BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

MOLECULAR DYNAMICS STUDIES OF PROTEIN STABILITY AND FOLDING KINETICS

by

Jose M. Borreguero

B.S., Basque Country University, 1997
M.A., Boston University, 2001

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
2004

Approved by

First Reader

H. Eugene Stanley, Ph.D.
University Professor,
Professor of Physics

Second Reader

Shyamsunder Erramilli, Ph.D.
Professor of Physics

MOLECULAR DYNAMICS STUDIES OF PROTEIN STABILITY AND FOLDING KINETICS

(Order No.)

Jose M. Borreguero

Boston University Graduate School of Arts and Sciences, 2004

Major Professor: H. Eugene Stanley, Professor of Physics

ABSTRACT

To be written at the latest stages of the draft.

Contents

List of Tables

List of Figures

List of Abbreviations

DMD	discrete molecular dynamics
TSE	transition state ensemble
MD	molecular dynamics
MC	Monte Carlo
NMR	nuclear magnetic resonance
p_{FOLD}	probability to fold
PDB	protein databank
R_G	radius of gyration
RMSD	root-mean-square deviation

Chapter 1

Protein World

1.1 The Biology of Proteins

Proteins are the nanomachines of living organisms, carrying out an overwhelming number of menial tasks which are the elementary steps of the complex cellular processes. These tasks include such different categories as: *(i)* transport and storage of chemicals; *(ii)* becoming building material for muscles and cell membranes; *(iii)* defend the organism agains external agents; and *(iv)* catalize the vast array of chemical reactions that regulate the life of the cell.

Despite their ubiquitous nature, proteins are cunning in their simplicity at the atomic level. All proteins are built with only five different types of atoms (N,O,H,C,S)! It is the different arrangements of these atoms which make the proteins so versatile. Through some unresolved evolution process, we find ourselves at present that these atoms organize in only twenty stable chemical compounds, termed *amino acids* (see Fig. 1.1). All amino acids can be divided into two parts: *(i)* the *backbone*, which is identical for all the twenty types; and *(i)* the *residue*, which is different and identifies each type.

The amino acids form and effective alphabet, and their different combinations in a linked chain of amino acids, one after the other, produce the different proteins. However, just in the same way as a random combination of letters does not produce a word, a random combination of amino acids does not produce a protein. A word is defined by its meaning,

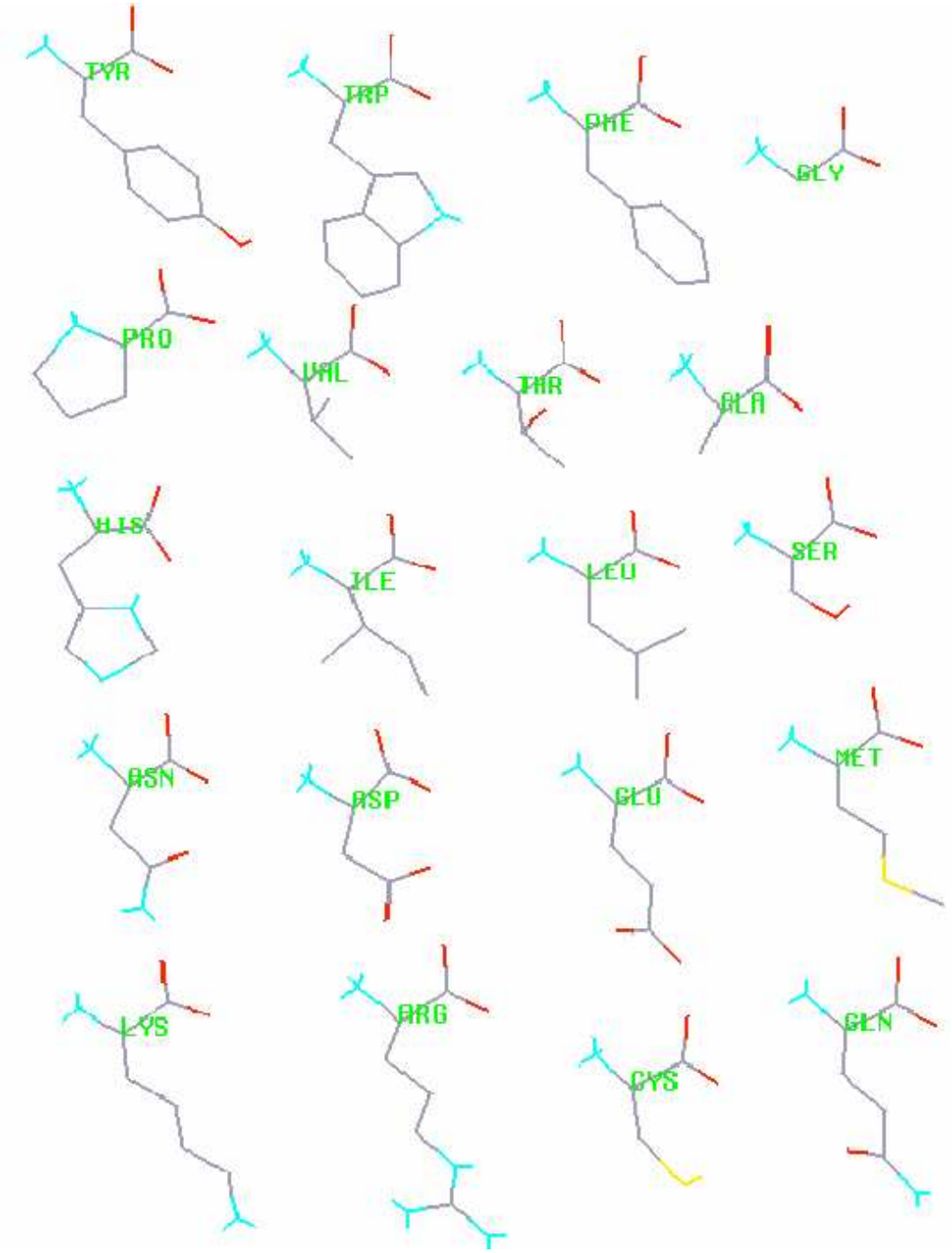

Figure 1.1: Wireframe representation of the twenty natural amino acids. Atoms are colored as follows: Nitrogen–Cyan, Oxygen–Red, Carbon–Grey, Sulfur–Yellow. Hidrogens not shown for simplicity.

and a protein is defined by its *function*. The genes, embedded in the DNA sequence, store the precise order in which the amino acids must be linked to produce the desired protein. This particular order is called the *primary structure* of the protein. The cell machinery, i.e the proteins, are in charge of producing other proteins from the DNA sequence starting point. The complexity of primary structure production makes the artificial assembly of proteins unfeasible with current technologies. Standar methods[1] resort to "slave cells" which receive an input gene and then produce the corresponding protein.

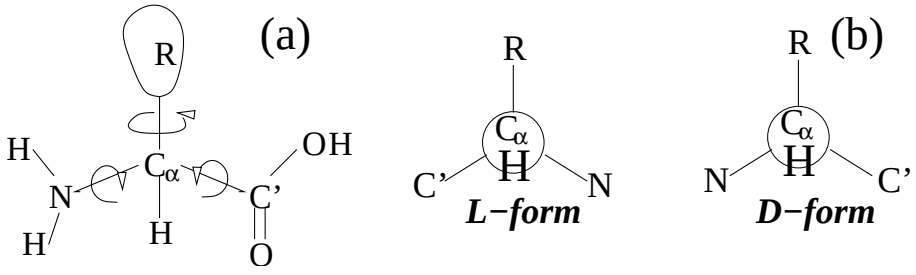

Figure 1.2: (a)Atomic arrangement of the amino acid backbone. Key atoms are the central C_α and the covalently linked N and C'. We represent all atoms making up the residue part by R. The curved arrows indicate axis of free rotation. (b) The two chiralities of amino acids, looking down the H-C_α bond from the hydrogen atom.

1.2 The Structure of Proteins

It is critical to understand the atomic structure of proteins, prior to any attempt to produce any type of protein model. We have discussed already that any amino acid can be divided into the backbone and the residue parts. When the amino acid is not linked to other amino acids, its backbone is composed of a central carbon atom (C_α) covalently linked to a carboxilic group ($COOH^-$), an amino group (NH_2) and one hydrogen (H). In addition, C_α is covalently linked to the beginning of the residue part (see Fig. 1.2a). The four groups attached to the C_α atom are chemically different for all amino acids types except for Glycine, which has only one hydrogen atom as the residue. Thus, amino acids are chiral molecules that can exist with two different "hands", L- or D-form (see Fig. 1.2b).

However, all natural amino acids produced in the cell have the L-form. Currently there is no satisfactory explanation on the origin of this symmetry breaking.

Amino acids link together to form the protein chain via the so called "peptide" bond. After the linking process, six atoms (C_α-C'-O-N-H-C_α) form a rigid plane termed *peptide-plane* (see Fig. 1.3a). The peptide-plane is connected to the rest of the chain via its two C_α atoms. We can view the protein chain as a connection of rigid peptide-planes. Peptide-planes can rotate around themselves through two axis of rotation with nexus at the C_α atoms (see Fig. 1.3b). The angle of rotation around the N–C_α axis is denoted by the "Φ" angle, and the angle of rotation around the C_α–C' axis is denoted by the "Ψ" angle. Thus, we can fully specify the conformation of the backbone to the atomic detail if we provide the series of Φ and Ψ angles $\{(\Phi_1,\Psi_1),...,(\Phi_N,\Psi_N)\}$.

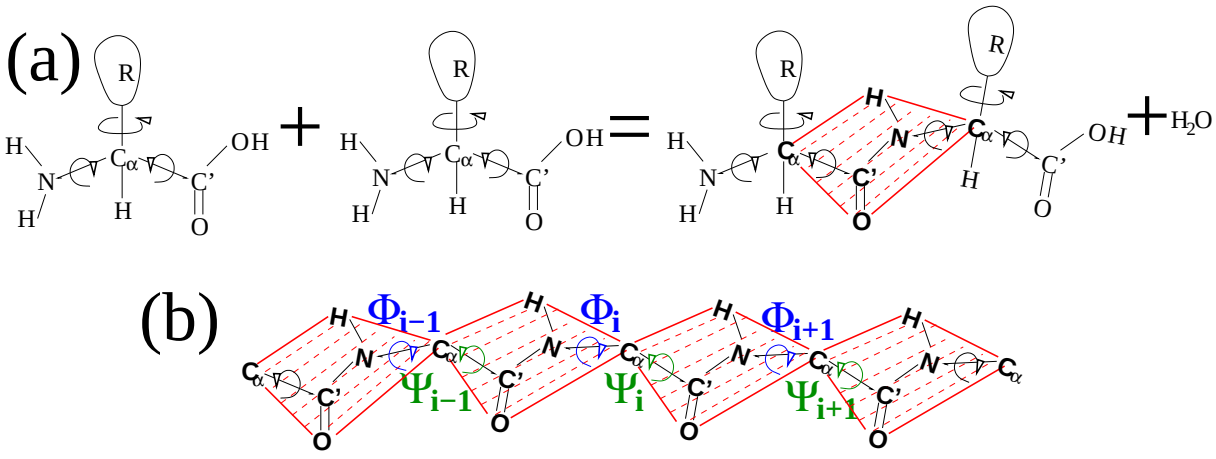

Figure 1.3: (a) Formation of a protein chain with a peptide bond. The six atoms in bold face form a rigid peptide plane, shown in red. (b) Backbone chain as a linkage of peptide-planes with specific Φ (blue) and Ψ (green) angles.

Not all Φ and Ψ values are allowed because of the excluded volume of neighboring atoms. The plot of all allowed Φ and Ψ values is usually termed the *Rachandran plot* (see Fig. 1.4a). There is a subset of Φ and Ψ values within the Rachandran plot which correspond to the values found with high probability in proteins. Important values are ($\Phi \simeq -139$, $\Psi \simeq 135$) corresponding to a *β-sheet* conformation (depicted by arrows), and ($\Phi \simeq -57$, $\Psi \simeq -47$)

corresponding to the α-*helix* conformation (see Fig. 1.4b). These conformations are referred to as the *secondary structure* of proteins.

