First-Principles Modeling of the Amyloid-Forming Peptide GNNQQNY

by

Chen Li

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Bachelor of Science in Physics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2007
[June 2007]

Author .....
Department of Physics
May 18, 2007

Certified by .....
Professor Nicola Marzari
Associate Professor in Computational Materials Science
Thesis Supervisor

Accepted by .....
Professor David E. Pritchard
Senior Thesis Coordinator, Department of Physics

First-Principles Modeling of the Amyloid-Forming Peptide GNNQQNY

by

Chen Li

Submitted to the Department of Physics
on May 18, 2007, in partial fulfillment of the
requirements for the degree of
Bachelor of Science in Physics

Abstract

This thesis presents an *ab initio* study of biological molecules using first-principles molecular dynamics. Density functional theory and Car-Parrinello molecular dynamics are used in the computational modeling of the water molecules and the amyloid-forming peptide GNNQQNY derived from the yeast prion protein Sup35. The Young's modulus of the fibril obtained from the *ab initio* method is larger than the experimentally reported value, which can be improved with a more complete understanding of the physical properties of the fibril.

Thesis Supervisor: Professor Nicola Marzari
Title: Associate Professor in Computational Materials Science

Acknowledgments

I thank Arash Mostofi and Nichola Marzari for discussions about this thesis. I owe much to my advisor and Arash, who spent many hours helping me with the details of my work. Finally, I would like to thank Young-Su Lee and other people in the group for their aid with the computing resource.

I thank the engineers of MIT beyond.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Understanding the electronic structures of complex biological molecules is essential in the modeling of biocatalytic systems. Early studies of biological molecules is based on the classical molecular dynamics, which assumes that the atoms follow the Newton's equations of motion under empirical potentials. The classical description fails to capture the quantum behaviors of the system arising from the many-body electronic wavefunctions. An accurate modeling must incorporate the quantum mechanical interactions.

Density functional theory describes the essential quantum mechanical interactions of a system through its electronic distribution. Density functional theory provides a good compromise between accuracy and computational cost [1]. It is capable of describing diverse interactions and has been successfully applied to the study of many-body systems, including complex protein peptides. It yields good agreement with the experimental studies of lattice vibrations, bulk moduli, equilibrium geometries, and binding energies of large systems [2]. Using density functional theory, we wish to study water molecules, which is the typical medium of most biological molecules, and the amyloid-forming peptide GNNQQNY that is essential to the understanding of amyloid related diseases.

Amyloid-like fibrils are unbranched protein peptides formed from different types of protein monomer. They are the causes of many amyloid-related diseases including Alzheimer's disease and Type II diabetes. Amyloid fibrils of different proteins

have a common structural pattern that suggests the existence of a common molecular structure [3]. Much experiments have been done on the amyloid-forming peptide GNNQQNY. They showed that the formation of the fibril depends on the nucleation. Their strength and stiffness are believed to be important to the growth and propagation of amyloid fibrils in disease [4].

The paper is organized as follows: In Chap. II, the density functional theory and Car-Parrinello molecular dynamics are presented. Kohn-Sham equations and Perdew-Burke-Ernzerhof exchange-correlation potentials are also described. In Chap. III, the computational methods used in the modeling are introduced. Chapter IV contains a study of the water molecules and amyloid-forming peptide GNNQQNY using Car-Parrinello molecular dynamics.

Chapter 2

Theory

2.1 Many-body Schrödinger Equation

A many-body system containing N particles is fully characterized by the solutions to the Schrödinger equation. Once the many-body wavefunctions that are solutions to the Schrödinger equation are known, all macroscopic properties of the system can be calculated.

A neutral system of atoms contains the electrons and the ionic cores. The Hamiltonian of the entire system includes the kinetic energy of the electrons (T_e), the kinetic energy of the ions (T_N), the electron-electron interaction (V_{e-e}), the electron-ion interaction (V_{e-N}), and the ion-ion interaction (V_{N-N}):

$$H = T_e + T_N + V_{e-N} + V_{e-e} + V_{N-N} \tag{2.1}$$

The first simplification to Eq. (2.1) is the Born-Oppenheimer approximation. Assuming that the time scale of the electronic motion is magnitudes smaller than the time scale of the ionic motion, it is to a good degree of accuracy that the ionic motion is treated classically, such that the time-average of the sum $T_e + V_{N-N}$ is constant over the time scale of the electronic motion. The ionic kinetic energy is determined by Newton's equation, and the ionic potential energy is determined by Coulomb's law. As a result, solving the Schrödinger equation is reduced to finding the electronic

many-body wavefunctions in the presence of an external potential $\sum_I V(\vec{R}_I - \vec{r}_i)$,

$$\begin{aligned} H &= T_e + V_{e-N} + V_{e-e} \Rightarrow \\ H &= -\frac{\hbar^2}{2m_e}\sum_i \nabla^2 + \sum_i [\sum_I V(\vec{R}_I - \vec{r}_i)] + \sum_i \sum_{j>i} \frac{1}{|\vec{r}_i - \vec{r}_j|} \end{aligned} \tag{2.2}$$

The summation of index I is performed over all the ions, and the summation of index i is performed over all the electrons in the system. For the rest of this paper, we will set $e = \hbar = m_e = 4\pi\epsilon_0 = 1$.

If the system has N electrons existing in 3-dimensional space, solving the eigenproblem of Eq. (2.2) would require simultaneous variation of $3N$ variables. This approach is not computationally feasible for dealing with large systems such as crystals or biological molecules. A more feasible approach is accomplished through density functional theory.

2.2 Density Functional Theory

The density functional theory was developed following the Thomas-Fermi model [5]. The key idea to density functional theory is the first Hohenberg-Kohn theorem [6], which is derived from variational principle[1]. The theorem states that an interacting system of fermions (in most systems, electrons) is completely and uniquely characterized by its density. This implies that if the external potential and the number of electrons are known, all system properties can be formulated as functionals of the charge density. Following this idea, the complexity of the problem is greatly reduced because the system now depends on only three parameters, the spatial x, y, and z.

The first Hohenberg-Kohn theorem allows the Hamiltonian in Eq. (2.2) to be rewritten as a functional of the charge density of the system $\rho_0(\mathbf{r}) = \rho_0(\mathbf{r_1}, \mathbf{r_2}, ..., \mathbf{r_N})$, in an external potential $v(\mathbf{r})$

$$E_{tot}[\rho_0(\mathbf{r})] = F[\rho_0(\mathbf{r})] + \int d\mathbf{r}\rho_0(\mathbf{r})v(\mathbf{r}) \ , \tag{2.3}$$

[1]see Parr [5] for the (short) derivation

$\mathbf{r_i}$ is the coordinate vector of i-th electron. $F[\rho_0(\mathbf{r})]$ is a functional that is the same for all systems with the same number of electrons, and is the sum of the kinetic energy $T[\rho_0(\mathbf{r})]$ and Hartree energy $E_H[\rho_0]$,

$$F[\rho_0(\mathbf{r})] = T[\rho_0(\mathbf{r})] + E_H[\rho_0(\mathbf{r})] \ . \tag{2.4}$$

