ENERGY LANDSCAPES : APPLICATION TO SILICON NANOCLUSTERS AND PROTEIN STABILITIES

by

Soohaeng Yoo

A DISSERTATION

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirements

For the Degree of PhD of Science

Major: Chemistry

Under the Supervision of Professor X.C. Zeng

Lincoln, Nebraska

December, 2004

DISSERTATION TITLE

Energy Landscapes : Application to Silicon Nanoclusters and

Protein Stabilities

BY

Soohaeng Yoo

SUPERVISORY COMMITTEE:

Approved	Date
Signature	11-30-04
Professor Xiao Cheng Zeng Typed Name	
Signature	11-30-'04
Professor Lawrence Parkhurst Typed Name	
Signature	11/30/04
Professor John Stezowski Typed Name	
Signature	11/30/04
Professor Hong Jiang Typed Name	
Signature	11/30/04
Professor Dennis J. Diestler Typed Name	
Signature	
Typed Name	

University of Nebraska Lincoln

ENERGY LANDSCAPES : APPLICATION TO SILICON NANOCLUSTERS AND PROTEIN STABILITIES

Soohaeng Yoo, Ph.D.

University of Nebraska, 2004

Advisor: X.C. Zeng

Statistical thermodynamics provides the relations between thermodynamical information and microscopic information. Energy landscapes provide the information of macroscopic phenomena, *i.e* phase transitions, protein folding problems, via the statistical thermodynamics. Among energy landscapes, potential energy surfaces, which are determined by the exact positions of all particles, are useful to investigate the global minima of clusters in conjunction with the global search techniques while free energy surfaces, which depend on the additional parameters such as temperature T, have been used to study the free-energy barrier between two thermal (*meta*)stable states as a function of order parameters or reaction coordinates. In this work, the concept of the energy landscape is applied to nanoscaled semiconductor clusters, phase transitions, and thermal stabilities of proteins. The quantum-mechanical calculations combined with the global search techniques have been performed to unveil the structural information of low-lying silicon clusters Si_n ($13 \leq n \leq 39$) since their detail morphologies still cannot be inferred directly from experiments. Free energy surface as a function of a proper order parameter is employed for the temperature dependence of protein's thermal stability resembling the macroscopic first-order phase transition behavior.

ACKNOWLEDGEMENTS

First and foremost, to my advisor, Prof. Xiao Cheng Zeng, for his patience, faith, superb guidance, and insightful discussion on this work. His confidence in me enabled me to bring this dissertation to fruition. And to committee members, Prof. Lawrence Parkhurst, Prof. John Stezowski, Prof. Hong Jiang, and Prof. Dennis J. Diestler for their support in the completion of this work.

I also owe my gratitude to the exceptional members at the Prof. Zeng's group, Dr. Ruben Parra, Dr. Guangtu Gao, Dr. Kwang Jin Oh, Dr. Vadim Warshavsky, Dr. Tikhon Bykov, Dr. Jaeil Bai, and Satya Bulusu for their roles on part of my work. Thank you especially to Dr. Xiaolei Zhu, Dr. Jijun Zhao, Dr. Jinlan Wang, Dr. Y.A. Lei, Dr. James R. Morris, Dr. Takahiro Koishi, and Prof. Kenji Yasuoka for their generous assistance in my researches.

My deepest gratitude goes to my parents and younger brothers in Korea, who always pray for me and my family. And finally, to my extraordinary family who have touched my life. First, my wife Hae-Yeon Jung has believed and supported me tirelessly from the start. My son Davin J. Yoo and unborn baby have brought happiness to my life.

Contents

Chapter 1

Introduction

Statistical thermodynamics is used to build a bridge between classical thermodynamics and molecular physics for the systems in thermal equilibrium. By counting the properties of individual molecules consisting of the system, statistical thermodynamics can provide the information of macroscopic phenomena, *i.e.* phase transitions, protein folding problems, and chemical reactions. The calculation of microscopic properties depends highly on the energy landscape. Thus, the energy landscape plays a key role in explanation of macroscopic thermodynamic properties.

