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

2

3

4

5

6

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 \*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  $a \to A$  and  $b \to B$  in  $\Delta G_{obs}$  as the difference in the effect of  $a \to A$  in the *ab* and *aB* backgrounds (Fig 1A):

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

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

$$\Delta G_{obs}^{genotype} = G_i^{genotype} - \left\langle G_{j,k}^{genotype} \right\rangle \tag{21}$$

where

$$\left\langle G_{j,k}^{genotype} \right\rangle = -RTln\left(e^{-G_{j}^{genotype}/RT} + e^{-G_{k}^{genotype}/RT}\right).$$
(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  $\Delta G_{abc}^{genotype}$ .

1 0 11	<b>1</b> 005	
Genotype	$\Delta G_{obs}^{genotype}$	$\left\langle G_{jk}^{genotype} \right\rangle$
ab	$G^{ab}_i - \left\langle \Delta G^{ab}_{j,k}  ight angle$	$-RTln\left(e^{-\left(G_{j}^{ab}\right)/RT}+e^{-\left(G_{k}^{ab}\right)/RT}\right)$
Ab	$G_{i}^{ab}+\Delta G_{A,i}^{ab}-\left\langle \Delta G_{j,k}^{Ab} ight angle$	$-RTln\left(e^{-\left(G_{j}^{ab}+\Delta G_{Aj}^{ab}\right)/RT}+e^{-\left(G_{k}^{ab}+\Delta G_{A,k}^{ab}\right)/RT}\right)$
aB	$G^{ab}_i + \Delta G^{ab}_{B,i} - \left\langle \Delta G^{aB}_{j,k}  ight angle$	$-RTln\left(e^{-\left(G_{j}^{ab}+\Delta G_{B,j}^{ab}\right)/RT}+e^{-\left(G_{k}^{ab}+\Delta G_{B,k}^{ab}\right)/RT}\right)$
AB	$G_{i}^{ab} + \Delta G_{A,i}^{ab} + \Delta G_{B,i}^{ab} - \left\langle \Delta G_{j,k}^{AB} \right\rangle$	$-\mathrm{RTln}\left(e^{-\left(G_{j}^{ab}+\Delta G_{A,j}^{ab}+\Delta G_{B,j}^{ab}\right)/RT}+\right.$
		$e^{-\left(G_k^{ab}+\Delta G_{A,k}^{ab}+\Delta G_{B,k}^{ab} ight)/RT}$

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

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

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

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

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

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

10

9

2

3

4

5

6

7

8

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

<sup>2</sup> We can next consider the two-conformation case, where *k* is not populated. In this case:

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

3 simplifies to:

$$\left\langle G_{j}^{genotype} \right\rangle = -RTln\left(e^{-G_{j}^{genotype}/RT}\right) = G_{j}^{genotype}.$$
 (28)

<sup>4</sup> 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	$G^{ab}_i - G^{ab}_j$
Ab	$G^{ab}_i + \Delta G^{ab}_{A,i} - \left(G^{ab}_j + \Delta G^{ab}_{A,j} ight)$
aB	$G_i^{ab} + \Delta G_{B,i}^{ab} - \left(G_j^{ab} + \Delta G_{B,j}^{ab}\right)$
AB	$G_{i}^{ab} + \Delta G_{A,i}^{ab} + \Delta G_{B,i}^{ab} - \left(G_{j}^{ab} + \Delta G_{A,j}^{ab} + \Delta G_{B,j}^{ab}\right)$

