ELSEVIER

Biophysical Chemistry 115 (2005) 153–158

Biophysical Chemistry

http://www.elsevier.com/locate/biophyschem

Analysis of PIN1 WW domain through a simple statistical mechanics model

Pierpaolo Bruscolini[a,*], Fabio Cecconi[b]

[a]*Instituto BIFI, Universidad de Zaragoza, c/ Corona de Aragón 42, E-50009 Zaragoza, Spain*
[b]*Dipartimento di Fisica Universitá "La Sapienza" and INFM Unitá di Roma1, P.le A. Moro 2, I-00185 Roma*

Received 30 June 2004; received in revised form 29 November 2004; accepted 10 December 2004
Available online 25 December 2004

Abstract

We have applied a simple statistical mechanics Gō-like model to the analysis of the PIN1 WW domain, resorting to mean field and Monte Carlo techniques to characterize its thermodynamics, and comparing the results with the wealth of available experimental data. PIN1 WW domain is a 39-residue protein fragment which folds on an antiparallel β-sheet, thus representing an interesting model system to study the behavior of these secondary structure elements. Results show that the model correctly reproduces the two-state behavior of the protein, and also the trends of the experimental ϕ_T values. Moreover, there is a good agreement between Monte Carlo results and the mean field ones, which can be obtained with a substantially smaller computational effort.

PACS: 87.15.Aa; 87.15.Cc
Keywords: Protein folding; PIN1 WW domain; Statistical mechanics models; Monte Carlo simulations; Mean field approximations; ϕ_T values

1. Introduction

Understanding the folding process of proteins is one of the most challenging issues of biochemistry which requires sophisticated simulations at atomic resolution generally referred as all-atom methods. At present, the large incompatibility between folding time scales and regimes explored by all-atom simulations makes the folding process not yet accessible to these powerful computational approaches. Even though very encouraging progress has been achieved, their applicability remains restricted to the study of peptides and fragments of proteins [1,2]. In addition, the comparison to experiments requires an accumulation of folding events to gain an enough large statistics furtherly narrowing the route to the full-atom techniques. These limitations suggest resorting to minimalist models which adopt a less accurate description of protein chains, residue–residue and residue–solvent interactions [3–7]. Approximate representations reduce the computational costs and, with a certain amount of uncertainty, allow to follow all the stages which bring a protein into its native fold. The use of simplified models within a statistical mechanical approach to protein folding is grounded on the assumption that not all the chemical details need to be retained to understand and describe the basic properties of folding processes. Of course the approximations that this kind of approach introduces must ensure that the basic principles of biochemistry are fulfilled to keep a correct description of the real molecules. Several years ago, a simple model was proposed by Gō [8] to attain a phenomenological but complete description of the folding reaction. The model replaces all non-bonded interactions by attractive native-state contact energies. This recipe, which can be applied only when native structure is known, implements the idea that a reasonable energy bias toward the native state could capture the relevant features of the folding process. This kind of modelling removes high energetic barriers along the pathways towards the native conformation (which lies in a deep minimum), and produces

* Corresponding author. Tel.: +34 976562214; fax: +34 976562215.
E-mail address: pier@unizar.es (P. Bruscolini).

0301-4622/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2004.12.020

relatively smooth energy landscapes. As a result the folding "funnel" [9,10] leading to the native state is very smooth so the folding process results "ideal". Folding events simulated through Gō-like potentials take only few nanoseconds making possible to obtain statistically meaningful results for generic proteins and polypeptide chains. Since Gō-like models lack any energetic frustration, the scope of their applications is related to the investigation of the role of geometric frustration and configurational entropy in the folding process. Their success in providing a reasonable account for kinetic properties of the folding process is related to the assumption that folding kinetics is mainly determined by native geometry, together with native state stability, and this view is indeed supported by several experimental works [11–15]. Along the lines indicated by the Gō-phylosophy, other simplified models exploiting the information present in the native state have been proposed [16–19]. In this paper we continue our analysis [20] of one of this Gō-like models, the Finkelstein model [19,21], and apply it to the study of the PIN1 WW domain (pdb code 1I6C) which has a well-defined and simple native structure made of two slightly bent antiparallel β-sheets. Its distinctive feature, which is also reflected in its name, is the presence of two triptophanes (W), located 20 residues apart from one another. Its structure, with a simple topology, lacks all those features that can complicate the modeling. Thus this molecule represents a suitable candidate to explore the kinetic and thermodynamic factors responsible for the formation of β-sheets and their stability, and is also a suitable benchmark through which models and theories are validated. The Finkelstein model is particularly suitable for analyzing the folding thermodynamics of two-state proteins and the WW domain is known to fold in a two-state scenario so we can test whether the model can faithfully reproduce the known experimental data [22] about WW domain folding.

