Protein Science (1999), *8*:1166–1180. Cambridge University Press. Printed in the USA.

REVIEW

Polymer principles and protein folding

KEN A. DILL[1]
University of California, San Francisco, 3333 California Street, Ste. 415, San Francisco, California 94118

(Received January 25, 1999; Accepted March 4, 1999)

Abstract

This paper surveys the emerging role of statistical mechanics and polymer theory in protein folding. In the polymer perspective, the folding code is more a solvation code than a code of local $\phi\psi$ propensities. The polymer perspective resolves two classic puzzles: (1) the Blind Watchmaker's Paradox that biological proteins could not have originated from random sequences, and (2) Levinthal's Paradox that the folded state of a protein cannot be found by random search. Both paradoxes are traditionally framed in terms of random unguided searches through vast spaces, and vastness is equated with impossibility. But both processes are partly *guided*. The searches are more akin to balls rolling down funnels than balls rolling aimlessly on flat surfaces. In both cases, the vastness of the search is largely irrelevant to the search time and success. These ideas are captured by energy and fitness landscapes. Energy landscapes give a language for bridging between microscopics and macroscopics, for relating folding kinetics to equilibrium fluctuations, and for developing new and faster computational search strategies.

Keywords: new view; polymer; protein folding; statistical mechanics

This paper describes a perspective on protein folding that derives in part from simple statistical mechanical and polymer models. As with any perspective, this one is a personal opinion, with all the limitations that implies. The first part of this paper explores the folding code. (1) *Structure*: How is the native structure encoded in the amino acid sequence? (2) *Thermodynamics*: Why is folding so cooperative? (3) *Kinetics*: What determines the speed and the rate-limiting steps of folding? Polymer modeling suggests that the folding code is more a solvation code and less a linear encoding of torsion angles along the peptide bond, even though the latter is not negligible. The second part explores the energy landscape perspective on folding kinetics. Polymer modeling suggests that the folding process more closely resembles balls rolling down bumpy funnels than balls rolling aimlessly on flat surfaces or rolling single file along identical trajectories.

DISCUSSION

Side-chain interactions contribute to architecture, just as backbone interactions do

The backbone forces of folding

Table 1 compares two different perspectives on the folding code. A backbone-centric, helix-centric perspective arose over the 50 year time frame, from the 1930s to 1980s. It originated with Mirsky and Pauling in 1936 (Mirsky & Pauling, 1936), who proposed that backbone hydrogen bonding is a prominent folding force. During the next 15 years, Pauling's group used the structures of small molecule hydrogen-bonding compounds to predict that folded proteins would have α-helical and β-sheet structures (Pauling & Corey, 1951a, 1951b, 1951c, 1951d; Pauling et al., 1951). The first X-ray crystal structures of globular proteins gave strong support to this view by confirming the existence of the predicted α-helices and β-sheets (Kendrew et al., 1958). Hydrogen bonding was seen to be an important structure-causing force in proteins.

During the same period, a step was taken toward understanding folding cooperativity through an understanding of the helix-coil transition. For many years it had been known that protein folding is cooperative, i.e., that there is a dramatic transition from denatured to native states upon only small changes in solvent, pH, or temperature. In the 1950s and 1960s, theoretical work particularly of Schellman (1958), Zimm and Bragg (1959), Poland and Scheraga (1970), and experiments (Doty & Yang, 1956; Doty et al., 1956) showed that long peptide chains can undergo a helix-coil transition that is cooperative. The helix-coil transition is driven by hydrogen bonding and $\phi\psi$ propensities among near-neighbor groups along the chain. For many years, this has been the main model for conformational cooperativity in biomolecules.

To complete the picture of structure, thermodynamics, and kinetics, experiments beginning in the 1970s showed that helices can form rapidly (Kim & Baldwin, 1982; Williams et al., 1996). One inference was that folding is hierarchical and can be explained by a scheme $1° \rightarrow 2° \rightarrow 3°$: the primary structure leads to secondary structure (fast), which is then assembled into tertiary structure (slower). Hierarchical assembly was seen as a solution to the prob-

Reprint requests to: Ken A. Dill, University of California, San Francisco, 3333 California Street, Ste. 415, San Francisco, California 94118; e-mail: dill@maxwell.ucsf.edu.

[1] The author is grateful to Hans Neurath, the Protein Society, and *Protein Science*, for the opportunity to present this overview, which is largely taken from a talk given on the occasion of the Hans Neurath Award lecture, at the Protein Society meeting, July 27, 1998.

Table 1.

	Backbone-centric view	Side-chain centric view
Dominant force	ΦΨ, Hydrogen bonds	Hydrophobicity, hydrogen bonds
Thermodynamic cooperativity	Helix-coil transition	Collapse transition
Kinetics	Helix formation is fast	Desolvation is fast
Role of hydrophobicity	Nonspecific	Drives specific architecture
Folding code	ΦΨ-centric (1° → 2° → 3°)	Solvation code

lem of how the protein sorts through conformational space haystack quickly on its way to finding the native state needle. The same hierarchy has been widely explored as a computational strategy for predicting native states from amino acid sequences: use local helix and sheet propensities to predict secondary structures, then assemble them into tertiary structures.

The upshot was a perspective in which the backbone interactions—hydrogen bonding and $\phi\psi$ propensities—have been seen as a large part of the explanation of the structures, thermodynamics, and kinetics of protein folding (Honig & Cohen, 1996; Aurora et al., 1997; Baldwin & Rose, 1999a, 1999b). The $\phi\psi$ propensities are not equivalent to hydrogen bonding, since hydrogen bonds are involved in nonlocal interactions, whereas $\phi\psi$ interactions, by definition, are not. Nevertheless, from the perspective of the *sequence-dependent* interactions, and sequence-structure relationships, a backbone-centric view is largely a $\phi\psi$-centric view, since there has been little basis for believing one amide-carbonyl backbone hydrogen bond has a substantially different strength than another in a sequence-dependent way. On the other hand, hydrophobic interactions, which were first identified as important for protein folding by Kauzmann (1959), were seen as a nonspecific glue that aided collapse but otherwise played little role in dictating the specific architectures of native proteins (Anfinsen & Scheraga, 1975). Hydrophobic interactions are mainly expressed by the side chains. They are "through-space" and solvent-mediated contact interactions, rather than "through-nearest-neighbor-bonds," as are $\phi\psi$ interactions. This distinction between torsion-based nearest-neighbor "through-chain" interactions that involve $\phi\psi$ angles, and contact-based "through-space" interactions, that involve displacement or exchange of solvent, seems less ambiguous than distinctions between secondary vs. tertiary forces, or local vs. nonlocal forces. The $\phi\psi$ interactions are mainly steric torsional constraints captured in Ramachandran plots. Contact interactions, such as side-chain contacts, include hydrogen bonding and hydrophobic interactions and van der Waals interactions among non-neighboring monomers.

The $\phi\psi$ perspective does not address a key issue. As with most other polymers, a large conformational space is a consequence of weak preferences of each monomer unit for one region of torsion-angle space relative to another region. But if they are to account for the folding code, $\phi\psi$ propensities must be different in the native state than in the denatured state. In particular, $\phi\psi$ interactions must change when folding conditions are turned on. But there is little evidence that tri- and tetra-peptides adopt native-like conformations and overcome the chain entropy, when the solvent or temperature are changed.

The side-chain forces of folding

A different perspective has developed from polymer modeling over about the past 15 years. The polymer perspective is side-chain centric, rather than backbone centric. The idea is that folding is dictated not so much by the propensities for nearest neighbor amino acids to favor particular $\phi\psi$ values (α-helix or β-sheet propensities), even though there is abundant evidence for such preferences (Honig & Cohen, 1996; Aurora et al., 1997). Rather, in the side-chain-centric view the greater contribution to the free energy of folding is encoded in a more delocalized "solvation" code: there are very few conformations of the full chain that can bury nonpolar amino acids to the greatest possible degree (Dill, 1985; Dill et al., 1995). Even short peptides, such as amphipathic helices, can be driven by solvation. Hydrophobic interactions, however they are defined, are among the strongest interactions among amino acids in water. And in large proteins, there are many of them. In this view, hydrophobic interactions are not nonspecific glue, but a crucial structure-determining driving force. In this view, folding cooperativity more closely resembles a process of polymer collapse in a poor solvent than a helix-coil transformation. In this view, fast secondary structure formation is less a consequence of strong helix propensities, and more an indirect consequence of a drive toward nonpolar desolvation.

The true balance between side-chain and backbone forces is not yet known. The side-chain-centric view has been based on the following logic. Simplified models that include side-chain interactions, but have the $\phi\psi$ preferences "turned off," predict many properties of globular proteins. In contrast, models that keep $\phi\psi$ propensities and turn off side-chain interactions predict only helices or strands and no compact folded state (Thomas & Dill, 1993). Indeed, helix-coil experiments show that $\phi\psi$ propensities control structures for sequences that are unable to collapse. For example highly charged poly-benzyl-L-glutamate is the classical helix-former.

It follows that a *minimal model* of globular protein behavior can be constructed from a side-chain-centric perspective but not from a backbone-centric perspective. This means that it may be possible to design polymers that could fold and perform protein-like functions, even without peptide backbones. RNA molecules already provide some proof of this principle. Minimal models are guides for such general principles.

But minimal models do not tell us the actual balance of forces in real proteins. If our goal is an accurate model of proteins, we undoubtedly cannot ignore backbone interactions (Honig & Cohen, 1996) or details of steric packing, or the different $\phi\psi$ interactions among the amino acids. In the end, since protein stability is a small difference of large interactions, all interactions can contribute to structure, thermodynamics, and kinetics.

What is the evidence for the side-chain-centric view? (1) A backbone centric view does not predict collapse. A coordination of $\phi\psi$ choices to cause collapse would be extraordinarily fortuitous. (2) Helix and strand propensities tend to be weak. Excepting poly-alanine-based sequences (Scholtz & Baldwin, 1992), peptides that are found to be in helices or strands in globular proteins are unstable when isolated in solution. Moreover, most helices and strands are amphipathic (Eisenberg et al., 1984; Bowie et al., 1990; Branden & Tooze, 1999), implicating solvation forces. The α-helical and β-strand propensities are context dependent (Kabsch & Sander, 1984; Minor & Kim, 1994), and the nonlocal interactions in β-sheets are large (Smith & Regan, 1995) and numerous. (3) In a globular protein, the number of local interactions is proportional to the number of amino acids *N*, but the number of nonlocal interactions is proportional to about 2*N*, so the latter should dominate in larger proteins. (4) Helices and strands often take their conformational instructions from their context or from the solvent (Kuroda et al., 1996; Predki et al., 1996). (5) To a first approximation, a fold is determined by the binary sequence of hydrophobic/polar monomers, even when $\phi\psi$ propensities are largely chosen randomly (Reidhaar-Olson & Sauer, 1988; Bowie et al., 1990; Lim & Sauer, 1991; Gassner et al., 1992; Lim et al., 1992; Kamtekar et al., 1993; Matthews, 1993; Munson et al., 1994, 1996; Lazar et al., 1997; Roy et al., 1997; Schafmeister et al., 1997; Wu & Kim, 1997). (6) Protein folds are less affected by mutations on their surfaces than in their hydrophobic cores (Lim & Sauer, 1991; Matthews, 1993). (7) Some experiments show that protein folding is not hierarchical, implying that secondary structures are not preassembled and used as building blocks in tertiary assembly. For example, a β-sheet protein can fold via a helical intermediate (Shiraki et al., 1995; Hamada et al., 1996). (8) Hydrophobic clustering, like secondary structure formation, can be very fast (Chan et al., 1997; Ramachandra Shastry & Roder, 1998; Ramachandra Shastry et al., 1998), and it can drive helix and sheet formation.

