Cooperativity in protein-folding kinetics

KEN A. DILL, KLAUS M. FIEBIG, AND HUE SUN CHAN

Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143-1204

ABSTRACT How does a protein find its native state without a globally exhaustive search? We propose the "HZ" (hydrophobic zipper) hypothesis: hydrophobic contacts act as constraints that bring other contacts into spatial proximity, which then further constrain and zip up the next contacts, etc. In contrast to helix–coil cooperativity, HZ–heteropolymer collapse cooperativity is driven by nonlocal interactions, causes sheet and irregular conformations in addition to helices, leads to secondary structures concurrently with early hydrophobic core formation, is much more sequence dependent than helix–coil processes, and involves compact intermediate states that have much secondary—but little tertiary—structure. Hydrophobic contacts in the 1992 Protein Data Bank have the type of "topological localness" predicted by the hypothesis. The HZ paths for amino acid sequences that mimic crambin and bovine pancreatic trypsin inhibitor are quickly found by computer; the best configurations thus reached have single hydrophobic cores that are within about 3 kcal/mol of the global minimum. This hypothesis shows how proteins could find globally optimal states without exhaustive search.

What is the origin of cooperativity in protein folding? Some proteins fold reversibly (1, 2), independently of pathway (3–5), to a unique native state. These native states must be at the global minimum of free energy that is accessible on the experimental time scale. To guarantee that a computational strategy finds the global optimum requires a search time that scales exponentially with chain length (6, 7). However, Levinthal (8) and Wetlaufer (9) pointed out that proteins fold much too fast (by at least tens of orders of magnitude) to involve an exhaustive search. Hence, the kinetics/thermodynamics paradox: how can a protein find a globally optimal state without a globally exhaustive search? It follows that the folding problem is not so much a problem of exhaustive computational searching as it is the question of what is the physical basis of cooperativity by which proteins avoid exhaustive searching of conformational space. This view has led to important experiments on protein-folding mechanisms and pathways (10–23).

The main paradigm for cooperativity in biopolymers is the helix–coil theory (24–27). "Cooperativity" describes how a globally optimal state can be found without a global search, hence cooperativity involves conformational choices that must be local in some sense. The nature of "localness" is at the heart of cooperativity. For helix–coil processes, the global minimum (the helix) is found by local processes by which each individual tetrapeptide in the sequence finds a hydrogen-bonded helical conformation; the partition function for the whole chain is a product of partition functions for individual helical units (24–27). Such processes are local in two respects. (i) Monomers i and j that seek to form a helical hydrogen-bonding contact are "sequence-local" (S-local)—i.e., near neighbors in the sequence ($j = i + 3$ for α-helices,
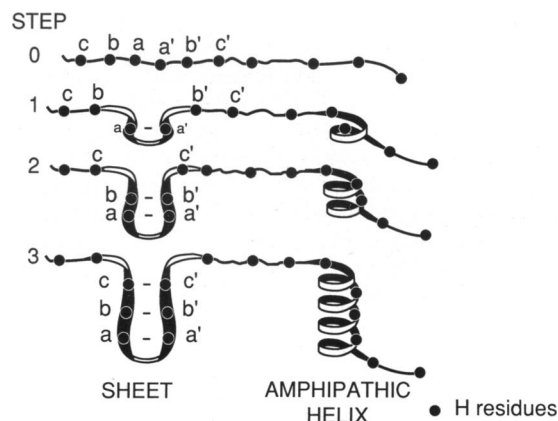

FIG. 1. HZ model of protein-folding pathways. The closest hydrophobic (H) residues (solid dots) in sequence pair together first, e.g., a and a' in step 0. They constrain the chain and bring other H monomers, such as the (b,b') pair, into spatial proximity. Now (b,b') further constrains the chain and brings the (c,c') pair into spatial proximity, etc. As H contacts form and develop a core, helices and sheets zip up if they have appropriate H sequences.

for example). (ii) Because of their sequence proximity, the conformational search for i to find j is small.

