Exploring the Propensities of Helices in PrP^C to Form β Sheet Using NMR Structures and Sequence Alignments

R. I. Dima and D. Thirumalai

Institute for Physical Science and Technology, and Department of Chemistry and Biochemistry, University of Maryland, College Park, Maryland 20742 USA

ABSTRACT Neurodegenerative diseases induced by transmissible spongiform encephalopathies are associated with prions. The most spectacular event in the formation of the infectious scrapie form, referred to as PrP^{Sc}, is the conformational change from the predominantly α-helical conformation of PrP^C to the PrP^{Sc} state that is rich in β-sheet content. Using sequence alignments and structural analysis of the available nuclear magnetic resonance structures of PrP^C, we explore the propensities of helices in PrP^C to be in a β-strand conformation. Comparison of a number of structural characteristics (such as solvent accessible area, distribution of (Φ, Ψ) angles, mismatches in hydrogen bonds, nature of residues in local and nonlocal contacts, distribution of regular densities of amino acids, clustering of hydrophobic and hydrophilic residues in helices) between PrP^C structures and a databank of "normal" proteins shows that the most unusual features are found in helix 2 (H2) (residues 172–194) followed by helix 1 (H1) (residues 144–153). In particular, the C-terminal residues in H2 are frustrated in their helical state. The databank of normal proteins consists of 58 helical proteins, 36 $\alpha+\beta$ proteins, and 31 β-sheet proteins. Our conclusions are also substantiated by gapless threading calculations that show that the normalized Z-scores of prion proteins are similar to those of other $\alpha+\beta$ proteins with low helical content. Application of the recently introduced notion of discordance, namely, incompatibility of the predicted and observed secondary structures, also points to the frustration of H2 not only in the wild type but also in mutants of human PrP^C. This suggests that the instability of PrP^C proteins may play a role in their being susceptible to the profound conformational change. Our analysis shows that, in addition to the previously proposed role for the segment (90–120) and possibly H1, the C-terminus of H2 and possibly N-terminus may play a role in the $\alpha \rightarrow \beta$ transition. An implication of our results is that the ease of polymerization depends on the unfolding rate of the monomer. Sequence alignments show that helices in avian prion proteins (chicken, duck, crane) are better accommodated in a helical state, which might explain the absence of PrP^{Sc} formation over finite time scales in these species. From this analysis, we predict that correlated mutations that reduce the frustration in the second half of helix 2 in mammalian prion proteins could inhibit the formation of PrP^{Sc}.

INTRODUCTION

Prions are infectious particles that are an abnormal isoform, PrP^{Sc}, of the normal host-encoded cellular prion protein PrP^C (Prusiner, 1997; Prusiner, 1998; Cohen, 1999). They are believed to be associated with neurodegenerative diseases in humans and many mammals, which are caused by transmissible spongiform encephalopathies (TSE). Examples of TSEs are familial Creutzfeldt–Jakob disease (CJD) in humans, scrapie of sheep, and bovine spongiform encephalopathy (BSE). Because no nucleic acid is implicated in the transformation from PrP^C to PrP^{Sc}, the "protein-only" hypothesis was proposed (Prusiner, 1997; Prusiner, 1998; Cohen, 1999), which argues that the same amino-acid sequence adopts two distinct structures. A wealth of experimental data on mammalian prions supports this hypothesis (Cohen and Prusiner, 1998). The protein-only hypothesis has also been demonstrated in recent studies on *Saccharomyces cerevisae* by Weissman and coworkers (Sparrer et al., 2000). They showed that introduction of preconverted

Sup35 leads to the formation of self-propagating [PSI$^+$], which apparently has prion-like characteristics.

The protein-only hypothesis implies that the conformational change leading to the PrP^{Sc} formation from the normal cellular form PrP^C may be spontaneous or might involve interactions with unidentified protein X (Telling et al., 1995). Although the mechanism of conversion of PrP^C into PrP^{Sc} and the tertiary structure of PrP^{Sc} are not known, numerous studies (Pan et al., 1993; Pergami et al., 1996) have shown that the two isoforms have identical amino-acid sequences. However, the structures of PrP^C and PrP^{Sc} are completely different. Nuclear magnetic resonance (NMR) structures of several mammalian prion proteins show that PrP^C(121–231) is ~45% helical with a very low (3–8%) β-sheet content (Riek et al., 1996, 1997; James et al., 1997; Donne et al., 1997; Zahn et al., 2000). Using Fourier transform infrared spectroscopy, it has been inferred that the secondary structure of PrP^{Sc} (formed from PrP^C(90–231)) has ~50% β-sheet content (Caughey et al., 1991; Gasset et al., 1993). An extraordinary conformational change takes place in the transition from the normal state to the infectious form. Determination of the NMR structures of PrP^C for a number of species (syrian hamster, mouse, bovine, and human) is significant because it potentially offers hope of understanding, at the molecular level, the mechanism of PrP$^C \rightarrow$ PrP^{Sc} conversion.

GLGGYMLGSAMSRPMIHFGNDWEDRYYRENMYRYPNQVYYRPVDQYSNQN

E1 | H1 → | E2

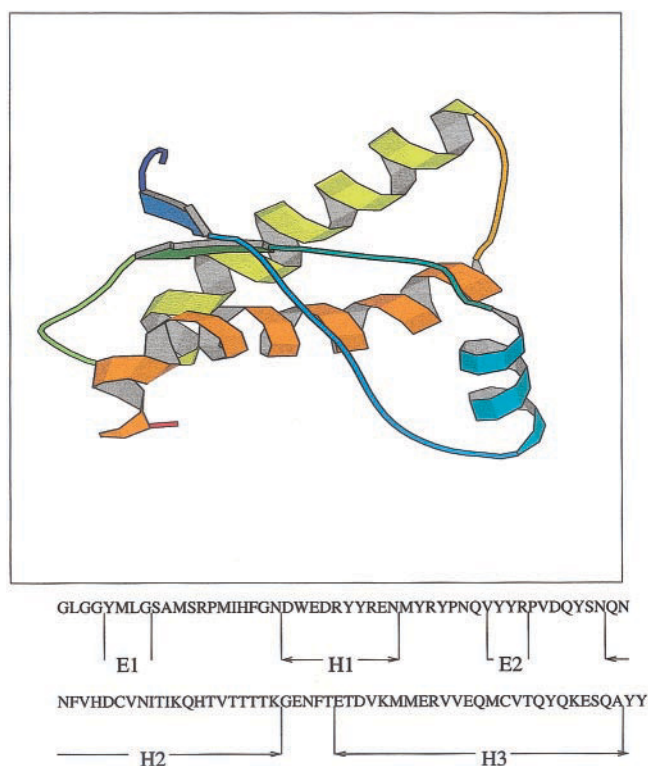NFVHDCVNITIKQHTVTTTTKGENFTETDVKMMERVVEQMCVTQYQKESQAYY

H2 | H3

FIGURE 1 Ribbon diagram of mPrP(121–231) (produced with the program MolScript). The sequence with the secondary structure assignments (given below) are taken from the header of the PDB file. The sequence numbering begins with Gly at 124 and ends with Tyr at 226. The amino acids are given in the 1-letter format. The two anti-parallel β-strands are between residues 128–131 and 161–164.
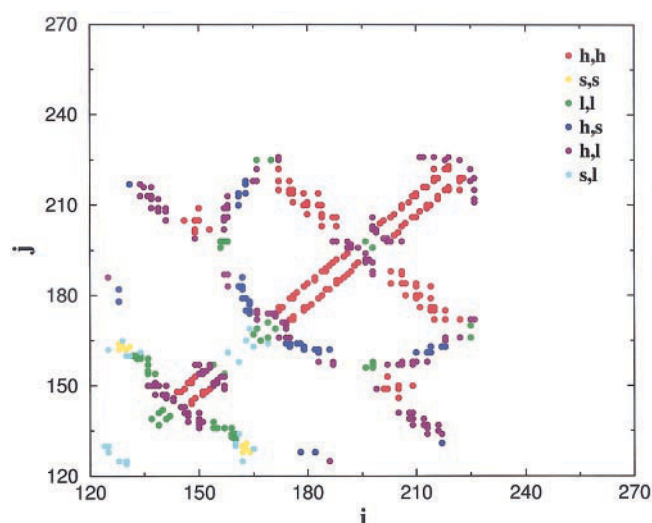

FIGURE 2 Contact map for the mPrP structure shown in Fig. 1. Two residues are assumed to be in contact if any of their heavy atoms are within 5.2 Å. In the legends, h stands for helices, s for strands, and l for loops. The contact map shows that H1 does not interact with H2. From the sequence (Fig. 1) and the contact map, it is clear that the few interactions between H1 and H3 are relatively destabilizing. In contrast, H2 and H3 make many stabilizing contacts with each other.

Prion proteins, encoded by a single gene, consist of about 250 residues of which the first 22 form a signal sequence. This is followed by unstructured, but likely helical, Cu^{2+} binding octarepeats rich in glycine (Prusiner, 1998). The NMR structure of the remaining protein PrP(90–231) shows that the N-terminus is disordered (Donne et al., 1997). The most ordered portion of the structure consists of ~103 or 104 residues, PrP^C(121–231). The structures from this region from three species (mouse, hamster, and human) show that PrP^C consists of three helices and two short β-strands (Riek et al., 1996, 1997; James et al., 1997; Donne et al., 1997; Zahn et al., 2000). Using the mouse prion protein as the reference (Fig. 1), the three helices H1, H2, and H3 span residues 144–153, 172–194, and 200–224, respectively. The NMR structure (Fig. 1) shows that there are two small anti-parallel β-strands (residues 128–131 and 161–164). The contact map for the mPrP(121–231) structure (Fig. 2) shows that PrP^C has the characteristics of an $\alpha+\beta$ protein.