<sup>5</sup> If we substitute the relevant expressions for  $\Delta G_{obs}^{genotype}$  into our expression for  $\varepsilon$ , we get

$$\left(\left[G_{i}^{ab} + \Delta G_{A,i}^{ab} + \Delta G_{B,i}^{ab} - \left(G_{j}^{ab} + \Delta G_{A,j}^{ab} + \Delta G_{B,j}^{ab}\right)\right] - \left[G_{i}^{ab} + {}^{ab}_{B,i} - \left(G_{j}^{ab} + \Delta G_{B,j}^{ab}\right)\right]\right) - \left(\left[G_{i}^{ab} + {}^{ab}_{A,i} - \left(G_{j}^{ab} + {}^{ab}_{A,j}\right)\right] - \left[G_{i}^{ab} - G_{j}^{ab}\right]\right)$$

$$(29)$$

$$G_{i}^{ab} + \Delta G_{A,i}^{ab} + \Delta G_{B,i}^{ab} - G_{j}^{ab} - \Delta G_{A,j}^{ab} - \Delta G_{B,j}^{ab} - G_{i}^{ab} - \Delta G_{B,i}^{ab} + G_{j}^{ab} + \Delta G_{B,j}^{ab} \setminus -G_{i}^{ab} - \Delta G_{A,i}^{ab} + G_{j}^{ab} + \Delta G_{A,j}^{ab} + G_{j}^{ab} - G_{j}^{ab}$$
(30)

 $\epsilon =$ 

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

## $\tau$ 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.

<sup>8</sup> 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 <sup>9</sup> 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 <sup>10</sup> 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 <sup>11</sup> for both mutations to have differential effects to observe ensemble epistasis.

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

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

<sup>13</sup> Because  $\Delta G_{B,ik}^{ab}$  is shared among terms, we can factor it out:

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

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

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

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

$$\left\langle \Delta G_{j,k}^{aB} \right\rangle = \Delta G_{B,jk}^{ab} + \left\langle G_{j,k}^{ab} \right\rangle. \tag{36}$$

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

$$\left\langle \Delta G_{j,k}^{AB} \right\rangle = \Delta G_{B,jk}^{ab} + \left\langle G_{j,k}^{Ab} \right\rangle. \tag{37}$$

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

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

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

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

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

2

3

4

11

12

13

14

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:

1 0 51	<b>y i</b> 005	0 1
genotype	genetic model	$\Delta G_{obs}^{genotype}$
abc	$lpha_{abc}$	$G_i^{abc} - \left\langle G_{j,k}^{abc} \right\rangle$
Abc	$\alpha_{abc} + \beta_A$	$G_{i}^{abc} + \Delta G_{A,i}^{abc} - \left\langle G_{j,k}^{Abc} \right\rangle$
aBc	$\alpha_{abc} + \beta_B$	$G_i^{abc} + \Delta G_{B,i}^{abc} - \left\langle G_{j,k}^{aBc} \right\rangle$
abC	$\alpha_{abc} + \beta_C$	$G_i^{abc} + \Delta G_{C,i}^{abc} - \left\langle G_{j,k}^{abc} \right\rangle$
ABc	$\alpha_{abc} + \beta_A + \beta_B + \varepsilon_{AB}$	$G_{i}^{abc} + \Delta G_{A,i}^{abc} + \Delta G_{B,i}^{abc} - \left\langle G_{j,k}^{ABc} \right\rangle$
AbC	$\alpha_{abc} + \beta_A + \beta_C + \varepsilon_{AC}$	$G_{i}^{abc} + \Delta G_{A,i}^{abc} + \Delta G_{C,i}^{abc} - \left\langle G_{j,k}^{AbC} \right\rangle$
aBC	$\alpha_{abc} + \beta_B + \beta_C + \varepsilon_{BC}$	$G_{i}^{abc} + \Delta G_{B,i}^{abc} + \Delta G_{C,i}^{abc} - \left\langle G_{j,k}^{aBC} \right\rangle$
ABC	$\alpha_{abc} + \beta_A + \beta_B + \beta_C + \varepsilon_{AB} + \varepsilon_{AC} + \varepsilon_{BC} + \varepsilon_{ABC}$	$G_{i}^{abc} + \Delta G_{A,i}^{abc} + \Delta G_{B,i}^{abc} + \Delta G_{C,i}^{abc} - \left\langle G_{j,k}^{ABC} \right\rangle$

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

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:

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

$$\tag{42}$$

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:

$$\alpha_{abc} = G_i^{abc} - \left\langle G_{j,k}^{abc} \right\rangle \tag{43}$$

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

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

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

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

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

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

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

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 *kT* and a standard deviation of 1 *kT*—and used these values to calculate  $\varepsilon_{ABC}$ . The resulting distribution of  $\varepsilon_{ABC}$  has a mean of 0 *kT* and a standard deviation of 1.67 *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.