In discussion of energy landscape, energy includes the potential energy, Helmholtz free energy, and Gibbs free energy. The relation of potential energy surface with free energy surfaces will be discussed in Chapter 2 since potential energy surface depends only on the the coordinates of all the particles, whereas free energy surface is varied as a function of external parameters, i.e. temperature T and pressure p. A thermal stability, which is correlated with thermal equilibrium, will be considered since it will be based on the discussion of phase transitions, chemical reactions, and also protein folding/unfolding problems. In short, the barrier ΔF^* between two thermal stable phases (reactants and products for chemical reactions, or folded and unfolded proteins for protein folding

problem) can be measured as a rate constant k via the relation of $k \propto exp(-\Delta F^*/k_B T)$, where k_B is the Boltzmann constant.

In the first part of this thesis (Chapters 3 – 5), I will focus mainly on the potential energy surface of silicon clusters Si_n. Unveiling of low-lying geometric structures of Si_n ($13 \leq n \leq 39$) will be the important step to identify the chemical and physical properties of nanoscale clusters furthermore. Since the empirical potentials have the limit in the description of *realistic* potential energy surface, the *ab initio* calculations have been used in this thesis to have reasonable low-lying geometric structures resulted from searching on *realistic* potential energy surface.

In Chapters 6-7, I turn to free energy surfaces and investigate the calculation of transition barrier (or nucleation barrier) for the vapor-liquid phase transition and the protein unfolding reaction. Both the phase transition and the protein unfolding reaction have the common macroscopic first-order transition as a function of temperature T. Specially, in Chapter 7, the qualitative and quantitative behaviors will be provided for the temperature dependence of thermal stability of protein.

Chapter 2

Energy Landscape

2.1 Introduction

Energy landscapes hold a key to understand a wide range of molecular phenomena.[1] Here, energy is considered one of the Gibbs free energy $G(N, P, T)$, Helmholtz free energy $A(N, V, T)$, and potential energy $U(N)$. For example, the potential energy surface (PES) itself is often investigated in studies of clusters, whereas for proteins and other biomolecules it is also common to define free energy surfaces, which are expressed in terms of order parameters or reaction coordinates. Energy Landscape is used to describe the qualitative and quantitative behaviors of various systems from molecular clusters, through protein folding problems, to liquid-liquid phase transition.[1]

2.2 Potential Energy Surface

For molecular clusters and viscous liquids, the system point depends on how a collection of N particles are arranged, because the potential energy $U(N)$ is determined by the exact positions of all the particles. The potential energy landscape $U(N, \mathbf{r}_1, \cdots, \mathbf{r}_N)$,

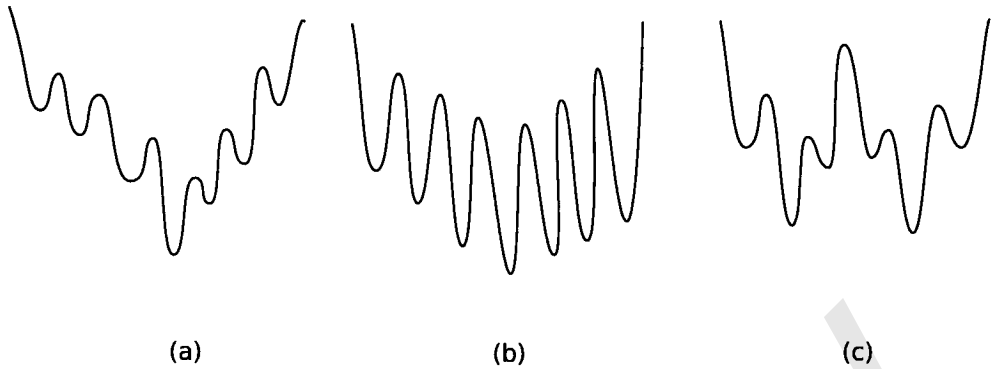

Figure 2.1: The possible one-dimensional potential energy surfaces for not only one funnel with (a) low downhill barriers and (b) higher downhill barriers but also (c) two funnels.

where $\mathbf{r}_i$ is the coordinates for the particle i, is a surface with $3N + 1$ dimensions, and impossible to conceptualize properly. Nevertheless, a one-dimensional representation is useful and can illustrate the distinctions between all possible (*meta*)stable states of liquid, crystal, and glass (see Figures 2.1 and 2.3).