We selected (from the Protein Data Bank (PDB)) four prion proteins with known tertiary structures: 1ag2 (*mouse prion protein*, 103 residues), 1b10 (*syrian hamster prion protein*, 104 residues), 1qlz (*human prion protein*, 104 residues, the first of 20 NMR models), 1qlx (*human prion*

protein, 104 residues, representative). According to Structural Classification of Proteins (SCOP) (Murzin et al., 1995), they belong to the $\alpha+\beta$ class of proteins, and, according to Homology Derived Secondary Structure of Proteins (Sander and Schneider, 1991), they are very similar because of 90% sequence identity between mPrP and h1PrP, an 86% sequence identity between shPrP and h1PrP, and a 94% sequence identity between mPrP and shPrP. There is some variability in the precise assignments of residues in the helices and the strands, which depends on whether NMR structures correspond to (90–231) or (121–231). The location of secondary structural elements also differs depending on the refinements used. In this study, we use the classification provided in the header of the PDB (Berman et al., 2000) file.

The purpose of our study is to identify those features of PrP that might play an important role in its conversion to PrP^{Sc} using sequence alignments and NMR structures. To pinpoint the plausible regions of the PrP^C structures that are susceptible to conformational fluctuations, we compare a number of characteristics of the structural motifs in PrP^C with those of "normal" proteins. We determined, from a databank of proteins from the PDB, the degree of solvent accessibility and the typical (ϕ, ψ) angles for the 20 types of amino acids in α-helices and in β-sheets. Comparison between the typical values and those for the amino acids of the sequences of PrP^C revealed that some positions in the helices of prion proteins exhibit characteristics that are distinct from what is normally observed. Distribution of contacts in PrP^C shows that they differ from normal proteins especially

in regard to the short-range contacts. Most of these differences are localized in helix 2 (H2). Our results suggest that the C-terminus of H2 is frustrated in the experimentally found α-helical state. Because the bulk of the stability of PrP^C comes from the core, our analysis suggests that nearly the whole protein might be involved in the transition to PrP^{C*} (a species that can nucleate and polymerize). This implies that there is a large barrier that separates PrP^{C*} from PrP^C.

MATERIALS AND METHODS

Choice of databanks

The four NMR structures of PrP^C analyzed are the mouse mPrP (103 residues, 1ag2), syrian hamster shPrP (104 residues, 1b10), human h1PrP (104 residues, 1qlz), and human h2PrP (104 residues, 1qlx). The PDB entries follow the lengths of the proteins. In the case of the entry h1PrP, we used the first of 20 NMR structures, and a representative structure was chosen for the entry h2PrP.

Because we are interested in comparing the prion proteins with normal proteins (for which no alternative conformations leading to fibril formation has been detected), we assembled databanks of mainly α, mainly β and $\alpha+\beta$ proteins (based on the SCOP classification) from the PDB. From the resulting sets, we kept only those proteins that satisfied the following criteria: 1) the experimental method for determining the three-dimensional (3D) structure of the protein was x-ray crystallography; 2) the protein is single-chain; 3) there are not many (more than 2) missing residues from the protein or many missing atoms from the amino acids of the protein; and 4) the experimental information about the secondary structure of the protein appears in the PDB header.

With these criteria, we were left with 58 mainly α proteins whose PDB entries are 1a32(OB), 1a43(OB), 1aa2(OB), 1aep, 1ail(OB), 1aum(OB), 1axn, 1ayi(OB), 1bbc, 1bd8, 1bea(OB), 1beo(OB), 1bfa(OB), 1bgc, 1bgd, 1bgf(OB), 1cei(OB), 1cem, 1csh, 1eca(OB), 1ecd(OB), 1fps(OB), 1gcb, 1hbg(OB), 1hik, 1huw, 1hyp(OB), 1hyt, 1ilk, 1ith, 1lh1(OB), 1lis, 1lki, 1lpe, 1lrv, 1mba(OB), 1mzl(OB), 1nfn, 1pbv, 1poa, 1r69(OB), 1rcb, 1utg(OB), 1vlk, 1vls, 2abk, 2asr, 2bct, 2cro(OB), 2end(OB), 2fha, 2int, 2lhb(OB), 2mhr, 451c(OB), 4cpv(OB), 4icb(OB), 5cyt(OB). The OB indicates that the architecture of the protein is orthogonal bundle according to Orengo et al. (1997).

The 36 normal $\alpha+\beta$ proteins in our comparison set have the PDB entries 119l, 1ah6, 1ako, 1cby, 1cof, 1ctf, 1fkb, 1gcb, 1han, 1kuh, 1lba, 1lbu, 1mrj, 1npk, 1opd, 1pgx, 1pkp, 1ptf, 1sap, 1snc, 1tif, 1uae, 1ubi, 1vcc, 1vhh, 2aak, 2hpr, 2phy, 2sn3, 3cla, 3lzm, 3ssi, 4bp2, 5pti, 7rsa, 9rnt.

The PDB entries of 31 mainly β proteins in our databank are 1aac, 1amm, 1aol, 1bfg, 1eur, 1fna, 1gpc, 1gpr, 1hoe, 1idk, 1jpc, 1knb, 1kum, 1lcl, 1mai, 1nif, 1pdr, 1 pmi, 1srl, 1tul, 1wba, 1whi, 1who, 1xnb, 2ayh, 2cba, 2cna, 2cpl, 2mcm, 2prd, 2sns.

Structural analysis

Using these three databanks, we determined some of the attributes, namely, the degree of solvent exposure and the (ϕ, ψ) angle combination, of the typical amino acid (for each of the 20 types of amino acids) in helices and in strands. Because the N-cap and C-cap of helices present special features compared to all other positions in a helix (see, for example, the study of Aurora and Rose, 1998, and references therein) and we were interested in obtaining the information about a typical residue in a helix, we treated these two positions as belonging to the loop category. We also treated helices shorter than six residues as loops.

Regular density

To decipher residue–residue interactions Baud and Karlin (1999) introduced several density measures that give an indication of the environments. Following these authors, we calculated the regular density for the nine structural classes formed from the three secondary structures (α helix, β-strands, and loops) and the degree of solvent exposure (buried (b), partially buried (pb), and exposed (e)). The regular density for an amino acid i ($i = 1, 2, \ldots, 20$) in an environmental class λ ($\lambda = 1, 2, \ldots, 9$) is (Baud and Karlin, 1999)

$$\bar{\rho}_i^\lambda = \frac{1}{N_T} \sum_{\alpha=1}^{N_T} \sum_j \Theta(T - d_{m,ij}^\alpha), \tag{1}$$

where $d_{m,ij}^\alpha$ is the minimum distance between two heavy atoms in a residue pair i and j, α is the structure, T is the threshold value ($= 5$ Å), $\Theta(x)$ is the Heaviside function, and N_T is the number of structures in the dataset. We calculated the distribution of $\bar{\rho}_i^\lambda$ and the associated dispersion $\delta\bar{\rho}_i^\lambda$. A mismatch for an amino acid in a class (hydrophobic (H), polar (P), positively charged ($+$), negatively charged ($-$)) is measured in terms of the variable

$$\Delta\rho_i^{\lambda,\alpha} = \frac{\rho_i^{\lambda,\alpha} - \bar{\rho}_i^\lambda}{\delta\bar{\rho}_i^\lambda}, \tag{2}$$

where $\rho_i^{\lambda,\alpha}$ is the regular density in the structure α. If $\Delta\rho_i^{\lambda,\alpha}$ lies outside the interval $-1 \leq \Delta\rho_i^{\lambda,\alpha} \leq 1$, then there is significant deviation from normal behavior. If a sequence harbors many mismatches, then we consider it to be frustrated in that structural element. We used the following classification of the amino acids:

1) hydrophobic: Cys, Phe, Leu, Trp, Val, Ile, Met, Tyr, Ala
2) polar: Gly, Pro, Asn, Thr, Ser, Gln, His
3) charged: Arg, Lys, Asp, Glu

RESULTS

Characteristics of normal proteins

To ascertain whether PrP^C proteins have any unusual structural characteristics, we first computed several characteristics of proteins in our database. These serve as a basis to illustrate the unusual features, if any, of the prion proteins.

Sequence composition

An obvious property to consider is the overall amino-acid composition. This is relevant because a proteomic sequence analysis of 31 organisms (Michelitsch and Weissman, 2000) revealed a large content of glutamine/asparaginine residues in yeast prions. The unusually high content of Gln and Asn suggests that the regions containing these residues may trigger the initial nucleation in the transition to the β-structures that leads to self-propagation in yeast prions (Michelitsch and Weissman, 2000). The composition of PrP^C is closer to that of all β proteins or to those $\alpha+\beta$ proteins that have a higher strand content than helices. In many $\alpha+\beta$ proteins, Ala is a typical helix former (Jiang et al., 1998). However, in PrP^C only H3 contains a single Ala, and none of the helices is amphipathic. In mPrP, H1 is made up of three hydrophobic, one polar and six charged residues and

has 3.66 residues/turn, and H2 is made up of six hydrophobic, two charged and 15 polar residues and has 3.92 residues/turn. Helix 3 is made up of seven hydrophobic, eight charged and ten polar residues and has 5.06 residues/turn. Although prion proteins have slightly unusual composition compared to a typical $\alpha+\beta$ protein, the differences are not dramatic.

