Consequences of Stability-Induced Epistasis for Substitution Rates

Noor Youssef,*[1,2] Edward Susko,[2,3] and Joseph P. Bielawski[1,2,3]

[1]Department of Biology, Dalhousie University, Halifax, Nova Scotia, Canada

[2]Centre for Genomics and Evolutionary Bioinformatics, Dalhousie University, Halifax, Nova Scotia, Canada

[3]Department of Mathematics and Statistics, Dalhousie University, Halifax, Nova Scotia, Canada

***Corresponding author**: E-mail: n.youssef@dal.ca.

Associate editor: Jeffrey Thorne

Abstract

Do interactions between residues in a protein (i.e., epistasis) significantly alter evolutionary dynamics? If so, what consequences might they have on inference from traditional codon substitution models which assume site-independence for the sake of computational tractability? To investigate the effects of epistasis on substitution rates, we employed a mechanistic mutation-selection model in conjunction with a fitness framework derived from protein stability. We refer to this as the stability-informed site-dependent (S-SD) model and developed a new stability-informed site-independent (S-SI) model that captures the average effect of stability constraints on individual sites of a protein. Comparison of S-SI and S-SD offers a novel and direct method for investigating the consequences of stability-induced epistasis on protein evolution. We developed S-SI and S-SD models for three natural proteins and showed that they generate sequences consistent with real alignments. Our analyses revealed that epistasis tends to increase substitution rates compared with the rates under site-independent evolution. We then assessed the epistatic sensitivity of individual site and discovered a counterintuitive effect: Highly connected sites were less influenced by epistasis relative to exposed sites. Lastly, we show that, despite the unrealistic assumptions, traditional models perform comparably well in the presence and absence of epistasis and provide reasonable summaries of average selection intensities. We conclude that epistatic models are critical to understanding protein evolutionary dynamics, but epistasis might not be required for reasonable inference of selection pressure when averaging over time and sites.

Key words: **epistasis, d*N*/dS, protein stability, substitution rates, mutation-selection model, protein evolution.**

Introduction

Most proteins must fold into a native structure in which they are moderately stable before they are able to perform their biological function. Protein stability depends on the sequence of amino acids and their interactions in the folded three-dimensional structures. Because of these interactions, evolutionary selective constraints to maintain adequate stability result in epistatic dependencies between residues. Specifically, epistasis manifests as a dependency in the fitness effect of a mutation on the background protein sequence in which it arose. For example, let $f_a^h(S)$ be the fitness of the protein provided amino acid a is occupying site h in the context of background sequence S. Then, $F^h(S) = \langle f_1^h(S), \ldots, f_{20}^h(S)\rangle$ is the site-specific vector of amino acid fitness values specifying the fitness landscape at site h. Following a substitution at another position in the protein, so that the background sequence changes from S to X, the fitness of the same amino acid will subsequently change, $f_a^h(S) \neq f_a^h(X)$. Therefore, in the presence of epistatic dependencies, the fitness landscape at a site is subject to fluctuations as substitutions occur at other sites (fig. 1A). Stability constraints typically result in global epistasis, meaning that a change in the incumbent amino acid at one site induces shifts in the fitness landscapes at many, often all, other sites in the protein (Starr and Thornton 2016). Although such interdependencies inevitably occur, the magnitude and frequency of these shifts, and their impact on protein evolution, remain controversial.

Using extensive computational experiments, Pollock et al. (2012) found that stability-induced epistasis results in frequent and substantial shifts in amino acid fitness landscapes. To the contrary, Ashenberg et al. (2013) used computational and experimental approaches and reported that although stability-induced fluctuations in site-specific amino acid fitness landscapes do occur, they are relatively minor in magnitude and are therefore inconsequential with regards to long-term evolutionary dynamics. This controversy has spurred multiple follow-up experiments, finding support for both claims and little consensus (Risso et al. 2015; Shah et al. 2015; Starr et al. 2018; Ferrada 2019). It remains unclear if and how stability-induced epistasis influences protein evolution.

Models used to infer evolutionary parameters from natural protein alignments commonly assume site-independence and other simplifying assumptions (e.g., time-stationary substitution rates, and low levels of among-site rate

doi:10.1093/molbev/msaa151 Advance Access publication September 8, 2020

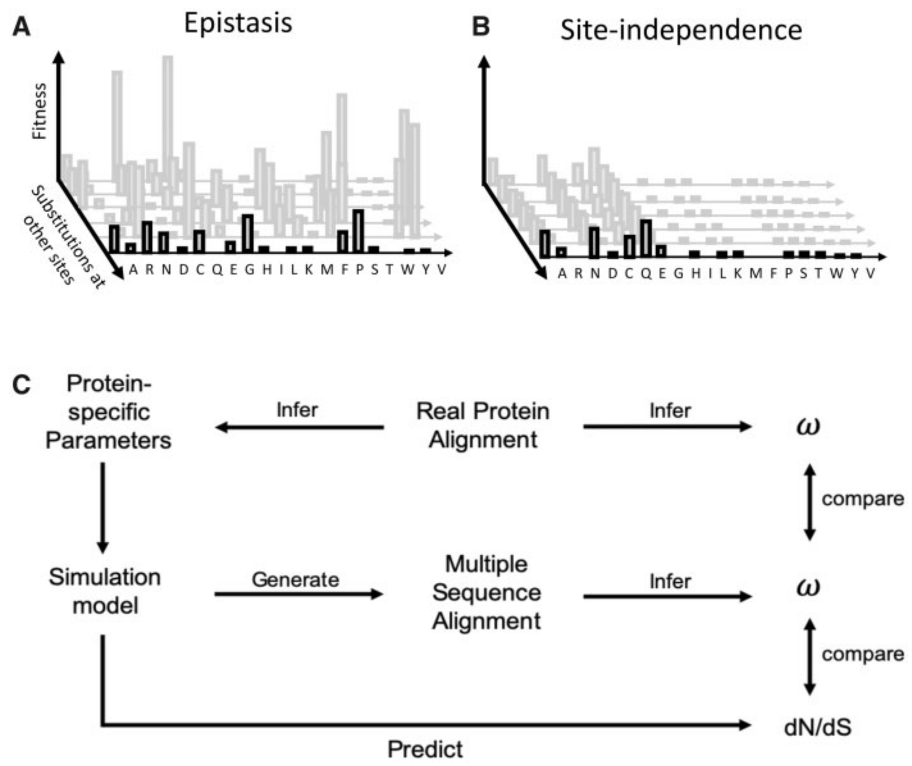
A
Epistasis
Fitness
Substitutions at other sites
A R N D C Q E G H I L K M F P S T W Y V
B
Site-independence
A R N D C Q E G H I L K M F P S T W Y V
C
Protein-specific Parameters
Infer
Real Protein Alignment
Infer
ω
compare
Simulation model
Generate
Multiple Sequence Alignment
Infer
ω
compare
dN/dS
Predict

Fig. 1. Site-specific fitness landscape dynamics (*A*, *B*) and flowchart of method design (*C*). (*A*) Epistasis results in a changing site-specific fitness landscape as substitutions occur at other positions in the protein. (*B*) Site-independent evolution implies a static (constant) fitness landscape. (*C*) Real protein alignments were fitted to M-series models to obtain maximum likelihood estimates of substitution rates (ω) and estimates of protein-specific parameters (phylogeny, κ, π_A, π_C, π_G, and π_T). The protein-specific parameters were then used to generate 50 alignments under each of the simulation models: C-SI, S-SI, and S-SD. The validity of the simulation model was assessed by comparing the inferred ω rates from the simulated alignments to the ω estimates from the corresponding real protein alignment. To assess the performance of inference models, expected substitution rates, dN^h/dS^h, were calculated directly from the simulation models and compared with the inferred ω values. Diagram modified from Spielman and Wilke (2015).

heterogeneity) for the sake of computational tractability. In this study, we focus on the widely used codon substitution models which infer selection pressure as ω, the normalized ratio of nonsynonymous substitutions to the ratio of synonymous substitutions (Goldman and Yang 1994; Muse and Gaut 1994); we refer to these as ω-based models. Natural proteins evolve under complex evolutionary dynamics that are not entirely captured by traditional ω-based models (e.g., epistatic interactions between sites). If epistasis between residues in a protein does have a dramatic effect on protein evolution, then the validity of inference from site-independence models might be negatively impacted.

Does epistasis substantially influence the rate at which proteins evolve? And if so, how reliable are inferences from traditional ω-based models which assume that sites evolve independently? Addressing these questions is our main objective. To do this, we model the evolutionary process from first principles of population genetics theory using the mutation-selection (MutSel) framework (Halpern and Bruno 1998; Yang and Nielsen 2008). Unlike ω-based models, MutSel models account for differences in amino acids fitness values and allow for more realistic levels of among-site rate heterogeneity by assigning each site a unique fitness landscape (F^h). MutSel frameworks are commonly used as a method for simulating plausible evolutionary dynamics (Rodrigue et al. 2010; Ashenberg et al. 2013; Spielman and Wilke 2015; Jones et al. 2017, 2018, 2020). These are site-independent models and therefore do not directly model the dynamics of epistasis. With appropriate fitnesses, they can in theory be used to model the marginal effects of stability and/or other selective pressures on a site. The challenge then lies in determining plausible site-specific fitness landscapes.

Several ways of calculating amino acid fitness values have been proposed. For example, Spielman and Wilke (2015) derived amino acid fitness values based on empirical site-specific frequencies from large alignments of homologous proteins. Alternatively, Jones et al. (2018) assigned amino acid fitness values such that the estimated probability density function of the scaled fitness effects ($2N_e[f_x - f_y]$ for amino acids x and y and effective population size N_e) matches the distribution inferred from empirical data. Hereafter, these approaches are referred to as site-wise MutSel. Under the site-wise MutSel formulations, site-specific fitness landscapes average the selective pressure acting on a site, assuming site-independent evolution, and therefore time-stationary fitness landscapes (fig. 1*B*). Changes in site-specific fitness landscape are interpreted as a change in selection pressure (due to either a change in environment or a change in protein function).

Determining fitness landscapes has also been addressed mechanistically by combining the MutSel approach with biophysical models of protein folding where fitness values depend on protein stability or the proportion of correctly folded proteins at thermodynamic equilibrium (Pollock et al. 2012; Ashenberg et al. 2013; Goldstein and Pollock 2016, 2017). Although comparable at first glance, the biophysical approaches differ extensively from the site-wise MutSel applications. Importantly, the biophysical models account for temporal variation in site-specific fitness landscapes that emerges as a consequence of global stability-induced epistasis (fig. 1A). Accounting for these temporal dynamics is essential for understanding how epistasis influences protein evolution. Although the evolution of natural proteins is certainly shaped by additional structural and functional constraints, for most proteins, proper folding into a native structure is prerequisite to being able to carry out their biological function.

To investigate the influence of epistasis on protein substitution rates, we use the MutSel evolutionary model in conjunction with a biophysical model of protein folding. We refer to this as the stability-informed site-dependent (S-SD) model since stability calculations inherently take into account epistatic interactions between sites. We develop an analogous stability-informed site-independent (S-SI) model where proteins evolve under equivalent stability-mediated selection pressures but having independent and constant fitness landscapes (fig. 1B). Specifically, from each S-SD evolutionary simulation, we calculated the average fitness landscapes at each site over different background sequences. We then use these site-specific average landscapes as the unique and constant landscapes for each site in the S-SI simulations (fig. 2). Therefore, for each S-SD alignment, we generated an analogous S-SI alignment under the same average selection constraints but without the temporal dynamics characteristic of epistasis. The S-SI versus S-SD model comparison allows for a novel and direct way of investigating the influence of stability-induced epistasis on evolutionary dynamics. To permit comparison with models that do not account for stability, we include a third independent and identically distributed across sites framework where site-specific fitness landscapes are derived from the C-series frequency profiles (Quang et al. 2008); we refer to this as the C-series site-independent (C-SI) model.

The conditions of our simulations are derived from multiple sequence alignments for three natural protein-coding genes with PDB structures 1QHW, 2PPN, and 1PEK. The three protein structures differ in important ways. The 2PPN protein folds following a two-state folding process and therefore conforms to one of core thermodynamic model assumptions. The 1QHW structure was used to maintain consistency with previous studies which used the same structure (Pollock et al. 2012; Goldstein and Pollock 2016, 2017). Lastly, the 1PEK protein is comparable in length to the 1QHW protein, however, the 1PEK protein is more densely packed. We begin by validating the stability-informed models and show that simulated alignments are phenomenologically comparable to the real protein alignments based on various metrics. We then use the S-SI and S-SD models to investigate the difference in dynamics when sites evolve with epistatic interactions or independently. We find that epistasis results in minor elevations in substitution rates over the whole protein. However, site-wise analysis reveals that the impact of epistatic interactions on substitution rates can be substantial at individual sites. We describe a mechanism whereby epistasis increases substitution rates compared with the rates under site-independent evolution. Lastly, we report that although models that treat site-wise variation in ω as a random variable underestimate the degree of among-site rate heterogeneity, the estimated ω rates tend to accurately identify the most common substitution rates across sites. Despite their simplicity, ω-based inference models preformed comparably well in the presence and absence of epistasis.

