Biophysical Journal Volume 83 December 2002 3533–3541

Folding Pathways of Prion and Doppel

Giovanni Settanni,* Trinh Xuan Hoang,*† Cristian Micheletti,* and Amos Maritan*†
*International School for Advanced Studies (S.I.S.S.A.) and INFM, 34014 Trieste, Italy; and †The Abdus Salam International Center for Theoretical Physics (ICTP), 34100 Trieste, Italy

ABSTRACT The relevance of various residue positions for the stability and the folding characteristics of the prion protein in its normal cellular form are investigated by using molecular dynamics simulations of models exploiting the topology of the native state. These models allow for reproducing the experimentally validated two-state behavior of the normal prion isoform. Highly significant correlations are found between the most topologically relevant sites in our analysis and the single point mutations known to be associated with the arousal of the genetic forms of prion disease. Insight into the conformational change is provided by comparing the folding process of cellular prion and doppel that share a similar native state topology: the folding pathways of the former can be grouped in two main classes according to which tertiary structure contacts are formed first enroute to the native state. For the latter a single class of pathways leads to the native state again through a two-state process. Our results are consistent and supportive of the recent experimental findings that doppel lacks the scrapie isoform and that such remarkably different behavior involves residues in the region containing the two β-strands and the intervening helix.

INTRODUCTION

Neurodegenerative diseases causing transmissible spongiform encephalopathies (TSE) are the subject of intense research. They affect humans and animals and include scrapie in sheep, mad-cow disease in bovines, and Creutzfeldt-Jacob disease in humans. Most of them arise sporadically, the rest of the instances are inherited or transmitted by inoculation through the dietary or infected tissues.

The belief is that the normal, and benign, form of the prion protein PrP^{c}, which is found in various organs of vertebrates, including the brain, can undergo a posttranslational process leading to a conformational change of its native state (Prusiner, 1982). This new form is designated PrP^{sc} and, contrarily to PrP^{c}, it is insoluble. The NMR investigation of the three-dimensional structure of recombinant PrP^{c} of various species (Goldgaber et al., 1989; Kretzschmar et al., 1991; Hsiao et al., 1989; Gabizon et al., 1993; Calzolai et al., 2000; Garcia et al., 2000; Liu et al., 2000; Zahn et al., 2000; James et al., 1997) has revealed that the N-terminus is unstructured, whereas the C-terminus consists of three α-helices and two short β-strands, but the structure of the malign form, PrP^{sc}, is mostly unknown. However, spectroscopic investigations, mainly based on circular dichroism, showed that PrP^{c} contains ~40% α-helices and little β-sheet, whereas PrP^{sc} is composed of ~30% α-helices and 45% β-sheet (Pan et al., 1993; Pergami et al., 1996).

Submitted February 19, 2002, and accepted for publication July 3, 2002.

Address reprint requests to Amos Maritan, International School for Advanced Studies (SISSA/ISAS), via Beirut 2-4, 34014 Trieste, Italy. Tel.: +39-040-2240462; Fax: +39-040-3787528; E-mail: maritan@sissa.it.

Dr. G. Settanni's present address is Biochemische Institut, Universitaet Zuerich, Winterthurerstrasse 190, 8057 Zuerich, Switzerland.

0006-3495/02/12/3533/09 $2.00

Unfolding and refolding experiments suggest that the conversion between the α-rich and β-rich form occurs through a complete unfolding of the protein (Baskakov et al., 2001). Experiments and clinical tests suggest two main processes for the onset and spreading of the prion infection. In the first case, the spreading of the scrapie form in healthy tissues contaminated with PrP^{sc} reveals that the presence of PrP^{sc} helps further conversion from normal to scrapie form, i.e., PrP^{sc} acts as a template for restructuring PrP^{c}. In the second case, the genetic influence on the propensity of cellular PrP^{c} to rearrange into PrP^{sc} is revealed by the discovery of at least 20 single point mutations in the *PrP* gene in humans that favor the spontaneous onset of the disease (Hsiao et al., 1989; Gabizon et al., 1993; Dlouhy et al., 1992; Petersen et al., 1992; Poulter et al., 1992).

Within the picture presented in Baskakov et al., 2001, the role of the mutations in the arousal of the disease could be determined by several factors ranging from the lowering of a free energy barrier in the conversion process to the increase in the oligomerization rate. One of the goals of the present study is to understand which are the sites where mutations are expected to have the major impact in the stabilization/destabilization of the PrP^{c} structure and/or its folding process.

Recently, some experiments on transgenic mice lacking the *PrP* gene ($PrnP^{0/0}$) have revealed the onset of neurodegeneration (Purkinje cell death) (Sakaguchi et al., 1996) completely different from usual prion disease. This disease was ultimately traced back to a prion homologous gene, *Prnd*, (not expressed in normal subjects), which encoded a protein named doppel, *Dpl*, which has almost the same topology of PrP^{c}, despite the low sequence identity (25%).

A striking difference between *Dpl* and PrP^{sc} is that the former affects the central nervous system while retaining its native structure, i.e., without a need to convert to a scrapie-

like conformation. Various hypotheses have been formulated (Mo et al., 2001) to explain these different behaviors.

The theoretical framework we elaborated and adopted in our study helps to shed light on the putative causes for such remarkably different behavior of doppel and prion. Indeed, within our simplified approach, the different folding behavior and stability of prion and doppel appear to be related to the small, but important, differences in the topology of the native state that affect, in an amplified fashion, the folding routes and ultimately facilitate structural rearrangements of prion.

The model we adopt builds on the importance of the native state topology in steering the folding process; that is, in bringing into contact pairs of amino acids that are found in interaction in the native state. In the past few years an increasing amount of experimental and theoretical evidence has confirmed this view (Riddle et al., 1999; Martinez and Serrano, 1999; Plaxco et al., 1998; Ferrara and Caflisch, 2001; Micheletti et al., 1999; Munoz and Eaton, 1999; Alm and Baker, 1999; Galzitskaya and Finkelstein, 1999; Maritan et al., 2000a,b; Clementi et al., 2000; Baker, 2000; Hoang and Cieplak, 2000a,b; Micheletti, 2001; Gsponer and Caflisch, 2001).

A further confirmation of this view is provided by the remarkable accord of the key folding stages predicted by topology-based models and the available experimental results (Cecconi et al., 2001; Settanni et al., 2001). Indeed, the success of the topological picture itself helps to explain why the folding process is not too sensitive to the detailed chemical composition of most residues in a protein. The dependence of the folding process on the detailed chemistry is much more subtle than that given by the native state topology.

Because traditional topology-based approaches provide a structural bias toward a single native conformation, they are not suited to model the misfolding of prion. Despite this, a considerable insight into prion misfolding can still be obtained by modeling the folding process of the cellular prion and doppel proteins into their similar native structures. The comparison of the pathways in the two cases reveals important differences which can then be associated with the possibility for prion to misfold, at variance with doppel.

