The role of hydrophobicity patterns in prion folding as revealed by recurrence quantification analysis of primary structure

Joseph P. Zbilut¹, Charles L. Webber Jr², Alfredo Colosimo³ and Alessandro Giuliani⁴

Department of Molecular Biophysics and Physiology, Rush Medical College, 1653 W. Congress, Chicago, IL 60612, ²Department of Physiology, Loyola University, Stritch School of Medicine, 2160 S. First Avenue, Maywood, IL 60153, USA, ³Department of Biochemical Sciences, University of Rome ‘La Sapienza’, P.le A. Moro, 5, 00185 Rome and ⁴Istituto Superiore di Sanità, Lab. TCE, Viale Regina Elena, 299, 00161 Rome, Italy

It has been suggested that the number and strength of local contacts are the major factors governing conformation accessibility of model two ground-state polypeptide chains. This phenomenology has been posed as a possible factor influencing prion folding. To test this conjecture, recurrence quantification analysis was applied to two model 36mers, and the Syrian hamster prion protein. A unique divergence of the radius function for the recurrence quantification variable %DET of hydrophobicity patterns was observed for both 36mers, and in a critical region of the hamster prion protein. This divergence suggests a partition between strong short- and long-range hydrophobicity patterns, and may be an important factor in prion phenomenology, along with other global thermodynamic factors.

Keywords: hydrophobicity/lattice simulations/prions/protein folding/recurrence quantification

Introduction

Simple models of protein folding on a lattice have been useful for the understanding of basic organizing principles (Sali et al., 1994; Thirumalai and Klímov, 1998). One important idea gleaned from these studies has been the conjecture that polymers may have multiple ground states, and thus may fold into different structures (Abkevich et al., 1998).

Abkevich et al. (1998) have extended this idea in combination with kinetic partitioning to suggest a possible phenomenology for the conformational flips of prions. They designed sequences of lattice model proteins which exhibited two different conformations of equal energy corresponding to the global energy minimum. Folding simulations demonstrated that one of these ground states was much more accessible than the other. A critical factor in determining the accessibility was the number and strength of local contacts in the ground state conformation. Although it is recognized that this may not be the only factor involved in such a phenomenology, it does provide some basic understanding of the process. To explore this possibility, as well as the feasibility of deriving an empirical, hydrophobicity based phenomenology, we applied recurrence quantification analysis (RQA) of hydrophobicity values along the sequence of the two given model 36mers described in Abkevich et al. (1998). We then compared the results to the recombinant prion protein (PrP) of the Syrian hamster, shPrP(90–231) (PDB ID code 1B10), which corresponds to the infectious fragment of the scrapie isoform (James et al., 1997).

Materials and methods

Recurrence quantification analysis was performed on the hydrophobicity values (Kyte and Doolittle, 1982) of the residue sequence of the simulated 36mers (Table I). Details of the analysis can be found in Zbilut et al. (1998a) (see also Giuliani and Manetti, 1996; Manetti et al., 1998; Giuliani et al., 2000), but briefly, RQA is a form of contact map (Vendruscolo et al., 1999). It differs from both traditional contact maps and hydropathy plots in that it uses a form of embedding as developed in the nonlinear physical sciences to simulate multidimensional processes (Webber and Zbilut, 1994; Zbilut et al., 1998b). Distances (D) in terms of hydrophobicity values in n-space between individual i-j pairs are calculated, and a Heavyside function is applied if the distances are within a determined error (ε), i.e. a radius (Dij = 1 if ≤ ε; Dij = 0 if > ε). This, of course, is similar to nearest neighbor calculations in n dimensions. Points which form contiguous segments represent ‘deterministic’ processes as opposed to purely random scattering of recurrent points which do not form such continuous segments. Thus multidimensionality as well as nearest-neighbor statistics give added information not available on a traditional hydrophobicity plot (Figures 1–3).

