

# Supplement: Ensemble epistasis: thermodynamic origins of non-additivity between mutations

*Anneliese J. Morrison<sup>\*†</sup>, Daria R. Wonderlick<sup>\*†</sup>, and Michael J. Harms<sup>\*†,1</sup>*

<sup>1</sup>Corresponding author : [harms@uoregon.edu](mailto:harms@uoregon.edu)

<sup>\*</sup>Department of Chemistry and Biochemistry, University of Oregon, Eugene, OR 974032.

<sup>†</sup>Institute for Molecular Biology, University of Oregon, Eugene, OR 97403.

# 1 Necessary conditions for ensemble epistasis

## 1.1 Ensemble epistasis appears between two mutations in a three-conformation ensemble

We define epistasis between mutations  $a \rightarrow A$  and  $b \rightarrow B$  in  $\Delta G_{obs}$  as the difference in the effect of  $a \rightarrow A$  in the  $ab$  and  $aB$  backgrounds (Fig 1A):

$$\varepsilon = \left( \Delta G_{obs}^{AB} - \Delta G_{obs}^{aB} \right) - \left( \Delta G_{obs}^{Ab} - \Delta G_{obs}^{ab} \right). \quad (20)$$

$\Delta G_{obs}^{genotype}$  is given by Equation 9:

$$\Delta G_{obs}^{genotype} = G_i^{genotype} - \langle G_{j,k}^{genotype} \rangle \quad (21)$$

where

$$\langle G_{j,k}^{genotype} \rangle = -RT \ln \left( e^{-G_j^{genotype}/RT} + e^{-G_k^{genotype}/RT} \right). \quad (22)$$

We model mutations as having additive effects within each conformation  $i$ ,  $j$ , or  $k$ .  $\Delta G_{obs}$  for each genotype is shown below (reproducing Table 1 in the main text):

**Table 2** Map between genotype and the thermodynamic description of  $\Delta G_{obs}^{genotype}$ .

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

If we substitute the relevant expressions for  $\Delta G_{obs}^{genotype}$  into our expression for  $\varepsilon$  (20), we get:

$$\varepsilon = \left( \left[ G_i^{ab} + \Delta G_{A,i}^{ab} + \Delta G_{B,i}^{ab} - \langle G_{j,k}^{AB} \rangle \right] - \left[ \Delta G_i^{ab} + \Delta G_{B,i}^{ab} - \langle G_{j,k}^{aB} \rangle \right] \right) - \left( \left[ \Delta G_i^{ab} + \Delta G_{A,i}^{ab} - \langle G_{j,k}^{Ab} \rangle \right] - \left[ \Delta G_i^{ab} - \langle G_{j,k}^{ab} \rangle \right] \right) \quad (23)$$

$$\varepsilon = \Delta G_i^{ab} + \Delta G_{A,i}^{ab} + \Delta G_{B,i}^{ab} - \langle G_{j,k}^{AB} \rangle - \Delta G_i^{ab} - \Delta G_{B,i}^{ab} + \langle G_{j,k}^{aB} \rangle - \Delta G_i^{ab} - \Delta G_{A,i}^{ab} + \langle G_{j,k}^{Ab} \rangle + \Delta G_i^{ab} - \langle G_{j,k}^{ab} \rangle \quad (24)$$

$$\varepsilon = - \langle G_{j,k}^{AB} \rangle + \langle G_{j,k}^{aB} \rangle + \langle G_{j,k}^{Ab} \rangle - \langle G_{j,k}^{ab} \rangle \quad (25)$$

$$\varepsilon = - \left[ \left( \langle G_{j,k}^{AB} \rangle - \langle G_{j,k}^{aB} \rangle \right) - \left( \langle G_{j,k}^{Ab} \rangle - \langle G_{j,k}^{ab} \rangle \right) \right] \quad (26)$$

This cannot be simplified further, implying that  $\varepsilon$  may be non-zero.