But protein-folding cooperativity must be different from helix–coil cooperativity because of the (i) driving forces and (ii) structures involved. (i) Protein folding is driven by nonlocal interactions to form a compact hydrophobic core (H core); these are different from the local interactions that form helices (28, 29). (ii) Globular proteins have regular secondary structures, such as helices and parallel and antiparallel sheets, irregular conformations, and various forms of tertiary organization, depending on the amino acid sequence. Helix–coil cooperativity leads only to helices, relatively independently of the sequence. The main obstacle to understanding cooperativity in globular proteins is how the chain searches the prohibitively large number of possible nonlocal contacts to assemble a H core.

Hydrophobic zipper (HZ) model of protein cooperativity

The HZ model is the idea, defined in Fig. 1, that H monomers pair together, one after another, like the zipping of a zipper. At time $t = 0$, folding conditions are "turned on" by removing denaturant for example. Any two H monomers near enough in the sequence (a and a' in Fig. 1) can readily find each other through a small conformational search, with little loss of conformational entropy. [H is used here to refer to the favorable contact interactions, mainly among nonpolar monomers, including polar and hydrogen-bond contributions.] Contact between a and a' now serves as a constraint, forcing other H monomers (b and b') into spatial proximity. Contact (b,b') then forms, with only a small additional

---

Abbreviations: H, hydrophobic; HP, H–polar; BPTI, bovine pancreatic trypsin inhibitor; HZ, hydrophobic zipper; S-local, sequence local; T-local, topologically local.

search, and it further constrains the chain, bringing c and c' into spatial proximity. Then contact (c,c') forms, etc. Contact (c,c') would be unfavorable at earlier times because it would involve a large conformational search in the absence of the (b,b') constraint. Contact (c,c') is nonlocal, since monomers c and c' are not near neighbors in the sequence, but it becomes local in a topological or conditional sense (T-local) in the presence of the constraint (b,b') (30). Fig. 1 shows how helices and sheets develop as HH contacts accrete in the process of chain collapse. For any particular HP sequence (H = hydrophobic, P = polar), chain folding may nucleate at different sites and follow different zipper paths in parallel, which converge to form a core.

The aim of the present paper is to explore the biochemical consequences of the HZ model; the details of the HZ path construction are given elsewhere (30). Briefly the recipe for determining an HZ path for an HP sequence is as follows. First form a contact between any $(i, i + 3)$ HH monomer pair. Then form any HH contact that either involves another $(i, i + 3)$ HH pair or an HH pair caused to be T-local by a preceding contact. Continue this process until no further HH contacts are possible. At each step, this process involves choices among alternatives; we choose randomly. The result is one plausible sequence of events, which we call a T-local string, leading to a compact state. We then construct another such sequence of events, making another random set of choices. A T-local string is a process of assembly, not a process of physical kinetics, because this scheme specifies no reverse rate or exit from local minima (H.S.C. and K.A.D., unpublished data) or temperature dependence, etc. But it is our hypothesis that folding pathways follow the same average order of events as T-local strings (30).

The HZ model has its basis in three observations. (*i*) Exhaustive simulations and path-integral theory show that internal constraints in polymers force some monomer pairs into spatial proximity, reducing the conformational entropy cost for forming helices, sheets, and other structures (7, 29, 32–34). (*ii*) For some HP sequences, copolymer chains collapse to unique native structures with single H cores (35–38). (*iii*) HP copolymers collapse via specific pathways, according to lattice-model Monte Carlo kinetics studies (39). The steps along these folding paths often involve minimal configurational entropy loss (30).

Predictions of the HZ model

In this section, we consider several aspects of how HZ–heteropolymer collapse cooperativity differs from helix–coil cooperativity and how the model bears on protein-folding puzzles and experiments. Then we discuss similarities with prior work.

Nonlocal Driving Force. Whereas helix–coil processes are driven by local interactions among near neighbors in the sequence, HZ collapse is driven mainly by contact interactions—i.e., among monomers that can be distant in the sequence but that are brought together by earlier events in the folding process. This conforms with evidence that HH interactions are important determinants of structure (40–43), stability (28), and early-nucleation folding events (44, 45) of globular proteins. Whereas the sequence determinant of the folding pathway is presumed to be mainly the nonpolar interactions, we note that the most T-local contacts invariably also involve hydrogen bonds and are the elementary building blocks of secondary structures.