In the present analysis, the determinism (%DET, percentage of recurrent points forming line segments) was calculated for a radius from 1 to 100% (the maximum; distances being rescaled on the unit interval) with an embedding of 3 to simulate a chemical environment in which each residue ‘views’ adjacent residues in simulated three dimensions. (It is emphasized that these dimensional perspectives should not be confused with real coordinates. The dimensions are a result of the mathematical ‘embedding’ procedure. See Figure 3.) As a control, the sequences were randomized 25 times, with a resultant loss of the divergence.

A similar RQA was also performed for the Syrian hamster PrP sequence of hydrophobicity values. In order to see the change of determinism along the entire shPrP sequence, a form of RQA was performed similar to the windowing procedure common in spectral analysis (Figure 4). Windows of 36 residue values (to effect a 36mer) were stepped through the sequence, overlapping one residue at a time.

Results and discussion

For the two model 36mers, the results demonstrate a shelf-like divergence between relatively linear constant %DET values in the low radius region, which quickly drop off to

---

Table I. Model 36mer sequences used by Abkevich et al. (1998).

| Model 1 | MIEGGSLWSTQTTPKHVWWWWEDWYAYGTKFPYGE |
| Model 2 | RGREPQLMILMLWQKEKMRISARGMEMEEPMHWGP |

© Oxford University Press
Fig. 1. Schematic of RQA. An ordered series is first embedded to create new vectors by lagging, representing a multidimensional $n$-space for the original scalar number. In this case, for an embedding of 3, the series is lagged 3 times. To calculate the distances the normed vectors are subtracted from each other. If the distances fall within a certain error threshold, which we term ‘radius,’ (see Figure 2), they can be tallied as recurrent points. Graphically, they can be represented as a recurrence plot (RP). Recurrent points which form line segments (circles in RP) are tallied as deterministic points (%DET). [These points can also be plotted creating a symmetric matrix (a recurrence plot, with an identity line separating two mirror image triangles). Recurrent points which form diagonal line segments (circled areas at left), are counted as being ‘deterministic.’]

In one sense, this singularity is unstable, not necessarily in the thermodynamic sense, but in the topological understanding of the graph. If there are perturbations which can sufficiently destabilize (e.g. $\Delta$pH, $\Delta$temperature or mutations) this arrangement, the observed ordered hydrophobicity can be easily destroyed. A different folding could then develop with ‘access’ to %DET patterns beyond the shelf. Presumably, this would increase the time to reach such a different state, as pointed out by Abkevich et al. (1998).

The RQA results for the recombinant Syrian hamster PrP revealed a divergence similar to one found in the model proteins, and is immediately adjacent to the flexible region of residues 29–124 (Donne et al., 1997) (Figure 8). This again is important since Harrison et al. (1999) have remarked that residues 90–231 are neither particularly hydrophobic nor hydrophilic.

Further inspection of the recurrence plots of the sliding windows demonstrated that the main line segments contributing to the elevated %DET was a result of residues 112–113 and 117–118 (Figure 9). Interestingly enough, these residues are conformationally plastic as well as in the most conserved regions as viewed from an evolutionary perspective (Cohen and Prusiner, 1998). This is consonant with the findings of James et al. (1997) who noted sparse structure between 90–112, but 36 long range NOE (nuclear Overhauser effect) crospeaks involving side-chain resonances for residues 113–125.

In addition, this area contains the theoretically important


\[
\begin{align*}
\text{Ordered Series} & \quad \text{Embedding} & \quad \text{Normed Vector} \\
1 & 1234 & \vec{x}_1 (1,2,3,4) \\
2 & 2345 & \vec{x}_2 (2,3,4,5) \\
3 & 3456 & \vec{x}_3 (3,4,5,6) \\
\end{align*}
\]

\[\text{where } \epsilon \text{ is a defined radius}\]