Solvent exposure ratio for the amino acids in a helix or strand in normal proteins

To characterize the degree of solvent exposure of an amino acid in a protein, we used $R = A_S/A_R$ where A_S is the solvent-accessible area calculated using the Lee and Richards (1971) algorithm, and A_R is the area of the same type (X) of amino acid in a Gly-X-Gly extended conformation. The values of A_R were taken from the literature (Creighton, 1993). Each position, excluding the ends of the chain, in a protein in its folded conformation has R between 0 and 1. We computed R for the 20 amino acids from the helices of the 58 mainly α proteins, and of the 36 $\alpha+\beta$ proteins. These distributions obey the known percentages of buried positions (defined by R being less than 0.05) for each type of amino acid. For example, Val is buried in 55% of cases, whereas Lys is buried in a proportion of only 3% in the mainly α proteins. The distribution of R values for the hydrophobic residues is peaked at small values of R, with negligible small peaks at larger values. For polar and charged amino acids, the distribution is much broader with a peak at larger values than the ones for the hydrophobic amino acids.

For hydrophobic amino acids, the typical value of R is identified with the median value, whereas, for polar and charged amino acids, the average of the distribution gives the typical ratio (the actual values are in the caption for Fig. 3). For Ala, Gly, and Met, none of these two classifications led to a reliable typical value.

Distribution of the (ϕ, ψ) angles for the typical amino acid in helix or strand

We computed the distribution of the (ϕ, ψ) angles for each type of amino acid in the helices and strands from the three datasets described in Methods. The limits of the (ϕ, ψ) pair for an amino acid (other than Gly and Pro) in an α helix are $-80 \leq \phi \leq -48$, $-59 \leq \psi \leq -27$. The range for a β sheet is $-150 \leq \phi \leq -90$, $90 \leq \psi \leq 150$ (Srinivasan and Rose, 1995).

We divided the range for each of the two angles in a helix in intervals of 10°, which led to nine different classes of (ϕ, ψ) angles. We assigned to each amino acid a number between 1 and 9, based on its (ϕ, ψ) angle. If the angles were outside the specified range of all classes, an arbitrarily large number (50) was assigned. We obtained the distribu-
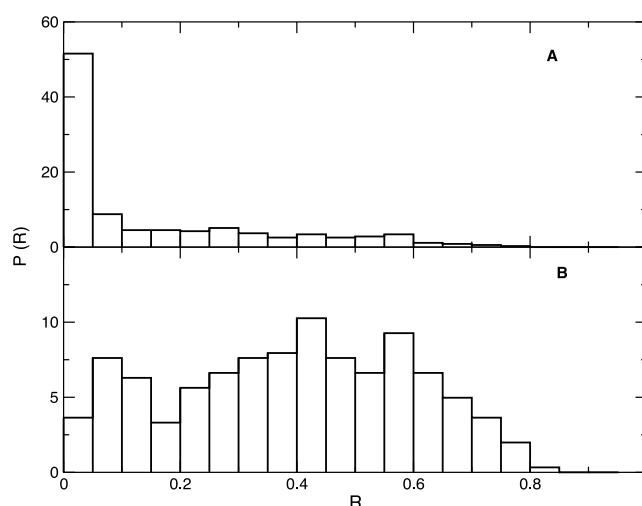

FIGURE 3   (A) Distribution of R, the ratio of accessibility of surface area of a given amino acid to the area of the same amino acid in a Gly-X-Gly extended conformation, for Val. (B) Same distribution as in (A) for Arg. The mean (median) values R and the corresponding dispersion for the amino acids are: Cys ($R = 0.016$, tol $= 0.100$), Phe ($R = 0.042$ tol $= 0.100$), Leu ($R = 0.019$ tol $= 0.100$), Trp ($R = 0.095$ tol $= 0.100$), Val ($R = 0.046$ tol $= 0.100$), Ile ($R = 0.036$ tol $= 0.100$), His ($R = 0.272$ tol $= 0.200$), Tyr ($R = 0.086$ tol $= 0.100$), Pro ($R = 0.404$ tol $= 0.253$), Asn ($R = 0.348$ tol $= 0.195$), Thr ($R = 0.262$ tol $= 0.232$), Ser ($R = 0.317$ tol $= 0.231$), Arg ($R = 0.394$ tol $= 0.203$), Gln ($R = 0.347$ tol $= 0.198$), Asp ($R = 0.371$ tol $= 0.232$), Lys ($R = 0.492$ tol $= 0.182$), Glu ($R = 0.405$ tol $= 0.211$).

tion of these numbers for the 20 amino acids in the helices of the 58 mainly α proteins, and of the 36 $\alpha+\beta$ proteins. The median value, computed using the distribution, was chosen to be the typical angle for that amino acid. This method of computing the typical angles is more meaningful than obtaining averages over the distribution of (ϕ, ψ) angles. The typical (ϕ, ψ) angles obtained from the 58 all α proteins using this procedure are: for Cys, Phe, Leu, Trp, Val, Ile, Met, His, Tyr, Asn, Thr, Arg, Gln, Asp, and Glu $-70 < \phi \leq -59$, $-49 < \psi \leq -38$, and for Ala, Ser, and Lys $-70 < \phi \leq -59$, $-38 < \psi \leq -27$. The corresponding values from the $\alpha+\beta$ databank are very similar.

For each of the two angles in a strand, we divided the above limits in intervals of 10°, which led to 25 different classes of (ϕ, ψ) angles. We attributed to each amino acid a number between 1 and 25, based on which class corresponded to its (ϕ, ψ) angles. If the angles fell outside the range, a score of 50 was given. From the distribution of these numbers for the 20 types of amino acids in the strands of the 31 mainly α proteins and of the 36 $\alpha+\beta$ proteins, the median value was calculated. The typical angle for a given amino acid in these structures corresponds to the median value.

Structural mismatches in helices of PrP^C

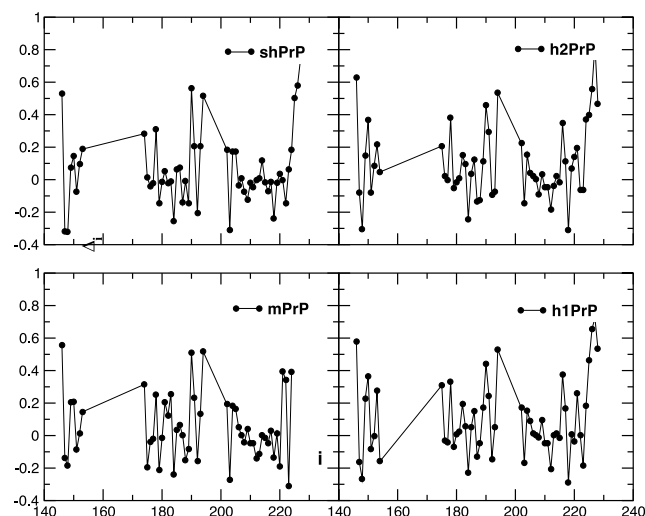The distribution of R values and the angles adopted by amino acids in the helical and strand structures give a

FIGURE 4 Mismatch values $\Delta_i = R_i - \bar{R}_i$ where $\bar{R}_i$ is the typical (mean or median) value (Fig. 3) and i is the position of the amino acid along the sequence. For all species, there are considerable deviations from the typical values even in the core of PrP^C.

calibration of what is typically expected in the database of normal structures. A comparison of these values with those adopted by amino acids in the helices and strands of prion proteins gives hints of plausible structural anomalies.

If R (a measure of solvent accessibility) for an amino acid in PrP^C is different from the typical values $R_\alpha + \delta R_\alpha$ (the typical values and the values of the tolerance δR_α are given in the caption of Fig. 3), then we consider it a mismatch. Similarly, if the (ϕ, ψ) angle for an amino acid falls outside the typical values, it is a mismatch. Using these criteria, we find that between 13 and 16 amino acids are mismatches in the four prion proteins (16 out of 52 positions in helices in the mPrP, 16 out of 56 positions in helices in h1PrP, 13 out of 55 positions in helices in shPrP and 15 out of 56 positions in helices in h2PrP). The difference between R for each position and the typical value of R for the amino acid at that position in the helices of the four prion proteins are shown in Fig. 4. We repeated this analysis for all the helices in 58 proteins from the mainly α databank and in the 36 proteins from the $\alpha+\beta$ databank. The resulting histogram of all the percent mismatches in angles (that is, the number of mismatches N_{mism} divided by the length of the helix, L_{helix}) shows that the prion proteins belong to the tail of the distribution (Fig. 5). A histogram for the mismatches in R also shows that the prion proteins belong to the tail of the distribution corresponding to values larger than the average of 20% (data not shown). An analysis of the amino acids in the β strands did not reveal any anomalies in prion proteins compared to the typical $\alpha+\beta$ protein.

