Minireview

Folding nuclei in proteins

Oxana V. Galzitskaya, Dmitry N. Ivankov, Alexei V. Finkelstein*

Institute of Protein Research, Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russia

Received 20 November 2000; revised 13 December 2000; accepted 18 December 2000

First published online 15 January 2001

Edited by Vladimir Skulachev

Abstract When a protein folds or unfolds, it passes through many half-folded microstates. Only a few of them can accumulate and be seen experimentally, and this happens only when the folding (or unfolding) occurs far from the point of thermodynamic equilibrium between the native and denatured states. The universal features of folding, though, are observed just close to the equilibrium point. Here the 'two-state' transition proceeds without any accumulation of metastable intermediates, and only the transition state ('folding nucleus') is outlined by its key influence on the folding–unfolding kinetics. Our aim is to review recent experimental and theoretical studies of the folding nuclei.

Key words: Protein folding; Folding intermediate; All-or-none transition; Two-state kinetics; Folding nucleus; Transition state; Rate of folding

1. Introduction

Protein physics is based on two fundamental experimental facts: (i) protein chains (at least those that are not too long or excessively modified after the biosynthesis) are capable of forming their native structures spontaneously [1,2] in appropriate environment and (ii) the native state is divided from the unfolded state of the chain by a co-operative transition [3] of the 'all-or-none' type [4,5]. The latter ensures robustness of protein action. It appears that biological evolution selects only those sequences, which can fold into well-defined three-dimensional (3D) native structures with the free energy significantly lower than that of the alternative structures: only such sequences are capable of the 'all-or-none' type folding and unfolding [6,7].

In this paper we consider only in vitro folding (for recent reviews on in vivo folding, see [8,9]), and only folding of water-soluble globular proteins. Although the membrane proteins and the fragments of fibrous proteins are also capable of spontaneous renaturation [10,11], the basics of folding are better studied just for water-soluble globular proteins (and for relatively small proteins, since large proteins tend to aggregate even at a very low concentration).

*Corresponding author. Fax: (7)-095-924 0493.
E-mail: afinkel@vega.protres.ru

Abbreviations: 3D, three-dimensional; TS, transition state

The fundamental folding problem has come to be known as the Levinthal paradox. It reads as follows [12]. On the one hand, the existence of renaturation suggests that the native state is thermodynamically the most stable state under the 'biological' conditions. On the other hand, a chain has zillions of possible conformations ($>2^{100}$ for a 100-residue chain, since at least two conformations are possible for each residue). Thus, the chain needs at least $\sim 2^{100}$ ps, or $\sim 10^{10}$ years to sample them all and to find the most stable fold. Then, how can the chain find its most stable structure within a 'biological' time (minutes)?

The initial hypothesis was that there exists a specific folding pathway, and the native fold is simply the end of this pathway rather than the most stable chain fold [12]. If this pathway was narrow, only a small part of the conformational space would be sampled, and the Levinthal paradox would be avoided.

This hypothesis of a 'kinetic control' has initiated numerous studies of intermediates in protein folding. The idea was that these intermediates would help trace the folding pathway, as the intermediates in a complicated biochemical reaction trace the pathway of this reaction [13].

To observe an intermediate, it either has to be stable enough to live for a time comparable with the time of its formation (thus, it must be more stable than the unfolded state and precede the rate-limiting folding step), or it has to be quenched and trapped. The trapping is usually done either using the quench-flow hydrogen-exchange pulse labeling [14,15] (with a subsequent nuclear magnetic resonance (NMR) investigation of the trapped product), or using the disulfide bond formation [16]. It should be noted that the kinetic studies are often hampered by heterogeneity of the unfolded state [17]: the 'fast-folding' unfolded chains seem to have an already 'correct' state of prolines (*cis* or *trans*), while the 'slow-folding' forms have the 'wrong' proline isomers (and proline is known to spend many seconds for each *cis–trans* isomerization [18]).

It has been shown that many average size proteins (of $\sim$100–200 residues) generally fold via intermediates which accumulate during the early stages of folding (Fig. 1). The burst (millisecond) stage is characterized by partial formation of the secondary structure and partial condensation, typical of the 'pre-molten' globule (see [19,20] for review). The next intermediate (formed within many milliseconds) usually demonstrates the molten globule state: it has rather native-like and rather protected secondary structures, and even some native-like tertiary contacts [15,19,21]. The further formation of the native state can take from a fraction of a second to

0014-5793/01/$20.00
PII: S0014-5793(01)02092-0

hours (the latter seems to be due to the slow rearrangement of packing of non-native proline isomers, non-native ligandings or non-native SS bonds [16,19]).

