Unfolding of proteins : Thermal and mechanical unfolding

By Joe S. Hur & Eric Darve

1. Motivation and Objectives

Over the past few decades, researchers have sequenced the human genome which is a blueprint for proteins. However, given only the information of the sequence we cannot yet accurately predict specific structural information such as the secondary and tertiary structure of a given protein, let alone its functionality in biological processes. Recent theoretical and experimental findings have shown that the topology or conformation of the native structure of small proteins plays a critical role in determining its biological function (Baker 2000; Alm & Baker 1999). In order to carry out their biological function properly, proteins must assume a shape, assembling themselves into an ordered structure. It is now well known that depending on the topology of proteins, for example, when proteins do not fold correctly, there can be serious medical complications, such as Parkinson's disease, Alzheimers to name a few. Furthermore, more biophysical manifestations of protein-protein interactions are reflected in processes related to phase equilibria such as crystallization, and in the marked dependence of the diffusion coefficients of macromolecules on their concentration (Price *et al.* 1999). The latter aspect is currently becoming of much interest to biologists and chemists with advances in techniques for monitoring protein diffusion in the crowded cellular environment (Dayel *et al.* 1999).

With ever increasing areas of interest in biomedical applications as mentioned above, much effort has been put into understanding the mechanism of protein folding. However, at present, there exists neither a simple and universally applicable theoretical framework nor an efficient and accurate computational framework that can account for many experimental findings. Current theories resort to mathematics that vary greatly in complexity and analytic tractablity in order to solve technical difficulties inherent in the problem or start from a phenomenological model (Wolynes *et al.* 1995; Clementi *et al.* 2000). On the computation frontier, with advances in computational resources, we are now equipped with better tools to tackle complex problems that are numerically expensive. However, for example, there is much controversy over the correct form of potential in the force field in molecular dynamics (Baker 2000) and folding proteins using Monte-Carlo simulation is still expensive to be applied to bigger proteins - the folding time scales for these proteins lie between a few milliseconds to minutes. Solving for the stable structures of the isolated proteins should not be the final aim for computer simulations. After all, proteins perform their biological function by interacting with other macromolecules through, for example, electrostatic and non-polar interactions, and thus a clear physical understanding of such phenomena is needed.

Our goal is to understand the mechanisms of protein folding-unfolding in the presence of an external force field - e.g. mechanical, and electrostatic fields - and to investigate the differences in the pathways of force-induced unfolding and thermally or denaturant-induced unfolding. For example, mechanical stability is very important for proteins that form muscle fibres, like myosin and kinesin, and for those that have to withstand forces

like cell-adhesion molecules that stabilize and form the contact between cells in tissues. Having a clear understanding of the mechanisms of unfolding will provide us with a means to design proteins to carry out specific functions as well as to tackle more complex problems that remain unsolved. A few examples of the latter include understanding how a substrate may be attracted to the active site of an enzyme by electrostatic interactions and how the local and global structures of proteins change upon ligand docking.

Recently, relevant to this work, researchers have used atomic force microscopy (AFM) and optical tweezers for dynamic measurement of mechanical unfolding of individual Titin immunoglobulin domains at the single molecule level (Rief *et al.* 1997; Kellermayer *et al.* 1997; Tskhovrebova *et al.* 1997). The widely studied protein Titin is a giant 3 MDalton muscle protein and a major constituent of the sarcomere in vertebrate striated muscle. It is a multi-domain protein which forms filaments approximately 1 μm in length spanning half a sarcomere and has a number of functions including the control of assembly of muscle thick filaments, a role in muscle elasticity and the generation of passive tension. Of the two regions - the I-band and the A-band - we are interested in the I-band of Titin because it plays an important role in muscle elasticity. The I-band is composed of a head to tail linear array of immunoglobulin domains interrupted at intervals by less highly structured linker sequences. All of the immunoglobulin domains are predicted to have the same basic structure. The notation 'I27' is used to represent the 27th domain of the Titin I-band. The wild type Titin protein is far too large for its thermodynamic and kinetic properties to be studied in detail using current techniques, however its multi-domain structure allows investigation of its properties by characterization of the constituent domains in isolation. This approach is frequently used for proteins where, to a first approximation, the domains behave independently.

An interesting aspect of the recent studies of Titin concerns the highly cooperative manner in which the domains break. However, the structural changes under mechanical forces could only be inferred from the force-extension curves in the experiment and for example, the stability of the individual beta-sheets could not be examined. To support the experimental findings using Titin, steered molecular dynamics (SMD) simulations showed that the force-induced unfolding of Titin domains is an all-or-none event, lacking stable intermediates (Lu *et al.* 1998). However, a more recent study by Paci *et al.* showed in their simulation a more complex behavior of force-induced unfolding in contrast to the simple sawtooth profile observed in the unfolding experiments and claimed that in those experiments only the onset of the unfolding of individual domains in multi-domain proteins was revealed (Paci & Karplus 2000).