The "Chicken-and-Egg Problem": Which Comes First, the H Core or Secondary Structure? Fig. 1 shows that the cascade of HH contacts that cause chain collapse also force the development of secondary structure, for appropriate HP sequences. Both helices and sheets form early and concurrently along HZ paths, consistent with experiments showing

both H clustering and secondary structures appearing within a few milliseconds (18–23). The observation that H monomers commonly repeat periodically in globular protein sequences, about every 3.6 monomers in amphipathic helices and 2.3 monomers in sheets (46–51), is consistent with a hydrophobic driving force for secondary-structure formation.

Initiation Is S-Local. The first steps of HZ chain collapse involve near-neighbor HH contacts. Local interactions, including helical (24, 52, 53) and turn (54) propensities and hydrogen bonds (55), will contribute to the energetic favorability of these nucleating conformations. Hence, propensities for peptides to be helical in solution [autonomous folding units (56, 57)] should correlate with helices in globular states (58); and yet, since the folding patterns are dominated by the nonlocal hydrophobic interactions, the helical propensities can be overridden (59, 60).

HP Sequence Encodes the Folding Pathway. Folding rates and paths are strongly dependent on sequence. We have studied rates of heteropolymer collapse to unique native states by Monte Carlo dynamics in the two-dimensional HP sequence lattice model (39). Fig. 2 shows more extensive results, for all the singly degenerate HP sequences with a chain length of $n = 13$ monomers. To ensure native states are stable, an HH attraction energy of $\varepsilon = -9.3kT$ was used. To explore the long folding times required, we have used a transition matrix method (31). Fig. 2 shows that the rate of folding varies widely from one sequence to another. Fig. 2 also shows that the shape of the conformational landscape depends on the amino acid sequence of the chain (see also refs. 61 and 62). We characterize the landscape by a quantity, the "native-state protectedness," which is defined as the sum:

$$\sum_{d>0} g(h_N - 1, d)e^{-\varepsilon d/kT}, \qquad [1]$$

where $g(h, d)$ is the number of conformations with h HH contacts that must break d HH contacts (i.e., $-\varepsilon d$ is the energetic barrier height) to reach the native state. The number of native HH contacts is h_N, so $h_N - 1$ represents conformations of the lowest nonnative minima. The protectedness measures the "bumpiness" of the conformational space—i.e., the difficulty of overcoming energetic barriers to escape kinetic trap conformations and reach the global minimum. Fig. 2 shows that slow folders have conformational spaces with many deep nonnative minima, whereas fast folders have conformational spaces with fewer and shallower nonnative minima. Slow folders have conformational spaces with ups and downs like the Himalayan mountains. Fast
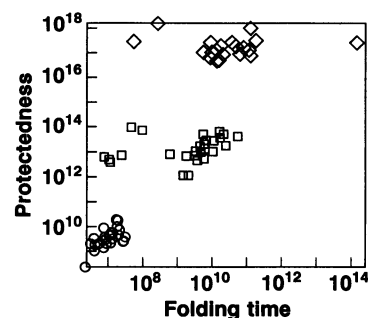
FIG. 2. "Protectedness" (see Eq. 1) of the native state vs. folding time (in units of Monte Carlo steps) for all $n = 13$ sequences on the two-dimensional HP model in Monte Carlo kinetics simulations (31, 39). The symbols $\bigcirc$, $\square$, and $\diamond$ represent sequences with $d_{max} = 2, 3,$ and 4, respectively, where d_{max} is the number of HH contacts a chain needs to break to escape from its deepest local minima. Different sequences fold at different rates. Fast folders have low protectedness.

folders have conformational spaces more like ski slopes, with mostly downhills toward the native state.

For Some HP Sequences There Is No Kinetics/Thermodynamics (Levinthal) Paradox. HZ paths are not exhaustive searches. Fig. 3 shows that if contact (k,l) is given, then contact (i,j) is more probable than other HH contacts, since (i,j) involves the least loss of conformational entropy. Monomer i largely makes no attempt to pair with more distant H monomers. HZ paths are cooperative: the probability of contact C_2 given C_1 is greater than the unconditional probability of contact C_2 alone. Hence, the cooperativity in heteropolymer collapse originates in the conformational entropy of the increasingly constrained chain.

