Cracking the folding code. Why do some proteins adopt partially folded conformations, whereas other don't?

Vladimir N. Uversky[a,b,*]

[a] *Institute for Biological Instrumentation, Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russia*
[b] *Department of Chemistry and Biochemistry, University of California, Santa Cruz, CA 95064, USA*

Received 11 December 2001; revised 16 January 2002; accepted 17 January 2002

First published online 12 February 2002

Edited by Thomas L. James

Abstract Many, but not all, globular proteins have been shown to have compact intermediate state(s) under equilibrium conditions in vitro, giving rise to the question: why do some proteins adopt partially folded conformations, whereas other do not? Here we show that charge to hydrophobicity ratio of a polypeptide chain may represent a key determinant in this respect, as proteins known to form equilibrium partially folded intermediates are specifically localized within a unique region of charge–hydrophobicity space. Thus, the competence of a protein to form equilibrium intermediate(s) may be determined by the bulk content of hydrophobic and charged amino acid residues rather than by the positioning of amino acids within the sequence.

Key words: Protein folding; Partially folded intermediate; All-or-none transition; Net charge; Hydrophobicity

1. Introduction

A variety of different physico-chemical forces play a role in stabilizing the unique three-dimensional structure of a protein. Both the strength and specificity of many of these forces are strongly dependent on environmental conditions in such a way that changes in the environment can reduce or even eliminate part of the conformational interactions, while the remainder are unchanged, or even intensified. Under some environmental conditions, the native protein structure can be transformed into new conformations with properties intermediate between those of the native and the completely unfolded states. Thus, the ability of a protein to adopt different stable partially folded conformations should be considered an intrinsic property of a polypeptide chain. It is known that many globular proteins may exist in at least four different conformations: the native (ordered), molten globule, pre-molten globule, and unfolded states [1–5]. Different partially folded conformations play crucial roles in the birth (synthesis and folding), life (function) and death (degradation) of globular proteins [1,6–10]. Moreover, the aggregation of partially folded proteins is responsible for a number of human diseases [11,12] and is a significant problem in biotechnology [13]. Since all the necessary and sufficient information to fold into the native, biologically active conformation is thought to be present in protein amino acid sequence [14], the capability of a given protein to adopt equilibrium partially folded conformation(s) may also be encoded in specific features of its amino acid sequence.

Interestingly, it has been shown that not all proteins (even homologous ones) have an identical response to changes in their environment. For example, hen egg white lysozyme represents a textbook illustration for the two-state model of denaturant-induced unfolding [15], whereas accumulation of classical molten globule under different experimental conditions was described for its homologue α-lactalbumin [16]. We show here that the charge to hydrophobicity ratio of a polypeptide chain may represent a key factor, determining competence of a protein to form equilibrium intermediate(s).

2. Materials and methods

Literature data on equilibrium unfolding of globular proteins (with unfolding induced by changes in pH, temperature or by increase in the concentration of strong denaturants, such as urea or guanidinium chloride) were analyzed. A test set of 154 globular proteins, for which data are readily available, was chosen based on this analysis. This test set was subdivided into two subsets based on the published unfolding behavior. In the first subset, 115 globular proteins were included with each being shown to adopt equilibrium partially folded conformation(s). The second subset contained 39 globular proteins, each of which has been shown to unfold without the formation of any intermediate state. Using the Swiss Institute of Bioinformatics (SIB) server, ExPASy [17], the following information was extracted for each individual protein: (i) number of amino acid residues; (ii) molecular mass; and (iii) total number of negatively (Asp+Glu) and positively charged (Arg+Lys) residues. The hydrophobicity of each amino acid sequence was calculated by the Kyte and Doolittle approximation [18], using a window size of five amino acids. The hydrophobicity of individual residues was normalized to a scale of 0–1 in these calculations. The mean hydrophobicity is defined as the sum of the normalized hydrophobicities of all residues divided by the number of residues in the polypeptide. The mean net charge is defined as the net charge at pH 7.0, divided by the total number of residues.

3. Results and discussion

Our analysis of literature data on equilibrium unfolding of globular proteins induced by changes in pH, temperature, or strong denaturants (urea or guanidinium chloride) revealed that unfolding in 115 proteins is accompanied by accumulation of equilibrium intermediate states of one sort or another. Another set comprises 39 proteins, which were shown to unfold according to a simple two-state model, i.e. no equilibrium intermediate of any kind was formed during their unfolding. The full list of proteins from both groups will be published elsewhere.

In an attempt to understand which factors may be respon-

*Fax: (1)-831-459 2935.
E-mail address: uversky@hydrogen.ucsc.edu (V.N. Uversky).