The organization of the paper is as follows. In Section 2 we discuss the model and its assumptions. In Section 3 we present the Monte Carlo and mean field methods we adopt, and in Section 4 we report and discuss our results. Finally, Section 5 is dedicated to the concluding remarks.

2. Description of Finkelstein Model

Finkelstein model assumes a simple description of the polypeptide chain, where residues can stay only in an ordered (native) or disordered (non-native) state. Then, each micro-state of a protein with L residues is encoded in a sequence of L binary variables $s=\{s_1,s_2,\ldots,s_L\}$, $s_i=\{0,1\}$. Residues with $s_i=1$ ($s_i=0$) are in their native (non-native) conformation. When all variables take on the value 1, the protein is considered folded, whereas the random coil corresponds to all 0's. Because each residue can be in one of the two states, ordered or disordered, the free energy landscape consists of 2^L configurations. This enormous reduction in the number of configurations available to a protein is a quite delicate point because it is a restrictive feature of the model. However this crude assumption, already employed in Ref. [23], is the simplest one leading to a two-state behaviour of the folding. The effective Hamiltonian (indeed, a free energy function) is

$$H(s) = \varepsilon \sum_{i<j} \Delta_{ij} S_i S_j - TS(s), \tag{1}$$

where $S(s)$ is given by:

$$S(s) = R\left[q \sum_{i=1}^{L} (1 - S_i) + S_{\text{loop}}(s)\right]. \tag{2}$$

R is the gas constant and T the absolute temperature. The first term in Eq. (1) is the energy associated to native contact formation. Non-native interactions are neglected: this further assumption can be just tested a posteriori and it is expected to hold if, during the folding process, the progress along the reaction coordinate is well depicted on the basis of the native contacts. That is, the reaction coordinate(s) must be related to just the native contacts. Moreover, such progress must be slow with respect to all other motions, so that all non-native interaction can be "averaged-out" when considering the folding pathways. Δ_{ij} denotes the element i,j of the contact matrix, whose entries are the number of heavy-atom contacts between residues i and j in the native state. Here we consider two amino acids to be in contact if there are at least two heavy atoms (one from amino acid i and one from j) separated by a distance less than 5 Å. The matrix Δ embodies the geometrical properties of the protein.

The second term in Eq. (1) is the conformational entropy associated to the presence of unfolded regions along the chain, and vanishes in the native state.

More precisely, the first term in Eq. (2) is a sort of "internal" entropy of the residues: qR represents the entropic difference between the coil and the native state of a single residue. This can be noticed by considering that in the fully unfolded state S_{loop} vanishes and the remaining entropy is qLR only.

The term RS_{loop} in Eq. (2) is the entropy pertaining to the disordered closed loops protruding from the globular native state [18]; it reads:

$$S_{\text{loop}}(s) = \sum_{i<j} J(r_{ij}) \prod_{k=i+1}^{j-1} (1 - s_k) s_i s_j. \tag{3}$$

According to [19], we take:

$$J(r_{ij}) = -\frac{5}{2}\ln|i-j| - \frac{3}{4}\frac{r_{ij}^2 - d^2}{Ad|i-j|}. \tag{4}$$