Usually, the folding mechanism is affected only when mutations occur in a small set of key residues. Within our theoretical framework those key residues take part in contacts that are crucial for the folding process (Cecconi et al., 2001). The establishment of the those contacts leads to a rapid formation of further interactions. Interestingly, the folding bottlenecks can be identified by just knowing the topology of the native conformation (Cecconi et al., 2001; Settanni et al., 2001). Thus, the topology itself also dictates the impact of chemistry on the folding process.

The purpose of the present work is threefold: 1) determination of the folding bottlenecks for the cellular prion and the identification of the key amino acids taking part in the corresponding crucial contacts; 2) making connections between the set of such key residues with those that are known to be associated with harmful PrP^{sc} mutations. As argued before, a random mutation on key positions will usually result in a disruption of the folding process. Only fine-tuned mutations can lead to a wild-type-like native state (as for viral enzymes mutating under drug attack (Cecconi et al., 2001)) or in another viable structure (as postulated by Prusiner for the prion (Prusiner, 1982)); and 3) furthermore, we will focus our attention on how the topological differences between the native states of *Dpl* and PrP^{c} may have an impact on the folding process and how these differences may result in misfolding propensities. It is important to stress that our study is based on the native-state structures of PrP^{c} and *Dpl*. Hence, although we can confidently identify the crucial folding residues, we cannot confirm explicitly that their mutation results in a different native state form.

MATERIALS AND METHODS

The model for prion

Our model builds on a schematic representation of the three-dimensional structure of the protein, where the amino acids are replaced by effective centroids identified with the C_α atoms. The bias toward the native state is introduced through a Go-like (Go, 1983) energy scoring function that rewards the formation of native contacts between the centroids. The list of native interactions was compiled from the knowledge of the coordinates of atoms of the human prion protein (PDB code 1qlx). A pair of amino acids is considered in interaction if any pair of their heavy atoms, i and j, have a native separation smaller than the distance $1.244(R_i + R_j)$, where the point of inflection of the van der Waals interaction occurs (R_i denotes the van der Waals radius of atom i). The values of R_i are taken from Tsai et al. (1999). In this way one defines a symmetric matrix, known as contact map, Δ, whose entries, Δ_{ij}, are equal to 1 if the ith and jth centroids interact, and zero otherwise.

The energy function for our system includes terms that are routinely used in standard molecular dynamics (MD) simulations on biopolymers (Pearlman et al., 1995; Brooks et al., 1983). It is composed by a "bonded" and a "non-bonded" term: $E = V_B + V_{NB}$. The former accounts for the constraints such as the peptide bond length and Ramachandran angle bias acting on the amino acids at a local level. Its explicit form is:

$$V_B = g \cdot V_p + h \cdot V_a + k \cdot V_d, \tag{1}$$

$$V_p = \sigma \sum_{i=1}^{N-1} (r_{i,i+1} - r_{i,i+1}^{(n)})^2, \tag{2}$$

$$V_a = \sigma \sum_{i=1}^{N-2} (\theta_{i,i+1,i+2} - \theta_{i,i+1,i+2}^{(n)})^2, \tag{3}$$

$$V_d = \sigma \sum_{i=1}^{N-3} (1 - \cos(\tau_{i,i+1,i+2,i+3} - \tau_{i,i+1,i+2,i+3}^{(n)})), \tag{4}$$

where $r_{i,j}$ is the distance between residues i and j, $\theta_{i,j,k}$ is the angle with the jth amino acid as vertex and the ith and the kth as edges, and $\tau_{i,j,k,l}$ is the dihedral generated by the ith, the jth, the kth, and the lth amino acid. The n as superscript denotes the native state value; σ represents a suitable scale factor to fix the temperature scale (set to 10 in our simulations) which is given in dimensionless units.

The minimum of expression (1) is precisely attained in correspondence of the native state. However, the cooperative character of the folding process (Kaya and Chan, 2000a,b) can be captured only by introducing an explicit bias toward the formation of native interactions (Micheletti et al., 2001). This is accomplished through the second term, V_{NB}, which weights the interaction of any pair of nonconsecutive centroids with a van der Waals-like potential:

$$V_{NB} = \sigma \sum_{i<j-3}^{N} V_{ij} = \sigma \sum_{i<j-3}^{N} \left(5\left(\frac{r_{ij}}{r_{ij}^{(n)}}\right)^{-12} - 6\left(\frac{r_{ij}}{r_{ij}^{(n)}}\right)^{-10} \Delta_{ij} \right). \quad (5)$$

This choice rewards native interactions and disfavors non-native ones. The form of the energy scoring function ensures that the global minimum is attained in correspondence of the native state, regardless of the precise values for the (positive) coupling parameters g, h, k. The values used in the present study are $g = 50$, $h = 5$, and $k = 0.3$; we have checked that the results are robust to their variation within a certain range (Settanni et al., 2001).

The dimensionless unit of time in our MD simulations is $5 \cdot 10^{-3}$. In principle, one may tune the scales for the absolute temperature and time in the model to reproduce experimental folding times. This would, however, require the introduction of additional parameters for the sole purpose of reproducing known experimental behavior (see Munoz and Eaton, 1999)). For this reason we choose to keep the model as transparent as possible by using the least number of parameters and focus on novel characterization of prion and doppel. The different conformations generated along the MD trajectory were sampled every 500 time steps to allow a sufficient uncorrelation at temperatures below the folding transition. The multiple-histogram technique (Ferrenberg and Swendsen, 1989) has been applied to reweight quantities such as the average internal energy collected in more than 50 equilibrated runs at different temperatures in the range 2–8. The optimal reweighting allows calculating thermodynamic quantities, such as the average internal energy and the specific heat for an arbitrary temperature, T.

As visible in Fig. 1 *A*, the specific heat exhibits a single peak, which signals a folding transition (Hao and Scheraga, 1997) compatible with a two-state mechanism (see Results section). In addition to this overall characterization of the equilibrium thermodynamics, it is possible to monitor how the individual interactions that are present in the native state are formed as the folding progresses. Indeed, although all native contacts are energetically favored in the same way, their entropic cost of formation may significantly vary according to their locality, burial/exposure, etc.