1 **1.2 To see ensemble epistasis, it is necessary to have three or more conformations**

2 We can next consider the two-conformation case, where  $k$  is not populated. In this case:

$$\Delta G_{obs}^{genotype} = G_i^{genotype} - \langle G_j^{genotype} \rangle \langle G_j^{genotype} \rangle \quad (27)$$

3 simplifies to:

$$\langle G_j^{genotype} \rangle = -RT \ln \left( e^{-G_j^{genotype}/RT} \right) = G_j^{genotype}. \quad (28)$$

4 As in Section 1.1, we can write a table showing  $\Delta G_{obs}$  for each genotype:

**Table 3 Map between genotype and the thermodynamic description of  $\Delta G_{obs}^{genotype}$  for a two-conformation ensemble.**

Genotype	$\Delta G_{obs}^{genotype}$
$ab$	$G_i^{ab} - G_j^{ab}$
$Ab$	$G_i^{ab} + \Delta G_{A,i}^{ab} - (G_j^{ab} + \Delta G_{A,j}^{ab})$
$aB$	$G_i^{ab} + \Delta G_{B,i}^{ab} - (G_j^{ab} + \Delta G_{B,j}^{ab})$
$AB$	$G_i^{ab} + \Delta G_{A,i}^{ab} + \Delta G_{B,i}^{ab} - (G_j^{ab} + \Delta G_{A,j}^{ab} + \Delta G_{B,j}^{ab})$

5 If we substitute the relevant expressions for  $\Delta G_{obs}^{genotype}$  into our expression for  $\epsilon$ , we get

$$\begin{aligned} \epsilon = & \left( \left[ G_i^{ab} + \Delta G_{A,i}^{ab} + \Delta G_{B,i}^{ab} - (G_j^{ab} + \Delta G_{A,j}^{ab} + \Delta G_{B,j}^{ab}) \right] - \left[ G_i^{ab} + \Delta G_{B,i}^{ab} - (G_j^{ab} + \Delta G_{B,j}^{ab}) \right] \right) \\ & - \left( \left[ G_i^{ab} + \Delta G_{A,i}^{ab} - (G_j^{ab} + \Delta G_{A,j}^{ab}) \right] - \left[ G_i^{ab} - G_j^{ab} \right] \right) \end{aligned} \quad (29)$$

$$\begin{aligned} \epsilon = & G_i^{ab} + \Delta G_{A,i}^{ab} + \Delta G_{B,i}^{ab} - G_j^{ab} - \Delta G_{A,j}^{ab} - \Delta G_{B,j}^{ab} - G_i^{ab} - \Delta G_{B,i}^{ab} + G_j^{ab} + \Delta G_{B,j}^{ab} \\ & - G_i^{ab} - \Delta G_{A,i}^{ab} + G_j^{ab} + \Delta G_{A,j}^{ab} + G_i^{ab} - G_j^{ab} \end{aligned} \quad (30)$$

6 All terms cancel, demonstrating it is necessary to have at least three conformations to observe ensemble epistasis.

7 **1.3 To see ensemble epistasis, it is necessary for mutations  $a \rightarrow A$  and  $b \rightarrow B$  to have different effects on conformations  $j$  and  $k$ .**

8 To test the necessity of mutations having differential effects on conformations  $j$  and  $k$ , we set  $\Delta G_{B,j} = \Delta G_{B,k} = \Delta G_{B,jk}^{ab}$ . This means  
9 mutation  $b \rightarrow B$  has the same effect on conformations  $j$  and  $k$ . In contrast, we left  $\Delta G_{A,j} \neq \Delta G_{A,k}$ , meaning  $a \rightarrow A$  has different effects on  
10 conformations  $j$  and  $k$ . Because  $b \rightarrow B$  has identical effects and  $a \rightarrow A$  has differential effects, this analysis tests whether it is necessary  
11 for both mutations to have differential effects to observe ensemble epistasis.

12 Consider the expression for  $\langle G_{j,k}^{aB} \rangle$ :

$$\langle G_{j,k}^{aB} \rangle = -RT \ln \left( e^{-(G_j^{ab} + \Delta G_{B,jk}^{ab})/RT} + e^{-(G_k^{ab} + \Delta G_{B,jk}^{ab})/RT} \right) \quad (31)$$

13 Because  $\Delta G_{B,jk}^{ab}$  is shared among terms, we can factor it out:

$$\langle G_{j,k}^{aB} \rangle = -RT \ln \left( e^{-G_j^{ab}/RT} e^{-\Delta G_{B,jk}^{ab}/RT} + e^{-G_k^{ab}/RT} e^{-\Delta G_{B,jk}^{ab}/RT} \right) \quad (32)$$

$$\langle \Delta G_{j,k}^{aB} \rangle = -RT \ln \left( e^{-\Delta G_{B,jk}^{ab}/RT} \left( e^{-G_j^{ab}/RT} + e^{-G_k^{ab}/RT} \right) \right) \quad (33)$$

$$\langle \Delta G_{j,k}^{aB} \rangle = -RT \ln \left( e^{-\Delta G_{B,jk}^{ab}/RT} \right) - RT \ln \left( e^{-G_j^{ab}/RT} + e^{-G_k^{ab}/RT} \right) \quad (34)$$

$$\langle \Delta G_{j,k}^{aB} \rangle = \Delta G_{B,jk}^{ab} - RT \ln \left( e^{-G_j^{ab}/RT} + e^{-G_k^{ab}/RT} \right) \quad (35)$$

$$\langle \Delta G_{j,k}^{aB} \rangle = \Delta G_{B,jk}^{ab} + \langle G_{j,k}^{ab} \rangle. \quad (36)$$