In this way the configuration of a disordered loop going from residues $(i+1)$ to $(j-1)$, with i and j in their native positions, is assimilated to a random walk with end-to-end distance r_{ij}, the latter being the distance between C_α atoms

of residues i and j in the native state. The parameters d=3.8 Å and A=20 Å are the average distance of consecutive C_α along the chain and persistence length, respectively. The entropy of one loop closure (Eq. (4)) differs from the classical result $-3R/2\ln(N)$ pertaining to a free Gaussian chains [24]. The presence of the factor 5/2, instead of 3/2, stems from the fact that a loop exiting the globule must lie completely outside of it, to account for the self-avoidance. Thus, the spatial domain occupied by the globule results in a forbidden region for the disordered loop, and this simple sterical constraint, reducing the number of accessible conformations, increases the entropy loss obtained from the closure of the loop [18].

3. Methods

A direct comparison between model predictions and experimental results requires a tuning of the coefficients q and ε in the energy function Eq. (1). In our computation we set q=2.31 and regarded ε as an adjustable parameter. We determined it by imposing that the mean field specific heat exhibits its "collapse" peak in correspondence to the experimental transition temperature T=332 K [22]. Despite the use of a simple MF approach, we expect that this procedure yields a correct estimate for ε, since the MF is known to reproduce the thermodynamics properties of the Finkelstein model pretty faithfully [20]. Once we determined the optimal choice of q and ε, we performed Monte Carlo simulations to investigate the thermal folding of the WW domain. We implemented a Metropolis algorithm with transition rates between states j and k

$$w(j\to k) = \exp\left[\left(H_j - H_k\right)/RT\right]$$

R being the gas constant, T the temperature and H_j the Finkelstein energy of state j, according to Eq. (1).

We applied the multiple histogram technique (MHT) [25] to reconstruct the system density of states (DOS) in the full range of accessible energies. To this end, we carried out MC runs at 50 equally spaced temperatures in the range 273–383 K, and for each run we collected the energy histogram to estimate the statistical weight of all configurations with a certain energy. Through the Swendsen–Ferremberg procedure [25], these histograms were optimally linearly combined to extract the whole DOS and thus compute the entropy $S(E)=R\ln[g(E)]$ up to an additive constant. The knowledge of entropy allows evaluating the free energy profiles $F(E)=E-TS(E)$, and other relevant thermodynamical quantities for the folding, such as the specific heat.

In its variational formulation [26], mean field approximation, for a system with Hamiltonian H and corresponding free energy F, amounts to minimizing

$$F_{\mathrm{var}} \le F_0 + \langle H - H_0\rangle_0, \tag{5}$$

where H_0 is a solvable trial Hamiltonian, F_0 is the corresponding free energy, both depending on free parameters $x=\{x_1 \dots x_L\}$ (variational parameters). Minimization leads to the self-consistent equations that in their general form read

$$\left\langle \frac{\partial H_0}{\partial x_l} \right\rangle_0 \langle H - H_0\rangle_0 - \left\langle (H - H_0)\frac{\partial H_0}{\partial x_l} \right\rangle_0 = 0, \tag{6}$$

with l=1, …,L. We have implemented different versions of the MFA for the model that differ each from the other by the choice of the trial Hamiltonian.

The standard MFA employs as the trial Hamiltonian:

$$H_0 = \sum_{i=1}^{L} x_i S_i, \tag{7}$$

with x_i to be determined by minimizing the variational free energy [26]

$$F_{\mathrm{var}}(x,T) = \sum_{i=1}^{L} f_0(x_i,T) + \langle H - H_0\rangle_0, \tag{8}$$

where $\sum_i f_0(x_i,T)$ is the free energy associated to H_0,

$$f_0(x_i,T) = -\frac{1}{\beta}\ln\left\{1+\exp(-\beta x_i)\right\}. \tag{9}$$

Thermal averages, performed through the Hamiltonian H_0, factorize $\langle s_i s_j \dots s_k\rangle_0 = \langle s_i\rangle_0\langle s_j\rangle_0 \dots \langle s_k\rangle_0$. The approximate average site "magnetization" $m_i=\langle s_i\rangle_0$ depends only on the field x_i, and is given by

$$m_i = \frac{\partial F_0}{\partial x_i} = \frac{1}{1+\exp(\beta x_i)}. \tag{10}$$

