

Ensemble epistasis: thermodynamic origins of non-additivity between mutations

Anneliese J. Morrison^{*†}, Daria R. Wonderlick^{*†} and Michael J. Harms^{*†,1}

^{*}Institute of Molecular Biology, University of Oregon, Eugene, OR, USA 97403, [†]Department of Chemistry and Biochemistry, University of Oregon, Eugene, OR, USA 97403

ABSTRACT Epistasis—when mutations combine non-additively—is a profoundly important aspect of biology. It is often difficult to understand its mechanistic origins. Here we show that epistasis can arise from the thermodynamic ensemble, or the set of interchanging conformations a protein adopts. *Ensemble epistasis* occurs because mutations can have different effects on different conformations of the same protein, leading to non-additive effects on its average, observable properties. Using a simple analytical model, we found that ensemble epistasis arises when two conditions are met: 1) a protein populates at least three conformations and 2) mutations have differential effects on at least two conformations. To explore the relative magnitude of ensemble epistasis, we performed a virtual deep-mutational scan of the allosteric Ca^{2+} signaling protein S100A4. We found that 47% of mutation pairs exhibited ensemble epistasis with a magnitude on the order of thermal fluctuations. We observed many forms of epistasis: magnitude, sign, and reciprocal sign epistasis. The same mutation pair could even exhibit different forms of epistasis under different environmental conditions. The ubiquity of thermodynamic ensembles in biology and the pervasiveness of ensemble epistasis in our dataset suggests that it may be a common mechanism of epistasis in proteins and other macromolecules.

KEYWORDS epistasis; thermodynamic ensemble; protein evolution; predictability; thermodynamics

Introduction

Epistasis—when the effect of a mutation depends on the presence or absence of other mutations—is a common feature of biology. Epistasis can hint at biological mechanism [Maisnier-Patin et al. \(2007\)](#); [Alexander et al. \(2009\)](#); [Yang et al. \(2019\)](#); [Ortlund et al. \(2007\)](#); [Yokoyama et al. \(2014\)](#); [Baier et al. \(2019\)](#), profoundly shape evolution [Sailer and Harms \(2017b\)](#); [Weinreich et al. \(2005\)](#); [Poelwijk et al. \(2007\)](#), and complicate bioengineering that involves simultaneously introducing multiple mutations [Miton and Tokuriki \(2016\)](#); [Sykora et al. \(2014\)](#); [Giger et al. \(2013\)](#). It is therefore important to understand the general mechanisms by which epistasis can arise. Such knowledge will help us better understand biological systems, explain historical evolutionary trajectories, and improve models to predict the combined effects of mutations.

One important class of epistasis is that which occurs between mutations within a single protein. The magnitude of such epistatic interactions, ϵ , can be quantitatively described as shown in Fig 1A; it simply represents the difference in the effect of mutation $a \rightarrow A$ in the ab and aB backgrounds. Sometimes, such epistasis can be understood intuitively. In Fig 1B, epistasis arises because the positive charge of mutation $a \rightarrow A$ is adjacent to the negative charge of mutation $b \rightarrow B$. Epistasis occurs as a result of an electrostatic interaction between charged residues. Sometimes, however, epistasis can be difficult to rationalize. Fig 1C shows epistasis between two positions distant in the structure. Where does such epistasis come from? Can it be predicted from an understanding of protein biochemistry?

We and others noted previously that the thermodynamic ensemble of a protein could potentially give rise to non-additive interactions between mutations [Sailer and Harms \(2017c\)](#); [Ancel and Fontana \(2000\)](#). Proteins exist as ensembles of interchanging conformations, where the probability of seeing an individual conformation is determined by its relative energy. The functional output of a protein is averaged over the functional properties and populations of all individual ensemble conformations [Tsai](#)

doi: 10.1534/genetics.XXX.XXXXXX

Manuscript compiled: Tuesday 22nd June, 2021

¹Corresponding author: Corresponding author: 1370 Franklin Blvd., Room 269A Klamath Hall, Molecular Biology, University of Oregon, Eugene, OR 97403. E-mail: harms@uoregon.edu

and Nussinov (2014); Wei *et al.* (2016); Motlagh *et al.* (2014). Mutations can have different effects on each conformation, redistributing their relative probabilities in a nonlinear fashion. The effects of such mutations with respect to an observable would not sum additively, leading to *ensemble epistasis*.

Many important questions about ensemble epistasis remain unanswered. Under what conditions is ensemble epistasis expected to arise? Can it lead to different classes of evolutionarily-relevant epistasis, i.e. magnitude, sign, reciprocal-sign, and high-order? Is it plausible that such epistasis could occur in a real protein, rather than the highly simplified lattice models we used previously? And, finally, are there signals for ensemble epistasis that one might detect experimentally?

To address these questions, we set out to rigorously describe the thermodynamic and mechanistic basis for ensemble epistasis. We identified the minimal set of conditions that are necessary to observe ensemble epistasis: 1) a protein populates three or more conformations and 2) mutations have differential effects on two or more conformations within the ensemble. We found that this can lead to many types of epistasis, including magnitude, sign, reciprocal sign, and high-order epistasis. From structure-based calculations on the allosteric S100A4 protein, we predict that a large fraction of mutant pairs in real proteins will exhibit ensemble epistasis. We also found that varying the concentration of allosteric effectors could tune epistasis, suggesting one might experimentally detect ensemble epistasis by measuring epistasis at different concentrations of allosteric effectors. We conclude that ensemble epistasis is likely an important determinant of non-additivity between mutations in proteins.

Materials and Methods

For the S100A4 epistasis analysis, we used three published structures for S100A4: the apo structure (PDB 1M31), the Ca^{2+} bound structure (PDB 2Q91), and the structure bound to both Ca^{2+} and a peptide extracted from Annexin A2 (PDB 5LPU). We removed all non- Ca^{2+} small molecules (including waters) and edited the files to have an identical set of non-hydrogen atoms for the S100A4 chains (trimming any residues before alanine 2 and after phenylalanine 93 in the uniprot sequence, P26447). We arbitrarily selected the first NMR model for the apo structure. Using ROSETTA (Linux build 2018.33.60351), we generated five independent, pre-minimized structures for each of the conformations (*apo*, *ca*, and *capep*). We then used the “cartesian_ddg” binary to introduce each mutation three times into each of these five pre-minimized structures, yielding 15 calculated ΔG values for each mutation in each of the three conformations Park *et al.* (2016). Finally, we averaged the 15 values for each mutation in each conformation. We assumed the units of these ΔG values were in $kcal \cdot mol^{-1}$ Alford *et al.* (2017).

For a given genotype, we described the free energy of the calcium-bound form as a function of calcium chemical potential ($\mu_{Ca^{2+}}$) with the expression $G_{ca}(\mu_{Ca^{2+}}) = G_{ca}^{\circ} - 4\mu_{Ca^{2+}}$. G_{ca}° is a constant describing both the relative stability of the “open” form of the protein relative to the “closed” form and the affinity of the open form for Ca^{2+} . We treated the free energy of the apo form as $G_{apo}^{\circ}(\mu_{Ca^{2+}}) = G_{apo}^{\circ}$, where G_{apo}° measures the free energy of the apo form. For convenience, we set $G_{apo}^{\circ,ab} = 0 kcal \cdot mol^{-1}$ and $G_{ca}^{\circ,ab} = 10 kcal \cdot mol^{-1}$ for $\mu = 0 kcal \cdot mol^{-1}$. This models the fact that, at some reference $[Ca^{2+}]$, the “closed” form is favored over the “open” form. As $[Ca^{2+}]$ increases, $G_{ca}(\mu_{Ca^{2+}})$ becomes more negative and eventually becomes more favorable than

G_{apo} . To verify that this result was not due to the choice of G_{ca}° , we re-ran our analysis for different values of G_{ca}° . We found that changing the value of G_{ca}° has little impact on the magnitude of epistasis we observe. Its main effect is changing the $\mu_{Ca^{2+}}$ value at which the maximum magnitude of epistasis is observed (see supplemental Fig S1).

We modeled the effects of mutations as changes to G_{ca}° and G_{apo}° . For the *Ab* genotype, for example, we would write:

$$G_{ca}^{Ab}(\mu_{Ca^{2+}}) = G_{ca}^{\circ} - 4\mu_{Ca^{2+}} + \delta G_{ca}^{a \rightarrow A} \quad (1)$$

$$G_{apo}^{Ab} = G_{apo}^{\circ} + \delta G_{apo}^{a \rightarrow A} \quad (2)$$

$$\langle G_{ca,apo}^{Ab} \rangle(\mu_{Ca^{2+}}) = -RT \ln \left(e^{-(G_{ca}^{\circ} - 4\mu_{Ca^{2+}} + \delta G_{ca}^{a \rightarrow A})/RT} + e^{-(G_{apo}^{\circ} + \delta G_{apo}^{a \rightarrow A})/RT} \right) \quad (3)$$

where $\delta G_{ca}^{a \rightarrow A}$ and $\delta G_{apo}^{a \rightarrow A}$ are the energetic effects of mutation $a \rightarrow A$ on the *ca* and *apo* conformations, respectively. See section 5 of the supplemental text for further information, including a derivation of the model.

Results

Defining the three-conformation ensemble

To understand how the thermodynamic ensemble might lead to epistasis, we first defined a simple quantitative model of a protein exchanging between three conformations *i*, *j*, and *k*. We defined *i* as the “active” conformation in equilibrium with two “inactive” conformations *j* and *k*. This is a generic model that describes, in broad strokes, a wide variety of functions that depend on conformational change (Fig 2A). For example, conformation *i*, but not conformations *j* and *k*, could be capable of catalysis.

We will analyze epistasis in the free energy difference between the active *i* conformation and the inactive conformations, *j* and *k* (ΔG_{obs}). This quantifies how much the active form of the enzyme is favored over the inactive forms. We define ΔG_{obs} as follows:

$$\Delta G_{obs} = G_i - \langle G_{j,k} \rangle, \quad (4)$$

where G_i is the energy of conformation *i* and $\langle G_{j,k} \rangle$ is the Boltzmann-weighted average of the free energies of conformations *j* and *k* (Fig 2B). Importantly, the free energy scale is linear, meaning—in the absence of epistasis—we expect the effects of mutations to sum.

We will now describe the origin of equation 4. (Some readers may wish to proceed to the next section, “Mutations can affect multiple conformations in the ensemble”.)

Due to thermal fluctuations, an individual protein molecule will flip between conformations *i*, *j*, and *k* over time. As a consequence, a population of many protein molecules will exhibit a mixture of conformations. Factors such as the number of favorable chemical bonds within each conformation determine the frequency of that conformation in the protein population.

The favorability of each conformation can be quantified by its free energy (*G*). Fig 2B shows a free energy landscape for a three-conformation ensemble. The large energy wells correspond to conformations *i*, *j*, and *k*, while the smaller wells correspond to small structural fluctuations within each conformation, such as side-chain rearrangements. Because conformation *i* has a low free energy in this hypothetical example, it will have a much higher frequency in the population than conformations *j* or *k*.

