Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 4195–4199, May 1991
Biochemistry

The protein-folding problem: The native fold determines packing, but does packing determine the native fold?

(protein packing/protein engineering)

MICHAEL J. BEHE*†, EATON E. LATTMAN‡, AND GEORGE D. ROSE*§

*Department of Biological Chemistry, Pennsylvania State University, College of Medicine, Hershey, PA 17033; and ‡Department of Biophysics and Biophysical Chemistry, The Johns Hopkins University, School of Medicine, Baltimore, MD 21205

Communicated by Bruno H. Zimm, January 31, 1991 (received for review January 3, 1991)

ABSTRACT **A globular protein adopts its native three-dimensional structure spontaneously under physiological conditions. This structure is specified by a stereochemical code embedded within the amino acid sequence of that protein. Elucidation of this code is a major, unsolved challenge, known as the protein-folding problem. A critical aspect of the code is thought to involve molecular packing. Globular proteins have high packing densities, a consequence of the fact that residue side chains within the molecular interior fit together with an exquisite complementarity, like pieces of a three-dimensional jigsaw puzzle [Richards, F. M. (1977) *Annu. Rev. Biophys. Bioeng.* 6, 151]. Such packing interactions are widely viewed as the principal determinant of the native structure. To test this view, we analyzed proteins of known structure for the presence of preferred interactions, reasoning that if side-chain complementarity is an important source of structural specificity, then sets of residues that interact favorably should be apparent. Our analysis leads to the surprising conclusion that high packing densities—so characteristic of globular proteins—are readily attainable among clusters of the naturally occurring hydrophobic amino acid residues. It is anticipated that this realization will simplify approaches to the protein-folding problem.**

It is well-known that a protein molecule will adopt its native three-dimensional structure spontaneously under normal physiological conditions (1). The transition to the native state from a denatured state is called protein folding. Despite intense research, a generalized mechanistic understanding of the folding transition remains obscure. This important question is called the protein-folding problem.

A key question—perhaps *the* key question—is the extent to which protein conformation is determined by packing interactions within the hydrophobic core. This question has its origins in the seminal work of Kauzmann (2), who used model compounds to argue that the burial of hydrophobic groups serves as a primary source of stabilization energy in folded proteins. Later, Richards showed that these buried groups are as well packed, on average, as crystals of small organic molecules, with packing densities more reminiscent of solids than of oil (3, 4). The inside of a typical protein contains side chains that fit together with a striking complementarity, like pieces of a three-dimensional jigsaw puzzle.

The high packing densities seen in globular proteins are an experimental fact (3–6). This fact has been interpreted to mean that protein conformation is linked tightly to internal packing. According to this interpretation, for example, lysozyme does not have the same folded conformation as ribonuclease, although both proteins have approximately the same size and composition, because the lysozyme sequence cannot achieve efficient internal packing when organized into a ribonuclease fold. Such an interpretation of packing is consistent with classical studies of protein evolution, where the most conserved residues are found in the buried interior (7).

Paradoxically, recent mutational studies of proteins have demonstrated that the hydrophobic core can tolerate a broad diversity of residue substitutions, usually with only minor effects on structure, stability, and function (8–12). Such results have prompted us to conduct a simple test of whether protein conformation is determined primarily by packing interactions. In this test we analyzed 67 proteins of known structure for preferred interactions, reasoning that if side-chain complementarity is an important source of conformational discrimination, then sets of residues that interact favorably should be readily apparent. Equivalently, if no interactions are found to be especially favorable, then efficient packing—an undeniable experimental fact—is achieved without severe limitation of the individually allowed side-chain orientations (13, 14); in which case it follows that packing and conformation are *not* tightly linked. In the ensuing analysis, conspicuous side-chain complementarity is distinguished from the broad, nonrandom distributions observed in earlier studies (15, 16).

Our analysis of packing is conducted in two parts. First, all residue pairs, x–y, are assayed for the existence of particularly favorable interactions between x and y (i.e., binary interactions). The existence of a covalent bond between pairs of half-cystine residues renders them a conspicuous outlier among pair-wise interactions and serves as a natural control for this stage of the analysis. It is conceivable that preferred higher level packing arrangements (i.e., tertiary, quaternary, . . .) can exist despite an absence of preferred binary interactions. To address this second possibility, the interaction of all x–y pairs with the remainder of the protein is assessed. In this latter stage, a lumped interaction term is computed for each x–y pair and its complementary protein-binding pocket. In other words, the first step measures the affinity within all pairs of residues, whereas the second step measures the affinity between all residue pairs and their microenvironment within the protein. In this experiment, the area buried between or among residues is used to evaluate affinity.

