Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 638–642, January 1990
Biophysics

Theory for protein mutability and biogenesis

(lattice model/exhaustive simulation/random sequences/protein origins/protein folding)

KIT FUN LAU AND KEN A. DILL

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

Communicated by Frederic M. Richards, October 17, 1989 (received for review June 14, 1989)

ABSTRACT **Using an elementary physical model for protein folding, of self-avoiding short copolymer chains on two-dimensional square lattices, we address two questions regarding the evolution and origins of globular proteins. (*i*) How will protein native structures and stabilities be affected by single- and double-site mutations? (*ii*) What is the probability that a randomly chosen sequence of amino acids will be compact and globular under folding conditions? For a large number of different sequences, we search the conformational space exhaustively to find unequivocally the "native" conformation(s), of global minimum free energy, for each sequence. We find that replacing nonpolar residues in the core by polar residues is generally destabilizing, that surface sites are less sensitive than core sites, that some mutations increase the degeneracy of native states, and that overall it is most probable that a mutation will be neutral, having no effect on the native structure. These results support a "Continuity Principle," that small changes in sequence seldom have large effects on structure or stability of the native state. The simulations also show that (*i*) the number of "convergent" sequences (different sequences coding for the same native structure) is extremely large and (*ii*) most sequences become quite dense under folding conditions. This implies that the probability of formation of a globular protein from a random sequence of amino acids by prebiotic or mutational methods is significantly greater than zero.**

The central question of protein evolution is how mutational change in the amino acid sequence leads to change in the structure and stability, and thereby to change in protein function. A fundamental question of protein origins is how different sequences, and how many of them, could fold up to stable globular conformations with active sites. To address these questions of protein biogenesis and change at the molecular level requires a "protein folding" algorithm, a predictor of the native structure of a protein from its amino acid sequence. It further requires an algorithm sufficiently fast so as to be capable of exploring a suitably large fraction of "sequence space," the set of all possible sequences of amino acids (Fig. 1). No such algorithm yet exists at atomic resolution. Our purpose here is to address these questions through use of a recently developed lower-resolution lattice statistical mechanics model of protein folding (1).

The Lattice Model of Protein Sequence/Structure

The model has been described in detail elsewhere (1); we just summarize the principal aspects here. Each protein is a linear chain of a specific sequence of n amino acids. Each amino acid can be either of two types: H (nonpolar) or P (other). The fraction of the n residues which are of type H is Φ; hence the fraction of type P is $1 - \Phi$. A chain conformation is represented as a self-avoiding walk on a two-dimensional

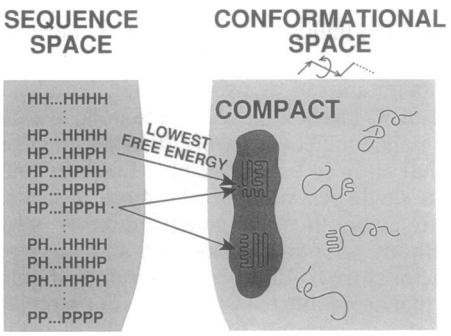

FIG. 1. Mapping of sequence space to native structure(s). Sequence space is the set of all possible sequences; conformational space is the set of all possible conformations. Two arrows to a native state from different sequences represents "convergence;" two arrows to different native structures from a given sequence represents "degeneracy." H, nonpolar (hydrophobic) residue; P, other type of residue (see text).

(2D) square lattice. Thus, each amino acid is represented as simply occupying one lattice site, connected to its chain neighbor(s), and unable to occupy a site filled by any other residue. For each monomer, we define "connected" neighbors as units j and $j + 1$ adjacent along the chain, and "topological" neighbors as units (i, j) adjacent in space but not adjacent in position along the sequence, $i \neq j - 1, j + 1$. Every HH contact between topological neighbors is assigned a contact energy equal to ε (<0) and every other interaction among all other possible types of neighbor pairs (H, P, or solvent, S) has energy equal to 0. For sufficiently short chains every accessible conformation is then surveyed by computer, and its energy (number of HH contacts) is evaluated to find the native structure(s), those with the maximum number of HH contacts. For $n < 11$, native states are found in this manner for every possible sequence. For longer chains, we randomly choose 200 different sequences, and native states are found by exhaustive exploration of all maximally compact conformations (1).