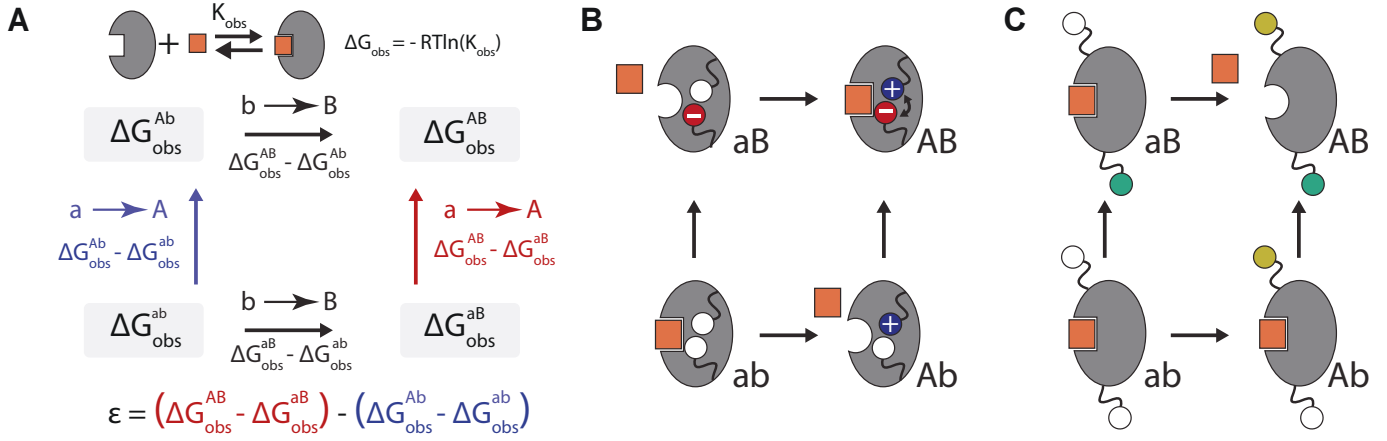


Figure 1 Mechanistic and mathematical descriptions of epistasis. A) The mathematical description of epistasis (ϵ) in ligand binding free energy (ΔG_{obs}) for the mutant cycle between genotypes ab and AB . ΔG_{obs} measures the strength of the binding interaction between protein (gray) and ligand (orange). We indicate genotypes as superscripts. ϵ is defined as the difference in the effect of mutation $a \rightarrow A$ in the aB background (red text), versus its effect in the ab background (blue text). B) Mutant cycle where epistasis is readily understood: the $a \rightarrow A$ and $b \rightarrow B$ mutations introduce charges into the hydrophobic core, destabilizing the protein and disrupting binding of the orange square. Mutations $a \rightarrow A$ and $b \rightarrow B$ lead to a new electrostatic interaction when introduced together (minus and plus signs) restoring stability and binding. C) Mutant cycle with difficult-to-understand epistasis. Mutations at two distant sites (green and yellow spheres) have no effect on binding of the orange square when introduced independently, but disrupt binding when introduced together.

1 The statistical weight for a given conformation is related to
2 its free energy by the Boltzmann distribution:

$$w_c = e^{-G_c/RT} \quad (5)$$

3 where c indicates a conformation with free energy G_c , R is the
4 gas constant and T is the temperature in Kelvin. In the three-
5 conformation ensemble, the frequency of conformation i is given
6 by:

$$f_i = \frac{w_i}{w_i + w_j + w_k} = \frac{e^{-G_i/RT}}{e^{-G_i/RT} + e^{-G_j/RT} + e^{-G_k/RT}} \quad (6)$$

7 Importantly, the frequencies of the conformations are coupled.
8 For example, making conformation j more stable (by decreasing
9 G_j) will lower f_i , even if G_i remains the same. This is because
10 individual protein molecules will spend more time in conformation
11 j and thus less time, on average, in conformation i .

12 As noted above, we are modeling an ensemble in which
13 conformation i is active and conformations j and k are not. A
14 typical way to quantify activity in such a system is with an
15 equilibrium constant, describing the frequency of i relative to j
16 and k :

$$K_{obs} = \frac{f_i}{f_j + f_k} = \frac{e^{-G_i/RT}}{e^{-G_j/RT} + e^{-G_k/RT}} \quad (7)$$

17 Equilibrium constants follow a multiplicative scale, mean-
18 ing that the effects of mutations are expected to multiply rather
19 than add. We will take logarithm of K_{obs} and place the observ-
20 able on a free-energy scale, where—in the absence of epista-
21 sis—mutational effects are expected to add:

$$\Delta G_{obs} = -RT \ln(K_{obs}) = G_i + RT \ln(e^{-G_j/RT} + e^{-G_k/RT}). \quad (8)$$

ΔG_{obs} measures the difference in the free energy, at equilibrium,
of the active i conformation and the inactive j and k conforma-
tions (Fig 2B). We will write the second term as:

$$\langle G_{j,k} \rangle \equiv -RT \ln(e^{-G_j/RT} + e^{-G_k/RT}) \quad (9)$$

where the brackets denote the Boltzmann-weighted average.
This gives us, finally:

$$\Delta G_{obs} = G_i - \langle G_{j,k} \rangle. \quad (10)$$

Mutations can affect multiple conformations in the ensemble

We next considered the effects of mutations. Because each con-
formation may have different physical interactions, the same
mutation may have different effects on different conformations.
For the three-conformation ensemble in Fig 2B, we thus need
terms to describe the effect of the mutation on conformations i , j
and k . To keep track of these effects, we will use the following
notation:

- The observable energy for genotype g is ΔG_{obs}^g (e.g. ΔG_{obs}^{ab}).
- The energy of conformation c is G_c^g (e.g. G_i^{ab}).
- The energetic effect of mutation $x \rightarrow X$ on conformation c is $\delta G_c^{x \rightarrow X}$ (e.g. $\delta G_j^{a \rightarrow A}$). Unless indicated, mutations are always introduced into the ab genetic background.
- Epistasis within a conformation—meaning the difference in the effect of $a \rightarrow A$ on the energy of conformation c in the ab and aB backgrounds—is $\delta \delta G_c^{ab \rightarrow AB}$.

We will now consider the effect of mutation $a \rightarrow A$ on ΔG_{obs}^{ab}
(Fig 2C). The three terms that describe its effect are $\delta G_i^{a \rightarrow A}$,
 $\delta G_j^{a \rightarrow A}$, and $\delta G_k^{a \rightarrow A}$. Fig 2C shows how a hypothetical muta-
tion $a \rightarrow A$ might change the ensemble: it has a small effect
on conformation i , stabilizes j , and destabilizes k . We would
describe the effect of the mutation mathematically as:

$$\Delta G_{obs}^{Ab} = \left(G_i^{ab} + \delta G_i^{a \rightarrow A} \right) - \left\langle G_{j,k}^{Ab} \right\rangle \quad (11)$$

where

$$\left\langle G_{j,k}^{Ab} \right\rangle \equiv -RT \ln \left(e^{-(G_j^{ab} + \delta G_j^{a \rightarrow A})/RT} + e^{-(G_k^{ab} + \delta G_k^{a \rightarrow A})/RT} \right). \quad (12)$$

The mutation in Fig 2C stabilizes $\left\langle G_{j,k}^{Ab} \right\rangle$ relative to $\left\langle G_{j,k}^{ab} \right\rangle$ because conformation j becomes so much more favorable. As a result, the ΔG_{obs}^{Ab} is lower than ΔG_{obs}^{ab} (Fig 2C).

The next step is to describe the effect of introducing two mutations simultaneously. To isolate epistasis that arises solely from changes to the thermodynamic ensemble, we will start by assuming that mutations are additive within each conformation. By this we mean that $G_c^{AB} = G_c^{ab} + \delta G_c^{a \rightarrow A} + \delta G_c^{b \rightarrow B}$. There are no epistatic contributions of the form $\delta \delta G_c^{ab \rightarrow AB}$ reflecting physical interactions within each conformation of the sort seen in Fig 1B. This means any epistasis we observe arises solely from the ensemble. We will revisit this simplifying assumption later.

Using this framework, we can describe the combined effects of mutations $a \rightarrow A$ and $b \rightarrow B$ on ΔG_{obs} as the following:

$$\Delta G_{obs}^{AB} = \left(G_i^{ab} + \delta G_i^{a \rightarrow A} + \delta G_i^{b \rightarrow B} \right) - \left\langle G_{j,k}^{AB} \right\rangle \quad (13)$$

where

$$\left\langle G_{j,k}^{AB} \right\rangle \equiv -RT \ln \left(e^{-(G_j^{ab} + \delta G_j^{a \rightarrow A} + \delta G_j^{b \rightarrow B})/RT} + e^{-(G_k^{ab} + \delta G_k^{a \rightarrow A} + \delta G_k^{b \rightarrow B})/RT} \right). \quad (14)$$

The thermodynamic ensemble can lead to epistasis

To understand the nature of epistasis arising from such a system, we must map the thermodynamic model in Equation 13 to epistasis. Table 1 shows the mapping between each genotype and its thermodynamic description, $\Delta G_{obs}^{genotype}$. We will treat epistasis as the quantitative difference between the effects of mutation $a \rightarrow A$ in the ab and aB backgrounds (Fig 1A):

$$\varepsilon = \left(\Delta G_{obs}^{AB} - \Delta G_{obs}^{aB} \right) - \left(\Delta G_{obs}^{Ab} - \Delta G_{obs}^{ab} \right). \quad (15)$$

We can substitute the thermodynamic equations for each ΔG_{obs} from Table 1 into Equation 15. Upon simplifying this expression (supplementary text, section 1.1), we obtain:

$$\varepsilon = - \left[\left(\left\langle G_{j,k}^{AB} \right\rangle - \left\langle G_{j,k}^{aB} \right\rangle \right) - \left(\left\langle G_{j,k}^{Ab} \right\rangle - \left\langle G_{j,k}^{ab} \right\rangle \right) \right]. \quad (16)$$

All terms associated with conformation i cancel. We are left with a description of ε that is only in terms of mutational effects on conformations j and k .

Our expression for ε is determined by the effects of mutations $a \rightarrow A$ and $b \rightarrow B$ on conformations j and k , not their effects on conformation i . Perturbations to the relative populations of j and k necessarily lead to nonlinear changes in ΔG_{obs} because the logarithmic term in $\left\langle G_{j,k} \right\rangle$ cannot be simplified further.

Conditions necessary for ensemble epistasis

We next used the thermodynamic description of ensemble epistasis derived above (Equation 16) to ask under what conditions ensemble epistasis is expected to arise. In the supplemental text, we show that there are two necessary conditions for ensemble epistasis:

1. The protein populates at least three conformations (supplemental text, section 1.2)
2. Mutations have differential effects on conformations j and k (supplemental text, section 1.3).

To understand what these conditions mean in practice, we calculated ensemble epistasis using equation 16 as a function of the difference in the stabilities of conformations j and k ($G_j^{ab} - G_k^{ab}$) and the difference in the effects of mutations on conformations j and k ($\delta G_j^{x \rightarrow X} - \delta G_k^{x \rightarrow X}$) (Fig 3A). In panels B-D, we reveal the underlying ensemble that leads to the epistasis observed in Fig 3A. The length of the pink arrows illustrates the effect of mutation $a \rightarrow A$ in each genetic background, ab or aB . The difference in the length of the pink arrows for the $ab \rightarrow Ab$ and $aB \rightarrow AB$ genotypes measures epistasis, ε .

We can see why multiple conformations are required for ensemble epistasis by comparing points B and C on Fig 3A. At point B, only conformation j is appreciably populated for all genotypes (pie charts, Fig 3B); at point C, conformations j and k have equal starting populations (pie charts, Fig 3C). This difference in the starting populations of j and k leads to different epistatic outcomes. At point B, both $ab \rightarrow Ab$ and $aB \rightarrow AB$ depend only on the effect of the mutation on conformation j because it is the only conformation appreciably populated. The lengths of the pink arrows are equal, indicating that there is no epistasis. At point C, the effect of $ab \rightarrow Ab$ on $\left\langle G_{j,k} \right\rangle$ is moderate because the stabilization of conformation j is offset by the entropic cost of depopulating conformation k . This results in epistasis because when $a \rightarrow A$ is introduced into the aB background, mutation $b \rightarrow B$ has already depopulated conformation k . As a result, the effect of $aB \rightarrow AB$ is determined solely by its stabilization of conformation j , and is thus larger than $ab \rightarrow Ab$.