The Hartree energy describes the electron-electron interaction and is given by

$$E_H[\rho_0] = \frac{1}{2}\int\int d\mathbf{r}d\mathbf{r}' \frac{\rho_0(\mathbf{r})\rho_0(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \ . \tag{2.5}$$

The second Hohenberg-Kohn theorem states that the minimum of $E_{tot}[\rho]$ occurs at the ground state density $\rho_0(\mathbf{r})$ [5]. Therefore, if $\rho_0(\mathbf{r})$ is known, $\frac{\delta E_{tot}}{\delta \rho_0(\mathbf{r})} = 0$. Let γ be the Lagrange multiplier associated with the constraint that there are N electrons in the system,

$$\int d\mathbf{r}\rho_0(\mathbf{r}) = N \tag{2.6}$$

then the density is related to the external potential through

$$\gamma = v(\mathbf{r}) + \frac{\delta(T[\rho_0(\mathbf{r})] + V_{ee}[\rho_0(\mathbf{r})])}{\delta\rho_0(\mathbf{r})} \ . \tag{2.7}$$

Let $\{\phi_i(\mathbf{r}_i, \mathbf{s}_i)\}$ be the one-electron wavefunction to Eq. (2.2), where n_i ($0 \leq n_i \leq 1$) is their occupation numbers, $\mathbf{r}_i$ is the three spatial coordinates, and $\mathbf{s}_i$ is the spin. The density and the kinetic energy of the N electron system are

$$\rho_0(\mathbf{r}) = |\Psi(\mathbf{r}, \mathbf{s})|^2 = \sum_i n_i \sum_s |\phi_i(\mathbf{r}_i, \mathbf{s}_i)|^2 \tag{2.8}$$

$$T[\rho_0(\mathbf{r})] = \sum_i n_i \langle\phi_i| -\frac{1}{2}\nabla^2 |\phi_i\rangle \ , \tag{2.9}$$

where the sum is performed over N one-body wavefunctions that are occupied [5]. The many-body wavefunction $\Psi(\mathbf{r}, \mathbf{s})$, where $\mathbf{r} = \{\mathbf{r_i}\}$ and $\mathbf{s} = \{\mathbf{s_i}\}$, is a solution to

Eq. (2.2). The one-body wavefunction is the solution to the equation

$$H_i = -\frac{\hbar^2}{2m_e}\nabla_i^2 + v_{\text{eff}} \tag{2.10}$$

where v_{eff} is the potential seen by an electron. The problem still difficult to solve because there are infinitely many number of terms in Eq. (2.9).

2.3 Kohn-Sham Equations

Kohn and Sham showed that it is possible construct a system of non-interacting electrons such that the ground state density ρ is equal the ground state of the interacting system ρ_0 [7]. Let $\{\psi_i(\mathbf{r}, \mathbf{s})\}$ be the one-body, spin-orbital wavefunctions of the non-interacting system where $\mathbf{r}$ is the three spatial coordinates. The density and the kinetic energy from Eq. (2.9) can be simplified

$$\rho(\mathbf{r}) = \sum_i \sum_s |\psi_i(\mathbf{r}, \mathbf{s})|^2 \tag{2.11}$$

$$T[\rho(\mathbf{r})] = \sum_i \sum_s \langle \psi_i | -\frac{1}{2}\nabla^2 | \psi_i \rangle \ . \tag{2.12}$$

The single-body orbitals $\{\psi_i(\mathbf{r}, \mathbf{s})\}$ satisfy the Kohn-Sham equations, in which the Hamiltonian from Eq. (2.2) reduces to that of a single-particle Schrödinger equation with an effective potential v_s,

$$H_{KS}\psi_i = (-\frac{1}{2}\nabla^2 + v_K S)\psi_i \tag{2.13}$$

where

$$v_{KS} = v_H + v_{xc} + v_{ions} \ . \tag{2.14}$$

The kinetic energy and the Hartree energy of the non-interacting system are known functionals, but they do not capture all the correlation and exhange of the fully interacting system. The total energy of the non-interacting system must have an additional term, the exchange-correlation energy E_{xc} that takes into account the

difference between the kinetic energies, $T[\rho_0]$ and $T[\rho]$.

$$E_v[\rho] = T[\rho] + \int d\mathbf{r} v(\mathbf{r})\rho(\mathbf{r}) + \frac{1}{2}\int d\mathbf{r}d\mathbf{r}' \frac{\rho(\mathbf{r})\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + E_{xc}[\rho] \ . \tag{2.15}$$

The effective potential v_s is the sum of the external potential, the Hartree potential, and the exchange correlation potential,

$$v_s(\mathbf{r}) = v(\mathbf{r}) + \int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + \frac{\delta E_{xc}[\rho]}{\delta\rho(\mathbf{r})} \ . \tag{2.16}$$

If E_{xc} is known exactly, then solving Kohn-Sham equations self-consistently would give us the exact many-body density and energy of the ground state. In practice, we approximate E_{xc} using local density approximation or generalized gradient approximation.

2.4 Car-Parrinello Molecular Dynamics

We want to find the global minimized energy $E_{tot}[\{\psi_i\}, \{\mathbf{R_I}\}]$. This is done by solving Kohn-Sham equations from the previous section self-consistently. We begin with an initial value of $\rho(\mathbf{r})$, calculate the potential v_s using Eq. (2.16). Next we use v_s to diagonalize the Hamiltonian in Eq. (2.13) in a chosen basis to find the eigenvectors $\{\psi_i\}$. The local minimization is performed on the potential energy, $E_0(\mathbf{R_I}) = min_{\psi_i} E_{tot}[\{\psi_i\}, \{\mathbf{R_I}\}]$, by varying $\{\psi_i\}$ at the fixed ionic positions. From them, a new $\rho(\mathbf{r})$ is calculated, and the process repeats until self-consistency [13].

In this way, the minimization of energy functional with respect to charge density is reduced to the minimization with respect to the one-body wavefunctions. Once the ground state ρ is found, the electrons and the ions are allowed to move. Electronic and ionic motions can be described in a single Lagrangian [2]. This is done by introducing a fictitious dynamics that describes the electronic motion.

To begin, let Λ be the Lagrange multiplier similar to the one defined in Eq. (2.6). The holonomic constraint is due to the orthonormality of the one-body wavefunctions,

which requires

$$\frac{\delta E}{\delta \psi_i^*(\mathbf{r},t)} = \sum \Lambda_{ij}\psi_j \ . \tag{2.17}$$

Since ions are treated classically, their dynamics can be formulated using the classical Lagrangian,

$$\mathcal{L}^{\mathrm{cl}} = \frac{1}{2}\sum_I M_I \dot{\mathbf{R}}_I^2 - V[\mathbf{R}_I] \tag{2.18}$$

Car-Parrinello technique allows the minimization on $\{\psi_i\}$ and $\mathbf{R}_I$ to be performed simultaneously by introducing a fictitious mass parameter μ_i. A new classical Lagrangian is defined to include both ionic and electronic dynamics. The Lagrangian is the sum of the electronic kinetic energy, ionic kinetic energy, potential energy of the coupled electron-ion fictitious system,

$$\mathcal{L} = \sum_i \int d\mathbf{r}\mu_i |\dot{\psi}_i(\mathbf{r})|^2 + \frac{1}{2}\sum_I M_I \dot{\mathbf{R}}_I^2 - E[\{\psi_i\},\{\mathbf{R}_I\}] \ . \tag{2.19}$$

The holonomic constraint is the orthonormality condition

$$\int d\mathbf{r}\psi_i^*(\mathbf{r})\psi_j(\mathbf{r}) = \delta_{ij} \ . \tag{2.20}$$

and the Euler equations are

$$\frac{\delta E}{\delta \psi_i^*} = \sum_j \Lambda_{ij}\psi_j \tag{2.21}$$

$$M_I \ddot{\mathbf{R}}_I = -\frac{\partial E}{\partial \mathbf{R}_I} \ . \tag{2.22}$$

Diagonalizing Λ yields the eigenstates and eigenvalues of the KS equations. In this way, the energy is optimization with respect to the electronic wavefunctions and all the interactions in the system. μ is chosen such that Born-Oppenheimer condition is satisfied: The fluctuation in the electronic kinetic energy is smaller than the energy gap between the ground state and the first excited state. This allows electrons to follow adiabatically the ionic motion, while remaining close to the ground states.