Simplified models are hypothesis generators

The predictions described above come, in part, from models that involve considerable simplification. An example is the HP model, in which each amino acid is represented as a bead, each bond is a straight line, bond angles are a few discrete options rather than a continuum, different conformations conform to lattices in two or three dimensions, and the 20 amino acids are condensed into a two-letter alphabet: H (hydrophobic) or P (polar) (Dill, 1985; Dill et al., 1995).

While statistical mechanical models are *simplified* in their representation of energies and atomic details, they are *more refined* in other respects (Camacho & Thirumalai, 1993; Bryngelson et al., 1995; Dill et al., 1995; Karplus, 1997; Onuchic et al., 1997): (1) their full conformational space can be explored extensively, sometimes without sampling or approximation, and (2) sometimes the full sequence space can be explored. For some questions, it is more important to get right the representation of conformational or sequence spaces than it is to get right the atomic details. Many questions of structure, stability, and kinetics are not about the locations of the hydrogen bonds in native lysozyme. They are not questions that are answerable by crystallography. They are about distributions and ensembles, flexibilities and entropies, energy landscapes, folding kinetics, big conformational changes, or sequence space. They are low-resolution questions that have low-resolution answers. Apart from X-ray crystallography and NMR, the workhorses of biomolecule science for many years have been low-resolution experiments—CD, fluorescence, small-angle scattering, some NMR experiments, calorimetry, chromatography, ANS binding, melting curves, etc.

For questions involving conformational ensembles, conformational entropies, sequence space, long time and large spatial scales, ensemble averaging, or non-native states like transition states, molten globules, intermediates or denatured states, there is currently little alternative to some degree of simplification in models. It is sometimes helpful *not* to have atomic details, picosecond by picosecond, because it is hard to see the forest of principles through the trees of detail. It would be a mistake to believe that any model is "improvable" by adding structural detail. Sometimes details are the problem, not the solution. This is a key message from the successes of the two-dimensional lattice Ising model in the revolution that took place in understanding critical phenomena (Stanley, 1971). The inability of earlier models of phase transitions to capture subtle critical behavior was attributable, not to the lack of realism and atomic detail, but to a lack of rigor in the mathematics of the models.

Mathematician Mark Kac once said that the purpose of models is "to polarize our thinking," to help us formulate questions. A model manifests a point of view; it regards certain components of a problem as relevant, important, or dominant, and other components as irrelevant, unimportant, or negligible, and then devises a chain of logic leading to predictions from those premises. Most broadly, the point of a model is to make decisive and testable predictions, regardless of whether its fine structure looks realistic. A key advantage of simplified models is that their parameters are physical and minimal in number. The chain of logic from premises to conclusions is direct. Simplified models serve to generate hypotheses that often cannot be generated in any other way, but that can then be tested by experiments or refined simulations.

Simplified models have been useful for exploring entropies and combinatoric principles of conformational and sequence spaces. Two problematic paradoxes of protein science have been shown by polymer modeling to be neither problematic nor paradoxical. (1) *The Blind Watchmaker Paradox*: The probability that natural proteins could be found in a random search of sequence space was seen to be impossibly small. (2) *The Levinthal Paradox*: The probability that a protein could find its native state by random search was seen to be impossibly small. Both paradoxes have been framed in terms of random unguided processes that search for a single point, the *endstate*, in a vast space. Biological evolution searches through sequence space; the endstate is a single protein having a particular function. Protein folding searches through conformational space; the endstate is the single native structure of a given protein. In both cases, the vastness of the search (i.e., the size of the search space) is taken, according to the paradox, to be the key to the impossibility of reaching the endstate.

Creationists, evolutionists, and blind watchmakers: Can proteins arise from random sequences?

Sequence space is a large place. For protein chains of 100 amino acids, the number of possible sequences of the 20 different amino acids is $20^{100} = 10^{130}$ (see Fig. 1). Creationists have used such numbers to argue the impossibility that proteins, and life, could

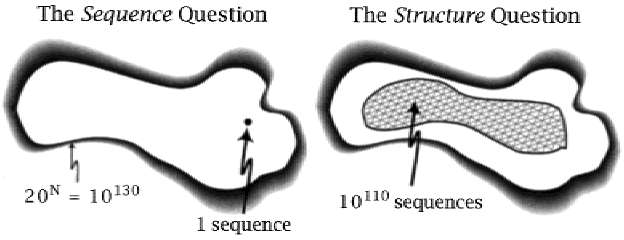

Fig. 1. Sequence space is large. There are 20^{100} different 100-mer sequences. The probability of finding one particular sequence is 20^{-100}, but the probability of finding any sequence that folds to a particular structure is predicted to be more than 100 orders of magnitude larger.

have arisen from the random sequences that were plausibly on the prebiotic earth. Creationists solve the large numbers problem by invoking divine intervention. Evolutionists solve the large numbers problem instead by the accretion of advantage that happens through natural selection (Dawkins, 1996). But both evolutionists and creationists start from the same premise, the large numbers problem. Evolutionists too assume that natural proteins are infinitesimal specks in an impossibly vast and meaningless sequence space, as indicated by the following quotes: "Only a very small fraction of this unimaginably large number of polypeptide chains would adopt a single stable three-dimensional conformation" (Alberts et al., 1998), and "It is certain that we need a hefty measure of cumulative selection in our explanations of life" (Dawkins, 1996). It is this numbers problem that I refer to, with the help of the wonderful metaphor of Richard Dawkins, as the Blind Watchmaker's paradox.

But statistical mechanical modeling (Lau & Dill, 1990; Chan & Dill, 1991; Lipman & Wilbur, 1991) shows that there is very little numbers problem in the first place. Reach into a soup of random amino acid sequences. The chance of pulling out a biologically important molecule depends on what is meant by "biologically important." The following two questions are vastly different (see Fig. 1): (1) What is the probability of pulling from that soup *a specific sequence*? (2) What is the probability of pulling from that soup *any sequence that folds to a specific structure*? The answer to question (1) is 10^{-130} for a 100-mer. The chance of pulling out a polypeptide having, say, the lysozyme sequence, is essentially zero. But to achieve biological function, we care only about finding a particular *fold*, not a particular *sequence*. And modeling shows that the probability of finding a structure is likely to be more than 100 orders of magnitude larger than the probability of finding a sequence (Lau & Dill, 1990; Chan & Dill, 1991). The chance of pulling out *any sequence* that folds to roughly lysozyme's structure is closer to 10^{-10} to 10^{-20}. While this number too may seem impossibly small, nature works with these sorts of numbers all the time. These numbers would imply about one such sequence in a liter of random sequences at nanomolar concentrations! And the probability of finding *any* chain fold, not just lysozyme, is even higher.

Why does the numbers problem disappear when we seek *structures* rather than *sequences*? There is an enormous "degeneracy" in sequence space: many different sequences can fold to the same native structure. A protein can be mutated substantially without changing its fold. The explanation for this is simple. If, as noted above, a fold is primarily determined by the binary sequence of hydrophobic/polar monomers, then the essential features of the full 20^{100} sequence space are found by searching a space of only about the $2^{100} = 10^{30}$ sequences that are written in a binary alphabet (H = hydrophobic, P = polar), a reduction of 100 orders of magnitude. Degeneracy means that hydrophobic monomers are largely interchangeable with each other, and polar monomers are interchangeable with each other, for determining a fold. (Function may have additional requirements, but estimates indicate that these do not change the numbers much (Lau & Dill, 1990).)

Moreover, modeling (Lau & Dill, 1990; Chan & Dill, 1991) and experiments (Reidhaar-Olson & Sauer, 1988; Matthews, 1993) show that the relevant space is even smaller, because only about $N/3$ of the residues are crucial for folding—those that define the hydrophobic core. To first approximation, most surface sites can be mutated without changing structure or function. (The factor of $N/3$ comes from the geometry of surface/volume ratios. Pack 100 amino acid spheres together to make a protein; 67 of them will be on the surface and 33 will be in the core.) Therefore, the real search for protein structure takes place, not in a space of 20^{N}, but in a space nearer in size to $2^{N/3} = 2^{33} = 10^{10}$ for $N = 100$. The other 120 orders of magnitude in sequence space are highly degenerate; the folded states of those sequences will look much like ones already found in the search of the smaller space.

Sequence space is therefore not likely to be vast darkness with infinitesimal specks of protein-like light spots. It is not perfectly light either. On a logarithmic scale, sequence space is predicted to be more like a beige sea in which virtually all molecules are "nearly folded." A typical random chain of 100 amino acids is predicted to be highly compact in water, have considerable secondary structure, and be structured much like a molten globule (Lau & Dill, 1990; Chan & Dill, 1991). This is a far better starting point for natural selection than are specks in astronomically large sequence spaces.

But whatever the starting point, there remains the need for a process of improvement. Evolution can improve proteins by natural selection. Richard Dawkins has explained natural selection using the metaphor of a Blind Watchmaker. By invoking "Watchmaker," he means the endstate has the appearance of having been designed. The traditional inference is that an object that appears designed was built by a systematic step by step procedure, as in building a watch. But Dawkins' term "Blind" means that, on the contrary, the natural selection process is not so systematic and does not involve a specific pre-ordained sequence of events. Natural selection is a Blind Watchmaker that improves proteins through incremental steps, each of which involves some bias, however small, at the same time it also involves considerable random choice among alternatives.

The Blind Watchmaker is also a useful metaphor for protein folding kinetics, the systematic accretion of native structure over the time course of a folding experiment (Zwanzig et al., 1992). A native protein has the appearance of design. For example, the steric fit of side chains in a protein core is as precise as that of a jigsaw puzzle. But tight packing of irregular objects can also be achieved by shaking up nuts and bolts in a jar, with no design involved (Bromberg & Dill, 1994). The appearance of design does not mean that folding happens in serial step by step fashion. Even a jigsaw puzzle can be constructed through different parallel sequences of events (Harrison & Durbin, 1985). The native structure can be reached, over the time course of folding, by a process that (1) starts from different initial conformations and (2) proceeds by incremental improvements, each of which has some bias but also involves considerable random choice among alternatives. Even a very small bias (deviation from randomness) in choosing among alternatives can speed up the search time (compared to a random search) by tens to hundreds of orders of magnitude (see below). When perfect randomness is not the driver, the vastness of the search becomes irrelevant to the search time (Dill, 1993). These ideas are captured in *landscapes*: fitness landscapes in sequence space or energy landscapes in conformational space. We now focus on the latter.

The old and new views of folding kinetics: Different questions

Protein folding kinetics has been described in terms of so-called Old and New Views (Baldwin, 1994, 1995; Dill & Chan, 1997). To define these views, we first distinguish *models*, old and new, from *views of microscopic folding processes*, old and new. For models, the terms old and new is too stark a contrast. It leads to the perception that we should be asking: Why do we need new models? What was wrong with the old ones? Which models are better? But these questions miss the mark. Old and new models do not address the same questions. The old models are mass action models used to fit experimental data on folding relaxation times and amplitudes. The new view is not a denial of these models. Mass action models remain valid for representing such data. While mass action modeling gives a *macroscopic* description of experimental data, the new statistical mechanical models give a *microscopic framework* to explain that data.