We can see why differential effects for each mutation are required by comparing points C and D on Fig 3A. At both points, conformations j and k have equal starting populations (pie charts, Fig 3 C-D). At point C, the mutations have opposite effects on conformations j and k (Fig 3C); at point D, the mutations have identical effects on conformations j and k (Fig 3D). This means that for point D the introduction of $a \rightarrow A$ or $b \rightarrow B$ shifts the total energy landscape, but does not change the relative proportions of j and k . As a result, mutation $a \rightarrow A$ has the same effect regardless of background (compare pink arrows, Fig 3D).

Ensembles can lead to magnitude epistasis, sign-epistasis, and reciprocal sign-epistasis

We next asked if the ensemble could lead to different evolutionarily-relevant classes of epistasis: magnitude, sign, and reciprocal sign epistasis. In magnitude epistasis, only the magnitude of a mutation's effect changes when another mutation is introduced. In sign epistasis, the same mutation has a positive effect in one background and a negative effect in another. Finally, in reciprocal sign epistasis, both mutations exhibit sign epistasis.

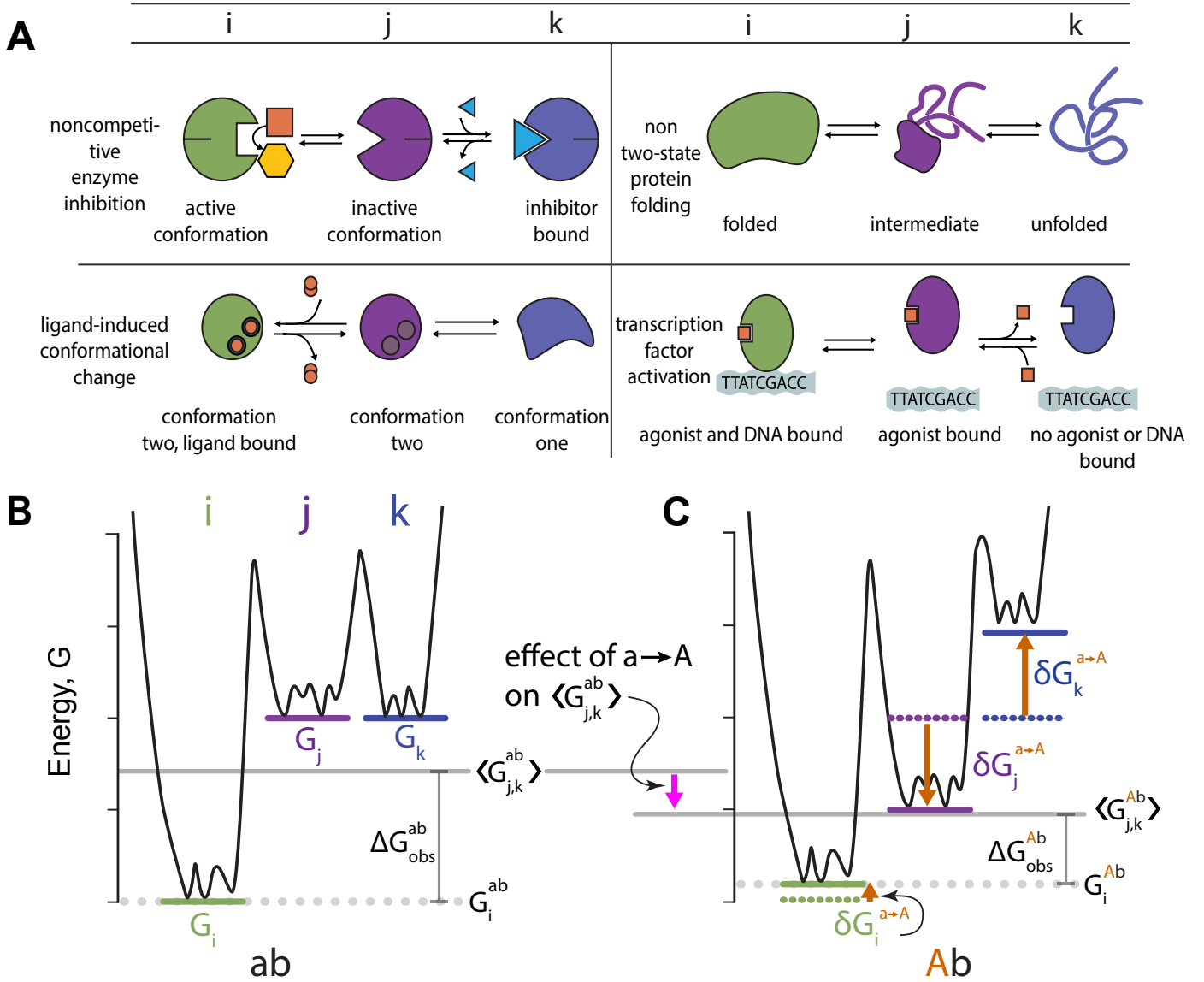


Figure 2 Mutations affect multiple ensemble conformations. A) Schematic examples of biological mechanisms in which a protein populates at least three conformations. Columns indicate conformation labels—*i* (green), *j* (purple), or *k* (blue). B) Energy diagram for a hypothetical protein with the *ab* genotype that adopts conformations *i* (green line), *j* (purple line), and *k* (blue line). The solid gray line indicates $\langle G_{j,k}^{ab} \rangle$ (the average energy of the inactive conformations *j* and *k*) and the dotted gray line indicates G_i^{ab} (the energy of the active conformation *i*). The difference between the solid and dotted gray lines is the observable, ΔG_{obs}^{ab} . C) Hypothetical mutation $a \rightarrow A$ changes the energies of conformations *i*, *j*, and *k* and thus ΔG_{obs} . Orange arrows represent the effect of mutation $a \rightarrow A$ on individual conformations. For example, $\delta G_j^{a \rightarrow A}$ shows the effect on conformation *j*. The mutation has a small effect on *i*, stabilizes *j*, and destabilizes *k*. This leads to a net decrease in $\langle G_{j,k}^{Ab} \rangle$ relative to $\langle G_{j,k}^{ab} \rangle$ (pink arrow), and thus a decrease in ΔG_{obs}^{Ab} relative to ΔG_{obs}^{ab} .

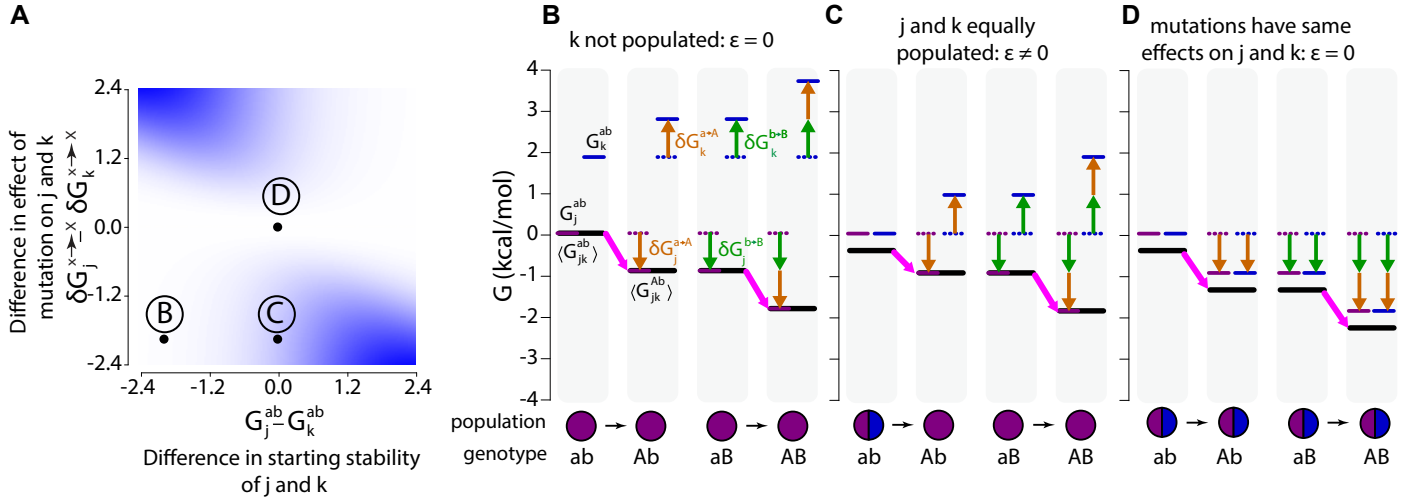


Figure 3 Ensemble epistasis arises from redistributed conformational probabilities. A) Epistasis as a function of the difference in the effects of the mutations $a \rightarrow A$ and $b \rightarrow B$ on conformations j and k ($\delta G_j^{a \rightarrow A} - \delta G_k^{a \rightarrow A}$ in $\text{kcal} \cdot \text{mol}^{-1}$, y-axis) and the difference in the stability of conformations j and k for the ab genotype ($G_j^{ab} - G_k^{ab}$ in $\text{kcal} \cdot \text{mol}^{-1}$, x-axis). Color indicates the magnitude of epistasis, ranging from 0 (white) to $1.6 \text{ kcal} \cdot \text{mol}^{-1}$ (blue). For the whole plot, $a \rightarrow A$ and $b \rightarrow B$ had identical effects ($\delta G_j^{a \rightarrow A} = \delta G_j^{b \rightarrow B}$ and $\delta G_k^{a \rightarrow A} = \delta G_k^{b \rightarrow B}$). We set $G_j^{ab} = 0 \text{ kcal} \cdot \text{mol}^{-1}$ and $\delta G_j^{x \rightarrow X} = -0.96 \text{ kcal} \cdot \text{mol}^{-1}$ and then varied G_k^{ab} and $\delta G_k^{x \rightarrow X}$ to sample parameter space. All calculations were done at $T = 298 \text{ K}$. Panels B-D show the thermodynamic origins for the epistasis at points B, C, and D indicated on panel A. The color scheme is consistent throughout: purple and blue lines are the energies of conformations j and k , respectively; orange arrows show the effects of mutation $a \rightarrow A$; green arrows show the effects of mutation $b \rightarrow B$; heavy black lines are the Boltzmann-weighted average energies of j and k , $\langle G_{j,k} \rangle$; heavy pink arrows are the observed effect of mutation $a \rightarrow A$ in the genotype indicated below the plot. The difference between the length of the pink arrows in the $ab \rightarrow Ab$ and $aB \rightarrow AB$ genotypes measures ϵ . The relative populations of conformations j and k are shown as a pie chart below the energy diagram.

We surveyed the parameter space for the effects of mutations on each conformation while tracking the magnitude and type of epistasis observed (Fig 4A). We set the initial energies of conformations j and k to be equal ($G_j^{ab} = G_k^{ab} = 0$). We then calculated epistasis using equation 16 as a function of the difference in the effects of mutations $a \rightarrow A$ and $b \rightarrow B$ on j and k .

We found four regimes, corresponding to magnitude, sign, reciprocal sign, and no epistasis. To understand the origins of these three regimes, we studied the thermodynamic ensembles that lead to epistasis at the points indicated C, D, and E. At this slice of parameter space, mutation $a \rightarrow A$ destabilizes conformation j by $0.35 \text{ kcal} \cdot \text{mol}^{-1}$ and stabilizes conformation k by $-0.35 \text{ kcal} \cdot \text{mol}^{-1}$. The effect of this mutation on the ensemble in the ab background is shown in Fig 4B: the mutation mildly stabilizes $\langle G_{j,k} \rangle$.

