Beilstein-Institut The Chemical Theatre of Biological Systems, May 24th - 28th, 2004, Bozen, Italy

EVOLUTIONARY PERSPECTIVES ON PROTEIN STABILITY

RICHARD A. GOLDSTEIN

Division of Mathematical Biology, National Institute for Medical Research, Mill Hill, London, NW7 1AA, U.K.
E-Mail: richard.goldstein@nimr.mrc.ac.uk

Received: 6th September 2004 / Published: 22nd July 2005

ABSTRACT

Biochemists, molecular biologists, and biophysicists, when confronted with the exquisite matching of natural organisms to their environment, have generally interpreted the properties of proteins as resulting from adaptation. Conversely, since 1962, evolutionary theory has been emphasizing the stochastic nature of neutral evolution and random drift. It is actually possible to explain many of the seemingly adaptive features of proteins as resulting through neutral evolution. In this paper, we use a simple computational model to demonstrate how marginal stability as well as the robustness of proteins to site mutations can be explained by neutral evolution of populations.

INTRODUCTION

If we were to analyse a computer looking only at its constituent parts, we might be able to characterize the physical properties of the case, the chemical properties of the circuit boards, and the ferromagnetic properties of the rotating disks. We would, however, be unable to understand why the components were the way they are, and thus would have trouble understanding how they worked together to give the collective behaviour observed with computers. This is because we had ignored the *functional* description, the discussion of the role and purpose for each of these components. This functional definition is appropriate because the computer is a result of a design process, where teams of engineers have created the computer as a tool for specific purposes, and the components reflect their use in these tasks. Conversely, it would be quite difficult to analyse the workings of a computer without any information about the physical, chemical, etc., properties of the components.

We could not understand how the various parts of the tasks were organized or implemented without knowing the opportunities, constraints, and limitations of the materials and processes available to the engineers. An understanding of a modern computer then requires both a description of the properties of the components in the context of the roles for which the computer was designed.

There has not been a corresponding 'design process' for living systems, no teams of engineers that have drawn schematics for how the cell should be constructed. Yet there is an overarching biological process that in some ways can resemble engineering design. This is, of course, the process of evolution, where living systems are evaluated based on their ability to survive and reproduce, with the 'winners' continuing into the next round. For this reason, we can ask 'why' questions: what is the 'reason' for the existence of an organ, tissue, cell, enzyme, etc. Understanding how these components have developed so as to contribute to survival and reproduction can give us many insights into their role and purpose, and how they work together to fulfil the required tasks. We cannot neglect, however, the opportunities, constraints, and limitations imposed on the process of evolution based on the physical and chemical properties of the components. By looking at biological systems in this evolutionary context, we can combine these two perspectives: understanding how evolution adapts the various components to the needs of survival and reproduction, while being constrained by the (constantly changing) properties of these components. This is in some ways the essence of the Chemical Theatre of Biological Systems: how the chemical properties of molecular systems determine the biological properties of living systems, while the biological process of evolution underlies the resulting chemical properties of the biological components.

It is important to keep in mind how evolution differs from engineering design. While randomness occurs in engineering work (a designer reads an inspiring article in a magazine found in the train), it is a central and indispensable aspect of evolution. Evolution can only work with the raw materials placed in its hands, the variations that occur at each generation, variations that result from the stochastic chemical reactions that produce site mutations and the random process of recombination, gene conversion, chromosome sorting, etc. All of this is constrained by the need for each intermediate individual to be capable of reproduction, for there to be an acceptable path from one genotype to another, viable at every point in between.

Selection, the mechanism, which governs the number of offspring an individual contributes to the next generation, is also extremely fickle. Favourable mutations can be eliminated, deleterious mutations can become fixed.

It is impossible to look at living systems without being amazed at the efficiency and power of evolution, how exquisitely well adapted such systems are. As a result, evolutionary thinking has often concentrated on the role of 'positive adaptation', the process by which increasingly fit members of the population are generated and selected embodied by the phrase, 'survival of the fittest'. There are dangers to neglecting evolutions' stochastic nature, that we can fall into what Gould calls the 'Panglossian paradigm', analysing how all is for the best in this 'best of all possible worlds' [1]. In contrast to the models that emphasize the role of adaptation, the 'neutral theory' developed in the 1960s [2-4] focuses on the consequences of the random processes of variation and drift. According to this theory, while an adaptive mutation is more likely to be accepted in a population, all but a small fraction of mutations are deleterious or neutral, all with corresponding probabilities of acceptance. As a result, most accepted changes are actually neutral or slightly deleterious. The stochastic nature of these changes does not preclude the derivation of important principles and concepts, any more than the stochastic nature of the motion of gas particles invalidates the ideal gas law. We need to use a different type of thinking, of looking at the evolutionary equivalents for concepts such as entropy. While evolutionary biology has been debating the importance and consequences of adaptation and neutral drift, biochemistry and molecular biology has generally analysed biological systems in terms of a purely adaptionist model. Little has been done looking at the role of neutral drift in determining the nature of biological systems at a molecular level.

