The Levinthal paradox: yesterday and today

Martin Karplus

A change in the perception of the protein folding problem has taken place recently. The nature of the change is outlined and the reasons for it are presented. An essential element is the recognition that a bias toward the native state over much of the effective energy surface may govern the folding process. This has replaced the random search paradigm of Levinthal and suggests that there are many ways of reaching the native state in a reasonable time so that a specific pathway does not have to be postulated. The change in perception is due primarily to the application of statistical mechanical models and lattice simulations to protein folding. Examples of lattice model results on protein folding are presented. It is pointed out that the new optimism about the protein folding problem must be complemented by more detailed studies to determine the structural and energetic factors that introduce the biases which make possible the folding of real proteins.

Addresses: Laboratoire de Chimie Biophysique, Institut le Bel, Universite Louis Pasteur, 67000 Strasbourg, France. Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138, USA.
E-mail: marci@brel.u-strasbg.fr or marci@tammy.harvard.edu

Electronic identifier: **1359-0278-002-S0069**

Folding & Design 01 Aug 1997, **2**:S69–S75

© Current Biology Ltd ISSN 1359-0278

An understanding of the mechanism by which a polypeptide chain folds from the denatured coil state to the native protein structure is an essential element of structural biology. Despite the considerable effort, both theoretical and experimental, that has been devoted to this problem, we are still not able to give a detailed description of the mechanism by which any protein folds [1]. Recently there has been a major change in the perception of the inherent difficulty of what is often called the ‘protein folding problem’. Actually, there are two ‘protein folding problems’. The first is concerned with the prediction of the three-dimensional structure of a protein from its sequence and the second, which is the one discussed here, is concerned with the kinetics and dynamics of the actual folding process. A complete solution of the second problem would, of course, simultaneously solve the first. However, it is more likely that the prediction of the native structure will be achieved by other methods, such as threading [2] and homology modelling [3,4], that are based mainly on the analysis of known structures, rather than by directly folding a polypeptide chain.

It is not a great exaggeration to say that the pessimistic viewpoint of yesterday (“It is impossible for proteins to fold to the native state, even though they do so readily in solution”) has been replaced by the optimistic viewpoint of today (“It is obvious that proteins should be able to fold rapidly to the native state”). In this brief report, I shall try to make clear the nature of this dramatic change, to indicate how it came about and to outline what remains to be done to obtain an understanding of protein folding.

Changes in perception are an essential element in the advancement of science. Often, the key to progress is not that a given view has been disproved and that the view that replaces it has been proved. Instead, the important element is the acceptance of the new view by the scientific community. That is what appears to be happening in the case of the protein folding problem. The present situation is somewhat analogous to the recognition nearly 20 years ago of the role of protein motion in protein function [5]. The classic view of biological macromolecules in their native state was static in character. The remarkable detail evident in crystal structures led to an image of biomolecules with every atom fixed in place. DC Phillips, who determined the first enzyme crystal structure, has written: “The period 1965–75 may be described as the decade of the rigid macromolecule. Brass models of DNA and a variety of proteins dominated the scene and much of the thinking” [6]. Even in 1980, Tanford stated that as a result of packing considerations “the structure of proteins must be quite rigid” [7]. Although it should have been self evident that proteins, like other polymers, undergo significant fluctuations at room temperature, only with the advent of molecular dynamics simulations was the static view of the structure of biomolecules replaced by a dynamic picture [5,8,9]: the atoms of which biopolymers are composed are in a state of constant motion at ordinary temperatures and the X-ray structure provides the average atomic positions in the crystal. Crystallographers acceded to this viewpoint after a number of years and sometimes even emphasize the parts of a molecule they do not see in a crystal structure as evidence of motion [10].

The Levinthal paradox [11] dominated ideas about protein folding until very recently. The essential concept introduced by Levinthal is that the appropriate point of reference for protein folding is a random search problem. Taken literally, as it was by many people, this means that all conformations of the polypeptide chain (except the native state) are equally probable, so that the native state can be found only by an unbiased random search. This has been referred to as the ‘golf course’ model of the

protein potential energy surface [12]. For such a surface, the time to find the native state is given by the number of configurations of the polypeptide chain (say, 10^{70} for a 100-residue protein) multiplied by the time required to find one configuration (say, 10^{-11} seconds). This leads to an enormously long folding time (say, 10^{59} seconds or about 10^{52} years). Since proteins generally fold in times on the order of milliseconds to seconds (except for special factors that slow the folding, such as proline isomerization [13]), there was indeed a paradox. The ultimate statement of the paradox has been given recently in the language of computational complexity, i.e. the demonstration that the random search problem posed by Levinthal is NP hard [14–17]. An important point that emerged from these studies is that an essential element of the complexity is the presence of long-range interactions, which lead to the well-known cooperative character of the folding transition. The problem would not be NP hard if each amino acid could find its native conformation independently of the others or if only near-neighbor interactions were involved [17].