Nevertheless, even though HZ paths are fast and nonexhaustive, short-chain lattice studies show that about 70% of all the HP sequences that have a single globally optimal native state can fold to it by following HZ paths (30). For these sequences, there is no paradox, because HZ paths can lead to global minima.

However, Some HP Sequences Do Not Fold to Global Minima by HZ Paths. This is according to the same lattice model studies (30) and implies (*i*) that such sequences might require additional mechanisms, such as diffusion–collision (63), to fold; or (*ii*) that if proteins follow HZ paths, then some sequences may not fold to thermodynamically stable states; or (*iii*) both of the above. Examples of proteins with high kinetic barriers include α-lytic protease (64) and plasminogen activator inhibitor (65, 66).

Many Paths or Few Paths? The HZ model shows that the "many-path" (jigsaw puzzle) (67) and "few-path" (sequential assembly) (11, 68) viewpoints are not necessarily inconsistent with each other. HZ paths encompass a large fraction of all possible conformations but a small fraction of all possible topologies (30). At the first step, when an $(i,i + 3)$ HH contact forms, the rest of the chain is irrelevant and, therefore, can have very broad conformational diversity. This diversity diminishes as the chain funnels (62, 69) to fewer compact states. Hence the chain conformations follow many paths, whereas the chain topologies (i.e., contact sets) follow only a few. For example, for a chain with m hydrophobic monomers, the first step is a choice between $m(m - 1)/2$ different HH pairings, but the HZ model chooses from fewer than $m - 1$ of them [i.e., only those with $(i,i + 3)$ spacing]. Most experiments monitor topologies; secondary structures, disulfide bonds, or excimer distances "see" only some contacts, while the rest of the chain can be conformationally diverse. The experimental observation that there are relatively few pathways (11, 12, 16, 20, 70) reflects the small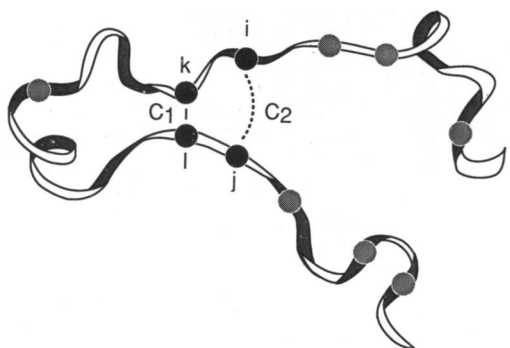 topological diversity. Nevertheless, since HZ processes usually involve more than one topological path and nucleation site, a protein could find the same native state (*i*) regardless of whether it comes off a ribosome or refolds from solution or (*ii*) regardless of some mutations (70).

Some Sequences with High HP Periodicity Should Zip up to Nonglobular Structures. Such structures are leucine zippers, coiled coils, and β-sheets (71–73).

HH Contacts in Proteins Form T-Local Strings. If proteins fold by HZ paths, then their native structures should reflect their HZ path histories. HH contacts in proteins should appear as T-local strings. We find this is the case. We study 52 proteins* in the Protein Data Bank (74, 75) with chain lengths between 70 and 200 amino acids. HH contacts are defined as pairs of alanine, isoleucine, leucine, methionine, phenylalanine, tryptophan, tyrosine, valine, and cysteine residues whose α-carbon coordinates are <7 Å apart. Fig. 4 shows (*i*) the contact order (i.e., loop length, reflecting the localness or nonlocalness of contacts along the sequence; see refs. 32 and 33), indicating that about half of all HH contacts are among monomers distant in the sequence, and (*ii*) the "effective contact order" (30), which measures the degree to which contacts are T-local. The leftmost peak in Fig. 4*B* indicates that most HH contacts in proteins can be formed via T-local strings.

Essentially the same results are obtained if we use a 6.5-Å neighbor criterion or β-carbon coordinates. The shape of this plot is not specific to native proteins; in other tests, we find that any compact chain with sufficiently high nonpolar composition will have an effective-contact-order plot resembling this. For example, by using the same native conformations but with scrambled sequences, there are far fewer total HH contacts but otherwise the same distribution shape. The shape of this plot is not an identifier of native proteins; the main point is that HZ processes can assemble many different compact conformations for chains that have sufficiently high H composition.

