Sequence entropy of folding and the absolute rate of amino acid substitutions

Richard A. Goldstein[1] and David D. Pollock[2*]

[1]Division of Infection & Immunity, University College London, London, WC1E 6BT, UK. r.goldstein@ucl.ac.uk; [2]Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, CO 80045 USA. David.Pollock@ucdenver.edu.

[*]Correspondence to: David.Pollock@ucdenver.edu.

Keywords: protein evolution, epistasis, coevolution, transition state theory, statistical mechanics, substitution rate, evolutionary Stokes shift, entrenchment, contingency

Abstract

An adequate representation of protein evolution needs to consider how the acceptance of new mutations depends on the overall context in which they arise. Epistatic interactions between sites in a protein link the substitutions at one site with the changes occurring at other sites, resulting in time and spatial rate heterogeneity beyond the capabilities of current models. Here, we exploit parallels between amino acid substitutions and chemical reaction kinetics to develop a new theory of protein evolution. This theory was developed by constructing a mechanistic framework for modelling amino acid substitution rates that employs the formalisms of statistical mechanics, with population genetics principles underlying the analysis. We use theoretical analyses and computer simulations of proteins under purifying selection for thermodynamic stability to show that substitution rates and the stabilisation of the currently resident amino acid (the 'evolutionary Stokes shift') can be predicted from biophysics and the effect of sequence entropy alone. Furthermore, we demonstrate that substitutions predominantly occur when epistatic interactions result in near neutrality of that substitution; substitution rates are thus determined by how often epistasis results in such nearly neutral conditions. Our theory provides a general framework for understanding and modelling protein sequence change under purifying selection, explains patterns of convergence and mutation rates in real proteins that are incompatible with previous models of protein evolution, and provides a better null model for the detection of adaptive changes.

Introduction

Protein sequences, like all biological systems, are continuously sculpted through the evolutionary process. Mutations that arise at the DNA level are generally fixed or eliminated with a probability that depends on the effect of that mutation on the protein's structure, stability, functionality, intermolecular interactions and other properties. These properties depend on complex networks of interacting amino acids throughout the protein so that the effect of a mutation depends upon the background sequence in which it occurs; selection on such holistic properties of the protein sequence induces epistatic interactions (coevolution) among sites. Because of the complexity of the epistatic interactions, it has been difficult to identify what determines substitution rates at a site, to characterise how these rates depend on the rest of the sequence, and to understand how they vary with time and location in the protein.

The standard approach to studying protein evolution is to employ empirical substitution rate models that neglect epistatic interactions and the resultant rate heterogeneity beyond simple scaling factors. In such models, the parameters are adjusted to best represent observed differences between related protein sequences[1-3]. Although these models have had a major impact in many areas of the life sciences, they cannot estimate the effect of epistatic interactions on stability, function or fitness, predict the role of compensatory substitutions in protein evolution[4,5], predict which of the 10% of deleterious mutations in humans are harmless in other species[6], or accurately represent the rate and time dependence of convergence and homoplasy[5]. Empirical models have been developed that include rate heterogeneity[7-10], but their utility and accuracy are limited by the sequence information required to estimate the resulting explosion of adjustable parameters. One attractive possibility is to create more mechanistic substitution models that better represent the underlying process of molecular evolution and protein biophysics. This would allow more accurate models to be constructed using a limited set of biologically meaningful parameters. Current attempts in this direction, however, are hindered by the lack of a deeper understanding of the process of sequence change, especially the characteristics and effects of epistasis.

We have previously demonstrated that computational simulations of protein evolution, with fitness determined by thermodynamic stability, can reproduce many of the puzzling aspects of protein evolution including the rate- and time-dependence of convergence[5] and the site- and time-dependence of substitution rates[11]. In particular, these models exhibit a phenomenon we named the 'evolutionary Stokes shift'[12], the tendency for the newly resident amino acid at a site to be stabilised, or 'entrenched'[13] over evolutionary time following a substitution. We also observed a tendency for the new amino acid to be pre-stabilised prior to the substitution by chance or contingency[13,14]. Consequentially, the evolutionary Stokes shift process can proceed through entirely neutral or nearly-neutral substitutions[12]. The pre- and post-adjustment of the protein to the new amino acid occurs without a corresponding changes in fitness, distinguishing this process from compensatory substitutions which generally involve a fitness increase[15].

Until now we could say little about the mechanism by which selection for stability determine substitution rates at individual sites in a protein, and what drives the pre- and post-substitution stabilisations in the absence of changes in fitness. To remedy this, we describe a mechanistic framework for modelling amino acid substitution rates employing the formalisms of statistical mechanics. Although our theoretical framework is based on the need of most proteins to be in a specific structure in order to function, analogous models can be applied to other forms of selection at this and other organismal levels; an example is the application of cellular Potts models to morphological development[16]. We find that average substitution rates can be explained by the evolutionary equivalent of transition state theory, with fluctuations in amino acid preferences due to epistatic interactions representing an essential aspect of the substitution process. Just as entropy plays a preeminent role in statistical mechanics, the sequence entropy of folding, defined as the log of the number of possible sequences that fold with the evolutionarily determined degree of stability, is central to evolutionary mechanics. We test our mathematical approximations and predictions using computational simulations of protein evolution. We demonstrate that average substitution rates at a site can be predicted from site-specific stability distributions estimated in the absence of selection on that site and the relative sequence entropy of folding associated with different overall protein stabilities. The effect of other global factors such as effective population size, protein structure designability, and selective strength are combined in the entropy term and do not need to be considered or estimated separately. This provides a powerful approach to understanding the determinates of substitution rates in the presence of epistasis.

Results

Site-specific stabilities and relative substitution rates: To develop a mechanical theory of protein evolution, we considered how purifying selection for stability determines site-specific substitution rates. Real proteins are under selection for a range of properties; we choose this specific form of selective pressure because it is well defined, theoretically tractable, and a common constraint for a large number of different types of proteins. We expect the insights from this analysis to be applicable whenever there is purifying selection for a holistic protein property resulting from a large number of modest contributions. In addition, analysing selection on stability provides a 'null model' to examine the effect of other forms of selection acting on specific proteins.

The stability $\Phi(\mathbf{X})$ of a protein sequence $\mathbf{X} = \{x_1, x_2, x_3 \dots x_n\}$ was defined as the negative of the free energy of folding, so that more positive values indicate greater stability. This redefinition allows simpler descriptive language and a more direct relationship to population genetics terms. The fitnesses of sequences were set equal to the probability that the encoded proteins would be folded at thermodynamic equilibrium (Equation (1))[12,17,18]. Thus, increases in stability correspond to increases in fitness.

To understand how the rest of the protein influences substitution rates at individual sites, we focus our attention on a focal site k, currently occupied by amino acid α. Relative to site k, the protein stability, $\Phi(\mathbf{X})$, can be partitioned into the sum of two contributions $\Phi(\mathbf{X}) =$

$\phi_{k,\alpha}(\mathbf{X}_{\not\ni k}) + \Phi_{k,\mathrm{Bath}}(\mathbf{X}_{\not\ni k})$. The first term, $\phi_{k,\alpha}(\mathbf{X}_{\not\ni k})$, includes the contribution to the stability from the site-specific interactions between the amino acid α at site k and the amino acids at all other sites excluding k, $\mathbf{X}_{\not\ni k}$; in standard statistical mechanics analyses, this term represents the system of interest. The second term, $\Phi_{k,\mathrm{Bath}}(\mathbf{X}_{\not\ni k})$, includes interactions among amino acids at all sites excluding the focal site. Because the vast majority of interactions do not involve site k, this second term corresponds to the thermodynamic bath in statistical mechanics, as indicated by the subscript. Both contributions include interactions in the folded state as well as unfolded states. For simplicity, in the Results and Discussion sections we omit the functional dependence of these values on $\mathbf{X}_{\not\ni k}$ and the specification of the site k when it is clear from context, and use Φ, ϕ_α and Φ_{Bath} to represent $\Phi(\mathbf{X})$, $\phi_{k,\alpha}(\mathbf{X}_{\not\ni k})$ and $\Phi_{k,\mathrm{Bath}}(\mathbf{X}_{\not\ni k})$, respectively. (For clarity, the explicit representation is maintained in the Methods section.)