Even though the experimental and theoretical investigations up to date have given us valuable information on force-induced unfolding of small proteins, the tools to probe multiple kinetic pathways, more complex structures and stability of domains in bigger proteins that carry out biological functions upon docking (where they have to assume a particular partially unfolded shape) are yet to be developed. In this article, we present a Hamiltonian model which builds on the important interactions of the native-state topology. We make a Gaussian approximation within the model such that the contact probabilities are determined self-consistently, in a spirit similar to a local mean-field approximation.

The paper is organized as follows. In §2, we describe our mathematical formulation. The Hamiltonian is defined and the self-consistent Gaussian approximation is introduced in calculating the contact pair probabilities. In §3.1 we present results for the equilibrium properties of several globular proteins and the immunoglobulin domain of Titin. Both

structural and thermodynamic quantities are examined. In §3.2 the thermal denaturation of the immunoglobulin domain of Titin is investigated. In particular, the stability of the six sheet interactions are examined. In §3.3 the results for pulling the ends with constant mechanical forces of a single immunoglobulin domain of Titin are presented and compared to the experimental findings as well as the results in §3.2. We conclude with a brief summary in §4.

2. Method

In our formulation, the Hamiltonian for a given protein of N residues consists of three energy contributions as shown in eqn.(2.1). The first term refers to the energy fluctuations of adjacent residues with respect to the native state. The second term denotes pairwise interactions between non-adjacent (non-consecutive) residue pairs and the last term is the imposed external force field which is assumed to be linear for direct comparison to experiment by Rief *et al.* (1997); Kellermayer *et al.* (1997); Tskhovrebova *et al.* (1997).

$$H = \frac{k_B \cdot \max(T, T_{min})}{2} \sum_{i=1}^{N-1} K(\vec{t_{i+1}} - \vec{t_i})^2 - \sum_{i,j} \frac{\Delta_{ij}}{2} [R^2 - (\vec{t_i} - \vec{t_j})^2] \Theta[R^2 - (\vec{t_i} - \vec{t_j})^2] + F^T t \tag{2.1}$$

In eqn.(2.1), $\vec{t_i}$ denotes the position vector of residue i relative to its native state. Δ_{ij} is the contact matrix of all residues in the native state and Θ is the Heaviside step function, which is defined as :

$$\Theta(x) = \begin{cases} 0 & \text{if } x \leq 0, \\ 1 & \text{if } x > 0. \end{cases} \tag{2.2}$$

We determine Δ_{ij} from the native state structures of proteins from the Protein Data Bank (PDB). We assign a value of 1 to residue pairs whose α-carbons(C_α) are separated by less than $6.5\mathring{A}$ and 0 otherwise.

In our formulation, pairwise interactions that are separated beyond a cut-off radius of R do not contribute the total energy of the system. We set R to $3\mathring{A}$ in all calculations. Note that the distance between two adjacent α-carbons is $3.78\mathring{A}$ and $3.63\mathring{A}$ in a trans and cis configuration respectively. On the other hand, pairwise interactions within the cut-off radius are modelled to vary quadratically with Δ_{ij}, the contact matrix, given as the weight factor. The requirement for a minimum cut-off temperature, T_{min}, is to ensure that at low temperatures the ratio between the energies arising from the fluctuations of adjacent residues and other energy contributions is finite. Note that in our formulation without specifying a cut-off temperature, C_p diverges as temperature approaches zero. We determine T_{min} from the mimina ($\frac{\partial C_v}{\partial T} = 0$) in the heat capacity constant C_v - temperature curves (C_v versus T) for each protein. We calculate the heat capacity from the following relation,

$$C_v = -T \frac{\partial^2 F}{\partial T^2}, \tag{2.3}$$

where F is the total free energy. The second derivative is calculated using a fourth order finite difference scheme.

Due to the symmetry of the Hamiltonian, the corresponding partition function Z is

singular.

$$Z = \int \Pi_i d\vec{r_i} \exp(-\beta H(\vec{r})) \tag{2.4}$$

By adding an additional spring to the first term in eqn. (2.1), one can break the symmetry in the Hamiltonian and easily decouple the extra energy contribution in the total free energy, where the free energy is related to the partition function Z as :

$$F = -T\ln(Z) = -T\ln\left(\int \Pi_i d\vec{r_i} \exp(-\beta H(\vec{r}))\right). \tag{2.5}$$