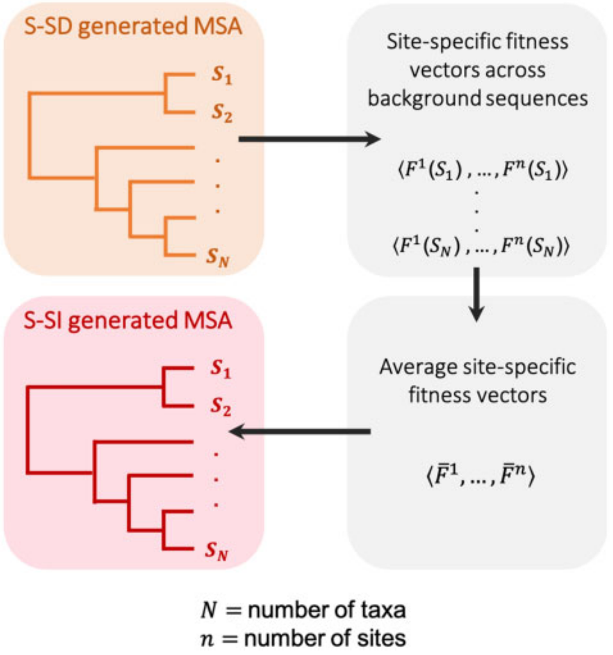
S-SD generated MSA
S_1
S_2
S_N
Site-specific fitness vectors across background sequences
$\langle F^1(S_1), \ldots, F^n(S_1)\rangle$
$\langle F^1(S_N), \ldots, F^n(S_N)\rangle$
S-SI generated MSA
S_1
S_2
S_N
Average site-specific fitness vectors
$\langle \bar{F}^1, \ldots, \bar{F}^n\rangle$
N = number of taxa
n = number of sites

FIG. 2. Derivation of the stability-informed site-independent (S-SI) model. First, we generated multiple sequence alignments (MSA) under the epistatic stability-informed site-dependence (S-SD) model (see Materials and Methods section for details). Then, at each site, we calculated $F^h(S) = \langle f_1^h(S), \ldots, f_{20}^h(S)\rangle$, the site-specific fitness vector where $f_a^h(S)$ is the fitness of amino acid a at site h given background sequence S. This was repeated across all extant sequences $S_1, \ldots, S_N$. Next, we calculated $\bar{F}^h$ the average fitness landscape at site h across background sequences. We generate under S-SI with $\bar{F}^h$ as the independent and constant fitness landscapes (see Materials and Methods section for details). N is the number of taxa in the protein-specific alignment (14, 14, and 12 for proteins 1QHW, 2PPN, and 1PEK), and n is the number of sites in the protein-specific alignment (300, 107, and 279 for proteins 1QHW, 2PPN, and 1PEK).

Results

Stability-Informed Models Generate Sequence Alignments Consistent with Real Data

Evaluating the Relationship between Substitution Rates and Structural Features

Buried residues, toward the core of the protein, are more densely packed having higher weighted contact number (WCN) and lower relative solvent accessibility (RSA) compared with surface residues. Analyses of natural protein alignments often reveal significant correlations between site-specific substitution rates and structural properties such as RSA and WCN: Buried sites tend to be more conserved with lower substitution rates compared with exposed sites (Shahmoradi et al. 2014; Yeh et al. 2014; Echave et al. 2015; Marcos and Echave 2015). We were interested in assessing if any of the generative models recapitulate this phenomenon. We measured the expected site-specific substitution rate (dN^h/dS^h) directly from the fitness landscapes using equation (4) for C-SI and S-SI and equation (5) for the S-SD. We refer to dN^h/dS^h as the expected substitution rate throughout the study since it represents the theoretically predicted substitution rate at evolutionary equilibrium (Spielman and Wilke 2015).

Under both stability-informed frameworks (S-SI and S-SD), a significant positive correlation was found between RSA and dN^h/dS^h, and a significant negative correlation was found between WCN and dN^h/dS^h (fig. 3). The correlations between RSA and dN^h/dS^h were slightly higher for rates predicted under the S-SD framework compared with the correlations based on the S-SI simulations. Similarly, correlations between WCN and rates predicted under the S-SD were more negative compared with rates predicted under S-SI. In contrast, the site-specific rates expected under the C-SI framework did not correlate significantly with RSA or WCN. Correlations and P values between structural properties and dN^h/dS^h are reported in supplementary table S1, Supplementary Material online.

Since the true substitution rates are unknown for the natural proteins, we used traditional codon models to infer substitution rates ω, measured as the normalized ratio of nonsynonymous to synonymous substitutions. The ω-based codon models use the maximum likelihood framework to estimate rate parameter (ω) conditioned on a known phylogeny and multiple sequence alignment. Briefly, the M-series ω-based models partition sites into k categories and estimate substitution rates $\omega_1 < .. < \omega_k$, and proportions $p_1, \ldots, p_k$ (Yang et al. 2000) (the models are described in more detail in the Materials and Methods section). In order to assess the correlation between RSA (and WCN) and substitution rates in our real alignments, we use the posterior mean ω^h from the best fitting M-series model as the site-specific rate estimate. The posterior mean ω^h at a site is calculated as $(\omega_1 \times P_1^h) + (\omega_2 \times P_2^h) + \cdots + (\omega_k \times P_k^h)$, where P_k^h is the posterior probability of the site corresponding to rate class ω_k. We found a significant positive correlation between posterior mean ω^h and RSA in the 1QHW and 1PEK real protein alignments (correlation coefficient was 0.39 and 0.53, respectively; both P values <1.0e-10) and a significant negative correlation between rates and WCN (correlation coefficient was −0.35 and −0.43 for the 1QHW and 1PEK alignments, respectively; both P values <1.0e-10). We found no significant correlation between rates and structural properties (RSA or WCN) for the 2PPN alignment. The small size of the 2PPN gene, and the unusual mixture of long and short edges in its phylogeny (supplementary fig. S1, Supplementary Material online), is likely problematic for posterior estimation of ω, which could explain the insignificant correlations.

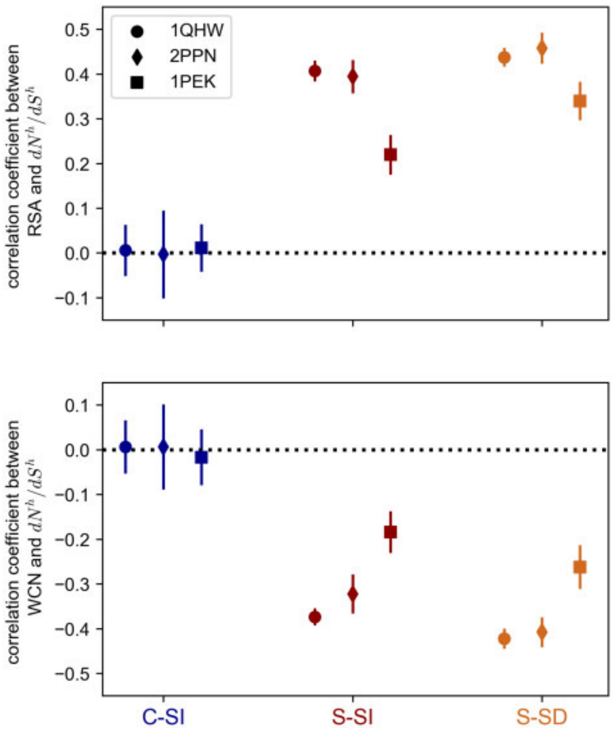

FIG. 3. Stability-informed models (S-SI and S-SD) reproduce empirically observed correlations between substitution rates and structural features. Fifty alignments were generated with three protein-specific parameters (1QHW, 2PPN, and 1PEK) under models C-SI, S-SI, and S-SD. For each alignment, we calculated the Pearson correlation between the expected site-specific substitution rates dN^h/dS^h and RSA (top panel) and WCN (bottom panel). Plotted are the mean correlation coefficients (and standard deviation) across trials.

Various alternative methods have been developed to infer site-specific substitution rates from multiple sequence alignments (e.g., Meyes and vonHaeseler 2003; Kosakovsky Pond and Frost 2005; Massingham and Goldman 2005; Murrell et al. 2012). However, the estimated rates are subject to large variability when the number of taxa is relatively small. These methods are therefore not suitable to infer site-specific rates for the alignments used here (number of taxa = 14, 14, and 12 for 1QHW, 2PPN, and 1PEK). Using large alignments (number of taxa = 300) of more than 200 proteins, Marcos and Echave (2015) estimated the correlations between rates and RSA and between rates and WCN. The range of correlations coefficients between RSA and site-specific rates was between 0.26 and 0.75; the range of correlation coefficients between WCN and site-specific rates was −0.19 and −0.73. The correlation coefficients we report for both rate measures (dN^h/dS^h and posterior mean ω^h) are within the range reported in Marcos and Echave (2015). Overall, we found that the stability-informed models are able to recapitulate the empirically observed correlations between structural properties and rates, which suggests that accounting for folding stability captures important structural features that are absent in the stability-naïve C-SI framework derived from the widely used C-series profiles.

Comparing Inferred Substitution Rates and Sequence Variability between Real and Simulated Data

In order to use the simulations as a means of investigating the influence of epistasis on rates, we needed to first verify that the generative models produce plausible substitution rates. In other words, we needed to compare substitution rates from the generative models with the rates experienced by real proteins. We fitted simulated and real alignments to codon model M3 ($k = 2$) to obtain estimates of substitution rates. A value of $\omega \approx 1$ is indicative of neutral or nearly neutral evolution where nonsynonymous mutations are fixed at an equal rate to synonymous mutations. An ω value <1 is representative of purifying selection, and $\omega > 1$ is indicative of positive selection.

Analyses of the natural 1QHW, 2PPN, and 1PEK alignments revealed evidence for purifying selection with $\omega_1 < \omega_2 < 1$ for all three natural alignments (fig. 4). The maximum likelihood estimates are reported in supplementary table S2, Supplementary Material online. The 2PPN protein alignment had the lowest rate estimates with $\omega_1 = 0.00$ and $\omega_2 = 0.09$ and respective proportions $p_1 = 0.67$ and $p_2 = 0.33$. The 1QHW and 1PEK alignments had comparable rate estimates with $\omega_1 = 0.01$ and 0.02 and $\omega_2 = 0.30$ and 0.24, respectively; however, the proportion of sites belonging to the more stringent selection regime (ω_1) was ~10% higher for the 1QHW alignment ($p_1 = 0.71$) compared with the 1PEK alignment ($p_1 = 0.64$).

Alignments generated under the stability-informed models (S-SI and S-SD) were also consistent with purifying selection, with $\omega_1 < \omega_2 < 1$ for all simulated protein-specific alignments (fig. 4, first row). The average maximum likelihood estimates are reported in supplementary table S3, Supplementary Material online. The ω values inferred from the S-SI-generated alignments were on average significantly lower than rates estimated from the analogous protein-specific S-SD simulations and more consistent with the ω values estimated from the natural protein alignments (fig. 4; Bonferroni corrected P values <1.0e-05 for all comparisons, supplementary table S4, Supplementary Material online). With the exception of the 1PEK protein, the natural alignments were consistently inferred to be under more stringent selection regimes with slightly lower substitution rates. For the 1PEK simulations, the ω_2 estimate from the real alignment ($\omega_2 = 0.24$) alignment was higher than the distribution of estimates from the S-SI alignments (fig. 4, first row). Nonetheless, the proportion of quickly evolving sites (p_2) was lower in the real alignment (fig. 4, second row). This suggests that in the real 1PEK protein, a small proportion of sites are evolving faster than expected under stability constraints. However, when considering all sites in the alignment, by comparing the single rate estimated under M0, we find that the rates are largely comparable: ω was 0.06 for the real 1PEK alignment and the mean ω estimate over the 50 S-SI trials was 0.07 (supplementary table S3, Supplementary