This model has the following basic features in common with real proteins (1). For small HH attraction, the predominant populations of chain conformations are unfolded. For molecules composed of certain sequences (which we refer to as "folding" sequences), increasing the HH attraction leads to a transition to a native state characterized by a very small number of conformations (most often, only one). The native conformations have the properties that they (*i*) are maximally compact, (*ii*) are composed of a core of predominantly H residues, and (*iii*) have considerable amounts of 2D equivalents of secondary structures: helices and sheets, principally antiparallel (2, 3). Moreover, their distributions of secondary structures are approximately the same as those observed in three-dimensional (3D) protein structures (H. S. Chan and K.A.D., unpublished results).

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.

Abbreviations: 2D and 3D, two- and three-dimensional or dimensions.

There are two principal virtues of this approach. (*i*) The model is physical; it is based on the dominant driving forces for folding: the solvent aversion of nonpolar residues in water and the conformational freedom and excluded volume of the chain. It has no adjustable parameters or additional *ad hoc* assumptions. (*ii*) It is sufficiently simple that for any given sequence, the entire conformational space can be searched exhaustively to determine unequivocally the "native" structure(s), those which are at the global minimum of free energy. Moreover, the native structures can be found for a large number of different sequences, permitting a broad exploration (exhaustive, in some cases) of the full sequence space.

This model is intended to address questions of general principle rather than of atomic detail. For questions of this type, justification is given elsewhere (1) that the shortness of the chains and 2D nature of the model should not be significantly limiting. We consider here only the nature of the compact folded conformations of chains; we do not address the thermodynamics of folding, which will be more sensitive to simplifications of this type. The native conformations are simply those of lowest energy, found by taking the limit $\varepsilon \rightarrow -\infty$ (1), which we refer to as "folding conditions." The distinction that ε is actually a free energy rather than an energy is irrelevant for present purposes.

To decide whether a sequence is a "folder" or not, we use the compactness (1–3) $\rho = t/t_{max}$, where t is the total number of topological neighbors in a given conformation and t_{max} is the maximum possible number of topological neighbors which could be achieved by any conformation of a given chain length. Below we consider different criteria for globularity, ranging from the most strict ($t = t_{max}$) to less strict ($t \geq t_{max} - 2$); a conformation is considered folded if it satisfies this criterion under folding conditions.

We have previously found that some folding sequences have only one native state (i.e., they are singly degenerate), and other folding sequences have more than one (multiply degenerate) (1). For long chains, it is most probable that a sequence will have only one native state (1). Few folding sequences have more than 20 native states, although the nonfolding sequences have thousands of conformations of lowest equal energy, few of which are compact. In the present work we focus largely on singly degenerate folding sequences. These may be of greater biological relevance, since otherwise structural ambiguity is likely to impose functional ambiguity. Singly degenerate sequences occur largely in the composition range of 30–70% H. Sequences with too few H residues do not have compact native structures, and sequences with too many H residues have many different compact structures of equal lowest energy and thus are not singly degenerate.

Plasticity of Native Structures to Sequence Mutations

First, we ask: if a sequence is changed at one or two residue positions, then how much will the native structure and energy change? The question is whether proteins are either (*i*) highly "plastic," whereby single-site mutations would seldom lead to a change of the native structure, or (*ii*) so precisely crafted that a single mutation would usually significantly alter the native structure. We test this here by choosing a sequence with known native structure, and making each possible single-site change, from H to P, or P to H, whichever is appropriate, then determining the "mutant" native structure and energy. For this purpose, from about 1000 random sequences, 251 were chosen as good folders with singly degenerate native states. We characterize structural change by a quantity Δd (defined by equation 13 of ref. 1), which simply counts the number of bond rotation changes, weighted by the relative magnitude of each change.

