Structural Domains and Main-Chain Flexibility in Prion Proteins[†]

N. Blinov,[‡,§] M. Berjanskii,[||] D. S. Wishart,[‡,⊥] and M. Stepanova*[,‡]

National Institute for Nanotechnology NRC, Edmonton, Alberta T6G 2M9, Canada, and Departments of Mechanical Engineering, Computing Sciences, and Biological Sciences, University of Alberta, Edmonton, Alberta, Canada

Received November 3, 2008; Revised Manuscript Received December 22, 2008

ABSTRACT: In this study we describe a novel approach to define structural domains and to characterize the local flexibility in both human and chicken prion proteins. The approach we use is based on a comprehensive theory of collective dynamics in proteins that was recently developed. This method determines the essential collective coordinates, which can be found from molecular dynamics trajectories via principal component analysis. Under this particular framework, we are able to identify the domains where atoms move coherently while at the same time to determine the local main-chain flexibility for each residue. We have verified this approach by comparing our results for the predicted dynamic domain systems with the computed main-chain flexibility profiles and the NMR-derived random coil indexes for human and chicken prion proteins. The three sets of data show excellent agreement. Additionally, we demonstrate that the dynamic domains calculated in this fashion provide a highly sensitive measure of protein collective structure and dynamics. Furthermore, such an analysis is capable of revealing structural and dynamic properties of proteins that are inaccessible to the conventional assessment of secondary structure. Using the collective dynamic simulation approach described here along with a high-temperature simulations of unfolding of human prion protein, we have explored whether locations of relatively low stability could be identified where the unfolding process could potentially be facilitated. According to our analysis, the locations of relatively low stability may be associated with the β-sheet formed by strands S1 and S2 and the adjacent loops, whereas helix HC appears to be a relatively stable part of the protein. We suggest that this kind of structural analysis may provide a useful background for a more quantitative assessment of potential routes of spontaneous misfolding in prion proteins.

Prion proteins (PrPs)[1] are highly conserved cellular glycoproteins that reside at the surface of the cellular membrane (*1*). The functions of prion proteins are largely unknown, although there is good evidence of their role in the transfer or transport of copper ions (*2*). Prion proteins are also known to be the causative agents for a number of neurodegenerative diseases. According to the protein only hypothesis (*1*), misfolded prion proteins are the sole infectious agents responsible for the development of bovine spongiform encephalopathy (mad-cow disease), Creutzfeldt−Jakob disease, chronic wasting disease, and other transmissible spongiform encephalopathies in animals and humans. Conversion of the normal cellular isoform of the prion protein (PrP^C) into the misfolded scrapie isoform (PrP^{Sc}) leads to substantially reduced protein solubility and may be accompanied by the formation of amyloid fibrils. The insolubility of misfolded prion proteins makes it very challenging to prepare samples and, as a result, to study their structure by such standard techniques of protein structure determination as X-ray crystallography and solution NMR spectroscopy. As a result, the high-resolution molecular structure of PrP^{Sc} as well as the pathways leading to conversion from the normal PrP^C to the infectious scrapie isoform remains largely unknown.

Despite the above-mentioned difficulties, many experimental facts about the structure of misfolded prion proteins, amilogenic prion oligomers, and amyloid fibrils have been well documented. For instance, circular dichroism experiments have revealed that the conversion from the cellular to the scrapie isoform is accompanied by a reduction in α-helical content and an increase in β-sheet content of the protein. The importance of intermolecular cross-β binding in the structure of PrP^{Sc} aggregates and amyloid fibrils has also been well established by X-ray scattering experiments (*3*). However, the actual regions of the prion protein that are involved in these interactions are still a matter of some dispute within the prion community. Recent experimental results using electron paramagnetic resonance and hydrogen/deuterium exchange (*4, 5*) suggest that for the recombinant prion protein, for example, the C-terminal part of the protein may be involved in the formation of the fibril core. At the same time, many existing theoretical models (*6*−*9*) assume that the enrichment in β-structure and formation of the protofibril core are initiated largely at the N-terminus.

[†] Funded by the Alberta Prion Research Institute.

* To whom correspondence should be addressed. Phone: 780-641-1717. Fax: 780-641-1601. E-mail: maria.stepanova@nrc-cnrc.gc.ca.

[‡] National Institute for Nanotechnology NRC.

[§] Department of Mechanical Engineering, University of Alberta.

[||] Department of Computing Sciences, University of Alberta.

[⊥] Department of Biological Sciences, University of Alberta.

[1] Abbreviations: BFGS, Broyden−Fletcher−Goldfarb−Shanno (minimization); BRMB, Biological Magnetic Resonance Bank; MD, molecular dynamics; NMR, nuclear magnetic resonance; PDB, Protein Data Bank; PrP, prion protein; PCA, principal component analysis; RCI, random coil index; SPC, simple point charge.

10.1021/bi802043h CCC: $40.75
Published on Web 01/29/2009

Theoretical models and numerical simulations are potentially capable of providing crucial information that will eventually lead to a better understanding of the molecular mechanisms behind prion diseases, such as the conversion of the normal prion protein into the infectious scrapie isoform. In particular, valuable information on misfolding of prion proteins can be obtained using a variety of all-atom molecular dynamics (MD) methods (*7, 8, 10*−*22*). However, the major theoretical challenges are the large size of the system of interest and the long time scale of the unfolding processes, which reaches milliseconds regimes and beyond. Furthermore, formation of a mature prion fibril may take hours and days (*23, 24*), if not even longer. Since all-atom molecular dynamics simulation beyond the nanosecond time scale is a computationally very demanding task, simulating the prion protein conversion process is simply too difficult to model from the first principles with today's computers. At the same time, it is certainly quite feasible to monitor the early stages of prion protein misfolding via molecular dynamics simulations. For instance, such MD simulations have demonstrated that prion proteins show a tendency for decreasing α-helical content and expanding its β-structure under simulated conditions that are known to promote aggregation *in vitro* (*7, 8, 11, 12, 17*−*22*). The β-sheet-rich misfolded states of prion proteins detected by MD can be regarded as possible intermediates on the pathway to the scrapie isoform, and thus, they can provide important information about the process of misfolding. A solution that we would like to explore is development of robust numeric approaches that can use relatively short molecular dynamics simulations (at the nanosecond time scale) to forecast structural transitions of prion proteins on much longer time scales. Such an approach would be crucial for understanding the conversion of prion proteins. Further progress in this area must be made in order to help to identify the protein regions that may be involved in stabilization of unfolding intermediates and formation of fibril core.

Numeric analysis of collective dynamic behaviors is one of the most promising approaches being developed toward identifying large-scale structural and longer term dynamic trends in proteins. The development of techniques to analyze collective (or essential) protein dynamics dates back to the early 1990s (*25*−*27*). In this method, collective coordinates are first derived by the principal component analysis (PCA) of the protein's MD trajectories. These components are then ranked according to the mean square displacement along them, after which a truncated set of essential collective coordinates is identified. The amplitudes of these concerted motions can be further analyzed by calculating the average displacements along the essential coordinates. It is generally thought that the essential degrees of freedom describe the most pronounced collective motions of the protein. These approaches have been applied toward the characterization of peptide flexibility by calculating the average displacements along the essential coordinates (*28, 29*). They have also been used for noise filtering to improve the identification of domains of collective motions (*30*).

Recently, a novel theoretical framework has been developed for building coarse-grained models of protein dynamics by employing the essential collective coordinates derived from short productive MD runs (*31*). These coordinates are used to construct the Mori projection operator, which is further employed to derive the respective generalized Langevin equations describing the dynamics of the protein. This particular formalism rigorously identifies the relationship between the essential coordinates and the protein's major dynamic properties, such as local flexibility. Furthermore, the formalism allows for a rigorous and dynamically consistent definition of collective motion "domains". The regions of rigidity and flexibility predicted by the theory (*31*) have been shown to match dynamic data generated from NMR experiments. This particular approach provides an advanced extension of the conventional interpretation of NMR data based on the molecular dynamics simulations (*32*).