The following analysis leads to the surprising conclusion that high packing densities—so characteristic of globular proteins—are readily attainable among clusters of the naturally occurring apolar residues. If true, then packing of the hydrophobic core is not the principal factor that discriminates between native and nonnative conformations.

Experimental Design. The crystal structures of 67 x-ray-elucidated proteins were analyzed, and all pair-wise interactions were identified for the 190 possible nonglycine pair-wise combinations of residues. Proteins used and their Brookhaven file names in parentheses (17) are as follows: cy-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

†Permanent address: Department of Chemistry, Lehigh University, Mountaintop Campus, Bethlehem, PA 18015.
§To whom reprint requests should be addressed.

tochrome b_{562} (156B), cytochrome c_{550} (155C), cytochrome c_{551} (351C), L-arabinose-binding protein (1ABP), actinidin (2ACT), alcohol dehydrogenase (4ADH), adenyl kinase (3ADK), α-lytic protease (2ALP), penecillopepsin (2APP), rhizopus acid protease (2APR), azurin (1AZA), azurin (1AZU), cytochrome b_5 (2B5C), bacteriochlorphyll-A protein (3BCL), bovine phospholipase A_2 (1BP2), cytochrome c_2 (3C2C), carbonic anhydrase B (2CAB), carbonic anhydrase C (1CAC), cytochrome c_3 (2CDV), concanavalin A (2CNA), carboxypeptidase A (5CPA), carp Ca^{2+}-binding protein (1CPV), crambin (1CRN), ribosomal protein (1CTF), α cobratoxin (1CTX), cytochrome *c* (3CYT), dihydrofolate reductase (4DFR), erythrocruorin (1ECD), elastase (3EST), ferredoxin (4FD1), ferredoxin (3FXC), flavodoxin (4FXN), γ chymotrypsin A (2GCH), glutathione peroxidase (1GP1), glutathione reductase (3GRS), high potential iron protein (1HIP), hemerythrin (1HMQ), α-amylase inhibitor (1HOE), insulin (1INS), lactate dehydrogenase (5LDH), leghemoglobin (1LH1), lamprey hemoglobin (2LHB), lysozyme (1LZ1), T_4 phage lysozyme (2LZM), sperm whale myoglobin (1MBN), horse hemoglobin (2MHB), snake neurotoxin (1NXB), ovomucoid third domain (2OVO), prealbumin (2PAB), papain (9PAP), plastocyanin (1PCY), phosphoglycerate mutase (3PGM), avian pancreatic polypeptide (1PPT), pancreatic trypsin inhibitor (5PTI), Bence–Jones immunoglobulin (1REI), rhodanese (1RHD), ribonuclease A (5RSA), rubredoxin (3RXN), subtilisin (1SBT), *Streptomyces griseus* protease (2SGA), scorpion neurotoxin (1SN3), staphylococcal nuclease (2SNS), Cu, Zn superoxide dismutase (2SOD), *Streptomyces subtilisin* inhibitor (2SSI), thermolysin (3TLN), troponin C (4TNC), and β-trypsin (1TPP).

In this study, a pair-wise interaction is said to occur between two residues whenever they bury at least 1 Å^2 of side-chain surface area between them. For each protein in the data base, all potential pairs were evaluated systematically, and those satisfying the definition were selected.

Solvent-accessible surface areas *A* were calculated with an algorithm developed by Lee and Richards (18). Atomic radii used, in Å, were as follows: tetrahedral C = 2.0, trigonal C = 1.7, carbonyl O = 1.4, hydroxyl O = 1.6, carboxyl O = 1.5, tetrahedral N = 2.0, trigonal N = 1.7, divalent S = 1.85, and sulfhydryl S = 2.0. The probe radius was 1.4 Å.

The area buried by an *x*–*y* pair is denoted $[A^b_{xy}]$ and is reckoned as the difference between the sum of the solvent-accessible areas for each residue individually (i.e., $A_x + A_y$) and the solvent-accessible area for the pair (i.e., A_{xy}). When evaluating a pair, only side-chain atoms were included in the calculation; backbone atoms of the two residues, as well as the remainder of the protein, were excluded. The standard-state area of an *x*–*y* pair, A^0_{xy}, is taken as the sum of the individual standard states, $A^0_x + A^0_y$. For a residue *x*, the individual standard-state area is given as the mean accessibility of an ensemble of Gly-Xaa-Gly tripeptides with dihedral angles taken from the observed distribution in proteins (19).