At point C, we see no epistasis (Fig 4A). We can see why this occurs in Fig 4C. Mutation $b \rightarrow B$ destabilizes both j and k by $0.35 \text{ kcal} \cdot \text{mol}^{-1}$. Because mutation $b \rightarrow B$ does not have differential effects on each conformation, $\langle G_{j,k} \rangle$ is globally shifted by $+0.35 \text{ kcal} \cdot \text{mol}^{-1}$. Introducing $a \rightarrow A$ and $b \rightarrow B$ together yields no epistasis because both the ab and aB genotypes have identical configurations—the observed effect comes only from mutation $a \rightarrow A$ (compare pink arrows in Fig 4B and Fig 4C).

At point D, we observe magnitude epistasis (Fig 4A). We can see why this occurs in Fig 4D. Mutations $a \rightarrow A$ and $b \rightarrow B$ have synergistic effects on each conformation: k is stabilized while j is destabilized. We see magnitude epistasis because although the relative population of j is reduced, it still has weight in the Boltzmann-weighted average stability (compare pink arrows in Fig 4B and 4D).

At point E, we see reciprocal sign epistasis (Fig 4A). We can see why this occurs in Fig 4E. $a \rightarrow A$ and $b \rightarrow B$ have opposite effects on j and k : $a \rightarrow A$ destabilizes j and stabilizes k , while $b \rightarrow B$ stabilizes j and destabilizes k . The effects are equal in magnitude but opposite in sign so their combined effects cancel, yielding $\langle G_{j,k}^{AB} \rangle$ equal to that of the ab genotype (compare pink arrows in Fig 4B and 4E). As a result, mutations $a \rightarrow A$ and $b \rightarrow B$ have individually stabilizing effects on $\langle G_{j,k} \rangle$ but are destabilizing when combined.

The magnitude and sign regions of Fig 4A show distinct patterns with regard to the sign of epistasis observed: mutations in the magnitude region are more stabilizing (positive epistasis) and those in the sign region are more destabilizing (negative epistasis) than anticipated based on single mutational effects. The magnitude region results in positive epistasis because mutations work synergistically to hyper-stabilize one conformation, while greatly destabilizing the other. This results in one conformation having very little weight in the Boltzmann distribution such that the remaining stabilized conformation determines the observable value. In the sign region, each mutation preferentially stabilizes a different conformation when introduced alone. However, when introduced together, they have opposing effects within a single conformation. The stabilizing effects of each mutation alone on $\langle G_{j,k} \rangle$ cancel, resulting in a less stable double mutant than anticipated.

The thermodynamic ensemble can lead to high-order epistasis

In addition to magnitude, sign, and reciprocal sign epistasis, high-order epistasis is evolutionarily important [Weinreich et al. \(2013\)](#); [Sailer and Harms \(2017b\)](#). In high-order epistasis, the effect of a three-way mutant cannot be explained by the indi-

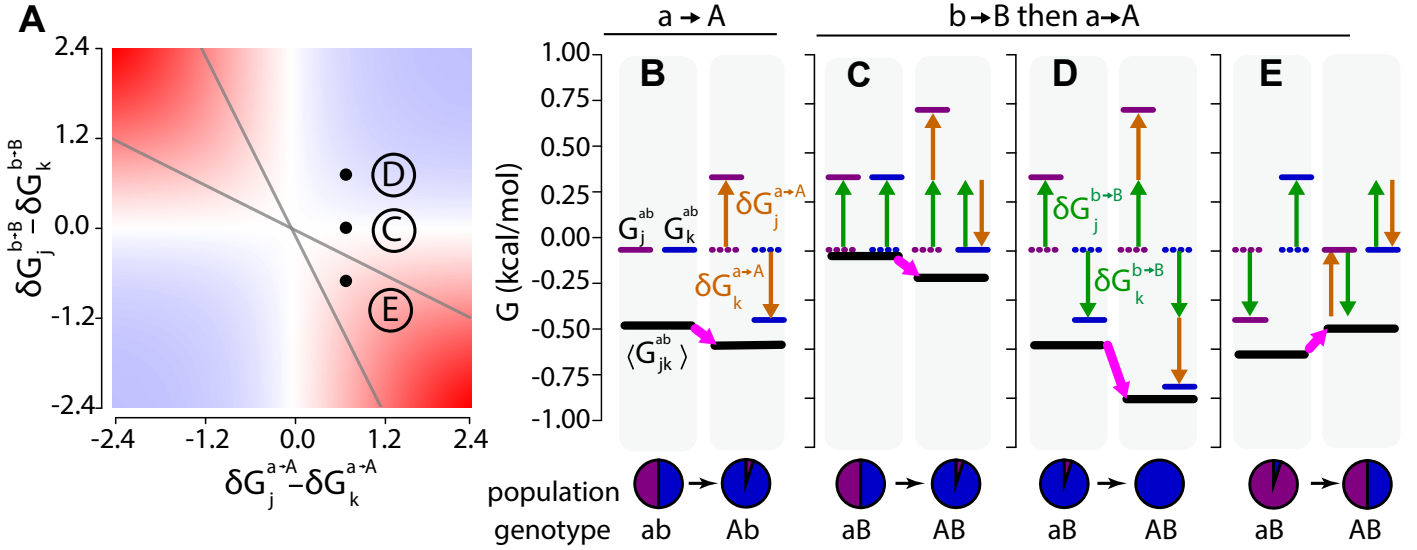


Figure 4 Ensemble epistasis arises when mutations have different effects on different conformations. A) Epistasis calculated for a three-conformation ensemble that starts with $G_j^{ab} = G_k^{ab} = 0$. The differences in the effects of mutations $a \rightarrow A$ and $b \rightarrow B$ on conformations j and k are indicated on the x- and y-axes. The magnitude of epistasis is indicated by the color, ranging from +1.6 (dark red) to 0 (white) to $-1.6 \text{ kcal} \cdot \text{mol}^{-1}$ (dark blue). Gray lines delineate regions of reciprocal sign (red regions within the lines) and sign epistasis (red regions outside of the lines). All calculations were done at $T = 298 \text{ K}$. Panels B-E show the thermodynamic origins of the epistasis indicated by points C, D, and E on panel A. The effect of mutation $a \rightarrow A$ is constant in all panels; the effect of mutation $b \rightarrow B$ differs depending on the scenario. The color scheme is consistent with Fig 2. B) The effect of $a \rightarrow A$ in the ab background. $a \rightarrow A$ destabilizes j and stabilizes k , stabilizing $\langle G_{jk}^{AB} \rangle$. C) Scenario C: no epistasis. $b \rightarrow B$ has the same effect on conformations j and k . D) Scenario D: $a \rightarrow A$ and $b \rightarrow B$ act synergistically to destabilize j and stabilize k . E) Scenario E: $a \rightarrow A$ and $b \rightarrow B$ have opposite effects on conformations j and k .

vidual and pairwise effects of its constituent mutations. In the supplement we find that high-order epistasis may arise by redistributing the relative populations of conformations j and k (see supplemental text, section 2). We anticipate that the results we have found for pairwise epistasis—the importance of differential mutational effects on different conformations, for example—will apply to high-order ensemble epistasis, but further work is needed to clarify the necessary and sufficient conditions to observe high-order ensemble epistasis.

Ensemble epistasis is not due to simplifying assumptions

We next wanted to relax two major assumptions we made above. The first assumption was that there were no epistatic interactions within conformations (as in Fig 1B). We show in the supplemental text (section 3), that epistasis within each conformation can co-exist alongside ensemble epistasis. We also revisit this question empirically in the following section, finding that ensemble epistasis and within-conformation epistasis have similar magnitudes.

The second assumption made above was that the ensemble could be described with only three conformations i , j and k (Fig 2). We asked what the form of ensemble epistasis would be if we considered an equilibrium between two sub-ensembles, X and Y , each of which could have many different conformations. The free energy difference between these sub-ensembles would be given by:

$$\Delta G_{obs} = \left[-RT \ln \left(\sum_{m \in X} e^{-G_m/RT} \right) \right] - \left[-RT \ln \left(\sum_{n \in Y} e^{-G_n/RT} \right) \right] \quad (17)$$

where m indexes over all conformations in X and n indexes over all conformations in Y . In more compact form, this would be:

$$\Delta G_{obs} = \langle G_X \rangle - \langle G_Y \rangle. \quad (18)$$

We show in the supplemental text (section 4) that for such a system, epistasis becomes:

$$\varepsilon = \left[\left(\langle G_X^{AB} \rangle - \langle G_X^{aB} \rangle \right) - \left(\langle G_X^{Ab} \rangle - \langle G_X^{ab} \rangle \right) \right] - \left[\left(\langle G_Y^{AB} \rangle - \langle G_Y^{aB} \rangle \right) - \left(\langle G_Y^{Ab} \rangle - \langle G_Y^{ab} \rangle \right) \right]. \quad (19)$$

Thus, we expect to see ensemble epistasis in such a system—for certain conformational energies and mutational effects, at least—because we cannot simplify the expression for ε further.

Ensemble epistasis may be a common feature in protein mutant cycles

Above we showed mathematically that ensemble epistasis can arise when multiple conformations are populated and mutations have different effects on different conformations. We next wanted to address whether these requirements are met in real systems. Multi-conformation ensembles are common in biology and we expect that the first requirement is often met (Fig 2A). However, it is not obvious that the requirement for differential effects of mutations is commonly satisfied. We designed a computational test to ask if it was plausible that both of these conditions are met simultaneously in a protein.

We investigated these questions using the allosteric Ca^{2+} signaling protein, human S100A4. S100A4 adopts a three-conformation ensemble, meeting our first requirement to ob-

