Biophysical Journal Volume 86 May 2004 2758–2764

Stability and the Evolvability of Function in a Model Protein

Jesse D. Bloom,*† Claus O. Wilke,† Frances H. Arnold,‡ and Christoph Adami†§
*Department of Chemistry, †Digital Life Laboratory; ‡Division of Chemistry and Chemical Engineering; and §Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91125

ABSTRACT Functional proteins must fold with some minimal stability to a structure that can perform a biochemical task. Here we use a simple model to investigate the relationship between the stability requirement and the capacity of a protein to evolve the function of binding to a ligand. Although our model contains no built-in tradeoff between stability and function, proteins evolved function more efficiently when the stability requirement was relaxed. Proteins with both high stability and high function evolved more efficiently when the stability requirement was gradually increased than when there was constant selection for high stability. These results show that in our model, the evolution of function is enhanced by allowing proteins to explore sequences corresponding to marginally stable structures, and that it is easier to improve stability while maintaining high function than to improve function while maintaining high stability. Our model also demonstrates that even in the absence of a fundamental biophysical tradeoff between stability and function, the speed with which function can evolve is limited by the stability requirement imposed on the protein.

INTRODUCTION

For nearly all proteins found in nature, there is a unique mapping from the linear protein sequence to a thermodynamically stable three-dimensional native structure, with the mapping determined by the laws of physics (Anfinsen, 1973). However, this unique mapping from sequence to conformation is not a general property of polypeptide sequences, because most randomly generated sequences do not have stable folded structures (Keefe and Szostak, 2001; Davidson et al., 1995). In other words, natural protein sequences exist in the space of foldable sequences, which is but a small subset of the space of all possible sequences. Therefore, evolution must have acted heavily on natural proteins to select those with stable native structures.

Although natural proteins possess stable native structures, the evolutionary fitness of a protein depends not on the stability of the native structure per se, but rather on the stability of this structure being appropriate to allow the protein to perform a function such as catalyzing a chemical reaction or binding to a ligand. Stability is therefore under selection only insofar as it is necessary for biochemical function, and most natural proteins are only marginally stable at their physiologically relevant temperatures (Fersht, 2002).

In protein mutagenesis studies, stability and function can appear to be competing properties, with mutations that increase stability often reducing function (Shoichet et al., 1995; Schreiber et al., 1994), and mutations that improve or alter function often decreasing stability (Wang et al., 2002). However, several lines of evidence demonstrate that high stability and high functionality are not inherently incompatible. In nature, there is a strong correlation between the temperature of an organism's environment and the stability of its proteins, indicating that natural evolution is able to create functional and highly stable proteins if there is sufficiently strong selection pressure (Somero, 1995; Rees and Adams, 1995).

In the laboratory, protein engineers have also demonstrated that natural proteins are not maximally stable by using directed evolution to find mutations that make proteins more stable without sacrificing enzymatic function (Giver et al., 1998; Arnold, 1998; Serrano et al., 1993; Arnold et al., 2001). These results show that high functionality and high stability can coexist, suggesting that the marginal stabilities of natural proteins are due primarily to the simple fact that highly stable sequences are rare (Taverna and Goldstein, 2002), and therefore that most mutations to an evolved protein will decrease its stability. For this reason, proteins will tend to be no more stable than is required by their environment, because any extra stability that confers no further selective advantage will be eliminated by mutations.

Comprehensive experimental examinations of protein evolution are limited by the vast number of possible sequences and the difficulties in rapidly assaying protein properties. However, simple protein models originally developed to study protein folding (Dill et al., 1995; Hinds and Levitt, 1994; Shakhnovich and Gutin, 1993; Socci et al., 1998) provide a useful tool for studying protein evolutionary dynamics (Chan and Bornberg-Bauer, 2002). Although these models are gross oversimplifications of real proteins, their tractability allows for a far more extensive exploration of sequence space than can be done experimentally. Previous studies using model proteins have focused on the evolution of stable structures (Xia and Levitt, 2002; Cui et al., 2002; Bastolla et al., 2000; Taverna and Goldstein, 2000; Tiana et al., 2000; Bornberg-Bauer and Chan, 1999) or fast-folding (Gutin et al., 1995; Mirny et al., 1998) proteins, whereas with few exceptions (Williams et al., 2001; Hirst, 1999) the interplay between the evolution of stability and function has gone unexamined. Here we use a model protein to investigate

Submitted November 7, 2003, and accepted for publication January 12, 2004.
Address reprint requests to Jesse D. Bloom, California Institute of Technology, 210-41, Pasadena, CA 91125. Tel.: 626-354-2565; Fax: 626-568-8743; E-mail: bloom@caltech.edu.
© 2004 by the Biophysical Society

how selection for stability affects the evolution of function. In our model, we describe the function of a protein as its ability to bind to a rigid ligand molecule. The fitness of a protein depends on its ability to perform its function of binding to a ligand, which in turn depends on its ability to fold to a native structure with some minimal stability. We can increase the minimal stability requirement by increasing the temperature parameter, allowing us to explore the relationship between stability and the evolvability of function.

