 0.80. On the other hand, D₂ (Figure 31(b)) does not show any discernible increase until 0.80.

In stage (II), sudden kinks in the D₁ curve similar to the TE curve appear at 0.82. From 0.82 to 1.30, both D₁^T and D₁^A continuously rise to the maximum value ~ 6.9 Å. Meanwhile, D₂ undergoes a slow and smooth increase in both strands simultaneously. Just after 1.30, D₂^T suddenly increases to about 2.2 Å indicating at least partial breaking of the C''-C' single-bonds in strand T (see Figure 32). At 1.34, D₂^A also increases to a much larger value meaning that both strands have underwent these changes. This fact clearly indicates that overstretching will result in the ring opening reactions (RORs) of the sugar's five-

member ring. Corresponding to the increase of D_2 , D_1 shows a significant drop. The subsequent stretching steps show a consistent increase in D_2 with $D_2^T > D_2^A$. Starting from 1.42, we observed nicks in strand T due to the breaking of a C-O ester bond (Figure 32). As the stretching proceeds, similar C-O bonds break on strand A as well. Our simulation stops at 1.50.

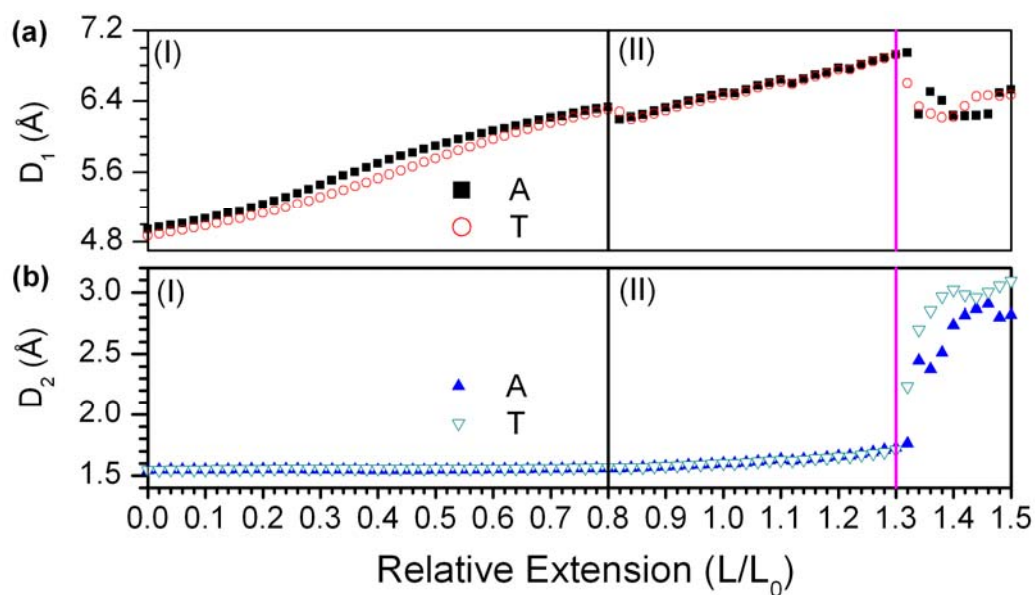


Figure 31. Average distance between C' and C'' in the direction of (a) $C' \rightarrow PO_4 \rightarrow CH_2 \rightarrow C''$ (denoted as D_1) and $C'' \rightarrow C'$ (denoted as D_2).

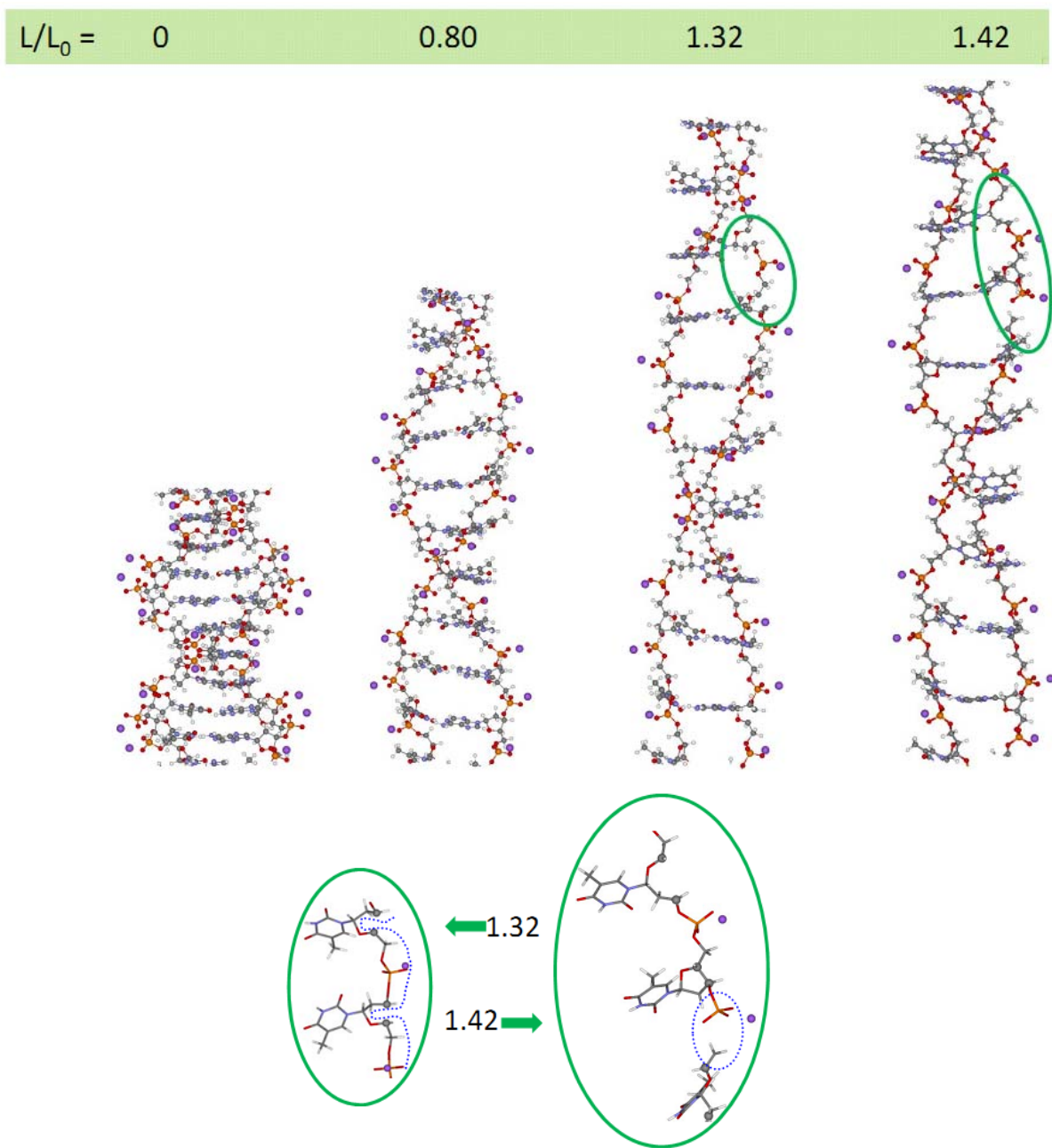


Figure 32. Displays of the elongated model at relative extension of 0, 0.80, 1.32, and 1.34. At 1.32, C'-C'' single bonds in the sugar group of strand T break resulting in the ring opening reactions (RORs). At 1.40, C-O ester bonds in strand T break resulting nicks in the backbone.

2. Nucleobases

The nucleobases also show some noticeable positional adjustments during stretching. They are best described by the changes in the dihedral angles α between the planes of basepairs (Figure 33(a)). Arrow A (T) labeled as $\uparrow A$ ($\uparrow T$) is the normal vector of adenine's (thymine's) ring plane. $\uparrow T'$ is the translation of $\uparrow T$. If the angle from $\uparrow A$ to $\uparrow T'$ is counter-clock wise, it is defined to be positive. Otherwise, negative. Figure 33(b) shows the average α at every 10% elongation within the range $\sim -25^\circ$ – 45° , which indicates relative rotation of the nucleobases during stretching. The increasing rate of α gradually slows when approaching 1.30 where the RORs occur. Then α drops to $\sim 36^\circ$ and stays stable to the end of the stretching.

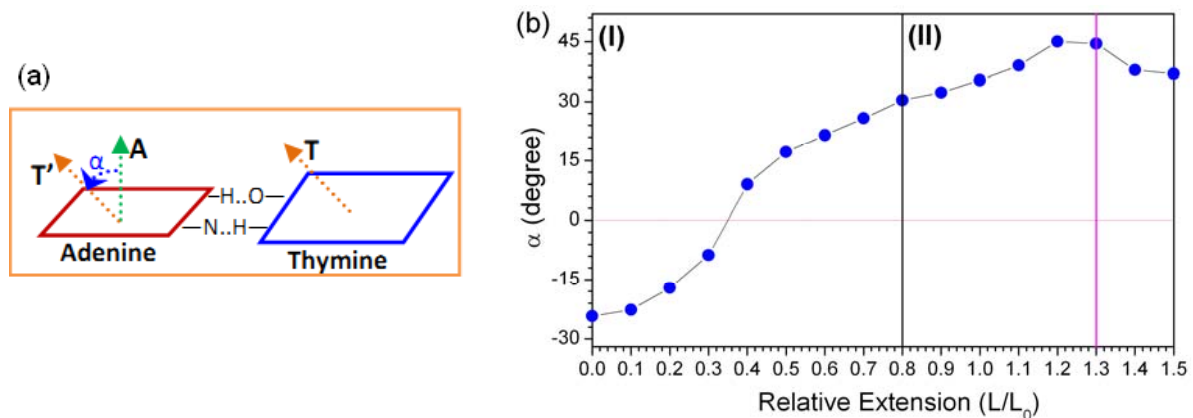


Figure 33. (a) Schematic display of the dihedral angle α between the nucleobases. $\uparrow A$ ($\uparrow T$) is the normal vector of the ring plane of adenine (thymine). $\uparrow T'$ is the translation of $\uparrow T$. (b) Variations of dihedral angle α between the base planes at each 10% elongation.

In the meantime, the HB bond lengths between the basepairs vary as well. Normally, possible HB BLs in biomolecules stay in the range $\sim 1.5\text{--}2.2$ Å [134]. Figure 34 shows the average BL values of the O..H (blue diamonds) and N..H (orange dots) pairs at each stretching step. By setting a cutoff at 2.1 Å (horizontal dash line in Figure 34), we found that the O..H HBs have already become very weak, almost approaching the breaking point at the end of stage (I). From 0.80 onward, O..H BL increases more steeply to a maximum of ~ 2.8 Å in the range $\sim 1.20\text{--}1.30$, indicating that the O..H HBs gradually break during this process. But immediately after 1.30, the values drop steeply to ~ 2.1 Å, and then stay stable till the end of simulation. This BL behavior can be interpreted as a consequence of the RORs. Once the originally-confined lengths in sugar rings are released, the backbone atoms instantly obtain more freedoms in movements. Their coordination adjustments in turn affect the helix radius and the relative position of the nucleobases. Different from the O..H HBs, the N..H HBs keep intact during the entire overstretching process. N..H BL first reaches a minimum at ~ 0.76 , then increase from 0.80 onward. The following features are rather similar with those of the O..H bonds — with a maximum (~ 1.8 Å) at ~ 1.20 followed by a steep decrease to some stable values. The relation between HB BL and bond strength is very sensitive as can be seen in the following electronic structure analysis.