The crucial contacts for the folding process are identified according to the method given in Cecconi et al., 2001; Settanni et al., 2001; and Micheletti et al., 2001, which singles out the contacts giving the largest contribution to the nonbonded specific heat at the folding transition temperature, T_F. Such contribution is obtained by first writing the thermodynamic average of the nonbonded energy as

$$\langle V_{NB} \rangle = \frac{\mathrm{Tr}\ V_{NB} e^{-\sigma \Sigma (V_{ij}/k_B T_{ij}) - (V_B/k_B T)}}{\mathrm{Tr}\ e^{-\sigma \Sigma (V_{ij}/k_B T_{ij}) - (V_B/k_B T)}}. \quad (6)$$

where the trace is taken over the system microstates. The average value of $\langle V_{NB} \rangle$ at temperature T is obtained by setting $T_{ij} = T$ for all contacting pairs, i and j. Upon differentiation of $\langle V_{NB} \rangle$ with respect to T_{ij} we obtain the contribution of the *i-j* interaction to the nonbonded specific heat:

$$c_{ij} = \frac{d\langle V_{NB} \rangle}{dT_{ij}} = \sigma \frac{\langle V_{NB} V_{ij} \rangle - \langle V_{NB} \rangle \langle V_{ij} \rangle}{k_B T^2}. \quad (7)$$

The physical interpretation of this procedure is that these special contacts are precisely those that, having a significant and sudden formation at T_F, act as bottlenecks for the folding process. It has been shown in our previous studies (Cecconi et al., 2001; Settanni et al., 2001; Micheletti et al., 2001) that $c_{i,j}$ measures the sensitivity of the potential energy of the system to perturbation occurring at the contact (i, j).

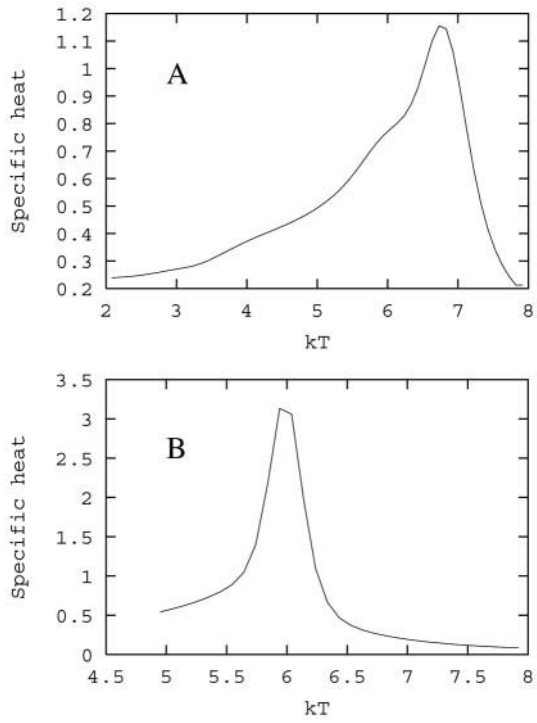

FIGURE 1 Temperature-dependent specific heat for the prion (*A*) and doppel (*B*) models. The folding transition temperature is identified at the point where the specific heat peak occurs.

We have checked that, for most of the contacts, c_{ij} is peaked at a temperature that closely coincides with T_F. For some others this may occur well above or below the folding transition temperature. Those, however, correspond either to the local contacts, which can be easily established even at high temperature, or to the fluctuation of the protein conformation within the native basin (at $T \ll T_F$), and are not important for folding. Thus we confine our analysis of the contacts to the events happening at T_F.

In our studies, a long run carried out at T_F of $2 \cdot 10^7$ time steps allowed a detailed analysis for all contacts. In addition to the identification of crucial contacts, it is useful to introduce a similar measure for the individual amino acids through the single residue specific heat (SRSH) resulting from the contribution of all contacts to which a given site, i, takes part to:

$$c_i = \sum_j \Delta_{ij} c_{ij} \quad (8)$$

SRSH values have been used to identify folding and unfolding phases and to pinpoint residues that play a relevant role in the process (Cecconi et al., 2001; Settanni et al., 2001). The crucial folding steps can thus be identified with two (related) criteria: namely either through the contacts or the sites with the largest specific heat of formation.

Folding pathways

Another issue that we have addressed in our study is to establish the presence of alternative pathways taken from unfolded conformations to the native state. The presence of alternative pathways, which describes how the native state can be progressively established, does not necessarily imply the

presence of folding intermediates (Hoang and Cieplak, 2000a,b; Ozkan et al., 2001).

Again, this is accomplished by examining in detail both the thermodynamic and the kinetic relevant features extracted from the model. The comparison of dynamic trajectories connecting unfolded and native-like conformations has revealed the presence of several folding events that have to occur for the protein to fold. The main ones that emerged from our analysis, and that are particularly apt for characterizing the folding process, correspond to the formation of contacts involving three sets of contiguous residues. The first set, *Bh*, from residue 125 to residue 172, comprises all N-terminal residues (β-sheet included) before helix 2 (defined according to the indications in the PDB file); the second set, *h*1, from residue 173 to 194, comprises residues of helix 2; and finally the third set, *h*2, from residue 200 to 228, is composed of residues from helix 3.

The four observed events are 1) formation of contacts between residues within *Bh*; 2) formation of contacts between *Bh* and *h*1; 3) formation of contacts between *Bh* and *h*2; and 4) formation of contacts between *h*1 and *h*2.

For convenience of presentation, to describe the four events on a similar footing, we focus on the top 10 contacts (in terms of specific heat contributions c_{ij}) in each interacting group. It is interesting to note that, among these top contacts, those occurring within an α-helix almost never occur. This can be explained because local contacts, among others, are easily formed, even at high temperature, and hence do not constitute a bottleneck for overcoming the configurational barrier to the native state.

We computed the fraction of formed contacts, Q_i ($i = Bh$, *Bh*-*h*1, *Bh*-*h*2, *h*1-*h*2), within each of the four groups along the dynamical trajectories of $2 \cdot 10^7$ steps at T_F. The natural binning size for each of the Q_i was 1/10. From this, we obtained a vivid picture of the dynamical effects of the free energy landscape at the folding transition (see Fig. 5 and the Results section). We also carried out 100 ("quenched") folding runs at $T = (1 - \epsilon)T_F$, where $\epsilon = 1/30$, slightly below T_F, starting from completely unfolded conformations (thermalized at high-temperature runs). In all cases, the native state was reached in $<2 \cdot 10^5$ steps. The average Q_i value as a function of the elapsed time using all 100 trajectories was also calculated (see Fig. 6).

The choice of the reaction coordinates used here was made from the analysis of several unfolding/refolding trajectories. This allowed us to identify a set of contacts whose progressive formation is sensitive to the progress toward the native state. Another aspect of our choice that is paramount in this context is that a perfectly analogous set of "sensitive" contacts can be defined for the doppel protein, thus allowing a comparison of the main folding events in the two cases. In principle one should recourse to an optimized and objective choice of the reaction coordinates to highlight the main folding differences between the two proteins. However, the number of degrees of freedom involved is huge, and makes the task arduous. For these reasons we will consider only the reduced reaction-coordinate space identified above which, though not optimized, appears to be quite effective in discriminating the folding landscape of cellular prion and doppel.

Doppel protein

We carried out a parallel analysis for the doppel protein, *Dpl*, whose NMR-resolved structure was retrieved from the PDB (PDB code 1i17, model 1). For this protein, too, we built the contact map considering all-atom distances, and obtained both the overall specific heat (Fig. 1 *B*) and the one associated to individual contacts using the same methods described above. A crucial difference in the folding events of the two proteins was revealed by monitoring the four events, analogous to those of the previous section.