Most of the individual predictions of the HZ hypothesis are not new. Earlier work (44, 45, 63, 76–88) concludes that a folding chain nucleates where hydrophobic residues are local in the sequence (44, 45) and proceeds by hierarchical condensation (77, 78) or by "crumpling" (83, 84) or by "on-site" construction (85–88). The present model differs in several respects. First, it provides a specific recipe for computing the folding pathway from the amino acid sequence. Second, it is a demonstration that such processes can find conformations at the global minimum of free energy for a broad range of sequences. Third, it shows that protein-like kinetics could arise from only nonlocal interactions, through a mechanism that is very different from classical helix–coil cooperativity. Most closely related to the present hypothesis is the on-site construction mechanism of Sikorski and Skolnick and their colleagues (85–88), which is also a zipping process specified by the amino acid sequence. But their model has a large driving force due to local interactions (89). It is well-established (24) that zipping and cooperativity can arise from local interactions. The present hypothesis is a rather different model of cooperativity. The present work simply suggests that a dominant factor in protein-like kinetics may be just solvent-aversion and therefore might be observed in simpler polymers.

FIG. 3. Origin of cooperativity. Given constraint C_1 (k,l), monomer i explores mainly the possibility of pairing with j, not with all other H monomers; thus, this search is nonexhaustive. It is cooperative because the probability of forming contact $C_2 = (i,j)$ is much higher if C_1 is formed than in the absence of C_1 because the conformational entropy is higher for the latter.

Biophysics: Dill *et al.*

Proc. Natl. Acad. Sci. USA 90 (1993) 1945

FIG. 4. Distributions of HH contacts in proteins in the Protein Data Base; "percent" refers to the fraction of contacts. (*A*) "Contact order" = $|i - j|$ for contact (i,j). About half the hydrophobic contacts are in helices and turns, half in nonlocal contacts. (*B*) "Effective contact order" is equal to the maximal spatial separation of residues i and j that can be achieved in the presence of other existing constraints; for details see ref. 30. HH contacts in proteins are T-local (i.e., have small effective contact order).

HZ as a protein-folding strategy

A successful protein-folding algorithm will ultimately require two components: a folding strategy (for searching conformational space) and accurate potential functions. Any recipe for cooperativity is a protein-folding strategy. Here, our aim is to ask whether the HZ strategy can construct an optimal hydrophobic core for a long chain of amino acids in reasonable computer time. But for long chains, how can we know how close we are to the global optimum? Analytical theory can estimate a tight upper bound on the maximum number of HH contacts for any HP sequence on simple lattices (K. Yue and K.A.D., unpublished data). Therefore, the HP lattice model allows us to ask how close (in energy) the HZ search strategy can come to the global optimum of the HP potential, even though the HP lattice model does not correctly predict structures. We compute HZ paths for HP sequences resembling bovine pancreatic trypsin inhibitor (BPTI) and crambin (Fig. 5), and we study the end-state conformations of these paths. From this simulation we learn about search efficiency and not about accuracy of representing real protein structures.

We find that the search strategy is fast. One HZ path for model crambin or BPTI can be generated in 3–6 min on a Sun-4 workstation. Fig. 6 shows the distribution of numbers of HH contacts obtained from 13,853 and 13,217 HZ end-states for model crambin and BPTI, respectively, on the three-dimensional simple cubic lattice. The "best" end-states (i.e., ones with a maximum number of HH contacts) from these paths have 32 HH contacts for crambin and 40 HH contacts for BPTI. Each "best" conformation has a single hydrophobic core and many S-local and S-nonlocal contacts. The conformations corresponding to the peaks of the two distributions look much like compact denatured states: they are compact (but not native), they have many HH contacts (sometimes as two cores or an elongated core), and they have

Crambin (46 Residues):
PPHHHPHHHPPPPHPHHPHHPPHHPHHHHHPPHHHHHHPHPHHPPHHP

BPTI (58 Residues):
PHPHHHPHHHPPHHPHPHHPHHHPHPHPHHPPHHHPPHPHPPPPHPPHPPHHPPHPPH

FIG. 5. HP sequences for model BPTI and crambin.