0014-5793/02/$22.00
PII: S0014-5793(02)02359-1

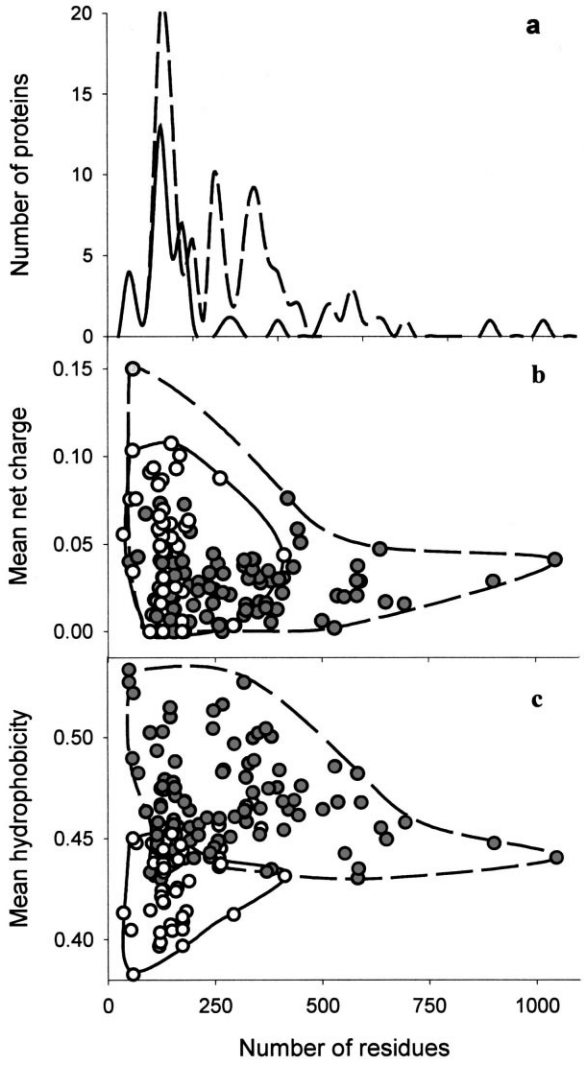

Fig. 1. Comparison of the general amino acid sequence features of the set of 115 proteins able to form equilibrium intermediates (open symbols, black lines) and the set of 39 proteins shown to unfold without accumulation of partially folded conformations (gray symbols and lines). Data are presented as: (a) the distribution of proteins as a function of length of amino acid sequences; (b) dependence of the mean net charge (net charge at pH 7.0/number of residues) on the length of the polypeptide chain; (c) length dependence of the mean hydrophobicity (Kyte and Doolittle approximation, normalized from 0 to 1).

sible for such tremendous difference in the formation of equilibrium partially folded intermediates, the general sequence features of proteins from both groups have been analyzed using a simple method comparing global sequence charge and hydrophobicity in a set of 154 globular proteins. Fig. 1a shows the differences in sequence length distribution of proteins from both groups. Those proteins observed to form equilibrium intermediates are generally larger than those proteins shown to unfold without accumulation of partially folded conformations, with mean lengths of 270 ± 176 and 145 ± 66 amino acid residues, respectively. However, Fig. 1a shows that length distributions for both classes overlap, suggesting that chain length alone cannot be used to discriminate proteins from the two classes. Further, Fig. 1b shows that there is essentially no difference between the two groups in terms of their mean net charge as a function of sequence length. However, statistical analysis reveals that globular proteins which are able to adopt partially folded intermediates have a lower overall net charge (0.027 ± 0.022) compared to proteins which unfold according to a two-state model (0.051 ± 0.032).

A different situation is observed when the hydrophobicity of the polypeptide chain is taken into account. Fig. 1c shows that the set of intermediate forming proteins is relatively well separated from the set of proteins unable to adopt a partially folded conformation. Statistical analysis of these data gives a mean hydrophobicity of 0.446 ± 0.023 and 0.422 ± 0.017 for the members of the two groups, respectively. Thus, the mean hydrophobicity may be a significant contributing factor in determining whether a protein will form equilibrium intermediates or not. Finally, combining both factors, mean hydrophobicity and mean net charge, allows for the reliable separation of both groups of proteins; they occupy different areas within the charge–hydrophobicity phase space (Fig. 2). These data imply that the competence of a protein to form equilibrium intermediate(s) may be predetermined by the bulk content of hydrophobic and charged amino acid residues.

Patterns of hydrophobic and hydrophilic residues are very important for protein folding and function. For instance, the burial of hydrophobic residues is considered to be the major factor determining the formation of the cores of globular proteins [19]. Further, alternating hydrophobic and hydrophilic residues in β-strands represents the major factor stabilizing proteins rich in these structures. Interestingly, it has been recently shown that sequences of three or more consecutive hydrophobic residues are significantly less common in globular proteins than would be predicted if residues were selected independently, which was interpreted as evolutionary selection against long blocks of hydrophobic residues within globular proteins [20]. Thus, the spatial distribution of hydrophobic and hydrophilic residues within polypeptide chains is a critical feature of amino acid sequences to direct the folding of proteins.