The constant of motion of the CP Lagrangian in Eq. (2.19) is defined to be

$$U_I = K_I + K_e + E[\{\psi_i\}, \{\mathbf{R}_I\}] \ . \tag{2.23}$$

If the electronic kinetic energy is small, $E[\{\psi_i\}, \{\mathbf{R}_I\}] \approx V[\{\mathbf{R}_I\}]$, and U_I is conserved approximately [13].

2.5 Kohn-Sham Equations with Ultrasoft Vanderbilt Pseudopotentials

It is useful to expand the single-particle wavefunctions $\{\psi_i\}$ of the Car-Parrinello method in a plane-wave basis. This allows the use of fast Fourier transform (FFT) in the calculation of the total energy and the forces on the ions. A disadvantage of using the plane-wave basis is its slow convergence in approximating the parts of the electronic wavefunctions that are close to the ions. This is dealt with using norm-conserving or ultrasoft Vanderbilt pseudopotentials that separately describes the valence electrons, the ionic core (the nucleus plus the core electrons), and their interactions [10].

Let ψ_i be the single-electron wavefunction of the Kohn-Sham Hamiltonian. The ultrasoft pseudopotential separates into two parts, a nonlocal part V_{NL} and a local part $V_{\text{loc}}^{\text{ion}}$,

$$E_{tot}[\{\psi_i\}, \{\mathbf{R}_I\}] = \sum_i \langle\psi_i| -\frac{1}{2}\nabla^2 + V_{NL} |\psi_i\rangle + E_H[\rho] + E_{xc}[\rho] \\ \int d\mathbf{r} V_{\text{loc}}^{\text{ion}}(\mathbf{r})\rho(\mathbf{r}) + U(\mathbf{R}_I) \ . \tag{2.24}$$

As in the previous section, $E_{xc}[\rho]$ is the exchange correlation energy, $U(\mathbf{R}_I)$ is the ion-ion interaction energy, and $E_H[\rho]$ is the Hartree energy. The ultrasoft pseudopotential is fully specified by $V_{\text{loc}}^I(\mathbf{r})$, $D_{\text{nm}}^{(0)}$, $Q_{\mathbf{nm}}(\mathbf{r})$, and $\beta_n(\mathbf{r})$. The local part $V_{\text{loc}}^{\text{ion}}$ is a sum of

atom-centered radial potentials,

$$V_{loc}^{ion}(\mathbf{r}) = \sum_I V_{\rm loc}^I(|\mathbf{r} - \mathbf{R}_I|) \ , \tag{2.25}$$

The nonlocal part,

$$V_{NL} = \sum_{{\rm nm},I} D_{\rm nm}^{(0)} \left|\beta_n^I\right\rangle \left\langle\beta_m^I\right| \ . \tag{2.26}$$

act as an augmented charge in the Hamiltonian. β_n^I depends on the displacement from the ionic positions,

$$\beta_n^I(\mathbf{r}) = \beta_n(\mathbf{r} - \mathbf{R}_I) \tag{2.27}$$

$\beta_n(\mathbf{r})$ and $D_{\rm nm}^{(0)}$ specify the atomic species. The single-electron wavefunction obeys the generalized orthonormality condition on ψ_i,

$$\langle\psi_i| S |\psi_j\rangle = \delta_{ij} \ , \tag{2.28}$$

and S is given by

$$S = 1 + \sum_{{\bf nm},I} q_{\rm nm} \left|\beta_n^I\right\rangle \left\langle\beta_m^I\right| \tag{2.29}$$

$$q_{\rm nm} = \int d\mathbf{r} Q_{\rm nm}(\mathbf{r}) \ . \tag{2.30}$$

The electron density is related to S through

$$\rho(\mathbf{r}) = \sum_i \langle\psi_i| S |\psi_i\rangle \ . \tag{2.31}$$

2.6 Car-Parrinello Molecular Dyanmics with Ultrasoft Vanderbilt Pseudopotentials

The Lagrangian is the same as Eq. (2.19). The modified holonomic orthonormality constraint is the generalized orthonormality condition given by Eq. The modified

Euler equations are

$$\mu_i \ddot{\psi}_i = -\frac{\delta E_{tot}}{\delta \psi_i^*} + \sum_j \Lambda_{ij} S \psi_j \tag{2.32}$$

$$M_I \ddot{\mathbf{R}}_I = -\frac{\partial E_{tot}}{\partial \mathbf{R}_I} + \sum_{ij} \Lambda_{ij} \langle \psi_i | \frac{\partial S}{\partial \mathbf{R}_I} | \psi_j \rangle \ . \tag{2.33}$$

Similarly, the equilibrium is achieved when the ionic acceleration $\ddot{\mathbf{R}}_I$ and electronic acceleration $\ddot{\psi}_i$ both vanish [10].

2.7 Perdew-Burke-Ernzerhof Exchange-Correlation Functional

The exchange-correlation energy E_{xc} is approximated using local density approximation (LDA). E_{xc} is given by weighting the exchange-correlation functional of the electron gas $e_{xc}[\rho(\mathbf{r})]$ with the density $\rho(\mathbf{r})$ of the non-interacting system,

$$E_{xc} = \int d\mathbf{r} \rho(\mathbf{r}) e_{xc}[\rho(\mathbf{r})] \ . \tag{2.34}$$

PerdewBurkeErnzerhof (PBE or General Gradient Approximation) exchange-correlation functional is frequently used in the CP simulations [11] [12], and will be used in this paper. PBE functional depends on the electronic density ρ, the reduced density gradient s, and the spin-polarization ξ [8]:

$$s = \frac{|\nabla \rho|}{2 k_F \rho} \tag{2.35}$$

$$\xi = \frac{\rho_\uparrow - \rho_\downarrow}{\rho} \tag{2.36}$$

k_F is the Fermi radius $(3\pi^2\rho)^{\frac{1}{3}}$. The Wigner-Seitz radius $r_s = (\frac{4\pi\rho}{3})^{-\frac{1}{3}}$ is used to perform calculation in the $\mathbf{r}$-space using

$$E_{XC}^{PBE} = \int d^3\mathbf{r} \rho(\mathbf{r}) \epsilon_{XC}^{PBE}(r_s(\mathbf{r}, s(\mathbf{r}), \xi(\mathbf{r})) \ . \tag{2.37}$$

Chapter 3

Computational Physics

3.1 Plane-wave Expansion

The system studied under the CP simulation can be a single unit of structure or a periodic structure. A single structure may be one water molecule. A 3-D periodic structure includes silicon crystals with long-range interactions, or water molecules in a large container. An example of a 1-D periodic structure is a one-branch polypeptide constructed from a single self-repeating unit of monomer. The interactions among many such monomers include the long-range hydrogen bonding. The short-range interactions include the covalent bonds.