14 Using the same reasoning, we can factor  $\Delta G_{B,jk}^{ab}$  out of the expression for  $\langle \Delta G_{j,k}^{AB} \rangle$ :

$$\langle \Delta G_{j,k}^{AB} \rangle = \Delta G_{B,jk}^{ab} + \langle G_{j,k}^{Ab} \rangle. \quad (37)$$

We can then substitute these simplified expressions for  $\langle \Delta G_{j,k}^{aB} \rangle$  and  $\langle \Delta G_{j,k}^{AB} \rangle$  into the expression for  $\varepsilon$ :

$$\varepsilon = \left( \langle G_{j,k}^{aB} \rangle - \langle G_{j,k}^{AB} \rangle \right) - \left( \langle G_{j,k}^{ab} \rangle - \langle G_{j,k}^{Ab} \rangle \right) \quad (38)$$

$$\varepsilon = \left( \left[ \Delta G_{B,jk}^{ab} + \langle G_{j,k}^{ab} \rangle \right] - \left[ \Delta G_{B,jk}^{ab} + \langle G_{j,k}^{Ab} \rangle \right] \right) - \left( \langle G_{j,k}^{ab} \rangle - \langle G_{j,k}^{Ab} \rangle \right) \quad (39)$$

$$\varepsilon = \Delta G_{B,jk}^{ab} + \langle G_{j,k}^{ab} \rangle - \Delta G_{B,jk}^{ab} - \langle G_{j,k}^{Ab} \rangle - \langle G_{j,k}^{ab} \rangle + \langle G_{j,k}^{Ab} \rangle \quad (40)$$

$$\varepsilon = 0. \quad (41)$$

All terms cancel, demonstrating that it is necessary for both  $a \rightarrow A$  and  $b \rightarrow B$  to have differential effects on conformations  $j$  and  $k$  to observe ensemble epistasis.

## 2 Ensembles can lead to high-order epistasis

In this section, we will investigate potential ensemble epistasis between three mutations. For this, we use a perturbative genetic model that describes the effects of mutations and combinations of mutations as perturbations away from phenotype of the reference genotype  $abc$  Weinreich *et al.* (2013); Sailer and Harms (2017a). The individual effects of mutations are described by  $\beta_A$ ,  $\beta_B$ , and  $\beta_C$ . Pairwise interactions between mutations are given by  $\varepsilon_{AB}$ ,  $\varepsilon_{AC}$ , and  $\varepsilon_{BC}$ . The three-way interaction is given by  $\varepsilon_{ABC}$ . In this formulation,  $\varepsilon_{ABC}$  captures the quantitative change in phenotype for mutating  $abc \rightarrow ABC$  that is not captured by  $\beta_A + \beta_B + \beta_C + \varepsilon_{AB} + \varepsilon_{AC} + \varepsilon_{BC}$ . We can write the genetic and thermodynamic descriptions of each genotype in a table:

**Table 4** Map between genotype and the thermodynamic description of  $\Delta G_{obs}^{genotype}$  for high-order epistatic interactions.

genotype	genetic model	$\Delta G_{obs}^{genotype}$
$abc$	$\alpha_{abc}$	$G_i^{abc} - \langle G_{j,k}^{abc} \rangle$
$Abc$	$\alpha_{abc} + \beta_A$	$G_i^{abc} + \Delta G_{A,i}^{abc} - \langle G_{j,k}^{Abc} \rangle$
$aBc$	$\alpha_{abc} + \beta_B$	$G_i^{abc} + \Delta G_{B,i}^{abc} - \langle G_{j,k}^{aBc} \rangle$
$abC$	$\alpha_{abc} + \beta_C$	$G_i^{abc} + \Delta G_{C,i}^{abc} - \langle G_{j,k}^{abC} \rangle$
$ABc$	$\alpha_{abc} + \beta_A + \beta_B + \varepsilon_{AB}$	$G_i^{abc} + \Delta G_{A,i}^{abc} + \Delta G_{B,i}^{abc} - \langle G_{j,k}^{ABc} \rangle$
$AbC$	$\alpha_{abc} + \beta_A + \beta_C + \varepsilon_{AC}$	$G_i^{abc} + \Delta G_{A,i}^{abc} + \Delta G_{C,i}^{abc} - \langle G_{j,k}^{AbC} \rangle$
$aBC$	$\alpha_{abc} + \beta_B + \beta_C + \varepsilon_{BC}$	$G_i^{abc} + \Delta G_{B,i}^{abc} + \Delta G_{C,i}^{abc} - \langle G_{j,k}^{aBC} \rangle$
$ABC$	$\alpha_{abc} + \beta_A + \beta_B + \beta_C + \varepsilon_{AB} + \varepsilon_{AC} + \varepsilon_{BC} + \varepsilon_{ABC}$	$G_i^{abc} + \Delta G_{A,i}^{abc} + \Delta G_{B,i}^{abc} + \Delta G_{C,i}^{abc} - \langle G_{j,k}^{ABC} \rangle$