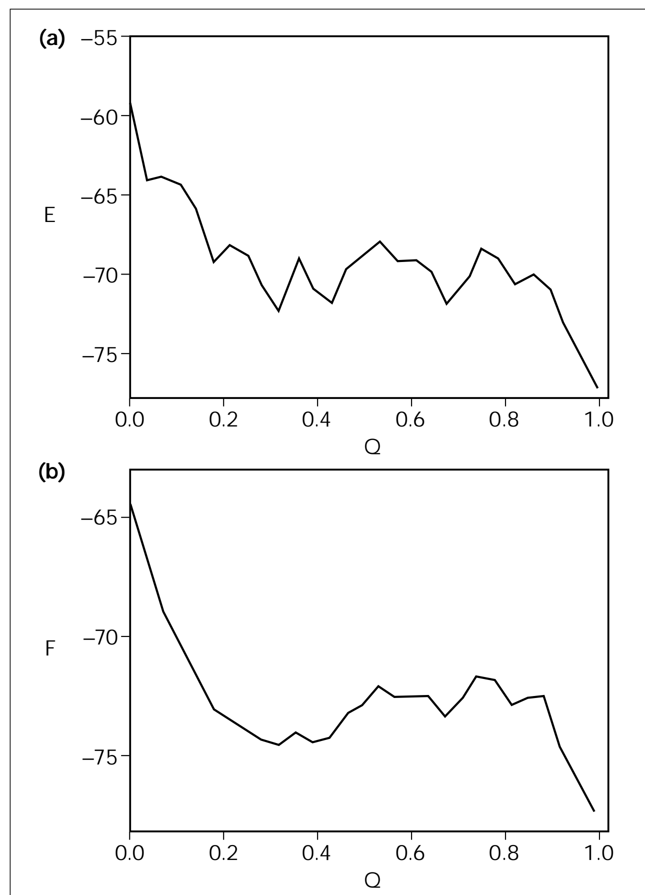
Levinthal's solution to the protein folding problem was that there were well-defined pathways to the native state [18], so that protein folding was under 'kinetic' control; a modern pedagogical description of this viewpoint is given by Dill and Chan in [19]. Too little was known in 1969 to make the pathway concept more specific or to test it experimentally. Since then, there have been many proposals, based mainly on the known structures of proteins, to restrict the conformational space that is searched and to reduce the folding time to the experimental range. Examples include the nucleation-growth mechanism [20–22], the diffusion-collision model [23,24], the framework model [25], and the jigsaw-puzzle model [26]. Most of these models are descriptive in character and do not provide a means for estimating the folding time, which is clearly an important element in a resolution of the Levinthal paradox. The diffusion-collision model [23,24] is an exception that has been used to relate the folding time to certain system parameters [27,28]. Although the various phenomenological models are discussed in the experimental literature, they seem to have had relatively little direct impact on experiments; there was no way to determine which one, if any of them, was correct. Moreover, none of the models appears to have been accepted as a resolution of the Levinthal paradox.

In the past few years, the focus of approaches to the protein folding problem has shifted from phenomenological models to consideration of the general characteristics of the energy surface of a polypeptide chain. This is eminently reasonable since the energy surface is one of the fundamental determinants of any reaction, whether a small molecule reaction [29] or protein folding. The change in focus is based on theoretical considerations [12,30,31] and the use of lattice simulations to study folding [32–34]. It makes explicit a simple concept that is implicit in the phenomenological models. This is the fact that, for any of the models to work, there must be energetic factors that bias the folding process. For example, only if a nucleus is stable, relative to the random coil structures, can it play a role in folding. This means that the difference in the energy and free energy between the denatured and native state is reflected in some way not only in the neighborhood of the native state ('golf course' surface) but over a significant portion of the surface sampled during the folding process.

Go and Abe in 1981 [35] made a pioneering study in this area. By doing lattice Monte Carlo simulations for a bead model of the polypeptide chain in two dimensions, they showed that fast folding occurred if the stabilizing interactions corresponded to those present in the native state. This concordance between stabilizing interactions and the native structure was referred to by Go as the 'harmony principle' of protein folding [21]. Bryngelson and Wolynes [12] developed a statistical mechanical model of protein folding in 1989 that embodied the harmony principle of Go; they used the term 'principle of minimum frustration', based on the concept of frustration in spin glass theories [31], to describe the importance of the relation between the structural features sampled in folding and those of the native state. The works of Go and Abe and Bryngelson and Wolynes contained many insights that are important for the present-day view of protein folding. However, they appear to have had rather little impact on experimentalists. As sometimes happens in science, these papers were published before their time had come.