However, the entropy of even the most structured intermediate, the molten globule, is still large [19,22]: it is not greatly reduced as compared with that of the unfolded state. This means that the folding intermediates per se only allow division of the Levinthal paradox into sub-problems (e.g. formation of the native-like folding pattern and native packing), but do not allow understanding of *how* the main Levinthal problem, the problem of sampling the enormous conformational space, is solved at each (or at least the last and the main) of these stages.

2. Re-discovery of the two-state ('all-or-none') transitions in kinetics

It is worth mentioning that the kinetic investigation of protein folding was conventionally held under (or extrapolated to) 'biological' conditions (25–37°C, neutral pH, absence of denaturant) [17]. In addition to the biological relevance of these conditions, the folding intermediates are most readily accumulated and interpreted when the (re)folding takes place under these conditions, where the native state (together with the folding intermediates) is much more stable than the unfolded state.

However, it is noteworthy that folding can occur also in the zone of thermodynamic equilibrium, where the unfolded state is almost as stable as the native state of a protein [22,23], while all the intermediates are unstable here and thus cannot be accumulated in principle (by the definition of the 'all-or-none' transition, observed in thermodynamics [4,5]).

Besides, there are many ~100-residue proteins which fold, and very rapidly fold, without any visible folding intermediates under a wide range of conditions, from the conditions of thermodynamic equilibrium up to 'biological' conditions [25]. The discovery of the simple 'two-state' folding, proceeding without any visible (accumulating) intermediates, was actually a re-discovery of the thermodynamic 'all-or-none' transitions in kinetics.

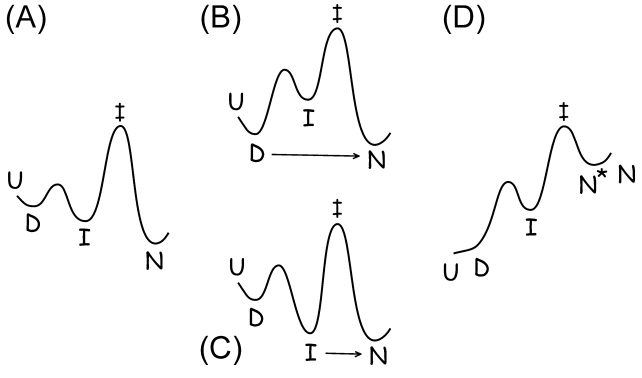

Fig. 1. Schemes of the free energy profiles for protein folding and unfolding under different conditions. U, completely unfolded chain under extreme unfolding conditions; D, denatured (mostly unfolded, but partially condensed and partially structured) state; I, major folding intermediate (molten globule); ‡, major free energy barrier (TS); N, native state. (A) Folding of the initially unfolded chain under extreme folding ('biological') conditions proceeds with observable intermediates D and I. (B, C) Two-state folding under moderate folding conditions (in or close to the transition zone): (B) the case when the D→N (or U→N) transition is observed, and I is much less stable (as in barnase at normal pH and high temperature [78]); (C) the case when the I→N transition is observed, and D is much less stable (as in barnase at low pH [78]). (D) Unfolding N→U under extreme unfolding conditions. In this case, the partially unfolded native state (N*) is sometimes observed as a burst unfolding intermediate [79].

The recent great progress in the understanding of protein folding [26,27] has been achieved just by investigation of the proteins, which fold without 'unnecessary complications' (previously widely used to trace the folding pathway): without accumulation of any intermediates at the folding and the unfolding pathways, without *cis–trans* proline isomerization, and without SS bond formation. The folding (and the unfolding) kinetics looks very simple in this case: all the properties of the native (or denatured) protein are restored synchronically, following the single exponential kinetics [28]. For some proteins, this simplicity is observed under a wide range of conditions, including in the zone of the reversible thermodynamic transition between two phases (the native and the denatured state); for others, only in the transition zone (Figs. 1 and 2A,B). Thus, the universal features of folding (and unfolding) can be observed just in the transition zone, while the moving off this zone towards the 'biological' conditions reveals individualities of various proteins (which are the 'unnecessary complications', when we try to understand the basics of protein folding).

3. Folding nuclei and folding rates: experimental studies

The above statement looks, in a sense, paradoxical. Indeed, what can we get from investigation of folding (or unfolding) in the transition zone, where we cannot accumulate any transition intermediates? The answer is: just here we can most readily, though indirectly, observe the folding transition state (TS), whose stability (or, more exactly, instability) determines the folding (and unfolding) rate [25–27,29]. The TS corresponds to the free energy maximum on the folding–unfolding pathway (Fig. 1) or, better said, to the free energy saddle point on the network of these pathways.