The modified Hamiltonian is given by,

$$\begin{aligned} H = \frac{k_B \cdot \max(T, T_{min})}{2}[\sum_{i=1}^{N-1} K(\vec{t_{i+1}} - \vec{t_i})^2 + K_s t_N^2] \\ -\sum_{i,j} \frac{\Delta_{ij}}{2}[R^2 - (\vec{t_i} - \vec{t_j})^2]\Theta[R^2 - (\vec{t_i} - \vec{t_j})^2] + F^T t \end{aligned} \tag{2.6}$$

where K_s is the spring constant for the artificially added spring t_N^2. By making use of the self-consistent Gaussian approximation,

$$p_{ij} =< \Theta[R^2 - (\vec{t_i} - \vec{t_j})^2] >, \tag{2.7}$$

Eqn.(2.6) can be rewritten in a compact matrix form as follows:

$$\beta H = \frac{1}{2}(t - \beta MF)^T M^{-1}(t - \beta MF) - \frac{1}{2}\beta^2 F^T MF. \tag{2.8}$$

β is the inverse temperature $(1/k_B T)$ where k_B is set to unity for convenience and the inverse of matrix M is defined as :

$$\begin{aligned} M^{-1} = \delta_{ij}(K(2 - \delta_{i1} - (1 + \frac{K_s}{K})\delta_{iN}) + \frac{2}{T}\sum_k \Delta_{ik} p_{ik}) \\ +(1 - \delta_{ij})(\frac{-2}{T} p_{ij}\Delta_{ij} + K(\delta_{i,j-1} - \delta_{i,j+1})). \end{aligned} \tag{2.9}$$

In eqn.(2.9), δ_{ij} is the Kronecker delta and p_{ij} denotes the contact probability of the residue pair i and j. The probability density function of $\vec{t}$ is then given by,

$$P(\vec{t}) = (2\pi)^{-\frac{3N}{2}} (\det M)^{-\frac{3}{2}} \exp(-\frac{1}{2}(t - \beta MF)^T M^{-1}(t - \beta MF)). \tag{2.10}$$

Previously, Micheletti et al. have proposed a self-consistent pair probability approximation for examining equilibrium properties of proteins, and we will follow the same methodology (Micheletti *et al.* 2001). First the contact pair probability is given by

$$p_{ij} = (2\pi)^{-\frac{3N}{2}} (\det M)^{-\frac{3}{2}} \int_{|t_i - t_j| < R} \Pi_k du_k \exp(-\frac{1}{2}(t - \beta MF)^T M^{-1}(t - \beta MF)). \tag{2.11}$$

Physically, p_{ij} corresponds to the pair contact probability of residue i and j that are separated by less than $R = 3\mathring{A}$, and thus is a measure of the structural and conformational deviation from the native state. By making a self-consistent Gaussian approximation for p_{ij} with the given Hamiltonian in eqn.(2.6), we can calculate p_{ij} in an iterative manner.

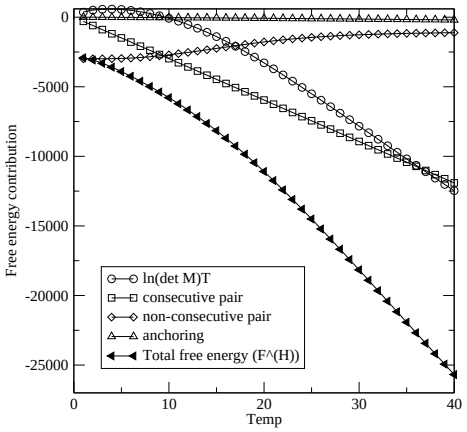

FIGURE 1. Free energy (F^H) versus temperature. Each free energy contribution is also plotted.

First, we transform eqn.(2.11) into spherical coordinates,

$$p_{ij} = \left(\frac{1}{2\pi G_{ij}}\right)^{\frac{3}{2}} \int \Theta(R^2 - |\vec{r} + \vec{L_{ij}}|^2) \exp^{-\frac{r^2}{2G_{ij}}} d\vec{r}. \tag{2.12}$$

In eqn.(2.12), G_{ij} is defined in terms of the matrix M_{ij},

$$G_{ij} = M_{ii} + M_{jj} - 2M_{ij}, \tag{2.13}$$

and $\vec{L}_{ij}$ is the imposed force field defined as

$$\vec{L_{ij}} = \frac{\vec{F}}{T}(M_{i1} - M_{iN} - M_{j1} + M_{jN}) \tag{2.14}$$

for the case of pulling the ends of a protein equally with a force vector $\vec{F}$. By utilizing the isotropicity of spherical coordinates, eqn.(2.12) can be further simplified and expressed as a series of Incomplete Gamma function of order $\frac{1}{2}$ i.e., $P_{\frac{1}{2}}(x)$ which is defined as :

$$P_{\frac{1}{2}}(x) = \frac{1}{\sqrt{\pi}} \int_0^x \exp^{-t} t^{-\frac{1}{2}} dt, \tag{2.15}$$

and p_{ij} is given by :

$$\begin{aligned} p_{ij} = {} & \frac{\sqrt{G_{ij}}}{\sqrt{2\pi}|L|} \left[\exp^{-\frac{(R+|L|)^2}{2G_{ij}}} - \exp^{-\frac{(R-|L|)^2}{2G_{ij}}}\right] \\ & + \frac{1}{2}\left[P_{\frac{1}{2}}\left(\frac{(R-|L|)^2}{2G_{ij}}\right) + P_{\frac{1}{2}}\left(\frac{(R+|L|)^2}{2G_{ij}}\right) - 2P_{\frac{1}{2}}\left(\frac{|L|^2}{2G_{ij}}\right)\right] \end{aligned} \tag{2.16}$$

