Molecular ensembles make evolution unpredictable

Zachary R. Sailer[a,b] and Michael J. Harms[a,b,1]

[a]Institute of Molecular Biology, University of Oregon, Eugene, OR 97403-1253; and [b]Department of Chemistry and Biochemistry, University of Oregon, Eugene, OR 97403-1253

Edited by Jerrold Meinwald, Cornell University, Ithaca, NY, and approved September 29, 2017 (received for review July 4, 2017)

Evolutionary prediction is of deep practical and philosophical importance. Here we show, using a simple computational protein model, that protein evolution remains unpredictable, even if one knows the effects of all mutations in an ancestral protein background. We performed a virtual deep mutational scan—revealing the individual and pairwise epistatic effects of every mutation to our model protein—and then used this information to predict evolutionary trajectories. Our predictions were poor. This is a consequence of statistical thermodynamics. Proteins exist as ensembles of similar conformations. The effect of a mutation depends on the relative probabilities of conformations in the ensemble, which in turn, depend on the exact amino acid sequence of the protein. Accumulating substitutions alter the relative probabilities of conformations, thereby changing the effects of future mutations. This manifests itself as subtle but pervasive high-order epistasis. Uncertainty in the effect of each mutation accumulates and undermines prediction. Because conformational ensembles are an inevitable feature of proteins, this is likely universal.

protein evolution | epistasis | ensemble | predictability | thermodynamics

Is evolution predictable? This is a fundamental question in evolutionary biology both for philosophical (1–4) and for practical reasons (5, 6). Deep mutational scanning experiments provide an intriguing new avenue to think about evolutionary prediction. These experiments reveal the effects of huge numbers of mutations and thus, provide rich information about local adaptive landscapes (7–9). This leads to a simple question: if we know the effect of every mutation in an ancestral genotype, can we predict future evolutionary trajectories? If not, what limits our ability to make predictions?

To pose this question, we attempted to predict the evolution of a simple physical protein model given a virtual deep mutational scan. In this context, prediction is knowing which mutations would accumulate, in what order, given knowledge of the effects of the mutations in the ancestral background. We attempted an "easy" prediction, reasoning that it could act as a starting point for more difficult scenarios involving more complex evolutionary processes. To maximize predictive success, we studied adaptive trajectories in which the environment was stable, there was consistent directional selection, and mutations were fixed by selection rather than drift.

We studied the evolution of improved thermodynamic stability—a shared feature of all folded proteins and a target of natural selection in many contexts (10–12). This is a useful phenotype for a number of reasons. First, because stability is a thermodynamic quantity, studying it may reveal features common to the evolution of other thermodynamic properties, such as allostery and ligand binding. Second, biological systems—from molecules to ecosystems—are ultimately physical; therefore, insights at the physical level may provide insights for higher levels of biological organization (13).

Surprisingly, we found that our predictions were quite poor. We even added all pairwise epistatic effects of mutations to our predictive model, requiring a massive virtual deep mutational scan of all possible pairs of mutations. Even this did not allow robust predictions of evolutionary trajectories. We find that the unpredictability arises directly from the thermodynamic ensemble of conformations populated by macromolecules, revealing a profound link between protein physics and the evolutionary process.

Results

We set out to predict trajectories that increased the stability of a lattice protein. Lattice models have been used extensively in studies of protein folding and evolution (13–17). A lattice model captures the fact that the weak interactions that define the structure of a protein stochastically break and form under cellular conditions, causing proteins to fluctuate between multiple conformations (18–20). These structural ensembles are critical to functions, such as allostery (20, 21), enzyme activity (19), complex assembly (22), and regulation (23).

A lattice model describes a protein ensemble as a collection of conformations on a grid. Some conformations will be favored, and others will be disfavored. The favorability of each conformation is quantified by its internal energy (E_c), which depends on the contacts between amino acids in that conformation. Conformations with more favorable contacts are more likely than those with fewer contacts. The overall stability of the protein is described by the free energy of the native conformation (ΔG_N°), which quantifies the population of the native conformation relative to all other conformations in the ensemble (Fig. 1*A*). Using reduced temperature units, this is given by

$$\Delta G_N^\circ = E_N + \ln\left(-e^{-E_N} + \sum_{i=1}^{C} e^{-E_i}\right), \qquad [1]$$

Significance

A long-standing goal in evolutionary biology is predicting evolution. Here, we show that the architecture of macromolecules fundamentally limits evolutionary predictability. Under physiological conditions, macromolecules, like proteins, flip between multiple structures, forming an ensemble of structures. A mutation affects all of these structures in slightly different ways, redistributing the relative probabilities of structures in the ensemble. As a result, mutations that follow the first mutation have a different effect than they would if introduced before. This implies that knowing the effects of every mutation in an ancestor would be insufficient to predict evolutionary trajectories past the first few steps, leading to profound unpredictability in evolution. We, therefore, conclude that detailed evolutionary predictions are not possible given the chemistry of macromolecules.

Author contributions: Z.R.S. and M.J.H. designed research; Z.R.S. performed research; Z.R.S. analyzed data; and Z.R.S. and M.J.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This is an open access article distributed under the PNAS license.