Where old and new views differ, however, is in their interpretation of the microscopic processes of folding. In the hope that changing terminology can help untangle some confusion, I will replace "old view" with "Sequential Micropath view" and "New view" with "Ensemble view."

Table 2 summarizes the differences between the two views of folding kinetics (Dill & Chan, 1997). The language of the Sequential Micropath view—pathways, transition states, reaction coordinates, on-path and off-path intermediates—is intended to explain *what exponentials do* (i.e., what you see in experiments). Experimental relaxation data are interpreted in terms of mass-action diagrams having arrows that connect symbols like *D* (denatured), *I* (intermediate), and *N* (native). Nothing in this language says what any one molecule is doing at any given time, or how the kinetics of folding is related to the monomer sequence, or how to assign microscopic chain conformations to labels *I* or *D* or transition state, etc. But the language of the Ensemble view—landscapes, folding funnels—is intended to describe *what molecules do* (i.e., how individual molecules progress toward the folded state), and how different monomer sequences lead to different kinetics.

Sequential micropath perspective

The main problem, according to the Sequential Micropath view, was the *search problem*, which has been called the *Levinthal paradox*. As Levinthal posed it (Levinthal, 1968; Wetlaufer, 1973), a random search of conformations would take a protein forever. Levinthal saw folding as a search through a vast conformational space, the haystack, for the native structure, the needle. Suppose the conformational space is represented by four preferred $\phi\psi$ angles for each peptide bond: α-helical, β-strand, and two others. In terms of those discrete options, the size of the space for a 100 residue chain is $4^{100} \approx 10^{60}$ chain conformations. Only one of these is the native structure. Levinthal's proposed solution for finding the needle in the haystack was that all chains must follow the same microscopic pathway, like ants single file on a trail (Levinthal, 1968). By "same pathway," he specified that every chain follows the same sequence of bond angle changes, in the same order, to reach the native state. In the Sequential Micropath view, kinetic intermediates (if they were on-pathway) were seen as helpful mileposts because they would show what route was taken, and therefore what routes were avoided, and therefore how the haystack was searched efficiently. Two-state kinetics was seen as uninformative about the mileposts of folding.

Table 2.

	Sequential micropath view	Ensemble view
Language	Paths, intermediates, transition states, reaction coordinates	Landscapes, funnels
Explains	What exponentials do (what you see)	What molecules do (how it works)
Main problem	Search problem	Trap problem
Proposed solution	Sequential pathways	Funnels
Intermediates	Mileposts	Traps
Two-state kinetics	No information	Implies fast folding

Ensemble perspective

In the Ensemble view, the vastness of the search is largely irrelevant. The more important problem is kinetic traps (Chan & Dill, 1994, 1998). Chains can sort very quickly through vast stretches of conformational space. In this view, chains fall energetically downhill, as when balls roll down bumpy funnels. Chains do not fold by random searches on level energy landscapes. In this view, two-state kinetics often means the chain is folding at nearly its maximum possible diffusion-limited speed, without kinetic traps. In this view, stable intermediates are mainly seen as kinetic traps that slow down the folding process.

Here's why the vastness of the search is irrelevant. Even a very small bias, in the form of the forces of protein folding, can be the difference between folding times measured in "lifetimes of the universe" vs. milliseconds (Bryngelson & Wolynes, 1987; Dill, 1987; Zwanzig et al., 1992). (On a large flat golf course, a golf ball will "never" find the hole by random processes, but if the golf course has even a small tilt that funnels toward the hole, no problem!)

What causes the funnel-like tilt on a folding landscape? The first estimates of the shapes of folding energy landscapes were based on mean-field theories (see Fig. 2) (Dill, 1985; Bryngelson & Wolynes, 1987). Hydrophobic collapse leads to compact chain conformations. The funnel arises because the drive to collapse is also a drive toward a reduced ensemble of conformations. There are many non-native states (high energy), but only one native state (low energy). The fraction of conformations that are compact is infinitesimal compared to the total conformational space. If there are 4^{100} conformations of a chain, Flory–Huggins-like theories predict that only about $(4/e)^{100} \approx 10^{17}$ of those conformations are compact (Dill, 1985). More accurate recent estimates predict a number that is even smaller (Yue & Dill, 1995): the number of compact conformations having a hydrophobic core may even be as small as 1 for some sequences. This estimate is supported by experiments on reduced alphabets based on hydrophobicity codes; a small fraction of sequences appear to fold relatively uniquely (Riddle et al., 1997; Roy et al., 1997; Schafmeister et al., 1997).

In short, the Ensemble view is a reversal of the sequential micropath view. What was seen as the slow step—the search through the huge haystack of non-native chain conformations—is now seen as happening at near diffusion-limited speed. Collapse can be fast. Sifting through most of the haystack is fast; the slow part is the endgame of reconfiguring a very small set of near-native conformations. In the past few years, new fast experimental methods (Burton et al., 1997; Callendar et al., 1999) have shown that proteins can fold at nearly diffusion-limited rates, on submillisecond time scales (Huang & Oas, 1995; Ballew et al., 1996a, 1996b; Pascher et al., 1996; Burton et al., 1997; Chan C-K et al., 1997;

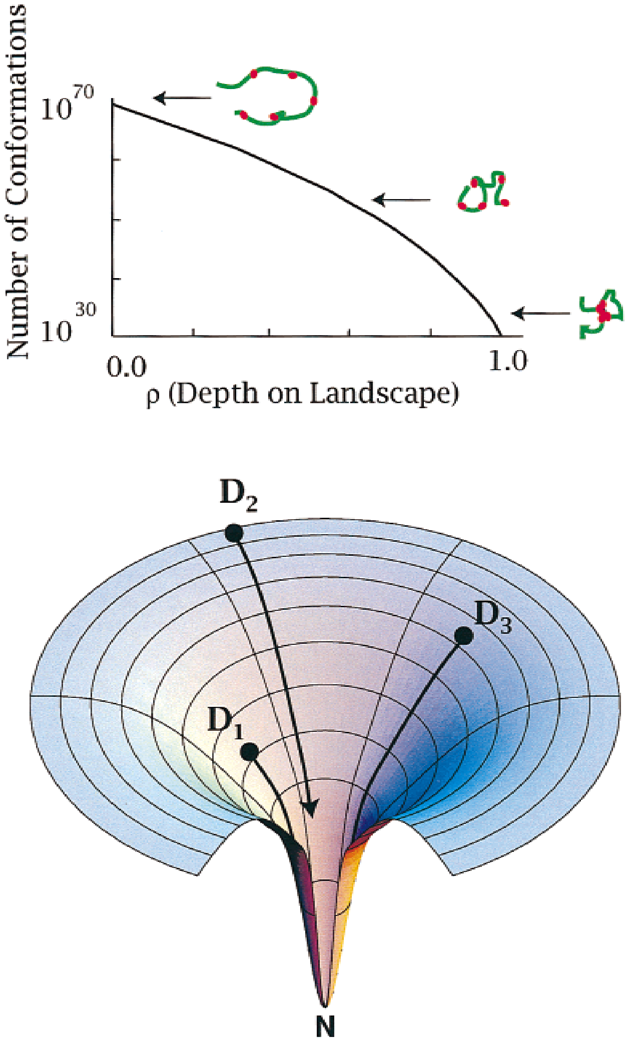

Fig. 2. Smooth funnel landscape (bottom). Denatured conformations follow different folding routes to the native state. The top figure shows the Flory–Huggins excluded volume estimate for the landscape shape (turned sideways): $\Omega \sim (N/\rho)!/[(N/\rho)^N(N/\rho - N)!]$, where Ω is the number of chain conformations, N is chain length, and $0 \leq \rho \leq 1$ is the compactness of the chain, an approximate measure of the depth on the landscape (Dill, 1985).

Gilmanshin et al., 1997a, 1997b, 1998; Ramachandra Shastry & Roder, 1998; Ramachandra Shastry et al., 1998). Energy landscapes provide the language that can help describe folding events at any level, from the microscopic to the macroscopic.

Energy landscapes connect single-chain microscopics to experimental macroscopics

One long-term goal of protein folding experiments has been to help understand the microscopic basis for the folding code. But this remains a promise, not a reality. Why? Prior to energy landscapes, there has been no way to connect the macroscopics that experiments measure to the microscopics that are needed in computational folding algorithms. Here is the problem.

Figure 3 illustrates the kind of folding kinetics data that is traditionally measured, and the mass-action models that are used to interpret them. When a single exponential decay is observed in both folding and unfolding directions, it is described as "two-state" kinetics, because two mass action symbols, such as *N* (native) and *D* (denatured), and an arrow interconnecting them, provide the simplest scheme that can model the data. But when multiple exponentials are observed, at least one additional symbol must be invoked in a mass-action law. When there are three such symbols, say *N*, *D*, and *I* (intermediate), there are two main ways those symbols have been interconnected by arrows: *I* is called an "On-pathway" intermediate or *I* is an "Off-pathway" intermediate.

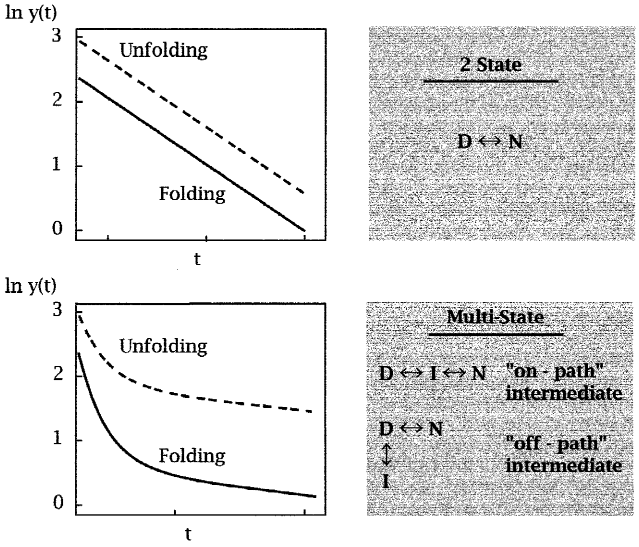

Fig. 3. Simple mass-action schemes describe observed relaxation rates and amplitudes, using symbols such as *N* (native), *D* (denatured), and *I* (intermediate).

But even when experiments provide a perfectly accurate mass action model for the folding and unfolding kinetics of a particular protein, it does not give enough information to make a microscopic model of how folding takes place. Experimental data are too averaged to inform the local decisions that must be made in conformational searching. A *microstate* is a single chain conformation. A *macrostate*—such as the unfolded state *U*, an intermediate state *I*, a molten globule *M*, or a transition state *T*—is some collection of individual conformations. The native state *N* is often appropriately regarded as both a microstate and a macrostate. To construct a folding algorithm requires a computational recipe that will begin with a microstate—some particular chain conformation—then evaluate its energy, then choose which specific bonds to change and by how much, in order to take a computational step to make it a more native conformation.

But experimentally obtained mass action models give only recipes for dealing with *macrostates*, such as *I* (intermediate), *D* (denatured state), *T* (transition state), etc., and not for dealing with microstates. Here are macrorecipes for how to move a chain conformation toward the native structure. From an on-pathway intermediate state, move uphill along *the reaction coordinate* in the *forward direction*. From a transition state conformation, move downhill. From an off-pathway intermediate, go back to the denatured state and try again to go forward along the reaction coordinate.