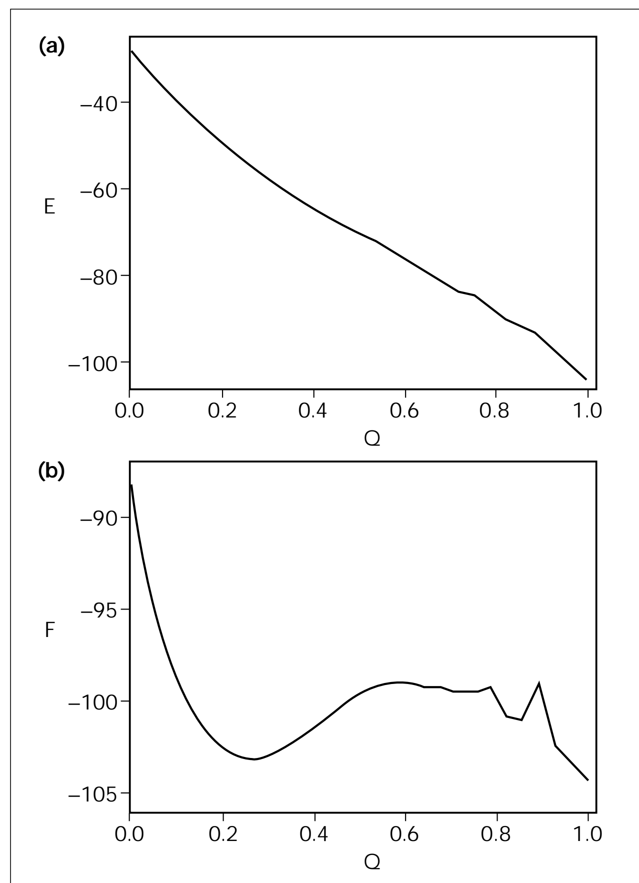
Although the focus on energy and free energy surfaces is thus not really new, it is only in the past two or three years that approaches that embody this concept have begun to play a more central role in the field of protein folding. This change is the result of the conjuncture of several factors. The first factor is the growth of interest in protein folding engendered to a large extent by the human and other genome projects [36]. The ever-increasing number of known protein sequences has raised a demand for understanding how proteins fold and for the ability to predict the native structure from the amino acid sequence. This has led many people, including physicists and mathematicians as well as chemists and biologists, to become involved in the protein folding problem. The second factor is that experiments on protein folding and unfolding have begun to provide detailed structural information that can be used to test theoretical descriptions. Important examples are data from NMR [37] and from protein engineering [38]. Also very recently, it has become possible to rapidly trigger folding and unfolding so that measurements can be extended from milliseconds down to the microsecond [39] and nanosecond timescales [40,41].

Figure 1



Folding of a 27-bead heteropolymer on a cubic lattice at low temperature. **(a)** The average effective energy, E, as a function of Q, the fraction of native contacts; there are 28 in the native state, which is a $3 \times 3 \times 3$ cube. **(b)** The average effective free energy, F, as a function of Q. Results adapted from Figure 4 of [42] at T = 0.7. Details of the definition and methods are given in [42].

Figure 2



Folding of a 27-bead heteropolymer on a cubic lattice at high temperature. **(a,b)** See Figure 1 for explanation. The sharp peak in (b) near 0.9 appears to be an artefact of the lattice simulation. The same sequence as in Figure 1 was used except that the native state was stabilized by offsetting the energies of all 28 native contacts by a random number from a Gaussian distribution with a mean of 0.0–0.8 k_BT and a standard deviation of 0.1 k_BT; T = 2.

The third factor was the publication of a Letter to *Nature* [42] which gave the first full analysis of the effective energy, entropy and free energy surface for the folding reaction of a protein modelled by a Monte Carlo simulation of a 27-bead heteropolymer with random interactions on a 3D lattice. It provided an explicit resolution of the Levinthal paradox: for this model, at least 30 out of 200 random sequences that were studied folded on the lattice in times many orders of magnitude shorter than that required to find the native state by sampling all of the configurations; e.g. with 'time' measured in Monte Carlo steps, only on the order of 5×10^7 steps were required for folding by the 30 sequences, while there are on the order of 10^{16} configurations in the conformational space. The effective energy and free energy sampled at different temperatures during the folding reaction are shown in Figures 1 and 2 as a function of the fraction of native contacts, Q, the progress variable for the reaction employed for this case; other measures of structural change may also be useful for describing the folding reaction [43]. At low temperatures (Figure 1), folding begins with a rapid hydrophobic collapse that leads to a decrease in energy but does not increase the value of Q significantly from that of the random coil state. In the region between Q = 0.2 and 0.7, the energy does not decrease (in fact, the surface is seen to be 'rough') and the polypeptide chain makes a slow stochastic search within the collapsed state to find the transition region; the transition region is fairly close to the native state, as indicated by the free energy profile (Figure 1b). From this region, the chain folds rapidly to the native state. By contrast, at high temperatures (Figure 2) there is a nearly monotomic decrease in the effective energy (an energy 'basin' or 'funnel') and there is an entropic barrier (a folding 'bottleneck') in the free energy surface prior to reaching the native state. As discussed in [42], the scenarios in Figures 1 and 2 may

describe the folding of small proteins; more complex scenarios are expected for larger proteins from lattice simulations [44–46] and from experiment [47].