Data deposition: All software reported in this paper is available on Github: https://github.com/harmslab/notebooks-epistasis-ensembles.

[1]To whom correspondence should be addressed. Email: harms@uoregon.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1711927114/-/DCSupplemental.

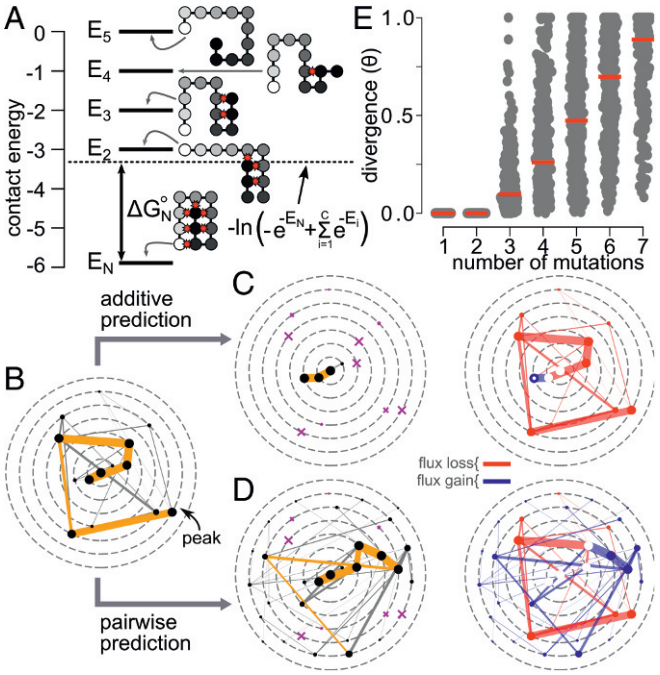

Fig. 1. Evolution is unpredictable in protein lattice models. *A* shows the meaning of ΔG_N° using five lattice model conformations of the many thousands possible. Each conformation is a single nonintersecting chain on a 2D grid. Amino acids are shown as circles colored by position in the sequence from black to white. Peptide bonds are dark bars. Noncovalent interactions are red stars. The strength of each noncovalent interaction depends on the identities of the interacting amino acids. The contact energy of each conformation is shown as a dark line; the Boltzmann-weighted average of the nonnative conformations is shown as a dashed line. (*B*) Relative probabilities of evolutionary trajectories starting from an ancestral genotype (center). Circles indicate genotypes; lines indicate mutational steps. The size of each circle and the width of each line indicate its probability. Dashed gray lines indicate increasing number of sequence differences from the ancestor. The orange trajectory is the highest probability trajectory. (*C*) Predicted trajectories using an additive predictive model. *C*, *Left* is colored the same as *B*. The genotypes visited in the actual trajectories but not the predicted trajectories are shown as purple ×, with sizes proportional to their probability. *C*, *Right* shows the difference between the predicted and actual trajectories. Red lines indicate trajectories missed in the prediction; blue lines indicate trajectories incorrectly added by the prediction. (*D*) Predicted trajectories using a pairwise epistatic predictive model; colors are the same as in *C*. (*E*) Divergence between predicted and actual trajectories for 1,000 starting genotypes as a function of number of mutations. Gray points are individual simulations. Red bars are the means for all genotypes after that number of steps.

where E_N is the internal energy of the native conformation, and the sum on the right goes over all C conformations in the ensemble.

We studied evolution in a strong selection, weak mutation regime (24). This assumes that the population size is large and that mutations fix sequentially—a reasonable assumption for a single gene for which recombination is rare. This also removes uncertainty caused by drift. We defined fitness as proportional to the fraction of the molecules in the native conformation, $w = 1/[1 + exp(\Delta G_N^\circ)]$, because the fraction of molecules folded—not the free energy—is under selection in most biological contexts (11).

We generated a random 12-amino acid protein sequence with $w \approx 0.7$ as our starting point (*Materials and Methods*). We calculated w for all genotypes differing by a single mutation relative to the ancestor and then determined the relative fixation probability for all point mutants. We then stepped out to all accessible genotypes and repeated the protocol. Any mutation with a nonzero fixation probability was considered accessible. Iterating this procedure generates a branching set of trajectories that improve the stability of the original native conformation. By comparing the relative fixation probabilities for each mutation along each trajectory, we can calculate the total probability flux through each possible trajectory (*Materials and Methods*).

We started by calculating ground truth evolutionary trajectories against which to compare our predictions (Fig. 1*B*). For the sequence in Fig. 1*B*, we found one main evolutionary trajectory (shown in orange in Fig. 1*B*) leading to a fitness peak six mutations away. There were also several lower probability trajectories accessible (shown in gray in Fig. 1*B*).

We next set out to predict these trajectories using information extracted from a virtual deep mutational scanning experiment. We calculated the change in ΔG_N° for all 228 possible point mutants to the ancestral genotype. Using this information, we could then predict ΔG_N° for any genotype as the free energy of the ancestor plus the sum of the effects of all mutations in the genotype. Finally, we could use these predicted ΔG_N° values to calculate probable evolutionary trajectories.

These predictions were quite poor (Fig. 1*C*). While the first move is correctly identified, the next move is incorrect. For the sequence shown in Fig. 1*C*, our predicted trajectories are limited to a peak directly adjacent to the ancestral genotype rather than the actual peak six mutations away.