For *Dpl* the sets of amino acids structurally homologous of the prion ones are as follows: the set *Bh* corresponds to residues from 1 to 50 before helix 3; *h*1 to residues from 51 to 76 (comprising helices 3 and 4); and the set *h*2 to residues from 78 to 96 (comprising helix 5). The numeration of the helices is consistent with the secondary structure indications present in the PDB file. Free energy profiles and time-dependent averages of the reaction coordinate are shown in Figs. 5 and 6, respectively.

TABLE 1 Top 30 sites in prion ranked according to the single residue specific heat (SRSH)

Rank	Site	Rank	Site	Rank	Site
1	137	11	213	21	214
2	210	12	165	22	163
3	175	13	160	23	212
4	158	14	162	24	184
5	141	15	206	25	187
6	183	16	205	26	179
7	209	17	157	27	208
8	198	18	139	28	217
9	161	19	150	29	159
10	211	20	134	30	180

RESULTS

Determination of key residues

The specific heat curves as a function of the temperature for PrP^c and *Dpl* are shown in Fig. 1, *A* and *B*, respectively. The single peak in each of them allows identifying unambiguously the folding transition in these systems. The two-state behavior of the system is supported by the plot showing the effective free energy profiles of the two systems as a function of the reaction coordinates Q_i that we have defined in the previous section; in these plots only two minima are present that correspond to the folded and the unfolded conformation (Fig. 5).

From the long run of $2 \cdot 10^7$ time steps performed at T_F, the SRSH values for prion have been computed according to Eq. 8. We have ranked the residues according to their SRSH (see Table 1). It is natural to compare the top key sites from our analysis with those whose mutation is known to favor the emergence of PrP^{sc} (highlighted in Table 2). In the top 10 amino acids of our list we found 4 of the known key mutating sites; whereas in the top 30 amino acids their number increases to 12. The statistical relevance of this match can be obtained from a combinatory calculation of the probability to have at least the observed number of matches by pure chance. For the top 10 amino acids the probability to get a better result by chance is 6.8%, whereas for the top 30 amino acids that probability becomes 0.02%.

TABLE 2 Sites where mutations (in the NMR-solved part of Human *PrP*) have been observed to cause/stop the prion disease or be fundamental for correct folding of PrP^c

145, 160, 178, 179, 180, **183**, 187, 188, 196
198, 200, 202, 208, **210**, **211**, 212, 214, 217

Sites 214 and 179 represent cysteine residues that form a disulfide bond needed to observe the PrP^c α-form of prion. Single residue specific heat (SRSH), as determined by our model, ranking within the top 10 are in boldface; those ranking within the top 30 are underlined (see Table 1).

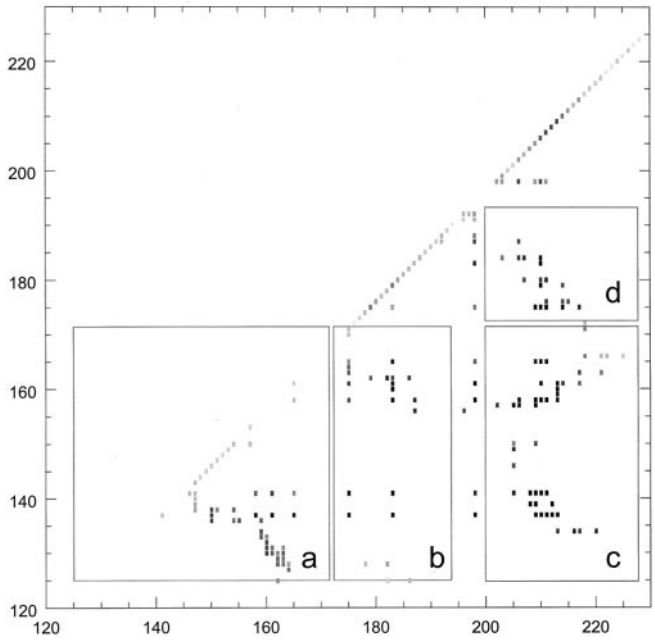

FIGURE 2 Contact map of PrP^c (PDB code 1qlx) where contacts are highlighted according to their specific heat c_{ij} of Eq. 7. Darker spots correspond to contacts with higher c_{ij}, while blue ones correspond low SRSH. The relevant sets of contacts (as defined in the text) are framed within the rectangles; (*A*) set *Bh*, (*B*) set *Bh*-*h*1, (*C*) set *Bh*-*h*2, (*D*) set *h*1-*h*2. Notice that red spots are prevalent in all sets but *Bh*.

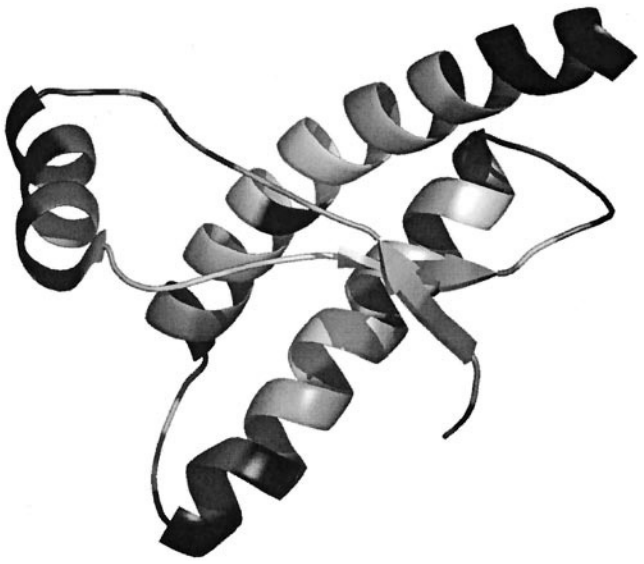

FIGURE 3 The three-dimensional structure of prion protein where amino acids are shown in gray-scale according to the value of their SRSH. Lighter gray corresponds to residue with higher SRSH. Hot sites are located mainly on the C-terminal strands of the β-sheet, on helix 2, and helix 3. This figure has been produced by using PyMOL (DeLano, 2002).

Thus, the criterion based on the SRSH allows confidently identifying amino acids important for the folding process of the prion protein.

Their structural location and role is made apparent in Fig. 2, where the prion contact map has been colored according to the value of the SRSH. As visible, crucial contacts connect secondary-structure elements thus forming the protein tertiary structure. In particular, they connect the β-sheet–helix 1 region to the helix 2 and helix 3 (*Bh*-*h*1 and *Bh*-*h*2 sets, respectively) and helix 2-helix 3 (*h*1-*h*2 set). The parts of the protein contributing more to the specific heat, within our model, are located at the C-terminal strand of the β-sheet, and at the two C-terminal helices, especially helix 3 (see Fig. 3).

Notwithstanding an overall structural homology between PrP^c and *Dpl*, which allows finding a good correspondence between their structural motifs, *Dpl* shows a different packing of the latter that results in a contact map containing significant differences with respect to prion (Fig. 4).