In this paper I look at how proteins can be considered in an evolutionary context, in particular, in how much can be explained by the process of neutral evolution. While obviously adaptation occurs, such a perspective offers a number of advantages. Firstly, because neutral drift is always taking place, a demonstration that a characteristic of proteins can result from neutral drift provides evidence that neutral evolution is a *sufficient* explanation, and is in fact a more *parsimonious* explanation, that should only be rejected when it is shown to be clearly inadequate. Secondly, considering neutral drift protects us against the 'Panglossian paradigm'. Thirdly, failure to be able to explain a phenomenon based on neutral evolution highlights those aspects that may require an adaptionist explanation. And lastly, such a perspective helps to remind us that evolution is *not* an optimization procedure.

Populations change and vary according to stochastic, discrete differential equations oblivious to any terms such as 'fitness'. There is no teleological underpinning to biological evolution except that imposed by the observer for their own intellectual convenience, or our psychological need to imagine ourselves at the peak of perfection in *some* space.

Proteins are a convenient choice for such investigations, in that they span a simplified form of the genotype/phenotype divide. Evolution involves the changing of the sequential information contained in the genotype as represented by the composition of biomolecules, generally DNA. Selection, however, acts on the resulting phenotypic traits representing the *interpretation* of the genotype through the process of development. Proteins can be characterized by their amino acid sequence (a close reflection of the genotype), as well as by their physical and chemical properties, a simplified form of phenotype. In this way they can provide insight into the evolution of higher organizational forms. In addition, proteins are important and interesting in their own right. They are involved in essentially every biological process, are basic to understanding all of these processes at a molecular level, and form the targets of most pharmaceutical products. Altered, 'engineered' proteins have the potential to fulfil many different tasks for us. Various aspects of proteins form the basis for many pathological conditions. Many processes involving proteins (such as protein folding) are still poorly understood.

Proteins are still extremely complicated. They are large molecules which involve a number of different types of atoms and functional groups, interacting with solvent, membranes, ions, etc. They have many characteristics, including structural, functional, kinetic and thermodynamic properties. They exist in a complicated biological context. To analyse the general principles of a realistic model of protein evolution is beyond our capabilities. One approach is to use realistic models but to focus our attention to more manageable properties of specific proteins in specific situations. Alternatively, we can use highly simplified models, restricting our inquiries to questions for which these models are appropriate, making all possible connections with experimental observations. We choose the latter approach here. In particular, we represent proteins as two-dimensional lattice models. We focus our attention on the stability of proteins, both to thermal fluctuations and mutations.

MODELS

The model used for these calculations involves a) the model of the evolving proteins, and b) the model used for the evolutionary behaviour, explicitly including population effects.

Models of the Proteins

Figure 1 shows the model of the protein used. Proteins are represented as 25 residue polypeptides confined to a 5 × 5 square lattice, where each residue occupies one lattice point.

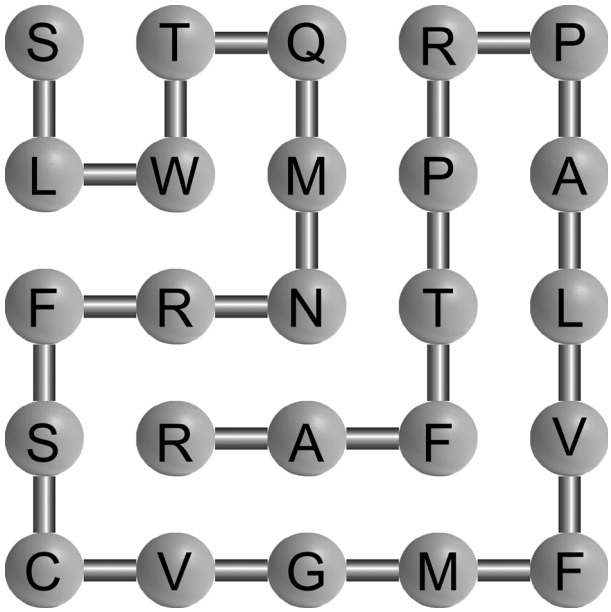

Figure 1. Model of lattice protein.

