Proceedings of the 2006 WSEAS International Conference on Mathematical Biology and Ecology, Miami, Florida, USA, January 18-20, 2006 (pp31-36)

A molecular dynamics investigation of the kinetic bottlenecks of the hPin1 WW domain. II: simulations with the Go model

FABIO CECCONI
Istituto dei Sistemi Complessi
(ISC-CNR)
(also INFM),
Via dei Taurini 19
I-00185 Roma,
ITALY
Fabio.Cecconi@roma1.infn.it

CARLO GUARDIANI
Centro CSDC
(also INFN *)
Università di Firenze,
Via Sansone 1,
I-50019 Sesto Fiorentino,
ITALY
carlo.guardiani@unifi.it

ROBERTO LIVI
Dipartimento di Fisica
(also CSDC, INFN and INFM),
Università di Firenze,
Via Sansone 1,
I-50019 Sesto Fiorentino,
ITALY
Livi@fi.infn.it

Abstract: - The present paper is the second of a series in which we perform molecular dynamics simulations on the WW domain of Pin1 protein. The aim of the work is the reconstruction of the folding/unfolding pathway with a special interest for the kinetic bottlenecks. In the first paper of the series we showed that unfolding simulations using the Sorenson/Head-Gordon (SHG) model correctly identify the kinetic bottlenecks in agreement with the experimental data. In the present paper we repeat the same simulation protocol using the Go model. The simulations show that the unfolding mechanisms reconstructed by the two models are consistent with each other, but the Go process is much less cooperative and thus less accurate in the identification of the kinetic bottlenecks. The poor performance of the Go model in the case of Pin1 WW domain, can be related to the absence of angular potentials, which makes the protein conformations more flexible: such an effect is presumably amplified by the small size and by the particular shape of the protein.

Key-Words: - WW domains, Pin1 protein , kinetic bottlenecks, Go model, Sorenson/Head-Gordon model, Molecular Dynamics

1 Introduction

The WW domains are a family of fast-folding, compact, modular domains featuring a triple-stranded, antiparallel beta-sheet. In particular, the human Pin1 protein WW domain, due to the availability of a large amount of structural [1, 2], thermodynamical and kinetic [3] experimental data, represents an excellent benchmark to test computational methods. The structure of this domain was resolved both through NMR [2] and X-ray [1] diffraction techniques. For a detailed structural description we refer the reader to the first paper of this series [4] and to References [1, 3].

The purpose of the present work is to identify the bottlenecks in the folding process of WW domains

*We gratefully acknowledge partial financial support from the INFN TO61 Experiment

through Molecular Dynamics simulations of thermal denaturation, using simplified protein models. The kinetic bottlenecks are related to the establishment of those specific interactions requiring the overcoming of large free-energy barriers. The formation of such interactions acts as a nucleus for the establishment of further contacts and accelerates the searching of the native state.

In the previous paper of this series [4] we reported the results of simulations carried on within the frame of the Sorenson/Head-Gordon (SHG) model [5]. This off-lattice, minimal model portrays the protein as a chain of beads of three different flavours: hydrophobic (B), hydrophilic (L) and neutral (N). The folding is driven by the attraction between hydrophobic beads while the secondary structural elements appear as the result of a bias on the dihedral angles. Despite the extreme simplification related to the lack of side-chain packing and hydrogen-bonding, the model correctly identified the kinetic bottlenecks of the folding/unfolding process in agreement with the results of the Φ-value analysis performed by Gruebele *et al.* [3].

It must be stressed, however, that the effectiveness of the SHG model strongly depends on the sequence optimization procedure, that compensates for the model limitations [6, 7]. This is why it is interesting to compare the performance of the SHG model with that of the Go model [8]: being sequence-independent, the latter does not require any optimization procedure. The theoretical basis of the Go model relies on the observation that the topology of the native state can play a crucial role in driving the folding process [9, 10, 11, 12, 13, 14, 15]. The key experimental findings which support the above statement are: (1) the close similarity of the transition-state conformations of proteins having structurally related native states (despite the very poor sequence similarity) [16], [17] and (2) the strong influence that certain simple topological properties, such as the contact order, have on protein folding rates [18]. These experimental observations are consistent with the growing evidence reported in several recent papers [9, 10, 11, 12, 13, 14, 15] that the Go model can be confidently used for the characterization of transition states. In this paper we report on the results of our unfolding simulations of the WW domain of the human Pin1 protein within the frame of the Go model. The results will be systematically compared with those obtained using the SHG model and described in the first paper [4] of this series.