As in the main text and sections above,  $\langle G_{j,k}^{genotype} \rangle$  is the Boltzmann-weighted average of the energies of all conformations. For the  $ABC$  genotype, for example:

$$\langle G_{j,k}^{ABC} \rangle = -RT \ln \left( e^{-\left( G_j^{abc} + \Delta G_{A,j}^{ab} + \Delta G_{B,j}^{ab} + \Delta G_{C,j}^{ab} \right) / RT} + e^{-\left( G_k^{abc} + \Delta G_{A,k}^{ab} + \Delta G_{B,k}^{ab} + \Delta G_{C,k}^{ab} \right) / RT} \right). \quad (42)$$

Using the map between the genetic model and  $\Delta G_{obs}$  for each genotype, we can algebraically solve for each coefficient in the genetic model in thermodynamic terms:

$$\alpha_{abc} = G_i^{abc} - \langle G_{j,k}^{abc} \rangle \quad (43)$$

$$\beta_A = \Delta G_{A,i}^{abc} - \left( \langle G_{j,k}^{Abc} \rangle - \langle G_{j,k}^{abc} \rangle \right) \quad (44)$$

$$\beta_B = \Delta G_{B,i}^{abc} - \left( \langle G_{j,k}^{aBc} \rangle - \langle G_{j,k}^{abc} \rangle \right) \quad (45)$$

$$\beta_C = \Delta G_{C,i}^{abc} - \left( \langle G_{j,k}^{abC} \rangle - \langle G_{j,k}^{abc} \rangle \right) \quad (46)$$

$$\varepsilon_{AB} = - \left( \langle G_{j,k}^{ABc} \rangle - \langle G_{j,k}^{Abc} \rangle \right) + \left( \langle G_{j,k}^{aBc} \rangle - \langle G_{j,k}^{abc} \rangle \right) \quad (47)$$

$$\varepsilon_{AC} = - \left( \langle G_{j,k}^{AbC} \rangle - \langle G_{j,k}^{Abc} \rangle \right) + \left( \langle G_{j,k}^{abC} \rangle - \langle G_{j,k}^{abc} \rangle \right) \quad (48)$$

$$\varepsilon_{BC} = - \left( \langle G_{j,k}^{aBC} \rangle - \langle G_{j,k}^{aBc} \rangle \right) + \left( \langle G_{j,k}^{abC} \rangle - \langle G_{j,k}^{abc} \rangle \right) \quad (49)$$

$$\varepsilon_{ABC} = - \left( \left( \langle G_{j,k}^{ABC} \rangle - \langle G_{j,k}^{ABc} \rangle - \langle G_{j,k}^{AbC} \rangle - \langle G_{j,k}^{aBC} \rangle \right) + \left( \langle G_{j,k}^{Abc} \rangle + \langle G_{j,k}^{aBc} \rangle + \langle G_{j,k}^{abC} \rangle - \langle G_{j,k}^{abc} \rangle \right) \right) \quad (50)$$

As expected,  $\varepsilon_{AB}$ ,  $\varepsilon_{AC}$ , and  $\varepsilon_{BC}$  have the same form as above. The high-order term is more complex, but has similar features. It consists of a collection of Boltzmann-weighted averages of conformations  $j$  and  $k$  that cannot be immediately simplified. This shows, directly analogous to the pairwise epistatic case, that mutations perturbing conformations  $j$  and  $k$  can lead to a potentially non-zero three-way interaction term  $\beta_{ABC}$ .