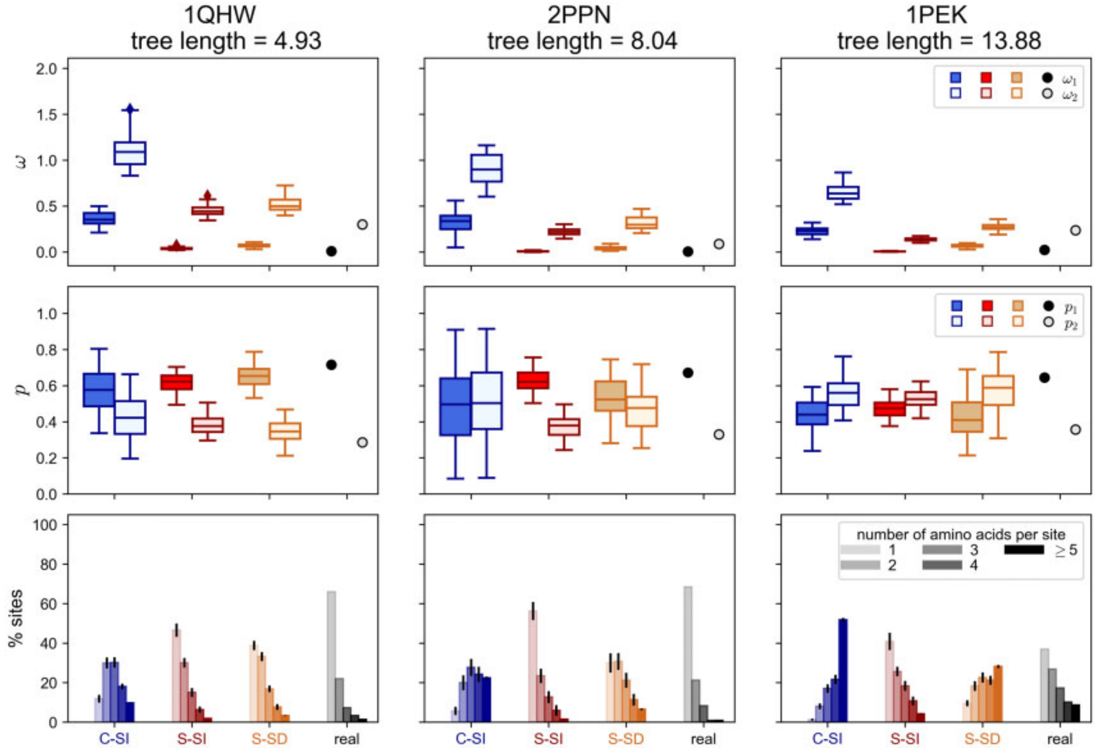

Fig. 4. Stability-informed models (S-SI and S-SD) generate alignments consistent with real data with respect to substitution rates and amino acid variability. For each of three natural protein (1QHW, 2PPN, and 1PEK corresponding to the three columns), we generated 50 protein-specific alignments under simulation models C-SI, S-SI, and S-SD. The first row reports the estimated substitution rates ω_1 (dark) and ω_2 (light) inferred from M3 ($k = 2$). The ω distributions are of the 50 model- and protein-specific alignments; the dots are the estimates from the real protein alignments. The second row reports the proportion of sites in each rate category, p_1 (dark) and p_2 (light). The third row plots the distributions of the number of amino acids observed per alignment site.

Material online). In contrast, rates inferred from the C-SI simulations were significantly higher than estimates from the other simulations and from the real proteins (Bonferroni corrected P values $<$1.0e-10 for all comparisons, supplementary table S4, Supplementary Material online). For the C-SI-generated alignments, the ω estimates were suggestive of neutral or weak selection regimes (fig. 4, first row).

Consistent with having the highest ω rate estimates, the C-SI-generated alignments were the most variable with regards to the number of amino acids observed per site (fig. 4, third row). Across the three protein-specific simulations, the proportion of fully conserved sites (one amino acid per site) were significantly lower than those observed from the stability-informed simulations (Bonferroni corrected P values $<$1.0e-10 for all comparisons, supplementary table S4, Supplementary Material online). Furthermore, the average fraction of sites with $\geq$5 amino acids were significantly higher. Although the S-SD-generated alignments were more conserved than the analogous C-SI simulation, the alignments were more variable compared with the corresponding S-SI simulations and real alignment. For the 1QHW and 2PPN alignments generated under S-SD, the distributions of the number of amino acids per site were largely consistent with the corresponding real protein alignment; however, the 1PEK-specific S-SD simulations were strikingly more variable (fig. 4, third row). This is consistent with results from Goldstein et al. (2015) which showed that under the S-SD model, the number of amino acids per site is expected to increase with tree length (branch lengths are measured as the expected number of single nucleotide substitutions per codon site). In general, we found that the S-SI simulations were the most consistent with the real alignments. In both the S-SI-simulated alignments and the natural alignments, 1) the most common site pattern included only one amino acid for all protein alignments and 2) the 2PPN proteins were the most conserved compared with the 1QHW and 1PEK proteins. The number of amino acids per site was on average slightly more conserved for the real alignments than the S-SI simulations which is consistent with the natural proteins being subject to additional selective constraints beyond folding stability.

Epistasis Increases Substitution Rates Compared with Site Independent Evolution

Comparing Expected Substitution Rates in the Presence and Absence of Epistatic Interactions

Values of ω estimated from the S-SD alignments were on average higher than estimates from the S-SI simulations (fig. 4, first row). This suggests that epistasis, as modeled in the S-SD framework, might lead to an increase in substitution rates compared with site-independent evolution. However, it remains unclear if the observed increase in rates is a genuine outcome of epistasis or a consequence of inference model misspecification. To address this, we compared the expected site-specific substitution rates calculated directly from the S-SI and S-SD generating frameworks. Consistent with our finding that epistasis increased the inferred substitution rates, the distributions of expected dN^h/dS^h were more positively skewed (higher) when epistasis was included (S-SD) for all three protein-specific simulations compared with the rates expected had sites evolved independently (S-SI; fig. 5). Rate distributions predicted from the S-SI model often displayed three peaks at dN^h/dS^h values representative of highly stringent selection regimes ($dN^h/dS^h \approx 0.00$), moderate selection pressures ($dN^h/dS^h \approx 0.25$), and more relaxed selection ($dN^h/dS^h \approx 0.4$). The position of the peaks differed only slightly depending on the protein-specific simulation (fig. 5, second row). Rate distributions estimated from S-SD were bimodal with considerably fewer sites under highly stringent selection ($dN^h/dS^h \approx 0$) compared with the analogous S-SI protein-specific distribution (fig. 5). Furthermore, more sites were under weak selection pressures under S-SD compared with S-SI; the percentage of sites with $dN^h/dS^h > 0.5$ under (S-SI, S-SD) were (8.5%, 17.2%), (2.9%, 4.2%), and (3.9%, 10.8%) for the 1QHW, 2PPN, and 1PEK simulations, respectively.

An advantage of the S-SI and S-SD frameworks is that for each site evolving with epistatic dependencies (under the temporally dynamic S-SD), we are able to model an analogous site evolving independently and under the same average stability restrictions (under the time-homogenous S-SI). To assess the magnitude of the effect of epistasis on evolutionary rates, we calculated the difference in substitution rates under epistasis (S-SD) and site-independence (S-SI). Averaged over all sites in the alignment, the mean differences in rates were 0.07, 0.08, and 0.11 for the 1QHW, 2PPN, and 1PEK simulations, respectively, implying that across the whole protein epistasis had a modest effect on substitution rates. However, site-wise analyses of rate differences revealed that epistasis increased the expected substitution rate at 88.8%, 89.5%, and 84.3% of sites in the 1QHW, 2PPN, and 1PEK simulations. The largest differences in dN^h/dS^h rates were observed at sites subject to stringent selection regimes under site-independence ($dN^h/dS^h < 0.2$, fig. 6). The less frequent and more minor reductions in rates due to epistasis occurred at sites evolving close to neutrality with $dN^h/dS^h \approx 1$ under site-independence.

Evaluating the Relationship between Epistatic Sensitivity and Structural Features

The previous result suggests that epistasis has a variable impact across sites. We were therefore interested in assessing the properties which made a site more or less sensitive to epistasis. To do this, we calculated a site's "epistatic sensitivity" by measuring the variability in the expected substitution rate given different background sequences. Since the vast majority of randomly generated sequences have zero probability of folding correctly, we used the sequences from the S-SD protein-specific alignments as the set of possible background sequences. Therefore, the number of background sequences was $50 \times N$, where $N = \{14, 14, 12\}$ is the number of taxa for the 1QHW, 2PPN, and 1PEK simulations, respectively.

If the substitution rate at a site was minimally influenced by the background sequence, then we expect little variation in dN^h/dS^h values. Alternatively, if the rate at a site was heavily influenced by the residues present at other positions,

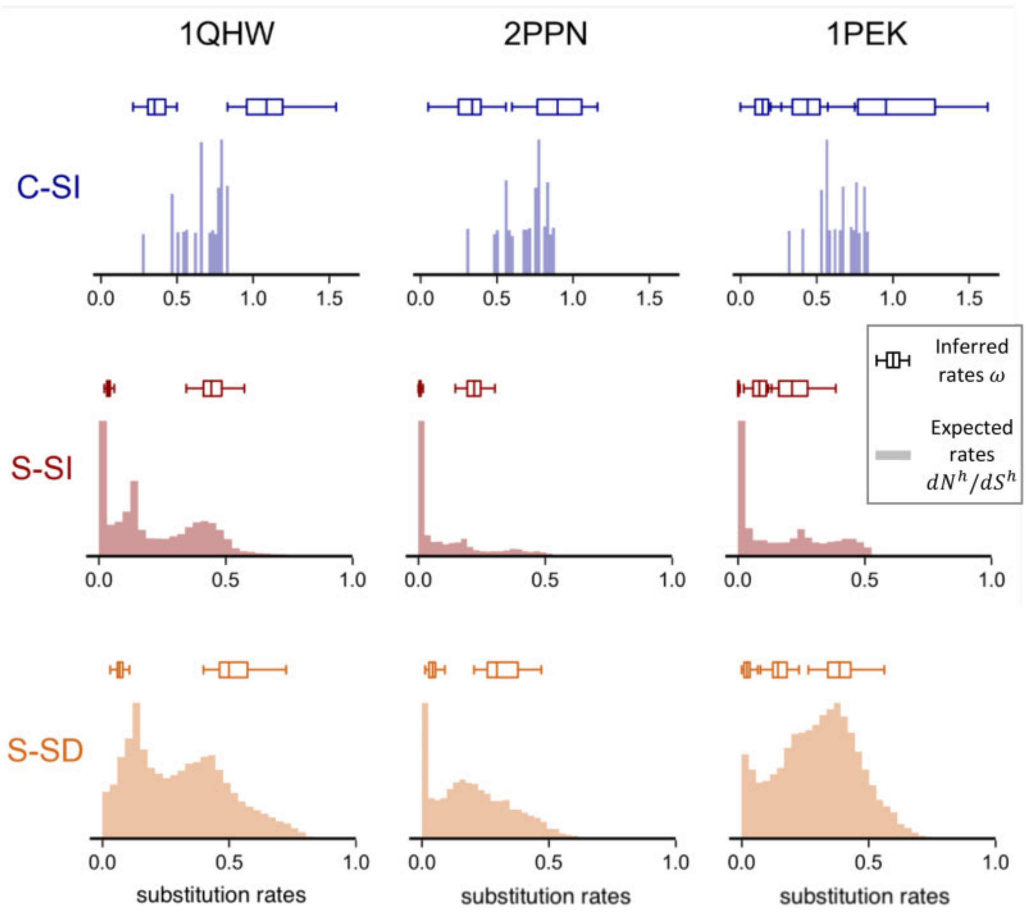

Fig. 5. M-series inference models capture the most common substitution rates across sites. Histograms represent the distributions of expected site-specific substitution rates, dN^h/dS^h, calculated from simulation models C-SI, S-SI, and S-SD (row) with protein-specific parameters (column). The boxplots represent the distribution of maximum likelihood rate estimates, $\omega_1 < \omega_2$, under M3 ($k = 2$) for proteins 1QHW and 2PPN and M3 ($k = 3$) for protein 1PEK ($\omega_1 < \omega_2 < \omega_3$). Note the difference in *x* axis range in the top row (0.0–1.5) and the bottom rows (0.0–1.0).

