Darin M. Taverna[1]
Richard A. Goldstein[1,2]
[1] *Biophysics Research Division, University of Michigan, Ann Arbor, MI 48109-1055*

[2] *Department of Chemistry, University of Michigan, Ann Arbor, MI 48109-1055*

Received 4 February 1999; accepted 19 July 1999

The Distribution of Structures in Evolving Protein Populations

Abstract: *Proteins exhibit a nonuniform distribution of structures. A number of models have been advanced to explain this observation by considering the distribution of designabilities, that is, the fraction of all sequences that could successfully fold into any particular structure. It has been postulated that more designable structures should be more common, although the exact nature of this relationship has not been addressed. We find that the nonuniform distribution of protein structures found in nature can be explained by the interplay of evolution and population dynamics with the designability distribution. The relative frequency of different structures has a greater-than-linear dependence on designability, making the distribution of observed protein structures more uneven than the distribution of designabilities. The distribution of structures is also affected by additional factors such as the topology of the sequence space and the similarity of other structures.* © 2000 John Wiley & Sons, Inc. Biopoly 53: 1–8, 2000

Keywords: *protein folding; protein designability; protein foldability; lattice models; molecular evolution; population dynamics*

INTRODUCTION

The uneven distribution of protein families among the various possible folds has been repeatedly noted.[1–5] That all folds are not equally represented provides insights into the relationship between sequence and structure, important for applications in both structure prediction and design. A number of theoretical models have been advanced to explain this uneven distribution by considering protein "designability," the relative number of sequences that would result in the formation of a specific structure.[6–14] For instance, Govindarajan and Goldstein made a theoretical connection between how much a given structure could be optimized for folding and the number of sequences that would successfully fold into that structure.[10,11] According to this work, those structures that were the most optimizable could be formed by many sequences

Correspondence to: Richard A. Goldstein, Department of Chemistry, University of Michigan, Ann Arbor, MI 48109-1055; email: richardg@umich.edu

Contract grant sponsor: NIH and NSF
Contract grant number: LM05770 and GM08270 (NIH) and BIR9512955 (NSF)

CCC 0006-3525/00/010001-08

that are far from optimal, while the less optimizable structures could only be formed by the few sequences with close-to-optimal interactions.

While it seems reasonable to presume that the more designable structures should be overrepresented among observed proteins, the exact nature of the relationship between designability and frequency has not been adequately addressed. The current distribution of protein structures is the result of the long period of evolution, inherently a dynamic, nonequilibrium process. In addition, evolution occurs in populations; the role of the population has been shown to have a dramatic impact on the evolutionary process of RNA.[15] Designability, however, represents a static property of individual protein structures. What is the relationship between the relative rate of occurrence of the various possible folds and the relative designabilities? What other factors are important? How do the dynamics of population-based evolution interact with factors such as designability and structural similarity to affect the structure's frequency in nature?

We consider how the results of our previously developed computational model of protein designability are affected by the dynamical aspects of population evolution. Using simulations with lattice proteins, we consider the differences between selecting sequences at random, the random walks of a single protein, and the dynamics of a population of proteins acted upon by random mutations, reproduction, and death. We find that population effects cause highly designable structures to be even more overrepresented, exaggerating the nonuniform distribution of designabilities. Furthermore, structures with similar designabilities can have significantly different occupancies. In particular, populations are enhanced with structures that have similarities to other overrepresented structures.

METHODS

The Model

We consider a highly simplified representation of evolving proteins. Our model proteins consist of chains of $n = 25$ monomers, confined to a 5×5 two-dimensional maximally compact square lattices, with each monomer located at one lattice point. This provides us with 1081 possible conformations represented by the 1081 self-avoiding walks on this lattice, neglecting structures related by rotation, reflection, or inversion.

We assume that the energies of any sequence in conformation k is given by a simple contact energy of the form:

$$E = \sum_{i<j} \gamma(\mathcal{A}_i \mathcal{A}_j) \Delta_{ij}^k \qquad (1)$$

Here, Δ_{ij}^k is equal to 1 if residues i and j are not covalently connected but are on adjacent lattice sites in conformation k, and $\gamma(\mathcal{A}_i \mathcal{A}_j)$ is the contact energy between amino acids $\mathcal{A}_i$ at location i and $\mathcal{A}_j$ at location j in the sequence. We use the contact energies derived by Miyazawa and Jernigan based on a statistical analysis of the database of known proteins that implicitly includes the effect of interactions of the protein with the solvent.[16] There are 132 pairs of residues that can possibly come into contact, with 16 of these contacts present in any given structure.

