10

Protein Folding, Unfolding and Aggregation Studied Using an All-Atom Model with a Simplified Interaction Potential

Anders Irbäck

Computational Biology and Biological Physics Group, Department of Theoretical Physics, Lund University, Sölvegatan 14A, S-223 62, Sweden
`anders@thep.lu.se`

Abstract. Finding a suitable transferable energy function for modeling of how different proteins fold into their respective native states is a major challenge in biophysics. Here, we discuss an all-atom protein model with implicit water and some studies based on this model. The model has a simplified and computationally convenient energy function. Despite its simplicity, the model has been found to quite successfully describe the structure and melting behavior of several peptides with about 20 amino acids. The same model, with unchanged parameters, has also been used to investigate the aggregation behavior of a fragment of Alzheimer's Aβ peptide and the mechanical properties of the 76-residue protein ubiquitin.

10.1 Introduction

How a protein interacts with other biological molecules, and therefore how it functions, depends crucially on its three-dimensional structure [1]. Knowledge of protein structures is therefore central to understanding biology. A protein structure is, on the other hand, not a completely static object, so having a single picture of the structure need not be sufficient. In fact, there are proteins that are unstructured or only partially structured and yet functional [2]. Bioinformatic analyses of genomes suggest that such proteins, which may adopt a specific structure upon binding to their biological targets, are more common than previously thought [3]. Another manifestation of the flexibility of protein chains is when misfolded proteins aggregate to form so-called amyloid fibrils [4]. Amyloid fibrils have been observed in many human disorders, including Alzheimer's disease and type II diabetes.

These examples clearly illustrate that being able to simulate protein dynamics on the computer is important in itself and not "only" as a means for structure prediction. Protein simulations are, however, notoriously difficult. The mere size of a protein molecule makes a quantum-mechanical calculation

unfeasible, so the models currently used are classical, and their level of detail varies. Studies at different levels of resolution serve complementary roles. Coarse-grained approaches, in the form of lattice or reduced off-lattice models, offer computational convenience and transparency and have been used to gain many useful insights into general aspects of protein folding (for a review, see [5]; see also the contribution by Bachmann and Janke in Chap. 8 of this volume). To study more detailed questions regarding specific proteins, it is often natural, and perhaps even necessary, to turn to atomistically detailed models. For all-atom simulations of proteins, there exists a number of widely used force fields, including AMBER [6], CHARMM [7], GROMOS [8] and OPLS-AA [9]. These force fields contain a large number of parameters, which typically are determined using experimental data or quantum-mechanical calculations on smaller molecules. The surrounding water molecules[1] are sometimes explicitly included in the calculations, but far from always, because it is computationally very expensive.

All-atom protein simulations are being used to study a wide range of problems, from ligand binding to unfolding and refolding studies. To fold an arbitrary amino acid sequence to its native state, from a random initial conformation, remains, however, a challenge. Such simulations are nontrivial also for small proteins [10], or peptides, for which the calculations are feasible from a purely computational perspective, which indicates that force-field development and refinement is a problem that needs continued attention.

This article discusses a simplified all-atom model for protein folding studies [11, 12, 13]. The model is based on a simple ansatz for the interaction potential. The determination of the potential is different compared to many other models because the potential is calibrated against data pertaining to folding properties of whole chains rather than properties of small groups of atoms. The calibration involves high-statistics folding simulations for a set of experimentally well-characterized peptides. The idea is to successively refine the potential by studying more and more sequences, which will impose new constraints on the model. It should be stressed that changes of the potential are accepted only if they are backward compatible, so that the model does not lose its ability to fold previously studied sequences.

The current version of this model is able to fold several sequences with about 20 amino acids. There are also examples of sequences of this size that the model fails to fold. One example is the tryptophan zipper β-hairpins [14], which make β-hairpins in the model but with the wrong topology. The set of sequences that the model can fold includes both α-helical and β-sheet peptides. Furthermore, for these sequences, the model provides a good description not only of the folded structure but also of the folded population and its temperature dependence [13]. Without changing any parameters, the model was also used to study the aggregation properties of the 7-amino acid fragment

---

[1] Throughout the article, it is assumed that the proteins are in aqueous solution. Membrane proteins are not considered.

Aβ_{16-22} of the amyloid-β peptide associated with Alzheimer's disease, with very promising results [15]. Recently, the mechanical [16] and the thermal [17] unfolding of ubiquitin, a protein with 76 amino acids, was investigated, again using exactly the same model. All the simulations of this model were performed using Monte Carlo methods.

The article is organized as follows. Section 10.2 gives a brief description of the model and the Monte Carlo methods used. Section 10.3 discusses the results obtained in the different studies mentioned above. A brief summary can be found in Sect. 10.4.

10.2 Model and Methods

The model studied contains all atoms of the protein chains, including hydrogen atoms, but no explicit water molecules. It assumes fixed bond lengths, bond angles and peptide torsion angles ($180°$), so that each amino acid has only the Ramachandran torsion angles ϕ, ψ and a number of side-chain torsion angles as its degrees of freedom.

The energy function of the model is intended to provide an effective description of the most relevant interactions. Many electrostatic interactions are ignored; they are assumed to be less important due to screening by the surrounding water. In the future, the energy function may be further developed with the inclusion of new terms such as Coulomb interactions between side-chain charges, but not before it becomes clear that they are needed.

A basic and somewhat unusual feature of the model is that its energy function is determined by studying folding properties of whole chains. For this approach to be computationally feasible, it is desirable to work with a simple potential with relatively few parameters to tune. At first glance, it may seem strange to combine such a potential with a detailed all-atom description of the protein chains; why spend time on a complicated all-atom description when the potential anyhow is simplified? It should therefore be stressed that including more atoms generally tends to make the potential simpler. To illustrate this point, consider the local interactions along the backbone, which are known to be important for the overall shape of the energy landscape [18]. It is well known that simple steric clashes explain much of the shape of the Ramachandran ϕ, ψ map [19]. In a model where some atoms are missing, the corresponding clashes must be accounted for through effective and generally more complicated interactions.

10.2.1 Energy Function

The energy function

$$E = E_{\text{loc}} + E_{\text{ev}} + E_{\text{hb}} + E_{\text{hp}} \qquad (10.1)$$

is composed of four terms. The term E_{loc} is local in sequence and represents an electrostatic interaction between adjacent peptide units along the chain. The other three terms are nonlocal in sequence. The excluded volume term E_{ev} is a $1/r^{12}$ repulsion between pairs of atoms.

The terms responsible for driving folding in this model are E_{hb} and E_{hp}, which represent hydrogen bonds and an effective hydrophobic attraction between nonpolar side chains, respectively. E_{hb} contains two kinds of hydrogen bonds, backbone–backbone bonds and bonds between charged side chains and the backbone. The functional form of the hydrogen bond potential is $u(r)v(\alpha, \beta)$, where the radial part is a 12,10-Lennard-Jones potential:

$$u(r) = 5 \left(\frac{\sigma}{r} \right)^{12} - 6 \left(\frac{\sigma}{r} \right)^{10}. \tag{10.2}$$

The angular dependence, $v(\alpha, \beta)$, has the form

$$v(\alpha, \beta) = \begin{cases} (\cos \alpha \cos \beta)^{1/2} & \text{if } \alpha, \beta > 90°, \\ 0 & \text{otherwise} \end{cases} \tag{10.3}$$

where $\alpha = \beta = 180°$ corresponds to a perfectly aligned hydrogen bond.

The effective hydrophobic attraction, E_{hp}, has the pairwise additive form

$$E_{\text{hp}} = -\sum_{I<J} M_{IJ} C_{IJ}, \tag{10.4}$$

where C_{IJ} is a measure of the degree of contact between side chains I and J, and M_{IJ} sets the energy that a pair in full contact gets. Specifically, C_{IJ} is defined as

$$C_{IJ} = \frac{1}{N_I + N_J} \left[\sum_{i \in A_I} f(\min_{j \in A_J} r_{ij}^2) + \sum_{j \in A_J} f(\min_{i \in A_I} r_{ij}^2) \right], \tag{10.5}$$

where A_I and A_J are predetermined sets of nonpolar side-chain atoms and $f(x)$ is a piecewise linear function which is 1 if $x < (3.5 \text{ Å})^2$ and 0 if $x > (4.5 \text{ Å})^2$. Roughly speaking, C_{IJ} is the fraction of atoms in A_I or A_J that are in contact with some atom from the other side chain.