The area buried between each interacting pair was calculated and used to obtain the mean area buried by pair-wise type. The mean area buried between residue type *x* and residue type *y*, $\langle A^b_{xy}\rangle$, is simply $\Sigma^n_{k=1}[A^b_{xy}]_k/n$, where $[A^b_{xy}]_k$ is the area buried in the *k*th *x*–*y* pair, and *n* is the total number of *x*–*y* pairs.

Area Buried in Pair-Wise Interactions. Fig. 1 is a plot of the mean area buried by each pair versus the total area of that pair. Residues are subdivided into three types: hydrophobic (alanine, cysteine, isoleucine, leucine, methionine, phenylalanine, tryptophan, valine); polar (asparagine, glutamine, serine, threonine, tyrosine); and charged (asparagine, aspartic acid, glutamic acid, and lysine). These categories are cross-compared in Fig. 1. Pairs containing proline or histidine exhibit excessive scatter and have been excluded from the figures. The atypical behavior of proline and histidine can be rationalized: Histidine exists in both protonated and deprotonated forms. For steric reasons proline, though apolar, is typically situated in peptide-chain turns on the surface of a protein (20). Thus, both histidine and proline span two of the three categories used here.

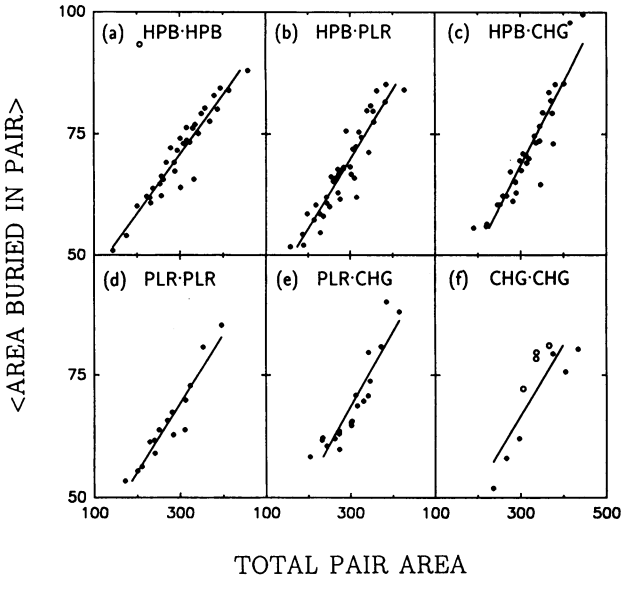

FIG. 1. Plot of the mean area buried within a pair, $\langle A^b_{xy}\rangle$, versus the total area of the pair. Total pair-wise area A^0_{xy} is taken as the sum of individual standard-state areas, $A^0_x + A^0_y$ (19). Areas are in Å^2. Each point represents a separate pair-wise category. Residues are subdivided into three classes—hydrophobic (HPB); polar, uncharged (PLR); and charged (CHG)—which are cross compared in (*a*–*f*). Respective least-squares lines, given by Eq. **1a**–**f**, are shown for each. In *a*, open circle corresponds to Cys-Cys, an internal control, as described in text; this point was excluded from the regression line. In *f*, open circles correspond to pairs bearing opposite charges, and closed circles correspond to pairs bearing like charges.

The straight line of best fit was derived for each pair of categories in Fig. 1. The linear equations together with standard errors and correlation coefficients, ρ, are as follows, respectively:

$$\langle A^b_{xy}\rangle = 0.12\ (\pm 0.01)\cdot A^0_{xy} + 34.08\ (\pm 2.11) \quad \rho = 0.95 \quad \textbf{[1a]}$$

$$\langle A^b_{xy}\rangle = 0.14\ (\pm 0.01)\cdot A^0_{xy} + 26.88\ (\pm 2.61) \quad \rho = 0.93 \quad \textbf{[1b]}$$

$$\langle A^b_{xy}\rangle = 0.17\ (\pm 0.01)\cdot A^0_{xy} + 16.74\ (\pm 3.90) \quad \rho = 0.93 \quad \textbf{[1c]}$$

$$\langle A^b_{xy}\rangle = 0.14\ (\pm 0.01)\cdot A^0_{xy} + 27.07\ (\pm 3.53) \quad \rho = 0.95 \quad \textbf{[1d]}$$

$$\langle A^b_{xy}\rangle = 0.15\ (\pm 0.02)\cdot A^0_{xy} + 22.00\ (\pm 4.83) \quad \rho = 0.92 \quad \textbf{[1e]}$$

$$\langle A^b_{xy}\rangle = 0.15\ (\pm 0.03)\cdot A^0_{xy} + 22.44\ (\pm 11.14) \quad \rho = 0.85 \quad \textbf{[1f]}$$

All interaction categories have similar slopes, each residue pair burying ≈0.14 Å^2 for every 1 Å^2 of available area, regardless of category. That is, each residue pair buries approximately one-seventh of its available area within the pair-wise interaction, approximating hexagonal close packing. The intercepts in these equations are also similar, and they are positive because, by definition, residues must bury some area between them to be classified as a pair.