In nature, not all of the sequences possible represent viable proteins, a characteristic we must include in our model. One universal property shared by essentially all proteins is that they need to fold into regular, compact shapes rapidly enough to avoid proteolysis and aggregation. The ability of various sequences to fold is obviously a constraint of great importance in considering protein evolution. In order to model the requirements placed on biologically viable proteins by their need to fold, we use the approach developed by Wolynes and co-workers who borrowed ideas from the physics of spin glasses.[17]

In particular, Wolynes and co-workers concluded based upon a particularly simple description of the energy landscape that the ability of a protein to fold is a function of the protein's "foldability" $\mathcal{F}$, defined as $\mathcal{F} = \Delta/\Gamma$, where Δ measures the difference of the free energy of the native state with respect to the average of the ensemble of random states, and Γ is the standard deviation of free energies of the random ensemble.[18,19] Based on the results of lattice simulations, we assume that a sequence will be able to fold if the foldability $\mathcal{F}$ is larger than a critical value $\mathcal{F}_{\text{crit}}$.[20–23] All sequences with a foldability above the $\mathcal{F}_{\text{crit}}$ value are considered to be equally fit. The value of $\mathcal{F}_{\text{crit}}$ might be a function of the aggregation rate and the concentration of proteolytes, and could presumably be altered during the process of evolution. In this work, we consider it a parameter that can be adjusted to simulate differing degrees of selective pressure. Given a protein sequence, we can calculate the energy of all possible conformations and find the native conformation, assumed to be the state of lowest energy,[24] and the corresponding foldability $\mathcal{F}$. We can then compare the foldability of that sequence with the assumed value of $\mathcal{F}_{\text{crit}}$ to see if it corresponds to a viable protein. While more sophisticated models of protein folding have been proposed,[25–29] the foldability criterion provides a rapidly computable measure of folding ability, making the population simulations described below computationally tractable. Previous investigators have performed evolutionary analysis based on minimizing folding time in Monte Carlo simulations; this, however, assumes an a priori selective advantage for faster folding rates and requires extensive computational time.[30] Other investigators have considered different thermodynamic measures to be indicative of folding ability[31]; it can be shown that these measures are highly correlated to foldability.[32]

Four evolutionary models were explored and contrasted. The first model ignored the dynamical aspects of evolution and simply computed structure designability by choosing

approximately one billion sequences at random, evaluating their viability, and measuring the resulting distribution of ground-state structures. This provided us with the designability $\mathcal{V}_k$ of each structure k, defined as the proportion of all viable sequences that formed into that structure.

The second and third methods attempted to capture the dynamic aspects of evolution by considering the evolutionary trajectory of a single sequence diffusing through the space of viable protein sequences. Starting from an initial sequence chosen at random from among all of the viable sequences (that is, with $\mathcal{F} > \mathcal{F}_{\text{crit}}$), amino acids were randomly mutated with the number of mutations chosen from a Poisson distribution with an average of one mutation per sequence per generation. Generations where no mutations occurred were not counted. The foldability of the new sequence was calculated; if the foldability was greater than $\mathcal{F}_{\text{crit}}$, the mutation was accepted. For the second model, if the resulting foldability was less than $\mathcal{F}_{\text{crit}}$, the mutation was rejected and the original sequence retained. This is analogous to random-walk model in which the particle has average zero velocity when a boundary is encountered. In the third model, mutations were attempted until a new viable sequence was found. This is defined as a myopic walk. Note that these later two models are identical when $\mathcal{F}_{\text{crit}} = 0$ and all sequences are viable. In either case, the evolving sequence and corresponding ground state were recorded for fifteen million generations.

The fourth model of evolution explored how the dynamics changed when a population of proteins was allowed to evolve, based on the reactor flow model of Eigen.[33] An initial population of $N = 500$ identical sequences of length n was selected containing a total of $N \times n$ residues. For all subsequent generations, each residue in every protein sequence was chosen with probability $1/n$ to be mutated to another random residue. The distribution of mutations per protein as well as the total number of mutations in the population followed a Poisson distribution, with an average of one mutation per sequence per generation. The foldability of all of the resulting sequences were calculated, and the N' sequences with $\mathcal{F} > \mathcal{F}_{\text{crit}}$ were considered viable and capable of reproducing. The mutational death rate represented the fraction of sequences whose mutation led to a subcritical fitness. The next generation of N sequences was chosen from the N' surviving sequences randomly with replacement, representing the stochastic process of reproduction. The number of offspring from any viable sequence could again be represented with a Poisson distribution, with mean N/N'. The population's dynamics were allowed to equilibrate for 50,000 to 200,000 generations, depending upon the value of $\mathcal{F}_{\text{crit}}$, and then the members of the population were recorded every subsequent 100 generations for approximately another 200,000 generations. Such evolving populations have been modeled by a number of investigators, who observed the presence of clusters of sequences that diffuse through the sequence space.[15] The clustering is induced by the reproduction process, and by the high rate of extinction of individual sequences that are not similar to others in the population.