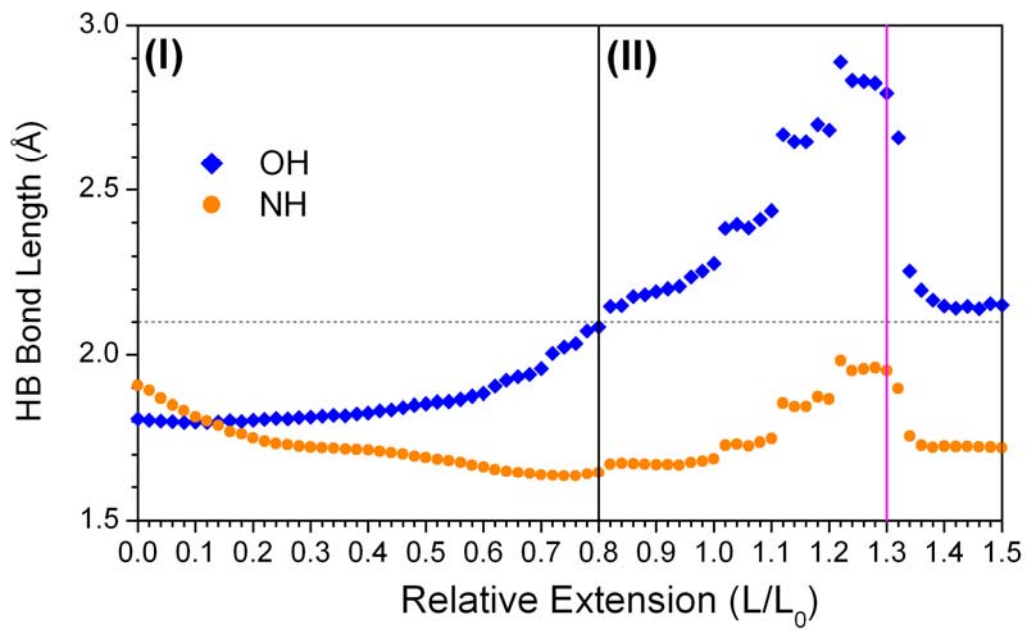


Figure 34. Variation of hydrogen bond (HB) bond length of O..H (blue diamond) and N..H (orange dot) between the basepairs.

Variation of Electronic Structure

1. Hydrogen Bond Order

The force-induced variation of HB BL is further confirmed by the calculated BO values (Figure 35) which are inversely related to BL. Corresponding to the maximum HB BL (2.1 Å) set in Figure 34, we choose HB BO cutoff at ~ 0.02 (e^-). The average BO values (black triangle) are obtained by adding all O..H and N..H BOs together and then divided by the total HB numbers of twenty. It can be used to evaluate the strength of nucleobase-connections within the whole system. As we all know, bonding strength decides the structural stability hence the electron TE. Although the specific quantitative value of HB energy may not be easy to obtain, and it only accounts for a small portion of the TE, we can still get some interesting relation between TE (Figure 30) and HB BO.

In stage (I), TE varies in such a way: slight decrease \rightarrow stay stable \rightarrow smooth increase. Although the N..H and O..H HBs behave differently, their total effects, corresponding to the average HB BO, varies in an opposite trend to the TE curve. It is quite likely that due to the relative translation and rotation of the basepairs, the HBs absorb some stretching energy to alleviate the load on the backbones, help to keep the system stable, hence postpone the occurrence of significant structural transformations such as backbone nucleations. In the first part (0.80-1.30) of stage (II), average HB strength drops quickly mainly because O..H HBs already enter the breaking region, and N..H HB strength starts to decrease as well. Meanwhile, TE increases much faster. This could be attributed to the steric effects between the basepairs which limit their freedoms in movement, and weaken the role

they played in stabilize the structure. After 1.30, the main factor affecting TE is the opened sugar rings in the backbones. The V-type change in TE is due to the structural transformation from the original B-DNA to the new structure resulted from the RORs. In comparison to this significant change, the effect of HB strength variations is negligible.

The analysis above clearly shows *ab initio* BO calculations provide an easy way to track specific bonds of interests, especially weak ones like HBs, to resolve the mechanism of DNA overstretching deformation from the atomic scale. It is also feasible to perform such analysis on other biomolecular systems for better understanding of their structural changes.

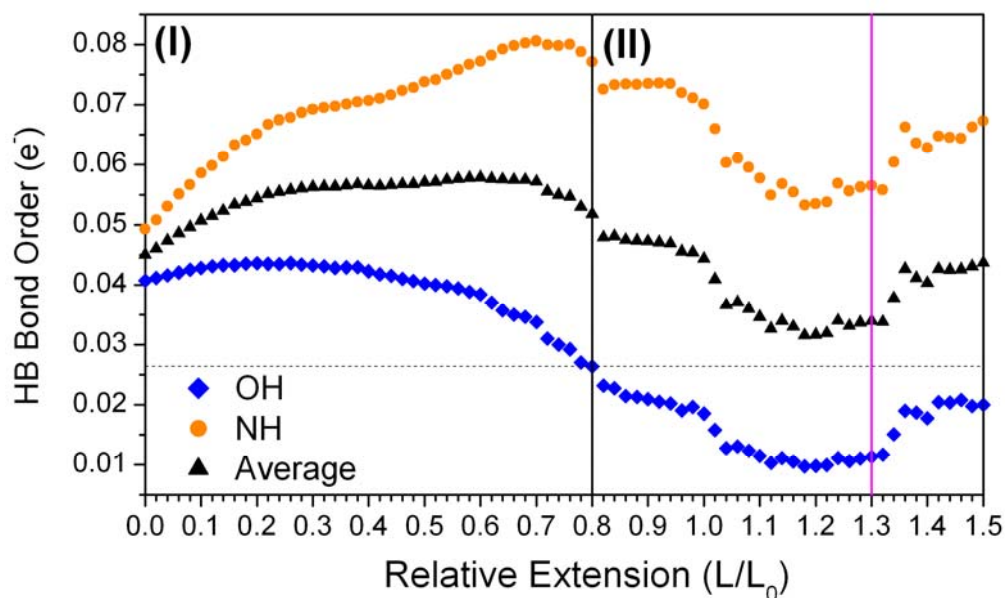



Figure 35. Variation of HB bond order of O..H (blue diamond) and N..H (orange dot) pairs. The average values are shown as black triangles.

2. Effective charge

We have also analyzed the charge transfer properties (ΔQ^*) in this process. ΔQ^* represents the difference between the effective charge Q^* of each atom in the molecule and that in the associated neutral atom. Positive values imply gaining charge. All atoms are categorized according to which functional group and strand they belong to. The average ΔQ^* of each functional group with different extensions are displayed in Figure 36. There are twelve bars in either panel (left: strand A, right: strand T), each corresponding to the relative extension of 0, 0.30, 0.50, 0.70, 0.80, 0.90, 1.10, 1.30, 1.32, 1.34, 1.40 and 1.42 (left to right). In each bar, we use different colors to distinguish the functional groups (nucleobase: green, PO_4+Na^+ : yellow, and sugar: magenta).

Generally, nucleobase A gains less charge than T, whereas the backbone groups do not differ much between strands. The shadows () represent the average net ΔQ^* of one nucleotide (PO_4+Na^+ +sugar+A(T)). We see that both strands are neutral at the beginning. But as the stretching proceeds, strand A loses some charge whereas strand T gains some charge. It indicates that stretching results in a positively charged strand A and a negatively charged strand T. Hence, we can infer that there are inter-strand electron redistributions accompanying the stretching process. The net quantity of ΔQ^* reaches a maximum at 0.80, and then reduces a little bit and maintains at a relatively stable value till the end of our simulation. The variation of ΔQ^* in the strands is achieved by the coordination of all functional groups. Considering the fact that O..H HBs start to break just after 0.80, we may conclude that the HBs which connect the two strands also act as an electron pathway to

adjust the electron distribution hence the electron TE of the system. Once O..H HBs break, these functionalities are weakened. Consequently, the net charge on the strands gradually stops to vary, and the TE starts to increase steeply. One thing worth mention is that even with the significant structural deformation of open sugar rings after 1.32, the net charge does not change any more, which further emphasizes the important roles of HBs in charge transfer activities.

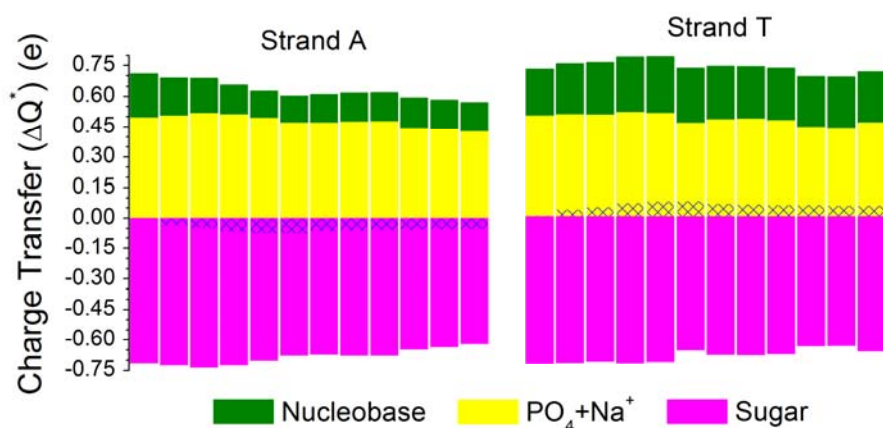



Figure 36. Charge transfer properties (ΔQ^*) categorized according to different strands (left panel: strand A; right panel: strand T) and functional groups (nucleobases: green; yellow: PO_4+Na^+ ; magenta: sugar) at relative extension of 0, 0.30, 0.50, 0.70, 0.80, 0.90, 1.10, 1.30, 1.32, 1.34, 1.40 and 1.42 (the bars from left to right). The shadow  represent the average ΔQ^* of one nucleotide.

3. Density of States

Figure 37 displays the calculated TDOS of the initial structure ($L/L_0 = 0$) in the energy range -25 – 15 eV. Generally speaking, the peak features are complex because DNA is such a big molecule with little structural symmetry. The states below -15 eV are mainly contributed by the deep core orbitals from C, O, and N atoms. Compared to other pseudo-potential methods, it is much easier to obtain such deep energy states from all-electron OLCAO calculations. We will resolve the TDOS of models at several extensions into the PDOS according to their functional groups, and focus our discussion on the energy range -4 – 8 eV, since the most interesting properties associated with electron transfer and chemical reactivity come from the HOMO-LUMO states.

The DOS and decomposed PDOS of the initial structure ($L/L_0 = 0$) and the structure just before the occurrence of RORs ($L/L_0 = 1.30$) are shown in Figure 38. There are some similarities between these two models: (1) clear HOMO-LUMO energy gap of ~ 3 eV; (2) the HOMO on top of the valence band (VB) mainly arise from the atoms in the backbone. On the other hand, some differences also exist. For the initial model, both sugar group and nucleobases contribute to the LUMO at bottom of the conduction band (CB), while at $L/L_0 = 1.30$, the LUMO are almost solely comprised of the states from nucleobase A.