Instead of working with external fields x_i's, it is more intuitive to use the corresponding "magnetizations" m_i's, writing F_{var} as a function of the m_i's. Due to the choice of H_0, Eq. (7), and to the expression Eq. (10), evaluating the thermal average $\langle H\rangle_0$ amounts to replacing, in the Hamiltonian Eq. (1), each variable s_i by its thermal average m_i. In the end we get:

$$F_{\mathrm{var}}(\mathbf{m},T) = \varepsilon \sum_{ij} \Delta_{ij} m_i m_j - TS(\mathbf{m}) + RT \sum_{i=1}^{L} g(m_i), \tag{11}$$

where $g(u)=u\ln(u)+(1-u)\ln(1-u)$ and $S(\mathbf{m})$ is obtained from Eq. (2) by substituting $s_i\to m_i$. The last term corresponds to $F_0-\langle H_0\rangle_0$ in Eq. (5): it is the entropy associated to the system with Hamiltonian H_0 and is the typical term that stems from this kind of MFA [26]. The minimization of function Eq. (11) with respect to $\mathbf{m}$ leads to self-consistent equations:

$$g'(m_i) = \varepsilon \sum_j \Delta_{ij} m_j - RT\left(q - \frac{\partial S_{\mathrm{loop}}(\mathbf{m})}{\partial m_i}\right). \tag{12}$$

Eq. (12) can be solved numerically by iteration and provide the optimal values of the magnetizations that we denote by $\mathbf{m}^*$. Once the set of solutions $\mathbf{m}^*$ is available, we can compute the variational free energy $F_{\mathrm{var}}(\mathbf{m}^*)$ that

represents the better estimate of the system free energy F. Free energy profiles are evaluated performing the minimization after the introduction of Lagrange multipliers, corresponding to the constraint of considering states with a fixed number of native residues.

A different MFA consists in taking a trial Hamiltonian that accounts exactly for the entropic term of the original one, resorting to the procedure introduced in Ref. [27], and approximates the interactions by introducing a weight dependent on the number of native residues in the configuration. Namely, we consider the set of configurations of the proteins with M native residues (M=0, ..., L) and take as the trial Hamiltonian

$$H_0(\mathbf{x}) = \sum_{M=0}^{L} \delta\left(M - \sum_i s_i\right) H_0^{(M)}(\mathbf{x}), \tag{13}$$

where $\delta(\bullet)$ is the Kronecker delta, and $H_0^{(M)}$ is the Hamiltonian restricted to the configurations with M natives:

$$H_0^{(M)}(\mathbf{x}) = \sum_{i=1}^{L} \tilde{\varepsilon}_i x_i \frac{M-1}{L-1} s_i - TS(s), \tag{14}$$

with $\tilde{\varepsilon}_i = (1/2)\sum_{j=1}^{N} \varepsilon_{i,j}\Delta_{i,j}$. Each residue i, in a generic configuration with M native residues, feels an interaction $\tilde{\varepsilon}_i$ which it would feel in the native state, weakened by a factor $(M-1)/(L-1)$ (accounting for the fact that not all the residues are native), times the external field x_i, to be fixed by the mean field procedure.

The mean field equations for this case can be found in Ref. [20].

4. Results and discussion

The folding transition is signalled by the behavior of the specific heat, which develops a peak identifying the T_f.