FIG. 6. HZ end states for folding HP model of BPTI (*Upper*) and crambin (*Lower*) on three-dimensional cubic lattices, showing the distribution of the number of HH contacts.

many helices and turns. Two hydrophobic cores in compact denatured states have been observed by small-angle x-ray scattering (90) and may be a general feature of compact nonnative states.

The best structures found by the HZ strategy have energies close to native. For the crambin and BPTI sequences given above, an upper bound on the maximum number of HH contacts is found to be 35 and 43, respectively (90); thus, the HZ search strategy comes to within 3 HH contacts [about 1.5 kcal/mol (31)] of the native structure in the present simulations. Therefore, we believe the HZ search strategy is a promising approach for reaching quickly and deeply into regions of conformational space to find compact and native-like structures.

Conclusions

We present the HZ hypothesis of protein-folding cooperativity. It constructs folding pathways from amino acid sequences. The cooperativity in heteropolymer collapse and protein folding is different than for helix–coil processes. Helix–coil cooperativity is driven by interactions that are local in the sequence (S-local) and leads only to helical structures, relatively independently of the amino acid sequence. HZ cooperativity for heteropolymer collapse is driven by nonlocal (but T-local) interactions mediated by the solvent. It leads to helical, sheet, and irregular structures, depending on the amino acid sequence. The folding process is postulated to be a fast nonexhaustive search, in which, for appropriate sequences, the early stages of H core formation are concurrent with formation of helices and sheets. The folding pathways involve formation of HH contacts that are brought into close spatial proximity by virtue of preceding contact constraints. Using short-chain lattice heteropolymer models, we find that HZ folding pathways lead to the

conformations of global minimum in free energy for many—but not all—HP sequences.