METHODS

The protein model

We use a highly simplified model of a protein consisting of a chain of $N = 18$ monomers on a two-dimensional lattice that we allow to occupy any compact or noncompact conformation.

The monomers can be of 20 types, corresponding to the 20 amino acids. Each monomer on the lattice has four nearest neighbor sites, of which as many as two can be occupied by nonbonded neighboring residues (three in the case of terminal residues). The energy $E(\mathcal{C})$ of a protein conformation $\mathcal{C}$ is the sum of the nearest-neighbor interactions of nonbonded residues,

$$E(\mathcal{C}) = \sum_{i=1}^{N} \sum_{j=1}^{i-2} C_{ij}(\mathcal{C}) \times \varepsilon(\mathcal{A}_i, \mathcal{A}_j),$$

where $C_{ij}(\mathcal{C})$ equals one if residues i and j are nearest neighbors in conformation $\mathcal{C}$ and zero otherwise, and $\epsilon(\mathcal{A}_i, \mathcal{A}_j)$ is the interaction energy between residue types $\mathcal{A}_i$ and $\mathcal{A}_j$. The interaction energies $\epsilon(\mathcal{A}_i, \mathcal{A}_j)$ are based on a widely used statistical analysis of real proteins by Miyazawa and Jernigan (1985) (Table 5). All energies are given in reduced units such that one energy unit equals $k_B T$ at room temperature (298 K). Temperatures are given in units such that $T = 1.0$ at room temperature.

Folding the proteins

The native structure and stability of the protein can be determined by finding the lowest energy conformation, $\mathcal{C}_{low}$, and the partition function. Computation of the partition function requires defining a temperature parameter T. This temperature parameter represents the thermodynamic temperature, however, because the model protein interaction energies are independent of temperature, the temperature parameter does not capture behaviors of real proteins that are caused by the temperature dependence of the interaction energies (for example, cold denaturation). To avoid confusion, we refer to T as the temperature parameter rather than as the temperature.

The partition function at a temperature parameter of T is:

$$Q(T) = \sum_{\{\mathcal{C}_i\}} \exp[-E(\mathcal{C}_i)/T],$$

where the sum is taken over all conformations $\{\mathcal{C}_i\}$. The free energy of folding $\Delta G_f(T)$ to $\mathcal{C}_{low}$ is then the difference between $E(\mathcal{C}_{low})$ and the free energy of the ensemble of all other conformations,

$$\Delta G_f(T) = E(\mathcal{C}_{low}) + T \ln\{Q(T) - \exp[-E(\mathcal{C}_{low})/T]\}.$$

The fraction of proteins $f(T)$ that are expected to be folded to $\mathcal{C}_{low}$ at equilibrium is given by

$$f(T) = \frac{1}{1+\exp[\Delta G_f(T)/T]}.$$

Exact calculation of $Q(T)$ requires enumeration of all 5.81×10^6 unique conformations corresponding to all of the self-avoiding walks that are not related by symmetry (Rapaport, 1987). Many of these walks have very few contacts, and so make only a small contribution to the partition function. We only explicitly considered the 7.95×10^5 conformations with more than four contacts. The remaining 5.01×10^6 conformations were treated by a crude mean-field model, estimating the partition sum contribution of all conformations with n contacts ($0 \le n \le 4$) as

$$Q_n(T) = \exp\left(\frac{-n\langle\epsilon\rangle}{T}\right) \times \exp\left(\frac{-n^2\sigma_\epsilon^2}{2T^2}\right) \times \mathcal{N}(n),$$

where $\langle\epsilon\rangle$ is the average residue-residue contact energy for the given protein sequence assuming any residue is equally likely to be in contact with any other nonadjacent residue, σ_ϵ^2 is the variance in the residue-residue contact energy, and $\mathcal{N}(n)$ is the number of conformations with n contacts. This approximation introduces only a very small error—a test of 10^3 random sequences at $T = 1.0$ showed that the root-mean-square error and maximum differences between the approximate and exact values of $\Delta G_f(T)$ were 1.6×10^{-4} and 2.8×10^{-3}, respectively. This error had no effect on the evolutionary trajectories, because running a sample trajectory with and without the approximation led to identical results. Folding a protein took roughly 0.3 s on a 2-GHz processor.

Modeling the protein function

We introduce the concept of function by considering the binding of a ligand to the protein, an idea that to our knowledge was first introduced by Miller and Dill (1997). We define the function of a protein as its ability to bind to a rigid ligand when the protein is in its lowest energy conformation $\mathcal{C}_{low}$. The binding energy $BE(\mathcal{C}_{low})$ is the interaction energy between the folded protein and the rigid ligand in the lowest energy binding position, found by searching all translational and rotational positions of the ligand relative to the protein. Fig. 1 shows the ligand used in the simulations bound to a protein in its lowest energy conformation.