This statistical mechanics formalism can now be applied to understanding the amino acid substitution rate under purifying selection in the low mutation rate regime where polymorphisms are negligible. Still considering a specific focal site, the instantaneous rate of substitution from a resident amino acid α to a new amino acid β is equal to the rate of mutation from α to β times the fixation probability. The fixation probability depends on the difference in fitness between proteins with amino acid α or β at that site with the rest of the sequence unchanged; in our thermodynamic model, fitness differences specifically depend on the impact of the two amino acids on the protein's stability. Prior to the mutation, when amino acid α is resident, the protein stability is equal to $\Phi = \phi_\alpha + \Phi_{\mathrm{Bath}}$ with corresponding fitness $m(\Phi)$, given by Equation 1 (Methods). After a mutation to amino acid β, the stability is equal to $\Phi' = \phi_\beta + \Phi_{\mathrm{Bath}} = \Phi + \phi_\beta - \phi_\alpha$, corresponding to fitness $m(\Phi') = m\left(\Phi + \phi_\beta - \phi_\alpha\right)$, where we have used the fact that Φ_{Bath} is unchanged by the mutation. The situation is complicated by the non-linear relationship between fitness and stability (Equation 1, Methods), but can be greatly simplified by noting that real proteins, as well as proteins from this and other evolutionary simulations under purifying selection for thermostability, evolve within a narrow range of stability values around an average value $\overline{\Phi}$[17,19-22]; see Supplementary Fig. S1. This narrow stability range occurs where the effectiveness of selection for greater stability is balanced by large numbers of slightly destabilising mutations fixed by genetic drift[23,24]. We therefore approximate the protein's stability prior to the mutation as equal to $\Phi = \overline{\Phi}$; the resulting change in fitness is then equal to $\Delta m_{\alpha\to\beta} = m\left(\overline{\Phi} + \phi_\beta - \phi_\alpha\right) - m(\overline{\Phi})$. The value of $\overline{\Phi}$ depends on factors such as temperature[17], effective population size (as shown in[17] and Fig. S1), and protein structure and function, but will be constant as long as these factors are approximately constant. With these assumptions, the change in fitness and thereby the probability of fixation of the mutation is therefore determined by the difference between the current values of ϕ_α and ϕ_β.

While the total stability value of $\overline{\Phi}$ is a constant, the manner in which this stability is distributed amongst the various interactions, and therefore the values of ϕ_α and ϕ_β as well as the corresponding substitution rate, will vary as substitutions occur along the rest of the protein sequence. The nature of this variation depends on which amino acid occupies position k because that amino acid affects the evolution in the rest of the protein[12]. In order to compute the estimated average substitution rate, we assume that the other sites are sufficiently numerous and change sufficiently rapidly that the protein is always fully adjusted to the current amino acid at site k. (This assumption is most likely to break down following non-conservative substitutions, as discussed below.) The joint probability distribution of ϕ_α and ϕ_β given total stability $\bar{\Phi}$ and the occupation of site k by amino acid α can then be described by the stationary distribution $\rho(\phi_\alpha, \phi_\beta | \Phi(\mathbf{X}) = \bar{\Phi}, x_k = \alpha)$, which we simplify to $\rho(\phi_\alpha, \phi_\beta | \alpha)$. The average substitution rate from α to β at a site is completely determined by the mutation rate and this marginal distribution. Because of its predicted importance in determining the average substitution rate, the rest of this paper will focus primarily on characterising this resident-dependent joint stability probability density.

With the approximations that proteins with the same structure and population conditions evolve to the same stability, and that the fitness is completely specified by this stability, all of the sequences represented by $\rho(\phi_\alpha, \phi_\beta | \alpha)$ have identical fitnesses. Under such circumstances, $\rho(\phi_\alpha, \phi_\beta | \alpha)$ is simply proportional to the *number* of sequences with those values of ϕ_α and ϕ_β, amino acid α at site k, and total stability $\bar{\Phi}$. In analogy to Boltzmann's description of entropy as proportional to the log of the number of microscopic representations of a system corresponding to a specified macroscopic description, we refer to the log of the number of sequences corresponding to specific values of ϕ_α, ϕ_β, $x_k = \alpha$ and $\bar{\Phi}$ as the 'sequence entropy of folding' $S(\phi_\alpha, \phi_\beta | \alpha)$. We note that this quantity is very different from the 'sequence entropy' derived from information theory and commonly used to represent site-specific variability[25]. The average substitution rates are determined by the dependence of the sequence entropy of folding on ϕ_α and ϕ_β, as reflected in $\rho(\phi_\alpha, \phi_\beta | \alpha) \propto e^{S(\phi_\alpha, \phi_\beta | \alpha)}$.

To explore and evaluate our theoretical analysis, we simulated the evolution of a 300-residue protein under selection for thermodynamic stability. In these simulations, fitness was equal to the probability of the protein being folded at thermodynamic equilibrium so as to match our theoretical model. These simulations are not meant to make quantitative predictions in particular cases, but rather to predict general characteristics of evolutionary behaviour for proteins that require a native confirmation to carry out some critical biological function. They have demonstrated their ability to reproduce fundamental aspects of protein evolution[12,17,18]. By using a simple pair-contact model of protein thermodynamics, we were able to perform replicate simulations corresponding to about 5 billion years given typical eukaryotic substitution rates.

We first examined the joint probability distributions $\rho(\phi_\alpha, \phi_\beta | \alpha)$ observed in the simulated proteins over long periods of time. Although the pair-contact potential does not vary amongst sites, these models still produce different average substitution rates at different sites, as with real proteins. To analyse these differences, we grouped sites with similar substitution patterns into four different site classes, with class 1 the most exposed and class 4 the most buried. Figs. 1A-D show the resulting joint probability distributions $\rho_{\mathbb{C}}(\phi_{\mathrm{Glu}}, \phi_{\mathrm{Lys}} | \mathrm{Glu})$ and $\rho_{\mathbb{C}}(\phi_{\mathrm{Glu}}, \phi_{\mathrm{Lys}} | \mathrm{Lys})$, the distribution of site-specific contributions of glutamic acid and lysine conditional on glutamic acid or lysine being resident, for the set of sites $\{k\}_{\mathbb{C}}$ belonging to site class $\mathbb{C} = \{1,2,3,4\}$. Corresponding distributions of population

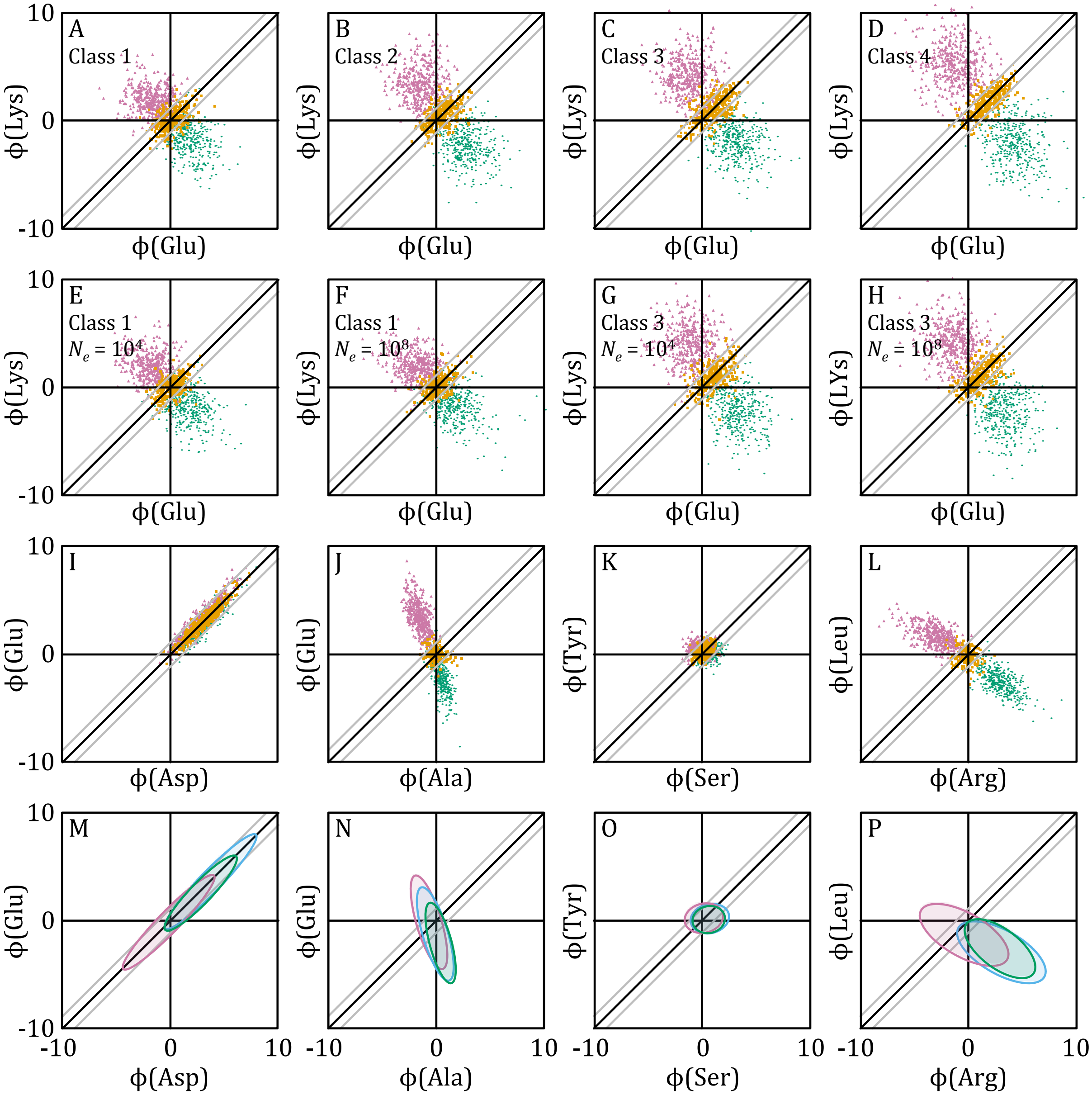

Figure 1: A-L) Relative local contributions to stability for glutamic acid and lysine in different site classes (A-D), or for different population sizes for site class 1 and 3 (E-H), or various amino acid pairs in site class 3 (I-L). Points were sampled when the amino acid in the abscissa was resident (green), when the amino acid in the ordinate was resident (pink), or during transitions between the two (yellow). M-P) Distributions of local contributions to stability in reference state when the non-interacting null amino acid was present ($\rho(\phi_\alpha, \phi_\beta|\emptyset)$, pink), when the amino acid in the abscissa was present as predicted using Equation (7) ($\tilde{\rho}(\phi_\alpha, \phi_\beta|\alpha)$, cyan), or as observed ($\rho(\phi_\alpha, \phi_\beta|\alpha)$, green). Grey diagonal lines mark the boundaries of regions of near-neutral substitutions.