Folding pathways

The analysis of the long run of $2 \cdot 10^7$ time steps carried out at T_F has led to the construction of an *effective* free energy landscape. It is based on the reaction coordinates Q_i as defined in the Methods section. The free energy as a function of pairs (Q_i, Q_j) for all six combinations of *i* and *j* have been calculated as

$$F(Q_i, Q_j) = -\ln(P(Q_i, Q_j)) \quad (9)$$

where $P(Q_i, Q_j)$ is the fraction of time the dynamical trajectory spends in the bin around (Q_i, Q_j). The contour plots of *F* are shown in Fig. 5, *A* and *B* for PrP^c and *Dpl*, respectively. The two energy minima around $Q = 0$ and in the higher *Q* region correspond to the unfolded and folded conformations, respectively. For prion the plots involving set *h*1-*h*2 show the presence of multiple pathways connecting these two minima. For example, the contour plot in the bidimensional space spanned by $Q_{h1\text{-}h2}$ and $Q_{Bh\text{-}h1}$ presents two possible successions of events: either the *Bh*-*h*1 set is formed followed by formation of the *h*1-*h*2 set, or vice versa (see Fig. 5 *A*). The same observation can be made for the other plots involving set *h*1-*h*2. Such pathway ambiguity is absent for the case of doppel (Fig. 5 *B*), where trajectories can be grouped essentially in a single set. Indeed, for doppel, on average, the formation of the *h*1-*h*2 contacts follows the other sets. This result is further confirmed by the analysis of the time-dependent average of the reaction coordinates Q_i values as obtained from the 100 nonequilibrium folding runs. Fig. 6, *A* and *B* show the results for PrP^c and *Dpl*, respectively. For *Dpl* the order in the formation of the contacts is the same detected by the equilibrium run: set *Bh* is formed before set *Bh*-*h*1, which is followed by *Bh*-*h*2, eventually followed by *h*1-*h*2 interhelical contacts. However, such a definite succession of events is not observed for prion. It is also interesting to check whether the nonequi-

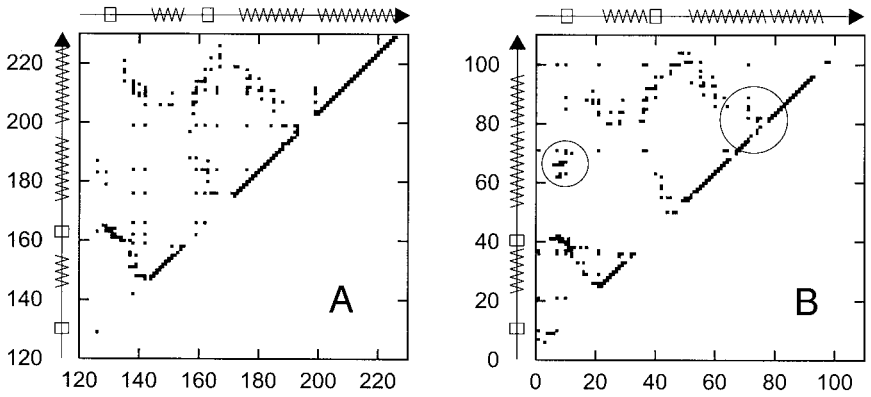

FIGURE 4 The contact maps of (*A*) cellular prion and (*B*) doppel computed according to the definition of contact given in the Methods section. On the edge of the squares, the secondary structures are reported. Circled in (*B*) are the contacts that mainly differ in the two structures. These result from the different packing of the β-sheet on the terminal helices (in prion mainly mediated by the C-terminal β-strand, while in doppel by the N-terminal one) and by the different angle between the terminal helices (in doppel they are almost parallel, making contacts all along their length, while in prion they are less parallel and interact more specifically in their middle).

librium data are consistent with the two-state characteristics as shown in the free energy landscape. Fig. 7 shows the time dependence of the partial fraction of the native contacts, Q', obtained by averaging over 100 nonequilibrium runs for *PrP*. Q' is computed by considering 40 native contacts with the highest SRSH values. It is shown that the data are remarkably well-fitted by a single exponential function. Such a behavior is an indispensable property of a two-state system with a single barrier in the free energy.

DISCUSSION

The goal of our numerical study was to characterize the folding process of the normal isoform of the prion protein,

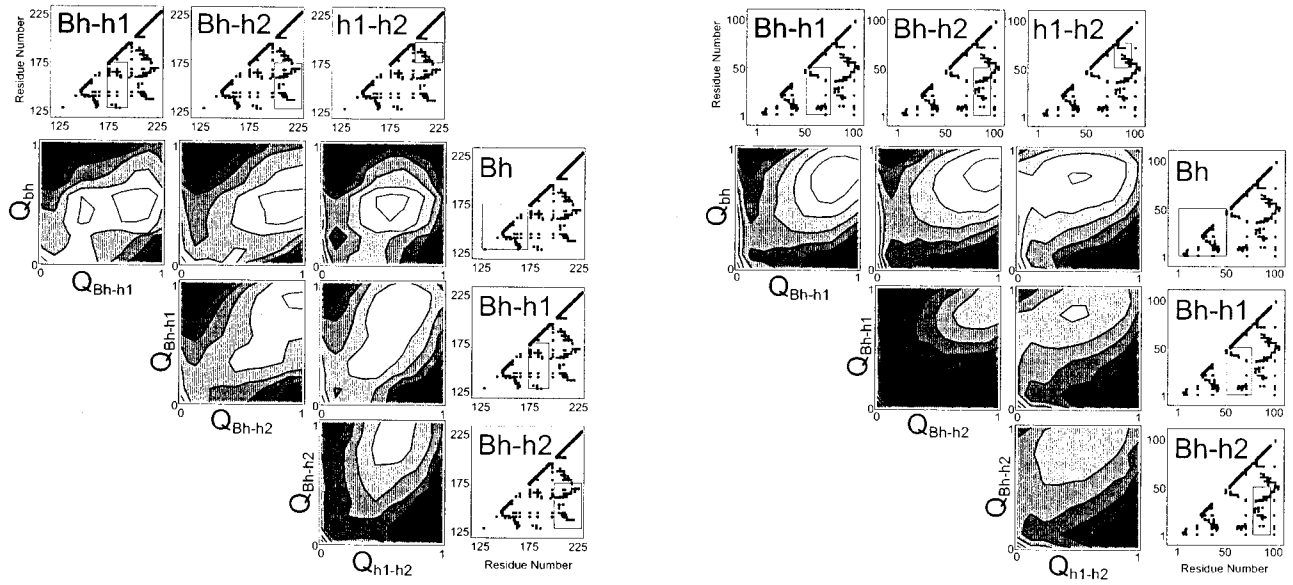

FIGURE 5 Effective free energy landscape for prion (*A*) and doppel (*B*) model. Upper and right squares represent the contact maps where the representative set of contacts defining the reaction coordinates are highlighted (see Methods section). Contour plots are shown for each of the six pairs of reaction coordinates (Q_i, Q_j). White (black) spots represent low (high) values of the free energy. The minimum close to (0, 0) corresponds to the unfolded state, whereas the minimum closer to (1, 1) corresponds to the folded state. In most of the contour plots for prion (*A*) more than one class of favorable paths from the unfolded to the folded conformation is possible, whereas for doppel (*B*) only one class of paths is the most favorable. Following these pathways allows assigning a preferential order to the events. Thus, on average, the sequence of contact formation for doppel is *Bh*, *Bh*-*h*1, *Bh*-*h*2, and *h*1-*h*2 (see also Fig. 6). For prion, two classes of equivalent pathways are present: in one case the *h*1-*h*2 contacts are formed before the other sets of contacts, in the other case *h*1-*h*2 contacts follow the formation of all the other contacts, as for *Dpl*.

