HOW FAST A PROTEIN CHAIN CAN FOLD TO ITS MOST STABLE STRUCTURE?

Alexei V Finkelstein[1]   Azat Ya. Badretdinov[1,2]

Abstract

Having $\sim 10^{100}$ of possible folds [1], how does the protein chain spontaneously [2] choose its native structure: as the most stable fold - but how the chain can find time to single it among 10^{100} of other folds? as the metastable fold which forms rapidly enough - but why then is protein folding so perfectly reversible [3] and why only the most stable structures of model protein chains demonstrate a reliable folding [4]? A purpose of this paper is show that a protein chain can find its most stable fold fast and without sorting out of all its other folds, i.e. to elucidate the "Levinthal paradox" [1].

We will consider the chains and conditions providing the "all-or-none" thermodynamic transition from the coil to the native state. It is known that such a thermodynamic behavior is typical [3] but requires specially selected sequences with a large energy gap between the lowest-energy fold and its competitors [5,6]. Our aim is to prove that in this case the most stable fold of an N-residue chain is achieved much rapidly, i.e., that a stable structure automatically forms a fast folding pathway to this structure..

---

[1] Institute of Protein Research, Russian Academy of Sciences, 142292, Pushchino, Moscow Region, Russian Federation; afinkel@sun.ipr.serpukhov.su

[2] Present address: Rockefeller University, Box 270, 1230 York Ave., New York, NY 10021, USA; azat@guitar.rockefeller.edu

1   Folding pathway

An N-residue chain can fold in N steps, each of which adds of one residue to a growing native structure (Fig.1). Here we will consider folding pathways of this kind. The additional pathways can only accelerate the folding since the rates of parallel reactions are additive.

If the free energy would be downhill along all the pathway, a 100-residue chain would fold in ~ 100 nsec, since the growth of a structure (e.g., an α-helix) by 1 residue takes a few nanoseconds [7]. If protein folding takes more than 100 ns, this is because the free energy increases at some steps of folding, and most of the folding time is spent climbing the free energy barrier and falling back, rather than moving along the folding pathway. A simple, based on the transition state theory [8] estimate of the transition time is:

$$\text{FOLDING TIME} \sim t \Big/ \exp\left(- \Delta G^{\#}\big/RT\right)$$

Here T is the absolute temperature, R – the gas constant, $\Delta G^{\#}$ – the free energy of the transition state counted off the initial free energy minimum, and $t \sim 1$ ns is the time of adding of one residue to the growing structure.

Let ΔE_n, ΔS_n and $\Delta G_n = \Delta E_n - T\Delta S_n$ be the interaction energy, the conformational entropy and the free energy, respectively, of the intermediate where n links are already fixed in the final positions while other $N-n$ links are disordered counted off those of the disordered chain (hence, all the $\Delta S_n < 0$). One can see that

$$\Delta G_n \big/ RT = \left(- \Delta S_N \big/ R\right) \times$$
$$\left(\Delta S_n \big/ \Delta S_N - \Delta E_n \big/ \Delta E_N\right) + \qquad (1)$$
$$\left(\Delta G_N \big/ RT\right) \times \left(\Delta E_n \big/ \Delta E_N\right)$$

Here ΔE_N, ΔS_N and ΔG_N are, respectively, the energy, entropy and free energy differences between the final structure and the disordered state. At the point of thermodynamic equilibrium between the final structure and the disordered state the $\Delta G_N = 0$, and the free energy barrier is

$$\Delta G^{\#}/RT = \max_n \{\Delta G_n/RT\} =$$
$$\left(-\Delta S_N/R\right) \cdot \max_n \left\{\Delta S_n/\Delta S_N - \right. \qquad (2)$$
$$\left. \Delta E_n/\Delta E_N\right\}$$