scaled selective coefficients are shown in Supplementary Fig. S2. The distributions are broad, consistent with earlier results demonstrating that selective pressures vary over a wide range as substitutions occur elsewhere in the protein[12]. Exposed, rapidly evolving sites with few selective constraints (site class 1, Fig. 1A) generally have more compact distributions with smaller variances in ϕ_{Glu} and ϕ_{Lys} compared to buried, slowly evolving sites (site class 4, Fig. 1D). The distributions also strongly depend on whether the glutamic acid or lysine is the resident amino acid. In particular, the potential contribution of an amino acid to the protein stability tends to be greater when that amino acid is resident at a site (e.g., $\rho_{\mathbb{C}}\left(\phi_{\mathrm{Glu}}, \phi_{\mathrm{Lys}} \middle| \mathrm{Glu}\right)$ is centred on higher values of ϕ_{Glu} than is $\rho_{\mathbb{C}}\left(\phi_{\mathrm{Glu}}, \phi_{\mathrm{Lys}} \middle| \mathrm{Lys}\right)$), a reflection of the 'evolutionary Stokes shift'[12]. The amount of this increase appears to be correlated with the variance in ϕ_{α}.

Joint distributions at different population sizes (over 4 orders of magnitude) and for other pairs of amino acids are shown in Figs. 1E-L. The bivariate distributions are surprisingly independent of the population size (Figs. 1E-H, S2), but are highly dependent on which amino acids are being compared (Figs. 1I-L, S2). Distributions for physicochemically similar amino acids (e.g., glutamic acid versus aspartic acid, Fig. 1I) are highly correlated, while those for dissimilar amino acids (e.g., glutamic acid versus alanine, Fig. 1J) are anti-correlated. A non-resident amino acid is generally stabilised if the distributions are correlated (e.g. ϕ_{Glu} is positive when aspartic acid is present), but destabilised if the distributions are anti-correlated (e.g. ϕ_{Glu} is negative when alanine is present).

Predicting relative substitution rates: As described above, substitution rates should be predictable from knowledge of $\boldsymbol{\rho}\left(\boldsymbol{\phi}_{\alpha}, \boldsymbol{\phi}_{\beta} \middle| \boldsymbol{\alpha}\right)$. To test this, we modelled $\boldsymbol{\rho}\left(\boldsymbol{\phi}_{\alpha}, \boldsymbol{\phi}_{\beta} \middle| \boldsymbol{\alpha}\right)$ as bivariate normal distributions based on the observations in Fig. 1, and numerically integrated over these distributions to calculate substitution rates using Kimura's formula for the probability of fixation[26-28] (Equation (2)). There is extremely good agreement between expected substitution rates and those obtained by counting substitutions that occurred during simulations, for all site classes over a four order of magnitude range of population sizes, as shown in Figs. 2A-C. This supports our use of the approximation $\boldsymbol{\Delta m_{\alpha\rightarrow\beta}} = \boldsymbol{m}\left(\boldsymbol{\bar{\Phi}} + \boldsymbol{\phi}_{\beta} - \boldsymbol{\phi}_{\alpha}\right) - \boldsymbol{m}(\boldsymbol{\bar{\Phi}})$. The population size independence of the predicted (Figs. 2A-C) and observed substitution rates (Fig. S1B) matches previous observations[29], and corresponds to our use of a concave-down fitness function (Equation 1)[30].

We next investigated whether substitution rate calculations could be simplified by considering the dynamics of the substitution process. As described above, the values of ϕ_{α} and ϕ_{β} vary as the rest of the protein sequence changes. A part of the evolutionary trajectory before and after a glutamic acid to lysine substitution is shown in Fig. 3. Most of the time prior to substitution when glutamic acid is resident, glutamic acid is stabilised by the evolutionary Stokes shift, while lysine is slightly destabilised, reflecting the physicochemical differences between these two amino acids. The pattern is reversed after the substitution when lysine is resident. Strikingly, the substitution occurs when fluctuations in the values of ϕ_{Glu} and ϕ_{Lys} take the site into a narrow overlap region between $\rho\left(\phi_{\mathrm{Glu}}, \phi_{\mathrm{Lys}} \middle| \mathrm{Glu}\right)$ and $\rho\left(\phi_{\mathrm{Glu}}, \phi_{\mathrm{Lys}} \middle| \mathrm{Lys}\right)$, along the diagonal $\phi_{\mathrm{Glu}} = \phi_{\mathrm{Lys}}$, where substitutions are nearly neutral (i.e. $m(\bar{\Phi} + \phi_\beta - \phi_\alpha) \approx m(\bar{\Phi})$). The general tendency for substitutions to occur under conditions of near neutrality is also supported by the data shown in Figs. 1 and S2.

These observations suggest the possible applicability of transition state theory (TST), a method for predicting the rate of chemical reactions[31]. TST focuses on how the energies of the reactants and products vary as the reactants undergo conformational fluctuations. The reaction is assumed to be possible only when the reactants are in a 'transition state' in

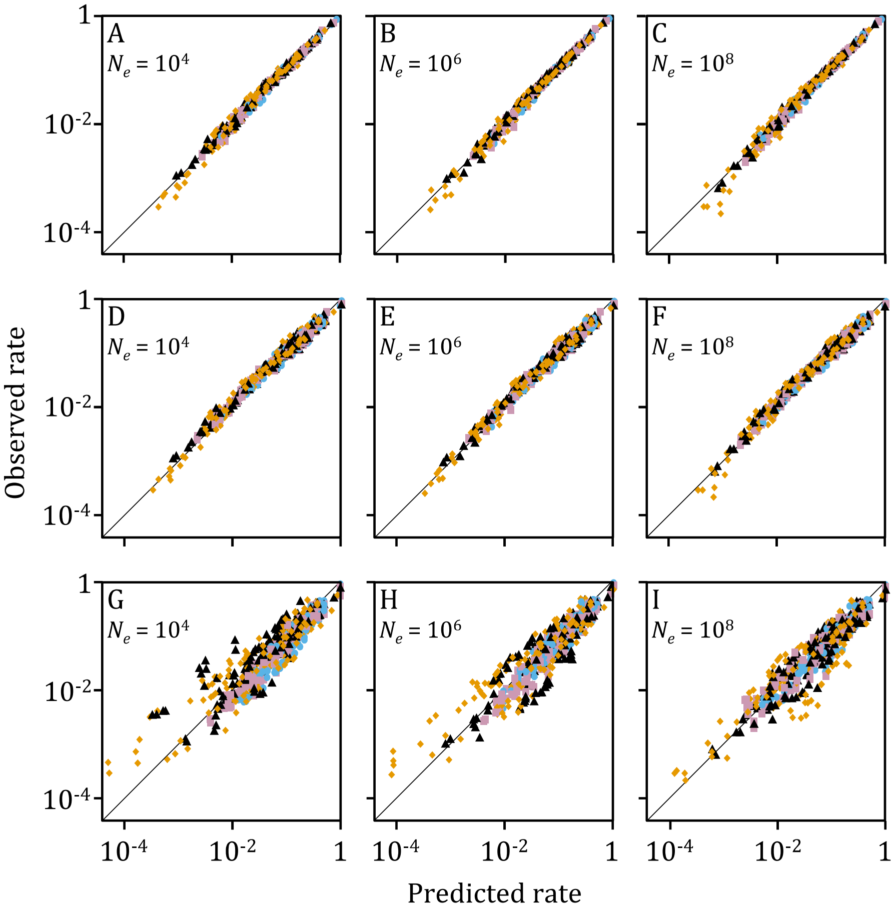

Figure 2: Comparison of predicted and observed substitution rates, for all pairs of amino acids separated by a single base change for all sites in the different site classes (Class 1, blue circles; Class 2, pink squares; Class 3, black triangles; Class 4, orange diamonds). A-C: predicted substitution rates calculated by integrating over $\rho(\phi_\alpha, \phi_\beta|\alpha)$ for three different population sizes. D-F: Predicted substitution rates calculated using transition state theory (Equation (6)), which assumes only near-neutral substitutions occur. G-I: Predicted substitution rates calculated using transition state theory with parameters estimated using Equation 7.

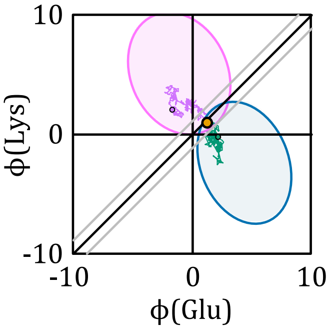

Figure 3: Example of a trajectory before and after a substitution from glutamic acid to lysine. Local contribution to stability when either is resident is shown for before (green) and after the substitution (pink) (green). Values during the substitution shown in yellow; beginning and end points are shown as black circles. The observed distributions over the simulations when glutamic acid or lysine is resident shown as shaded region. Grey diagonal lines mark regions of near-neutral substitutions.

which the energies of reactant and product are equal. The predicted reaction rate is then equal to the probability that the reactants are in the transition state, times the rate of conversion from reactants in the transition state to products.