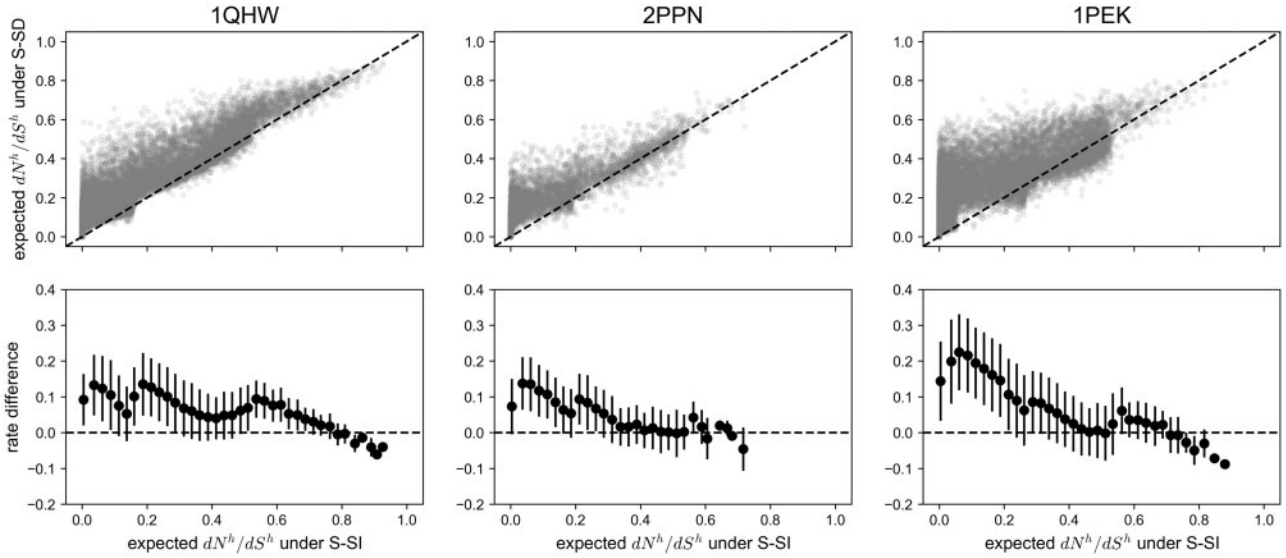

Fig. 6. Epistasis results in an increase the expected substitution rate at a site, dN^h/dS^h, compared with the expectation under site-independent evolution. Analysis was completed for three protein structures: 1QHW, 2PPN, and 1PEK (columns). Top panels show the relationship between dN^h/dS^h under a S-SI model (rates calculated using eq. 4) and a S-SD model (rates calculated using eq. 5). Epistasis increased substitution rates at 88.8%, 89.5%, and 84.3% of sites in the 1QHW, 2PPN, and 1PEK proteins. Bottom panels show the difference in dN^h/dS^h under S-SD compared with the rate under S-SI. Positive values indicate that rates are expected to be higher when epistatic interactions are included. The mean differences in rates were 0.07, 0.08, and 0.11 for the 1QHW, 2PPN, and 1PEK simulations, respectively.

we expect higher variance in the dN^h/dS^h values depending on the background protein sequence. We found that the degree of epistatic sensitivity correlated significantly with structural properties, specifically RSA and WCN. The correlation coefficient (*P* value) between RSA and epistatic sensitivity was 0.34 (1.00e-09), 0.39 (3.84e-05), and 0.32 (4.6e-08) for the 1QHW, 2PPN, and 1PEK protein structures. Similarly, a significant correlation was observed between WCN and epistatic sensitivity with $r = -0.38$ (1.03e-11), -0.42 (7.64e-06), and -0.22 (2.1e-04) for the 1QHW, 2PPN, and 1PEK protein structures. Note that the correlations are not due to increased variability at sites with higher mean rates (see supplementary table S5, Supplementary Material online, for additional analyses of the correlation between RSA (and WCN) and epistasis sensitivity calculated as the standard deviation in the log of the rate). Therefore, the results suggest that sites near the core of the protein structure, with low solvent exposure (RSA) and high packing density (WCN), were more robust to changes in the background protein sequence compared with solvent-exposed residues (high RSA and low WCN). Supplementary figure S2, Supplementary Material online, shows the relationship between epistatic sensitivity and number of contacts for all three proteins.

The observation that highly connected sites are less influenced by epistasis may initially appear counterintuitive. However, consider a highly connected site at which the fitness landscape needs to be compatible with the amino acid residues present at several interacting positions. A change at a few of the many neighboring amino acids has negligible effect on a fitness landscape that is otherwise highly constrained by its many contacts; hence, there are minimal impacts on dN^h/dS^h values. We illustrate this using a buried site and an exposed site in the 1QHW protein (fig. 7*A*). For buried site 41 (RSA = 0.01 and WCN = 1.27), the standard deviation in dN^h/dS^h was 0.04 across all 50 × 14 background sequences. The fitness landscape at site 41 given four background sequences with increasing divergence levels is plotted in figure 7*B* (top panels). Amino acid isoleucine (I) was consistently the fittest at site 41, followed by amino acids valine (V) and leucine (L) across the different background sequences. At equilibrium, the site will primarily be occupied by the optimal amino acid (I) and most nonsynonymous mutations will be deleterious resulting in a low dN^h/dS^h as expected given the correlations between RSA (or WCN) and dN^h/dS^h (fig. 3). By contrast, consider a surface site which tends to have fewer contacts. A substitution at one of the few interacting positions is more likely to induce a larger shift in amino acid preferences and consequently alter the expected substitution rate. This is illustrated in the bottom panels of figure 7*B*, which show the fitness landscapes at surface site 73 of the 1QHW protein (RSA = 0.82, WCN = 0.79, standard deviation in $dN^h/dS^h = 0.11$).

Traditional ω-Based Codon Substitution Models Perform Well despite Their Site-Independence Assumption

Assessing the Accuracy of Substitution Rate Inference under M-Series Codon Models

We have thus far shown that epistasis impacts substitution rates; however, traditional codon models used to infer selection pressures assume that sites evolve independently. Does neglecting to account for epistasis bias inference from traditional ω-based models? Furthermore, ω-based models assume that a small number of rate categories are sufficient to account for the among-site rate heterogeneity. It is therefore important to compare errors in estimation due to epistasis with the baseline estimation errors arising from unmodeled variability in rates across sites. Comparing the inferred substitution rates (ω) from the S-SI simulations to the theoretical rate expectations dN^h/dS^h, allows us to assess the inference of rates in the presence of among-site rate heterogeneity but without temporal changes in rate due to epistasis. The S-SD simulations allow us to assess the performance of ω-based models in the presence of among-site rate heterogeneity and epistasis.

First, we used the M3 (k) versus M3 ($k + 1$) likelihood ratio test to determine the number of significant rate categories from each alignment (table 2). Three factors influence the number of significant rate categories: simulation model,

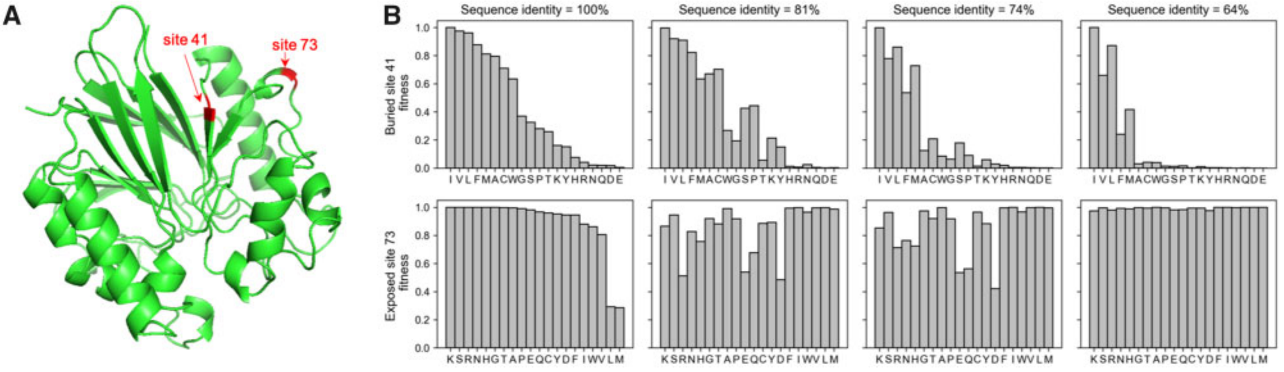

Fig. 7. Buried sites are more robust to changes in the background protein sequence compared with exposed sites. (*A*) The structure of the 1QHW protein. Arrows indicate the location of buried site 41 (RSA = 0.01 and WCN = 1.27) and exposed site 73 (RSA = 0.82 and WCN = 0.79). (*B*) The fitness landscapes at buried site 41 (top panels) and exposed site 73 (bottom panels) given different background sequences (columns). The reported sequence identities are in reference to the background sequence used to determine the landscapes in the left-most column.

protein length, and tree length. Within each protein-specific simulation, we found that the C-SI alignments had the lowest number of significant tests for three rate categories compared with S-SI and S-SD simulations. This is perhaps expected since the C-SI simulations had less heterogeneity in rates across sites compared with the stability-informed models. Each C-SI alignment had at most 20 unique rate categories, whereas under S-SI and S-SD each site had a unique fitness landscape(s) (see Materials and Methods section for details). Second, within each generating framework, the 2PPN-specific simulations had the lowest number of significant results for three rate categories. The 2PPN alignments were much smaller with only 107 codon sites compared with the 1QHW (300 codon sites) and 1PEK (279 codon sites) alignments. This suggests that there is less power to detect additional rate components with fewer sites. Lastly, despite similar numbers of codon sites, a larger number of the 1PEK-specific simulations displayed significant evidence for three rate categories compared with the 1QHW-specific simulations. There are two potential reasons for this observation: 1) the number of rate categories is influenced by the protein structure such that the 1PEK contact map induces more variation in rates across sites compared with the 1QHW structure or 2) there is more power to identify rate heterogeneity with deeper trees (1PEK tree length = 13.88, 1QHW tree length = 4.93). To distinguish between these two possibilities, we conducted an additional experiment: We generated 1QHW-specific alignments under the three generative frameworks (C-SI, S-SI, and S-SD) along the 1QHW phylogeny with double the branch length (blx2, table 2) and 1QHW-specific mutation parameters (table 1). From these additional simulations, we found an increase in detection of three rate categories across all generative models. More importantly, the number of significant tests for three rate categories was now comparable to those from the 1PEK-specific simulations (table 2). These results support the notion that deeper trees provide more informative site patterns for the detection of among-site rate heterogeneity.

Overall, we found that the number of rate categories inferred using the M3 (k)–M3 ($k+1$) likelihood ratio test was consistent with the number of peaks observed in the corresponding dN^h/dS^h distribution. We next asked whether the inferred substitution rates (ω) corresponded to the expected rates (dN^h/dS^h). For the 1QHW- and 2PPN-specific simulation, two rate categories were most commonly detected in the S-SI simulations. The first rate category was reflective of the sites subject to highly stringent selection regimes with low substitution rates ($\omega_1 \approx 0$). The second rate category often took on values representative of the average of the tail of the dN^h/dS^h distribution (fig. 5, second row). For the S-SD simulations, the inferred ω values were consistent with the most common rates with ω_1 values comparable to the first peak in the dN^h/dS^h distribution and ω_2 approximating the second peak (fig. 5, third row). More than half of the 1PEK-simulated alignments showed significant evidence for three rate categories; 28/50, 39/50, and 43/50 under C-SI, S-SI, and S-SD, respectively (table 2). Consequently, for the 1PEK simulations, we compared the distributions of expected dN^h/dS^h rates with the ω_1, ω_2, and ω_3 distributions estimated under M3 ($k = 3$) and found that the rates inferred using traditional codon models tended to capture the most common rate categories (i.e., the distribution of ω values corresponded to peaks in the dN^h/dS^h distributions, fig. 5). Therefore, in the presence and absence of epistasis, the ω estimates were consistent with the most common rate expectations.

Table 1. Protein-Specific Mutation Parameters Estimated from the Natural Alignments for Proteins 1QHW, 2PPN, and 1PEK under ω-Based Model M3 ($k = 3$).

	1QHW	2PPN	1PEK
κ	4.372	2.503	0.904
π_A	0.205	0.268	0.188
π_C	0.318	0.245	0.346
π_G	0.280	0.294	0.258
π_T	0.197	0.192	0.208
Number of taxa	14	14	12
Number of sites	300	107	279
Tree length	4.93	8.04	13.88

Note.—κ is the transition-to-transversion ratio and π_j is the stationary frequency of nucleotide j.

Table 2. Model Contrasts for Real and Simulated Alignments from Three Proteins (1QHW, 2PPN, and 1PEK).

Model Contrast	1QHW	1QHW blx2	2PPN	1PEK
Real				
M0 versus M3 ($k = 2$)	Yes	—	Yes	Yes
M3 ($k = 2$) versus M3 ($k = 3$)	Yes	—	No	Yes
M3 ($k = 3$) versus M3 ($k = 4$)	No	—	No	No
M3 ($k = 2$) versus CLM3	Yes	—	No	Yes
BUSTED ($\omega_3 < 1$) versus BUSTED	No	—	No	Yes
Tree length	4.93	—	8.04	13.88
C-SI				
M0 versus M3 ($k = 2$)	50	50	50	50
M3 ($k = 2$) versus M3 ($k = 3$)	6	19	1	28
M3 ($k = 3$) versus M3 ($k = 4$)	0	0	0	3
M3 ($k = 2$) versus CLM3	7	30	17	33
BUSTED ($\omega_3 < 1$) versus BUSTED	0	0	0	0
Mean tree length	5.27	10.48	7.55	13.32
S-SI				
M0 versus M3 ($k = 2$)	50	50	50	50
M3 ($k = 2$) versus M3 ($k = 3$)	21	42	7	39
M3 ($k = 3$) versus M3 ($k = 4$)	0	15	0	3
M3 ($k = 2$) versus CLM3	10	23	14	22
BUSTED ($\omega_3 < 1$) versus BUSTED	0	0	0	0
Mean tree length	4.99	9.35	7.15	12.45
S-SD				
M0 versus M3 ($k = 2$)	50	50	50	50
M3 ($k = 2$) versus M3 ($k = 3$)	15	42	16	43
M3 ($k = 3$) versus M3 ($k = 4$)	2	0	0	4
M3 ($k = 2$) versus CLM3	25	47	35	50
BUSTED ($\omega_3 < 1$) versus BUSTED	0	0	1	0
Mean tree length	5.04	9.65	7.57	14.18

Note.—The 1QHW blx2 results are from simulations on the 1QHW tree with double the branch length. Reported are the number of alignments out of 50 for which the specified likelihood ratio test was significant. Alignments were generated under simulation models C-SI, S-SI, and S-SD. The mean total tree lengths from M3 ($k = 3$) are also reported.