The plane-wave basis introduces periodic boundary conditions in all cases. The system studied is defined in a computational supercell through the positions of the ions. The atoms are positioned such that the structure is close to the system in equilibrium. For the study of single-unit systems, the potential at the edges of the supercell must be small enough such that the system is prevented from interacting with its periodic images. This can be done by introducing a vacuum region a few-angstroms thick wrapped around the cell boundaries. For the study of the periodic structures, the periodic boundary conditions guarentee that the atoms will never wander out of the cell.

The Hamiltonian from Eq. (2.13) satisfies the Bloch theorem. The single-particle

wavefunction expanded in the plane-wave basis is

$$\psi_i^{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\mathbf{r}} \sum_{\mathbf{g}} c_i^{\mathbf{k}}(\mathbf{g}) e^{i\mathbf{g}\mathbf{r}} \tag{3.1}$$

The kinetic energy E_k is evaluated in the reciprocal space,

$$E_k = \frac{1}{2} \sum_{i,\mathbf{g}} g^2 c_i^*(\mathbf{g}) c_i(\mathbf{g}) \tag{3.2}$$

Assuming that the electronic wavefunctions are close to the ground states, the computation is truncated for the plane waves with a kinetic energy $E_k = \frac{1}{2}(\mathbf{k}+\mathbf{g})^2$ less than the threshold energy E_{cut}. Given a kinetic energy cutoff E_{cut}^{wf} on the wavefunctions, the charge density cutoff E_{cut}^{ρ} is usually made 2-3 times larger than $4E_{cut}^{wf}$. The three-dimensional FFT grid for charge density, the exchange-correlation, and Hartree potential calculations are determined by E_{cut}^{ρ}.

Convergence of the calculation depends on the cutoff wavelength, the fictitious electronic mass, and the time steps of integration. The accuracy of the calculation depends on E_{cut}, which is determined by the pseudopotential of the system. Norm-conserving, soft pseudopotentials require a E_{cut} of more than 70 Ry when elements such as O, N, F, and transition metals are used. Ultrasoft Vanderbilt pseudopotentials allow the electronic structures of O and N to converge efficiently ($E_{\text{cut}} \sim 25$ Ry) [8]. Ultrasoft pseudopotentials are used in our study of the protein molecules.

3.2 Electrostatic Energy of the Ions

The ion-ion energy is evaluated using Ewald summation [9][1]. Together with the KS energy, this completely characterizes the total energy of the system. Computational *ab initio* DFT is able to accurately and efficiently characterize the system properties such as chemical bonds, regardless whether they are metallic or covalent, or in the case of covalent bonds, whether they are single, double, or hybrid.

[1] see Mareschal [14] for a (complete) derivation

3.3 Molecular dynamics

3.3.1 Ion Dynamics

When the ions are allowed to move, the velocity is drawn from the Boltzmann distribution. Using the accelerations obtained from Eq. (2.33), Verlet algorithm performs the integration by expanding the ion position in power series one forward and one backward in time [13]:

$$r(t+\delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 + \frac{1}{6}b(t)\delta t^3 + O(\delta t^4) \tag{3.3}$$

$$r(t-\delta t) = r(t) - v(t)\delta t + \frac{1}{2}a(t)\delta t^2 - \frac{1}{6}b(t)\delta t^3 + O(\delta t^4)\ . \tag{3.4}$$

Substituting Eq. (3.3) into Eq. (3.4) gives

$$r(t+\delta t) = 2r(t) - r(t-\delta t) + a(t)\delta t^2 + O(\delta t^4)\ , \tag{3.5}$$

Verlet algorithm generates an accuracy of $O(\delta t^4)$ in position and $O(\delta t^2)$ in velocity. The accuracy depends on the size of the discrete time steps.

3.3.2 Electron Dynamics

If the energy gap between the ground state and the first excited state is large enough, the kinetic energy of the electronic energy is proportional to μ. We can choose a large value of μ and dt for the integration. The previous section indicates that dt should be chosen as large as possible, so that the computation can proceed quickly. In our CP simulation, each time step is 0.024189 fs. However, the atoms should not be moving so quickly that the adiabatic condition, in which the electronic density always remains in its ground state, is no longer satisfied. This might result in a diagonalization error on Λ.

In this case, dt needs to be decreased and electronic mass increased. Smaller dt improves the accuracy of the ion dynamics and ensures that the constant of motion of

the CP Lagrangian is conserved. The adiabaticity is better conserved in systems with a large electronic energy gap. When the dynamics is adiabatic, there is no transfer of energy between electronic and ionic degrees of freedom.

3.3.3 Damping

Stable biological molecules are often found at the minimum of the potential energy landscape. If they are not at the minimum, there are nonzero kinetic energy associated with both electronic and ionic motions. The system can be relaxed to attain a minimal potential energy through damping. Damping rate is applied externally to the system by taking the energy out of the system. There is a unique configuration of the system such that the potential energy is at the minimum, and none of the ions experience acceleration. At other configurations of the ions, the ions have acceleration, which results in velocity that change the the configuration at the next iteration in the CP molecular dynamics simulation.

A damping term can be added to either the ionic equation of motion or the electronic equation of motion. The damp force is proportional to the damping rate and the velocities. An initial estimate of the damping rate can be calculated according to

$$\text{rate} = \sqrt{\frac{1}{2}\log|\frac{E_1 - E_2}{E_2 - E_3}|}\,, \tag{3.6}$$

where E_1, E_2, E_3 are the successive values of the energy.

3.4 Car-Parrinello Simulation

To introduce the external stimuli to the systems such as a group of biomolecules, the thermodynamical properties of the biomolecules can be better described using statistical ensemble. External stimuli such as temperature can be incorporated into the CP Lagrangian [16].

A Nosé thermostat can be used to maintain the ions in the desired temperature and the electrons in the ground states. It uses velocity-dependent forces in the electronic

motion. The temperature we use is a Nosé thermostat that is applied to the system at a fixed frequency. The frequency is set at the vibrational spectrum of the system.

All four atomic species (H, O, N, C) in our simulation are described using ultrasoft pseudopotential. To satisfy the orthonormality condition, the tolerance for iterative orthonormalization is set to 10^{-8}, and the maximum number of iterations is set between 250-350. E_{cut}^{wf} is set to 25 Ry, and E_{cut}^{ρ} is set to 200 Ry. The pseudo-atomic radius of the atomic species used in Ewald summation are all set to 1. All generated by CP technique is measured in the atomic units: energy (1 Hartree energy or 1 H_h = 4.35974 $\times$ 10^{-18} Joule), electronic mass in a.u. and ionic mass in a.m.u. (1 a.u. of mass = 1/1822.9 a.m.u. = 9.10939 $\times$ 10^{-31} kg), length (1 bohr radius = 1 a_0 = 5.29177 $\AA$), and temperature (Kelvin).