## 10 3 Epistasis between mutations within conformations can co-exist with ensemble epistasis

#### 11 3.1 Effect of adding epistasis within conformations

<sup>12</sup> We next investigated the effect of including epistasis between mutations within each conformation *i*, *j*, and *k*. Imagine, for example, <sup>13</sup> 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 <sup>14</sup> be given by  $G_j^{ab} + \Delta G_{A,j}^{ab} + \Delta G_{AB,j}^{ab}$  where the final term measures the interaction energy of the ion pair. In this section, we <sup>15</sup> will denote all epistatic terms within a conformation—or summary terms that contain at least one such term—using bold-faced  $\Delta\Delta G$ .

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

Genotype	$\Delta G_{obs}^{genotype}$	$\left\langle G_{jk}^{genotype} \right\rangle$
ab	$G^{ab}_i - \left\langle \Delta G^{ab}_{j,k}  ight angle$	$-RTln\left(e^{-\left(G_{j}^{ab}\right)/RT}+e^{-\left(G_{k}^{ab}\right)/RT}\right)$
Ab	$G^{ab}_i + \Delta G^{ab}_{A,i} - \left\langle \Delta G^{Ab}_{j,k}  ight angle$	$-RTln\left(e^{-\left(G_{j}^{ab}+\Delta G_{Aj}^{ab}\right)/RT}+e^{-\left(G_{k}^{ab}+\Delta G_{A,k}^{ab}\right)/R}\right)$
aB	$G_{i}^{ab}+\Delta G_{B,i}^{ab}-\left\langle \Delta G_{j,k}^{aB} ight angle$	$-RTln\left(e^{-\left(G_{j}^{ab}+\Delta G_{B,j}^{ab}\right)/RT}+e^{-\left(G_{k}^{ab}+\Delta G_{B,k}^{ab}\right)/R}\right)$
AB	$G^{ab}_i$ + $\Delta G^{ab}_{A,i}$ + $\Delta G^{ab}_{B,i}$ + $\Delta \Delta \mathbf{G}^{ab}_{\mathbf{AB},i}$ +	$- \left( \left( C^{ab} + A C^{ab} + A C^{ab} + A A C^{ab} \right) / DT \right)$
	$\left\langle G_{ik}^{AB} \right\rangle$	-RTln $\left( e^{-\left( G_{j} + \Delta G_{A,j} + \Delta G_{B,j} + \Delta \Delta G_{AB,j} \right) / KI} + \right)$

<b>Table 5</b> Map between genotype and the thermodynamic description of $\Delta G_{abc}^{genotype}$ with v	within conformation interactions.
---	-----------------------------------

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

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

<sup>19</sup> The two sources of epistasis now interact. The first is  $\Delta\Delta G_{AB,i}^{ab}$ , which measures any interactions within conformation *i*. This is not <sup>20</sup> ensemble epistasis, as it arises from within a given conformation rather than from redistribution of the probabilities of conformations. <sup>21</sup> The term  $\langle G_{j,k}^{AB} \rangle$ , on the other hand, is determined by both in-conformation epistasis and ensemble epistasis. The values of  $\Delta\Delta G_{AB,j}^{ab}$ <sup>22</sup> and  $\Delta\Delta G_{AB,k}^{ab}$ —interactions within conformations *j* and *k*, respectively—contribute to the redistribution of probabilities within <sup>23</sup> conformations, and thus contribute to ensemble epistasis.