For the basic estimate of $\Delta G^{\#}$, we consider a pathway proceeding via more or less compact intermediates and into account only the main free energy constituents; they are proportional to the volume n and surface $n^{2/3}$ of the intermediate. Hence, ΔE_n is proportional to $n - \mu_n n^{2/3}$ and ΔS_n to $-n - \gamma_n n^{2/3}$ where $\mu_n \sim 1$ accounts for a usual energetic surface tension and γ_n for the entropic one [9]. The last is connected with the average entropy spent to close a loop protruding from the nucleus; $\gamma_n \sim 1$ since the loop entropy is $\sim \ln(m)$, where m is the loop length, and the loop probability is proportional to $m^{-3/2}$ [10]. Note that $\Delta G^{\#}$ would be

$$\left(-\Delta S_N/R\right) \times \max\left(n/N - n/N\right) = 0$$ if both the surface tension terms μ and γ were equal to zero. The expansion of Eq.(2) over small terms proportional to μ and γ gives

$$\Delta G^{\#}/RT \Big|_{\Delta G_N=0} = \sigma \cdot \max_n \Big\{ n^{2/3}$$
$$\left[\mu_n + \mu_n \mu_N/N^{1/3} + \gamma_n - \right. \qquad (3)$$
$$\left. - (n/N)^{1/3} \mu_N\right]\Big\} \sim N^{2/3}$$

Here $\sigma = -\Delta S_N/NR$ is a constant (2.3 according to [3]) determined by the entropy loss of one residue in the native protein relative to the coil.

Thus, the activation barrier is proportional to $N^{2/3}$ rather than to N (the latter is the Levinthal estimate). A physical reason for this effect is the entropy-by-energy compensation in the course of folding [11]. The sequence heterogeneity can lead to some change of the barrier due to the ruggedness of the folding pathway, but the corresponding term is smaller, $\sim N^{1/2}$ [12], and does not change the above basic estimate. Thus

$$\text{FOLDING TIME} \sim \exp(N^{2/3})$$
$$\text{nanoseconds} \qquad (4)$$

is, on the average, sufficient to achieve the most stable fold. Despite the non-polynomial (NP) dependence on the residue number N (which agrees with the general mathematical theory [13]), this folding time is not too long: a 100- or 150-residue protein finds its most stable fold within minutes (however, this can be a problem for a bigger protein, unless each of its domains searches for its stable fold separately, as it probably does [3]).

The above consideration neglects the entropy of possible knotting of the disordered loops: it seems to be very small as compared with the effects considered above [14]. However, this is correct only until the chains are not extremely long. Numerical experiments with the non-phantom polymers show that on can expect one knot for a chain region of about 100 links. More precise: for a region from 27 links for the most knotted chain to 335 links for a coil formed by an extremely thin chain, and for much longer for a thick chain (see [14,15] and the refs. therein). Thus, a necessity of correct knotting on the early folding steps (Fig.2) will introduce a multiplier of $\sim\exp(0.01N)$ in the above given estimate of the folding time, but this multiplier can become really important only for the chains of many thousands of residues which is far above the normal size of protein chains.

2 Conclusion

The main feature of folding intermediates shown in Fig.1 is that they consist of a compact nucleus with a native arrangement of involved residues, while the remaining chain is loose and capable of fast rearrangement. In this way the protein avoids a very slow rearrangement of a compact globule and such a folding pathway is not as ragged as it has been suggested for rearrangement of a dense globule [16].

Fast folding takes place only in that range of conditions where the misfolded states cannot trap the folding since they are less stable than both the final lowest-energy fold and the initial unfolded state (Fig.3). Hence, a fast and unambiguous folding can occur only in those chains which provide a large energy gap between the lowest-energy fold and the other folds [4,17].

3 Acknowledgments

References

[1] Levinthal, C. (1968). J. Chim. Phys., Chim. Biol. 65, 44-45.

[2] Anfinsen, C.B. (1973). Science 181, 223-230.

[3] Creighton, T.E. (1991). Proteins (2-nd ed.). W.H.Freeman, New York.