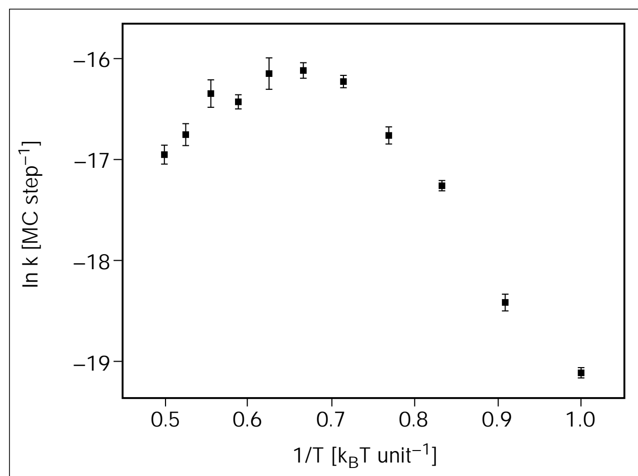
From a large number of trajectories calculated at a series of temperatures, it was shown that folding has the simple exponential behavior expected for a unimolecular reaction with an activation barrier [48]. An Arrhenius plot of the calculated rate constant as a function of temperature is shown in Figure 3. At low temperatures, the rate constant increases with temperature, as expected for an activation energy dominated reaction. As the temperature increases, there is a striking deviation from simple Arrhenius behavior. The rate reaches a maximum and then decreases with increasing temperature. This behavior results from the fact that the activation free energy of the folding reaction, which is energy dominated at low temperatures, becomes entropy dominated at high temperatures; see the effective energy and free energy surfaces shown in Figures 1 and 2. Interaction parameters can be chosen (e.g. for different 'sequences' or different solution conditions) so that behavior corresponding to Figure 1 or to Figure 2 occurs at the 'physiological' folding temperature.

The experimental rate constant for folding of several proteins [49] has been shown to have the behavior illustrated in Figure 3. At the normal folding temperature (i.e. in the neighborhood of 25°C), the temperature dependence for the fast-folding protein CI2, for example, shows Arrhenius-type behavior (see Figure 1 of [49]), suggesting that the activation barrier is energy dominated at that temperature. Comparison with the lattice results indicates that the effective energy surface is not a simple funnel, i.e. the surface may be intermediate between Figures 1 and 2. At higher temperature, inverse Arrhenius behavior is observed for CI2, as in Figure 3, which indicates that the effective energy surface is more funnel-like.

The awareness of lattice simulations in the wider community of experimentalists is, in part, an accident. It is likely to be due not to the Letter to *Nature* [42] itself, but to the publication of a News & Views article written about it by Baldwin [50]. It is recognized that more people read the News & Views articles than the Letters to which they refer. Baldwin pointed to a 'new view' of protein folding that emerged from the Letter [42], which showed that "the ability to fold rapidly is possessed by the same sequences that can form thermodynamically stable structures." To characterize the change in viewpoint, it is useful to quote from Baldwin's News & Views [50] and from one written by him only four years earlier [51] to comment on experiments by Fersht and co-workers [52,53] on a folding intermediate in barnase.

In 1990, Baldwin wrote [51]: "After many years of scepticism, it need no longer be doubted that proteins go through a series of identifiable intermediate changes as they fold up to assume their native conformations… the ghost of the 'jigsaw-puzzle' model, which postulates a random collection of folding intermediates, has now been laid to rest."

Figure 3



Arrhenius plot of the folding reaction rate constant as a function of temperature for the sequence used in Figure 1. Adapted from [48].

This quotation, which clearly supports the pathway concept of Levinthal, is to be contrasted with Baldwin's statement in 1994 [50]: "In the Monte Carlo simulations, molecules with a given sequence undergo folding on many different pathways at the same time. The simulations therefore indicate that there is no unique pathway of folding, and no unique transition state. On each of these pathways there is a high-energy intermediate that is close in structure to the native form, and thereby resembles the unique transition state sometimes invoked by experimentalists; however, the folding of a 27-bead chain involves about 10^3 such 'transition states'. Until now, experimentalists have drawn an analogy between protein folding and an ordinary chemical reaction that has a defined series of intermediates and a single rate-limiting step. An immediate challenge for them arising from the findings of Karplus and co-workers is to determine whether or not the folding of real proteins has a unique transition state."

The challenge was taken up quickly by Fersht *et al.* [54] in a paper comparing protein engineering experiments with the lattice model results in [42] (see also the comments in [34]).