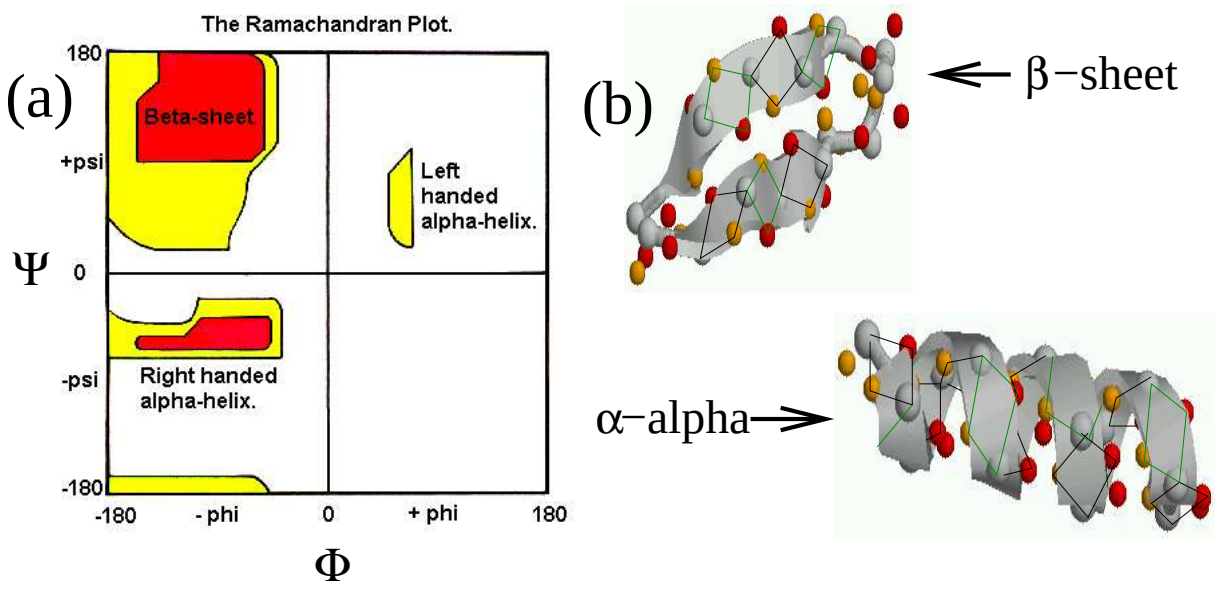

Figure 1.4: (a) Allowed Ψ and Φ within the yellow regions. Red regions characterize the values tipically found in proteins. (b) A tipical β-sheet and α-helix, with peptide-planes shown in black and green.

Small chains of amino acids, of about 20 amino acids or less, adopt either a random-coil conformation or an α-helix or a β-sheet. However, proteins are longer chains, with more than a hundred amino acids. At phisiological conditions, long helixes and sheets are not stable. Thus, proteins must arrange their amino acids in alternative ways. It has been observed that the protein chain arranges into a series of short and medium helixes and sheets, which then accomodate themselves to form a globular structure, called the *tertiary structure* or *fold* of the protein (see Fig. 1.5). The tertiary structure is stable at phisiological conditions and defines the funtion of the protein.

1.3 The Scientific Question

The advent of the genome projects [2], aiming to identify every single gene in the human DNA as well as in the DNA of some other organisms, has brought a wealth of protein

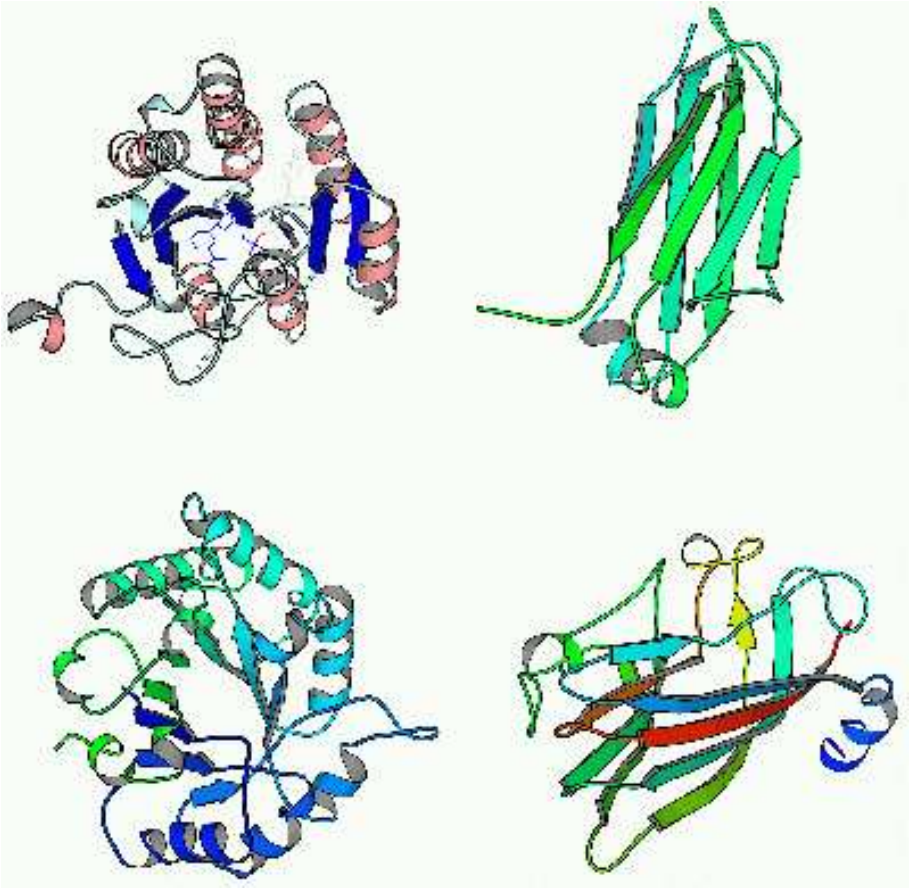

Figure 1.5: Schematic drawing of a number of common protein folds. Top left: a/b domain with Rossmann fold (3a, 20b-hydroxysteroid dehydrogenase), top right: immunoglobulin constant domain, bottom left: TIM barrel (triose phosphate dehydrogenase), bottom right: jelly roll (satellite tobacco necrosis virus coat protein)

primary structures. Current gene databases store more than ten billion entries, and the expected publication of new sequenced genome will sustain the growth rate in the number of entries over the next years. This means that for many proteins, we know the particular order in which amino acids are linked together. However useful, this information is not enough to predict *with our current knowledge* the particular function that a protein will perform in the organism. Rather, we can today univocally identify the function if and only if we are given the spatial arrangement of the atoms that make up the protein [3].

At physiological conditions[1], a protein has a preferred, stable atomic arrangement termed *tertiary structure*, *native state* or *folded state*. Tertiary structure determines function and *vice-versa*.

Ample experimental evidence[4, 5] supports the central hypothesis that for relatively small proteins (of the order of 100 amino acids) the primary structure encodes all the necessary information to predict the tertiary structure, and consequently, the protein function. The *protein folding question* is formulated in the following terms: given the primary structure, what is the tertiary structure?

1.4 A Physicist Viewpoint

From the physicist's point of view, the protein folding problem translates into understanding the basic principles by which amino acids arrange themselves via their physical interactions into the tertiary structure, or folded conformation. However, one begins to understand the difficulty of the problem when confronted with the multidimensionality of the protein phase-space. For instance, assuming that each amino acid has only two conformations, folded and random, and that the protein has 100 amino acids, the number of total conformations is 2^{100}. The sortest time for an amino acid to make a conformational change is of the order of the pico-second, thus taking a total of $2^{100} \cdot 10^{-12}\text{s} \simeq 40$ billion years to complety explore the phase-space of this over-simplified system and find the folded conformation. On the other hand, proteins need to fold in the order of the millisecond in order to avoid aggregation in the cell. This problem has been known as the Levinthal paradox, after Levinthal formulated it for the first time in the seventies [6]. Thus, it is critical to develop models as simple as possible, yet able to reproduce the main characteristics of protein folding. In addition, we need to develop computational tools that are efficient and capable of achieving the time scales neccessary to study the folding of the protein.

[1] $T = 298^{\circ}\text{K}$; $P = 1\text{at}$; $pH = 7$

Chapter 2

Computer Simulations

2.1 Introduction

There is a current enormous attempt in the experimental arena to explore the role of each amino acid in the folding process and final stabilization of the protein in its native conformation. Experiments based on site-directed mutagenesis [7–11] identify the free energy that individual amino acids contribute to the energy gap between folded and misfolded (or unfolded) states. However, the complexity and sophistication of such experiments make it merely impossible to massively use them to study the energetics of individual amino acids. There are several complications with the experiments at various stages due to the individuality of proteins — e. g. *purification* of the protein from the bacterial cell where the mutated protein is produced and *crystallization* of the purified proteins for the crystallographic studies, to mention a few. Such complications make it, in some cases, very difficult to study the proteins in vitro. In addition, the environment of the experiments usually drastically differs from the natural cell environment — the concentration of proteins in vitro is much lower than that in cells, raising a question of the experiments' reliability.

Computer simulations are an attractive alternative to these complex experiments. There have been many theoretically inspired computer experiments [12–16] attempting to predict the thermodynamics and kinetic role of amino acids in proteins. On the other hand, the complexity and vast dimensionality of the protein conformational space [6] makes the folding

time too long to be reachable by computational studies at the atomistic detail [17–19]. The shortest vibrational mode, corresponding to a perturbation of the atomic covalent bonds, has a femptosecond period while the fastest proteins are known to fold in the milisecond time scale. Current computing power allows one to approach the $10^{-7}s$ time scale with full atomistic detail.

Simplified models [12, 17, 20–27] became popular due to their ability to reach reasonable time scales and to reproduce the basic thermodynamic and kinetic properties of proteins [5, 18, 28], such as: *(i)* unique native state (NS), i. e. there should exist a single conformation with the lowest potential energy; *(ii)* cooperative folding transition (resembling first order transition); *(iii)* thermodynamical stability of the NS; *(iv)* kinetic accessibility, i.e. the NS should be reachable in a biologically reasonable time [25, 29].

2.2 Monte Carlo on-lattice

The simplicity of the Monte Carlo algorithm, and a significantly smaller conformational space of the protein models due to the lattice constraints, make Monte Carlo on-lattice simulations a powerful tool to study the thermodynamic properties of proteins [19, 20, 30, 31].

The Monte Carlo algorithm is based on a set of rules for the transition from one conformation C to another C'. These transitions are weighted by some transition matrix $W_{C\to C'}$ which reflects the phenomena under study. In the classical Monte Carlo scheme we have

$$W_{C\to C'} = min[\frac{e^{-\beta E_{C'}}}{e^{-\beta E_C}}, 1], \tag{2.1}$$

where E_C is the energy of the system when in conformation C, and $\beta = 1/k_BT$ is the inverse temperature (when the Boltzmann constant is taken as unity). A random walk through energy space will generate the canonical distribution of energies $H(E) = g(E)e^{-\beta E}$, where $g(E)$ being the density of states (DOS).

We allow the following configurational changes in the protein chain:

- translation of the whole protein-chain as a rigid body.

- pivoting around a randomly selected amino acid
- reflexion around a randomly selected amino acid and perpendicular to a selected axis (either the X-axis or Y-axis).
- kinkjump around a randomly selected amino acid.

All these changes must preserve protein-chain continuity. In addition, we avoid the overlapping of two particles on the same lattice site by interchanging their positions. We allow the following configuration changes in the aqueous solvent:

- select a water particle at random and change the value of its internal variable.
- select two neighboring water particles and interchange their positions.

2.2.1 Multicanonical Monte Carlo

Despite we find on-lattice models a big step down the degree of refinement of atomistic models, the resulting energy landscape of a protein plus solvent is characterized by a multitude of local minima. At low temperatures and with the classical Monte Carlo scheme, the protein spends all the simulation time in one these local minima after it is found. Thus, the classical Monte Carlo scheme makes unfeasible an efficient sampling of the protein conformational space.

Fortunately, the Multicanonical Monte Carlo (MMC) technique [32–34] is specially suited to avoid transient trapping in local energy minima at low temperatures. Specifically, we use the multiple-range random walk algorithm [35] to calculate the DOS. Unlike conventional Monte Carlo which generates a canonical destribution $H(E) = g(E)e^{-\beta E}$, the MMC rules for acceptance of a new system conformation generate a flat distribution $H(E) = constant$.