First example shown in Figure 2.1a is a PES with the well defined global minimum and relatively small downhill barriers. Since the downhill barriers are accessible kinetically, it might be expected that the global potential energy minimum is relatively easy to locate. These systems are therefore likely to be good 'structure-seekers' or capable of 'self-assembly'. Since the simple lattice-type protein models provide this PES with a single funnel as shown in Figure 2.1a, probably it has been considered that the energy landscapes for protein folding might have a single 'folding funnel' into a native protein.[2–5] In Figure 2.1b, the PES has a well defined global minimum with a single funnel, but the downhill barriers that separate adjacent minima are larger than kinetically accessible barriers. Due to higher downhill barriers, the global minimum will be rare to locate via simulated annealing method. In the last example shown in Figure 2.1c the PES have two more funnels. One funnel has different geometrical features from others as well

as is far from others in terms of reaction coordinates. Here, geometrical features might indicate liquid .vs. crystalline phases (see Section 2.4 and Figure 2.3) or elongated .vs. spherical isomers for silicon clusters (see Chapter 4).

Since semiconductor clusters are special class of matters with size in between a single atom and semiconductor quantum dots,[6] small-to-medium sized semiconductor clusters have received considerable attention due to their potential relevance and application to the silicon-based nanoelectronic industry. In this thesis, possible low-lying geometric structures of small-to-medium sized silicon clusters have been unveiled via searching on their PES (see Chapters 3 – 5). Searching of the global minimum on the multifunnel-like PES will be necessary to mount all of potential energy barriers between funnels. The computational scheme of global search method will be discussed in Section 2.3. Note that these potential energy barriers between funnels shown in Figure 2.1c have no relation with a rate constant measured by experiments (see Section 2.5).*

2.2.1 The Calculation of Potential Energy Surface

First-principle (or *ab initio*) techniques can be considered at first since they provide more *realistic* potential energy surfaces (PES) of clusters and also their ground state electron densities. The highest accuracy is achieved by Configuration Interaction (CI), Many Body Perturbation Theory (MBT) including properly both electronic exchange and correlation within a many-body treatment.[7] But these calculations are computationally very demanding and can be used only for molecules with up to about 20 atoms. An alternative scheme is Density Functional Theory (DFT). It is based on the electron density rather than wave functions and treat both exchange and correlation by local-density and generalized gradient approximations (LDA/GGA). The many-body problem of in-

*For the chemical reactions between two reactants in the gas phases, their free energy barriers are close to their potential energy barriers.

teracting electrons and nuclei is mapped to a series of one-electron equations, the so-called Kohn-Sham (KS) equations. A variety of implementation within DFT are from basis representations, including Gaussian-type orbitals, Slater-type orbitals, numerical basis sets, LMTOs, plane waves, or a mixed-type basis. Due to the high energy cut-off required for the plane-wave basis, these schemes become computationally very demanding for larger structures including transition metal clusters. DFTs are still too slow to perform global search directly. Since performing more than thousands of gradient minimizations requires fast and accurate evaluation of energies and atomic forces, the semi-empirical methods become the next choice. Among them, tight-binding (TB) models incorporates quantum effects beyond classical empirical potentials by direct modeling of the electronic structure. Recently density functional TB (DFTB) scheme[8] has been developed and applied successfully to silicon clusters.[9,10]