The distribution of structural change resulting from 3263 different single-site mutations on chains of length $n = 13$ is shown in Fig. 2*B*. It is clear from Fig. 2 that proteins are predicted to be nearly maximally insensitive to single-site change. In this case, nearly 90% of all possible single-site changes cause no change in the native structure. It is interesting that when a mutation does change the structure, it does so cooperatively: the small peak at $\Delta d = 3$–8 implies that the typical single-site-modified mutant native structure has several bonds all reconfigured concertedly. Most of the concerted reconfigurations arise from changing an H to a P in the core. Most surface mutations lead to no change in the structure. This observed insensitivity of native structures to single mutational change is particularly remarkable, since each single change is a potentially large fraction of the overall stabilization energy for these short model chains.

Fig. 2*A* shows that the native state energies, like native state structures, are relatively insensitive to single-site mutation. When a P residue replaces an H in the core, it destabilizes on average. When a P residue replaces an H on the surface of the protein, there is less average destabilization and a larger number of neutral mutations. When an H residue replaces a P at the surface, it is generally neutral; less often

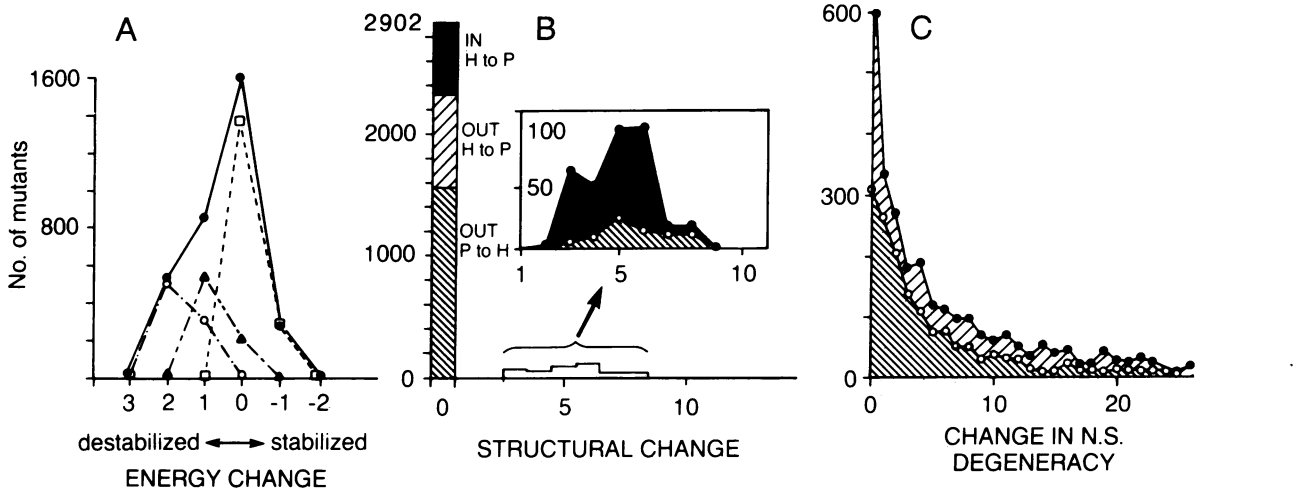

FIG. 2. Consequences of single mutations. (*A*) Distribution of energy changes in 3263 mutants. ●, Total energy change; ○, H changed to P inside; ▲, H to P outside; □, P to H outside. (*B*) Distribution of structural changes (x-axis is Δd; see text and ref. 1). (*Inset*) Expanded y-axis. Key for breakdown of changes in the 0-change column applies to *B* and *C*. (*C*) Distribution of the number of mutation-induced new native states (N.S.) (in addition to the original wild type).

it is stabilizing. There are too few instances of stable native structures with a P residue in the core for these chains of $n = 13$ to draw conclusions about that type of replacement. Some examples of the specific configurational changes which arise from these single-site mutations are shown in Fig. 3. These results are consistent with evidence that surface modifications generally have less effect on conformation or stability than interior changes and that interior changes are often destabilizing (4–9).