In Figure 39, we chose to discuss four characteristic models after RORs happening ($L/L_0 =$ (a) 1.32, (b) 1.34, (c) 1.40, and (d) 1.42). At 1.32, a few defect gap states start to show up on top of the VB of $(\text{Na}+\text{PO}_4)_T$ group as well as in the HOMO-LUMO gap of $(\text{Sugar})_T$ group. Then similar situation occurs on strand A at 1.34. These changes should be attributed to the force-induced RORs. Possible direct consequences of the breaking of C'-C''

single bonds include: (1) unpaired radical electrons associated with their original owner C atoms (noted as C \cdot , see Figure 40(a)); (2) one C atom carries more charge therefore negatively charged (C $^-$) while the other positively charged (C $^+$) (see Figure 40(b)). Both C \cdot radicals and C $^+(-)$ ions could be very reactive, thus adding counter ions should help to pacify the dangling bonds hence stabilize the structure. The open sugar rings could stay in either case leaving the whole backbone in random order of electron distribution as shown in Figure 41. It can be (a) all C \cdot , (b) alternative C-C $^+$ C-C $^-$..., (c) mixed C-C \cdot C $^+$ C $^-$ C \cdot ..., or (d) clustered C $^+$ C-C $^-$ C $^+$ C-C $^-$ C $^+$..., etc. It should be very difficult to have continuous or large scale charge flow in the backbone, since these radicals and ions are separated far away by the PO $_4$ groups such that they should prefer to form bonds with each other. But one thing for sure is that the DOS will be greatly affected. As the stretching proceeds, we see the intensity of the defect states also grows, indicating a growingly unstable structure. We also notice that the peak intensity and positions of the nucleobases show very little variation, while these features vary from one model to another for the backbone groups. This further confirms the conclusion previously obtained from the effective charge analysis that after the occurrence of RORs, there are only intra-strand electron redistributions, and the nucleobases are not involved in this process.

In general, the above detailed DOS analysis allows us to probe the very sensitive variations in electronic structure during the stretching simulation, and to easily make connections to the atomic structural deformations.

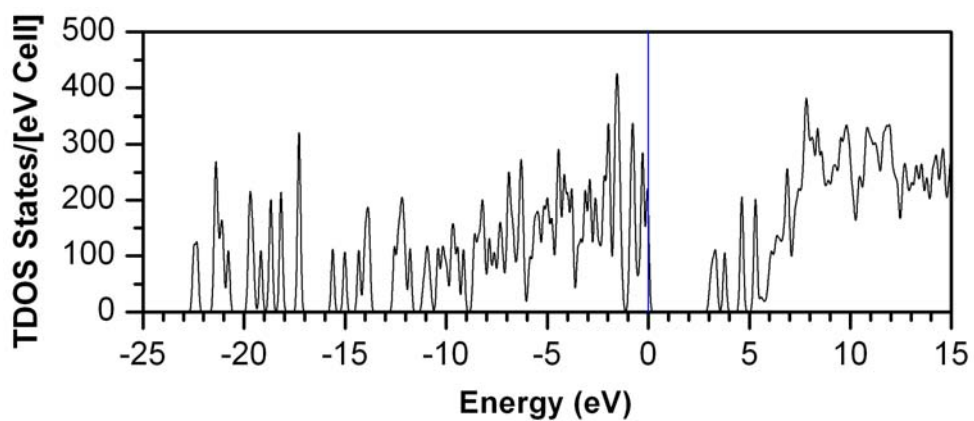


Figure 37. Calculated TDOS of the initial structure ($L/L_0 = 0$).

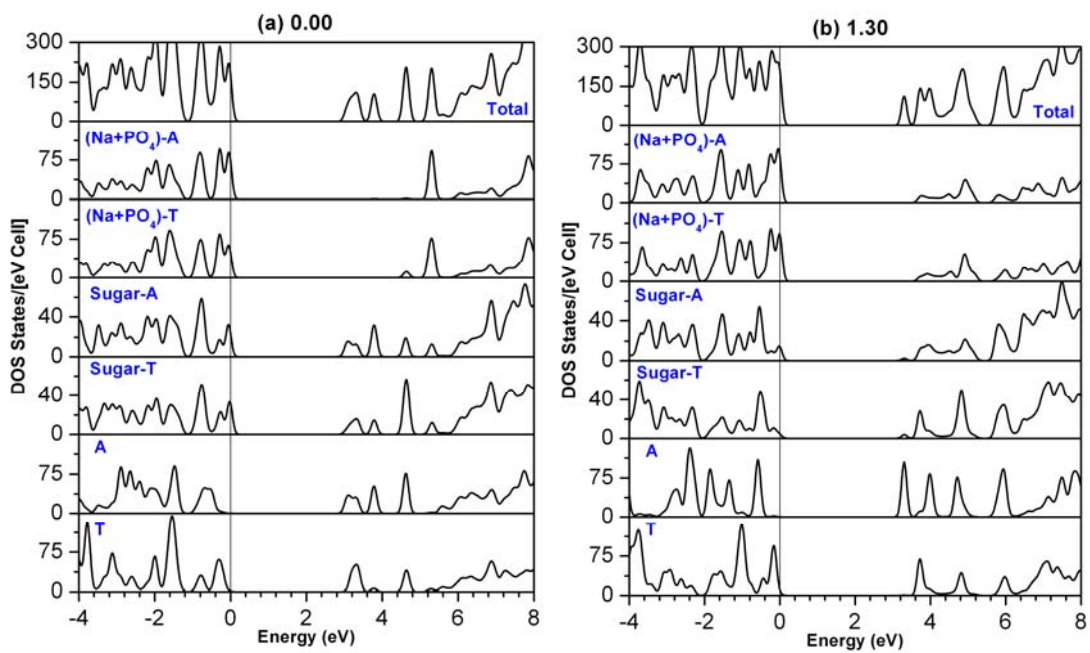


Figure 38. Calculated TDOS and decomposed PDOS of different functional groups at $L/L_0 =$ (a) 0 and (b) 1.30.

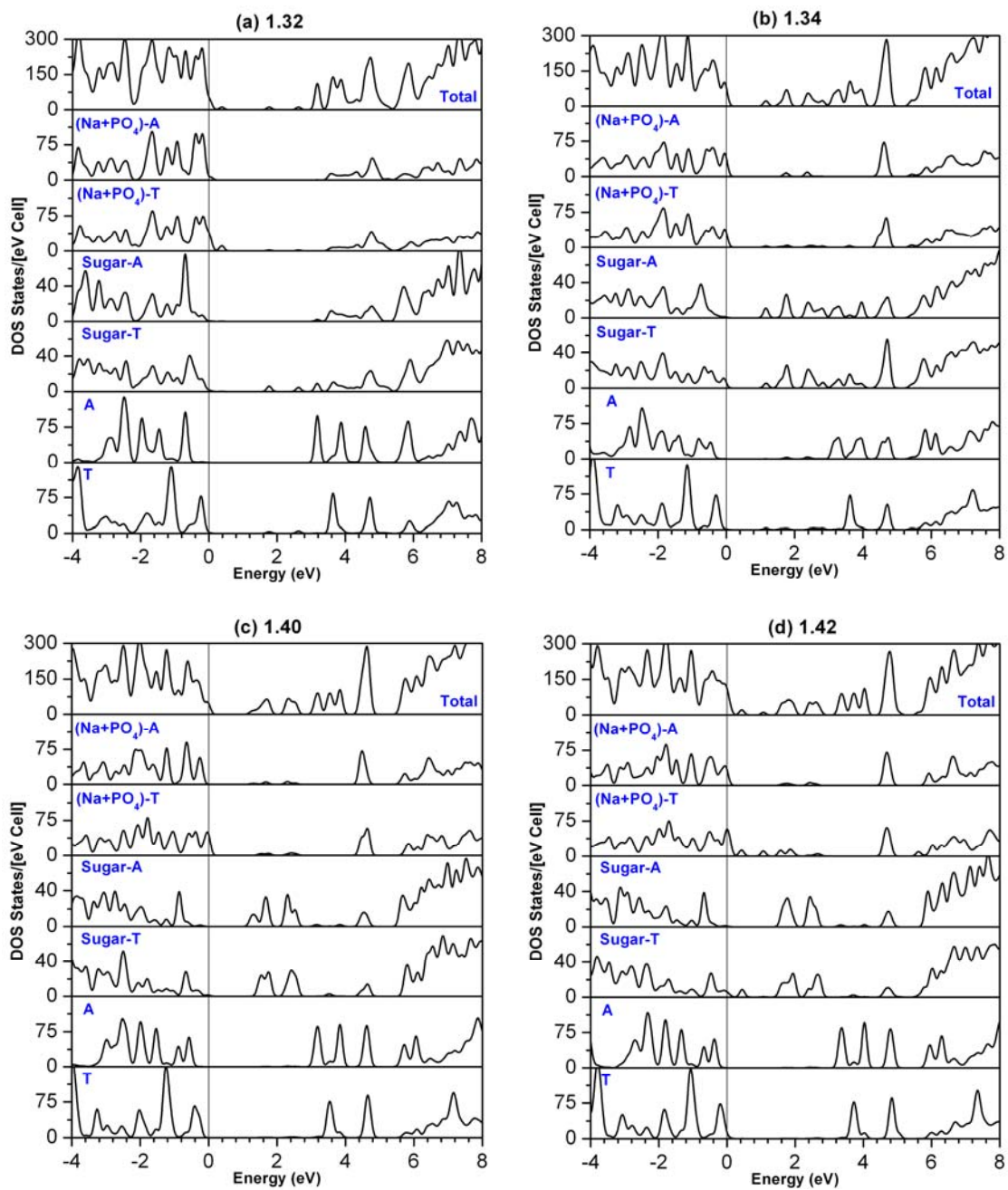


Figure 39. Calculated TDOS and decomposed PDOS of different functional groups at $L/L_0 =$ (a) 1.32, (b) 1.34, (c) 1.40, and (d) 1.42.

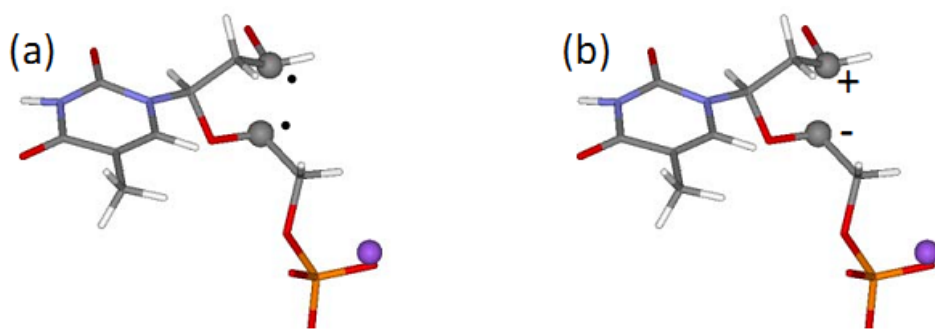


Figure 40. The breaking of C'-C'' single bond results in (a) two C· radicals (· represents the unpaired electron) or (b) C⁺(-) ions.