 $-\left(G_{k}^{ab}+\Delta G_{A,k}^{ab}+\Delta G_{B,k}^{ab}+\Delta \Delta G_{AB,k}^{ab}\right)/RT$ 

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

Genotype	$\Delta G_{obs}^{genotype}$
ab	$G_i^{ab} - G_j^{ab}$
Ab	$G^{ab}_i + \Delta G^{ab}_{A,i} - \left(G^{ab}_j + \Delta G^{ab}_{A,j} ight)$
aB	$G^{ab}_i + \Delta G^{ab}_{B,i} - \left(G^{ab}_j + \Delta G^{ab}_{B,j} ight)$
AB	$G^{ab}_i$ + $\Delta G^{ab}_{A,i}$ + $\Delta G^{ab}_{B,i}$ + $\Delta \Delta \mathbf{G}^{\mathbf{ab}}_{\mathbf{AB},\mathbf{i}}$ -
	$\left(G^{ab}_{j}+\Delta G^{ab}_{A,j}+\Delta G^{ab}_{B,j}+{f \Delta \Delta G^{ab}_{AB,j}} ight)$

# 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:

$$\varepsilon = \Delta \Delta \mathbf{G}_{\mathbf{AB},\mathbf{i}}^{\mathbf{ab}} - \Delta \Delta \mathbf{G}_{\mathbf{AB},\mathbf{j}}^{\mathbf{ab}} \tag{52}$$

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 i 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  $\Delta G_{Bi}^{ab}$  =  $\Delta G_{B,k}^{ab} = \Delta G_{B,ik}^{ab}:$ 

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

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

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

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

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

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

$$\left\langle \Delta G_{j,k}^{AB} \right\rangle = \Delta G_{B,jk}^{ab} + \left\langle \mathbf{G}_{j,k}^{AB*} \right\rangle \tag{57}$$

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

$$\varepsilon = \Delta \Delta G_{AB,i}^{ab} - \Delta G_{B,jk}^{ab} + \left\langle G_{j,k}^{ab} \right\rangle - \Delta G_{B,jk}^{ab} - \left\langle \mathbf{G}_{j,k}^{\mathbf{AB}*} \right\rangle - \left\langle G_{j,k}^{ab} \right\rangle + \left\langle G_{j,k}^{Ab} \right\rangle$$
(59)

$$\varepsilon = \Delta \Delta \mathbf{G}_{\mathbf{AB},\mathbf{i}}^{\mathbf{ab}} - \left( \left\langle \mathbf{G}_{j,k}^{\mathbf{AB}*} \right\rangle - \left\langle \mathbf{G}_{j,k}^{Ab} \right\rangle \right).$$
(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 15 did not specify that  $\Delta\Delta G_{AB,i}^{ab}$  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. 16 We therefore added a second constraint:

$$\Delta \Delta \mathbf{G}_{\mathbf{A}\mathbf{B},\mathbf{j}}^{\mathbf{a}\mathbf{b}} = \Delta \Delta \mathbf{G}_{\mathbf{A}\mathbf{B},\mathbf{k}}^{\mathbf{a}\mathbf{b}} = \Delta \Delta \mathbf{G}_{\mathbf{A}\mathbf{B},\mathbf{j}\mathbf{k}}^{\mathbf{a}\mathbf{b}}.$$
(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 18 can now factor out  $\Delta\Delta G_{AB,ik}^{ab}$ : 19