The structure formed by the protein chain in the TS is called the 'folding nucleus'. It plays a key role in protein folding. So far, there is only one, very difficult experimental method to identify the folding nuclei in proteins: to find the residues whose mutations affect the folding rate by changing the TS stability as strongly as that of the native protein [24,27] (Fig. 2C,D).

The participation of a residue in the folding nucleus is expressed by the residue's ϕ value. ϕ is defined as $\Delta \ln k_f / \Delta \ln K$, where k_f is the folding rate constant, K is the equilibrium constant (k_f/k_u) and Δ means the mutation-induced shift of the corresponding value. According to the model of a native-like folding nucleus [24,25], $\phi = 1$ means that the residue has its native conformation and environment already in the TS (i.e. that it is inside the folding nucleus), while $\phi = 0$ means that the residue remains in the unfolded state in the TS. The values of $\phi \approx 0.5$ are ambiguous: either the residue is at the surface of the nucleus, or it is inside one of the alternative nuclei, belonging to two different folding pathways. It is noteworthy that the values $\phi < 0$ and $\phi > 1$ (which would be inconsistent with the model of a native-like folding nucleus) are extremely rare and never refer to a residue with a reliable measured $\Delta \ln K$.

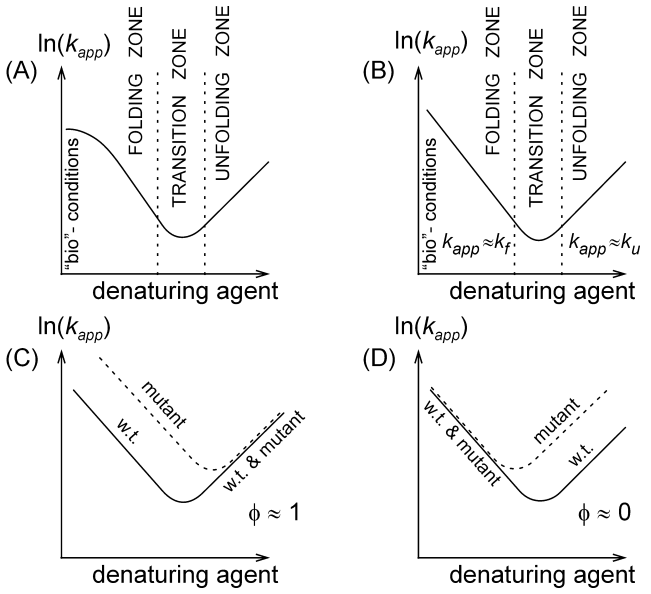

Fig. 2. Chevron plot: apparent rate of the folding–unfolding process (k_{app}) versus the denaturant concentration (or the temperature). (A) Typical plot: the protein has the two-state transition only at and close to the zone of thermodynamic equilibrium between the native and the denatured states, but not under the 'biological' conditions. (B) Plot for a protein having the two-state transition throughout the whole experimental conditions. For the two-state transitions, $k_{app} = k_f + k_u$, where k_f is the folding rate and k_u is the unfolding rate: thus, $k_{app} \approx k_f$ in the folding zone (where $k_f \gg k_u$), and $k_{app} \approx k_u$ in the unfolding zone (where $k_f \ll k_u$) [80]. (C) Mutation of a residue, having its native environment and conformation already in the TS, changes the mutant folding rate rather than its unfolding rate. (D) Mutation of a residue, which remains in the denatured state in the TS, has the opposite effect.

To estimate ϕ, all the rates k have to be measured at (or extrapolated to) the same conditions. Usually, being interested in the 'biologically relevant' nucleus, one extrapolates them to the zero denaturant concentration. However, it should be noted that the nucleus corresponding to the wild-type protein mid-transition is outlined more reliably: here the extrapolation is shorter and therefore more robust, especially when the branches of the chevron are curved; the letter suggests a change of the nucleus with the folding conditions [30].

One can estimate the position of the TS on the reaction coordinate from the dependence of the folding and unfolding rate constants on the denaturant concentration C [31]: $\beta_T = (\delta \ln k_f/\delta C)/(\delta \ln K/\delta C)$. When β_T is close to 1, the TS solvent-accessible surface area is close to that of the native protein; when β_T is close to 0, the TS is rather unfolded. As a rule, the observed values of β_T are close to 0.6–0.8 for small proteins; at the same time, the average value of ϕ is usually about 0.3–0.4. The difference between the average ϕ and β_T values means that there are some non-native interactions in the TS ensemble (cf. [32]).