Adapting this theory, the substitution rate from α to β at every site was estimated from the probability that the protein is in a transition state in which the fitnesses of the wild type and mutant are nearly equal, times the rate of substitution under neutral conditions. The probability that the protein is in a nearly neutral transition state was calculated by integrating $\rho\left(\phi_\alpha, \phi_\beta|\alpha\right)$ over a constant-width zone straddling the neutral line $\phi_\alpha = \phi_\beta$. As described in Methods, the width of the neutral zone is determined by how the numbers of sequences varies with the overall protein stability, which is represented in our analysis as an exponential $\rho_\Phi(\Phi) \propto \exp(-\gamma\Phi)$, where γ was estimated for our simulations by the relative numbers of destabilising and stabilising mutations. Under this assumption, the width of the neutral zone can be shown to approximately equal $\frac{2}{\gamma}$. Multiplying this probability by the neutral substitution rate results in a closed-form expression for substitution rates (Equation (6)). This approach produced strikingly accurate substitution rate predictions (Fig. 2D-F). Notably, because this calculation considers only neutral substitutions, Kimura's fixation fitness-dependent and population size-dependent fixation probability formula is not needed, greatly simplifying the calculations. The average substitution rate between amino acids is completely determined by the bivariate normal approximations of $\rho\left(\phi_\alpha, \phi_\beta|\alpha\right)$, with no other parameters needed besides the mutation rate and γ.

The equilibrium distributions of site-specific stabilities and the evolutionary 'Stokes Shift': As described above, the rate of amino acid substitutions is determined by $\boldsymbol{\rho}\left(\boldsymbol{\phi}_\alpha, \boldsymbol{\phi}_\beta|\boldsymbol{\alpha}\right)$ in the region where $\boldsymbol{\phi}_\alpha \approx \boldsymbol{\phi}_\beta$. A better mechanistic description requires that we understand how these distributions, and therefore the substitution rates, are determined; we now demonstrate how this goal can be advanced using the principles of statistical mechanics and sequence entropy of folding.

As discussed above, the distribution $\rho(\phi_\alpha, \phi_\beta|\alpha) \propto e^{S(\phi_\alpha, \phi_\beta|\alpha)}$ simply reflects the sequence entropy of folding for specified values of ϕ_α, ϕ_β, $x_k = \alpha$ and $\bar{\Phi}$. We approximate $\rho\left(\phi_\alpha, \phi_\beta|\alpha\right)$ by the product of two terms, $\rho_{\text{Loc}}\left(\phi_\alpha, \phi_\beta\right) \times \rho_{\text{Bath}}(\Phi_{\text{Bath}} = \bar{\Phi} - \phi_\alpha)$. The local term $\rho_{\text{Loc}}(\phi_\alpha, \phi_\beta$ represents the fraction of sequences with site-specific ϕ_α and ϕ_β, while the second term $\rho_{\text{Bath}}(\bar{\Phi} - \phi_\alpha)$ represents the fraction of sequences where the bath interactions provide sufficient contributions to the stability so that $\phi_\alpha + \Phi_{\text{Bath}} = \bar{\Phi}$. (Equivalently, we are approximating the sequence entropy of folding, $S(\phi_\alpha, \phi_\beta|\alpha)$, as the sum of local and bath terms, $S(\phi_\alpha, \phi_\beta|\alpha) \approx S_{\text{Loc}}(\phi_\alpha, \phi_\beta) + S_{\text{Bath}}(\Phi_{\text{Bath}} = \bar{\Phi} - \phi_\alpha)$.) This calculation assumes independence of the bath and local contributions to total stability, which is likely to be approximately true as the interactions involved in the two terms are different.

To estimate the first term, we calculated $\rho(\phi_\alpha, \phi_\beta|\emptyset)$, the distribution of site-specific stability contributions in the absence of selection at that focal site; this was accomplished by performing simulations in which a non-interacting amino acid $\emptyset$ was fixed at that site and all other sites were allowed to evolve freely. We then calculated the values of ϕ_α and ϕ_β that would result if amino acids α and β were substituted for $\emptyset$ in the sequences arising from the simulation. Interactions involving the focal amino acid represent a small fraction of total stability contributions, so the second term $\rho(\Phi_{\text{Bath}})$ was approximated by the

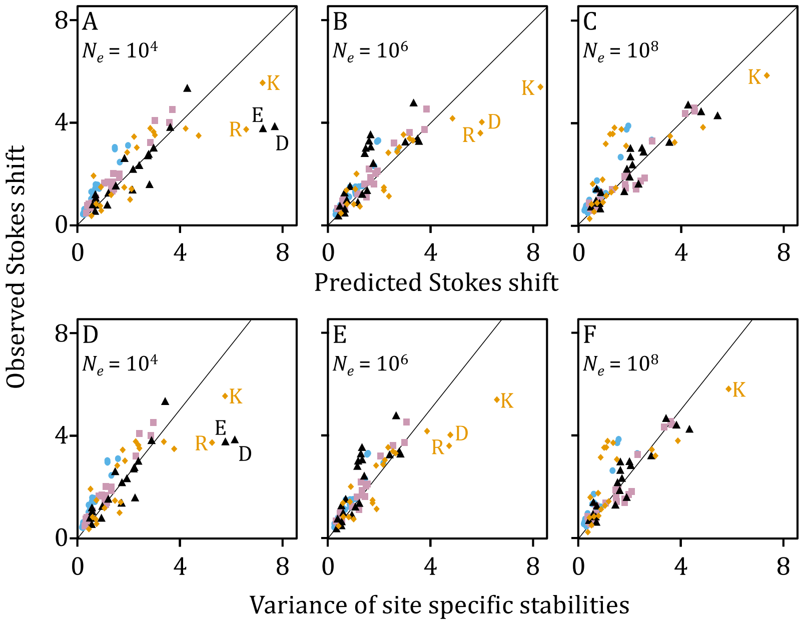

Figure 4: Accuracy of site-specific stability and evolutionary Stokes shift predictions. A-C) Observed versus estimated values of the Stokes shift ($\zeta_{\alpha|\alpha}$) for all four site rate classes (Class 1, blue circles; Class 2, pink squares; Class 3, black triangles; Class 4, orange diamonds), for three different population sizes. D-F) The linear relationship between the observed evolutionary Stokes shift and the variance in amino acid-specific stability contributions in the absence of selection on the site ($\sigma^2_{\alpha|\emptyset}$). The lines shown are theoretical predictions with $\gamma = 1.26$. Outliers are identified.

distribution of protein sequences with total stability Φ represented as before as the exponential $\rho_{\Phi}(\Phi) \propto \exp(-\gamma \Phi)$.

As derived in the Methods section, the evolutionary Stokes shift is expected to change the average value of ϕ_{α} by an amount $\zeta_{\alpha|\alpha} \equiv \overline{\phi}_{\alpha|\alpha} - \overline{\phi}_{\alpha|\emptyset} = \gamma\sigma^2_{\alpha|\emptyset}$, where $\overline{\phi}_{\alpha|\alpha}$ is the average value of ϕ_{α} when α is resident at that site and $\sigma^2_{\alpha|\emptyset}$ is the variance in the distribution of $\rho(\phi_{\alpha}|\emptyset)$. The mechanism of this shift can be understood by comparing the relative contributions of ϕ_{α} and Φ_{Bath} to $\overline{\Phi}$. Increasing values of ϕ_{α} correspond to decreasing values of Φ_{Bath} necessary to fulfill $\phi_{\alpha} + \Phi_{\mathrm{Bath}} = \overline{\Phi}$. As the number of possible sequences rapidly increases with decreasing Φ_{Bath}, the result is a strong bias towards increased values of ϕ_{α}. This stabilisation resulting from the large increase in sequence entropy of folding with decreasing Φ_{Bath} is precisely the evolutionary Stokes shift.

The predicted distributions of $\rho(\phi_{\alpha}, \phi_{\beta}|\alpha)$ versus those observed in thermodynamic simulations are shown in Figs. 1M-P. The estimated $\zeta_{\alpha|\alpha}$ obtained by this approximation matches values observed in the simulations surprisingly well given the approximations made (Fig. 4A-C). As predicted, the entropic stabilisation is approximately linear with $\sigma^2_{\alpha|\emptyset}$, and the slope is close to the estimated value of γ = 1.26 (kcal mol^{-1})$^{-1}$ (Fig. 4D-F), confirming the trends evident in Fig. 1. The observed entropic stabilisation is smaller than predicted for the two largest shifts in the two slowest rate classes, involving the charged lysine, arginine, aspartic acid and glutamic acid. Earlier work demonstrated that equilibration for the most buried states can be extremely slow[12]; these deviations may occur when the protein has had insufficient time to adjust to the presence of the new amino acid.