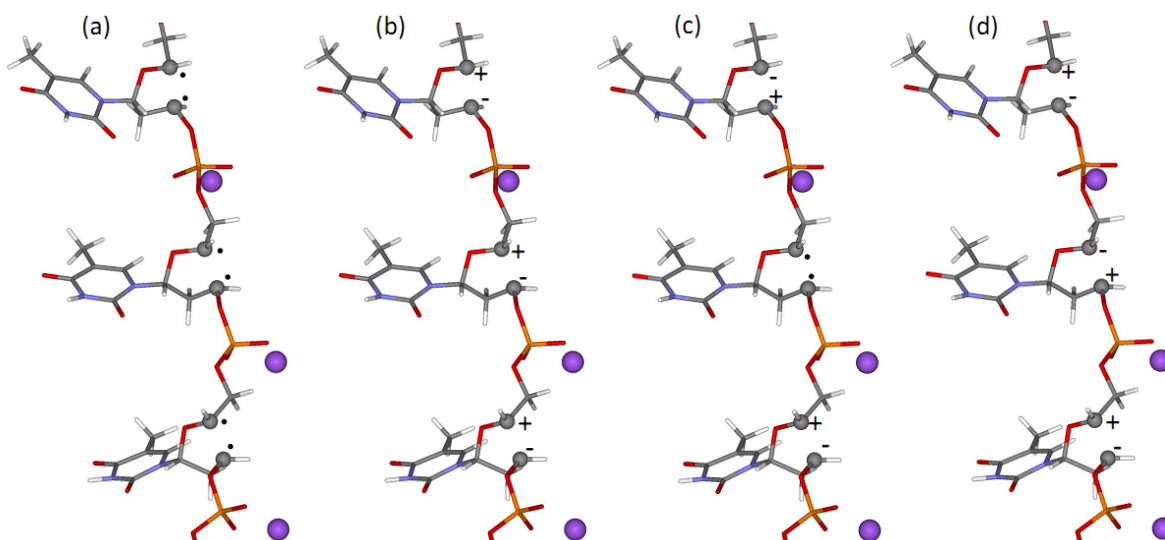


Figure 41. Possible cases of electron redistribution along the backbone after RORs: (a) all C·, (b) alternative C-C+C-C+..., (c) mixed C-C+C·C+C-..., and (d) clustered C+C-C-C+C+C-C-C+....

3.2.4 Discussions

General Picture from Overstretching Simulation

Now, we can try to generalize a concise picture of the entire process. In stage (I) ($\sim 0-0.80$), most stretching tension is absorbed by the C'-PO₄-CH₂-C'' segments in the backbone, and consequently D₁ continues to extend. The nucleobases rotate and push each other outwards of the groove to minimize the internal steric effect, which results in a decrease in N..H HB bond length as compared to an increase in O..H HB bond length. The O..H bond strength becomes progressively weaker. Generally, we did not see very significant structural deformations of the double-helix structure in stage (I).

After that, the cumulative effects of the applied tension start to exhibit. The first important change in stage (II) is the breaking of O..H HBs in the region $\sim 0.80-0.90$. Without the constraints from O..H, the nucleobases acquire more freedom in space, such that they can move even closer and cause the helical radius r to shrink. The decrease of D₁ at 0.82 is due to this reason. In the region $\sim 0.82-1.30$, which corresponds to be the stage between the entropic elasticity plateau and the overstretching plateau in the *force-extension* curve (Figure 24(b)), the C'-PO₄-CH₂-C'' segments are continuously stretched to the maximum extent, and the C''-C' covalent bonds in sugars are gradually weakened approaching the breaking point. At ~ 1.32 , C''-C' bonds become less strong than the adjacent C-O ester bonds. Hence, the RORs occur *before* the nicks appear resulting in significant structural transformation from the normal B-DNA to a new state due to the release of the confined length in the ring structure. We identify this state as O-DNA (Opened-DNA) to distinguish it from other

known or suggested states. The O-DNA ceases to exist at the moment the nicks appear in both strands. By this criterion, we see that O-DNA state of this AT10 model exists in the range ~ 1.32 - 1.48 , which in reality should be responsible for the ~ 115 pN plateau in Figure 24(b) under the torsionally constrained condition, or the short ~ 115 pN plateau labeled with (5) in Figure 23 under the unconstrained condition. Besides the structural deformations, there are electron redistributions between different functional groups along with the overstretching process as reflected by the variations DOS and effective charge.

Potential Doubts and Clarification

Some comments on the ssDNA formation from the experimentalists' viewpoint need to be addressed: ssDNA should initiate from free ends or nicks [113], because the C-O ester bonds in the DNA backbones are quite strong — known to withstand high forces up to at least 1000 pN — such that the attachment bonds between DNA and beads should break before the nicks appear on the backbone. This implies that the bonds between sugar and PO_4 break too early in our simulation. Possible clarifications to this doubt are as follows: (1) From Figure 32, we do notice that at the time the nicks appear, some of the freshly opened sugar rings have not fully released the originally confined length, and some sugar has larger deformation compared to others. It is possible that in real cases the deformation of the rings can be more even so that no particular sugar-phosphate bonds bear larger tension than the other regions; (2) The C-O breaking could be affected by some environmental factors like ion concentration in solution and temperature, etc.

Prediction of O-DNA Properties

We can now predict some properties of O-DNA and discuss the role it plays in the overall stretching process. First, the starting configuration of O-DNA should not differ much upon different DNA strands because RORs only depend on the bonding strength in the backbone. This has been confirmed by the fact that RORs occur on both A and T strands almost simultaneously. Moreover, similar simulations performed by us on another model (dC-dG)₁₀ (CG10) (will be presented in upcoming papers) gives the same result that O-DNA is formed at ~ 1.30 . Second, we do notice that the occurrence of nicks varies with sequence. The T strand breaks earlier (at 1.42) than the A strand (at 1.48). It is highly likely that for the C or G strands the nick point will also be different. Considering the experimental results that C-G rich DNA fragments show less hysteresis compared to A-T rich ones [117], we may conclude that the mechanical limit of the backbones in O-DNA are affected by the sequence order. Third, the breaking of C''-C' bonds introduces unpaired electrons which leaves O-DNA in an unstable state. In our simulations we have observed that the energy convergence rate became much slower after these structural transformations. It is conceivable that O-DNA may be stabilized and induced to last longer by increasing the ion concentration in solution or reducing the environmental temperature, to pacify the dangling bonds. This means that the retraction hysteresis caused by the unpeeling strands could be suppressed. Some of these effects have already been observed in experiments [117]. Fourth, the opened rings introduce more possible sites for nicks along the backbone. Under real experimental conditions which may involve solutions, it is very possible that nicks will lead to the breaking of all HBs, i.e.

the unpeeling of the double helix. Generally, O-DNA should be seen as a necessary state during DNA overstretching at high force plateau regardless of the attachment geometry and base-pair sequences. The ssDNA observed in experiments should be generated shortly after this state.

3.2.5 Conclusions

We believe O-DNA could be assumed as the currently missing state in explaining the overstretching mechanism at high force level. It is possible to generate S-DNA if the strands are torsionally free or if an external torque is applied to counteract DNA's own twist, but it may not be that easy through direct stretching under torsionally constrained condition, because the internal steric effects between nucleobases will force the HBs to break. One important comment is that the TE in this work cannot be directly compared to what if applied by the external stretching force because of the whole complicated process accompanied by structural transformation. Also of concern is the stability of O-DNA under real experimental conditions and different stretching force rates, but it is beyond the scope of the present *ab initio* simulation.

3.3 Summary

The simulation of water performed in this work mainly concern about the HB interactions from the quantum mechanical viewpoint, and the results have achieved nice

consistence with the experiments. From the DNA simulation, we have gained more complete understanding of the molecular mechanism underlying the complex overstretching properties.

As have mentioned before, these simulations are still under relatively ideal conditions, which simplifies the problem to some extent. Considering the simulation of wet DNA stretching, we may encounter two main difficulties in the model construction stage. First, the systems may easily contain several thousand of atoms and hence too large for first-principles calculations. Second, the solvent effect is not easy to determine since in real case DNA molecules will interact with the solvent through a large number of HBs as well as other long-range intermolecular forces. To obtain the optimized structure, the intrinsic pressure of solvent need to be fully equilibrated with the pure liquid state at each step during the stretching or compression process, which will significantly increase the simulation time. Generally, not only considerable computational resources are required but also special attentions need to be paid to avoid the dissociation of the molecules in the system. Regarding the stage of electronic structures analysis, the *ab initio* packages need to be chosen wisely for the most efficient interpretation of the system. For instance, the atom-centered OLCAO method has some unique features in the effective charge analysis. Equally important are the improvement of methods, such as more accurate evaluation, faster algorithm, better visualization method as well as sophisticated statistical models, which will also enable us to deal with complex systems.

CHAPTER 4

FUTURE WORK

4.1 Overview

Scientists believe that nanobiotechnology eventually will be indispensable in people's daily lives especially in the medical area. In fact, the convergence of the life sciences, physical sciences, computer sciences, and engineering [135] has had an immediate influence on the medical device research in preventing (e.g. short interfering RNA (siRNA) to silence the genes with the perpetuation of diseases [136]), diagnosing (e.g. circulating-tumor-cell (CTC) chips with extreme sensitivity to detect cancer metastases [137], imaging technology to spot the most subtle changes caused by diseases on retina [138]), and treating diseases (e.g. biocompatible coatings on body implants [139], extracellular biomaterial scaffold seeded with cells, tissues or drugs to speed up the healing process [140], biodegradable nanoparticles [141] and reprogrammed bacterial sensors [142] with precise targets at cancerous cells during drug delivery to eliminate the side effects of chemotherapy). With continuous efforts, more and more medical applications based on delicately tuned biomaterials and well-arranged bioprocesses are expected to come out thereby yield high economic and social value.

Equally important as the exploration for practical applications is the investigation for new mechanistic principles underlying the complex bioprocesses. All biofunctional processes rely on the cooperation of various components such as DNA, protein molecules and environmental solvent and ions. The stimulating and responding signals correctly generated

from a series of successive physical and chemical reactions guarantee the normal status of those processes hence the normal function of organisms. Gaining a complete understanding of the mechanisms may inspire us in various aspects not solely suggesting additional avenues for therapies and potential cures in the medical research.

Advances in computational simulation and modeling analysis achieved through supercomputing have opened the door for this purpose. The developing field of computational biology has attracted physicists, chemists and biologists to be actively involved in. For instance, a variety of professional MD packages with specially designed force field to describe biological systems, such as AMBER [143], NAMD [144], etc., have been developed and become routine tools to supplement the experimental studies of the conformations and functions of DNA and proteins. Interesting case studies include the simulation of DNA translocation through nanopores [145], which essentially may have the same mechanism as how DNA go into cells with the help of transmembrane cellular proteins. Besides MD, by implementing statistical mechanical methods into advanced computational models, researchers have revealed how T lymphocytes (cells essential for adaptive immunity) develop, detect, and recognize the molecular signatures of pathogens displayed on the surface of other cells [146, 147]. This groundbreaking theoretical work has had a major impact on experimental cellular and molecular immunology as well.