For the latter three methods, we compared the relationship between *occupancy* $\mathbb{O}_k$, defined as the fraction of all evolutionarily derived sequences folding into a structure k after preequilibration, with the *designability* $\mathcal{V}_k$, the proportion of randomly sampled, viable sequences that formed into a structure k. For all four models, three values of $\mathcal{F}_{\text{crit}}$ were used, with $\mathcal{F}_{\text{crit}} = 0$ (no foldability requirement), 3.5, and 4.5. Based on the results from random sequences, the proportion of all sequences that remain viable for these values of $\mathcal{F}_{\text{crit}}$ is 100, 14.3, and 0.15%, respectively. For the evolution runs, three random initial sequences were used, all with $\mathcal{F}_{\text{crit}} > 4.5$. Our $\mathcal{F}_{\text{crit}} = 0$, single-sequence evolution simulations are denoted by $\mathbb{O}_k^s$, whereas for $\mathcal{F}_{\text{crit}} > 0$ we differentiate the results for the two different types of walks with $\mathbb{O}_k^{s(r)}$ (for random walks) and $\mathbb{O}_k^{s(m)}$ (for myopic walks). The occupancies for the population simulations are denoted by $\mathbb{O}_k^p$.

RESULTS

Sensitivity to Initial Conditions

Each of the runs were repeated for three initial sequences. The relative occupancies of the various structures were largely independent of the choice of initial sequence for $\mathcal{F}_{\text{crit}} = 0$ and $\mathcal{F}_{\text{crit}} = 3.5$ (average correlation coefficients between the occupancies for the different runs were 1.00 and 0.98 for the single-sequence and population simulations at $\mathcal{F}_{\text{crit}} = 0$, respectively, and 1.00, 1.00, and 0.97 for the random walk, myopic walk, and population simulations at $\mathcal{F}_{\text{crit}} = 3.5$), indicating the simulations had sufficient time to adequately explore the sequence space and reach steady state. In contrast, the relative occupancies were highly dependent on the initial sequence for $\mathcal{F}_{\text{crit}} = 4.5$, indicating insufficient sampling. This finding agrees with earlier work in which we demonstrated that larger values of $\mathcal{F}_{\text{crit}}$ induce glassy behavior, and the evolutionary dynamics become slow, nonexponential, and nonself-averaging.[34,35] Under these conditions, the evolutionary dynamics become largely confined to "neutral nets" where the sequence and interactions changes while retaining the initial structure. For these reasons, we concentrate for the rest of the paper on results obtained for the two smaller values of $\mathcal{F}_{\text{crit}}$.

Distribution of Designabilities and Occupancies

As previous investigators have noted, we observe a broad distribution of the designability, $\mathcal{V}_k$. As shown in Figure 1 and as demonstrated previously, this distribution becomes even broader when a foldability requirement is imposed.[11,36] Figures 1 and 2 show

how $\mathbb{O}_k^{s(r)}$, the relative occupancy of various structures during the random walk of a single sequence, closely follows the designability distribution for $\mathcal{F}_{\text{crit}} = 0$ as well as $\mathcal{F}_{\text{crit}} = 3.5$.

The results displayed in Figures 1 and 3 demonstrate that the relative occupancies of the various structures during population evolution match the single-sequence occupancies and structure designabilities in the case of $\mathcal{F}_{\text{crit}} = 0$ when all sequences are considered viable. In contrast, the distribution of relative occupancies is significantly broader than the distribution of designabilities when $\mathcal{F}_{\text{crit}} = 3.5$, with many structures with near-zero occupancies and some with extremely high occupancies. As is clear from the insert to Figure 3b, there is a more-than-linear dependence between these quantities, with the population occupancies roughly proportional to the designability raised to the 1.45 power. An exponential fit was clearly inadequate (data not shown). As a result, the highly designable structures are even more overrepresented and the lesser-designable structures less frequent than in cases where population effects are neglected.