[4] Sali, A., Shakhnovich, E.I. & Karplus, M. (1994). J. Mol. Biol. 235, 1614-1636.

[5] Goldstein, R.A., Luthey-Schulten, Z.A. & Wolynes, P.G. (1992). Proc. Natl. Acad. Sci. USA 89, 4918-4922.

[6] Finkelstein, A.V., Gutin, A.M. & Badretdinov, A.Ya. (1995). In Subcellular Biochemistry. Proteins: Structure, Function and Protein Engineering. (Biswas, B.B. & Roy, S., eds.), pp.1-26, Plenum , NY.

[7] Zana, R. (1975). Biopolymers 14, 2425-2428.

[8] Moore, J.W. & Pearson, R.G. (1981). Kinetics and Mechanism. J.Wiley, New York.

[9] Finkelstein, A.V. & Badretdinov, A.Ya. (1997). Folding & Design 2, 115-121.

[10] Lifshitz, I.M., Grosberg, A.Y. & Khokhlov, A.R. (1979). Rev.Mod.Phys. 50, 683-713.

[11] G , N. (1983). Ann. Rev. Biophys. Bioeng. 12, 183-210.

[12] Thirumalai, D. (1995). J. de Phys. (Orsay, France) 15, 1457-1467.

[13] Ngo, J.T. & Marks, J. (1992). Protein Eng. 5, 313-321.

[14] Frank-Kamenetskii, M.D. & Vologodskii, A.D. (1981). Uspekhi Fiz. Nauk (USSR) 134, 641-674.

[15] Grosberg, A.Y. (1997). Uspekhi Fiz. Nauk (Russia) 167, 129-166.

[16] Bryngelson, J.D., Onuchic, J.N., Socci, N.D. & Wolynes, P.G. (1995). Proteins 21, 167-195.