2.3 Global Search Algorithms

A number of methods have been developed for seeking global minima.[11] An early one is the simulated annealing method, which attempts to mimic real annealing experiments, namely, the target system is gradually cooled toward the zero temperature after being equilibrated at high temperatures. It is known that the simulated annealing method can be inefficient to locate the global minima of nanoclusters since the system can be easily trapped in some metastable configurations when temperature becomes too low in comparison with downhill barriers existing between adjacent minima (see Figure 2.1b).[12]

2.3.1 Genetic Algorithm

One of mostly used global search methods is a genetic algorithm (GA) method.[13,14] This is based on Darwin's theory of biological evolution of real life. The GA optimization

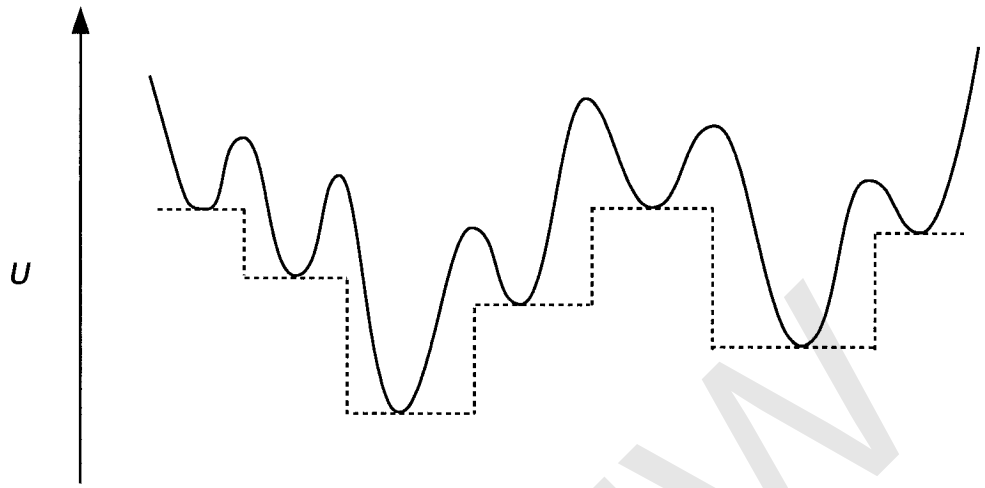

Figure 2.2: The effect of the basin transformation described by equation 2.1 on a one-dimensional PES. The solid and dashed lines represent the original and transformed potential U and $\tilde{U}$, respectively.

generally starts with a population of random structures (*parent*). Then three operations, *selection*, *crossover*, and *mutation* are used for the generation of *offspring* i.e. *mating*. The new population of *offsprings* are selected according to their 'fitness' (biological evolution). The fittest members of new offsprings are selected to be *parent* in the next generation. Even the GA method has been successful in searching the global minima for the multifunnel-like PES, the amount of computer memory and CPU time required for the GA calculation can increase very fast as the number of populations increases.

2.3.2 Basin Hopping Method

Recently, another global optimization technique, a basin-hopping (BH) method, has been developed and applied to the Lennard-Jones clusters[11,15,16] and a variety of atomic and molecular clusters as well as to peptides, polymers, and a glass-forming solid.[16–20] The efficiency of BH method has been proved since it can locate the global minima in case of the multifunnel-like PES, i.e. LJ_{38} and LJ_{75}. Here we give a brief summary of

the 'basin-hopping' global optimization technique. More details about this technique can be found elsewhere.[1,15,16] By using the 'basin-hopping' method, the transformed PES $\tilde{U}$ is generated via the mapping

$$\tilde{U}(N, \mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N) = \mathbf{min}\{U(N, \mathbf{r}_1, \mathbf{r}_2, \cdots, \mathbf{r}_N)\}, \tag{2.1}$$

where **min** denotes that the energy minimization is performed with starting configuration of $\{\mathbf{r}_1, \cdots, \mathbf{r}_N\}$ (see Figure 2.2). The topography of the transformed potential surface will resemble a multidimensional staircase, with each step corresponding to the basin of attraction The basin of attraction represents a set of geometries from which energy minimization always leads to the local minimum. With the transformed PES, the intra-potential-well vibration can be removed, thereby the system can 'hop' directly between local minima at each step.