The fitness function

The fitness $\mathcal{F}(T)$ of a protein at a temperature parameter of T is defined as the negative product of the fraction of the proteins that are folded $f(T)$ times the binding energy $BE(\mathcal{C}_{low})$ of the folded protein to the ligand, so that

$$\begin{aligned}\mathcal{F}(T) &= -f(T) \times BE(\mathcal{C}_{low}) \\ &= -\frac{1}{1+\exp[\Delta G_f(T)/T]} \times BE(\mathcal{C}_{low}).\end{aligned}$$

Because $\mathcal{F}(T)$ has a sigmoidal dependence on $\Delta G_f(T)$, the fitness of an unstructured protein is essentially zero, and a protein gains fitness as it achieves some minimal stability determined by the temperature parameter. Once a protein has achieved the minimal stability, it can only substantially improve its fitness by improving its ligand binding function. The stringency of the stability requirement depends on the temperature parameter T at which the fitness is computed, with higher temperature parameters favoring greater stabilities.

Strictly speaking, the fraction of proteins bound to the ligand also displays a sigmoidal dependence on the product of the ligand binding energy and the fraction of folded proteins. However, because we are only interested in differences in fitnesses rather than their magnitudes, any function that

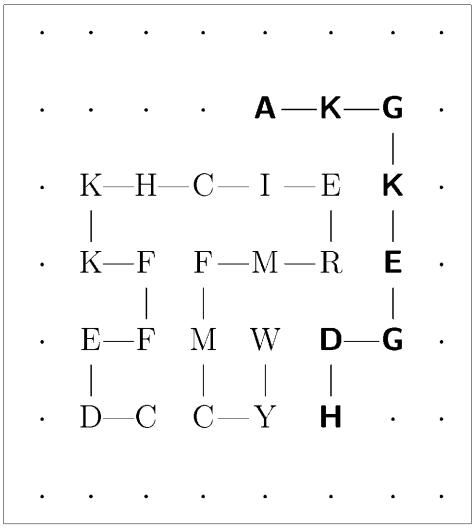

FIGURE 1 Lowest energy conformation of a protein (*left*) bound to the rigid ligand used in the simulations (*right*, shown in *bold*) in the lowest energy binding position. The stability of the protein at a temperature parameter of $T = 1.0$ is $\Delta G_f = -1.04$, and the binding energy of the protein to the ligand is $BE = -17.90$.

monotonically increases with this product will give the same results, and so we choose the simpler functional form defined above.

Evolving the proteins

Each evolutionary replicate began with a population of 99 random sequences. At each generation, the 33 most-fit sequences were selected, and each was used to generate two identical offspring. Random point mutations were made in all 99 resulting proteins with a per-site mutation rate of 3.3×10^{-2}, which corresponds to a per-protein mutation rate of 0.6. The mutated proteins were then refolded, and their fitnesses were calculated.

For the evolutionary trajectories that began with evolved proteins, we first evolved random populations for 250 generations to bind to each of the ligands shown in Fig. 6. The best binding sequences for ligands one, two, and three were FFKFKKFKIFMLKWMKMF, FMGFMIIFFLKFKKFGWF, and MFHVFCHFEWPKPMKCFM, respectively. These sequences were then used for the initial identical populations of 99 proteins for the evolutionary runs, which were otherwise carried out as before.

RESULTS

Rapid evolution of fitness at different stabilities

We carried out 50 evolutionary replicates each beginning with a different initial random population at five different reduced temperature parameters, $T = 0.8$, 0.9, 1.0, 1.1, and 1.2. All replicates exhibited similar evolutionary trajectories, with a rapid gain of fitness in the first few hundred generations followed by only small subsequent increases in fitness. Fig. 2 shows two typical replicates at a temperature parameter of 1.0. Both the stability and the binding energy improved over time, with improvements in binding energy usually associated with temporary small decreases in stability.

Fig. 2 also shows the lowest energy conformations at four different points in the evolutionary trajectory for the two proteins. The overall structures of the proteins were highly conserved, and the stabilities and binding energies adjusted primarily by changes in sequence that preserved these basic structures. This behavior is consistent with the evolution of