We can generate a constant $H(E)$, beginning with $H(E) \equiv 0$, by performing a random walk in energy space with transition matrix

$$W_{C \to C'} = min[\frac{1/g(E')}{1/g(E)}, 1], \tag{2.2}$$

the ultimate goal of MMC being the computation of the DOS. Since we do not know *a priori* the DOS, we initialize $g(E) = 1$ and perform the following algorithm:

1. at step i, generate new conformation C_i' from old conformation C_i.
2. if the fraction $F \equiv \frac{1/g(E')}{1/g(E)} \geq 1$, accept the move and $C_{i+1} \equiv C_i'$. Otherwise, compare the fraction F to a random number r. If $F \geq r$, accept the move and $C_{i+1} \equiv C_i'$, otherwise reject the move and $C_{i+1} \equiv C_i$.
3. Multiply the DOS by a constant factor $f > 1$: $g(E_{C_i+1}) \rightarrow f \cdot g(E_{C_i+1})$.
4. Increase the energy distribution: $H(E_{C_i+1}) \rightarrow H(E_{C_i+1}) + 1$
5. If the ratio $\frac{\langle H^2(E)\rangle - \langle H(E)\rangle^2}{\langle H(E)\rangle^2} \simeq (\log f)^2$ stop the simulation, otherwise repeat first step.

Once the simulation is finished, we obtain an energy distribution $H(E)$ which is constant with accuracy proportional to $\log f$. If we desire higher accuracy, we perform a new simulation with a new factor f', $1 < f' < f$ and initialize the DOS with the output $g(E)$ from the old simulation.

We can calculate equilibrium averages, or higher moments, for quantities that are functions of the system-energy and at *any* temperature,

$$\langle A\rangle = \frac{1}{Z(\beta)} \int dE g(E) A(E) e^{-\beta E}, \tag{2.3}$$

$$Z(\beta) = \int dE g(E) e^{-\beta E}. \tag{2.4}$$

If we are interested in a quantity $A(x_1, .., x_N)$ that is are not a function of the system-energy, then we calculate the extended density of states (EDOS) $g(x_1, .., x_N, E)$. The algorithm to calculate the EDOS is the same as for the DOS.

Once we obtain a constant $H(x_1, .., x_N, E)$ up to accuracy $logf$, we calculate the thermodynamics averages,

$$\langle A\rangle = \frac{1}{Z(\beta)} \int dx_1 .. \int dx_N \int dE g(x_1, .., x_N, E) A(x_1, .., x_N) e^{-\beta E}. \tag{2.5}$$

2.3 Discrete Molecular Dynamics

Monte Carlo simulations may not be a reliable tool to study the protein kinetics. The time in Monte Carlo algorithms is estimated as the average number of moves. It was pointed out [36] that Monte Carlo simulations are equivalent to the solution of the master equation for the dynamics and, hence, there is a relation between physical time and computer time, which is counted as the number of Monte Carlo steps. However, a number of delicate issues, such as the dependence of the dynamics on Monte Carlo move set, remain outstanding and hence an independent test of the dynamics using the molecular dynamics (MD) approach is preferred.

To address the questions sensitive to protein kinetics, it would be useful to study MD models of proteins. Thus far, several MD simulations have been performed [37–39], which demonstrate the ability of the simplified models to study the thermodynamics and kinetics properties of proteins. More detailed models have proven to be useful only in the shortest time scales, well below the required time for the protein fold.

We propose to implement the discrete molecular dynamics (DMD) algorithm [40, 41] to study proteins. The earliest molecular dynamics simulations were performed with the discrete algorithm, before the advent of the nowadays conventional, or contiunous molecular dynamics (CMD). DMD has a higher speed performance than CMD, in exchange for a higher demand on computer memory. With current computer memory, a typical PC platform exceeds by far the DMD requirements, making DMD a choice tool to simulate the folding of proteins.

2.3.1 Temporal Evolution

The DMD central point is the use of discontinous potentials that model interactions between pair of particles. We depict two typical potentials in Fig. 2.1. The first potential correspond to a non-bonded interaction: the two particles may overcome a finite potential barrier located at the interaction range r^* and become free particles. The second potential correspond to a bonded pair: particles move freely only whithin a certain range of

distances, bonded by two infinite potential barriers. In addition, we can approximate the dynamics of DMD and CMD in the short time scale by including additional potential barriers (see Fig. 2.1) so that the discrete potential matches more closely the continous potential function. However, we loose all the speed advantages of DMD in doing so.

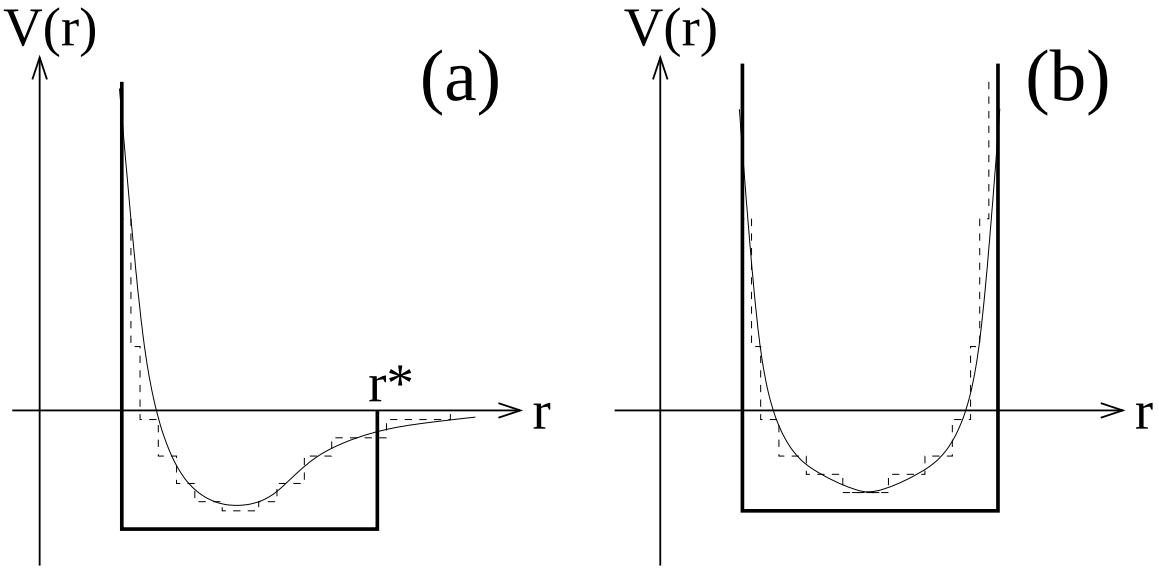

Figure 2.1:

Figure 2.2: (a) Typical non-bonded and (b) bonded discontinous potentials for two interacting particles (bold lines). We can approximate a discrete potential to a continuous potential function by including additional steps (dashed lines).

With discontinous potentials, a particle is not subject to any force unless its distance to other particle happens to be exactly one of the distances for which there exits a potential barrier. Otherwise, the particle moves freely and we can easily predict its position for future times. Because we can predict the future positions of all particles, we can also compute for any pair of particles the time when their mutual distance will correspond to one of those distances for which there exist a potential barrier. Thus, for each pair of particles we calculate a *collision time*, t_{coll}. Once we know all possible t_{coll} in the system, we sort the minimal one, t_{min}, and evolve the position of the particles from t_0 to $t_0 + t_{min}$. After the algorithm has performed this step, all particles have changed their positions and there will

be only one particular pair of particles (p_i, p_j) for which their mutual distance will exactly be one of the distances for which there exits a potential barrier. Thus, we have to solve collision equations for this particle-pair (p_i, p_j), which consist of conservation of energy, momemtum and angular momentum equations. The solution of these equations are new velocities for the two particles. After calculating the new velocities, we repeat the whole process by computing new t_{coll}'s.

This algorithm can be greatly improved, provided we notice three observations: *(i)* leaving aside the particular colliding-pair (p_i, p_j), all other particles have not yet collided and therefore are still moving freely. Thus, there is no necessity of updating their positions since we can still calculate their positions for future times; *(ii)* We only need to calculate new t_{coll}'s only for all particle-pairs for which either one of the two particles is p_i or p_j. This is so because p_i and p_j have changed their velocities as the result of their previous collision; *(iii)* We do not need to calculate t_{coll}'s for any two particles far away from each other, because it is very likely that before this collision happens, one of the two particles will collide with a closer particle. Thus, we divide the system volume into cells, and for any given particle we compute t_{coll}'s only with particles that are in the same cell and in neighboring cells.

2.3.2 Temperature Control

Our systems of interes, namely proteins and proteins in aqueous solution, undergo changes resembling a first order transition close to ambient temperature. Upon transition, the sytem either releases or absorbs large quantities of energy. It is therefore critical to implement an auxiliary methodology that controls the system temperature in a way that mimics natural temperature control.

Rare Gas Thermostat

The simplest temperature control to implement in the DMD algorithm consists of the inclusion of extra particles. These particles interact with the amino acids via ellastic collisions,

drawing or adding energy to the protein. Extra particles also undergo ellactic collisions among themselves, in order to maintain the Maxwellian velocity distribution through time. We define the protein temperature T_{prot} and the system temperature T with the kinetic energy:

$$\frac{3}{2}k_B T_{prot} = \frac{1}{N_{aa}} \sum_{i=1}^{N_{aa}} \frac{m v_i^2}{2} \tag{2.6}$$

$$\frac{3}{2}k_B T = \frac{3}{2}k_B T_{prot} + \frac{1}{N_{xp}} \sum_{i=1}^{N_{xp}} \frac{m v_i^2}{2}, \tag{2.7}$$

where N_{aa} is the number of amino acids and N_{xp} is the number of extra particles. At equilibrium conditions, $T_{prot} \equiv T$.

Upon a folding transition, the potential energy of the protein decreases abruptly and its kinetic energy increases in the same measure, thus increasing T_{prot}. We need $N_{xp} \gg N_{aa}$ so that the extra particles will absorb the excessive kinetic energy of the amino acids, thus decresaing T_{prot}, while at the same time the system temperature T will not change appreciably. In addition, N_{xp} must be big enough so that the rate at which the extra particles absorb the heat will be bigger than the typical folding transition rate.

While most easy to implement, the drawback of this temperature control method is the overhead of computational cost due to collisions between the extra particles. These collisions have no bearing in the properties of the protein. Thus, one desires to make N_{xp} as small as possible. After several try-outs, we find a value of $N_{xp} \simeq 40 N_{aa}$ to be a good compromise between computational overhead and temperature control effectiveness. These number of extra particles whithin the system volumes simulated correspond to densities typical of the rare gas.

The Andersen Thermostat

We can eliminate most of computational overhead incurred by the rare gas method by removing the collisions between the extra particles. To achieve this goal, we implement the Andersen themostat (AT) [42]. In AT, each amino acid is immersed in a constant temperature bath of imaginary (ghost) particles. Ellastic collisions between amino acids

and the particles of this imaginary bath maintain the temperature of the amino acids. Thus, we schedule one ghost collision for each amino acid at a future time $t_0 + \delta t$ extracted from a Poisson distribution of times

$$P(t) = \nu E^{-\nu \delta t}, \tag{2.8}$$

where ν is the ghost collision rate. When a ghost collision needs to be computed, we simply assign a new velocity for the amino acid. The new velocity is extracted from a Maxwellian distribution at system temperature T,

$$P(\vec{v}) = \frac{m}{(2\pi k_B T)^{3/2}} e^{-m\vec{v}^2/2k_B T}. \tag{2.9}$$

The ghost collision rate ν must be bigger than the folding transition rate in order to rapidly absorb the released energy upon folding. While in the folding transition, many amino acids that are separated along the protein chain become into contact in space. Thus, an acceptable value for ν is the typical long-range collision rate, i.e the typical rate at which two amino acids that are separated along the protein chain collide.

While this method resolves the computational overhead of the rare gas, it imposes somewhat unrealistic collisions for some amino acids. In particular, we should implement a smaller number of ghost collisions for amino acids in the core of the protein. In the rare gas method, these amino acids are shielded from the extra particles by the surrounding amino acids. In AT, amino acids are assigned ghost collisions with no regard to the local environment.

The Berendsend Thermostat

If computational speed is a critical requirement of the simulations, the Berendsend [43] thermostast allows one to remove all collisions, real or ghost type, between protein and surrounding solvent. This is done by coupling the system to a heat bath via a heating rate coefficient α. We re-scale the system temperature T at regular time intervals Δt,

$$\frac{T(t + \Delta t) - T(t)}{\Delta T} = -\alpha(T(t) - T_{HB}), \tag{2.10}$$

where T_{HB} is the heat bath temperature. Since we define the system temperature via equation 2.3.2, we have to rescale the particle speeds by factor $\sqrt{T(t+\Delta t)/T(t)}$.

