Contingency and entrenchment in protein evolution under purifying selection

Premal Shah[1], David M. McCandlish[1], and Joshua B. Plotkin[2]

Department of Biology, University of Pennsylvania, Philadelphia, PA 19104

Edited by Christopher Wylie, Brown University, Providence, RI and accepted by the Editorial Board May 7, 2015 (received for review July 9, 2014)

The phenotypic effect of an allele at one genetic site may depend on alleles at other sites, a phenomenon known as epistasis. Epistasis can profoundly influence the process of evolution in populations and shape the patterns of protein divergence across species. Whereas epistasis between adaptive substitutions has been studied extensively, relatively little is known about epistasis under purifying selection. Here we use computational models of thermodynamic stability in a ligand-binding protein to explore the structure of epistasis in simulations of protein sequence evolution. Even though the predicted effects on stability of random mutations are almost completely additive, the mutations that fix under purifying selection are enriched for epistasis. In particular, the mutations that fix are contingent on previous substitutions: Although nearly neutral at their time of fixation, these mutations would be deleterious in the absence of preceding substitutions. Conversely, substitutions under purifying selection are subsequently entrenched by epistasis with later substitutions: They become increasingly deleterious to revert over time. Our results imply that, even under purifying selection, protein sequence evolution is often contingent on history and so it cannot be predicted by the phenotypic effects of mutations assayed in the ancestral background.

intragenic epistasis | coevolution | near neutrality | protein stability

Whether a heritable mutation is advantageous or deleterious to an organism often depends on the evolutionary history of the population. A mutation that is beneficial at the time of its introduction may confer its beneficial effect only in the presence of other potentiating or permissive mutations (1–9). Thus, the fate of a mutation arising in a population may be contingent on previous mutations (10–13). Conversely, once a mutation has fixed in a population, the mutation becomes part of the genetic background onto which subsequent modifications are introduced. Because the beneficial effects of the subsequent modifications may depend on the focal mutation, as time passes reversion of the focal mutation may become increasingly deleterious, leading to a type of evolutionary conservatism, or entrenchment (14–18).

In the context of protein evolution, the effects of contingency and entrenchment are most easily studied by considering a sequence of single amino acid changes (19) that extends both forward and backward in time from some focal substitution. To assess the roles of contingency and entrenchment we can study the degree to which each focal substitution was facilitated by previous substitutions, and the degree to which the focal substitution influences the subsequent course of evolution (Fig. 1*A*).

Dependencies within a sequence of substitutions are closely connected to the concept of epistasis—that is, the idea that the phenotypic effect of a mutation at a particular genetic site may depend on the genetic background in which it arises (20–24). In the absence of epistasis, a mutation has the same effect regardless of its context and therefore regardless of any prior history or subsequent evolution. By contrast, in the presence of epistasis, each substitution may be contingent on the entire prior history of the protein, and it may constrain all subsequent evolution.

The potential for epistasis to play an important role in evolution, including protein evolution, has not been overlooked by researchers (1, 8, 25–34), nor have the concepts of contingency (3, 4, 9, 12, 35–38) and, more recently, entrenchment (18, 39, 40). However, most studies have addressed the role of epistasis in the context of adaptive evolution (1–9, 27, 30, 31, 36, 38), whereas the consequences of epistasis under purifying selection have received less attention (18, 41–44). Indeed, although some more sophisticated models have been proposed (e.g., refs. 45–50), all commonly used phylogenetic models of long-term protein evolution assume that epistasis is absent so that sites evolve independently (51–56).

Here we explore the relationships between epistasis, contingency, and entrenchment under long-term purifying selection on protein stability. Our analysis combines computational models for protein structures with population-genetic models for evolutionary dynamics. We use a force-field-based model, FoldX (57), to characterize the effects of point mutations on a protein's stability and fitness. This approach allows us to simulate evolutionary trajectories of protein sequences under purifying selection, by the sequential fixation of nearly neutral mutations. We can then dissect the epistatic relationships between these substitutions by systematically inserting or reverting particular substitutions at various time points along the evolutionary trajectory.

Our analysis considers epistasis both at the level of protein stability and at the level of fitness. Whereas empirical studies in diverse proteins have demonstrated that the stability effects of point mutations are typically additive across sites (58, 59), in this study we are specifically interested in epistasis for stability among the mutations that fix during evolution. Even if most random mutations are virtually additive in their effects on stability, the mutations that fix under purifying selection are highly nonrandom, and so there is reason to suspect that epistasis for stability may be enriched among such mutations. Moreover, because the mapping

Significance

How large a role does history play in evolution? Do later events depend critically on specific earlier events, or do all events occur more or less independently? If a change occurs early in evolution, does it become easier or harder to revert the change as time proceeds? Here, we explore these ideas in the context of protein evolution, by simulating sequence evolution under purifying selection and then systematically permuting the order of amino acid substitutions. Our results suggest that the amino acid substitutions that occur in evolution are typically contingent on the presence of prior substitutions, and that substitutions that occur early in evolution become entrenched and difficult to modify as subsequent substitutions accrue.

Author contributions: P.S., D.M.M., and J.B.P. designed research; P.S. performed research; P.S., D.M.M., and J.B.P. contributed new reagents/analytic tools; P.S., D.M.M., and J.B.P. analyzed data; and P.S., D.M.M., and J.B.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. C.W. is a guest editor invited by the Editorial Board.

[1]P.S. and D.M.M. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: jplotkin@sas.upenn.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1412933112/-/DCSupplemental.

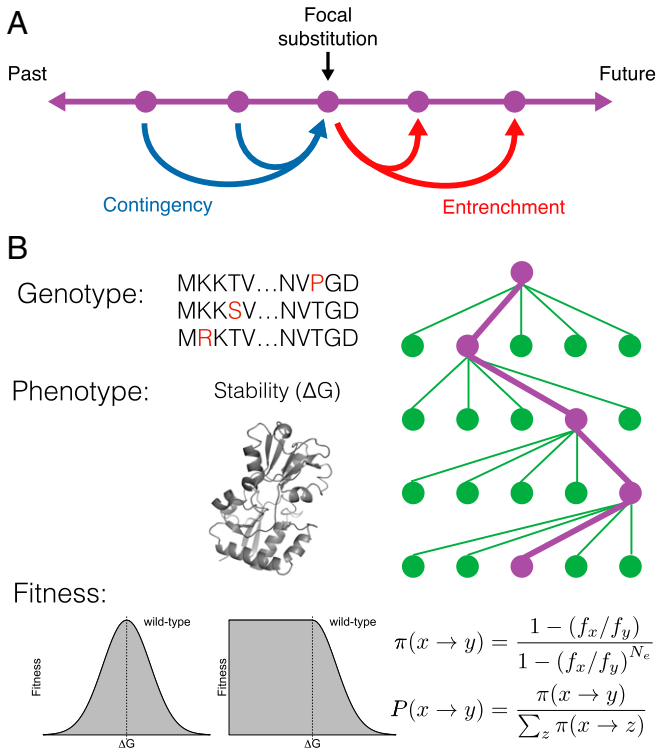

Fig. 1. (*A*) A schematic model indicating how a focal substitution may be contingent on prior substitutions and may constrain future substitutions along an evolutionary trajectory, owing to epistasis. (*B*) A model of protein evolution under weak mutation and purifying selection for thermodynamic stability. Starting from the wild-type sequence of *argT* we propose 10 random 1-aa point mutations. For each of the proposed mutants we compute its predicted stability (ΔG) using FoldX, and its associated fitness. The fitness function is assumed to be either Gaussian or semi-Gaussian, with a maximum at the wild-type stability. One of the proposed mutants fixes in the population, based on its relative fixation probability under the Moran model with effective population size N_e. This process is iterated for 30 consecutive substitutions to produce an evolutionary trajectory. We simulate 100 replicate trajectories, each initiated at the wild-type *argT* sequence.

from stability to fitness is itself nonlinear (18, 26, 60, 61) and because selection is sensitive to selection coefficients as small as the inverse of the population size (62), even slight variation in the stability effects of mutations across different genetic backgrounds may be sufficient to influence the course of evolution.

Using the computational approach summarized above, we will demonstrate that the nearly neutral mutations that fix under purifying selection are, indeed, often epistatic with each other for both stability and fitness. In particular, we find that each mutation that fixes is typically permitted to fix by the presence of preceding substitutions—that is, most substitutions would be too deleterious to fix were it not for epistasis with preceding substitutions. Conversely, we also find that mutations that fix typically become entrenched over time by epistasis—so that a substitution that was nearly neutral when it fixed becomes increasingly deleterious to revert as subsequent substitutions accumulate (18, 39). These results imply an important role for epistasis in shaping the course of sequence evolution in a protein under selection to maintain thermodynamic stability.

Model

Evolutionary Model. We explore the evolution of a protein sequence in the weak-mutation regime, so that each new mutation introduced into the population either is lost or goes to fixation, with probabilities that depend upon the mutant's fitness, before another mutation is introduced (see ref. 63 for a review). Fixation or loss are considered instantaneous so that the population is always monomorphic for a particular protein sequence. We study the 238-aa lysine-arginine-ornithine-binding periplasmic protein (*argT*) from *Salmonella typhimurium* as a model system, chosen because its crystal structure is known (1LAF) and is simple enough that computational predictions for the stability effects of mutations are feasible. We estimate the stability of each proposed mutant sequence using the force-field approach FoldX. We first use the RepairPDB function of FoldX to iteratively remove bad torsion angles and van der Waals clashes, and we then use the BuildModel function to compute stabilities of mutants.

The relevance of our study to protein evolution in nature is intrinsically limited by the accuracy of FoldX in computing the stability effects of mutations. Force-field methods such as FoldX provide only modest accuracy in predicting the effects of specific mutations (64, 65), in part because they approximate multibody interactions by sums of pairwise interactions. Nevertheless, the stability effects of random mutations to the *argT* sequence, as predicted by FoldX, are almost entirely additive (discussed below), in accordance with experimental data (58, 59), and they are also influenced by the native 3D structure (*SI Appendix*). Furthermore, we will compare the magnitude of epistasis observed in our evolutionary simulations using FoldX to empirical data on variation in the stability effects of mutations across different genetic backgrounds.

Although most studies of protein evolution assume that destabilizing mutations decrease protein activity and fitness, the effects of overstabilizing mutations remain unclear. For most of the results presented in the main text, we model purifying selection on protein stability by assuming a Gaussian fitness function centered around the ΔG of the wild-type *argT* sequence (Fig. 1*B*), so that both destabilizing and overstabilizing mutations produce variants with lower fitness than the wild type. This assumption is consistent with empirical measurements on several families of proteins (66–69). In addition, we consider an alternative, semi-Gaussian fitness function that penalizes only destabilizing mutations (discussed below). We assume an effective population size of $N_e = 10^4$ for the purpose of computing the fixation probabilities of mutants. The SD of the fitness function is fixed at 37.75 ΔG kcal/mol, whose value is chosen so that roughly 25% of all possible one-step mutations from the wild-type *argT* sequence have a scaled selection coefficient $|N_e s| < 1$ and about 38% of all mutations are virtually lethal, $N_e s < -20$ (*SI Appendix*, Fig. S1*A*). Here the selection coefficient, s, denotes the difference in log fitness. This choice of fitness function is thus consistent with experimental data on the distribution of fitness effects of mutants (70–73).

We implement evolution under weak mutation as follows. We initialize the population fixed for a starting sequence, always chosen to be the wild-type (Protein Data Bank) *argT* sequence. At each discrete time step we propose a set of 10 point mutations to the current sequence, x. We compute the fixation probability for each of the mutants, y, according to the standard Moran process (74):

$$\pi(x \to y) = \frac{1-(f_x/f_y)}{1-(f_x/f_y)^{Ne}}, \quad [1]$$

where f_x denotes the fitness of genotype x and $\pi(x \to y)$ denotes the fixation probability of a mutant genotype y introduced into a population fixed for genotype x. Next, we let genotype y fix according to its fixation probability relative to all proposed mutants,

$$P(x \to y) = \frac{\pi(x \to y)}{\sum_z \pi(x \to z)}, \quad [2]$$

and we update the state of the population from sequence x to sequence y. We iterate this process for a total of 30 discrete time

steps, each corresponding to a substitution event, so that the final protein sequence is achieved by an evolutionary trajectory of 30 substitutions starting from the initial, wild-type *argT* sequence (Fig. 1*B*). The timescale of our simulations therefore represents roughly 13% divergence at the protein sequence level, which is similar to divergences often studied by comparative sequence analysis. We simulate 100 replicate trajectories, started from the same initial sequence, and we typically report results on the ensemble average.

Quantifying Epistasis, Contingency, and Entrenchment. We seek to understand the structure of epistasis between substitutions along evolutionary trajectories of protein sequences under purifying selection. To quantify epistasis we use a standard definition for pairs of subsequent mutations, as well as a natural generalization of this definition for longer trajectories.

Consider first the case in which the population starts at some genotype S_0 with fitness f_0. Upon fixation of the first substitution the population moves to genotype $S_{0,1}$ with fitness $f_{0,1}$. Upon fixation of the second substitution the population moves to genotype $S_{0,1,2}$ with fitness $f_{0,1,2}$. In the absence of the first mutation, the second mutation would have moved the population to genotype $S_{0,2}$ with fitness $f_{0,2}$. The standard measure of epistasis between these two substitutions is defined as

$$E = \left[\log(f_{0,1,2}) - \log(f_0)\right] - \left(\left[\log(f_{0,1}) - \log(f_0)\right] + \left[\log(f_{0,2}) - \log(f_0)\right]\right). \quad [3]$$

Writing the definition in this way suggests that we view epistasis as the deviation between the fitness effect of the double mutant and the sum of the fitness effects of the single mutants.

This definition of epistasis can alternatively be interpreted in terms of the order in which substitutions occurred along the evolutionary trajectory. For instance, in the above scenario mutation 1 fixes before mutation 2 and it therefore has fitness effect $\log(f_{0,1}) - \log(f_0)$. However, we can also ask what the fitness effect of mutation 1 would have been had the two mutations fixed in the opposite order. In this alternative scenario, the fitness effect of mutation 1 would have been $\log(f_{0,1,2}) - \log(f_{0,2})$. The standard definition of epistasis between a pair of mutants can be rewritten as the difference between these two fitness effects:

$$E = \left[\log(f_{0,1,2}) - \log(f_{0,2})\right] - \left[\log(f_{0,1}) - \log(f_0)\right]. \quad [4]$$

Thus, the standard measure of epistasis can be seen as a measure of how much larger the fitness effect of the first substitution would be if the order of the two substitutions were reversed.