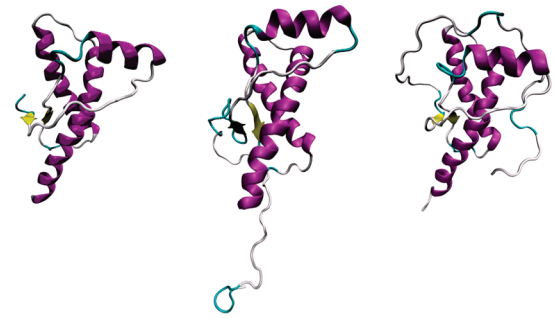

FIGURE 1: Model systems for human prion protein: model 1, short construct 125−230; models 2 and 3, longer constructs 89−228 and 90−230, respectively.

In the work presented here, we have applied the concept of essential dynamics toward modeling the dynamic domain structure of human and chicken prion proteins. For this purpose, we conducted short (2−3 ns) MD calculations and then used the resulting trajectories to analyze the local flexibility and identify the structural domains based on the general dynamic theory (*32*). We validated our model predictions by comparing the calculated results with the experimentally measured protein flexibility as derived by the random coil index (RCI) and NMR chemical shifts (*33*). We also explore the relevance of our results to prion protein misfolding and their potential to identify intermediates in the PrP^{C} to PrP^{Sc} conversion process.

METHODS

Molecular Dynamics Protocol. MD simulations and analysis were performed with Gromacs 3.2.1 (*34*) using the GROMOS96 43a1 force field (*35*). For the chicken prion protein, residues 12−242 (PDB ID 1U3M (*36*)) were used. For human prion proteins, three model constructs were employed as illustrated in Figure 1. In model 1, the folded structure encompassing residues 125−238 was used as described in the Protein Data Bank (PDB ID 1QM3 (*37*)), whereas in models 2 and 3, the construct PrP 88−230 (PDB ID 1QM1 (*37*)) was employed with the coordinates of the residues from 125 to 228 taken from the PDB file. Model 2 represents the sequence PrP 89−228 with a torsion angle based model of the 89−124 region. The torsion angles for the 89−125 region were predicted from NMR chemical shifts (BMRB ID 4434) (*37*) using the Preditor program (*38*) and used to build a protein model with our in-house program Pepmake (*39*). Steric clashes between regions 89−124 and 125−228 were initially eliminated by adjusting the torsion angles of residues 123−125 with the program MOLMOL

(*40*). Further model optimization was done using Gromacs 3.2.1 with the GROMOS96 43a1 force field (*34*). In model 3, the sequence 90−230 was constructed with the coordinates of the missing atoms from the disordered regions being built using the DiscoveryStudio 2.0/Biopolymer (Accelrys Inc.). Then, both regions 90−124 and 229−230 were optimized with the program MODELLER (*41, 42*). Figure 1 shows the resulting secondary structures.

Each protein was placed in a triclinic box with a distance between the protein and box edges of 1.5 nm and then solvated with SPC water molecules. If ions were present in original PDB models, they were also included in the starting models for MD simulations. Counterions (Na^+ or Cl^-) were added to the system to adjust the net charge of the system to zero. Solvent was minimized before addition of ions using 100 steps of a steepest descent algorithm followed by 100 steps of Broyden−Fletcher−Goldfarb−Shanno (*43*) (BFGS) minimization. Positional restraints with a force constant of 1×10^5 kJ mol^{-1} nm^{-2} were applied to all protein atoms during the minimization to prevent distortion of protein structure by nonequilibrated solvent. After the counterions had been added to the system, six cycles of short steepest descent and BFGS minimizations were performed with decreasing position restraints on protein atoms ($K_{posre} = 1 \times 10^5$, 1×10^4, 1000, 100, 10, and 0 kJ mol^{-1} nm^{-2}). Three 5 ps equilibration MD steps were performed. During the first step, positional restraints of 1×10^5 kJ mol^{-1} nm^{-2} were applied to all protein atoms. The second and the third steps were carried out with positional restraints of the same strength on all backbone atoms and then on backbone atoms in secondary structure elements, respectively. The temperature of the protein and the solvent was maintained separately at 300 K by coupling the systems with Berendsen thermostats (*44*). The coupling time was 0.1 ps. The pressure was maintained at 1 atm with isotropic pressure coupling using the Berendsen algorithm, a time constant of 1 ps, and a compressibility of 4.5×10^{-5} bar^{-1}. The actual MD production run was computed for 4 ns. An integration step of 2 fs was used in all MD simulations. Bond lengths were restrained with the LINCS algorithm (*45*) using a fourth order expansion, four iterations during minimization and two iterations during MD simulations. The radius for calculating short-range electrostatic interactions was 1.2 nm. Long-range electrostatic interactions were treated with a particle-mesh Ewald summation (*46*) that was performed with maximum spacing for the FFT grid of 0.12 nm and cubic interpolation. The neighbor list within the radius of 1.0 nm was updated every 20 fs. The cutoff radius for calculation of van der Waals interactions was 1.1 nm. Simulations of prion protein unfolding were conducted at 500 K using the aforementioned protocol.

Identification of Structural Domains and Characterization of Local Flexibility. The basic theory of essential collective protein dynamics (*31*) describes a macromolecule by a set of generalized Langevin equations derived through the Mori projection operator technique (*47*). The essential collective coordinates derived from MD trajectories by principal component analysis (*48*)

$$\vec{E}^k = \{E_1^k,\ E_2^k,\ ...,\ E_{3N}^k\} \quad k = 1, ..., K \tag{1}$$

are employed as the dynamic variables in the Mori projection formalism. In eq 1, $\vec{E}^k$ represent the normalized eigenvectors in the 3*N*-dimentional configuration space of the protein. *N* is the total number of atoms, and *K* is the number of essential coordinates that sample 90−95% of the total displacement (usually, $K = 10-30$). The values E_n^k represent the direction cosines of the eigenvectors $\vec{E}^k$ in the configuration space of the protein. These values can also be viewed as the projections of the eigenvectors $\vec{E}^k$ onto the Cartesian degrees of freedom of individual atoms. According to the theory (*31*), the degrees of freedom for which the direction cosines E_n^k adopt similar magnitudes for all essential coordinates *k*, $E_n^k \approx E_m^k$, are coupled dynamically. This leads to a robust and straightforward definition of dynamic domains of correlated motion, where atoms move coherently. Such domains can be defined as groups of atoms with similar values E_n^k. Unlike other existing approaches to identify protein domains (*30, 49*), no *a priori* assumption regarding the number of domains, their elementary building blocks, or interatomic interactions are made. Furthermore, no additional noise reduction is required, because the clustering is performed in the space of essential collective motions, where random fluctuations are eliminated. Since the essential dynamic formalism is based on a rigorous theoretical background, the outcomes are straightforwardly interpretable in terms of the collective dynamic properties of the protein.

In the framework of the theory (*31*), the dynamic domains are identified by the nearest-neighbor clustering (*50*) in the 3*K*-dimensional space of directional cosines of essential coordinates, where atoms are represented by *N* points with the coordinates E_i^k. A natural interpretation of these "domains of correlated motion" is that they identify regions of relative rigidity in the protein. The only variable parameter in the theory is the interdomain distance, *d*, which defines the maximum distance in the 3*K*-dimensional space, for the corresponding atoms to belong to the same domain (*51*). The identification of the domains is sensitive to the selection of *d*. In the limit of vanishing *d*, the number of atoms in a given domain decreases, and the domains tend to degenerate into individual atoms. On the other hand, for large *d*, all of the atoms in the protein belong to one large domain. Thus, the parameter *d* can be understood as a fundamental variable defining the level of correlation or "rigidity" within the dynamic domains. In section 3 it is demonstrated how the value of *d* can be selected in a particular protein in order to provide the most informative structural score, as well as how the relative rigidities in a protein can be compared by varying the value of *d*.