The major assumptions, underlying the ϕ analysis of the folding nucleus by point mutations [24], are that the mutations do not change substantially either the folding pathway, the nucleus, the structure of the folded state or the unfolded state ensemble. Experimentally, this is proved correct by double mutations in the protein [33]. However, some strong mutations can significantly affect the distribution of structures in the TS ensemble [34].

It has been shown that proteins with different sequences but similar 3D structures have similar folding nuclei [35–37]. However, there is at least one experimental example showing that the location of the nucleus differs among proteins with the same (flavodoxin-like) topology [38]. It has been shown also that circular permutation, changing the protein topology, sometimes changes [39] and sometimes does not change [40] the TS.

4. Folding nuclei and folding rates: theoretical studies

Theoretical studies of protein folding have gone in parallel and sometimes in cooperation with the experimental studies. They have contributed a great deal to the understanding of protein folding.

First, we should mention the folding simulations held on simple lattice protein models by the group of Shakhnovich. These works have a profound influence on the understanding of the experimental results. They show that, as a rule, a fast and reproducible folding leads only to the most stable chain fold (especially when its stability is reinforced by a special 'editing' of the sequence) [41], that a preliminary burst collapse is not an obligatory prerequisite of folding [42], and that such an obligatory prerequisite is the formation of a certain set of native contacts between remote regions of the folding chain [43]. These specific contacts (the 'nucleus') obligatorily appear in the TS (and lead to the subsequent rapid assembly of the native fold), whereas the other contacts are optional and occur in the TS with a low probability.

The nucleation gave a key to solve the Levinthal paradox and to estimate the characteristic protein folding rates. Although both the in vitro [29] and in silico [44,45] experiments show that further stabilization of the native state leads to an even faster folding, the nucleation is especially easily treated [46] near the above mentioned point of thermodynamic equilibrium between the native and the unfolded phases. At the equilibrium point, the rates and pathways of protein folding and unfolding are equal by definition, while the unfolding is much easier to imagine. Also one should not consider the misfolded structures at this point (i.e. the Levinthal paradox is avoided here): being less stable than both the native and the unfolded states, they simply cannot arise.

Since, at the equilibrium point, the free energies of the native and the unfolded phases are equal, the additional free energy caused by the nucleus (the free energy barrier of nucleation) is due only to the boundary between the native and

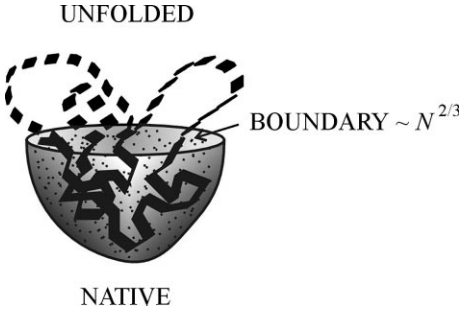

Fig. 3. Transient semi-unfolded (or semi-folded) state of a protein molecule. The unfolded part is shown by dashed lines, the folded structure is shown by solid lines. The unfolded closed loops protruding from the folded part (the nucleus) create an additional surface tension.

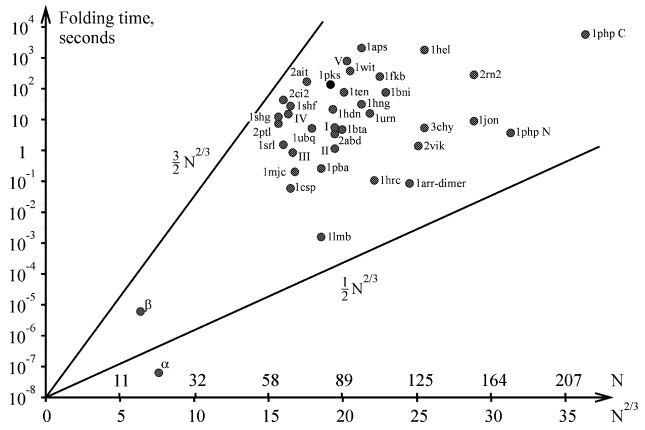

Fig. 4. Observed folding–unfolding time t (at the point of equilibrium between the native and unfolded states), presented on a logarithmic scale versus $N^{2/3}$ (where N is the number of residues in the chain). The points refer to the α-helical and the β-hairpin peptides [20] and to all 36 proteins, listed in [25], whose folding–unfolding time in mid-transition can be calculated from the presented data (where possible, the proteins are marked by their PDB codes given in [25]; the proteins, having no PDB codes: I, ACBP rat; II, ACBP yeast; III, CspB (*B. caldolyticus*); IV, CspB (*Thermotoga maritima*); V, Tenascin (long form)). The theoretically predicted region of $\ln(t)$ is limited by the lines $\ln(\tau)+0.5N^{2/3}$ and $\ln(\tau)+1.5N^{2/3}$ [46,47], where $\tau \approx 10$ ns is the time of α-helix elongation by one residue [81]. One can see that all the experimental points are within this range (except for that of the α-helix, which is a 1D rather than 3D object).