But these macrorecipes do not answer the following questions. (1) What is the macrostate to which a particular chain conformation should be assigned? Or choose a macrostate: what microscopic conformations are in it? What is the ensemble called “intermediate state,” or “denatured state,” or “transition state”? Currently, such assignments must be made arbitrarily. Macrostates are averages over many microscopic conformations; they are not descriptions of single chain conformations. (2) What series of chain conformations defines the reaction coordinate? A reaction coordinate is a macrovariable, not a microvariable (see below). For protein folding, the reaction coordinate is not known in microscopic terms. (3) Even if we knew the reaction coordinate, how do we know which way is forward? Which specific bond angles should we change to progress toward the native state? Every protein folding algorithm must make these kinds of microdecisions at every step. But no experiment yet gives such microinformation. Energy landscapes can provide the common language to bridge between micro- and macrodescriptions.

What is an energy landscape?

According to the principles of thermodynamics, if a system has n degrees of freedom $\boldsymbol{\phi} = [\phi_1, \phi_2, \ldots, \phi_n]$, the stable state of the system can be found by determining the set of values $\boldsymbol{\phi}^* = [\phi_1^*, \phi_2^*, \ldots, \phi_n^*]$ that gives the minimum value of the free energy function $F(\boldsymbol{\phi}) = F(\phi_1, \phi_2, \ldots, \phi_n)$, when explored over all possible values of $\boldsymbol{\phi}$ (see Fig. 4). Such functions $F(\boldsymbol{\phi})$ are called *energy landscapes*. Energy landscapes, *per se*, are neither new, nor controversial, nor limited to proteins. “Energy landscape” is nothing more than a name for this function. For protein folding, $\boldsymbol{\phi}$ may be the backbone and sidechain bond angles, for example.

Distinguishing between microscopics and macroscopics

The distinction between the old and new views is the distinction between an energy landscape and a reaction coordinate diagram, which, in turn, is a distinction between microscopics and macroscopics (see Fig. 5). The Sequential Micropath view postulates a simple relationship between these two types of diagrams. In the Ensemble view, the relationship can be complex, but, in general, is not known. A microstate is a single point on an energy landscape and has free energy $F_{micro} = F(\boldsymbol{\phi})$, which is also called the *internal free energy* (Dill & Chan, 1997). A macrostate has free energy $F_{macro} = F(\xi)$, where ξ is just a scalar quantity, such as a reaction coordinate or a progress variable. A given value of ξ represents some particular ensemble of microscopic conformations.

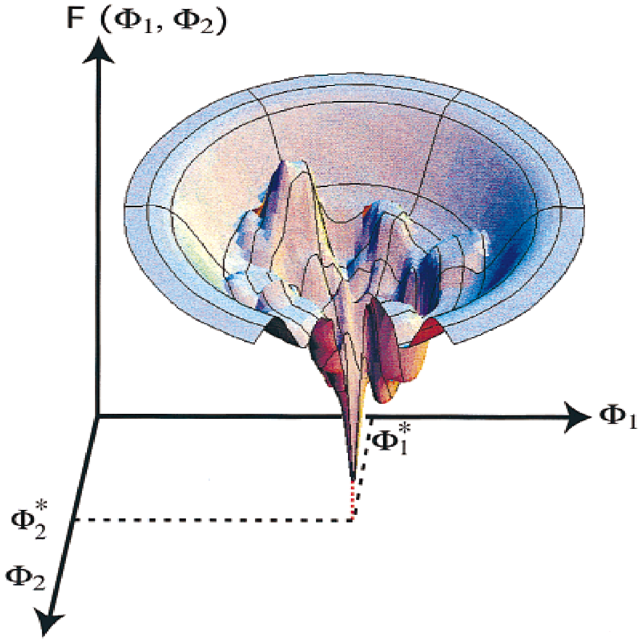

Fig. 4. Energy landscapes are free energies, $F_{micro}(\phi_1, \phi_2, \ldots)$, as a function of the degrees of freedom, $\phi_1, \phi_2, \ldots$, such as backbone and side-chain bond angles.

Figure 6 illustrates the difference between F_{micro} and F_{macro}, in a simple model. Suppose we choose as a progress variable the number of hydrophobic contacts, $\xi = 0, 1, 2, \ldots, m$, to reflect the extent of folding. This is just one of many possible progress variables; it is just chosen here for illustration because it simplifies the math. The *density of states* $g(\xi)$ is a count of the number of different microstates $\boldsymbol{\phi}$ that define a particular macrostate ξ. Figure 6 shows one of the $g(0) \approx 500{,}000$ conformations that have $\xi = 0$ hydrophobic contacts, one of the $g(4) = 67$ conformations that have $\xi = 4$ hydrophobic contacts, and the $g(6) = 1$ conformation that has $\xi = 6$ hydrophobic contacts; this is the native structure in this model.

To determine F_{micro}, focus on a particular conformation. For that conformation, sum all the energies due to bond angles, torsions, stretches, van der Waals interactions, hydrogen bonds, electrostatics, and include the solvation free energy due to the relative amounts of buried and exposed hydrophobic and polar surface. F_{micro} is a *free* energy, rather than just an energy, because it includes solvation and desolvation entropies and the hydrophobic effect. F_{micro} is not the *total* free energy, however, because it does *not* include the chain conformational entropy: it treats only a single conformation. In the HP model, in which hydrophobicity dominates, a given chain conformation has ξ hydrophobic contacts, so $F_{micro} = \xi\epsilon$, where $\epsilon < 0$ is the free energy of desolvating two nonpolar groups and bringing them into contact.

The relationship between energy landscapes and reaction diagrams is a relationship between F_{macro} and F_{micro}. F_{macro} *does* include the chain conformational entropy,

$$F_{macro}(\xi) = -kT \ln \Omega = -kT \ln[g(\xi)e^{-F_{micro}/kT}]$$

$$= F_{micro}(\xi) - kT \ln g(\xi), \tag{1}$$

where Ω is the partition function, $F_{micro}(\xi)$ is the internal free energy for each conformation that has ξ hydrophobic contacts, and $g(\xi)$ is the number of conformations having ξ hydrophobic contacts. Other progress variables can be more complex, but this simple model is sufficient for present purposes. If we express the conformational entropy of the macrostate ξ as $S_{conformational}(\xi) = \mathrm{k} \ln g(\xi)$, then Equation 1 becomes

$$F_{macro}(\xi) = F_{micro}(\xi) - TS_{conformational}(\xi), \tag{2}$$

The main point is that F_{micro} is the free energy of a single chain conformation whereas F_{macro} is the free energy of some ensemble of conformations that collectively have some macroscopic meaning, such as an intermediate, transition state, molten globule, or the denatured state. F_{macro} includes a conformational entropy ($k \ln g(\xi)$), due to the number of microscopic conformations in the particular

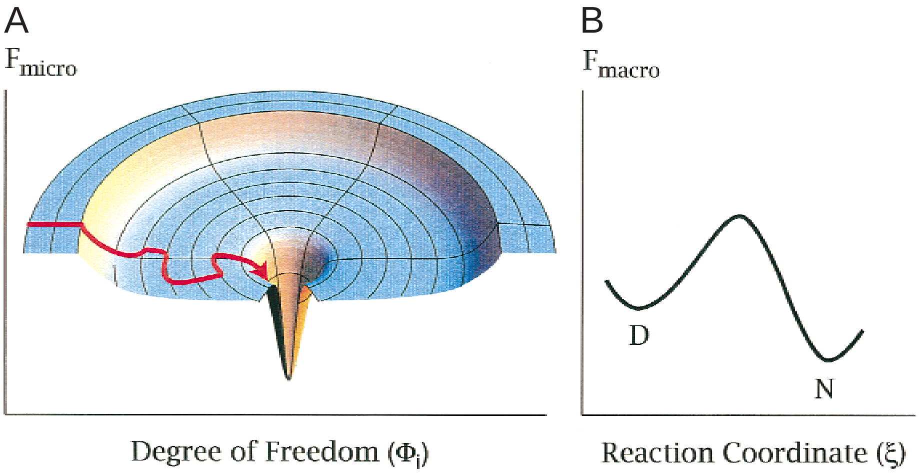

Fig. 5. (**A**) Energy landscape vs. (**B**) reaction diagram. A landscape is a free energy F_{micro} of each individual chain conformation vs. the many microscopic degrees of freedom. A reaction diagram is a free energy F_{macro} of an ensemble of molecules, and includes the chain conformational entropy. Here F_{macro} is a function of a single variable, ξ, such as a reaction coordinate. The reaction coordinate is usually not known for protein folding. The red arrow on the landscape indicates a possible micropath, an individual folding trajectory. In this case, the micropath never involves an uphill step, and yet the reaction diagram has a free energy barrier. The barrier is due to the slow entropic search of many different chains seeking the entry to the central steep funnel.

macrostate. $F_{macro}(\xi)$ is a function of a single variable ξ, and therefore it corresponds to just an ordinary two-dimensional plot, of the folding free energy vs. reaction coordinate ξ. This is the traditional reaction coordinate diagram (see Fig. 5B). In contrast, $F_{micro}(\boldsymbol{\phi})$ is the energy landscape; it is a function of many degrees of freedom. Landscapes are usually plotted in three dimensions, as a simplification, since it is impossible to draw high-dimensional surfaces. Traditional terms such as intermediate state, pathway, transition state, and free energy barrier refer to $F_{macro}(\xi)$. In contrast, computer simulations usually explore $F_{micro}(\boldsymbol{\phi})$.

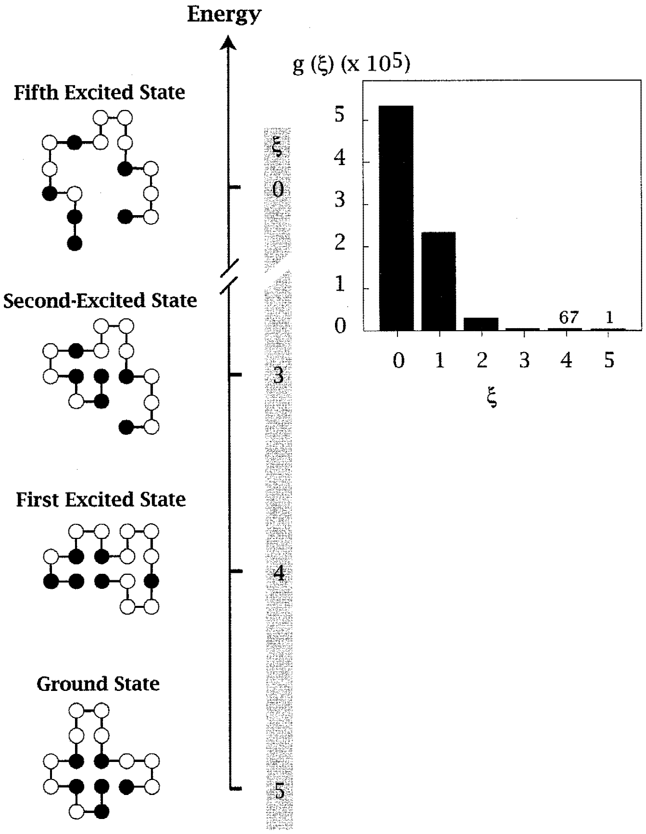

Fig. 6. The density of states $g(\xi)$ is a count of the number of chain conformations, in this case having ξ hydrophobic contacts. On the energy ladder, more hydrophobic contacts corresponds to lower energy.