This result is unsurprising: we did not include any epistasis in our predictions. Like real proteins, residues in a lattice model form direct contacts with each other. We would, therefore, expect to see pairwise epistasis—a difference in ΔG_N° when mutations are introduced together versus separately. We, therefore, reran our predictions of trajectories accounting for both the individual effects of mutations and pairwise epistasis between them. Practically, this involved another, larger, virtual deep mutational scanning experiment: we calculated ΔG_N° for all 228 possible single mutants and all 22,836 possible double mutants. By comparing the effects of each mutation together and in pairs, we could build a more sophisticated prediction model that accounts for pairwise interactions between mutations (*Materials and Methods*).

Addition of pairwise epistasis improved our predictions relative to the additive model, but we still performed quite poorly (Fig. 1*D*). Although the addition of pairwise epistasis allows the trajectories to escape the local region of the ancestral genotype, the predicted and actual trajectories diverge after the second step. Many genotypes are visited that were not seen in the actual trajectories, while many genotypes in the actual trajectories were missed (purple crosses in Fig. 1*D*). This includes the actual fitness peak.

To verify that this was a robust feature of lattice proteins, we then repeated our pairwise epistasis predictions for 1,000 different random starting sequences. To characterize the quality of our predictions, we calculated the difference in the probabilities of matched trajectories between our predicted and actual maps (θ). This metric ranges from 0.0 (no difference) to 1.0 (complete difference) (*Materials and Methods* and *SI Appendix*) (25). We then calculated θ as a function of the number of steps from the ancestral genotype for all 1,000 spaces (Fig. 1*E*). In all spaces, we correctly predict the first two moves. (This is because we built the prediction model using the fitness of these genotypes—hardly a difficult prediction.) Addition of the third mutation, however, causes immediate divergence between the predicted and actual trajectories. This divergence continues until, by the seventh step, the predicted and actual maps have an average divergence of 0.9.

Ensembles Induce Epistasis. Our predictions accounted for all pairwise epistasis but still failed. It follows that there must be three-way (or greater) epistatic interactions between mutations. This is surprising, as lattice models are built from pairwise

contacts alone. What is the source of this "high-order" epistasis? We know that these interactions must be indirect at a structural level, as the only direct interactions are pairwise. We, therefore, searched for mechanisms that would lead to indirect interactions between mutations.

Allostery, where binding at one site indirectly affects activity at a distant site, is a useful analog of this problem. One way that allostery can arise is through a conformational ensemble (20, 21). Binding at one site perturbs the relative populations of different structures in the conformational ensemble. This can, indirectly, change activity at another site. We hypothesized that a similar phenomenon was leading to evolutionary unpredictability.

Fig. 2 shows a highly simplified lattice model that illustrates indirect, "ensemble-induced" epistasis between a pair of mutations. (The logic can be extended to indirect multiway interactions between mutations, but visualization of the phenomenon is much easier for pairwise epistasis.) In this example, we have a 6-amino acid protein, where each site can be either hydrophobic (H) or polar (P). We will consider introducing two mutations that do not contact one another in the structure: H2P (Fig. 2, orange) and H4P (Fig. 2, purple). We start with the nonepistatic case (Fig. 2*A*). This ensemble has only two conformations: A and B. ΔG°_A is simply the difference in the contact energy between conformation A and conformation B (Fig. 2*A* and Table 1). H2P destabilizes conformation A, and H4P indirectly stabilizes conformation A by destabilizing conformation B. If we sum the effects of the mutations, we obtain the correct stability of the double mutant.