The energy parameters of the model, like the M_{IJ}s in (10.4), are internally in the simulations numbers on a dimensionless scale. The correspondence (a constant factor) of this scale to the physical energy scale is not obvious due to the effective character of the interaction terms. It can be determined by using the model prediction of the dimensionless energy value for an observable and the experimental value for the same. We chose to use the melting temperature of the Trp cage for this calibration, so the calculated melting temperature of this peptide was set to its experimental value, $T_{\text{m}} = 315$ K [20]. It must be emphasized that the energy scale was left unchanged when studying the other peptides.

For simplicity, this description of the potential was for a single protein chain. The model can be easily extended to multichain systems. Different chains then interact by excluded-volume interactions, hydrogen bonding and hydrophobic attraction between nonpolar side chains. The interchain interactions have the same form as the corresponding intrachain interactions.

10.2.2 Monte Carlo Methods

The thermodynamics of this model can be simulated by using simple single-variable Monte Carlo updates. The torsion angles are then turned one by one, and each change is followed by a Metropolis accept/reject question. This update can be applied to both backbone and side-chain degrees of freedom. However, when applied to a backbone angle, this method tends to lead to a large nonlocal deformation of the chain, which is likely to be rejected if the chain is compact.

Therefore, it is desirable to also include a backbone update less drastic than the single-angle update. One possibility is to use a strictly local method such as the concerted-rotation algorithm [21]. For long chains, such an update has the advantage of leaving a large part of the energy function unchanged. However, strictly local methods tend to be quite complex, and the proteins amenable to atomic-level simulations are in any case not very large. Instead, we have therefore chosen to use biased Gaussian steps [22], a semi-local method that is flexible and easy to implement. This method simultaneously turns n adjacent torsion angles τ_i along the backbone, where n typically is 7 or 8. Let $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_n)$. A tentative new set of angles, $\boldsymbol{\tau}'$, is generated with a bias toward local deformations. Specifically, using a conformation-dependent $n \times n$ matrix G such that $(\boldsymbol{\tau}' - \boldsymbol{\tau})^{\mathrm{T}} \mathsf{G}(\boldsymbol{\tau}' - \boldsymbol{\tau}) \approx 0$ for changes corresponding to local deformations, $\boldsymbol{\tau}'$ is drawn from the Gaussian distribution

$$W(\boldsymbol{\tau} \to \boldsymbol{\tau}') = \frac{(\det \mathsf{A})^{1/2}}{\pi^3} \exp\left[-(\boldsymbol{\tau}' - \boldsymbol{\tau})^{\mathrm{T}} \mathsf{A}(\boldsymbol{\tau}' - \boldsymbol{\tau})\right], \qquad (10.6)$$

$$\mathsf{A} = \frac{a}{2}(\mathbf{1} + b\mathsf{G}), \qquad (10.7)$$

where $\mathbf{1}$ denotes the $n \times n$ unit matrix and a and b are tunable parameters. The parameter a controls the acceptance rate, whereas b sets the degree of bias toward local deformations. Typical values are $a = 300\,(\mathrm{rad})^{-2}$ and $b = 10\,(\mathrm{rad}/\text{Å})^{-2}$. Finally, the new conformation $\boldsymbol{\tau}'$ is subject to an accept/reject step, with probability

$$P_{\mathrm{acc}}(\boldsymbol{\tau} \to \boldsymbol{\tau}') = \min\left(1, \frac{W(\boldsymbol{\tau}' \to \boldsymbol{\tau})}{W(\boldsymbol{\tau} \to \boldsymbol{\tau}')} \exp[-(E' - E)/kT]\right) \qquad (10.8)$$

for acceptance (k is Boltzmann's constant and T is the temperature). The factor $W(\boldsymbol{\tau}' \to \boldsymbol{\tau})/W(\boldsymbol{\tau} \to \boldsymbol{\tau}')$ is needed for detailed balance to be fulfilled, since G is conformation-dependent and W thereby asymmetric.

These two updates, the single-variable update and the biased Gaussian step, are updates for simulations in the canonical ensemble. In recent years, there has been an increasing interest in generalized-ensemble techniques for protein folding simulations [23]. The aim of these methods is to make it easier for the system to jump between different free-energy minima and thereby speed up the exploration of conformational space. A method of this type that has been used for the present model is simulated tempering [24, 25, 26]. In this method, the temperature is a dynamical variable that can take on a set of predetermined values, T_k. The extended ensemble studied by this method is defined by the partition function

$$Z = \sum_k \mathrm{e}^{-g_k} Z(T_k)\,, \tag{10.9}$$

where the g_ks are tunable parameters and $Z(T_k)$ denotes the canonical partition function at temperature T_k. The temperature jumps are controlled by a Metropolis accept/reject question.

10.2.3 Software

The program package PROFASI (PROtein Folding and Aggregation SImulator) implements the model and Monte Carlo methods described above [27] and is freely available to academic users. This package, written in C++, is designed to handle both single chains and multichain systems in a flexible and efficient manner. Using this package, the folding studies below require roughly a day on a cluster of 10 2.2 GHz processors for each peptide.

10.3 Results

This section presents results obtained using the model described above. The number of sequences that this model can fold is limited. On the other hand, it should be stressed that all the different sequences are studied for one and the same choice of model parameters. In other words, the only input to the calculations is the amino acid sequence. All our calculations, except in the unfolding studies of ubiquitin, were started from random initial configurations.

10.3.1 Peptide Folding

Folding studies based on this model have been performed [13] for the following set of peptides: the α-helical Trp cage [20] and F_s [28], the β-hairpins GB1p [29] and GB1m2/GB1m3 [30], and the three-stranded β-sheets Betanova [31] and LLM [32]. These peptides are schematically illustrated in Fig. 10.1.
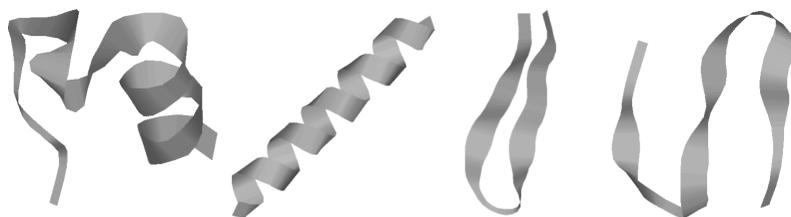
Fig. 10.1. Schematic illustration of the different geometries of the peptides studied. Shown from *left* to *right* are the Trp cage, F$_s$, a β-hairpin (GB1p, GB1m2 and GB1m3) and a three-stranded β-sheet (Betanova and LLM). Drawn with RasMol [33]

A common measure of similarity between protein conformations is the root-mean-square deviation (RMSD), minimized over rigid-body translations and rotations. Below, Δ_b and Δ denote RMSD values calculated over backbone atoms and all nonhydrogen atoms, respectively.