For instance, Let our system be composed of one particle moving freely with velocity $\vec{v}_0$ at $t = 0$. The initial temperature T_0 is defined as $3/2k_BT_0 = 1/2m\vec{v}_0^2$. If we change this system temperature to T_{HB} by a series of infinitesimal steps, the temperature and velocity time-evolution will be

$$T(t) = T_{HB} + (T_0 - T_{HB})e^{-\alpha t} \tag{2.11}$$

$$\vec{v}(t) = \vec{v}_0 \cdot \Lambda(t) = \vec{v}_0\sqrt{\frac{3k_BT(t)}{mv_0^2}} \tag{2.12}$$

Unfortunately, if we rescale the particle velocities in a more complicated system we have to recompute all collision times between pair of particles, thus imposing a high computational cost. To avoid this cost, we take an alternative strategy. Returning to our previous system of one particle, let's impose a transformation in the time coordinate via

$$dz = \Lambda(t)dt. \tag{2.13}$$

After the trasformation, the new velocity is constant

$$\vec{v}(z) \equiv \frac{d\vec{r}}{dz} = \frac{d\vec{r}}{\Lambda(t)dt} = \frac{1}{\Lambda(t)}\vec{v}(t) = \vec{v}_0 \tag{2.14}$$

and the equations of motion are invariant in form, thus giving the same trajectory,

$$0 = m\frac{d\vec{r}}{dt^2} \rightarrow 0 = m\frac{d\vec{r}}{dz^2}. \tag{2.15}$$

The equations of motion are not invariant in form for more complicated systems whith interactions,

$$\nabla U = m\frac{d\vec{r}}{dt^2} \rightarrow \nabla U = \frac{1}{\Lambda(t)^2}m\frac{d\vec{r}}{dz^2}. \tag{2.16}$$

Thus, we gerenarilize the transformation by changing the potential via

$$U(z) = \frac{1}{\Lambda(t(z))^2} \cdot U \rightarrow \nabla U(z) = m\frac{d\vec{r}}{dz^2}. \tag{2.17}$$

Our strategy consists in simulating the system in the z-time coordinate for a series of *small* time steps Δz_i. After every time step, we calculate quantitites pertaining to the Berendsend thermostat

$$\Lambda(t+\Delta t_i) = \Lambda(t)\sqrt{\frac{T(t+\Delta t_i)}{T(t)}} \simeq \Lambda(t)\sqrt{1-\alpha\frac{\Delta z_i}{\Lambda(t)}\frac{T(t)-T_{HB}}{T(t)}} \tag{2.18}$$

$$\Delta t_i = \frac{\Delta z_i}{\Lambda(t+\Delta t_i)} \tag{2.19}$$

$$T(t+\Delta t_i) = \Lambda(t+\Delta t_i)^2 T(t). \tag{2.20}$$

In addition, we rescale the potential energy of the system so that we can simulate the next Δz-time step,

$$U(z+\Delta z_i) = \frac{U(z)}{\Lambda(t+\Delta t_i))^2}. \tag{2.21}$$

The Berendsend thermostat gains computational speed but imposes unrealistic updates in the velocities of the particles, because all particles have their velocities updates at the same time. This method underestimates the role of the big fluctuations that arise due to sudden transfer of energy to a particular particle.

Chapter 3

Protein Modelling

3.1 Introduction

3.2 On-Lattice Model

We implement a protein plus aqueous solvent system on the square lattice with periodic boundary conditions. Every amino acid occuppies only one lattice-site, and every water molecule occuppies one lattice-site. No two water molecules can occuppy the same lattice-site. The same rule applies for two amino acids, and for one amino acid and one water molecule. We model the hydrogen-bond interactions between water molecules only to the next-neighbor range. A water molecule can have up to four next-neighbors if no amino acids occupy a neighboring lattice-site (Fig. 3.2). We model the possible orientations of a water molecule by the allowed values of a q-state Potts variable σ_i. Only when two neighbor molecules $\langle i, j\rangle$ are in the correct orientation ($\sigma_i = \sigma_j$) does a hydrogen bond (HB) that increases the volume of the system by ΔV form. This interaction mimics the increment on volume due to the incipient formation of a tetrahedral structure in the real hydrogen bond network. If the two neighbor molecules $\langle i, j\rangle$ are not in the correct orientation, the interaction of the particles does not imply any increment in volume. The Hamiltonian for our model of water-water interaction may be written as [44]

$$\mathcal{H}_{HB} = -J \sum_{\langle i,j\rangle} \delta_{\sigma_i,\sigma_j} \ , \tag{3.1}$$

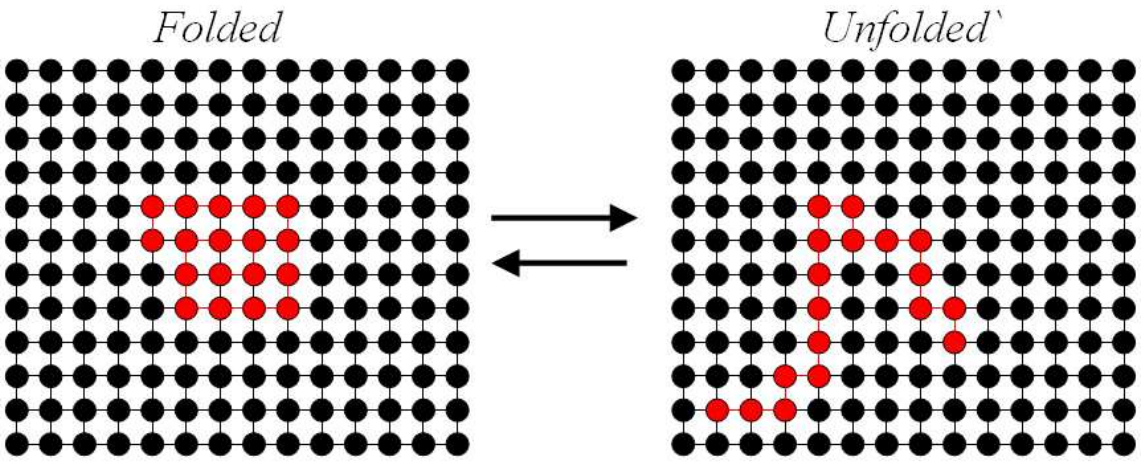

Figure 3.1: Bidimensional water-protein model. The protein amino acids (red) occupy contigous lattice sites and are surrounded by water molecules (black).

were $J > 0$ is the interaction strenght between water molecules upon tetrahedral network formation. The total volume of the system is given by $V = V_0 + N_{HB}\Delta V$, where $N_{HB} = \sum_{\langle i,j \rangle} \delta_{\sigma_i,\sigma_j}$ is the total number of hydrogen bonds with $\Delta V > 0$ in the system. The sum $\sum_{\langle i,j \rangle}$ extends only to nearest neighbors. The enthalpy of the system ($\mathcal{H}_{HB} + PV$) is given by

$$\mathcal{H}_{HB} + PV = -(J - P\Delta V) \sum_{\langle i,j \rangle} \delta_{\sigma_i,\sigma_j} \ , \tag{3.2}$$

where P is the pressure applied to the system. Our model solvent features a pressure $P_c = J/\Delta V$, above which formation of a hydrogen bond with $\Delta V > 0$ is not favorable from the energetic point of view.

We model the protein as a self-avoiding random walk embedded in the water bath. For simplicity, we consider a non-polar homopolymer that interacts with water via the partial ordering of water molecules forming hydrogen bonded structures around the homopolymer. We mimic the interaction using the Hamiltonian,

$$\mathcal{H}_p = J_r n_{HB}(n_{max} - \sum_{\langle i,j \rangle} n_i n_j) \ , \tag{3.3}$$

where the parameter $J_r > 0$ is the strength of the repulsive interaction and $n_{HB} = N_{HB}/N_{water}$ is the number density of hydrogen bonds with $\Delta V > 0$ (N_{water} is the number

of water molecules). The water–polymer repulsion increases as the water molecules tend to form the tetrahedral, low–density, hydrogen bonded network, where an unfolded apolar macromolecule is unlikely to be embedded. n_{max} is the maximum number of monomer–monomer contacts and $\sum_{\langle i,j\rangle} n_i n_j$ is the number of monomer–monomer contacts, where $n_i = 1$ if the lattice position i is occupied with a monomer, and zero otherwise. Therefore, $n_{max} - \sum_{\langle i,j\rangle} n_i n_j$ is a measure of polymer compactness, and equals zero when the polymer is maximally compact. Thus, Eq. (3.3) states that the hydrophobic repulsion driving the polymer to a compact state is equal to zero when the polymer is maximally compact ($\sum_{\langle i,j\rangle} n_i n_j \approx n_{max}$) or when the water forms the high–density bond network ($n_{HB} \approx 0$).

We hypothesize that the inability of water molecules to arrange in the low density ice–like structures is the principal mechanism responsible for the experimentally observed protein unfolding at very low temperatures [45–47]. At low pressures ($P < P_c$) and low temperatures, water molecules form a low density hydrogen bonded network, so the polymer is forced to adopt a compact state. At high pressures ($P > P_c$), the water is not able to form the low density network and arranges in a more dense state. In this case the effective repulsion between the apolar monomers and the solvent decreases, and water molecules penetrate into the homopolymer core, unfolding the compact state. Our hypothesis is supported by the experimental observation that unfolding at low temperatures exists mainly at high pressures (of the order of kbar), where water only freezes in the dense ice II phase.

3.3 Off-Lattice Models

Lattice models can not assess the role of each particular amino acid in the thermodynamics and kinetic properties of the protein because of the unrealistic constraints that the lattice model imposes on the real conformational degrees of freedom. These constraints translate into a poor capability of the lattice models to discern the topological properties of a protein, which are central for determining its function in the organism.

In addition, it is crucial to model correctly the protein-backbone flexibility. The reason is that the folding transition of some proteins can be represented as a two-stage process:

first, the unfolded protein approaches its transition state, where a few, well-defined amino acids form contacts [19] and drastically reduce the entropy. Second, the protein goes to the folded state, where energy is reduced drastically with less change in entropy. Flexibility is the main factor in calculation of the entropy of the model proteins, and thus it is crucially important to find an appropriate set of constraints among amino acids to mimic the real flexibility of the protein. A simple "beads-on-a-string" model of a protein will not reproduce the experimentally observed folding transition. There are many constraints in real proteins, such as the excluded volume of side-chains and the rigidity and orientation of covalent bonds. All these factors strongly reduce the conformational space of real proteins and significantly separate the "beads-on-a-string" protein model from real proteins.

We propose to tackle the problem of protein modelling through a series of increasingly complex models, each one suited for a different question. We begin with the one-bead model where each amino acid is modeled by a single bead. This model is useful to study the general properties of the folding transition kinetics. The next step differenciate between the backbone and side-chain components of each amino acid. Thus, we assign two beads per amino acid in the two-bead model. We study the details of the transition state ensemble with this model. Finally, we increase the complexity of the description of the protein backbone by modelling three-beads per amino acid backbone. We study the conformational changes of the protein backbone with this model.

3.3.1 One-bead Model

A simple beads-on-a-string model cannot reproduce the experimentally observed folding process, whereby the protein undergoes a folding transition without accumulation of intermediate conformations. The beads-on-a-string model features an excess of flexibility when compared to the flexibility of real proteins. Thus, we introduce a set of additional constraints among amino acids in order to mimic the flexibility of proteins. We obtain the set of constrains from structure of the native state of the protein that we study. Let's assume that in the native state, two amino acids with positions i and j along the protein sequence

are separated by a distance d_{ij} bigger than a predefined value D. Then we impose a repulsive interaction of the hard-core type between these two amino acids. This interaction is effective at all times, and not only when the protein adopts the native structure. Thus, we make use of a particular conformation, the native state, to define interactions that will be valid regardless of the particular conformation that the protein will adopt. We set D as the hard-core diameter for the two particular amino acids i and j.

We coarse grain all atoms belonging to one amino acid into a single bead. We place the center of the bead in the C_β carbon of the amino acid (see Fig. 3.2). If the particular amino acid is Glycine, which has no C_β carbon, then we place the center of the bead in the C_α carbon.