A principal conclusion from these single-site mutation experiments is that, taken over all possible types of mutations at both core and surface positions, it is most probable that a mutation will be neutral, causing no change in native structure or energy. That mutations are generally neutral for structure and stability is a necessary but not sufficient condition to account for the observations that mutations are also generally neutral for biological function (10).

There are two interesting caveats to the conclusion that most mutations are neutral. First, with increasing chain length, the surface-to-volume ratio of the native state decreases, so that interior changes become more numerous relative to surface mutations and therefore statistically more important. Second, some mutations have no effect on either the structure or the energy of the native state. These mutations are somewhat more subtle, inasmuch as they affect only the *degeneracy* of native states. This is a situation in which the wild-type native state remains a native structure for the mutated sequence, but the mutant now also has additional native states of the same energy. Fig. 2*C* shows that most mutations induce these additional native states, often many of them. This degeneracy generally arises from sequence changes at surface positions (particularly P changed to H), whereby a surface loop becomes "floppy" (i.e., it develops multiple isoenergetic conformations; see Fig. 3 *C* and *D*). Large changes in degeneracy occur when mutation results in fewer possible HH contacts, often due to H to P change in the core (not shown). The practical effect of increased-degeneracy mutations in real proteins would be decreased enzyme activity, since some of the mutation-induced alternative native structures may not be active.

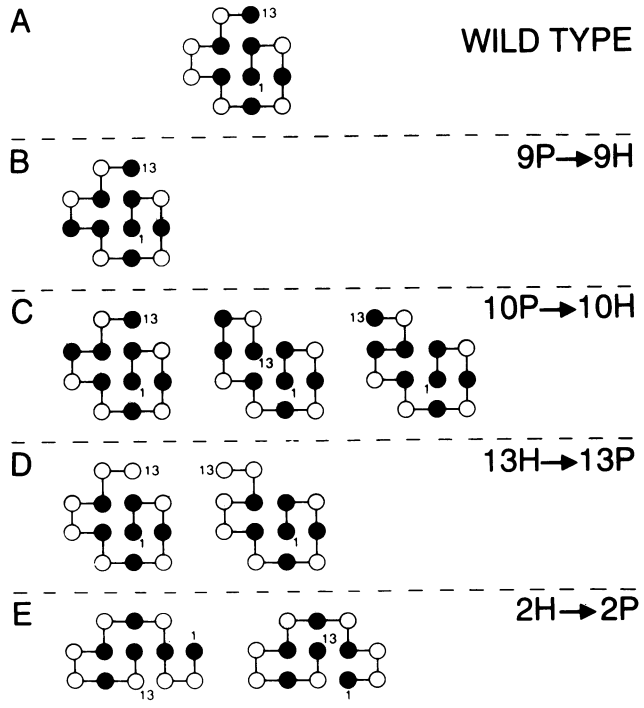

FIG. 3. Effects of single mutations. ●, H residues; ○, P residues. (*A*) Wild type. This sequence has only this one maximally compact native state. (*B*) Change of residue 9 from P to H causes no change in native structure. (*C* and *D*) Single mutations which increase the degeneracy of native states; the wild type is among them. (*E*) Single mutation which destroys the native structure.

We have also performed 3012 double-site mutations on 251 different sequences of length $n = 13$. The first site for mutation was chosen randomly. The second site was chosen to be every one of the 12 remaining sites on that chain. The conclusions are quite similar to those of the single-site changes: (*i*) most are structurally neutral, (*ii*) there is cooperativity when they are not neutral, (*iii*) the distribution of effects by type and location of the mutation is similar, (*iv*) the same effect of surface-to-volume ratio is observed for the one longer chain length we have tested ($n = 18$). The principal difference of the double mutants relative to the single mutants is, not surprisingly, simply that there is overall somewhat more net change in structure and native energy from two mutations than from one.

In addition, we observe a small fraction of revertants—second-site mutations that cancel the effect of the first, either on structure or on energy of the native state. Two percent of the double-site mutants are structural revertants; the first mutation modifies the native structure, and the second, at a different site, reestablishes the original native structure (see Fig. 4). We always observe multiple native states in revertants from singly degenerate wild types; hence a general consequence of reversion may be reduced enzymatic activity.