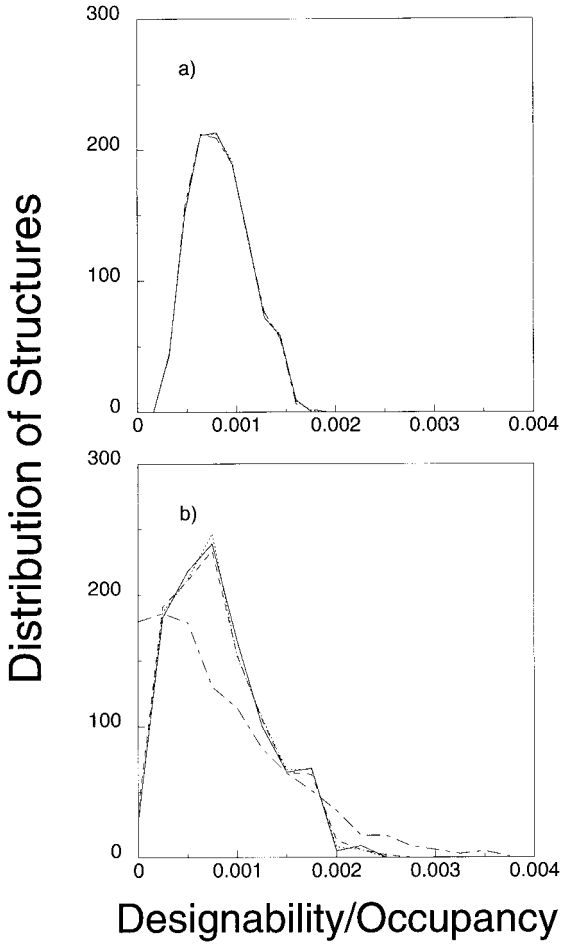

FIGURE 1 (a) Distribution of the designabilities $\mathcal{V}_k$ (——), single-sequence occupancies $\mathbb{O}_k^s$ (----), and population occupancies $\mathbb{O}_k^p$ (–––) for $\mathcal{F}_{\text{crit}} = 0.0$. (b) $\mathcal{V}_k$ (——), random walk occupancies $\mathbb{O}_k^{s(r)}$ (----), myopic-walk occupancies $\mathbb{O}_k^{s(m)}$ (——), and $\mathbb{O}_k^p$ (–––) for $\mathcal{F}_{\text{crit}} = 3.5$.

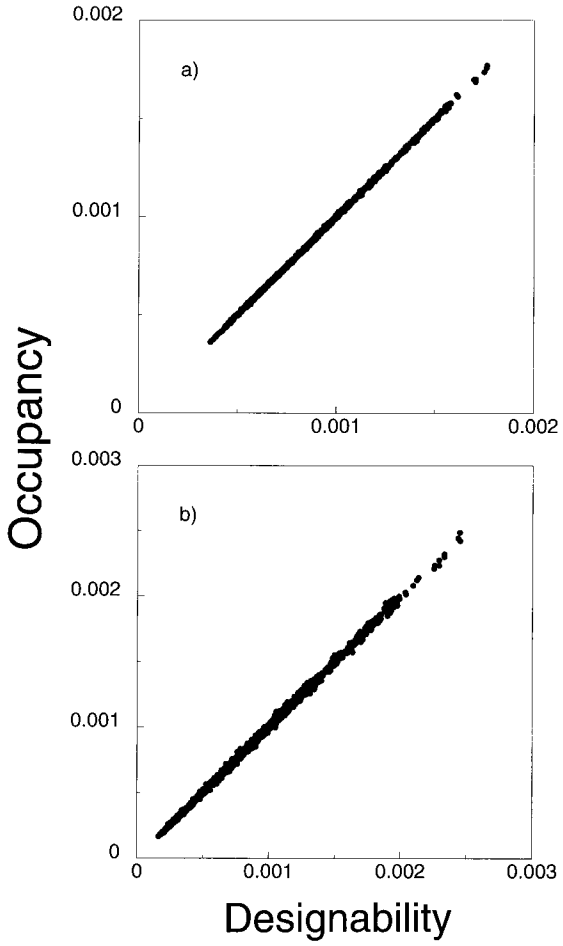

FIGURE 2 Scatter plots displaying corresponding designability values ($\mathcal{V}_k$) and random-walk occupancy values [$\mathbb{O}_k^{s(r)}$] for identical structures, for $\mathcal{F}_{\text{crit}} = 0.0$ (a) and for $\mathcal{F}_{\text{crit}} = 3.5$ (b).