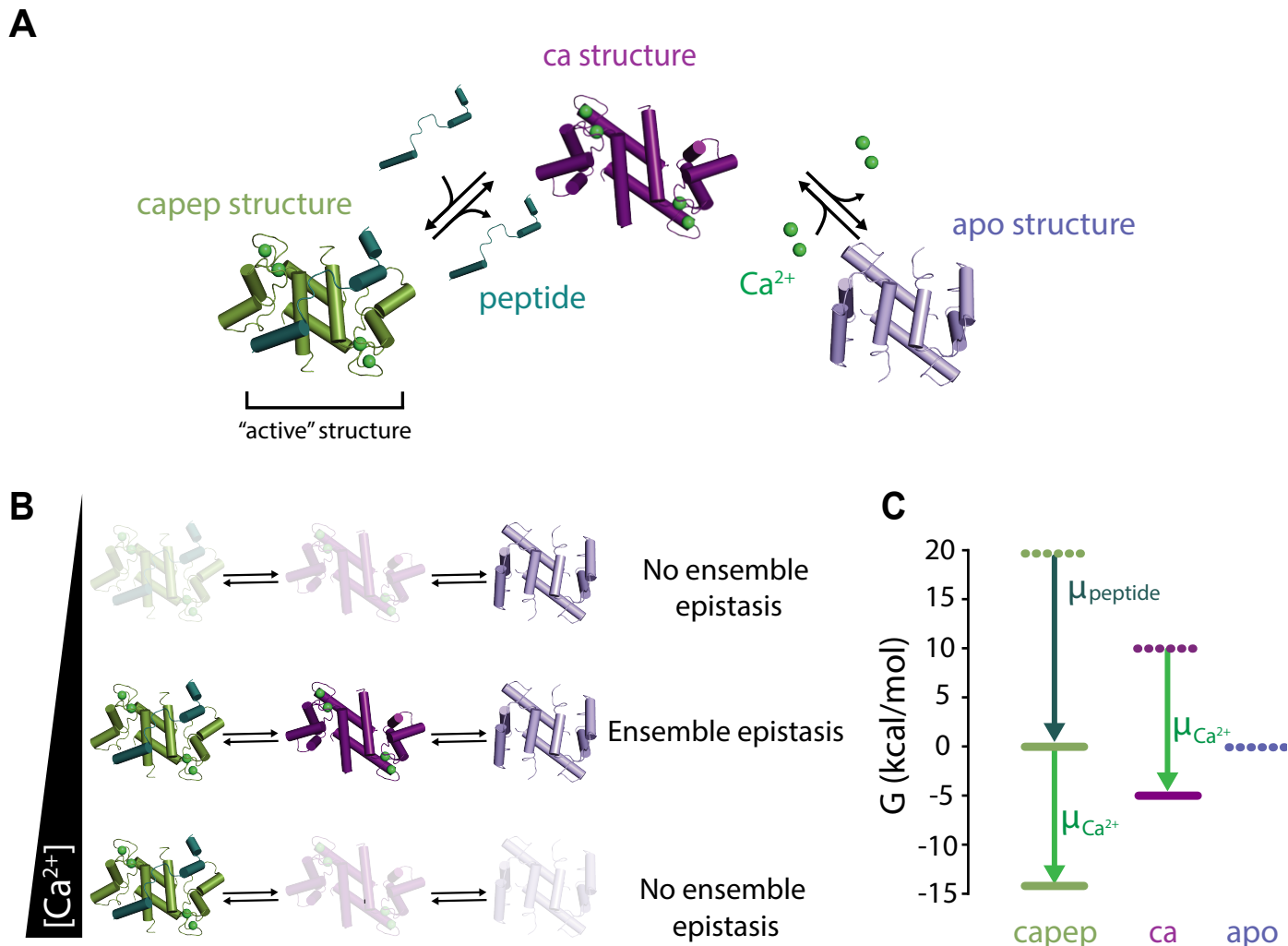


Figure 5 Testing for ensemble epistasis in the S100A4 protein. A) Three-conformation ensemble of the S100A4 protein. The apo conformation (*apo*, slate, PDB: 1M31) is in equilibrium with the Ca²⁺ bound (*ca*, purple, PDB: 2Q91) and Ca²⁺ / *peptide* bound (*capep*, green, PDB: 5LPU) conformations when Ca²⁺ (lime green spheres) and *peptide* (dark green) are present. B) The relative populations of the *apo*, *ca*, and *capep* conformations change as Ca²⁺ concentration increases in the presence of saturating peptide. The magnitude of ensemble epistasis observed is Ca²⁺-dependent, because only some Ca²⁺ concentrations lead to multiple populated conformations. C) Assigned energies (kcal · mol⁻¹) of S100A4 conformations. *Apo* is most stable when *peptide*, μ_{peptide} , and Ca²⁺ chemical potentials, $\mu_{\text{Ca}^{2+}}$, are zero (dashed lines). *Capep* is stabilized by increasing $\mu_{\text{peptide}} = 20 \text{ kcal} \cdot \text{mol}^{-1}$ (dark green arrow, solid green line). Increasing $\mu_{\text{Ca}^{2+}}$ alters the energies of both *ca* and *capep* (lime green arrow, solid lines). All calculations were done at $T = 298 \text{ K}$.

serve ensemble epistasis (Fig 5A) Valley *et al.* (2002); Malashkevich *et al.* (2008); Ecsédi *et al.* (2017). In the absence of Ca^{2+} , it favors the “apo” conformation (Fig 5A, slate); addition of Ca^{2+} stabilizes the “ca” conformation with an exposed hydrophobic peptide-binding surface (Fig 5A, purple); finally, addition of peptide leads to formation the “capep” conformation that has both Ca^{2+} and peptide bound (Fig 5A, green). These structures can be assigned indices, as in our analytical model: *capep* (i), *ca* (j), and *apo* (k).

We used software for structure-based energy calculations (ROSETTA) to estimate the stability effects of all 3,382 possible single point mutations to the *capep*, *ca*, and *apo* conformations of S100A4. This gives us $\delta G_{\text{capep}}^{x \rightarrow X}$, $\delta G_{\text{ca}}^{x \rightarrow X}$, and $\delta G_{\text{apo}}^{x \rightarrow X}$ for every mutation $x \rightarrow X$.

We then exploited the allosteric nature of S100A4 to switch between conditions where only single conformations are appreciably populated and where multiple conformations are populated. To model the ensemble, we selected reference concentrations of Ca^{2+} and peptide such that $G_{\text{capep}}^{\circ} \gg G_{\text{ca}}^{\circ} \gg G_{\text{apo}}^{\circ}$ (Fig 5C; see methods). We know experimentally that the protein favors the *apo* conformation in the absence of Ca^{2+} and peptide Garrett *et al.* (2008). We modeled the signaling behavior of S100A4 by changing the concentrations of Ca^{2+} and peptide: $G_{\text{capep}} = G_{\text{capep}}^{\circ} - 4\mu_{\text{Ca}^{2+}} - \mu_{\text{peptide}}$ and $G_{\text{ca}} = G_{\text{ca}}^{\circ} - 4\mu_{\text{Ca}^{2+}}$, where $\mu_{\text{Ca}^{2+}}$ and μ_{peptide} are the chemical potentials of Ca^{2+} and peptide relative to their reference concentrations (Fig 5C). Depending on our choice of $\mu_{\text{Ca}^{2+}}$ and μ_{peptide} , we can observe different relative populations of the *capep*, *ca*, and *apo* conformations. For ΔG_{obs} , we used:

$$\Delta G_{\text{obs}}^{\text{genotype}} = G_{\text{capep}}^{\text{genotype}} + RT \ln \left(e^{-G_{\text{ca}}^{\text{genotype}}/RT} + e^{-G_{\text{apo}}^{\text{genotype}}/RT} \right). \quad (20)$$

By analogy to what we derived in Equation 16, epistasis is calculated as:

$$\varepsilon = - \left[\left(\langle G_{\text{ca,apo}}^{AB} \rangle - \langle G_{\text{ca,apo}}^{aB} \rangle \right) - \left(\langle G_{\text{ca,apo}}^{Ab} \rangle - \langle G_{\text{ca,apo}}^{ab} \rangle \right) \right]. \quad (21)$$

We constructed all 5.6 million pairs of mutations by treating the $\delta G_{\text{capep}}^{x \rightarrow X}$, $\delta G_{\text{ca}}^{x \rightarrow X}$, and $\delta G_{\text{apo}}^{x \rightarrow X}$ ROSETTA values as additive within each conformation, meaning that we calculated the effect of two mutations $a \rightarrow A$ and $b \rightarrow B$ in combination on the *apo* conformation, for example, as $G_{\text{apo}}^{AB} = G_{\text{apo}}^{ab} + \delta G_{\text{apo}}^{a \rightarrow A} + \delta G_{\text{apo}}^{b \rightarrow B}$. We made this assumption to isolate epistasis arising solely from changes to the ensemble, as we did in our general thermodynamic model in Equation 13.

Under the assumption of within-conformation additivity, we calculated epistasis in $\langle G_{\text{ca,apo}} \rangle$ using Equation 21 as a function of $\mu_{\text{Ca}^{2+}}$ at a fixed μ_{peptide} (see methods for more details). We observed peaks in epistasis at intermediate values of $\mu_{\text{Ca}^{2+}}$, where the *capep*, *ca*, and *apo* conformations may all be populated. In contrast, we observed no epistasis at low $\mu_{\text{Ca}^{2+}}$ (where only the *apo* conformation is populated) or high $\mu_{\text{Ca}^{2+}}$ (where only the *capep* conformation is populated). We observed three basic patterns of $\mu_{\text{Ca}^{2+}}$ -dependent epistatic magnitude, as exemplified by the three mutant pairs shown in Fig 6A: F145R/L109I had no epistasis (left panel) while F145R/F78A had negative epistasis (middle panel) and F145R/M85K had positive epistasis (right panel). Interestingly, the type of epistasis observed—magnitude (dark blue), sign (gold), or reciprocal sign (green)—was also dependent upon $\mu_{\text{Ca}^{2+}}$ (Fig 6A). This was quite common in our dataset: approximately 61% of pairs with an epistatic magnitude

above $0.6 \text{ kcal} \cdot \text{mol}^{-1}$ switched epistatic type at least once as $\mu_{\text{Ca}^{2+}}$ increased.

We next looked at the magnitude and type of epistasis for all 5.6 million mutation pairs at their peak values over the range of $\mu_{\text{Ca}^{2+}}$. We found that 47% of the 5.6 million pairs exhibited epistasis at or above the order of thermal fluctuation, $0.6 \text{ kcal} \cdot \text{mol}^{-1}$ (Fig 6B). We found that 34% of pairs exhibited magnitude, 12% sign, and 1% reciprocal-sign epistasis at this cutoff. Approximately 11% of pairs exhibited epistasis with a magnitude above $2 \text{ kcal} \cdot \text{mol}^{-1}$.

To understand the structural origins of the observed epistasis, we compared the positions of each mutation from Fig 6A in the *apo* (slate, Fig 6C) and *ca* (purple, Fig 6C) conformations. We first consider F145R. This position is solvent exposed in the *apo* conformation but buried in the *ca* conformation. As a consequence, introducing Arg mildly stabilizes the *apo* conformation, but dramatically destabilizes the *ca* conformation due to burying its charge. Next, L109I is a conservative mutation at a site whose environment is essentially unchanged between the *apo* and *ca* conformations. F78A is solvent exposed in the *apo* conformation but buried in the *ca* conformation. The Phe to Ala mutation is destabilizing to the *ca* conformation due to the loss of hydrophobic contacts. Finally, M85K is buried in the *apo* conformation, but exposed in the *ca* conformation. Mutation to Lys introduces a buried charge, greatly destabilizing it due to the cost of ion desolvation. The differences in the effects of L109I, F78A, and M85K on the *apo* and *ca* conformations cause them to exhibit different types of epistasis when paired with F145R.

F145R exhibits no epistasis when paired with L109I at $\mu_{\text{Ca}^{2+}} = 3.5 \text{ kcal} \cdot \text{mol}^{-1}$ (Fig 6E). The L109I mutation has a negligible effect on the *apo* and *ca* conformations (genotype *ab*, Fig 6E). As a result, F145R has the same effect on $\langle G_{\text{ca,apo}} \rangle$ when introduced into both *ab* and L109I (*ab*) backgrounds (compare pink arrows in Fig 6D and 6E).

Pairing F145R with F78A results in sign epistasis. F78A is destabilizing to both conformations, but much more so to the *ca* conformation (genotype *ab*, Fig 6F). Both F78A and F145R preferentially destabilize the *ca* structure, leading to a dramatic decrease in its relative population when introduced together (green arrows, Fig 6F). We see sign epistasis because the synergistic destabilization of the *ca* conformation makes $\langle G_{\text{ca,apo}}^{AB} \rangle$ only dependent on the stability of the *apo* conformation (compare pink arrows in Fig 6D and 6F).

F145R exhibits magnitude epistasis when paired with M85K. The M85K mutation is greatly destabilizing to the *apo* conformation and slightly destabilizing to the *ca* conformation (green arrows, Fig 4G). Combining both mutations causes a decrease in the stability of both conformations and a net destabilization of $\langle G_{\text{ca,apo}}^{AB} \rangle$, leading to the observation of magnitude epistasis (pink arrows, Fig 6G).