In practice, the transformed PES $\tilde{U}$ can be explored via canonical Monte Carlo (MC) simulation. For example, at each MC step all coordinates are randomly displaced with an adjustable step size to yield an acceptance ratio of 0.5. The energy change $\Delta\tilde{U}$ for hopping between two minima is accepted with the probability of $\exp(-\Delta\tilde{U}/k_BT)$.

2.4 Free Energy Surface

In the previous Sections, we give attention to the PES and the ways of how to access the global minimum point. Since free energy surfaces depend on the external parameters such as temperature T and pressure p, it will be necessary to underlie how free surface surface changes as a function of external parameters. In this Section, we will consider mainly a system at the constant volume V and constant temperature T, which is known as the canonical ensemble. Figure 2.3 shows the simple one-dimensional schematic plot

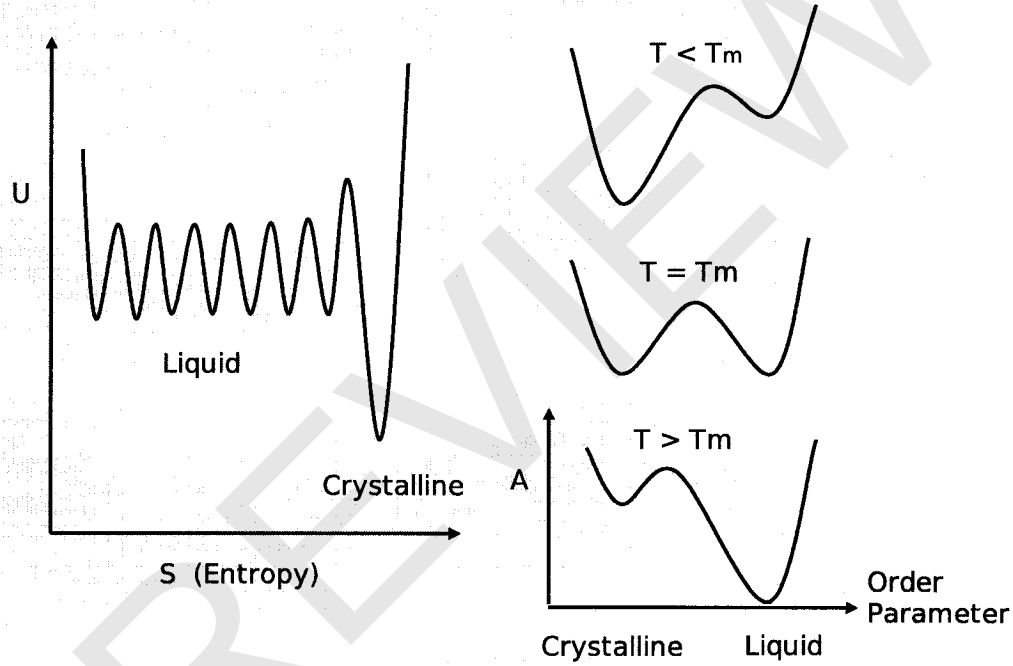

Figure 2.3: While one-dimensional potential energy surface $U(N, \cdots)$ (left) is independent of temperature, one-dimensional Helmholtz free energy surface $A(N, V, T)$ (right) depends on temperature. Here, we assume that the possible states existing on liquid phases and crystalline phase are at minima of free energy surface.

of the PES for liquid and crystalline phase. The local minima shown in Figure 2.3 are considered as the possible states Ω. Thus the relation of entropy S with the possible states Ω is deduced via Boltzmann's entropy formula, $S = k_B \ln \Omega$. S will be larger when the number of states that the system can visit is larger. In Figure 2.3, the entropy of liquid phase is larger than that of crystalline phase since the accessible states of liquid is much larger than those of crystalline.