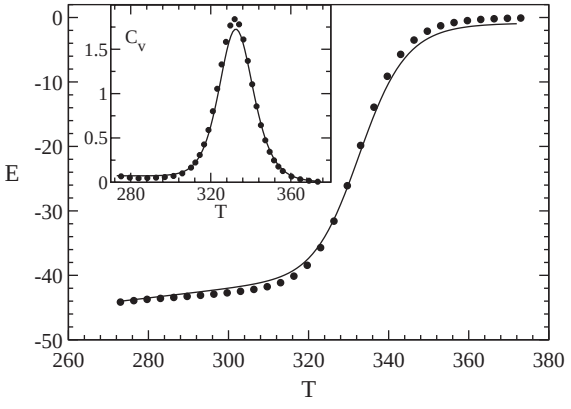

Fig. 1. Specific heat in kcal mol^{-1} T^{-1} (inset) and energy (in kcal mol^{-1}) as function of temperature, computed through MC simulations (points) and standard mean field approach (line).

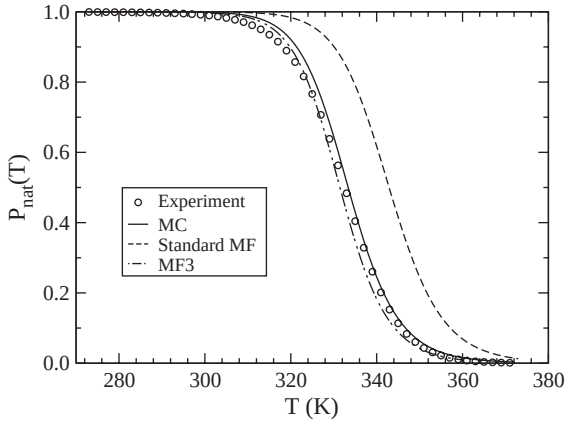

Fig. 2. Fraction of native protein as a function of temperature: MC simulation, standard mean field, mean field 3 of Ref. [20] compared with experimental fit in Ref. [22].

Standard MF peak position is imposed to the correct experimental folding temperature to fit the parameters; notice though that MC peak is correctly found at the same position, providing a consistency check between the two methods (Fig. 1).

PIN1 WW domain is reported to be a two-state folder [22]: this is recovered by both the MC and the MF approximations, as can be seen in Fig. 2. MC and the more complicated MF approach reproduce with reasonable accuracy the experimental signal.

The two-state nature of the protein can also be seen in the free energy profiles (Figs. 3 and 4). It is remarkable that the barrier separating folded from unfolded conformations is quite flat, especially in the MC case, so that mutations could likely induce relevant changes in its position with just a slight change in the energies, a scenario which is indeed suggested in Ref. [22].

Monte Carlo and mean field free energy profiles allow to estimate the stability gap ΔG and the folding barrier $\Delta G^{\dagger}$ as

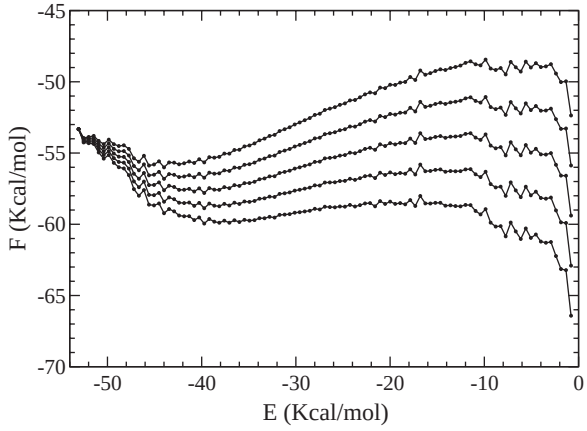

Fig. 3. MC free energy profiles, with energy as the coordinate of reaction, at different temperatures: from top to bottom T=292, 312, 332, 352, 372 K.