Trp Cage

The 20-amino acid Trp cage is a "miniprotein" with a compact folded state and a melting temperature of 315 K [20]. Its NMR-derived native structure [20] contains an α-helix (amino acids 2–8), a single turn of 3_{10}-helix (amino acids 11–14) and a hydrophobic core consisting of three prolines (Pro12, Pro18, Pro19) and two aromatic amino acids (Tyr3, Trp6). The Trp cage has been studied using several different models [34, 35, 36, 37, 38]. Many groups obtained structures similar to the experimental structure, but it was not shown that the melting behavior is well described by any of these models. Figures 10.2 and 10.3 show results obtained using the model of this article. The first of these figures shows the free energy $F(\Delta_b, E)$ calculated as a function of backbone RMSD from the experimental structure, Δ_b, and energy, E, at 275 K. $F(\Delta_b, E)$ has a simple shape with one dominating minimum located at $\Delta_b \approx 2.3$ Å, showing that this peptide, in the model, adopts a structure similar to the experimental structure. Figure 10.3 illustrates the melting behavior. The melting temperature of the Trp cage was, as mentioned earlier, used to set the energy scale of the model and is therefore by construction equal to its experimental value. The rate at which the native structure melts with temperature is, by contrast, a prediction of the model. Figure 10.3(a) shows the temperature dependence of the helix content. A simple two-state fit provides an excellent description of the data. Figure 10.3(b) shows the native population obtained from this fit along with experimental data. The calculated native population agrees very well with experimental data over the entire temperature range, although the agreement is not prefect at low temperature. The model thus gives a realistic melting behavior.
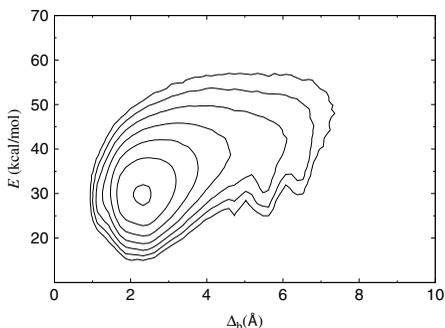
Fig. 10.2. Free energy $F(\Delta_b, E)$ calculated as a function of backbone RMSD (amino acids 2–19), Δ_b, and energy, E, for the Trp cage at 275 K. The contours are spaced at intervals of $1kT$. Contours more than $6kT$ above the minimum free energy are not shown. The free energy $F(\Delta_b, E)$ is defined by $\exp[-F(\Delta_b, E)/kT] \propto P(\Delta_b, E)$, where $P(\Delta_b, E)$ denotes the joint probability distribution of Δ_b and E at temperature T
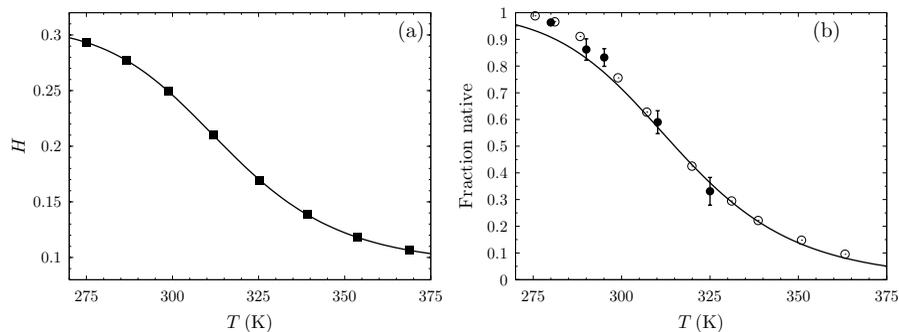

Fig. 10.3. Melting behavior of the Trp cage. (**a**) Helix content H against temperature. The line is a fit to the two-state expression $H(T) = (H_u + H_n K(T))/(1 + K(T))$, where H_u and H_n are the values in the unfolded and native states, respectively, and $K(T)$ is the effective equilibrium constant. $K(T)$ is given by $K(T) = \exp[(1/kT - 1/kT_m)\Delta E]$, where T_m is the melting temperature and ΔE is the energy difference between the unfolded and native states. The helix content H is, for an N-amino acid chain, defined as the fraction of the $N - 2$ inner amino acids with their Ramachandran (ϕ, ψ) pair in the α-helix region $-90° < \phi < -30°$, $-77° < \psi < -17°$. (**b**) Native population against temperature. The line represents the model and was obtained from the fit of data for H. Plot symbols show experimental results [20] based on CD ($\circ$) and NMR ($\bullet$), respectively

F$_\mathrm{s}$

The F$_\mathrm{s}$ peptide is a designed alanine-based peptide with 21 amino acids, which makes an α-helix [28]. It has been extensively studied both experimentally [28, 39, 40] and theoretically [41, 42, 43]. Its melting behavior was investigated experimentally using circular dichroism (CD) as well as infrared (IR) spectroscopy, with slightly different results. The melting temperature measured by IR was 334 K [39], whereas two CD-based studies obtained $T_\mathrm{m} = 308$ K [28] and $T_\mathrm{m} = 303$ K [40].

Figure 10.4(a) shows a $F(\Delta_\mathrm{b}, E)$ plot for F$_\mathrm{s}$, analogous to that in Fig. 10.2 for the Trp cage. In the calculations for F$_\mathrm{s}$, Δ_b denotes backbone RMSD from an ideal α-helix. The free energy has its global minimum at $\Delta_\mathrm{b} \approx 0.5$ Å, which indeed corresponds to an α-helix. In addition, there are two local minima with $\Delta_\mathrm{b} \approx 7$ Å and $\Delta_\mathrm{b} \approx 11$ Å, respectively, both of which correspond to β-sheet structures. These two local minima are very weakly populated compared to the α-helix minimum.

Figure 10.4(b) shows the temperature dependence of the helix content. A two-state fit of the data gave a melting temperature of $T_\mathrm{m} = 304 \pm 1$ K and an energy difference of $\Delta E = 11.9 \pm 0.3$ kcal/mol. This melting temperature is lower than what was found by IR, but in good agreement with the CD-based results. The result for ΔE is furthermore in perfect agreement with the CD-based result of Thompson et al. [40], namely $\Delta E = 12 \pm 2$ kcal/mol. The helix-coil transition of F$_\mathrm{s}$ is thus very similar in both location and sharpness to what was found in this experimental study.

GB1p, GB1m2 and GB1m3

Let us now turn from α-helical structure to β-sheet structure. The GB1p, GB1m2 and GB1m3 peptides make β-hairpins, which is the minimal unit of β-sheet structure. A β-hairpin has two strands which are connected by hydrogen
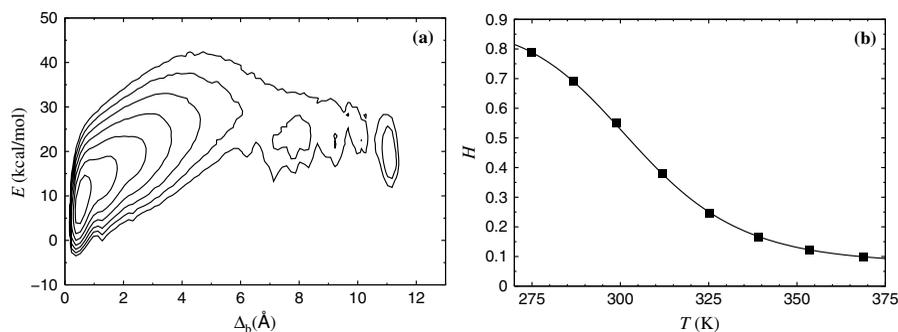

Fig. 10.4. The F$_\mathrm{s}$ peptide. (**a**) Free energy $F(\Delta_\mathrm{b}, E)$ (see Fig. 10.2). (**b**) Helix content H versus temperature (see Fig. 10.3(a))

bonds. GB1p is a fragment from the protein G B1 domain (amino acids 41–56) and is known to make a β-hairpin on its own [29]. Many groups have performed simulations of this particular peptide [44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54]. GB1m2 and GB1m3 are two sequences recently derived from GB1p [30]. By CD and NMR, it was found that the folded populations of GB1p, GB1m2 and GB1m3 at 298 K are ~30, 74 ± 5 and $86 \pm 3\%$, respectively [30]. The stability of GB1p was also studied using Trp fluorescence [55]. A two-state analysis of these data gave $T_m = 297$ K and $\Delta E = 11.6$ kcal/mol [55]. These values correspond to a native population higher than that found by CD and NMR.