The relation of potential energy $U(N, \mathbf{r}_1, \cdot, \mathbf{r}_N)$ with Helmholtz free energy $A(N, V, T)$ via the partition function $Q(N, V, T)$ of the system is given by

$$A(N, V, T) = -k_B T \ln Q(N, V, T), \tag{2.2}$$

$$Q(N, V, T) = \int_V \frac{1}{\Lambda^{3N} N!} \exp(-\beta U(N, \mathbf{r}_1, \cdots, \mathbf{r}_N)) d\mathbf{r}_1 \cdots d\mathbf{r}_N, \tag{2.3}$$

where $\Lambda = h/(2\pi m k_B T)^{1/2}$, m is the mass of a particle, and $\beta = 1/k_B T$. The thermal stability of the system is correlated with the thermal equilibrium since the system will go into the condition of thermal equilibrium spontaneously. At constant T and V, we note from thermodynamics not only that the condition for thermal equilibrium is that A is its minimum value, but also that $\Delta A \leq 0$ for a spontaneous process. In Figure 2.3, the Helmholtz free energy surface is plotted as a function of a order parameter. We can see two possible stable phases[†] : one is the crystalline phase and the other is the liquid phase since both phases satisfy the condition of thermal equilibrium with $\Delta A \leq 0$. The relative stability, however, depends on temperature. For example, under supercooled temperature $T < T_m$, the supercooled liquid is *metastable* with respect to the crystalline, so it is driven to crystallize if crystallization barrier is accessible kinetically. At melting temperature $T = T_m$, the Helmholtz free energy of crystalline phase A_{cry}

[†]Other possible stable phases of amorphous, glass, low-density liquid, and high-density liquid are ignored in order to make it simplify

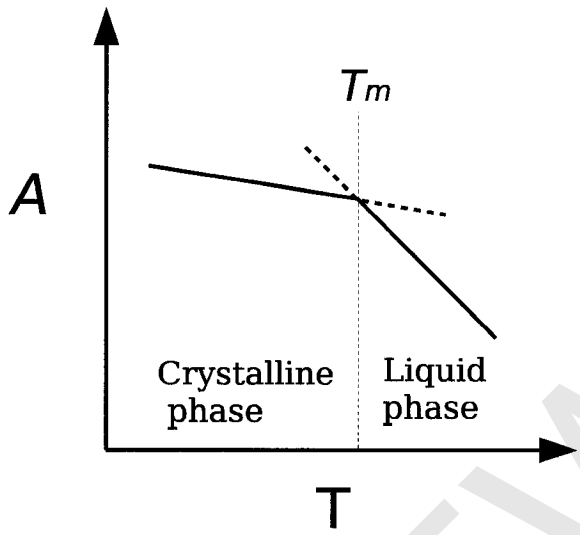

Figure 2.4: The schematic temperature dependence of Helmholtz free energy $A(N, V, T)$ of the crystalline and liquid phases. The phase with the lowest Helmholtz free energy (dash lines) at a specified temperature is the most stable one at that temperature. At transition temperature, here melting temperature, two phases have the same Helmholtz free energies.

becomes the same as that of liquid phase A_{liq}. Once the temperature is higher than melting temperature ($T > T_m$), the liquid phase becomes stable and the crystalline phase *metastable* due to $A_{liq} < A_{cry}$. Figure 2.4 shows the schematic temperature dependence of the Helmholtz free energy of crystalline and liquid phases at thermal equilibrium. We note that the slope of each phase is correlated with its entropy via $(\partial A/\partial T)_{N,V} = -S$, that is, the entropy of liquid phase is larger than that of crystalline (which is showed as the broadness in one-dimensional PES in Figure 2.3).