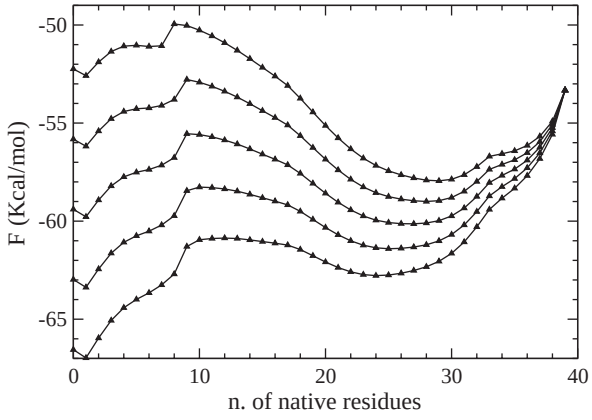

Fig. 4. Standard MF free energy profiles, in the number of native residues, at different temperatures: from top to bottom T=292, 312, 332, 352, 372 K.

a function of temperature. The comparison with the corresponding experimental curves [22]

$$\Delta G_{\mathrm{ex}}(T) = \Delta G_0 + \Delta G_1(T - T_{\mathrm{f}}) + \Delta G_2(T - T_{\mathrm{f}})^2$$

$$\Delta G^{\dagger}_{\mathrm{ex}}(T) = \Delta G^{\dagger}_0 + \Delta G^{\dagger}_1(T - T_{\mathrm{f}}) + \Delta G^{\dagger}_2(T - T_{\mathrm{f}})^2$$

where T_{f}=332 K, $\Delta G_{0,1,2}$={−0.062,0.105,6.244·10^{-4}} kcal/mol and $\Delta G^{\dagger}_{0,1,2}$={5.089,0.0568,1.232·10^{-3}} kcal/mol. The result of this comparison is reported in Fig. 5.

Notice that all methods compare most favorably with the experimental results in the vicinity of T_{f}, which is to be expected, since the model only accounts for the geometry, and not for the details of the interactions, with their temperature dependence in the hydrophobic contributions. MC gives a good estimate of both the stability gap and the barrier, while standard mean field gives a reasonable description of the folding barrier, but overestimates the stability. On the other hand, the more complicated MF scheme recovers correctly the stability, but it overestimates the barrier, at least if we consider, as we did in Ref. [20], just the profile of F_0 (relying on the good approximation that F_0 provides to F_{var}), without resorting to the more correct, but computationally expensive minimization of a constrained F_{var}. A more accurate analysis of free energy profiles within this MF scheme is left for future work. In the following, we analyze standard MF and MC results concerning another important experimental quantity, namely the ϕ_{T} values (Fig. 6). ϕ_{T} values are defined as

$$\phi_{\mathrm{T}} = \frac{\partial \Delta G^{\dagger}}{\partial T}\,\frac{1}{\frac{\partial \Delta G}{\partial T}} = \frac{\Delta S^{\dagger}}{\Delta S}, \qquad (15)$$

and give an idea of the entropy of the barrier compared to that of the native state, providing a measure of the proximity of the barrier to the folded state. The experimental results show a monotonically increasing, continuous function, spanning a wide range of values. MC and MF results indeed agree in the monotonically increasing behavior, reflecting thus the Hammond behavior [28,29], even if in a discretized version. Indeed, they show a series of discrete jumps that, in the case of MC simulations, are not simply an effect of the binning in the reaction coordinate, but seem to suggest sharp movements in the barrier position: sudden changes in ϕ_{T} are in complete correspondence to shifts in the position of the barrier, as reported in Fig. 6.

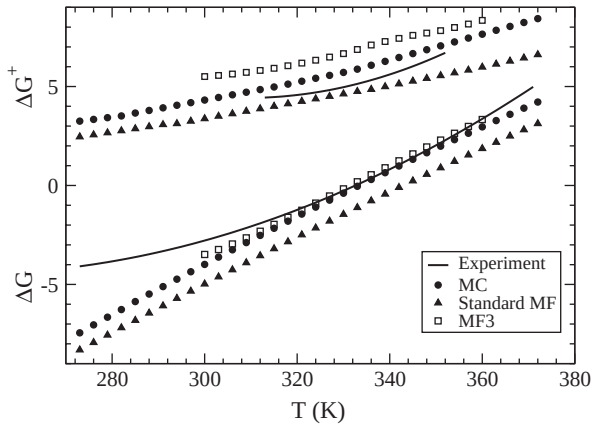

Fig. 5. Folding barrier (top set of curves) and stability of the native state (bottom set) as a function of T, from experiment and simulations. Data are reported in kcal/mol.