2 Methods

The Go model [8] portrays the protein as a chain of beads centered on the α-carbon positions. The model provides a bias towards the native state by promoting the formation of native pairwise interaction. Given a reference native structure and chosen a distance cutoff (in our case $R_c = 6.5$ Å), those residue pairs, whose distance R_{ij} is less than R_c, are considered to be in native contact and throughout the simulation they interact through an attractive Lennard-Jones potential:

$$V_{nat}(r_{ij}) = \epsilon \left[\left(\frac{R_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{ij}}{r_{ij}} \right)^{6} \right] \qquad (1)$$

Conversely, if two residues are not in native contact, they interact through an excluded-volume repulsive potential:

$$V_{nnat}(r_{ij}) = \epsilon \left(\frac{\sigma}{r_{ij}} \right)^{12}$$

with $\sigma = 4.5$ Å and $\epsilon = 2.1 Kcal\, mol^{-1}$. The force-field is completed by an harmonic potential describing the interaction between consecutive residues and mimicking the covalent bonds.

We performed constant temperature MD simulations within the isokinetic scheme [19], by using reduced units. The unfolding simulations started from the PDB conformation stored in the file 1NMV.pdb (low temperature, $T = 0.01$) and we gradually heated the system to the value $T = 1$ in 50 temperature jumps. For each temperature we equilibrated the system over 6×10^6 time steps. Sampling of observables was performed for further 6×10^6 time steps of the dynamics.

The course of the denaturation process was monitored through several thermodynamic and structural parameters. The energy E and the specific heat C_V were computed through the multiple histogram technique [20]. The method was also applied for the computation of the structural overlap Q, *i.e.* the average fraction of native contacts in the protein. Further structural indicators employed in our study are

the radius of gyration R_g and the root mean square deviation $rmsd$. Due to the importance of the two hydrophobic clusters in Pin1 WW domain, we also introduced specific order parameters:

$$Q_{CLk} = \frac{\sum_{ij \in CLk} R_{ij}}{\sum_{ij \in CLk} r_{ij}}, k = 1, 2$$

where R_{ij} and r_{ij} are the native and current distances respectively, between residues i and j belonging to the same cluster. Small values of Q_{CLk} indicate that the cluster is ill-formed, because its residues are far apart. For further information on the methodology followed in our computations, we refer the reader to the first paper of this series [4].

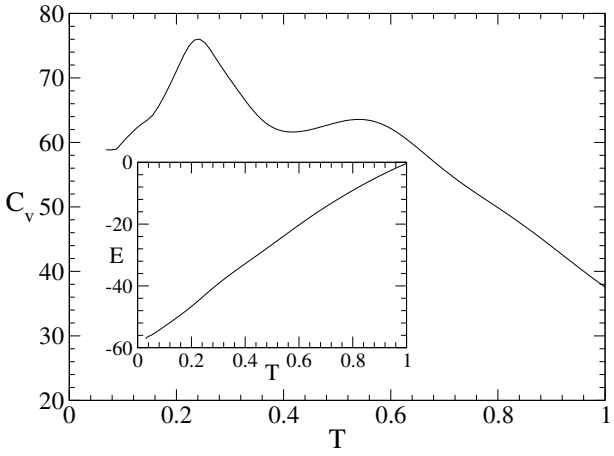

Figure 1: Thermal behaviour of the specific heat in the Go unfolding. Inset: plot of energy in the thermal Go denaturation of the Pin1 WW domain. The results were obtained with the weighted histogram method.

3 Results

The specific heat profile (Fig. 1) of an unfolding simulation from the native structure (1NMV.pdb) using the Go model is characterized by two peaks: a narrow and high peak (P1) at $T_1 = 0.23$ and a wider but less pronounced peak (P2) at $T_2 = 0.54$ reminiscent of the shoulder in the SHG simulations. The importance of P1 is highlighted by the behaviour of two structural parameters: Q_{CL1} and R_g. Parameter Q_{CL1} decreases significantly at T_1, thus signalling the breakdown of hydrophobic cluster CL1. This results in a swelling of the protein shown by a steep increase of R_g. The structural parameter of the second hydrophobic cluster, Q_{CL2}, instead, decreases at higher temperatures in proximity of P2, where $\beta 1 - \beta 2$ contacts in the central parts of the strands are cleft. Remarkably, the thermal behaviours of both *rmsd* and Q do not exhibit sudden changes as they are characterized by gradual variations, which reflect a progressive denaturation.

The graduality of contact breakdown appears to be a peculiar feature of the Go model as opposed to SHG model, where the specific heat plot is characterized by a single peak and native contacts break down abruptly in a narrow temperature range around the peak. This difference is confirmed by the free energy profiles that in the Go simulation do not show the double-well shape, typical of an abrupt transition as in the case of the SHG simulation.