2.5 Transition State and Free Energy Surface

The kinetic rates for phase transitions, protein folding/unfolding, and chemical reactions are the experimental observable values. These kinetic rates k are connected to the free

energy barriers ΔF^* via

$$k(T) \propto \exp(\frac{-\Delta F^*(T)}{k_B T}). \tag{2.4}$$

Thus, the calculation of free energy barrier between two states is one of the most attractive techniques.[21,22] There are several available schemes for the calculation of the free energy barrier via computational approach. Two general classes are available in the computational approach. One is called molecular dynamics methods. The classical molecular dynamics method generates the trajectory via integrating Newton's equations of motions. The other classes is called the Monte Carlo method. The convenient Monte Carlo method generates the trajectory that samples configurations in accordance with the canonical Boltzmann distribution. These methods provide meaningful statistics for the thermal stable states. For sampling configuration of the transition states, the statistic errors becomes no more negligible. To obtain significant statistics for rare events without wasting time, histogram methods combined with non-Boltzmann (or biased) sampling have been employed. This technique, sometimes called *umbrella sampling*, can be profitably used to calculate the free energy change as a function of one or more order parameters. For a single order parameter Φ, the free energy change is defined via

$$\Delta F = F(\Phi) - F(\Phi_0) = -k_B T \ln \frac{P(\Phi)}{P(\Phi_0)}, \tag{2.5}$$

where $P(\Phi)$ is the probability distribution as a function of Φ and Φ_0 is the order parameter at the reference state (or thermal stable state). The derivation and application of Eq. (2.5) will be discussed in Chapters 6 and 7.

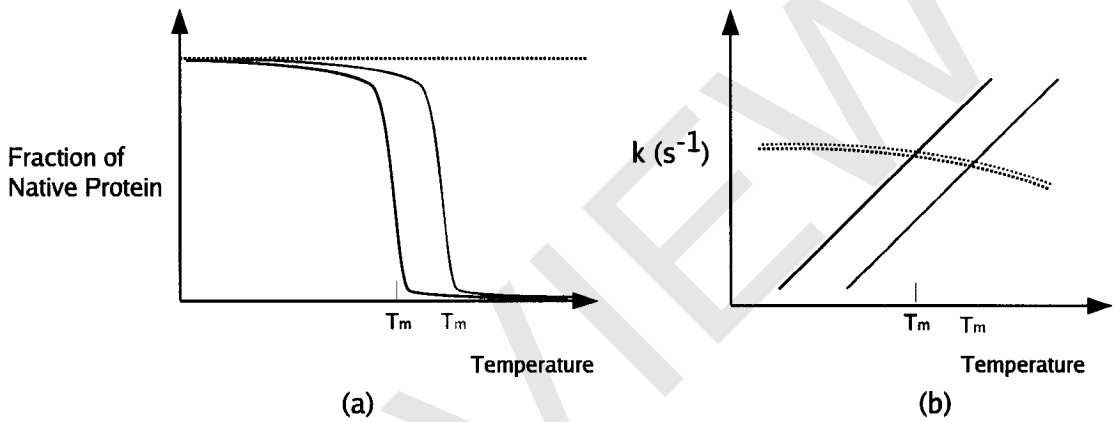

Figure 2.5: The schematic plots for experimental information of temperature dependence of protein folding/unfolding reactions. (a) Equilibrium unfolding transition of mesophilic (dark line) and thermophilic (red line) proteins as a function of temperature. Thermophilic protein has higher melting temperature T_m than mesophilic protein. The unfolding reactions of both proteins have some resemblance to the macroscopic first-order transition. (b) Folding (k_f, dashed line) and unfolding (k_u, continuous line) rate constants are extracted from a temperature-jump experiments via relaxation rate constant $k_{obs} = k_f + k_u$. Unfolding kinetics of the thermophilic protein (red line) becomes significantly slower than that of the mesophilic (dark line), while folding kinetics of both proteins are very close.

2.6 Thermal Stability of Proteins

In this Section, we will discuss the thermal stability of soluble proteins. One of the key characterizations of thermal stabilities of soluble proteins is the protein melting temperature T_m as shown in Figure 2.5. For example, thermophilic proteins have higher T_m than mesophilic proteins.[‡] Figure 2.5a show the experimental data from Circular Dichroism spectra for protein thermal stabilities as a function of temperature. The temperature dependence of the protein thermal stability resembles the macroscopic first-order transition. Figure 2.5b plots the folding/unfolding rate constants from temperature-jump experiments.[23] Folding rate constants k_f are fast with less dependence of temperature, while unfolding rate constants k_u show the huge dependence of temperature. In comparison with the mesophilic protein, the thermophilic protein exhibits that unfolding rate constants are shifted significantly toward slower, while folding rate constants are very close. Two experimental informations indicate that the thermal stability of the soluble protein has the highly relation with the unfolding reaction.