In earlier work, we described how the evolutionary Stokes shift results in stabilisation of amino acids that are similar to the current resident, and destabilisation of amino acids that have large physicochemical differences. We can now understand the basis of this effect. According to our new theory, the presence of α at the site shifts the average values of ϕ_{β} by $\zeta_{\beta|\alpha} = \gamma\, \varphi_{\alpha\beta|\emptyset}\, \sigma_{\alpha|\emptyset}\, \sigma_{\beta|\emptyset}$, where $\varphi_{\alpha\beta|\emptyset}$ is the correlation between ϕ_{α} and ϕ_{β} in $\rho(\phi_{\alpha}, \phi_{\beta}|\emptyset)$; these shifts can be to either higher or lower values depending on whether the physicochemical properties of the amino acids are similar or different (positive or negative $\varphi_{\alpha\beta|\emptyset}$, respectively), increasing or decreasing the density of the distribution in the region $\phi_{\alpha} \approx \phi_{\beta}$ and the corresponding substitution rate. Substitution rates estimated with the TST approximation (Equation (6)) using the site-specific stabilities calculated using Equation (7) are remarkably accurate for all four site classes over four orders of magnitude (Fig. 2G-I).

Discussion

The evolutionary mechanics developed here represents a fundamental shift in how we conceptualise the process of protein evolution. It allows us to understand how sequence entropy and epistasis determine the relative magnitudes of substitution rates and how these rates fluctuate over time. We provide a mechanistic explanation for the known predominance of nearly neutral evolution and a better understanding of what happens during purifying selection. We and others have previously shown that evolutionary simulations based on protein thermodynamics produce patterns of epistasis, convergence, and entrenchment that are qualitatively similar to patterns in real proteins; the current research provides a clear explanation why these patterns were produced, how they result from statistical mechanics considerations.

A central result of the theory is an understanding of how relative substitution rates among amino acids arise. The effect of an amino acid-altering mutation on protein stability depends on the relative contributions to stability made by the resident and mutant amino acids (ϕ_α and ϕ_β, respectively), while the probability that a mutation will fix depends on the impact of that stability change on fitness, along with the current effective population size, as described by Kimura[26-28]. This at first seems to suggest that we need to determine the difference in fitnesses and the effective population size in order to calculate the fixation probability and the rate of substitution. However, such an agenda is compromised by epistatic interactions connecting the site of interest to other sites throughout the protein. As substitutions occur at these other sites, the stability contribution of a resident amino acid at a site fluctuates, as would the contribution of a new amino acid at that site resulting from a mutation, resulting in variations in the fixation probability that are difficult to predict. In this paper, we show that this complication leads to an even greater simplification. Occasionally, these fluctuations will equalise stability contributions among pairs of amino acids, in which case substitutions from one to the other are nearly neutral; the substitutions that occur under these conditions dominate the evolutionary process, shifting our focus from how to estimate the fitness change resulting from a substitution to calculating the fraction of the time that epistatic interactions make that substitution nearly neutral. Thus, although evolutionary mechanics theory fully incorporates population genetics theory and Kimura's equation for the probability of a substitution, systems near equilibrium do not require Kimura's formula to predict and explain substitution rates among amino acids. Fluctuations in contributions to stability cannot be ignored because they are the essential element necessary to create the conditions under which substitutions occur.

According to this new perspective, the relative rates of substitutions among different amino acids result from differences in the frequencies of nearly neutral conditions. Amino acids with similar physicochemical properties can make correlated contributions to stability. Such correlations will increase the probability of near-neutrality, providing a mechanistic explanation for higher rates of conservative change, a phenomenon first described by Fisher[32]. The multiplicity of interactions at buried sites increase the variances of ϕ_α and ϕ_β, reducing the probability of near neutrality and thus the substitution rate, as shown in Figs. 1A-1D, consistent with observed slower substitution rates observed at internal (buried) sites compared with external (exposed) sites.

Characterising the frequency of nearly neutral conditions requires an understanding of the joint distribution of ϕ_α and ϕ_β. This is complicated by the tendency for the resident amino acid (and similar amino acids) to be stabilised, what we call the 'evolutionary Stokes shift'. By developing a statistical mechanical view of protein evolution, this shift can be seen as a direct consequence of sequence entropy of folding. Increases in the stabilising contributions of an amino acid occupying a given site increase the affinity of that amino acid for that site. Given that total protein stability is approximately constant, this reduces the amount of stabilisation required from interactions among the remaining amino acids (the 'bath'). Because more sequences are able to fulfil this reduced stabilisation requirement, the contributions of the bath to the sequence entropy of folding is larger, and higher affinities for the resident amino acid are entropically favoured. In this context, the sequence adjusts to the current amino acid at the focal site in an analogous manner to the way that memory foam pillows adjust to the head of a slumberer, by distributing the air in the latex foam. The sequence adjustment acts to prevent non-conservative changes, and instead substitutions tend to occur preferentially between amino acids sharing physicochemical properties. Although describable as an adjustment, this evolutionary mechanism can be fully reversible, as are the simulations described here, with similar processes of moving into and away from the neutral zone[12]. These processes, called 'contingency' and 'entrenchment' by Plotkin and colleagues[13], are mirrors of each other, so that if the substitution were reversed the dissipation (entrenchment) process, played backwards, would have the same statistical properties as the pre-adaptation (contingency) process played forwards.

A key result is that the magnitude of the entropic stabilisation that drives the evolutionary Stokes shift is proportional to the variance of the underlying site-specific stability distribution in the absence of selection at the focal site times a protein-wide constant characterising the decrease in the number of sequences with increasing protein stability: the effect can be understood purely in terms of biophysics and sequence entropy. As with the average entropic stabilisation, the predicted average substitution rates *can be estimated solely based on these distributions and the mutation rates,* with *no adjustable parameters*. Surprisingly, the strength of selection and the effective population size do not affect the steady-state evolutionary Stokes shift, in agreement with theoretical predictions of the size-independence of substitution rates demonstrated in Supplementary Fig. S1[29,30]. Details of the protein structure, function, and context can influence these distributions, but otherwise do not affect the substitution rates, as long as the assumptions and approximations of the analysis remain valid. In particular, other forms of constant selection acting on the protein such as interactions with other proteins, ligand binding, catalysis, and avoiding proteolysis and aggregation would restrict the number of acceptable sequences and the form of the distributions, but would not otherwise affect the theory or calculations. Such additional selective pressures may also occasionally force adaptive, non-neutral substitutions if external pressures change. When an outside change compels such a substitution, an evolutionary Stokes shift still occurs, except the process is no longer reversible[12]. The interaction of fluctuating selection and fluctuating population size is another area requiring further investigation.

The pre- and post-adjustment of proteins to a substitution is explicitly time-dependent. Here, we addressed only the theoretical equilibrium predictions and results from simulations designed to be near equilibrium. Some discrepancies between the predicted and observed Stokes shifts for charged residues in buried sites, however, may be explained by time dependence. Individual sites at specific time points may be constrained by conserved neighbouring sites in the structure as well as the conserved structural context of their interactions with those sites. Such effects may influence the time-dependent probability of back mutations as well as subsequent substitutions, an important topic for further investigation.

We do not intend to imply that the theory developed here explains everything about molecular evolution, as we have considered only a simple situation of purifying selection in which fitness is based on the ability to fold and population size does not fluctuate. This theory provides an improved conceptual basis to understand what we should expect to happen in the absence of further complications, and should allow more accurate and confident prediction of non-structural functional constraints, adaptation, and fluctuating population sizes when they do occur. Although the current work is focused on fitness defined by protein stability, we expect other kinds of selection to fit into this framework, either by defining a large nearly neutral landscape in their own right, or by constraining the stability-based nearly neutral network.

In conclusion, the work described here sets up a theory of evolutionary mechanics, and simulations demonstrate that this theory can be used to predict substitution rates from the basic properties of how amino acids interact. As with the role of statistical mechanics in thermodynamics, we can apply the theory of evolutionary mechanics to understand how the microscopic events of evolutionary mechanics (mutation rates, fitness differences, and fixation probabilities) lead to the macroscopic events of molecular evolution (relative rates of substitution, and distributions of fluctuating rates across sequences and over time).

Methods

Simulations of protein evolution: The methods used to simulate protein evolution have been described previously[12,17,18]. Our simulations modelled proteins evolving under selection for a common requirement for globular proteins, stability of the native conformation. The free energy $G(\mathbf{X}, \mathbf{r})$ of a protein sequence $\mathrm{X} = \{x_1, x_2, x_3 \dots x_n\}$ in conformation $\mathbf{r}$ was calculated by summing the pairwise energies of amino acids in contact in that conformation, using the contact potentials derived by Miyazawa and Jernigan[33]. The free energy of folding $\Delta G_{\mathrm{Folding}}(\mathbf{X})$ was computed by first determining the free energy of the sequence in a pre-chosen native state, the conformation of the 300-residue purple acid phosphatase, PDB 1QHW[34]. The energies of unfolded states were assumed to follow a Gaussian distribution; the parameters characterising this distribution were estimated by calculating the free energies of the sequence in a widely diverse set of 55 different protein structures. The energy of the unfolded state was then calculated by assuming a large set (10^{160}) of possible unfolded structures with free energies drawn from this distribution. The free energy of folding $\Delta G_{\mathrm{Folding}}(\mathbf{X})$ was calculated as the difference between the two, and stability was $\Phi(\mathbf{X}) = -\Delta G_{\mathrm{Folding}}(\mathbf{X})$. The Malthusian fitness of a sequence $m(\mathbf{X})$ was defined as the fraction of that sequence that would be folded to the native state at equilibrium

$$m(\Phi(\mathbf{X})) = \frac{\exp\left(\frac{\Phi(\mathbf{X})}{T}\right)}{1 + \exp\left(\frac{\Phi(\mathbf{X})}{T}\right)} \tag{1}$$

where T is temperature in units of energy, 0.6 kcal mol^{-1}.