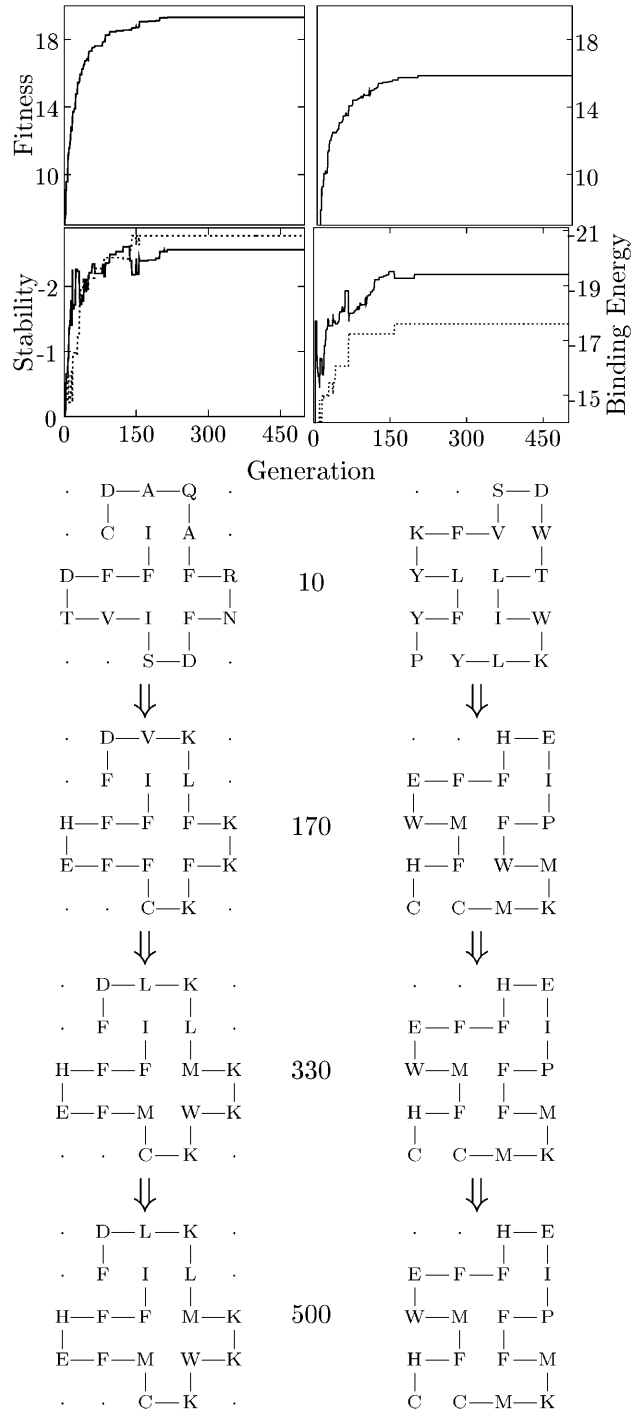

FIGURE 2 Two typical replicates performed at a temperature parameter of 1.0. The plots at the top show the evolution of fitness, of stability (*solid lines*), and binding energy (*dotted lines*) of the most fit member of the population. Structures at bottom are of the most-fit sequence at 10, 170, 330, and 500 generations.

real proteins, which is believed to involve changes in sequence that conserve the structural scaffold of the protein. The ligand binding arrangement shown in Fig. 1 is typical of that of an evolved protein. The most-fit proteins usually evolved to fold to compact conformations with a binding region that fits into the cavity of the ligand.

Proteins evolve ligand binding function more efficiently at lower stability

To determine how selection for stability affected the evolution of ligand binding function, we examined the binding energies achieved after 500 generations of evolution at all five temperature parameters. Fig. 3 shows the distribution of binding energies for all cases. At least a few replicates evolved strong binding proteins at all of the temperature parameters. However, the frequency of evolution of strong binders was much higher at lower temperature parameters, whereas at higher temperature parameters many of the evolutionary trajectories became stuck at weak binding proteins. The binding energy distributions are statistically different for temperature parameters that varied by >0.2 with confidences of >0.95 (Kolmogorov-Smirnov test, D and P values for comparison of $T = 0.8$ and $T = 1.0$, $T = 0.8$ and 1.1, $T = 0.8$ and 1.2, $T = 0.9$ and $T = 1.1$, $T = 0.9$ and 1.2, and $T = 1.0$ and $T = 1.2$ are 0.64 and 7.8×10^{-10}, 0.58 and 3.7×10^{-8}, 0.68 and 4.8×10^{-11}, 0.34 and 4.4×10^{-2}, 0.60 and 1.08×10^{-8}, and 0.52 and 1.2×10^{-6}, respectively; Press et al., 2002).

Table 1 shows the mean binding energies, stabilities, and fitnesses evolved by the proteins. At higher temperature parameters, the proteins evolved higher stability but weaker binding. These results indicate that strong selection for stability inhibits the evolution of strong binding. The evolution of a few strong binders at high temperature parameters shows that it is fundamentally possible to evolve good binding under strong selection for stability, however the results clearly indicate that the statistical likelihood of evolving strong binding is decreased by increasing the selection for stability.