The 1081 possible conformations for the protein represent all self-avoiding walks through the lattice, not counting rotations, reflections and inversions. The two-dimensional lattice is obviously a highly simplified representation, and would be inappropriate for dynamics studies (where the set of conformations are generally non-ergodic, depending upon the move-set [5]) but this model does allow us to have a reasonable surface-to-volume ratio with a moderately-sized protein.

The energy $E(S,C)$ of a given sequence S in conformation C is a pairwise contact energy, given by

$$E(S,C) = \sum_{<i,j>} \gamma(A_i, A_j)\, u_{i,j} \qquad (1)$$

where $\gamma(A_i,A_j)$ is the interaction energy between the amino acid types located at locations i and j in the sequence (such as between the serine and threonine in the upper-left corner of the protein shown in Fig. 1) and $u_{i,j}$, is 1 if residues i and j are in contact (that is, are not covalently connected but are on nearest neighbouring lattice points), and zero otherwise. We used the values of $\gamma(A_i,A_j)$ calculated by Miyazawa and Jernigan based on their statistical analysis of a database of known protein structures [6]. These interaction terms represent 'potentials of mean force' that implicitly include the interaction of the residues with the solvent.

We assume, based on earlier work, that the conformation of lowest energy represents the native state C_{ns} [7]. The probability P_{ns} that a protein at equilibrium at temperature T would be in this state is

$$P_{\text{ns}} = \frac{\exp\left(E(S, C_{\text{ns}})/kT\right)}{\sum_{C} \exp\left(E(S, C)/kT\right)} \tag{2}$$

where the lower sum is over all conformations. This allows us to compute $\Delta G_{folding}$

$$\Delta G_{\text{folding}} = -kT \log\left(\frac{P_{\text{ns}}}{1 - P_{\text{ns}}}\right) \tag{3}$$

$$= -kT \log\left(\frac{\exp\left(E(S, C_{\text{ns}})/kT\right)}{\sum_{C \neq C_{\text{ns}}} \exp\left(E(S, C_{\text{ns}})/kT\right)}\right) \tag{4}$$

The interaction parameters are in kcal/mol, so all free energies are in these units.

Modelling Population Evolution

Population dynamics form a central element in the evolutionary process. In particular, it is the fact that populations are finite (and much smaller than, for instance, Avogadro's number) that introduces the large stochastic element. For these reasons, we include such dynamics in our simulations. In general, we start with a population of identical protein sequences. Given a specified average mutation rate, we select the number of mutations in each particular generation with the approximate Poisson distribution. These residues as well as the proteins in which they belong are selected at random, and each amino acid is changed at random to one of the other 19. The various properties of the resulting proteins are computed, including the ground state and the thermal stability. We then eliminate 'non-viable' sequences, the sequences that fail to meet some appropriately chosen selection criterion, such as minimum acceptable thermostability. In order to represent the random process or reproduction, we choose new sequences at random from the set of surviving sequences from the previous generation, with replacement, until we have the same number of sequences as we had initially. This represents the population at the next generation. The population size remains fixed during the simulation.

RESULTS

Protein Thermostability

It has been long noted that proteins are marginally stable, with stabilities ($-\Delta G_{folding}$) typically observed of around 10 kcal/mol. The most general explanation for this effect is that proteins have evolved to be marginally stable, that marginally stable proteins are more 'fit', that marginal stability represents an adaptation. A number of different reasons have been given why marginal stability would be advantageous. For instance, protein functionality might require a degree of flexibility incompatible with high stability [8,9] Alternatively, protein flexibility would destabilize ligand binding, allowing greater control and better abilities to modulate the affinity through mutation, post-translational modification, or other slight changes in chemical properties [10-12]. Finally, marginal stability might increase the ability to degrade proteins, an important aspect of cellular control. An alternative explanation for the observed marginal stability is that this involves 'optimization given constraints', that proteins have to fulfil so many other selective criteria involving functionality, rigidity, solubility, etc., that proteins can only increase their stability to a certain point without compromising these other factors. This second explanation is also adaptionist in nature, but emphasizes the more complicated nature of fitness, and that all components to the fitness cannot be independently optimized.

Both of these explanations - that proteins that are marginally stable are more fit, or that stability is optimised given constraints - are adaptionist in nature. What would occur during neutral evolution, where neither of these mechanisms is active? To answer this question, we modelled a population of 3000 sequences, collecting data for 30,000 generations following an initial 30,000 generations for equilibration ('burn-in') [13]. The mutation rate was maintained at a rate of 0.2% mutations per protein per generation. Proteins were considered viable if they were 'adequately' stable, that is, with a ΔG of folding $\Delta G_{folding}$ less than some 'critical' ΔG_{crit}.