The simulations implemented a Gillespie algorithm[35] representing the evolution of a protein in the low mutation rate limit where the monomorphic population is represented by a single sequence. Starting from a single randomly chosen nucleotide sequence encoding a 300 amino-acid protein, we simulated evolution by considering in each step all possible nucleotide mutations with rates given by the K80 nucleotide model ($\kappa = 2$)[36]. The fixation probability of each mutation was calculated based on the Kimura formula for diploid organisms[26-28],

$$P_{\mathrm{Fix}}(\mathbf{X}, \mathbf{X}') = \frac{1 - \mathrm{e}^{-2(m(\Phi(\mathbf{X}')) - m(\Phi(\mathbf{X})))}}{1 - \mathrm{e}^{-4N_e(m(\Phi(\mathbf{X}')) - m(\Phi(\mathbf{X})))}} \tag{2}$$

where $\mathbf{X}$ and $\mathbf{X}'$ are the sequences before and after the mutation. The total substitution rate was set equal to the product of the mutation rate times the fixation probability, summed over all possible mutations. At each step, the evolutionary time was advanced by an amount chosen from an exponential distribution based on the total substitution rate, and one substitution was chosen to be fixed at random with relative probabilities determined by the product of the mutation rates times the acceptance probabilities.

Sequence evolution was simulated for a sufficient number of generations such that protein stability was roughly constant, representing mutation-drift selection balance. 100 equilibrated proteins were chosen, and two longer simulations were performed using each these equilibrated proteins as initial starting sequences, for a total of 200 simulations. The evolution of each lineage was simulated for an evolutionary distance of approximately seven amino acid replacements per amino acid position. The sequence and energy were sampled at regular time intervals.

Grouping of sites: For ease of analysis, we divided protein sites into four classes with similar substitution rates. Substitution matrices were calculated individually for each site; due to the length of simulations, we had on average over 1400 substitutions at each site. Sites were then clustered based on the off-diagonal elements of the substitution matrices using K-means clustering[37,38]. The resulting clusters were approximately equal in size, and class membership strongly depended on how buried or exposed sites were in the native state (as indicated by number of contacts). We ranked clusters by surface exposure, where class 1 is the most exposed and 4 is the most buried.

Calculating the site-specific contribution to protein stability: The site-specific contribution $\phi_{k,\alpha}(\mathbf{X}_{\not\ni \boldsymbol{k}})$ of amino acid α at focal site k as a function of the amino acids $\mathbf{X}_{\not\ni \boldsymbol{k}}$ at all sites excluding k is equal to $\Phi\{x_1, x_2, x_3 \dots x_{k-1}, \alpha, x_{k+1} \dots x_n\}$, the stability when the focal site is occupied by α, minus $\Phi\{x_1, x_2, x_3 \dots x_{k-1}, \emptyset, x_{k+1} \dots x_n\}$, the stability of a reference state when α is replaced by a non-interacting amino acid $\emptyset$, with the rest of the sequence and thus all other interactions unchanged. The part of the stability unaffected by this replacement is represented by the 'bath' interactions $\Phi_{k,\mathrm{Bath}}(\mathbf{X}_{\not\ni \boldsymbol{k}}) = \Phi(\mathbf{X}) - \phi_{k,\alpha}(\mathbf{X}_{\not\ni \boldsymbol{k}})$ so that $\Phi(\mathbf{X}) = \phi_{k,\alpha}(\mathbf{X}_{\not\ni \boldsymbol{k}}) + \Phi_{k,\mathrm{Bath}}(\mathbf{X}_{\not\ni \boldsymbol{k}})$.

Calculating the substitution rate integrating over distributions of local contributions: The average rate for substitution α → β at site k, $Q_{k,\alpha\to\beta}$, is equal to the neutral substitution rate $\boldsymbol{\upsilon}_{\boldsymbol{\alpha\to\beta}}$ times the average probability of fixation, which is a function of the stability of the protein before and after the substitution. The standard deviation of observed values of Φ, 0.71 kcal mol^{-1}, was small compared with the range of values of $\phi_{k,\alpha}(\mathbf{X}_{\not\ni \boldsymbol{k}})$, allowing us to represent the distribution Φ by its average, $\Phi \simeq \overline{\Phi}$ = 9.15 kcal mol^{-1}. We assumed that the stability before the substitution was $\overline{\Phi}$ and afterwards was $\overline{\Phi} + (\phi_{k,\beta}(\mathbf{X}_{\not\ni \boldsymbol{k}}) - \phi_{k,\alpha}(\mathbf{X}_{\not\ni \boldsymbol{k}}))$. The average substitution rate was then estimated as

$$Q_{k,\alpha\to\beta} = \nu_{\alpha\to\beta} \iint \frac{1 - \mathrm{e}^{-2\Delta m(\phi_{k,\alpha}, \phi_{k,\beta})}}{1 - \mathrm{e}^{-4N_e\, \Delta m(\phi_{k,\alpha}, \phi_{k,\beta})}} \rho\big(\phi_{k,\alpha}, \phi_{k,\beta} | x_k = \alpha\big)\, \mathrm{d}\phi_{k,\alpha}\, \mathrm{d}\phi_{k,\beta}. \tag{3}$$

where $\Delta m\big(\phi_{k,\alpha}, \phi_{k,\beta}\big) = m\left(\overline{\Phi} + \big(\phi_{k,\beta} - \phi_{k,\alpha}\big)\right) - m(\overline{\Phi})$ and $\rho(\phi_{k,\alpha}, \phi_{k,\beta} | x_k = \alpha)$ is the joint distribution of $\phi_{k,\alpha}(\mathbf{X}_{\not\ni k})$ and $\phi_{k,\beta}(\mathbf{X}_{\not\ni k})$ for the equilibrium distribution of sequences $\mathbf{X}_{\not\ni k}$ when α occupies site k.

Based on observations in Fig. 1, $\rho(\phi_{k,\alpha}, \phi_{k,\beta}|x_k = \alpha)$ was modeled as a bivariate normal distribution of the form $\rho(\phi_{k,\alpha}, \phi_{k,\beta}|x_k = \alpha) = \mathcal{N}(\bar{\phi}_{k,\alpha|\alpha}, \bar{\phi}_{k,\beta|\alpha}, \sigma^2_{k,\alpha|\alpha}, \sigma^2_{k,\beta|\alpha}, \varphi_{k,\alpha\beta|\alpha})$. Parameters were calculated directly from evolutionary simulation, and Equation **Error! Reference source not found.**) was integrated numerically. The neutral substitution rate was calculated using the same K80 nucleotide model ($\kappa = 2$)[36] as used in the simulation, with all non-nonsense codons considered equally likely.

Calculating the substitution rate integrating assuming only neutral substitutions: As observed in Fig. 1, substitutions generally occur in a neutral region in which $\mathbf{\Delta}\Phi_{k,\alpha\to\beta} = \phi_{k,\beta}(\mathbf{X}_{\not\ni k}) - \phi_{k,\alpha}(\mathbf{X}_{\not\ni k}) \approx 0$, so that

$$\frac{1 - \mathrm{e}^{-2\Delta m(\phi_{k,\alpha},\phi_{k,\beta})}}{1 - \mathrm{e}^{-4N_e\,\Delta m(\phi_{k,\alpha},\phi_{k,\beta})}} \approx \frac{1}{2\,N_e}. \tag{4}$$

This condition is satisfied in a band of width 2ε centred on $\phi_{k,\beta}(\mathbf{X}_{\not\ni k}) - \phi_{k,\alpha}(\mathbf{X}_{\not\ni k})$, where ε represents a deviation from strict neutrality that is sufficiently close for Equation (4) to be sufficiently accurate.

A natural scale for ε was obtained by considering the 'free fitness' $\Gamma(\Phi)$ of the protein equal to $\Gamma(\Phi) = m(\Phi) + \frac{S(\Phi)}{4N_e}$ where $S(\Phi)$ is the sequence entropy of folding, equal to the log of the number of sequences corresponding to a given total stability Φ[39,40]. Free fitness is analogous to thermodynamic free energy but with temperature T replaced by $4N_e$, and encompasses contributions from both fitness and sequence entropy to determine the distribution of states; evolutionary dynamics moves towards maximising this quantity. As the stability represents the sum of many small interactions, we would expect the distribution of stabilities to obey the central limit theorem and to resemble a Gaussian distribution. We are, however, on the tail of the distribution where the Gaussian is indistinguishable from an exponential, with one additional unidentifiable parameter. We instead assume $S(\Phi) = \ln(\Omega_0\, e^{-\gamma\Phi})$ where Ω_0 is constant. Noting that the system is at equilibrium with $\frac{\partial\Gamma(\Phi)}{\partial\Phi} = 0$ when $\Phi = \bar{\Phi}$, it can be demonstrated that

$$\left.\frac{\partial\, 4N_e m(\Phi)}{\partial\Phi}\right|_{\Phi=\bar{\Phi}} = \gamma \tag{5}$$

Thus, γ defines the rate of change of the population-weighted fitness $4N_e m(\Phi)$ with stability. Alternatively, a change in stability of $\frac{1}{\gamma}$ corresponds to a unit change in population-weighted fitness. In our calculations, we equated $\varepsilon = \frac{1}{\gamma}$; the estimation of γ is described below. Note that this calculation demonstrates that ε is, surprisingly, independent of effective population size N_e. This is a result of the balance between selection and mutational drift at equilibrium; for fixed effect of mutational drift, the degree of selection ($\frac{\partial m(\Phi)}{\partial \Phi}$) adjusts to changes in effective population size so that their product is constant[29,30].