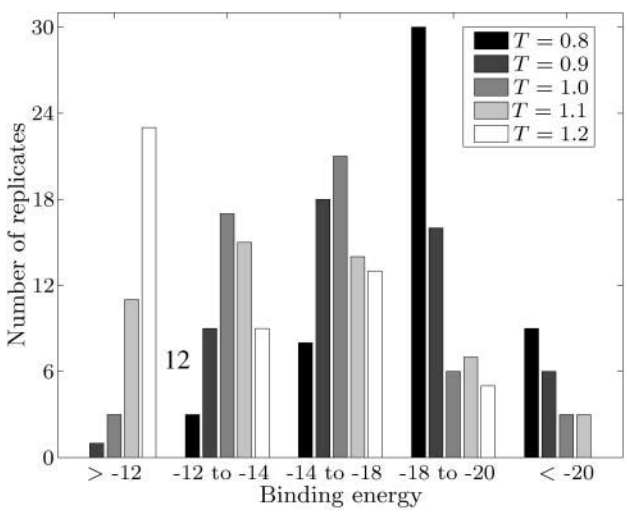

FIGURE 3 Ligand binding function evolves more efficiently at lower temperature parameters. The histogram shows the distribution of the best binding energies after 500 generations of evolution for all 50 replicates at each temperature parameter. Binding energies are of the most-fit member of the population.

TABLE 1 Average binding energies, stabilities, and fitnesses after 500 generations

Temperature	$\langle BE \rangle$	$\langle \Delta G_f \mid_{T=1.0} \rangle$	$\langle \mathcal{F} \mid_{T=1.0} \rangle$
0.8	−18.47 ± 0.32	−1.60 ± 0.06	14.75 ± 0.29
0.9	−16.99 ± 0.37	−1.95 ± 0.07	14.57 ± 0.32
1.0	−15.78 ± 0.37	−2.29 ± 0.07	14.18 ± 0.32
1.1	−15.08 ± 0.42	−2.65 ± 0.09	13.99 ± 0.37
1.2	−13.78 ± 0.38	−2.82 ± 0.09	12.97 ± 0.32

The stabilities and fitnesses are computed at a reference temperature parameter of 1.0 to allow comparison. Mean ± SE. Values are averages of the most-fit member of all 50 populations.

Low-stability evolution as a route to high-stability fitness

Because sequences tend to evolve stronger binding at lower temperature parameters, we speculated that it might be possible to evolve high stability and strong binding proteins more efficiently by performing the initial generations at low temperature parameters. This approach is analogous to simulated annealing, except that in this case the temperature parameter is being increased, because performing the initial generations at a low temperature parameter helps the proteins escape weak binding traps. We tested this idea by performing 50 replicates in which the temperature parameter was set at 0.8 for the first 200 steps, then increased in a linear gradient from 0.8 to 1.2 for 200 steps, and then kept at 1.2 for the final 100 steps. We compared the average final fitness of the proteins after this 500-generation gradient to the final fitness after 500 generations at a constantly high temperature parameter of 1.2 to determine which approach better allows the proteins to optimize the combination of stability and ligand binding function. Fig. 4 shows that the proteins tended to evolve higher fitness with the gradient than with the constantly high temperature parameter. The two distributions are statistically different with high confidence (Kolmogorov-Smirnov test, $D = 0.40$, $P = 4.2 \times 10^{-4}$; Press et al., 2002).

The gradient approach is more efficient at evolving high fitness because it prevents the proteins from becoming trapped in regions of high stability but weak binding by allowing them to first evolve strong binding and then improve their stability. This can be seen in the two typical trajectories shown in Fig. 5. The immediate selection for high stability of the constant temperature parameter approach locks the proteins into stable structures that cannot easily improve their binding energy, whereas gradual selection for

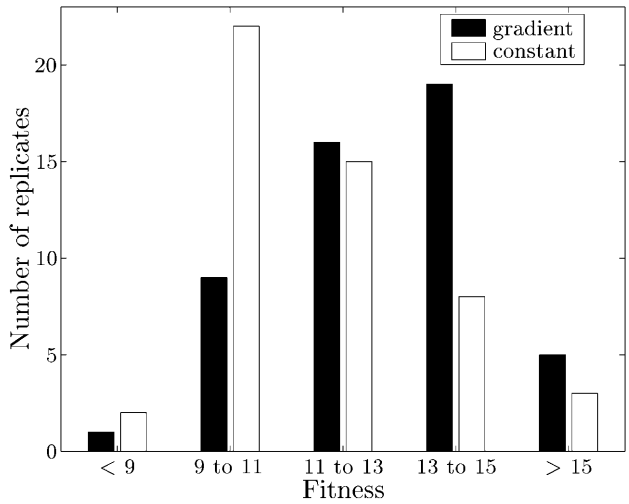

FIGURE 4 High temperature parameter fitness evolves more efficiently with a gradient than with constant selection at a high temperature parameter. The histogram shows the distribution of fitnesses after 500 generations for evolution at a constant temperature parameter of 1.2 and evolution with the gradient from 0.8 to 1.2. Fitnesses are of the most fit member of the population.

stability allows the proteins to first achieve strong binding and then optimize their stability. These results indicate that it is easier to improve stability while maintaining strong binding than it is to improve binding while maintaining high stability.