The results of these mutation experiments can be summarized by what might be referred to as the "Continuity Principle," the idea that the lowest free energy surface in sequence space is relatively smooth. That is, small perturbations of sequence lead to small perturbations of native structure, and larger perturbations of sequence generally lead to larger perturbations of native structure. In this regard, this model is consistent with what is now a large body of experimental evidence that protein native structures are relatively insensitive to single and double mutations (4, 5, 8, 9, 11, 12). It is hard to imagine how biological evolution could succeed otherwise. If the lowest free-energy surface of sequence space were predominantly discontinuous rather than continuous, then mutations would have arbitrary consequences, largely destroying the structure and function of a protein.

Globular Molecules from Random Amino Acid Sequences?

It has been generally believed that the evolutionary origin of an enzyme from a random sequence of amino acids is exceedingly improbable. The number of different primary sequences of length $n = 100$ residues, composed at each position of the 20 different amino acids, is $20^{100} = 1.27 \times 10^{130}$. Hence, the probability that a single protein with a specified sequence

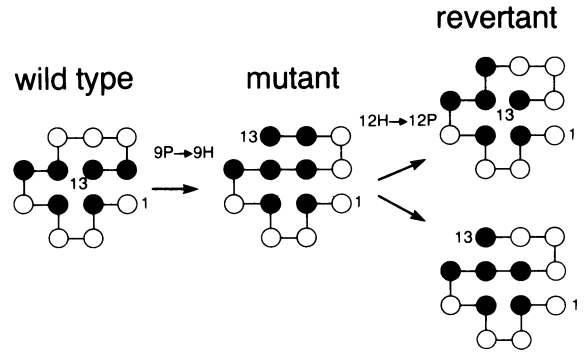

FIG. 4. Revertant. Singly degenerate wild-type sequence is mutated at position 9, resulting in a different native structure. A second mutation, at position 12, reestablishes the wild type as one of two native structures.

would have arisen by random chance through independent selection of a sequence of amino acids is 7.89×10^{-131}. This number is so exceedingly small that, as with the chance of producing a Shakespearean play from a monkey dancing on a typewriter, it is essentially impossible. A variation of this argument is that a given protein function is achieved if only a few of the n residues are specified precisely, rather than all of them. On this basis, estimates range from 10^{-20}, if 10 residues are essential (13), to 2.1×10^{-65}, based on sequence variation in cytochrome c (refs. 14–16, but see also ref. 17). This general argument has become of some importance as support for the view that proteins could not have arisen from natural prebiotic chemical processes on earth (13) and as support for creationism (18–20).

The alternative to this "sequence" hypothesis, that nature "seeks" a particular sequence, is the "structure" hypothesis, that nature seeks merely any compact conformation with the proper active site. We explore the structure hypothesis below. We believe that evolution "cares" only about the biological function, and therefore the native conformation of the molecule, no matter what sequence is required to achieve it (21–23). The sequence and structure views are extremely different: it is shown below that any single given native structure will arise from an extraordinarily large number of different sequences.

In this section, we first ask: What fraction of sequence space corresponds to compact globular molecules under folding conditions? Let $N(n)$ equal the number of sequences of chain length n in the sequence space, and let $N_f(n)$ equal the number of sequences which fold—i.e., those in which the conformations of lowest energy (native states) are maximally compact or nearly so. The simulations show that the fraction of sequence space corresponding to folding molecules diminishes approximately as (see Fig. 5)

$$\frac{N_f(n)}{N(n)} = ka^n, \qquad [1]$$

where the constants are $k = 2.04$ and $a = 0.792$ for $t = t_{max}$, $k = 4.87$ and $a = 0.813$ for $t \geq t_{max} - 1$, and $k = 4.06$ and $a = 0.860$ for $t \geq t_{max} - 2$. On this basis, extrapolation shows that for chains of $n = 100$ monomer units, the fraction of these 2D sequences which fold is in the range of 10^{-6} to 10^{-10}, depending on the strictness of the criterion used to classify a sequence as a folding molecule.