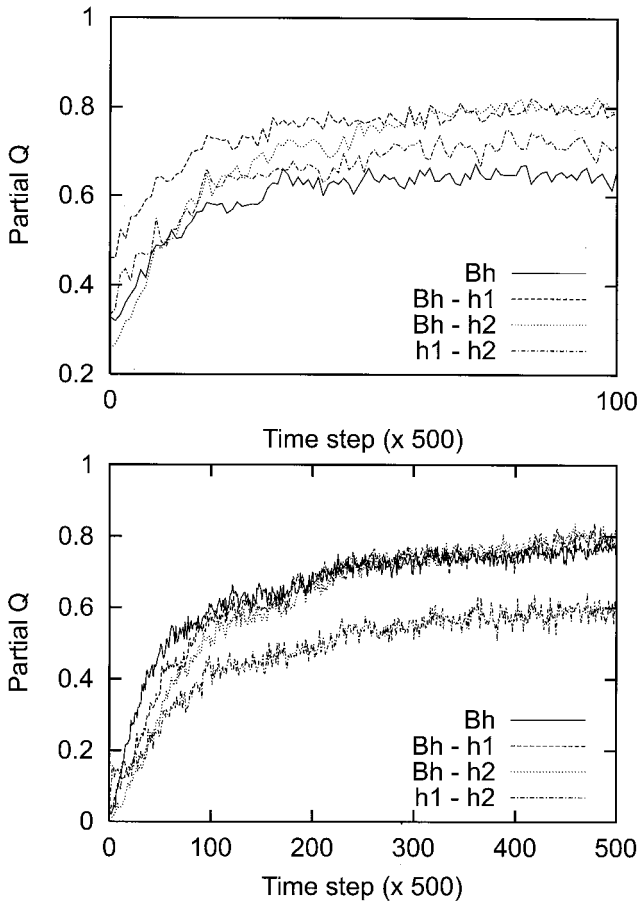

FIGURE 6 Average time dependence of reaction coordinates Q_i values for prion (*A*) and doppel (*B*) computed on 100 folding runs starting from completely unfolded conformations. (*A*) In the case of prion the curves corresponding to the different sets (*solid line* for *Bh*, *dashed line* for *Bh-h*1, *dotted line* for *Bh-h*2, and *dash-dotted line* for *h*1-*h*2) present intersections. Indeed, the average over the different folding pathways results in a non-well-defined sequence of folding events. (*B*) The folding events in the case of doppel protein have a well-defined sequence: the *Bh* set (*solid line*) is the first to form, then *Bh-h*1 (*dashed line*), then *Bh-h*2 (*dotted line*), and finally *h*1-*h*2 (*dash-dotted line*).

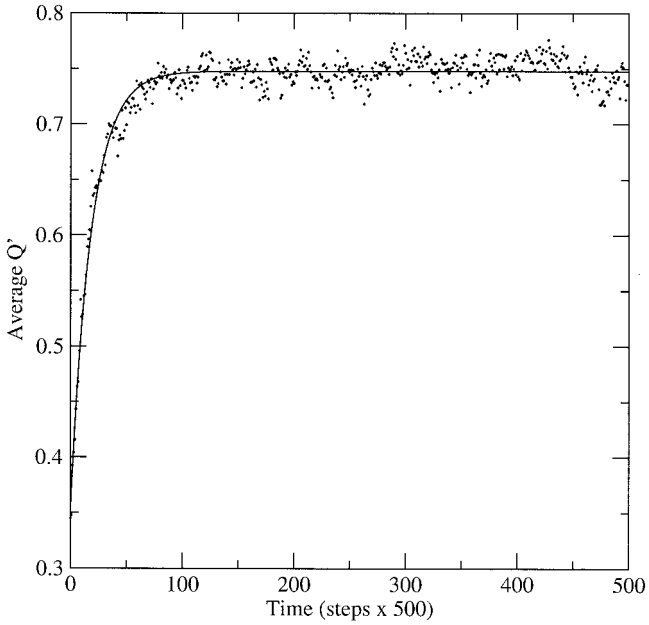

FIGURE 7 Time dependence of the partial fraction of native contacts, Q', computed by averaging over 100 independent folding runs starting from unfolded conformations for *PrP*. The data are fitted with a single exponential function of the form $f(t) = f(0) \cdot e^{-k_F \cdot t} + f(\infty) \cdot (1 - e^{-k_F \cdot t})$. The correlation coefficient found for the fit is 0.98.

pointing out some putative structural mechanisms responsible for the propensity of PrP^c to mutate into the harmful scrapie form, PrP^{sc}.

At the heart of our analysis is a topology-based energy function that, by rewarding the formation of native-like interactions among the residues, allows mimicking the build-up of native structure in a folding process. The progress toward the native state occurs through overcoming a configurational barrier, which is responsible for the appearance of a marked peak in the specific heat curve. The folding transition revealed by this peak has well-defined two-state characteristics, consistent with several experimental results (Wildegger et al., 1999; Baskakov et al., 2001; Glockshuber, 2001). Crucial contacts (and sites) for the folding process have thus been identified as those giving the largest contributions to the specific heat peak (Fig. 2).

The top sites isolated by our procedure contain a highly significant fraction of residues that are known to enhance the misfolding propensity of PrP^c. This led us to conclude that the mutations that led to the disease, affecting the topologically relevant positions that we have pointed out, most probably result in a destabilization of the native PrP^c folding process as opposed to other factors that concur in determining the emergence of the disease (i.e., an enhancement of the aggregation propensity or of the stability of the PrP^{sc} isomer (Baskakov et al., 2001)). This indicates that there is little freedom in choosing the chemical identity of the amino acids at the sites that have a special structural role during the folding process (Settanni et al., 2001).