What are "folding pathways"? Micropaths and microbarriers vs. macropaths and macrobarriers

The distinction between micro and macro also applies to folding kinetics. A micropath is one trajectory that one protein follows as it folds. At time t, the degrees of freedom have the value $\boldsymbol{\phi}(t)$. That is, $\boldsymbol{\phi}(0) = [\phi_1(0), \phi_2(0), \ldots, \phi_n(0)]$ at time $t = 0$, then $\boldsymbol{\phi}(t_1) = [\phi_1(t_1), \phi_2(t_1), \ldots, \phi_n(t_1)]$ at time $t = t_1$, etc. $\boldsymbol{\phi}(t)$ describes the path a fly might take in a multidimensional space. Most computer simulations have explored one or a few micropaths, although a few modeling efforts have been able to explore more complete ensemble averages (Chan & Dill, 1994, 1998). Because proteins are subject to Brownian motion, a micropath involves much motion that would seem pointless to an observer. For example, a chain can pass back and forth through a given configuration many times. In contrast, a macropath describes some progress variable $\xi(t)$, which involves different ensembles at different times during the folding or unfolding process. Experiments have given information only about macropaths, whereas simulations usually give only information about micropaths.

The key to the Sequential Micropath view is an implicit assumption of equivalence between *macropaths* and *micropaths*. The premise of the Sequential Micropath view is that there is a simple and direct relationship between $\boldsymbol{\phi}(t)$ and $\xi(t)$, just as there is in traditional chemical kinetics (see Fig. 7). If an energy landscape has an energy well corresponding to reactants A, another energy well corresponding to products B, and a lowest-energy "superhighway," which defines the route that most molecules take from A to B, then a one-dimensional reaction pathway ξ can be obtained by painting a stripe along the centerline of the superhighway through the multi-

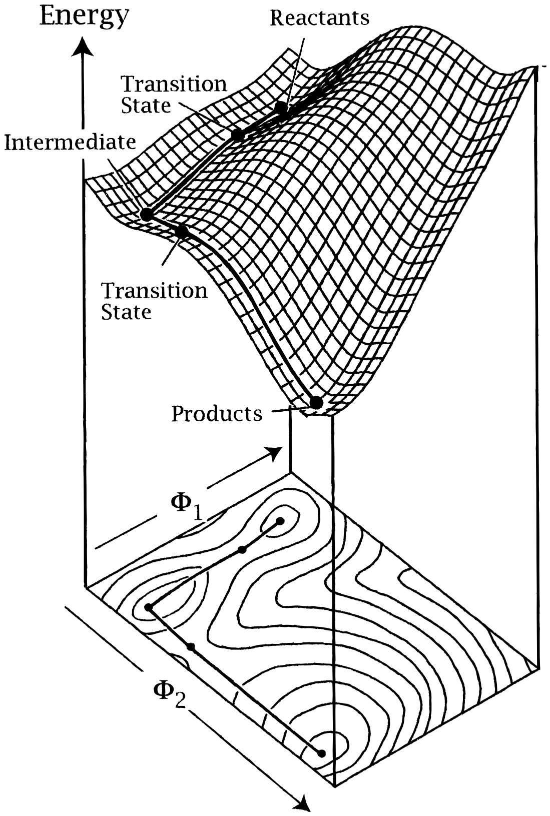

Fig. 7. Classical energy landscape for chemical reactions. Reactants, products, and intermediate states are low-energy depressions. The reaction pathway is a lowest energy highway from reactants to products. Transition states are peaks along the pathway. For simple chemical reactions, most molecules follow essentially the same reaction path.

dimensional $\boldsymbol{\phi}$ space from A to B. In some cases, particularly near the end of the folding process, this way of defining reaction pathway may be useful and adequate. But this direct relationship between $\boldsymbol{\phi}$ and ξ is valid only when molecules, like ants along a single file trail, all follow essentially the same route. That is, if one molecule folds by first forming a helix at the N-terminal end, then forming a contact between residues 1 and 27, then undoing the helix, then forming a contact between monomers 3 and 18, etc., then the equivalence of micropaths and macropaths would mean that all the other molecules will undergo exactly the same sequence of events too.

But while micropaths in chemical reaction kinetics overwhelmingly overlap with each other, the micropaths in protein folding can be very different. Chemical bonding involves energies much greater than kT, whereas each interaction in a folding process is not much larger than kT, so thermal motions can cause much larger variations in folding than in chemical reactions. One molecule may form its N-terminal helix first, while another molecule in solution, bombarded differently by Brownian motion, may form its C-terminal contacts first. In the end, both molecules will fold, but each follows a different micropath. Simulations usually show some degree of preference among microroutes, particularly in late stages of folding, but there cannot be perfect registry in the early stages of the micropaths because the starting points (the denatured conformations) are so different.

Figure 8A illustrates that for traditional chemical reactions (Fig. 7), there is a direct correspondence between the macrolevels (reaction coordinate diagram) and microlevels (energy landscape). Peaks and valleys along the reaction profile represent microscopic milestones along the energy landscape. But Figure 8B illustrates that sometimes there may be no such correspondence for some folding processes. For folding, a given experimental observation (as manifested in the reaction coordinate profile) can arise from many different landscape shapes. A landscape uniquely specifies a reaction diagram, but a reaction diagram does not uniquely specify a landscape. Figure 8B illustrates that folding becomes increasingly pathway-like at late stages, because the molecules become localized near the native state in conformational space. When chemical reactions have a single exponential time dependence, it implies an identifiable energy barrier. But for a single exponential, or any other particular time dependence in folding processes, no direct inference about microscopic bottlenecks is possible, as shown below.

Sometimes micropaths can be very different from macropaths. Figures 5 and 9 show two landscape features in which micropaths do not coincide with macropaths. (1) *A downhill micropath contributes to an uphill macropath*. A downhill micropath means that the chain does not break favorable contacts, say hydrophobic contacts. But this can involve a barrier on a reaction diagram because the microscopic meandering on flat plains on an energy diagram can be slow (Fig. 5). This will be manifested as a conformational entropy barrier (uphill) on the reaction diagram but only as a slight downhill slope on the energy landscape. (2) *A downhill macropath can include some uphill micropaths*. An uphill micropath can arise when one chain breaks favorable contacts, while most other chains find lower energy routes that avoid breaking contacts (Fig. 9). It is because we do not yet know the relationships between micropaths and macropaths that we cannot use experimental data and mass-action models (macropaths), to help us forge folding algorithms, which require knowledge of microscopic details.

Energy landscapes are funnels: The bottom is smaller than the top

While the shapes of folding energy landscapes are not yet known in detail, it is uncontestable that they are funnel like, in the sense of the term that we use here (Dill & Chan, 1997). Here, "funnel" means that many conformations have high energy and few have low energy. More specifically, conformations having high F_{micro} (denatured states) have high conformational entropy and states having low F_{micro} (native state and other deep minima) have low conformational entropy. (By some definitions, "funnel" also carries the connotation of *smooth* landscapes, so it also has implications about dynamics and time dependence, namely that barriers are small so the process happens quickly. Here, the term funnel carries no such implication about kinetics or barrier heights or smoothness or any landscape shape feature other than: there are many conformations of high free energy (F_{micro}) and few conformations of low free energy.)

Energy landscapes also have funnel-like shapes for processes of ligand binding to biomolecules: there are few tightly bound conformations, and many unbound or weakly bound conformations (Frauenfelder et al., 1991; Miller & Dill, 1997; Tsai et al., 1999).

The chicken–egg problem: Collapse first or secondary structure first?

Which comes first in the folding process, collapse or secondary structure? Just as the answer to where chickens come from is more complex than "chicken" or "egg," so also folding is undoubtedly

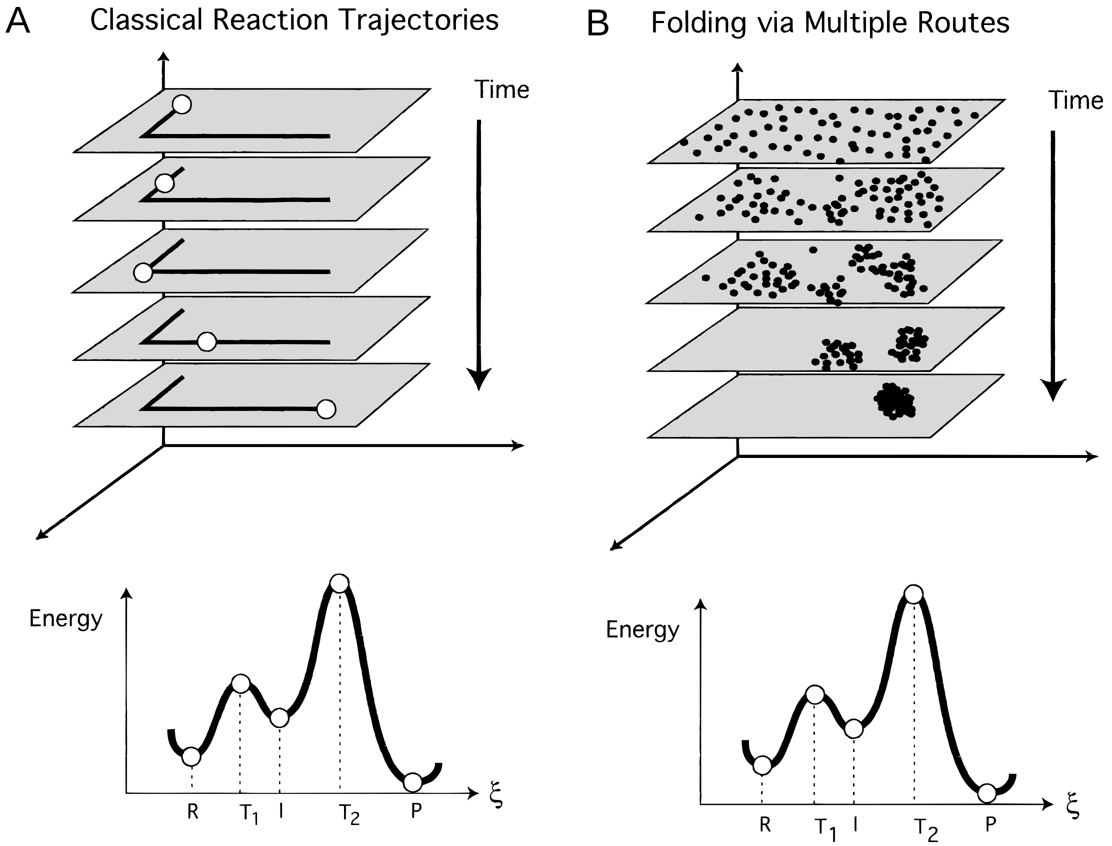

Fig. 8. A: For chemical reactions (energies $\gg kT$), the macrostates on reaction coordinate diagrams correspond to the time series of microstates on the energy landscape. **B:** For folding processes (energies per interaction $\approx kT$), the observed macrostates may not uniquely specify the time series of microstates on the energy landscape.

more complex than "collapse first" or "secondary structure first." "Collapse," "secondary structure," and "hierarchical" assembly (Baldwin & Rose, 1999a, 1999b) are macroterms, like "reaction coordinate," since each describes an ensemble property. Hierarchic folding has been recently defined "as a process in which folding begins with structures that are local in sequence and marginal in stability; these local structures interact to produce intermediates of ever-increasing complexity and grow, ultimately, into the native conformation" (Baldwin & Rose, 1999a, 1999b). It is proposed that hierarchical folding involves multiple folding routes, rather than a unique sequential pathway. By these criteria, there is little to distinguish hierarchical folding from the Ensemble view. Growing stability corresponds to a downhill flow on a landscape, and the early preference for local contacts is similar to that found in energy-based microscopic models, such as the following. The diffusion-collision model is based on assuming that $\phi\psi$ preferences are established early, then secondary structures assemble into tertiary structures (Karplus & Weaver, 1976; Lee et al., 1987; Yapa et al., 1992). A zippers model also proposes that local contacts form earlier, on average, than the nonlocal contacts. But the zippers model supposes that structure development is driven by solvation forces (Dill et al., 1993; Fiebig & Dill, 1993).