Figure 10.5 shows the free energy $F(\Delta, E)$ for GB1p and GB1m3 at 275 K, as obtained using the model of this article. Here, an RMSD Δ calculated over all heavy atoms is needed because the backbone RMSD Δ_b cannot distinguish between the two possible β-hairpin topologies (with similar backbone folds but oppositely oriented side chains). A complete three-dimensional structure has not been determined for any of these peptides. Reference structures for the RMSD calculation were therefore derived from the NMR structure for the full protein G B1 domain [56]. For GB1p, the free energy $F(\Delta, E)$ has two low-lying minima, both of which are similar to the reference structure ($\Delta \approx 2.0$ Å and $\Delta \approx 3.2$ Å, respectively). Both these minima correspond to a β-hairpin with native hydrogen bonds. The difference between the two minima lies in the shape of the turn region. In addition to these minima, there are two weakly populated local minima at $\Delta \approx 5.3$ Å and $\Delta \approx 8$–10 Å, which correspond to a β-hairpin with the opposite topology (flipped side chains) and α-helix, respectively. The free-energy plot for GB1m3 (see Fig. 10.5(b)) has a simpler shape with only one detectable minimum. This sequence is more stable than the original GB1p sequence, as it should. The results for GB1m2 (not shown) are very similar to those for GB1m3.

Different experiments on GB1p have, as mentioned above, obtained different results for the native population. To see whether the model can explain this discrepancy, two different analyses were performed. First, the temperature
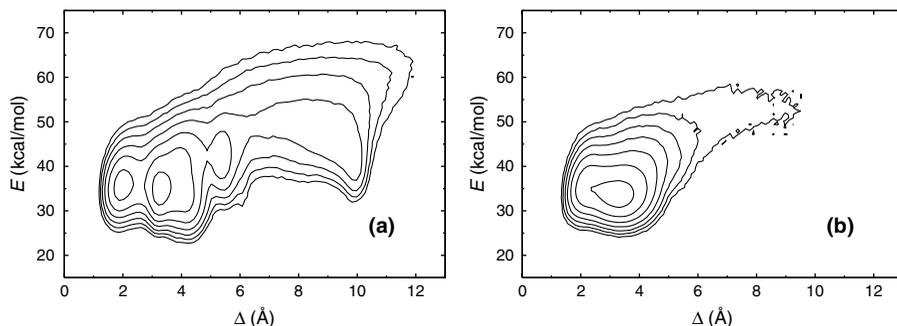

Fig. 10.5. Free energy $F(\Delta, E)$ at 275 K for (**a**) GB1p and (**b**) GB1m3. Contour levels are as in Fig. 10.2

dependence of the hydrophobicity energy E_{hp} was studied. This quantity should be strongly correlated with Trp fluorescence because the Trp amino acid (Trp43) is part of a hydrophobic cluster in the folded structure. A two-state fit of the E_{hp} data (see Fig. 10.6(a)) gave $T_m = 297 \pm 1$ K and $\Delta E = 14.2 \pm 0.2$ kcal/mol, which indeed is in good agreement with the Trp fluorescence results. Figure 10.6a also shows the corresponding results for GB1m3. As expected, the melting curve is shifted toward higher temperature for GB1m3. A second and independent estimate of the native population can be obtained by studying native hydrogen bonds. The number of native hydrogen bonds should be strongly correlated with the native population as obtained by CD and NMR. For a given conformation, let N_{hb}^{nat} denote the number of native hydrogen bonds that are present (a hydrogen bond is considered formed if its energy is lower than a cutoff). Figure 10.6(b) shows the probability distribution of N_{hb}^{nat} for GB1p and GB1m3 at 299 K. The distribution has a clear bimodal shape for both peptides, with one native and one unfolded peak. The folded population can be directly estimated from the area of the native peak. With $N_{hb}^{nat} \geq 3$ as criterion for native, one finds that GB1p and GB1m3 have folded populations of 27 ± 2 and 82 ± 1 %, respectively. These results are in good agreement with the above-mentioned CD- and NMR-based results at 298 K [30] (~ 30 % for GB1p; 86 ± 3 % for GB1m3).

The discrepancy between the experimental results for the native population of GB1p is thus very similar to what one finds in the model when analyzing different observables. This finding suggests that the difference between the experimental results reflects a true dependence of the apparent folded population on the observable studied. Such a dependence does not occur for ideal two-state systems, but a simple two-state picture might very well be inadequate for a small peptide such as GB1p. For GB1m3, there is no Trp
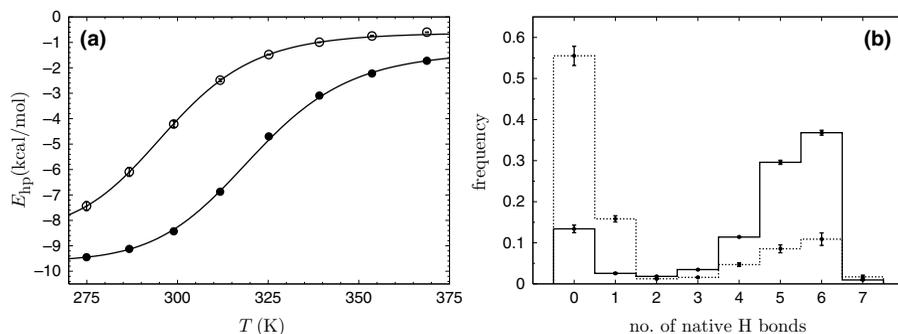

Fig. 10.6. (a) The hydrophobicity energy E_{hp} against temperature for GB1p (○) and GB1m3 (●). The lines are two-state fits ($T_m = 297 \pm 1$ K, $\Delta E = 14.2 \pm 0.2$ kcal/mol for GB1p; $T_m = 321 \pm 1$ K, $\Delta E = 15.0 \pm 0.4$ kcal/mol for GB1m3). (b) Probability distribution of the number of native hydrogen bonds, N_{hb}^{nat}, for GB1p (*dotted line*) and GB1m3 (*full line*) at 299 K

fluorescence study available. The simulation results suggest that the difference between results based on Trp fluorescence and results based on CD and NMR should be smaller for this peptide. Whether or not this prediction is correct remains to be seen.

Betanova and LLM

Betanova is a designed three-stranded β-sheet peptide with 20 amino acids [31]. It is only marginally stable, but Betanova mutants with higher stability have been developed, such as the triple mutant LLM (Val5Leu, Asn12Leu, Thr17Met) [32]. Betanova has been studied by both NMR [32] and computer simulations [57, 58, 59]. The NMR-based native populations of Betanova and LLM are 9 and 36%, respectively, at 283 K.

It turns out that Betanova and LLM make three-stranded β-sheets similar to the experimental structures in the model of this article. Both peptides have low stability, as they should, as can be seen from Fig. 10.7. This figure shows the probability distribution of N_{hb}^{nat} for these peptides at 287 K, N_{hb}^{nat} being the number of native hydrogen bonds. In both cases, the distribution exhibits three peaks: a native peak at high N_{hb}^{nat}, an unfolded peak at low N_{hb}^{nat} and a third peak at $N_{hb}^{nat} \approx 4$. It was found that conformations in the intermediate state with $N_{hb}^{nat} \approx 4$ tend to contain the first (N-terminal) β-hairpin but not the second (C-terminal) one, which is in agreement with experimental data [32]. With $N_{hb}^{nat} \geq 6$ as the definition of native, one finds that Betanova and LLM are 6 ± 1 and $38 \pm 2\%$ folded, respectively, at this temperature. These values are in good agreement with the NMR-based native populations.