One useful measure of the evenness of the distribution of designabilities and occupancies among the various possible structures is the effective population size Ω_{eff}, defined as

$$\Omega_{\text{eff}} = \frac{1}{\sum_k P_k^2} \tag{2}$$

where P_k is the designability or occupancy of structure k. If these quantities are distributed evenly among Ω structures, then $\Omega_{\text{eff}} = \Omega$. Uneven distributions will reduce the value of Ω_{eff}. The effective population sizes for $\mathcal{F}_{\text{crit}} = 0$ as computed using the designabilities, single-sequence evolutionary trajectories, and

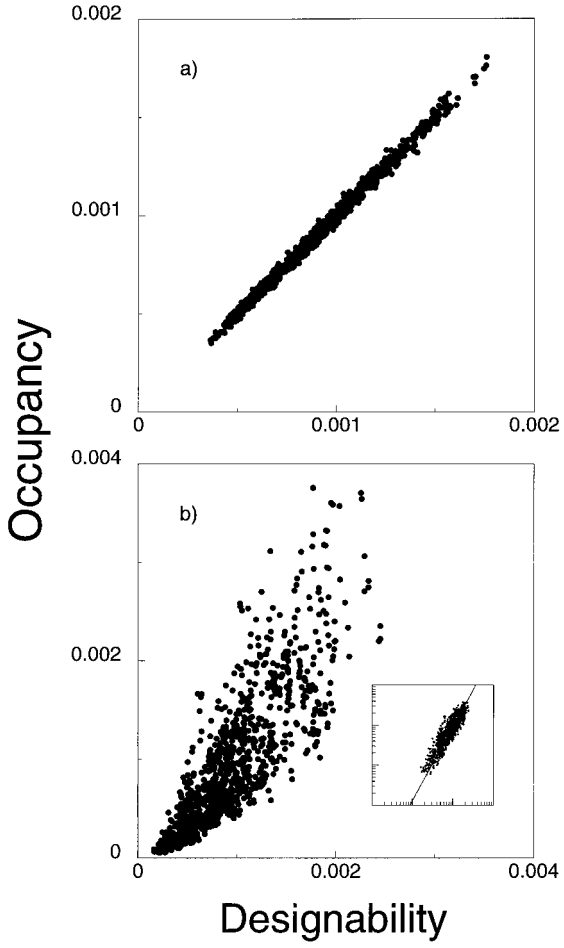

FIGURE 3 Scatter plots displaying corresponding designability values ($\mathcal{V}_k$) and population-simulation occupancy values ($\mathcal{O}_k^p$) for identical structures, for $\mathcal{F}_{\text{crit}} = 0.0$ (a) and for $\mathcal{F}_{\text{crit}} = 3.5$ (b). Inset to (b) shows the corresponding $\mathcal{V}_k$ and $\mathcal{O}_k^p$ values in a log–log plot. Also shown is a best-fit line with slope 1.45.

population simulations, are 985 ($\mathcal{V}_k$), 985 ($\mathcal{O}_k^s$), and 984 ($\mathcal{O}_k^p$), respectively, within the random scatter of individual runs. The effective population sizes for the corresponding runs of $\mathcal{F}_{\text{crit}} = 3.5$ are 863 ($\mathcal{V}_k$), 864 [$\mathcal{O}_k^{s(r)}$], 852 [$\mathcal{O}_k^{s(m)}$], and 683 ($\mathcal{O}_k^p$), respectively, highlighting the similarities between the single-sequence occupancies and the designabilities, the broader distribution of designabilities for higher $\mathcal{F}_{\text{crit}}$ values, and the more extreme distribution of occupancies during the population simulations for $\mathcal{F}_{\text{crit}} = 3.5$.

DISCUSSION

Population Dynamics Have an Effect on Observed Occupancies

We would expect the distribution of the different structures' occupancies to mirror the distribution of their designabilities as long as the evolutionary dynamics sample the space of possible sequences in a random and unbiased way. Unsurprisingly, the evolutionary trajectories of single sequences undergoing random walks fulfill this requirement, providing the simulations are sufficiently long for adequate sampling to be achieved. This requirement is similarly fulfilled during population evolution when $\mathcal{F}_{\text{crit}} = 0$ and all sequences represent viable proteins. In these cases, all sequences have an equal probability of being sampled, and so the relative occupancies simply reflect the relative degeneracies of the mapping of sequence to structure. The situation is different during population evolution for nonzero values of $\mathcal{F}_{\text{crit}}$, when a large fraction of the sequences are nonviable. There is an appreciable probability that any sequence will mutate to a new sequence with $\mathcal{F} < \mathcal{F}_{\text{crit}}$, die, and not contribute to the next generation. When the probability of death is not uniform for all possible structures, the distribution of occupancies changes significantly.