the unfolded phases (Fig. 3). The maximal size of this boundary is $\sim N^{2/3}$ for the N-residue chain. Depending on the protein topology, the boundary of the native part of the nucleus may or may not be covered by the unfolded closed loops, whose entropy adds to the conventional surface energy of the boundary. It can be estimated that the free energy of the boundary is $1.5N^{2/3}k_BT$ (where k_B is the Boltzmann constant and T the temperature), when all the boundary is covered by loops, and $0.5N^{2/3}k_BT$, when the boundary is free of loops [46,47]. Since a characteristic time of rearrangement of one residue is ~ 1 ns (see legend to Fig. 4), it takes up to $\exp(1.5N^{2/3})$ ns to overcome the free energy barrier of nucleation in the first case, and only $\sim \exp(0.5N^{2/3})$ ns in the second [46,47]. This range is exactly consistent with the observed times of protein folding (Fig. 4).

An overview of the suggested scaling laws for protein folding rates under various conditions can be found in [48].

It is worthwhile to point out that Fig. 4, so successful in prediction of the range of the folding rates of small proteins, predicts a very slow folding of larger proteins, unless they are divided into separately folding domains (foldons) and/or unless they have a special architecture with a small number of loops going from one part of the large domain to another.

A more accurate theory of protein folding rates requires a more detailed analysis of protein structures. A high negative correlation exists between the protein folding rate and the 'contact order', i.e. the average sequence separation of residues, close in the 3D protein structure [49]. Since the contact order refers to the whole native protein, the existing correlation implies that the topology of the TS resembles that of the native protein. The other successful methods are based on the nucleation–collision model of protein folding [50] and on solution of kinetic equations for the networks of protein folding–unfolding pathways [51].

As regards the theoretical search for folding–unfolding nuclei in proteins, several different approaches have been suggested recently.

The idea of specific nuclei, suggested by the lattice simulations of protein folding [43], generated an evolutionary approach to the prediction of the nuclei. It is based on the search for a set of highly conserved residues having no obvious functional role [52–55]. It should be mentioned that this approach, at best, can give only the common part of the nuclei existing in homologous proteins. However, some recent observations show that the residue conservatism across the homologous proteins correlates with deep immersion into the hydrophobic core of a protein [55] rather than into the folding nucleus [56].

It should be noted that some correlation between the nuclei (the regions of high ϕ values) and the hydrophodic cores and secondary structures exists [57,58], but it is rather low [57].

The most direct approach generates a plausible TS using the all-atom molecular dynamic simulations of protein unfolding [59–62]. According to these simulations, the unfolding is hierarchic [63,64] (at least when it occurs far from the equilibrium): tertiary interactions break early, whereas secondary structures remain. The repeated trajectories show a statistical distribution around the experimentally found TSs and dem-

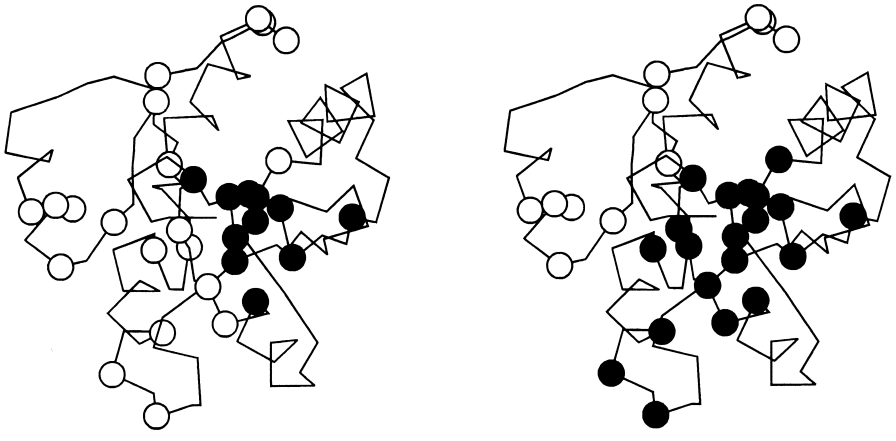

Fig. 5. Folding nuclei in the CheY protein: experiment [82] (left) and theory [73] (right). The experimentally studied residues are shown by circles: those constituting the nucleus (with $\phi > 0.3$) are denoted by closed circles, the others (with $\phi < 0.3$) by open circles.

onstrate a broad ensemble of the TS structures. However, these simulations need extremely denaturing conditions (600 K, etc.) to be completed. Therefore, the TSs found for such an extreme *un*folding can be, in principle, rather different from those existing for folding [65].