[17] Wolynes. P.G. (1997) Proc. Natl. Acad. Sci. USA 94, 6170-6175.
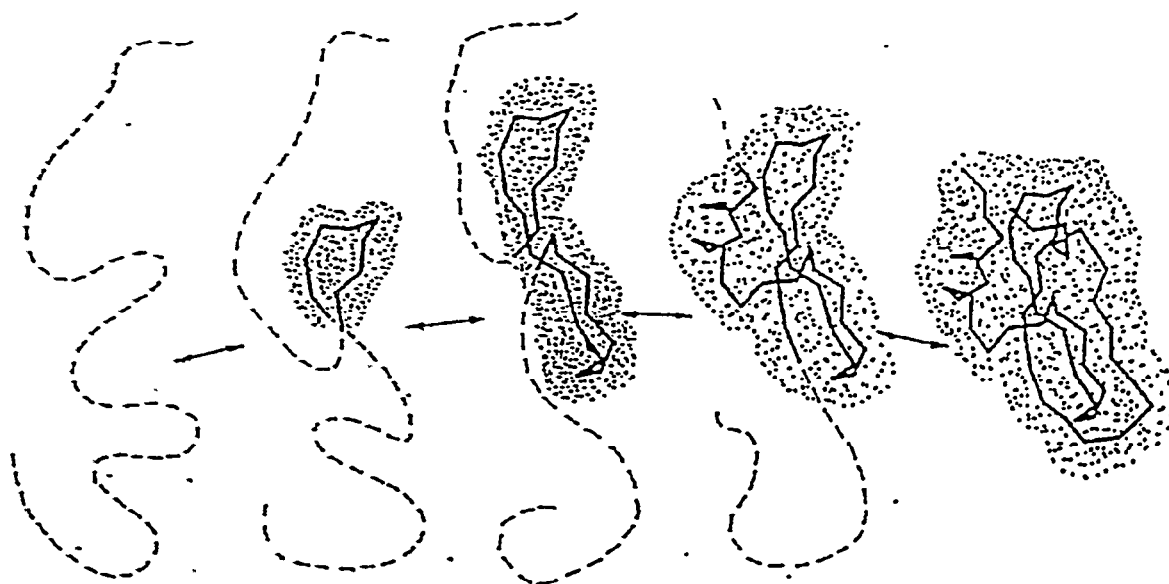
Figure 1. Possible pathway of a sequential folding. Solid line shows the backbone fixed in its native conformation (the fixed side-chains are not shown; the region occupied by them is shown by dots). Note that it is always possible to choose such a pathway where compact nuclei are decorated by the protruding loops from only one side. All the disordered chain regions are shown by the dashed line.
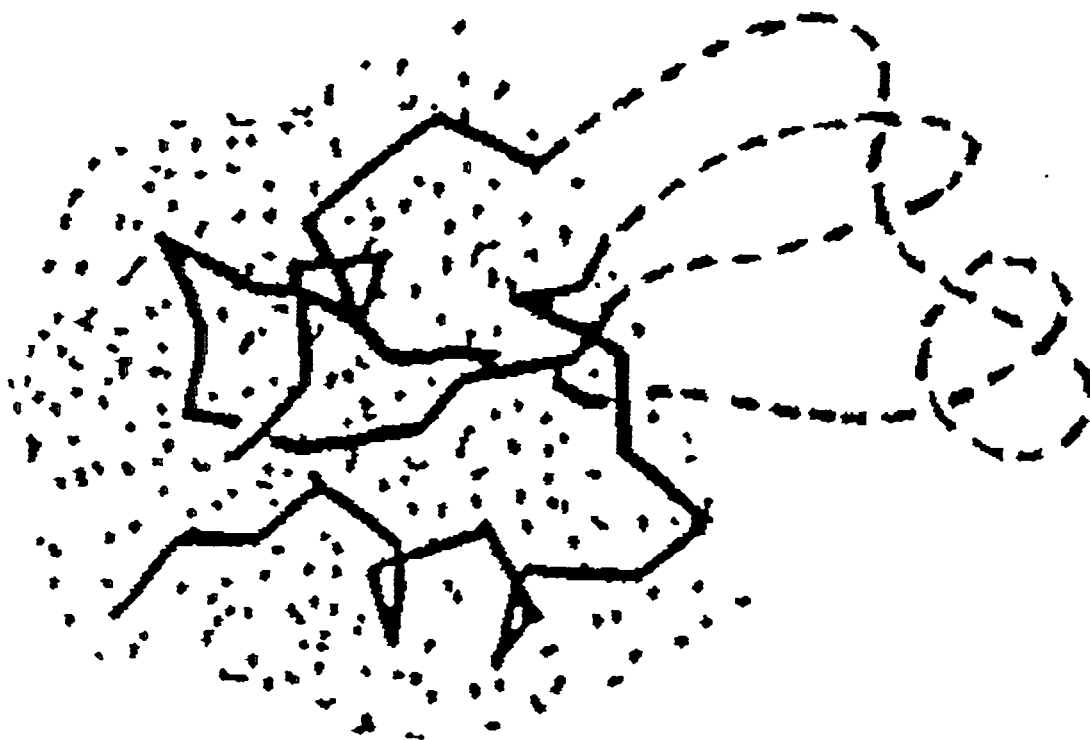