become exponentially increasing values (Figure 5). Such a profile is unusual in our experience, and is confirmed by the randomization results (Figure 6). The shelf is constant from residues 127–149. The implications would seem straightforward: contrary to the impressions of hydrophobicity plots (Figure 7), which suggest no remarkable features, RQA demonstrates that there is a definite, pronounced structured area (high values of %DET) if one considers their apparent closeness (low radius) in embedding space. This structuring should be understood in terms of a repetitive hydrophobic/hydrophilic pattern, and not simply as a region of uniform hydrophobicity values. What is more striking is the narrowness of the shelf and concomitant drop off. This would imply that local contacts predominate as Abkevich et al. (1998) have proposed. As such, this area may be termed singular, insofar as after the drop, the %DET values increase slowly with no unique profile. In one sense, this singularity is unstable, not necessarily in the thermodynamic sense, but in the topological understanding of the graph. If there are perturbations which can sufficiently destabilize (e.g. $\Delta$pH, $\Delta$temperature or mutations) this arrangement, the observed ordered hydrophobicity can be easily destroyed. A different folding could then develop with ‘access’ to %DET patterns beyond the shelf. Presumably, this would increase the time to reach such a different state, as pointed out by Abkevich et al. (1998).

The RQA results for the recombinant Syrian hamster PrP revealed a divergence similar to one found in the model proteins, and is immediately adjacent to the flexible region of residues 29–124 (Donne et al., 1997) (Figure 8). This again is important since Harrison et al. (1999) have remarked that residues 90–231 are neither particularly hydrophobic nor hydrophilic.

Further inspection of the recurrence plots of the sliding windows demonstrated that the main line segments contributing to the elevated %DET was a result of residues 112–113 and 117–118 (Figure 9). Interestingly enough, these residues are conformationally plastic as well as in the most conserved regions as viewed from an evolutionary perspective (Cohen and Prusiner, 1998). This is consonant with the findings of James et al. (1997) who noted sparse structure between 90–112, but 36 long range NOE (nuclear Overhauser effect) crospeaks involving side-chain resonances for residues 113–125. In addition, this area contains the theoretically important
Recurrence quantification of prions

Fig. 3. Plotting %DET in a given range. It should be noted that RQA searches for patterns irrespective of actual values in the series. In (a), the series has two sequences beginning with –1 (arrows). These sequences are identified in plot (b) as open circles if a strict identity is sought (radius = 0). If, on the other hand, the radius is expanded to 1, additional points appear (c), however, none of the new recurrent points form line segments. If a Heavyside function is applied to these points, and plotted on a square plot, it appears as in (d).

Fig. 4. Illustration of the windowed RQA. A sliding window based on an RP analysis for a defined subseries is moved through the original series. Calculated values of, for example, %DET for the individual windows are then plotted to form a new series. Values for multiple radius values can also be plotted along a y-axis (radius) and z-axis (actual values of %DET). See Figure 8 below.

residue 129, which has been suggested to play an important role in prion diseases through mutations (James et al., 1997). To determine what would happen with a mutation of residue 129, Val was substituted (Figure 10). This mutation occurs in patients with the dementing illness, familial Creutzfeldt-Jakob disease. Interestingly, the singular region was preserved, but

Fig. 5. %DET of two model 36mers (from Abkevich et al., 1998) in the range 0–100 of radius. (a) Model 1; (b) model 2. Note the linear, constant portion of high %DET (approximate radius values 5–10), which then diverge to very low values, followed by an exponential increase.
Fig. 6. (a, model 1, b, model 2). Results from shuffling (without replacement) of the two model 36mers. Results are mean ± standard deviation.

Fig. 7. (a, model 1, b, model 2). Hydrophobicity plots for the two model 36mers.

Fig. 8. (a) Syrian hamster prion protein windowed RQA for %DET. Note the ‘shelf’ of linear, high constant values of %DET, ranging from 127 to 149 (arrow), and again from 158 to 170 with an intervening interval from 150–157 devoid of ‘local’ %DET. (There is some distortion of the Residue axis due to the 3D perspective which results in a ‘floating’ object.) For a correct alignment of residues, see the contour plot in (b). Note that although the given values for the Syrian hamster range from 90 to 231, the plot begins with 127 to account for the windowing procedure (arrow = shelf). Thus, at point 127, the %DET values are derived from the residues starting at 90.