The 'new view' expressed by Baldwin represents a change in the reference for protein folding studies from the random search paradox of Levinthal to a search biased toward the native state in an essential way by the variation of the effective energy of the polypeptide chain as a function of its

conformation. In Figure 2 there is a monotomic energy bias toward the native state, while in Figure 1 there is not. However, in both limits there is a sufficient reduction in the space that is sampled by the polypeptide chain to make folding possible on a reasonable timescale; simple arguments suggest that the folding time for a protein whose effective energy surface is represented by Figure 2 would be on the order of a millisecond or less, while it would be ~1 second for a protein corresponding to Figure 1. This 'new view' is no different, in principle, from the idea embodied in the harmony principle of Go [21] and the minimum frustration principle of Bryngelson and Wolynes [12,31], nor, as stated above, from elements implicit in phenomenological folding models.

The omission in the News & Views article by Baldwin of references to the work of the many other people (e.g. [12,21,26,33,55]) who contributed to developing the 'new view' led to a personal, as well as a published, correspondence [56,57] that appears to have contributed to the dissemination of the conclusions. Moreover, the impact of lattice-based folding simulations has continued to grow with a veritable deluge of papers published recently. They have demonstrated the versatility of lattice simulations and their ability to mimic a wide range of folding behavior [19,32–34,43–46,55,58–60]. An important element in the change in the perception of the protein folding problem has been the felicitous introduction of the 'folding funnel' [60–62] as an idealized construct for the free energy surface of the polypeptide chain. The funnel paradigm [60,62] was based on calculations [43] with a 27-mer heteropolymer model that essentially reproduced the results of Šali *et al.* (Figure 4 of [42]) with a somewhat different choice of interactions; that two different models for the interaction give very similar surfaces helps to establish the generality of the original result. The 'folding funnel', as first drawn in [60], represents the energy and entropy of the polypeptide chain as a function of the progress variable Q. (See also [19], which shows colorful illustrations of idealized funnels in three dimensions without specification of the nature of the reaction coordinates that are involved.) In [60,62], there is a funnel-like narrowing as Q increases that indicates a reduction of the configuration space (decrease in entropy). Depending on the system parameters, this could be a 'trivial' consequence of the topology of the system (i.e. there are fewer states for larger Q) or it could be the result of energy biases that lead to a physically significant reduction in the number of states accessible at a given temperature; an example of the latter is the initial hydrophobic collapse that occurs in the folding scenario shown in Figure 1. The other aspect of the folding funnel is that there is a more-or-less monotomic decrease in effective energy as Q increases; this corresponds to the surface in Figure 2a. A funnel that leads to fast folding must have a decrease in the effective energy as a function of Q that is sufficient to balance most of the entropy decrease. In the lattice simulation shown in Figure 2, there results a free energy with an entropic barrier at high Q (Figure 2b). The lattice results have been supported by model calculations [63–65] which suggest that the expected timescale for folding on simple funnel-like energy surfaces can be on the order of that observed experimentally.

Since the new view is based on general theoretical considerations and on simplified generic lattice simulations, it does not (cannot) provide information on the actual effective energy surface for a protein. Thus, the protein folding problem has now gone full circle or, perhaps better, a full turn of a spiral toward real understanding. We are much more optimistic about being able to solve the folding problem because the Levinthal paradox is no longer a concern. However, we are now faced with the issue of how the energy bias toward the native state is made to extend over a sufficient portion of the configuration space to make folding possible on the experimental timescale. There is as yet no direct evidence (it will be very hard to obtain it!) that the landscape paradigm embodied in a simple funnel applies to any specific protein. Funnel-like biased energy surfaces have been used to analyze experimental data on the folding of several proteins [66,67], but alternative interpretations of the results are generally possible. One recent example is the interesting study of a series of Ala→Gly mutants of the λ repressor, a small fast-folding protein [68]. The paper points out that the results are consistent with funnel landscapes, the diffusion-collision model [23], the framework model [25], and the minimal kinetic description of a single transition state barrier that is altered by the mutations.

To know what is going on in real proteins requires an understanding of the features of the potential energy function and the aspects of the native and intermediate structures of proteins that are involved in achieving the required energy bias, on the one hand, and avoiding deep traps, on the other. It is likely that certain of the phenomenological models will be resurrected in achieving this understanding [68,69]. However, their use will now often be phrased in terms of how they affect the energy landscape of the folding polypeptide chain [70]. Simulation techniques based on realistic models of proteins with all-atom potential functions and explicit or implicit solvent models can aid in relating the 'new view' to real proteins. Such simulations have already been used to probe certain aspects of the unfolding dynamics [71–73] and the free energy surface of a small three-helix bundle protein [74]. Improved simulations which are becoming feasible with advances in computer technology, when combined with the new wave of experiments that are providing more detailed structural and kinetic information, can be expected to solve the real protein folding problem ("*Le bon Dieu est dans le détail*" [75]) in the next few years.