In this paper, we determine and analyze the regions of local flexibility in human and chicken prion proteins under thermally denaturating conditions. Furthermore, we extend the theory (*31*) to characterize the local flexibility of intermolecular bonds within the same theoretic framework. For this purpose, we introduce a quantitative descriptor of the relative local flexibility, *F*, as the distance in the 3*K*-dimentional space between the point representing the atom for which the flexibility is to be evaluated and some reference point. In this work, the reference point is defined as a centroid calculated over the coordinates of all α-carbon atoms. The resulting coordinates of the reference point in the 3*K*-dimensional space are given by

$$\varepsilon^\kappa = \frac{1}{N_{C_\alpha}} \sum_{i=1}^{N_{C_\alpha}} E_i^\kappa \tag{2}$$

where the sum is taken over the coordinates of all α-carbon atoms and N_{C_α} is their total number. The point with the coordinates $\{\varepsilon^1, \varepsilon^2, ..., \varepsilon^{3K}\}$ represents the motion of the entire protein molecule. We define our descriptor of the main-chain flexibility for α-carbon atoms in each residue

$$F_{C_\alpha}(i) = \sum_{\kappa=1}^{3K} (E_i^\kappa - \varepsilon^k)^2 \quad (3)$$

where i denote α-carbon atoms and E_i^κ are the corresponding projections of essential eigenvectors. Obviously, a wide range of choices can be introduced for the flexibility descriptor. The one given by eqs 2 and 3 characterizes the deviation in the motion of each α-carbon atom with respect to the averaged motion of the entire main chain. Note that the descriptor F does not depend on the interdomain parameter d. A low value for F can be interpreted as a strong correlation with motion of the entire main chain, e.g., low flexibility, whereas high F values identify more mobile and/or flexible locations. In this work, we obtain the flexibility descriptors for each α-carbon atom and normalize the dependence $F(i)$ so that F does not exceed 1. The resulting descriptor of relative local rigidity is used to analyze the dynamics behaviors of the prion proteins in this study.

Random Coil Index Protocol. We verified our theoretic approach by comparing the aforementioned descriptor of relative local flexibility with protein flexibility determined from experimental chemical shifts using the random coil index or RCI (*52, 53*). The RCI describes the chemical shift "proximity" of the polypeptide structure and dynamics to that observed of a random coil. It was optimized to correlate with amplitudes of MD motions on a picosecond to nanosecond time scale. However, chemical shifts that are used by the RCI protocol can be sensitive to conformational exchange on a much longer time scale (microsecond to millisecond). This makes the RCI theoretically capable of detecting motions over a very large range of frequencies (*52*), including those typical for some collective, segmental motions in proteins.

More specifically, we compared our MD-derived descriptor of the relative local flexibility with the RCI as follows: Sets of experimental chemical shifts with BMRB accession numbers 4379 (human prion protein 121−230), 4434 (human prion protein 90−230), and 6269 (chicken prion protein) were selected for calculating the RCI of human and chicken prion proteins, respectively. The NMR assignments were rereferenced by a previously described procedure (*53*). Standardized random coil chemical shifts (*54*) were determined based on protein primary sequences and corrected with neighboring residue correction factors for the $i \pm 1$ and $i \pm 2$ residues (*55*). The adjusted reference chemical shifts were subtracted from the experimental chemical shifts according to previously published methods (*52, 53*). Figure 2 compares the RCI profile for the folded part of human prion protein 90−230 with that for 121−230. As anticipated, the two flexibility profiles are very close. For this reason, in the following discussion we employ the RCI profile for PrP 121−230 for comparison with our numerical results.

RESULTS AND DISCUSSION

Dynamic Structural Domains Identify Rigid Regions in Prion Proteins. We start with the identification of the dynamic domains based on the procedure described in section 2. For this, we employ 0.2 ns segments of MD trajectories containing 2000 snapshots. These trajectories were processed by PCA taking into account all atoms. For calculating the essential collective coordinates, 20 principal components with the highest eigenvalues were used that sample more than 90% of the total displacement. The direction cosines of the collective eigenvectors $\vec{E}^k$ have been represented by N points, each corresponding to an individual atom. In this space (*56*), points that are located close to each other correspond to a strong correlation in the directions of motion of the corresponding atoms. To obtain dynamic domains, the N points were clustered using the nearest-neighbor technique. No structural property, such as the number of domains, was assumed *a priori*. The only restriction in the formalism is that the domains are assumed to contain more than two atoms. Figure 3 shows the number of domains identified by this technique as a function of the interdomain distance d for human prion protein, model 1. The complementary information, the total number of atoms in all domains N_{tot}, the number of atoms in the largest domain N_{max}, and their difference $N_{tot} - N_{max}$ are also represented in panels a and b of Figure 4, respectively. It is clearly seen from these figures that the set of domains is sensitive to the selection of the

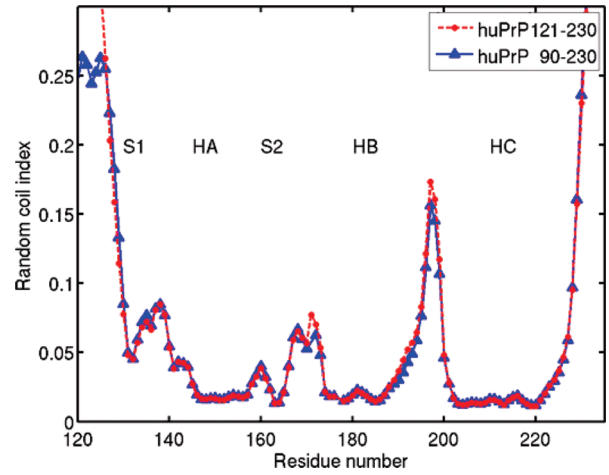

FIGURE 2: RCI profiles for human prion protein PrP 121−230 (red) and the folded part of PrP 90−230 (blue). The letters indicate the locations of β-sheets S1 and S2 and α-helices HA, HB, and HC.

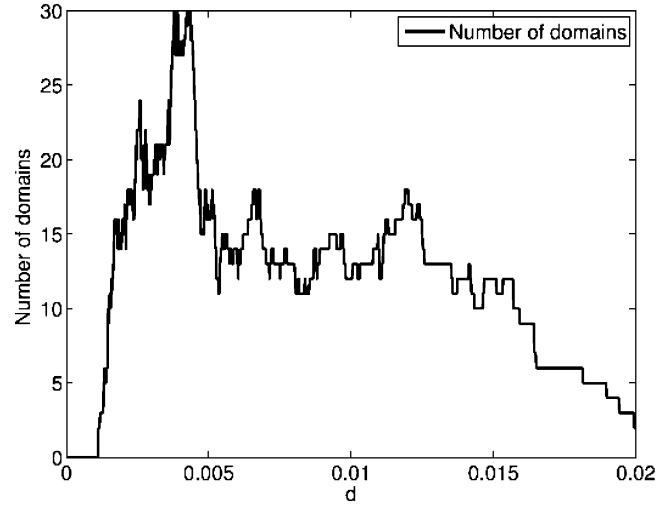

FIGURE 3: Number of dynamic domains in the human prion protein (model 1) as a function of the dimensionless interdomain distance d (*51*).

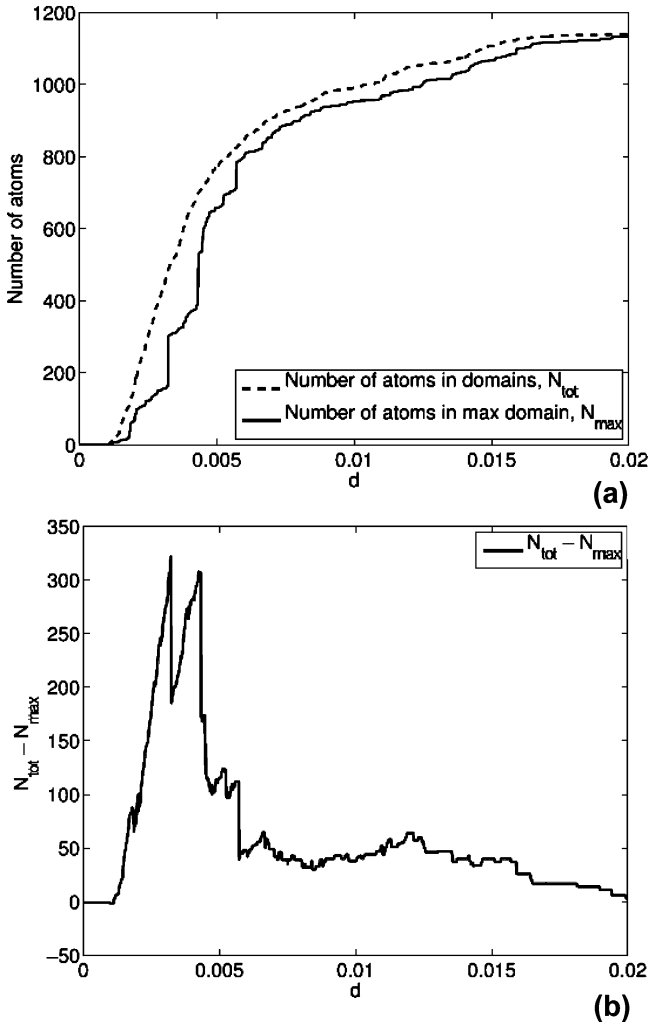

FIGURE 4: Total number of atoms in all domains N_{tot} and number of atoms in the largest domain N_{max} in human prion protein (model 1) (a) and the difference $N_{tot} - N_{max}$ (b) shown as functions of the dimensionless interdomain distance d for the human prion protein (*51*).