Evolution from different initial proteins

The results we have described thus far are from evolutionary trajectories that began with random protein sequences. Biological and laboratory protein evolution do not start with random sequences, but instead modify the properties of existing proteins. To test whether the trends we observed depend on the initial populations, we repeated our experiments beginning with proteins that had been evolved to bind to different ligands by first evolving random protein populations for 250 generations to bind to the three ligands shown in Fig. 6. We then used the most fit protein from each of these runs as the beginning sequence for 50 runs of evolution for binding to the original ligand used above (Fig. 1).

All of the trends we found by beginning with random populations were preserved when we started from these evolved proteins. We again found that evolutionary trajectories at lower temperature parameters yielded stronger binding final proteins, and that the gradient was more effective at evolving high-temperature fitness than a constant high temperature parameter (Fig. 7). The different initial populations do lead to different final binding energies and fitnesses, with random populations tending to lead to better values. However, the trends of lower temperature parameter leading to stronger binding and a gradient approach leading to higher fitness hold for all four initial populations.

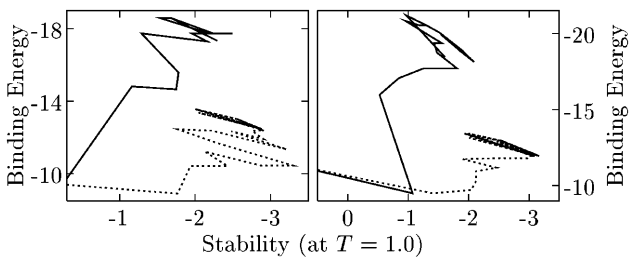

FIGURE 5 Evolution occurs more efficiently when proteins can first evolve strong ligand binding function at low temperature parameters. Solid lines show the stability-function trajectories with a gradient from 0.8 to 1.2, and dotted lines show trajectories with a constant high temperature parameter of 1.2. At left the final fitnesses are 14.01 for the gradient and 10.52 for the constant temperature parameter; at right, the values are 13.92 and 10.60, respectively. Plots are for the most-fit member of the population sampled every 10 generations. Stabilities are computed at a reference temperature parameter of $T = 1.0$ to allow comparison.

DISCUSSION

We have shown that strong selection for stability inhibits the evolution of ligand binding function by the proteins in our model. The ability of a few proteins to evolve strong binding at high temperature parameters shows that there are sequences that exhibit both good stability and strong binding. However, at high temperature parameters, the evolving proteins are more likely to become trapped in regions of sequence space that correspond to highly stable but weakly binding proteins. Presumably this trapping is due to that fact that stronger selection for stability reduces the network of sequences that are essentially neutral because there are fewer highly stable sequences compatible with a given fold (Shakhnovich, 1998; Koehl and Levitt, 2002). Therefore, at high temperature parameters, mutations that increase binding are more likely to lead to an unacceptably large drop in stability. The net effect is a roughening of the fitness landscape that makes it more difficult to escape local optima.

We present a strategy to overcome this problem of the evolutionary trajectories becoming trapped at high stability but weak binding proteins. Performing the initial rounds of evolution at a low temperature parameter decreases the

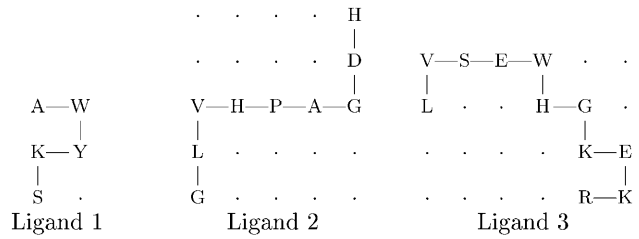

FIGURE 6 Proteins were evolved for 250 generations to bind to these ligands. We then modified the function of these evolved proteins by evolving them to bind to the original ligand shown in Fig. 1.

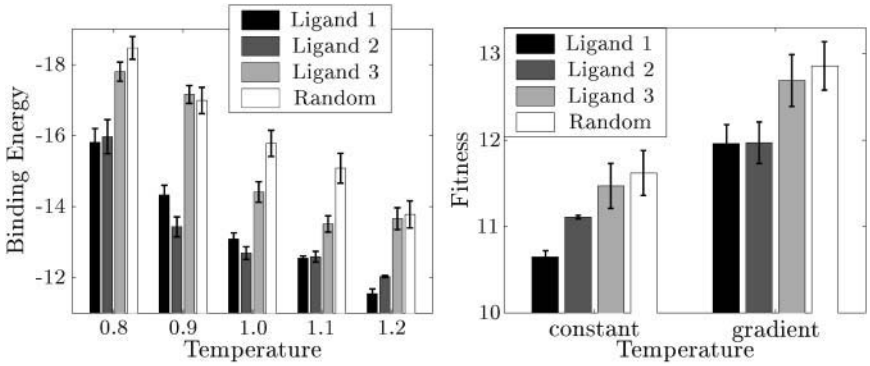

FIGURE 7 Lower temperature parameter evolution leads to stronger binding, and a gradient approach leads to better high-temperature parameter fitness regardless of the initial population. At left are the binding energies for evolution at the indicated temperature parameter with the indicated initial populations. At right are the fitnesses at a temperature parameter of 1.2 with evolution at a constant temperature parameter or with a gradient. Values are the mean ± SE for the most-fit member of the population for all 50 runs after 500 generations.