1. Anfinsen, C. B. (1973) *Science* **181**, 223–230.
2. Privalov, P. L. (1979) *Adv. Protein. Chem.* **33**, 167–241.
3. Garel, J.-R., Nall, B. T. & Baldwin, R. L. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 1853–1857.
4. Kato, S., Okamura, M., Shimamoto, N. & Utiyama, H. (1981) *Biochemistry* **20**, 1080–1085.
5. Denton, J. B., Koniski, Y. & Scheraga, H. A. (1982) *Biochemistry* **21**, 5155–5163.
6. Dill, K. A. (1985) *Biochemistry* **24**, 1501–1509.
7. Chan, H. S. & Dill, K. A. (1989) *Macromolecules* **22**, 4559–4573.
8. Levinthal, C. (1968) *J. Chim. Phys.* **65**, 44–45.
9. Wetlaufer, D. B. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 697–701.
10. Ptitsyn, O. B. & Semisotnov, G. V. (1991) in *Conformations and Forces in Protein Folding*, eds. Nall, B. T. & Dill, K. A. (Am. Assoc. Adv. Sci., Washington, DC), pp. 155–168.
11. Kim, P. S. & Baldwin, R. L. (1982) *Annu. Rev. Biochem.* **51**, 459–489.
12. Kim, P. S. & Baldwin, R. L. (1990) *Annu. Rev. Biochem.* **59**, 631–660.
13. Baum, J., Dobson, C. M., Evans, P. A. & Hanley, C. (1989) *Biochemistry* **28**, 7–13.
14. Roder, H., Elove, G. A. & Englander, S. W. (1988) *Nature (London)* **335**, 700–704.
15. Udgaonkar, J. B. & Baldwin, R. L. (1988) *Nature (London)* **335**, 694–699.
16. Creighton, T. E. (1990) *Biochem. J.* **270**, 1–16.
17. Weissman, J. S. & Kim, P. S. (1991) *Science* **253**, 1386–1393.
18. Briggs, M. S. & Roder, H. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 2017–2021.
19. Matouschek, A., Serrano, L. & Fersht, A. R. (1992) *J. Mol. Biol.* **224**, 819–835.
20. Serrano, L., Matouschek, A. & Fersht, A. R. (1992) *J. Mol. Biol.* **224**, 847–859.
21. Semisotnov, G. V., Rodionora, N. A., Kutyskenko, V. P., Ebert, B., Blanck, J. & Ptitsyn, O. B. (1987) *FEBS Lett.* **224**, 9–13.
22. Gilmanshin, R. I. & Ptitsyn, O. B. (1987) *FEBS Lett.* **223**, 327–329.
23. Kuwajima, K., Garvey, E. P., Finn, B. E., Matthews, C. R. & Sugai, S. (1991) *Biochemistry* **30**, 7693–7703.
24. Zimm, B. H. & Bragg, J. K. (1959) *J. Chem. Phys.* **31**, 526–535.
25. Schellman, J. A. (1958) *J. Phys. Chem.* **62**, 1485–1494.
26. Poland, D. C. & Scheraga, H. A. (1970) *Theory of the Helix-Coil Transition in Biopolymers, Statistical Mechanical Theory of Order-Disorder Transitions in Biological Macromolecules*, (Academic, New York).
27. Flory, P. J. (1969) *Statistical Mechanics of Chain Molecules*, (Interscience, New York).
28. Dill, K. A. (1990) *Biochemistry* **29**, 7133–7155.
29. Chan, H. S. & Dill, K. A. (1991) *Annu. Rev. Biophys. Biophys. Chem.* **20**, 447–490.
30. Fiebig, K. M. & Dill, K. A. (1993) *J. Chem. Phys.*, in press.
31. Dill, K. A., Alonso, D. O. & Hutchinson, K. (1989) *Biochemistry* **28**, 5439–5449.
32. Chan, H. S. & Dill, K. A. (1989) *J. Chem. Phys.* **90**, 492–509.
33. Chan, H. S. & Dill, K. A. (1990) *J. Chem. Phys.* **92**, 3118–3135.
34. Chan, H. S. & Dill, K. A. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6388–6392.
35. Lau, K. F. & Dill, K. A. (1989) *Macromolecules* **22**, 3986–3997.
36. Lau, K. F. & Dill, K. A. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 638–642.
37. Chan, H. S. & Dill, K. A. (1991) *J. Chem. Phys.* **95**, 3775–3787.
38. Shortle, D., Chan, H. S. & Dill, K. A. (1992) *Protein Sci.* **1**, 201–215.
39. Miller, R., Danko, C. A., Fasolka, M. J., Balazs, A. C., Chan, H. S. & Dill, K. A. (1991) *J. Chem. Phys.* **96**, 768–780.
40. Lim, W. A. & Sauer, R. T. (1989) *Nature (London)* **339**, 31–36.
41. Bowie, J. U., Reidhaar-Olson, J. F., Lim, W. A. & Sauer, R. T. (1990) *Science* **247**, 1306–1310.
42. Sweet, R. M. & Eisenberg, D. (1983) *J. Mol. Biol.* **171**, 479–488.
43. Bashford, D., Chothia, C. & Lesk, A. M. (1987) *J. Mol. Biol.* **196**, 199–216.
44. Matheson, R. R., Jr., & Scheraga, H. A. (1978) *Macromolecules* **11**, 819–829.
45. Moult, J. & Unger, R. (1991) *Biochemistry* **30**, 3816–3824.
46. Schiffer, M. & Edmundson, A. B. (1967) *Biophys. J.* **7**, 121–135.
47. Eisenberg, D., Weiss, R. M. & Terwilliger, T. C. (1982) *Nature (London)* **299**, 371–374.