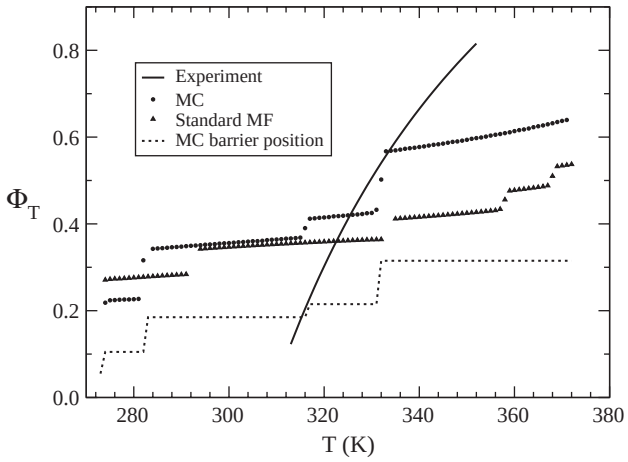

Fig. 6. ϕ_{T} values from experiments and simulation, together with the barrier position for the MC case. Barrier position values at a given T are evaluated as the energy coordinate (x-axis in Fig. 3) corresponding to the barrier top at that temperature, normalized to the total contact energy in the native state (independent from T: E=−53.32 kcal/mol with our choice of the parameters). Notice how the shifts in ϕ_{T} correspond to those in the barrier position.

5. Conclusions

The application of the Finkelstein model to protein PIN1 WW domain reveals that this model, after fitting the parameter ε in order to reproduce the correct transition temperature, is able to describe correctly the thermodynamics of the folding process, at least in the case of simple two-state behavior. Indeed, the estimate of the folding barrier, both in the case of MF approximation as well as for MC

simulations, lies within a relative error of about 15% from the experimental estimate in all the region of experimental measures. This is indeed interesting, as the model lacks every detail about the nature of the residues, dealing with all atomic contacts in the ground-state on the same footing. Moreover, the estimate of the entropy is based on the theory of noninteracting polymers, and neglects possible clashes of the protruding unfolded loops with the folded part of the protein.

Another important result concerns the ϕ_T values: both MF and MC results recover the non-decreasing nature of experimental values, with MC providing a better estimate of the slope than MF. At difference with the experimental values, though, theoretical ϕ_T values increase in a discontinuous fashion, with abrupt changes followed by steady plateaus. This behavior is related to the fact that the transition state is quite broad, so that the actual free energy maximum, determining the barrier, jumps through different values of the reaction coordinate (the number of native residues or the energy). This is an aspect that deserves further analysis, also because the simple three-state model, with a negligible intermediate, put forward by the authors of Ref. [22] does not seem to be able to reproduce the experimental results with sufficient accuracy, and a satisfactory description of the transition state of this protein has still to be found. Probably, it will require the introduction of residue heterogeneities and more accurate studies on the dynamics of the system.

References

[1] A.R. Dinner, T. Lazaridis, M. Karplus, Understanding beta-hairpin formation, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 9068.

[2] V.S. Pande, D.S. Rokhsar, Molecular dynamics simulations of unfolding and refolding of a beta-hairpin fragment of protein G, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 9602.

[3] J.D. Bryngelson, P.G. Wolynes, Spin glasses and the statistical mechanics of proteins, Proc. Natl. Acad. Sci. U. S. A. 84 (1987) 7524.

[4] K.F. Lau, K. Dill, A lattice statistical mechanics models of the conformational and sequence spaces of proteins, Macromolecules 22 (1989) 3986.

[5] E.I. Shaknovich, A.V. Gutin, A new approach to the design of stable proteins, Protein Eng. 6 (1993) 793.

[6] Z. Guo, D. Thirumalai, Kinetics and thermodynamics of folding of a de-novo designed four-helix bundle protein, J. Mol. Biol. 263 (1996) 323.