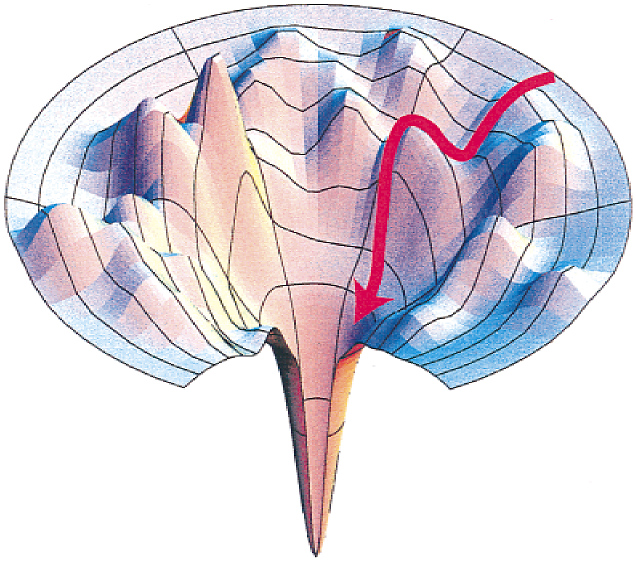

Fig. 9. An uphill micropath (red line) is surrounded by more favorable routes that do not involve uphill steps to reach the native state.

Alternatives to the sequential micropath and ensemble views?

There have been efforts to marry old and new views (Pande et al., 1998; Pande & Rokhsar, 1999). Those efforts aim to reconcile how there can be preferred folding routes at the same time that individual chains follow different micropaths. But no marriage is needed. Preferred routes and states are part and parcel of the Ensemble perspective. In my opinion, the Ensemble perspective is not one model, one result, or one energy landscape shape. It is not a denial of patterns, pathways, uniqueness, or structure. It is just a perspective based on recognizing the general funnel-like nature of the energy function $F_{micro}(\boldsymbol{\phi})$ with bumps and wiggles and shapes that have yet to be determined. The funnel perspective is universally captured in many different models, monomer sequences, potential functions, move sets, and definitions of transition states. While particular results can depend on model details, the funnel concept is a broad brush picture of how a large ensemble leads to a small ensemble, how an unstructured population changes through time to become a single structure, and how the degrees of freedom diminish from being many and uncoupled and unsynchronized to being few and coupled and synchronized. This process is bound to involve preferences.

Indeed, at the end of the folding process, it would be remarkable—and maybe impossible—to have large diversity in conformations or trajectories. Most of the simulations that have led to the Ensemble perspective have found preferred folding routes in the late stages of folding (Miller et al., 1992; Lazaridis & Karplus, 1997). What is new in the new view, and what was the essence of Levinthal's problem, was what happens in the *early stages* of folding, not the late stages. Levinthal's concern was how to search the huge space of denatured conformations. The Ensemble view merely asserts that molecules cannot be synchronized at the beginning of folding because different chains have such different unfolded conformations. Although the denatured state is a single macrostate, it is a very heterogeneous collection of microstates.

It will surely remain a matter of opinion for any given simulation whether what is interesting is the pathway or the variance from it. In either case, energy landscapes provide the basis for calculating any property of interest.

Why do we need energy landscapes?

Once energy landscapes are better understood, particularly for more realistic models of proteins, they should be able to serve several purposes. First, they should provide a consistent and rigorous language for interrelating macroscopics to microscopics. Simulators' micropaths, when properly averaged, can teach us about experimentalists' macropaths, and experimentalists' macropaths can test simulators' models. Second, landscapes provide a link between thermodynamics and kinetics, described below. And third, landscapes may provide the bridge so that folding kinetics can be brought to bear on speeding up conformational search strategies, also described below.

Relating thermodynamics and kinetics: A fluctuation–dissipation relationship for proteins?

A most remarkable result in statistical mechanics is the fluctuation–dissipation theorem (Chandler, 1987). This theorem relates a kinetic property of systems (the rate of approach to equilibrium), to an equilibrium property (the nature of the equilibrium fluctuations). Energy landscapes provide the framework for relating the thermodynamics and kinetics of protein folding. Figure 10 shows two landscapes: one is a smooth funnel, the other is a rugged funnel. For the smooth funnel, folding kinetics should be fast and two state. For the rugged landscape, folding kinetics will be slower and more complex.

The shape of the landscape also describes the fluctuations at equilibrium. Fluctuations are interesting for two reasons. First, these are the motions that are important for protein function, such as when an enzyme enters a transition state for catalyzing a reaction. Second, fluctuations can be measured by NMR or thermal factors in X-ray crystallography. The fluctuations are those conformations having energies only one or two kT higher than the native conformation and are therefore transiently populated due to occasional Brownian bombardments. If a protein has a smooth landscape, the motions of the protein are mostly small wiggles, never deviating much from the native structure because to do so would require a high energy. But for a bumpy landscape, very non-native-like conformations can occasionally be populated under native conditions because the energies of such conformations are not much higher than those of the native molecule (Miller & Dill, 1995; Tang & Dill, 1998). During those fluctuations, protons or ligands could exchange in or out, or the protein could have other transiently different properties than the native molecule. If we knew the shapes of energy landscapes, we could better understand the relationship between folding kinetics and equilibrium fluctuations around the native state.

Landscape-ology can help in developing new conformational search strategies

Knowing the shapes of energy landscapes should also help to create faster computer conformational search methods. In the Sequential Micropath view, on-pathway intermediates are held in special regard because of how they might illuminate the folding

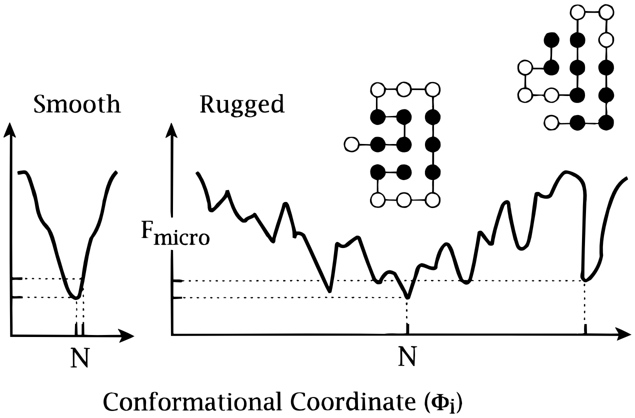

Fig. 10. Comparing fluctuations on smooth vs. rugged landscapes. The state of lowest free energy is native (N), indicated as the lower tick mark on the y-axis. Normal fluctuations increase the energy, as indicated by the higher tick mark. Thermal fluctuations lead to only small conformational deviations from the native structure on smooth landscapes, but can lead to larger deviations on rough landscapes. The native lattice conformation has six hydrophobic contacts, whereas a conformation having only one unit higher energy (five hydrophobic contacts), has a completely different conformation. Rugged landscapes mean that small excursions in energy (from native) can lead to large excursions in structure.

path. But, as noted above, computer folding algorithms have not been able to use such macro-information. The landscape view is more egalitarian. Every conformation, no matter how distant from the native state, can give some useful information *about the native state*, as described below. New and faster conformational search strategies are emerging that are based on what rudimentary knowledge is currently available of the shapes of energy landscapes.

Current search methods, such as Monte Carlo (MC), Simulated Annealing (SA), and Molecular Dynamics (MD), explore energy surfaces and are slow because they get caught in kinetic traps. We call these *local* search methods; they do not make use of global information about the shape of the underlying energy landscape. In a local search method, some very small change in a conformation is considered. Such changes are highly localized on the energy landscape. An energy is evaluated, and some decision is made whether to take that step uphill or downhill. The move is accepted or rejected, usually either based on Metropolis criteria or Newton's laws. Such strategies are very slow because they are unguided by global information, they involve much randomness, and they usually terminate in kinetic traps.

Here is an analogy. One way to find the lowest point on the Himalayan mountains is to always walk downhill until you can go down no further. Then go uphill until you can go down again. This is the Monte Carlo and SA approach. Such random walking is much slower and more haphazard than if you used a contour map to guide your journey. For example, if a protein folding algorithm creates a structure that does not have a hydrophobic core, it should not keep changing one bond angle at a time as many current methods do; it should stop wasting its time and move to a very different part of conformation space.

New methods are developing for speeding up conformational searching, based on emerging knowledge of the shapes of energy landscapes (Maranas et al., 1995; Phillips et al., 1995). For example, the idea behind the Convex Global Underestimator (CGU) method (Phillips et al., 1995; Dill et al., 1997; K.W. Foreman, A.J. Phillips, J.B. Rosen, & K.A. Dill, unpubl. comm.) is to sample a few conformations chosen randomly from the conformational space, find the nearest local energy minimum for each one, then construct a multi-dimensional parabolic "underestimator" surface $U(\boldsymbol{\phi})$ underneath all the minima that are known so far (see Fig. 11). $U(\boldsymbol{\phi})$ serves as a predictor for where the global minimum might be found, if the energy landscape is funnel-like. Subsequent underestimator surfaces are constructed iteratively for increasingly narrowed regions around the native state. In this way, *every* chain conformation that is sampled—no matter how non-native—contributes some information about the landscape shape, and contributes to an estimate of where the native state will be found. In contrast, local search methods make no such use of collective information about all other conformations that have been sampled before a given step.

The CGU and other underestimator methods look promising, on the following bases. (1) Starting from different initial starting points on the landscape, the CGU usually reaches the same final point, indicating that it finds global minima and does not get stuck in kinetic traps. (2) An advantage of the CGU over MC and SA is that no problem-dependent adjustment is required, as when devising temperature schedules or proper move sets. (3) Tests so far in a simple protein folding model and on van der Waals clusters up to 21 atoms shows that CGU reaches much lower on energy landscapes in a given time than MC or SA (see Fig. 12), and the advantage increases with chain length (K.W. Foreman, A.J. Phillips, J.B. Rosen, & K.A. Dill, unpubl. comm.). The only knowledge the CGU currently uses is just that landscapes are funnel-like. As we learn more about the shapes of protein energy landscapes, it should be possible to create faster conformational search strategies.

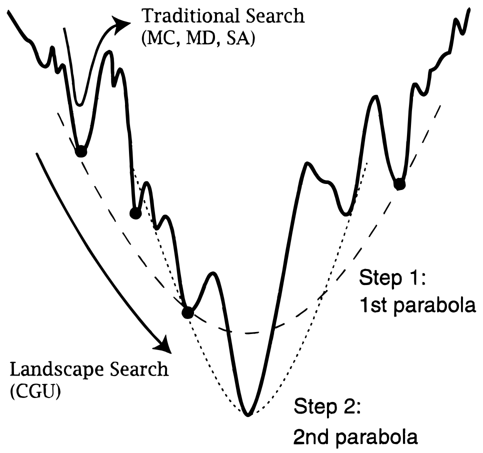

Fig. 11. Convex global underestimator (CGU) conformational search strategy. Traditional methods, such as Monte Carlo (MC), molecular dynamics (MD), and simulated annealing (SA), search over the tops of energy landscapes and can get caught in kinetic traps. The CGU searches underneath the landscape instead by using a few sampled local minima (indicated by dots) to generate a series of underestimating parabolic surfaces to locate the global minimum (Dill et al., 1997).