Other approaches are based on a simplified modelling of the protein folding landscapes and trajectories [66,67]. The lattice simulations [68], as well as experiment [69], show that the TS is an ensemble rather than a single conformation and that it can be described by an order parameter, such as the fraction of native contacts. The TS search is based on the projection of the folding trajectories onto a single reaction coordinate (the fraction of the native contacts) and investigation of barriers at the obtained free-energy profiles (like those shown in Fig. 1) [70–72]. Although such a projection is not a rigorous procedure (e.g. the structure which is 'nearly native' in terms of contacts, can contain an additional knot of the chain, and need complete unfolding before coming to the native state), these studies were able to outline, though crudely, the folding nuclei of some small proteins.

Further progress is due to the analysis of multidimensional networks of the protein folding–unfolding trajectories done by various algorithms [57,73,74]. All these approaches (as well as the studies of the folding rates [46,50]) consider only the attractive native interactions (the 'Go model' [75]) to reduce the energy frustrations and heterogeneity of interactions, and model the trade-off between the formation of attractive interactions and the loss of conformational entropy during protein folding. These works allow the folding nuclei to be outlined in more detail (Fig. 5). Despite the relative simplicity of these models, they give a good correlation with experiment [76].

The success of the reviewed theoretical studies suggests that the topological features (and the size of the protein) play a much more important role in folding than the high resolution details of protein structure [46,73,76,77].

5. Conclusions

We have discussed the folding rates and the folding nuclei for the 'unfolded chain → native state' transitions.

However, one can also expect the existence of nuclei for the 'unfolded chain → molten globule' and for the 'molten globule → native state' transitions. In principle, these nuclei can be studied in the same way, as the nuclei for the 'unfolded chain → native state' transitions. It will be interesting to outline these nuclei, both experimentally and theoretically, and especially to see if the same or different residues participate in all of them.

Acknowledgements: We are grateful to N.S. Bogatyreva for drawing Fig. 3. This work was supported by the Russian Foundation for Basic Research, by the CRDF grant RB2-2022 and by an International Research Scholar's Award to A.V.F. from the Howard Hughes Medical Institute.