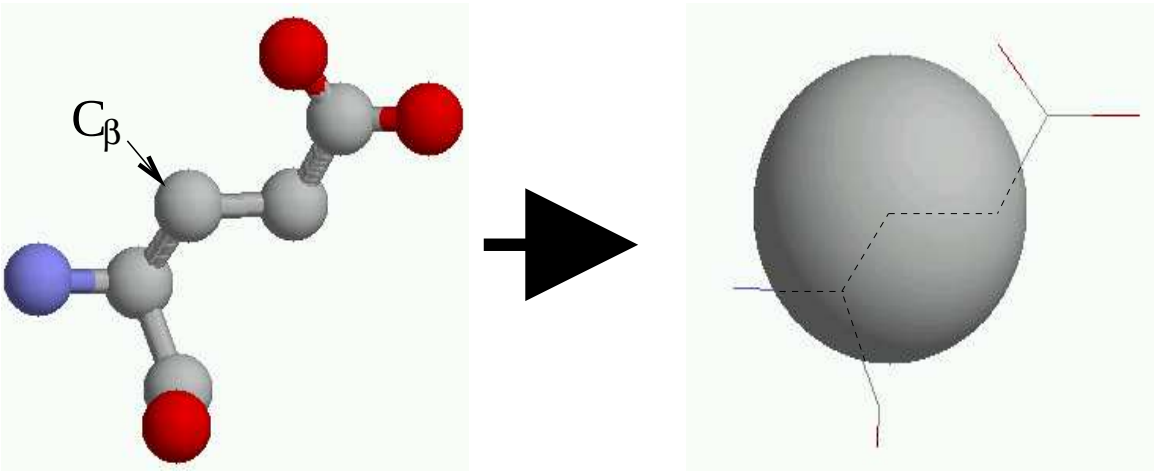

Figure 3.2: Coarse-grain process from all-atom to one-bead. We substitute all atoms of the amino acid by a single sphere centered in the C_β carbon

We connect consecutive beads i and $i+1$ into a single polymer chain with the help of auxiliary "peptide" bonds (see Fig. 3.3). We allow i and $i+1$ to wiggle around a distance $d_{i,i+1}$ whithin certain limits. We choose the distance $d_{i,i+1}$ to be precisely the distance between i and $i+1$ when the protein adopts the native state conformation. All these specifications we gather in the following potential energy between beads i and $i+1$,

$$U(\mid \vec{r}_i - \vec{r}_{i+1} \mid) = \begin{cases} \infty & \text{if} \quad \mid \vec{r}_i - \vec{r}_{i+1} \mid \leq d_{i,i+1}(1-\sigma/2) \\ 0 & \text{if} \quad d_{i,i+1}(1-\sigma/2) < \mid \vec{r}_i - \vec{r}_{i+1} \mid < d_{i,i+1}(1+\sigma/2) \\ \infty & \text{if} \quad \mid \vec{r}_i - \vec{r}_{i+1} \mid \geq d_{i,i+1}(1+\sigma/2) \quad , \end{cases} \tag{3.4}$$

where σ controls the flexibility of the bond.

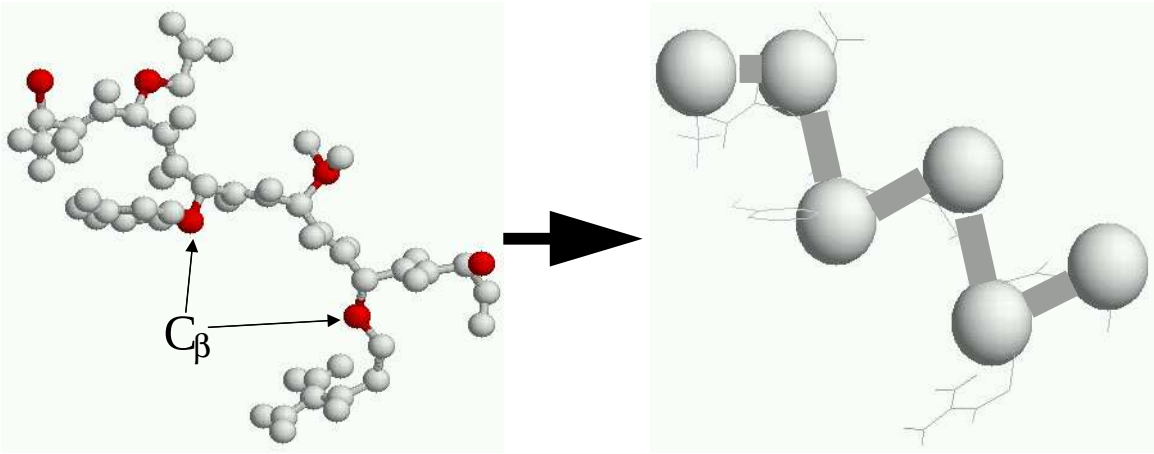

Figure 3.3: Auxiliary peptide bonds between consecutive C_β carbons make up the the polymeric chain.

The simplicity one-bead model allows one to simulate the time evolution of the protein long enough to attain statistical significance in thermodynamical averages, as well as to reproduce many folding transitions to study the kinetics of folding.

3.3.2 Two-bead Model

The next step above the one-bead model is the explicit segregation of the two parts of the amino acid, the backbone and the residue, into two beads (see Fig. 3.4). This segregations allows one to separate the backbone dynamics from the tipical scales corresponding to interactions among residues. In addition, it allows one to mimic the protein flexibility with higher fidelity.

We model a set of constrains to mimic the flexibility of proteins (see Fig. 3.5). The firsts group of contrains bonds consecutive C_α's to model the "peptide" bond. The second

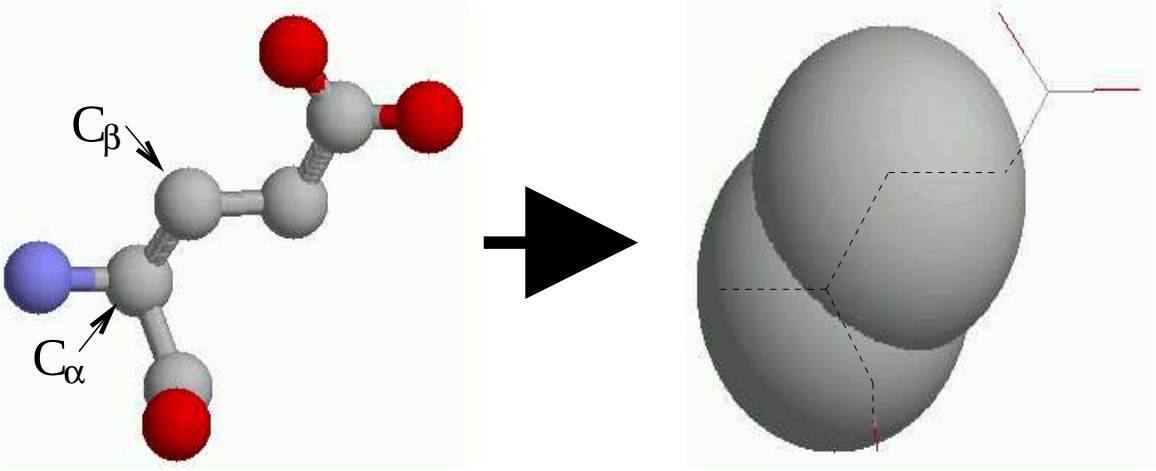

Figure 3.4: Coarse-grain process from all-atom to two-bead. We substitute all atoms of the amino acid by a two spheres centered in the C_α and C_β carbons

group bonds next to consecutive C_α's to model preferred angles between peptide bonds. The third group of constrains bonds the C_β of each amino acid to its C_α. Finally, a fourth group of constrains bonds each C_β to the the C_α's of the previous and next amino acids.

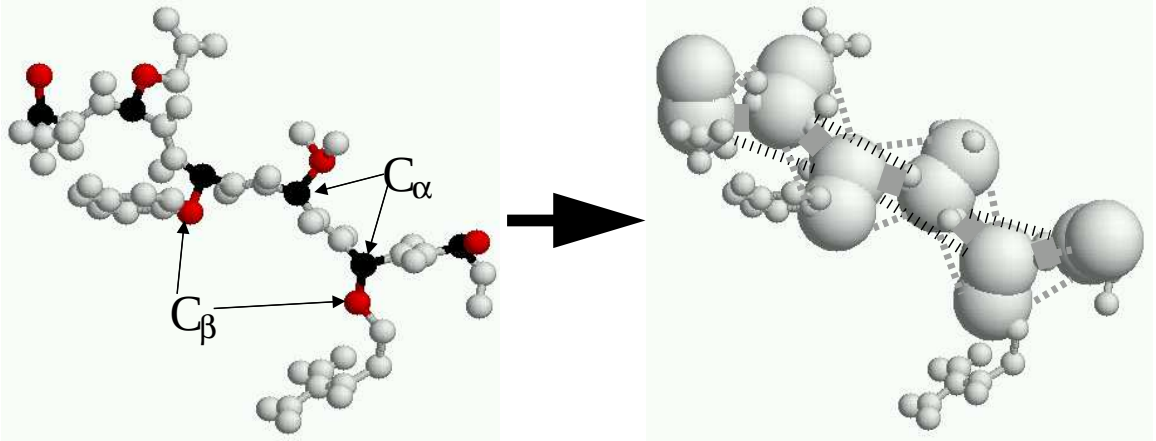

Figure 3.5: Auxiliary peptide bonds between consecutive C_α and C_β carbons make up the the polymeric chain.

A word about the distances. They are retrieved from the protein database.

3.3.3 Four-bead Model

The four bead model targets the backbone conformational changes, allowing to investigate rearrangements of the backbone to the atomic detail. The protein model employs three beads to represent the backbone and one bead to represent the residue [48, 49]. Molecular dynamics studies in such a polypeptide system have shown a sharp helix-coil transition [49], which suggests that it is possible to study other coformational changes of the protein backbone such as the α-helix to β-sheet transition.

(see Fig. 3.6).

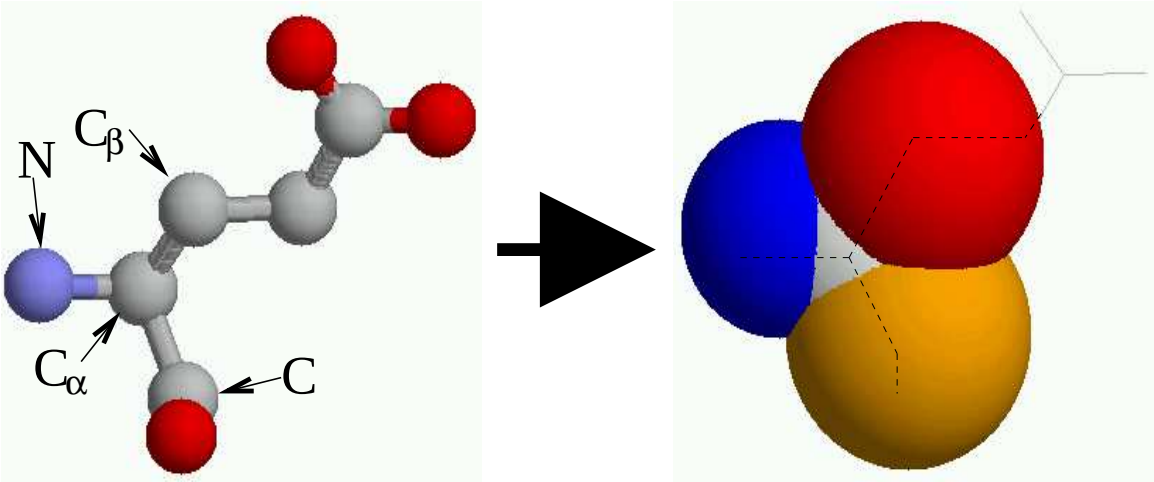

Figure 3.6: Coarse-grain process from all-atom to two-bead. We substitute all atoms of the amino acid by a two spheres centered in the C_α and C_β carbons

(see Fig. 3.7).

A word about the distances. They are retrieved from the protein database.