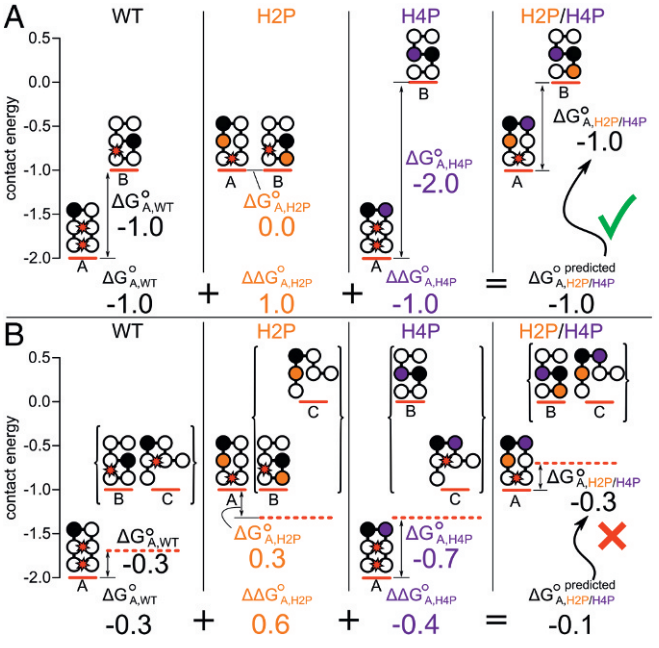

Fig. 2. Conformational ensembles induce epistasis. The panels show how epistasis arises from the conformational ensemble of a simple, six-residue lattice protein. In this model, residues can be either H (white circles) or P (filled circles). Favorable H-H contacts are worth −1 and are denoted by a red star. Colors denote mutations: H2P (orange) and H4P (purple). Solid red lines indicate the contact energies of each conformation. Genotypes are denoted above each subpanel. The thermodynamic stability of state *A* is shown for each genotype (for example, $\Delta G^\circ_{A,\,WT}$). The information used to calculate $\Delta G^{\circ\,predicted}_{A,\,H2P/H4P}$ is shown along the bottom of *A* and *B*. (*A*) In the two-state system, the effects of H2P and H4P sum in the H2P/H4P mutant; therefore, $\Delta G^{\circ\,predicted}_{A,\,H2P/H4P}$ is correct (green check mark). In *B*, we see that addition of a third state in the ensemble leads to epistasis. ΔG°_A for each genotype is now the difference in the contact energy of conformation *A* and the Boltzmann-weighted sum of the contact energies of conformations *B* and *C* (dashed red line). Because of this nonlinearity, $\Delta G^\circ_{A,\,WT} + \Delta\Delta G^\circ_{A,\,H2P} + \Delta\Delta G^\circ_{A,\,H4P} \neq \Delta G^\circ_{A,\,H2P/H4P}$ (red ×).

Table 1. Mathematical formulation of lattice phenotypes

Ensemble complexity	Phenotype
Two state	$E_N - E_U$
Three state	$E_N + \ln\left(e^{-E_U} + e^{-E_{U'}}\right)$
Full ensemble	$E_N + \ln\left(-e^{-E_N} + \sum_{i=1}^{C} e^{-E_i}\right)$

What if we add a third conformation (C) to the thermodynamic ensemble? This is shown in Fig. 2*B*. ΔG°_A is now the difference between the contact energy of A and the log of the Boltzmann-weighted sum of the contact energies of B and C (Fig. 2*B* and Table 1). The mutations now no longer behave additively. In the WT background, H2P is destabilizing by 0.6 (reduced energy units). In the H4P background, H2P is destabilizing by only 0.4. This is indirect, ensemble-induced epistasis. In the WT background, H2P destabilizes conformation A, such that it has the same contact energy as conformation B. These states strongly compete with one another, causing H2P to have a relatively large effect (0.6). H4P alters this effect by destabilizing conformation B. This means that, when H2P is introduced, conformation B does not compete effectively with conformation A. As a result, H2P has a more mild effect (0.4).

The presence of the third state in the ensemble leads to epistasis between these mutations and an incorrect prediction of the double-mutant stability. Predicting the effect of a mutation in a future genetic background, therefore, requires knowing its effect on every member of the ensemble, not simply its aggregate effect on the entire population of states. This is, in practice, impossible to measure. Ensemble-induced epistasis is directly analogous to ensemble-induced allostery (20, 21); however, we have now substituted mutations for binding events and binding sites.

Evolutionary Trajectories Exhibit Extensive Ensemble-Induced Epistasis. From our reasoning above, we would predict that ensemble-induced epistasis would arise for any conformational ensemble with more than two states and that it could lead to epistasis of any order. We, therefore, set out to quantify the epistasis present in our evolutionary trajectories. Quantitatively, epistasis accounts for variation not accounted for by lower ordered effects of mutations. Pairwise epistasis is the difference in the effects of two mutations introduced together versus separately. Three-way epistasis is the difference in ΔG°_N for a triple mutant versus the predicted ΔG°_N from the individual and pairwise epistatic coefficients. In thermodynamic terms, individual effects are $\Delta\Delta G^\circ_N$, pairwise interactions are $\Delta\Delta\Delta G^\circ_N$, and three-way interactions are $\Delta\Delta\Delta\Delta G^\circ_N$. This can be extended to any order of interaction (26–28).

Measuring an Lth-order interaction requires characterizing 2^L combinations of mutations. To access a collection of 2^L genotypes, we constructed binary genotype–phenotype maps containing all possible combinations of the mutations between the ancestral genotypes and their highest probability genotype six mutations away (e.g., between the ancestor and the peak in Fig. 1*B*). We decomposed epistasis in ΔG°_N using a linear model capturing the individual effects of mutations and any interactions between them (26–28) using the ancestral genotype as the reference state (27).

We detected high-order epistasis in every calculated trajectory. Fig. 3*A* shows the average magnitude of epistatic