For hydrophobic pairs, in particular, values of $\langle A^b_{xy}\rangle$ scale linearly with total pair area, as seen in Fig. 1*a*. The pair with the smallest total area, Ala-Ala, and that with the largest total area, Trp-Trp, are both well-represented by the least-squares line through the data. The only significant exception to linear scaling is seen for Cys-Cys pairs, which serve as a control. Such pairs are often covalently bonded and, thus, held in

close proximity, with a consequent increase in $\langle A^b_{CysCys} \rangle$. The point corresponding to Cys-Cys in Fig. 1*a* is shown as an open circle.

It is conceivable that preferred pair-wise interactions do exist but are concealed within the envelope of average values in Fig. 1 and Eqs. **1a–f**. To test this possibility, histograms that plot the distribution of pairs as a function of area buried were prepared for all 190 pair-wise interactions; the 36 hydrophobic pairs are shown in Fig. 2. The broad distributions seen in these histograms are not suggestive of any markedly preferred modes of interaction. To further pursue this conclusion, many specific pairs with pair-wise buried areas that rank in the upper 5% of their class (i.e., the right-most tails of the histograms in Fig. 2) were inspected for preferred interaction geometries, since visual inspection of the data might have revealed patterns refractory to statistical analysis. A sample of 5–10 of the most buried pairs was collected from every category containing two large hydrophobic residues (e.g., Leu-Leu, Val-Ile, etc.) as well as selected categories containing polar or charged residues (for example, Asp-Arg, Lys-Trp, etc.). Each pair in this sample was examined individually using an Evans and Sutherland PS 330 graphics system with the INSIGHT molecular modeling language (21). No regularities were evident upon visual inspection, consistent with the preceding interpretation of Fig. 1.

Area Buried Between Residue Pairs and the Protein. No preferred pair-wise interactions are apparent in the plots of Fig. 1, except for Cys-Cys pairs, which serve as a control. The remaining question is whether any of the 190 pairs are preferred by the host protein. That is, are some pairs more efficiently buried than other pairs of the same chemical type?

To answer this question, the surface area buried between each type of residue pair and its host protein was calculated. In Fig. 3, the mean total area buried for each pair, $\langle A^t_{xy} \rangle$, is plotted against the standard-state area of the pair, A^0_{xy}. Explicitly, $\langle A^t_{xy} \rangle$ is the component of side-chain area buried by residue pair x–y and the remainder (i.e., non-x, non-y) of the protein.

As in Fig. 1, residues are subdivided into three categories (hydrophobic, polar, and charged), which are cross-compared in Fig. 3. In all categories, $\langle A^t_{xy} \rangle$ scales linearly with total pair area. Hydrophobic pairs are especially well-fit by a straight line; both the smallest (Ala-Ala) and largest (Trp-Trp) fall near the line. Only one infrequently occurring pair, Met-Met, is found $>2\sigma$ from its predicted position. Interestingly, Cys-Cys pairs are no longer singular, burying as much total surface as would be expected for any other hydrophobic–hydrophobic pair of corresponding size.