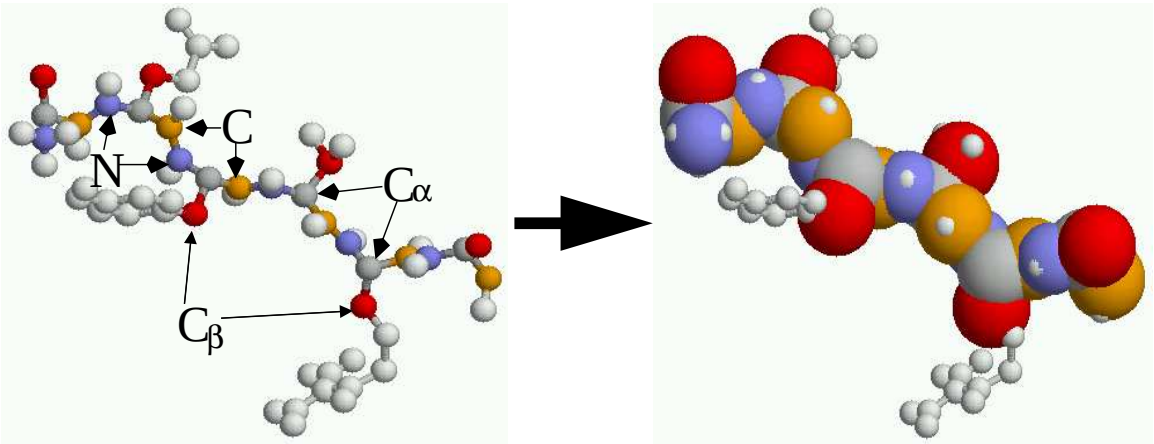

Figure 3.7:

Chapter 4

Protein Stability Studies

4.1 Introduction

Some proteins become thermodynamically unstable at low temperatures, a phenomenon called cold denaturation [45–47, 50]. This phenomenon has been mainly observed at high pressures, ranging from approximately 200 MPa up to 700 MPa [51, 52]. An explanation of the $P - T$ phase diagram of a protein with cold denaturation has been proposed [53, 54], but a microscopic understanding of the mechanisms leading to cold denaturation has yet to be developed, due in part to the complexity of protein-solvent interactions.

Existing theories of folding and unfolding of diluted proteins consider hydrophobicity as the driving force of protein stability [22, 55–64]. In the case of apolar macromolecules, hydrophobicity has been identified with the assembly and segregation of the macromolecule to minimize the disruption of hydrogen bonds among water molecules [55, 61–65]. Water tends to be removed from the surface of apolar molecules, forming a cage composed of highly organized water molecules around the molecule, where the disruption of hydrogen bonds is minimized [66]. The simplest hydrophobic model features an effective attraction between hydrophobic molecules [67], but does not reproduce cold denaturation. In order to obtain cold denaturation with this model, new studies [56, 68] had to insert a temperature-dependent attraction derived from experimental observations at ambient pressure [69]. An explicit account of water around the hydrophobic molecules has also been considered in

order to understand the cold denaturation process with temperature-independent interactions. Theoretical attempts modeled the effective water-protein interactions with the free energy cost of excluding the solvent around the nonpolar molecule [65, 70, 71]. Numerical simulations based on these attempts have been applied to study the pressure denaturation found in proteins [60].

Not until recently has cold denaturation been studied at the microscopic level. Simplified models [72–75], based on a bimodal description of the energy of water in the shell around the hydrophobic molecule [76–78], predicted cold denaturation. Similar results were obtained using a lattice model of a random hydrophobic-hydrophilic heteropolymer interacting with the solvent [79]. Several models mimicking the interaction between water molecules and non-polar monomers have also been applied to the study of cold denaturation. [80–85]. One possible reason for the inability of the previous models to capture both the molecular details of cold denaturation and the effect of pressure is the neglect of (i) correlations among water molecules near the freezing point, and (ii) the density anomaly due to the tetrahedral structure of the hydrogen bonded network.

4.2 Methods

We implement the two-dimensional water lattice model discussed in Sec. 3.2. In addition, we also treat the protein according to the model presented in the same section. Thus, we write the enthalpy of protei plus water system as

$$\mathcal{W} = \mathcal{H}_p + \mathcal{H}_{HB} + PV \tag{4.1}$$

Since the energy landscape of the protein interacting with the water network is characterized by a multitude of local minima, we perform multicanonical Monte Carlo simulations to avoid transient trapping of our water-protein system in local energy minima at low temperatures (Sec. 2.2.1). We adopt the algorithm in order to embed the self-avoiding protein into the lattice, and to calculate the two-parameter density of states $g(N_{HB}, N_c)$, where

$N_c \equiv \sum_{\langle i,j \rangle} n_i n_j$ is the number of residue-residue contacts. From the density of states we calculate the temperature and pressure dependence of the average number of residue-residue contacts

$$\overline{N}_c = \sum_{N_{HB}} \sum_{N_c} N_c g(N_{HB}, N_c) \frac{e^{-\mathcal{W}(N_{HB},N_c)/T}}{Z}, \tag{4.2}$$

where Z is the partition function. We perform Monte Carlo simulations of a system of 383 water molecules and a protein consisting of 17 non-polar residues with periodic boundary conditions.

4.3 Results

4.3.1 Water Model

Above P_c, we find that N_{HB} decreases as we decrease the temperature, and the water model undergoes a transition to a state where all hydrogen bonds with $\Delta V > 0$ are broken. Below P_c, we find that N_{HB} increases as we decrease the temperature, and the water model undergoes a sharp transition, at $T = T_c = (J - P\Delta V)/(\ln(1+\sqrt{q}))$ [44], to a state where all hydrogen bonds with $\Delta V > 0$ are formed. Thus, our water model reproduces the freezing of water to low and high density ice, since for $P < P_c$ ($P_c \approx 200$MPa in real water) water freezes to the low density ice Ih, and for $P > P_c$, water freezes to the high density ice II [86]. A relation between these two phases of ice and protein folding has already been suggested from a thermodynamic point of view [87, 88].

4.3.2 Cold Denaturation

We ploy in Fig. 4.1 the dependence of $\overline{N}_c/n_{max}$ on temperature for different values of the pressure both above and below P_c. The calculated density of states $g(N_{HB}, N_c)$ converges to the true value with an accuracy of the order of 10^{-5}. The value of $\overline{N}_c/n_{max}$ ranges from one (maximally compact protein) to approximately 0.71 (which is the average number of residue–residue contacts found at high temperatures). Only when $P > P_c$ do we observe the

cold denaturation of the protein. Above P_c, n_{HB} decreases monotonically with decreasing temperature, thus the repulsion term $J_r N_{HB}$ also decreases monotonically. We find then that a re-entrant transition as the one found in cold denaturation does not require a non-monotonic function of the temperature for the polymer–water repulsion.

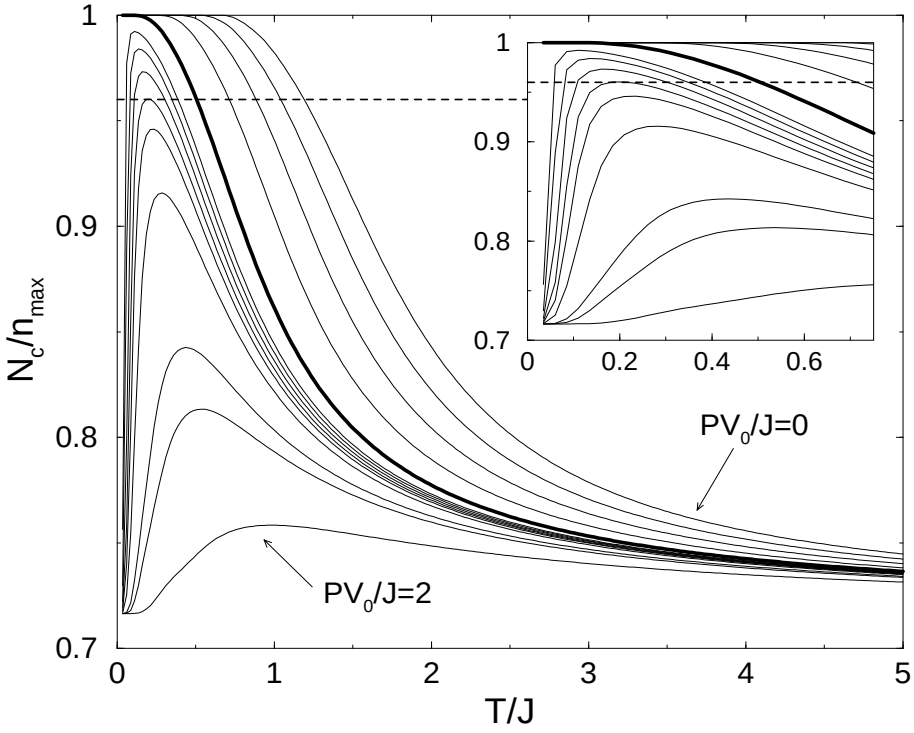

Figure 4.1: Normalized number of residue-residue contacts vs. temperature for pressure values $PV_0/J = 0, 0.25, 0.5, 0.75, 1, 1.1, 1.125, 1.15, 1.175, 1.2, 1.25, 1.4, 1.5, 2$. A magnified region where the cold denaturation takes place is shown in the inset. The dotted line corresponds to $\overline{N}_c/n_{max} > 0.96$. We represent the curve corresponding to $P = P_c = J/\Delta V$ by a bold line. The values of the parameters used are: $J = 1, J_r = 10, \Delta V = 1$ and $q = 10$.

We also reconstruct the phase diagram of the water-protein system in the $P - T$ plane (Fig. 4.2). We consider the protein to be in the collapsed state if 96% of all possible contacts are formed, i.e., if $\overline{N}_c/n_{max} > 0.96$. For each pressure value, the freezing lines of water shown in Fig. 4.2 are given by the temperature at which we observe a maximum in the specific heat of the water bath. We compare our findings to experimental observations [89]

for bovine pancreatic ribonuclease A studied by ^{1}H NMR spectroscopy. We find remarkable qualitative agreement between the experimental and numerical $P-T$ phase diagrams. In both experimental and numerical $P-T$ phase diagrams, we observe that cold denaturation occurs at high pressures and, as we lower the temperature, close to the water–ice II freezing line and in the region where water molecules are not capable of forming low density ice–like structures. In addition to the study of ribonuclease A [89], cold denaturation at very high pressures in the kbar range has also been observed in chymotrypsinogen [53, 54], myoglobin [53, 54], staphylococcal nuclease [90] and has been proposed as the principal mechanism for the observed pressure-inactivation of bacteria, such as *Escherichia coli*.

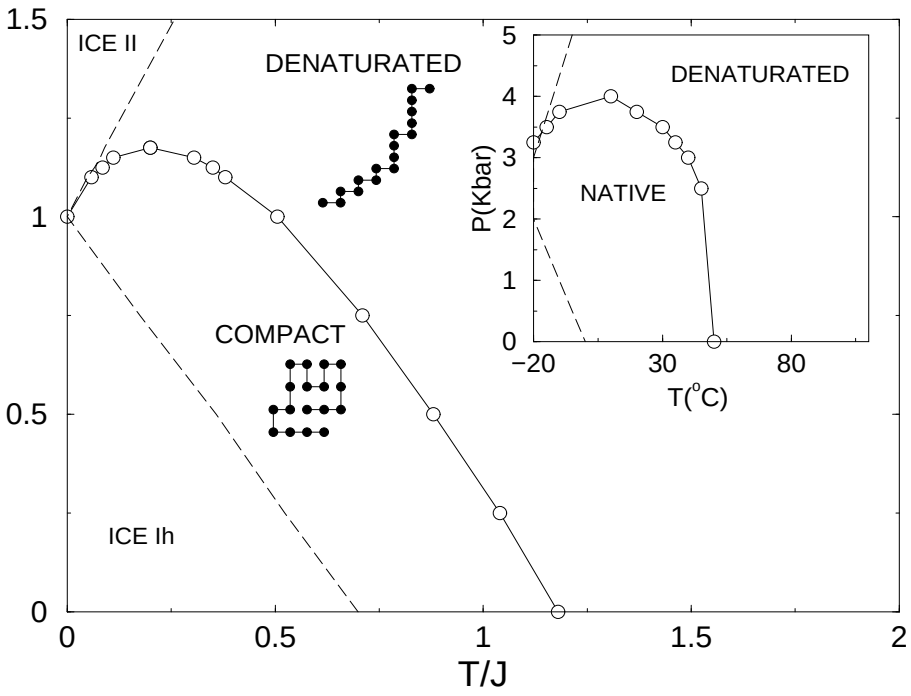

Figure 4.2: P-T phase diagram for the protein derived from Fig. 4.1. Dashed lines indicate the freezing lines for model water. Water freezes in low density ice Ih for $PV_0/J < 1$ and in dense ice II for $PV_0/J > 1$. In the inset we present the experimental results obtained by Zhang et al. [89] for the bovine pancreatic ribonuclease A. Two typical configurations of the protein are shown, one in the compact state and the other in the denaturated state.