This interpretation of epistasis in terms of substitution order suggests a natural generalization, which will allow us to quantify epistasis in longer evolutionary trajectories. Consider a trajectory starting at the wild-type sequence and then subsequently fixing mutations 1,2,3, $\ldots, n$. For any mutation i, we can ask how much larger the fitness effect of mutation i would have been under the alternative trajectory in which mutation i is removed from position i along the trajectory—where it actually occurred—and instead inserted at some other position j along the trajectory. More formally, in such a trajectory we define the following measure to quantify epistasis between substitutions i and j:

$$E_{(i,j)} = \begin{cases} \left[\log(f_{0,1,\ldots,j-1,i}) - \log(f_{0,1,\ldots,j-1})\right] - \left[\log(f_{0,1,\ldots,i}) - \log(f_{0,1,\ldots,i-1})\right], & \text{for } i \geq j \\ \left[\log(f_{0,1,\ldots,j}) - \log(f_{0,1,\ldots,i-1,i+1,\ldots j})\right] - \left[\log(f_{0,1,\ldots,i}) - \log(f_{0,1,\ldots,i-1})\right], & \text{for } i < j, \end{cases} \quad [5]$$

It is easy to verify that $E_{(i,i+1)}$ reduces to the standard measure of epistasis between two subsequent substitutions.

This generalized definition of epistasis allows us to define what we mean by contingency and entrenchment. A substitution is contingent on previous substitutions if it is more likely to fix as a result of the substitutions that preceded it. More precisely, for $i > j$ we define substitution i to be contingent on the preceding substitutions $j, \ldots, i-1$ if $E_{(i,j)} < 0$. The condition $E_{(i,j)} < 0$ means that substitution i is relatively more beneficial when it actually occurs than it would have been had it occurred at some earlier time step, j. Conversely, we say that a substitution i is entrenched by subsequent substitutions if it becomes relatively more deleterious to revert as a result of the subsequent substitutions. More precisely, for $i < j$ we say a substitution i is entrenched by subsequent substitutions $i+1, \ldots, j$ if $E_{(i,j)} > 0$. The condition $E_{(i,j)} > 0$ means that the effect of reverting substitution i at time j is relatively more deleterious than it would have been to revert substitution i immediately after it initially occurred.

Results

Mutational Effects on Protein Stability. Random mutations in a protein-coding sequence typically destabilize the protein structure (26, 72, 75–80). Thus, if protein evolution proceeded solely via random substitutions, without any selection, we would expect a decrease in protein stability over time. However, under purifying selection to maintain a given degree of thermodynamic stability, strongly destabilizing (or overstabilizing) mutations will have low fitness and correspondingly low fixation probability, so that the only mutations that substitute will tend to produce stabilities similar to that of the wild-type sequence.

We simulated the evolution of the *argT* protein sequence under selection for its native stability, starting from the wild-type sequence, computing stabilities (ΔG) and fixation probabilities of mutants as described above. Starting from the wild-type sequence, most one-step mutations are destabilizing (84%, binomial test $P < 10^{-15}$). However, among the one-step mutations that fix in our simulations of purifying selection, there is no significant bias toward destabilization (54%, binomial test $P = 0.48$). This is due to the fact that the average destabilizing effect is significantly greater than the average stabilizing effect (t test $P < 10^{-15}$), and so the average fitness of a destabilizing mutation is significantly lower than that of a stabilizing mutation (t test $P < 10^{-15}$). More generally, we find that the mean stability effect of all substitutions at their time of fixation is quite small, mean $|\Delta\Delta G| = 0.58$ kcal/mol, with almost an equal number of stabilizing (48%) and destabilizing (52%) mutations fixing along evolutionary trajectories. These substitutions are typically nearly neutral (mean $|N_e s| = 2.34$) (*SI Appendix*, Figs. S1*B* and S2*B*), such that the fitness of the protein decreases by only ∼0.04% on average after 30 substitutions. In addition to having mild effects on stability and fitness, substitutions are distributed nonrandomly in the protein structure. We find more substitutions at sites with greater solvent-accessible surface area (Pearson's correlation $\rho = 0.54$, $P < 10^{-15}$; see *SI Appendix*) as well as at residues occupying small volumes in the protein (Pearson's correlation $\rho = -0.22$, $P = 0.0008$; see *SI Appendix*), consistent with biophysical expectations (46, 81–83).

By contrast, when we simulate protein sequence evolution via the fixation of random point mutations—that is, without any

selection at all—then the stability for the native structure decreases along evolutionary trajectories, as illustrated by the ensemble mean trajectory shown in *SI Appendix*, Fig. S2*A*. Likewise, in the absence of selection substitutions are more often destabilizing than stabilizing (binomial test, $P < 10^{-15}$), as expected from empirical studies on the effects of random mutations (61, 72, 75, 76, 80).

Epistasis Along Evolutionary Trajectories: Contingency. We quantified the structure of epistasis between substitutions along evolutionary trajectories of *argT* sequences simulated under purifying selection. We used a generalized definition of epistasis $E_{(i,j)}$ that applies to any pair of substitutions i and j along a trajectory (Model). We first studied the degree of contingency between substitutions in these trajectories. For $i > j$ we say that substitution i is contingent on the preceding substitutions $j, \ldots, i-1$ if the condition $E_{(i,j)} < 0$ holds. This contingency condition means that substitution i is relatively more beneficial at the time of its actual fixation than it would have been had it been introduced at some earlier step, j.

We find that substitutions in *argT* under purifying selection are often epistatic and they tend to be contingent on earlier substitutions. Fig. 2*A* (left side) illustrates this phenomenon by focusing on contingency between the substitutions that occur at step $i = 16$ and the substitutions that occur at earlier steps $j < 16$, among an ensemble of 100 replicate evolutionary trajectories. The mean epistasis measure $E_{(16,j)}$ is significantly less than zero for each step $j < 16$ (t test, $P < 0.002$ for each j)—indicating that the substitutions that fix at step $i = 16$ are contingent on earlier substitutions.

There is a subtlety associated with the contingency condition $E_{(i,j)} < 0$, which compares the selection coefficient of substitution i when it fixes versus the selection coefficient of the same mutation had it fixed at some earlier step, j. These two selection coefficients can each be negative or positive. The condition $E_{(i,j)} < 0$ means that substitution i is "relatively more beneficial" at time i compared with at a prior time; this includes the possibility that substitution i is in fact deleterious, but less deleterious at the time of its actual fixation compared with having fixed at some earlier time. In practice, in simulations under purifying selection most of the mutations that fix along the evolutionary trajectory are neutral or nearly neutral at the time of their fixation (*SI Appendix*, Fig. S1*B*). So, in these simulations the condition $E_{(i,j)} < 0$ typically means that substitution i would have been deleterious had it occurred at the earlier step j.

The extent of contingency in our simulations is illustrated in Fig. 2*B, Top*, which compares the selection coefficients of the mutations that fix at all steps $i = 2 \ldots 30$ with the selection coefficients of the same mutations had they been introduced at earlier steps $j < i$ along their evolutionary trajectories. When considering all pairs of ordered substitutions $j < i$ in our simulations under purifying selection we find a mean value $\overline{N_e E} = -5.86$ and that $\sim 70\%$ of pairs exhibit $E_{(i,j)} < 0$. In other words, the great majority of mutations that fix are contingent on earlier substitutions—that is, the same mutations would typically be deleterious were they introduced in prior genetic backgrounds. These results imply that, even under purifying selection for stability, the mutations that fix during the evolution of a protein sequence are typically contingent on the history of prior substitutions.

Epistasis Along Evolutionary Trajectories: Entrenchment. We have shown that mutations that fix under purifying selection are contingent on earlier substitutions. Now we ask the converse question: What is the effect of later substitutions on the fitness effects of

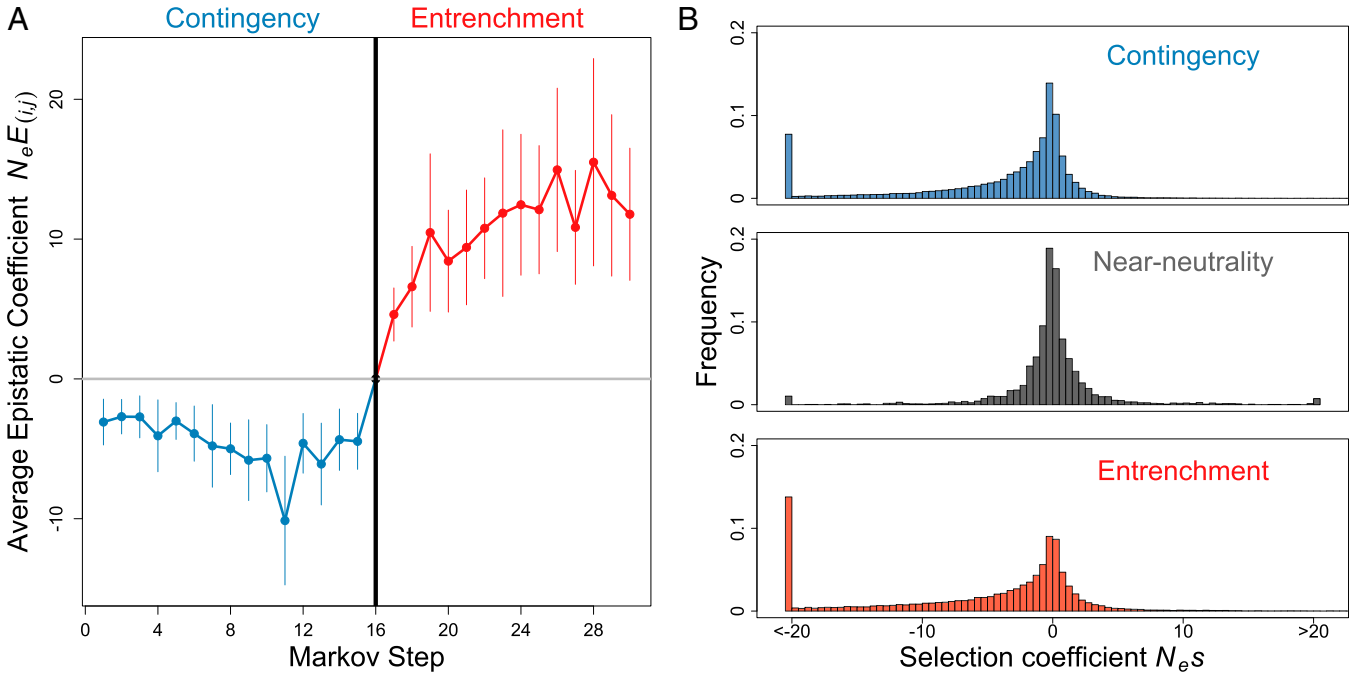

Fig. 2. (*A*) Substitutions that accrue under purifying selection are typically epistatic: They exhibit both contingency with earlier substitutions and entrenchment by later substitutions. *A* indicates the fitness effects of the mutations that fix at step $i = 16$ if they were introduced into earlier (contingency $j < 16$) or later (entrenchment $j > 16$) genetic backgrounds. Under purifying selection, the epistatic coefficients $E_{(16,j)}$ are significantly less than zero, on average, for all $j < 16$ and significantly greater than zero for all $j > 16$. Thus, the substitutions under purifying selection, which are nearly neutral when they fix, are contingent on earlier substitutions, and they become more deleterious to revert as later substitutions accrue. Vertical bars indicate ±2 SE around the ensemble mean of 100 replicate simulated populations. (*B*) Distribution of scaled selection coefficients ($N_e s$) for all substitutions that fix along evolutionary trajectories. The gray histogram shows the distribution of selection coefficients of these mutations at the time that they fix ("near-neutrality"), the blue histogram shows the distribution of selection coefficients for the same mutations i if they were introduced in earlier backgrounds $j = 0, \ldots, i-1$ ("contingency"), and the red histogram shows the distribution of selection coefficients for the same mutations i if they are removed from later backgrounds $j = i+1, \ldots, 30$ ("entrenchment").

substitutions that have already fixed? In particular, we ask whether mutations that are nearly neutral when they fix subsequently become deleterious to revert later in the trajectory—a phenomenon that Pollock et al. (18) have called an "evolutionary Stokes shift."

A positive value of $E_{(i,j)}$ for $j > i$ means that reverting a focal substitution i in a later background containing mutations $1, \ldots, j$ is relatively more deleterious than reverting it immediately after it fixes in the population. Thus, $E_{(i,j)} > 0$ indicates entrenchment of substitution i by the following substitutions $i+1, \ldots, j$.

We find that substitutions under purifying selection are typically entrenched by later substitutions. Fig. 2*A* (right side) illustrates this phenomenon by focusing on entrenchment of substitutions that occur at time $i = 16$ by substitutions that occur at later time points $j > i$ along the same evolutionary trajectories. The mean entrenchment coefficient $E_{(16,j)}$ is significantly greater than zero for each subsequent step $j > 16$ (t test, $P < 10^{-3}$ for each j). In other words, even though most of these mutations are nearly neutral at the time of fixation, reverting the same mutations from later genetic backgrounds is typically deleterious.

More generally, when considering all ordered pairs of substitutions under purifying selection, the epistatic values $E_{(i,j)}$ for $j > i$ are significantly greater than zero on average (t test, $P < 10^{-15}$) with a mean value $\overline{N_e E} = 9.96$, meaning that substitutions are more deleterious to revert in later backgrounds. In particular, we find that ~72% of pairs $j > i$ exhibit positive values $E_{(i,j)} > 0$, indicating a strong tendency for later substitutions to entrench earlier substitutions.

Moreover, the degree to which a substitution becomes entrenched by epistasis tends to increase with each subsequent substitution that accrues. A positive slope of $E_{(i,j)}$ versus j, for $j > i$, indicates that the focal substitution i becomes increasingly deleterious to revert as subsequent substitutions accumulate. We estimated the slope of $E_{(i,j)}$ versus j using least squares and found that this slope is significantly positive, on average, across all steps i in our simulations (one-tailed t test, $P < 10^{-15}$). Likewise, over 80% of substitutions exhibit positive slopes, indicating a tendency for the strength of entrenchment to increase over time (see also *SI Appendix*, Fig. S3). Thus, even under purifying selection, we find that protein-coding substitutions are rendered "irreversible" by subsequent substitutions and that the strength of irreversibility tends to increase with time.

The trend of increasing entrenchment that we have observed in our simulations has an intuitive explanation. After a focal mutation fixes in a protein, subsequent substitutions are typically contingent on its presence. As a result, reverting the focal substitution at a later point along the evolutionary trajectory becomes increasingly deleterious, because it interacts with a greater number of intervening substitutions. Therefore, at least on the timescale of divergence we have studied, we naturally expect that the degree of a substitution's entrenchment should increase over time. Over very long time scales, however, as substitutions begin to saturate, the degree of entrenchment will likely level off or perhaps even decrease.