In the case of no applied force, eqn.(2.16) reduces to

$$p_{ij} = \frac{2}{\sqrt{\pi}} P_{\frac{3}{2}}\left[\frac{R^2}{2G_{ij}}\right], \tag{2.17}$$

where $P_{\frac{3}{2}}$ is the incomplete Gamma function of order $\frac{3}{2}$.

We first start each calculation by initializing all p_{ij}'s to 0.5. Two different initial values of 0.25 and 0.75 were used to ensure that the results are independent of initial conditions. The matrix M_{ij}^{-1} is then constructed using the initial values of p_{ij}'s and the contact

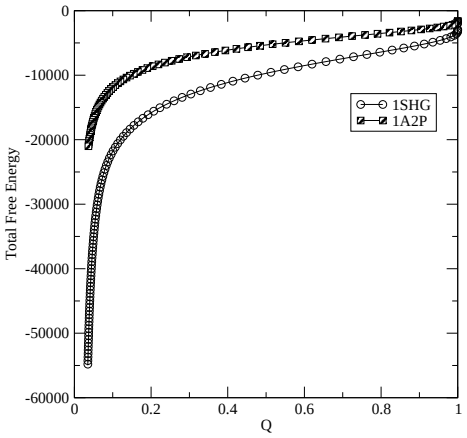

FIGURE 2. Total free energy (F^T) versus the fraction of native contacts (Q) for protein 1SHG and 1A2P.

matrix Δ_{ij}. The native state structure for each protein was taken from the Brookhaven Protein Data Bank (PDB) and was first analyzed in a separate subroutine where the types of amino-acid were identified and the distance between all pair residues was calculated and recorded to construct the contact matrix Δ_{ij}. We have employed a more refined discrete scheme of constructing the contact matrix - where depending on the type of amino-acid and distance between two residues, discrete leveled amino-acid interactions are taken into account - but only minor quantitative differences were observed between the two schemes. LU decomposition is employed to invert the symmetric matrix M_{ij}^{-1}(Press *et al.* 1992). The new p_{ij}'s are then calculated using eqn.(2.16) with the inverted matrix M_{ij} where the incomplete Gamma functions are evaluated using two different schemes, i.e. a fast converging series expansion and 50-point Gaussian quadrature (Press *et al.* 1992) to check numerical accuracy. The iteration continues until the differences between the previous and current values for *all* p_{ij}'s are below a prescribed tolerance δ. We used $\delta = 10^{-6}$ for equilibrium and thermal unfolding, and 10^{-5} for mechanical unfolding calculations and convergence was achieved within a dozen $(10 - 40)$ iterations.

3. Results

In this section, we apply the model to examine the equilibrium and non-equilibrium properties of several proteins. First, the equilibrium properties - both structural and thermodynamic - are presented. We determine the folding temperature in a systematic way as outlined in the previous section and calculate each free energy contributions arising from the Hamiltonian model. By varying the temperature, we further investigate the thermal denaturation of the immunoglobin (Ig) domains of Titin (1TIT). Finally the ends of the Ig domains are pulled mechanically and the unfolding mechanism as well as the stability of six key β-strands are studied.

3.1. *Equilibrium*

Following the steps in §2, we examine the equilibrium properties of the globular protein 2CI2, 1SHG and barnase (1A2P) and the immunoglobulin domain of Titin (1TIT). The remaining parameters we have to set in the Hamiltonian in eqn.(2.6) is the strength of the peptide strength K and K_s for the artificially added spring. Physically, a larger K

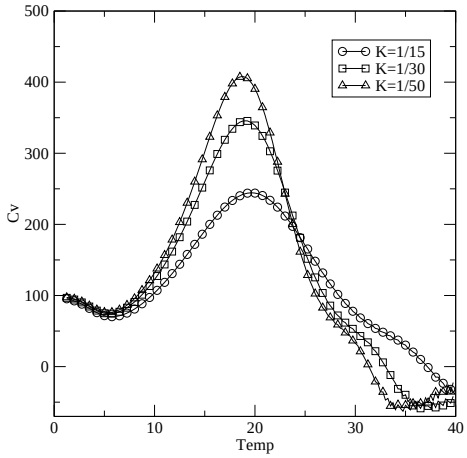

FIGURE 3. Heat capacity versus temperature for 2ci2 with varying peptide strength parameter K.