Figure 2. Successful folding of the semi-folded protein to the final structure requires a correct knotting of the disordered loops already in the folding intermediate.
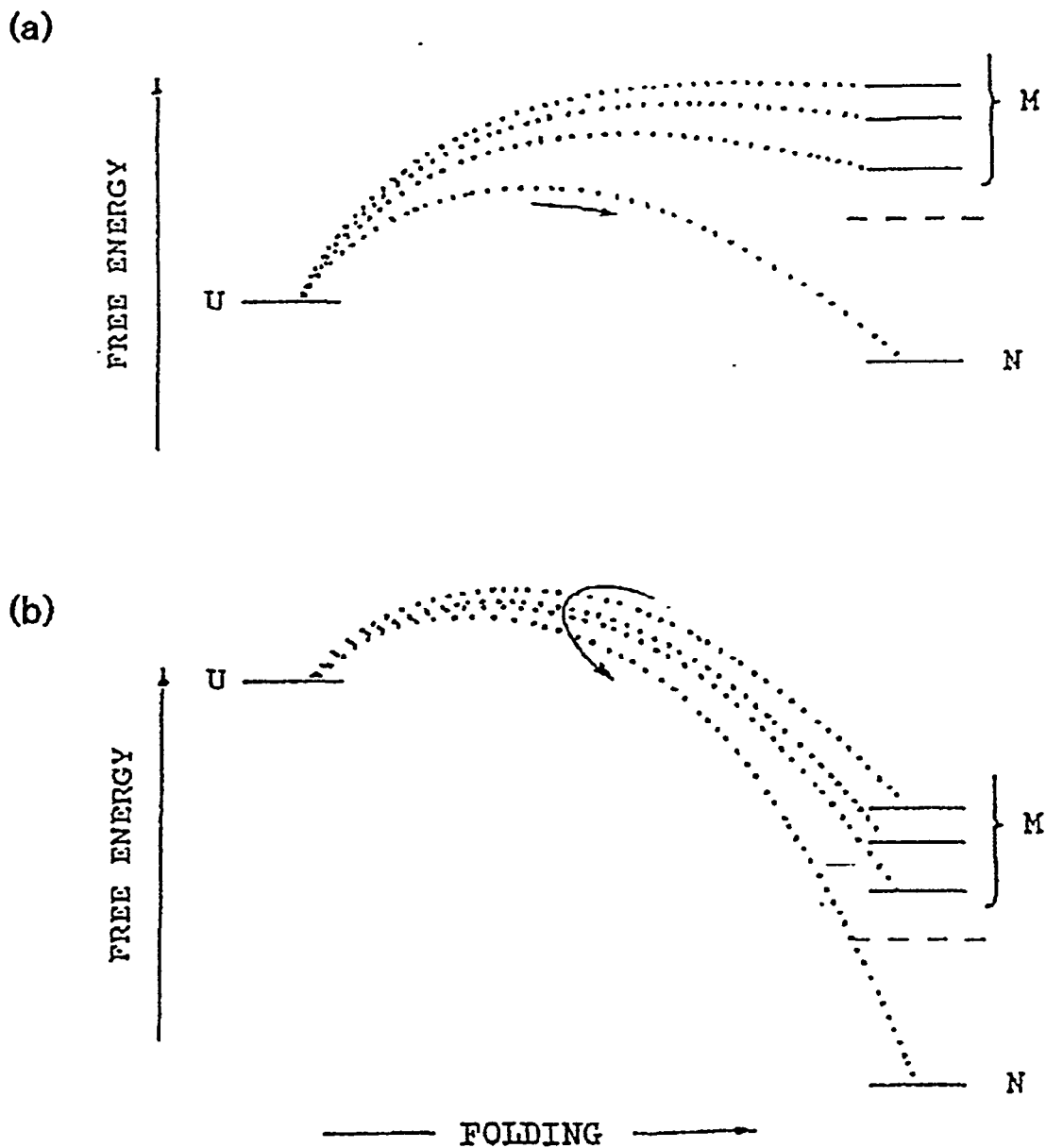
Figure 3. Folding under different conditions. Bold lines show the free-energies of the native fold (N), of the unfolded chain (U), and of the misfolded structures (M); dashed line shows the free energy of the totality of all the misfolded folds. Dotted lines show schematically the behavior of free energy along the folding pathways; a small ruggedness of the free energy profiles is not shown. (A) The native fold is more stable than the coil, and the coil is more stable than all the misfolded forms taken together; rapid folding to the native state is not hindered by kinetic traps. (B) Both native and misfolded folds are more stable than the coil (which is a phase where fast rearrangement occurs); the chain rapidly forms the misfolded forms and then slowly undergoes a transition to the stable state N via the unfolded state. The arrows show the mainstream of the folding process.