Environment-dependent structural characteristics reveal unusual number of mismatches in PrP^C

Baud and Karlin (1999) introduced nine structural categories to quantify the environmental propensities of amino
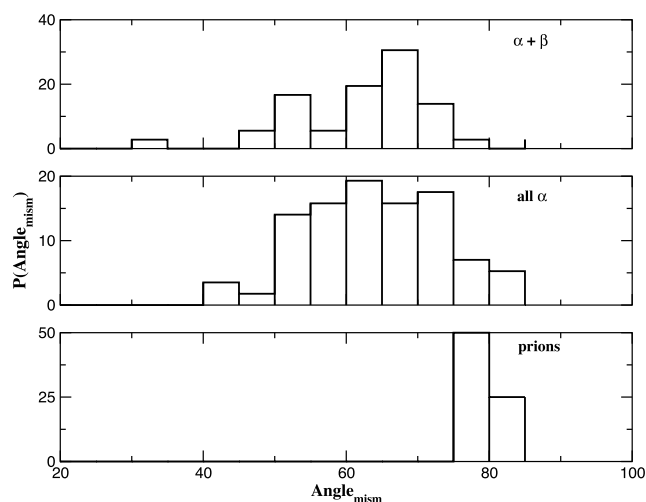

FIGURE 5 Distribution of the number of mismatches for amino acids in helices (normalized by helix length and expressed as percentages) for the (ϕ, ψ) angle. The value of a (ϕ, ψ) angle at a position is considered a mismatch if it is different from the typical value. The top panel shows the distribution for the amino acids in the helices of the $\alpha+\beta$ proteins, the middle one corresponds to the proteins from the all α databases, and the bottom panel corresponds to the four prion proteins. Prion proteins contain more mismatches than regular proteins that do not aggregate. Helices in PrP^C have many more positions with (ϕ, ψ) angles lying outside the range of a right-handed α helix. This shows that there are local instabilities in the secondary structures of prion proteins. The corresponding distribution for R values shows no such dramatic differences between the prion proteins and other proteins.

acids in folded proteins. They are based on the three standard secondary structure states (helix, strand, and loop), and on three side-chain solvent-accessibility levels: $A_s \leq 10\%$ for the buried state (b), $10\% < A_s \leq 40\%$ for the partly buried state (pb), and $A_s > 40\%$ for the exposed state (e).

There are clear differences among the various types of amino acids in these nine structural classes in terms of number of neighbors and their identities (aromatic, hydrophobic, polar, positively, and negatively charged). Using this approach, we calculated the typical number of neighbors (within 5 Å) and the corresponding standard deviations for the amino acids in each of the nine structural classes from the database of proteins. We compared these reference results with the structural characteristics of amino acids in PrP^C in terms of mismatches. The resulting histogram of such mismatches shows that there is a similarity between this analysis and the one in Fig. 5 (data not shown). There are significant deviations from the expected behavior of the regular density (defined in Methods) for amino acids exposed in the helices of PrP^C. This is particularly dramatic for exposed hydrophobic residues (Fig. 6 A), buried negatively charged residues (data not shown) and exposed negatively charged residues (Fig. 6 B). These two figures suggest that, in PrP^C, there are more mismatches between the structural preferences of amino acids and the actual structural environment that is found in normal proteins. Thus,
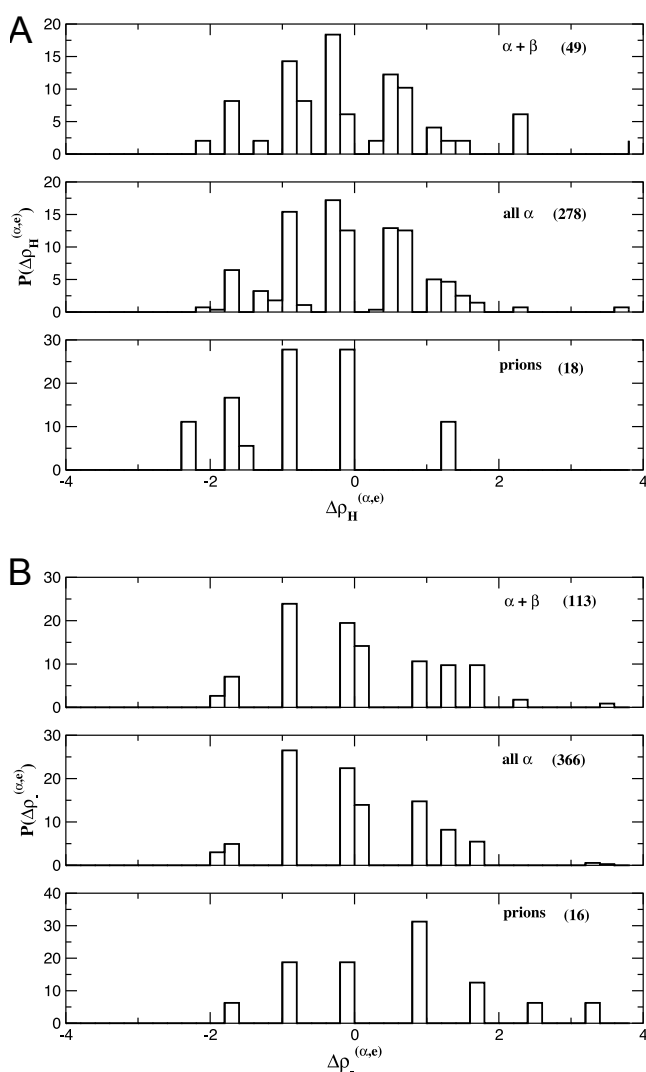
FIGURE 6 (A) Distribution of the normalized deviations of the regular density (Eq. 2 in Methods) for solvent-exposed hydrophobic amino acids in PrP^C that are in helices. For comparison, the results from helices in our dataset (i.e., from the all α and the $\alpha+\beta$ proteins) are also shown. The numbers in parenthesis correspond to the total number of residues in the given environmental class. For example, there are 18 exposed hydrophobic amino acids in the helices of the four prion proteins. (B) Same as in (A) except that it is for the solvent-exposed negatively charged residues.

both secondary-structure analysis and environmental analysis that reflects tertiary structure preferences suggest anomalies in PrP^C.

Hydrogen bonds

The present analysis shows that amino acids at some sites in the PrP^C present unusual properties compared to the average amino acid of the same type in normal proteins. This is exemplified in the degree of solvent accessibility, i.e., some polar amino acids (e.g., Glu-146 and Asp-147 in the first helix of mPrP), which are typically exposed, are buried in

TABLE 1 Statistics of local and nonlocal contacts

Proteins	Short-range number (type)*	Long-range number (type)
All α	10639 (Leu-Leu)	5429 (Leu-Leu)
$\alpha + \beta$	7155 (Leu-Val)	3796 (Leu-Val)
All β	6661 (Leu-Ile)	3684 (Leu-Ile)
Prions	398 (Tyr-Asn)	305 (Met-Tyr)

*The parenthesis gives the residues in most probable contacts.

PrP^C. It follows that the α-isoform of PrP^C should have many unsatisfied buried hydrogen-bond donors/acceptors. To assess this, we examined the hydrogen-bond characteristics of normal proteins. The analysis of the all α protein 1aa2 (108 residues) with the WHAT CHECK program (Hooft et al., 1996) reveals seven unsatisfied buried hydrogen-bond donors/acceptors, whereas, for the $\alpha+\beta$ protein 1fkb (107 residues), the number is eight, and for the all β protein 1aac (105 residues), the corresponding number is eight. Mc-Donald and Thornton (1994) showed that, in normal proteins, only a low percentage ($\sim$6%) of the total number of residues have unsatisfied buried hydrogen-bond donors/acceptors. The WHAT CHECK analysis for prion proteins shows 15 (14%) unsatisfied buried hydrogen-bond donors/acceptors in the mPrP (1ag2), also 15 (14%) in the shPrP (1b10), 6 (5.8%) and 9 (8.7%) in the human prion (h2PrP and h1PrP, respectively). We find that mPrP and shPrP have more than twice the usual amount of unsatisfied buried hydrogen-bond donors/acceptors, whereas the human prion protein behaves more like an average protein in this respect. There is good agreement between the identity of the sites revealed by this analysis and the problem sites identified by mismatches in R and the distribution of (ϕ, ψ) angles. As an example, the WHAT CHECK analysis for mPrP reveals unsatisfied hydrogen bonds at positions 130, 139, 141, 142, 143, 145, 151, 155, 161, 166, 170, 174, 183, 187, and 219 (Fig. 4) and we found 130, 145, 151, 174, 183, 187, and 219 to have mismatches in terms of R and the (ϕ, ψ) angles.

Local versus nonlocal contacts

We consider two residues to be in contact if they have at least two of the heavy atoms from their side-chain within 5.2 Å. Contacts among residues that are less than (greater than) 20% of the length of the protein along the sequence are classified as local (nonlocal). The tendency of PrP^C to undergo $\alpha \rightarrow \beta$ transition suggests that these structures may only be marginally stable (Cohen et al., 1994), and hence the distribution of contacts, which points to the intrinsic stability of proteins, is useful.

We calculated the number and type of the short- and long-range order contacts in prion proteins and in the proteins from the three databases. The results, shown in Table 1 and Fig. 7, show that, for a typical protein (whether all α, $\alpha+\beta$, or all β), the number of short-range (or local) contacts is about twice as large as the number of long-range (or
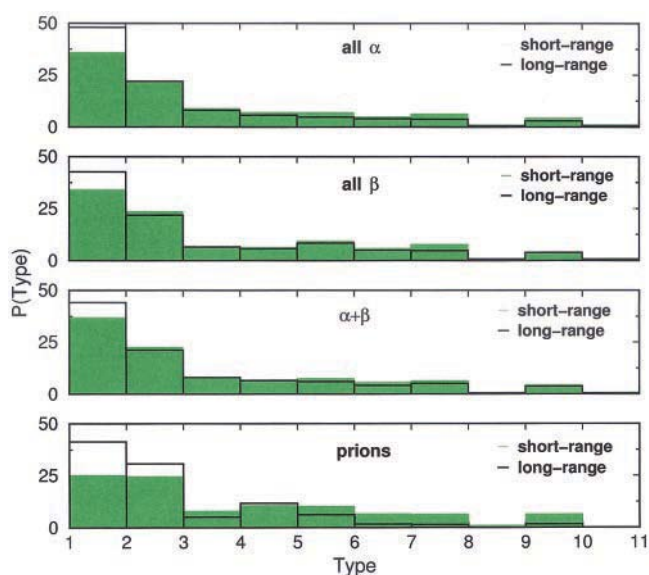
FIGURE 7 Percentage of each of the possible types of local and nonlocal contacts in PrP^C and in all the proteins from our database. Local contacts are defined as those that are within 20% of the length of the protein along the sequence. All other contacts are considered nonlocal. The 20 types of amino acids are divided into four classes: hydrophobic (H) (as above), polar (P) (as above), positively charged (+) (Arg and Lys), and negatively charged (−) (Asp and Glu). This leads to 10 types of possible interactions, which are indicated in the ordinate: 1 (H, H), 2 (H, P), 3 (H, +), 4 (H, −), 5 (P, P), 6 (P, +), 7 (P, −), 8 (+, +), 9 (+, −), and 10 (−,−).