Since the secondary and tertiary structures of folded proteins are stabilized via hydrogen bonding interactions, the importance of hydrogen bonding has been studied. The classical view point is that the net effect produced by hydrogen bonds approximates zero, on consideration that molecular groups involved in the folded state can establish the same kind of interaction with water molecules in the unfolded state. It has been suggested earlier that the primary driving force might be a hydrophobicity.[24–34] Due to ordering of water around hydrophobic molecular groups, folding processes occur in a direction of increase of water entropy by collapsing of hydrophobic molecular groups into the core of protein. The long-range hydrophobic attraction could not be explained sim-

[‡]Mesophilic proteins are proteins existing in mesophilic microorganisms (optimal living temperature is about 25 degree) and thermophilic proteins from thermophilic microorganism living at very high temperature more than 50 degree)

ply by water entropy. The drying behavior of liquid water on the hydrophobic surface induced nanoscopic or microscopic bubbles between the hydrophobic surfaces which result in the long-range hydrophobic attraction.[35–40] But the contribution of both hydrophobicity and hydrogen-bonding interaction to the first-order transition behavior is still in infancy and many limitations and obstacles remain. Chapter 7 will discuss the relation of hydrophobicity with the first-order transition behavior for understanding of the thermal stability of proteins.

Chapter 3

Global geometry optimization of silicon clusters described by three empirical potentials

3.1 Introduction

The global minima of silicon clusters have also been studied on the basis of the Stillinger-Weber (SW) and Gong empirical potentials.[41–45] The SW potential was developed to reproduce a variety of bulk solid and liquid properties of silicon,[44] and thus the global minima of SW clusters are not expected to be the same as the realistic global minima, especially for small-sized silicon clusters. Recently, Mousseau and coworkers suggested a slightly modified Stillinger-Weber (MSW) potential to simulate properties of amorphous silicon.[46] Again, the global minima of MSW are not expected to be the same as those based on ab initio calculations. Gong also proposed a modified SW potential (hereafter called the Gong potential) in order to capture certain structural features of small-sized silicon clusters based on ab initio calculations.[45] Thus, it will be interesting

to examine how well the Gong potential can describe the mid-sized clusters.

The simulated annealing method has been tried to the SW potential for $3 \leq n \leq 17$.[41] It is possible that the simulated annealing method allows the geometries trapped in the metastable configurations in case that temperature becomes too low to pass over the vibrational barriers between local minima. The performing quenching methods in parallel with the molecular dynamics simulation of SW potential revealed that the simulated annealing method failed to find global minima of Si_6, Si_{11}, and Si_{13}.[42] Using the SW and Gong potentials, Iwamatsu has calculated the global minima for $3 \leq n \leq 15$ by using the genetic algorithm (GA) global optimization method.[13] He found that the global-minimum structures of the SW potential are identical to those obtained via the MDQ method.[42]

Here, we apply the basin-hopping technique[20] (see Section 2.3.2) to locate the global minima of silicon clusters for $3 \leq n \leq 30$. First, we will employ both the SW and Gong potentials to compare the calculated global minima with those (for $n \leq 15$) based on the GA method and those (for $n \leq 20$) based on the simulated annealing technique. We will then use the MSW potential to examine the effects of changing three-body part of potential function on the global-minimum structures. Next, for $n > 20$, the global minima obtained via the basin-hopping method will be compared with the available ab initio or TB calculations, as well as with the mobility experiments. We will monitor the first appearance of the endohedral atom in the cluster and discuss the possible prolate-to-spherical-like structural transition.