Intriguingly, a slight decrease from $\mu_{\text{Ca}^{2+}} = 3.5 \text{ kcal} \cdot \text{mol}^{-1}$ to $\mu_{\text{Ca}^{2+}} = 2.2 \text{ kcal} \cdot \text{mol}^{-1}$ switches the type of epistasis from magnitude to sign for the F145R/M85K pair (compare Fig 6D/G to Fig 6H/I). The switch is solely due to the change in the relative energies of the *ca* and *apo* conformations in the *ab* genotype: the *ca* conformation is slightly stabilized relative to the *apo* conformation. The introduction of F145R stabilizes the *apo* conformation, resulting in net stabilization of $\langle G_{\text{ca,apo}}^{ab} \rangle$. M85K destabilizes both conformations, destabilizing $\langle G_{\text{ca,apo}}^{ab} \rangle$. When both mutations are combined, $\langle G_{\text{ca,apo}}^{AB} \rangle$ is further destabilized,

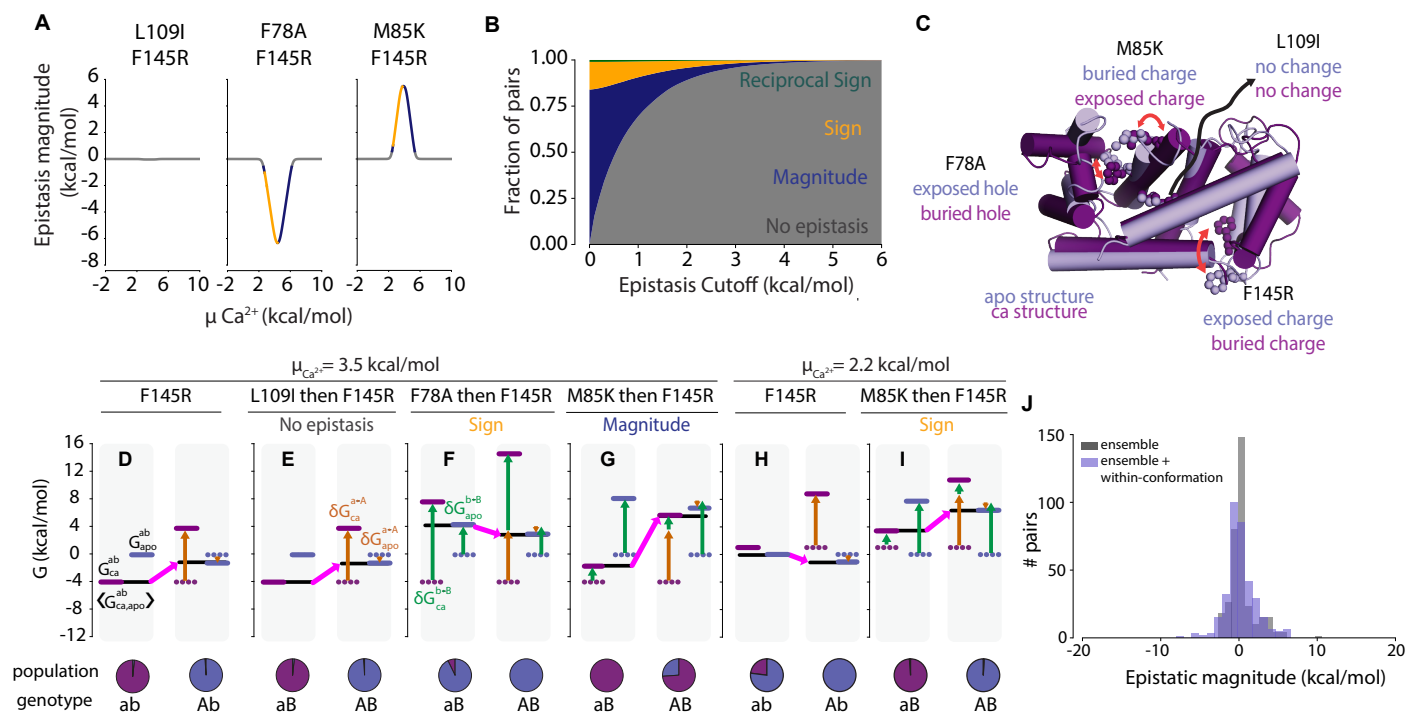


Figure 6 The ensemble of S100A4 exhibits ensemble epistasis. A) Epistatic magnitude ($\text{kcal} \cdot \text{mol}^{-1}$, y-axis) as a function of $\mu_{\text{Ca}^{2+}}$ ($\text{kcal} \cdot \text{mol}^{-1}$, x-axis) for three mutation pairs: L109I/F145R (left panel), F78A/F145R (middle panel), and M85K/F145R (right panel). Color is consistent with epistatic type in panel B. B) Fractional contribution of each epistatic type (y-axis) as a function of epistatic magnitude cutoff ($\text{kcal} \cdot \text{mol}^{-1}$, x-axis), colored by type: reciprocal sign (green), sign (gold), and magnitude (dark blue). Pairs with epistasis below the cutoff are considered non-epistatic (gray). C) Positions of mutations in the *ca* (purple) and *apo* (slate) conformations. Text indicates their relative environments in each conformation. Red arrows indicate changes in position between the *ca* and *apo* conformations. D-I) Thermodynamic origins of epistasis for three mutation pairs at $\mu_{\text{Ca}^{2+}} = 3.5 \text{ kcal} \cdot \text{mol}^{-1}$ (D-G) or $\mu_{\text{Ca}^{2+}} = 2.2 \text{ kcal} \cdot \text{mol}^{-1}$ (H-I). Ca^{2+} chemical potential is indicated above the panel. Mutation $a \rightarrow A$ (F145R) is constant; mutation $b \rightarrow B$ differs in panels E-G and I. The color scheme is consistent throughout: purple and blue lines are the energies of *ca* and *apo*, respectively, while black lines represent $\langle G_{\text{ca,apo}}^{\text{genotype}} \rangle$; all other colors are consistent with Fig 2-3. Specific mutations and epistatic classes are indicated at the top of the panel; genotypes and relative populations are below. G) Introduction of mutation F145R ($a \rightarrow A$) into the *ab* background at $\mu_{\text{Ca}^{2+}} = 3.5 \text{ kcal} \cdot \text{mol}^{-1}$. E) No epistasis scenario: mutations F145R ($a \rightarrow A$) and L109I ($b \rightarrow B$). F) Sign epistasis scenario: mutations F145R ($a \rightarrow A$) and F78A ($b \rightarrow B$). G) Magnitude epistasis scenario: mutations F145R ($a \rightarrow A$) and M85K ($b \rightarrow B$). H) Introduction of mutation F145R ($a \rightarrow A$) into the *ab* background at $\mu_{\text{Ca}^{2+}} = 2.2 \text{ kcal} \cdot \text{mol}^{-1}$. I) Sign epistasis scenario: mutations F145R ($a \rightarrow A$) and M85K ($b \rightarrow B$). J) Histogram showing the distribution of epistasis between 344 mutant pairs assuming no epistasis between mutations within each conformation (gray), or using calculated epistasis between mutations within each conformation (slate blue).

1 resulting in the observation of sign epistasis (compare pink ar-
2 rows in Fig 6H and Fig 6I).

3 **Ensemble epistasis is robust to addition of epistasis from** 4 **structural contacts**

5 We next wanted to ask how the relative magnitude of epistasis
6 changes when we allow epistasis to arise from both the ensemble
7 and structural contacts. We used ROSETTA to calculate the
8 within-conformation interaction energies of 344 mutant pairs.
9 We then re-calculated the stability of each conformation c as:

$$G_c^{AB} = G_c^{ab} + \delta G_c^{a \rightarrow A} + \delta G_c^{b \rightarrow B} + \delta \delta G_c^{ab \rightarrow AB}, \quad (22)$$

10 where $\delta \delta G_c^{ab \rightarrow AB}$ is the interaction energy within the conforma-
11 tion calculated by ROSETTA. The values of $\delta \delta G_c^{ab \rightarrow AB}$ had a
12 mean and standard deviation of $9.3 \pm 9.8 \text{ kcal} \cdot \text{mol}^{-1}$. We used
13 these new values to calculate ϵ in $\langle G_{ca,apo} \rangle$. Fig 6J shows how
14 the distribution of epistatic magnitude changes when we allow
15 non-additivity to arise from the ensemble alone versus both the
16 ensemble and structural contacts. We found that 24% of the 344
17 mutation pairs exhibit epistasis on the order of $0.6 \text{ kcal} \cdot \text{mol}^{-1}$,
18 with an average magnitude of $0.97 \text{ kcal} \cdot \text{mol}^{-1}$ when we allow
19 epistasis to arise only from the ensemble. When we allowed epis-
20 tasis to arise from structural contacts in addition to the ensemble,
21 we found that 35% of pairs exhibited epistasis on the order of
22 $0.6 \text{ kcal} \cdot \text{mol}^{-1}$, with an average magnitude of $1.4 \text{ kcal} \cdot \text{mol}^{-1}$.
23 The addition of within-conformation contacts widens the distri-
24 bution relative to the ensemble-only dataset, yielding a modest
25 increase in the average epistatic magnitude. Ensemble epistasis
26 thus seems to be an important source of epistasis, even for pro-
27 teins that also exhibit epistasis from structural contacts within
28 each conformation.

29 **Discussion**

30 We found that epistasis can arise from a fundamental property
31 of proteins and other macromolecules: the thermodynamic en-
32 semble. Previously we observed ensemble epistasis using lattice
33 models, but the conditions under which it arises and if they are
34 plausibly met in more realistic models of proteins remained un-
35 resolved Sailer and Harms (2017c). Here we used a simple—but
36 general—thermodynamic model to study the how the ensemble
37 leads to epistasis. Ensemble epistasis arises because mutations
38 can affect any conformation in the ensemble. Since observables
39 are averaged over the entire ensemble, they cannot be separated
40 into additive components.

41 **Ensemble epistasis should be pervasive in biology**

42 We expect ensemble epistasis in systems where 1) at least three
43 conformations are populated and 2) mutations have differential
44 effects on at least two conformations. The first requirement may
45 be common: multi-conformation ensembles often underlie bio-
46 logical function, from allostery to fold-switching (Fig 2A) Wei
47 et al. (2016). The commonality of the second requirement, how-
48 ever, is not as obvious. We tested for the plausibility of meeting
49 the second requirement by modeling the effects of mutations
50 on different conformations of the S100A4 protein. S100A4 is a
51 Ca^{2+} signaling protein that adopts three conformations, meet-
52 ing the requirement for multiple populated conformations (Fig
53 5A). We identified mutations that had differential effects on
54 both inactive conformations, which satisfied the second require-
55 ment. Nearly half of the mutant pairs exhibited epistasis above
56 $0.6 \text{ kcal} \cdot \text{mol}^{-1}$, suggesting that—at least in principle—ensemble
57 epistasis should be detectable in real proteins (Fig 6A).

58 There is mounting indirect evidence of links between epis-
59 tasis and thermodynamic ensembles. For example, in TEM-1
60 β -lactamase, two adaptive mutations were identified that in-
61 dependently increased structural heterogeneity and function.
62 Together the mutations exhibited epistasis, shifting the ensem-
63 ble into a dominantly non-productive structure Dellus-Gur et al.
64 (2015). Epistasis also underlies changes in dynamics that caused
65 functional divergence between Src and Abl kinases and the evo-
66 lution of fold-switching proteins Wilson et al. (2015); Seeliger
67 et al. (2007).

68 Recently, a thermodynamic model was used to decompose
69 mutational effects on the GB1 protein Otwinowski (2018). A
70 three-structure ensemble model was able to explain much of
71 the epistasis observed in the dataset. The remaining epistasis
72 pointed towards residues that contribute to functionally impor-
73 tant structural dynamics. This approach yielded mechanistic
74 information about the system. Notably, the mathematical frame-
75 work of the thermodynamic ensemble is not limited to proteins
76 and other macromolecules—it has been used to describe much
77 more complex biological systems like signaling networks and
78 bacterial communities Khazaei et al. (2012); Tran et al. (2008);
79 Hameri et al. (2019); Lu et al. (2013); Bessonard et al. (2014);
80 Venturi et al. (2010).