Acknowledgements
The work reported in this paper was supported in part by a grant from the National Science Foundation. It describes a personal view of recent developments in the theory of protein folding that was presented at the symposium on 'Distance-Based Approaches to Protein Structure Determination III' (Copenhagen, November 5–8, 1996). I wish to thank particularly A Šali and CM Dobson for helpful comments and A Šali for assistance with the figures.

References
1. Karplus, M. & Shakhnovich, E. (1992). Protein folding: theoretical studies of thermodynamics and dynamics. In *Protein Folding*. (Creighton, T., ed.), pp. 127–195, W.H. Freeman & Sons, New York.
2. Elofsson, A., Fischer, D., Rice, D.W., LeGrand, S.M. & Eisenberg, D. (1996). A study of combined structure/sequence profiles. *Fold. Des.* **1**, 451–461.
3. Šali, A. (1995). Modelling mutations and homologous proteins. *Biotechnology* **6**, 437–451.
4. Moult, J., Judson, R., Fidelis, K. & Pedersen, J. (1995). A large scale experiment to assess protein structure prediction methods. *Proteins* **23**, 454–460.
5. Brooks, C.L., III, Karplus, M. & Pettitt, B.M. (1988). *Proteins: A Theoretical Perspective of Dynamics, Structure, & Thermodynamics. Adv. Chem. Phys. LXXI*, John Wiley & Sons, New York.
6. Phillips, D.C. (1981). Concluding remarks. In *Biomolecular Stereodynamics **II***. (Sarma, R.H., ed.), pp. 497–498, Adenine Press, Guilderland, New York.
7. Tanford, C. (1980). *The Hydrophobic Effect*. (2nd edn), John Wiley & Sons, New York.
8. McCammon, J.A., Gelin, B.R. & Karplus, M. (1977). Dynamics of folded proteins. *Nature* **267**, 585–590.
9. Karplus, M. & Petsko, G.A. (1990). Molecular dynamics simulations in biology. *Nature* **347**, 631–639.
10. Marquart, M., Deisendorfer, J., Huber, R. & Palm, W. (1980). Crystallographic refinement and atomic models of the intact immunoglobulin molecule Kol and its antigen-binding fragment at 3.0 Å and 1.9 Å resolution. *J. Mol. Biol.* **141**, 369–381.
11. Levinthal, C. (1969). How to fold graciously. In *Mossbauer Spectroscopy in Biological Systems, Proceedings of a Meeting held at Allerton House, Monticello, Illinois*. (Debrunner, P., Tsibris, J.C.M. & Münck, E., eds.), p. 22, University of Illinois Press, Urbana.
12. Bryngelson, J.D. & Wolynes, P.G. (1989). Intermediates and barrier crossing in a random energy model (with applications to protein folding). *J. Phys. Chem.* **93**, 6902–6915.
13. Schmid, F.X. (1992). Kinetics of unfolding and refolding of single domain proteins, In *Protein Folding*. (Creighton, T., ed.), pp. 197–241, W.H. Freeman & Sons, New York.
14. Ngo, J.T. & Marks, J. (1992). Computational complexity of a problem in molecular structure prediction. *Protein Eng.* **5**, 313–321.
15. Unger, R. & Moult, J. (1993). Finding the lowest free energy conformation of a protein is an NP-hard problem: proof and implications. *Bull. Math. Biol.* **55**, 1183–1198.
16. Fraenkel, A.S. (1993). Complexity of protein folding. *Bull. Math. Biol.* **55**, 1199–1210.
17. Ngo, J.T., Marks, J. & Karplus, M. (1994). Computational complexity, protein structure prediction, and the Levinthal paradox. In *The Protein Folding Problem and Tertiary Structure Prediction*. (Merz, K. Jr. & Le Grand, S., eds.), pp. 435–508, Birkhäuser, Boston, MA.
18. Levinthal, C. (1968). Are there pathways for protein folding? *J. Chim. Physique*. **65**, 44–45.
19. Dill, K. & Chan, H.S. (1997). From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10–19.
20. Wetlaufer, D.B. (1973). Nucleation, rapid folding and globular intrachain regions in proteins. *Proc. Natl Acad. Sci. USA* **70**, 697–701.
21. Go, N. (1983). Theoretical studies of protein folding. *Annu. Rev. Biophys. Bioeng.* **12**, 183–210.
22. Tsong, T.Y., Baldwin, R.L. & McPhie, P. (1972). A sequential model of nucleation-dependent protein-folding: kinetic studies of ribonuclease A. *J. Mol. Biol.* **63**, 453–475.
23. Karplus, M. & Weaver, D.L. (1976). Protein-folding dynamics. *Nature* **260**, 404–406.
24. Karplus, M. & Weaver, D.L. (1994). Folding dynamics: the diffusion-collision model and experimental data, *Protein Sci.* **3**, 650–668.
25. Kim, P.S. & Baldwin, R.L. (1982). Specific intermediates in the folding reactions of small proteins and the mechanism of folding. *Annu. Rev. Biochem.* **51**, 459–489.
26. Harrison, S.C. & Durbin, R. (1985). Is there a single pathway for the folding of a polypeptide chain? *Proc. Natl Acad. Sci. USA* **82**, 4028–4030.