corresponds to a more rigid rod-like bond. The parameter K_s for the artificially added spring can be set to any non-zero value. The extra free energy due to this spring is F_A:

$$F^A = -T\ln(Z) = -T\ln\left[\frac{2\pi T}{K_s}\right]^{\frac{3}{2}}. \tag{3.1}$$

In all our calculations the total free energy F^T is calculated by subtracting the extra free energy F^A :

$$F^T = F^H - F^A, \tag{3.2}$$

where F^H is the corresponding total free energy to the Hamiltonian in eqn.(2.6) given by,

$$F^H = -\frac{3N}{2}\ln(2\pi)T - \frac{3}{2}\ln(\det M)T - \frac{R^2}{2}\sum_{i,j}\Delta_{ij}p_{ij} - T\ln\left[\frac{2\pi T}{K_s}\right]^{\frac{3}{2}}. \tag{3.3}$$

The free energy (F^H) and the individual energy terms in eqn.(3.3) for the protein 1A2P are shown in Fig.(1). Given our Hamiltonian, there are free energies associated with the consecutive pair interactions (the first term in eqn.(3.3)), and those arising from the non-consecutive pairwise interactions (the second and the third term) as well as those due to the artificial anchoring (the last term). In Fig.(2), the total free energy (F^T) is plotted versus the fraction of native contacts Q for the protein 1SHG and 1A2P, where Q is defined as :

$$Q = \frac{\Sigma'_{i,j}\Delta_{ij}p_{ij}}{\Sigma'_{i,j}\Delta_{ij}}. \tag{3.4}$$

In eqn.(3.4) the prime denotes that sum is carried out over only non-consecutive pairs. Previously, Micheletti *et al.* have used used a value of $K = \frac{1}{15}$ by inspecting the behavior of the fraction of native contacts (Q) as defined in eqn.(3.4)(Micheletti *et al.* 2001). In this work, three values of K were used - $\frac{1}{15}$, $\frac{1}{30}$ and $\frac{1}{50}$. In Fig.(3), we plot the heat capacity as a function of temperature with varying peptide strength parameter K for the protein 2CI2. The maximum of the heat capacity corresponds to the folding temperature which is defined as the temperature at which a protein is equally likely to be either in a

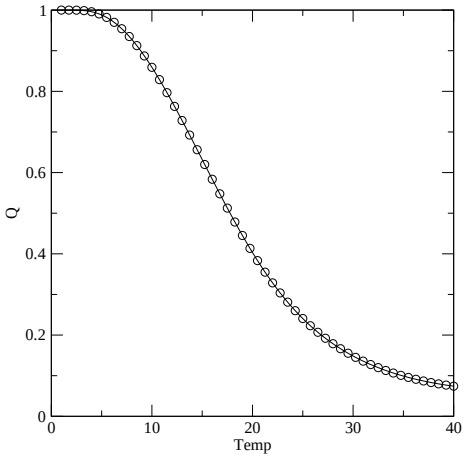

FIGURE 4. The fraction of native contacts (Q) versus temperature for the protein 1A2P.

folded or unfolded state. At the folding temperature the fraction of native contacts was shown to be around 0.5 in previous studies (Clementi *et al.* 2000; Micheletti *et al.* 2001; Plaxco *et al.* 1998; Chan & Dill 1998). To verify if the chosen range of the K parameter yielded consistent results with previous observations, we calculate the fraction of native contacts Q for 1A2P whose folding temperature was $T = 16.25$ for $K = \frac{1}{15}$. At the folding temperature Q is 0.56 as shown in Fig.(4). Note that the folding temperature is not a sensitive function of the peptide strength parameter K as the location of the maxima in heat capacity in Fig.(3) did not shift with different values of K. Next we calculate the equilibrium properties of the immunoglobulin domain of the protein Titin. The folding temperature was determined as above from the C_v-T plot and in Fig.(4) the fraction of native contacts Q is plotted versus temperature. Whereas Q is a global ordering parameter that measures the deviation of an entire folded protein domain from the native state, a more relevant local ordering parameter that monitors the deviation of each residue in the protein from its native state is P_i defined as :

$$P_i = \frac{\Sigma_j' \Delta_{ij} p_{ij}}{\Sigma_j' \Delta_{ij}} \tag{3.5}$$

The local ordering parameter P_i is shown at the folding temperature in Fig.(5). Note that the eight β strands are (in the order of first residue number - last residue number in the strand for each strand and in the parenthesis the type of amino acid) : 4(VAL)-7(PRO), 11(VAL)-15(VAL), 18(THR)-25(LEU), 32(GLY)-36(LEU), 47(CYS)-52(ASP), 55(LYS)-61(HIS), 69((GLY)-75(ALA) and 78(ALA)-88(GLU). We can clearly see that each strand exhibits different degrees of ordering at the folding temperature. In the next section, we examine the thermal denaturation of the same domain of Titin (1TIT).