The time needed for the systems to reach equilibrium in the presence of damping is controlled by the damping of the electrons and the damping of the ions. To accelerate the process, the initial configurations of the systems are made as close to the real physical systems as possible. Tangney et al. [17] showed that an exact adiabatic separation in a Car-Parrinello simulation can never be achieved even for small values of μ because the high frequency components in the ionic spectra that leads to direct coupling between the ionic and electronic degrees of freedom and the μ-dependent errors associated with the low-frequency components of the ionic motion.

Chapter 4

Simulations

4.1 Water

4.1.1 Motivation

Biological systems such as amyloid fibrils are often found in aqueous media, where they interact with the water molecules through hydrogen bonding. Therefore, understanding the properties and the dynamics of the liquid water is important in the study of the biological systems in an aqueous environment. Computational studies of water molecules using first principles molecular dynamics such as Car-Parrinello technique are found to be in good agreement with the experimental data [19].

We study the dynamical properties of liquid water using Car-Parrinello technique and PBE pseudopotentials. It is shown that a system of 32 water molecules per computational cell is sufficient to simulate the liquid state of water [19], and the electronic ground state can be accurately described in microcanonical CP simulations with a ratio of the electronic mass to ionic mass of $\frac{\mu}{M} < \frac{1}{5}$. The radial distributions of the wavefunctions and the behaviors of the water molecules were investigated using different pseudopotentials including PBE [18]. It is shown that the liquidlike state occurs at the elevated temperature of 400 K and above, and the computation time is approximately 50 ps to achieve well-converged thermal averages. A small cutoff for the plane-wave basis is sufficient for obtaining good results [19].

PAGES (S) MISSING FROM ORIGINAL

Pages 32 thru 34 missing

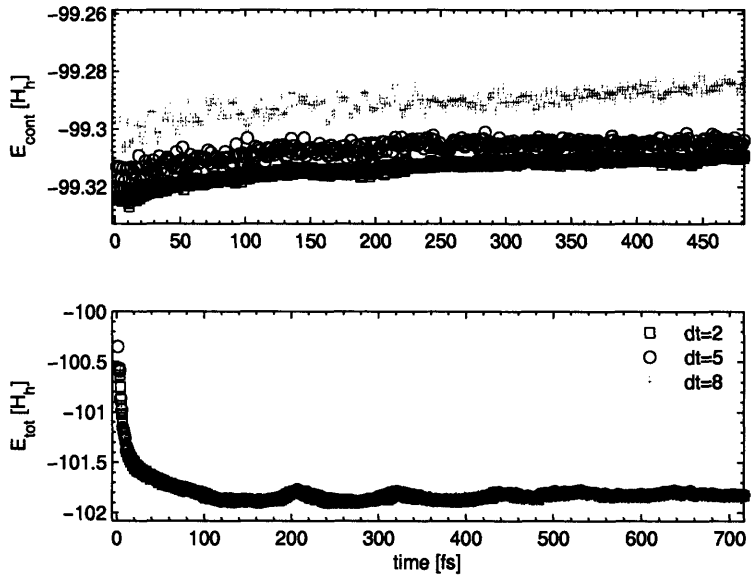

Figure 4-3: (a) Constant of motion and (b) Total energy when the Nosé thermostat is applied. A comparison of the data shows that the larger values of *dt* yield a faster convergence, but smaller values are better at preserving the adiabatic condition.

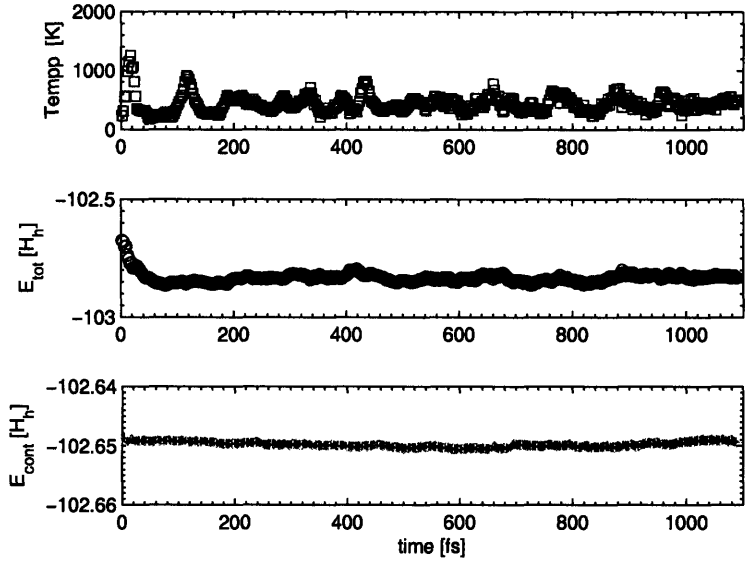

Figure 4-4: The constant of motion and the total energy with μ = 450 a.u and *dt* = 10 when the Nosé thermostat is removed. They are approximately constant, as expected of an isolated system.

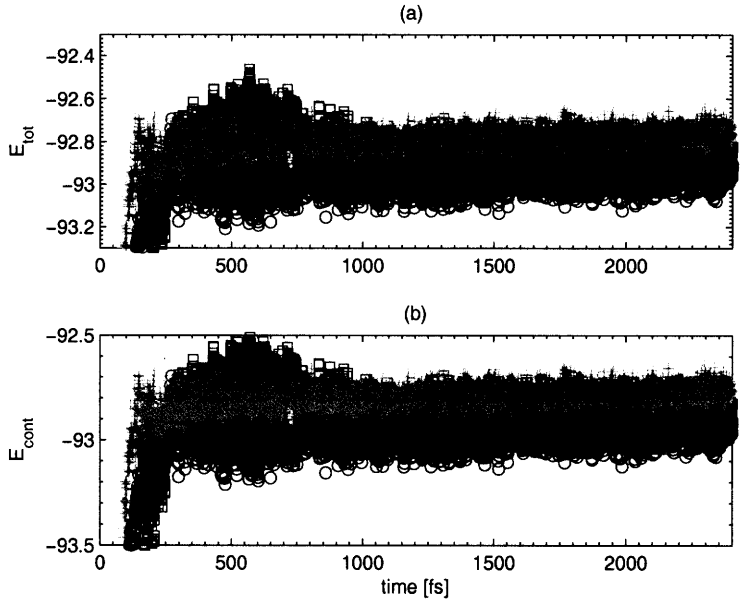

Figure 4-5: (a) Total energy (b) Constant of motion. The convergence of different electronic masses, under no external damping or temperature. The constant of motion are approximately constant.

4.2 Amyloid-Forming Peptide

4.2.1 Motivation

The amyloid-forming peptide GNNQQNY is a seven-residue, fibril-forming peptide from yeast prion. It is soluble in water at a low concentration, in which the interaction between the monomers and the water molecules is predominantly hydrogen bonding. A pair of GNNQQNY molecules bind together through the Asn and Gln side chains and form a basic unit called the steric zipper. At a high concentration, the steric zippers form insoluble amyloid-like fibrils that are the causes of amyloid-related diseases [4]. The neighboring steric zippers interact through weak van der Waals forces as shown in Figure 4-7.