selection for stability, and so allows the proteins to more easily find strong binding regions of sequence space. The temperature parameter can then be increased, which leads to the selection of more stable sequences. Our results indicate that this approach is more effective for evolving highly stable and strong binding proteins than constant selection for both high stability and strong binding. This strategy takes advantage of the fact that it is easier to maintain strong ligand binding while improving stability than to maintain high stability while improving binding.

Our results fit into the framework of current theories about the distributions of proteins in sequence space that has emerged from other lattice protein studies. These studies have shown that protein structures are coded for by structurally neutral networks spanning many diverse sequences, and that these networks are structured as superfunnels, with the most stable sequences also possessing the most connections in the networks (Bornberg-Bauer and Chan, 1999; Broglia et al., 1999; Bastolla et al., 2000, 1999). Our work suggests that a protein evolves function most effectively when it can freely explore in its structurally neutral network, rather then when it is trapped in a small number of highly stable sequences. Our initial relaxation of the stability requirement facilitates exploration of the structurally neutral network, and once highly functional sequences are found, they can be optimized for stability. Although we do not consider recombination in our current study, other work (Cui et al., 2002; Xia and Levitt, 2002) has shown that whereas structurally neutral networks can easily be explored locally by point mutations, moves between networks or to distant regions of the same network are facilitated by crossover-induced sequence space jumps. Therefore, we suggest that addition of recombination to our evolutionary protocol may further assist in the evolution of function.

The evolution of our model proteins also has strong parallels with real protein evolution. As with real proteins, our model proteins evolve primarily by structurally conservative mutations that tinker with the contacts in a preserved structural scaffold, rather than by mutations that cause wholesale structural changes. The interplay between the evolution of stability and function in our model is also reminiscent of real protein evolution; for example, in the evolution of new function in TEM-1 β-lactamase, gains in function were correlated with drops in stability, followed by gradual regaining of the lost stability (Wang et al., 2002).

Our model points to general trends that are important in both natural and experimental protein evolution, where different structural and functional properties are under different selection pressures. Protein evolution involves concurrent selection for stability and function, and productive mutations must improve one of these properties without excessively damaging the other. Because most mutations to evolved proteins will be deleterious to at least one of these properties, strong selection for both stability and function will limit the number of productive mutations, and so lead to trapping at local fitness optima. Protein evolution therefore occurs most efficiently when the temporary drops in stability associated with gains in function are buffered by mild selection for stability.

We thank Hue Sun Chan, George Somero, Jeff Endelman, and Chris Voigt for their helpful comments on the manuscript.

J.D.B. is supported by a Howard Hughes Medical Institute predoctoral fellowship. C.O.W. and C.A. were supported by the National Science Foundation under contract No. DEB-9981397. Part of this work was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

REFERENCES

Anfinsen, C. B. 1973. Principles that govern the folding of protein chains. *Science*. 181:223–230.

Arnold, F. H. 1998. Enzyme engineering reaches the boiling point. *Proc. Natl. Acad. Sci. USA*. 95:2035–2036.

Arnold, F. H., P. L. Wintrode, K. Miyaszki, and A. Gershenson. 2001. How enzymes adapt: lessons from directed evolution. *Trends Biochem. Sci.* 25:100–106.

Bastolla, U., H. E. Roman, and M. Vendruscolo. 1999. Neutral evolution of model proteins: diffusion in sequence space and overdispersion. *J. Theor. Biol.* 200:49–64.

Bastolla, U., M. Vendruscolo, and H. E. Roman. 2000. Structurally constrained protein evolution: results from a lattice simulation. *Eur. Phys. J. B.* 15:385–397.

Bornberg-Bauer, E., and H. S. Chan. 1999. Modeling evolutionary landscapes: mutational stability, topology, and superfunnels in sequence space. *Proc. Natl. Acad. Sci. USA*. 96:10689–10694.

Broglia, R., G. Tiana, H. Roman, E. Vigezzi, and E. Shakhnovich. 1999. Stability of designed proteins against mutations. *Phys. Rev. Lett.* 82: 4727–4730.

Chan, H. S., and E. Bornberg-Bauer. 2002. Perspectives on protein evolution from simple exact models. *Applied Bioinformatics.* 1:121–144.