To determine whether this could, in fact, lead to non-zero high-order epistasis, we randomly assigned values to each of the  $\langle G_{j,k}^{genotype} \rangle$  terms—drawing from a normal distribution with a mean of  $0 \text{ kT}$  and a standard deviation of  $1 \text{ kT}$ —and used these values to calculate  $\varepsilon_{ABC}$ . The resulting distribution of  $\varepsilon_{ABC}$  has a mean of  $0 \text{ kT}$  and a standard deviation of  $1.67 \text{ kT}$ . Thus, while it remains to be seen whether high-order ensemble epistasis occurs in real systems, there do exist combinations of values for  $\langle G_{j,k}^{genotype} \rangle$  that lead to non-zero high-order ensemble epistasis.

### 3 Epistasis between mutations within conformations can co-exist with ensemble epistasis

#### 3.1 Effect of adding epistasis within conformations

We next investigated the effect of including epistasis between mutations within each conformation  $i$ ,  $j$ , and  $k$ . Imagine, for example, that mutations  $a \rightarrow A$  and  $b \rightarrow B$  introduced a new ion pair in conformation  $j$ , but not conformation  $i$  or  $k$ . This would mean  $G_j^{AB}$  would be given by  $G_j^{ab} + \Delta G_{A,j}^{ab} + \Delta G_{B,j}^{ab} + \Delta \Delta \mathbf{G}_{AB,j}^{ab}$ , where the final term measures the interaction energy of the ion pair. In this section, we will denote all epistatic terms within a conformation—or summary terms that contain at least one such term—using bold-faced  $\Delta \Delta \mathbf{G}$ .

If we allow for epistasis between mutations within each conformation, we have to modify our expression form  $\Delta G_{obs}^{genotype}$  as follows:

**Table 5** Map between genotype and the thermodynamic description of  $\Delta G_{obs}^{genotype}$  with within conformation interactions.

Genotype	$\Delta G_{obs}^{genotype}$	$\langle G_{jk}^{genotype} \rangle$
$ab$	$G_i^{ab} - \langle \Delta G_{j,k}^{ab} \rangle$	$-RT \ln \left( e^{-G_j^{ab}/RT} + e^{-G_k^{ab}/RT} \right)$
$Ab$	$G_i^{ab} + \Delta G_{A,i}^{ab} - \langle \Delta G_{j,k}^{Ab} \rangle$	$-RT \ln \left( e^{-G_j^{ab} + \Delta G_{A,j}^{ab}/RT} + e^{-G_k^{ab} + \Delta G_{A,k}^{ab}/RT} \right)$
$aB$	$G_i^{ab} + \Delta G_{B,i}^{ab} - \langle \Delta G_{j,k}^{aB} \rangle$	$-RT \ln \left( e^{-G_j^{ab} + \Delta G_{B,j}^{ab}/RT} + e^{-G_k^{ab} + \Delta G_{B,k}^{ab}/RT} \right)$
$AB$	$G_i^{ab} + \Delta G_{A,i}^{ab} + \Delta G_{B,i}^{ab} + \Delta \Delta \mathbf{G}_{AB,i}^{ab} - \langle G_{j,k}^{AB} \rangle$	$-RT \ln \left( e^{-G_j^{ab} + \Delta G_{A,j}^{ab} + \Delta G_{B,j}^{ab} + \Delta \Delta \mathbf{G}_{AB,j}^{ab}/RT} + e^{-G_k^{ab} + \Delta G_{A,k}^{ab} + \Delta G_{B,k}^{ab} + \Delta \Delta \mathbf{G}_{AB,k}^{ab}/RT} \right)$

If we substitute these into the expression for  $\varepsilon$  and simplify, we obtain:

$$\varepsilon = \Delta \Delta \mathbf{G}_{AB,i}^{ab} + \left( \langle G_{j,k}^{aB} \rangle - \langle \mathbf{G}_{j,k}^{AB} \rangle \right) - \left( \langle G_{j,k}^{ab} \rangle - \langle G_{j,k}^{Ab} \rangle \right). \quad (51)$$

The two sources of epistasis now interact. The first is  $\Delta \Delta \mathbf{G}_{AB,i}^{ab}$ , which measures any interactions within conformation  $i$ . This is not ensemble epistasis, as it arises from within a given conformation rather than from redistribution of the probabilities of conformations. The term  $\langle \mathbf{G}_{j,k}^{AB} \rangle$ , on the other hand, is determined by both in-conformation epistasis and ensemble epistasis. The values of  $\Delta \Delta \mathbf{G}_{AB,j}^{ab}$  and  $\Delta \Delta \mathbf{G}_{AB,k}^{ab}$ —interactions within conformations  $j$  and  $k$ , respectively—contribute to the redistribution of probabilities within conformations, and thus contribute to ensemble epistasis.

