Thermodynamics of Model Prions and its Implications for the Problem of Prion Protein Folding

Paul M. Harrison¹, Hue Sun Chan², Stanley B. Prusiner³,⁵ and Fred E. Cohen¹,²,³,⁴,*

Departments of

¹Cellular & Molecular Pharmacology
²Pharmaceutical Chemistry
³Biochemistry & Biophysics
⁴Medicine and ⁵Neurology
University of California
San Francisco, 513 Parnassus Ave, San Francisco CA 94143, USA

Prion disease is caused by the propagation of a particle containing PrPSc, a misfolded form of the normal cellular prion protein (PrP C). PrP C can re-fold to form PrPSc with loss of α-helical structure and formation of extensive β-sheet structure. Here, we model this prion folding problem with a simple, low-resolution lattice model of protein folding. If model proteins are allowed to re-fold upon dimerization, a minor proportion of them (up to ~17%) encrypts an alternative native state as a homodimer. The structures in this homodimeric native state re-arrange so that they are very different in conformation from the monomeric native state. We find that model proteins that are relatively less stable as monomers are more susceptible to the formation of alternative native states as homodimers. These results suggest that less-stable proteins have a greater need for a well-designed energy landscape for protein folding to overcome an increased chance of encrypting substantially different native conformations stabilized by multimeric interactions. This conceptual framework for aberrant folding should be relevant in Alzheimer’s disease and other disorders associated with protein aggregation.

Introduction

A substantial body of evidence demonstrates that the neurodegenerative disorders known as prion diseases are caused by an agent composed largely, if not entirely, of a misfolded form (PrPSc) of the cellular prion protein (PrP C) (Prusiner & DeArmond, 1994; Cohen et al., 1994; Harrison et al., 1997). Three forms of prion disease occur: inherited (e.g. familial Creutzfeldt-Jakob disease (CJD), Gerstmann-Straussler-Scheinker syndrome and fatal familial insomnia in humans), infectious (e.g. kuru and iatrogenic CJD in humans, bovine spongiform encephalopathy in cattle and scrapie in sheep) and sporadic (e.g. sporadic CJD in humans).

The kinetics and thermodynamics of prion misfolding and their relationship to the three forms of prion diseases have been discussed extensively (for reviews, see Cohen & Prusiner, 1998; Prusiner et al., 1998). The link between the misfolding of the prion protein and the actual molecular pathologic events that give rise to prion disease remains to be elucidated fully (Hegde et al., 1998; Prusiner et al., 1998).

PrP C is a predominantly α-helical molecule, whereas PrPSc is rich in β-sheet structure (Pan et al., 1994). No chemical modification has yet been identified by which PrP C and PrPSc differ (Stahl et al., 1992). Experimental studies of a recombinant molecule corresponding to residues 90-231 of PrP indicate that PrP C is marginally stable relative to a folding intermediate (James et al., 1997; Zhang et al., 1997). At high concentrations at 35°C, recombinant PrP(90-231) forms a multimeric complex with substantial β-sheet structure (James et al., 1997; Zhang et al., 1997). Ionizing radiation studies suggest that the minimally infectious multimer of PrPSc is a dimer (Bellinger-Kawahara et al., 1988). These data together with an array of transgenic animal experiments and scrapie inactivation studies (Prusiner et al., 1990; Oesch et al., 1994; Telling et al., 1996a)
et al have argued that PrPSc is stabilized by multimeric interactions either upon conversion and/or through subsequent multimerization (Cohen et al., 1994; Lansbury & Caughey, 1995). A host-specific co-factor, dubbed protein X (which may act as a chaperone), is required for conversion of PrPc to PrPSc (Telling et al., 1996a; Kaneko et al., 1997). Strains of prions for PrPSc have been identified with distinct incubation times that result in different neurohistopathology (Prusiner & DeArmond, 1994). Recent work has shown that the biological properties of prion strains are likely to be enciphered in the tertiary structure of PrPSc (Telling et al., 1996a; Collinge et al., 1996). There are at least two to four PrPSc strain conformations and possibly more (Bessen et al., 1995; Collinge et al., 1996; Parchi et al., 1996; Safar et al., 1998).

Further examples of prions, the [PSI] and [URE3] non-Mendelian inheritance phenomena, have been found in the yeast Saccharomyces cerevisiae (Lindquist, 1997), and a likely prion candidate [Het-s] has been observed in the fungus Podospora anserina (Wickner, 1997). Experimental evidence suggests that the [PSI] and [URE3] phenomena in yeast are caused by propagatable isoforms of normal functioning cellular proteins (Sup35p and Ure2p, respectively; Lindquist, 1997). Despite the very different biological contexts, there are some notable specific similarities between mammalian PrP prion and yeast prion propagation. As in PrP prion propagation, the [PSI] phenomenon requires a co-factor in vivo. Intermediate concentrations of the Hsp104 chaperone protein are necessary for efficient propagation of [PSI] in vivo (Chernoff et al., 1995). In vitro, the Sup35p/[PSI] prion domain converts into β-sheet-rich amyloid-like filaments (King et al., 1997). Furthermore, there is evidence for different strains of [PSI] that can be induced by overproduction of the same Sup35p protein (Derkatch et al., 1996). For the [URE3] prion, partial protease resistance is observed as in the case of PrPSc, but not for the non-prion phenotype of the Ure2p protein (Masison et al., 1997).