distance d. Thus, no correlated domains containing more than two atoms can be identified below a minimum distance $d_{min} \approx 0.001$, whereas most of the macromolecule is recognized as a single domain beyond a maximum distance $d_{max} \approx 0.017$. The most informative breakdown of the molecule into domains is reached in the interval $d = 0.002-0.005$, where the dependencies in Figures 3 and 4b reach their peak values. For illustration, in Figure 5 we show the dynamic domains in human PrP, model 1, calculated for values of d equal to 0.002, 0.003, 0.0035, and 0.004. The domains are highlighted with different colors. Blue corresponds to the largest domain followed by red, green, and yellow in order of decreasing domain size. Only the four largest domains are shown. The parts of the protein outside of the four largest domains are colored in gray. It can be seen that the overall number of atoms in the domains, as well as the size of the largest domain, increases with d in accordance with Figure 4. Another important result that emerges from Figure 5 is that the domains identified by this process form compact groups of atoms partly coinciding with some elements of the secondary structure, although the formalism used in the domain identification does not assume any proximity in the atoms' locations in the primary, secondary, or tertiary

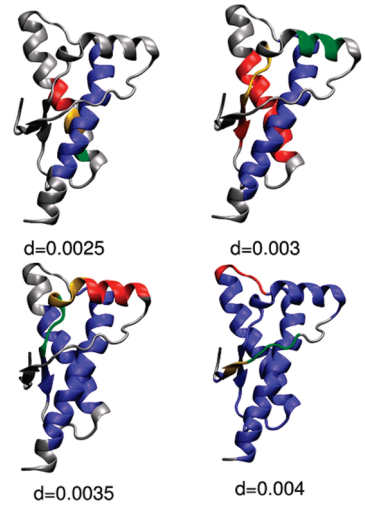

FIGURE 5: Four largest dynamic domains identified in the human prion protein (model 1) for different interdomain distances. The blue color corresponds to the largest domain; it is followed by red, green, and yellow in order of decreasing domain size.

structure. By the definition, proximity in the $3K$-dimensional space of the essential collective coordinates reveals only the correlations in atomic motion. The fact that these correlations identify compact groups of atoms confirms the consistency of our clustering formalism.

As mentioned in section 2, the largest dynamic domains identify the largest regions of relative rigidity in a protein, whereas off-domain regions, or regions containing small domains, are relatively soft or flexible. Analyzing the suite of domain systems by varying the interdomain distance d, as shown in Figure 5, reveals the relative flexibilities (or stabilities) of various parts of the protein. For example, the domains that are identified with relatively low values of d indicate the most rigid and stable parts of the molecule. Thus, from the domain system shown in Figure 5 (for $d = 0.0025$) it follows that the most rigid and stable part is the helix HC. In contrast for high d values, regions that do not belong to any large domain identify the locations of enhanced softness or increased flexibility. According to the domain system for $d = 0.004$ seen in Figure 5, relatively flexible regions are found in β-strand S1 and the loops connecting α-helix HA with the rest of the molecule. The comparison of all the cases shown in Figure 5 indicates that the regions of relative rigidity and softness identified for different d values are consistent with each other, regardless of particular domain breakdowns arising from the choice of d. In the following discussion, we have analyzed mostly the domain system identified with $d = 0.0035$, since this d value provides the most detailed and informative structural score (see also Figures 3 and 4b). It should be noted however, that the particular choice of d does not affect the conclusions, since a variation of d within reasonable limits reveals similar structural behaviors.

In addition to the analysis of model 1, illustrated by Figure 5, we also analyzed the dynamic domains in models 2 and 3. The corresponding secondary structures for all models are shown in Figure 1. Figure 6 demonstrates the dynamic domains obtained for $d = 0.0035$ in the folded part of these longer constructs. Despite some difference in detail (discussed in the next section), the locations of the two largest

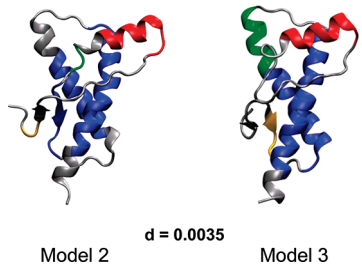

FIGURE 6: Four largest dynamic domains identified for $d = 0.0035$ in the human prion protein for models 1 and 2 (unfolded chains not shown).

domains (colored in blue and red) in models 2 and 3 are similar to those in model 1.

As we have demonstrated, the dynamic domains identify relatively large rigid segments in the macromolecule. However, the local rigidity/flexibility can also be characterized with atomic level resolution, if the flexibility descriptor introduced in section 2 is employed. Panels a, b, and c of Figure 7 show the flexibility descriptor F computed by eq 3 individually for each residue in models 1, 2, and 3, respectively. Note that low values of F identify relatively rigid locations. Clearly seen are the low values of F in the regions corresponding to α-helices HB and HC and to β-strand S2. In addition, residues that belong to the four largest domains identified for $d = 0.0035$ are shown in Figure 7 using the same coloring schemes as in Figures 5 and 6. As expected, the largest domain (shown in blue) identifies α-helices HB and HC in all three figures and also β-strand S2 in Figure 7a,b. In these regions, the F values are lower compared to the rest of the polypeptide. In contrast, relatively flexible regions that correspond to high F levels either do not belong to any domains or contain smaller domains such as the loop between HB and HC. Overall, Figure 7 demonstrates that the set of dynamically coupled domains identified by this reference-free clustering technique (*31*) is fully consistent with the complementary estimate of the local residue-level relative flexibility.

Our descriptor of the local flexibility given by eq 3 is representative of the level of correlation between the motion of individual C_α atoms and the entire main chain. It is interesting and informative to compare this descriptor with the experimental results obtained from an RCI analysis of the chemical shifts for these proteins. In Figure 8, our normalized F values computed for the same three models as in Figure 7, are compared with the RCI values calculated for the human prion protein. In all the cases considered, the predicted positions of major maxima and minima match remarkably well with the NMR results. There is a difference in the relative heights of individual maxima, which arises from a difference in the definition of the F and the RCI. However, despite this difference, the two methods show an excellent match in the locations of relative rigidity and softness.