**Table 6** Map between genotype and the thermodynamic description of  $\Delta G_{obs}^{genotype}$  for a two conformation ensemble with within-conformation interactions.

Genotype	$\Delta G_{obs}^{genotype}$
$ab$	$G_i^{ab} - G_j^{ab}$
$Ab$	$G_i^{ab} + \Delta G_{A,i}^{ab} - (G_j^{ab} + \Delta G_{A,j}^{ab})$
$aB$	$G_i^{ab} + \Delta G_{B,i}^{ab} - (G_j^{ab} + \Delta G_{B,j}^{ab})$
$AB$	$G_i^{ab} + \Delta G_{A,i}^{ab} + \Delta G_{B,i}^{ab} + \Delta \Delta G_{AB,i}^{ab} - (G_j^{ab} + \Delta G_{A,j}^{ab} + \Delta G_{B,j}^{ab} + \Delta \Delta G_{AB,j}^{ab})$

### 3.2 To see ensemble epistasis, it is necessary to have three or more conformations even with a epistasis within conformations

We showed before that two conformations along could not lead to ensemble epistasis. We next tested whether this result held even if we allowed for epistasis within conformations. If we remove conformation  $k$ , reducing the ensemble to conformations  $i$  and  $j$  we obtain the following expressions for  $\Delta G_{obs}^{genotype}$ :

In this case,  $\varepsilon$  simplifies to:

$$\varepsilon = \Delta \Delta G_{AB,i}^{ab} - \Delta \Delta G_{AB,j}^{ab} \quad (52)$$

Thus, any epistasis we observe arises from interactions within conformations  $i$  and  $j$ —not from redistribution of the probabilities of conformations that characterizes ensemble epistasis.

### 3.3 To see ensemble epistasis, it is necessary for mutations $a \rightarrow A$ and $b \rightarrow B$ to have different effects on conformations $j$ and $k$ , even with interactions within conformations

In this analysis, we will repeat the work done in 1.3, but now allowing for interactions within conformations. We will set  $\Delta G_{B,j}^{ab} = \Delta G_{B,k}^{ab} = \Delta G_{B,jk}^{ab}$ :

$$\langle G_{j,k}^{aB} \rangle = -RT \ln \left( e^{-(G_j^{ab} + \Delta G_{B,j}^{ab})/RT} + e^{-(G_k^{ab} + \Delta G_{B,k}^{ab})/RT} \right) \quad (53)$$

$$\langle \Delta G_{j,k}^{aB} \rangle = \Delta G_{B,jk}^{ab} + \langle G_{j,k}^{ab} \rangle. \quad (54)$$

Things are more complicated for  $\langle \Delta G_{j,k}^{AB} \rangle$ :

$$\langle \Delta G_{j,k}^{AB} \rangle = -RT \ln \left( e^{-(G_j^{ab} + \Delta G_{A,j}^{ab} + \Delta G_{B,jk}^{ab} + \Delta \Delta G_{AB,i}^{ab})/RT} + e^{-(G_k^{ab} + \Delta G_{A,k}^{ab} + \Delta G_{B,jk}^{ab} + \Delta \Delta G_{AB,i}^{ab})/RT} \right) \quad (55)$$

$$\langle \Delta G_{j,k}^{AB} \rangle = \Delta G_{B,jk}^{ab} - RT \ln \left( e^{-(G_j^{ab} + \Delta G_{A,j}^{ab} + \Delta \Delta G_{AB,i}^{ab})/RT} + e^{-(G_k^{ab} + \Delta G_{A,k}^{ab} + \Delta \Delta G_{AB,i}^{ab})/RT} \right) \quad (56)$$

We can call the term on the right  $\langle G_{j,k}^{AB*} \rangle$ :

$$\langle \Delta G_{j,k}^{AB} \rangle = \Delta G_{B,jk}^{ab} + \langle G_{j,k}^{AB*} \rangle \quad (57)$$

$$\varepsilon = \Delta \Delta G_{AB,i}^{ab} - \left( [\Delta G_{B,jk}^{ab} + \langle G_{j,k}^{ab} \rangle] - [\Delta G_{B,jk}^{ab} + \langle G_{j,k}^{AB*} \rangle] \right) - \left( \langle G_{j,k}^{ab} \rangle - \langle G_{j,k}^{AB} \rangle \right) \quad (58)$$

$$\varepsilon = \Delta \Delta G_{AB,i}^{ab} - \Delta G_{B,jk}^{ab} + \langle G_{j,k}^{ab} \rangle - \Delta G_{B,jk}^{ab} - \langle G_{j,k}^{AB*} \rangle - \langle G_{j,k}^{ab} \rangle + \langle G_{j,k}^{AB} \rangle \quad (59)$$

$$\varepsilon = \Delta \Delta G_{AB,i}^{ab} - \left( \langle G_{j,k}^{AB*} \rangle - \langle G_{j,k}^{AB} \rangle \right). \quad (60)$$