If we assume that $\rho(\phi_{k,\alpha}, \phi_{k,\beta}|x_k = \alpha)$ is broader than ε, and that Equation (4) is satisfied, Equation **Error! Reference source not found.**) becomes

$$
\begin{aligned}
Q^{\mathrm{TST}}_{k,\alpha\to\beta} &= 2\varepsilon\, \upsilon_{\alpha\to\beta} \iint \rho\left(\phi_{k,\alpha}, \phi_{k,\beta}\middle|x_k = \alpha\right) \delta\left(\phi_{k,\alpha} - \phi_{k,\beta}\right) \mathrm{d}\phi_{k,\alpha}\, \mathrm{d}\phi_{k,\beta} \\
&= \frac{\upsilon_{\alpha\to\beta}}{\gamma} \frac{\exp\left(-\frac{(\bar{\phi}_{k,\alpha|\alpha} - \bar{\phi}_{k,\beta|\alpha})^2}{2(\sigma^2_{k,\alpha|\alpha} + \sigma^2_{k,\beta|\alpha} - 2\varphi_{k,\alpha\beta|\alpha}\sigma_{k,\alpha|\alpha}\sigma_{k,\beta|\alpha})}\right)}{\sqrt{2\pi\left(\sigma^2_{k,\alpha|\alpha} + \sigma^2_{k,\beta|\alpha} - 2\varphi_{k,\alpha\beta|\alpha}\sigma_{k,\alpha|\alpha}\sigma_{k,\beta|\alpha}\right)}}
\end{aligned}
\tag{6}
$$

where $\delta\left(\phi_{k,\alpha} - \phi_{k,\beta}\right)$ is the Dirac delta function.

For highly similar amino acids the entire distribution of $\rho\left(\phi_{k,\alpha}, \phi_{k,\beta}\middle|x_k = \alpha\right)$ may be contained in a region significantly narrower than the neutral zone, resulting in an overestimation of $Q_{k,\alpha\to\beta} > \upsilon_{\alpha\to\beta}$. For this reason, the estimated rate was capped at the neutral rate $\upsilon_{\alpha\to\beta}$.

Estimating $\boldsymbol{\rho\left(\phi_{k,\alpha}, \phi_{k,\beta}\right)}$: As described in the Results section, we approximate $\rho(\phi_{k,\alpha}, \phi_{k,\beta}|x_k = \alpha)$ as the product of two terms, $\rho_{\mathrm{Loc}}\left(\phi_{k,\alpha}, \phi_{k,\beta}\right) \times \rho_{\mathrm{Bath}}(\Phi_{k,\mathrm{Bath}} = \bar{\Phi} - \phi_{k,\alpha})$, where $\rho_{\mathrm{Loc}}(\phi_{k,\alpha}, \phi_{k,\beta})$ represents the fraction of sequences with given values of $\phi_{k,\alpha}$ and $\phi_{k,\beta}$ independently of how the rest of the protein adjusts to the current amino acid resident at site k, while $\rho_{\mathrm{Bath}}(\Phi_{k,\mathrm{Bath}} = \bar{\Phi} - \phi_{k,\alpha})$, represents the fraction of sequences where the bath interactions contribute sufficiently to the stability so that $\phi_{k,\alpha} + \Phi_{k,\mathrm{Bath}} = \bar{\Phi}$.

$\rho_{\mathrm{Loc}}\left(\phi_{k,\alpha}, \phi_{k,\beta}\right)$ was approximated by $\rho\left(\phi_{k,\alpha}, \phi_{k,\beta}\middle|x_k = \emptyset\right)$, the observed distribution observed when site k was occupied by a non-interacting amino acid $\emptyset$. We assumed that the contribution to the stability was small and approximated the distribution of Bath contributions with the distribution of total protein stabilities, $\rho_{\mathrm{Bath}}\left(\Phi_{k,\mathrm{Bath}} = \bar{\Phi} - \phi_{k,\alpha}\right) \simeq \rho_{\Phi}\left(\Phi_{k,\mathrm{Bath}} = \bar{\Phi} - \phi_{k,\alpha}\right) \propto \exp\left(-\gamma\left(\bar{\Phi} - \phi_{k,\alpha}\right)\right)$.

Because the number of possible sequences is immense, and because $\phi_{k,\alpha}$ and $\phi_{\mathrm{k},\beta}$ are the result of many interactions, the central limit theorem suggests that $\rho\left(\phi_{k,\alpha}, \phi_{k,\beta}\middle|x_k = \emptyset\right)$ can be approximated by a bivariate normal distribution $\rho\left(\phi_{k,\alpha}, \phi_{k,\beta}\middle|x_k = \emptyset\right) \propto \mathcal{N}\{\bar{\phi}_{\mathrm{k},\alpha|\emptyset}, \bar{\phi}_{\mathrm{k},\beta|\emptyset}, \sigma^2_{\mathrm{k},\alpha|\emptyset}, \sigma^2_{\mathrm{k},\beta|\emptyset}, \varphi_{\mathrm{k},\alpha\beta|\emptyset}\}$. The normalised product of $\rho\left(\phi_{k,\alpha}, \phi_{k,\beta}\middle|x_k = \emptyset\right)$ and $\rho_{\mathrm{Bath}}\left(\Phi_{k,\mathrm{Bath}} = \bar{\Phi} - \phi_{k,\alpha}\right) \propto \exp\left(-\gamma\left(\bar{\Phi} - \phi_{k,\alpha}\right)\right)$ results in an estimated shifted bivariate normal distribution $\tilde{\rho}_{k,\alpha}\left(\phi_{k,\alpha}, \phi_{k,\beta}\right) = \mathcal{N}\left\{\tilde{\phi}_{k,\alpha|\alpha}, \tilde{\phi}_{k,\beta|\alpha}, \tilde{\sigma}^2_{k,\alpha|\alpha}, \tilde{\sigma}^2_{k,\beta|\alpha}, \tilde{\varphi}_{k,\alpha\beta|\alpha}\right\}$ with

$$
\begin{gathered}
\widetilde{\Phi}_{k,\alpha|\alpha} = \overline{\Phi}_{\mathrm{k},\alpha|\emptyset} + \gamma\sigma^2_{k,\alpha|\emptyset} \\
\widetilde{\sigma}^2_{k,\alpha|\alpha} = \sigma^2_{k,\alpha|\emptyset} \\
\widetilde{\Phi}_{k,\beta|\alpha} = \overline{\Phi}_{\mathrm{k},\beta|\emptyset} + \gamma\,\varphi_{\mathrm{k},\alpha\beta|\emptyset}\,\sigma_{k,\alpha|\emptyset}\,\sigma_{k,\beta|\emptyset} \\
\widetilde{\sigma}^2_{k,\beta|\alpha} = \sigma^2_{k,\beta|\emptyset} \\
\widetilde{\varphi}_{k,\alpha\beta|\alpha} = \varphi_{k,\alpha\beta|\emptyset}
\end{gathered}
\tag{7}
$$

Substituting these results into Equation (6) yields

$$
Q^{\mathrm{TST},\emptyset}_{k,\alpha\to\beta} = \frac{\upsilon_{\alpha\to\beta}}{\gamma}\frac{\exp\left(-\frac{\left(\overline{\Phi}_{\mathrm{k},\alpha|\emptyset} - \overline{\Phi}_{\mathrm{k},\beta|\emptyset} + \gamma\sigma^2_{k,\alpha|\emptyset}\left(1-\varphi_{\mathrm{k},\alpha\beta|\emptyset}\,\frac{\sigma_{k,\beta|\emptyset}}{\sigma_{k,\alpha|\emptyset}}\right)\right)^2}{2(\sigma^2_{k,\alpha|\emptyset} + \sigma^2_{k,\beta|\emptyset} - 2\varphi_{k,\alpha\beta|\emptyset}\sigma_{k,\alpha|\emptyset}\sigma_{k,\beta|\emptyset})}\right)}{\sqrt{2\pi\left(\sigma^2_{k,\alpha|\emptyset} + \sigma^2_{k,\beta|\emptyset} - 2\varphi_{k,\alpha\beta|\emptyset}\sigma_{k,\alpha|\emptyset}\sigma_{k,\beta|\emptyset}\right)}}
\tag{8}
$$

Characterising the bath state distribution: As described above, we assume that the number of protein sequences with a given value of Φ in the range of interest around $\Phi = \overline{\Phi}$ is approximately exponential $\Omega(\Phi) \sim e^{-\gamma\Phi}$. We estimated γ from the average change in stability resulting from random mutations, $\langle \rho_{\mathrm{mut}}(\Delta\Phi)\rangle$, which is negative due to the greater number of sequences coding for proteins with lower stability. This suggests that by correcting for the dependence of Ω on Φ by multiplying $\rho_{\mathrm{mut}}(\Delta\Phi)$ and $e^{\gamma\Delta\Phi}$, this bias would disappear. We adjusted γ so that $\langle \Delta\Phi e^{\gamma\Delta\Phi}\rangle = 0$ where the average was over all possible mutations during the simulations, yielding $\gamma = 1.26$ (kcal mol^{-1})$^{-1}$.