The Flexibility Analysis of Human PrP 125−228 (Model 1) Is Representative of the Folded Part in Longer Constructs. Comparison of the results in Figures 5, 6, and 7 for the three different models indicates that the major motif elements of the folded structure in human prion protein are not sensitive to the model employed for the unstructured N-terminus. Thus, the three graphs in Figure 7 consistently demonstrate rigid and stable domains corresponding to helices HB and

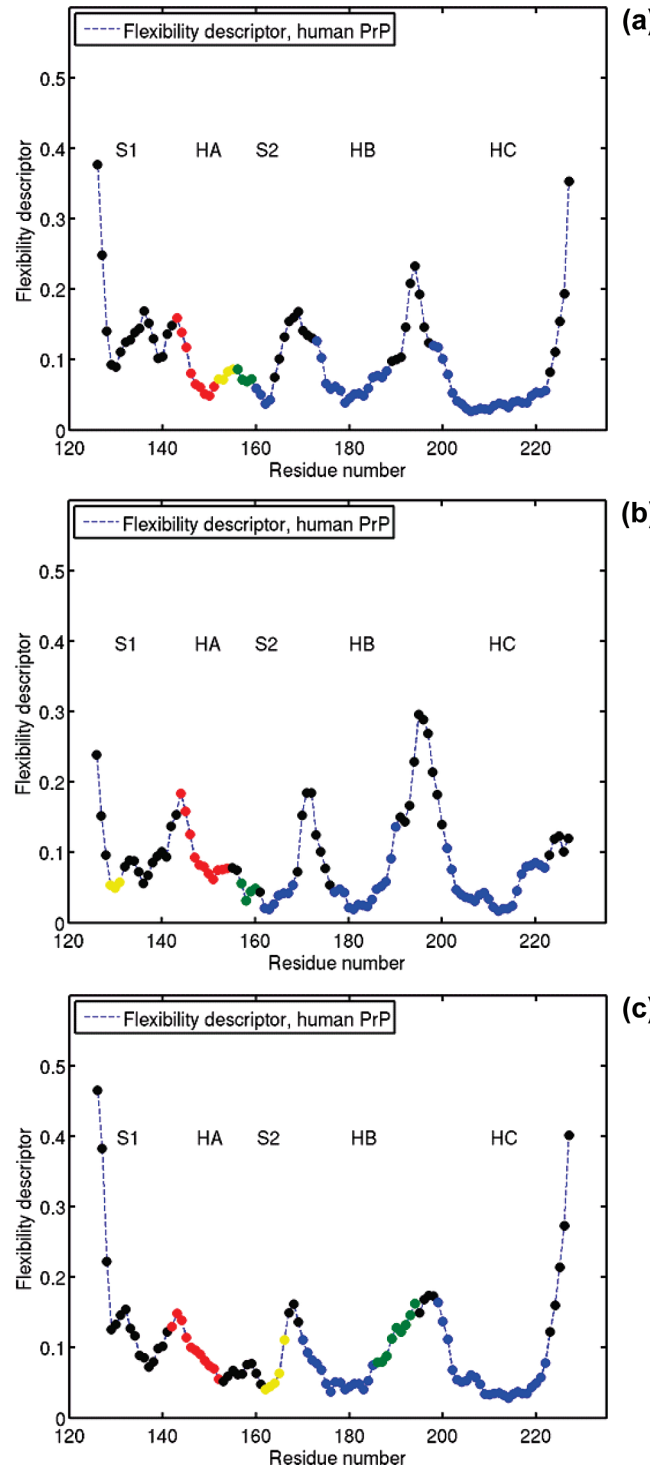

FIGURE 7: The predicted main-chain flexibility descriptor F as a function of the residue number for the human prion protein: (a) model 1; (b) model 2; (c) model 3. The colors indicate the four largest dynamic domains identified for $d = 0.0035$ (see Figures 5 and 6). Black points correspond to off-domain residues. The letters indicate the locations of β-sheets S1 and S2, and α-helices HA, HB, and HC.

HC neighboring the flexible loops between helices HB and HC and between β-strand S2 and helix HB. Also in all models, the second largest domain (red) is associated with helix HA. The corresponding F value tends to increase with decreasing residue number, indicating an increase in the mobility in the region of the loop connecting HA with β-strand S1, while helix HA maintains its structural identity.

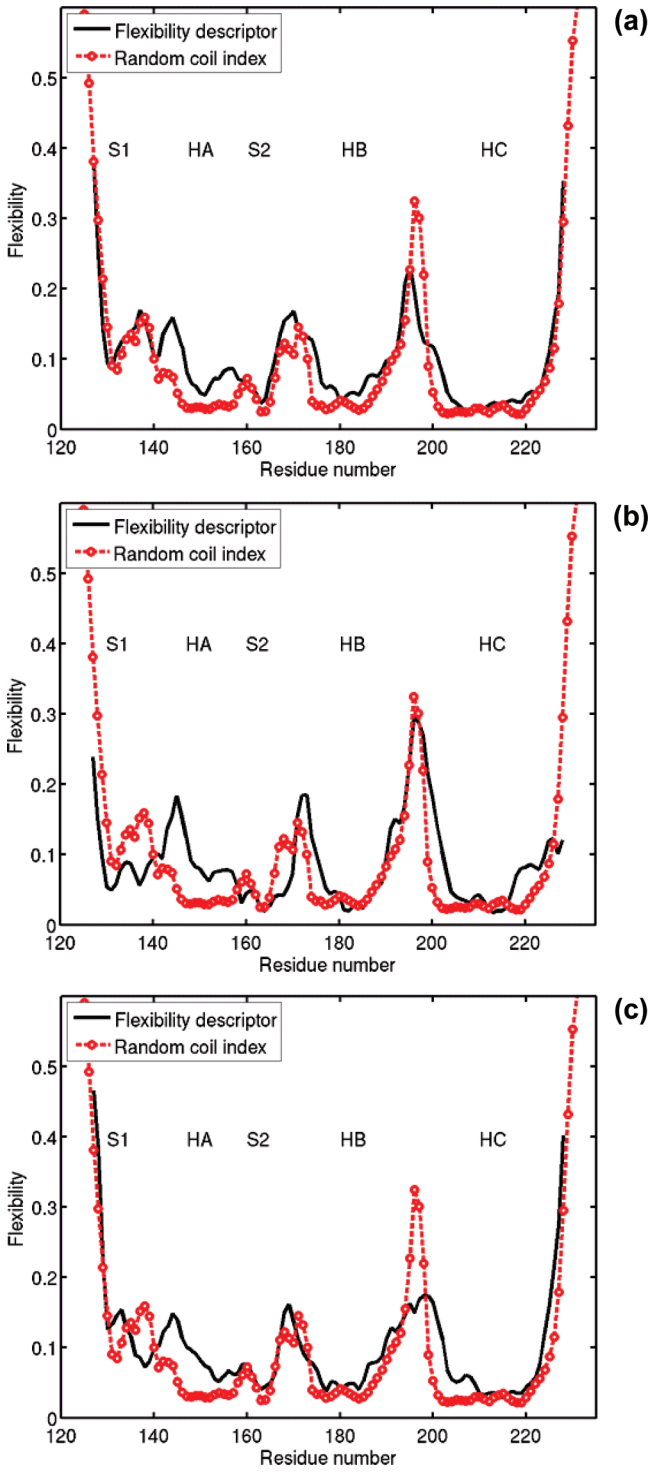

FIGURE 8: NMR-derived random coil index (RCI) for human PrP 121−230 (circles) and the normalized main-chain flexibility descriptor F (solid line) for model 1 (a), model 2 (b), and model 3 (c) as functions of the residue number.

The most affected region is, quite naturally, β-strand S1. Analysis of the F profiles in Figure 7a,b shows that the N-terminal region in model 1 is slightly more flexible than the corresponding region in model 2. However, Figure 7c shows that model 3 results in an even more flexible β-strand S1 than model 1. One can conclude that the unfolded chain must perturb the region around β-strand S1, depending on the conformation adopted by the chain. Other parts of the protein that appear to be in proximity of the unfolded chain are the C-terminus (model 2) and the loop between helices