These folding criteria are the strictest possible and are thus highly conservative. For example, even the least strict folding criterion, $t \geq t_{max} - 2$, is considerably more restrictive than would be appropriate to characterize real proteins—i.e., longer chains that generally have rough surfaces, active site cavities, and sometimes floppy loops. In addition, it has previously been shown that in a "sequence space soup," a medium containing an ensemble of every possible sequence of a given length, it is most probable that a molecule will form approximately half the maximum possible number of HH contacts under folding conditions (see figures 8 and 12–15 of ref. 1).

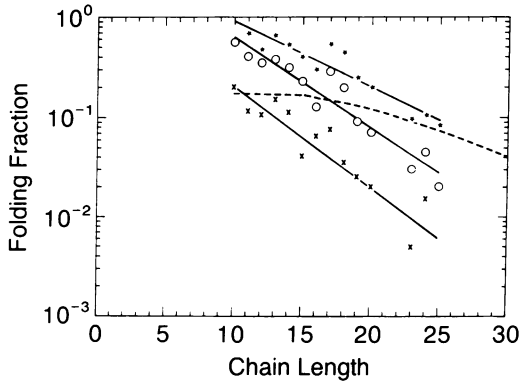

FIG. 5. Fraction of all sequences which fold, as a function of chain length. Continuous lines are the best fits through calculated points. X, $t = t_{max}$; ○, $t \geq t_{max} - 1$; *, $t \geq t_{max} - 2$. Broken line represents prediction from the "critical core" model, Eq. 4.

In other words, most molecules in sequence space soup are predicted to be extremely compact under folding conditions (1).* Because molecular compactness drives the formation of secondary structures (3), these many globular structures should also contain considerable amounts of secondary structure. These predictions are supported by the experiments of Rao *et al.* (24) on random terpolymer sequences of lysine, alanine, and glutamic acid. They observe that their terpolymer sequence-space soup is composed of highly globular molecules with overall 46% helix content, as determined by circular dichroism.

Whereas the fraction of all possible sequences which can fold to *some* compact conformation is $N_f(n)/N(n)$, the fraction $g(n)$ which can fold to *one particular* compact conformation is approximately

$$g(n) = \frac{N_f(n)\langle\omega\rangle}{N(n)\Omega(n)}, \qquad [2]$$

where $\Omega(n) = \Omega_{2D}(n) = 0.226(1.40)^n$ or $\Omega(n) = \Omega_{3D}(n) = 0.0345(1.74)^n$ are the numbers of compact conformations [shape-averaged "magic" numbers (3)] for the 2D square lattice (3), or the 3D simple cubic lattice, respectively (H. S. Chan and K.A.D., unpublished results) and $\langle\omega\rangle$ is the average number of native states per folding sequence. This factor of $\langle\omega\rangle$ accounts for sequences which code for more than one compact native conformation; $\langle\omega\rangle = 9.4$ for $n = 10$.

The simulation results are useful for testing approximate analytical models for $g(n)$, the simplest of which is the "critical core" model: to specify a particular native structure simply requires H residues in particular critical positions in the sequence, largely located in the core of the protein. Suppose there are n_c "critical" residues, positions at which replacement of an H will lead to loss of that native conformation. If all other residues are completely irrelevant for the determination of the structure, and if H and P are drawn with equal likelihood from a pool of monomers, then the fraction of sequence space which can achieve this sequence is $g = (\frac{1}{2})^{n_c}$.

If, in addition, we assume that these critical residues are predominantly located in a "core," the geometric interior of the molecule, and if the compact molecules are taken to be spherical (in 3D), or circular disks (in 2D), then

$$n_c = n_{interior} = \begin{cases} \pi\left[\left(\frac{n}{\pi}\right)^{1/2} - 1\right]^2 & \text{in 2D} \\ \frac{4\pi}{3}\left[\left(\frac{3n}{4\pi}\right)^{1/3} - 1\right]^3 & \text{in 3D} \end{cases}, \qquad [3]$$

where the radius $(n/\pi)^{1/2}$ is in units of monomer size.