The distributions of dN^h/dS^h under S-SD and S-SI are rich distributions showing variation like that of a continuous

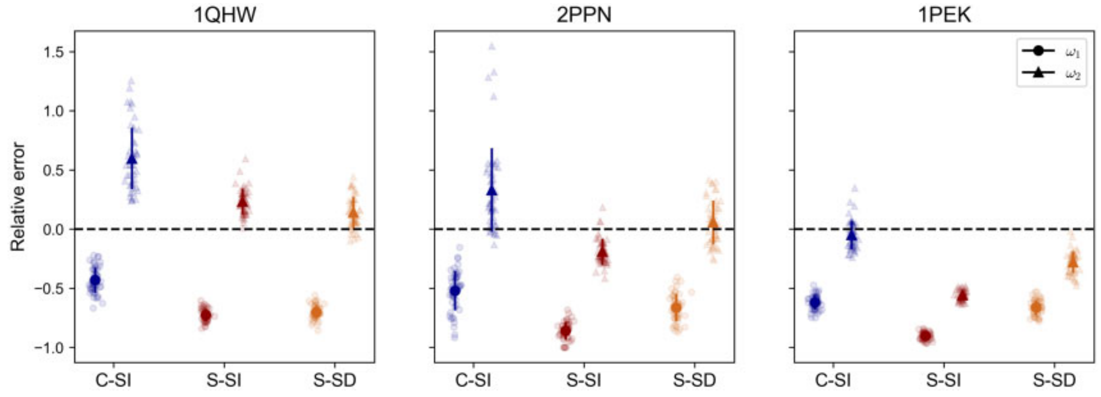

Fig. 8. The accuracy of rate estimation under M-series model is comparable when alignments are generated with and without epistasis. Plotted is the relative error ($\omega_c/(\mathrm{d}N_c/\mathrm{d}S_c) - 1$) in rate estimation under M3 ($k = 2$) for alignments generated under C-SI model, S-SI, and S-SD for each of the three proteins (1QHW, 2PPN, and 1PEK). The lighter points represent the relative error from each of the 50 model- and protein-specific trials. The darker points are the average values across trials and the bars are the standard deviation.

distribution (fig. 5). Due to computational limitations (related to use of the pruning algorithm), ω-based models can only approximate these distributions discretely. Some care is thus required in defining the target of ω-based model estimation. We assessed the performance of ω-based models in two additional ways. First, we looked at the correlations between expected site-specific rates ($\mathrm{d}N^h/\mathrm{d}S^h$) and the posterior mean ω^h inferred based on the best fitting M-series model. For rates calculated based on the stability-informed models (S-SI and S-SD), the correlations were significant in all 50 model- and protein-specific trials. The mean correlation coefficients and P values are reported in supplementary table S6, Supplementary Material online. The average correlation coefficients between $\mathrm{d}N^h/\mathrm{d}S^h$ and the posterior mean ω^h ranged from 0.24 to 0.40 under C-SI, 0.68 to 0.79 under S-SI, and 0.67 to 0.74 under S-SD.

Second, under M3 ($k = 2$), ω_c is interpretable as the substitution rate averaged over time and across sites belonging to the rate class $c = 1$ or 2. Therefore, a potential way of addressing the performance of M3 ($k = 2$) is by resolving sites according to the posterior probability of belonging to rate class (P_c) and calculating the average expected rate $\mathrm{d}N_c/\mathrm{d}S_c = 1/n\sum_h P_c^h\ \mathrm{d}N^h/\mathrm{d}S^h$. We compared the excepted $\mathrm{d}N_c/\mathrm{d}S_c$ with the inferred ω_c values for respective rate class c; the relative error in rate estimates is plotted in figure 8. As expected, the errors were lowest for alignments generated under C-SI, since the generating model was the most consistent with inference model assumptions (rates under C-SI are independent and identically distributed). Nonetheless, the ω_1 values were often underestimated. Based on the results of Spielman and Wilke (2015), we suspect that the underestimation is due to the asymmetry in the mutation models ($\mu_{ij} \neq \mu_{ji}$) present in all protein-specific simulations (table 1). Importantly, and consistent with results from figure 5, the relative error in ω estimates was comparable across S-SI and S-SD simulations. This supports the previous conclusion that the performance of ω-based models is somewhat robust to epistatic effects.

Detecting Temporal Fluctuations in Substitution Rates and Positive Selection

By framing the S-SI and S-SD models within the MutSel framework, differences in site-wise evolutionary dynamics between the site-independence assumption and epistatic evolution become apparent. Under the traditional site-wise MutSel framework, the substitution process is modeled independently at each position and hence the fitness effect of a mutation is not influenced by the background protein sequence with fixed site-specific fitness landscapes (fig. 1B). Shifts in fitness landscapes (nonstationary dynamics) are interpreted as evidence of adaptive events where external changes in environment or gene function result in changes in the amino acid preferences at the site (dos Reis 2015; Jones et al. 2017). However, if a site is subject to epistatic interactions, the site-specific fitness landscape, and hence the expected substitution rate at the site, is influenced by the residues present at other positions. Epistasis, therefore, implies a nonstationary substitution process over time such that the fitness landscape at a site constantly changes because of substitutions at other positions (fig. 1A), even when there are no adaptive events.

We were therefore interested in assessing whether traditional ω-based inference models are able to detect temporal rate fluctuations due to epistasis. However, it is important to note that using the MutSel framework, Jones et al. (2017) previously observed that site-independent evolution can result in a detectable signal for temporal variation in substitution rates (at evolutionary equilibrium) by a process reminiscent of Wright's nonadaptive phase of shifting. This occurs when a site accepts a mutation due to drift to a suboptimal amino acid which is then followed by a transient period of higher rates of nonsynonymous fixations as the site evolves toward the peak of the landscape. Additionally, they found that these dynamics can result in site patterns consistent with positive selection when tested using the BUSTED ($\omega_3 < 1$) versus BUSTED likelihood ratio test. It is therefore important to compare the results due to epistasis with the

Table 3. Mean Maximum Likelihood Estimate (MLE) under CLM3 from 50 Simulated Alignments under Models (C-SI, S-SI, or S-SD) with Protein-Specific Parameters (1QHW, 2PPN, or 1PEK).

Simulation Model	1QHW Mean MLE	2PPN Mean MLE	1PEK Mean MLE
C-SI	$\omega_1 = 0.268, \omega_2 = 0.983$ $p_1 = 0.440, \delta = 0.385$	$\omega_1 = 0.321, \omega_2 = 5.458$ $p_1 = 0.579, \delta = 0.267$	$\omega_1 = 0.232, \omega_2 = 2.743$ $p_1 = 0.449, \delta = 0.159$
S-SI	$\omega_1 = 0.028, \omega_2 = 0.449$ $p_1 = 0.589, \delta = 0.062$	$\omega_1 = 0.006, \omega_2 = 0.290$ $p_1 = 0.649, \delta = 0.046$	$\omega_1 = 0.004, \omega_2 = 0.181$ $p_1 = 0.489, \delta = 0.031$
S-SD	$\omega_1 = 0.052, \omega_2 = 0.520$ $p_1 = 0.635, \delta = 0.148$	$\omega_1 = 0.024, \omega_2 = 0.424$ $p_1 = 0.587, \delta = 0.182$	$\omega_1 = 0.033, \omega_2 = 0.314$ $p_1 = 0.363, \delta = 0.106$

baseline detection rates expected due to nonadaptive shifting balance.

We used the M3 ($k = 2$)-CLM3 model comparison to test for temporal variations in rates. M3 ($k = 2$) serves as the null model, whereas the covarion-like CLM3 accounts for temporal switches between ω_1 and ω_2 by estimating a δ parameter interpretable as the expected number of rate switches per substitution. We found that the number of significant tests for temporal rate shifts was mainly influenced by two factors: the tree length and the generative model. Consistent with the results reported in Jones et al. (2017), we found that the number of trials for which CLM3 was the better fitting model increased with tree length (table 2), this was true for all generative models and all protein-specific simulations. In regard to the generative model, within each set of protein-specific simulations, the number of trials with evidence for temporal switching was highest for the S-SD simulations compared with alignments generated from the site-independent frameworks (C-SI and S-SI). Furthermore, δ was estimated to be at least two times higher in the S-SD simulations compared with the S-SI simulations (table 3). For the 1QHW-, 2PPN-, and 1PEK-simulated alignments, δ was estimated to be (0.062, 0.148), (0.046, 0.182), and (0.031, 0.106) when simulated under (S-SI, S-SD). These results suggest that temporal variations in rates due to stability-induced epistasis produce a detectable signal in excess of the baseline signal expected due to nonadaptive shifting balance on static fitness landscapes.

Surprisingly, none of the simulated alignments showed significant evidence of positive selection using the BUSTED ($\omega_3 < 1$)–BUSTED likelihood ratio test, with the exception of only 1/50 S-SD-generated alignments with 2PPN-specific parameters (table 2). This is in contrast with previous results where nonadaptive shifting balance produced evidence of positive selection in up to 40% of trials (Jones et al. 2017). This suggests that shifting balance dynamics can be sufficiently different when fitness landscapes are informed by stability constraints rather than being randomly drawn from a normal distribution. However, this hypothesis warrants further analyses, which are beyond the scope of this article, since the lack of detection could be a consequence of the range of simulation parameters evaluated here.

Discussion

We have examined the influence of stability-induced epistasis on expected and inferred substitution rates and assessed the accuracy of rate estimation from traditional ω-based models. We found that epistasis resulted in minor elevations in substitution rates considering sites across the whole protein. However, the impact of epistasis on site-specific dynamics was prominent. A site evolving with epistatic effects on fitness had higher substitution rates compared with an analogous site evolving independently and under the same average stability constraints. Under site-independence, theory predicts that purifying selection will maintain the site on or near the fitness optima of the fixed fitness landscape (i.e., the site will predominantly be occupied by the optimal amino acid). Most nonsynonymous mutations will be deleterious and are eliminated from the population resulting in low rates of nonsynonymous substitutions relative to the rates of synonymous substitutions (low dN^h/dS^h). In comparison, consider an epistatic site h and suppose that the site is occupied by the fittest residue, a, given the current background sequence S. Following a substitution at another position in the protein (so that the background sequence changes from S to X), the fitness landscape at site h will change (fig. 1B). If the change maintains a as the fittest residue, then the substitution rate will remain low. On the other hand, if the change in landscape renders amino acid a suboptimal, then over some period of time the site will be occupied by a suboptimal amino acid. Therefore, the change in fitness landscape induces a change in the amino acid equilibrium frequencies (supplementary fig. S3, Supplementary Material online). Since the expected substitution rate, dN^h/dS^h, is a function of the equilibrium frequencies (eqs. 4 and 5), and since epistatic sites are more likely to be occupied by suboptimal amino acid (supplementary fig. S4, Supplementary Material online), the expected substitution rate will consequently be higher compared with site-independence. In other words, in the presence of epistasis, sites must constantly adapt to amino acid replacements occurring at other positions in the protein which results in higher substitution rates.

The observation that epistasis increased substitution rates contrasts with previous results discussed in Rodrigue and Lartillot (2017), which found that epistasis most often decreased substitution rates compared with site-independence. The discrepancy between our results and theirs is likely because of differences in the way epistatic interactions are modeled and because of differences in expectations of what the rate would have been under site-independent evolution. Rodrigue and Lartillot (2017) model epistasis as random deviations from multiplicative fitness and consider the effect of an epistatic landscape by comparison with a randomly assigned fixed fitness landscape. Here, we implicitly model epistasis as a by-product of protein stability, and we compare the rates from a model that accounts for protein stability but no epistasis (S-SI) to a model that accounts for stability and includes temporal rate fluctuations due epistasis (S-SD). As such, both an epistatic and an independently evolving stability-informed site are subject to the same average stability constraints, however, the epistatic site experiences fluctuating fitness landscapes whereas the independent site is evolving on a fixed landscape (fig. 1*A* and *B*). Our approach allows for a direct way of investigating the influence of epistasis on protein evolution.