48. Eisenberg, D., Weiss, R. M. & Terwilliger, T. C. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 140–144.
49. Cornette, J. L., Cease, K. B., Margalit, H., Sponge, J. L., Berzofsky, J. A. & DeLisi, C. (1987) *J. Mol. Biol.* **195**, 659–685.
50. Finer-Moore, J., Rubin, B. & Stroud, R. L. (1989) in *Prediction of Protein Structure and the Principles of Protein Conformation*, eds. Fasman, G. D. (Plenum, New York), p. 719.
51. Hazes, B. & Hol, W. G. J. (1992) *Proteins* **12**, 278–298.
52. Baldwin, R. L. (1989) *Trends Biochem. Sci.* **14**, 291–294.
53. Scholtz, J. M. & Baldwin, R. L. (1992) *Annu. Rev. Biophys. Biomol. Struct.* **21**, 95–118.
54. Wright, P. E., Dyson, H. J. & Lerner, R. A. (1988) *Biochemistry* **27**, 7167–7175.
55. Stickle, D. F., Presta, L. G., Dill, K. A. & Rose, G. D. (1992) *J. Mol. Biol.* **226**, 1143–1159.
56. Yu, M. H. & King, J. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6584–6588.
57. Feng, M.-F. & Englander, S. W. (1991) *J. Mol. Biol.* **221**, 1045–1061.
58. Lyu, P. C., Liff, M. I., Marky, L. A. & Kallenbach, N. R. (1990) *Science* **250**, 669–673.
59. Kabsch, W. & Sander, C. (1983) *Biopolymers* **22**, 2577–2637.
60. Argos, P. (1987) *J. Mol. Biol.* **197**, 331–348.
61. Shakhnovich, E., Farztdinov, G., Gutin, A. M. & Karplus, M. (1991) *Phys. Rev. Lett.* **67**, 1665–1668.
62. Leopold, P. E., Montal, M. & Onuchic, J. N. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 8721–8725.
63. Karplus, M. & Weaver, D. L. (1976) *Nature (London)* **260**, 404–406.
64. Baker, D., Sohl, J. L. & Agard, D. A. (1992) *Nature (London)* **356**, 263–265.
65. Carrell, R. W., Evans, D. L. & Stein, P. E. (1991) *Nature (London)* **353**, 576–578.
66. Banzon, J. A. & Kelly, J. W. (1992) *Protein Eng.* **5**, 113–115.
67. Harrison, S. C. & Durbin, R. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 4028–4030.
68. Ptitsyn, O. B. (1987) *J. Protein Chem.* **6**, 273.
69. Dill, K. A. (1986) in *Protein Engineering*, eds. Oxender, D. L. & Fox, C. F. (Liss, New York), p. 187.
70. Serrano, L., Matouschek, A. & Fersht, A. R. (1992) *J. Mol. Biol.* **224**, 805–818.
71. Karle, I. L., Flippen-Anderson, J. C., Sukumar, M. & Balaram, P. (1992) *Proteins* **12**, 324–330.
72. Cohen, C. E. & Parry, D. A. (1990) *Proteins* **7**, 1–15.
73. Rasmussen, R., Benvegnu, D., O'Shea, E. K., Kim, P. S. & Alber, T. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 561–564.
74. Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F. & Weng, J. (1978) in *Crystallographic Databases—Information Content, Software Systems, Scientific Applications*, eds. Allen, F. H., Bergerhoff, G. & Seivers, R. (Data Commission Int. Union of Crystallogr., Bonn, F.R.G.), pp. 107–132.
75. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) *J. Mol. Biol.* **112**, 535–542.
76. Ptitsyn, O. B. & Rashin, A. A. (1975) *Biophys. Chem.* **3**, 1–20.
77. Crippen, G. (1978) *J. Mol. Biol.* **126**, 315–332.
78. Rose, G. (1979) *J. Mol. Biol.* **134**, 447–470.
79. Cohen, F. E., Sternberg, M. J. E., Phillips, D. C., Kuntz, I. D. & Kollman, P. A. (1980) *Nature (London)* **286**, 632–634.
80. Lesk, A. & Rose, G. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 4304–4308.
81. Richardson, J. (1981) *Adv. Protein. Chem.* **34**, 167–339.
82. Fetrow, J. S., Zehfus, M. H. & Rose, G. D. (1988) *Bio/Technology* **6**, 167–171.
83. de Gennes, P. G. (1985) *J. Phys. Lett.* **46**, L-639.
84. Grosberg, A. Y., Nechaev, S. K. & Shakhnovich, E. I. (1988) *J. Phys. (Paris)* **49**, 2095–2100.
85. Skolnick, J. & Kolinski, A. (1990) *J. Mol. Biol.* **212**, 787–817.
86. Sikorski, A. & Skolnick, J. (1990) *J. Mol. Biol.* **215**, 183–198.
87. Skolnick, J., Kolinski, A. & Yaris, R. (1989) *Biopolymers* **28**, 1059–1095.
88. Sikorski, A. & Skolnick, J. (1989) *Biopolymers* **28**, 1097–1113.
89. Skolnick, J. & Kolinski, A. (1990) *Science* **250**, 1121–1125.
90. Flanagan, J. M., Kataoka, M., Shortle, D. & Engelman, D. M. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 748–752.