The bath state distribution determines the equilibrium stabilities through Equation (5). Substituting Equation (1) into Equation (5) yields $\overline{\Phi} \approx T \ln\left(\frac{4N_e}{\gamma T}\right)$. This expression results in estimations for $\overline{\Phi}$ of 6.53, 9.27, and 12.05 for N_e equal to 10^4, 10^6, and 10^8, respectively. These agree well with the average of the distributions shown in Supplementary Figure S1: 6.40, 9.15, and 11.90.

We note that under this model, the population scaled fixed load $2N_e\left(1 - m(\overline{\Phi})\right)$ is equal to

$$
2N_e\left(1 - m(\overline{\Phi})\right) = 2N_e\left(\frac{1}{1 + \left(\frac{4N_e}{\gamma T}\right)}\right) \approx \frac{\gamma T}{2}
\tag{9}
$$

that is, it only depends on the dependence of the sequence entropy on the stability and the temperature. For our system, $2N_e\left(1 - m(\overline{\Phi})\right) \approx 0.38$.

Data Availability: Data will be made available on Dryad, including structures used, contact potentials, tables of outcomes, and raw program data output.

Code Availability: All simulations and analysis software will be made available on GitHub.

Acknowledgments

We thank Bhavin Khatri for helpful discussions. We acknowledge the support of the Medical Research Council (UK) (MC_U117573805) to RAG and the National Institutes of Health (NIH; GM083127 and GM097251) to DDP.

References

1 Muse, S. V. & Gaut, B. S. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol Biol Evol* **11**, 715-724 (1994).

2 Nielsen, R. & Yang, Z. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**, 929-936 (1998).

3 Tamuri, A. U., dos Reis, M., Hay, A. J. & Goldstein, R. A. Identifying Changes in Selective Constraints: Host Shifts in Influenza. *Plos Comput Biol* **5**, e1000564, doi:Artn E1000564, Doi 10.1371/Journal.Pcbi.1000564 (2009).

4 Castoe, T. A. *et al.* Evidence for an ancient adaptive episode of convergent molecular evolution. *Proc Natl Acad Sci USA* **106**, 8986-8991 (2009).

5 Goldstein, R. A., Pollard, S. T., Shah, S. D. & Pollock, D. D. Nonadaptive Amino Acid Convergence Rates Decrease over Time. *Mol Biol Evol* **32**, 1373-1381, doi:10.1093/molbev/msv041 (2015).

6 Kondrashov, A. S., Sunyaev, S. & Kondrashov, F. A. Dobzhansky-Muller incompatibilities in protein evolution. *Proc Natl Acad Sci USA* **99**, 14878-14883, doi:10.1073/pnas.232565499 (2002).

7 Halpern, A. L. & Bruno, W. J. Evolutionary distances for protein-coding sequences: modeling site-specific residue frequencies. *Mol Biol Evol* **15**, 910-917 (1998).

8 Koshi, J. M. & Goldstein, R. A. Context-dependent optimal substitution matrices. *Protein Eng* **8**, 641-645 (1995).

9 Koshi, J. M., Mindell, D. & Goldstein, R. A. Using physical-chemistry based mutation models in phylogenetic analyses of HIV-1 subtypes. *Mol Biol Evol* (1999).

10 Lartillot, N. & Philippe, H. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol* **21**, 1095-1109 (2004).

11 Goldstein, R. A. & Pollock, D. D. The tangled bank of amino acids. *Protein Sci*, doi:10.1002/pro.2930 (2016).

12 Pollock, D. D., Thiltgen, G. & Goldstein, R. A. Amino acid coevolution induces an evolutionary Stokes shift. *Proc Natl Acad Sci USA* **109**, E1352-1359, doi:10.1073/pnas.1120084109 (2012).

13 Shah, P., McCandlish, D. M. & Plotkin, J. B. Contingency and entrenchment in protein evolution under purifying selection. *Proc Natl Acad Sci USA* **112**, E3226-3235, doi:10.1073/pnas.1412933112 (2015).

14 Pollock, D. D., Zwickl, D. J., McGuire, J. A. & Hillis, D. M. Increased taxon sampling is advantageous for phylogenetic inference. *Syst Biol* **51**, 664-671 (2002).

15 Kimura, M. The role of compensatory neutral mutations in molecular evolution. *Journal of Genetics* **64**, doi:doi:10.1007/BF02923549 (1985).

16 Izaguirre, J. A. *et al.* CompuCell, a multi-model framework for simulation of morphogenesis. *Bioinformatics* **20**, 1129-1137, doi:10.1093/bioinformatics/bth050 (2004).

17 Goldstein, R. A. The evolution and evolutionary consequences of marginal thermostability in proteins. *Proteins* **79**, 1396-1407, doi:10.1002/prot.22964 (2011).

18 Williams, P. D., Pollock, D. D., Blackburne, B. P. & Goldstein, R. A. Assessing the accuracy of ancestral protein reconstruction methods. *PLoS Comput Biol* **2**, e69, doi:10.1371/journal.pcbi.0020069 (2006).

19 Privalov, P. L. Stability of proteins: small globular proteins. *Adv Protein Chem* **33**, 167-241 (1979).

20 Privalov, P. L. & Gill, S. J. Stability of protein-structure and hydrophoboc interaction. *Advances in Protein Chemistry* **39**, 191-234 (1988).

21 Taverna, D. M. & Goldstein, R. A. Why are proteins marginally stable? *Proteins* **46**, 105-109 (2002).

22 Zeldovich, K. B. & Shakhnovich, E. I. Understanding protein evolution: from protein physics to Darwinian selection. *Annu Rev Phys Chem* **59**, 105-127, doi:10.1146/annurev.physchem.58.032806.104449 (2008).

23 Iwasa, Y. Free fitness that always increases in evolution. *J Theor Biol* **135**, 265-281 (1988).

24 Sella, G. & Hirsh, A. E. The application of statistical physics to evolutionary biology. *Proc Natl Acad Sci USA* **102**, 9541-9546 (2005).

25 Shenkin, P. S., Erman, B. & Mastrandrea, L. D. Information-theoretical entropy as a measure of sequence variability. *Proteins* **11**, 297-313, doi:10.1002/prot.340110408 (1991).

26 Crow, J. F. & Kimura, M. *An introduction to population genetics theory*. (Harper & Row, 1970).

27 Kimura, M. Some problems of stochastic processes in genetics. *Ann. Math. Stat.* **28**, 882–901 (1957).

28 Kimura, M. On the probability of fixation of mutant genes in a population. *Genetics* **47**, 713-719 (1962).

29 Goldstein, R. A. Population size dependence of fitness effect distribution and substitution rate probed by biophysical model of protein thermostability. *Genome Biol Evol* **5**, 1584-1593, doi:10.1093/gbe/evt110 (2013).

30 Cherry, J. L. Should we expect substitution rate to depend on population size? *Genetics* **150**, 911-919 (1998).

31 Eyring, H. The Activated Complex in Chemical Reactions. *J Chem Phys* **3**, 107-115 (1935).

32 Fisher, R. *The Genetic Theory of Natural Selection*. (Oxford University Press, 1930).

33 Miyazawa, S. & Jernigan, R. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules* **18**, 534-552 (1985).

34 Lindqvist, Y., Johansson, E., Kaija, H., Vihko, P. & Schneider, G. Three-dimensional structure of a mammalian purple acid phosphatase at 2.2 A resolution with a mu-(hydr)oxo bridged di-iron center. *Journal of Molecular Biology* **291**, 135-147 (1999).

35 Gillespie, D. T. Exact Stochastic Simulation of Coupled Chemical Reactions. *J Phys Chem* **81**, 2340-2361 (1977).

36 Kimura, M. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *J Mol Evol* **16**, 111-120 (1980).

37 Forgy, E. W. Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics* **21**, 768-769 (1965).

38 Lloyd, S. Least squares quantization in PCM. *IEEE Transactions on Information Theory* **28**, 129-137, doi:10.1109/TIT.1982.1056489 (1982).

39 Khatri, B. S. & Goldstein, R. A. A coarse-grained biophysical model of sequence evolution and the population size dependence of the speciation rate. *J Theor Biol* **378**, 56-64, doi:10.1016/j.jtbi.2015.04.027 (2015).

40 Khatri, B. S., McLeish, T. C. & Sear, R. P. Statistical mechanics of convergent evolution in spatial patterning. *Proc Natl Acad Sci USA* **106**, 9564-9569, doi:10.1073/pnas.0812260106 (2009).