3.2. *Thermal Unfolding*

The eight β-strands in the immunoglobulin domain of Titin are stabilized via hydrogen bonding. The six interacting residue pairs are : 6-24, 11-85, 19-60, 35-72, 48-59 and 69-84. The corresponding strand-strand or sheet interactions are denoted as : AB, A'G, BE, CF, DE and FG. First we vary the temperature to examine both the global (Q) and local (P_i) ordering. As shown in Fig.(6) as the temperature increases, the global ordering is slowly destroyed and at about twice the folding temperature ($T_f = 17.75$), the protein

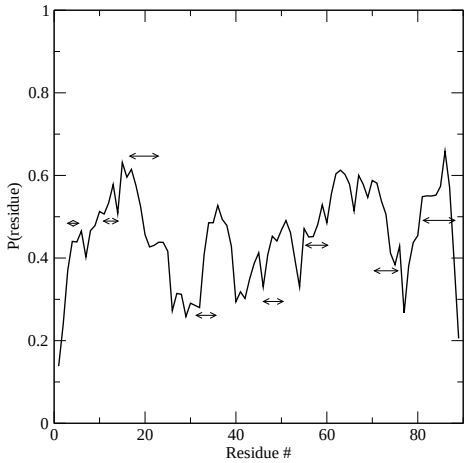

FIGURE 5. Local ordering parameter P_i for 1TIT at the folding temperature. The arrows denote the location of the eight β-strands.

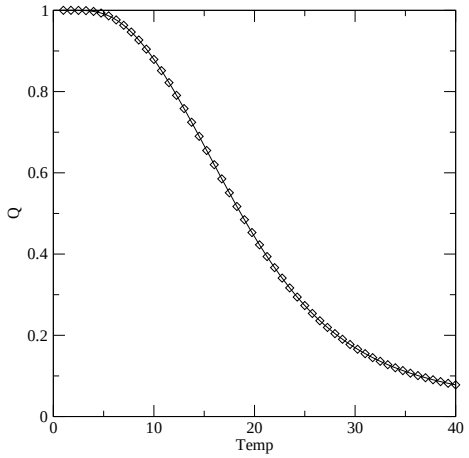

FIGURE 6. The fraction of native contacts (Q) versus temperature for the protein 1TIT.

has almost denatured. The individual residues which are in the ordered state at low temperatures (for example at $0.35T_f = 6.21$) slowly denature from the native state with varying degrees of ordering at high temperatures. Another global parameter to monitor the denaturation process is the size of the domain as a function of temperature which can be calculated by the following relation,

$$S^2(T) = 3G_{1N}(T) + S_o^2 \tag{3.6}$$

where S^2 is the mean square of the end-to-end distance of the protein and the subscript o denotes the native state, and G_{1N} is the component $(1, N)$ of the matrix G_{ij} as defined in eqn.(2.13). In Fig.(8), the extension versus temperature is shown and we see a smooth growth of the domain with increasing temperature. Note that at $2T_f = 35.5$, the domain size has increased only by 5% compared to $T_f = 17.75$. The six sheet interactions represented by the six individual pair interactions are shown in Fig.(9). The sheet interaction FG (69-84) is the most stable interaction over the entire temperature range and A'G is

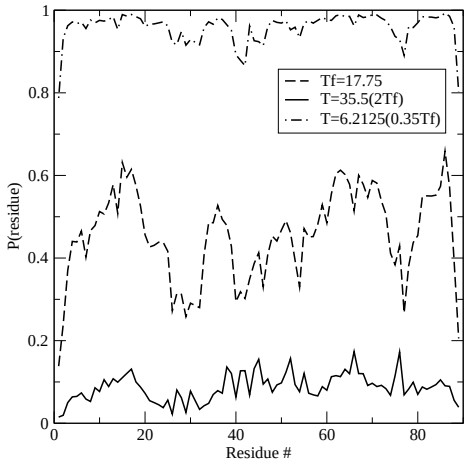

FIGURE 7. Local ordering parameter P_i for 1TIT with varying temperature near the folding temperature.

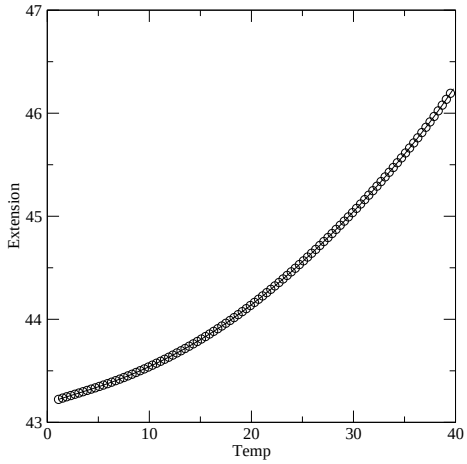

FIGURE 8. Domain size as a function of temperature for 1TIT.

the least stable interaction at low temperatures. Near the folding temperature all five interactions except for FG are comparable in strength.