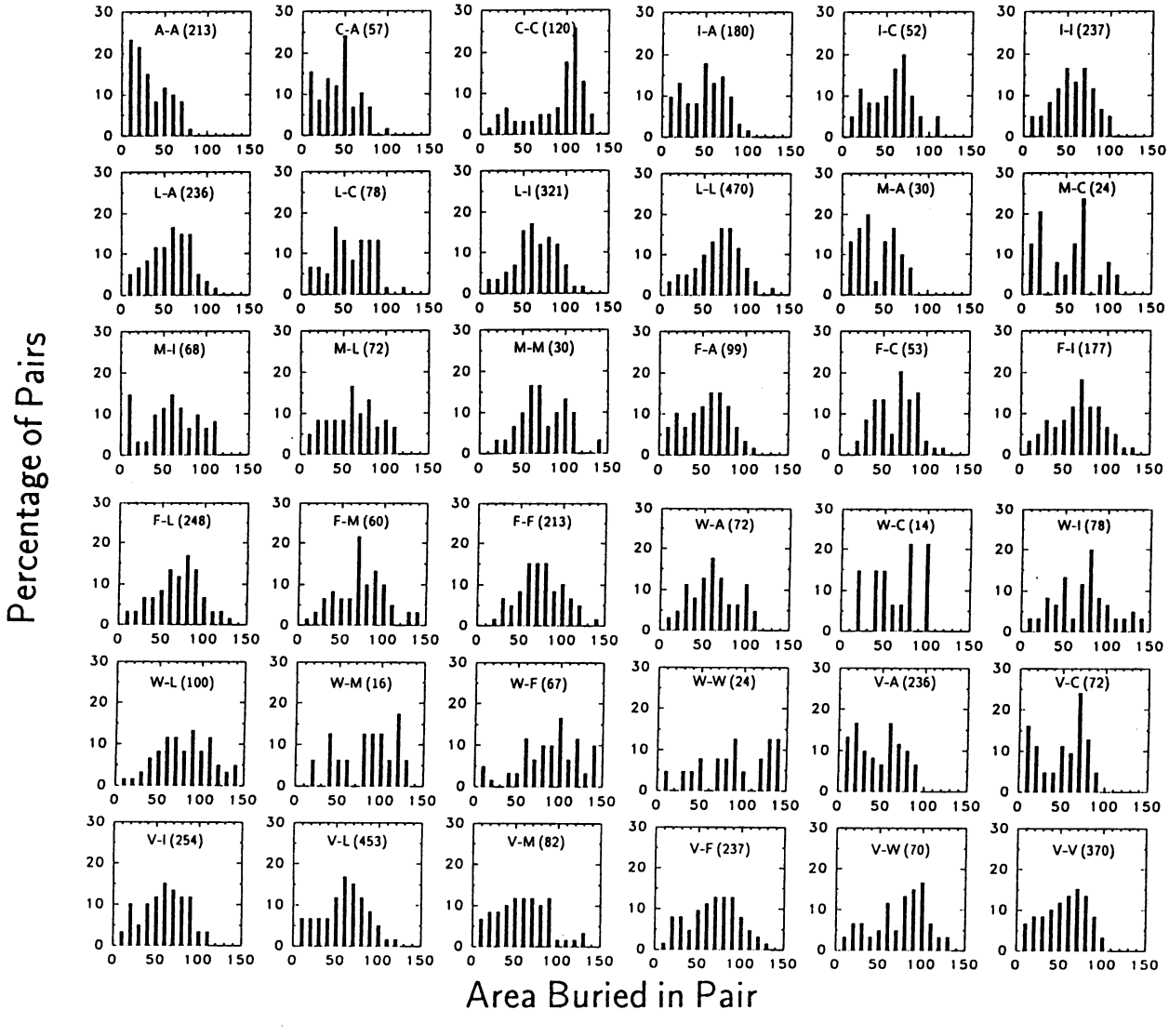

FIG. 2. Each histogram, representing one pair-wise interaction between hydrophobic residues, plots the observed distribution of a pair as a function of pair-wise area buried. Histograms are identified by a pair of single-letter amino acid codes together with number of occurrences of that pair, shown in parentheses.

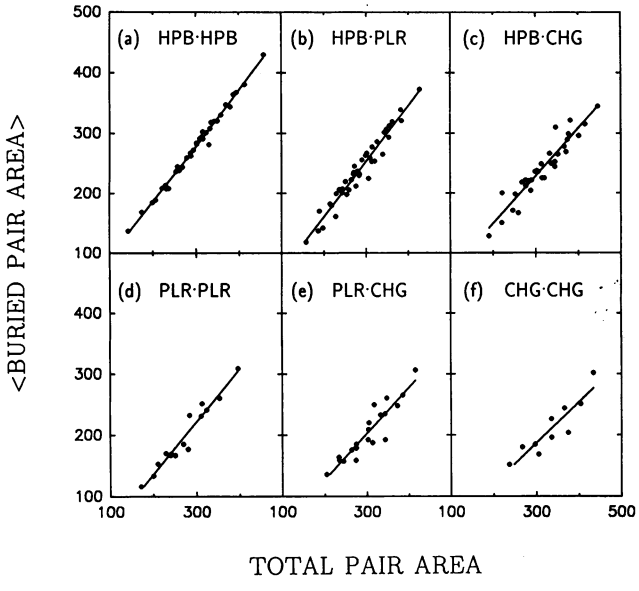

FIG. 3. Plot of the mean area buried between each type of residue pair and its host protein, $\langle A^{t}_{xy} \rangle$, versus the total area of the pair, A^{0}_{xy}. Areas are in Å^2. Categories, abbreviations, and division into *a–f* are identical to Fig. 1. Respective least-squares lines are shown for *a–f*.