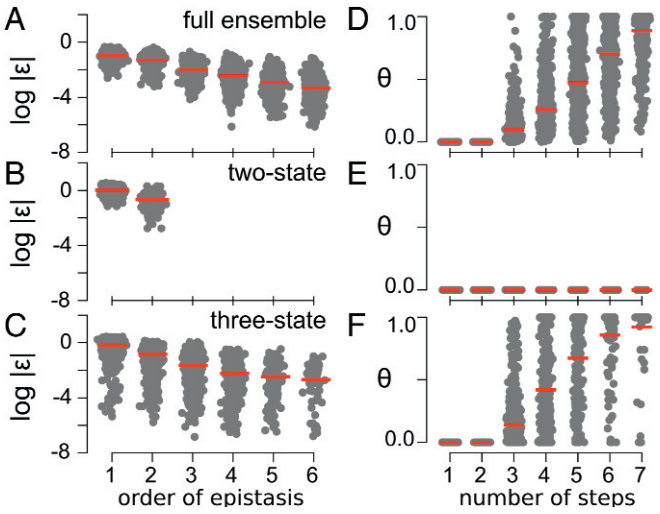

Fig. 3. Ensemble-induced epistasis leads to unpredictability. *A–C* are jitter plots that show the average magnitude of epistasis $|\varepsilon|$ observed for increasing orders of epistasis in 1,000 maps generated from full (*A*), two-state (*B*), and three-state (*C*) ensembles. Gray points represent the average magnitude of the epistatic coefficients at a given order for a single map. Red bars indicate the means. *D–F* show divergence between the "true" trajectories and predicted trajectories for full (*D*), two-state (*E*), and three-state maps (*F*). For clarity, *D* reproduces the data in Fig. 1*E*.

coefficients of increasing order for all spaces. The magnitude decays with increasing order but is still detectable up to the sixth order in all spaces.

If the ensemble is, indeed, the source of this epistasis, we predicted that it would disappear if we removed the ensemble. We, therefore, generated truncated lattice models that had only two or three conformations in their ensemble. The two-state ensemble had the native state and the lowest energy nonnative conformation (N and U). The three-state ensemble had the native state and the two lowest energy nonnative conformations (N, U, and U').

We predicted that the two-state model would exhibit only pairwise epistasis, while the three-state model would exhibit higher ordered epistasis. This can be understood from the energy function for each ensemble. ΔG_N° for the two-state model is linear with respect to contact energy (Table 1). The only epistasis that arises is via direct interactions encoded in the contact energy. In a three-state (or higher) ensemble, ΔG_N° no longer reduces to a linear difference in contact energies (Table 1). This means that mutations have nonlinear effects on the probabilities of conformations within the ensemble, leading to ensemble-induced epistasis (Fig. 2).

To investigate epistasis in these reduced ensembles, we generated binary maps as before: we used either a two-state or three-state ensemble to calculate ΔG_N° for each genotype, calculated evolutionary trajectories, and then built binary maps between the ancestor and most probable final genotype. We then decomposed these maps to extract epistasis. As predicted, the two-state ensembles exhibited only pairwise epistasis (Fig. 3*B*). In contrast, the three-state ensemble exhibited extensive high-order epistasis (Fig. 3*C*)—just like the full ensemble (Fig. 3*A*).

We next asked whether reducing the ensembles altered predictability. If the unpredictability that we initially observed arises from epistasis induced by the ensemble, we would predict high predictability for two-state ensemble but poor predictability for the three-state ensemble. We observed precisely this pattern. A pairwise prediction model was able to perfectly predict evolutionary trajectories for two-state ensemble (Fig. 3*E*) but failed to predict evolutionary trajectories for the three-state ensemble (Fig. 3*F*). The unpredictability observed for the three-state ensemble is directly comparable with the unpredictability observed for the full ensemble (Fig. 3*D*).

Predictions Fail Even When High-Order Epistasis Is Included. If molecular ensembles lead to epistasis, which undermines evolutionary prediction, an obvious solution is to characterize even higher orders of epistasis. What if we knew all three-way interactions? Or four-way interactions? Is there some order of epistasis that, when characterized, allows long-range evolutionary predictions?

We cannot ask this question for open-ended evolutionary trajectories, as it rapidly becomes intractable. (Even for our 12-site lattice protein, quantifying all possible three-way interactions would require characterizing 1,533,034 genotypes.) Instead, we chose to try to predict trajectories from ancestral to derived genotype through the binary maps above, each containing 2^6 genotypes. We built increasingly complex epistatic models ranging from first order (constructed from characterization of six point mutants in the ancestral background) to sixth order (constructed from characterization of all combinations of point mutants in the ancestral background).

We found that incorporation of high-order epistasis led to little improvement in our predictions. Fig. 4 shows the deviation between our predicted and actual trajectories for models incorporating increasingly higher orders of epistasis. Fig. 4*A* shows predictions using the full ensemble. The additive model begins to deviate from the actual trajectories after the first step, the pairwise model begins to deviate after the second step, and the three-way model begins to deviate after the third step. As soon as the model has to make a prediction beyond the phenotypes that were used to build the model, trajectories begin to deviate. Even our fifth-order model—which required knowing the phenotypes of 63 of 64 genotypes in the space–does not always correctly predict the final step (Fig. 4, purple curve).

This unpredictability arises from the thermodynamic ensemble. Fig. 4 *B* and *C* shows the same analysis for the two-state and three-state ensembles, respectively. For the two-state ensemble, we were able to predict trajectories perfectly with the addition of pairwise epistasis. For the three-state model, we see similar behavior to the full ensemble: inclusion of high-order epistasis does not improve predictions. This is because the epistasis does not capture specific interactions but instead, reveals that the ensemble is changing quantitatively and nonlinearly as mutations accumulate. No matter what order of epistasis is characterized, the future remains obscure.