Survival Rates Cause Occupancies to Vary More Strongly Than Designability

Even modest differences in survival rates can have an observable effect on population dynamics (see, for example, Ref. 37). In Figure 3b we show that the relationship between occupancy and designability has a greater-than-linear, power law dependence. This represents the strong dependence of occupancy on survival rate. At $\mathcal{F}_{\text{crit}} = 3.5$, the 10% of our structures with the lowest population occupancies have an 18% higher death rate than structures with occupancies in the top 10%. Figure 4a shows the average survival rate of a mutating sequence, defined as the proportion of all mutated sequences that have $\mathcal{F} > \mathcal{F}_{\text{crit}}$, as a function of the foldability of the sequence prior to mutation. We note that the sequences resulting from the mutation of sequences with high foldabilities are more viable that those from sequences with lower foldabilities. Figure 4b demonstrates that highly designable structures contain more highly foldable sequences than structures with lower designabilities. Therefore, on average, the less designable structures have a higher rate of mutating to an nonviable sequence. Conversely, as highly designable structures are more robust to death via lethal mutation, these structures become more overrepresented in the population. Such mutational robustness due to well-defined ground states has also been noted for RNA structures.[38]

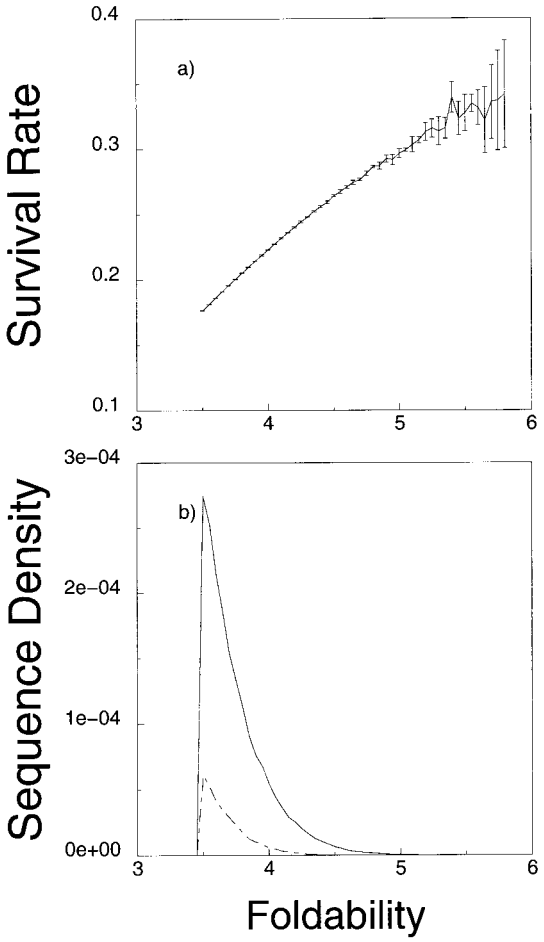

FIGURE 4 (a) Average survival rate of random sequences mutated with $\mathcal{F}_{\text{crit}} = 3.5$ as a function of the initial foldability. (b) Distribution of foldabilities for random sequences for a structure in the top 10% of designable structures (——) and a structure in the bottom 10% of designable structures (– – –). The area under each curve represents the designability for that particular structure.

These Results Cannot be Fully Explained on the Basis of Connectivity of the Sequence Space

We included the myopic walk to see if the occupancies observed in the population dynamics could be explained by the connectivity of the sequence space. Single sequence myopic walks, where there is a successful mutation at each generation, are sensitive to this connectivity in a way that random walks are not. As shown in Figure 1b, the distribution of occupancies for a myopic walk closely matches those of a random walk. Although we note that the effective population size for $\mathbb{O}_k^{s(m)}$ is lower than that of $\mathcal{V}_k$ (indicating a slightly broader distribution of occupancies), this change is small relative to the effect of including population dynamics. This indicates that the changes observed with the inclusion of population dynamics cannot only be the result of sequence space connectivity.