Since substitution rates are primarily determined from the fitness coefficients, we expect that dynamic fitness landscapes due to epistasis will induce fluctuations in substitution rates over time. The variation in rate may be transient, where preferences at the site shift from some subset of amino acids to another, for example, polar residues might be preferred in one background sequence whereas nonpolar residues might be preferred given another sequence. For a short period of evolutionary time, the substitution rate will be transiently high as the site adjusts to the new peak (dos Reis 2015). Alternatively, a shift from a more-uniform to a more-rugged landscape (or vice versa) would result in a sustained difference in rate from low to high (or high to low). To test if such dynamics are detectable using traditional ω-based inference models, we conducted the M3-CLM3 likelihood ratio test on all simulated alignments. Although we were able to detect evidence of temporal rate variations under epistasis, it is important to note that Jones et al. (2017) showed that evolution on fixed fitness landscapes can also result in detectable signal for temporal variations in rates. They described a process reminiscent of the nonadaptive phase of Wright's shifting balance where a deleterious substitution due to drift moves a site away from its fitness peak and is followed by a transient period of high rates of nonsynonymous substitutions as the site evolves back to the fitness optima. In this way, epistasis and shifting balance result in similar temporal rate dynamics; a site becomes occupied by a suboptimal amino acid and subsequent nonsynonymous mutations are fixed in order to readjust to the fitness peak. The difference, however, is that under site-independence the site is destabilized due to a chance deleterious substitution at the site. In contrast, under epistasis, the site is destabilized because of a substitution at another position causing a shift in the underlying fitness landscape. We found that the intensity of temporal rate switching was on average at least two times higher because of epistasis compared with the switching rates due to shifting balance. The higher switching rates is perhaps expected since shifting balance dynamics are contingent on the rare fixation of deleterious mutations by drift, whereas epistasis subjects sites to continuous changes in fitness landscapes.

Moreover, nonadaptive shifting balance dynamics were previously shown to elevate ω rates to values >1 (Jones et al. 2017), resulting in the canonical signal for positive selection. Specifically, Jones et al. (2017) reported significant evidence for positive selection at 10–40% of trials when branch lengths were sufficiently long (total tree length was at least seven substitutions per codon site). Here, two of the three phylogenies used for simulations had a total tree length >7 substitutions per codon site (the 2PPN and 1PEK phylogenies). However, we found no evidence for positive selection when alignments were generated with stability-informed fitness landscapes (with and without epistasis). Importantly, these results suggest that realistic fitness landscapes based on stability constraints are not a source of conflation for the canonical signal for adaptive evolution ($\omega > 1$) when tested using traditional ω-based inference models.

Inference models operate on a set of assumptions that are certainly incorrect for real protein evolution. Two of the most pervasive assumptions are that sites evolve independently, and that the variability in rates among site is accurately approximated by a small number of rate categories. We find that, despite not accounting for epistasis, ω-based inference models perform comparably well when alignments are generated with and without epistatic interactions. A potential explanation for the comparability in model performance is that the magnitude or frequency (or both) of changes in amino acid preferences as a by-product of stability-induced epistasis are minor throughout evolutionary history. This supports previous computational and experimental work showing that, with respect to their impact on protein stability, amino acid fitness effects tend to remain relatively well conserved over long evolutionary times (Ashenberg et al. 2013; Risso et al. 2015). Our results suggest that, while accounting for epistasis is essential for understanding how proteins evolve, the site-independence assumption does not appear to limit the utility or accuracy of traditional inference models at identifying average selective pressures acting on natural proteins.

To address the concern that among-site rate variation might not be well approximated by a small number of rate categories, more sophisticated inference models based on the MutSel framework were developed that permit a unique substitution process at each alignment site (Rodrigue et al. 2010; Tamuri et al. 2012, 2014; Rodrigue and Lartillot 2014). However, these frameworks are generally only applicable when large phylogenies (>100 taxa) are available in order to reliably estimate site-specific parameters (e.g., the amino acid frequencies at each site, 19 parameters per site). Therefore, inference from smaller data sets must rely on traditional ω-based inference models which group sites into a small number of categories and estimate a much smaller number of parameters. Although we found that the full extent of site-wise rate heterogeneity was not detectable by traditional models, the number of significant rate categories was widely consistent with the number of peaks in the distributions of expected rates. This suggests that traditional inference models are capable of detecting among-site heterogeneity when a sufficient number of sites share similar rates. Additionally, and perhaps more importantly, the ω values estimated were comparable to the theoretical rate expectations at the two or three clusters of sites. Furthermore, we found that the posterior mean ω^h calculated from simple M-series models correlated significantly with the expected rates. Overall, our results suggest that ω-based models sufficiently describe average selective pressures.

The MutSel framework and biophysical models are a step toward more mechanistically plausible generative frameworks. Nonetheless, our models are limited by any underlying assumptions about the evolutionary process that are inconsistent with real protein evolution. The population genetics theory underlying the MutSel framework assumes mutations enter a population at an extremely low rate followed by a near-instantaneous fixation or loss. As such, a system might not be well modeled by MutSel when the dynamics of standing polymorphism can impact substitution rates (e.g., extended residence times for polymorphism, selective interference, and stochastic tunneling in large population), or the mutation rate is high (e.g., viral systems). As our goal was to model an evolutionarily conserved property (stability constraints) for lineages having low mutation rates and relatively small effective sizes, MutSel substitution dynamics are expected to be appropriate.

The principles of thermodynamics underlying the biophysical model assume a simple two-state folding process where proteins are either correctly folded or unfolded. Small monomeric proteins ($<$ 100 amino acids) can fold in this way (Jackson 1998); however, larger proteins require stable intermediate structures to fold properly. Of the protein structures used here, and previously within this framework (Goldstein 2011, 2013; Pollock et al. 2012; Goldstein and Pollock 2016, 2017), only the 2PPN protein has been experimentally shown to fold following the two-state process (Jackson 1998). In fact, although it is the largest protein known to fold without the need of intermediate structures, it is the smallest protein to ever be used within this thermodynamic framework. More generally, the three structures used here differ in important ways (e.g., biological function, protein length, and packing density); nonetheless, we observed similar consequences of epistasis on substitution rates which suggests that the results may be generalized across stable, globular proteins.

The current formulation of the biophysical model is limited to stable proteins with a known three-dimensional structure and therefore does not characterize the evolutionary dynamics of intrinsically disordered proteins or proteins with multiple conformations. The three-dimensional structure is used to approximate the free energy of a sequence in a given native state. Various methods have been developed to estimate stability values upon mutations (e.g., FoldX [Guerois et al. 2002] and Rosetta [Rohl et al. 2004]). In this study, we used the Miyazawa–Jernigan contact potentials with the pairwise energy approximation for its computational manageability and because even the most sophisticated models at best only moderately predict mutational effects (Potapov et al. 2009). Furthermore, this model was sufficient because we did not require exact amino acid sequences that can be folded in the native structure; that is a demanding task even when more computationally exhaustive methods are used. Instead, our objective was to simulate plausible evolutionary dynamics, and we have shown that the modeling framework is sufficient for this purpose. In addition, the models used here assume selection acting only on protein stability, whereas natural proteins are subject to additional functional and structural constraints. A recent approach was presented by de la Paz et al. (2020) using multiple sequence alignments of natural protein families (1,000 sequences) to estimate global epistatic contributions. The approach reproduces empirical and theoretical phenomena and is a promising tool for improving our understanding of protein evolution.

To conclude, we have found that epistasis alters the dynamics of how proteins evolve. It is therefore important to model epistatic interactions when the objective is to gain intuition and develop a deeper understanding of how protein sequences change over time. However, with regards to inference of selective pressures, our analysis suggests that explicit modeling of epistasis might not be of paramount importance. Instead, accounting for the phenomenological outcomes of epistasis, in allowing for more diversity in among-site amino acid preferences (Tamuri et al. 2014; Rodrigue and Lartillot 2017) and/or accounting for temporal fluctuations in substitution rates (Murrell et al. 2015; Jones et al. 2017), offers a promising avenue for the future development of inference models.

Materials and Methods

Natural Protein Alignments

Three globular, monomeric proteins were used throughout this study with PDB codes 1QHW, 1PEK, and 2PPN. The 1QHW structure is from a purple acid phosphatase protein extracted from rat bone and is likely involved in bone resorption (Lindqvist et al. 1999). The 2PPN protein is a peptidyl-prolyl cis–trans isomerase extracted from human cells which facilities the folding of other proteins (Szep et al. 2009). The 1PEK protein is a proteinase K used in protein digestion. The structure was extracted from *Engypdontium album* (Betzel et al. 1993). The three protein structures differ in important ways. First, we included the 1QHW protein for consistency since it is the only protein to have previously been used in this modeling framework. We included the 2PPN protein because of its smaller size (it is approximately a third of the length of the other two proteins) and, more importantly, because it has been shown to fold following the two-state folding (Jackson 1998) and therefore does not violate one of the core thermodynamic model assumption. Lastly, we selected the 1PEK protein because, although it is comparable in length to the 1QHW protein, it is a more densely packed protein. The average number of contacts per site was 8.39 for the 1PEK protein compared with 7.5 for the 1QHW protein (and 6.9 for the 2PPN structure).

For each of the three proteins, we created a multiple sequence alignment of orthologous gene sequences using MUSCLE (Sievers et al. 2011). Protein sequences were chosen if there were no insertions or deletions since that will likely imply changes in the protein structure which are not accounted for in the modeling framework. The accession numbers for the gene sequences are reported in supplementary table S7, Supplementary Material online. The 1QHW and 2PPN alignments included gene sequences from fourteen taxa, whereas the 1PEK alignment was made up of 12 sequences. The length of the 1QHW, 2PPN, and 1PEK alignments were 300, 107, and 279 codon sites, respectively (table 1).

For each protein alignment, we inferred a phylogenetic tree using IQ-TREE (Nguyen et al. 2015) with ModelFinder (Kalyaanamoorthy et al. 2017) and ultrafast bootstrapping (Minh et al. 2013) (supplementary fig. S1, Supplementary Material online). Maximum likelihood estimates yielded a wide range of tree lengths (table 1) which allowed us to investigate how the relationship between model assumptions and substitution rate was affected by tree length.

Following the protocol outlined in Sydykova et al. (2018), we calculate RSA and WCN for all sites in each of the protein structures. RSA is the ratio of a residue's solvent-accessible surface area, calculated using DSSP (Kabsch and Sander 1983), to its maximum solvent-accessible surface area. WCN is calculated as $\sum_{j\neq j} 1/r_{ij}^2$ where r_{ij} is the distance between the geometric centers of the side chains of residues occupying sites i and j.

Simulation Models

Mutation-Selection

We generated sequence alignments using three simulation models: C-SI, S-SI, and S-SD. The simulation models differ in how fitness values are calculated (stability-informed, S-, or estimated from C-series profiles, C-) and whether they model sites as evolving independently or with epistatic interaction (-SI vs. -SD, respectively). We used the phylogenetic trees (supplementary fig. S1, Supplementary Material online) and mutation parameters (table 1) estimated from the real protein alignments to generate 50 protein-specific alignments under C-SI, S-SI, and S-SD, for a total of 150 simulated alignments per protein structure (fig. 1C). The evolutionary process, for all the simulation models, was based on the MutSel framework (Halpern and Bruno 1998). MutSel assumes a Wright–Fisher population with fixed effective population size (N_e) and a weak mutation, strong selection regime such that a mutation is either fixed (or eliminated) before the introduction of a second mutant into the population. From population genetics theory, the probability of a mutation y going to fixation in a diploid population currently fixed at variant x depends on N_e and the relative fitness effect, $s_{xy} = f_y - f_x$ (Kimura 1962):

$$P_{\text{fix}} = \frac{1 - \exp\left(-2\, s_{xy}\right)}{1 - \exp\left(-4 N_e s_{xy}\right)}. \quad (1)$$

For our models, a variant x either represents an entire sequence (S-SD) or the amino acid at a single site (C-SI and S-SI). The substitution process is modeled as a continuous-time Markov chain which is fully specified by the instantaneous rate matrix Q with elements:

$$q_{xy} \propto 2N_e \mu_{xy} P_{\text{fix}}. \quad (2)$$

q_{xy} is the substitution rate from x to y which is equal to the rate of a novel mutation y occurring in the population, $2N_e\mu_{xy}$, and its subsequent rate of fixation, P_{fix}. Mutations arise at the DNA-level following the HKY model (Hasegawa et al. 1985) allowing only single nucleotide changes.

$$\mu_{xy} = \begin{cases} 0, & \text{if } x \text{ and } y \text{ differ by more than one nucleotide} \\ \pi_j, & \text{if } x \text{ and } y \text{ differ by a synonymous transversion}, \\ \kappa\pi_j, & \text{if } x \text{ and } y \text{ differ by a synonymous transition} \end{cases} \quad (3)$$

where μ_{xy} is the mutation rate from codon x to y, κ is the transition–transversion rate ratio, and π_j is the stationary frequency of the substituted nucleotide j for $j \in \{A, C, G, T\}$. When generating protein-specific alignments, we used the nucleotide frequencies π_j and κ values estimated from the corresponding real alignment under inference model M3 ($k = 3$) (table 1). All models assume that selection acts on the final protein product. The models therefore assign all synonymous codons the same fitness.