Discussion

Our work shows that the physical properties of proteins can lead to profound evolutionary unpredictability. Because each

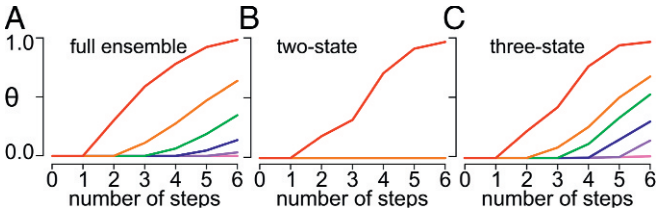

Fig. 4. Addition of high-order epistasis does not lead to predictability. The panels show the deviation between predicted and actual trajectories through binary genotype–phenotype maps using predictive models with increasing orders of epistasis: additive (red), pairwise (orange), three way (green), four way (blue), five way (purple), and six way (pink). The panels correspond to maps using a full ensemble (*A*), a two-state ensemble (*B*), and a three-state ensemble (*C*) maps. Each curve is averaged over 1,000 maps.

mutation alters the relative probabilities of all conformations of a protein, the quantitative effect of a mutation is different in every genetic background. As a result, the effect of a mutation early in a trajectory does not predict its effect later, and evolutionary trajectories become unpredictable. Because thermodynamic ensembles are a natural aspect of molecular architecture and ubiquitous for function, we expect that this is a universal link between the biochemistry of macromolecules and their evolution.

A key point from our work is that unpredictability can arise even in this extraordinary simple system. The problem of predicting evolution will only become harder as the complexity and realism of the models increase. Using a larger protein, for example, would increase the number of possible options and degeneracy of trajectories, making predictions more challenging. Likewise, constructing a more realistic evolutionary model—incorporating drift for example—increases the number of available trajectories and makes evolutionary prediction more challenging than the strong selection case (*SI Appendix*).

Ensemble-Induced Epistasis Is Likely Common. Our work suggests that any macromolecule that populates three or more conformations can exhibit ensemble-induced epistasis. This is an extraordinarily common set of conditions, as most macromolecular functions require populating multiple states (18–21). For example, consider an allosterically inhibited enzyme that takes two conformations: E (inactive) and E^* (active). An inhibitor I binds to E, shifting the population from E^* to E. The fraction of the enzyme in the active form given $[I]$ is

$$E \cdot I \rightleftarrows I + E \rightleftarrows E^*$$
$$f_{active} = \frac{[E^*]}{[E^*] + [E] + [E \cdot I]}.$$

This protein has three distinct states—E^*, E, and $E \cdot I$—that have different structures and would, thus, respond differentially to mutations. We would, therefore, expect ensemble-induced epistasis in f_{active}.

In addition to theoretical considerations, there is experimental evidence that ensemble-induced epistasis shapes evolution (29, 30). The most direct is an engineered evolutionary trajectory that converts a protein from one fold to another (13, 29). Midway through the trajectory, a single mutation switches the fold. If introduced earlier in the trajectory, the mutation does not have the same effect. The fold-switching mutation has its singular effect, because other mutations have prestabilized the alternate fold. This is ensemble-induced epistasis: mutations perturb the relative stability of a nonnative conformation, opening up a new evolutionary trajectory.

Another observation is the presence of high-order epistasis in every combinatorial protein genotype–phenotype studied (28, 31). The phenotypes studied are diverse, including binding affinity, spectroscopic properties, and enzyme activity. Although there is no direct evidence that the high-order epistasis in these maps arises from underlying ensembles, ensemble-induced epistasis provides a simple, universal explanation that unites these disparate observations of high-order epistasis.

One question that remains is how our observations in lattice models map quantitatively to real proteins. What is the magnitude of ensemble-induced epistasis in real systems? How many steps can be predicted before real evolutionary predictions diverge? This will depend on the details of the sequence and its associated ensemble. Our results suggest, however, that ensemble-induced epistasis will eventually lead to divergence between predicted and actual trajectories in any protein genotype–phenotype map.

It also remains to be seen if something analogous to ensemble-induced epistasis exists on a larger scale, such as in a signaling network. Such networks do exhibit ensemble-like behavior, populating a collection of different configurations that rearrange in response to stimuli and sometimes even exhibiting stable three-state character (32, 33) We might, therefore, be able to explain high-order epistasis in such systems using an ensemble framework (28, 31).

Interpreting Epistasis. Our analysis also sheds light on the question of the origin and interpretation of high-order epistasis. First, our work shows that it is relatively easy to create a system with irreducible high-order epistasis, even with a very simple lattice protein. There is no simple scale that reduces the epistasis (28, 34): it is an integral part of the system.

Second, our work shows that there is no mechanistic interpretation for epistatic coefficients that arise by such a process. The epistasis is fundamentally statistical rather than biological (34). The ensemble effectively encrypts the interactions that give rise to the epistasis. A three-way interaction cannot be interpreted in a direct physical manner or in a way to predict which conformations changed. It quantifies the effect of the mutation, integrated over its effect on all conformations in the ensemble. In our view, the best interpretation of epistasis in macroevolutionary trajectories is as a means to quantify uncertainty in future predictions—not necessarily as a way to gain mechanistic insight into the system.