Epistasis Between Consecutive Substitutions. We have shown that the selection coefficient of a given substitution is contingent on prior substitutions and becomes entrenched by subsequent substitutions, constraining evolution against reversions as time proceeds. However, does epistasis constrain the paths available to evolution on shorter time scales as well—that is, between consecutive substitutions?

To address this question we consider an evolutionary trajectory starting at genotype A followed by subsequent substitutions B and C, producing the trajectory $A \to AB \to ABC$. We ask how likely is the observed path compared with the alternative path $A \to AC \to ACB$. Assuming no back mutations, the probabilities of the two paths are determined solely by the probability of the first substitution. We calculate the probability of seeing one path versus the other based on their fixation probabilities:

$$P(A \to AB \to ABC) = \frac{\pi(A \to AB)}{\pi(A \to AB) + \pi(A \to AC)}. \quad [6]$$

A value $P(A \to AB \to ABC) > 1/2$ indicates that the actual path taken during evolution ($A \to AB \to ABC$) is more favorable than the alternative path ($A \to AC \to ACB$), and vice versa.

We calculated the relative probabilities of actual and alternate paths for all pairs of consecutive substitutions in the ensemble of simulated evolutionary trajectories. These probabilities, whose distribution is shown in Fig. 3, exhibit an interesting bimodal pattern. For a large portion of consecutive substitutions, including those whose effects are additive, the actual and alternative paths were almost equally probable, producing a mode near 0.5. By contrast, for another large portion of consecutive substitutions (>26% of pairs), the actual path was more than 30 times as likely as the alternate path, producing a mode near 1. This second mode indicates a high degree of epistasis: Many substitutions are conditional on the presence of the immediately preceding substitution. Indeed, 19% of consecutive substitutions are highly contingent ($N_e E_{(i+1,i)} < -10$). Thus, even over short timescales, epistasis plays a large role in shaping the paths taken by evolution under purifying selection.

Sources of Epistasis. The high degree of epistasis for fitness observed in our simulations under purifying selection could result from two alternative sources: epistasis in the computationally predicted protein stabilities themselves, or epistasis in the nonlinear mapping from stability to fitness (or both). If the combined effect of multiple substitutions on predicted stabilities does not equal the sum of their individual effects, then this form of epistasis for stability would induce epistasis in the fitness effects

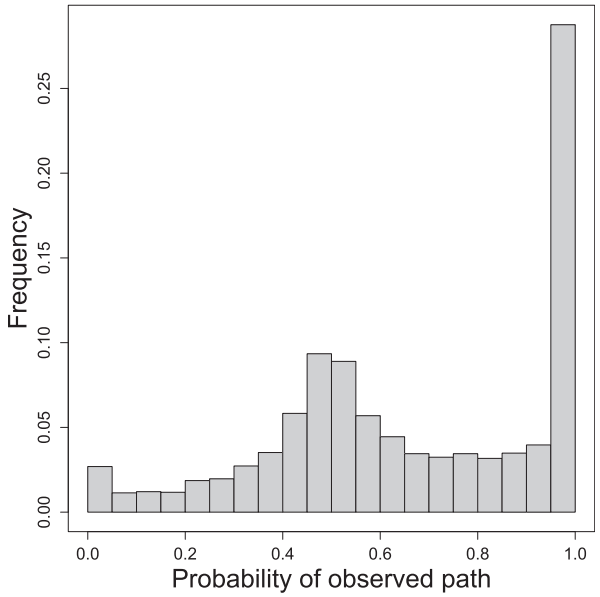

Fig. 3. Epistasis constrains paths available to evolution. The figure shows the relative probability of fixing two consecutive substitutions (*B* and *C*) in their observed order in simulated evolution ($A \to AB \to ABC$) compared with the reversed order ($A \to AC \to ACB$). Under purifying selection for stability, the distribution of relative fixation probabilities is distinctly bimodal. A large proportion of substitutions have almost equal probability of taking either path, producing a mode near 0.5. For another large portion (>26%) of pairs, the observed path is more than 30 times as likely as the alternate path (producing a mode near 1), indicating that many substitutions are highly contingent on the immediately preceding substitution.

of substitutions. Alternatively, even in the absence of epistasis for protein stability, epistasis in fitness may arise from the nonlinear mapping between stability and fitness. To resolve which of these two effects dominates we undertook additional analyses.

Briefly, we defined two measures that quantify the degree of epistasis in protein stabilities themselves, and the degree of epistasis arising from the stability-to-fitness mapping (*SI Appendix*). We found that epistasis for stability explains a large proportion of the observed variance in epistasis for fitness ($R^2 = 0.517$, *SI Appendix*, Fig. S6*A*). By contrast, epistasis arising solely from the stability-to-fitness map explains very little variance in epistasis for fitness [$R^2 = 0.02$, *SI Appendix*, Fig. S6*B*; the two R^2 values reported here are not expected to sum to 1 (see *SI Appendix*)]. Thus, our results on epistasis for fitness (Fig. 2) are driven primarily by epistasis in the effects of mutations on protein stabilities themselves, rather than nonlinearities in the stability-fitness map.

The strong influence of epistasis for protein stability in our simulations is surprising in light of experimental data showing that the effects of mutations on stability are typically additive (58, 59). To determine whether the nonadditivity we detect is an artifact of our computational procedure for estimating the stability of mutations, we constructed 1,000 pairs of single mutations, and their corresponding double mutants, around the wild-type *argT* sequence. We found that the stability effects of these double mutants were very closely predicted by the summed effects of their corresponding single mutants ($R^2 = 0.96$, Fig. 4*A*). Similarly, fitting an additive model for stability to each individual pair of mutations shows that most pairs are very nearly additive (median $R^2 = 0.999$; see *SI Appendix*). We performed the same exercise, constructing 1,000 pairs of single mutations and their corresponding double mutations, around an evolved *argT* sequence which differs at 16 sites from the wild type. Once again, we found that the stability effects of double mutants are well predicted by the summed effects of single mutants ($R^2 = 0.92$, Fig. 4*B*), and additive models typically explain most of the variation in stability (median $R^2 = 0.974$ for individual pairs, see *SI Appendix*). Furthermore, the stability effects of all single mutations in the wild-type *argT* sequence are highly correlated with their effects in the evolved *argT* sequence ($R^2 = 0.83$, Fig. 4*C*). This correlation, produced by FoldX, is comparable to the correlation of stability effects across two genetic backgrounds with the same level of divergence as measured experimentally by Ashenberg et al. (61) ($R^2 = 0.90$). All of these results confirm that the effects of random mutations on stability predicted by FoldX are almost entirely additive, in accordance with experimental data (58, 59, 61).

However, when we repeat the same tests for additivity on the consecutive substitutions that occur in our simulated evolutionary trajectories, a very different picture emerges. These substitutions that occur under purifying selection are much less additive ($R^2 = 0.26$ when predicting the stability effects of double mutants

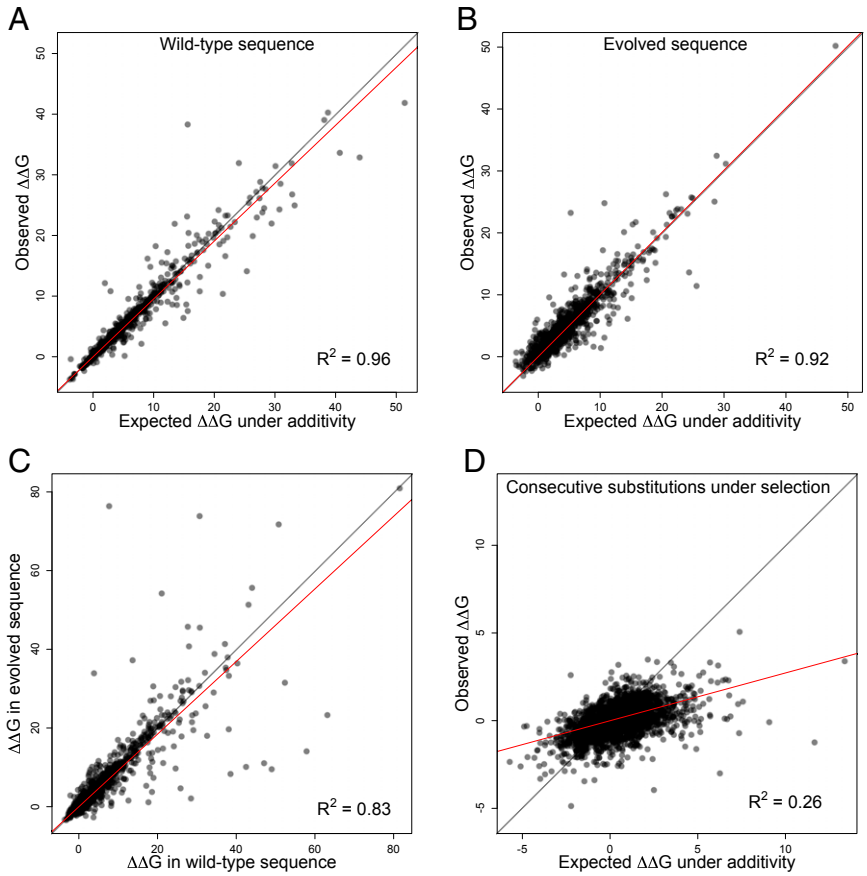

Fig. 4. Additivity of stability effects. (*A*) The effects of random mutations on protein stability as calculated by FoldX. The ΔΔG of double mutants in the wild-type *argT* sequence are highly correlated with the summed effects of their corresponding single mutations. (*B*) Starting from an evolved *argT* sequence, which differs from the wild type by 16 substitutions, the ΔΔG of double mutants are, again, highly correlated with the summed effects of their corresponding single mutations. (*C*) The stability effects of all point mutations around the wild-type *argT* sequence are highly correlated with their effects in the evolved *argT* sequence. (*D*) By contrast, the stability effects of consecutive substitutions along evolutionary trajectories simulated under purifying selection are only weakly additive: the effects of double mutants correlate weakly with the summed effects of single mutants. The line $y = x$ is represented in black and the best-fit regression line with zero intercept ($y = \beta x$) is represented in red in each panel.

from the summed effects of single mutants, Fig. 4*D*; median $R^2 = 0.82$ when fitting a linear model to each pair of mutations individually). This suggests that, whereas random mutations have nearly additive effects on stability (Fig. 4 *A*–*C*), evolution under purifying selection enriches for substitutions with epistatic effects on stability (Fig. 4*D*).

Magnitude of Epistasis. To quantify the magnitude of epistasis for stability in our simulations, we examined the effects of the mutations that fixed across a range of different genetic backgrounds. In particular, for each mutation that fixed along an evolutionary trajectory, we assayed its stability effect in all 30 genetic backgrounds from the same trajectory, calculating both the the SD of ΔΔG, to determine the across-background variation in stability effects, and the mean |ΔΔG|, to determine the across-background mean stability effect.

We found that the across-background SD in stability effects has an ensemble mean of 0.80 kcal/mol. This value is small compared with the average magnitude of stability effects of random mutations in the wild-type background (mean $|\Delta\Delta G| = 2.98$ kcal/mol). In other words, the degree of epistasis for stability along the evolutionary trajectories is small compared with the typical effects of random mutations (see also *SI Appendix*, Fig. S8). This degree of variation observed in our simulations is roughly consistent with the experimental results of Risso et al. (84), who report changes in the stability effects of mutations between modern and reconstructed ancestral backgrounds in the range of ±1 kcal/mol, as well as with the variation in stability effects reported by Ashenberg et al. (61).

The across-background mean effect on stability (mean $|\Delta\Delta G| = 1.02$ kcal/mol) was also smaller than the mean effect of random mutations in the wild-type background (mean $|\Delta\Delta G| = 2.98$ kcal/mol). Moreover, when introducing observed substitutions across all 30 genetic backgrounds from the same trajectory, the frequency of absolute stability effects exceeding 2.98 kcal/mol was only 0.04. Thus, the mutations that fix in our simulations of purifying selection tend to have relatively mild effects on stability, across many backgrounds.

Whereas the mutations that fix during our simulations have mild effects across many backgrounds, they have yet milder effects on the genetic background in which they actually fixed: mean $|\Delta\Delta G| = 1.02$ kcal/mol across genetic backgrounds versus mean $|\Delta\Delta G| = 0.58$ kcal/mol at the time of fixation. This value is also consistent with the results of an experimental study by Serrano et al. (59), who found that mutations that fix along a trajectory tend to have $|\Delta\Delta G| < 1$ kcal/mol.

Taken together, the results above help to resolve the apparent contradiction between the lack of epistasis in the stability effects of random mutations, compared with the prevalence of epistasis for the mutations that fix under simulated purifying selection. Natural selection in our simulations permits only mutations of very small effect to fix. A mutation can have a very small effect either because it has a very small effect in all backgrounds, or because epistatic interactions make its effect especially small in the background in which it fixes. The analyses above show that the mutations that fix tend to have small effects in most backgrounds, but they have yet smaller effects on the particular background in which they fix. Because purifying selection enriches for mutations of small effect, it therefore also enriches for mutations with epistatic interactions that ameliorate its stability effects at the time of its fixation. Thus, even though there is only a small amount of epistasis between random mutants, purifying selection on protein stability will enrich for epistasis among the mutations that fix. This phenomenon reflects the general principle of "regression to the mean" (85): Choosing observations based on the high value of a response variable enriches for both observations whose predictor variables produce a high response and for observations with large positive error terms (86, 87). In other words, purifying selection on protein stability is expected to enrich for epistasis, even though the effects of random mutations on stability are virtually additive.

Robustness of Simulation Results

Alternate Fitness Function. Our model of purifying selection assumes that overstabilizing mutations are as deleterious as destabilizing mutations, so that only the wild-type stability has optimal fitness. However, several studies have shown that overstabilizing mutations can be neutral under stabilizing selection (75, 77, 80, 88). Therefore, we also considered an alternative, semi-Gaussian fitness landscape in which *argT* sequences more stable than the wild type are just a fit as the wild type (Fig. 1*B*). We chose the variance of the semi-Gaussian to ensure, as with the Gaussian, that roughly roughly 25% of all possible one-step mutations from the wild-type *argT* sequence are nearly natural ($|N_e s| < 1$, see *SI Appendix*, Fig. S4). We ran the same set of simulations (100 replicate trajectories, each for 30 substitutions) on this alternative fitness landscape, and we found that our results remain qualitatively unchanged.

The absolute effects of all substitutions that accrue on the semi-Gaussian landscape (mean $|\Delta\Delta G| = 0.77$ kcal/mol) are slightly higher on average than under the Gaussian landscape (mean $|\Delta\Delta G| = 0.58$ kcal/mol), owing to the lack of fitness penalty for large stabilizing substitutions. Nonetheless, evolved proteins that have accrued 30 substitutions on the semi-Gaussian landscape are only marginally more stable (~0.5 kcal/mol) on average than the initial, wild-type sequence.