C-Series Site-Independent Model

Under C-SI, amino acid fitness values were approximated from the C-series empirical frequency profiles (Quang et al. 2008), commonly used in phylogenetic inference. The C-series model capture among-site variation in amino acid preferences (and hence frequencies) by assuming that a site belongs to one of 20 different frequency profiles. In the MutSel framework, the frequency of amino acid a is related to its fitness f_a by the following relationship $\pi_a \propto \pi_a^{(0)} \exp(2N_e f_a)$, where $\pi_a^{(0)}$ is the stationary frequency in the absence of selection pressure (dos Reis 2015). We use this to convert each of the 20 C20 frequency profiles to 20 fitness vectors. Note that the amino acid frequencies in the absence of selection pressures, $\pi_a^{(0)}$, reflect underlying biases in the mutation process since, without selection, the stationary frequency of a codon (or nucleotide triple ijk) is proportional to $\pi_i\pi_j\pi_k$. Then, $\pi_a^{(0)}$ is calculated as the sum of the stationary frequencies of synonymous codons corresponding to amino acid a. Because the three proteins studied here had different mutational parameters (table 1), the C20 profiles translated to 20 protein-specific fitness landscapes. When generating alignments under C-SI, each site was randomly assigned one of the 20 protein-specific fitness vectors. As such, the C-SI model assumes that sites evolve independently and are identically distributed.

Stability-Informed Models (S-SI and S-SD)

Alternatively, the stability-informed models (S-SI and S-SD) define fitness as the proportion of correctly folded proteins at thermodynamic equilibrium, which is a nonlinear function of the protein's folding stability. Details for stability calculations are provided in the Thermodynamic Model of Protein Folding section.

Epistasis refers to the dependence of the fitness effect of a mutation on the background genetic sequence. To account for epistasis within the MutSel framework, each site was assigned a vector of amino acid fitness values $F^h(S) = \langle f_1^h(S), \ldots, f_{20}^h(S)\rangle$ where $f_a^h(S)$ is the fitness of the protein calculated using equation (6) given amino acid a at site h and background sequence S. Throughout the evolution of the protein, all site-specific fitness vectors were recalculated following a nonsynonymous substitution somewhere in the protein.

To assess if and how epistasis influences substitution rates, we developed an analogous S-SI model where epistatic effects on folding stability were marginalized such that the fitness landscape at a site, F^h, is independent of the background sequence and is therefore constant across time. To allow for a direct comparison between alignments generated with and without epistasis, we used the S-SD simulations to estimate the independent fitness landscapes, F^h (fig. 2). In other words, let $\{S_1, \ldots, S_N\}$ be the extant sequences in an S-SD-simulated alignment, where N is the number of taxa. We calculated f_a^h as the average fitness value for amino acid a over sequences $\{S_1, \ldots, S_N\}$ ($f_a^h = (1/N)\sum_{t=1}^{N} f_a^h(S_t)$). The average fitness values were used to specify the independent site-specific fitness vectors, F^h, under S-SI.

Scaling Branch Lengths

In order for branch lengths to have the desired interpretation as the mean number of single nucleotide substitution per codon site, the substitution rates must be rescaled. For -SI-generated alignments, we rescaled the rate matrices in the conventional way by dividing all Q^h by the mean expected rate of change, $(1/n)\sum_h \sum_x - \pi_x$ q_{xx}^h where n is the number of sites and $q_{xx}^h = -\sum_{y \neq x} q_{xy}^h$ (Jones et al. 2017). Alternatively, to obtain the appropriate scaling factor for the S-SD alignments, we ran the simulation for 1,000 substitutions using the Gillespie algorithm (Gillespie 1977). We recorded the overall time T required for 1,000 substitutions to occur by summing over the waiting times between substitutions, $T = \sum_{t=0}^{1{,}000}$ $\sum_h \tau_t^h$ where τ^h is the waiting time until the next substitution event at site h which is exponentially distributed with mean $1/q_{xx}^h$. Branch lengths, b, were then rescaled such that $b = n(T/1,000)$. We validated the scaling methods by comparing the inferred branch lengths from the simulated alignments to the true generating branch lengths (mean tree lengths from each set of simulations are reported in table 2). To avoid nonequilibrium behavior, each of the protein-specific simulations was initiated at amino acid sequences with fitness values >0.99 given the respective protein structure. The algorithm used to explore the sequence space to find sequences with high fitness values is reported in supplementary table S8, Supplementary Material online.

Expected Substitution Rate d*N*/d*S* Calculations

The evolutionary rate at a site is commonly defined as the ratio of nonsynonymous to synonymous substitutions rates (N^h/S^h) normalized by the ratio of nonsynonymous to synonymous mutations rates ($N_{\mathrm{mut}}^h/S_{\mathrm{mut}}^h$). Assuming selection acting at the protein-level such that synonymous codons have the same fitness value, the rate of fixation of a synonymous mutation will be equal to its underlying mutation rate, $S^h = S_{\mathrm{mut}}^h$. Therefore, the expected substitution ratio simplifies to $\mathrm{d}N^h/\mathrm{d}S^h = N^h/N_{\mathrm{mut}}^h$. In the traditional MutSel framework (i.e., assuming site-independence as done in simulation models C-SI and S-SI), the evolutionary rate at a site, $\mathrm{d}N^h/\mathrm{d}S^h$, can be calculated directly from the site-specific fitness coefficients and the protein-specific mutation rates:

$$\mathrm{d}N^h/\mathrm{d}S^h = \frac{N^h}{N_{\mathrm{mut}}^h} = \frac{\sum_x \sum_{y \in \mathcal{N}_x} \pi_x^h q_{xy}^h}{\sum_x \sum_{y \in \mathcal{N}_x} \pi_x^h \mu_{xy}}, \qquad (4)$$

where $\mathcal{N}_x$ is the set of codons that are nonsynonymous to codon x and differ by a single nucleotide, q_{xy}^h is the substitution rate from codon x to codon y calculated using equation (2), μ_{xy} is the mutation rate calculated using equation (3), and π_x^h is the stationary frequency for codon x at site h. We note that dos Reis (2015) presented an alternative way of calculating $\mathrm{d}N^h/\mathrm{d}S^h$ where the nonsynonymous mutation rate, N_{mut}^h, was calculated in reference to the neutral stationary frequencies $\pi_x^{(0)}$. Although the interpretation of the $\mathrm{d}N^h/\mathrm{d}S^h$ values differ (as discussed in Jones et al. [2017]), we found that both formulations resulted in highly comparable rate values (Pearson correlation coefficient = 0.99, P value <0.05; supplementary fig. S5, Supplementary Material online). The $\mathrm{d}N^h/\mathrm{d}S^h$ rates reported in the main article were calculated using equation (4).

When epistatic dependencies between sites are modeled within the MutSel framework, the average substitution rate at a site can in principle be calculated as

$$\mathrm{d}N^h/\mathrm{d}S^h = \frac{\sum_S N^h(S)}{\sum_S N_{\mathrm{mut}}^h(S)}, \qquad (5)$$

where the sum is over all possible background sequences S. However, the number of possible sequences is very large, 20^n where n is the length of the protein. For a relatively small protein of length 100, the number of possible sequences is larger than the estimated number of atoms in the observable universe. Since the evolution of natural proteins billions of years ago, natural proteins are evolving on a relatively small, localized portion of sequence space. Therefore, although $\mathrm{d}N^h/\mathrm{d}S^h$ averaged over all 20^n background sequences is the theoretical rate expectation, it is impossible to calculate and likely does not reflect the rates for real proteins. Instead, for S-SD simulations, we define the evolutionary rate at a site as the mean substitution rate observed throughout the evolution of a protein over a defined length of time. Specifically, for each S-SD alignment i (for $i = 1, \ldots, 50$), we approximate the rate at a site ($\mathrm{d}N_i^h/\mathrm{d}S_i^h$) by summing over the extant sequences $\{S_1, \ldots, S_N\}_i$. To address the robustness of our results to a more extensive sampling of sequences in the local space, we compared the $\mathrm{d}N_i^h/\mathrm{d}S_i^h$ with the rate $\mathrm{d}N_{ij}^h/\mathrm{d}S_{ij}^h$ by leaving out the jth sequence. Then, we calculated the bias and mean-squared error (MSE) as described in supplementary equations (S1) and (S2), Supplementary Material online. We found that the bias and MSE in $\mathrm{d}N^h/\mathrm{d}S^h$ estimates were minor, suggesting that calculating rates as the average over the extant sequences has minimal consequences on rate expectations. Results are discussed in more detail in supplementary Assessing Sample Size section, Supplementary Material online, and distributions of bias and MSE are plotted in supplementary figure S6, Supplementary Material online.

Thermodynamic Model of Protein Folding

The stability-informed-generating models estimate the fitness of a protein sequence based on the biophysical model described in Goldstein and Pollock (2017). The fitness of an amino acid sequence is assumed to be equal to the proportion proteins in the native (folded) structure at thermodynamic equilibrium. The method assumes a two-state system (Jackson 1998) where the protein molecule can occupy one of two possible macrostates, the folded (F) and unfolded (U) configurations. From thermodynamic theory, the probability of a system occupying a macrostate i is $P_i = e^{-\beta E_i}/q$ where E_i is the free energy associated with state i, $\beta = 1/kT$ (k is the Boltzmann constant and T is the absolute temperature), and q is the normalizing partition function ($q = \sum_i e^{-\beta E_i}$). With only two possible macrostates, $q = e^{-\beta E_F} + e^{-\beta E_U}$. The fitness of a sequence $S = \{a^1, a^2, \ldots, a^n\}$ (which is equal to the probability of the sequences being in the folded state) can be calculated as

$$\mathrm{fitness}(S) = P_F(S) = \frac{e^{-\beta E_F(S)}}{e^{-\beta E_F(S)} + e^{-\beta E_U(S)}}. \tag{6}$$

$P_F(S)$ can be rewritten in terms of the folding stability, ΔG, of the amino acid sequence measured as the difference in energies between the folded and unfolded states, $\Delta G(S) = E_F(S) - E_U(S)$:

$$P_F(S) = \frac{e^{-\beta \Delta G(S)}}{e^{-\beta \Delta G(S)} + 1}. \tag{7}$$

The free energy $E_k(S)$ associated with sequence S in a given structure k is approximated as the sum of pairwise potentials for amino acids in contact,

$$E_k(S) = \sum_{i<j} \varepsilon_{MJ}(a^i, a^j)\mathrm{CM}_k^{i,j}, \tag{8}$$

where ε_{MJ} are the contact potentials determined by Miyazawa and Jernigan (1985), and CM_k is the contact matrix specifying interactions between sites in structure k such that $\mathrm{CM}_k^{i,j} = 1$ if site i and j are in contact and 0 otherwise. Sites are considered to be in contact if the C_β atoms of the amino acids in the observed sequence are within 7 Å of each other. If the amino acid present is glycine, distance is considered with reference to the C_α atom.

Given the PDB protein structures, the free energy in the folded structure $E_F(S)$ can be directly calculated using equation (5). $E_U(S)$, the free energy of the unfolded macrostate, can be calculated from thermodynamic theory using $E_U(S) = -\ln Z_U(S)/\beta$, where $Z_U(S) = \sum_i^{Nu} e^{-\beta E_i(S)}$ and Nu is the number of unfolded microstates set equal to 3.4^n allowing for $\sim$3.4 conformations per residue. The distribution, $\rho(E)$, of free energies, $E = E_U(S)$, over unfolded states is approximated by

$$\rho(E) = \frac{1}{\sqrt{2\pi\Delta E^2(S)}} \exp \frac{-[E(S) - \bar{E}(S)]^2}{2\Delta E^2(S)} \tag{9}$$

such that

$$\begin{aligned} Z_U &= N_U \int \rho[E(S)] e^{-\beta E(S)}\, dE \\ &= N_U \exp\left[\frac{1}{2}\beta^2 \Delta E^2(S) - \beta \bar{E}(S)\right]. \end{aligned} \tag{10}$$

For a sequence S, we obtain estimates of $\bar{E}(S)$ and $\Delta E^2(S)$ by using equation (8) to calculate the free energies in a predefined set of 55 alternative structures (PDB codes reported in supplementary table S9, Supplementary Material online). With the approximations of $E_U(S)$ and $E_F(S)$, the stability of a sequence can be expressed as

$$\Delta G(S) = E_F(S) - E_U(S), \tag{11}$$

$$= E_F(S) + \beta^{-1} \ln Z_U(S), \tag{12}$$

$$= E_F(S) - \bar{E}(S) + \frac{1}{2}\beta \Delta E^2(S) + \beta^{-1} \ln N_U, \tag{13}$$

which is then used to calculate the fitness of an amino acid sequence S using equation (6).