To understand prion propagation better, we have designed the simplest thermodynamic situation that applies to any prion propagation model: monomeric native conformations and re-folded dimeric native conformations stabilized by interfacial interactions for the same protein sequence. This thermodynamic situation is implemented using a simple model of protein folding, the HP lattice model. While this model is low-resolution both spatially and energetically, it has demonstrated many of the characteristics of protein folding and has been useful in understanding the energy landscapes that govern how a protein folds (Dill et al., 1995). Model proteins are defined as sequences that have a unique monomeric native conformation. Those model proteins that have conformationally distinct monomeric and homodimeric native states are used as model prions. When the number of conformations in the homodimeric native state is >1, it is said to comprise a set of prion strain conformations. All other model proteins are treated as canonical or Anfinsenian proteins. Anfinsen’s hypothesis states that “the three-dimensional structure of a native protein in its normal physiological milieu is the one in which the Gibbs free energy of the whole system is lowest; that is, the native conformation is determined by the totality of inter-atomic interactions and hence by the amino-acid sequence in a given environment.” (Anfinsen, 1973). Prions are non-Anfinsenian because they are sensitive to fluctuations in their “physiological milieu” that may make dimerization/multimerization more likely. For conditions of the prion sequence where dimerization may be more unlikely (such as low concentration), prions will adopt their monomeric native state.

Our goal is not to study the specific features of prion protein folding, but rather to take a more general perspective and examine unexpected observations that arise from our initial model design. Re-folded homodimeric conformations are studied in particular, but we believe that the effects of higher-order multimerization would, in principle, be similar. We examine properties such as conformational similarity and thermodynamic stability for Anfinsenian and prion model proteins with unique stable monomeric conformations†. The biological relevance of our results is examined. Thermodynamic stability is found to be a key factor in determining susceptibility to an alternative native state.

### Results and Discussion

To define our model proteins and prions, we have studied 3D 27-mer maximally compact cubic conformations (Chan & Dill, 1990; Skhakhnovich & Gutin, 1990; Sali et al., 1994; Li et al., 1996) and 2D n-mers for a range of chain lengths n from 11 to 16 (Chan & Dill, 1990, 1991, 1996). These systems have previously been examined extensively with regard to other properties of protein folding (Dill et al., 1995). The 2D studies allow us to examine the effects of variation of chain length and of surface complementarity in the dimer, whereas the cubic conformations have the dimensionality of real proteins but the possible dimer packing geometries are much more limited. Unless specified otherwise, results below are for the 27-mer cubes.

For the 3D cubic conformations, 14,906 model proteins were found in a sample of 250,000 ran-

---

† Here the term unique refers to the existence of a single lowest-energy native conformation for a sequence. It does not imply that the sequence always adopts that particular conformation. Even under strongly folding conditions, conformational fluctuation is possible and the sequence can sometimes adopt non-native conformations (Dill & Chan, 1997).
domly generated HP sequences. When these model proteins were re-evaluated to find homodimeric native conformations, 1110 prions were found, with the remainder of the model proteins comprising Anfinsenian sequences. Similarly in 2D, all of the model proteins for a particular chain length were re-evaluated for homodimer native conformations (the enumeration procedures are described in Methods). These enumeration results are detailed in full in Table 1. The prion and Anfinsenian sequences have been analyzed and compared for the properties of hydrophobicity, secondary structure, conformational similarity and thermodynamic stability.

**Proportion of sequences that are prions**

Model protein sequences that exhibit prion-like behaviour make up a minor fraction of all possible sequences in the HP model (2 to 17%; Table 1). Those that encrypt a single strain conformation account for 2 to 7% of model-protein sequences. Interestingly, in the present model, this is of the same order as the percentage of all possible sequences that turn out to have unique native-state conformations as monomers (Table 1). Examples of model 3D and 2D prions are depicted in Figures 1 and 2, respectively.