No Preferred Interactions Found Among Hydrophobic Side Chains. Globular proteins have characteristically high packing densities and few internal cavities of atomic dimension (4). It seems plausible that such efficient packing originates in natural side-chain complementarity, and, indeed, the shapes of many individual residues do seem potentially complementary. For example, the planar rings of aromatic residues—phenylalanine, tryptophan, and tyrosine—could stack like aromatic bases in nucleic acids; leucine side chains might nestle together like spoons, and oppositely charged residues, such as arginine and glutamic acid could align so as to simultaneously foster both electrostatic and hydrophobic interactions.

A number of investigators have analyzed proteins for inter-residue and inter-atomic contacts (22–28). It was observed that approximately one-third of charged residues are positioned so as to form ion pairs (29–31), that aromatic groups are often aligned with an inter-group distance between 4.5–7 Å and a dihedral angle near 90° (32), and that sulfur (33) and oxygen (34) atoms interact preferentially with the edges of aromatic groups. These discovered preferences are all influenced primarily by electrostatic attraction and not by hydrophobic packing.

In this study we ask a different question: do any two residue side chains bury more surface between them or have more of their conjoint surface buried by the rest of the protein than other pairs of comparable chemical type? This question is related intrinsically to quantitative measures of hydrophobic interaction and packing (3–6, 19, 35–39). Except for controls, no preferred packing interactions were evident in our analysis, either within pairs of residues (Fig. 1) or between residue pairs and their respective protein-binding pockets (Fig. 3).

Implications for Protein Folding and Packing. This study emphasizes the concept of conformational specificity, defined as the property that enables a globular protein to discriminate the native fold from other conformations. Despite considerable heterogeneity, the protein interior is exceedingly well packed (3–6), and this fact has prompted the often-held view that packing and conformation are linked by cause and effect.

One can make a logical distinction between two possible conditions: packing as an indispensable prerequisite for the native conformation (*conditio sine qua non*) and packing as the causal agent of the native conformation (*conditio per quam*). Applying this distinction, Richards' observation of characteristically high packing densities in proteins (4) is experimental support for the view that packing is a necessary prerequisite for native conformation. This view is consistent with studies of folding thermodynamics (40–42) because well-packed conformations—given that they are attainable—will be energetically favored over more loosely packed alternatives. Indeed, protein–protein dispersion forces within the molecular interior must be at least as favorable as corresponding water–protein interactions in the unfolded state, or else dispersion forces would favor denaturation.

It need not follow, however, that packing also serves as the causal agent for conformational specificity, and, in fact, the foregoing analysis indicates that it does not. Our results lend themselves to either of two extreme interpretations: (*i*) Each protein represents a singular microcosm. In this view, the side chains can potentially populate any conformation permitted by rotation about their dihedral angles (13, 14). If the set of proteins under consideration is small relative to the ensemble of pair-wise side-chain conformers, then no preferred pattern of interaction would be found. (*ii*) The naturally occurring apolar residues can pack together efficiently in a large number of ways. The energy differences between alternate arrangements are slight, and no small number of patterns would predominate.

Results from mutational studies of proteins (8–12, 43) impose constraints upon interpretation *i*. If this interpretation is correct, then such studies imply that conformation space contains a virtual continuum of well-packed arrangements. Were this not the case, a large class of mutations would be expected to result in either a major conformational shift or a large reduction in stability. In fact, mutational studies demonstrate that proteins can tolerate sweeping mutations, typically with only minor changes in both structure and stability (8–12, 43). This view is corroborated by the very existence of protein families—e.g., the globins, which can be regarded as mutational studies of nature. To the degree that these results are general, a given fold must be near other well-packed alternatives of similar conformation. In such a conformational "landscape," conformation is not a sensitive function of internal packing.

If interpretation *ii* is correct, then, in general, allowed conformers (13, 14) can pack together efficiently. In this case, energy differences between conformers would not be due primarily to packing.

Both interpretations *i* and *ii* lead to the conclusion that packing is not the principal cause of conformational specificity. This conclusion is surprising to the degree that our expectations have been conditioned by the low packing densities seen for liquid hydrocarbons (44). As noted by Richards, the protein interior is not comparable to oil (4). However, comparison can also be made with crystalline hydrocarbons (44), which fall into two distinct classes. The first class crystallizes in oblique or rectangular lattices and is more tightly packed than the second class, which crystallizes in hexagonal or nearly hexagonal lattices. The mean volume per -CH_2- group is ≈24 Å^3 in the former class and ≈25.5 Å^3 in the latter. Using Richards' value (3) of 18.23 Å^3 per -CH_2- group, the corresponding packing densities are 0.76 and 0.71, respectively. Similar packing densities are found for -CH_3 groups. These packing densities resemble the values in globular proteins (4).