Other Factors Are Important in Determining Population Occupancies

Figure 3b also shows that structures with similar designabilities may have occupancies that differ by a relatively large factor. The scatter is present for two reasons. First, a structure's designability and average death rate are imperfectly correlated, suggesting that the death rate may depend on the specific topology of the sequence space. Occupancy is very sensitive to small differences in death rates. Second, there is a small yet significant population flux between different structures. Just as sequence distribution within a structure is important, the viability of nearby sequences in different structures is also significant. This flux actually reduces the distribution of occupancies, as structures with high death rates and low occupancies tend to have a net population influx, while structures with high occupancies have a net population outflow. The magnitude of this flux depends on the presence of alternative structures that are nearby in sequence space, so that a structure with low designability surrounded by many neighboring structures with high designabilities will show a higher occupancy than another structure with similarly low designabilities without a similar number of neighbors. As has been observed with simulations of RNA populations, it is not necessarily the most fit sequence that become the most populated. Sequences surrounded by especially fit neighbors can have larger-than-expected populations.[15] Again, the result is a larger scatter in the relationship between designability and occupancy.

The population flux effect is reduced because of the correlation between a structure's designability and its similarity to other possible structures. In earlier work that considered the maximum foldability possible for each structure, we demonstrated that highly optimizable, and thus highly designable, structures would be less likely to be similar to other highly designable structures.[34] This is because the most designable structures would have many contacts between residues far apart in the sequence, which impose strong constraints on the possible structures. In contrast, structures with low designabilities would have many contacts between residues near each other in the sequence, contacts that can be shared with other structures. An inverted form of this argument was advanced by Tang and co-workers, who concluded that having few similar possible structures would in

itself increase the designability.[39] As a result, those structures with many structural neighbors generally have low designabilities and have neighbors with low designabilities, reducing the opportunity for substantial redistribution of population. Conversely, those structures with higher designabilities and thus higher occupancies tend to have fewer structural neighbors, and again less opportunity for population flux.

We have assumed that the proteins remain viable as long as the foldability $\mathcal{F}$ remains larger than the critical foldability $\mathcal{F}_{\text{crit}}$, even if the ground-state structure changes. This population flux is essential to the simulation in order to provide sampling of the space of possible structures. It is unclear how much this assumption reflects the situation for biological proteins, in that they must fulfill functional constraints that often rely on maintaining a particular geometrical arrangement of specific functional residues. On the other hand, native-state changes during the evolutionary simulation described in this paper tend to be largely conservative, as small to modest changes in the sequence tend to result in evolution to highly similar structures, conserving many local configurations. Finally, allowing structural changes during the simulation may be more appropriate in modeling the beginning stages of protein evolution when the distribution of proteins among the possible folds was being determined, as at that time sequences may not have been as "finely tuned" and catalysis may have taken advantage of more generic aspects of the protein structure and amino acid interactions.

Highly Designable Structures Are More Likely to be Observed as a Direct Result of Population Evolution

The overrepresentation of certain structural folds is one of the more striking aspects of observed protein structures. Our results indicate we could interpret this as resulting from both convergent and divergent evolution. We expect to see convergent evolution as it would be more likely for a sequence to evolve into a highly designable fold than into a fold with lesser designability. This is clearly evident in categorization systems such as the SCOP database, where proteins with seemingly no evolutionary relationship share similar structures.[4] We would also expect to observe divergent evolution because the greater available sequence space and smaller death rate of highly designable sequences would give more flexibility to the evolution of proteins with novel functions. An example of this is demonstrated in yeast by the case of the Cdc25 fold appearing in three evolutionarily related proteins with quite different functions; rhodanese, the Map Kinase phosphatase noncatalytic domain, and the arsenate resistance protein ACR2.[40]

We would like to thank Nicolas Buchler and Sridhar Govindarajan for helpful comments and Todd Raeker for computational assistance. Financial support was provided by NIH grants LM05770 and GM08270, and NSF shared-equipment grant BIR9512955.