Not all proteins behave equally as we decrease temperature at high pressure. In par-

ticular, there are some proteins that do not exhibit cold denaturation. We reproduce the variability of protein dynamics at high pressure and low temperature by varying the hydrophobic parameter J_r to lower values, effectively impeding a stable compact state for pressures above the $P = P_c$ line. In Fig. 4.3 we present the phase diagrams obtained for different values of J_r, ranging from two to 20. The shape of the phase diagram changes as we increase the value of the repulsive interaction J_r, allowing stabilization of the compact state and cold denaturation above the $P = P_c$ line.

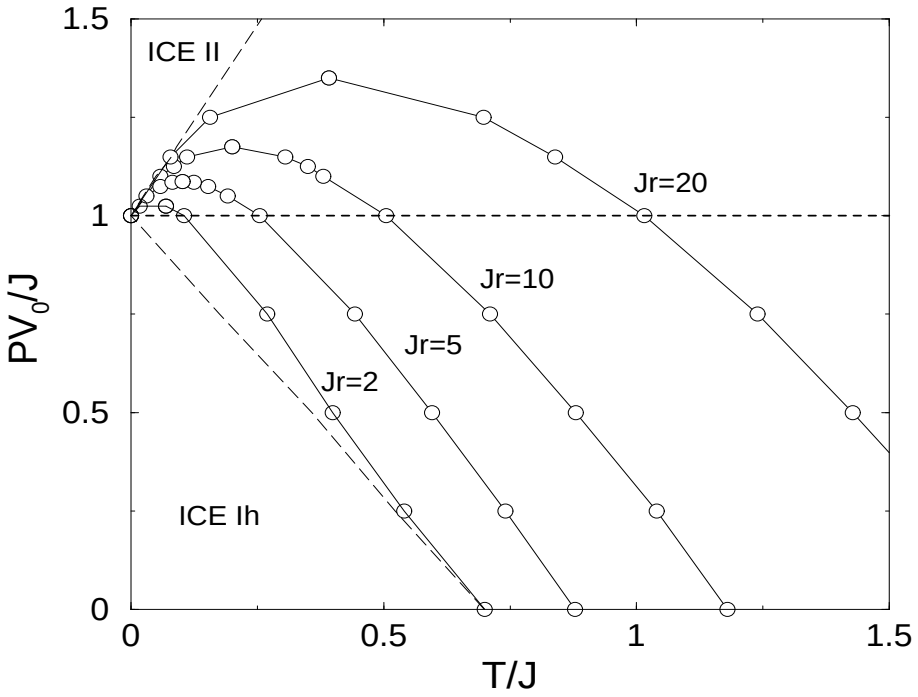

Figure 4.3: P-T Phase diagram for proteins with $J_r = 2, 5, 10, 20$. Dashed lines indicate the computed freezing lines for water. Dotted line indicates the $P = P_c = J/\Delta V$ line.

4.4 Discussion

Within the framework of our model, we reproduce the experimentally-observed *thermodynamics* of cold denaturation. However, we cannot address the *kinetics* of this process. It could be also feasible to investigate in future work the dynamics with a more sophisticated

model, where we consider not just the average number of hydrogen bonds with $\Delta V > 0$, but the actual numbers for each water molecule that is a neighbor to a residue.

Recent computer simulations studies [91] with all-atom models have studied the effect of pressure and average density on the hydrophobic effect. The authors performed simulations of two Lennard-Jones (LJ) particles in the TIP4P water model [92]. At constant temperature, they found that the aggregation of the two particles is favored with a moderate increase of pressure, or analogously, with a moderate increase of water density. However, this effect is reversed for pressures in the kbar range, so that aggregation becomes unstable with increasing water density. These studies support our results of a critical pressure above which an increase of pressure leads both to an increase of water density (because of the reduction of the number of hydrogen bonds with $\Delta V > 0$) and to destabilization of the model protein. Following these results and based on previous studies [61–64], Shimizu et al. [68] have shown that three-body interactions have a destabilization effect on the aggregation of three LJ particles. However, the inclusion of these interactions into a model of a more complex protein did not lead to significant changes. Finally, all-atom simulations recently addressed the pressure denaturation of proteins [93], but they have not provided conclusive evidence.

We conclude that the effect of pressure on water density is key for understanding cold denaturation of proteins. The density anomaly of water arises from the low density hydrogen bonded structures responsible for the hydrophobic effect, driving the protein to a compact state [22, 46, 50, 59, 61–64, 66]. At extreme pressures above P_c, lowering the temperature implies an increasing free energy cost to form a hydrogen bond with $\Delta V > 0$, so the density anomaly disappears. In this scenario, the hydrophobic effect decreases and cold denaturation occurs. Our model supports this mechanism. Also, a specific arrangement of amino acids in the protein structure, determined by amino acid interactions, dictates the dynamics of proteins at low temperature and high pressure, thus making some proteins more stable than others at these $P - T$ conditions.

Chapter 5

Protein Kinetics Studies

To be writtein after Stability Chapter is completed.

Bibliography

[1] T. Creighton. *Proteins: structures and molecular properties, second edition*. W. H. Freeman and Co., New York, 1993.

[2] D. MacEr. *The Human Genome Project*. JAI Press, Stamford, CT, 2004.

[3] C. Branden and J. Tooze. *Introduction to Protein Structure*. Garland Publishing Inc, New York, 1999.

[4] C. Ghelis and J. Yon. *Protein Folding*. Academic Press, New York, 1982.

[5] T. E. Creighton. *Protein Folding*. W. H. Freeman & Co., New York, 1992.

[6] C. Levinthal. Are there pathways for protein folding? *J. Chem. Phys.*, 65:44–, 1968.

[7] V. Villegas, J. C. Martinez, F. X. Aviles, and L. Serrano. Structure of the transition state in the folding process of human procarboxypeptidase A2 activation domain. *J. Mol. Biol.*, 283:1027–1036, 1998.

[8] F. Chiti, N. Taddei, P. M White, M. Bucciantini, F. Magherini, M. Stefani, and C. M. Dobson. Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nature Struct. Biol.*, 6:1005–1009, 1999.

[9] M. Lorch, J. Mason, A. Clarke, and M. Parker. Effects of core mutations on the folding of a beta–sheet protein: implications for backbone organization in the i–state. *Biochemistry*, 38:1377–1385, 1999.

[10] D. E. Kim, C. Fisher, and D. Baker. A breakdown of symmetry in the folding transition state of protein L. *J. Mol. Biol.*, 298:971–984, 2000.

[11] E. L. McCallister, E. Alm, and D. Baker. Critical role of β–hairpin formation in protein G folding. *Nature Struct. Biol.*, 7:669–673, 2000.

[12] K. A. Dill. Theory for the folding and stability of globular proteins. *Biochemistry*, 24:1501–1509, 1985.

[13] N. V. Dokholyan, S. V. Buldyrev, H. E. Stanley, and E. I. Shakhnovich. Molecular dynamics studies of folding of a protein-like model. *Folding & Design*, 3:577–587, 1998.

[14] E. Alm and D. Baker. Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures. *Proc. Natl. Acad. Sci. USA*, 96:11305–11310, 1999.

[15] O. V. Galzitskaya and A. V. Finkelstein. A theoretical search for folding/unfolding nuclei in three-dimensional protein structures. *Proc. Natl. Acad. Sci. USA*, 96:11299–11304, 1999.

[16] V. Muñoz and W. A. Eaton. A simple model for calculating the kinetics of protein folding from three-dimensional structures. *Proc. Natl. Acad. Sci. USA*, 96:11311–11316, 1999.

[17] N. Gō and H. Abe. Non-interacting local-structure model of folding and unfolding transition in globular proteins. I. formulation. *Biopolymers*, 20:991–1011, 1981.

[18] M. Karplus and E. I. Shakhnovich. Protein folding: theoretical studies of thermodynamics and dynamics. [5].

[19] E. I. Shakhnovich. Theoretical studies of protein-folding thermodynamics and kinetics. *Curr. Opin. Struc. Biol.*, 7:29–40, 1997.

[20] H. Taketomi, Y. Ueda, and N. Gō. Studies on protein folding, unfolding and fluctuations by computer simulations. *Int. J. Peptide Protein Res.*, 7:445–459, 1975.

[21] J. D. Bryngelson and P. G. Wolynes. Intermediates and barrier crossing in a random energy model (with applications to protein folding). *J. Phys. Chem.*, 93:6902–6915, 1989.

[22] K. A. Dill. Dominant forces in protein folding. *Biochemistry*, 29:7133–7155, 1990.

[23] E. I. Shakhnovich. Proteins with selected sequences fold into unique native conformation. *Phys. Rev. Lett.*, 72:3907–3910, 1994.

[24] V. I. Abkevich, A. M. Gutin, and E. I. Shakhnovich. Free–energy landscape for protein–folding kinetics — intermidiates, traps, and multiple pathways in theory and lattice model simulations. *J. Chem. Phys.*, 7:6052–6062, 1994.

[25] A. M. Gutin, V. I. Abkevich, and E. I. Shakhnovich. Evolution–like selection of fast–folding model proteins. *Proc. Natl. Acad. Sci. USA*, 92:1282–1286, 1995.

[26] E. I. Shakhnovich, V. I. Abkevich, and O. Ptitsyn. Conserved residues and the mechanism of protein folding. *Nature*, 379:96–98, 1996.

[27] J. Saven and P. Wolynes. Statistical mechanics of the combinatorial synthesis and analysis of folding molecules. *J. Chem. Phys.*, 101:8375–8389, 1997.

[28] P. L. Privalov. Thermodynamic problems of protein structure. *Ann. Rev. Biophys. Biophys. Chem.*, 18:47–69, 1989.

[29] D. K. Klimov and D. Thirumalai. Factors governing the foldability of proteins. *Proteins*, 26:411–441, 1996.

[30] K. Binder and D. W. Heerman. *Monte Carlo simulations in statistical physics.* Springer-Verlag, Berlin, 1992.

[31] M. E. J. Newman and G. T. Barkema. *Monte Carlo methods in statistical physics.* Clarendon Press, Oxford, 1999.

[32] B A. Berg and T. Neuhaus. Multicanonical algorithms for first order phase transitions. *Phys. Lett. B*, 267:249–253, 1991.

[33] B. A. Berg and T. Neuhaus. Multicanonical ensemble: A new approach to simulate first-order phase transitions. *Phys. Rev. Lett.*, 68:9–12, 1992.

[34] J. Lee. New monte carlo algorithm: Entropic sampling. *Phys. Rev. Lett.*, 71:211?214, 1993.

[35] F. Wang and D. P. Landau. Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.*, 86:2050–2053, 2001.

[36] A. Baumgartner. *Applications of the Monte Carlo simulations in statistical physics*. Springer-Verlag, New York, 1987.

[37] A. Irback and H. Schwarze. Sequence dependence of self–interacting random chains. *J. Phys. A: Math. Gen.*, 28:2121–2132, 1995.

[38] G. F. Berriz, A. M. Gutin, and E. I. Shakhnovich. Cooperativity and stability in a Langevin model of protein like folding. *J. Chem. Phys.*, 106:9276–9285, 1997.

[39] Z. Guo and C. L. Brooks. Thermodynamics of protein folding: a statistical mechanical study of a small all–β protein. *Biopolymers*, 42:745–757, 1997.

[40] B. J. Alder and T. E. Wainwright. Studies in molecular dynamics. I. general method. *J. Chem. Phys.*, 31:459–466, 1959.

[41] D. C. Rapaport. *The art of molecular dynamics simulation*. Cambridge University Press, Cambridge, 1997.

[42] A. C. Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, 72:2384–2393, 1980.

[43] H. J. C. Berendsen, J. Poastma, W. Van Gunsteren, A. DiNola, and J. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–3690, 1984.

[44] G. Franzese, M. I. Marqués, and H. E. Stanley. Intramolecular coupling as a mechanism for a liquid-liquid phase transition. *Phys. Rev. E*, 67:011103, 2003.