[7] P. De Los Rios, G. Caldarelli, Putting proteins back into water, Phys. Rev. E 62 (2000) 8449.

[8] N. Go, Theoretical studies of protein folding, Annu. Rev. Biophys. Bioeng. 12 (1983) 183–210.

[9] P.G. Wolynes, J.N. Onuchic, D. Thirumalai, Navigating the folding routes, Science 267 (1995) 1619.

[10] K.A. Dill, H.S. Chan, From Levinthal to pathways and funnels, Nat. Struct. Biol. 4 (1997) 10–19.

[11] J.N. Onuchic, P.G. Wolynes, Theory of protein folding, Curr. Opin. Struct. Biol. 14 (2004) 70–75.

[12] J. Chen, L.X. Zhang, L. Jing, Y.X. Wang, Z.T. Jiang, D.L. Zhao, Predicting protein structure from long-range contacts, Biophys. Chemist. 105 (2003) 11–21.

[13] K.W. Plaxco, K.T. Simons, I. Ruczinski, D. Baker, Topology, stability, sequence, and length: defining the determinants of two-state protein folding kinetics, Biochemist 39 (2000) 11177–11183.

[14] D.S. Riddle, V.P. Grantcharova, J.V. Santiago, E. Alm, I. Ruczinski, D. Baker, Experiment and theory highlight role of native state topology in SH3 folding, Nat. Struct. Biol. 6 (1999) 1016–1024.

[15] F. Chiti, N. Taddei, P. Webster, D. Hamada, T. Fiaschi, G. Ramponi, C.M. Dobson, Acceleration of the folding of acylphosphatase by stabilization of local secondary structure, Nat. Struct. Biol. 6 (1999) 380–387.

[16] E. Alm, D. Baker, Prediction of protein-folding mechanisms from free-energy landscapes derived from native structures, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 11305–11310.

[17] V. Muñoz, W.A. Eaton, A simple model for calculating the kinetics of protein folding from three-dimensional structures, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 11311–11316.

[18] A.V. Finkelstein, A.Y. Badretdinov, Physical reason for fast folding of the stable spatial structure of proteins: a solution of the Levinthal paradox, Mol. Biol. 31 (1997) 391–398.

[19] O.V. Galzitskaya, A.V. Finkelstein, A theoretical search for folding/unfolding nuclei in three-dimensional protein structures, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 11299–11304.

[20] P. Bruscolini, F. Cecconi, Mean-field approach for a statistical mechanical model of proteins, J. Chem. Phys. 119 (2003) 1248–1256.

[21] S.O. Garbuzynskiy, A.V. Finkelstein, O.V. Galzitskaya, Outlining folding nuclei in globular proteins, J. Mol. Biol. 336 (2004) 509–525.

[22] M. Jäger, H. Nguyen, J.C. Crane, J.K. Kelly, M. Gruebele, The folding mechanism of a β-sheet: the WW domain, J. Mol. Biol. 311 (2001) 373–393.

[23] R. Zwanzig, Simple model of protein-folding kinetics, Proc. Natl. Acad. Sci. U. S. A. 92 (1995) 9801–9804.

[24] H. Jacobson, W.H. Stockmayer, Intramolecular reaction in polycondensations: I. The theory of linear systems, J. Chem. Phys. 18 (1950) 1600.

[25] A.M. Ferrenberg, R.H. Swendsen, Optimized Monte Carlo data analysis, Phys. Rev. Lett. 63 (1989) 1195–1198.

[26] M. Plischke, B. Bergersen, Equilibrium Statistical Physics, World Scientific, Singapore, 1989.

[27] P. Bruscolini, A. Pelizzola, Exact solution of the Munoz–Eaton model for protein folding, Phys. Rev. Lett. 88 (2002) 258101.

[28] G.S. Hammond, J. Am. Chem. Soc. 89 (1967) 2915–2927.

[29] I.E. Sánchez, T. Kiefhaber, Non-linear rate-equilibrium free energy relationships and Hammond behavior in protein folding, Biophys. Chemist. 100 (2003) 397–407.