Many experiments have been done in the study of the structure and the properties of the amyloid fibrils including the amyloid-forming peptide GNNQQNY . X-ray microscrystalllography shows that the amyloid fibrils formed from different pro-

teins share a structural cross-β pattern that suggests a common molecular structure. Nuclear mangetic resonance (NMR) spectroscopy shows that there are four atomic species (H, O, N, C) totaling 214 atoms in a steric zipper. The micromechanical experiments on the fibril suggest that the fibril has a strength comparable to that of steel and a stiffness comparable to that of silk. In particular, Smith et. al. have shown experimentally that the fibril has a strength of 0.6 ± 0.4 GPa and a Young's modulus of 3.3 ± 0.4 GPa [4].

The amyloid formation has been studied using classcial molecular dyanimcs, and recently Tsemekhman et. al. showed that CP technique is in good agreement with the classical technique. Although a fibril contains a finite number of GNNQQNY molecules, it is well-approximated if assumed to be infinite in length [20]. We wish to use first principles molecular dynamics to estimate the Young's modulus of the GNNQQNY fibril-forming peptide. The objective is to estimate the Young's modulus and check its agreement with the experimentally measured values [21].

4.2.2 Structure of the Cross-β Spine

We use the experimentally determined parameters of the cross-β spine of given by Nelson et. al. [21]. The supercell is defined to be a monoclinic lattice of 21.94 $\mathring{A} \times$ 23.48 $\mathring{A} \times$ 4.87 $\mathring{A}$. The monomer has 180° rotation symmetry in the plane spanned by two lattice vectors that are perpendicular to the axis of elongation, as shown in Figure 4-7. The angle between the two lattice vectors is 107.08°.

The computational cell size along the axial direction is changed to simulate the stretching of the fibril. The stretching increases or decreases the distances between two neighboring GNNQQNY molecules, leaving the bonding within a GNNQQNY molecule unchanged. Based on the experimental studies of the fibril's micromechanical properties, we expect that there is an optimal cell size for which the total energy is the minimum.

The total energy as a function of the elongation can be expanded in a power series about the minimum of the total energy. Young's modulus is related to the second derivative through a constant. To estimate the precision required of the total energy,

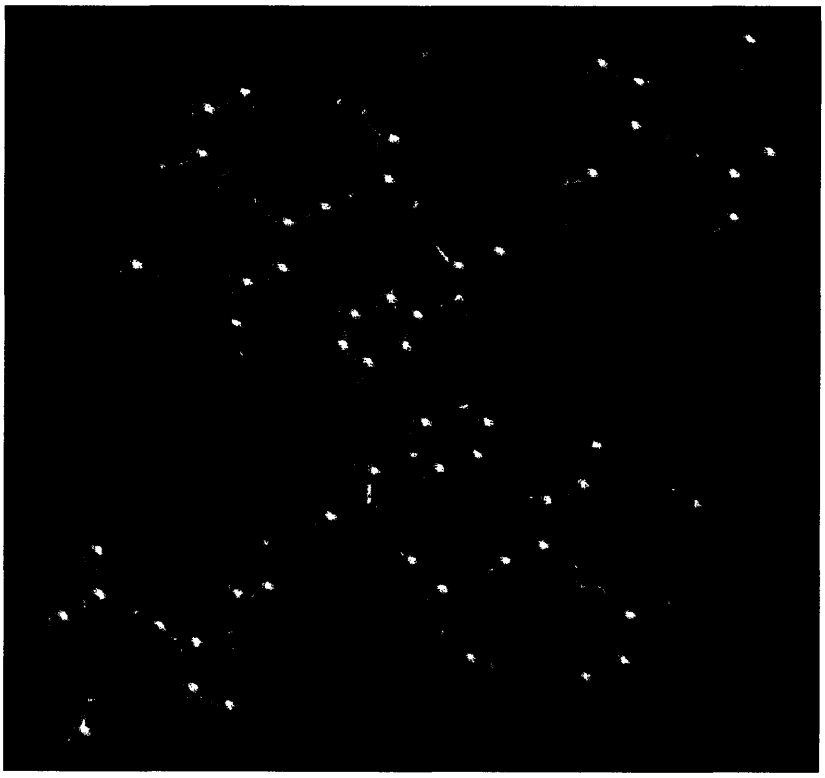

Figure 4-6: The initial configuration of a pair of GNNQQNY confined in a computational supercell.

let δE be the change in total energy, Δ be the change in $\frac{d^2E}{dc^2}$,

$$E = E_0 + \frac{1}{2}\frac{d^2E}{dc^2}(c - c_0) \tag{4.1}$$

$$E + \delta E = E_0 + \frac{1}{2}(\frac{d^2E}{dc^2} + \Delta)(c - c_0)\ . \tag{4.2}$$

This implies that $\delta E = \frac{1}{2}\Delta(c - c_0)$.

Using CP molecular dynamics, the electron density of the GNNQQNY molecules is first relaxed to the ground state by fixing the ionic positions and setting the electron damping rate to a finite value. The CP simulation of the water molecules suggests an electronic damping rate between 0.5 and 1. With the fictitious electronic mass μ set to 350 a.u., the CP technique converges when dt is 2. The relaxation takes less than 700 time steps.

After electronic relaxation, the damping rate of ion and the electrons are set to

Figure 4-7: The periodic images of a pair of GNNQQNY molecule. The direction of elongation is perpendicular to the page.

non-zero values. The computational medium is T = 0 K in vacuum. The initial guess of the damping rate is estimated from the CP simulation of the water molecules, and after a hundred time steps, from Eq. (3.6). Different rates of convergence are found after a few thousand times steps by trying out different combinations of electronic and ionic damping rates about the intial estimated values. To generate the fastest convergence, the combinations of the electronic and ionic damping rates are found such that the energy is critically damped. Figure 4-9(b) shows that the combination of electron rate of 0.03 and and ion rate of 0.001 gives the fastest convergence.

4.2.3 Relaxation

The constant of motion always has a higher energy since it is the sum of the electronic kinetic energy and the total energy of the Kohn-Sham Hamiltonian.

Damping rates are optimized when the system is critically damped. The damping rate of the electrons is varied between 0.01 and 0.1. The damping rate of ions is varied between 0.001 and 0.01. The damping rates might be too high such that forces and accelerations of the ions become zero. In this case, the system has the apparance of having converged. To check if it is truly converged, the damping rates on both the