**Hydrophobicity of the sequences**

Is a predominantly hydrophobic or hydrophilic sequence more likely to form an alternative native state? Fractional hydrophobicities (Φ) were calculated for the prion-determinant sequence fragments of PrP, Sup35p and Ure2p, and for a representative sample of protein sequences from the PDB database (Bernstein et al., 1977; Table 2). The minimal infectious sequence fragment of PrP (residues 90-231) is not particularly hydrophobic or hydrophilic, compared to the representative sample of protein sequences (Table 2). The monomeric form of PrP is soluble at concentrations sufficient for NMR spectroscopy (Donne et al., 1997; Riek et al., 1997), but the PrPSc aggregate is highly insoluble. By contrast, the yeast prion sequence fragments for

---

**Table 1. Numbers of model prions for 2D 11 to 16-mers and 3D 27-mers and their hydrophobicities**

<table>
<thead>
<tr>
<th>Chain length</th>
<th>Total number of sequences</th>
<th>Number of model-protein sequences</th>
<th>Number of prions with only one strain</th>
<th>Mean fractional hydrophobicity (Φ)</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>N&lt;sub&gt;M&lt;/sub&gt; (N&lt;sub&gt;total&lt;/sub&gt;)</td>
<td>N&lt;sub&gt;prions&lt;/sub&gt; (N&lt;sub&gt;M&lt;/sub&gt;)</td>
<td>Φ&lt;sub&gt;all&lt;/sub&gt; (prions)</td>
<td>Φ&lt;sub&gt;1&lt;/sub&gt; (prions, strains = 1)</td>
</tr>
<tr>
<td>3D data</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27 (28)</td>
<td>14,906 (7.4)</td>
<td>236/14,906 (1.6)</td>
<td>0.48 ± 0.08</td>
<td>0.48 ± 0.06</td>
</tr>
<tr>
<td>16 (7)</td>
<td>1539 (6.5)</td>
<td>54/1539 (3.5)</td>
<td>0.54 ± 0.12</td>
<td>0.54 ± 0.12</td>
</tr>
<tr>
<td>15 (6)</td>
<td>952 (5.8)</td>
<td>44/952 (4.6)</td>
<td>0.55 ± 0.09</td>
<td>0.55 ± 0.09</td>
</tr>
<tr>
<td>14 (6)</td>
<td>893 (5.2)</td>
<td>40/893 (4.5)</td>
<td>0.56 ± 0.11</td>
<td>0.56 ± 0.11</td>
</tr>
<tr>
<td>13 (5)</td>
<td>873 (4.9)</td>
<td>38/873 (4.3)</td>
<td>0.57 ± 0.12</td>
<td>0.57 ± 0.12</td>
</tr>
<tr>
<td>12 (5)</td>
<td>846 (4.5)</td>
<td>33/846 (3.9)</td>
<td>0.58 ± 0.13</td>
<td>0.58 ± 0.13</td>
</tr>
<tr>
<td>11 (4)</td>
<td>745 (4.1)</td>
<td>27/745 (3.6)</td>
<td>0.59 ± 0.14</td>
<td>0.59 ± 0.14</td>
</tr>
</tbody>
</table>

---

*a For n = 11, 12 and 13 in 2D, homodimer enumerations have also been extended to include all possible conformations. For these, the prion fraction of model-protein sequences increases to: 12% (11-mers), 13% (12-mers) and 25% (13-mers).

*b Chain length n and (in parentheses) minimum number of intra-chain contacts in the unique monomeric native conformations.

*c The number of these that are model-protein sequences (N<sub>M</sub>).

*d The number of the model-protein sequences that are prions (N<sub>prions</sub>).

*e The number of the model-protein sequences that are prions with number of strains = 1.

*f Φ for all prions.

*g Φ for prions with number of strains = 1.
Sup35p and Ure2p are dominated by polar/charged residues. These sequences have very high proportions of glutamine (27 % for Sup35p and 11 % for Ure2p) and asparagine (18 % for Sup35p and 37 % for Ure2p), giving them a hydrophilicity that is unusual for globular protein domains (Table 2). Thus, the composition of the small set of currently known sequences capable of forming prions shows no general obvious distinction in terms of sequence hydrophobicity.

We also examined the model prion and Anfinse-nian sequences for compositional bias. The fractional hydrophobicity (Φ) of an HP model sequence is simply the number of H residues in the sequence divided by the chain length. Mean hydrophobicity values (Φ) for the prions and for the Anfinsen proteins as populations were calculated (Table 1). The prion sequences are generally no more or less hydrophobic than Anfinseinian protein sequences. However, each model prion sequence uses what hydrophobic residues it possesses to stabilize its homodimeric native-state conformations by forming more hydrophobic contacts at a binding interface. For example, for the 3D model prions with a single strain, on average 24(±3) % of the hydrophobic contacts in prion homodimers are interfacial, compared to 18(±3) % for lowest-energy homodimers made from the monomeric native-state conformation.