27. Bashford, D., Weaver, D.L. & Karplus, M. (1984). Diffusion-collision model for the folding kinetics of the λ-repressor operator-binding domain, *J. Biomol. Struct. Dyn.* **1**, 1243–1255.
28. Bashford, D., Cohen, F.E., Karplus, M., Kuntz, I.D. & Weaver, D.L. (1988). Diffusion-collision model for the folding kinetics of myoglobin. *Proteins* **4**, 211–227.
29. Levine, R.E. & Bernstein, R.B. (1987). *Molecular Reaction Dynamics and Chemical Reactivity*. Oxford University Press, Oxford.
30. Dill, K.A. (1985). Theory for the folding and stability of globular proteins. *Biochemistry* **24**, 1501–1509.
31. Bryngelson, J.D. & Wolynes, P.G. (1987). Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. USA* **84**, 7524–7528.
32. Bryngelson, J.D., Onuchic, J.N., Socci, N.D. & Wolynes, P.G. (1995). Funnels, pathways and the energy landscape of protein folding: a synthesis. *Proteins* **21**, 167–195.
33. Dill, K.A., *et al.*, & Chan, H.S. (1995). Principles of protein folding – a perspective from simple exact models. *Protein Sci.* **4**, 561–602.
34. Karplus, M. & Šali, A. (1995). Theoretical studies of protein folding and unfolding. *Curr. Opin. Struct. Biol.* **5**, 58–73.
35. Go, N. & Abe, H. (1981). Noninteracting local-structure model of folding and unfolding transition in globular proteins. I. Formulation. *Biopolymers* **20**, 991–1011.
36. Woodhead, A.D. & Barnhart, B.J., eds. (1988). *Biotechnology and the Human Genome. Innovations and Impact 46*. Plenum Press, New York.
37. Miranker, A.D. & Dobson, C.M. (1996). Collapse and cooperativity in protein folding. *Curr. Opin. Struct. Biol.* **6**, 31–42.
38. Fersht, A.R. (1994). Characterizing transition states in protein folding: an essential step in the puzzle. *Curr. Opin. Struct. Biol.* **5**, 79–84.
39. Nolting, B., Golbik, R. & Fersht, A.R. (1995). Submillisecond events in protein folding. *Proc. Natl Acad. Sci. USA* **92**, 10668–10672.
40. Ballew, R.M., Sabelko, J. & Gruebele, M. (1996). Direct observation of fast protein folding: the initial collapse of apomyoglobin. *Proc. Natl Acad. Sci. USA* **93**, 5759–5764.
41. Phillips, C.M., Mizutani, Y. & Hochstrasser, R.M. (1995). Ultrafast thermally induced unfolding of RNase A. *Proc. Natl Acad. Sci. USA* **92**, 7292–7296.
42. Šali, A., Shakhnovich, E. & Karplus, M. (1994). How does a protein fold? *Nature* **369**, 248–251.
43. Socci, N.D., Onuchic, J.N. & Wolynes, P.G. (1996). Diffusion dynamics of the reaction coordinate in protein folding. *J. Chem. Phys.* **104**, 5860–5868.
44. Shakhnovich, E.I. (1994). Proteins with selected sequences fold into unique native conformation. *Phys. Rev. Lett.* **72**, 3907–3910.
45. Abkevich, V.I., Gutin, A.M. & Shakhnovich, E.I. (1995). Domains in folding of model proteins. *Protein Sci.* **4**, 1167–1177.
46. Dinner, A.R., Šali, A. & Karplus, M. (1996). The folding mechanism of larger model proteins: role of native structure, *Proc. Natl Acad. Sci. USA* **93**, 8356–8361.
47. Dobson, C.M., Evans, P.A. & Radford, S.E. (1994). Understanding how proteins fold: the lysozyme story so far. *Trends Biochem. Sci.* **19**, 31–37.
48. Karplus, M., Caflisch, A., Šali, A. & Shakhnovich, E. (1995). Protein dynamics: from the native to the unfolded state and back again. In *Modelling of Biomolecular Structures and Mechanisms*. (Pullman, A. *et al.*, eds.), pp. 69–84, Kluwer Academic Publishers, Amsterdam.
49. Oliveberg, M., Tan, Y.-J. & Fersht, A.R. (1995). Negative activation enthalpies in the kinetics of protein folding. *Proc. Natl Acad. Sci. USA* **92**, 8926–8929.
50. Baldwin, R.L. (1994). Matching speed and stability. *Nature* **369**, 183–184.
51. Baldwin, R.L. (1990). Pieces of the folding puzzle. *Nature* **346**, 409–410.
52. Bycroft, M., Matouschek, A., Kellis, J.T. Jr., Serrano, L. & Fersht, A.R. (1990). Detection and characterization of a folding intermediate in barnase by NMR. *Nature* **346**, 488–490.
53. Matouschek, A., Kellis, J.T. Jr., Serrano, L., Bycroft, M. & Fersht, A.R. (1990). Transient folding intermediates characterized by protein engineering. *Nature* **346**, 440–445.
54. Fersht, A.R., Itzhaki, L.S., ElMasry, N.F., Matthews, J.M. & Otzen, D.E. (1994). Single versus parallel pathways of protein folding and fractional formation of structure in the transition state. *Proc. Natl Acad. Sci. USA* **91**, 10426–10429.