4.2 Prospects of Quantum Mechanical Modeling in Biology

Now, we may start to think what quantum mechanical modeling can contribute to this fascinating area. Some scientist has proposed the idea of quantum proteomics [148], which studies the electronic structure of each protein that belongs to an organism's proteome. The information such as the energy levels, charges on the amino acid side chains, and permanent dipole moments may unveil the electronic and magnetic environment of the bioprocesses and subsequently connect to biofunctions [148-150]. The *ab initio* DFT calculations presented in this thesis also partially characterize the hydrogen bonding interaction and mechanical properties in DNA. However, all these calculations only provide ground state information of nano-scale molecules surrounded by vacuum at ultra-low temperatures. With all these crucial conditions applied, it might not be very easy to make sound connections between the computational results and the mechanisms underlying real bioprocesses, which usually involve different types of energy conversion through chemical reactions, light absorption, formation of excited electronic states, transfer of excitation energy, and transfer of electrons and protons (hydrogen ions) [151]. There is no doubt that more sophisticated models and proper simulation schemes need to be introduced when studying bioprocesses, but more importantly, we should bear in mind that no single method can comprehensively interpret the phenomena related to the most complicated thing in the world — life. Quantum mechanical computational scientists need to pay close attentions to what are missing or puzzling in the experimental research, and try to bridge the gap.

One mystery successfully clarified by the collaboration of experimentalists and theorists is photosynthesis. Photosynthesis is a multistage process that must be described by quantum mechanics. The initial stage (light reactions) taking place in the reaction center (assembled complex of several proteins, pigments and other cofactors) [152] converts solar energy into chemical usable form (ATP), and releases byproducts such as O₂ (or H₂), and NADPH (coenzyme acting as reducing agent) [153]. The following stage (carbon reactions or the Calvin Cycle) uses ATP and NADPH to fix carbon dioxide into organic compounds, especially sugars. The light reactions start with photons hitting an antenna molecule and kicking up ripples of energized electrons — excitons — like a rock splashing water from a puddle. These excitons then pass from one molecule to the next until they reach the reaction center displaying a wave-like character even at relatively high temperatures of 100 ~ 200K [154]. It is just this process that amplifies the microscopic molecular photon absorption into the macroscopic signaling state at the protein level. Some scientists have pointed out that, instead of following random, undirected paths, excitons could be coherent, with their waves extending to more than one molecule while staying in step and reinforcing one another. That means they would be able to move through the forest of antenna molecules by two or more routes at once, and furthermore automatically select the most efficient path to the reaction center [154]. Recent studies have found evidences to back up this hypothesis by using photosynthesis experiments of bacterial complex and algae at both cryogenic and ambient temperatures [155-158]. But here rises another question: how can the coherence last long enough to be useful in photosynthesis instead of being destroyed instantly by the surrounding

molecular chaos at ambient temperatures? Perhaps this can never be answered by experiments since nature simply does so. Computational simulations using continuous-time quantum walks suggest an answer: random noise in the environment might actually enhance the efficiency of the energy transfer in photosynthesis rather than degrade it [159]. The noise can free up an exciton trapped on particular sites to avoid destroying its coherence and allow it to get to where it is going [159]. Scientists have already started considering learning from the biological systems and taking advantages of the environmental noise to build solar cells with improved energy-conversion efficiency [154].

In fact, there are other known or suspected examples of quantum effects in nature. In some enzyme-catalysed reactions, protons move from one molecule to another by the quantum-mechanical tunneling rather than having to muster the energy to climb over it [160]. Another interesting one is about the odor sensing of human beings. The molecules identifying certain type of odor are recognized not only by their shape but also by their phonon frequencies because of the electron tunneling between the molecules and the receptors in people's nose [161]. Also found is that the retinal isomerization can be steered by modulating the phases and amplitudes of the spectral components in the photoexcitation pulse through constructive and destructive interference effects [162]. Besides these solved ones, researchers suspect that mechanisms underlying the long-term puzzles — birds' eye compass, and the growth-inhibiting influence of blue light on plants — could be decoded by the theory of quantum-assisted magnetic sensing which relies on some entangled radical pairs created by the incoming photons [163-166]. The two unpaired electrons in the radical pair

exist in a state that the spin of one electron remains correlated with that of the other no matter how far they are. The environmental magnetic fields shift the radical pair between two quantum states with different chemical reactivity such that some chemicals are generated with different concentration at different locations to reflect Earth's field orientation. The potential influence of this assumption is substantial: by understanding how nature maintains the coherence of these two quantum states at ambient conditions, we may be able to find clues in designing systems suitable for quantum computing which highly relies on the coherence of qubit information [167]. On talking about quantum computing, the idea of learning from DNA structure for long-term qubit-data storage has also been proposed recently [168]. Honestly speaking, all the aforementioned robust quantum effects in warm live organisms may have already challenged our original thinking, but it is just the confusion that gives birth to an emerging discipline called *quantum biology*, which talks about functions of the non-trivial quantum features such as superposition, nonlocality, coherence, entanglement, and tunneling in biological objects and problems [169].

To clarify the underlying mechanisms of those bioprocesses, quantum mechanical modeling in elaborate scheme is indispensable. The first key point of a desirable computational scheme is a clearly defined model which consists of the basic molecules participating in specific quantum effects. With this solid information, the seeming complexity could be break down to some fundamental process feasible for computational modeling. The second point is that various quantum mechanical formulations (e.g. path integral, electron density, phase space, etc.) may be required to describe one single system.

Normally, *ab initio* studies on condensed matter systems will adopt the frozen-core approximation and only treat the valence electrons in either wavefunction or density matrix formulation, since the nuclei can be assumed as stationary and acting like classical particles compared to the electrons. But this is not suitable for studying some bioprocesses in which the effects from the highly correlated motion of heavy atoms are non-negligible [170]. Some investigation on proton transfer in HBs has demonstrated that by combining an *ab initio* path integral representation of the quantized nuclei and a DFT representation of the valence electrons, the calculated free energy barrier is a lot more accurate compared to purely classical treatment of the nuclei [171]. These kinds of sophisticated studies are instructive but their quantities are still rare. Therefore it has not yet been the moment to talk about some generalized methodology capable of handling every biological effect. An even higher criterion on the computational efficiency can only be achieved through tireless studies case by case.

4.3 Final Comments

The advancement of natural science and technology has brought us into a new research era in which we can study everything from the deepest bottom, appreciate the delicateness of life, and eventually build up novel applications to benefit our world. *ab initio* modeling will no doubt be a powerful tool in this exploring journey. To contribute as a computational scientist, we should break through the originally confined scope, creatively design better simulation schemes, and delicately pursue for the underlying mechanisms.

APPENDIX A

VASP RELAXATION RUNNING TIME COMPARISON

Before:	After:
<pre>PBS job submitting file: #!/bin/bash -l #PBS -q regular #PBS -N test #PBS -l walltime=6:00:00 #PBS -l mppwidth=768 #PBS -j oe #PBS -V cd \$PBS_O_WORKDIR module load vasp/5.2 aprun -n 512 -N16 gvasp Running time: LOOP: cpu time 236.71: real time 236.84 LOOP: cpu time 307.94: real time 307.98 LOOP: cpu time 252.94: real time 252.95 LOOP: cpu time 249.16: real time 249.20 LOOP: cpu time 322.15: real time 322.21 LOOP: cpu time 279.92: real time 279.99 LOOP: cpu time 287.14: real time 287.19 LOOP: cpu time 285.03: real time 285.06 LOOP: cpu time 302.00: real time 302.06 LOOP: cpu time 286.12: real time 286.17 LOOP: cpu time 291.82: real time 291.89 LOOP: cpu time 288.59: real time 288.87 LOOP: cpu time 295.42: real time 295.49 LOOP: cpu time 285.90: real time 285.92 LOOP: cpu time 291.63: real time 291.72 LOOP: cpu time 287.91: real time 287.97 LOOP: cpu time 282.06: real time 282.11 LOOP: cpu time 286.01: real time 286.08</pre>	<pre>PBS job submitting file: #!/bin/bash -l #PBS -q regular #PBS -N test #PBS -l walltime=6:00:00 #PBS -l mppwidth=768 #PBS -j oe #PBS -V cd \$PBS_O_WORKDIR module load vasp/5.2 aprun -n 384 -N12 -S3 -cc cpu gvasp Running time: LOOP: cpu time 116.42: real time 116.52 LOOP: cpu time 148.80: real time 148.80 LOOP: cpu time 122.04: real time 122.03 LOOP: cpu time 120.16: real time 120.16 LOOP: cpu time 152.31: real time 152.32 LOOP: cpu time 96.28: real time 96.29 LOOP: cpu time 98.36: real time 98.35 LOOP: cpu time 97.60: real time 97.61 LOOP: cpu time 102.41: real time 103.01 LOOP: cpu time 98.42: real time 98.42 LOOP: cpu time 99.54: real time 99.54 LOOP: cpu time 98.88: real time 98.88 LOOP: cpu time 100.47: real time 100.50 LOOP: cpu time 98.11: real time 98.12 LOOP: cpu time 99.73: real time 99.74 LOOP: cpu time 98.59: real time 98.61 LOOP: cpu time 97.05: real time 97.07 LOOP: cpu time 97.98: real time 97.98</pre>

LOOP: cpu time 288.97: real time 289.07	LOOP: cpu time 99.27: real time 99.29
LOOP: cpu time 277.37: real time 277.41	LOOP: cpu time 95.84: real time 95.84
LOOP: cpu time 282.30: real time 282.38	LOOP: cpu time 96.53: real time 96.54
LOOP: cpu time 252.65: real time 252.73	LOOP: cpu time 88.27: real time 88.27
LOOP: cpu time 234.25: real time 234.28	LOOP: cpu time 83.17: real time 83.18
LOOP: cpu time 232.34: real time 232.47	LOOP: cpu time 82.75: real time 82.75
LOOP: cpu time 215.19: real time 215.22	LOOP: cpu time 77.94: real time 77.94
LOOP: cpu time 212.41: real time 212.45	LOOP: cpu time 76.94: real time 76.95
LOOP: cpu time 210.19: real time 210.24	LOOP: cpu time 76.42: real time 76.43
LOOP: cpu time 207.87: real time 207.94	LOOP: cpu time 75.92: real time 75.93
LOOP: cpu time 202.70: real time 202.77	LOOP: cpu time 74.48: real time 74.48
LOOP: cpu time 176.58: real time 176.89	LOOP: cpu time 66.51: real time 66.52
LOOP+: cpu time 8219.41: real time 8221.95	LOOP+: cpu time 3052.79: real time 3053.71
Elapsed time (sec): 8631.025	Elapsed time (sec): 3327.090

APPENDIX B

TYPICAL INITIAL DATASETS OF OXYGEN ATOM POTENTIAL.