TABLE 2 Nature of the most probable local and nonlocal contacts

Helix	mPrP	shPrP	h1PrP	h2PrP
H1	(+−, HH)*	(+−, HH)	(+−, HH)	(HH, HH)
H2	(HP, HH)	(P+, HH)	(PP, HH)	(PP, HH)
H3	(HH, HH)	(HH, HH)	(H−, HH)	(HP, HH)

*The first entry corresponds to local and the second to nonlocal contacts. H, P, +, − refer to hydrophobic, polar, positively charged, and negatively charged, respectively.

are 2cpl and 1hoe, but neither of them has such a large percentage of +− contacts as prion proteins. Note that all of the above-mentioned proteins from the $\alpha+\beta$ class have a much lower helical content than PrP^C: 1fkb has 1 helix, the rest being sheet, 1vhh is 19% helical and 16% sheet, 2cpl is 12% helical and 27% sheet and 9rnt has just one helix, the rest being sheet.

This analysis reveals some important characteristics of PrP^C. They have an unusually large number of +− local contacts; They also have far from typical percentages of both local and nonlocal HP, H−, PP, P− contacts; and The percentages of local H− and P− contacts are similar to what is seen in proteins that have a much higher sheet content than prion proteins.

Nature of contacts in H1, H2, and H3 in PrP^C

The unusual local and nonlocal contacts (+−, HP, H−, PP, P−) in PrP^C are localized in the three helices. Because we are interested in regions of instability, we examined the nature of local and nonlocal contacts in H1, H2, and H3. The idea is to compare the pair of the most probable type of short-range and the most probable type of long-range contact in each helix from PrP^C with the same pair from the helices of all other proteins. We define as short-range any contact between an amino acid of a helix with any other amino acid belonging to the same helix (including the N- and the C-cap of the helix), and as long-range the contacts with the amino acids outside the helix. We extracted, from the prion proteins and from all the proteins in the all α and $\alpha+\beta$ datasets, the pair of the most probable types of short and long-range contacts for every helix. We then counted how often each of the pairs seen in PrP^C appears in the other proteins. The combined results, presented in Tables 2 and 3, suggest that H2 is the most different from average, followed by H1 and then by H3.

Contacts among H1, H2, and H3

The $\alpha \rightarrow \beta$ structural transition in PrP^C(90–231) leads to ~50% β-sheet content in PrP^{Sc}, which suggests that a relatively large number of the residues in the α-helices must rearrange into a β-strand conformation. Previous studies,

nonlocal) contacts, whereas, in PrP^C, the number of short-range contacts is approximately equal to that of the number of long-range contacts. More importantly, the nature of residues that are involved in local contacts in PrP^C is strikingly different from the ones in normal proteins. In normal proteins the most probable local contacts are exclusively made up of hydrophobic residues, which is not the case in prion proteins (Table 1). The numbers of HP (hydrophobic (H) and polar (P)) and HH contacts are approximately equal in prion proteins, whereas, in normal proteins, there are more HH contacts than HP contacts. The percentages of local contacts of types +−, H−, and PP are much higher in PrP^C than in normal proteins.

To put the above striking observations about PrP^C in perspective, it is necessary to assess whether similar characteristics can be found in normal proteins. We searched the dataset to find instances of local and nonlocal contacts involving amino acids that are not typically associated with each other. There are a few proteins with similar characteristics as PrP^C. They are 1axn, 1mzl, 2asr, and 5cyt from the all α database; 1fkb, 1vhh, 2cpl, and 9rnt from the $\alpha+\beta$ database; and 1aac and 1hoe from the all β database. Most of these proteins have, at short-range, comparable numbers of HP and HH contacts and many PP contacts, just like PrP^C. The all α proteins also have many local H+ and P+ type of contacts, but none of them has as many +− and H− contacts as PrP^C. Proteins that appear most similar to PrP^C

TABLE 3 **Joint probability of occurrence of various types of contacts**

Type	In α (394)	In $\alpha+\beta$ (150)	Total (544)	Percentage
$(+-,\ \text{HH})^*$	29	21	50	9.2%
(HP, HH)	70	24	94	17.3%
(HH, HH)	158	43	201	36.9%
(P+, HH)	13	14	27	4.9%
(PP, HH)	23	3	26	4.8%
(H-, HH)	22	6	28	5.1%

*The contact types are described in Fig. 7.

based in part on the observation that H1 is unusually hydrophilic (Billeter et al., 1995; Morrisey and Shakhnovich, 1999), have already proposed that, in PrP^{Sc}, H1 is likely to play a role in the α transition. Here, we find that the unusual nature of short- and long-range contacts in H2 makes it frustrated in the helical state. Therefore, it is interesting to analyze the contacts among the three helices to estimate the effects that a structural transition ($\alpha \rightarrow \beta$) in a helix might have on the other helices. The arrangement of the helices in PrP^C gives them an orthogonal bundle (OB) architecture (Orengo et al., 1997). Therefore, we selected 47 proteins with this type of architecture from the CATH database (Orengo et al., 1997). Of these, 27 are among the 58 all α proteins from our dataset. We calculated the number of interhelical contacts in each of these 47 proteins and in the four prion proteins. Plot of the number of contacts formed by a helix versus its length and the percentage of stabilizing (that is, HH or $+-$) contacts among these (data not shown) shows that all three helices from PrP^C are similar (in this respect) to helices from other proteins.

The analysis of the interhelical contacts reveals that there are many more (H$-$) interhelical contacts in prion proteins than in other proteins (most of which are due to the contacts between H1 and H3), and that the percentage of HH type of contacts in PrP^C is somewhat increased (due to the contacts between H2 and H3). The largest majority of stabilizing (HH and $+-$) contacts occur between the amino acids of the first half of H2 (172–183) and those of the second half of H3 (213–224) (around the disulfide bond between Cys-179 and Cys-214). Contacts between H2 and H3 are very similar (in terms of number and type) with the interhelical contacts seen in other proteins having the same architecture as the PrP^C, suggesting that any structural transformation involving H2 (especially its first half, i.e., residues 172–183) is likely to affect H3 also (especially its second half, i.e., residues 213–224).

Clustering of hydrophilic and hydrophobic residues suggests that H2 is frustrated

Istrail et al. (1999) have noted that one of the main features of an aggregation-prone sequence is that the hydrophobic and hydrophilic residues are clustered into a few large
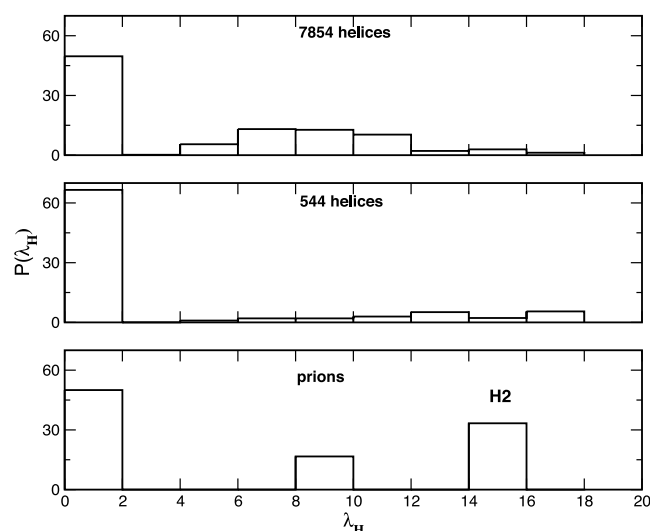

FIGURE 8 Histogram of the λ_H (Eq. 3) values for the sequences of helices of PrP^C, the 57 all α proteins and of the 37 $\alpha+\beta$ proteins (544 in total). The middle panel shows $P(\lambda_H)$ for all 7854 helices with at least 23 residues. These were extracted from the PDB using the DSMP database.

groups. To assess the clustering of hydrophobic residues, we looked for how homogeneously the hydrophobic residues are distributed in each of the two halves of the three helices in PrP^C in comparison to the other proteins in the dataset. Because there is a rather broad distribution of helix lengths in proteins, to describe the distribution of hydrophobic amino acids in each helix, we used the quantity

$$\lambda_H = \left| \frac{1}{2} - \frac{S_{max}}{S_H} \right|, \qquad (3)$$

where S_{max} is the maximum number of hydrophobic residues that is found in the two halves of the helix, and S_H is the total number of hydrophobic residues in the helix. The histogram of λ_H is presented in Fig. 8. In the helices of normal proteins, the hydrophobic residues are uniformly distributed in the two halves. Two of the helices in prion proteins (H1 and H3) obey this rule very well. The clear difference is in H2, which has nearly all its hydrophobic amino acids clustered in one half. To check the generality of this observation, we calculated λ_H for all the helices (with minimum length of 23 amino acids) in PDB. Using the Database of Structural Motifs in Proteins (DSMP), we found 7854 such helices among the 12,904 proteins in the PDB. The corresponding histogram of λ_H (Fig. 8) is very similar to the one obtained using only the helices in our dataset of proteins.