### Conformational similarity of the monomeric and homodimeric native states

Spectroscopic and immunologic studies have shown that there is an extensive conformational rearrangement between PrP\(^{C}\) and PrP\(^{Sc}\) that is largely confined to the middle third of the PrP sequence (Pan et al., 1994; Peretz et al., 1997). The conformational rearrangement of our lattice prions on going from the monomeric native state to the homodimeric native state has been examined. Figure 3(a) shows the trend in conformational similarity (defined in Methods) between the monomeric and homodimeric native states (S\(_{3D}\)\(^{mono-di}\)) for the 3D 27-mer sequences, with a single homodimeric native-state (degeneracy = 1). The first peak at S\(_{3D}\)\(^{mono-di}\) = 1.0 is for the Anfinse-nian proteins, whereas the second, much smaller peak is for the prions. The lattice prion monomeric and homodimeric native-state conformations tend to be substantially different. In 3D, they are almost as dissimilar as any two randomly chosen compact conformations (Table 3). This trend in S\(_{3D}\)\(^{mono-dir}\) with a peak at close to random values, changes little for the prions with multiple stable dimeric native-state conformations or strains. The mean S\(_{3D}\)\(^{mono-di}\) values range between 0.28 and 0.31 for sequences capable of forming two to eight strains. A similar trend in conformational similarity (S\(_{2D}\)\(^{mono-di}\)) was observed

### Table 2. Fractional hydrophobicities (Φ) for the prion-determinant sequence fragments and the PDB

<table>
<thead>
<tr>
<th>Protein or protein set</th>
<th>Φ (Miyazawa &amp; Jernigan, 1996)</th>
<th>Φ (Fauchere &amp; Pliska, 1983)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDB July 1997 (non-redundant set) (mean)</td>
<td>0.411(±0.037)</td>
<td>0.416(±0.036)</td>
</tr>
<tr>
<td>PrP(^{C}) (90-231)</td>
<td>0.370</td>
<td>0.372</td>
</tr>
<tr>
<td>Sup35p (1-125)</td>
<td>0.269*</td>
<td>0.335*</td>
</tr>
<tr>
<td>Ure2p (1-90)</td>
<td>0.282</td>
<td>0.286</td>
</tr>
</tbody>
</table>

* The values differ here because of the proportion of proline in the sequence, which is ranked differently in the two scales because of its secondary-structure propensities.
for the longer 2D chains (15-mers and 16-mers) examined. However, the similarity values for the prions in 2D are not as close to the random values as their 3D counterparts (Table 3). In 2D, some sequences that form distinct conformations can have the same contact maps. This situation occurs only for enantiomeric pairs in 3D. Also, there may be a dimensionality effect for this similarity measure. These changes in conformation observed in the lattice setting are consistent with the spectroscopic results on the structures of PrPSc and PrPSc (Pan et al., 1994).

**Table 3.** Conformational similarities between the monomeric and homodimeric native states for 3D 27-mer cubes and the longer 2D chains

<table>
<thead>
<tr>
<th>Chain length</th>
<th>PrPSc (3D)</th>
<th>PrPSc (2D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>27-mer cubes/native</td>
<td>0.31±0.10</td>
<td>0.18±0.08</td>
</tr>
<tr>
<td>27-mer cubes/random</td>
<td>0.19±0.08</td>
<td>0.18±0.15</td>
</tr>
<tr>
<td>15-mers/native</td>
<td>0.57±0.19</td>
<td>0.59±0.17</td>
</tr>
<tr>
<td>15-mers/random</td>
<td>0.18±0.16</td>
<td>0.18±0.15</td>
</tr>
<tr>
<td>16-mers/native</td>
<td>0.59±0.17</td>
<td>0.59±0.17</td>
</tr>
<tr>
<td>16-mers/random</td>
<td>0.18±0.15</td>
<td>0.18±0.15</td>
</tr>
</tbody>
</table>

The Thermodynamics of Being a Model Prion

Conformational similarity of prion strain conformations

The PrPSc molecules in different prion strains are conformationally distinct as judged by their stability to limited proteolysis (Collinge et al., 1996; Telling et al., 1996a). PrPSc isolated after proteolysis and deglycosylation of infectious material derived from patients with Creutzfeldt-Jakob disease and fatal familial insomnia, had different sizes (19 kDa and 21 kDa, respectively) that bred true upon subsequent passage through mice carrying identical prion transgenes (Telling et al., 1996a). Collinge et al. (1996) have shown that, after treatment with proteinase K, PrPSc from the new variant CJD strain in humans was smaller than PrPSc from previously characterised prion strains.

How similar do multiple lattice prion strain conformations tend to be? The trend in $S_{\text{di-di}}^{\text{3D}}$ for lattice prion strain conformations is illustrated by Figure 3(b). The distribution shows two major peaks, one for slight relative rearrangements of conformation and the other for alternative packings of the same conformation. When compared with Figure 3(a), this shows that lattice prion strains tend to be substantially more similar in conformation with each other than with the monomeric native structure. For the 3D cubic conformations, the distributions for $S_{\text{di-di}}^{\text{3D}}$ for prions capable of forming three to eight strains were also derived. These also demonstrate two major peaks. Ignoring the example of alternative half-dimer packing, where $S_{\text{di-di}}^{\text{3D}} = 1.0$, $S_{\text{di-di}}^{\text{3D}}$ consistently tends toward larger mean values than $S_{\text{mono-di}}^{\text{3D}}$ ($S_{\text{di-di}}^{\text{3D}}$ in the range 0.48 ± 0.19 to 0.64 ± 0.19 for two to eight strains).