3.3. *Mechanical Unfolding*

In this section we apply mechanical forces to the immunoglobulin domain of Titin. The experiments mentioned in §1 used two different tools to probe the unfolding pathways of the same molecule. Pulling the ends of it showed that the series of individual immunoglobulin domains open one by one. Also the protein domains were shown to resist a mechanical force of the order of a few hundreds of piconewtons (pN) after which one domain unfolds causing a reduction in the applied force. Further extension gradually increases the applied force until another domain unfolds. This stepwise single domain unfolding results in a so called saw-tooth pattern. The six sheet pair contact probabilities are shown in Fig.(10). The interaction A'G is the most unstable one below $F = 3.7$, which corresponds to a dimensional force of $150pN$. The interaction AB is lost at $F \approx 150pN$,

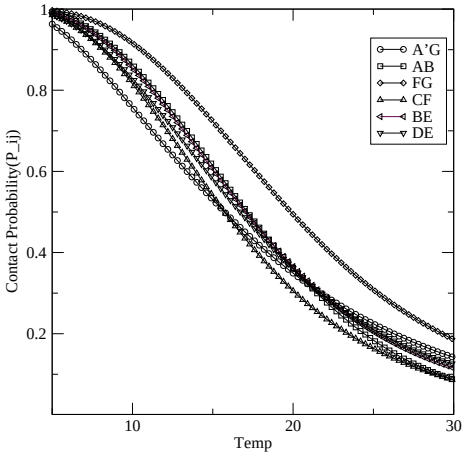

FIGURE 9. Six sheet contact probabilities as a function of temperature for 1TIT.

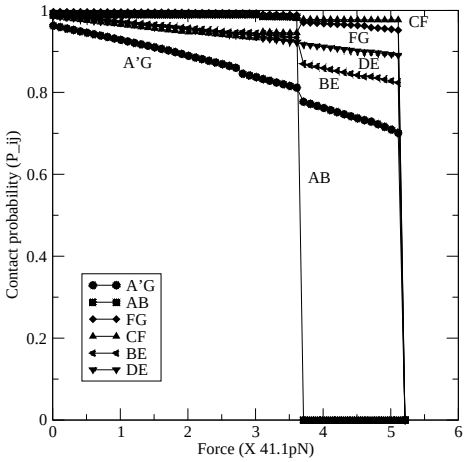

FIGURE 10. Six sheet contact probabilities as a function of pulling force for 1TIT.

suggesting the two β-strands A and B are no longer in contact. At $F = 5.16(212pN)$ the entire immunoglobulin domain is unfolded as seen from the vanishingly small contact pair probabilities for all six sheet interactions. This finding that the sheet interaction A'G and AB are the least stable ones is consistent with the results obtained from steered molecular dynamics by (Lu *et al.* 1998). The fraction of native contacts (Q) as larger forces are applied is shown in Fig. (11). Compared to Fig. (6), we see discrete jumps in the total nativeness of the protein domain.

To directly compare to previous experimental findings (Rief *et al.* 1997; Kellermayer *et al.* 1997; Tskhovrebova *et al.* 1997), we calculate the mean square end-to-end distance in the case of a constant pulling force $\vec{F}$ at the ends of a protein, which is given by,

$$S^2(T) = 3G_{1N}(T) + |\vec{S_o} + \vec{L_{1N}}|^2 \tag{3.7}$$

where $\vec{L}$ is defined in eqn.(2.14). In Fig.(12), the end-to-end distance is plotted as a function of the applied pulling force. We observe similar stalls in the plot where the domain size increases but the force remains constant. Related to the stability of each

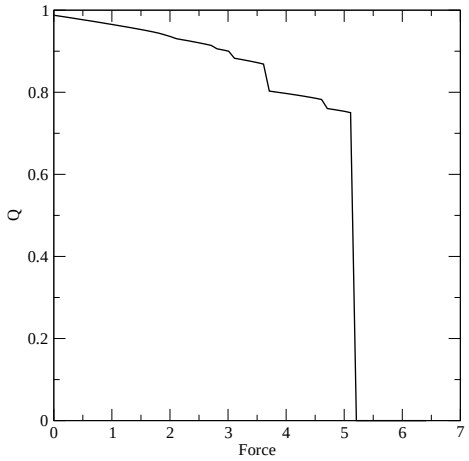

FIGURE 11. The fraction of native contacts (Q) versus pulling force for the protein 1TIT.