Unlike on the Gaussian landscape, where strict neutrality is extremely rare, 25% of consecutive substitutions are strictly neutral on the semi-Gaussian landscape (*SI Appendix*, Fig. S5). Despite this difference, the overall fraction of highly epistatic consecutive substitutions—substitutions for which evolution is 30 times more likely to proceed via the observed path than the alternate path—is similar for both Gaussian (~26%) and semi-Gaussian (~23%) fitness landscapes (*SI Appendix*, Fig. S5). All of our other results on epistasis in the Gaussian simulations are also similar to the semi-Gaussian simulations: In ~53% of pairs, later substitutions were contingent on earlier substitutions ($E_{(i,j)} < 0$), with mean value $\overline{N_e E} = -11.22$, and in 52% of pairs earlier substitutions were entrenched by subsequent substitutions ($E_{(i,j)} > 0$), with mean value $\overline{N_e E} = 18.22$ (*SI Appendix*, Fig. S4*D*). Finally, we find that 76% of substitutions show increasing entrenchment in semi-Gaussian simulations (Binomial test, $P < 10^{-15}$, slope based on 20 substitutions or more), similar to the Gaussian case.

As in the Gaussian case, epistasis for fitness observed during evolution on the semi-Gaussian fitness landscape is primarily due to nonadditivity in ΔΔG of nearly neutral substitutions. Consecutive substitutions along semi-Gaussian evolutionary trajectories are less additive for stability ($R^2 = 0.4$, *SI Appendix*, Fig. S7) than random mutations around either the wild-type sequence or around an evolved sequence 16 substitutions away ($R^2 > 0.9$, Fig. 4 *A* and *B*). Furthermore, epistasis in stability explains a large proportion of epistasis in the fitness effects of substitutions ($R^2 = 0.33$), whereas the nonlinear mapping from stability to fitness accounts for a very small fraction of epistasis in fitness ($R^2 = 0.03$, *SI Appendix*, Fig. S6 *C* and *D*).

As in the Gaussian case, the average magnitude of stability effects of fixed substitutions across backgrounds (mean $|\Delta\Delta G| =$ 1.10 kcal/mol) for the semi-Gaussian landscape is smaller than the average magnitude of stability effects of random mutations in the wild-type background (mean $|\Delta\Delta G| = 2.98$ kcal/mol). Likewise, the frequency of mutations in the across-background semi-Gaussian dataset with an absolute stability effect larger than 2.98 kcal/mol was only 0.05. Thus, the mutations that fix in our semi-Gaussian simulations tend to have relatively mild effects on stability across many backgrounds.

In summary, our results are qualitatively the same under both the Gaussian and semi-Gaussian fitness landscapes. This concordance reflects the simple fact that mutations increasing stability beyond that of wild type are extremely rare (80), and so the shape of the fitness function for stabilities greater than wild-type has little effect on the evolutionary dynamics.

Larger Sample of Random Mutations. The results reported above are based on 100 replicate simulations of *argT* evolution under weak mutation. At each discrete step in these simulations we proposed 10 point mutations, for reasons of computational tractability, from which one was chosen to substitute. To verify that our results are not influenced by the relatively small sample of mutations, we ran a set of shorter simulations (100 replicates, each for 20 substitutions) proposing in this case 100 point mutations at each step. All of our qualitative results remain unchanged under this larger sampling scheme (*SI Appendix*, Fig. S9 and Table S1).

Discussion

We have developed a computational framework for studying the evolution of protein sequences under purifying selection for native structure and stability. Using the ligand-binding protein *argT* as a representative example, our results reveal extensive epistasis between the mutations that fix under selection. These results suggest a coherent picture of the role of epistasis in protein evolution under long-term purifying selection.

We find that although most mutations are nearly neutral when they fix, the same mutations would typically be deleterious if introduced on earlier genetic backgrounds. Thus, the substitutions that accrue along an evolutionary trajectory are typically contingent on epistatic interactions with earlier substitutions. In fact, a sizable fraction of substitutions are contingent upon the presence of the immediately preceding substitution.

We also find that once a mutation fixes in a protein, the fitness effect of reverting the mutation becomes more deleterious over time. That is, after a mutation fixes it becomes entrenched and difficult to remove due to epistatic interactions with subsequent substitutions. In addition, the degree of entrenchment tends to increase over time.

Taken together, our computational studies of protein evolution under purifying selection suggest that epistasis induces both contingency and entrenchment. There are also theoretical reasons to expect that these two phenomena will occur generically in any fitness landscape that combines the conditional neutrality of mutations with a mode of evolution in which substitutions fix sequentially. In particular, both of these phenomena are consequences of the fact that the fitness effects of a substitution depend on substitutions that precede it.

The quantitative approach used here also allows us to dissect the sources of epistasis causing contingency and entrenchment in our simulations. We find that epistasis is due, in large part, to nonadditivity in the effects of mutations on protein stability. This result is surprising because, both empirically and in our own simulations, the effects of most mutations on protein stability are nearly additive (58, 59). The resolution to this apparent paradox comes from recognizing that natural selection can detect very small differences in fitness. Thus, even a small amount of epistasis at the level of stability can have a profound effect on the evolutionary process. Furthermore, only those mutations with very small effects are permitted to fix under purifying selection. This form of selection enriches for epistasis because, although the mutations that fix tend to have small stability effects in most backgrounds, these mutation have particularly small stability effects in the backgrounds in which they fix, owing to epistasis. Indeed, the observed enrichment for epistasis is simply an example of the principal of regression to the mean, which has previously been implicated in shaping the frequency of epistasis in adaptive evolution (86, 87).

Our study provides insight into a recent debate concerning the degree to which amino acid preferences at a site change as a protein evolves (18, 61, 84, 89). Based on simulations of long-term protein evolution similar to those conducted here, Pollock et al. (18) argued that coevolution between sites would result in site-specific amino acid preferences that change substantially over time. In particular, they suggested that the longer an amino acid remains fixed at a site, the more deleterious it should become to revert. Ashenberg et al. (61) and Risso et al. (84) responded with empirical evidence that the stability effects of mutations are largely conserved over time. Our results suggest a possible resolution to this debate, by showing that even a relatively small degree of nonadditivity in the stability effects of mutations can have a large effect on the evolutionary process. If, as we observe, only those mutations with small stability effects (e.g., $|\Delta\Delta G| < 1$ kcal/mol) can fix, then nonadditivity on the order of 1 kcal/mol [comparable to that reported by Risso et al. (84)] is sufficient to render a substantial fraction of mutations that fix effectively irreversible except on a subset of genetic backgrounds. Thus, epistasis for stability may still produce increasing entrenchment over time even if the stability effects of mutations remain largely conserved.

Our analysis is also consistent with the results of two recent comparative studies of sequence evolution that sought to evaluate whether site specific preferences change over time. Naumenko et al. (39) studied the rate of reversion and found that the longer an amino acid had been present at a site, the lower the reversion rate to the ancestral amino acid. This is precisely what would be expected if the entrenchment observed in our simulations occurs in nature. Similarly, Goldstein et al. (90) studied the probability of parallel evolution along two lineages as a function of the evolutionary distance between those two lineages. They found that the larger the evolutionary distance between the two lineages, the lower the rate of convergent evolution. This is also consistent with our hypothesis that the mutations permitted to fix at a particular site are contingent on earlier mutations: The more diverged a pair of lineages, the fewer preceding substitutions they share, and the greater the difference between the set of substitutions that are acceptable at a site.

Our results are also related to several other recent studies on protein evolution under purifying selection. Breen et al. (32) have recently argued that epistasis of the form described here—where some substitutions are only permissible due to preceding substitutions—is the primary factor in molecular evolution. Although the formal validity of their inference has been the topic of debate (91), our simulation results are in accordance with their basic contention and provide a detailed view of the form of epistasis in proteins under purifying selection. Our results are also consistent with the results of both theoretical (60) and empirical studies (6, 72) showing that epistatic interactions are in large part governed by underlying biophysical interactions between substitutions.

All of our analysis has been enabled by formulating a model that assigns fitness effects to mutations based on computational predictions of protein stabilities. This approach accounts for possible dependencies among sites, whereas most models of protein sequence evolution along a phylogeny assume that sites evolve independently (51–55). Such phylogenetic models necessarily disregard any possible epistatic interactions between sites. Although convenient for reconstructing phylogenies or calculating simple summary statistics, such as dN/dS, we know that proteins are in fact highly coordinated structures whose residues often experience physiochemical interactions that fundamentally determine fold, stability, and function. Our results suggest that incorporating these biophysical factors, and the resulting nonindependence between sites, may produce more accurate models of protein evolution (92–94).

The approach we have used here nonetheless makes a number of simplifying assumptions. In particular, our evolutionary simulations do not allow cosegregating mutations—that is, we assume weak mutation. Although this assumption is typical in models of long-term molecular evolution (but see refs. 95–97 for some exceptions), it is known that polymorphism can substantially affect the dynamics of an evolving population because a compensatory mutation can occur on the background of a segregating deleterious allele (98–100). More work is required to understand how polymorphism in a population might affect the prevalence of contingency and entrenchment.

We have also assumed that purifying selection acts on the global stability of a protein. In reality, however, it is likely that the strength of selection on stability varies within a protein—so that the protein core experiences stronger purifying selection than the periphery (46, 81–83). Incorporating local stability requirements would certainly improve our understanding of selective constraints, but it seems unlikely to qualitatively change our results on the dominant sources of epistasis that modulate substitutions.

Our analysis has neglected other aspects of purifying selection on thermodynamic aspects of proteins—in particular, selection against adopting alternative structures (101, 102). Ideally, one could incorporate negative selection against alternative structures by threading sequences against a large "decoy" database of alternative structures. Even though decoy datasets do not always represent all competing folds, leading to errors in energy analysis, adding this additional constraint, when it becomes computationally feasible, may yield important insights into the action of selection as a protein sequence moves away from the wild type, as well as insights into the origins of novel protein folds.

Finally, selection for stability is not the only source of selection on a protein. A ligand-binding protein, such as considered here, also experiences selection for its function—namely, binding its target. Substitutions that are nearly neutral with respect to stability might significantly alter the function and will be unlikely to fix, or vice versa (77). However, the number of residues directly involved in a ligand-binding protein's function is typically small in comparison with those that predominantly influence its stability (77). Hence our conclusions regarding epistatic nature of substitutions are unlikely to be altered substantially by incorporating constraints on ligand-binding function.

Our approach to studying epistasis in protein evolution is fundamentally limited by our ability to computationally estimate the stability effects of mutations. Although FoldX is one of the state-of-the art force-field methods for such computations, and it likely provides greater accuracy than computations based on lattice structures or simple contact potentials (18, 61, 77, 78), the ability to accurately predict the effects of specific mutations is still quite limited (64, 65). Even though we are interested in aggregate patterns across many substitutions, rather than the effects of individual mutations, our exploration of protein evolution has still been restricted by computational cost, which required us to sample a relatively small subset of proposed mutations. (We have, at least, shown that are results remain unchanged by increasing sample size 10-fold.) In addition, the accuracy of computational predictions is reduced further as protein sequences diverge from the wild type. We do, however, find results for the patterns of epistasis between substitutions near the wild-type sequence similar to those we find toward the end of our simulated evolutionary trajectories. This suggests that there is no systematic bias in our results introduced by decaying accuracy of stability predictions.

Our simulation results on epistasis provide a clear direction for future experimental investigation. Whereas we have provided a general statistical explanation for the increased prevalence of epistasis among mutations that fix during protein evolution, empirical studies may elucidate the specific biophysical mechanisms underlying such nonadditive interactions.

ACKNOWLEDGMENTS. J.B.P. acknowledges funding from the Burroughs Wellcome Fund, the David and Lucile Packard Foundation, US Department of the Interior Grant D12AP00025, and Foundational Questions in Evolutionary Biology Fund Grant RFP-12-16. D.M.M. acknowledges funding from NIH training Grant 2T32AI055400-11. D.M.M. and J.B.P. acknowledge funding from US Army Research Office Grant W911NF-12-1-0552. We acknowledge computational support from Open Science Grid and Extreme Science and Engineering Discovery Environment, which is supported by National Science Foundation Grant ACI-1053575.