The nucleated polymerization model for prion propagation suggests that different strains arise from different alternative packings of PrPSc molecules (Lansbury & Caughey, 1995). In general here, strains do not arise solely from alternative packing arrangements of the same half-dimer conformation (Figure 3(b)). The peak for alternative packing arrangements (at $S_{\text{di-di}}^{\text{3D}} = 1.0$) is virtually absent for the 2D investigations, because they
entail more complex surface complementarity. We expect that the surface complementarity of proteins that dimerize will be better represented by the irregular interfaces seen in 2D, rather than by the smooth packing of two cubes. For sequences that support two strains, there are two examples for the 2D 16-mers (out of 60, 3.3%) and no example for the 15-mers (out of 25). To the extent that this model is relevant, it seems unlikely that multiple packing arrangements that differ at a low degree of resolution can solely explain the existence of multiple prion strains. A more likely scenario to explain the experimental observation that some prion strains have distinct proteolytic degradation patterns is that a local rearrangement of the chain preserves the overall morphology of the multimer (Cohen & Prusiner, 1998).

A well-described example of conformational degeneracy in multimeric proteins is domain-swapping (Bennett et al., 1995). In domain swapping, a segment of one protein chain is replaced with the corresponding segment of a second protein chain, and vice versa. This has been suggested as an explanation of some protein aggregation phenomena (Bennett et al., 1995), where alternative conformations that arise in aggregates are formed from various domain-swapping arrangements. It also arises for lattice prion strains, as indicated in the example in Figure 4. However, different lattice prion strain conformations may also involve different overall shapes for the dimerized structures that entail some other form of symmetric rearrangement (Figure 4).

**Secondary structure**

Substantially increased β-sheet content is observed in the PrP<sup>C</sup>-PrP<sup>Sc</sup> conversion. The change in β-sheet is from 3% to 43% for converting PrP<sup>C</sup> to PrP<sup>Sc</sup> as determined by FTIR spectroscopy. There is a concomitant loss of α-helical structure from 42% to 30% during the transform-

---

**Figure 4.** A 2D prion-strain set example with six conformations that demonstrates two ways to make distinct native-state strain conformations (domain or residue-swapping, or some other form of symmetric rearrangement within the homodimer). The monomeric native-state conformation is also illustrated in (a) and the set of six strain conformations is illustrated in (b). This example was found by exhaustive enumerations over all possible chain conformations for 13-mers (see Methods).
ation (Pan et al., 1994). In other cases of monomer-multimer dimorphism, such as inclusion body formation with β-lactamase or in amyloidogenesis of mutant lysozymes, an increase in β-sheet formation is noted (Przybycien et al., 1994; Booth et al., 1997).

Sheet formation was examined for the lattice prions (Figure 5). Definitions of sheet contacts are taken from previous lattice work (Chan & Dill, 1990, 1991). A sheet is simply defined as two consecutive parallel or antiparallel contacts on the lattice in 2D and three in 3D (Chan & Dill, 1990, 1991). For the 3D 27-mer cubes, the prion homodimer binding interfaces exhibit a modest increase in sheet content (3.4 ± 2.3 sheet contacts per nine-contact interface). The distribution of the change in sheet content for the remainder of the conformation is approximately Gaussian about zero (data not shown). However, examination of the data for the longer 2D chains (15 and 16-mer) indicates that extensive dimer sheet is enabled by formation of a large dimer interface and the change in shape that accompanies it (Figure 5). The mean change in the percentage of total sheet from the monomer to the homodimer half-dimer conformation is +26(±26)% for the 16-mers, and +27(±24)% for the 15-mers.

**Thermodynamic stability**

We have evaluated the importance of stability for the behaviour of our model proteins and prions. Stabilities (ΔG\textsubscript{monomer} defined in Methods) were calculated for the 27-mer cubes for the monomeric native states of all of the model proteins (Figure 6). The stability results presented here are independent of the value of the weighting factor α (defined in Methods), which is inversely proportional to the simulation temperature as used in folding simulations such as those performed by Sali et al. (1994) and Unger & Moult (1996).

For Anfinsenian lattice proteins, the energy difference between the native state and the lower-energy non-native conformations is substantial (Figure 6). However, the lattice prions in their monomeric native state tend to be less stable than their Anfinsenian counterparts (Figure 6). The vast majority (98.0%) of the homodimeric native states have their conformations residing in the lowest non-native energy level (E\textsubscript{native}+1) of their respective monomeric native-state energy spectra. Also, stabilities in this model are predominantly determined by the number of conformations in the lowest non-native energy levels (particularly the E\textsubscript{native}+1 level; Shortle et al., 1992).