The unfolding mechanism proceeds through three different stages. Before and along the increasing branch of P1, the contacts belonging to the hydrophobic clusters break down. This pattern is consistent with the one observed in the SHG unfolding simulations, where the breakdown of the first cluster occurs just before the specific heat peak. The following stage of the unfolding pathway, corresponding to the increasing branch of P2, is characterized by the breakdown of $\beta_2 - \beta_3$ contacts and $\beta_1 - \beta_2$ contacts in the region of the strands distal from loop L1. These events are thus equivalent to those characterizing the increasing branch of the C_V peak in the SHG unfolding simulations. The final stage of unfolding, in correspondence of the decreasing branch of P2, features the breakdown of $\beta_1 - \beta_2$ contacts and it is thus equivalent to events taking place at the top and shoulder of the C_V plot in the SHG simulation. The most notable difference in the unfolding mechanism in the Go and SHG simulations relies on the fact that, while in the SHG model, the $\beta_1 - \beta_2$ contacts break down only at high temperatures (top and shoulder of the C_V peaks), in the Go model, many of them are broken also at low temperatures (before and at peak 1). Figure 2 summarizes the steps of the unfolding process through a color-coded contact map.

Tables 1 and 2 list the main stages of the unfolding process in the Go and SHG simulations. The differences in the first stage are mainly due to the different compactness of the reference state structures.

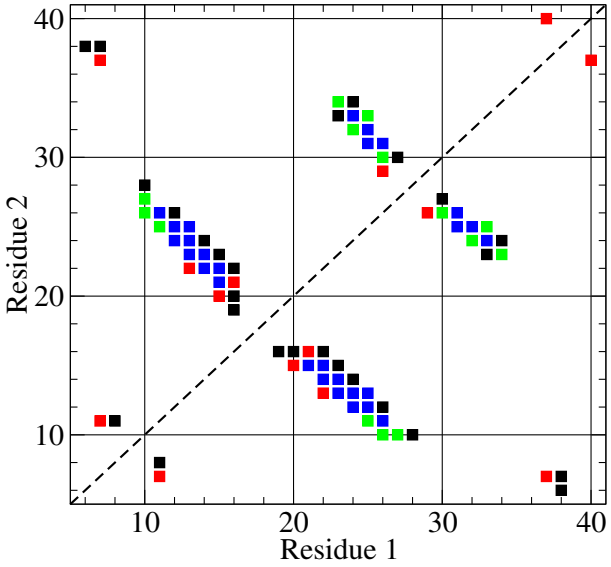

Figure 2: Color-coded contact map of the PDB structure used as the reference conformation in the Go simulations. In the map, black squares represent contacts broken at low temperatures, before P1; the red squares represent the contacts broken at peak P1; the green squares refer to the contacts broken in correspondence of the increasing branch of P2, and the blue squares refer to the decreasing branch of P2.

A difference can also be noticed in the second and third stages, which in the Go simulations appear to be reversed as compared to SHG simulations. Apart from these differencies, the unfolding pathways, reconstructed by the two models, are consistent with each other and they identify three key steps:

1. Breakdown of cluster CL1
2. Breakdown of β_2-β_3 and loop I contacts
3. Breakdown of β_1-β_2 contacts

4 Conclusions

We performed molecular dynamics unfolding simulations of the Pin1 WW domain with the aim of identifying the kinetic bottlenecks in the reaction pathway. We systematically compared the performance of two different protein models, namely the SHG model [5] (see paper 1 of this series [4]) based on the hydrophobicity character of the residues of the protein chain, and the Go model [8] that relies on the topology of the native state.

GO-UNFOLDING	
stage	breakdown event
Before P1	$\beta_1 - \beta_2$
P1	CL1 and T1
Incr. Branch P2	$\beta_2 - \beta_3$ + $\beta_1 - \beta_2$ (distal)
Decr Branch P2	$\beta_1 - \beta_2$ (interm.)

Table 1: Summary of the events occurring during the thermal denaturation of Pin1 WW domain using the Go model.

HG-UNFOLDING	
stage	breakdown event
Before P	$\beta_1 - \beta_3$, $\beta_2 - tail$ and CL1.
Incr. branch of P	$\beta_2 - \beta_3$ $\beta_1 - \beta_2$ (distal)
Peak	T1
Shoulder	$\beta_1 - \beta_2$ (interm.)

Table 2: Overview of the events taking place during the thermal unfolding of Pin1 WW domain using the HG model.

Our simulations showed that the SHG model leads to an unfolding mechanism much more cooperative than the one yielded by the Go model. Several indicators such as the specific heat, the free-energy profiles and some structural parameters, support this conclusion.

Despite these differences, the unfolding mechanisms occurring in both models are consistent with each other from the point of view of the patterns of contact breakdown. In fact, in both cases the unfolding pathway proceeds as follows: (1) breakdown of the first hydrophobic cluster; (2) breakdown of $\beta_2 - \beta_3$ and L1 contacts; (3) breakdown of the contacts in the intermediate region of β_1 and β_2 strands. The most important difference is the more gradual opening of $\beta_1 - \beta_2$ contacts in the Go model. As a consequence, the SHG model appears to be more realistic in the identification of folding bottlenecks, as shown by the good agreement with the Φ-value data analysis by Gruebele *et al.* [3].