1. Weinreich DM, Delaney NF, Depristo MA, Hartl DL (2006) Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312(5770):111–114.
2. Ortlund EA, Bridgham JT, Redinbo MR, Thornton JW (2007) Crystal structure of an ancient protein: Evolution by conformational epistasis. *Science* 317(5844):1544–1548.
3. Blount ZD, Borland CZ, Lenski RE (2008) Historical contingency and the evolution of a key innovation in an experimental population of Escherichia coli. *Proc Natl Acad Sci USA* 105(23):7899–7906.
4. Bridgham JT, Ortlund EA, Thornton JW (2009) An epistatic ratchet constrains the direction of glucocorticoid receptor evolution. *Nature* 461(7263):515–519.
5. Bloom JD, Gong LI, Baltimore D (2010) Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. *Science* 328(5983):1272–1275.
6. McLaughlin RN, Jr, Poelwijk FJ, Raman A, Gosal WS, Ranganathan R (2012) The spatial architecture of protein function and adaptation. *Nature* 491(7422):138–142.
7. Natarajan C, et al. (2013) Epistasis among adaptive mutations in deer mouse hemoglobin. *Science* 340(6138):1324–1327.
8. Gong LI, Suchard MA, Bloom JD (2013) Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife* 2:e00631.
9. Harms MJ, Thornton JW (2014) Historical contingency and its biophysical basis in glucocorticoid receptor evolution. *Nature* 512(7513):203–207.
10. Gould SJ (1990) *Wonderful Life: The Burgess Shale and the Nature of History* (Norton, New York).
11. Mani GS, Clarke BC (1990) Mutational order: A major stochastic process in evolution. *Proc R Soc Lond B Biol Sci* 240(1297):29–37.
12. Travisano M, Mongold JA, Bennett AF, Lenski RE (1995) Experimental tests of the roles of adaptation, chance, and history in evolution. *Science* 267(5194):87–90.
13. Beatty J (2006) Replaying life's tape. *J Philos* 103(7):336–362.
14. Muller HJ (1939) Reversibility in evolution considered from the standpoint of genetics. *Biol Rev Camb Philos Soc* 14:261–280.
15. Riedl R (1977) A systems-analytical approach to macro-evolutionary phenomena. *Q Rev Biol* 52(4):351–370.
16. Wimsatt WC (1986) *Integrating Scientific Disciplines* (Springer, Dordrecht, The Netherlands), pp 185–208.
17. Szathmáry E (2006) Path dependence and historical contingency in biology. *Understanding Change: Models, Methodologies, and Metaphors* (Palgrave Macmillan, New York), pp 140–157.
18. Pollock DD, Thiltgen G, Goldstein RA (2012) Amino acid coevolution induces an evolutionary Stokes shift. *Proc Natl Acad Sci USA* 109(21):E1352–E1359.
19. Smith JM (1970) Natural selection and the concept of a protein space. *Nature* 225(5232):563–564.
20. Whitlock MC, Phillips PC, Moore FBG, Tonsor SJ (1995) Multiple fitness peaks and epistasis. *Annu Rev Ecol Syst* 26:601–629.
21. Wolf JB, Brodie ED, Wade MJ (2000) *Epistasis and the Evolutionary Process* (Oxford Univ Press, Oxford).
22. Cordell HJ (2002) Epistasis: What it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum Mol Genet* 11(20):2463–2468.
23. Phillips PC (2008) Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nat Rev Genet* 9(11):855–867.
24. de Visser JAGM, Krug J (2014) Empirical fitness landscapes and the predictability of evolution. *Nat Rev Genet* 15(7):480–490.
25. Kondrashov AS, Sunyaev S, Kondrashov FA (2002) Dobzhansky-Muller incompatibilities in protein evolution. *Proc Natl Acad Sci USA* 99(23):14878–14883.
26. DePristo MA, Weinreich DM, Hartl DL (2005) Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat Rev Genet* 6(9):678–687.
27. Lunzer M, Golding GB, Dean AM (2010) Pervasive cryptic epistasis in molecular evolution. *PLoS Genet* 6(10):e1001162.
28. Burke MK, et al. (2010) Genome-wide analysis of a long-term evolution experiment with Drosophila. *Nature* 467(7315):587–590.
29. Kryazhimskiy S, Dushoff J, Bazykin GA, Plotkin JB (2011) Prevalence of epistasis in the evolution of influenza A surface proteins. *PLoS Genet* 7(2):e1001301.
30. Khan AI, Dinh DM, Schneider D, Lenski RE, Cooper TF (2011) Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* 332(6034): 1193–1196.
31. Chou HH, Chiu HC, Delaney NF, Segrè D, Marx CJ (2011) Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* 332(6034):1190–1192.
32. Breen MS, Kemena C, Vlasov PK, Notredame C, Kondrashov FA (2012) Epistasis as the primary factor in molecular evolution. *Nature* 490(7421):535–538.

33. Blount ZD, Barrick JE, Davidson CJ, Lenski RE (2012) Genomic analysis of a key innovation in an experimental Escherichia coli population. *Nature* 489(7417):513–518.
34. Wiser MJ, Ribeck N, Lenski RE (2013) Long-term dynamics of adaptation in asexual populations. *Science* 342(6164):1364–1367.
35. Losos JB, Jackman TR, Larson A, Queiroz K, Rodriguez-Schettino L (1998) Contingency and determinism in replicated adaptive radiations of island lizards. *Science* 279(5359):2115–2118.
36. Salverda MLM, et al. (2011) Initial mutations direct alternative pathways of protein evolution. *PLoS Genet* 7(3):e1001321.
37. Meyer JR, et al. (2012) Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science* 335(6067):428–432.
38. Dickinson BC, Leconte AM, Allen B, Esvelt KM, Liu DR (2013) Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution. *Proc Natl Acad Sci USA* 110(22): 9007–9012.
39. Naumenko SA, Kondrashov AS, Bazykin GA (2012) Fitness conferred by replaced amino acids declines with time. *Biol Lett* 8(5):825–828.
40. Soylemez O, Kondrashov FA (2012) Estimating the rate of irreversibility in protein evolution. *Genome Biol Evol* 4(12):1213–1222.
41. Bershtein S, Segal M, Bekerman R, Tokuriki N, Tawfik DS (2006) Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* 444(7121): 929–932.
42. Wang ZO, Pollock DD (2007) Coevolutionary patterns in cytochrome c oxidase subunit I depend on structural and functional context. *J Mol Evol* 65(5):485–495.
43. Povolotskaya IS, Kondrashov FA (2010) Sequence space and the ongoing expansion of the protein universe. *Nature* 465(7300):922–926.
44. Xu J, Zhang J (2014) Why human disease-associated residues appear as the wild-type in other species: Genome-scale structural evidence for the compensation hypothesis. *Mol Biol Evol* 31(7):1787–1792.
45. Pollock DD, Taylor WR, Goldman N (1999) Coevolving protein residues: Maximum likelihood identification and relationship to structure. *J Mol Biol* 287(1):187–198.
46. Robinson DM, Jones DT, Kishino H, Goldman N, Thorne JL (2003) Protein evolution with dependence among codons due to tertiary structure. *Mol Biol Evol* 20(10): 1692–1704.
47. Rodrigue N, Philippe H, Lartillot N (2006) Assessing site-interdependent phylogenetic models of sequence evolution. *Mol Biol Evol* 23(9):1762–1775.
48. Choi SC, Hobolth A, Robinson DM, Kishino H, Thorne JL (2007) Quantifying the impact of protein tertiary structure on molecular evolution. *Mol Biol Evol* 24(8):1769–1782.
49. Arenas M, Dos Santos HG, Posada D, Bastolla U (2013) Protein evolution along phylogenetic histories under structurally constrained substitution models. *Bioinformatics* 29(23):3020–3028.
50. Usmanova DR, Ferretti L, Povolotskaya IS, Vlasov PK, Kondrashov FA (2015) A model of substitution trajectories in sequence space and long-term protein evolution. *Mol Biol Evol* 32(2):542–554.
51. Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol* 11(5):725–736.
52. Kosiol C, Holmes I, Goldman N (2007) An empirical codon model for protein sequence evolution. *Mol Biol Evol* 24(7):1464–1479.
53. Yang Z, 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.
54. Rodrigue N, Philippe H, Lartillot N (2010) Mutation-selection models of coding sequence evolution with site-heterogeneous amino acid fitness profiles. *Proc Natl Acad Sci USA* 107(10):4629–4634.
55. 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.
56. Bloom JD (2014) An experimentally determined evolutionary model dramatically improves phylogenetic fit. *Mol Biol Evol* 31(8):1956–1978.
57. 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.
58. Wells JA (1990) Additivity of mutational effects in proteins. *Biochemistry* 29(37): 8509–8517.
59. Serrano L, Day AG, Fersht AR (1993) Step-wise mutation of barnase to binase. A procedure for engineering increased stability of proteins and an experimental analysis of the evolution of protein stability. *J Mol Biol* 233(2):305–312.
60. Wylie CS, Shakhnovich EI (2011) A biophysical protein folding model accounts for most mutational fitness effects in viruses. *Proc Natl Acad Sci USA* 108(24):9916–9921.
61. Ashenberg O, Gong LI, Bloom JD (2013) Mutational effects on stability are largely conserved during protein evolution. *Proc Natl Acad Sci USA* 110(52):21071–21076.
62. Kimura M (1968) Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genet Res* 11(3):247–269.
63. McCandlish DM, Stoltzfus A (2014) Modeling evolution using the probability of fixation: History and implications. *Q Rev Biol* 89(3):225–252.
64. 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.
65. Kellogg EH, Leaver-Fay A, Baker D (2011) Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins* 79(3): 830–838.
66. Somero GN (1995) Proteins and temperature. *Annu Rev Physiol* 57:43–68.
67. Shoichet BK, Baase WA, Kuroki R, Matthews BW (1995) A relationship between protein stability and protein function. *Proc Natl Acad Sci USA* 92(2):452–456.
68. Teilum K, Olsen JG, Kragelund BB (2011) Protein stability, flexibility and function. *Biochim Biophys Acta* 1814(8):969–976.
69. Howell SC, Inampudi KK, Bean DP, Wilson CJ (2014) Understanding thermal adaptation of enzymes through the multistate rational design and stability prediction of 100 adenylate kinases. *Structure* 22(2):218–229.
70. Sanjuán R, Moya A, Elena SF (2004) The distribution of fitness effects caused by single-nucleotide substitutions in an RNA virus. *Proc Natl Acad Sci USA* 101(22): 8396–8401.
71. Eyre-Walker A, Keightley PD (2007) The distribution of fitness effects of new mutations. *Nat Rev Genet* 8(8):610–618.
72. Jacquier H, et al. (2013) Capturing the mutational landscape of the beta-lactamase TEM-1. *Proc Natl Acad Sci USA* 110(32):13067–13072.
73. Bank C, Hietpas RT, Wong A, Bolon DN, Jensen JD (2014) A bayesian MCMC approach to assess the complete distribution of fitness effects of new mutations: Uncovering the potential for adaptive walks in challenging environments. *Genetics* 196(3):841–852.
74. Ewens WJ (2004) *Mathematical Population Genetics: I. Theoretical Introduction* (Springer, New York).
75. Arnold FH, Wintrode PL, Miyazaki K, Gershenson A (2001) How enzymes adapt: Lessons from directed evolution. *Trends Biochem Sci* 26(2):100–106.
76. Taverna DM, Goldstein RA (2002) Why are proteins marginally stable? *Proteins* 46(1): 105–109.
77. Bloom JD, Labthavikul ST, Otey CR, Arnold FH (2006) Protein stability promotes evolvability. *Proc Natl Acad Sci USA* 103(15):5869–5874.
78. Tokuriki N, Stricher F, Schymkowitz J, Serrano L, Tawfik DS (2007) The stability effects of protein mutations appear to be universally distributed. *J Mol Biol* 369(5): 1318–1332.
79. Bloom JD, Raval A, Wilke CO (2007) Thermodynamics of neutral protein evolution. *Genetics* 175(1):255–266.
80. Zeldovich KB, Chen P, Shakhnovich EI (2007) Protein stability imposes limits on organism complexity and speed of molecular evolution. *Proc Natl Acad Sci USA* 104(41):16152–16157.
81. Koshi JM, Goldstein RA (1995) Context-dependent optimal substitution matrices. *Protein Eng* 8(7):641–645.
82. Mirny LA, Shakhnovich EI (1999) Universally conserved positions in protein folds: Reading evolutionary signals about stability, folding kinetics and function. *J Mol Biol* 291(1):177–196.
83. Bloom JD, Drummond DA, Arnold FH, Wilke CO (2006) Structural determinants of the rate of protein evolution in yeast. *Mol Biol Evol* 23(9):1751–1761.
84. Risso VA, 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.
85. Edward JD, Mishra SN (1988) *Modern Mathematical Statistcs* (Wiley, New York).
86. Draghi JA, Plotkin JB (2013) Selection biases the prevalence and type of epistasis along adaptive trajectories. *Evolution* 67(11):3120–3131.
87. Greene D, Crona K (2014) The changing geometry of a fitness landscape along an adaptive walk. *PLOS Comput Biol* 10(5):e1003520.
88. Chen P, Shakhnovich EI (2009) Lethal mutagenesis in viruses and bacteria. *Genetics* 183(2):639–650.
89. Pollock DD, Goldstein RA (2014) Strong evidence for protein epistasis, weak evidence against it. *Proc Natl Acad Sci USA* 111(15):E1450.
90. Goldstein RA, Pollard ST, Shah SD, Pollock DD (2015) Nonadaptive amino acid convergence rates decrease over time. *Mol Biol Evol*, 10.1093/molbev/msv041.
91. McCandlish DM, Rajon E, Shah P, Ding Y, Plotkin JB (2013) The role of epistasis in protein evolution. *Nature* 497(7451):E1–E2, discussion E2–E3.
92. Rodrigue N, Philippe H (2010) Mechanistic revisions of phenomenological modeling strategies in molecular evolution. *Trends Genet* 26(6):248–252.
93. Thorne JL, Lartillot N, Rodrigue N, Choi SC (2012) Codon models as a vehicle for reconciling population genetics with inter-specific sequence data. *Codon Evolution: Mechanisms and Models*, eds Cannarozzi GM, Schneider A (Oxford Univ Press, Oxford), pp 97–110.
94. Wilke CO (2012) Bringing molecules back into molecular evolution. *PLOS Comput Biol* 8(6):e1002572.
95. Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A (2012) Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Mol Biol Evol* 29(8):1917–1932.
96. De Maio N, Schlötterer C, Kosiol C (2013) Linking great apes genome evolution across time scales using polymorphism-aware phylogenetic models. *Mol Biol Evol* 30(10):2249–2262.
97. Nasrallah CA, Huelsenbeck JP (2013) A phylogenetic model for the detection of epistatic interactions. *Mol Biol Evol* 30(9):2197–2208.
98. Kimura M (1985) The role of compensatory neutral mutations in molecular evolution. *J Genet* 64:7–19.
99. Carter AJR, Wagner GP (2002) Evolution of functionally conserved enhancers can be accelerated in large populations: A population-genetic model. *Proc Biol Sci* 269(1494):953–960.
100. Iwasa Y, Michor F, Nowak MA (2004) Stochastic tunnels in evolutionary dynamics. *Genetics* 166(3):1571–1579.
101. Grahnen JA, Nandakumar P, Kubelka J, Liberles DA (2011) Biophysical and structural considerations for protein sequence evolution. *BMC Evol Biol* 11:361.
102. Liberles DA, et al. (2012) The interface of protein structure, protein biophysics, and molecular evolution. *Protein Sci* 21(6):769–785.

Contingency and Entrenchment in Protein Evolution Under Purifying Selection

Premal Shah, David M. McCandlish, and Joshua B. Plotkin

University of Pennsylvania, Philadelphia PA 19104, USA

Supplementary Appendix

Sources of epistasis

The high degree of epistasis for fitness observed in our simulations under purifying selection could result from two alternative sources: epistasis in the computationally predicted protein stabilities themselves, or epistasis in the non-linear mapping from stability to fitness, or both. If the combined effect of multiple substitutions on predicted stabilities ($\Delta\Delta G$) does not equal the sum of their individual effects, then this form of epistasis for stability would induce epistasis in the fitness effects of substitutions. Alternatively, even in the absence of epistasis for protein stabilities, epistasis in fitness may arise from the non-linear (Gaussian or semi-Gaussian) mapping between stability and fitness. To resolve which of these two effects dominates we undertook additional analyses, applied to our simulations under each of the two fitness functions (Gaussian and semi-Gaussian).

First, we quantified the degree of epistasis in protein stabilities themselves using the following metric, directly analogous to Eqn. 5.

$$S_{(i,j)} = \begin{cases} \Big[D_{0,1,\ldots,j-1,i} - D_{0,1,\ldots,j-1}\Big] - \\ \qquad \Big[D_{0,1,\ldots,i} - D_{0,1,\ldots,i-1}\Big], & \text{for } i \geq j \\ \Big[D_{0,1,\ldots,j} - D_{0,1,\ldots,i-1,i+1,\ldots j}\Big] - \\ \qquad \Big[D_{0,1,\ldots,i} - D_{0,1,\ldots,i-1}\Big], & \text{for } i < j, \end{cases} \tag{S1}$$

where D_x denotes the protein stability (ΔG) of genotype x.