![Figure 5](image-url)

**Figure 5.** The change in the total proportion of sheet contacts (Δsheet) for the monomer and half-dimer conformations, for the 2D 16-mers, including the interface contacts. This is calculated as: Δ\text{sheet} = (number of sheet contacts made by the half-dimer/total number of contacts made by the half-dimer) – (number of sheet contacts made by the monomer/total number of contacts made by the monomer). Sheets have a minimum size of two parallel or antiparallel consecutive contacts (see Chan & Dill, 1991). A similar trend is observed for the 2D 15-mers.
These observations taken together imply that sequences are more likely to form prions if the lowest non-native energy levels are relatively highly populated. There is therefore a statistical link between monomeric native-state stability and susceptibility to the prion phenomenon. Although we do not include kinetic simulation in the current study, previous studies using similar models show that these low-energy non-native conformations (i.e. conformationally distant relatives; see Dill & Chan, 1997) would likely correspond to kinetic traps in which relatively poorly designed sequences tend to accumulate during folding (Thirumalai & Woodson, 1996). Therefore, well-designed folding kinetics for the monomers is necessary to avoid these low-energy conformations that tend to be substantially different from the monomer, but may be encryptable in a multimeric form with the formation of a few intermolecular interactions. A dynamic evaluation of the kinetic barriers on this landscape will be required to address this issue.

Comprehensive analysis of single-site mutations

Point mutations to the PrP gene are known to cause prion diseases with an autosomal dominant pattern of inheritance (Prusiner & DeArmond, 1994). The simplicity of the HP model enables examination of all possible single-site mutations for a sequence (27 in total for the cubic model proteins). The effect of these mutations on stability has been studied for the model cubic prions and Anfinsenian proteins. Only coding mutations for the monomeric native state are examined (Figure 7). Coding mutations for the monomeric native state are mutations that maintain this conformation as the unique native state conformation. A minor fraction (15-16%) of all single-site mutations are coding in this model.

For monomeric native states of the prion sequences studied, ~70.0% of the coding single-site mutations are stabilizing and ~30.0% are destabilizing (Figure 7). For the Anfinsen sequences, ~50.0% are stabilizing (Figure 7). Stabilizing mutations for prions are linked to loss of prion status (and conversely, destabilizing mutations for Anfinsenian sequences predispose towards prion formation; Figure 7). Mutations from prion to prion and from Anfinsenian sequence to Anfinsenian sequence tend to be equally stabilizing/destabilizing (Table 4). Conversely, mutations from prion to Anfinsenian sequence tend to be stabilizing ~70% of the time and mutations from Anfinsenian sequence to prion
Bernoulli trials were performed to compare the proportion of (de)stabilizing mutations between either of the prions populations (i) and (ii), and either of populations (iii) and (iv). They show that the difference in this proportion is highly significant ($p < 0.005$, where $p$ is the probability that the proportions are the same). These proportions are not dependent on the number of strain conformations (as the percentages for number of strains = 1 and > 1 are similar).

Conclusions

The substantial conformational change that underlies prion diseases arises quite logically in a simple model of protein folding. In the HP model, a minor proportion of model proteins can encrypt an alternative native multimeric state and do not necessarily need a very hydrophobic or hydrophilic sequence to effect this change. This is consistent with observations on real prions: while PrP is of intermediate hydrophobicity, the yeast prions Ure2p and Sup35p are more hydrophilic than the average protein of known structure.

We have shown that model prion monomeric native states tend to be less stable. Since lack of stability implies the existence of more low-energy meta-stable conformations, there is a greater chance that some of these meta-stable conformations will be stabilized by multimeric interactions. This implies that susceptibility to prion formation has a probabilistic basis and that more examples of prion-like behaviour are likely to be found for real proteins, particularly those of marginal stability. Prion sequences may sample non-native/unfolded conformations with greater frequency. Factors that increase the likelihood of multimer formation, including the introduction of a multimeric template, or mutations or the expression of high concentrations of the protein, will distort the energy landscape (Dill & Chan, 1997) for the folding/refolding of the chain for prion sequences. An experimental analogy for PrP would be that inoculation of PrPSc and overexpression of PrP in transgenic animals lead to experimental disease (Cohen & Prusiner, 1998).