[45] C. N. Pace and Ch. Tanford. Thermodynamics of the unfolding of β-lactoglobulin A in aqueous urea solutions between 5 and 55°. *Biochemistry*, 7:198–208, 1968.

[46] P. L. Privalov. Cold denaturation of proteins. *Crit. Rev. Biochem. Mol. Biol.*, 25:281–305, 1990.

[47] J. Jonas. Cold denaturation of proteins. *ACS Symp. S*, 676:310–323, 1997.

[48] S. Takada, Z. Luthey-Schulten, and P. G. Wolynes. Folding dynamics with nonadditive forces: A simulation study of a designed helical protein and a random heteropolymer. *J. Chem. Phys.*, 110:11616–11629, 1999.

[49] A. V. Smith and C. K. Hall. α-helix formation: Discontinuous molecular dynamics on an intermediate-resolution protein model. *Proteins: Struct. Func. & Genet.*, 4:344–360, 2001.

[50] G. P. Privalov and P. L. Privalov. Problems and prospects in microcalorimetry of biological macromolecules. *Methods. Enzymol.*, 323:31–62, 2000.

[51] S. Kunugi and N. Tanaka. Cold denaturation of proteins under high pressure. *BBA–Prot. Struct. Mol. Enz.*, 1595:329–344, 2002.

[52] L. Smeller. Pressure-temperature phase diagrams of biomolecules. *BBA–Prot. Struct. Mol. Enz.*, 1595:11–29, 2002.

[53] S. A. Hawley. Reversible pressure-temperature denaturation of chymotrypsinogen. *Biochemistry*, 10:2436–2442, 1971.

[54] A. Zipp and W. Kauzmann. Pressure denaturation of metmyoglobin. *Biochemistry*, 12:4217–4228, 1973.

[55] W. Kauzmann. *Adv. Prot. Chem.*, 14:1–, 1959.

[56] K. A. Dill, D. 0. V. Alonso, and K. Hutchinson. Thermal stabilities of globular proteins. *Biochemistry*, 28:5439–5449, 1989.

[57] K. F. Lau and K. A. Dill. A lattice statistical-mechanics model of the conformational and sequence-spaces of proteins. *macromolecules*, 22:3986–3997, 1989.

[58] K. A. Dill and D. Stigter. Modeling protein stability as heteropolymer collapse. *Adv. Prot. Chem.*, 46:59–104, 1995.

[59] H. S. Chan and K. A. Dill. The protein folding problem. *Phys. Today*, 46:24–32, 1993.

[60] G. Hummer, S. Garde, A. E. García, M. E. Paulaitis, and L. R. Pratt. The pressure dependence of hydrophobic interactions is consistent with the observed pressure denaturation of proteins. *Proc. Natl. Acad. Sci. USA*, 95:1552–1555, 1998.

[61] D. Chandler. Hydrophobicity: Two faces of water. *Nature*, 417:491–491, 2002.

[62] P. R. ten Wolde and D. Chandler. Drying-induced hydrophobic polymer collapse. *Proc. Natl. Acad. Sci. USA*, 99:6539–6543, 2002.

[63] K. Lum, D. Chandler, and J. D. Weeks. Hydrophobicity at small and large length scales. *J. Phys. Chem. B*, 103:4570–4577, 1999.

[64] P. R. ten Wolde. Hydrophobic interactions: an overview. *J. Phys: Cond Matt.*, 14:9445–9460, 2002.

[65] F. H. Stillinger. *J. Solut. Chem.*, 2:141–, 1973.

[66] H. S. Frank and M. W. Evans. *J. Chem. Phys.*, 13:507–, 1945.

[67] C. Vanderzande. *Lattice Models of Polymers.* Cambridge University Press, Cambridge, 1998.

[68] S. Shimizu and H. S. Chan. Anti-cooperativity and cooperativity in hydrophobic interactions: Three-body free energy landscapes and comparison with implicit-solvent potential functions for proteins. *Proteins: Struct. Func. & Genet.*, 48:15–30, 2002.

[69] Y. Nozaki and C. Tanford. *J. Biol. Chem.*, 246:2211–2217, 1971.

[70] L. R. Pratt and D. Chandler. *J. Chem. Phys.*, 67:3683–, 1977.

[71] R. A. Pierotti. *J. Chem. Phys.*, 67:1840–, 1963.

[72] P. De los Rios and G. Caldarelli. Cold and warm swelling of hydrophobic polymers. *Phys. Rev. E*, 63:Art. No. 031802, 2001.

[73] P. De los Rios and G. Caldarelli. Putting proteins back into water. *Phys. Rev. E*, 62:8449–8452, 2000.

[74] G. Caldarelli and P. De los Rios. Cold and warm denaturation of proteins. *J. Biol. Phys.*, 27:229–241, 2001.

[75] A. Hansen, M. H. Jensen, K. Sneppen, and G. Zocchi. Statistical mechanics of warm and cold unfolding in proteins. *Eur. J. Phys. B*, 6:157–161, 1998.

[76] N. Muller. Search for a realistic view of hydrophobic effects. *Acc. Chem. Res*, 23:23–28, 1990.

[77] B. Lee and G. Graziano. A two-state model of hydrophobic hydration that produces compensating enthalpy and entropy changes. *J. Am. Chem. Soc.*, 118:5163–5168, 1996.

[78] K. A. T. Silverstein, A. D. J. Haymet, and K. A. Dill. Molecular model of hydrophobic solvation. *J. Chem. Phys.*, 111:8000–8009, 1999.

[79] A. Trovato, J. van Mourik, and A. Maritan. Swollen-collapsed transition in random hetero-polymers. *Eur. J. Phys. B*, 6:63–73, 1998.

[80] P. Bruscolini and L. Casetti. Lattice model for cold and warm swelling of polymers in water. *Phys. Rev. E*, 61:R2208–R2211, 200.

[81] P. Bruscolini, C. Buzano, and A. Pelizzola A. Pretti. Bethe approximation for a model of polymer solvation. *Phys. Rev. E*, 64:Art. No. 050801, 2001.

[82] A. Ball. Two-state protein model with water interactions: Influence of temperature on the intrinsic viscosity of myoglobin. *Phys. Rev. E*, 63:Art. No. 061906, 2001.

[83] A. Bakk, J. S. Hoye, and A. Hansen. Heat capacity of protein folding. *Biophys. J.*, 81:710–714, 2001.

[84] A. Bakk, J. S. Hoye, and A. Hansen. Apolar and polar solvation thermodynamics related to the protein unfolding process. *Biophys. J.*, 82:713–719, 2002.

[85] O. Collet. Warm and cold denaturation in the phase diagram of a protein lattice model. 53:93–99, 2001.

[86] V. F. Petrenko and R. W. Whitworth. *Physics of Ice.* Oxford University Press, 1999.

[87] G. W. Robinson and C. H. Cho. Role of hydration water in protein unfolding. *Biophys. J.*, 77:3311–3318, 1999.

[88] C. J. Tsai, J. V. Maizel, and R. Nussinov. The hydrophobic effect: A new insight from cold denaturation and a two-state water structure. *Crit. Rev. Biochem. Mol. Biol.*, 37:55–69, 2002.

[89] NMR-study of the cold, heat, and pressure unfolding of ribonuclease A. *Biochemistry*, 34:8631–8641, 1995.

[90] Differences between the pressure- and temperature-induced denaturation and aggregation of β-lactoglobulin A, B, and AB monitored by FT-IR spectroscopy and small-angle X-ray scattering. *Biochemistry.*

[91] V. A. Payne, N. Matubayasi, L. R. Murphy, and R. M. Levy. Monte Carlo study of the effect of pressure on hydrophobic association. *J. Phys. Chem. B*, 101:2054–2060, 1997.

[92] M. W. Mahoney and W. L. Jorgensen. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.*, 112:8910–8922, 2000.

[93] E. Paci. High pressure simulations of biomolecules. *Biochim. Biophys. Acta.*, 1595:185–200, 2002.

Curriculum Vitæ

Jose M. Borreguero

Boston University, Physics Department
Center For Polymer Studies
590 Commonwealth Avenue
Boston, Massachusetts 02215 USA

Telephone: 617/353-8051
Facsimile: 617/353-9393 or 617/353-3783
E-mail: jmborr@bu.edu
WWW: http://polymer.bu.edu/~jmborr

EDUCATION

- Ph.D., 2004 - Physics; Boston University, Boston, MA
- M.A., 2001 - Physics; Boston University, Boston, MA
- B.S., 1997 - Physics; Basque Country University, Spain

EMPLOYMENT

- Research Assistant, Physics Department, Boston University. Spring 2002 - present.
- Teaching Assistant, Physics Department, Boston University. Fall 1998 - Spring 2001.

HONORS, AWARDS

- NECSI Travel Award for International Conference on Complex Systems, 2002.
- UNESCO Student Award for International Conference on Theoretical Physics, 2002.
- ACS Petroleum Research Fund awarded to H. E. Stanley (principal co-author). 2001 — 2003.
- NSF Student Award for APS meeting 2002.
- NSF Student Award for STATPHYS 21, 2001.

COMPUTER SKILLS

Platforms: SGI ORIGIN2000, IBM RS/6000 SP, DEC, PC, BEOWULF.

OS: DOS, Windows, UNIX, LINUX.

languages: C/C++, Fortran9x, Perl.

LANGUAGES

English (proficient), Spanish (native), French (conversant), Basque (conversant).

PRESENTATIONS

- International Conference on Complex Systems, Nashua, USA, 2002 - *Fluctuation analysis in the Transition State Ensemble of the SH3 domain* (oral presentation).
- International Conference on Theoretical Physics, Paris, 2002 - *Topological Frustration of the SH3 domain* (poster presentation).
- American Physical Society meeting, Indianapolis, USA, 2002 - *Thermodynamic and Folding Kinetic Analysis of the SH3 Fold* (poster presentation).
- Statphys 21, Cancun, Mexico, 2001 - *The Folding Nucleus of c–Crk SH3 Domain* (poster presentation).
- International School of Physics "Enrico Fermi" Course CXLV: "Protein folding, evolution and design", Varenna, Italy, 2000 - *The Transition State Ensemble of c–Crk SH3 Domain* (poster presentation).

REFERENCES

- H. Eugene Stanley: University Professor of Physics, Professor of Physiology, Director of Center for Polymer Studies; Department of Physics, Boston University, 590 Commonwealth Avenue, Boston, MA 02215. **Tel:** (617) 353 2617, **FAX:** (617) 353 3783,

  Email: hes@bu.edu.

- Eugene I. Shakhnovich: Professor of Chemistry; Department of Chemistry, Harvard University, 12 Oxford St., Cambridge, MA 02138. **Tel:** (617) 495 4130, **FAX:** (617) 496 5948

  Email: eugene@belok.harvard.edu

- Nikolay V. Dokholyan: Assistant Professor; Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, 303 Mary Ellen Jones, CB#7260 Chapel Hill, NC 27599. **Tel:** (919) 843 2513, **Fax:** (919) 966 2852

Email: dokh@med.unc.edu

PUBLICATIONS

Original Articles

[1] Jose M. Borreguero, Nikolay V. Dokholyan, Sergey V. Buldyrev, Eugene I. Shakhnovich, H. Eugene Stanley, "Thermodynamics and Folding Kinetics Analysis of the SH3 Domain from Discrete Molecular Dynamics", *J. Mol. Biol.* **318**, 863-876 (2002).

[2] Manuel I. Marques, Jose M. Borreguero, H. Eugene Stanley, and Nikolay V. Dokholyan, "A possible mechanism for cold denaturation of proteins at high pressure", Phys. Rev. Lett (2003), accepted for publication.

[3] F. Ding, J.M. Borreguero, S.V. Buldyrev, H.E. Stanley, and N.V. Dokholyan, "Mechanism for the alpha-helix to beta-hairpin transition", *Proteins: Structure, Function, and Genetics* **53**, 220-228 (2003).

[4] J.M. Borreguero, F. Ding, S.V. Buldyrev, H.E. Stanley, and N.V. Dokholyan, "Multiple Folding Pathways of the SH3 domain", *Biophysical Journal*, submitted (2004).

Review Articles

N.V. Dokholyan, J.M. Borreguero, S.V. Buldyrev, F. Ding, H.E. Stanley, and E.I. Shakhnovich, "Identifying the importance of amino acids for protein folding from crystal structures." Methods in Enzymology, Vol. 374: Macromolecular crystallography D. Editors: C.W. Carter Jr. and R.M. Sweet (2003).