81 **Relationship to threshold epistasis**

82 Ensemble epistasis is related to—but conceptually distinct from—
83 threshold epistasis. Threshold epistasis describes non-additivity
84 arising from the accumulation of destabilizing mutations. Below
85 some threshold stability, the fraction of folded protein molecules
86 drops and any function encoded by the folded structure is lost
87 Bershtein et al. (2006); Kumar et al. (2017); Gong et al. (2013);
88 Petrović et al. (2018); Bloom et al. (2007). The same mutation
89 could have no effect on a high stability protein, but be highly
90 deleterious to a low stability protein. Both ensemble and thresh-
91 old epistasis arise because the protein can populate more than
92 one conformation; however, at this point, the two mechanisms
93 for epistasis diverge.

94 To make this concrete, consider the activity of an enzyme. En-
95 zyme activity is proportional to the fraction of enzyme molecules
96 that are in the active form. Mutations that have an additive, lin-
97 ear effect on thermodynamic stability will have a non-additive,
98 nonlinear effect on the fractional population of the active form
99 (equation 6). As such, we can observe epistasis between muta-
100 tions at the level of enzyme activity simply because we are de-
101 scribing a nonlinear function (activity) with a linear model (equa-
102 tion 16) Sailer and Harms (2017a); Otwinowski et al. (2018). If we
103 transform the nonlinear fractional population scale (equation 6)
104 onto a linear free energy scale (equation 8), threshold epistasis
105 disappears. One can describe the non-additive, nonlinear effects
106 of mutations on activity as additive, linear effects on stability.
107 This is not to say threshold epistasis does not matter—phenotype
108 and fitness often depend on nonlinear fractional populations—
109 but rather that it is possible to analyze the data in a way that
110 removes epistasis.

111 Ensemble epistasis, however, cannot be removed by trans-
112 forming the data onto a linear scale. We describe the observable
113 (ΔG_{obs}) and the effects of mutations ($\delta G_c^{x \rightarrow X}$) on the same lin-
114 ear free energy scale. But because mutations have different
115 effects on different conformations, these linear perturbations are
116 re-weighted in nonlinear fashion, thus leading to irreducible
117 epistasis.

Ensemble epistasis may shape evolution

Though it remains to be seen, we expect that ensemble epistasis plays an important role in shaping protein evolution. We have shown that simple ensembles give rise to magnitude, sign, and reciprocal sign epistasis (Fig 4), and that they may give rise to high-order epistasis (supplemental text, section 3). Sign and reciprocal sign epistasis are particularly important; they can decrease accessible evolutionary trajectories and are required for the presence of multiple peaks in fitness landscapes Weinreich *et al.* (2005); Poelwijk *et al.* (2007, 2011); Weinreich (2006); Lunzer *et al.* (2005); Bridgham *et al.* (2006); Kvittek and Sherlock (2011); Chiotti *et al.* (2014); Salverda *et al.* (2011); Palmer *et al.* (2015). High-order epistasis can alter accessibility and can facilitate the bypassing of evolutionary dead-ends in genotype-phenotype maps, making evolution deeply unpredictable Sailer and Harms (2017b); Weinreich *et al.* (2013); Palmer *et al.* (2015); Wu *et al.* (2016).

Aside from giving rise to evolutionarily-relevant classes of epistasis, we anticipate that ensemble epistasis occurs under physiologically relevant—and thus evolutionarily important—conditions. Ensemble epistasis is maximized when multiple conformations are populated (Fig 6A): exactly within the concentration regime where macromolecules act as molecular switches. Further, we found in our S100A4 calculations that we could see changes in the type of epistasis observed as we changed the amount of allosteric effector, $\mu_{Ca^{2+}}$ (Fig 6A). This suggests that ensemble epistasis could play a critical role in shaping the availability of evolutionary trajectories—possibly even in an environment-dependent manner. A small change in the concentration of an effector could open or close new evolutionary trajectories. A similar phenomenon has been observed in allosteric proteins where ligands can act as agonists or antagonists in response to changes in environment, ultimately via changes in the thermodynamic ensemble Motlagh and Hilser (2012).

Detecting ensemble epistasis

Our work predicts ensemble epistasis is common. How would one detect it experimentally? Effector- or environment-dependent epistasis may be a signal of ensemble epistasis. One straightforward experimental test for ensemble epistasis would be to perturb the thermodynamic ensemble by tuning environmental factors such as effector concentration (Fig 5B). For S100A4, we observed distinct effector-dependent patterns of epistasis for mutation pairs, where the amount of epistasis we observed changed with the addition of Ca^{2+} (Fig 6A). Ensemble epistasis should be maximized at concentrations where many distinct conformations are populated (i.e. at concentrations where functional transitions occur) and minimized when mutations can impact only a single conformation. (i.e. low $\mu_{Ca^{2+}}$). Environmental-dependent epistasis has been noted previously, possibly pointing to an underlying ensemble epistasis Barker *et al.* (2015); Samir *et al.* (2015); Joshi and Prasad (2014); Remold and Lenski (2004); Guerrero *et al.* (2019); Flynn *et al.* (2013); Chiotti *et al.* (2014); Nosil *et al.* (2020).

Additionally, one might test for ensemble epistasis by measuring the temperature dependence of epistasis. If the free energy of each conformation does not change with temperature, the predictions are straightforward. For very low temperatures, only the deepest energy well—corresponding to the most stable conformation—should be populated, preventing ensemble epistasis. At very high temperature, all conformations will have the same statistical weight, and thus will be equally populated regardless of free energy (Equation 6). But, because of this fact, mutations will not redistribute the populations of the conformations—meaning there will be no ensemble epistasis. For intermediate temperature values, we might expect appreciable temperature-dependent effects on ensemble epistasis. Unfortunately, the free energy of each conformation is not constant with temperature for most proteins Dill (1990). As such, we would expect the effects of ensemble epistasis are convolved with changes in the enthalpy and entropy of each conformation—making temperature-dependent experiments difficult to interpret.

Conclusion

Our results reveal that a universal property of proteins and other macromolecules, the thermodynamic ensemble, can lead to epistasis. While the pervasiveness of ensemble epistasis in biology remains unknown, we anticipate that it is widespread. First, ensemble epistasis is maximized under the physiological conditions where biologically important, ensemble-mediated functions occur. Second, even a simple, three-conformation system can lead to a rich variety of epistasis, suggesting that the necessary conditions for ensemble epistasis are met for many proteins. And, third, structure-based calculations using experimentally solved protein structures revealed the potential for rampant ensemble epistasis. As such, we anticipate that ensemble epistasis plays important roles in shaping protein biology and evolution.

Data Availability

Supplemental files available at FigShare. The file "Supplemental derivations and proofs" has all referenced derivations and proofs in the text. Fig S1 demonstrates that our epistatic analysis of human S100A4 is not sensitive to our assumptions about the affinity of the protein for calcium. Supplemental Material available at figshare: <https://doi.org/10.25386/genetics.14377394>.

Table 1 Map between genotype and the thermodynamic description of $\Delta G_{obs}^{genotype}$.

Genotype	$\Delta G_{obs}^{genotype}$	$\langle G_{j,k}^{genotype} \rangle$
<i>ab</i>	$G_i^{ab} - \langle G_{j,k}^{ab} \rangle$	$-RT \ln \left(e^{-G_i^{ab}/RT} + e^{-G_k^{ab}/RT} \right)$
<i>Ab</i>	$(G_i^{ab} + \delta G_i^{a \rightarrow A}) - \langle G_{j,k}^{Ab} \rangle$	$-RT \ln \left(e^{-(G_j^{ab} + \delta G_j^{a \rightarrow A})/RT} + e^{-(G_k^{ab} + \delta G_k^{a \rightarrow A})/RT} \right)$
<i>aB</i>	$(G_i^{ab} + \delta G_i^{b \rightarrow B}) - \langle G_{j,k}^{aB} \rangle$	$-RT \ln \left(e^{-(G_j^{ab} + \delta G_j^{b \rightarrow B})/RT} + e^{-(G_k^{ab} + \delta G_k^{b \rightarrow B})/RT} \right)$
<i>AB</i>	$(G_i^{ab} + \delta G_i^{a \rightarrow A} + \delta G_i^{b \rightarrow B}) - \langle G_{j,k}^{AB} \rangle$	$-RT \ln \left(e^{-(G_j^{ab} + \delta G_j^{a \rightarrow A} + \delta G_j^{b \rightarrow B})/RT} + e^{-(G_k^{ab} + \delta G_k^{a \rightarrow A} + \delta G_k^{b \rightarrow B})/RT} \right)$

1 All analyses and ROSETTA input files can be downloaded di-
2 rectly from https://github.com/harmslab/ensemble_epistasis.

3 Acknowledgements

4 We thank everyone in the laboratory of M.J.H. for help develop-
5 ing theses ideas. We also want to thank David Pollock, Daniel
6 Herschlag, and Doug Barrick for helpful comments on earlier
7 drafts of the manuscript.

8 Funding

9 This work was supported by funds from the National Institutes
10 of Health NIH 7T32GM007759 (to A.J.M.) and National Science
11 Foundation CAREER Award DEB-1844963 (to M.J.H.).

12 Competing Interests

13 None declared.

14 Literature Cited

15 Alexander, P. A., Y. He, Y. Chen, J. Orban, and P. N. Bryan, 2009
16 A minimal sequence code for switching protein structure and
17 function. *Proceedings of the National Academy of Sciences of*
18 *the United States of America* **106**: 21149–21154.
19 Alford, R. F., A. Leaver-Fay, J. R. Jeliazkov, M. J. O’Meara, F. P.
20 DiMaio, *et al.*, 2017 The Rosetta All-Atom Energy Function for
21 Macromolecular Modeling and Design. *Journal of Chemical*
22 *Theory and Computation* **13**: 3031–3048.
23 Ancel, L. W. and W. Fontana, 2000 Plasticity, evolvability, and
24 modularity in RNA. *Journal of Experimental Zoology* **288**:
25 242–283.
26 Baier, F., N. Hong, G. Yang, A. Pabis, C. M. Miton, *et al.*, 2019
27 Cryptic genetic variation shapes the adaptive evolutionary
28 potential of enzymes. *eLife* **8**: e40789.
29 Barker, B., L. Xu, and Z. Gu, 2015 Dynamic epistasis under
30 varying environmental perturbations. *PloS One* **10**: e0114911.
31 Bershtein, S., M. Segal, R. Bekerman, N. Tokuriki, and D. S.
32 Tawfik, 2006 Robustness–epistasis link shapes the fitness land-
33 scape of a randomly drifting protein. *Nature* **444**: 929–932.
34 Bessonnard, S., L. De Mot, D. Gonze, M. Barriol, C. Dennis,
35 *et al.*, 2014 Gata6, Nanog and Erk signaling control cell fate
36 in the inner cell mass through a tristable regulatory network.
37 *Development* **141**: 3637–3648.
38 Bloom, J. D., F. H. Arnold, and C. O. Wilke, 2007 Breaking pro-
39 teins with mutations: threads and thresholds in evolution.
40 *Molecular Systems Biology* **3**: 76.
41 Bridgham, J. T., S. M. Carroll, and J. W. Thornton, 2006 Evolution
42 of Hormone-Receptor Complexity by Molecular Exploitation.
43 *Science* **312**: 97–101.