Surprisingly, the monomeric native-state conformations of the prions tend to undergo substantial rearrangement to form the conformations of their homodimeric native states. There is thus a greater need for a well-designed energy landscape for the folding of the normal monomeric form of a less-stable protein, so that the greater chance of becoming
<table>
<thead>
<tr>
<th>Sequence group</th>
<th>Class of mutation</th>
<th>Number and percentage that are stabilizing for prion-to-prion mutations</th>
<th>Number and percentage that are stabilizing for prion-to-Anfinsenian protein mutations</th>
<th>Number and percentage that are stabilizing for Anfinsenian protein-to-prion mutations</th>
<th>Number and percentage that are stabilizing for Anfinsenian protein-to-Anfinsenian protein mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prions forming a single strain conformation</td>
<td>97/188 (51.6%)</td>
<td>592/807 (73.4%)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>All prions</td>
<td>519/997 (52.1%)</td>
<td>2556/3514 (72.7%)</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Random sample of 1200 model-protein sequences</td>
<td>35/76 (46.0%)</td>
<td>181/263 (68.8%)</td>
<td>91/285 (31.9%)</td>
<td>2384/4769 (50.0%)</td>
<td>2462/5030 (48.9%)</td>
</tr>
<tr>
<td>Random sample of 1200 Anfinsenian proteins with a single dimer in the homodimeric native state</td>
<td>N/A</td>
<td>N/A</td>
<td>72/187 (38.5%)</td>
<td>2462/5030 (48.9%)</td>
<td>2462/5030 (48.9%)</td>
</tr>
</tbody>
</table>
ing a prion can be overcome. In a sense, the normal functioning native form of a protein can be seen as a kinetic trap on the energy landscape of the alternative multimeric native form.

A recent lattice investigation by Abkevich et al. (1998) has relevance to prions. In their study, they envision the misfolded form of prion proteins as originating from kinetic partitioning into possible iso-energetic conformations (i.e. the cellular and misfolded forms), which in their model are assumed to have essentially the same energetic favorability. In contrast to the present analysis, Abkevich et al. (1998) do not consider dimerization or aggregation as mechanisms for enhancing the thermodynamic favorabilities of the disease-causing forms of prion proteins.

Methods

Enumerations to find the native states for sequences as a monomer and as a homodimer

In the HP lattice model, sequences that are made from two residue types (hydrophobic H and polar P) are configured on a three-dimensional (3D) cubic or two-dimensional (2D) square lattice. The HP lattice model is implemented as previously described (Chan & Dill, 1991). For the HP model, in both 2D and 3D, interactions H-H, H-P and P-P are scored energetically as \( e_{HH} = 1; e_{HP} = e_{PP} = 0 \), where \( e_{AB} \) is the score assigned to the interaction/contact between residue types A and B. The total energy for a conformation, \( E \), is \( \Sigma e_{AB} \), where the sum is over all interactions for the conformation. HP sequences that have a single native-state conformation as a monomer are used as model proteins.

In 2D, these model protein sequences have previously been determined by exhaustive enumeration over all possible conformers (Chan & Dill, 1991, 1996). Here, all 2D HP model-protein sequences with chain length \( n = 11-16 \) are submitted to a second set of homodimer enumerations to determine native re-folded homodimeric conformations. The individual chain conformation within the dimer is termed the half-dimer. All possible dimer conformations that comprise a conformation docked against a copy of itself are enumerated to generate an ensemble of possible native homodimers. These homodimer enumerations are performed only for all conformations that have at least the minimum number of contacts for conformations coded for by model-protein sequences.

In 3D, enumerations are performed for a random sample of 250,000 HP 27-mer sequences over the maximally compact conformations to identify sequences that have unique monomeric native-state conformations, i.e. that can be used as model proteins. These sequences are then submitted to enumerations, again only over the maximally compact conformations to find the lowest-energy re-folded homodimeric conformation. For 3D 27-mers, all of the maximally compact conformations (i.e. those with the maximum number of contacts = 28) are cubic. There are 103,346 cubic conformations that are unrelated to each other by rotational or reflective symmetry (Chan & Dill, 1990; Shakhnovich & Gutin, 1990; Li et al., 1996). For the homodimerization enumeration stage, all possible cube dockings that do not entail an offset of the cube faces are included. This 3D model is a perturbed homopolymer model (Chan & Dill, 1990, 1996), wherein a strong background adhesion or sticking energy is assumed so that one has to consider only the maximally compact conformations.

Additionally, exhaustive enumerations over all possible conformations (not just the compact conformations) for very short 2D HP chains (11-mers to 13-mers) were performed. It has been demonstrated that restricting the conformational ensemble to the maximally compact conformations for computational efficiency as performed in our 3D studies here can modify the thermodynamic and kinetic properties of a model system relative to the corresponding model that accounts for all possible conformations (Chan & Dill, 1996; Klimov & Thirumalai, 1996). However, we believe that our 3D approach is adequate for the general, low-resolution thermodynamic trend we aim to determine here. This is confirmed by the qualitatively similar trends exhibited by our 3D results and corresponding 2D results that are obtained using more extensive conformational ensembles.

Model definitions

A model-protein sequence is defined as a prion if each and every conformation in the homodimeric native state is distinct from the monomeric native state. Other model-protein sequences are termed Anfinsenian sequences. We suggest that, under conditions conducive to dimerization, these sequences may transiently form homodimers made from the monomeric native state conformation to avoid encrypting an alternative native state.

For a prion sequence, when the number of conformations in the homodimeric native state is >1, it is said to comprise a set of prion strain conformations (Telling et al., 1996a). Prion strains are thus accommodated in this energetically low-resolution model as multiple conformations in the homodimeric native state. At low concentrations of the prion sequence where dimerization is unlikely, prions will adopt their monomeric native state. However, the homodimeric native state is said to be encrypted by the model-protein sequence under these conditions that favor the monomeric folding landscape.