$D_{j,A}$	$\beta_{j,A}$
-0.22887365E+00	0.30000000E+00
-0.17788331E+01	0.81654446E+00
-0.43081427E+01	0.22224829E+01
-0.10496011E+02	0.60491869E+01
0.25132887E+01	0.16464767E+02
0.25100771E+02	0.44814047E+02
-0.60836611E+01	0.12197554E+03
0.14746808E+02	0.33199484E+03
-0.13642236E+02	0.90362850E+03
0.19140063E+02	0.24595095E+04
-0.22412284E+02	0.66943295E+04
0.28526996E+02	0.18220726E+05
-0.34224721E+02	0.49593442E+05
0.39833950E+02	0.13498417E+06
-0.39534840E+02	0.36740192E+06
0.25108640E+02	0.10000000E+07

APPENDIX C

MINIMAL BASIS OF OXYGEN ATOMIC ORBITALS

NUM_ALPHA_S_P_D_F

20 20 0 0

ALPHAS

0.12000000E+00	0.23711531E+00	0.46853060E+00	0.92579816E+00
0.18293410E+01	0.36147064E+01	0.71425186E+01	0.14113338E+02
0.27887404E+02	0.55104421E+02	0.10888418E+03	0.21515090E+03
0.42512977E+03	0.84003981E+03	0.16598859E+04	0.32798696E+04
0.64808943E+04	0.12805994E+05	0.25304144E+05	0.50000000E+05

NUM_CORE_ORBITALS

1

LS_WAVEFUNCTIONS

0

-0.25566292E-04	0.16520969E-03	-0.58567369E-03	0.17417748E-02
0.45574060E-01	0.75880064E+00	0.29760591E+01	0.55324124E+01
0.67200232E+01	0.64080889E+01	0.53539631E+01	0.41722344E+01
0.30186184E+01	0.23210644E+01	0.14372317E+01	0.14189511E+01
0.36862504E+00	0.12505747E+01	-0.43684461E+00	0.10389304E+01

NUM_VALENCE_ORBITALS

2

LS_WAVEFUNCTIONS

0

-0.30323632E-01	-0.20292707E+00	-0.56092422E+00	-0.80574631E+00
-0.56237334E+00	0.41171595E+00	0.14150850E+01	0.18848229E+01
0.19365685E+01	0.16682652E+01	0.13246322E+01	0.10070134E+01
0.71757075E+00	0.54905918E+00	0.33799786E+00	0.33391133E+00
0.86310405E-01	0.29375958E+00	-0.10262723E+00	0.24384273E+00

1

0.29744514E-01	0.82012528E-01	0.31921377E+00	0.66207498E+00
0.13325156E+01	0.18016892E+01	0.20678049E+01	0.19335284E+01
0.17794474E+01	0.12727443E+01	0.11390169E+01	0.66505790E+00
0.66740023E+00	0.27459813E+00	0.41654273E+00	0.61007983E-01
0.28565655E+00	-0.23183485E-01	0.15051292E+00	0.51245066E-01

APPENDIX D

ABBREVIATIONS

BA: bond angle
BL: bond length
BO: bond angle
BP: Beche-Perdew
CB: covalent band
CTC: circulating-tumor-cell
DFT: density functional theory
DOS: density of states
DZP: double- ζ polarization
EB: extended basis
FB: full basis
FFT: fast Fourier transformation
FJC: freely-joined chain
FS: final structure
GGA: generalized gradient approximation
HB: hydrogen bond
HDF5: Hierarchical Data Format Version 5
HK: Hohenberg-Kohn
HOMO: highest occupied molecular orbitals
IS: initial structure
KS: Kohn-Sham
L(S)DA: local (spin) density approximation
LCAO: linear combination of atomic orbitals
LI: local index
LUMO: lowest unoccupied molecular orbitals
LYP: Lee-Yang-Parr
MB: minimal basis
MD: molecular dynamics
M-DNA: molten-DNA
NAB: Nucleic Acid Builder
NERSC: National Energy Research Scientific Computing Center
NNHO: near-neighbor hydrogen and oxygen
O-DNA: opened-DNA
OLCAO: orthogonalized linear combination of atomic orbitals
PAW: projector augmented wave
PBE: Perdew-Burke-Ernzerhof
PDOS: partial density of states

PES: photoelectron spectroscopy
pN: pico Newton
PW: plane wave
PW86: Perdew-Wang86
 Q^* : effective charge
RDF: radial distribution function
ROR: ring opening reaction
SCF: self-consistent field
siRNA: short interfering RNA
SIESTA: Spanish Initiative for Electronic Simulations with Thousands of Atoms
ssDNA: single-stranded DNA
SZ: single- ζ
 T_d : denaturation temperature
TE: total energy
US-PP: ultra-soft Vanderbilt pseudo-potential
VASP: Vienna Ab initio Simulation Package
VB: valence band
WF: wave function
WLC: worm-like chain
XANES: x-ray absorption near edge structure
XAS: x-ray absorption spectra
XC: exchange correlation
XRS: x-ray Raman scattering

REFERENCES

- [1] C. M. Niemeyer, and C. A. Mirkin, *Nanobiotechnology: Concepts, Applications and Perspectives* (Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2004).
- [2] P. W. K. Rothmund, *Nature* **440**, 297 (2006).
- [3] T. J. Macke, and D. A. Case, *Modeling Unusual Nucleic Acid Structures in Molecular Modeling of Nucleic Acids*, edited by N. B. Leontes, and J. J. SantaLucia (American Chemical Society, Washington, DC, 1998), pp. 379.
- [4] *MPI*, Argonne National Laboratory (<http://www.mcs.anl.gov/research/projects/mpi/>).
- [5] *OpenMP*, OpenMP.org (<http://openmp.org/wp/about-openmp/>).
- [6] *Advanced CompuTational Software (ACTS)*, DOE (<http://acts.nersc.gov/tools.html>).
- [7] R. C. Merkle, *Nanotechnology* **2**, 134 (1991).
- [8] A. S. Barnard, *Nature Nanotechnology* **4**, 332 (2009).
- [9] B. D. Dunietz, M. D. Beachy, Y. Cao, D. A. Whittington, S. J. Lippard, and R. A. Friesner, *Journal of the American Chemical Society* **122**, 2828 (2000).
- [10] R. Hoffmann, *J. Chem. Phys.* **39**, 1397 (1963).
- [11] P. Vashishta, A. Nakano, R. K. Kalia, and I. Ebbsjö, *Materials Science and Engineering B* **37**, 56 (1996).
- [12] M. F. Horstemeyer, *Multiscale Modeling: A Review in Practical Aspects of Computational Chemistry: Methods, Concepts and Applications*, edited by J. Leszczynski, and M. K. Shukla (Springer, New York, 2009).
- [13] S. F. Boys, *Proc. R. Soc. London Ser. A* **200**, 542 (1950).
- [14] A. Nel, T. Xia, L. Mädler, and N. Li, *Science* **311**, 622 (2006).
- [15] A. S. Barnard, *Nat Mater* **5**, 245 (2006).
- [16] J. Wong-Ekkabut, S. Baoukina, W. Triampo, I. M. Tang, D. P. Tieleman, and L. Monticelli, *Nat Nano* **3**, 363 (2008).
- [17] S. B. Smith, Y. Cui, and C. Bustamante, *Science* **271**, 795 (1996).

- [18] J. F. Léger, G. Romano, A. Sarkar, J. Robert, L. Bourdieu, D. Chatenay, and J. F. Marko, *Phys. Rev. Lett.* **83**, 1066 (1999).
- [19] G. Kresse, and J. Hafner, *Phys. Rev. B* **47**, 558 (1993).
- [20] P. Ordejón, E. Artacho, and J. M. Soler, *Physical Review B* **53**, R10441 (1996).
- [21] W.-Y. Ching, and P. Rulis, *Electronic Structure Method for Complex Materials: The orthogonalized linear combination of atomic orbitals* (Oxford University Press, Oxford, 2011).
- [22] W.-Y. Ching, and P. Rulis, *J. Phys.: Condens. Mat.* 104202 (2009).
- [23] C. C. J. Roothaan, *Reviews of Modern Physics* **23**, 69 (1951).
- [24] L. H. Thomas, *Proc. Camb. Phil. Soc.* **23**, 542 (1927).
- [25] E. Fermi, *Rend. Accad. Naz. Lincei* **6**, 602 (1927).
- [26] P. Hohenberg, and W. Kohn, *Physical Review B* **136**, 864 (1964).
- [27] W. Kohn, and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- [28] K. Burke, J. P. Perdew, and M. Levy, *Modern Density Functional Theory: A Tool for Chemistry* (Elsevier Press, Amsterdam, 2004).
- [29] D. C. Langreth, and M. J. Mehl, *Physical Review B* **28**, 1809 (1983).
- [30] J. P. Perdew, and W. Yue, *Physical Review B* **33**, 8800 (1986).
- [31] A. D. Becke, *Physical Review A* **38**, 3098 (1988).
- [32] C. Lee, W. Yang, and R. G. Parr, *Physical Review B* **37**, 785 (1988).
- [33] J. P. Perdew, P. Ziesche, and H. Eschrig, *Electronic Structure of Solids '91* (Akademie Verlag, Berlin, 1991).
- [34] J. P. Perdew, K. Burke, and M. Ernzerhof, *Physical Review Letters* **77**, 3865 (1996).
- [35] L. C. Balbás, M. José Luís, and J. M. Soler, *Phys. Rev. B* **64**, 165110 (2001).
- [36] D. M. Ceperley, and B. J. Alder, *Physical Review Letters* **45**, 566 (1980).

- [37] R. Car, M. Parrinello, A. Ppolian, P. Lebouyre, and N. Boccara, *Simple Molecular Systems at Very High Density* (Plenum Press, 1989).
- [38] R. Car, and M. Parrinello, *Physical Review Letters* **55**, 2471 (1985).
- [39] R. P. Feynman, *Physical Review* **56**, 340 (1939).
- [40] A. M. Rappe, K. M. Rabe, E. Kaxiras, and J. D. Joannopoulos, *Physical Review B* **41**, 1227 (1990).
- [41] N. Troullier, and J. L. Martins, *Physical Review B* **43**, 1993 (1991).
- [42] G. Kresse, J. Hafner, and R. J. Needs, *Journal of Physics: Condensed Matter* **36**, 7451 (1992).
- [43] D. Vanderbilt, *Physical Review B* **41**, 7892 (1990).
- [44] K. Laasonen, R. Car, C. Lee, and D. Vanderbilt, *Physical Review B* **43**, 6796 (1991).
- [45] K. Laasonen, A. Pasquarello, R. Car, C. Lee, and D. Vanderbilt, *Physical Review B* **47**, 10142 (1993).
- [46] P. E. Blöchl, *Physical Review B* **50**, 17953 (1994).
- [47] *Hopper Cray XE6*, National Energy Research Scientific Computing Center (<http://www.nersc.gov/systems/hopper-cray-xe6/>).
- [48] J. M. Soler, E. Artacho, J. D. Gale, A. J. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal, *J. Phys. : Condens. Mat.* **14**, 2745 (2002).
- [49] E. Artacho, E. Anglada, O. Diéguez, J. D. Gale, A. J. García, J. Junquera, R. M. Martin, P. Ordejón, J. M. Pruneda, D. Sánchez-Portal, and J. M. Soler, *J. Phys. : Condens. Matter.* **20**, 064208 (2008).
- [50] O. Pablo, *Computational Materials Science* **12**, 157 (1998).
- [51] S. Goedecker, *Reviews of Modern Physics* **71**, 1085 (1999).
- [52] P. Ordejón, *physica status solidi (b)* **217**, 335 (2000).
- [53] O. F. Sankey, and D. J. Niklewski, *Physical Review B* **40**, 3979 (1989).
- [54] M. P. Allen, and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, Oxford, 1987).