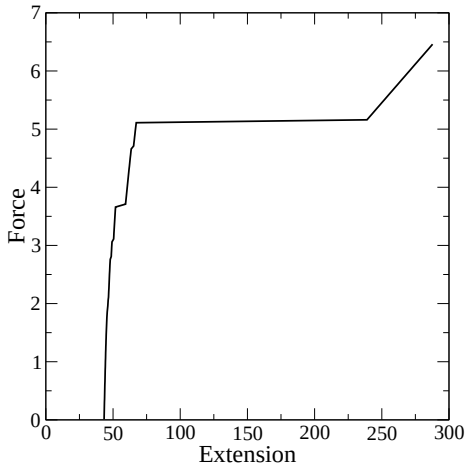

FIGURE 12. Force-extension plot for 1TIT.

sheet interaction, we conclude that the first force stall in Fig.(12) is due to the breakage of the hydrogen bonding pair in the AB sheet (residue pair 6-24). The AB sheet completely breaks at $F = 3.7$ and at $F = 5.16(210pN)$, the six sheet interactions are destabilized and the entire immunoglobulin domain is open. After the critical force of $F = 5.16$, the linear growth in extension isintrinsically due to the quadratic potential of the residues as given by the Hamiltonian (see eqn.(2.6)).

4. Conclusions

We have employed a Hamiltonian model based on a self-consistent Gaussian appoximation to examine the unfolding process of proteins in external - both mechanical and thermal - force fields. The motivation was to investigate the unfolding pathways of proteins by including only the essence of the important interactions of the native-state topology. Furthermore, if such a model can indeed correctly predict the physics of protein unfolding, it can complement more computationally expensive simulations and theoretical

work. The self-consistent Gaussian approximation by Micheletti et al. has been incorporated in our model to make the model mathematically tractable by significantly reducing the computational cost. All thermodynamic properties and pair contact probabilities are calculated by simply evaluating the values of a series of Incomplete Gamma functions in an iterative manner. We have compared our results to previous molecular dynamics simulation and experimental data for the mechanical unfolding of the giant muscle protein Titin (1TIT). Our model, especially in light of its simplicity and excellent agreement with experiment and simulation, demonstrates the basic physical elements necessary to capture the mechanism of protein unfolding in an external force field.

5. Acknowledgments

J. S. Hur would like to thank the Center for Turbulence Research at Stanford University for its support and Dr. Yves Dubief for helpful suggestions on the manuscript. E. Darve would like to thank the center for computational astrobiology at NASA Ames research center.

REFERENCES

Alm, E. & Baker, D. 1999 Matching theory and experiment in protein folding. *Curr. Opin. Struc. Biol.* **9**, 189–196.

Baker, D. 2000 A surprising simplicity to protein folding. *Nature* **405**, 39–42.

Chan, H. S. & Dill, K. A. 1998 Protein folding in the landscape perspective : Chevron plots and non-arrhenius kinetics. *Proteins* **30**, 2–33.

Clementi, C., Nymeyer, H. & Onuchic, J. N. 2000 Topological and energetic factors : What determines the structural details of the transition state ensemble and "enroute" intermediates for protein folding : An investigation for small globular proteins. *J. Mol. Biol.* **298**, 937–953.

Dayel, M. J., Hom, E. F. Y. & Verkman, A. S. 1999 Diffusion of green fluorescent protein in the aqueous-phase lumen of endoplasmic reticulum. *Biophys. J.* **76**, 2843–2851.

Kellermayer, M. S. Z., Smith, S. B., Granzier, H. L. & Bustamante, C. 1997 Folding-unfolding transitions in single titin molecules characterized with laser tweezers. *Science* **276**, 1112–1116.

Lu, H., Isralewitz, B., Krammer, A., Vogel, V. & Schulten, K. 1998 Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys. J.* **75**, 662–671.

Micheletti, C., Banavar, J. R. & Maritan, A. 2001 Conformations of proteins in equilibrium. *Phys. Rev. Lett.* **87**, 88102–88102.

Paci, E. & Karplus, M. 2000 Unfolding proteins by external forces and temperature : The importance of topology and energetics. *Proc. Natl. Acad. Sci.* **97**, 6521–6526.

Plaxco, K. W., Simons, K. T. & Baker, D. 1998 Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277**, 985–994.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. 1992 *Numerical Recipes in Fortan*, 2nd edn., , vol. 1. Cambridge: Cambridge Press.

Price, W. S., Tsuchiya, F. & Arata, Y. 1999 Lysozyme aggregation and solution

properties studied using pgse nmrdiffusion measurements. *J. Am. Chem, Soc.* **121**, 11503–11512.

Rief, M., Gautel, M., Oesterhelt, F., Fernandez, J. M. & Gaub, H. E. 1997 Reversible unfolding of individual titin immunoglobulin domains by afm. *Science* **276**, 1109–1112.

Tskhovrebova, L., Trinick, J., Sleep, J. A. & Simmons, R. M. 1997 Elasticity and unfolding of single molecules of the giant muscle protein titin. *Nature* **387**, 308–312.

Wolynes, P. G., Onuchic, J. N. & Thirumalai, D. 1995 Navigating the folding routes. *Science* **267**, 1619–1620.