The results for $\Delta G_{crit} = 0$ is shown in Fig. 2A. There are many extremely stable sequences that can be obtained through hill-climbing methods, but the evolutionary runs generally result in marginally-stable proteins. In these simulations, there were no constraints placed on the optimization, nor were there any advantages to marginal stability. Proteins are naturally marginally stable because the vast number of viable sequences ($\Delta G_{folding} < \Delta G_{crit}$) is marginally stable ($\Delta G_{folding} \approx \Delta G_{crit}$).

This can be seen by comparison with the distribution observed for random sequences with $\Delta G_{folding} < \Delta G_{crit}$, also shown in Fig. 2A. Again, the vast proportion of such sequences has $\Delta G_{folding} \approx \Delta G_{crit}$. This is because they represent the extreme tail of a broader distribution of random stabilities, as shown in Fig. 2B.

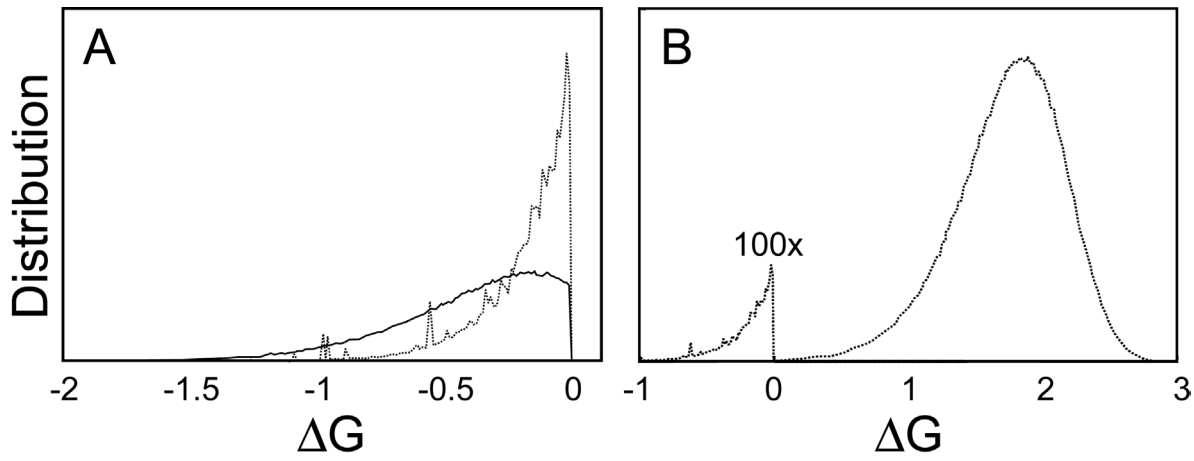

Figure 2. A) Distribution of stabilities observed for the proteins resulting from population dynamics with $\Delta G_{folding} < \Delta G_{crit} = 0$ (solid line), compared with the distribution for randomly selected proteins, also with $\Delta G_{folding} < 0$ (dashed line). **B)** The overall distribution of stabilities of random proteins.

One way to make intuitive sense of these results is to consider the abstract space of all possible sequences, represented in Fig. 3. This space is extremely high-dimensional (as many dimensions as the length of the sequences), but also extremely sparse (only 20 points along each dimension). The vast majority of this space consists of non-viable sequences, sequences that would not fold or would not be stable in any given fold. There are sequences in various regions of the space, shown in grey, which would be able to fold into one of a number of possible structures. One aspect of high dimensional spaces is the surface to volume ratio: the fraction of the volume of a hypersphere of dimension n and radius r occupied by a thin shell of radius δr goes as $\frac{n\,\delta r}{r}$, meaning that for a 100-dimensional space (corresponding to a 100-residue protein), 99% of the sequence volume would be in the outer 1% of the hypersphere. If the region inside the hypersphere represents viable proteins - foldable and stable - and the region outside the hypersphere represents non-viable proteins - unfoldable and unstable - this suggests that the vast majority of viable proteins are barely viable.

This is purely the result of the high dimensionality of the sequence space, assuming that the viability is slowly varying (so regions on the edge of the viable region are barely viable), and that the regions in sequence space corresponding to viable proteins form relatively compact objects.