Hence, using Eqs. 2 and 3, and substituting $\Omega = \Omega_{2D}(n)$ and $g = (\frac{1}{2})^{n_c}$, we have in 2D:

$$\frac{N_f(n)}{N(n)} = \langle\omega\rangle^{-1}\, 0.226(1.40)^n(\tfrac{1}{2})^z, \qquad [4]$$

*Although we do not address the issue of protein–protein interactions, we note that an important problem in selecting sequences for biological viability is that biological systems require sequences which do not lead to aggregation at low concentrations. This criterion will not be satisfied by many folding sequences of high H composition, and it imposes a further restriction on the biological suitability of a sequence.

where the exponent $z = \pi[(n/\pi)^{1/2} - 1]^2$.

This "critical core" model is compared with the simulations of $N_f(n)/N(n)$ in Fig. 5. It is found to be a reasonable model, and it improves in the limit of increasing n. For example, it predicts $N_f(100)/N(100) = 3.89 \times 10^{-8}$, in the same range as the simulation results. Nevertheless, there are two caveats relevant to the critical core model. First, the simulations show that in addition to critical and irrelevant residues, there is a large third class, not taken into account in this critical core model, of residues which sometimes affect the conformation and sometimes do not, depending on other residues in the sequence. Second, our present simulations also show that the critical residues are not always found in the geometric interior of the molecule; sometimes they are at the surface. These results are consistent with the elegant random cassette mutagenesis experiments of Reidhaar-Olson, Lim, and Sauer on λ phage repressor (4, 9). They observe that a large number of different sequences are functional, largely dependent only on conservation of H residues in a critical core. In addition they observe some residue interdependences, as we have noted above.

The general principles derived from this 2D model should also apply to chains in 3D. When the critical core model and Eq. 3 are used, the number of different sequences that code for any single compact native backbone conformation in 3D for a chain of length $n = 100$ is predicted to be 5.4×10^{21} in terms of 2 residue types, H and P (or 5.4×10^{121} in terms of 20 residue types). This number may vary considerably depending on details; the principal conclusion is that any single given native conformation is encoded in an extremely large number of different sequences. This is consistent with much evidence for "convergent evolution" (25–31).†

What fraction of all sequences will have active sites? If we define an active site as three prespecified residues in particular relative positions and orientations in a compact native state (following the example of serine proteases requiring Ser-195, His-57, and Asp-102), then 10^{-2} to 10^{-7} of folding sequences should have a given active site. This estimate derives from simulations of residue pairings in lattice models of compact chains (3) and from rotational restrictions estimated by using small molecule rigid rotor partition functions. The 10^5-fold variation in this estimate arises because of the difficulty in accurately estimating the latter. Although only crude estimates are currently possible, nevertheless, combined with the prediction that a considerable fraction of the molecules in sequence space soup are highly compact, the structure hypothesis of protein biogenesis differs qualitatively from the sequence hypothesis in predicting that there is a significantly nonzero probability of the origin of an enzyme from a random sequence of amino acids.

Conclusions

The properties of proteins change in response to mutations of their sequences. Thus the laws of protein mutability and evolution must be rooted in the nature of the folding forces. In the present work, we have explored these laws by using a model for these driving forces which we believe to be the simplest possible. It takes into account chain flexibility, excluded volume, and the solvent aversion of the nonpolar residues in water. Even this rudimentary model is sufficient to account for many of the mutability and biogenic properties observed in proteins: the greater sensitivity of interior sites than exterior sites; the increased degeneracy of native states, and hence decreased activity, due to mutation; the existence of double-site revertants; the considerable insensitivity to single-site changes; the continuity of the native state surface of the sequence space; and the large number of convergent sequences for any given native structure. It suggests that these properties may not necessarily be unique to polymers composed of amino acids. And it suggests how molecules as complex as catalytic globular proteins could have arisen so readily in simple prebiotic solutions, wherein only a virtually negligible fraction of all possible sequences would have been sampled during the origins of life.