The analysis presented here, based on topological arguments, provides a physically appealing interpretation for the connection between certain site mutations and the arousal of the prion disease. Conversely, the key mutating sites not captured by our analysis are presumably crucial because they affect some aspects of the process leading to the scrapie form that cannot be accounted for by the present method. For example, one of the mechanisms leading to misconformation is the template binding of PrP^c to PrP^{sc} involving the interconversion of the former into the latter. Consistent with this interpretation, amino acids that do not present high SRSH values and that are known to be involved in the development of the disease, specifically Met-129, Glu-200, and Glu-219, are mostly charged or polar residues with a large exposed surface area.

Despite the topological similarities of *Dpl* and PrP^c, their different spatial organization of secondary motifs can result in significant differences of their folding pathways, as revealed by our topology-based approach. Strikingly, for PrP^c the folding pathways can be classified into two main, distinct routes, whereas for *Dpl* only one of these is essentially present. For the latter, the series of folding steps is more distinctly marked than for the former. Such noticeably different behavior can be ascribed to the different set of contacts stabilizing the assembly of the α-helices and β-sheets in the two proteins. It is interesting to note that the relative contact order (Plaxco et al., 1998) of PrP^c is 0.2056 and is lower than that of *Dpl*, which is 0.2298, even though the former has a somewhat higher number of contacts which is, in turn, reflected by the shift of the folding transition peak toward higher temperatures (see Fig. 1). This difference could explain the different average folding time measured in our models for the two proteins. Indeed, according to Fig. 6, the sets of contacts in prion are all formed within the first 25,000 time steps, while contacts in doppel can take four to five times longer to form. This difference is of the same order of magnitude as that predicted by Plaxco (Plaxco et al., 1998) on the basis of the relative contact order of the two proteins. This finding is in agreement with experimental evidence for the fast folding kinetics of cellular prion (Wildegger et al., 1999).

Although the present picture would certainly be influenced by the introduction of amino acid-specific interactions, it is appealing to connect the different folding routes to the appearance of misfolded conformers. According to this view, the misfolded structures would result from following the folding route of PrP^c where the inter C-terminal helices contacts (set *h*1-*h*2) are formed while the contacts with the β-sheet (set *Bh*-*h*1 and *Bh*-*h*2) are not yet formed (Figs. 5 *A* and 6 *A*). Indeed, along this route, which is alternative to the one where the two events are interchanged as it occurs in the single route of doppel (Figs. 5 *B* and 6 *B*), the N-terminal part of the protein is free to rearrange its structure before reaching the native state. This scenario is consistent with several recent models (see, e.g., Peretz et al., 1997; Huang et al., 1996) that identify the N-terminal part of the protein as the one that undergoes the conformational change that leads the scrapie form of prion.

On the basis of native-state hydrogen exchange experiments, in a recent study Nicholson et al. (2002) have pointed out a fundamental difference in the free energy landscape of *PrP* and *Dpl*. In particular, the hydrogen exchange technique allows concluding that *PrP* folds through a partially structured state (not an intermediate) in which the region containing the two C-terminal helices is formed while the rest is largely unstructured. However, they did not detect this partially structured form in the case of *Dpl*. This suggested that the partial structure state of *PrP* may act as an important factor in the ability of prion to convert from PrP^c to PrP^{sc}. These experimental findings and their implications are in remarkable agreement with the results presented in our study. They also confirm that the topological features of *PrP* and *Dpl* are the underlying reasons that make the two proteins behave differently. [We became aware of the results by Nicholson et al. after the first revision of the present paper.]

REFERENCES

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

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

Baskakov, I. V., G. Legname, S. B. Prusiner, and F. E. Cohen. 2001. Folding of prion protein to its native alpha-helical conformation is under kinetic control. *J. Biol. Chem.* 276:19687–19690.

Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy minimization, and dynamics calculations. *J. Comp. Chem.* 4:187–217.

Calzolai, L., D. A. Lysek, P. Guntert, C. von Schroetter, R. Riek, R. Zahn, and K. Wuthrich. 2000. NMR structures of three single-residue variants of the human prion protein. *Proc. Natl. Acad. Sci. U.S.A.* 97:8340–8344.

Cecconi, F., C. Micheletti, P. Carloni, and A. Maritan. 2001. Molecular dynamics studies on HIV-1 protease drug resistance and folding pathways. *Proteins.* 43:365–372.

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

DeLano, W. L. 2002. The PyMOL Molecular Graphics System. DeLano Scientific, San Carlos, CA, USA. http://www.pymol.org.

Dlouhy, S. R., K. Hsiao, M. R. Farlow, T. Foroud, P. M. Conneally, P. Johnson, S. B. Prusiner, M. E. Hodes, and B. Ghetti. 1992. Linkage of the Indiana kindred of Gerstmann-Straussler-Scheinker disease to the prion protein gene. *Nat. Genet.* 1:64–67.

Ferrara, P., and A. Caflisch. 2001. Native topology or specific interactions: what is more important for protein folding? *J. Mol. Biol.* 306:837–850.

Ferrenberg, A. M., and R. H. Swendsen. 1989. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.* 63:1195–1198.

Gabizon, R., H. Rosenmann, Z. Meiner, I. Kahana, E. Kahana, Y. Shugart, J. Ott, and S. B. Prusiner. 1993. Mutation and polymorphism of the prion protein gene in Libyan Jews with Creutzfeldt-Jakob-disease (CJD). *Am. J. Hum. Genet.* 53:828–835.

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

Garcia, F. L., R. Zahn, R. Riek, and K. Wuthrich. 2000. NMR structure of the bovine prion protein. *Proc. Natl. Acad. Sci. U.S.A.* 97:8334–8339.

Glockshuber, R. 2001. Folding dynamics and energetics of recombinant prion proteins. *Adv. Protein Chem.* 57:83–105.

Go, N. 1983. Theoretical studies of protein folding. *Annu. Rev. Biophys. Bioeng.* 12:183–210.

Goldgaber, D., L. G. Goldfarb, P. Brown, D. M. Asher, W. T. Brown, S. Lin, J. W. Teener, S. M. Feinstone, R. Rubenstein, R. J. Kascsak, J. W. Boellaard, and D. C. Gajdusek. 1989. Mutations in familial Creutzfeldt-Jakob disease and Gerstmann-Straussler-Scheinker's syndrome. *Exp. Neurol.* 106:204–206.

Gsponer, J., and A. Caflisch. 2001. Role of native topology investigated by multiple unfolding simulations of four SH3 domains. *J. Mol. Biol.* 309:285–298.

Hao, M. H., and H. A. Scheraga. 1997. On foldable protein-like models: a statistical-mechanical study with Monte Carlo simulations. *Physica A.* 244:124–146.

Hoang, T. X., and M. Cieplak. 2000a. Molecular dynamics of folding of secondary structures in Go-like models of proteins. *J. Chem. Phys.* 112:6851–6862.

Hoang, T. X., and M. Cieplak. 2000b. Sequencing of folding events in Go-type proteins. *J. Chem. Phys.* 113:8319–8328.