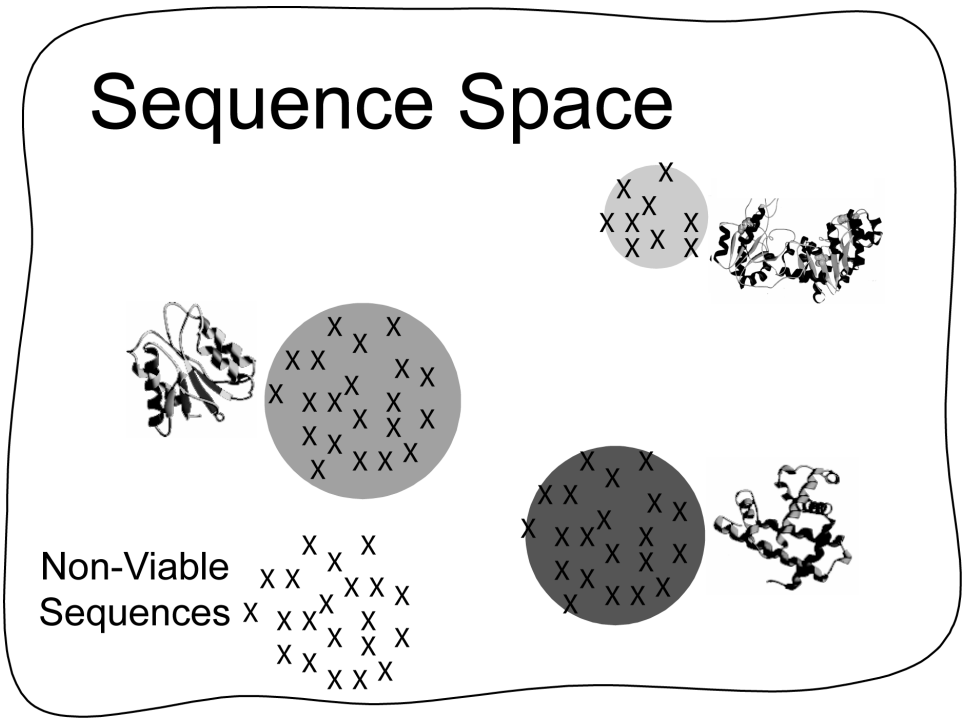

Figure 3. Representation of the space of all possible sequences.

In situations such as these, if the vast majority of viable protein sequences are marginally stable, there is no need to propose other evolutionary mechanisms (such as adaptation) to account for this observation in biological proteins. Neutral evolution is a sufficient, and therefore more parsimonious, explanation. In this case, the tendency for proteins to be marginally stable represents 'sequence entropy': in a random walk, the majority of sequences visited will mirror the location of the majority of random viable sequences. Sequence entropy can drive proteins towards marginal stability.

It has been observed that proteins that have modified in order to increase their thermostability sometimes display lower than native function [14-17]. Does this indicate that proteins have evolved for marginal stability in order to increase function? Firstly, the result is controversial. But even if the observation were correct and general, we can still analyse it through the Panglossian paradigm. According to Professor Pangloss, a character in Voltaire's Candide,

"…our nose was made to carry spectacles, so we have spectacles.

Legs were clearly intended for breeches, and we wear them." [18].

Removing our noses would greatly affect our ability to wear spectacles. Yet it still does not provide any evidence that this is actually what our noses are 'for'. Or to use another analogy, globular proteins have evolved to function in the aqueous environment of the cytosol. One with less understanding of evolutionary biology might incorrectly conclude that our cytosol is aqueous in order to provide the best environment for globular proteins. And to provide more evidence for this claim, they might note that these proteins do not function well in non-aqueous environments. But of course proteins have evolved in a certain context, and changing this context can change the ability of the proteins to function, regardless of whether the context preceded the protein evolution. Proteins evolved in the context of a highly-dimensional sequence space where marginal stability is most likely, and the various properties that emerged during evolution were compatible with this context.

We can see this more directly in an evolutionary competition between different sets of proteins [13]. Consider three different types of proteins, which fulfil the same function through three different mechanisms. One type of protein can function if it has high stability ($\Delta G_{folding} < -2$), one type of protein requires moderate stability to function ($-2 < \Delta G_{folding} < -1$), and the third type requires marginal stability ($-1 < \Delta G_{folding} < 0$). A highly stable protein of the first type has identical fitness to a moderately stable protein of the second type, which has the identical fitness as a marginally stable protein of the third type. Any protein that mutated into a sequence with a value of $\Delta G_{folding}$ not within its functional range would be considered non-viable and eliminated from the simulation, as described above. Three different populations, one of each type, were allowed to equilibrate separately, and then mixed together in a single evolutionary run. In 24 out of 25 runs, the proteins that required marginal stability to function dominated the population by the end. (One time the protein type that required moderate stability became dominant, emphasizing the stochastic nature of evolution.) This was not because there was any positive selection, or any selective advantage at all for the proteins that required marginal stability. It was just that the functionality arose that was most consistent with the type of proteins that are formed naturally during the runs, those that were marginally stable. In the non-intuitive world of evolution, proteins evolved to require marginal stability because they were naturally marginally stable!