Table 10.1 gives a summary of calculated and experimental native populations for the five β-sheet peptides studied. The model results are based on the number of native hydrogen bonds, whereas the experimental results are based on CD and NMR. The overall agreement is very good, and it should be
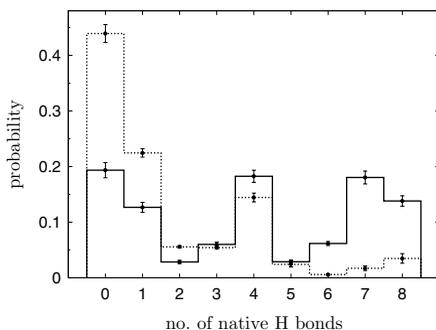

Fig. 10.7. Probability distribution of the number of native backbone hydrogen bonds, N_{hb}^{nat}, for Betanova (*dotted line*) and LLM (*full line*) at 287 K

Table 10.1. Experimental and calculated native populations of the β-sheet peptides studied

	exp.	model
GB1p	∼30 % (298 K)	27 ± 2 % (299 K)
GB1m2	74 ± 5 % (298 K)	84 ± 1 % (299 K)
GB1m3	86 ± 3 % (298 K)	82 ± 1 % (299 K)
Betanova	9 % (283 K)	6 ± 1 % (287 K)
LLM	36% (283 K)	38 ± 2 % (287 K)

stressed that this agreement was achieved without having to tune any model parameters. Good agreement with experimental stability data was also seen for the α-helical Trp cage and F_s peptides. This quantitative agreement indicates that factors such as Coulomb interactions between charged amino acids play a quite limited role in the folding thermodynamics of these peptides, compared to hydrogen bonding and hydrophobic attraction, which are the main driving forces of this model.

10.3.2 Aggregation of $A\beta_{16-22}$ Peptides

The fibrillar aggregates that characterize amyloid diseases are formed by specific proteins, but many other proteins are capable of forming similar aggregated structures [4, 60].

Among the known fibril-forming proteins are several fragments of the amyloid Aβ peptide associated with Alzheimer's disease. One such fragment is the 7-amino acid $A\beta_{16-22}$ peptide [61]. This peptide contains the hydrophobic 16–20-amino acid segment (Lys-Leu-Val-Phe-Phe), which is known to be important in the Aβ–Aβ interaction [62]. Its small size makes $A\beta_{16-22}$ well suited as a model system for studying the mechanisms of fibril formation. Computer simulations of $A\beta_{16-22}$ systems have been reported by several groups [63, 64, 65, 66].

An amyloid fibril is composed of β-sheets whose strands run perpendicular to the fibril axis [67]. For $A\beta_{16-22}$ fibrils, there is evidence from solid-state NMR that the β-strands have an antiparallel organization [61, 68].

The exact role of protein aggregation in amyloid diseases remains to be clarified. However, there is increasing evidence suggesting that the full fibrils are not the neurotoxic agent, whereas there are studies linking Aβ oligomers to neurotoxicity [69, 70, 71].

Using the model of this article, a study of small $A\beta_{16-22}$ systems was performed [15]. The energy function was exactly the same as in the folding studies above. Systems of one, three and six $A\beta_{16-22}$ peptides were studied by Monte Carlo simulations, which were started from random (nonaggregated) structures. The $N_c = 3$ and $N_c = 6$ systems, N_c being the number of chains, were contained in periodic boxes. The box sizes were $(35 \text{ Å})^3$ and $(44 \text{ Å})^3$ for

$N_c = 3$ and $N_c = 6$, corresponding to a constant peptide concentration. To the set of Monte Carlo moves, rigid-body translations and rotations of whole chains were added.

Secondary Structure

Figure 10.8 shows the α-helix and β-strand contents H and S against temperature for different N_c. For $N_c = 1$, both H and S are small at all temperatures studied. So, in this model, the $A\beta_{16-22}$ monomer is mainly a random coil throughout this temperature range. The $N_c = 3$ and $N_c = 6$ systems show a qualitatively different behavior; S increases sharply with decreasing temperature, to values of $S = 0.6$ and higher, whereas H is small. These results show that unless the temperature is too high, the three- and six-chain systems self-assemble into ordered structures with a high β-strand content.

The results for $N_c = 1$ and $N_c = 3$ can be compared with results from molecular dynamics simulations with explicit water by Klimov and Thirumalai [64]. The conclusion that the isolated $A\beta_{16-22}$ peptide is mainly a random coil is in agreement with their results. However, the $N_c = 3$ results disagree with theirs because they found a lower β-strand content for $N_c = 3$ than for $N_c = 1$.
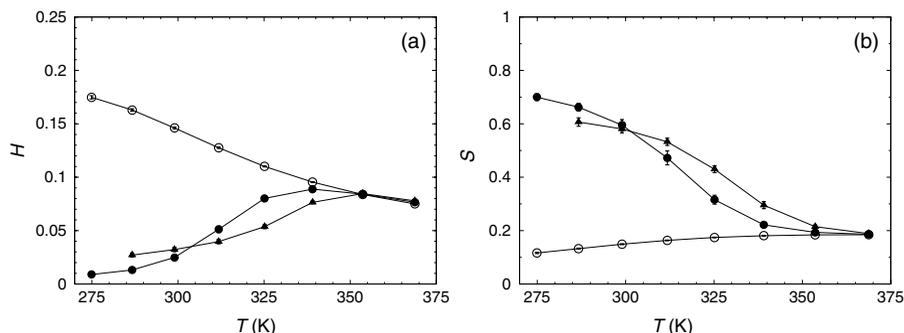

Fig. 10.8. Secondary structure of $A\beta_{16-22}$ systems. For a chain with N amino acids, the α-helix and β-strand contents are defined as the fractions of the $N-2$ inner amino acids with their (ϕ, ψ) pair in the α-helix and β-strand regions of the Ramachandran space, assuming that α-helix corresponds to $-90° < \phi < -30°$, $-77° < \psi < -17°$ and that β-strand corresponds to $-150° < \phi < -90°$, $90° < \psi < 150°$. The average α-helix and β-strand contents, over all the chains of the system, are denoted by H and S, respectively. (**a**) The α-helix content H against temperature T for $N_c = 1$ (○), $N_c = 3$ (●) and $N_c = 6$ (▲). Lines joining data points are only a guide for the eye. (**b**) Same for the β-strand content S

β-Strand Organization

As mentioned above, there is experimental evidence [61, 68] that the β-strands in full Aβ$_{16-22}$ fibrils have an antiparallel organization. To examine the β-strand organization in the model, parallel and antiparallel pairs of strands were identified and counted. Given a multichain configuration, all pairs of chains were identified such that (i) their interchain hydrogen bond energy was less than a cutoff (roughly corresponding to 2–3 hydrogen bonds), and (ii) both chains had a β-strand content higher than 0.5. For each such pair of chains, the scalar product of their normalized end-to-end unit vectors was calculated. If this scalar product was greater than 0.7 (less than −0.7), the two chains were taken as parallel (antiparallel). The numbers of such parallel and antiparallel pairs are denoted by n_+ and n_-, respectively.

Table 10.2 shows the joint probability distribution $P(n_+, n_-)$ for the $N_c = 6$ system at $T = 287$ K. From this table, it can be seen that many configurations contain mixed β-sheet structure. Nevertheless, there is a clear asymmetry between n_+ and n_-; large n_- values are much more probable than large n_+ values. The combination $(n_+, n_-) = (4, 0)$ is, e.g., very unlikely to occur, whereas $(n_+, n_-) = (0, 4)$ does occur with a significant frequency.

The results in Table 10.2 were obtained at the lowest temperature studied. With increasing temperature, the average n_+ and n_- steadily decrease. At the highest temperature studied, 369 K, about 99% of the conformations have $n_+ = n_- = 0$.

A clear trend seen in Table 10.2 is that large n_+ values are suppressed, which means that large parallel β-sheets are unlikely to form. The probability of having large antiparallel β-sheets is higher. Compared to purely antiparallel β-sheet structures, it is possible that mixed β-sheet structures are more difficult to extend to large stable structures. To be able to check whether or not this is the case, simulations of larger systems are required.