Maximum Likelihood Inference of Selection Pressure

M-Series Models

The M-series models assume a time-reversible, stationary, continuous-time Markov chain where the instantaneous substitution rate matrix A defines the rate of substitution between codon x and y as

$$a_{xy} \propto \begin{cases} 0, & \text{if } x \text{ and } y \text{ differ by more than one nucleotide} \\ \pi_j, & \text{if } x \text{ and } y \text{ differ by a synonymous transversion} \\ \kappa\pi_j, & \text{if } x \text{ and } y \text{ differ by a synonymous transition} \\ \omega\pi_j, & \text{if } x \text{ and } y \text{ differ by a nonsynonymous transversion} \\ \omega\kappa\pi_j, & \text{if } x \text{ and } y \text{ differ by a nonsynonymous transition} \end{cases}. \tag{14}$$

κ is the transition-to-transversion rate ratio, π_j is the stationary frequency of the target nucleotide j, and ω is the nonsynonymous to synonymous rate ratio. This describes MG (Muse and Gaut 1994) parameterization of M0, the simplest M-series model, with a single rate parameter estimated for all sites in the alignment. To account for variation in selection pressure across sites, M3 (k) extends M0 by allowing for k discrete number of rate categories, each with a rate parameter ω_k and corresponding proportion of sites p_k. M0 is analogous to M3 ($k = 1$). The M3 (k) versus M3 ($k + 1$) likelihood ratio test was used to determine the appropriate number of rate categories for each alignment.

CLM3

To test for variation in substitution rate across time, we used the covarion-like CLM3 as implemented by Jones et al. (2017) which assumes that the substitution process switches over time between one with an $\omega = \omega_1$ and another with $\omega = \omega_2$. The switching and substitution processes can be modeled as a two-dimensional Markov chain (X, Y) where X is the current codon and Y indicates the substitution process, 1 or 2. Ordering the possible states as (1,1), (2,1), . . ., the rate matrix is

$$A = \frac{1}{r_1}\begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} + \frac{\delta}{r_2}\begin{pmatrix} -p_2 I & p_2 I \\ p_1 I & -p_1 I \end{pmatrix},, \qquad (15)$$

where A_1 and A_2 are the substitution rate matrices constructed using equation (14) with ω_1 and ω_2, respectively. p_1 and p_2 are the expected proportion of time a site evolves under the respective ω, I is the identity matrix, and δ denotes the rate of change between selection regimes. r_1 and r_2 are scaling parameters such that time is measured as the expected number of single nucleotide changes per codon site and δ is the expected number of switches per unit time. The model contrast M3 ($k = 2$) versus CLM3 provides a likelihood ratio test for evidence of switching between rate categories ω_1 and ω_2 across the tree.

BUSTED

The branch-site unrestricted statistical test for episodic diversification, BUSTED (Murrell et al. 2015), is based on the BS-REL framework (Kosakovsky Pond et al. 2011) allowing for variations in rates across sites and branches. Specifically, BUSTED estimates three rate categories ($\omega_1 \leq \omega_2 \leq \omega_3$) where at each branch in the tree, a site belongs to one of the three ω categories. The model also estimates proportions p_1 and p_2 ($p_3 = 1 - p_1 - p_2$) shared across sites. If there is evidence for positive selection ($\omega_3 > 1$), then a likelihood ratio test of BUSTED with ω_3 constrained to be <1 against an unconstrained BUSTED is conducted.

Code and Data Availability

Real and simulated alignments, as well as the code used to generate, analyze, and plot, have been uploaded to GitHub (https://github.com/noory3/Consequences-of-stability-induced-epistasis, last accessed July 16, 2020.)

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank Scott McCain for insightful discussion about the project and for reviewing the manuscript, and Joseph Mingrone for installing and updating the necessary software. We thank the reviewers for helpful feedback that improved the manuscript. E.S. and J.P.B. were supported by a Discovery Grant awarded by the Natural Sciences and Engineering Council of Canada.

References

Ashenberg O, Gong LI, Bloom JD. 2013. Mutational effects on stability are largely conserved during protein evolution. *Proc Natl Acad Sci U S A*. 110(52):21071–21076.

Betzel C, Singh TP, Visanji M, Peters K, Fittkau S, Saenger W, Wilson KS. 1993. Structure of the complex of proteinase K with a substrate analogue hexapeptide inhibitor at 2.2-A resolution. *J Biol Chem*. 268(21):15854–15858.

de la Paz JA, Nartey C, Yuvaraj M, Morcos F. 2020. Epistatic contributions promote the unification of incompatible models of neutral molecular evolution. *Proc Natl Acad Sci U S A*. 117(11):5873–5882.

dos Reis M. 2015. How to calculate the non-synonymous to synonymous rate ratio of protein-coding genes under the Fisher–Wright mutation-selection framework. *Biol Lett*. 11(4):20141031.

Echave J, Jackson EL, Wilke CO. 2015. Relationship between protein thermodynamic constraints and variation of evolutionary rates among sites. *Phys Biol*. 12(2):025002.

Ferrada E. 2019. The site-specific amino acid preferences of homologous proteins depend on sequence divergence. *Genome Biol Evol*. 11(1):121–135.

Gillespie D. 1977. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem*. 81(25):2340–2361.

Goldman N, Yang ZH. 1994. Codon-based model of nucleotide substitution for protein-coding DNA-sequences. *Mol Biol Evol*. 11:725–736.

Goldstein RA. 2011. The evolution and evolutionary consequences of marginal thermostability in proteins. *Proteins* 79(5):1396–1407.

Goldstein RA. 2013. Population size dependence of fitness effect distribution and substitution rate probed by biophysical model of protein thermostability. *Genome Biol Evol*. 5(9):1584–1593.

Goldstein RA, Pollard ST, Shah SD, Pollock DD. 2015. Nonadaptive amino acid convergence rates decrease over time. *Mol Biol Evol*. 32(6):1373–1381.

Goldstein RA, Pollock DD. 2016. The tangled bank of amino acids. *Protein Sci*. 25(7):1354–1362.

Goldstein RA, Pollock DD. 2017. Sequence entropy of folding and the absolute rate of amino acid substitutions. *Nat Ecol Evol*. 1(12):1923–1930.

Guerois R, Nielsen JE, Serrano L. 2002. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol*. 320(2):369–387.

Halpern AL, Bruno WJ. 1998. Evolutionary distances for protein-coding sequences: modeling site-specific residue frequencies. *Mol Biol Evol*. 15(7):910–917.

Hasegawa M, Kishino H, Yano T. 1985. Dating of human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol*. 22(2):160–117.

Jackson S. 1998. How do small single-domain proteins fold? *Fold Des*. 3:81–91.

Jones CT, Youssef N, Susko E, Bielawski JP. 2017. Shifting balance on a static mutation-selection landscape: a novel scenario of positive selection. *Mol Biol Evol*. 34(2):391–407.

Jones CT, Youssef N, Susko E, Bielawski JP. 2018. Phenomenological load on model parameters can lead to false biological conclusions. *Mol Biol Evol*. 35(6):1473–1488.

Jones CT, Youssef N, Susko E, Bielawski JP. 2020. A phenotype-genotype codon model for detecting adaptive evolution. *Syst Biol*. 69:722–738.

Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22(12):2577–2637.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*. 14(6):587–589.

Kimura M. 1962. On the probability of fixation of mutant genes in a population. *Genetics* 47:713–719.

Kosakovsky Pond SL, Frost SDW. 2005. Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol*. 22(5):1208–1222.

Kosakovsky Pond SL, Murrell B, Fourment M, Frost SDW, Delport W, Scheffler K. 2011. A random effects branch-site model for detecting episodic diversifying selection. *Mol Biol Evol*. 28(11):3033–3043.

Lindqvist Y, Johansson E, Kaija H, Vihko P, Schneider G. 1999. Three-dimensional structure of a mammalian purple acid phosphatase at 2.2 A resolution with a mu-(hydr)oxo bridged di-iron center. *J Mol Biol*. 291(1):135–147.

Marcos ML, Echave J. 2015. Too packed to change: side-chain packing and site-specific substitution rates in protein evolution. *PeerJ* 3:e911.

Massingham T, Goldman N. 2005. Detecting amino acid sites under positive selection and purifying selection. *Genetics* 169(3):1753–1762.

Meyes S, vonHaeseler A. 2003. Identifying site-specific substitution rates. *Mol Biol Evol*. 20(2):182–189.

Minh BQ, Nguyen MA, von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol*. 30(5):1188–1195.

Miyazawa S, Jernigan R. 1985. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules* 18(3):534–552.

Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, Eren K, Pollner T, Martin DP, Smith DM, et al. 2015. Gene-wide identification of episodic selection. *Mol Biol Evol*. 32(5):1365–1371.

Murrell B, Wertheim J, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. 2012. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet*. 8(7):e1002764.

Muse SV, Gaut BS. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with applications to the chloroplast genome. *Mol Biol Evol*. 11:715–724.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. 32(1):268–274.

Pollock DD, Thiltgen G, Goldstein RA. 2012. Amino acid coevolution induces an evolutionary Stokes shift. *Proc Natl Acad Sci U S A*. 109(21):E1352–E1359.

Potapov V, Cohen M, Schreiber G. 2009. Assessing computational methods for predicting protein stability upon mutation: good on average but not in the details. *Protein Eng Des Sel*. 22(9):553–560.

Quang LS, Gascuel O, Lartillot N. 2008. Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics* 24(20):2317–2323.

Risso VA, Manssour-Triedo F, Delgado-Delgado A, Arco R, Barroso-delJesus A, Ingles-Prieto A, Godoy-Ruiz R, Gavira JA, Gaucher EA, Ibarra-Molero B, et al. 2015. Mutational studies on resurrected ancestral proteins reveal conservation of site-specific amino acid preferences throughout evolutionary history. *Mol Biol Evol*. 32(2):440–455.

Rodrigue N, Lartillot N. 2014. Site-heterogeneous mutation-selection models within PhyloBayes-MPI package. *Bioinformatics* 30(7):1020–1021.

Rodrigue N, Lartillot N. 2017. Detecting adaptation in protein-coding genes using a Bayesian site-heterogeneous mutation-selection codon substitution model. *Mol Biol Evol*. 34(1):204–214.

Rodrigue N, Philippe H, Lartillot N. 2010. Mutation-selection models for coding sequence evolution with site-heterogeneous amino acid fitness profiles. *Proc Natl Acad Sci U S A*. 107(10):4629–4634.

Rohl CA, Strauss CE, Misura KM, Baker D. 2004. Protein structure prediction using Rosetta. *Methods Enzymol*. 383:66–93.

Shah P, McCandlish DM, Plotkin JB. 2015. Contingency and entrenchment in protein evolution under purifying selection. *Proc Natl Acad Sci U S A*. 112(25):E3226–E3235.

Shahmoradi A, Sydykova DK, Spielman SJ, Jackson EL, Dawson ET, Meyer AG, Wilke CO. 2014. Predicting evolutionary site variability from structure in viral proteins: buriedness, packing, flexibility, and design. *J Mol Evol*. 79(3–4):130–142.

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, et al. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Mol Syst Biol*. 7(1):539.

Spielman SJ, Wilke CO. 2015. The relationship between d*N*/d*S* and scaled selection coefficients. *Mol Biol Evol*. 32(4):1097–1108.

Starr TN, Flynn JN, Mishra P, Bolon DNA, Thornton JW. 2018. Pervasive contingency and entrenchment in a billion years of Hsp90 evolution. *Proc Natl Acad Sci U S A*. 115(17):4453–4458.

Starr TN, Thornton JW. 2016. Epistasis in protein evolution. *Protein Sci*. 25(7):1204–1218.

Sydykova DK, Jack BR, Spielman SJ, Wilke CO. 2018. Measuring evolutionary rates of proteins in a structural context. *F1000Research* 6:1845.

Szep S, Park S, Boder E, Van Duyne G, Saven JG. 2009. Structural coupling between FKBP12 and buried water. *Proteins* 74(3):603–611.

Tamuri AU, dos Reis M, Goldstein RA. 2012. Estimating the distribution of selection coefficients from phylogenetic data using sitewise mutation-selection models. *Genetics* 190(3):1101–1115.

Tamuri AU, Goldman N, dos Reis M. 2014. A penalized-likelihood method to estimate the distribution of selection coefficients from phylogenetic data. *Genetics* 197(1):257–271.

Yang ZH, Nielsen R. 2008. Mutation-selection models of codon substitution and their use to estimate selective strengths on codon usage. *Mol Biol Evol*. 25(3):568–579.

Yang ZH, Nielsen R, Goldman N, Pedersen AMK. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431–449.

Yeh S-W, Liu J-W, Yu S-H, Shih C-H, Hwang J-K, Echave J. 2014. Site-specific structural constraints on protein sequence evolutionary divergence: local packing density versus solvent exposure. *Mol Biol Evol*. 31(1):135–139.