REFERENCES

1. Levitt, M.; Chothia, C. Nature (London) 1976, 261, 552–557.
2. Chothia, C. Nature (London) 1992, 357, 543–544.
3. Orengo, C. A.; Jones, D. T.; Thornton, J. M. Nature (London) 1994, 372, 631–634.
4. Murzin, A. G.; Brenner, S. E.; Hubbard, T. J. P.; Chothia, C. J Mol Biol 1995, 247, 536–540.
5. Govindarajan, S.; Recabarren, R.; Goldstein, R. A. Proteins 1999, 35, 408–414.
6. Finkelstein, A. V.; Ptitsyn, O. B. Prog Biophys Mol Biol 1987, 50, 171–190.
7. Lipman, D. J.; Wilbur, W. J. Proc R Soc Lond (Biol) 1991, 245, 7–11.
8. Finkelstein, A. V.; Gutin, A. M.; Badretdinov, A. Y. FEBS Lett 1993, 325, 23–28.
9. Finkelstein, A. V.; Gutin, A. M.; Badretdinov, A. Y. Subcell Biochem 1995, 24, 1–26.
10. Govindarajan, S.; Goldstein, R. A. Biopolymers 1995, 36, 43–51.
11. Govindarajan, S.; Goldstein, R. A. Proc Natl Acad Sci USA 1996, 93, 3341–3345.
12. Li, H.; Helling, R.; Tang, C.; Wingreen, N. Science 1996, 273, 666–669.
13. Bornberg-Bauer, E. Biophys J 1997, 73, 2393–2403.
14. Shakhnovich, E. I. Folding Design 1998, 3, R45–R58.
15. Schuster, P.; Stadler, P. F. Comput Chem 1994, 3, 295–324.
16. Miyazawa, S.; Jernigan, R. L. Macromolecules 1985, 18, 534–552.
17. Bryngelson, J. D.; Wolynes, P. G. Proc Natl Acad Sci USA 1987, 84, 7524–7528.
18. Goldstein, R. A.; Luthey-Schulten, Z. A.; Wolynes, P. G. Proc Natl Acad Sci USA 1992, 89, 4918–4922.
19. Goldstein, R. A.; Luthey-Schulten, Z. A.; Wolynes, P. G. Proc Natl Acad Sci USA 1992, 89, 9029–9033.
20. Šali, A.; Shakhnovich, E. I.; Karplus, M. J. J Mol Biol 1994, 235, 1614–1636.
21. Šali, A.; Shakhnovich, E. I.; Karplus, M. J. Nature (London) 1994, 369, 248–251.
22. Abkevich, V. I.; Gutin, A. M.; Shakhnovich, E. I. J Chem Phys 1994, 101, 6052–6062.
23. Betancourt, M. R.; Onuchic, J. N. J Chem Phys 1995, 103, 773–787.
24. Govindarajan, S.; Goldstein, R. A. Proc Natl Acad Sci USA 1998, 95, 5545–5549.

25. Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. Annu Rev Phys Chem 1997, 48, 545–600.
26. Plotkin, S. S.; Wang, J.; Wolynes, P. G. J Chem Phys 1997, 106, 2932–2948.
27. Pande, V. S.; Grosberg, A. Y.; Tanaka, T. Biophys J 1997, 73, 3192–3210.
28. Veitshans, T.; Klimov, D.; Thirumalai, D. Folding Design 1997, 2, 1–22.
29. Socci, N. D.; Onuchic, J. N.; Wolynes, P. G. J Chem Phys 1996, 104, 5860–5868.
30. Abkevich, V. I.; Gutin, A. M.; Shakhnovich, E. I. J Mol Biol 1995, 252, 460–471.
31. Mirny, L. A.; Shakhnovich, E. I. Proc Natl Acad Sci USA 1998, 95, 4976–4981.
32. Buchler, N. E. G.; Goldstein, R. A. J Chem Phys 1999, in press.
33. Eigen, M. Naturwissenschaften 1971, 10, 465–523.
34. Govindarajan, S.; Goldstein, R. A. Biopolymers 1997, 42, 427–438.
35. Govindarajan, S.; Goldstein, R. A. Proteins 1997, 29, 461–466.
36. Buchler, N. E. G.; Goldstein, R. A. Proteins Struct Funct Genet 1999, 34, 113–124.
37. Li, W. H. Molecular Evolution; Sunderland: Sinauer, 1997.
38. Wuchty, S.; Fontana, W.; Hofacker, I. L.; Schuster, P. Biopolymers 1999, 49, 145–165.
39. Li, H.; Tang, C.; Wingreen, N. Proc Natl Acad Sci USA 1998, 95, 4987–4990.
40. Fauman, E. B.; Cogswell, J. P.; Lovejoy, B.; Rocque, W. J.; Holmes, W.; Montana, V. G.; Piwnica-Worms, H.; Rink, M. J.; Saper, M. A. Cell 1998, 93, 617–625.