Again, it is important to emphasize that these simulations do not demonstrate that marginal stability results from neutral evolution. Only that it would result from neutral evolution, this is the most parsimonious explanation, and there is no need to look for an alternative explanation. The observation of marginal stability does not provide any evidence for any selective pressure for marginal stability.

Evolutionary Robustness

An alternative method to analyse the tendency for our evolutionary model to favour functional mechanisms consistent with marginal stability is to consider the role of mutational robustness. Fitness is related to the number of viable offspring likely in the next generation. A fully viable protein that requires high stability is quite likely to have mutated offspring that do not have the requisite stability. Proteins with marginal stability are more likely to have their mutated offspring having the marginal stability that they need for function. Proteins with functions that require high stability wander in a much more treacherous fitness landscape.

This can be seen directly in what has been described as 'the survival of the flattest' [19]. Consider a set of random protein sequences. For each sequence, we can calculate its initial stability, which we will refer to as ΔG_{wt}. We can also look at random mutations of this 'wild type' sequence. In general mutations will change the stability. Figure 4 shows the probability that a mutation is destabilizing, that is, $\Delta\Delta G_{folding} > 0$ [20].

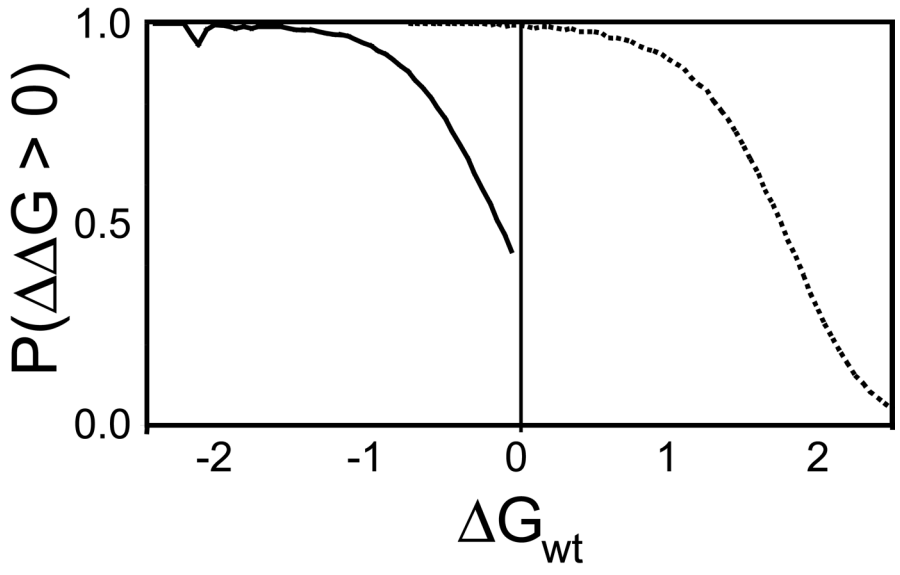

Figure 4. Probability a site mutation is destabilising as a function of the initial stability, for evolved proteins ($\Delta G_{crit} = 0$, solid line) compared with random proteins (dashed line).

We now look at a similar quantity for the proteins that result from population dynamics, where the proteins have evolved so that they survive only if $\Delta G_{folding} < 0$. If we take these sequences and make new, random mutations, we find that the probability of a destabilizing mutation, for a given value of ΔG_{wt}, is much lower. This means that two sequences with identical values of $\Delta G_{folding}$, possibly the same structure, one derived from a population simulation, the other chosen at random, can have vastly different responses to site mutations.