Chiotti, K. E., D. J. Kvitek, K. H. Schmidt, G. Koniges, 44
K. Schwartz, *et al.*, 2014 The Valley-of-Death: Reciprocal sign 45
epistasis constrains adaptive trajectories in a constant, nutrient 46
limiting environment. *Genomics* **104**: 431–437. 47
Dellus-Gur, E., M. Elias, E. Caselli, F. Prati, M. L. M. Salverda, 48
et al., 2015 Negative Epistasis and Evolvability in TEM-1 β - 49
Lactamase—The Thin Line between an Enzyme’s Conforma- 50
tional Freedom and Disorder. *Journal of Molecular Biology* 51
427: 2396–2409. 52
Dill, K. A., 1990 Dominant forces in protein folding. *Biochemistry* 53
29: 7133–7155, eprint: <https://doi.org/10.1021/bi00483a001>. 54
Ecsédi, P., B. Kiss, G. Gógl, L. Radnai, L. Buday, *et al.*, 2017 55
Regulation of the Equilibrium between Closed and Open Con- 56
formations of Annexin A2 by N-Terminal Phosphorylation 57
and S100A4-Binding. *Structure* **25**: 1195–1207.e5. 58
Flynn, K. M., T. F. Cooper, F. B.-G. Moore, and V. S. Cooper, 2013 59
The Environment Affects Epistatic Interactions to Alter the 60
Topology of an Empirical Fitness Landscape. *PLOS Genetics* 61
9: e1003426. 62
Garrett, S. C., L. Hodgson, A. Rybin, A. Touthkine, K. M. Hahn, 63
et al., 2008 A Biosensor of S100A4 Metastasis Factor Activa- 64
tion: Inhibitor Screening and Cellular Activation Dynamics. 65
Biochemistry **47**: 986–996. 66
Giger, L., S. Caner, R. Obexer, P. Kast, D. Baker, *et al.*, 2013 67
Evolution of a designed retro-aldolase leads to complete active 68
site remodeling. *Nature Chemical Biology* **9**: 494–498. 69
Gong, L. I., M. A. Suchard, and J. D. Bloom, 2013 Stability- 70
mediated epistasis constrains the evolution of an influenza 71
protein. *eLife* **2**: e00631. 72
Guerrero, R. F., S. V. Scarpino, J. V. Rodrigues, D. L. Hartl, and 73
C. B. Ogbunugafor, 2019 Proteostasis Environment Shapes 74
Higher-Order Epistasis Operating on Antibiotic Resistance. 75
Genetics **212**: 565–575. 76
Hameri, T., M.-O. Boldi, and V. Hatzimanikatis, 2019 Statistical 77
inference in ensemble modeling of cellular metabolism. *PLOS* 78
Computational Biology **15**: e1007536. 79
Joshi, C. J. and A. Prasad, 2014 Epistatic interactions among 80
metabolic genes depend upon environmental conditions. 81
Molecular BioSystems **10**: 2578–2589. 82
Khazaei, T., A. P. McGuigan, and R. Mahadevan, 2012 Ensemble 83
Modeling of Cancer Metabolism. *Frontiers in Physiology* **3**. 84
Kumar, A., C. Natarajan, H. Moriyama, C. C. Witt, R. E. Weber, 85
et al., 2017 Stability-Mediated Epistasis Restricts Accessible 86
Mutational Pathways in the Functional Evolution of Avian 87
Hemoglobin. *Molecular Biology and Evolution* **34**: 1240–1251. 88
Kvitek, D. J. and G. Sherlock, 2011 Reciprocal Sign Epistasis 89
Frequently Experimentally Evolved Adaptive Muta- 90
tions Causes a Rugged Fitness Landscape. *PLoS Genetics* **7**: 91
e1002056. 92

- 1 Lu, M., M. K. Jolly, R. Gomoto, B. Huang, J. Onuchic, *et al.*,
2 2013 Tristability in Cancer-Associated MicroRNA-TF Chimera
3 Toggle Switch. *The Journal of Physical Chemistry B* **117**: 13164–
4 13174.
- 5 Lunzer, M., S. P. Miller, R. Felsheim, and A. M. Dean, 2005 The
6 Biochemical Architecture of an Ancient Adaptive Landscape.
7 *Science* **310**: 499–501.
- 8 Maisnier-Patin, S., W. Paulander, A. Pennhag, and D. I. Anders-
9 son, 2007 Compensatory evolution reveals functional interac-
10 tions between ribosomal proteins S12, L14 and L19. *Journal of*
11 *Molecular Biology* **366**: 207–215.
- 12 Malashkevich, V. N., K. M. Varney, S. C. Garrett, P. T. Wilder,
13 D. Knight, *et al.*, 2008 Structure of Ca²⁺-Bound S100A4 and Its
14 Interaction with Peptides Derived from Nonmuscle Myosin-
15 IIA. *Biochemistry* **47**: 5111–5126.
- 16 Miton, C. M. and N. Tokuriki, 2016 How mutational epistasis
17 impairs predictability in protein evolution and design. *Protein*
18 *Science* **25**: 1260–1272.
- 19 Motlagh, H. N. and V. J. Hilser, 2012 Agonism/antagonism
20 switching in allosteric ensembles. *Proceedings of the National*
21 *Academy of Sciences of the United States of America* **109**:
22 4134–4139.
- 23 Motlagh, H. N., J. O. Wrabl, J. Li, and V. J. Hilser, 2014 The
24 ensemble nature of allostery. *Nature* **508**: 331–339.
- 25 Nosil, P., R. Villoutreix, C. F. de Carvalho, J. L. Feder, T. L.
26 Parchman, *et al.*, 2020 Ecology shapes epistasis in a geno-
27 type–phenotype–fitness map for stick insect colour. *Nature*
28 *Ecology & Evolution*.
- 29 Ortlund, E. A., J. T. Bridgham, M. R. Redinbo, and J. W. Thornton,
30 2007 Crystal structure of an ancient protein: evolution by
31 conformational epistasis. *Science (New York, N.Y.)* **317**: 1544–
32 1548.
- 33 Otwinowski, J., 2018 Biophysical Inference of Epistasis and the
34 Effects of Mutations on Protein Stability and Function. *Molec-
35 ular Biology and Evolution* **35**: 2345–2354.
- 36 Otwinowski, J., D. M. McCandlish, and J. B. Plotkin, 2018 Infer-
37 ring the shape of global epistasis. *Proceedings of the National*
38 *Academy of Sciences* **115**: E7550–E7558.
- 39 Palmer, A. C., E. Toprak, M. Baym, S. Kim, A. Veres, *et al.*, 2015
40 Delayed commitment to evolutionary fate in antibiotic resis-
41 tance fitness landscapes. *Nature Communications* **6**: 7385.
- 42 Park, H., P. Bradley, P. Greisen, Y. Liu, V. K. Mulligan, *et al.*, 2016
43 Simultaneous Optimization of Biomolecular Energy Functions
44 on Features from Small Molecules and Macromolecules. *Jour-
45 nal of Chemical Theory and Computation* **12**: 6201–6212.
- 46 Petrović, D., V. A. Risso, S. C. L. Kamerlin, and J. M. Sanchez-
47 Ruiz, 2018 Conformational dynamics and enzyme evolution.
48 *Journal of The Royal Society Interface* **15**: 20180330.
- 49 Poelwijk, F. J., D. J. Kiviet, D. M. Weinreich, and S. J. Tans, 2007
50 Empirical fitness landscapes reveal accessible evolutionary
51 paths. *Nature* **445**: 383–386.
- 52 Poelwijk, F. J., S. Tănase-Nicola, D. J. Kiviet, and S. J. Tans, 2011
53 Reciprocal sign epistasis is a necessary condition for multi-
54 peaked fitness landscapes. *Journal of Theoretical Biology* **272**:
55 141–144.
- 56 Remold, S. K. and R. E. Lenski, 2004 Pervasive joint influence
57 of epistasis and plasticity on mutational effects in *Escherichia*
58 *coli*. *Nature Genetics* **36**: 423–426.
- 59 Sailer, Z. R. and M. J. Harms, 2017a Detecting High-Order Epis-
60 tasis in Nonlinear Genotype-Phenotype Maps. *Genetics* **205**:
61 1079–1088.
- 62 Sailer, Z. R. and M. J. Harms, 2017b High-order epistasis shapes
evolutionary trajectories. *PLOS Computational Biology* **13**:
e1005541.
- Sailer, Z. R. and M. J. Harms, 2017c Molecular ensembles
make evolution unpredictable. *Proceedings of the National*
Academy of Sciences **114**: 11938–11943.
- Salverda, M. L. M., E. Dellus, F. A. Gorter, A. J. M. Debets,
J. van der Oost, *et al.*, 2011 Initial Mutations Direct Alternative
Pathways of Protein Evolution. *PLoS Genetics* **7**: e1001321.
- Samir, P., Rahul, J. C. Slaughter, and A. J. Link, 2015 Environmen-
tal Interactions and Epistasis Are Revealed in the Proteomic
Responses to Complex Stimuli. *PLOS ONE* **10**: e0134099.
- Seeliger, M. A., B. Nagar, F. Frank, X. Cao, M. N. Henderson, *et al.*,
2007 c-Src Binds to the Cancer Drug Imatinib with an Inactive
Abl/c-Kit Conformation and a Distributed Thermodynamic
Penalty. *Structure* **15**: 299–311.
- Sykora, J., J. Brezovsky, T. Koudelakova, M. Lahoda, A. Fortova,
et al., 2014 Dynamics and hydration explain failed functional
transformation in dehalogenase design. *Nature Chemical Bi-
ology* **10**: 428–430.
- Tran, L. M., M. L. Rizk, and J. C. Liao, 2008 Ensemble Modeling
of Metabolic Networks. *Biophysical Journal* **95**: 5606–5617.
- Tsai, C.-J. and R. Nussinov, 2014 A Unified View of “How Al-
lostery Works”. *PLOS Computational Biology* **10**: e1003394.
- Vallely, K. M., R. R. Rustandi, K. C. Ellis, O. Varlamova, A. R.
Bresnick, *et al.*, 2002 Solution Structure of Human Mts1
(S100A4) As Determined by NMR Spectroscopy. *Biochemistry*
41: 12670–12680, [_eprint: https://doi.org/10.1021/bi020365r](https://doi.org/10.1021/bi020365r).
- Venturi, V., Á. Kerényi, B. Reiz, D. Bihary, and S. Pongor, 2010
Locality versus globality in bacterial signalling: can local com-
munication stabilize bacterial communities? *Biology Direct* **5**:
30.
- Wei, G., W. Xi, R. Nussinov, and B. Ma, 2016 Protein Ensembles:
How Does Nature Harness Thermodynamic Fluctuations for
Life? The Diverse Functional Roles of Conformational Ensem-
bles in the Cell. *Chemical Reviews* **116**: 6516–6551.
- Weinreich, D. M., 2006 Darwinian Evolution Can Follow Only
Very Few Mutational Paths to Fitter Proteins. *Science* **312**: 111–
114.
- Weinreich, D. M., Y. Lan, C. S. Wylie, and R. B. Heckendorn, 2013
Should evolutionary geneticists worry about higher-order
epistasis? *Current opinion in genetics & development* **23**:
700–707.
- Weinreich, D. M., R. A. Watson, and L. Chao, 2005 Perspective:
Sign Epistasis and Genetic Constraint on Evolutionary Trajecto-
ries. *Evolution* **59**: 1165–1174.
- Wilson, C., R. V. Agafonov, M. Hoemberger, S. Kutter, A. Zorba,
et al., 2015 Using ancient protein kinases to unravel a modern
cancer drug’s mechanism. *Science* **347**: 882–886.
- Wu, N. C., L. Dai, C. A. Olson, J. O. Lloyd-Smith, and R. Sun,
2016 Adaptation in protein fitness landscapes is facilitated by
indirect paths. *eLife* **5**: e16965.
- Yang, G., D. W. Anderson, F. Baier, E. Dohmen, N. Hong, *et al.*,
2019 Higher-order epistasis shapes the fitness landscape of a
xenobiotic-degrading enzyme. *Nature Chemical Biology* **15**:
1120–1128.
- Yokoyama, S., J. Xing, Y. Liu, D. Faggionato, A. Altun, *et al.*, 2014
Epistatic Adaptive Evolution of Human Color Vision. *PLOS*
Genetics **10**: e1004884.