14

5

7

8

9

10

11

12

13

17

$$\varepsilon = \left\langle G_{j,k}^{Ab} \right\rangle + RTln \left( e^{-\left(G_{j}^{ab} + \Delta G_{A,j}^{ab} + \Delta \Delta G_{AB,jk}^{ab}\right)/RT} + e^{-\left(G_{k}^{ab} + \Delta G_{A,k}^{ab} + \Delta \Delta G_{AB,jk}^{ab}\right)/RT} \right).$$
(62)

$$\varepsilon = \left\langle G_{j,k}^{Ab} \right\rangle + \Delta \Delta G_{\mathbf{AB},j\mathbf{k}}^{\mathbf{ab}} + RTln \left( e^{-\left(G_{j}^{ab} + \Delta G_{A,j}^{ab}\right)/RT} + e^{-\left(G_{k}^{ab} + \Delta G_{A,k}^{ab}\right)/RT} \right)$$
(63)

$$\varepsilon = \left\langle G_{j,k}^{Ab} \right\rangle + \Delta \Delta \mathbf{G}_{\mathbf{AB},\mathbf{jk}}^{\mathbf{ab}} - \left\langle G_{j,k}^{Ab} \right\rangle \tag{64}$$

$$\varepsilon = \Delta \Delta \mathbf{G}_{\mathbf{AB},\mathbf{i}}^{\mathbf{ab}} - \Delta \Delta \mathbf{G}_{\mathbf{AB},\mathbf{jk}}^{\mathbf{ab}}.$$
(65)

We now reach a case similar to the two-conformation case with epistasis within conformations: we may still observe epistasis, but it is due to contacts within each conformation rather than ensemble epistasis. If one of the mutations has an identical effect on conformations j and k—including any epistatic interactions within the conformation—ensemble epistasis will not be observed.

#### 4 Ensemble epistasis can arise for ensembles with more than three conformations

<sup>5</sup> To simplify our analysis, we started by assuming we were interested in the equilibrium between a single confirmation *i* and two other conformations, *j* and *k*. In many instances, we are instead interested in the difference in free energy between two collections of conformations. For example, we might be interested in the free energy difference between all bound conformations and all unbound conformations for a protein interacting with a drug. We can denote these two sub-ensembles with *X* and *Y*. We are thus interested in:

$$\Delta G_{obs} = \langle G_X \rangle - \langle G_Y \rangle \tag{66}$$

9 where

$$\langle G_X \rangle = -RTln\left(\sum_{m \in X} e^{-G_m/RT}\right),$$
(67)

<sup>10</sup> and *m* indexes over all conformations in the *X* sub-ensemble. Analogously,

$$\langle G_Y \rangle = -RTln\left(\sum_{n \in Y} e^{-G_n/RT}\right),$$
(68)

and *n* indexes over all conformations in the *Y* sub-ensemble. If we substitute the relevant expressions for  $\Delta G_{obs}^{genotype}$  into our expression for  $\varepsilon$  60, we obtain:

$$\varepsilon = \left\langle G_X^{AB} \right\rangle - \left\langle G_Y^{AB} \right\rangle - \left\langle G_X^{aB} \right\rangle + \left\langle G_Y^{aB} \right\rangle - \left\langle G_X^{Ab} \right\rangle + \left\langle G_Y^{Ab} \right\rangle + \left\langle G_X^{ab} \right\rangle - \left\langle G_Y^{ab} \right\rangle$$
(69)

$$\varepsilon = \left\langle G_X^{AB} \right\rangle - \left\langle G_X^{aB} \right\rangle - \left\langle G_X^{Ab} \right\rangle + \left\langle G_X^{ab} \right\rangle - \left\langle G_Y^{AB} \right\rangle + \left\langle G_Y^{aB} \right\rangle + \left\langle G_Y^{Ab} \right\rangle - \left\langle G_Y^{ab} \right\rangle - \left\langle G_Y^{ab} \right\rangle$$
(70)

$$\varepsilon = \left( \left\langle G_X^{AB} \right\rangle - \left\langle G_X^{AB} \right\rangle - \left\langle G_X^{Ab} \right\rangle + \left\langle G_X^{ab} \right\rangle \right) + \left( - \left\langle G_Y^{AB} \right\rangle + \left\langle G_Y^{aB} \right\rangle + \left\langle G_Y^{Ab} \right\rangle - \left\langle G_Y^{ab} \right\rangle \right)$$
(71)