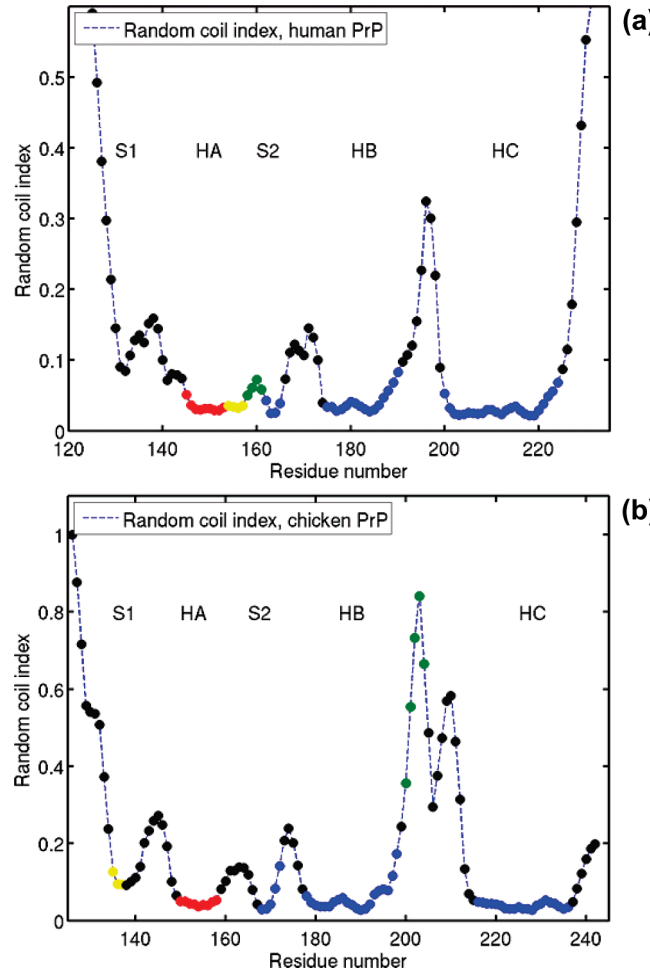

FIGURE 9: NMR-derived random coil index (RCI) with the four largest domains ($d = 0.0035$) indicated by colors locally for each residue: (a) for human prion protein, model 1; (b) for chicken prion protein PrP 128−242.

HB and HC (model 3). From Figure 7b,c, it can be seen that the corresponding F values are decreased, meaning that these locations are somewhat stabilized. Regardless of these differences, the overall shapes of the F profiles reveal similar structures in the three models being considered. This becomes particularly evident from the comparison with the RCI data in Figure 8. Thus in Figure 8a, which shows the F profile for model 1, the agreement with RCI in positions of major maxima and minima has not been impaired by the absence of the unfolded chain. We conclude that the structure and dynamics in PrP 125−228 are reasonably representative of what occurs in longer PrP constructs as well, so we will use this truncated model for our subsequent analysis.

The Domain Analysis Provides a Highly Sensitive Score of the Protein's Structure. As mentioned previously, the system of dynamic domains and the local flexibility descriptors provide complementary structural information. In particular, the F profiles evaluate local residue-specific flexibility relative to a reference, whereas the dynamic domain system provides a reference-free coarse-grained assessment of the major regions of relative rigidity and flexibility. These assertions can be tested by comparing them against experimental NMR data. In this regard, comparisons with the NMR-derived RCI values should be particularly revealing. In Figure 9, the RCI profiles for human and chicken prion proteins are compared with the corresponding domain

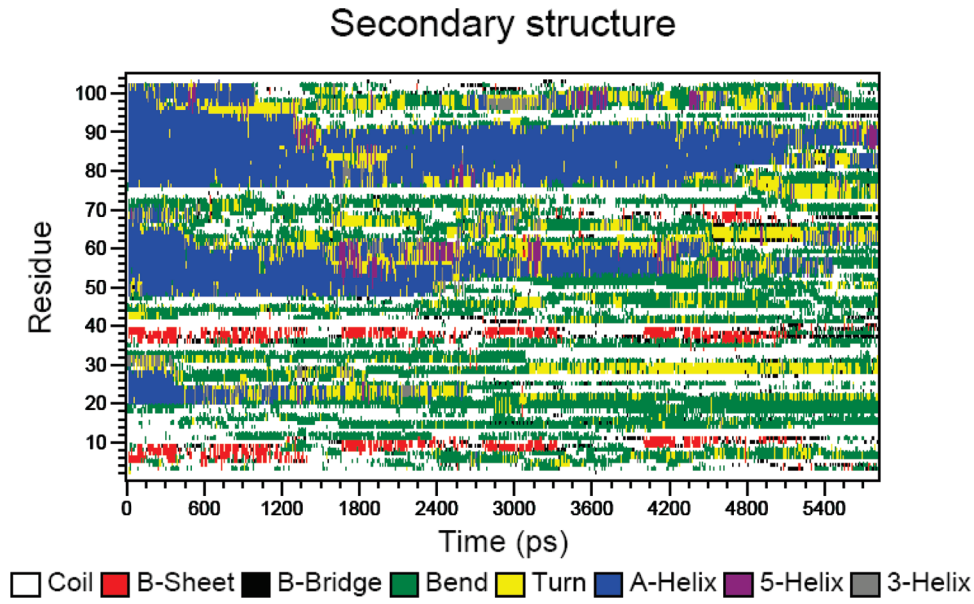

FIGURE 10: Evolution of the secondary structure in the course of the high-temperature MD simulation of the unfolding of the human prion protein (model 1).

systems identified for $d = 0.0035$. In both cases, the largest domain (shown in blue) matches very well with the observed regions of enhanced rigidity. These correspond to α-helices HB and HC and β-strand S2, whereas the second largest domain features helix HA. The primary difference between the human and chicken prion proteins is in the length of loop HB−HC and in the locations of smaller domains (shown by green and yellow). In human PrP^C, both domains are located between HA and S2, whereas in chicken PrP^C, one domain features β-strand S1 and another is located between helices HB and HC. In contrast, the levels of RCI or F do not show this difference. On the basis of these results, our conclusion is that the dynamic domain decomposition is a highly sensitive characteristic of the protein's structure and reveals subtle features that are inaccessible to other methodologies. In the next section, we discuss in more detail the dynamic domain decomposition in the context of unfolding of human prion protein.

The Dynamic Domain System Might Help To Indicate the Route of Prion Unfolding. Analysis of the domain systems identified with different interdomain distances d (shown in Figure 5) reveals important dynamic information. First, note that helix HC belongs to the largest domain in every case. This clearly indicates that HC is the most rigid part of the protein. For small d, helix HB falls apart into several dynamic domains. However, when d increases, these small domains merge together and form a single domain, first with β-strand S2 and then, for larger d, with helix HC as well as with β-strand S2 but not with HA. An interpretation of this is that helices HB and HC show correlated displacements, whereas helix HA tends to follow a different dynamic path. The fact that helices HB and HC show correlated motions can be explained by the presence of the disulfide bond between Cys-179 and Cys-214. In contrast, the dynamic specificity of β-strand S1 is a less obvious feature. It is known that β-strands S1 and S2 form a rather stable (at the time scale of simulations) β-sheet, whereas our domain analysis demonstrates that these strands are dynamically uncoupled (see Figure 5). This means that nonspecific van der Waals interactions between the side chains of the residues from helix HB and β-strand S1 may prevail over the network of the hydrogen bonds between S1 and S2.

An important implication of the results described here is that they might indicate the presence of a region of instability in the prion protein with an enhanced propensity for misfolding. This region is the β-sheet formed by strands S1 and S2 and two loops connecting the strands with helix HA. Conceivably, this could be a region of dynamic instability in this protein. One can expect that, in constructs longer than those shown in Figure 5, these regions may be even more disturbed, whereas helix HC remains stable. It should also be noted that we have employed relatively short MD trajectories. No actual unfolding/misfolding has been detected over this short time. Our analysis only reveals the early precursors of the possible PrP^C unfolding process.

A complementary insight into the pathways and mechanisms of prion protein unfolding/misfolding can be obtained from short molecular dynamics simulations performed at high temperature (*17*−*22*). Such an approach has been used to study the folding and unfolding of a number of different proteins (*57*). In Figure 10 we present the evolution of the secondary structure over the molecular dynamics trajectory for the human prion protein carried out at high temperature (500 K). Figure 11 provides snapshots illustrating the various stages of the unfolding. Evidently, a strong correlation exists between our predictions based on the dynamic domain decomposition in Figure 5 and the evolution of the secondary structure seen in the course of the high-temperature simulations in Figure 11. For example, helix HC shows no sign(*12*s of unfolding over this short time period. Furthermore, helix HC tends to maintain most of its structure over longer time simulations. In the snapshot for 5.8 ns shown in Figure 11, a partly unfolded helix HC still can be identified, although other elements of the secondary structure have completely degraded. Such a behavior is in agreement with our domain analysis, which suggests that HC is the most rigid part of the protein. In contrast, helices HA and HB are partly unfolded after a 2.7 ns molecular dynamics run, which correlates with the dynamic information we have inferred using the domain decomposition with relatively low d values.