Likewise, to quantify epistasis arising solely from the stability-to-fitness mapping, we developed a measure that isolates this effect by assuming no epistasis in protein stabilities ($S_{(i,j)} = 0$). This is equivalent to assuming that the effect on protein stability of substituting a focal substitution i in an earlier background j ($i \geq j$) is the same as when it just fixed. Likewise, we assume the effect on protein stability of reverting a focal substitution i from a later background j ($i < j$) is exactly opposite to its effect at the time when it fixed. To implement these assumptions we set the "artificial" ΔG of genotypes not observed directly along evolutionary trajectories under purifying

selection ($j-1, i$ and $i-1, i+1, \ldots j$) as follows:

$$D'_{0,1,\ldots,j-1,i} = D_{0,1,\ldots,j-1} + \Big[D_{0,1,\ldots,i} - D_{0,1,\ldots,i-1}\Big] \text{ for } i \geq j \tag{S2}$$

$$D'_{0,1,\ldots,i-1,i+1,\ldots j} = D_{0,1,\ldots,j} - \Big[D_{0,1,\ldots,i} - D_{0,1,\ldots,i-1}\Big] \text{ for } i < j \tag{S3}$$

We then quantify epistasis due solely to the mapping of stability $\rightarrow$ fitness by modifying Eqn. 5 as follows

$$M_{(i,j)} = \begin{cases} \Big[\log(f'_{0,1,\ldots,j-1,i}) - \log(f_{0,1,\ldots,j-1})\Big] - \\ \qquad \Big[\log(f_{0,1,\ldots,i}) - \log(f_{0,1,\ldots,i-1})\Big], \quad \text{for } i \geq j \\ \Big[\log(f_{0,1,\ldots,j}) - \log(f'_{0,1,\ldots,i-1,i+1,\ldots j})\Big] - \\ \qquad \Big[\log(f_{0,1,\ldots,i}) - \log(f_{0,1,\ldots,i-1})\Big], \quad \text{for } i < j, \end{cases} \tag{S4}$$

where $f'_{0,1,\ldots,j-1,i}$ and $f'_{0,1,\ldots,i-1,i+1,\ldots j}$ are the fitnesses of genotypes $j-1, i$ and $i-1, i+1, \ldots j$, respectively, under the Gaussian or semi-Gaussian fitness function, based on their "artificial" protein stabilities, $D'_{0,1,\ldots,j-1,i}$ and $D'_{0,1,\ldots,i-1,i+1,\ldots j}$.

We find that the epistasis coefficient for stability itself ($S_{(i,j)}$) explains a large proportion of the variance in epistasis for fitness ($E_{(i,j)}$): $R^2 = 0.517$ for Gaussian fitness function and $R^2 = 0.331$ for the semi-Gaussian fitness function (SI Appendix, Fig. S6). By contrast, epistasis arising solely from the stability-to-fitness map ($M_{(i,j)}$) explains very little of the variance in epistasis for fitness: $R^2 = 0.018$ for Gaussian and $R^2 = 0.029$ for semi-Gaussian fitness functions (The R^2 reported above are based on separate regressions and need not sum to one). Thus, our results on epistasis for fitness are primarily driven by epistasis in the effects of mutations on protein stabilities, rather than non-linearity in the stability-fitness mapping.

Additivity of $\Delta\Delta$G under `FoldX`

To test whether the effects of random mutations on protein stability estimated by `FoldX` are additive, we generated 1000 pairs of random mutations in the wild-type sequence, as well as double mutants of each mutant pair. We found that the expected $\Delta\Delta$G of a double mutant assuming additivity (namely, the sum of $\Delta\Delta$G of the two individual mutations, G_A+G_B) is highly correlated with the actual $\Delta\Delta$G of the double mutant provided by `FoldX`, G_{AB} ($R^2 = 0.96$, Fig. 5A). We performed the same test for 1000 pairs of point mutations around an evolved *argT* sequence, 16 mutations away from the wild-type sequence, which was observed along one of our evolutionary trajectories under purifying selection. As in the case of the wild-type sequence, we generated 1000 pairs of random mutations and their double mutants in the evolved sequence. Once again we found that the effects of random mutations remain almost completely additive even far away from the wild-type sequence ($R^2 = 0.92$, Fig. 5B).

In addition to comparing the observed stability effects of double-mutants to the expected effects under additivity in a correlation analysis, we also estimate the goodness-of-fit of a non-epistatic model fit to the observed $\Delta\Delta$G of mutations by calculating the coefficient of determination (R^2) for each pair of mutations. For each pair of mutations A and B, `FoldX` estimates the $\Delta\Delta$G of

the single mutants (G_A, G_B), the double mutant (G_{AB}) and the wild-type (G_{WT}=0). The sum of squared residuals of a best-fit non-epistatic model can be calculated as

$$SS_{res} = (G_A - \widehat{G_A})^2 + (G_B - \widehat{G_B})^2 + (G_{AB} - \widehat{G_A} - \widehat{G_B})^2 \tag{S5}$$

where $\widehat{G_A}$ and $\widehat{G_B}$ are the estimated ΔG of single mutations that minimize SS_{res}. The coefficient of determination of each pair can then be calculated as

$$R^2 = 1 - \frac{SS_{res}}{SS_{total}} \tag{S6}$$

$$\text{where } SS_{total} = (G_{WT} - \mu)^2 + (G_A - \mu)^2 + (G_B - \mu)^2 + (G_{AB} - \mu)^2 \tag{S7}$$

$$\text{and } \mu = \frac{G_{WT} + G_A + G_B + G_{AB}}{4} \tag{S8}$$

We calculated the R^2 for each pair of random mutations around both the wild-type sequence, and for each pair of mutations around the evolved sequence. For mutations around the wild-type sequence we found that the non-epistatic model typically accounts for almost all the variation in $\Delta\Delta$G (median $R^2 = 0.999$). Similarly, for mutations around evolved sequence 16 steps away from wild-type, the non-epistatic model explains most of the variation in $\Delta\Delta$G (median $R^2 = 0.974$).

In summary, we find that the effects of random mutations on protein stability as estimated by `FoldX` are almost entirely additive, with very little residual epistasis. Nonetheless, we find much more epistasis between consecutive substitutions along our evolutionary trajectories. The correlation between the effects of single and double mutants for consecutive substitutions was much weaker than for random mutations, with an R^2 of 0.26 for the Gaussian fitness function (Fig. 5D) and 0.38 for the semi-Gaussian fitness function (SI Appendix Fig. S7). Similarly, the median R^2 when considering each pair of mutants individually (as described above) was 0.82 for the Gaussian fitness function and 0.86 for the semi-Gaussian fitness function. Because the degree of additivity for consecutive substitutions is weaker than the degree of additivity for random mutations, we conclude that purifying selection for folding stability enriches for epistasis in stability along our simulated evolutionary trajectories.

Biophysical properties of nearly-neutral substitutions

We analyzed the biophysical properties of the mutations that fixed along our simulated evolutionary trajectories under purifying selection. We analyzed the solvent accessible surface area (SASA) and Laguerre-Voronoi volume. Prior studies have found a higher substitution rate at sites with higher SASA and at residues occupying smaller volumes in the protein [1, 2, 3, 4].

We estimated SASA and volume of each residue in the WT sequence (PDB:1LAF) using the VLDP server `http://www.dsimb.inserm.fr/dsimb_tools/vldp/` [5]. We compared these quantities to the number of times an amino acid at a site was substituted across our ensemble of 100 replicate evolutionary trajectories. (Very seldom did the same site substitute more than once in any given trajectory.) Consistent with expectations, we found that sites with larger SASA in the wild-type PDB tended to have a larger number of substitutions, under both Gaussian (Pearson's correlation $\rho = 0.536$, p$< 10^{-15}$) and semi-Gaussian fitness functions (Pearson's correlation $\rho = 0.487$, p$< 10^{-14}$). Similarly, residues that occupy smaller volumes in the wild-type PDB tend to have a larger number of substitutions, under both Gaussian (Pearson's correlation $\rho = -0.216$, p$= 0.0008$) and semi-Gaussian fitness functions (Pearson's correlation $\rho = -0.204$, p$= 0.0015$).

Additivity of $\Delta\Delta$G and structural distance between residues

Even though the effects of most random mutations are nearly additive in the wild-type sequence, we can nevertheless ask whether any residual epistasis in protein stabilities of double mutants is related to the biophysical distance between residues in the folded three-dimensional structure. We estimated the distance between the C_α atoms of the two amino acid residues using the R-package `bio3D`. We found a weak but significant effect of distance between residues and the absolute degree of non-additivity in their stabilities $|\Delta\Delta G_{1,2} - (\Delta\Delta G_1 + \Delta\Delta G_2)|$ (Pearson's correlation, $p < 10^{-4}$). In particular, mutations that are physically closer to each other tend to be more epistatic than distant mutations. For instance, we find mutations that lie within 5 Å of each other tend to have a higher fraction of strongly epistatic mutations ($|\Delta\Delta G_{1,2} - (\Delta\Delta G_1 + \Delta\Delta G_2)| > 1$ kcal/mol), compared to mutations that lie more than 20 Å apart (Fisher's exact test, $p = 0.002$).

Larger sample of random mutations

In order to verify that our results are not influenced by the relatively small sample of mutations proposed at each step (10) we ran a set of shorter simulations (100 replicates, each for 20 substitutions), proposing in this case 100 point-mutations at each step under both Gaussian and semi-Gaussian fitness landscapes. We then compared the results of these simulations to the results of our original simulations when truncated to the same length.

In particular, with a 10-fold higher number of sampled mutations, we find the absolute effects of substitutions on protein stability ($|\Delta\Delta$G$| = 0.508$ kcal/mol under Gaussian and $|\Delta\Delta$G$| = 0.694$ kcal/mol under semi-Gaussian) to be comparable to the stability effects of substitutions under lower sample trajectories of the same length ($|\Delta\Delta$G$| = 0.534$ kcal/mol under Gaussian and $|\Delta\Delta$G$| = 0.689$ kcal/mol under semi-Gaussian). The fraction of highly epistatic consecutive substitutions—substitutions for which evolution is 30-times more likely to proceed via the observed path than the alternate path—also remains similar across both sampling regimes (high-sample Gaussian 13% and semi-Gaussian 13%; low-sample Gaussian 14% and semi-Gaussian 17%). Finally, the average contingency and entrenchment epistatic coefficients remain approximately unchanged when increasing the number of sampled mutations (see SI Appendix Table S1, Fig. S9).

Code and supplementary data

The scripts and supplementary files for estimating $\Delta\Delta$G of mutants using `FoldX` can be accessed at `http://mathbio.sas.upenn.edu/contingency_entrenchment.zip`. The package also contains the list of mutants that fixed along our evolutionary trajectories, and their stability estimates when they fixed as well as in earlier (contingency) and later (entrenchment) backgrounds along the trajectories. All of the stability datasets are provided as .RData objects.

References

[1] Koshi JM, Goldstein RA (1995) Context-dependent optimal substitution matrices. *Protein Engineering Design and Selection* 8:641–645.

[2] Mirny LA, Shakhnovich EI (1999) Universally conserved positions in protein folds: reading evolutionary signals about stability, folding kinetics and function. *J Mol Biol* 291:177–196.

[3] Robinson DM, Jones DT, Kishino H, Goldman N, Thorne JL (2003) Protein evolution with dependence among codons due to tertiary structure. *Mol. Biol. Evol.* 20:1692–1704.

[4] Bloom JD, Drummond DA, Arnold FH, Wilke CO (2006) Structural determinants of the rate of protein evolution in yeast. *Mol. Biol. Evol.* 23:1751–1761.

[5] Esque J, Léonard S, de Brevern AG, Oguey C (2013) VLDP web server: a powerful geometric tool for analysing protein structures in their environment. *Nucl Acids Res* 41:W373–8.

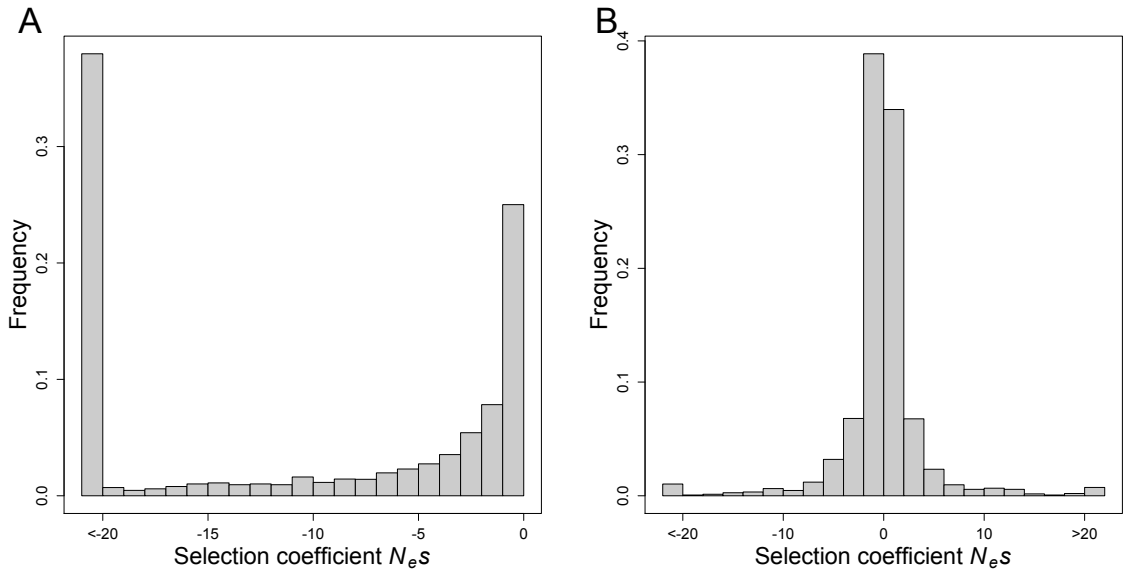

Figure S1: (**A**) The distribution of selection coefficients of all one-step mutations around the wild-type *argT* sequence, under a Gaussian fitness function. Roughly 25% of all one-step mutants are nearly-neutral, with $|N_e s| < 1$, while roughly 38% of mutants are strongly deleterious, with $N_e s < -20$. (**B**) Under purifying selection, most substitutions that accrue in our simulations are nearly-neutral. The histogram shows the scaled selection coefficients of all substitutions across 100 replicate evolutionary trajectories.