References

[1] Anfinsen, C.B., Haber, E., Sela, M. and White, F.H. (1961) Proc. Natl. Acad. Sci. USA 47, 1309–1314.
[2] Anfinsen, C.B. and Scheraga, H.A. (1975) Adv. Protein Chem. 29, 205–300.
[3] Tanford, C. (1968) Adv. Protein Chem. 23, 121–218; (1970) Adv. Prot. Chem. 24, 1–95. Protein Denaturation.
[4] Privalov, P.L. (1979) Adv. Protein Chem. 33, 167–241.
[5] Privalov, P.L. (1982) Adv. Protein Chem. 35, 1–104.
[6] Go, N. (1975) Int. J. Pept. Protein Res. 7, 313–323.
[7] Bryngelson, J.B. and Wolynes, P.G. (1990) Biopolymers 30, 177–188.
[8] Ellis, R.J. and Hartl, F.U. (1999) Curr. Opin. Struct. Biol. 9, 102–110.
[9] Hardesty, B., Tsalkova, T. and Kramer, G. (1999) Curr. Opin. Struct. Biol. 9, 111–114.
[10] London, E. and Khorana, H.G. (1982) J. Biol. Chem. 257, 7003–7011.
[11] Baum, J. and Brodsky, B. (1998) Fold. Des. 2, R53–R60.
[12] Levinthal, C. (1968) J. Chim. Phys. Chim. Biol. 65, 44–45.
[13] Leninger, A.L., Nelson, D.L. and Cox, M.M. (1993) Principles of Biochemistry, 2nd edn., Part III, Worth, New York.
[14] Udgaonkar, J.B. and Baldwin, R.L. (1988) Nature 335, 694–699.
[15] Dyson, J.H. and Wright, P.E. (1996) Annu. Rev. Phys. Chem. 47, 369–395.
[16] Creighton, T.E. (1992) in: Protein Folding (Creighton, T.E., Ed.), pp. 301–351, W.H. Freeman and Co, New York.
[17] Baldwin, R.L. and Creighton, T.E. (1980) in: Protein Folding (Jaenicke, R., Ed.), pp. 217–260, Elsevier, Amsterdam.
[18] Brandts, J.F., Halverson, H.R. and Brennan, M. (1975) Biochemistry 14, 4953–4963.
[19] Ptitsyn, O.B. (1995) Adv. Protein Chem. 47, 83–229.
[20] Eaton, W.A., Muñoz, V., Hagen, S.J., Jas, G.J., Lapidus, L.J., Henry, E.G. and Hofrichter, J. (2000) Annu. Rev. Biophys. Biomol. Struct. 29, 327–359.
[21] Roder, H., Elöve, G.A. and Englander, S.W. (1988) Nature 335, 700–704.
[22] Kuwajima, K. and Sugai, S. (1978) Biophys. Chem. 8, 247–254.
[23] Segawa, S.-I. and Sugihara, M. (1984) Biopolymers 23, 2473–2488.
[24] Matouschek, A., Kellis Jr., J.T., Serrano, L. and Fersht, A.R. (1989) Nature 340, 122–126.
[25] Jackson, S.E. (1998) Fold. Des. 3, R81–R91.
[26] Fersht, A.R. (1995) Curr. Opin. Struct. Biol. 5, 79–84.
[27] Dobson, C.M. and Karplus, M. (1999) Curr. Opin. Struct. Biol. 9, 92–101.
[28] Kragelund, B.B., Robinson, C.V., Knudsen, J., Dobson, C.M. and Poulsen, F.M. (1995) Biochemistry 34, 7217–7224.
[29] Fersht, A.R. (1997) Curr. Opin. Struct. Biol. 7, 3–9.
[30] Otzen, D.E., Kristensen, O., Proctor, M. and Oliveberg, M. (1999) Biochemistry 38, 6499–6511.
[31] Matouscheck, A., Otzen, D.E., Itzhaki, L.S., Jackson, S.E. and Fersht, A.R. (1995) Biochemistry 34, 13656–13662.
[32] Li, L., Mirny, L.A. and Shakhnovich, E.I. (2000) Nat. Struct. Biol. 7, 336–342.
[33] Fersht, A.R., Matouschek, A. and Serrano, L. (1992) J. Mol. Biol. 224, 771–782.
[34] Burton, R.E., Huang, G.S., Daugherty, M.A., Calderoni, T.L. and Oas, T.G. (1997) Nat. Struct. Biol. 4, 305–310.
[35] Martinez, J.C. and Serrano, L. (1999) Nat. Struct. Biol. 6, 1010–1016.
[36] Riddle, D.S., Grantcharova, V.P., Santiago, J.V., Alm, E., Ruczinski, I. and Baker, D. (1999) Nat. Struct. Biol. 6, 1016–1024.
[37] Perl, D., Welker, Ch., Schindler, T., Schroder, Marahiel, M.A., Jaenicke, R. and Schmid, F.X. (1998) Nat. Struct. Biol. 5, 229–235.
[38] Steensma, E. and van Mierlo, C.P.M. (1998) J. Mol. Biol. 282, 653–666.
[39] Viguera, A.R., Serrano, L. and Wilmanns, M. (1996) Nat. Struct. Biol. 3, 874–880.
[40] Otzen, D.E. and Fersht, A.R. (1998) Biochemistry 37, 8139–8146.
[41] Šali, A., Shakhnovich, E.I. and Karplus, M. (1994) J. Mol. Biol. 235, 1614–1636.
[42] Gutin, A.M., Abkevich, V.I. and Shakhnovich, E.I. (1995) Biochemistry 34, 3066–3076.
[43] Abkevich, V.I., Gutin, A.M. and Shakhnovich, E.I. (1994) Biochemistry 33, 10026–10036.
[44] Galzitskaya, O.V. and Finkelstein, A.V. (1995) Protein Eng. 8, 883–892.
[45] Gutin, A.M., Abkevich, V.I. and Shakhnovich, E.I. (1996) Phys. Rev. Lett. 77, 5433–5436.