Evaluating the structural properties of PrP^C using threading

Because of our reliance on environmental classes, a threading study that is based on profiles rather than on specific

interactions between proteins would supplement our conclusions. To this end, we use the standard 3D-1D scores introduced by Eisenberg and co-workers (Bowie et al., 1991).

To use the profile method for threading, we first determined for each protein from our dataset the environment in each of its positions. We performed gapless threading of all the prion sequences on all possible fragments of structures from the dataset. For scoring, we used the 3D-1D scores, which encode the likelihood of finding the twenty amino acids in the 18 possible environments (Bowie et al., 1991). For comparison purposes, we also did the threading for all the proteins in our dataset.

According to the scores from threading analysis, the prion proteins are very similar to those $\alpha + \beta$ proteins of similar length, but which have little helical content compared to the strand content. The $R_{\text{3D-1D}}$ scores for all α and $\alpha + \beta$ proteins having a high helical content (and of similar length as the prion proteins) are well above 30, which is large compared to the scores of 20 obtained for prion proteins. An estimator of the stability of proteins in their native-state conformations is the Z score (Bowie et al., 1991),

$$Z = \frac{N - \langle N \rangle}{\sigma}, \quad (4)$$

where N refers to the score of the sequence in its native state conformation, $\langle N \rangle$ is the average score of the sequence over all other conformations but the native one, and σ is the corresponding standard deviation. A stable protein is characterized by a large and positive Z.

The normalized Z scores for the majority of $\alpha + \beta$ proteins, with helical/strand content similar to that of PrP^C, are considerably higher than for prion proteins (data not shown). This suggests that PrP^C, which are rich in helices, are not stable in their native state conformations. The normalized Z of 0.7 for PrP^C resembles the normalized Z score for other $\alpha + \beta$ proteins of similar length and with very low helical content.

The goodness of the fit between a sequence and a structure can be assessed using 3D profile scores, which, for correct models, increase with the length of a protein (Luthy et al., 1992). The $R_{\text{3D-1D}}$ scores for the stable proteins in the PDB are proportional to the sequence length (L). A plot of the $R_{\text{3D-1D}}$ scores for the proteins in our dataset versus L can be fit using

$$R_{\text{3D-1D}} = 0.36 \, (\pm 0.01)L - 6.92 \, (\pm 2.25). \quad (5)$$

If this is applied to prion proteins, with $L = 104$ residues, we expect $R_{\text{3D-1D}}$ to be between 27.2 and 33.8. But the computed values of $R_{\text{3D-1D}}$ are 21.9, 19.9, and 23.6 for mPrP, shPrP, and h2PrP, respectively. This also shows that PrP^C proteins are not optimal in their native α helical state.

Helix 2 is frustrated in PrP^C mutants: Analysis using PHD

It is known that inherited human TSEs (familial CJD, GSS, and FFI) are associated with mutations in the PRNP gene. According to SWISS-PROT, seven of the point mutations (D178N, V180I, T183A, H187R, T188R, T188K, T188A) are found in H2 (172–194). A naive application of helical propensities due to Chou and Fasman (1978) would suggest that, except for D178N, all the other point mutations should lead to better helix formation. However, reliable secondary-structure prediction requires the use of context-dependent propensities based on multiple sequence alignments as used in PHD (Profile network from Heidelberg) (Rost and Sander, 1993). Kallberg et al. (2001) have also argued that H2 is "frustrated" in the helical conformation, whereas H1 and H3 are not. This conclusion was reached by looking for sequences with ≥ 7 residues from 1324 protein structures that are predicted by PHD to be in β-strands, but are experimentally determined to be in α-helices. This α/β discordance (or mismatch) was proposed to be associated with amyloid fibril formation (Kallberg et al., 2001). We measure the degree of α/β discordance or frustration using

$$S_{\alpha/\beta} = \frac{1}{L} \sum_{i=1}^{L} (R_i - 5), \quad (6)$$

where R_i is the reliability score predicted by PHD for a particular position along the sequence, 5 is the average score, and L is the length of the sequence. If $S_{\alpha/\beta} = 4$ and the sequence is experimentally determined to be in a helical conformation, then the particular sequence is frustrated (or maximally discordant) in the predicted secondary structure; $S_{\alpha/\beta} = 0$ is marginal. Negative values of $S_{\alpha/\beta}$ imply that PHD is unreliable in this prediction.

It has already been noted by Kallberg et al. that wild-type H2 is discordant in mouse and syrian hamster. We calculated the $S_{\alpha/\beta}$ scores for the predictions of PHD of secondary structure content due to point mutations in H2 (Fig. 9). The $S_{\alpha/\beta}$ scores are 1.83, 1.94, 1.80, 1.30, 1.80, 1.54, 1.94, and 1.94 for the WT, D178N, V180I, T183A, H187R, T188K, T188R, and T188A, respectively. Helix 2 for all these mutations is, just as in the WT, discordant or frustrated (Fig. 1). The differences in the $S_{\alpha/\beta}$ scores suggest that the biophysical characteristics of PrP^C mutants can differ greatly (Liemann and Glockshuber, 1999). However, in all cases, H2 is frustrated.

DISCUSSION

Experimental evidence

A key experimentally testable prediction of our analysis is that, in addition to the previously identified residues, even segments of the relatively rigid core of the PrP^C, especially
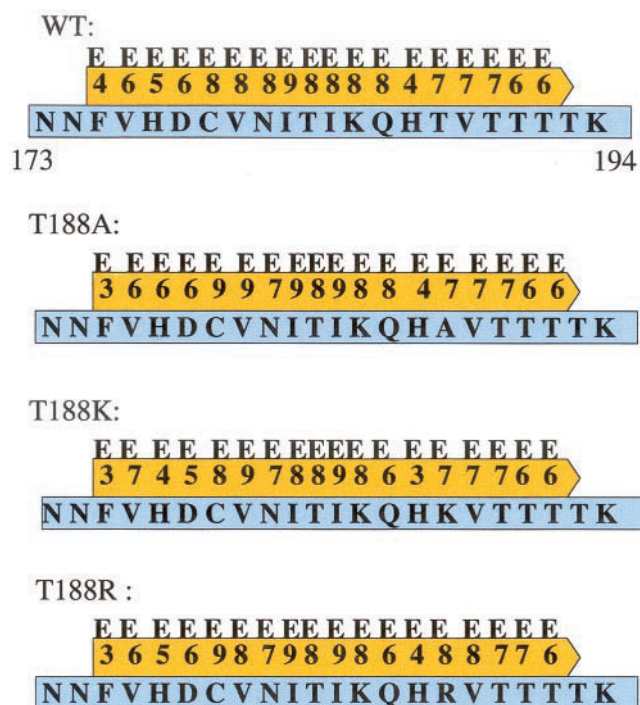
FIGURE 9   Assessment of the extent of frustration (or discordance) in H2 from hPrP^C and some disease-causing point mutations. The predicted structures using PHD are designated as strands (E and confidence scores in gold), whereas NMR shows that they are helical (blue). The frustration score, computed using Eq. 6, is given in the text. Examples of other frustrated helices are given in Kallberg et al. (2001).

the C-terminal residues of H2, might play a role in the transition to an assembly competent state, i.e., one that can nucleate and polymerize. This result and the observation that the secondary structure of PrP^{Sc} has a high percentage of β-sheet content imply that the whole protein molecule unfolds substantially in going from PrP^C to an aggregation prone state. Although direct and unequivocal experimental validation is currently not available, many distinct lines of evidence lend support to our theoretical analysis. These include backbone hydrogen/deuterium (H/D) exchange dynamics, NMR relaxation measurements, and CD spectroscopy of the states in the PrP^C→PrP^{Sc} transition.

Backbone H/D exchange dynamics

It has been postulated that the formation of fibrils in amyloids and prion proteins occurs from populated (at least partially) intermediates, i.e., PrP^C→[I]→PrP^{Sc} (Hornemann and Glockshuber, 1998; Morillas et al., 2001). To test this proposal, Clarke and coworkers (Hosszu et al., 1999) measured the backbone H/D exchange rates in human prion protein with the disulfide bond between Cys-179 and Cys-214 intact. The structural mobility of the protein fragments can be inferred from such measurements. Under the conditions of the experiment (pH = 5.5 and T = 303 K), PrP^C

unfolds in a single step upon addition of GuHcl with an equilibrium constant $K_{UN} \simeq 1.5 \times 10^3$ (Hosszu et al., 1999). The exchange process can be used to obtain the residue-dependent protection factor P, which gives an indication of the structural mobility. Efficient H/D exchange occurs from the solvent-exposed region. Thus, if $P \simeq K_{UN}$, then exchange occurs only from the unfolded PrP^C rather than from an equilibrium or off-pathway intermediate. Remarkably, it was found that only about 10 residues (Hosszu et al., 1999), clustered around the disulfide bond in the core of PrP^C, have P values greater than K_{UN}. Of these, only Cys-179 and Ile-182 from H2 have $p > K_{UN}$. The rest of the residues are in H3 (see Fig. 2 a of Hosszu et al., 1999). Using these observations, Clarke and coworkers ruled out the possibility that PrP^{Sc} forms from a native-like intermediate. Their experiments and the observation that PrP^{Sc} has predominantly β-sheet architecture led them to conclude that "complete or near-complete unfolding must precede rearrangement of the amyloidogenic intermediate."