Cui, Y., W. H. Wong, E. Bornberg-Bauer, and H. S. Chan. 2002. Recombinatoric exploration of novel folded structures: a heteropolymer-based model of protein evolutionary landscapes. *Proc. Natl. Acad. Sci. USA.* 99:809–814.

Davidson, A. R., K. J. Lumb, and R. T. Sauer. 1995. Cooperatively folded proteins in random sequence libraries. *Nat. Struct. Biol.* 2:856–864.

Dill, K. A., S. Bromberg, K. Yue, K. M. Fiebig, D. P. Yee, P. D. Thomas, and H. S. Chan. 1995. Principles of protein folding: a perspective from simple exact models. *Protein Sci.* 4:561–602.

Fersht, A. R. 2002. Structure and Mechanism in Protein Science. W. H. Freeman and Company, New York.

Giver, L., A. Gershenson, P. O. Freskgard, and F. H. Arnold. 1998. Directed evolution of a thermostable esterase. *Proc. Natl. Acad. Sci. USA.* 95:12809–12813.

Gutin, A. M., V. I. Abkevich, and E. I. Shakhnovich. 1995. Evolution-like selection of fast-folding model proteins. *Proc. Natl. Acad. Sci. USA.* 92:1281–1286.

Hinds, D. A., and M. Levitt. 1994. Exploring conformational space with a simple lattice model for protein-structure. *J. Mol. Biol.* 243:668–682.

Hirst, J. D. 1999. The evolutionary landscape of functional model proteins. *Protein Eng.* 12721–726.

Keefe, A. D., and J. W. Szostak. 2001. Functional proteins from a random-sequence library. *Nature.* 410:715–718.

Koehl, P., and M. Levitt. 2002. Protein topology and stability define the space of allowed sequences. *Proc. Natl. Acad. Sci. USA.* 993:1280–1285.

Miller, D. W., and K. A. Dill. 1997. Ligand binding to proteins: the binding landscape model. *Protein Sci.* 6:2166–2179.

Mirny, L. A., V. I. Abkevich, and E. I. Shakhnovich. 1998. How evolution makes proteins fold quickly. *Proc. Natl. Acad. Sci. USA.* 95:4976–4981.

Miyazawa, S., and R. L. Jernigan. 1985. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules.* 18:534–552.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, editors 2002. Are two distributions different? *In* Numerical Recipes in C. Cambridge University Press, Cambridge, UK. 620–628.

Rapaport, D. C. 1987. Algorithms for lattice statistics. *Comput. Phys. Rep.* 5:265–350.

Rees, D. C., and M. W. W. Adams. 1995. Hyperthermophiles: taking the heat and loving it. *Structure.* 3:251–254.

Schreiber, G., A. M. Buckle, and A. R. Fersht. 1994. Stability and function: two constraints in the evolution of barstar and other proteins. *Structure.* 2:945–951.

Serrano, L., A. G. Day, and A. R. Fersht. 1993. Step-wise mutation of barnase to binase: a procedure for engineering increased stability of proteins and an experimental analysis of the evolution of protein stability. *J. Mol. Biol.* 233:305–312.

Shakhnovich, E. I. 1998. Protein design: a perspective from simple tractable models. *Fold. Des.* 3:R45–R58.

Shakhnovich, E. I., and A. M. Gutin. 1993. Engineering of stable and fast-folding sequences of model proteins. *Proc. Natl. Acad. Sci. USA.* 90: 7195–7199.

Shoichet, B. K., W. A. Baase, R. Kuroki, and B. W. Matthews. 1995. A relationship between protein stability and protein function. *Proc. Natl. Acad. Sci. USA.* 92:452–456.

Socci, N. D., J. N. Onuchic, and P. G. Wolynes. 1998. Protein folding mechanisms and the multidimensional folding funnel. *Proteins.* 32:136–158.

Somero, G. N. 1995. Proteins and temperature. *Annu. Rev. Physiol.* 57: 43–68.

Taverna, D. M., and R. A. Goldstein. 2000. The distribution of structures in evolving protein populations. *Biopolymers.* 53:1–8.

Taverna, D. M., and R. A. Goldstein. 2002. Why are proteins marginally stable? *Proteins.* 46:105–109.

Tiana, G., R. Broglia, and E. Shakhnovich. 2000. Hiking in the energy landscape in sequence space: a bumpy road to good folders. *Proteins.* 39:244–251.

Wang, X., G. Minasov, and B. K. Shoichet. 2002. Evolution of an antibiotic resistance enzyme constrained by stability and activity trade-offs. *J. Mol. Biol.* 320:85–95.

Williams, P. D., D. D. Pollock, and R. A. Goldstein. 2001. Evolution of functionality in lattice proteins. *J. Mol. Graph. Model.* 19:150–156.

Xia, Y., and M. Levitt. 2002. Roles of mutation and recombination in the evolution of protein thermodynamics. *Proc. Natl. Acad. Sci. USA.* 16: 10382–10387.