[46] Finkelstein, A.V. and Badretdinov, A.Ya. (1997) Fold. Des. 2, 115–121.
[47] Finkelstein, A.V. and Badretdinov, A.Ya. (1997) Mol. Biol. (Russia Engl. Edition) 31, 391–398.
[48] Wolynes, P.G. (1997) Proc. Natl. Acad. Sci. USA 94, 6170–6175.
[49] Plaxo, K.W., Simons, K.T. and Baker, D. (1998) J. Mol. Biol. 277, 985–994.
[50] Munõz, V. and Eaton, W.A. (1999) Proc. Natl. Acad. Sci. USA 96, 11311–11316.
[51] Ivankov, D.N. and Finkelstein, A.V. (2001) FEBS Lett., submitted for publication.
[52] Shakhnovich, E., Abkevich, V. and Ptitsyn, O. (1996) Nature (London) 379, 96–98.
[53] Ptitsyn, O.B. (1998) J. Mol. Biol. 278, 655–666.
[54] Ptitsyn, O.B. and Ting, K.-L. (1999) J. Mol. Biol. 291, 671–682.
[55] Mirny, L.A. and Shakhnovich, E.I. (1999) J. Mol. Biol. 291, 177–196.
[56] Plaxco, K.W., Larson, S., Ruczinski, I., Riddle, D.S., Thayer, E.C., Buchwitz, B., Davidson, A.R. and Baker, D. (2000) J. Mol. Biol. 298, 303–312.
[57] Galzitskaya, O.V., Skoogarev, A.V., Ivankov, D.N. and Finkelstein, A.V. (1999) Proceedings of the Pacific Symposium on Biocomputing '2000.
[58] Nölting, B. and Andret, K. (2000) Proteins 41, 288–298.
[59] Li, A. and Daggett, V. (1996) J. Mol. Biol. 257, 412–429.
[60] Daggett, V., Li, A., Itzhaki, L.S., Otzen, D.E. and Fersht, A.R. (1996) J. Mol. Biol. 257, 430–440.
[61] Caflisch, A. and Karplus, M. (1995) J. Mol. Biol. 252, 672–708.
[62] Brooks III, C.L., Gruebele, M., Onuchic, J.N. and Wolynes, P.G. (1998) Proc. Natl. Acad. Sci. USA 95, 11037–11038.
[63] Lazaridis and Karplus, M. (1997) Science 278, 1928–1931.
[64] Isai, J., Levitt, M. and Baker, D. (1999) J. Mol. Biol. 291, 215–225.
[65] Finkelstein, A.V. (1997) Protein Eng. 10, 843–845.
[66] Onuchic, J.N., Socci, N.D., Luthey-Schulten, Z. and Wolynes, P.G. (1996) Fold. Des. (London) 1, 441–450.
[67] Dill, K.A. and Chan, H.S. (1997) Nat. Struct. Biol. 4, 10–19.
[68] Pande, V.S. and Rocksar, D.S. (1999) Proc. Natl. Acad. Sci. USA 96, 1273–1278.
[69] Burton, R.E., Huang, G.S., Daugherty, M.A., Calderoni, T.L. and Oas, T.G. (1997) Nat. Struct. Biol. 4, 305–310.
[70] Shoemaker, B.A., Wang, J. and Wolynes, P.G. (1999) J. Mol. Biol. 287, 675–694.
[71] Shoemaker, B.A. and Wolynes, P.G. (1999) J. Mol. Biol. 287, 657–674.
[72] Plotkin, S.S. and Onuchic, J.N. (2000) Proc. Natl. Acad. Sci. USA 97, 6509–6514.
[73] Galzitskaya, O.V. and Finkelstein, A.V. (1999) Proc. Natl. Acad. Sci. USA 96, 11299–11304.
[74] Alm, E. and Baker, D. (1999) Proc. Natl. Acad. Sci. USA 96, 11305–11310.
[75] Ueda, Y., Taketomi, H. and Go, N. (1975) Int. J. Pept. Protein Res. 7, 445–459.
[76] Baker, D. (2000) Nature 405, 39–42.
[77] Clementi, C., Jennings, P.A. and Onuchic, J.N. (2000) Proc. Natl. Acad. Sci. USA 97, 5871–5876.
[78] Oliveberg, M. and Fersht, A.R. (1996) Biochemistry 35, 2738–2749.
[79] Feng, Z., Ha, J.-H. and Loh, S.N. (1999) Biochemistry 38, 14433–14439.
[80] Matouscheck, A., Kellis Jr., J.T., Serrano, L., Bycroft, M. and Fersht, A.R. (1990) Nature (London) 346, 440–445.
[81] Zana, R. (1975) Biopolymers 14, 2425–2428.
[82] López-Hernández, E. and Serrano, L. (1996) Fold. Des. (London) 1, 43–55.