If packing is not the causal agent of conformational specificity, then it should be possible for a protein to attain native quality packing without adopting the native conformation. In fact, Novotný *et al.* (45) have shown that atomic packing fails to discriminate between deliberately misfolded proteins and their native counterparts. A similar test, which to our knowl-

edge has not been made, would use available algorithms (46, 47) to reconstruct residue side chains from the α-carbon coordinates of a deliberately misfolded protein.

Why do proteins have a unique conformation? A globular protein adopts its unique equilibrium conformation spontaneously under physiological conditions, guided along the folding pathway by information encoded within the linear sequence (1). In a bad solvent, the polypeptide chain is driven toward spatially compact states by forces that are familiar (36).

Why do proteins have a unique conformation? Highly selective internal packing seems an attractive way to rationalize conformational uniqueness (13), but in view of the preceding analysis, it is now deemed unlikely. An informative discussion of this issue is found in a recent review by Dill (48), who suggests that, over and above chain compactness, the most significant further restriction on conformation space is imposed simply by the hydrophobic effect. It has been shown by both theory and simulation (49, 50) that a heteropolymeric sequence of nonpolar and polar monomers (e.g., residues) will have few conformers, irrespective of monomer details. Also, it is known that local hydrophobic clusters, though randomly distributed (51), coincide well, albeit imperfectly, with buried regions of the three-dimensional structure (52–54). Similar ideas have been used to define a characteristic hydrophobicity pattern for a protein; such a pattern can, in turn, be used to selectively identify other proteins from the same family within a heterogeneous data base (55–57).

The foregoing ideas about packing are subject to experimental validation. For example, following the strategy of Ponder and Richards (13), sequences designed to be compatible with a particular fold could be devised. Design criteria for such sequences might include the linear pattern of hydrophobic and hydrophilic residues found in a natural protein but not the actual residues. A similar approach to the design of amphipathic helices has been successfully exploited by Kaiser and coworkers (58).

The evidence presented in this paper can be summarized in a sentence: in globular proteins, the native fold determines packing, but packing does not determine the native fold. We anticipate that this realization will simplify approaches to the folding problem.

This paper is dedicated to Frederic M. Richards. We are indebted to the many people who offered critical comments and helpful suggestions upon reading earlier versions of this manuscript, including Tom Alber, Robert Baldwin, William DeGrado, Ken Dill, David Eisenberg, S. Walter Englander, Jacquelyn Fetrow, Juliette Lecomte, Wendell Lim, Frederic Richards, Robert Sauer, Stephen H. White, and Bruno Zimm. This work was supported by National Institutes of Health Grant GM 29458.