We thank Phil Patten for many stimulating and insightful discussions which provided considerable motivation for this work, and Drs. Hue Sun Chan, Charly Craik, and Jan W. H. Schreurs for helpful comments.

†"Convergent evolution" in enzymology generally refers to two different sequences whose two different native structures may have only a small degree of common structure (e.g., the active sites of chymotrypsin and subtilisin); for our simulations it refers to completely identical native structures.

1. Lau, K. F. & Dill, K. A. (1989) *Macromolecules* **22,** 3986–3997.
2. Chan, H. S. & Dill, K. A. (1989) *J. Chem. Phys.* **90,** 492–509.
3. Chan, H. S. & Dill, K. A. (1990) *Macromolecules*, in press.
4. Reidhaar-Olson, J. F. & Sauer, R. T. (1988) *Science* **241,** 53–57.
5. Alber, T., Dao-Ping, S., Nye, J. A., Muchmore, D. C. & Matthews, B. W. (1987) *Biochemistry* **26,** 3754–3758.
6. Wright, C. S., Alden, R. A. & Kraut, J. (1969) *Nature (London)* **221,** 235–242.
7. Hecht, M. H., Sturtevant, J. M. & Sauer, R. T. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 5685–5689.
8. Alber, T., Bell, J. A., Dao-Ping, S., Nicholson, H., Wozniak, J. A., Cook, S. & Matthews, B. W. (1988) *Science* **239,** 631–635.
9. Lim, W. A. & Sauer, R. T. (1989) *Nature (London)* **339,** 31–36.
10. Kimura, M. (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge).
11. Alber, T., Dao-Ping, S., Wilson, K., Wozniak, J. A., Cook, S. P. & Matthews, B. W. (1987) *Nature (London)* **330,** 41–46.
12. Perutz, M. F. (1984) *Adv. Protein Chem.* **36,** 213–244.
13. Hoyle, F. & Wickramasinghe, C. (1981) *Evolution from Space: A Theory of Cosmic Creationism* (Simon & Schuster, New York).
14. Yockey, H. P. (1977) *J. Theor. Biol.* **67,** 337–343.
15. Yockey, H. P. (1977) *J. Theor. Biol.* **67,** 345–376.
16. Yockey, H. P. (1977) *J. Theor. Biol.* **67,** 377–398.
17. Viswanadhan, V. N. (1987) *J. Theor. Biol.* **124,** 123–125.
18. Gish, D. T. (1972) *Evolution: The Fossils Say No!* (Creation-Life, San Diego).
19. Wilder-Smith, A. E. (1981) *The Natural Sciences Know Nothing of Evolution* (Master Books, San Diego).
20. Wysong, R. L. (1976) *The Creation–Evolution Controversy* (Inquiry, San Diego).
21. Doolittle, R. F. (1984) in *Proceedings of the 63rd Annual Meeting of the Pacific Division of the American Association for the Advancement of Science*, eds. Aubrey, F. & Thwaites, W. (Am. Assoc. Adv. Sci., Washington, DC), p. 153.
22. Kaplan, R. W. (1974) *Radiat. Environ. Biophys.* **11,** 31–40.
23. Crick, F. H. C. (1981) *Life Itself* (Simon & Schuster, New York).
24. Rao, S. P., Carlstrom, D. E. & Miller, W. G. (1974) *Biochemistry* **13,** 943–952.
25. Almassy, R. J. & Dickerson, R. E. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 2674–2678.
26. Perutz, M. F., Kendrew, J. C. & Watson, H. C. (1965) *J. Mol. Biol.* **13,** 669–678.
27. Grütter, M. G., Weaver, L. H. & Matthews, B. W. (1983) *Nature (London)* **303,** 828–831.
28. Rao, S. T. & Rossmann, M. G. (1973) *J. Mol. Biol.* **76,** 241–256.
29. Richardson, J. (1982) *Adv. Protein Chem.* **34,** 167–339.
30. Lesk, A. M. & Chothia, C. (1980) *J. Mol. Biol.* **136,** 225–270.
31. Ptitsyn, O. B. (1985) *J. Mol. Struct.* **123,** 45–65.