55. Camacho, C.J. & Thirumalai, D. (1993). Kinetics and thermodynamics in model proteins. *Proc. Natl Acad. Sci. USA* **90**, 6369–6372.
56. Chan, H.S. (1995). Kinetics of protein folding. *Nature* **373**, 664–665.
57. Karplus, M., Šali, A. & Shakhnovich, E. (1995). Comment: kinetics of protein folding. *Nature* **373**, 664–665.
58. Shakhnovich, E. (1996). Modelling protein folding: the beauty and power of simplicity. *Fold. Des.* **1**, R50–R54.
59. Kolinski, A. & Skolnick, J. (1993). Lattice representations of globular proteins: how good are they? *J. Comp. Chem.* **14**, 1194–1202.
60. Onuchic, J.N., Wolynes, P.G., Luthey-Schulten, Z. & Socci, N.D. (1995). Toward an outline of the topology of a realistic protein-folding funnel. *Proc. Natl Acad. Sci. USA* **92**, 3626–3630.
61. Leopold, P.E., Montal, M. & Onuchic, J.N. (1992). Protein folding funnels: a kinetic approach to the sequence–structure relationship. *Proc. Natl Acad. Sci. USA* **89**, 8721–8725.
62. Wolynes, P.G., Onuchi, J.N. & Thirumulai, D. (1995). Navigating the folding routes. *Science* **267**, 1619–1620.
63. Zwanzig, R. (1995). Simple model of protein folding kinetics. *Proc. Natl Acad. Sci. USA* **92**, 9801–9804.
64. Finkelstein, A.V. & Badretdinov, A.Y. (1997). Rate of protein folding near the point of thermodynamic equilibrium between the coil and the most stable chain fold. *Fold. Des.* **2**, 115–121.
65. Durup, J. (1997). On "Levinthal paradox" and the theory of folding. *Theochem.* in press.
66. Onuchic, J.N., Socci, N.D., Luthey-Schulten, Z. & Wolynes, P.G. (1996). Protein folding funnels: the nature of the transition state ensemble. *Fold Des.* **1**, 441–450.
67. Schoemaker, B.A., Wang, J. & Wolynes, P.G. (1997). Structural correlations in protein folding funnels. *Proc. Natl Acad. Sci. USA* **94**, 777–782.
68. Burton, R.E., Huang, C.S., Daugherty, M.A., Calderone, T.L. & Oas, T.G. (1997). The energy landscape of a fast-folding protein mapped by Ala→Gly substitution. *Nat. Struct. Biol.* **4**, 305–310.
69. Baldwin, R.L. (1996). Why is protein folding so fast? *Proc. Natl Acad. Sci. USA* **93**, 2627–2628.
70. Eaton, W.A., Thomson, P.A., Chan, C.-K, Hagen, S.J. & Hofrichter, J. (1996). Fast events in protein folding. *Structure* **4**, 1133–1139.
71. Daggett, V. & Levitt, M. (1994). Protein folding→unfolding dynamics. *Curr. Opin. Struct. Biol.* **4**, 291–295.
72. Caflisch, A. & Karplus, M. (1994). Molecular dynamics studies of protein and peptide folding and unfolding. In *The Protein Folding Problem and Tertiary Structure Prediction.* (Merz, K., Jr. & Le Grand, S., eds.), pp. 195–232, Birkhäuser, Boston, MA.
73. Li, A. & Daggett, V. (1994). Characterization of the transition state of protein unfolding by use of molecular dynamics: chymotrypsin inhibitor 2. *Proc. Natl. Acad. Sci. USA* **91**, 10430–10434.
74. Boczko, E.M. & Brooks C.L., III (1995). First principles calculation of the folding free energy of a three-helix bundle protein. *Science* **269**, 393–396.
75. A quotation usually attributed to G Flaubert. It is interesting that in the United States, the expression "The Devil is in the details" is used instead.