Biophysical studies

NMR structures show that the N-terminal residues (90–120) in PrP^C are disordered (Riek et al., 1996, 1997; James et al., 1997; Donne et al., 1997; Zahn et al., 2000). It is also known that this region is resistant to protease digestion, which gives credence to the notion that it is structured in PrP^{Sc}. Indirect evidence for this finding comes from the work of Peretz et al. (1997), who identified PrP^C- and PrP^{Sc}-specific epitopes in the region (90–120) using binding of recombinant Fab fragments. This study suggests that the smallest region that undergoes structural transition upon prion formation is the disordered segment (90–120) in PrP^C. To test whether (90–120) is the only region with altered conformation in PrP^{Sc}, Hornemann and Glockshuber (1998) performed biophysical studies of the structured C-terminal domain (121–231) of PrP^C. Under acidic conditions (pH < 5) urea-induced unfolding of PrP^C exhibits a three-state transition with a well defined "equilibrium" intermediate. Far UV-CD spectrum of the acid intermediate revealed that it has the characteristics of a β-sheet structure (Hornemann and Glockshuber, 1998). These measurements suggest that, regardless of whether the intermediate is under equilibrium, certain regions besides the previously implicated (90–120) fragment (Peretz et al., 1997) must also be implicated in the conformational transition. This is not inconsistent with our analysis.

Unfolding of PrP^C

Because the core of PrP^C might play a role in the conformational change, it follows that substantial unfolding must precede the formation of a species, β-PrP (Baskakov et al., 2001), that can subsequently nucleate and polymerize. This result, when combined with the stability of PrP^C [(6–10)

kcal/mole depending on conditions and fragment length] suggests that a large barrier separates the α-helical PrP^C and β-PrP. The time scale for β-PrP, the α-helical isoform of PrP^C, to PrP^β conversion is $\tau_{\alpha\to\beta} \simeq \exp(\Delta G^\ddagger/RT)$ where an estimate for ΔG‡ is between (10–20) kcal/mole (see below and Baskakov et al., 2001). The presence of such a large barrier separating α-PrP and β-PrP implies that, in passing from one isoform to another (i.e., at the transition state), the protein should globally unfold.

Given that the N-terminus of PrP^C (90–231) is disordered and that H1 interacts at best weakly with the C-terminal domain, it follows that much of the stability of native α-PrP^C must come from its core. This implies that fluctuations in this region (H2 and H3) would occur on the longest time scale. Direct NMR relaxation dynamics probing the structural plasticity of prion protein show that, on the millisecond time scale, H2 and H3 are relatively rigid (Viles et al., 2001). The time scale for the fluctuations in the core that will nearly globally unfold PrP^C may be estimated using the data from Wildegger et al. (1999). They found that mPrP^C(121–231)-F175W (Phe at 175 replaced by Trp) refolds to α-Prp^C with a time constant of ~170 s at $T = 227$ K. Using $\Delta G_{UN} \simeq 8.4$ kcal/mole (taken from Table 1 of Wildegger et al., 1999), we estimate that the unfolding time $\tau_u \simeq 969$ s. Assuming that the pre-factor in the Arrhenius expression is given by the diffusion-limited estimate of $\simeq 1$ μs (Chan et al., 1997), we find that ΔG$^\ddagger \simeq 12$ kcal/mole. This estimate is not inconsistent with the one for mPrP(89–231) (Baskakov et al., 2001) especially when corrected for the differences in length. The long lifetime of the α-helical state of PrP^C together with the requirement that the assembly-competent state must have a fair amount of β-sheet structure implies that substantial unfolding of the protein must occur before polymerization.

Secondary structure content of PrP^{Sc}

At present, there are no high-resolution structures of PrP^{Sc}. Fourier transform infrared studies suggest that 48% of residues (or about 68 residues) in PrP^{Sc}(90–231) are in β-strands (Wille et al., 1996). Because the N-terminal region (90–120) in PrP^{Sc} is protease resistant, it is likely that this epitope is ordered in the fibrils. It is unlikely that all the residues in this region adopt a strand conformation for two reasons.

First, of the last nine residues (112–120) in mPrP^C, six are alanines, which tend to form α-helices, especially if there are many of them in a row. To assess whether this is peculiar to mPrP^C, we performed multiple sequence alignments that highlight the differences among various species. The degree of conservation in this region among the 150
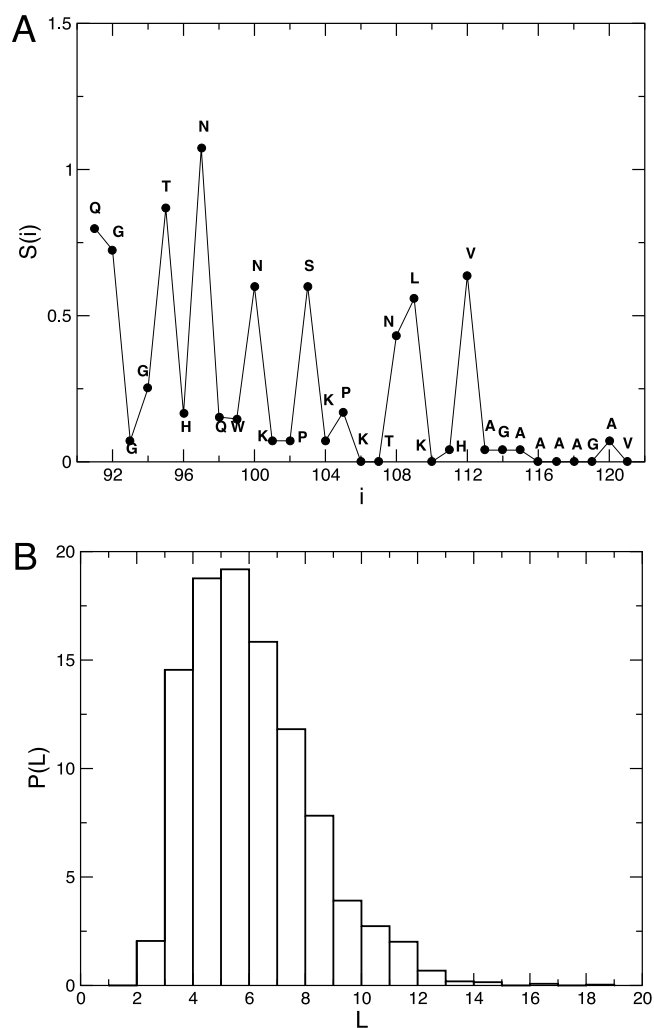
FIGURE 10   (A) Sequence entropy for 91–121 (mouse prion numbering) obtained from an alignment of 150 prion sequences using Eq. 7. The sequence is given in the one-letter code format. (B) The distribution of lengths in 5265 strands from 423 single-chain proteins in PDBselect.

prion sequences, aligned against mPrP, was measured using the sequence entropy,

$$S(i) = -\sum_{\alpha=1}^{20} p_\alpha(i)\ln p_\alpha(i), \qquad (7)$$

where $p_\alpha(i)$ is the probability of finding the amino acid type α at position i in the 150 aligned sequences. If a residue is perfectly conserved, then $S(i) = 0$, and we assume that a very highly conserved residue has $0 < S(i) \leq 0.1$. With this criterion, we find (Fig. 10 A) that in excess of 60% of residues in this region are either perfectly or very highly conserved. More importantly, the ends of this region (112–120), containing a series of Ala, are perfectly conserved. These are unlikely to be in a β-strand.

Second, to satisfy the secondary structure content of PrP^{Sc}(90–231), it would be necessary to have very long

strands. We computed the length distribution of 5265 β-strands (Fig. 10 *B*) in 423 single-chain proteins from PDBselect. The most likely strand length is about 6, with the fraction of strands with length >8 being only 9.8%. Because the strands have to be connected by loops, it follows that not every one of the residues in 90–120 is likely to be in a strand conformation. To account for the 68 residues in PrP^{Sc}(90–231) that need to be in the strand conformation, the region involving the transition to β-strand has to extend beyond the 90–120 peptide. Given the near-perfect conservation of the loop 1 residues (data not shown), the probability of a number of residues in H1 to be helical (according to Chou–Fasman rules or PHD prediction), and the length restriction of β-strands, it is logical to suggest that there has to be a number of β-strands connected by loops in the region 121–170 in the fibril. Because a certain number of residues has to be in loops, it is difficult to imagine having all 68 residues that have to be in strands confined to the region preceding H2. Such a stringent restriction can be relieved by having the most frustrated part of H2 (C-terminal residues) to be in an altered state in PrP^{Sc}. This accounting also provides indirect support to the proposal here and elsewhere (Kallberg et al., 2001) concerning the fate of H2 in the fibril.

Unlikelihood of PrP^{Sc} formation in avian prions

It has been suggested (Simonic et al., 2000) that it is unlikely that nonmammalian PrP^C undergo the α→β conformational transition. Our model would imply that, in these sequences, H2 has higher helical propensity than in mammalian prions. By sequence alignments, we have identified the helical region of avian prions (e.g., chicken, duck, crane, condor). The results show, for example, that H2 in the avian prions contains several alanines that are known to be excellent helix formers. The typical value of $\lambda_H \approx 0.08$ (Eq. 3) for these prions is consistent within range of what is expected for normal proteins (Fig. 8). According to AGADIR (Munoz and Serrano, 1994), the helical propensities for H1, H2, and H3 in chicken prions, using the default settings (278 K, pH = 7.0, ionic strength = 0.1) and extending each helix by three residues at each end, are 3.2%, 0.7%, and 9.4%, respectively. For mPrP, the helical contents for H1, H2, and H3 are predicted to be 27%, 0.3%, and 8.4%, respectively. There is a dramatic decrease in the helical content of H1 compared to the mPrP case, whereas the helical content of both H2 and H3 increases compared to mPrP. PHD prediction gives $S_{\alpha/\beta}$ (Eq.6) ≈ 0.57 for H2, which shows that avian helices are not as frustrated as their mammalian counterparts. The enhanced helical propensity of H2 in chicken prions compared to the mPrP explains the lack of α→β change in the former. This adds further support to our proposition that H1 and H2 serve as plausible regions for the α→β transition in mammalian prions.