Evolution Is Unpredictable. Epistasis makes a relatively small contribution to variation in our lattice models (Fig. 3*A*) as well as real datasets (28, 31). This allows prediction of phenotypes with relatively high accuracy, as has been noted before in lattice models (17). Approximate phenotypes are, however, insufficient for predicting trajectories. Because evolutionary trajectories are a contingent series of steps, small uncertainties in phenotypes are amplified into large uncertainties in trajectories (25). Practically, this means that you can predict a multimutation phenotype from a deep mutational scanning experiment but likely not its evolutionary accessibility from the ancestral state.

Many previous discussions of unpredictability have revolved around robustness of trajectories to external factors, such as environmental perturbation (2), genetic drift (4), or a change in the nature of selection (35). The unpredictability that we observe arises from the architecture of protein systems themselves. Our work indicates that the physical architecture of biomolecules naturally leads to ensemble-induced epistasis. Accumulating mutations thus alter the effects of future mutations, making evolution unpredictable given information about the effects of mutations in the ancestral state.

Materials and Methods

All of our analyses are contained in Python scripts and Jupyter notebooks available on Github (https://github.com/harmslab/notebooks-epistasis-ensembles). Full details are given in *SI Appendix*.

For the protein lattice model simulations, we extended the lattice proteins package originally written by Jesse Bloom (https://github.com/harmslab/latticeproteins) (17) using Miyazawa and Jernigan contact energies (from table V in ref. 36) and reduced temperature units (36). We randomly generated 12-site protein sequences and then evolved each sequence until the fraction folded was ≈ 0.7. We calculated the probability of a given evolutionary trajectory as a series of independent, sequential fixation events using a strong selection, weak mutation model (37). In this model, the fixation probability for going from genotype x to $x + 1$ is

$$\pi_{x \rightarrow x+1} = 1 - e^{-\left(\frac{w_{x+1}}{w_{x-1}}\right)},$$

where w_x and w_{x+1} are the relative fitness rates of the x and $x+1$ genotypes (*SI Appendix*).

To quantify the difference between the predicted and actual trajectories, we calculated the magnitude of the difference in probability of all observed trajectories through each space (*SI Appendix*) (25).

We predicted the ΔG°_{N} for each genotype and then used this to predict evolutionary trajectories. For the additive model, the predicted stability of a genotype with a set of $\{M\}$ mutations was

$$\Delta \hat{G}^{\circ}_{N,\{M\}} = \Delta G^{\circ}_{N,anc} + \sum_{i \in \{M\}} \Delta\Delta G^{\circ}_{N,i},$$

where $\Delta G^{\circ}_{N,anc}$ is the stability of the ancestor and $\Delta\Delta G^{\circ}_{N,i}$ is the effect of the ith mutation in the ancestral background. For the pairwise epistatic model, we added epistatic coefficients:

$$\Delta \hat{G}^{\circ}_{N,\{M\}} = \Delta G^{\circ}_{N,anc} + \sum_{i \in \{M\}} \Delta\Delta G^{\circ}_{N,i} + \sum_{i<j \in \{M\}} \Delta\Delta\Delta G^{\circ}_{N,ij},$$

where $\Delta\Delta\Delta G^{\circ}_{N,ij}$ accounts for any pairwise epistasis between the mutations i and j. We quantified this epistasis by introducing mutations i, j, and then, i and j together. We then took the difference:

$$\Delta\Delta\Delta G^{\circ}_{N,ij} = \Delta G^{\circ}_{N,ij} - \left(\Delta G^{\circ}_{N,anc} + \Delta\Delta G^{\circ}_{N,i} + \Delta\Delta G^{\circ}_{N,j}\right),$$

where $\Delta\Delta G^{\circ}_{N,i}$ and $\Delta\Delta G^{\circ}_{N,j}$ are the individual effects of mutations i and j, respectively, in the ancestral background and $\Delta G^{\circ}_{N,ij}$ is the stability of the ij double mutant.

To extract high-order epistasis, we used the epistasis package (https://github.com/harmslab/epistasis) (28). A genotype with L mutations is described by 2^L hierarchical epistatic coefficients (26–28, 31). We generated all 2^6 binary combinations of the substitutions that accumulated between the ancestor and most probable final sequence, calculating ΔG°_{N} for all 64 mutants. Because we were doing predictions starting from the ancestral state, we used the so-called "biochemical model" that uses the ancestral genotype as the reference state (27). *SI Appendix* has additional details.

ACKNOWLEDGMENTS. We thank Luke Wheeler, Tyler Starr, and Joe Thornton as well as members of the laboratory of M.J.H. for fruitful discussions when developing these ideas. This work was funded by funds from the University of Oregon (to Z.R.S.) and Alfred P. Sloan Fellowship FG-2015-65336 (to M.J.H.).