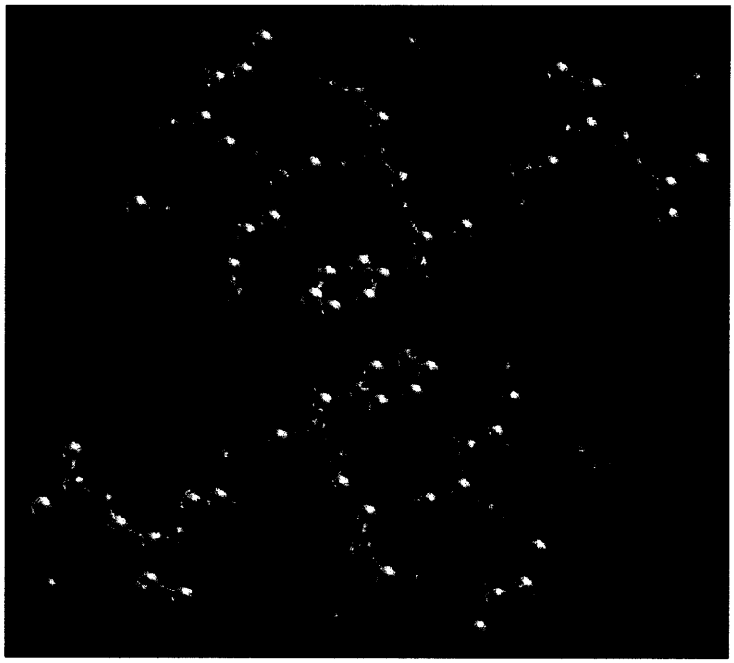

Figure 4-8: The configuration of the peptide after being relaxed to the minimum of the potential well. Since the initial configuration of the GNNQQNY molecule is well chosen, the final configuration is very close to the initial configuration. Atoms that follow random motion and wander out of a side of the cell re-appear at the other side.

ions and electrons must be removed. The energy is at the minimum of the energy landscape if the ions do not accelerate (above 10^{-4}). When the system is converged, the forces, the acceleration, the temperature of the ions, the kinetic energy of the electrons, and the mean square displacement are below 10^{-4}. The masses of the ions are given the same value of 12 a.u., which does not influence the total energy and makes the assessment on the magnitudes of the ionic acceleration easier. In the absence of a thermostat, the total energy of the Kohn-Sham Hamiltonian is equal to the constant of motion of the CP Lagrangian. The variation in total energy should be less than 10^{-5}. The mean square displacement and temperature are zero. The force is below 10^{-4}. As shown in Figure 4-10, the system is underdamped and critically damped.

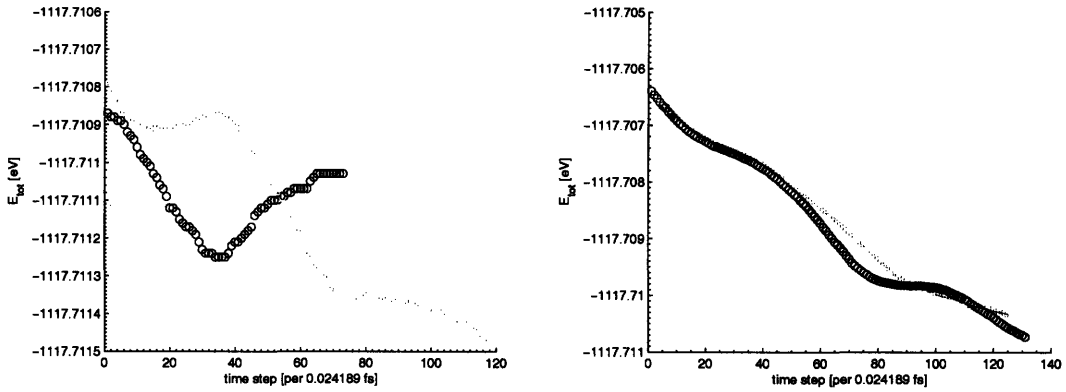

Figure 4-9: A comparison of different damping rates with c = 4.87 $\mathring{A}$. (a) $\gamma_{ion} = 0$ with $\gamma_e = 0.05, 0$. (b) $\gamma_e = 0.03$ with $\gamma_{ion} = 0.001, 0$. The energy that is overdamped is exponentially decaying. The energy that is underdamped is exponentially oscillatory.

4.2.4 Young's Modulus

Once the system is converged, we want to simulate the elongation of the fibril. For the purpose of computational efficiency, the supercell size is increased by a small percentage and the initial positions of the ions are the equilibrium positions from a previous simulation. Since the cell is stretched by a small amount, the system is not too far away from its minimal configuration. Same damping rates of 0.03 for the ions and 0.001 for the electrons are used in all the simulations.

Let c be the size of the computationl cell in the direction of elongation. If CP technique is precise and c corresponds to an energy that is close to the minimum of the potential well, it is possible to estimate the Young's modulus well enough with only a few data points. The three points are selected about 4.87$\mathring{A}$, which is experimentally found by Nelson et. al. [21]. The converged values of the total energy is shown in Table 4.1.

Let Y be Young's modulus, which is also the elastic constant of the potential energy of the system. Let c_0 be the size of the cell at the minimum potential energy, and A the cross sectional area of the fibril. Young's modulus is defined as

$$Y = \frac{\partial^2 E_{tot}}{\partial c_0} \frac{c_0}{A} \tag{4.3}$$

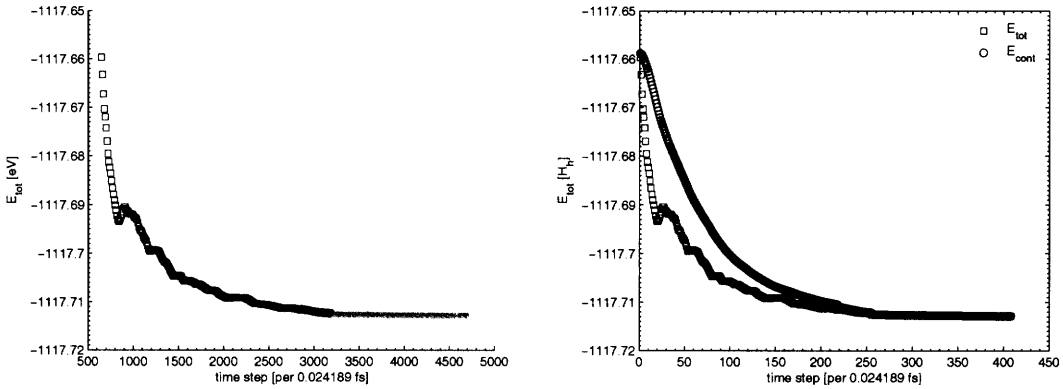

Figure 4-10: (a) c = 4.87 $\mathring{A}$, convergence of the total energy. The following damping rates γ_e and γ_{ion} are applied successively: γ_e= 0.02, γ_{ion} = 0; γ_e= 0.03, γ_{ion} = 0.001; γ_e= 0.03, γ_{ion} = 0. (b) A comparison between the convergence of total energy and the constant of motion.

Table 4.1: The system is consider to be converged when the forces on the ions are below 10^{-4}.

c	E_{tot} (H_h)	E_{cons} (H_h)	E_{kinc}(H$_h$)
4.67	-1117.75212	-1117.75212	0.0000
4.87	-1117.71290	-1117.71290	0.0000
4.97	-1117.68522	-1117.68522	0.0000

For our system, the total energy equals the potential energy. According to Figure 4-12, c_0 is close to 4.405$\mathring{A}$. The area of the fibril is approximately the area of the monoclinic computational cell that is perpendicular to the direction of elongation. This yields a Young's modulus of 20.0 GPa.