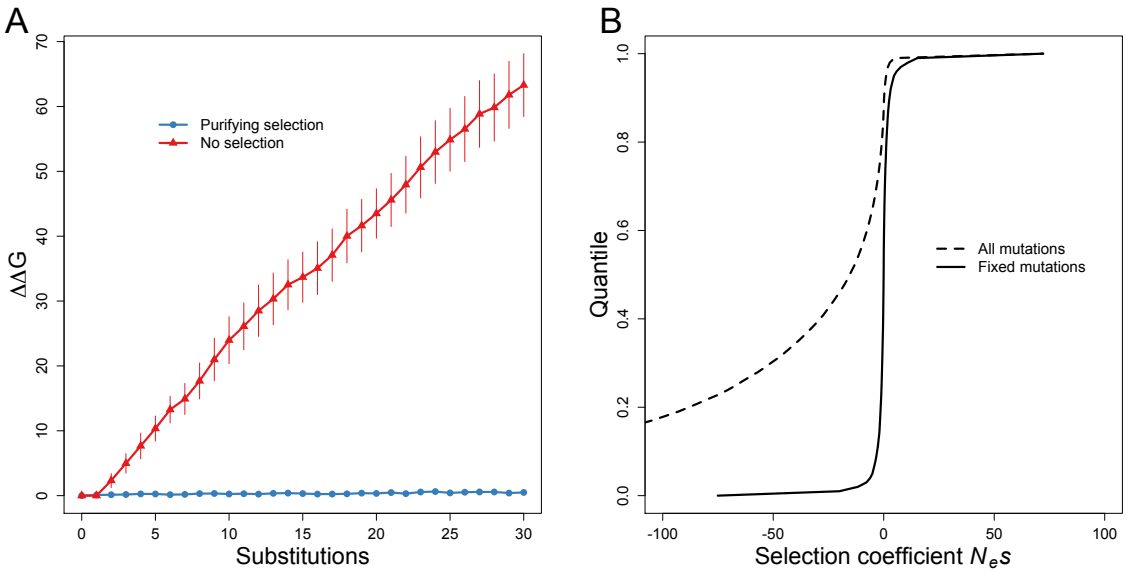

Figure S2: (**A**) Substitutions that accrue under simulated purifying selection in Gaussian fitness landscape do not change the overall stability of the protein; whereas the fixation of random point mutations, in the absence of selection, tends to decrease protein stability. Vertical bars indicate ±2 SE around the ensemble mean of 100 replicate simulated populations, each initiated at the wild-type *argT* sequence. (**B**) The distribution of selection coefficients for fixed (solid lines) and random (dashed lines) point mutations sampled along the evolutionary trajectories during purifying selection under a Gaussian fitness landscape. Most mutations that fix are nearly-neutral, whereas randomly sampled mutations tend to be strongly deleterious.

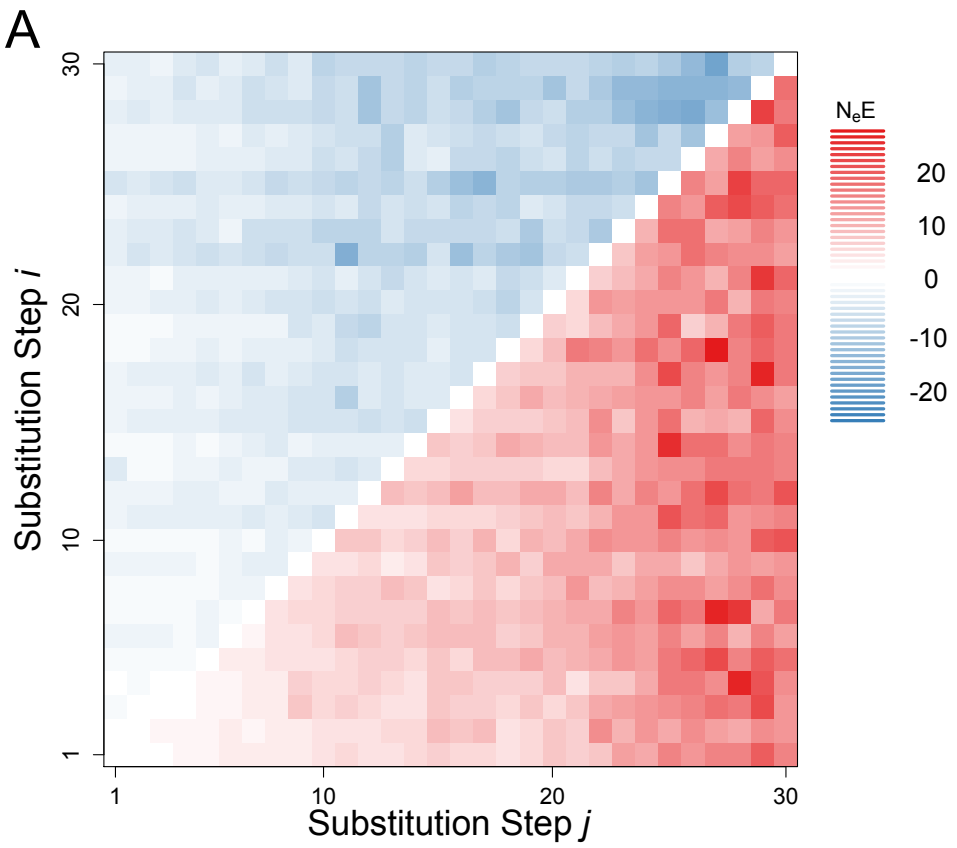

Figure S3: The mean epistatic effect ($N_eE_{(i,j)}$) between pairs of substitutions that fix at steps i and j in replicate simulations of $argT$ evolution under purifying selection (Gaussian fitness landscape). Each point in the grid represents the ensemble mean of $N_eE_{(i,j)}$ across 100 replicate simulations, each initiated at the wild-type $argT$ sequence. Contingency and entrenchment between substitutions are represented in shades of blue and red, respectively. The intensity of color reflects the magnitude of epistatic effects between pairs of substitutions. The mean $N_eE_{(i,j)}$ is negative for $i > j$ and positive for $i < j$ with the exception of three pairs where the mean epistatic coefficient did not differ significantly from zero (t-test, $p < 0.05$ for all but three pairs $i \neq j$).

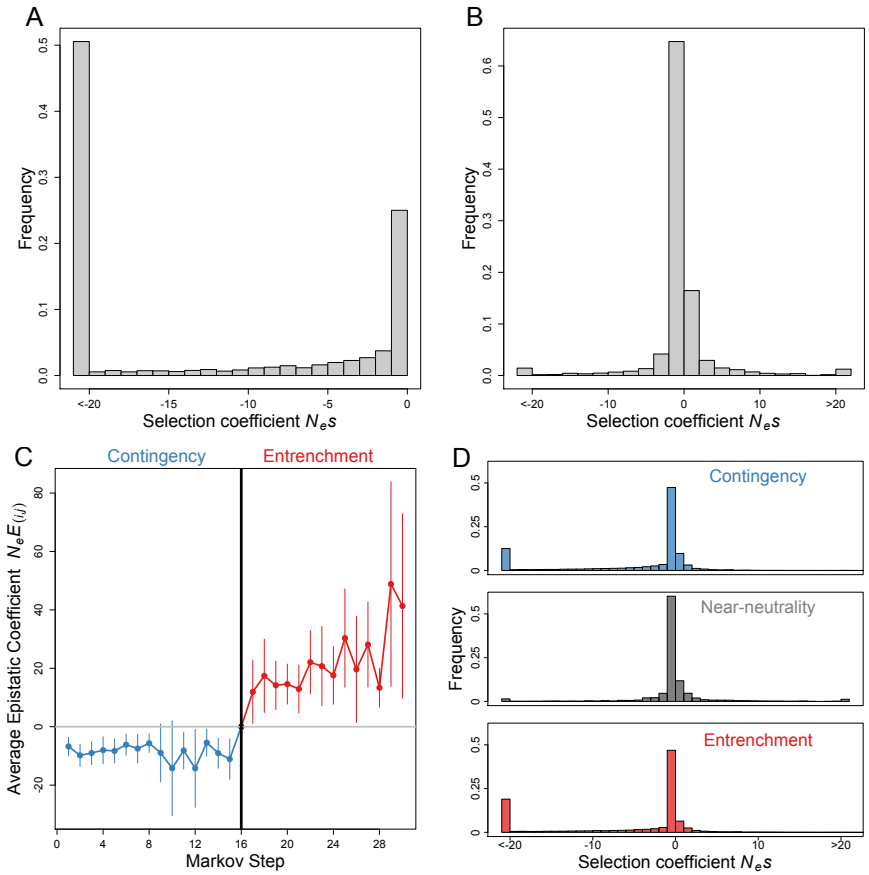

Figure S4: (**A**) The distribution of selection coefficients of all one-step mutations around the wild-type *argT* sequence under a semi-Gaussian fitness function. Roughly 25% of all one-step mutants are nearly-neutral with $|N_e s| < 1$, while roughly 50% of mutants are strongly deleterious with $N_e s < -20$. (**B**) Under purifying selection against destabilizing mutations (semi-Gaussian fitness function), most substitutions that accrue in our simulations are nearly neutral. The histogram shows the scaled selection coefficients of all substitutions among 100 replicates of simulated evolutionary trajectories. (**C**) Substitutions that accrue under purifying selection against destabilizing mutations (semi-Gaussian fitness function) are highly epistatic: they exhibit both contingency with earlier substitutions and entrenchment by later substitutions. The figure indicates the fitness effect of substitutions that fixed at step $i = 16$ in earlier (contingency $j < 16$) or later (entrenchment $j > 16$) genetic backgrounds across 100 independent evolutionary trajectories. Under purifying selection, the average epistatic coefficient $N_e E_{(16,j)}$ is significantly less than zero for all but two $j < 16$; and significantly greater than zero for all $j > 16$ (t-test, $p < 0.05$). Thus, substitutions that are nearly-neutral when they fix are typically contingent on earlier substitutions; and they become deleterious to revert as later substitutions accrue. Vertical bars indicate ± 2 SE around the ensemble mean of 100 replicate simulated populations. (**D**) The distribution of scaled selection coefficients ($N_e s$) for all substitutions that fix along evolutionary trajectories (semi-Gaussian fitness function). The gray histogram shows the distribution of selection coefficients of these mutations at the time that they fix ("near-neutrality"); the blue histogram shows the distribution of selection coefficients for the same mutations i if they were introduced in early backgrounds $j = 0, \ldots, i-1$ ("contingency"); and the red histogram shows the distribution of selection coefficients for the same mutations i if they are removed from later backgrounds $j = i+1, \ldots, 30$ ("entrenchment").

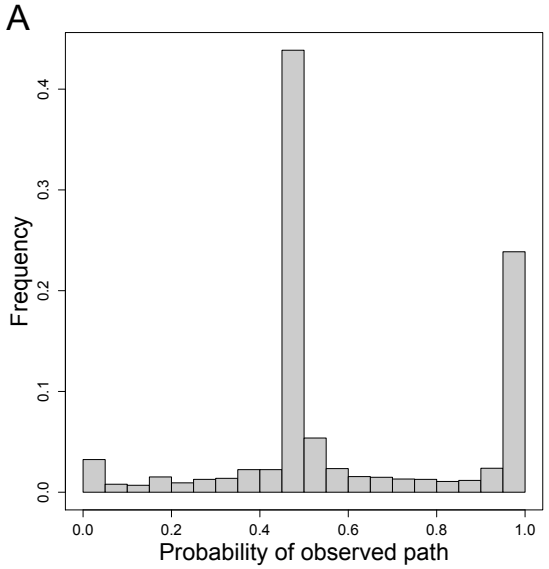

Figure S5: Purifying selection constrains paths available to evolution. The figure shows the probability of fixing two consecutive substitutions (B and C) in their observed order in simulated evolution ($A \to AB \to ABC$) compared to the reversed order ($A \to AC \to ACB$). Under purifying selection against destabilizing mutations (semi-Gaussian fitness function), the distribution of relative fixation probabilities is distinctly bimodal. A large proportion of substitutions have almost equal probability of taking either path, producing a mode near 0.5. For another large portion ($\sim 23\%$) of pairs the observed path is more than 30-times as likely as the alternate path (producing a mode near 1), indicating that many substitutions are highly contingent on the immediately preceding substitution.

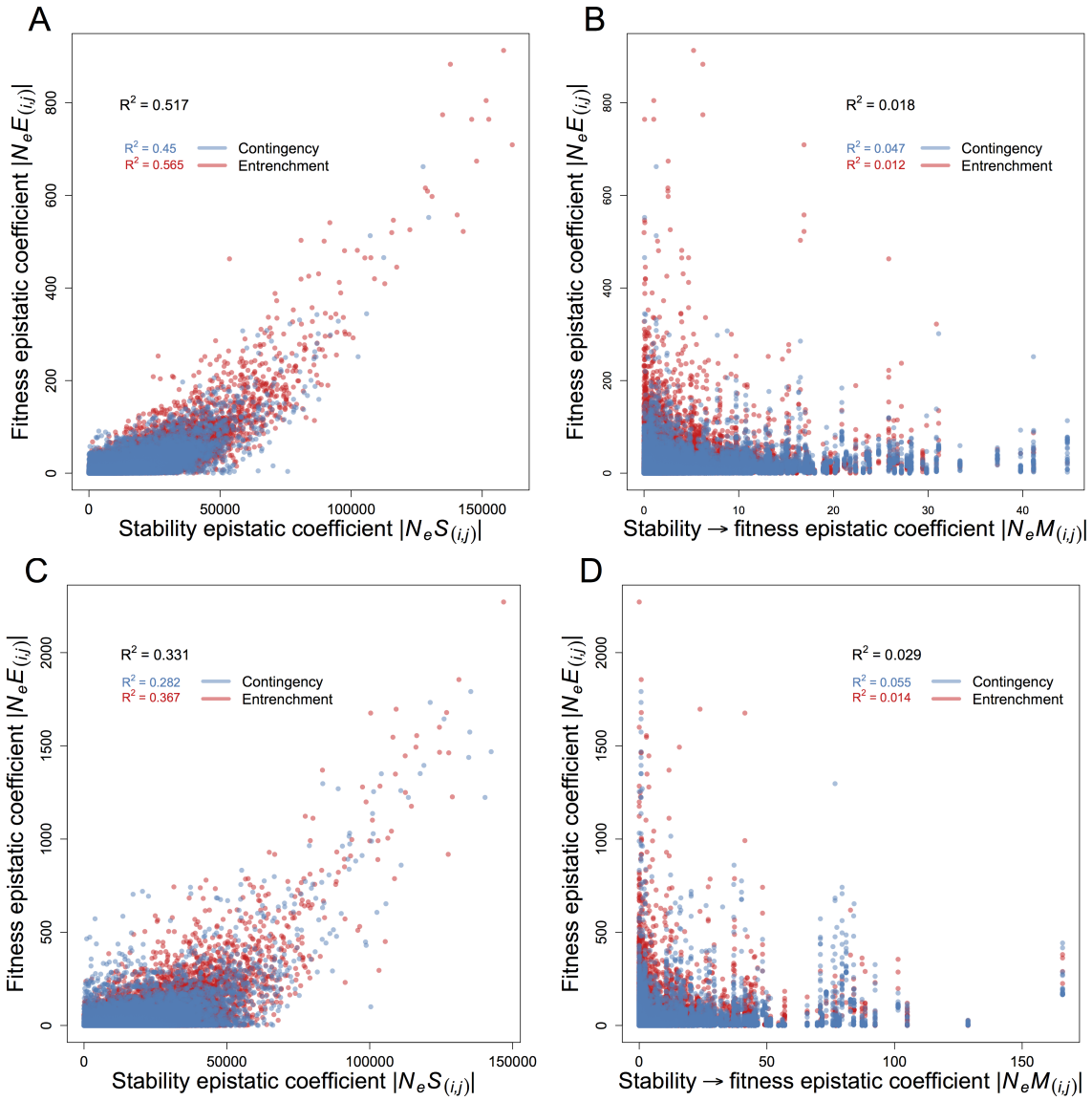

Figure S6: Sources of epistasis in fitness between substitutions that accrue under purifying selection, using either Gaussian (**A** & **B**) or semi-Gaussian (**C** & **D**) fitness functions. Epistasis in protein stabilities ($S_{(i,j)}$, Eqn. S1) explains a large portion of the variance in epistasis for fitness ($E_{(i,j)}$, Eqn. 4). By contrast, epistasis due to the non-linear mapping from stability-to-fitness ($M_{(i,j)}$, Eqn. S4) explains little of the variance in epistasis for fitness ($E_{(i,j)}$).