Why are antiparallel β-sheets favored over parallel ones? Klimov and Thirumalai [64] concluded that Aβ$_{16-22}$ peptides make antiparallel β-sheets because of Coulomb interactions between charged side chains; the two end

Table 10.2. The probability distribution $P(n_+, n_-)$ for $N_c = 6$ Aβ$_{16-22}$ peptides at $T = 287$ K (entries smaller than 10^{-3} are omitted)

n_+	n_-				
	0	1	2	3	4
0	0.028 (5)	0.059 (11)	0.08 (2)	0.06 (2)	0.030 (15)
1	0.038 (6)	0.12 (2)	0.16 (3)	0.10 (3)	0.006 (3)
2	0.026 (11)	0.11 (5)	0.14 (5)	0.004 (2)	
3	0.008 (5)	0.013 (9)	0.015 (12)		

n_+ and n_- are the numbers of parallel and antiparallel pairs of β-strands (see text). The numbers in parentheses are the statistical errors in the last digits

side chains of the $A\beta_{16-22}$ peptide carry opposite charges, which indeed should make the antiparallel orientation electrostatically favorable. However, our model completely ignores Coulomb interactions between side-chain charges and still strongly favors the antiparallel organization. Other mechanisms than Coulomb interactions between side-chain charges might therefore play a significant role, such as the geometry of backbone–backbone hydrogen bonds, steric effects, and the precise distribution of hydrophobicity along the chains.

Examples of Low-Energy Structures

What is the typical shape of the aggregated structures? In the $N_c = 6$ simulations, no single dominating shape was observed, but rather a number of more or less degenerate local free-energy minima. Figure 10.9 shows two snapshots of such minima.

In the simplest class of typical structures observed in the simulations, five of the chains form a relatively flat β-sheet, whereas the remaining chain is a random coil and held in contact with the β-sheet by hydrophobic attraction. Six-stranded β-sheets also occurred in the simulations, but with a very low frequency. Furthermore, for the six-chain system, there were new nontrivial structures with no analogs in the three-chain simulations. The second structure in Fig. 10.9 illustrates this. Here, stability is achieved by stacking two different, three-stranded, β-sheets together, which brings hydrophobic side chains from the two β-sheets in close contact. Such "sandwiches" occurred with a nonnegligible frequency in the simulations, although not as frequently as five-stranded β-sheets.

10.3.3 Mechanical Unfolding of Ubiquitin

The same model has also been used to study the mechanical unfolding of ubiquitin [16, 17], a protein with 76 amino acids. The unfolding behavior of
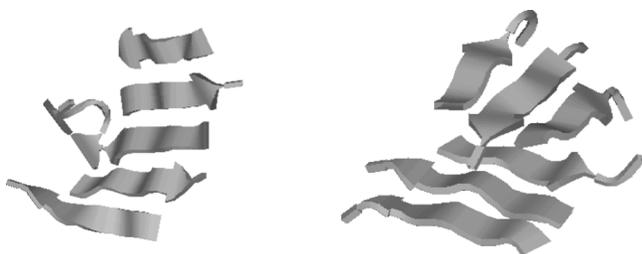

Fig. 10.9. Two typical low-energy structures from the simulations of six $A\beta_{16-22}$ peptides: a five-stranded β-sheet (*left*), and two three-stranded β-sheets "sandwiching" several of their hydrophobic side chains between them (*right*). Drawn with RasMol [33]

ubiquitin under the influence of a stretching force was recently investigated experimentally by single-molecule constant-force methods [72]. Many observed unfolding traces had a simple two-state character, whereas others showed clear evidence of intermediate states. The quantity monitored was the end-to-end distance. This quantity alone provides only limited information about the unfolding behavior and the structure of the intermediate states.

More detailed experimental information is available about the pathway followed by this protein in temperature-induced unfolding. In its native form, ubiquitin contains an α-helix packed against a five-stranded β-sheet (see Fig. 10.10). The thermally most stable parts of the native structure seem to be the α-helix and the first (N-terminal) β-hairpin [74, 75], which is part of the β-sheet. There is also evidence that the same two secondary-structure elements, the α-helix and the first β-hairpin, form early as ubiquitin folds to its native conformation. An extensive ϕ-value analysis, based on 27 mutations throughout the protein, found that both these structures are present in the transition state but that the rest of the molecule remains largely unstructured at this stage of folding [76]. Another study suggested that two additional β-strands are present in the transition state [77].

The simulations of the mechanical unfolding of ubiquitin were carried out at constant force [16]. As in the constant-force experiments [72], the forces acted on the chain ends. In the presence of these forces, the energy function becomes $\tilde{E} = E - \boldsymbol{F} \cdot \boldsymbol{R}$, where E is given by (10.1), $\boldsymbol{F}$ is the force, and $\boldsymbol{R}$ denotes the end-to-end vector. The forces studied were 100, 140 and 200 pN. For each force, a set of more than 500 unfolding simulations were carried out using Monte Carlo dynamics. For the backbone angles, to avoid large unphysical deformations of the chain, the single-variable update was not used, but only the semi-local biased-Gaussian-step method. All simulations were started from the native structure but with different random number seeds.
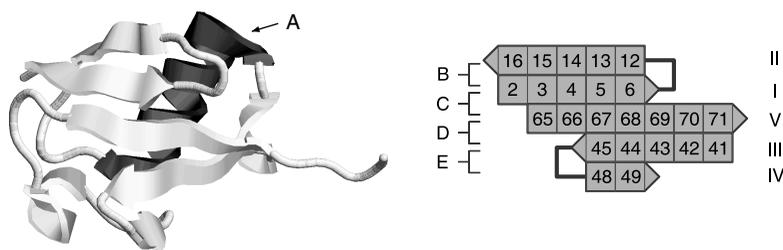

Fig. 10.10. Schematic illustrations of the native structure of ubiquitin with labels A–E for the secondary-structure elements. (*Left*) A 3D model (Protein Data Bank code 1d3z [73]) drawn with RasMol [33]. (*Right*) The organization of the β-sheet. Amino acid numbers, strand labels (I–V) and two β-hairpin turns are indicated

End-to-End Distance

Figure 10.11 shows the time evolution of the end-to-end distance $r = |\boldsymbol{R}|$ in two typical runs at 100 pN. From this figure, it can be seen that unfolding proceeds in a stepwise fashion in the simulations, although the steps are not as sharp as in the experiments. In many runs, the unfolding occurred in a single step, as in run R1 in Fig. 10.11. However, in agreement with the experiments [72], there were also several examples of unfolding via intermediate states, as in run R2 in Fig. 10.11. So, both one-step unfolding and unfolding through intermediate states indeed occurred in the simulations.

Also, properties such as the size of the unfolding step, the frequency of occurrence of intermediate states and the position of the typical intermediate state were all found to be in reasonable agreement with experimental data [72]. For the range of forces studied, it was found that the size of the unfolding step increased with force, whereas intermediate states occurred more frequently at lower force. These dependencies on the strength of the applied force remain to be verified experimentally.

The comparison of these results with experimental data serve as an important test of the model; the model was developed by folding studies of much smaller chains, and it is not at all clear that it can fold ubiquitin. Unfolding is, however, easier to simulate than folding to a unique native state, and atomic-level simulations of force-induced unfolding have been performed before for several proteins including ubiquitin [78] and the I27 Ig domain from titin [79, 80]. The results above suggest that the model of this article indeed is able to capture key features of the mechanical unfolding of ubiquitin. Having verified these properties, more detailed measurements were performed, which in particularly aimed at characterizing the typical intermediate state.
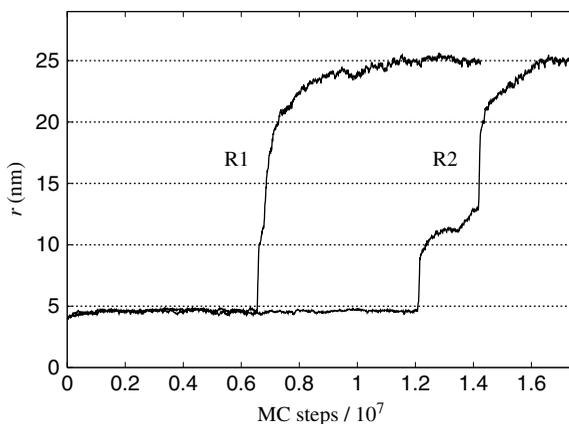