Analysis of the model proteins and prions arising from the 2D 15-mer and 16-mer enumerations and (particularly) from the 3D cubic 27-mer enumerations is discussed in more detail below.

Fractional sequence hydrophobicities (\( \Phi \))

The fractional hydrophobicity (\( \Phi \)) of a lattice HP sequence is the proportion of H residues in it. \( \Phi \) is calculated for all of the prion and model protein lattice HP sequences. Fractional sequence hydrophobicities for real proteins are similarly defined. Here, \( \Phi \) values are calculated for the prion-determinant sequence fragments for PrP, Ure2p and Sup35p. The Miyazawa & Jernigan (1996) and Fauchere & Pliska (1983) hydrophobicity scales have been normalized to the interval 0.0 to 1.0, with the most hydrophobic residue type given a value of 1.0 and the least hydrophobic residue type given a value of 0.0. The value of \( \Phi \) for each sequence is calculated as the sum of the individual residue hydrophobicities divided by the number of residues in the sequence. These fractional hydrophobicities are analogous to (though not directly comparable to) those calculated for the lattice HP sequences. A representative set of 756 protein chains taken from the
A modification of the algorithm described by Sander & Schneider (1992) was analyzed for hydrophobicity in this way.

**Conformational similarity**

The conformational similarity of the monomeric and homodimeric native-state conformations is analyzed. \( S_{3D}^{di-di} \) is the proportion of residue-residue contacts that are the same for the monomer and half-dimer conformations for the cubic enumerations (as originally defined by Shakhnovich & Gutin, 1990):

\[
S_{3D}^{di-di} = \frac{\text{(number of contacts that are the same between the two conformations)}}{\text{(number of contacts in conformation 1 + number of contacts in conformation 2)}}
\]

There is a one-to-one correspondence between the 103,346 maximally compact 3D 27-mer conformations and their contact maps. For the 2D enumerations, different conformations can have identical contact maps and not all conformations have the same number of contacts. The 2D similarity measure is termed \( S_{2D}^{ mono-di} \) and is given by:

\[
S_{2D}^{ mono-di} = \frac{2 \times \text{(number of contacts that are the same for the two conformations)}}{\text{(number of contacts in conformation 1 + number of contacts in conformation 2)}}
\]

The conformational similarity of multiple conformations in the homodimeric native state is examined and is termed \( S_{di-di}^{3D} \) for both 2D and 3D. The similarity \( S_{di-di}^{3D} \) is defined as the proportion of contacts that are the same between the conformations of two half-dimers and is calculated in the same manner as \( S_{mono-di}^{3D} \) in both 2D and 3D.

In both 2D and 3D, random similarity values are calculated from a sample of 10,000 comparisons between conformations that have the minimum number of contacts that are observed in the monomeric native-state conformations.

**Stability**

A stability for the monomeric native state \( (\Delta G_{\text{monomer}}^{native}) \) is calculated in a fashion similar to that used in a previous study (Shortle et al., 1992). This is an energetic measure of how the native conformation is selected from the ensemble of possible native conformations and is similar to thermodynamic stabilities that are calculated in the experimental study of protein folding. The ensemble of all conformations that are enumerated other than the native is termed the non-native state. This stability is given by:

\[
\Delta G^{\text{native}} = -\frac{1}{kT} \ln \left( \frac{P_{\text{native}}}{P_{\text{non-native}}} \right)
\]

where:

\[
P_{\text{native, monomer}} = \frac{\exp(-\beta E_{\text{native}})}{Q}
\]

\( E_{\text{native}} \) is the energy of the native conformation and \( \beta = \frac{1}{kT} \) is a weighting factor analogous to \( 1/kT \). Stability results are calculated using \( \beta = 5.0 \), which gives an overall negative trend for \( \Delta G \) values.

Since only relatively compact conformations are enumerated for most of the chain lengths in 2D and 3D studied here, the stability \( (\Delta G^{\text{monomer}}) \) calculated here is not equal to the free energy difference between the ground state and all other conformations, but is expected to correlate with it, as the equilibrium is dominated by that between the native and low-lying non-native conformations (Shortle et al., 1992).

**Acknowledgments**

Figures 1, 2 and 4 were made using the graphics program PREPI (supplied by Suhail Islam and Michael Sternberg, Biomolecular Modelling Laboratory, Imperial Cancer Research Fund, UK). Thanks to Kinkade Reiling for assistance with calculations and to Andrew Wallace and Marcin Joachimiak for comments on the manuscript. H.S.C. thanks Professor Ken Dill for support and encouragement. This work was supported by grants from the National Institutes of Health to F.E.C. and S.B.P.

**References**


*Edited by A. R. Fersht*

(Received 14 August 1998; received in revised form 9 December 1998; accepted 21 December 1998)