Although the distributions of the quantities used in our study to discriminate globular proteins overlap considerably (0.027 ± 0.022 vs. 0.051 ± 0.032 and 0.446 ± 0.023 vs. 0.422 ± 0.017 for the mean net charge and mean hydrophobicity calculated for the proteins that are able to adopt partially

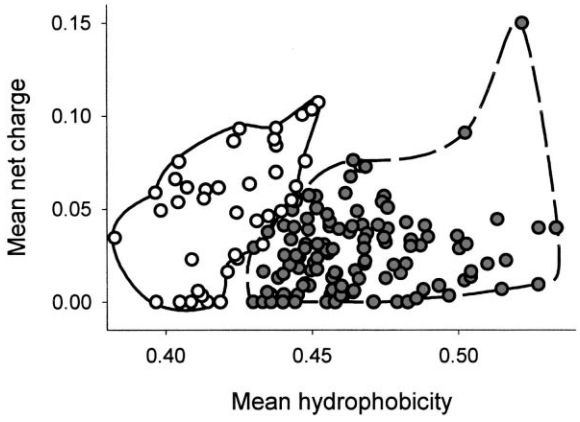

Fig. 2. Comparison of mean net charge vs. mean hydrophobicity for the set of 115 proteins able to form equilibrium intermediates (open symbols, black lines) and the set of 39 proteins shown to unfold without accumulation of partially folded conformations (gray symbols and lines).

folded intermediates and for those that unfold according to a two-state model, respectively), reliable separation of these two groups of proteins was achieved when both factors were considered simultaneously. In fact, these groups occupy different areas within the charge–hydrophobicity phase space (Fig. 2). Thus, our data show that the capability of a protein to adopt equilibrium partially folded conformation(s) may be encoded in the charge/hydrophobicity ratio of its polypeptide chain, not its sequence. This may mean that partially folded conformations are stabilized mostly by non-specific, side chain–side chain interactions of hydrophobic amino acid residues. Interestingly, proteins that do not have equilibrium intermediates are less hydrophobic and have, in general, a larger net charge than those competent to form discrete intermediate states. This may indicate that such proteins are less strengthened by hydrophobic interactions and more disturbed by electrostatic repulsion. Thus, smaller environmental changes may be required to overcome the marginal stabilization energies leading to immediate and complete unfolding of the protein.

Acknowledgements: I am very grateful to Prof. A.K. Dunker for valuable discussions. I wish to thank Dr. J.R. Gillespie, Prof. J. Goers and Dr. P.O. Souillac for their invaluable help with manuscript improvement.

References

[1] Ptitsyn, O.B. (1995) Adv. Protein Chem. 47, 83–229.
[2] Uversky, V.N. and Ptitsyn, O.B. (1994) Biochemistry 33, 2782–2791.
[3] Uversky, V.N. and Ptitsyn, O.B. (1996) J. Mol. Biol. 255, 215–228.
[4] Uversky, V.N. (1997) Protein Pept. Lett. 4, 355–367.
[5] Uversky, V.N. (1998) Biofizika (Moscow) 43, 416–421.
[6] Baldwin, R.L. (1991) Chemtracts Biochem. Mol. Biol. 2, 379–389.
[7] Bychkova, V.E. and Ptitsyn, O.B. (1993) Chemtracts Biochem. Mol. Biol. 4, 133–163.
[8] Baldwin, R.L. (1995) J. Biomol. NMR 5, 103–109.
[9] Fink, A.L. (1995) Methods Mol. Biol. 40, 343–360.
[10] Fink, A.L. (1995) Annu. Rev. Biophys. Biomol. Struct. 24, 495–522.
[11] Uversky, V.N., Talapatra, A., Gillespie, J.R. and Fink, A.L. (1999) Med. Sci. Monitor 5, 1001–1012.
[12] Uversky, V.N., Talapatra, A., Gillespie, J.R. and Fink, A.L. (1999) Med. Sci. Monitor 5, 1238–1254.
[13] Schein, C.H. (1989) BioTechnology 7, 1141–1149.
[14] Anfinsen, C.B., Haber, E., Sela, M. and White, F.N. (1961) Proc. Natl. Acad. Sci. USA 47, 1309–1314.
[15] Tanford, C. (1968) Adv. Protein Chem. 23, 121–282.
[16] Kuwajima, K. (1996) FASEB J. 10, 102–109.
[17] Appel, R.D., Bairoch, A. and Hochstrasser, D.F. (1994) Trends Biochem. Sci. 19, 258–260.
[18] Kyte, J. and Doolittle, B.F. (1982) J. Mol. Biol. 157, 105–132.
[19] Rose, G.D. and Roy, S. (1980) Proc. Natl. Acad. Sci. USA 77, 4643–4647.
[20] Schwartz, R., Istrail, S. and King, J. (2001) Protein Sci. 10, 1023–1031.