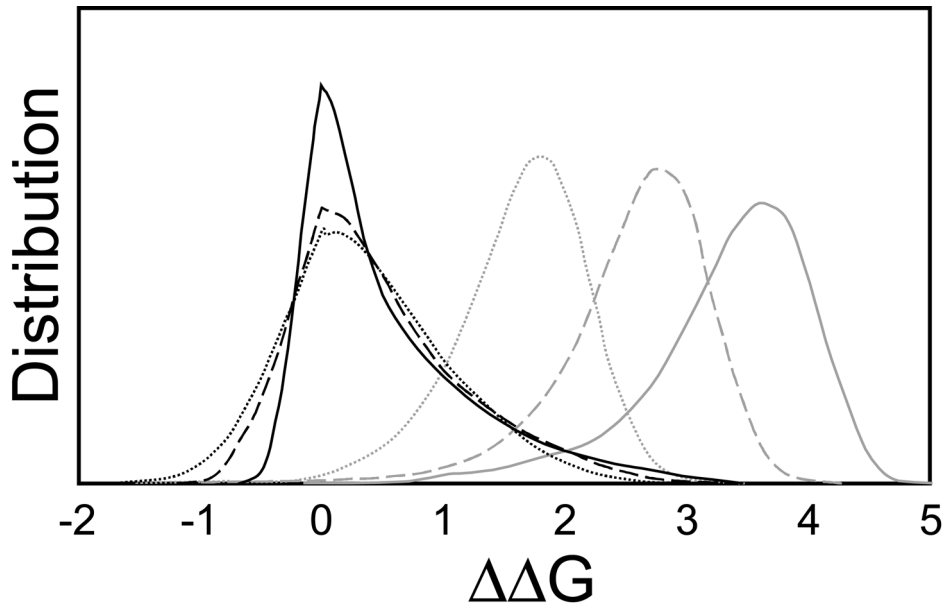

Figure 5. Distribution of changes in stability ($\Delta\Delta G$) for evolved proteins (black lines; $\Delta G_{crit} = 0$ (short dashes), -1 (long dashes), and -2 (solid)), compared with the distribution for random proteins (grey lines; $\Delta G_{folding} < 0$ (short dashes), -1 (long dashes), and -2 (solid)).

Figure 5 demonstrates this robustness in a different manner. In this figure, we show the distribution of $\Delta\Delta G$ values for random sequences (chosen with $\Delta G_{folding} < 0$, -1, and -2), compared with the corresponding distribution for evolved sequences (evolved with $\Delta G_{crit} = 0$, -1, and -2). For the random sequences, the more stable they are, the more destabilizing the average mutation. In fact, for every unit of increased initial stability, there is about an extra unit decrease in stability resulting from a random mutation. Conversely, for evolved sequences, the more stable the proteins, the more likely it is that the mutation will have little or no effect on the stability! Again, population dynamics work with evolution in order to select sequences that have fit offspring, that is, have robustness to mutations.

An alternative perspective can be obtained by considering the work of Eigen on biomolecular evolution [21]. Individuals do not evolve; populations do, representing a distribution of genotypes and their resultant phenotypes, what Eigen called a 'pseudospecies'. This distribution is maintained by the centrifugal nature of mutations. Rather than considering the fitness of a single wild-type sequence, we need to consider the averaged fitness over the entire distribution. A lower, flatter region of the fitness landscape might actually have a higher average fitness than a higher, more sharply peaked region. The population dynamics result in the selection for a network of related sequences that share a general level of fitness, and thus are robust to mutations through this network.

This perceived robustness of proteins to substitutions has been observed experimentally. For instance, Reddy *et al.*, catalogued a wide range of mutations, observing that approximately 25% actually increased thermal stability [22].

CONCLUSION

In contrast to the general tendency for biochemists and molecular biologists to imagine evolution as a march up the fitness landscape to higher and higher levels, modern evolutionary biology has emphasized the importance of evolution's stochastic nature, and the consequences of neutral evolution. While biochemists and molecular biologists are likely to try to explain an observed characteristic of living systems by asking how it serves to increase the fitness of those that possess it, many of these properties can be explained by neutral drift interacting with population dynamics. In this manner, sequence entropy - the number of sequences with a given characteristic - becomes important. Sequence entropy has been used to explain why some substructures and structures are over-represented in biological proteins [23-27] why proteins might fold to the structure of lowest free energy [7] and in work summarized here, why proteins are naturally marginally stable [13,28] and robust to site mutations [20].

One consequence of this perspective relates to the opportunities available in protein engineering. The work described above makes the prediction that biologically-derived proteins would be distinct from random sequences with similar properties by being especially robust to mutations. This suggests that it would be possible to modify naturally-occurring proteins to develop novel properties, and that other important qualities - structural rigidity, thermostability, etc. - would be maintained.

Conversely, the tendency of evolution to select flatter regions of the fitness landscape, ignoring potentially higher, narrower, peaks, means that there may be higher fitness peaks available to those who wish to find them through non-evolutionary means.

ACKNOWLEDGEMENTS

The work described here was largely performed by Sridhar Govindarajan, Darin Taverna, and Paul Williams.

REFERENCES

[1] Gould, S.J., Lewontin, R.C. (1979) The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proc. R. Soc. Lond. B* **205**:581-598.

[2] Sueoka, N. (1962) On the genetic basis of variation and heterogeneity of DNA base composition. *Proc. Natl. Acad. Sci. USA* **48**:582-592.