Fig. 10.11. Monte Carlo evolution of the end-to-end distance r in two representative runs. The protein chain unfolds in a single step in run R1 and via an intermediate state in run R2. The pulling force is 100 pN in both runs

Unfolding Pathway

To delineate the unfolding process, five key elements of the native structure, labeled A–E, were followed (see Fig. 10.10). The structure A is the α-helix, whereas B–E are the four different pairs of adjacent strands in the β-sheet. To examine the order in which the structures A–E break, the presence of their native hydrogen bonds was followed as a function of the end-to-end distance r (r shows an essentially monotonic increase with time). Figure 10.12 shows the result of this analysis for all the native backbone hydrogen bonds in the structures A–E at 100 pN. From this figure, it is immediately clear that the structures do not break in a random order but instead in a statistically preferred order, namely CBDEA. The structures C and B tend to break below the typical r for intermediate states, $r_I \approx 12$ nm, whereas D, E and A tend to break above r_I. The data in Fig. 10.12 thus suggest that the typical intermediate is composed of A, D and E.
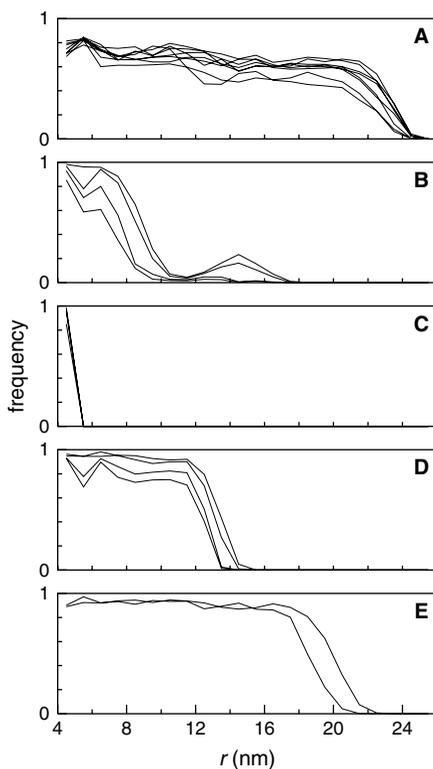
Fig. 10.12. The frequency of occurrence as a function of the end-to-end distance r for all native backbone hydrogen bonds in the respective structures **A–E**, at 100 pN. Each curve represent one hydrogen bond

This analysis does not tell how strong the statistical preference is for the unfolding order CBDEA, since it is based on averages over all events. To check this, the time of breaking of the structures A–E was directly analyzed, so that an unfolding path (a permutation of ABCDE) was obtained for each individual event. Disregarding interchanges of B and D due to partial refolding of the β-hairpin B, it was found that more than 80 % of the events followed the same unfolding path, CBDEA. The order was the same in events with and without intermediate states, and neither did it change between the three different forces studied.

This predicted unfolding order does not agree with the zero-force experiments mentioned above, in which A and B were found to be the most stable structures. Therefore, it should be stressed that the fact that B breaks early in the simulations appears physically reasonable. That C breaks first is inevitable because the other parts cannot sense the force until C is broken (see Fig. 10.10). The native state is mechanically resistant because C is pulled longitudinally so that several bonds must break simultaneously. Once C is gone, nothing keeps the β-hairpin B from unzipping, one bond at a time. Unzipping requires less force than separation by longitudinal pulling [81, 82], which makes B likely to break soon after C.

Recently, the thermal unfolding of ubiquitin was studied using the same model [17] to see whether the model actually predicts a different unfolding order in this case. This was found to be the case. In fact, in these simulations, the most long-lived structures were A and B, which experimentally have been found to be the thermally most stable parts of the molecule [74, 75]. A previous computational study compared thermal unfolding, folding and mechanical unfolding for the I27 Ig domain from titin [83].

10.4 Summary

With improved algorithms and faster computers, it is becoming computationally feasible to simulate how small proteins fold to their native states. This is an exciting development that will lead to a better understanding not only of protein folding but also of the interaction of proteins with other molecules and materials. For all-atom simulations of proteins, a number of interaction potentials, at different levels of sophistication, have been developed. The model discussed in this article has a crude but computationally convenient potential, which was developed through folding studies of selected peptides. It is encouraging that this model, despite its simplicity, could be used to study peptide folding and aggregation as well as the mechanical unfolding of ubiquitin, without changing any parameter of the model. The model will be further developed by studying new amino acid sequences, which will impose new conditions on the interaction potential. Hopefully, this will extend the applicability of the model to new interesting problems.

References

1. C. Brändén, J. Tooze: *Introduction to Protein Structure* (Garland, New York, 1991)
2. H. J. Dyson, P. E. Wright: Curr. Opin. Struct. Biol. **12**, 54 (2002)
3. A. K. Dunker, J. D. Lawson, C. J. Brown, R. M. Williams, P. Romero, J. S. Oh, C. J. Oldfield, A. M. Campen, C. R. Ratliff, K. W. Hipps, J. Ausio, M. S. Nissen, R. Reeves, C. H. Kang, C. R. Kissinger, R. W. Bailey, M. D. Griswold, M. Chiu, E. C. Garner, Z. Obradovic: J. Mol. Graph. Modell. **19**, 26 (2001)
4. C. M. Dobson: Nature **426**, 884 (2003)
5. H. S. Chan, H. Kaya, S. Shimizu: Computational methods for protein folding: scaling a hierarchy of complexities. In *Current Topics in Computational Molecular Biology*, edited by T. Jiang, Y. Xu, M. Q. Zhang (MIT press, Cambridge, Massachusetts, USA, 2002), pp. 403–447
6. J. M. Wang, P. Cieplak, P. A. Kollman: J. Comput. Chem. **21**, 1049 (2000)
7. A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus: J. Phys. Chem. B **102**, 3586 (1998)
8. U. Stocker, W. F. van Gunsteren: Proteins **40**, 145 (2000)
9. G. Kaminski, W. L. Jorgensen: J. Phys. Chem. **100**, 18010 (1996)
10. T. Yoda, Y. Sugita, Y. Okamoto: Chem. Phys. **307**, 269 (2004)
11. A. Irbäck, B. Samuelsson, F. Sjunnesson, S. Wallin: Biophys. J. **85**, 1466 (2003)
12. A. Irbäck, F. Sjunnesson: Proteins **56**, 110 (2004)
13. A. Irbäck, S. Mohanty: Biophys. J. **88**, 1560 (2005)
14. A. G. Cochran, N. J. Skelton, M. A. Starovasnik: Proc. Natl. Acad. Sci. USA **98**, 5578 (2001)
15. G. Favrin, A. Irbäck, S. Mohanty: Biophys. J. **87**, 3657 (2004); erratum **89**, 754 (2005)
16. A. Irbäck, S. Mitternacht, S. Mohanty: Proc. Natl. Acad. Sci. USA **102**, 13427 (2005)
17. A. Irbäck, S. Mitternacht: Proteins **65**, 759 (2006)
18. G. Chikenji, Y. Fujitsuka, S. Takada: Proc. Natl. Acad. Sci. USA **103**, 3141 (2006)
19. G. N. Ramachandran, V. Sasisekharan: Adv. Protein Chem. **23**, 283 (1968)
20. J. W. Neidigh, R. M. Fesinmeyer, N. H. Andersen: Nat. Struct. Biol. **9**, 425 (2002)
21. L. R. Dodd, T. D. Boone, D. N. Theodorou: Mol. Phys. **78**, 961 (1993)
22. G. Favrin, A. Irbäck, F. Sjunnesson: J. Chem. Phys. **114**, 8154 (2001)
23. U. H. E. Hansmann, Y. Okamoto: Curr. Opin. Struct. Biol. **9**, 177 (1999)
24. A. P. Lyubartsev, A. A. Martsinovski, S. V. Shevkunov, P. N. Vorontsov-Velyaminov: J. Chem. Phys. **96**, 1776 (1992)
25. E. Marinari, G. Parisi: Europhys. Lett. **19**, 451 (1992)
26. A. Irbäck, F. Potthast: J. Chem. Phys. **103**, 10298 (1995)
27. A. Irbäck, S. Mohanty: J. Comput. Chem. **27**, 1548 (2006)
28. D. J. Lockhart, P. S. Kim: Science **260**, 198 (1993)
29. F. J. Blanco, G. Rivas, L. Serrano: Nat. Struct. Biol. **1**, 584 (1994)