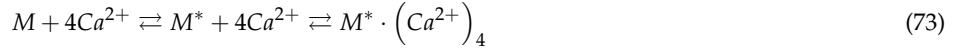
In this scenario, ensemble epistasis may still be observed, even if  $b \rightarrow B$  has the same effect on both  $j$  and  $k$ . This said, because we did not specify that  $\Delta \Delta G_{AB,j}^{ab}$  was the same as  $\Delta \Delta G_{AB,k}^{ab}$ , the  $b \rightarrow B$  mutation can, in fact, have different effects on  $j$  and  $k$  if  $A$  is present. We therefore added a second constraint:

$$\Delta \Delta G_{AB,j}^{ab} = \Delta \Delta G_{AB,k}^{ab} = \Delta \Delta G_{AB,jk}^{ab} \quad (61)$$

This constraint ensures that  $b \rightarrow B$  has the same effect on  $j$  and  $k$ , regardless of genetic background. Starting from Equation 60, we can now factor out  $\Delta \Delta G_{AB,jk}^{ab}$ :



With these assumptions, we can describe the system with the following scheme and equilibrium constants:



$$K_* = \frac{[M^*]}{[M]} \quad (74)$$

$$K_C = \frac{[M^* \cdot (Ca^{2+})_4]}{[M^*] [Ca^{2+}]^4}. \quad (75)$$

The stability of  $M^* \cdot (Ca^{2+})_4$  relative to the other protein conformations is given by:

$$\Delta G = -RT \ln \left( \frac{[M^* \cdot (Ca^{2+})_4]}{[M] + [M^*]} \right). \quad (76)$$

Substitute the equilibrium constants and simplify:

$$\Delta G = -RT \ln \left( \frac{[M^*] K_C [Ca^{2+}]^4}{[M] + [M^*]} \right), \quad (77)$$

$$\Delta G = -RT \ln \left( \frac{K_* [M] K_C [Ca^{2+}]^4}{[M] + K_* [M]} \right), \quad (78)$$

$$\Delta G = -RT \ln \left( \frac{K_* K_C [Ca^{2+}]^4}{1 + K_*} \right). \quad (79)$$

Assume that  $K_* \ll 1$ , meaning that  $M$  is highly favored over  $M^*$  in the absence of  $Ca^{2+}$ :

$$\Delta G \approx -RT \ln \left( \frac{K_* K_C [Ca^{2+}]^4}{1} \right) = -RT \ln \left( K_* K_C [Ca^{2+}]^4 \right) \quad (80)$$

$$\Delta G = -RT \ln (K_*) - RT \ln (K_C) - RT \ln \left( [Ca^{2+}]^4 \right) \quad (81)$$

$$\Delta G = -RT \ln (K_*) - RT \ln (K_C) - 4RT \ln \left( [Ca^{2+}] \right). \quad (82)$$

Setting  $\mu_{Ca^{2+}} = RT \ln([Ca^{2+}])$ :

$$\Delta G = \Delta G_* + \Delta G_C - 4\mu_{Ca^{2+}} \quad (83)$$

where  $\Delta G_*$  is the stability of  $M^*$  relative to  $M$  in the absence of  $Ca^{2+}$ .  $\Delta G_C$  describes the affinity of the  $M^*$  conformation for  $Ca^{2+}$ . The terms  $\Delta G_*$  and  $\Delta G_C$ , together, describe the intrinsic stability of the active, metal-bound "ca" complex at a reference  $[Ca^{2+}]$ . We therefore define a new constant:

$$G_{ca}^\circ \equiv \Delta G_* + \Delta G_C. \quad (84)$$

The final expression for  $G_{ca}(\mu_{Ca^{2+}})$  is:

$$G_{ca}(\mu_{Ca^{2+}}) = G_{ca}^\circ - 4\mu_{Ca^{2+}}. \quad (85)$$

The microscopic free energy of the *apo* ( $M$ ) conformation does not depend on the concentration of  $Ca^{2+}$ ; therefore,  $G_{apo}$  is a constant:

$$G_{apo}(\mu_{Ca^{2+}}) = G_{apo}^\circ. \quad (86)$$

## 5.2 Setting arbitrary offset

We do not know  $G_{ca}^\circ$  or  $G_{apo}^\circ$ . We do know, however, that at a low calcium concentration  $G_{apo}(\mu_{Ca^{2+}}) \ll G_{ca}(\mu_{Ca^{2+}})$  (meaning, the  $M$  form is favored over  $M^*$  at low calcium). We also know that  $G_{ca}(\mu_{Ca^{2+}})$  will increase linearly relative to  $G_{apo}^\circ$  as a function of  $\mu_{Ca^{2+}}$ . If we do not care about the absolute value of  $[Ca^{2+}]$  at which the system transitions between favoring *apo* and *pep*, we can choose arbitrary values for  $G_{ca}^\circ$  and  $G_{apo}^\circ$  and then still calculate how epistasis should change as a function of  $\mu_{Ca^{2+}}$  for the protein. For convenience, we set  $G_{apo}^\circ = 0$  and  $G_{ca}^\circ = 10$  at  $\mu_{Ca^{2+}} = 0$ . We tested the sensitivity of our results to our choice of  $G_{apo}^\circ$  (Fig S1).



1 **5.3 Modeling mutant cycles**

*ab* genotype:

$$G_{ca}^{ab}(\mu_{Ca^{2+}}) = G_{ca}^{\circ} - 4\mu_{Ca^{2+}} \quad (87)$$

$$G_{apo}^{ab} = G_{apo}^{\circ} \quad (88)$$

$$\langle G_{apo,ca}^{ab} \rangle (\mu_{Ca^{2+}}) = -RT \ln \left( e^{-(G_{ca}^{\circ} - 4\mu_{Ca^{2+}})/RT} + e^{-(G_{apo}^{\circ})/RT} \right) \quad (89)$$

2 *Ab* genotype:

$$G_{Ab,ca}(\mu_{Ca^{2+}}) = G_{ca}^{\circ} - 4\mu_{Ca^{2+}} + \Delta G_{A,ca}^{ab} \quad (90)$$

$$G_{Ab,apo} = G_{apo}^{\circ} + \Delta G_{A,apo}^{ab} \quad (91)$$

$$\langle G_{apo,ca}^{Ab} \rangle (\mu_{Ca^{2+}}) = -RT \ln \left( e^{-(G_{ca}^{\circ} - 4\mu_{Ca^{2+}} + \Delta G_{A,ca}^{ab})/RT} + e^{-(G_{apo}^{\circ} + \Delta G_{A,apo}^{ab})/RT} \right) \quad (92)$$

3 *aB* genotype:

$$G_{aB,ca}(\mu_{Ca^{2+}}) = G_{ca}^{\circ} - 4\mu_{Ca^{2+}} + \Delta G_{B,ca} \quad (93)$$

$$G_{aB,apo} = G_{apo}^{\circ} + \Delta G_{B,apo} \quad (94)$$

$$\langle G_{apo,ca}^{aB} \rangle (\mu_{Ca^{2+}}) = -RT \ln \left( e^{-(G_{ca}^{\circ} - 4\mu_{Ca^{2+}} + \Delta G_{B,ca})/RT} + e^{-(G_{apo}^{\circ} + \Delta G_{B,apo})/RT} \right) \quad (95)$$

4 *AB* genotype:

$$G_{AB,ca}(\mu_{Ca^{2+}}) = G_{ca}^{\circ} - 4\mu_{Ca^{2+}} + \Delta G_{A,ca} + \Delta G_{B,ca} \quad (96)$$

$$G_{AB,apo} = G_{apo}^{\circ} + \Delta G_{A,apo} + \Delta G_{B,apo} \quad (97)$$

$$\langle G_{apo,ca}^{AB} \rangle (\mu_{Ca^{2+}}) = -RT \ln \left( e^{-(G_{ca}^{\circ} - 4\mu_{Ca^{2+}} + \Delta G_{A,ca} + \Delta G_{B,ca})/RT} + e^{-(G_{apo}^{\circ} + \Delta G_{A,apo} + \Delta G_{B,apo})/RT} \right) \quad (98)$$

5 Final expression for  $\mu_{Ca^{2+}}$ -dependence of  $\varepsilon$ :

$$\varepsilon(\mu_{Ca^{2+}}) = - \left[ \left( \langle \Delta G_{ca,apo}^{AB} \rangle - \langle G_{ca,apo}^{aB} \rangle \right) - \left( \langle G_{ca,apo}^{Ab} \rangle - \langle G_{ca,apo}^{ab} \rangle \right) \right] \quad (99)$$