[3] Kimura, M. (1968) Evolutionary rate at the molecular level. *Nature* (Lond.) **217**:624-626.

[4] King, J.L., Jukes, T.H. (1969) Non-Darwinian evolution. *Science* **164**:788-798.

[5] Abkevich, V.I., Gutin, A.M., Shakhnovich, E.I. (1995) Impact of local and non-local interactions on thermodynamics and kinetics of protein folding. *J. Mol. Biol.* **252**:460-471.

[6] Miyazawa, S., Jernigan, R.L. (1985) Estimation of effective interresidue contact energies from protein crystal structures: Quasi-chemical approximation. *Macromolecules* **18**:534-552.

[7] Govindarajan, S., Goldstein, R.A. (1998) On the thermodynamic hypothesis of protein folding. *Proc. Natl. Acad. Sci. USA* **95**:5545-5549.

[8] Wagner, G., Wuthrich, K. (1979) Correlation between the amide proton exchange rates and the denaturation temperatures in globular proteins related to the basic pancreatic trypsin inhibitor. *J. Mol. Biol.* **130**:31-37.

[9] Tang, K.E.S., Dill, K.A. (1998) Native protein fluctuations: The conformational-motion temperature and the inverse correlation of protein flexibility with protein stability. *J. Biomol. Struct. Dyn.* **16**:397-411.

[10] Dunker, A.K., Garner, E., Guilliot, S., Romero, P., Albrecht, K., Hart, J., Obradovic, Z., Kissinger, C., Villafranca, J.E. (1998) Protein disorder and the evolution of molecular recognition: Theory, predictions and observations. *Pacific Symp. Biocomputing* **3**:473-484.

[11] Wright, P.E., Dyson, H.J. (1999) Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **293**:321-331.

[12] Dunker, A.K., Obradovic, Z. (2001) The protein trinity-linking function and disorder. *Nat. Biotechnol.* **19**:805-806.

[13] Taverna, D.M., Goldstein, R.A. (2002) Why are proteins marginally stable? *Proteins: Struct. Funct. Genet.* **46**:105-109.

[14] Alber, T., Wozniak, J.A. (1985) A genetic screen for mutations that increase the thermal-stability of phage-T4 lysozyme. *Proc. Natl. Acad. Sci. USA* **82**:747-750.

[15] Bryan, P.N., Rollence, M.L., Pantoliano, M.W., Wood, J., Finzel, B.C., Gilliland, G.L., Howard, A.J., Poulos, T.L. (1986) Proteases of enhanced stability: Characterization of a thermostable variant of subtilisin. *Proteins: Struct. Funct. Genet.* **1**:326-334.

[16] Liao, H., McKenzie, T., Hageman, R. (1986) Isolation of a thermostable enzyme variant by cloning and selection in a thermophile. *Proc. Natl. Acad. Sci. USA* **83**:576-580.

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

[18] Voltaire, F.M. (1759) (real name Arouet, F. M.), *Candide*, (John Butt, trans.) Penguin, London.

[19] Nimwegen, E. v., Crutchfield, J.P., Huynes, M. (1999) Neutral evolution of mutational robustness. *Proc. Natl. Acad. Sci. USA* **96**:9716-9720.

[20] Taverna, D.M., Goldstein, R.A. (2002b) Why are proteins so robust to site mutations? *J. Mol. Biol.* **315**:479-484.

[21] Eigen, M. (1971). Self organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **10**:465-523.

[22] Reddy, B.V.B., Datta, S., Tiwari, S. (1998) Use of propensities of amino acids to the local structure environment to understand effect of substitution mutations on protein stability. *Protein Engng* **11**:1137-1145

[23] Finkelstein, A.V., Ptitsyn, O.B. (1987) Why do globular proteins fit the limited set of folding patterns. *Prog. Biophys. Mol. Biol.* **50**:171-190.

[24] Govindarajan, S., Goldstein, R.A. (1995) Searching for foldable protein structures using optimized energy functions. *Biopolymers* **36**:43-51.

[25] Govindarajan, S., Goldstein, R.A. (1996) Why are some protein structures so common? *Proc. Natl. Acad. Sci. USA* **93**: 3341-3345.

[26] Li, H., Helling, R., Tang, C., Wingreen, N. (1996) Emergence of preferred structures in a simple model of protein folding. *Science* **273**:666-669.

[27] Shakhnovich, E.I. (1998) Protein design: A perspective from simple tractable models. *Folding & Design* **3**:R45-R58.

[28] Williams, P.D., Pollock, D.D., Goldstein, R.A. (2001) Evolution of functionality in lattice proteins. *J. Mol. Graphics Modell.* **19**:150-156.