$$\varepsilon = \left[ \left( \left\langle G_X^{AB} \right\rangle - \left\langle G_X^{aB} \right\rangle \right) - \left( \left\langle G_X^{Ab} \right\rangle - \left\langle G_X^{ab} \right\rangle \right) \right] - \left[ \left( \left\langle G_Y^{AB} \right\rangle - \left\langle G_Y^{aB} \right\rangle \right) - \left( \left\langle G_Y^{Ab} \right\rangle - \left\langle G_Y^{ab} \right\rangle \right) \right]$$
(72)

<sup>13</sup> The rightmost terms (involving sub-ensemble *Y*) are directly analogous to the ensemble we analyzed throughout this manuscript.

The leftmost terms (involving sub-ensemble *X*) are new ensemble epistasis that arises when we have multiple conformations on both sides of the equilibrium constant.

# <sup>16</sup> 5 Modeling the calcium-dependence of ensemble populations for S100A4.

#### 17 5.1 Deriving the model

<sup>18</sup> S100A4 populates both a closed conformation (*M*) and an open conformation (*M*\*), differentiated by exposure of a hydrophobic cleft <sup>19</sup> by rotation of two helices. In the absence of  $Ca^{2+}$ , *M* is favored over  $M^*$ .  $Ca^{2+}$  binds cooperatively to four sites in the  $M^*$  conformation <sup>20</sup> Garrett *et al.* (2008). The  $M^* \cdot (Ca^{2+})_4$  and *M* species correspond to the "*ca*" and "*apo*" species from the main text. Finally, peptide <sup>21</sup> binds preferentially to the  $M^*$  conformation. To model the system, we make the following assumptions:

- <sup>22</sup> 1. *M* is strongly favored over  $M^*$  in the absence of  $Ca^{2+}$ .
- 23 2.  $Ca^{2+}$  binds cooperatively at four equivalent sites on  $M^*$ .
- <sup>24</sup> 3.  $Ca^{2+}$  binds much more tightly to  $M^*$  than M, allowing us to neglect the  $M \cdot Ca_4^{2+}$  conformation.
- 4. *Peptide* binds much more tightly to  $M^*$  than M, allowing us to neglect any  $M \cdot peptide$  conformations.

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

$$M + 4Ca^{2+} \rightleftharpoons M^* + 4Ca^{2+} \rightleftharpoons M^* \cdot \left(Ca^{2+}\right)_4 \tag{73}$$

$$K_* = \frac{[M^*]}{[M]} \tag{74}$$

2

3

4

6

8

9

10

11

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

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

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

Substitute the equilibrium constants and simplify:

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

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

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

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

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

$$\Delta G = -RTln\left(K_{*}\right) - RTln\left(K_{C}\right) - RTln\left(\left[Ca^{2+}\right]^{4}\right)$$
(81)

$$\Delta G = -RTln\left(K_{*}\right) - RTln\left(K_{C}\right) - 4RTln\left(\left[Ca^{2+}\right]\right).$$
(82)

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

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

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*<sup>2+</sup>]. We therefore define a new constant:

$$G_{ca}^{\circ} \equiv \Delta G_* + \Delta G_C. \tag{84}$$

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

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

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

# 5.3 Modeling mutant cycles

ab genotype:

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

$$G^{ab}_{apo} = G^{\circ}_{apo} \tag{88}$$

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

<sup>2</sup> Ab genotype:

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

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

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

₃ *aB* genotype:

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

$$G_{aB,apo} = G_{apo}^{\circ} + \Delta G_{B,apo} \tag{94}$$

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

<sup>4</sup> *AB* genotype:

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

$$G_{AB,apo} = G_{apo}^{\circ} + \Delta G_{A,apo} + \Delta G_{B,apo}$$
<sup>(97)</sup>

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

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

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