Hsiao, K., H. F. Baker, T. J. Crow, M. Poulter, F. Owen, J. D. Terwilliger, D. Westaway, J. Ott, and S. B. Prusiner. 1989. Linkage of a prion protein missense variant to Gerstmann-Straussler syndrome. *Nature.* 338: 342–345.

Huang, Z., S. B. Prusiner, and F. E. Cohen. 1996. Structures of prion proteins and conformational models for prion diseases. *Fold. Des.* 1:13–19.

James, T. L., H. Liu, N. B. Ulyanov, S. Farr-Jones, H. Zhang, D. G. Donne, K. Kaneko, D. Groth, I. Mehlhorn, S. B. Prusiner, and F. E. Cohen. 1997. Solution structure of a 142-residue recombinant prion protein corresponding to the infectious fragment of the scrapie isoform. *Proc. Natl. Acad. Sci. U.S.A.* 94:10086–10091.

Kaya, H., and H. S. Chan. 2000a. Energetic components of cooperative protein folding. *Phys. Rev. Lett.* 85:4823–4826.

Kaya, H., and H. S. Chan. 2000b. Polymer principles of protein calorimetric two-state cooperativity. *Proteins.* 40:637–661.

Kretzschmar, H. A., G. Honold, F. Seitelberger, M. Feucht, P. Wessely, P. Mehraein, and H. Budka. 1991. Prion protein mutation in family 1st reported by Gerstmann, Straussler, and Scheinker. *Lancet.* 337: 1160–1160.

Liu, A. Z., R. Riek, G. Wider, C. von Schroetter, R. Zahn, and K. Wuthrich. 2000. NMR experiments for resonance assignments of C-13, N-15 doubly-labeled flexible polypeptides: application to the human prion protein hPrP(23–230) *J. Biomol. NMR.* 16:127–138.

Maritan, A., C. Micheletti, and J. R. Banavar. 2000a. Role of secondary motifs in fast folding polymers: a dynamical variational principle. *Phys. Rev. Lett.* 84:3009–3012.

Maritan, A., C. Micheletti, A. Trovato, and J. R. Banavar. 2000b. Optimal shapes of compact strings. *Nature.* 406:287–290.

Martinez, J. C., and L. Serrano. 1999. The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nat. Struct. Biol.* 6:1010–1016.

Micheletti, C. 2002. Optimal prediction of folding rates and transition state placement from native state geometry. *Cond. Mat.* 0202090.

Micheletti, C., J. R. Banavar, and A. Maritan. 2001. Conformations of proteins in equilibrium. *Phys. Rev. Lett.* 87:088102.

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

Mo, H. P., R. C. Moore, F. E. Cohen, D. Westaway, S. B. Prusiner, P. E. Wright, and H. J. Dyson. 2001. Two different neurodegenerative diseases caused by proteins with similar structures. *Proc. Natl. Acad. Sci. U.S.A.* 98:2352–2357.

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

Nicholson, E. M., H. Mo, S. B. Prusiner, F. E. Cohen, and S. Marqusee. 2002. Differences between the prion protein and its homolog doppel: a partially structured state with implications for scrapie formation. *J. Mol. Biol.* 316:807–815.

Ozkan, S. B., I. Bahar, and K. A. Dill. 2001. Transition states and the meaning of Φ-values in protein folding kinetics. *Nat. Struct. Biol.* 8:765–769.

Pan, K. M., M. Baldwin, J. Nguyen, M. Gasset, A. Serban, D. Groth, I. Mehlhorn, Z. W. Huang, R. J. Fletterick, F. E. Cohen, and S. B. Prusiner. 1993. Conversion of α-helices into β-sheets features in the formation of the scrapie prion proteins. *Proc. Natl. Acad. Sci. U.S.A.* 90: 10962–10966.

Pearlman, D. A., D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. Debolt, D. Ferguson, G. Seibel, and P. Kollman. 1995. AMBER, a package of computer-programs for applying molecular mechanics, normal-mode analysis, molecular-dynamics and free-energy calculations to simulate the structural and energetic properties of molecules. *Comp. Phys. Commun.* 91:1–41.

Peretz, D., R. A. Williamson, Y. Matsunaga, H. Serban, C. Pinilla, R. Bastidas, R. Rozenshteyn, T. L. James, R. A. Houghten, F. E. Cohen, S. B. Prusiner, and D. R. Burton. 1997. A conformational transition at the N terminus of the prion protein features in formation of the scrapie isoform. *J. Mol. Biol.* 273:614–622.

Pergami, P., H. Jaffe, and J. Safar. 1996. Semipreparative chromatographic method to purify the normal cellular isoform of the prion protein in nondenatured form. *Anal. Biochem.* 236:63–73.

Petersen, R. B., M. Tabaton, L. Berg, B. Schrank, R. M. Torack, S. Leal, J. Julien, C. Vital, B. Deleplanque, W. W. Pendlebury, D. Drachman, T. W. Smith, J. J. Martin, M. Oda, P. Montagna, J. Ott, L. Autiliogambetti, E. Lugaresi, and P. Gambetti. 1992. Analysis of the prion protein gene in thalamic dementia. *Neurology.* 42:1859–1863.

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

Poulter, M., H. F. Baker, C. D. Frith, M. Leach, R. Lofthouse, R. M. Ridley, T. Shah, F. Owen, J. Collinge, J. Brown, J. Hardy, M. J. Mullan, A. E. Harding, C. Bennett, R. Doshi, and T. J. Crow. 1992. Inherited prion disease with 144 base pair gene insertion. 1. Genealogical and molecular studies. *Brain.* 115:675–685.

Prusiner, S. B. 1982. Novel proteinaceous infectious particles cause scrapie. *Science.* 216:136–144.

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

Sakaguchi, S., S. Katamine, N. Nishida, R. Moriuchi, K. Shigematsu, T. Sugimoto, A. Nakatani, Y. Kataoka, T. Houtani, S. Shirabe, H. Okada, S. Hasegawa, T. Miyamoto, and T. Noda. 1996. Loss of cerebellar Purkinje cells in aged mice homozygous for a disrupted Prp gene. *Nature.* 380:528–531.

Settanni, G., A. Cattaneo, and A. Maritan. 2001. Role of native-state topology in the stabilization of intracellular antibodies. *Biophys. J.* 80:2935–2945.

Tsai, J., R. Taylor, C. Chothia, and M. Gerstein. 1999. The packing density in proteins: standard radii and volumes. *J. Mol. Biol.* 290:253–266.

Wildegger, G., S. Liemann, and R. Glockshuber. 1999. Extremely rapid folding of C-terminal domain of the prion protein without kinetic intermediates. *Nat. Struct. Biol.* 6:550–553.

Zahn, R., A. Z. Liu, T. Luhrs, R. Riek, C. von Schroetter, G. F. Lopez, M. Billeter, L. Calzolai, G. Wider, and K. Wuthrich. 2000. NMR solution structure of the human prion protein. *Proc. Natl. Acad. Sci. U.S.A.* 97:145–150.