- [55] *PHONON*, K. Parlinski, (<http://wolf.ifj.edu.pl/phonon/>).
- [56] J. Callway, *Energy Band Theory* (Academic Press, 1964).
- [57] R. C. Chaney, C. C. Lin, and E. E. Lafon, *Physical Review B* **3**, 459 (1971).
- [58] C. S. Wang, and J. Callaway, *Comput. Phys. Commun.* **14**, 327 (1978).
- [59] B. N. Harmon, W. Weber, and D. R. Hamann, *Physical Review B* **25**, 1109 (1982).
- [60] *Mean Field Theory*, Wikipedia (http://en.wikipedia.org/wiki/Mean_field_theory).
- [61] W.-Y. Ching, C. C. Lin, and D. L. Huber, *Physical Review B* **14**, 620 (1976).
- [62] M.-Z. Huang, L. Ouyang, and W.-Y. Ching, *Physical Review B* **59**, 3540 (1999).
- [63] J. P. Buban, K. Matsunaga, J. Chen, N. Shibata, W.-Y. Ching, T. Yamamoto, and Y. Ikuhara, *Science* **311**, 212 (2006).
- [64] J. Chen, L. Ouyang, P. Rulis, A. Misra, and W.-Y. Ching, *Physical Review Letters* **95**, 256103 (2005).
- [65] P. Rulis, H. Yao, L. Ouyang, and W.-Y. Ching, *Physical Review B* **76**, 245410 (2007).
- [66] P. Rulis, L. Ouyang, and W.-Y. Ching, *Physical Review B* **70**, 155104 (2004).
- [67] L. Ouyang, P. Rulis, W.-Y. Ching, M. Slouf, G. Nardin, and L. Randaccio, *Spectrochimica Acta Part A* **61**, 1647 (2005).
- [68] W.-Y. Ching, *J. Am. Ceram. Soc.* **73**, 3135 (1990).
- [69] L. Ouyang, *Parallel Computing and First-principles Calculations: Applications to Complex Ceramics and Vitamin B12* (Doctoral Dissertation, University of Missouri-Kansas City, 2000).
- [70] P. Rulis, *Computational Studies of Bioceramic Crystals & Related Materials* (Doctoral Dissertation, University of Missouri-Kansas City, 2005).
- [71] *LAPACK*, Netlib (<http://www.netlib.org/lapack/>).
- [72] *HDF5*, The HDF Group (<http://www.hdfgroup.org/HDF5/>).
- [73] J. A. White, and D. M. Bird, *Phys. Rev. B* **50**, 4954 (1994).

- [74] G. Lehmann, and M. Taut, *physica status solidi (b)* **54**, 469 (1972).
- [75] O. Jepsen, and O. K. Anderson, *Solid State Communications* **9**, 1763 (1971).
- [76] J. Rath, and A. J. Freeman, *Physical Review B* **11**, 2109 (1975).
- [77] R. S. Mulliken, *J. Chem. Phys.* **23**, 1833 (1955).
- [78] *XANES*, Wikipedia (<http://en.wikipedia.org/wiki/XANES>).
- [79] *Auger Effect*, Wikipedia (http://en.wikipedia.org/wiki/Auger_effect).
- [80] *Exciton*, Wikipedia (<http://en.wikipedia.org/wiki/Exciton>).
- [81] S.-D. Mo, and W.-Y. Ching, *Physical Review B* **62**, 7901 (2000).
- [82] T. Mizoguchi, A. Seko, M. Yoshiya, H. Yoshida, T. Yoshida, W.-Y. Ching, and I. Tanaka, *Phys. Rev. B* **76**, 195125 (2007).
- [83] T. Mizoguchi, T. Sasaki, S. Tanaka, K. Matsunaga, T. Yamamoto, M. Kohyama, and Y. Ikuhara, *Phys. Rev. B* **74**, 235408 (2006).
- [84] P. Rulis, W.-Y. Ching, and M. Kohyama, *Acta Mater.* **52**, 3009 (2004).
- [85] K. Tatsumi, T. Mizoguchi, S. Yoshioka, T. Yamamoto, T. Suga, T. Sekine, and I. Tanaka, *Phys. Rev. B* **71**, 033202 (2005).
- [86] T. Mizoguchi, J. P. Buban, K. Matsunaga, T. Yamamoto, and Y. Ikuhara, *Ultramicroscopy* **106**, 92 (2006).
- [87] *Phase Diagram of Water*, Wikipedia (http://en.wikipedia.org/wiki/File:Phase_diagram_of_water.svg).
- [88] G. A. Lyakhov, and D. M. Mazo, *Europhys. Lett.* **57**, 396 (2002).
- [89] P. Wernet, D. Nordlund, U. Bergmann, M. Cavalleri, M. Odelius, H. Ogasawara, L. A. Naslund, T. K. Hirsch, L. Ojamae, P. Glatzel, L. G. M. Pettersson, and A. Nilsson, *Science* **304**, 995 (2004).
- [90] A. K. Soper, *J. Phys.: Condens. Mat.* **17**, S3273 (2005).
- [91] M. Leetmaa, K. T. Wikfeldt, M. P. Ljungberg, M. Odelius, J. Swenson, A. Nilsson, and L. G. M. Pettersson, *J. Chem. Phys.* **129**, 084502 (2008).

- [92] M. Odelius, M. Cavalleri, A. Nilsson, and L. G. M. Pettersson, *Phys. Rev. B* **73**, 024205 (2006).
- [93] J. D. Smith, C. D. Cappa, K. R. Wilson, B. M. Messer, R. C. Cohen, and R. J. Saykally, *Science* **306**, 851 (2004).
- [94] T. Tokushima, Y. Harada, O. Takahashi, Y. Senba, H. Ohashi, L. G. M. Pettersson, A. Nilsson, and S. Shin, *Chem. Phys. Lett.* **460**, 387 (2008).
- [95] A. K. Soper, *Chem. Phys.* **258**, 121 (2000).
- [96] P. T. Kiss, and A. Baranyai, *J. Chem. Phys.* **131**, 204310 (2009).
- [97] G. Brancato, N. Rega, and V. Barone, *Phys. Rev. Lett.* **100**, 107401 (2008).
- [98] J. M. Sorenson, G. Hura, R. M. Glaeser, and T. Head-Gordon, *J. Chem. Phys.* **113**, 9149 (2000).
- [99] S. Yoo, X. C. Zeng, and S. S. Xantheas, *The Journal of Chemical Physics* **130**, 221102 (2009).
- [100] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez, *Journal of Computational Chemistry* **30**, 2157 (2009).
- [101] A. K. Soper, and C. J. Benmore, *Phys. Rev. Lett.* **101**, 065502 (2008).
- [102] E. Schwegler, G. Galli, and F. Gygi, *Phys. Rev. Lett.* **84**, 2429 (2000).
- [103] V. Garbuio, M. Cascella, L. Reining, R. Del Sole, and O. Pulci, *Phys. Rev. Lett.* **97**, 137402 (2006).
- [104] A. Nilsson, H. Ogasawara, M. Cavalleri, D. Nordlund, M. Nyberg, P. Wernet, and L. G. M. Pettersson, *The Journal of Chemical Physics* **122**, 154505 (2005).
- [105] D. Nordlund, M. Odelius, H. Bluhm, H. Ogasawara, L. G. M. Pettersson, and A. Nilsson, *Chemical Physics Letters* **460**, 86 (2008).
- [106] A. Rahman, and F. H. Stillinger, *J. Am. Chem. Soc.* **95**, 7943 (1973).
- [107] R. Kumar, J. R. Schmidt, and J. L. Skinner, *J. Chem. Phys.* **126**, 204107 (2007).
- [108] A. D. Hammerich, and V. Buch, *J. Chem. Phys.* **128**, 111101 (2008).
- [109] M. V. Fernández-Serra, and E. Artacho, *Phys. Rev. Lett.* **96**, 016404 (2006).

- [110] S. Myneni, Y. Luo, L. Å. Näslund, M. Cavalleri, and L. Ojamäe, *J. Phys.: Condens. Mat.* **14**, L213 (2002).
- [111] D. Nordlund, H. Ogasawara, K. J. Andersson, M. Tatarkhanov, M. SalmerÅ³n, L. G. M. Pettersson, and A. Nilsson, *Phys. Rev. B* **80**, 233404 (2009).
- [112] A. Balaeff, S. L. Craig, and D. N. Beratan, *The Journal of Physical Chemistry A* **115**, 9377 (2011).
- [113] J. van Mameren, P. Gross, G. Farge, P. Hooijman, M. Modesti, M. Falkenberg, G. J. L. Wuite, and E. J. G. Peterman, *Proceedings of the National Academy of Sciences* **106**, 18231 (2009).
- [114] M. Rief, H. Clausen-Schaumann, and H. E. Gaub, *Nat Struct Mol Biol* **6**, 346 (1999).
- [115] J. D. Watson, and F. H. C. Crick, *Nature* **171**, 737 (1953).
- [116] P. Cluzel, A. Lebrun, C. Heller, R. Lavery, J.-L. Viovy, D. Chatenay, and F. o. Caron, *Science* **271**, 792 (1996).
- [117] H. Fu, H. Chen, J. F. Marko, and J. Yan, *Nucleic Acids Research* **38**, 5594 (2010).
- [118] H. Fu, H. Chen, X. Zhang, Y. Qu, J. F. Marko, and J. Yan, *Nucleic Acids Research* **39**, 3473 (2011).
- [119] S. A. Harris, Z. A. Sands, and C. A. Laughton, *Biophysical Journal* **88**, 1684 (2005).
- [120] S. Whitelam, S. Pronk, and P. L. Geissler, *Biophys J* **94**, 2452 (2008).
- [121] S. Whitelam, S. Pronk, and P. L. Geissler, *The Journal of Chemical Physics* **129**, 205101 (2008).
- [122] M. C. Williams, and I. Rouzina, *Current Opinion in Structural Biology* **12**, 330 (2002).
- [123] M. C. Williams, I. Rouzina, and V. A. Bloomfield, *Accounts of Chemical Research* **35**, 159 (2002).
- [124] D. H. Paik, and T. T. Perkins, *Journal of the American Chemical Society* **133**, 3219 (2011).
- [125] C. P. Calderon, W.-H. Chen, K.-J. Lin, N. C. Harris, and C.-H. Kiang, *Journal of Physics: Condensed Matter* **21**, 034114 (2009).