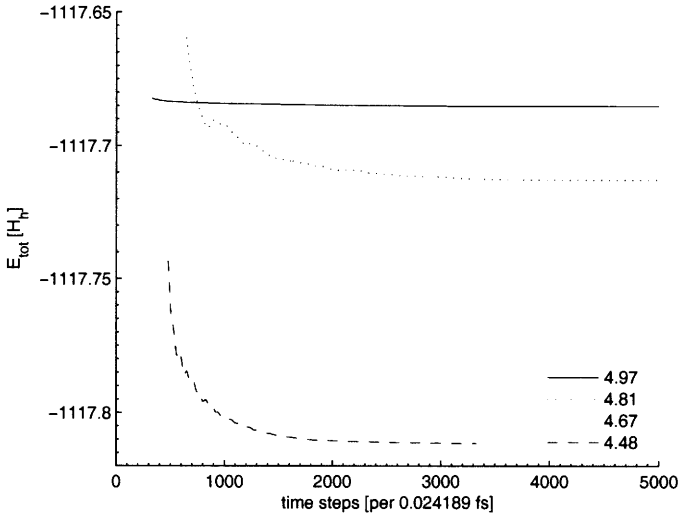

Figure 4-11: A comparison of the convergence of the total energy with different cell sizes c. The larger the difference between the initial energy and the final energy, the longer it takes the system to converge.

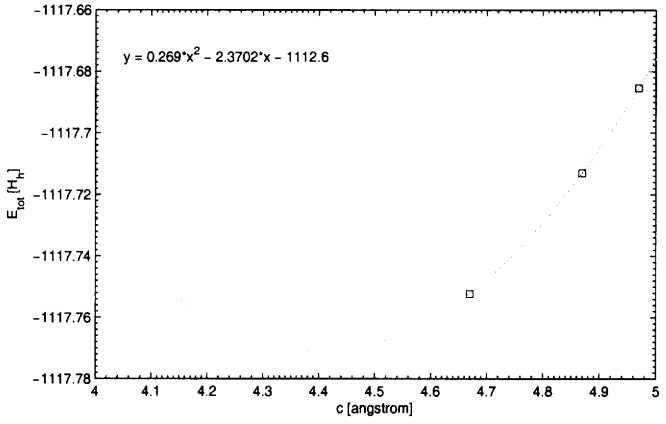

Figure 4-12: The data points are fitted with a parabola. The cell size c_0 that produces the minimum total energy is approximately $4.405\mathring{A}$.

Chapter 5

Conclusion

Car-Parrinello molecular dynamics are shown to be computationally efficient in the simulation of the structural and dynamical properties of complex biomolecules. Macroscopic properties such as Young's modulus are shown to be derived from first principles molecular dynamics. In particular, the study of the amyloid-forming peptide GNNQQNY using first-principles molecular dynamics generates a Young's modulus that is six times larger than the experimentally reported value. A good estimation of the cross sectional area of the fibril is needed for better estimations of the mechanical properties.

The study of water molecules and the amyloid-forming peptide GNNQQNY in this thesis demonstrates the effectiveness of using Car-Parrinello molecular dynamics to obtain the macroscopic properties of large, complex systems. Although Car-Parrinello molecular dynamics is not limited to the descriptions of biological molecules, an understanding of the electronic properties of such systems and a control over the systems' constituents can potentially aid the experimental studies. For example, it is very simple to introduce water molecules or transition metal atoms to the fibril. Their interactions might influence the fibril's micromechanical properties that can be medically beneficial. Such possibilities are endless and suggests that there is much promising research work to be done using first principle molecular dynamics.

Bibliography

[1] R. O. Jones and O. Gunnarsson. The density functional formalism, its applications and prospects. *Rev. Mod. Phys.*, 61:689-746, July 1989.

[2] R. Car and A. Pasquarello. Unified approach for molecular dynamics and density functional theory. *Phy. Rev. Let.*, 55:2471-2474, August 1985.

[3] P. C. A. Wel, K. N. Hu, J. Lewandowski, and R. G. Griffin. Dynamic Nuclear Polarization of Amyloidogenic Peptide Nanocrystals: GNNQQNY, a Core Segment of the Yeast Prion PRotein Sup35p. *J. Am. Chem. Soc.*, 128:10840 -10846, August 2006.

[4] J. F. Smith, T. P. J. Knowles, C. M. Dobson, C. E. MacPhee, and M. E. Welland . Characterization of the nanoscale properties of individual amyloid fibrils. *Proc. Natl. Acad. Sci. U.S.A.*, 103:1580615811, October 2006.

[5] R. G. Parr *Density-Functional Theory of Atoms and Molecules.* Oxford University Press, U.S.A., Reprint, June 2003.

[6] P. Hohenberg and W. Kohn. Inhomogeneous Electron Gas. *Phys. Rev.*, 136:B864 - B871, June 1964.

[7] W. Kohn and L. J. Sham. Self-Consistent Equations Including Exchange and Correlation Effects. *Phy. Rev.*, 140:A1133-A1138 , June 1965.

[8] P. Giannozzia, F. D. Angelis, and R. Car. First-principle molecular dynamics with ultrasoft pseudopotentials: Parallel implementation and application to extended bioinorganic systems. *J. Chem. Phys.*, 120:5903-5915, April 2004.

[9] J. M. Thijssen *Computational Physics.* Cambridge University Press, U.K., 2006.

[10] K. Laasonen, A. Pasquarello, R. Car, C. Lee, and D. Vanderbilt. Car-Parrinello molecular dynamics with Vanderbilt ultrasoft pseudopotentials. *Phy. Rev. B,* 47:10142-10153, April 1993.

[11] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.*, 77:3865-1996, May 1996.

[12] M. Ernzerhofa and G. E. Scuseria Assessment of the PerdewBurkeErnzerhof exchange-correlation functional. *J. Chem. Phys.*, 110:5029-5036, December 1999.

[13] D. J. Tildesley and M. P. Allen. *Computer Simulation in Chemical Physics.* Springer, London, 1st edition, 1993.

[14] M. Mareschal, G. Ciccotti, and P. Nielaba *Bridging the Time Scales: Molecular Simulations for the Next Decade.* Springer, U.S., 1st Edition, 2002.

[15] A. Leach. *Molecular Modelling: Principles and Applications.* Prentice Hall, Essex, England, 2nd edition, 2001.

[16] D. Frenkel and B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications.* Academic Press, London, 2nd edition, 2001.

[17] P. Tangney and S. Scandolo. How well do CarParrinello simulations reproduce the BornOppenheimer surface? Theory and examples. *J. Chem. Phys.*, 116:14-24, January 2002.

[18] E. Schwegler, G. C. Grossman, F. Gygi, and G. Galli. Towards an assessment of the accuracy of density functional theory for first principles simulations of water II. *J. Chem. Phys.*, 121:5400-5409, September 2004.

[19] H. L. Sit and N. Marzari. Static and Dynamic Properties of Heavy Water at Ambient Conditions from First-Principles Molecular Dynamics. *J. Chem. Phys.*, 122:204510, May 2005.

[20] K. Tsemekhman, L. Goldschmidt, D. Eisenberg, and D. Baker. Cooperative hydrogen bonding in amyloid formation. *Protein Sci.*, *doi:10.1110/ps.062609607*, February 2007.

[21] R. Nelson1, M. R. Sawaya, M. Balbirnie, A. Madsen, C. Riekel, R. Grothe1, and D. Eisenberg. Structures of the cross-β spine of amyloid-like fibrils. *Nature*, 435:773-778, June 2005.