The simulations also showed that, despite the rather high $rmsd$ between the structure produced by the SHG folding simulation and the PDB conformation, the SHG model was able to capture the essential features of the native topology. In conclusion, for the specific WW domain considered in this paper,

the SHG model clearly outperforms the Go model in the identification of the unfolding pathway and, in particular, in the location of the kinetic bottlenecks, while still retaining the most important topological features of the native structure. Our work thus provides evidence of the ability of coarse-grained, minimal models to capture key aspects of the protein folding process, thus representing valuable tools in the investigation of large-sized, slow-folding proteins that are too computationally demanding for all-atom simulations.

References:

[1] M.A. Verdecia, H.K. Huang, M.E. Bowmanm, K.P. Lu, T. Hunter, and J.P. Noel. Structural basis for phosphoserine-proline recognition by ww domains. *Nature Struct. Biol.*, 7:639–643, 2000.

[2] E. Bayer, S. Goettsch, J. W. Mueller, B. Griewel, E. Guiberman, L.M. Mayr, and P. Bayer. Structural analysis of the mitotic regulator hpin1 in solution: Insights into domain architecture and substrate binding. *J.Biol.Chem.*, 278:26183, 2003.

[3] M. Jaeger, H. Nguyen, J.C. Crane, J.W. Kelly, and M. Gruebele. The folding mechanism of a beta-sheet: the ww domain. *J. Mol. Biol.*, 311:373–393, 2001.

[4] F. Cecconi, C. Guardiani, and R. Livi. A molecular dynamics investigation of the kinetic bottlenecks of the hpin1 ww domain: simulations with the sorenson/head-gordon model. Submitted to the 2006 WSEAS International Conference on *Mathematical Biology and Ecology*, Miami, Florida, USA, January 18-20, 2006.

[5] J.M. Sorenson and T. Head-Gordon. Matching simulation and experiment: a new simplified model for simulating protein folding. *J. Comp. Bio.*, 7:469–481, 2000.

[6] S. Brown, J. Fawzi, and T. Head-Gordon. Coarse-grained sequences for protein folding and design. *Proc. Natl. Acad. Sci. USA*, 100:10712–10717, 2003.

[7] J.M. Sorenson and T. Head-Gordon. Redesigning the hydrophobic core of a model β-sheet protein: destabilizing traps through a threading approach. *Proteins*, 37:582–591, 1999.

[8] N. Go and H.A. Scheraga. On the use of classical statistical mechanics in the treatmentof polymer chain conformations. *Macromolecules*, 9:535–542, 1976.

[9] C. Micheletti, J.R. Banavar, A. Maritan, and F. Seno. Protein structures and optimal folding from a geometrical variational principle. *Phys. Rev. Lett*, 82:3372–3375, 1999.

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

[11] V. Munoz, E.R. Henry, J. Hofrichter, and W.A. Eaton. A statistical mechanical model for beta-hairpin kinetics. *Proc. Natl. Acad. Sci. USA*, 95:5872–5879, 1998.

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

[13] F. Chiti, N. Taddei, P.M. White, M. Bucciantini, F. Magherini, M. Stefani, and C.M. Dobson. *Nature Struct. Biol.*, 6:1005–1009, 1999.

[14] J.C. Martinez, M.T. Pisabarro, and L. Serrano. Obligatory steps in protein folding and the conformational diversity of the transition state. *Nature Struct. Biol.*, 5:721–729, 1998.

[15] C. Clementi, H. Nymeyer, and J.N. Onuchic. Topological and energetic factors: what determines the structural details of the transition state ensemble and en-route intermediates for protein folding ? *J. Mol. Biol.*, 298:937–953, 2000.

[16] J.C. Martinez and L. Serrano. The folding transition state between sh3 domains is conformationally restricted and evolutionarily con-

served. *Nature Struct. Biol.*, 6:1010–1016, 1999.

[17] D.S. Riddle, V.P. Grantcharova, J.V. Santiago, E. Alm, I. Ruczinski, and D. Baker. Experiment and theory highlight role of native state topology in sh3 folding. *Nature Struct. Biol.*, 6:1016–1024, 1999.

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

[19] D.J. Evans, W.G. Hoover, B.H. Failor, B. Moran, and A.J.C. Ladd. Non-equilibrium molecular dynamics via gauss's principle of least constraint. *Phys Rev A*, 28:1016–1021, 1983.

[20] A.M. Ferrenberg and R.H. Swendsen. Optimized monte carlo data analysis. *Phys. Rev. Lett.*, 63:1195–1198, 1989.