- [126] M. Doi, and S. F. Edwards, *The Theory of Polymer Dynamics* (Clarendon Press, Oxford, 1999).
- [127] S. Smith, L. Finzi, and C. Bustamante, *Science* **258**, 1122 (1992).
- [128] M. C. Williams, I. Rouzina, and M. J. McCauley, *Proceedings of the National Academy of Sciences* **106**, 18047 (2009).
- [129] A. Sarkar, J.-F. Léger, D. Chatenay, and J. F. Marko, *Physical Review E* **63**, 051903 (2001).
- [130] Z. Bryant, M. D. Stone, J. Gore, S. B. Smith, N. R. Cozzarelli, and C. Bustamante, *Nature* **424**, 338 (2003).
- [131] J. F. Allemand, D. Bensimon, R. Lavery, and V. Croquette, *Proceedings of the National Academy of Sciences of the United States of America* **95**, 14152 (1998).
- [132] L. Pauling, and R. B. Corey, *Proceedings of the National Academy of Sciences* **39**, 84 (1953).
- [133] C. Nisoli, and A. R. Bishop, *Physical Review Letters* **107**, 068102 (2011).
- [134] G. A. Jeffrey, *An Introduction to Hydrogen Bonding* (Oxford University Press, New York, 1997).
- [135] P. A. Sharp, and R. Langer, *Science* **333**, 527 (2011).
- [136] M. S. Goldberg, D. Xing, Y. Ren, S. Orsulic, S. N. Bhatia, and P. A. Sharp, *Proceedings of the National Academy of Sciences* **108**, 745 (2011).
- [137] L. V. Sequist, S. Nagrath, M. Toner, D. A. Haber, and T. J. Lynch, *J. Thorac. Oncol.* **4**, 281 (2009).
- [138] J. M. Schmitt, R. Huber, and J. G. Fujimoto, *Limiting ischemia by fast Fourier-domain imaging in Optical Coherence Tomography in Cardiovascular Research*, edited by E. Regar, T. G. van Leeuwen, and P. W. Serruys (Informa Healthcare, Abingdon Oxon, 2007).
- [139] H. Wieneke, T. Sawitowski, S. Wendt, O. Dirch, Y. L. Gu, U. Dahmen, A. Fischer, and R. Erbel, *Z. Kardiol* **91**, 66 (2002).
- [140] T. Dvir, B. P. Timko, D. S. Kohane, and R. Langer, *Nat Nano* **6**, 13 (2011).

- [141] A. Verma, O. Uzun, Y. Hu, Y. Hu, H.-S. Han, N. Watson, S. Chen, D. J. Irvine, and F. Stellacci, *Nature Materials* **7**, 588 (2008).
- [142] J. C. Anderson, E. J. Clarke, A. P. Arkin, and C. A. Voigt, *Journal of Molecular Biology* **355**, 619 (2006).
- [143] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, *Journal of Computational Chemistry* **26**, 1668 (2005).
- [144] M. T. Nelson, W. Humphrey, A. Gursoy, A. Dalke, L. V. Kalé, R. D. Skeel, and K. Schulten, *International Journal of High Performance Computing Applications* **10**, 251 (1996).
- [145] M. Muthukumar, *Annual Review of Biophysics and Biomolecular Structure* **36**, 435 (2007).
- [146] A. K. Chakraborty, and A. Košmrlj, *Annual Review of Physical Chemistry* **61**, 283 (2010).
- [147] A. Kosmrlj, E. L. Read, Y. Qi, T. M. Allen, M. Altfeld, S. G. Deeks, F. Pereyra, M. Carrington, B. D. Walker, and A. K. Chakraborty, *Nature* **465**, 350 (2010).
- [148] F. Pichierri, in *ArXiv:1107.5853v1* (Cornell University Library, 2011).
- [149] F. Pichierri, *Biosystems* **103**, 132 (2011).
- [150] F. Pichierri, *Journal of Molecular Structure: THEOCHEM* **950**, 79 (2010).
- [151] *Quantum Biology*, University of Illinois at Urbana-Champaign, Theoretical and Computational Biophysics Group (http://www.ks.uiuc.edu/Research/quantum_biology/).
- [152] *Photosynthetic Reaction Center*, Wikipedia (http://en.wikipedia.org/wiki/Photosynthetic_reaction_centre).
- [153] *Nicotinamide adenine dinucleotide phosphate (NADPH)*, Wikipedia (<http://en.wikipedia.org/wiki/NADPH>).
- [154] P. Ball, *Nature* **474**, 272 (2011).
- [155] G. S. Engel, T. R. Calhoun, E. L. Read, T.-K. Ahn, T. Mancal, Y.-C. Cheng, R. E. Blankenship, and G. R. Fleming, *Nature* **446**, 782 (2007).

- [156] E. Collini, C. Y. Wong, K. E. Wilk, P. M. G. Curmi, P. Brumer, and G. D. Scholes, *Nature* **463**, 644 (2010).
- [157] H. Lee, Y.-C. Cheng, and G. R. Fleming, *Science* **316**, 1462 (2007).
- [158] G. Panitchayangkoon, D. Hayes, K. A. Fransted, J. R. Caram, E. Harel, J. Wen, R. E. Blankenship, and G. S. Engel, *Proceedings of the National Academy of Sciences* **107**, 12766 (2010).
- [159] M. Mohseni, P. Rebentrost, S. Lloyd, and A. Aspuru-Guzik, *J. Chem. Phys.* **129**, 174106 (2008).
- [160] P. Ball, *Nature* **431**, 396 (2004).
- [161] J. C. Brookes, F. Hartoutsiou, A. P. Horsfield, and A. M. Stoneham, *Physical Review Letters* **98**, 038101 (2007).
- [162] V. I. Prokhorenko, A. M. Nagy, S. A. Waschuk, L. S. Brown, R. R. Birge, and R. J. D. Miller, *Science* **313**, 1257 (2006).
- [163] T. Ritz, P. Thalau, J. B. Phillips, R. Wiltschko, and W. Wiltschko, *Nature* **429**, 177 (2004).
- [164] K. Maeda, K. B. Henbest, F. Cintolesi, I. Kuprov, C. T. Rodgers, P. A. Liddell, D. Gust, C. R. Timmel, and P. J. Hore, *Nature* **453**, 387 (2008).
- [165] E. M. Gauger, E. Rieper, J. J. L. Morton, S. C. Benjamin, and V. Vedral, *Physical Review Letters* **106**, 040503 (2011).
- [166] M. Ahmad, P. Galland, T. Ritz, R. Wiltschko, and W. Wiltschko, *Planta* **225**, 615 (2007).
- [167] H. Neven, *Google Workshop on Quantum Biology: Welcome and Introduction* (2010) (<http://www.youtube.com/watch?v=pSe8mBWeeYM&feature=BFa&list=PLEEC8D055615B0F5B&lf=plpp>).
- [168] E. Rieper, *Google Workshop on Quantum Biology: Classical and Quantum Information in DNA* (2010) (<http://www.youtube.com/watch?v=2nqHOnVTxJE&feature=BFa&list=PLEEC8D055615B0F5B&lf=plpp>).
- [169] *Quantum Biology*, Wikipedia (http://en.wikipedia.org/wiki/Quantum_biology).

[170] A. Sergi, AAPP | PHYSICAL, MATHEMATICAL, AND NATURAL SCIENCES **87**, C1C0901001 (2009).

[171] M. E. Tuckerman, and D. Marx, Physical Review Letters **86**, 4946 (2001).

VITA

Lei Liang was born on March 24, 1984 in Anhui, People's Republic of China. She was educated in local public schools. In September 2002, she enrolled in the University of Science and Technology of China, majoring in physics. After four years of undergraduate study, she received the Bachelor degree of Science in July 2006.

In August 2006, Ms. Liang came to the United States as a graduate student in the Department of Physics at the University of Missouri-Kansas City. She earned her Master degree of Science in Physics in September 2009. She started her interdisciplinary Ph. D program with Physics as the coordinating unit and Statistics as co-discipline at the University of Missouri-Kansas City in September 2009. Meanwhile, she started her Master program in Statistics. In addition to completing the degree requirements, she also worked variously as teaching assistant in general physics laboratories and research assistant.

Ms. Liang is a member of the American Physical Society and American Ceramics Society. She has published five papers in several journals. She has also presented two conference talks and three posters. She won the Chancellor's Interdisciplinary Ph. D Fellowship in 2010 and 2011.

University of Missouri-Kansas City
School of Graduate Studies

**REPORT OF RESULTS OF FINAL DISSERTATION EXAMINATION
FOR INTERDISCIPLINARY Ph.D. CANDIDATES**

Candidate's Name: **Lei Liang**

Student ID Number: **12277111**

Date of examination: **Nov 29, 2011**

Candidate's Coordinating Unit Discipline: **Physics**

Candidate's Co-Discipline(s): **Math & Statistics**

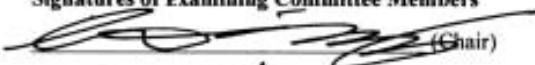

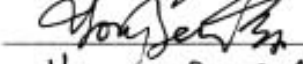
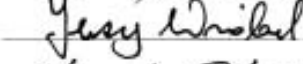
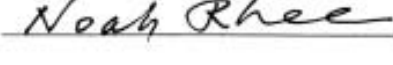
Title of Dissertation: **ab initio Computational Applications to Complex Biomolecular Systems**

Results of Examination:

Passed the Final Dissertation Examination

Failed the Final Dissertation Examination

Examining Committee Member Evaluations and Comments:

Signatures of Examining Committee Members	Member's Evaluation
 (Chair)	<input checked="" type="checkbox"/> Superior <input type="checkbox"/> Good <input type="checkbox"/> Fair <input type="checkbox"/> Inferior (failed)
 (Member)	<input checked="" type="checkbox"/> Superior <input type="checkbox"/> Good <input type="checkbox"/> Fair <input type="checkbox"/> Inferior (failed)
 (Member)	<input checked="" type="checkbox"/> Superior <input type="checkbox"/> Good <input type="checkbox"/> Fair <input type="checkbox"/> Inferior (failed)
 (Member)	<input checked="" type="checkbox"/> Superior <input type="checkbox"/> Good <input type="checkbox"/> Fair <input type="checkbox"/> Inferior (failed)
 (Member)	<input checked="" type="checkbox"/> Superior <input type="checkbox"/> Good <input type="checkbox"/> Fair <input type="checkbox"/> Inferior (failed)
_____ (Member)	<input type="checkbox"/> Superior <input type="checkbox"/> Good <input type="checkbox"/> Fair <input type="checkbox"/> Inferior (failed)

The defense of the dissertation is approved when a majority of the supervisory committee members recommend approval and sign the Report of the result of Final Doctoral Examination form. Within 48 hours of the defense, the supervisory committee chair shall report the results of the final dissertation examination in writing to the candidate and to the dean of the School of Graduate Studies. The dean of the School of Graduate studies, in turn, informs the registrar.

[Revised 3/8/2010]

PUBLICATIONS

1. **Lei Liang**, Paul Rulis, W. Y. Ching, "Mechanism of structural deformation in DNA double helix overstretching", (submitted).
2. **Lei Liang**, Paul Rulis, Lizhi Ouyang, W. Y. Ching, "Ab initio investigation of hydrogen bonding and network structure in a supercooled model of water", Phys. Rev. B 83 (2), 024201 (2011).
3. **Lei Liang**, Paul Rulis, W. Y. Ching, "*Mechanical properties, electronic structure, and bonding of α - and β -tricalcium phosphates with surface characterization*", Acta Biomaterialia 6, 3763 (2010).
4. **Lei Liang**, Paul Rulis, Bart Kahr, W. Y. Ching, "*Theoretical study of the large linear dichroism of herapathite*", Phys. Rev. B 80, 235132 (2009).
5. Wenbo Yu, **Lei Liang**, Zijing Lin, "*Comparison of some representative density functional theory and wave function theory methods for the studies of amino acids*", J. Comput. Chem. 30 (4), 589 (2009).
6. **Lei Liang**, Zhijing Liu, "*Novel Application of Dip-pen Nanolithography*", Modern Scientific Instruments 2, 10 (2006).