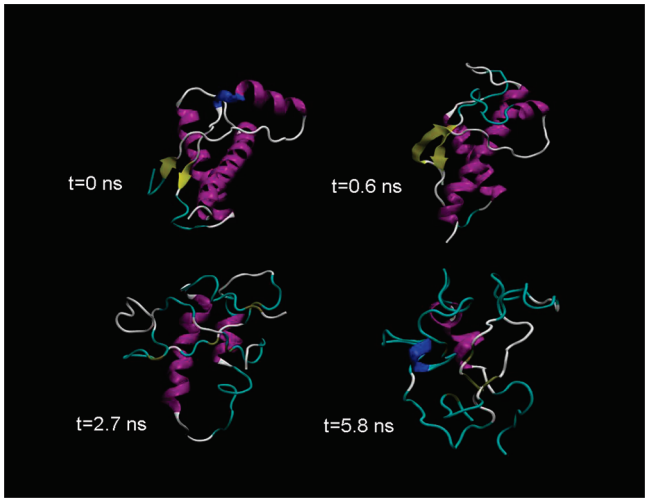

FIGURE 11: Snapshots of the high-temperature molecular dynamics calculations illustrating the different stages of the unfolding process over the trajectory from Figure 10.

Interestingly, the β-sheet structure does not unfold over these shorter times. In fact, it even extends after 0.6 ns but degrades completely after 2.7 ns, as can be clearly seen from Figure 11. This observation may have important implications for understanding the early stages of the prion misfolding and aggregation process. Such a modification of the secondary structure (even corresponding to an initial increase of the length of the β-strand) correlates with our finding that the β-sheet domain and the adjacent parts of this protein appear to be dynamically unstable. This is also in agreement with a previouly published MD study of the high-temperature unfolding of prion proteins (*18, 19*). Indeed, an elongation of the β-sheets was also observed in refs *20*−*22*. At the same time, in some studies (*20, 22*) an enhanced stability of helix HA was detected relative to the other elements to the secondary structure. Such contradictory conclusions from different MD studies can be partly attributed to the strong nonergodicity of the system under consideration (*21*). In our MD simulatuions, such a behavior was found to be relatively atypical and occurred in rare cases. At the same time we note, as it has been found recently) in long MD simulations on helical fragments from the mouse prion protein, that helix HA is indeed relatively stable while helices HB and HC can undergo transition from α-helical to β/random coil confromation.

In conclusion, our analysis of the dynamic domain structure of human prion proteins suggests that α-helix HC is the most stable part of the protein. This observation agrees with a direct simulation of prion protein unfolding at high temperature. In contrast, the β-sheet formed by strands S1 and S2 and the adjacent loops can be considered as dynamically unstable and may be potential locations facilitating the processes of unfolding and misfolding. These results suggest that spontaneous unfolding of the native prion protein would tend to occur at the N-terminus.

SUMMARY

In this report we have used relatively short molecular dynamics simulations (at the nanosecond time scale) to define dynamic structural domains and to characterize the local flexibility in both human and chicken prion proteins. This was done with the expectation that these properties may define slow motions occurring at time scales beyond nanoseconds. Our approach is based on a comprehensive theory of protein collective dynamics (*31*) which extracts essential collective coordinates derived from MD trajectories by PCA. Under this theoretical framework, we are able to identify the protein domains where atoms move coherently and to characterize the local main-chain flexibility for each residue. We have verified our results by comparing the structures of the predicted dynamic domain systems and the computed main-chain flexibility profiles with the NMR-derived (experimental) random coil indexes for both human and chicken PrP^C. The three sets of data show excellent agreement. Further, we expect that the dynamic domains calculated by this technique provide a highly sensitive measure of the protein's correlated structure and dynamics and that such analysis is capable of revealing the dynamic behaviors that are inaccessible by the conventional assessment of the secondary structure.

In addition to this work, we have also compared the dynamic domain structure with a high-temperature MD simulation of the unfolding of human PrP^C. These complementary studies have identified the β-sheet formed by strands S1 and S2 and the adjacent loops as the locations of relatively low stability. In contrast, helix HC was found to be the most stable part of the protein. We believe that these structural properties may be indicative of potential routes for the unfolding and subsequent misfolding in prion proteins. These results also suggest that spontaneous unfolding may be initiated at the N-terminus of native prion proteins.

ACKNOWLEDGMENT

The authors thank A. Kovalenko for helpful discussions. Figures 1, 5, 6, and 11 were created by using the VMD software (*58*).

REFERENCES

1. Prusiner, S. B. (1998) Prions. *Proc. Natl. Acad. Sci. U.S.A. 95*, 13363–13383.
2. Millhauser, G. L. (2007) Copper and the prion protein: Methods, structures, function, and disease. *Annu. Rev. Phys. Chem. 58*, 299–320.
3. Sunde, M., L., Serpell, L. C., Bartlam, M., Fraser, P. E., Pepys, M. B., and Blake, C. C. F. (1997) Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J. Mol. Biol. 273*, 729–739.
4. Cobb, N. J., Sonnichsen, F. D., Mchaourab, H., and Surewicz, W. K. (2007) Molecular architecture of human prion protein amyloid: A parallel, in-register β-structure. *Proc. Natl. Acad. Sci. U.S.A. 104*, 18946–18951.
5. Lu, X., Wintrode, P. L., and Surewicz, W. K. (2007) β-sheet core of human prion protein amyloid fibrils as determined by hydrogen/deuterium exchange. *Proc. Natl. Acad. Sci. U.S.A. 104*, 1510–1515.
6. Govaerts, C., Wille, H., Prusiner, S. B., and Cohen, F. E. (2004) Evidence for assembly of prions with left-handed β-helices into trimers. *Proc. Natl. Acad. Sci. U.S.A. 101*, 8342–8347.
7. DeMarco, M., Silveira, J., Caughey, B., and Daggett, V. (2006) Structural properties of prion protein protofibrils and fibrils: An experimental assessment of atomic models. *Biochemistry 45*, 15573–15582.
8. DeMarco, M. L., and Daggett, V. (2004) From conversion to aggregation: Protofibril formation of the prion protein. *Proc. Natl. Acad. Sci. U.S.A. 101*, 2293–2298.
9. Yang, S., Levine, H., Onuchic, J. H., and Cox, D. L. (2005) Structure of infectious prions: stabilization by domain swapping. *FASEB J. 19*, 1778–1782.
10. Ding, F., LaRocque, J. J., and Dokholyan, N. V. (2005) Direct observation of protein folding, aggregation, and a prion-like conformational conversion. *J. Biol. Chem. 280*, 40235–40240.