1. Anfinsen, C. B. (1973) *Science* **181,** 223–230.
2. Kauzmann, W. (1959) *Adv. Protein Chem.* **14,** 1–63.
3. Richards, F. M. (1974) *J. Mol. Biol.* **82,** 1–14.
4. Richards, F. M. (1977) *Annu. Rev. Biophys. Bioeng.* **6,** 151–176.
5. Chothia, C. (1975) *Nature (London)* **254,** 304–308.
6. Finney, J. L. (1975) *J. Mol. Biol.* **96,** 721–732.
7. Schulz, G. E. & Schirmer, R. H. (1979) *Principles of Protein Structure* (Springer, New York).
8. Estell, D. A., Graycar, T. P., Miller, J. V., Powers, D. B., Burnier, D. B., Ng, P. G. & Wells, J. A. (1986) *Science* **233,** 659–663.
9. Matthews, B. W. (1987) *Biochemistry* **26,** 6885–6888.
10. Alber, T. (1989) *Annu. Rev. Biochem.* **58,** 765–798.
11. Lim, W. A. & Sauer, R. T. (1989) *Nature (London)* **339,** 31–36.
12. Matouschek, A., Kellis, J. T., Serrano, L. & Fersht, A. R. (1989) *Nature (London)* **340,** 122–126.
13. Ponder, J. W. & Richards, F. M. (1987) *J. Mol. Biol.* **193,** 775–791.
14. Moult, J. & James, M. N. G. (1986) *Proteins* **1,** 146–163.
15. Bryant, S. H. (1989) *Proteins* **5,** 233–245.
16. Singh, J. & Thornton, J. M. (1990) *J. Mol. Biol.* **211,** 595–615.
17. Bernstein, F. C., Koetzle, T. G., Williams, G. J. B., Meyer, E. F., Jr., Brice, M. D., Rogers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) *J. Mol. Biol.* **112,** 535–542.
18. Lee, B. K. & Richards, F. M. (1971) *J. Mol. Biol.* **55,** 379–400.
19. Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H. & Zehfus, M. H. (1985) *Science* **229,** 834–838.
20. Rose, G. D., Gierasch, L. M. & Smith, J. A. (1985) *Adv. Prot. Chem.* **37,** 1–109.
21. Biosym Technologies (1991) INSIGHT Molecular Modeling System (Biosym, San Diego).
22. Manavalan, P. & Ponnuswamy, P. K. (1977) *Arch. Biochem. Biophys.* **184,** 476–487.
23. Warme, P. K. & Morgan, R. S. (1978) *J. Mol. Biol.* **118,** 273–287.
24. Warme, P. K. & Morgan, R. S. (1978) *J. Mol. Biol.* **118,** 289–304.
25. Kuntz, I. D. & Crippen, G. M. (1979) *Int. J. Peptide Protein Res.* **13,** 223–228.
26. Narayana, S. U. L. & Argos, P. (1984) *Int. J. Peptide Protein Res.* **24,** 25–39.
27. Miyazawa, S. & Jernigan, R. L. (1985) *Macromolecules* **18,** 534–552.
28. Bryant, S. H. & Amzel, L. M. (1987) *Int. J. Peptide Protein Res.* **29,** 46–52.
29. Barlow, D. J. & Thornton, J. M. (1983) *J. Mol. Biol.* **168,** 867–885.
30. Rashin, A. A. & Honig, B. (1984) *J. Mol. Biol.* **173,** 515–521.
31. Barlow, D. J. & Thornton, J. M. (1986) *Biopolymers* **25,** 1717–1733.
32. Burley, S. K. & Petsko, G. A. (1985) *Science* **229,** 23–28.
33. Reid, K. S. C., Lindley, P. F. & Thornton, J. M. (1985) *FEBS Lett.* **190,** 209–213.
34. Thomas, K. A., Smith, G. M., Thomas, T. B. & Feldmann, R. J. (1982) *Proc. Natl. Acad. Sci. USA* **79,** 4843–4847.
35. Lesser, G. J. & Rose, G. D. (1990) *Proteins* **8,** 6–13.
36. Tanford, C. (1980) *The Hydrophobic Effect* (Wiley, New York).
37. Nozaki, Y. & Tanford, C. (1971) *J. Biol. Chem.* **246,** 2211–2217.
38. Wolfenden, R., Andersson, L., Cullis, P. M. & Southgate, C. B. (1981) *Biochemistry* **20,** 849–855.
39. Eisenberg, D. & McLachlan, A. D. (1986) *Nature (London)* **319,** 199–203.
40. Sandberg, W. S. & Terwilliger, T. C. (1989) *Science* **245,** 54–57.
41. Malcolm, B. A., Wilson, K. P., Matthews, B. W., Kirsch, J. F. & Wilson, A. C. (1990) *Nature (London)* **345,** 86–89.
42. Connelly, P. R., Varadarajan, R., Sturtevant, J. M. & Richards, F. M. (1990) *Biochemistry* **29,** 6108–6114.
43. Lim, W. A. & Sauer, R. T. (1991) *J. Mol. Biol.*, in press.
44. Small, D. M. (1986) *The Physical Chemistry of Lipids* (Plenum, New York).
45. Novotný, J., Rashin, A. A. & Bruccoleri, R. E. (1988) *Proteins* **4,** 19–30.
46. Reid, L. S. & Thornton, J. M. (1989) *Proteins* **5,** 170–182.
47. Correa, P. (1990) *Proteins* **7,** 366–377.
48. Dill, K. A. (1990) *Biochemistry* **39,** 7133–7155.
49. Lau, K. F. & Dill, K. A. (1989) *Macromolecules* **22,** 3986–3997.
50. Shakhnovich, E. I. & Gutin, A. M. (1990) *Nature (London)* **346,** 773–775.
51. White, S. H. & Jacobs, R. E. (1990) *Biophys. J.* **57,** 911–921.
52. Rose, G. D. (1978) *Nature (London)* **272,** 586–590.
53. Rose, G. D. & Roy, S. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 4643–4647.
54. Kyte, J. & Doolittle, R. F. (1983) *J. Mol. Biol.* **157,** 105–132.
55. Sweet, R. M. & Eisenberg, D. (1983) *J. Mol. Biol.* **171,** 479–488.
56. Gribskov, M., McLachlan, A. D. & Eisenberg, D. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 4355–4358.
57. Bowie, J. U., Clarke, N. D., Pabo, C. O. & Sauer, R. T. (1990) *Proteins* **7,** 257–264.
58. Taylor, J. W. & Kaiser, E. T. (1987) *Methods Enzymol.* **154,** 473–498.