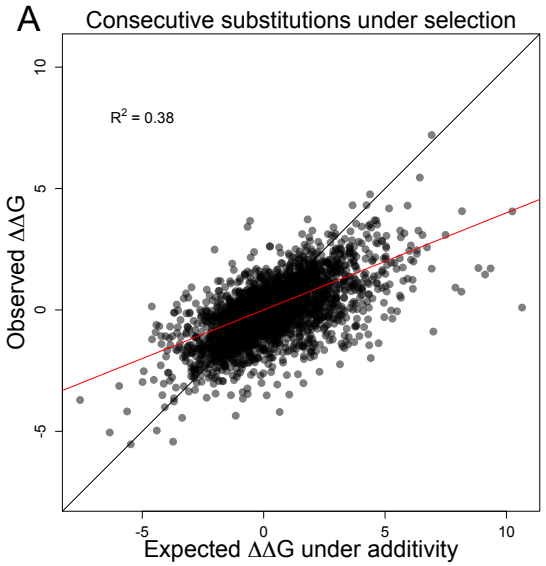

Figure S7: Additivity of $\Delta\Delta G$ during evolution under purifying selection. Consecutive substitutions along evolutionary trajectories under the semi-Gaussian landscape are only weakly additive ($R^2 = 0.38$). The line $y = x$ is represented in black and the best-fit regression line with zero intercept ($y = \beta x$) is represented in red.

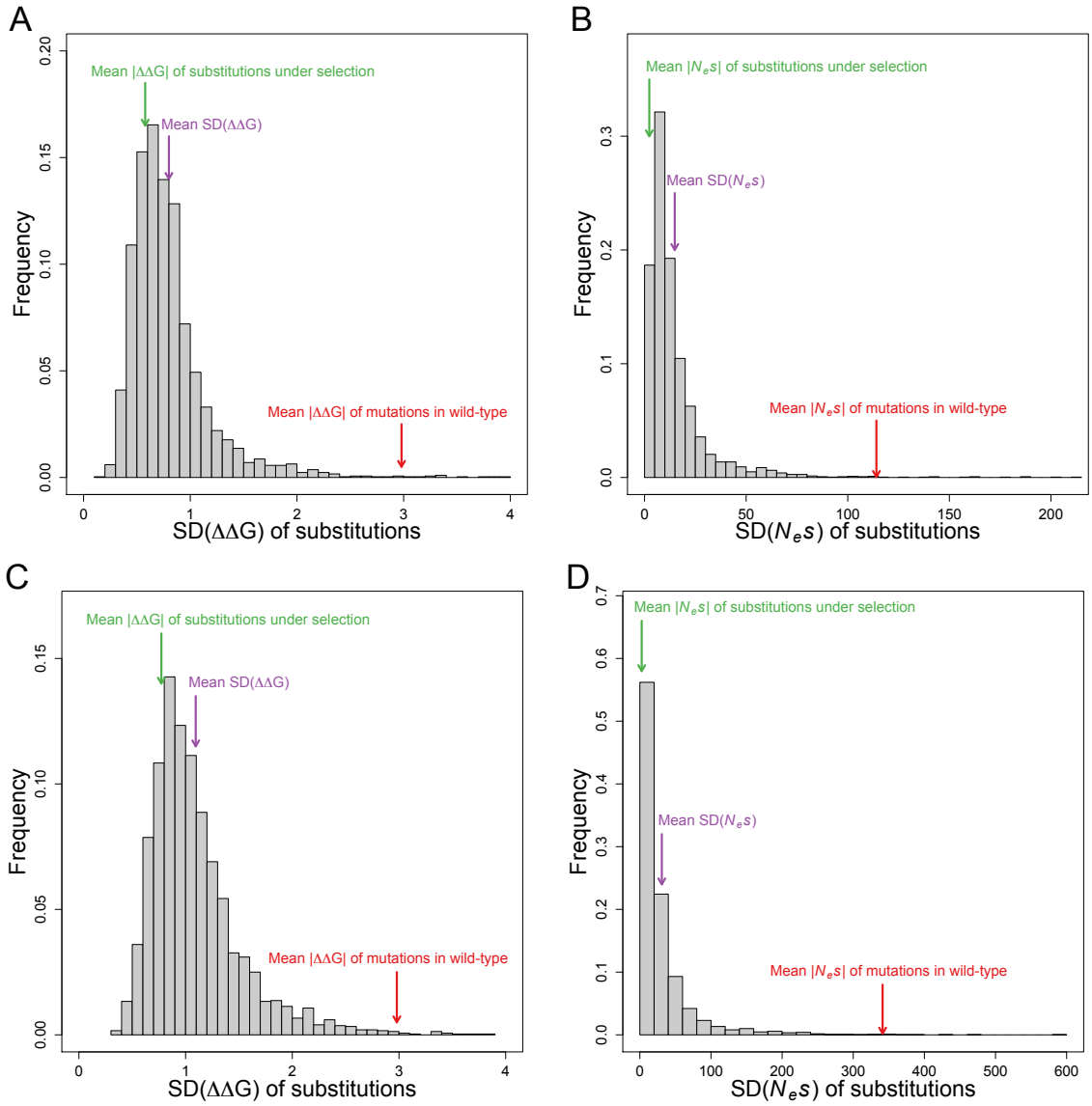

Figure S8: The effects on protein stability and fitness of mutations that fix along evolutionary trajectories simulated under Gaussian (**A** & **B**) or semi-Gaussian (**C** & **D**) fitness landscapes. The green arrows indicate the mean absolute effects on protein stability (left panels) and on fitness (right panels) of substitutions at the time of their fixation. The histograms represent the standard deviations of the effects of the same substitutions introduced in different backgrounds along their respective evolutionary trajectories. The red arrows, by contrast, indicate the mean absolute effects of random mutations introduced in the wild-type *argT* sequence. Under both fitness landscapes, substitutions have small absolute mean effects on protein stabilities at the time of fixation ($\overline{|\Delta\Delta G|}$ = 0.58 kcal/mol for Gaussian, and $\overline{|\Delta\Delta G|}$ = 0.77 kcal/mol for semi-Gaussian) relative to the mean absolute effect of random mutations ($\overline{|\Delta\Delta G|}$ = 2.98 kcal/mol). Moreover, the effects of these substitutions continue to remain fairly small and consistent across the 30 genetic backgrounds within their respective evolutionary trajectories. Similarly, the effects of substitutions on fitnesses ($N_e s$) both at the time of their fixation and in other genetic backgrounds are much smaller than the effects of random mutations (**B** & **D**).

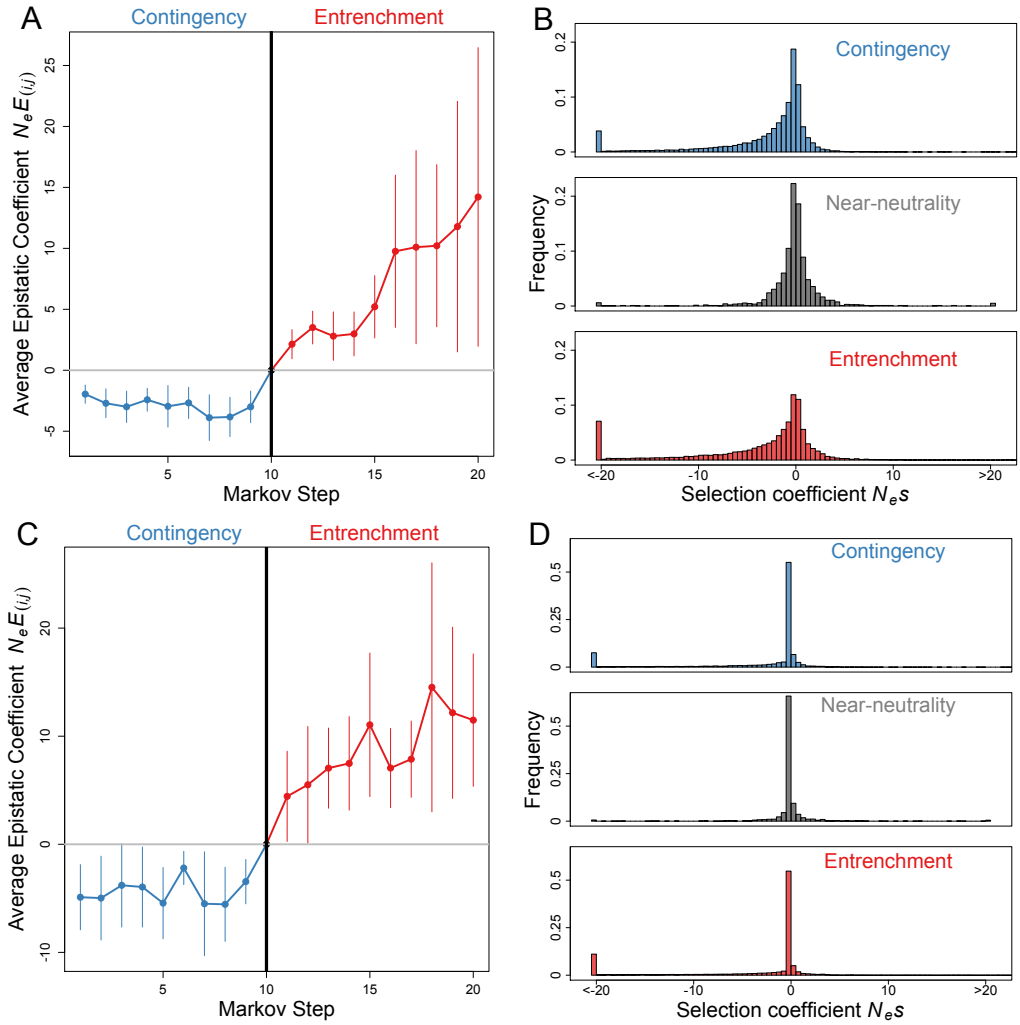

Figure S9: Substitutions that accrue along evolutionary trajectories under both Gaussian (**A & B**) and semi-Gaussian fitness landscapes (**C & D**) are highly epistatic even with a 10-fold higher rate of sampled mutations at each step. The figure indicates the fitness effect of substitutions that fixed at step $i = 10$ in earlier (contingency $j < 10$) or later (entrenchment $j > 10$) genetic backgrounds across 100 independent evolutionary trajectories. Under purifying selection, the average epistatic coefficient $N_eE_{(10,j)}$ is significantly less than zero for all $j < 10$ under Gaussian and all but one $j < 10$ under semi-Gaussian landscapes; and significantly greater than zero for all $j > 10$ (t-test, $p < 0.05$) under both fitness regimes. Thus, substitutions that are nearly-neutral when they fix are typically contingent on earlier substitutions; and they become deleterious to revert as later substitutions accrue. Vertical bars indicate ± 2 SE around the ensemble mean of 100 replicate simulated populations. (**B & D**) The distribution of scaled selection coefficients (N_es) for all substitutions that fix along evolutionary trajectories. The gray histogram shows the distribution of selection coefficients of these mutations at the time that they fix ("near-neutrality"); the blue histogram shows the distribution of selection coefficients for the same mutations i if they were introduced in early backgrounds $j = 0, \ldots, i-1$ ("contingency"); and the red histogram shows the distribution of selection coefficients for the same mutations i if they are removed from later backgrounds $j = i+1, \ldots, 20$ ("entrenchment").

	Gaussian fitness landscape		semi-Gaussian fitness landscape	
100 evolutionary trajectories of 20 substitutions	**Low sample (10 mutations)**	**High sample (100 mutations)**	**Low sample (10 mutations)**	**High sample (100 mutations)**
Stability effect of substitutions at their time of fixation (mean $\|\Delta\Delta G\|$).	0.534	0.508	0.689	0.694
Stability effect of fixed substitutions in other backgrounds (mean $\|\Delta\Delta G\|$).	0.825	0.808	0.922	0.895
Fraction of substitutions contingent with earlier substitutions ($E_{(i,j)}<0$).	0.68	0.69	0.52	0.49
Fraction of substitutions entrenched by later substitutions ($E_{(i,j)}>0$).	0.70	0.70	0.49	0.46
Contingency epistatic coefficient (mean N_eE).	-3.799	-3.527	-7.802	-5.427
Entrenchment epistatic coefficient (mean N_eE).	5.558	6.208	11.445	7.900
Epistatic coefficient between consecutive substitutions (mean $N_eE_{(i+1,j)}$).	-4.050	-4.282	-7.203	-4.649
Fraction of consecutive substitutions where path taken by evolution is 30 times more likely than the alternative.	0.21	0.19	0.20	0.16
Additivity of consecutive substitutions (R^2).	0.325	0.308	0.426	0.443

Table S1: Stability and epistatic effects of substitutions remain unchanged along evolutionary trajectories of the same length under both Gaussian and semi-Gaussian fitness landscapes when increasing the number of sampled mutations by 10-fold at each step.