11. DeMarco, M. L., and Daggett, V. (2007) Molecular mechanism for low pH triggered misfolding of the human prion protein. *Biochemistry 46*, 3045–3054.
12. Dima, R. I., and Thirumalai, D. (2004) Probing the instabilities in the dynamics of helical fragments from mouse PrP^C. *Proc. Natl. Acad. Sci. U.S.A. 101*, 15335–15340.
13. Langella, E., Improta, R., and Barone, V. (2004) Checking the pH-induced conformational transition of prion protein by molecular dynamics simulations: effect of protonation of histidine residues. *Biophys. J. 87*, 3623–3632.
14. De Simone, A., Dodson, G. G., Verma, C. S., Zagari, A., and Fraternali, F. (2005) Prion and water: Tight and dynamical hydration sites have a key role in structural stability. *Proc. Natl. Acad. Sci. U.S.A. 102*, 7535–7540.
15. De Simone, A., Spadaccini, R., Temussi, P. A., and Fraternali, F. (2006) Toward the understanding of MNEI sweetness from hydration map surfaces. *Biophys. J. 90*, 3052–3061.
16. De Simone, A., Zagari, A., and Derreumaux, P. (2007) Structural and hydration properties of the partially unfolded states of the prion protein. *Biophys. J. 93*, 1284–1292.
17. Langella, E., Improta, R., Crescenzi, O., and Barone, V. (2006) Assessing the acid-base and conformational properties of histidine residues in human prion protein (125−228) by means of pK_a calculations and molecular dynamics simulations. *Proteins: Struct., Funct., Bioinf. 64*, 167–177.
18. Shamsir, M. S., and Dalby, A. R. (2007) β-sheet containment by flanking prolines: Molecular dynamic simulations of the inhibition of β-sheet elongation by proline residues in human prion protein. *Biophys. J. 92*, 2080–2089.
19. Shamsir, M. S., and Dalby, A. R. (2005) One gene, two diseases and three conformations: Molecular dynamics simulations of mutants of human prion protein at room temperature and elevated temperatures. *Proteins: Struct., Funct., Bioinf. 59*, 275–290.
20. Sekijima, M., Motono, C., Yamasaki, S., Kaneko, K., and Akiyama, Y. (2003) Molecular dynamics simulation of dimeric and monomeric forms of human prion protein: Insight into dynamics and properties. *Biophys. J. 85*, 1176–1185.
21. Barducci, A., Chelli, R., Procacci, P., and Schettino, V. (2005) Misfolding pathways of the prion protein probed by molecular dynamics simulations. *Biophys. J. 88*, 1334–1343.
22. Gu, W., Wang, T., Zhu, J., Shi, Y., and Liu, H. (2003) Molecular dynamics simulation of the unfolding of the human prion protein domain under low ph and high temperature conditions. *Biophys. Chem. 104*, 79–94.
23. Apetri, A. C., Vanik, D. L., and Surewicz, W. K. (2005) Polymorphism at residue 129 modulates the conformational conversion of the D178N variant of human prion protein 90−231. *Biochemistry 44*, 15880–15888.
24. Klug, G. M. J. A., Losic, D., Subasinghe, S., Aguilar, M.-I. L., Martin, L., and Small, D. H. (2003) Amyloid protein oligomers induced by metal ions and acid pH are distinct from those generated by slow spontaneous ageing at neutral pH. *Eur. J. Biochem. 270*, 4282–4293.
25. Kitao, A., Hirata, F., and Go, N. (1991) The effects of solvent on the conformation and the collective motions of protein: Normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum. *Chem. Phys. 158*, 447–472.
26. Garcia, A. E. (1992) Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett. 68*, 2696–2699.
27. Amadei, A. A., Linsenn, B. H., and Berendsen, H. J. C. (1993) Essential dynamics of proteins. *Proteins 17*, 412–425.
28. Emberly, E. G., Mukhopadhyay, R., Wingreen, N. S., and Tang, C. (2003) Flexibility of α-helices: Results of a statistical analysis of database protein structures. *J. Mol. Biol. 327*, 229–237.
29. Emberly, E. G., Mukhopadhyay, R., Tang, C., and Wingreen, N. S. (2004) Flexibility of β-sheets: Principal component analysis of database protein structures. *Proteins: Struct., Funct., Bioinf. 55*, 91–98.
30. Zhuravleva, A., Korzhnev, D. M., Nolde, S. B., Kay, L. E., Arseniev, A. S., Billeter, M., and Orekhov, V. Yu. (2007) Propagation of dynamic changes in barnase upon binding of barstar: An NMR and computational study. *J. Mol. Biol. 367*, 1079–1092.
31. Stepanova, M. (2007) Dynamics of essential collective motions in proteins: Theory. *Phys. Rev. E 76*, 051918 (1−16).
32. Case, D. (2002) Molecular dynamics and NMR spin relaxation in proteins. *Acc. Chem. Res. 35*, 325–331.
33. Berjanskii, M. V., and Wishart, D. S. (2005) A simple method to predict protein flexibility using secondary chemical shifts. *J. Am. Chem. Soc. 127*, 14970–14971.
34. Lindahl, E., Hess, B., and van der Spoel, D. (2001) GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model. 7*, 306–317.
35. Scott, W., Hunenberger, P. H., Tironi, I. G., Mark, A. E., Billeter, S. R., Fennen, J., Torda, A. E., Huber, T., Kruger, P., and van Gunsteren, W. F. (1999) The GROMOS biomolecular simulation program package. *J. Phys. Chem. A 103*, 3596–3607.
36. Calzolai, L., Lysek, D. A., Pérez, D. R., Güntert, P., and Wüthrich, K. (2005) Prion protein NMR structures of chickens, turtles, and frogs. *Proc. Natl. Acad. Sci. U.S.A. 102*, 651–655.
37. Zahn, R., Liu, A., Lührs, T., Riek, R., von Schroetter, C., López García, F., Billeter, M., Calzolai, L., Wider, G., and Wuthrich, K. (2000) NMR solution structure of the human prion protein. *Proc. Natl. Acad. Sci. U.S.A. 97*, 145–150.
38. Berjanskii, M. V., Neal, S., and Wishart, D. S. (2006) PREDITOR: A web server for predicting protein torsion angle restraints. *Nucleic Acids Res. 34*, W63–W69.
39. http://redpoll.pharmacy.ualberta.ca/pepmake.
40. Koradi, R., Billeter, M., and Wüthrich, K. (1996) MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graphics 14*, 51–55.
41. Fiser, A., Do, R. K., and Sali, A. (2000) Modeling of loops in protein structures. *Protein Sci. 9*, 1753–1773.
42. Fiser, A., and Sali, A. (2003) ModLoop: Automated modeling of loops in protein structures. *Bioinformatics, 19*, 2500–2501.
43. Liu, D. C., and Nocedal, J. (1989) On the limited memory BFGS method for large scale optimization. *Math. Program. 45*, 503–528.
44. Berendsen, H. J. C., Postma, J. P. M., Vangunsteren, W. F., Dinola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys. 81*, 3684–3690.
45. Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997) LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem. 18*, 1463–1472.
46. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An $N\cdot\log(N)$ method for Ewald sums in large systems. *J. Chem. Phys. 98*, 10089–10092.
47. (a) Mori, H. (1965) Transport, collective motion, and Brownian motion. *Prog. Theor. Phys. 33*, 423–425. (b) Mori, H. (1965) A continued-fraction representation of the time-correlation functions. *Prog. Theor. Phys. 34*, 399–416.
48. Tatsuoka, M. M. (1988) Multivariate Analysis, New York.
49. Yesylevsky, S. O., Kharkyanen, V. N., and Demchenko, A. P. (2006) Dynamic protein domains: Identification, interdependence, and stability. *Biophys. J. 91*, 670–685.
50. Jain, A. K., Murty, M. N., and Flynn, P. J. (1999) Data clustering: A review. *ACM Comput. Surv. 31*, 264–323.
51. Note that the interdomain distance d is a dimensionless value, as this follows from the definition of the metric in the $3K$-dimensional space of directional cosines of the essential collective degrees of freedom.
52. Berjanskii, M. V., and Wishart, D. S. (2008) Application of the random coil index to studying protein flexibility. *J. Biomol. NMR 40*, 31–48.
53. Berjanskii, M. V., and Wishart, D. S. (2006) NMR: Prediction of protein flexibility. *Nat. Protoc. 1*, 683–688.
54. Schwarzinger, S., Kroon, G. J., Foss, T. R., Wright, P. E., and Dyson, H. J. (2000) Random coil chemical shifts in acidic 8 M urea: Implementation of random coil shift data in NMRView. *J. Biomol. NMR 18*, 43–48.
55. Schwarzinger, S., Kroon, G. J. A., Foss, T. R., Chung, J., Wright, P. E., and Dyson, H. J. (2001) Sequence-dependent correction of random coil NMR chemical shifts. *J. Am. Chem. Soc. 123*, 2970–2978.
56. The dynamic domains have been identified in the $3K$-dimensional space of essential collective motions, as described in ref *31*.
57. Shea, J.-E., and Brooks, C. L., III. (2001) From folding theories to folding proteins: A review and assessment of simulation studies of protein folding and unfolding. *Annu. Rev. Phys. Chem. 52*, 499–535.
58. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. *J. Mol. Graphics 14*, 33–38.

BI802043H