Historical parallels with polymer science?

For the past 40 years, a defining paradigm of protein science has been Structural Biology. Structural Biology has provided a framework for deciding what questions are important and how to answer them. Two key imperatives of Structural Biology are (1) *high resolution*, the importance of atomic detail and (2) *unique architectures*, the importance of specific geometric interrelationships among atoms. Protein structures have atomic resolution, and every

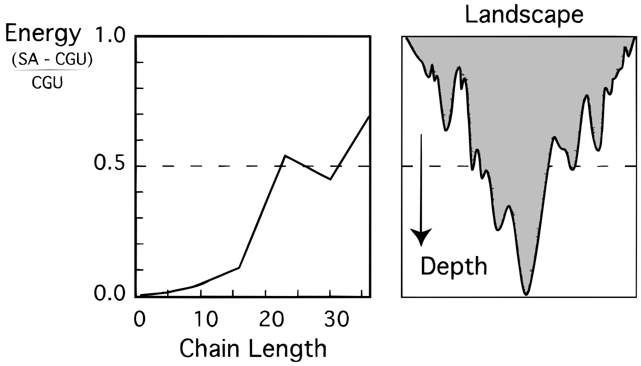

Fig. 12. Relative search depth (energy) of simulated annealing compared to the CGU, for different lengths of model proteins, up to 36 amino acids (K.W. Foreman et al., in prep.). For short chains, SA reaches the same depth on energy landscapes as the CGU, but for longer chains, SA gets stuck at increasingly higher altitudes on the energy landscape, where the relative depth is indicated by the cartoon on the right.

atom has its place. It has been considered important to know THE native structure, THE transition state, or THE intermediate structure.

Of course, it is clear that proteins wiggle and move; they are not perfectly static (Karplus, 1997). But even so, such fluctuations are often regarded as a sort of footnote to the main message, much like error bars in experimental data. According to this logic, folding pathways are less like a perfect train track, where no lateral variation is allowed, and more like a highway, where some small degree of weaving and lateral meandering can take place.

But in the polymer view, statistics can play a fundamentally different and deeper role. It is more like replacing a train track, not with a highway, but with a ski bowl. Driving on a highway from point A to point B can be described by average velocities, positions, and altitudes along the "reaction coordinate," the highway. But tracking an ensemble of skiers is quite a different business than tracking the flow of cars on a highway. Skiers can take different routes. The average position of skiers on a mountainside is a much more heterogeneous property, with less apparent meaning. What is THE structure, or even THE averaged structure, at any given time, is not yet clear, or necessarily always meaningful.

The importance of one particular structure and the neglect of statistics has a parallel in the history of polymer science. The breakthrough that founded polymer science was the macromolecular hypothesis, the idea that there were long chains covalently linked together (Flory, 1953; Morawetz, 1985). The huge resistance to this idea prior to the 1920s was due to a faith in the importance of specific structures and a reluctance to fully appreciate the statistics. According to Flory (1953):

> "Organic chemists were motivated by the desire to devise concise formulas and to isolate pure substances, the term *pure* . . . invariably implying a formula of convenient size. Hence the quest for *the* cellulose molecule or *the* rubber molecule continued. . . . By the turn of the century this objective had crystallized to a discipline which dominated synthetic organic chemistry. To be eligible for acceptance in the chemical kingdom, a newly created substance . . . had to be separated in such a state that it could be characterized by a molecular formula. The investigator was obliged to adduce elementary analyses to confirm the composition, and to supplement these with molecular weight determinations for the purpose of showing that the substance was neither more nor less complex than the formula proposed. Otherwise the fruits of his labors would not be elevated to an honored place in the immortal pages of the chemical compendiums. The successes of synthetic organic chemistry in creating the hundreds of thousands of different combinations and permutations of atoms must not be discounted. In magnitude of creative achievement, they are scarcely surpassed in any other field of science. While this discipline was strikingly successful, it also tended to narrow the outlook of contemporary researchers. They came to believe that every definable substance could be classified in terms of a single definite molecule capable of being represented by a concise formula."

With the macromolecular hypothesis came the recognition that *N* polymer molecules in solution, even when they are called by the same name, such as polyethylene, are not identical to each other. Each molecule in solution can have a different conformation and even, for synthetic polymers, a different chain length. Hence different experiments see different ensemble averages and give different perspectives on the same "molecule," polyethylene. Within only about a decade after the macromolecular hypothesis was accepted, quantitative statistical mechanical models began to successfully explain rubber elasticity, the viscosities and viscoelasticities of chain molecule liquids, the dependence of the physical properties of polymeric materials on molecular weight distributions, and the unusual thermodynamics of polymer solutions. Such statistical ideas now provide the foundation of modern polymer theory. For many properties of proteins, too, it seems clear that statistics is not just a caveat about small details but is at the very heart of the problems that proteins must solve.

Conclusions

Statistical mechanical models can give useful insights about proteins. While all-atom models sacrifice conformational sampling to gain atomic detail, statistical mechanical models do the reverse. Because simple models explore non-native states so effectively, have few parameters, and cost little computer time, they have been useful for exploring folding forces and principles. They have led to the perspective that the folding code is primarily a solvation code, rather than a local propensities code. Statistical mechanical models are well suited to addressing combinatoric problems, such as the Levinthal and Blind Watchmaker paradoxes. The conclusion is that we should beware of needle in a haystack arguments, because nature does not seem to work that way. Each step is not unguided. Conformational and sequence spaces are more like landscapes. Landscapes are funnel like, wide at the top and narrow at the bottom, sometimes with hills and valleys. All conformations—not just on-pathway intermediates for example—can give some guidance toward the global minimum. New computational search methods are drawing on this information to make better folding and docking algorithms. The energy landscape perspective may help connect the currently disjoint areas of kinetics experiments and conformational search strategies.

Acknowledgments

I thank Sarina Bromberg and Jack Schonbrun for very helpful comments. I also thank Hue Sun Chan for his insightful comments on this manuscript, for his many contributions over the past decade that are summarized in this work, and for teaching me a great deal about proteins. I appreciate support from the Lawrence Berkeley Labs and from NIH grant GM34993.

References

Alberts B, Bray D, Johnson A, Raff M, Roberts K, Walter P. 1998. *Essential cell biology: An introduction to the molecular biology of the cell*. New York: Garland.

Anfinsen CB, Scheraga HA. 1975. Experimental and theoretical aspects of protein folding. *Adv Prot Chem 29*:205–300.

Aurora R, Creamer TP, Srinivasan R, Rose GD. 1997. Local interactions in protein folding: Lessons from the α-helix. *J Biol Chem 272*:1413–1416.

Baldwin RL. 1994. Protein folding: Matching speed and stability. *Nature 369*:183–184.

Baldwin RL. 1995. The nature of protein folding pathways: The classic versus the new view. *J Biomol NMR 5*:103–109.

Baldwin RL, Rose GD. 1999a. Is protein folding hierarchic? I. Local structure and peptide folding. *Trends Biochem Sci 24*:26–33.

Baldwin RL, Rose GD. 1999b. Is protein folding hierarchic? II. Folding intermediates and transition states. *Trends Biochem Sci 24*:77–83.

Ballew RM, Sabelko J, Gruebele M. 1996a. Direct observation of fast protein folding: The initial collapse of apomyoglobin. *Proc Natl Acad Sci USA 93*:5759–5764.

Ballew RM, Sabelko J, Gruebele M. 1996b. Observation of distinct nanosecond and microsecond protein folding events. *Nature Struct Biol 3*:923–926.

Bowie J, Reidhaar-Olson J, Lim WA, Sauer RT. 1990. Deciphering the messages in protein sequences: Tolerance to amino acid substitutions. *Science 247*:1306–1310.

Branden C, Tooze J. 1999. *Introduction to protein structure*, 2nd ed. New York: Garland.

Bromberg S, Dill KA. 1994. Side chain entropy and packing in proteins. *Protein Sci 3*:997–1009.

Bryngelson J, Onuchic J, Socci ND, Wolynes PG. 1995. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins 21*:167–195.

Bryngelson J, Wolynes PG. 1987. Spin-glass and the statistical mechanics of protein folding. *Proc Natl Acad Sci USA 84*:7524–7528.

Burton RE, Huang GS, Daugherty MA, Calderone TL, Oas TG. 1997. The energy landscape of a fast-folding protein mapped by ala → gly substitutions. *Nature Struct Biol 4*:305–310.

Callendar RH, Gilmanshin R, Dyer RB, Woodruff WH. 1999. *Annu Rev Phys Chem*. Forthcoming.

Camacho CJ, Thirumalai D. 1993. Kinetics and thermodynamics of folding in model proteins. *Proc Natl Acad Sci USA 90*:6369–6372.

Chan C-K, Hu Y, Takahashi S, Rousseau DL, Eaton WA, Hofrichter J. 1997. Submillisecond protein folding kinetics studied by ultrarapid mixing. *Proc Natl Acad Sci USA 94*:1779–1784.

Chan HS, Dill KA. 1991. "Sequence space soup" of protein and copolymers. *J Chem Phys 95*:3775–3787.

Chan HS, Dill KA. 1994. Transition states and folding dynamics of proteins and heteropolymers. *J Chem Phys 100*:9238–9257.

Chan HS, Dill KA. 1998. Protein folding in the landscape perspective: Chevron plots and non-Arrhenius kinetics. *Proteins 30*:2–33.

Chandler D. 1987. *Introduction to modern statistical mechanics*. New York: Oxford Press.

Dawkins R. 1996. *The blind watchmaker: Why the evidence of evolution reveals a universe without design*. New York: Norton.

Dill KA. 1985. Theory for the folding and stability of globular proteins. *Biochemistry 24*:1501–1509.

Dill KA. 1987. The stabilities of globular proteins. In: Oxender DL, Fox CF, eds. *Protein Engineering*. Alan R. Liss Inc. pp 187–192.

Dill KA. 1993. Folding proteins: Finding a needle in a haystack. *Curr Op Struct Biol 3*:99–103.

Dill KA, Bromberg S, Yue K, Fiebig KM, Yee DP, Thomas PD, Chan HS. 1995. Principles of protein folding: A perspective from simple exact models. *Protein Sci 4*:561–602.

Dill KA, Chan HS. 1997. From Levinthal to pathways to funnels: The "new view" of protein folding kinetics. *Nature Struct Biol 4*:10–19.

Dill KA, Fiebig KM, Chan HS. 1993. Cooperativity in protein-folding kinetics. *Proc Natl Acad Sci USA 90*:1942–1946.

Dill KA, Phillips, AT, Rosen JB. 1997. Protein structure and energy landscape dependence on sequence using a continuous energy function. *J Comp Biol 4*:227–239.

Doty P, Bradbury JH, Holtzer AM. 1956. The molecular weight, configuration and association of poly-γ-benzyl-L-glutamate in various solvents. *J Am Chem Soc 78*:947–954.