30. R. M. Fesinmeyer, F. M. Hudson, N. H. Andersen: J. Am. Chem. Soc. **126**, 7238 (2004)
31. T. Kortemme, M. Ramírez-Alvarado, L. Serrano: Science **281**, 253 (1998)
32. M. López de la Paz, E. Lacroix, M. Ramírez-Alvarado, L. Serrano: J. Mol. Biol. **312**, 229 (2001)
33. R. Sayle, E. J. Milner-White: Trends Biochem. Sci. **20**, 374 (1995)
34. C. D. Snow, B. Zagrovic, V. S. Pande: J. Am. Chem. Soc. **124**, 14548 (2002)
35. C. Simmerling, B. Strockbine, A. E. Roitberg: J. Am. Chem. Soc. **124**, 11258 (2002)
36. J. W. Pitera, W. Swope: Proc. Natl. Acad. Sci. USA **100**, 7587 (2003)
37. R. Zhou: Proc. Natl. Acad. Sci. USA **100**, 13280 (2003)
38. A. Schug, W. Wenzel, U. H. E. Hansmann: J. Chem. Phys. **122**, 194711 (2005)
39. S. Williams, T. P. Causgrove, R. Gilmanshin, K. S. Fang, R. H. Callender, W. H. Woodruff, R. B. Dyer: Biochemistry **35**, 691 (1996)
40. P. A. Thompson, W. A. Eaton, J. Hofrichter: Biochemistry **36**, 9200 (1997)
41. J. A. Vila, D. R. Ripoll, H. A. Scheraga: Proc. Natl. Acad. Sci. USA **97**, 13075 (2000)
42. A. E. García, K. Y. Sanbonmatsu: Proc. Natl. Acad. Sci. USA **99**, 2782 (2002)
43. H. Nymeyer, A. E. García: Proc. Natl. Acad. Sci. USA **100**, 13934 (2003)
44. D. Roccatano, A. Amadei, A. Di Nola, H. J. C. Berendsen: Protein Sci. **8**, 2130 (1999)
45. V. S. Pande, D. S. Rokhsar: Proc. Natl. Acad. Sci. USA **96**, 9062 (1999)
46. A. R. Dinner, T. Lazaridis, M. Karplus: Proc. Natl. Acad. Sci. USA **96**, 9068 (1999)
47. A. E. García, K. Y. Sanbonmatsu: Proteins **42**, 345 (2001)
48. B. Zagrovic, E. J. Sorin, V. Pande: J. Mol. Biol. **313**, 151 (2001)
49. R. Zhou, B. J. Berne, R. Germain: Proc. Natl. Acad. Sci. USA **98**, 14931 (2001)
50. E. Kussell, J. Shimada, E. I. Shakhnovich: Proc. Natl. Acad. Sci. USA **99**, 5343 (2002)
51. R. Zhou: Proteins **53**, 148 (2003)
52. P. G. Bolhuis: Proc. Natl. Acad. Sci. USA **14**, 12129 (2003)
53. G. Wei, N. Mousseau, P. Derreumaux: Proteins **56**, 464 (2004)
54. P. H. Nguyen, G. Stock, E. Mittag, C. K. Hu, M. S. Li: Proteins **61**, 795 (2005)
55. V. Muñoz, P. A. Thompson, J. Hofrichter, W. A. Eaton: Nature **390**, 196 (1997)
56. A. M. Gronenborn, D. R. Filpula, N. Z. Essig, A. Achari, M. Whitlow, P. T. Wingfield, G. M. Clore: Science **253**, 657 (1991)
57. B. D. Bursulaya, C. L. Brooks III: J. Am. Chem. Soc. **121**, 9947 (1999)
58. G. Colombo, D. Roccatano, A. E. Mark: Proteins **46**, 380 (2002)
59. S. Y. Kim, J. Lee, J. Lee: J. Chem. Phys. **120**, 8271 (2004)
60. J. C. Rochet, P. T. Lansbury Jr.: Curr. Opin. Struct. Biol. **10**, 60 (2000)
61. J. J. Balbach, Y. Ishii, O. N. Antzutkin, R. D. Leapman, N. W. Rizzo, F. Dyda, J. Reed, R. Tycko: Biochemistry **39**, 13748 (2000)
62. L. O. Tjernberg, J. Näslund, F. Lindqvist, J. Johansson, A. R. Karlström, J. Thyberg, L. Terenius, C. Nordstedt: J. Biol. Chem. **271**, 8545 (1996)
63. B. Ma, R. Nussinov: Proc. Natl. Acad. Sci. USA **99**, 14126 (2002)
64. D. K. Klimov, D. Thirumalai: Structure **11**, 295 (2003)
65. S. Santini, N. Mousseau, P. Derreumaux: J. Am. Chem. Soc. **126**, 11509 (2004)
66. S. Gnanakaran, R. Nussinov, A. E. García: J. Am. Chem. Soc. **128**, 2158 (2006)
67. M. Sunde, C. Blake: Adv. Protein Chem. **50**, 123 (1997)

68. D. J. Gordon, J. J. Balbach, R. Tycko, S. C. Meredith: Biophys. J. **86**, 428 (2004)
69. M. P. Lambert, A. K. Barlow, B. A. Chromy, C. Edwards, R. Freed, M. Liosatos, T. E. Morgan, I. Rozovsky, B. Trommer, K. L. Viola, P. Wals, C. Zhang, C. E. Finch, G. A. Krafft, W. L. Klein: Proc. Natl. Acad. Sci. USA **95**, 6448 (1998)
70. D. M. Walsh, D. M. Hartley, Y. Kusumoto, Y. Fezoui, M. M. Condron, A. Lomakin, G. B. Benedek, D. J. Selkoe, D. B. Teplow: J. Biol. Chem. **274**, 25945 (1999)
71. D. M. Walsh, I. Klyubin, J. V. Fadeeva, W. K. Cullen, R. Anwyl, M. S. Wolfe, M. J. Rowan, D. J. Selkoe: Nature **416**, 535 (2002)
72. M. Schlierf, H. Li, J. M. Fernandez: Proc. Natl. Acad. Sci. USA **101**, 7299 (2004)
73. G. Cornilescu, J. L. Marquardt, M. Ottiger, A. Bax: J. Am. Chem. Soc. **120**, 6836 (1998)
74. F. Cordier, S. Grzesiek: J. Mol. Biol. **317**, 739 (2002)
75. H. S. Chung, M. Khalil, A. W. Smith, Z. Ganim, A. Tokmakoff: Proc. Natl. Acad. Sci. USA **102**, 612 (2005)
76. H. M. Went, S. E. Jackson: Protein Eng. Des. Sel. **18**, 229 (2005)
77. B. A. Krantz, R. S. Dothager, T. R. Sosnick: J. Mol. Biol. **337**, 463 (2004)
78. P. C. Li, D. E. Makarov: J. Chem. Phys. **121**, 4826 (2004)
79. H. Lu, K. Schulten: Biophys. J. **79**, 51 (2000)
80. S. B. Fowler, R. B. Best, J. L. T. Herrera, T. J. Rutherford, A. Steward, E. Paci, M. Karplus, J. Clarke: J. Mol. Biol. **322**, 841 (2002)
81. R. Rohs, C. Etchebest, R. Lavery: Biophys. J. **76**, 2760 (1999)
82. D. J. Brockwell, E. Paci, R. C. Zinober, G. S. Beddard, P. D. Olmsted, D. A. Smith, R. N. Perham, S. E. Radford: Nat. Struct. Biol. **10**, 731 (2003)
83. M. Cieplak, J. Sułkowska: J. Chem. Phys. **123**, 194908 (2005)