1. Monod J (1974) On chance and necessity. *Studies in the Philosophy of Biology*, eds Ayala FJ, Dobzhansky T (Macmillan Education, London), pp 357–375.
2. Gould SJ (1989) *Wonderful Life: The Burgess Shale and the Nature of Life* (Norton, New York).
3. Morris SC (2010) Evolution: Like any other science it is predictable. *Philos Trans R Soc Lond B Biol Sci* 365:133–145.
4. Harms MJ, Thornton JW (2014) Historical contingency and its biophysical basis in glucocorticoid receptor evolution. *Nature* 512:203–207.
5. Miton CM, Tokuriki N (2016) How mutational epistasis impairs predictability in protein evolution and design. *Protein Sci* 25:1260–1272.
6. Lässig M, Mustonen V, Walczak AM (2017) Predicting evolution. *Nat Ecol Evol* 1:0077.
7. Fowler DM, et al. (2010) High-resolution mapping of protein sequence-function relationships. *Nat Methods* 7:741–746.
8. Hietpas RT, Jensen JD, Bolon DNA (2011) Experimental illumination of a fitness landscape. *Proc Natl Acad Sci USA* 108:7896–7901.
9. Sarkisyan KS, et al. (2016) Local fitness landscape of the green fluorescent protein. *Nature* 533:397–401.
10. Gromiha MM, Oobatake M, Sarai A (1999) Important amino acid properties for enhanced thermostability from mesophilic to thermophilic proteins. *Biophys Chem* 82:51–67.
11. Taverna DM, Goldstein RA (2002) Why are proteins marginally stable? *Proteins* 46:105–109.
12. Couñago R, Chen S, Shamoo Y (2006) In vivo molecular evolution reveals biophysical origins of organismal fitness. *Mol Cell* 22:441–449.
13. Sikosek T, Chan HS (2014) Biophysics of protein evolution and evolutionary protein biophysics. *J R Soc Interface* 11:20140419.
14. Lau KF, Dill KA (1989) A lattice statistical mechanics model of the conformational and sequence spaces of proteins. *Macromolecules* 22:3986–3997.
15. Mirny LA, Abkevich VI, Shakhnovich EI (1998) How evolution makes proteins fold quickly. *Proc Natl Acad Sci USA* 95:4976–4981.
16. Bornberg-Bauer E, Chan HS (1999) Modeling evolutionary landscapes: Mutational stability, topology, and superfunnels in sequence space. *Proc Natl Acad Sci USA* 96:10689–10694.
17. Bloom JD, Wilke CO, Arnold FH, Adami C (2004) Stability and the evolvability of function in a model protein. *Biophysical J* 86:2758–2764.
18. Weber G (1975) Energetics of ligand binding to proteins. *Advances in Protein Chemistry*, eds Anfinsen CB, Edsall JT, Richards FM (Academic, London), Vol 29, pp 1–83.
19. Eisenmesser EZ, et al. (2005) Intrinsic dynamics of an enzyme underlies catalysis. *Nature* 438:117–121.
20. Motlagh HN, Wrabl JO, Li J, Hilser VJ (2014) The ensemble nature of allostery. *Nature* 508:331–339.
21. Gunasekaran K, Ma B, Nussinov R (2004) Is allostery an intrinsic property of all dynamic proteins? *Proteins* 57:433–443.
22. Marsh JA, Teichmann SA (2014) Protein flexibility facilitates quaternary structure assembly and evolution. *PLoS Biol* 12:e1001870.
23. Rousseau F, Schymkowitz J (2005) A systems biology perspective on protein structural dynamics and signal transduction. *Curr Opin Struct Biol* 15:23–30.
24. Gillespie JH (2010) *Population Genetics: A Concise Guide* (Johns Hopkins Univ Press, Baltimore).
25. Sailer ZR, Harms MJ (2017) High-order epistasis shapes evolutionary trajectories. *PLoS Comput Biol* 13:e1005541.
26. Heckendorn RB, Whitley D (1999) Predicting epistasis from mathematical models. *Evol Comput* 7:69–101.
27. Poelwijk FJ, Krishna V, Ranganathan R (2016) The context-Dependence of mutations: A linkage of formalisms. *PLoS Comput Biol* 12:e1004771.
28. Sailer ZR, Harms MJ (2017) Detecting high-order epistasis in nonlinear genotype-phenotype maps. *Genetics* 205:1079–1088.
29. Alexander PA, He Y, Chen Y, Orban J, Bryan PN (2009) A minimal sequence code for switching protein structure and function. *Proc Natl Acad Sci USA* 106:21149–21154.
30. Nelson ED, Grishin NV (2016) Long-range epistasis mediated by structural change in a model of ligand binding proteins. *PLoS One* 11:e0166739.
31. Weinreich DM, Lan Y, Wylie CS, Heckendorn RB (2013) Should evolutionary geneticists worry about higher-order epistasis? *Curr Opin Genet Dev* 23:700–707.
32. Lu M, et al. (2013) Tristability in cancer-associated microRNA-TF chimera toggle switch. *J Phys Chem B* 117:13164–13174.
33. Bessonnard S, et al. (2014) Gata6, Nanog and Erk signaling control cell fate in the inner cell mass through a tristable regulatory network. *Development* 141:3637–3648.
34. Cordell HJ (2002) Epistasis: What it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum Mol Genet* 11:2463–2468.
35. Kryazhimskiy S, Tkačik G, Plotkin JB (2009) The dynamics of adaptation on correlated fitness landscapes. *Proc Natl Acad Sci USA* 106:18638–18643.
36. Miyazawa S, Jernigan RL (1985) Estimation of effective interresidue contact energies from protein crystal structures: Quasi-chemical approximation. *Macromolecules* 18:534–552.
37. Gillespie JH (1984) Molecular evolution over the mutational landscape. *Evolution* 38:1116–1129.