Doty P, Yang JT. 1956. Polypeptides VII. Poly-gamma-benzyl-L-glutamate: The helix-coil transition in solution. *J Am Chem Soc 78*:498.

Eisenberg D, Weiss RM, Terwilliger C. 1984. The hydrophobic moment detects periodicity in protein hydrophobicity. *Proc Natl Acad Sci USA 81*:140–144.

Fiebig KM, Dill KA. 1993. Protein core assembly processes. *J Chem Phys 98*:3475–3487.

Flory PJ. 1953. *Principles of polymer chemistry*. Ithaca, New York: Cornell University Press. pp 8–19.

Frauenfelder H, Sligar SG, Wolynes PG. 1991. The energy landscapes and motions of proteins. *Science 254*:1598–1603.

Gassner NC, Baase WA, Matthews BW. 1992. A test of the jigsaw puzzle model for protein folding by multiple methionine substitutions within the core of T4 lysozyme. *Proc Natl Acad Sci USA 93*:12155–12158.

Gilmanshin R, Callendar RH, Dyer RB. 1998. Fast events in protein folding: The time evolution of primary processes. *Nature Struct Biol 5*:363–365.

Gilmanshin R, Williams S, Callendar RH, Woodruff WH, Dyer RB. 1997a. Fast events in protein folding: Relaxation dynamics of secondary and tertiary structure in native apomyoglobin. *Proc Natl Acad Sci USA 94*:3709–3713.

Gilmanshin R, Williams S, Callendar RH, Woodruff WH, Dyer RB. 1997b. Fast events in protein folding: Relaxation dynamics and structure of the I form of apomyoglobin. *Biochemistry 36*:15006–15012.

Hamada D, Segawa S, Goto Y. 1996. Non-native alpha-helical intermediate in the refolding of beta-lactoglobulin: A predominantly beta-sheet protein. *Nature Struct Biol 3*:868–873.

Harrison SC, Durbin R. 1985. Is there a single pathway for the folding of a polypeptide chain? *Proc Natl Acad Sci USA 82*:4028–4030.

Honig B, Cohen FE. 1996. Adding backbone to protein folding: Why proteins are polypeptides. *Folding Design 1*:R17–R20.

Huang GS, Oas TG. 1995. Submillisecond folding of monomeric λ repressor. *Proc Natl Acad Sci USA 92*:6878–6882.

Kabsch W, Sander C. 1984. On the use of sequence homologies to predict protein structure: Identical pentapeptides can have completely different conformations. *Proc Natl Acad Sci USA 81*:1075–1078.

Kamtekar S, Schiffer JM, Xiong H, Babik JM, Hecht MH. 1993. Protein design by binary patterning of polar and nonpolar amino acids. *Science 262*:1680–1685.

Karplus M. 1997. The Levinthal paradox: Yesterday and today. *Folding Design 2*:S69–S75.

Karplus M, Weaver DL. 1976. Protein-folding dynamics. *Nature 260*:404–406.

Kauzmann W. 1959. Some factors in the interpretation of protein denaturation. *Adv Prot Chem 14*:1–63.

Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. 1958. A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature 181*:662–666.

Kim PS, Baldwin RL. 1982. Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. *Annu Rev Biochem 51*:459–489.

Kuroda Y, Hamada D, Tanaka T, Goto Y. 1996. High helicity of peptide fragments corresponding to beta-strand regions of beta-lactoglobulin observed by 2D-NMR spectroscopy. *Folding Design 4*:255–263.

Lau KF, Dill KA. 1990. Theory for protein mutability and biogenesis. *Proc Natl Acad Sci USA 87*:638–642.

Lazar GA, Desjarlais JR, Handel TM. 1997. De novo design of the hydrophobic core of ubiquitin. *Protein Sci 6*:1167–1178.

Lazaridis T, Karplus M. 1997. "New view" of protein folding reconciled with the old through multiple unfolding simulations. *Science 278*:1928–1931.

Lee S, Bashford D, Karplus M, Weaver DL. 1987. Brownian dynamics simulation of protein folding: A study of the diffusion-collision model. *Biopolymers 26*:481–509.

Levinthal C. 1968. Are there pathways for protein folding? *J Chim Phys 65*: 44–45.

Lim WA, Farrugio DC, Sauer RT. 1992. Structural and energetic consequences of disruptive mutations in a protein core. *Biochemistry 31*:4324.

Lim WA, Sauer RT. 1991. The role of internal packing interactions in determining the structure and stability of a protein. *J Mol Biol 219*:359–376.

Lipman DJ, Wilbur WJ. 1991. Modeling neutral and selective evolution of protein folding. *Proc Roy Soc B245*:7–11.

Maranas DC, Androulakis IP, Floudas CA. 1995. A deterministic global optimization approach for the protein folding problem. *DIMACS series in Discrete Mathematics and Theoretical Computer Science 23*:133–150.

Matthews BW. 1993. Structural and genetic analysis of protein stability. *Annu Rev Biochem 62*:139–160.

Miller DW, Dill KA. 1995. A statistical mechanical model for hydrogen exchange in globular proteins. *Protein Sci 4*:1860–1873.

Miller DW, Dill KA. 1997. Ligand binding to proteins: The binding landscape model. *Protein Sci 6*:2166–2179.

Miller R, Danko CA, Fasolka MJ, Balacz AC, Chan HS, Dill KA. 1992. Folding kinetics of proteins and copolymers. *J Chem Phys 96*:768–780.

Minor DL, Kim PS. 1994. Context is a major determinant of beta-sheet propensity. *Nature 371*:264–267.

Mirsky AE, Pauling L. 1936. On the structure of native, denatured, and coagulated proteins. *Proc Natl Acad Sci USA 22*:439–447.

Morawetz H. 1985. *Polymers: The origins and growth of a science*. New York: Wiley.

Munson M, Balasubramanian S, Fleming KG, Nagi AD, et al. 1996. What makes a protein a protein: Hydrophobic core designs that specify stability and structural properties. *Protein Sci 5*:1584–1593.

Munson M, O'Brien R, Sturtevant JM, Regan L. 1994. Redesigning the hydrophobic core of a four-helix-bundle-protein. *Protein Sci 3*:2105–2022.

Onuchic JN, Luthy-Schulten Z, Wolynes PG. 1997. Theory of protein folding: The energy landscape perspective. *Annu Rev Phys Chem 48*:545–600.

Pande VS, Grosberg AY, Tanaka T, Rokhsar DS. 1998. Pathways for protein folding: Is a new view needed? *Curr Op Struc Biol 8*:68–79.

Pande VS, Rokhsar DS. 1999. Folding pathway of a lattice model for proteins. *Proc Natl Acad Sci USA 96*:1273–1278.

Pascher T, Chesick JP, Winkler JR, Gray HB. 1996. Protein folding triggered by electron transfer. *Science 271*:1558–1560.

Pauling L, Corey RB. 1951a. Atomic coordinates and structure factors for two helical configurations of polypeptide chains. *Proc Natl Acad Sci USA 37*:235–240.

Pauling L, Corey RB. 1951b. The pleated sheet, a new layer configuration of polypeptide chains. *Proc Natl Acad Sci USA 37*:251–256.

Pauling L, Corey RB. 1951c. The structure of fibrous proteins of the collagen-gelatin group. *Proc Natl Acad Sci USA 37*:272–281.

Pauling L, Corey RB. 1951d. Configurations of polypeptide chains with favored orientations around single bonds: Two new pleated sheets. *Proc Natl Acad Sci USA 37*:729–740.

Pauling L, Corey RB, Branson HR. 1951. The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci USA 37*:205–211.

Phillips AT, Rosen JB, Walke VH. 1995. Molecular structure determination by convex global underestimation. *DIMACS series in Discrete Mathematics and Theoretical Computer Science 23*:181.

Poland DC, Scheraga HA. 1970. *Theory of the helix-coil transition*. New York: Academic Press.

Predki PF, Agrawal V, Brunger AT, Regan L. 1996. Amino-acid substitutions in a surface turn modulate protein stability. *Nature Struct Biol 3*:54–58.

Ramachandra Shastry MC, Roder H. 1998. Evidence for barrier-limited protein folding kinetics on the microsecond time scale. *Nature Struct Biol 5*:385–392.

Ramachandra Shastry MC, Saunder JM, Roder H. 1998. Kinetic and structural analysis of submillisecond folding events in cytochrome *c*. *Acc Chem Res 31*:717–725.

Reidhaar-Olson JF, Sauer RT. 1988. Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. *Science 241*:53–57.

Riddle DS, Santiago JV, BrayHall ST, Doshi N, Grantcharova VP, Yi Q, Baker D. 1997. Functional rapidly folding proteins from simplified amino acid sequences. *Nature Struct Biol 4*:805–809.

Roy S, Ratnaswamy G, Boice JA, Fairman D, McLendon G, Hecht MH. 1997. A protein designed by binary patterning of polar and nonpolar amino acids displays native-like properties. *J Am Chem Soc 116*:5302–5306.

Schafmeister CE, LaPorte SL, Miercke LJW, Stroud RM. 1997. A designed four helix bundle protein with native-like structure. *Nature Struct Biol 4*:1039–1046.

Schellman JA. 1958. The factors affecting the stability of hydrogen-bonded polypeptide structures in solution. *J Chem Phys 62*:1485.

Scholtz JM, Baldwin RL. 1992. The mechanism of α-helix formation by peptides. *Annu Rev Biophys Biomol Struct 21*:95–118.

Shiraki K, Nishikawa K, Goto Y. 1995. Trifluoroethanol-induced stabilization of the α-helical structure of β-lectoglobulin: Implication for non-hierarchical protein folding. *J Mol Biol 245*:180–194.

Smith CK, Regan L. 1995. Guidelines for protein design: The energetics of beta-sheet side chain interactions. *Science 270*:980–982.

Stanley HE. 1971. *Introduction to phase transitions and critical phenomena*, Oxford: Oxford Press.

Tang KES, Dill KA. 1998. Native protein fluctuations: The conformational-motion temperature and the inverse correlation of protein flexibility with protein stability. *J Biomol Struct Dyn 16*:397–411.

Thomas PD, Dill KA. 1993. Local and nonlocal interactions in globular proteins and mechanisms of alcohol denaturation. *Protein Sci 2*:2050–2065.

Tsai C-J, Kumar S, Ma B, Nussinov R. 1999. Folding funnels, binding funnels, and protein function. *Protein Sci 8*:1179–1188.

Wetlaufer DB. 1973. Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc Natl Acad Sci USA 70*:697–701.

Williams S, Causgrove TP, Gilmanshin R, Fang KS, Callendar RH, Woodruff WH, Dyer RB. 1996. Fast events in protein folding: Helix melting and formation in a small peptide. *Biochemistry 35*:691–697.

Wu LC, Kim PS. 1997. Hydrophobic sequence minimization of the α-lactalbumin molten globule. *Proc Natl Acad Sci USA 94*:14314–14319.

Yapa K, Weaver DL, Karplus M. 1992. β-sheet coil transitions in a simple polypeptide model. *Proteins 12*:237–265.

Yue K, Dill KA. 1995. Forces of tertiary structural organization in globular proteins. *Proc Natl Acad Sci USA 92*:146–150.

Zimm BH, Bragg J. 1959. Theory of the phase transition between helix and random coil in polypeptide chains. *J Chem Phys 31*:526.

Zwanzig R, Szabo A, Bagchi B. 1992. Levinthal's paradox. *Proc Natl Acad Sci USA 89*:20–22.