CONCLUSIONS

We have explored plausible structural abnormalities at the residue level in PrP^C using available NMR structures and sequence alignments. Although clear signals of abnormal behavior at various locations can be gleaned, it is clear that this kind of analysis alone cannot explain all of the biophysical characteristics of PrP^C→PrP^{Sc} conversion. To account for the percentage of helix and sheet content in the protease-resistant prion particle, it is necessary to argue that the disordered N-terminus (residues 90–120) plays a key role in the PrP^C→PrP^{Sc} transition (Peretz et al., 1997). However, as shown here and by Hornemann and Glockshuber (1998), this alone is not sufficient to explain the gross secondary-structure contents of the prion particle. Conformational rearrangements in the ordered C-terminus region are required to produce a priogenic intermediate. The role played by the C-terminus region can be probed by searching for conditions under which the peptide fragment PrP(121–231) forms fibrils.

REFERENCES

Aurora, R., and G. D. Rose. 1998. Helix capping. *Protein Sci.* 7:21–38.

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

Baud, F., and S. Karlin. 1999. Measures of residue density in protein structures. *Proc. Natl. Acad. Sci. U.S.A.* 96:12494–12499.

Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Sindyalov, and P. E. Bourne. 2000. The Protein Databank. *Nucleic Acids Res.* 28:235–242.

Billeter, M., R. Riek, G. Wider, S. Hornemann, R. Glockshuber, and K. Wuthrich. 1995. Prion protein NMR structure and species barrier for prion diseases. *Proc. Natl. Acad. Sci. U.S.A.* 94:7281–7285.

Bowie, J. U., R. Luthy, and D. A. Eisenberg. 1991. Method to identify protein sequences that fold into a known three-dimensional structure. *Science.* 253:164–170.

Caughey, B. W., A. Dong, K. S. Bhat, D. Ernst, S. F. Hayes, and W. S. Caughey. 1991. Secondary structure-analysis of the scrapie-associated protein PrP 27–30 in water by infrared spectroscopy. *Biochemistry.* 30:7672–7680.

Chan, C. K., Y. Hu, S. Takahashi, D. L. Rousseau, W. A. Eaton, and J. Hofrichter. 1997. Submillisecond protein folding kinetics studied by ultrarapid mixing. *Proc. Natl. Acad. Sci. U.S.A.* 94:1779–1784.

Chou, P. Y., and G. D. Fasman. 1978. Empirical predictions of protein conformation. *Ann. Rev. Biochem.* 47:251–276.

Cohen, F. E. 1999. Protein misfolding and the prion diseases. *J. Mol. Biol.* 293:313–320.

Cohen, F. E., K. M. Pan, Z. Huang, M. Baldwin, R. J. Fletterick, and S. B. Prusiner. 1994. Structural clues to prion replication. *Science.* 264:530–531.

Cohen, F. E., and S. B. Prusiner. 1998. Pathological conformations of prion proteins. *Ann. Rev. Biochem.* 67:793–819.

Creighton, T. E. 1993. Proteins: Structures and Molecular Properties. W. H. Freeman and Co., New York. 142.

Donne, D. G., J. H. Viles, D. Groth, I. Mehlhorn, T. L. James, F. E. Cohen, S. B. Prusiner, P. E. Wright, and H. J. Dyson. 1997. Structure of the recombinant full-length hamster prion protein PrP(29–231): the N ter-minus is highly flexible. *Proc. Natl. Acad. Sci. U.S.A.* 94:13452–13457.

Gasset, M., M. A. Baldwin, R. J. Fletterick, and S. B. Prusiner. 1993. Perturbation of the secondary-structure of the scrapie prion protein under conditions that alter infectivity. *Proc. Natl. Acad. Sci. U.S.A.* 90:1–5.

Hooft, R. W. W., G. Vriend, and C. Sander. 1996. Errors in protein structures. *Nature (Lond.).* 381:272–273.

Hornemann, S., and R. Glockshuber. 1998. A scrapie unfolding interme-diate of the prion protein domain PrP(121–231) induced by acidic pH. *Proc. Natl. Acad. Sci. U.S.A.* 95:6010–6014.

Hosszu, L. P., N. J. Baxter, G. S. Jackson, A. Power, A. R. Clarke, J. P. Waltho, C. J. Craven, and J. Collinge. 1999. Structural mobility of the human prion protein probed by backbone hydrogen exchange. *Nature Struct. Biol.* 6:740–743.

Istrail, S., R. Schwartz, and J. King. 1999. Lattice simulations of protein aggregation. *J. Comp. Biol.* 6:143–162.

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.

Jiang, B., T. Guo, L. W. Peng, and Z. R. Sun. 1998. Folding type-specific secondary structure propensities of amino acids, derived from α-helical, β-sheet, α/β, and $\alpha+\beta$ proteins of known structures. *Biopolymers.* 45:35–49.

Kallberg, Y., M. Gustafsson, B. Persson, J. Thyberg, and J. Johansson. 2001. Prediction of amyloid fibril-forming proteins. *J. Biol. Chem.* 276:12945–12950.

Lee, B., and F. M. Richards. 1971. The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* 55:379–385.

Liemann, S., and R. Glockshuber. 1999. Influence of amino acid substi-tutions related to inherited human prion diseases on the thermodynamic stability of the cellular prion protein. *Biochemistry.* 38:3258–3267.

Luthy, R., J. U. Bowie, and D. Eisenberg. 1992. Assessment of protein models with three-dimensional profiles. *Nature (Lond.).* 365:83–85.

McDonald, I., and J. M. Thornton. 1994. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* 238:777–793.

Michelitsch, M. D., and J. S. Weissman. 2000. A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions. *Proc. Natl. Acad. Sci. U.S.A.* 97:11910–11915.

Morillas, M., D. L. Vanik, and W. K. Surewicz. 2001. On the mechanism of α-helix to β-sheet transition in the recombinant prion protein. *Bio-chemistry.* 40:6982–6987.

Morrisey, M. P., and E. I. Shakhnovich. 1999. Evidence for the role of PrP C helix 1 in the hydrophilic seeding of prion aggregates. *Proc. Natl. Acad. Sci. U.S.A.* 96:11293–11298.

Munoz, V., and L. Serrano. 1994. Elucidating the folding problem of helical peptides using empirical parameters. *Nature Struct. Biol.* 1:399–409.

Murzin, A. G., S. E. Brenner, T. Hubbard, and C. Chothia. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247:536–540.

Orengo, C. A., A. D. Michie, S. Jones, D. T. Jones, M. B. Swindells, and J. M. Thornton. 1997. CATH—a hierarchic classification of protein domain structures. *Structure.* 5:1093–1108.

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

Peretz, D., R. A. Williamson, Y. Matsunaga, H. Serban, C. Pinilla, R. B. 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. *Ann. Biochemistry.* 236:63–73.

Prusiner, S. B. 1997. Prion diseases and the BSE crisis. *Science.* 278:245–251.

Prusiner, S. B. 1998. Prions. *Proc. Natl. Acad. Sci. U.S.A.* 95:13363–13383.

Riek, R., S. Hornemann, G. Wider, M. Billeter, R. Glockshuber, and K. Wuthrich. 1996. NMR structure of the mouse prion protein domain PrP(121–231). *Nature (Lond.).* 382:180–182.

Riek, R., S. Hornemann, G. Wider, M. Billeter, R. Glockshuber, and K. Wuthrich. 1997. NMR characterization of the full-length recombinant murine prion protein, mPrP(23–231). *FEBS Lett.* 413:282–288.

Rost, B., and C. Sander. 1993. Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol.* 232:584–599.

Sander, C., and R. Schneider. 1991. Database of homology derived protein structures and the structural meaning of sequence alignment. *Proteins Struct. Funct. Genet.* 9:56–68.

Simonic, T., S. Duga, B. Strumbo, R. Asselta, F. Ceciliani, and F. Ronchi. 2000. cDNA cloning of turtle prion protein. *FEBS Lett.* 469:33–38.

Sparrer, H. E., A. Santoso, F. C. J. Szoka, and J. S. Weissman. 2000. Evidence for the prion hypothesis: induction of the yeast [Psi$^+$] factor by in vitro-converted Sup35 protein. *Science.* 289:595–599.

Srinivasan, R., and G. D. Rose. 1995. LINUS: a hierarchic procedure to predict the fold of a protein. *Proteins Struct. Funct. Genet.* 22:81–99.

Telling, G. C., M. Scott, J. Mastrianni, R. Gabizon, M. Torchia, F. E. Cohen, S. J. DeArmond, and S. B. Prusiner. 1995. Prion propagation in mice expressing human and chimaeric PrP trangenes implicates the interaction of cellular PrP with another protein. *Cell.* 83:79–90.

Viles, J. H., D. Donne, G. Kroon, S. B. Prusiner, F. E. Cohen, H. J. Dyson, and P. E. Wright. 2001. Local structural plasticity of the prion proteins: analysis of NMR relaxation dynamics. *Biochemistry.* 40:2743–2753.

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

Wille, H., G. Zhang, M. A. Baldwin, F. E. Cohen, and S. B. Prusiner. 1996. Separation of scrapie prion infectivity from prp amyloid polymers. *J. Mol. Biol.* 259:608–621.

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