Fig. 9. (a) Recurrence plot detailing the line segments (arrows) responsible for the shelf of Figure 8, residues 112–113 and 117–118; and (b) relevant portion of hydrophobicity plot. Note that although there are also other areas immediately adjacent that have relatively high hydrophobicities, none form a pattern.
several other changes were noted. Specifically, the drop off was more pronounced, while several adjacent regions developed increased %DET values. Again, if subjected to appropriate perturbations, a different profile of %DET hydrophobicity becomes ‘available’, which can then disrupt the short and long correlations of PrP. Features such as these may be important for the determination of specific symptoms such as insomnia versus dementia (James et al., 1997).

In another lattice simulation of model protein folding, Harrison et al. (1999) also demonstrated folding into alternate multimeric states. Their conclusion, however, was that this was a result of instability of monomeric native states. This lack of stability implies the existence of more low energy conformations which can be stabilized by monomeric interactions. The implication is that prion formation is essentially a probabilistic process depending upon factors such as high protein concentration or mutations which could distort the energy landscape. This requires a well-designed energy landscape, so that a greater chance of becoming a prion can be overcome. The authors view the normal functioning native protein form as a kind of ‘kinetic trap’. Such a scenario is in contrast to the Abkevich et al. (1998) study, which does not specifically focus on ‘physiological milieu’ mechanisms for prion folding.

Although these studies exhibit different phenomenologies for prion protein folding, they are not altogether that dissimilar. The Abkevich et al. (1998) model is operative only in the case where the native state is degenerate between two native conformations, and one with many local and the other with many nonlocal contacts. The rich local contacts must be broken, allowing for a greater global energy barrier thus slowing the process for folding into the conformation with more nonlocal contacts. Clearly, however, this ignores other factors such as dimerization, mutations or high protein concentrations, and it has been shown that such factors appear to be important in the configuration of the energy landscape (Dill and Chan, 1997). However, there is no apparent contradiction in combining both of these phenomenologies; the kinetic partitioning model may be operative, but altered when placed in the larger context of aggregations, mutations, etc., and involved in a ‘kinetic trap’. And certainly, kinetic partitioning itself is somewhat probabilistic.

The present study would tend to provide some support for both views. Taken by itself, a model protein with degenerate ground states does seem to exhibit a singularity emphasizing...
local (low radius) contacts. This singularity, however, placed in the context of a real protein, demonstrates that mutations can create a profound change. The %DET landscape (Figure 10) becomes more variegated, while other thermodynamic factors may allow for distortion of the energy landscape thus allowing for accessibility to other hydrophobicity patterns (higher radius). The work of Harrison et al. (1999) points to stabilizing or destabilizing effects on monomeric native state prion proteins in the case of mutations. A careful examination of these factors is beyond the scope of this paper.

We also analyzed the PrP of the mouse (Riek et al., 1998; GenBank, P04925) with similar results (not shown), which is to be expected given the high level of conservation of sequences among many mammalian species (Wopfner et al., 1999). The import of these findings suggests that a singular divergence in the RQA hydrophobicity profile may be a necessary (but not sufficient) concomitant to prion formation. Mutations and thermodynamic factors may be important to complete the picture. Given the ever more complex scenarios found for protein folding (Shakhnovich, 1999), RQA may reveal peculiarities or phenomenologies of hydrophobicity patterns not easily detected by traditional plots.

Software
Software used in this research is available through links at http://www.rushu.rush.edu/molbio/physiozbi.html

Acknowledgements
We would like to express our gratitude to M.Pocchiari of the Laboratorio di Virologia, Istituto Superiore di Sanità, Viale Regina Elena 299, 00161 Rome, for his advice and useful discussions regarding the complexities of prions.

References

Received August 10, 1999; revised September 28, 1999; accepted November 29, 1999