

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

3 Anneliese J. Morrison,^{1,2} Daria R. Wonderlick,^{1,2} and Michael J. Harms^{1,2,*}

4 * Corresponding author: harms@uoregon.edu

5 1. Department of Chemistry and Biochemistry, University of Oregon, Eugene, OR 97403

6 2. Institute for Molecular Biology, University of Oregon, Eugene, OR 97403

7 1 Ensemble epistasis requires three or more structures:

8 The definition of the Boltzmann-weighted stability of structures j and k for some genotype is (Equation 3):

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

9 If $\Delta G_k \gg \Delta G_j$, structure k is not appreciably populated; we set the second term to 0.

$$\langle \Delta G_{j,k}^{genotype} \rangle = -RT \ln \left(e^{-\Delta G_j^{genotype}/RT} \right)$$

10

$$\langle \Delta G_{j,k}^{genotype} \rangle = \Delta G_j^{genotype}$$

11 Now we have a two-state system, populating only structures i and j . ΔG_j^{Ab} is given by the free energy of the
12 wildtype structure perturbed by mutation A :

$$\Delta G_j^{Ab} = \Delta G_j^{ab} + \Delta \Delta G_{A,j}^{ab}.$$

13 Substituting expressions for the free energy of each genotype into our expression for β_{AB} , we find that all of
14 the terms cancel:

$$\beta_{AB} = (\Delta G_j^{ab} + \Delta \Delta G_{A,j}^{ab} - \Delta G_j^{ab}) + (\Delta G_j^{ab} + \Delta \Delta G_{B,j}^{ab} - \Delta G_j^{ab}) - (\Delta G_j^{ab} + \Delta \Delta G_{A,j}^{ab} + \Delta \Delta G_{B,j}^{ab} - \Delta G_j^{ab}),$$

15 yielding

$$\beta_{AB} = 0.$$

16 Thus, two-state systems cannot exhibit ensemble epistasis.

17 2 Ensemble epistasis requires mutations have differential effects:

18 We set $\Delta \Delta G_{B,j}^{ab} = \Delta \Delta G_{B,k}^{ab} = \Delta \Delta G_B^{ab}$ for mutation B , while allowing mutation A to have different effects
19 on each structure: $\Delta \Delta G_{A,j}^{ab} \neq \Delta \Delta G_{A,k}^{ab}$.

20 Consider the effect of B in the ab background:

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

21 We can factor out $\Delta\Delta G_B^{ab}$:

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

22

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

23

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

24

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

25

$$\langle \Delta G_{j,k}^{aB} \rangle = \Delta\Delta G_B^{ab} + \langle \Delta G_{j,k}^{ab} \rangle.$$

26 Consider the effect of B in the Ab background:

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

27 Again, we factor out $\Delta\Delta G_B^{ab}$:

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

28

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

29

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

30

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

31

$$\langle \Delta G_{j,k}^{AB} \rangle = \Delta\Delta G_{B,j}^{ab} + \langle \Delta G_{j,k}^{Ab} \rangle.$$

32 Substitute expressions for $\langle \Delta G_{j,k}^{aB} \rangle$ and $\langle \Delta G_{j,k}^{AB} \rangle$ into the expression for β_{AB} (Equation 9):

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

33

$$\beta_{AB} = (\langle \Delta G_{j,k}^{Ab} \rangle - \langle \Delta G_{j,k}^{ab} \rangle) + (\Delta\Delta G_B^{ab} + \langle \Delta G_{j,k}^{aB} \rangle - \langle \Delta G_{j,k}^{ab} \rangle) - (\Delta\Delta G_{B,j}^{ab} + \langle \Delta G_{j,k}^{Ab} \rangle - \langle \Delta G_{j,k}^{ab} \rangle).$$

34 All terms cancel, yielding:

$$\beta_{AB} = 0.$$

35 Mutations must have differential effects on two or more structures to observe ensemble epistasis.

3 Ensembles can lead to high-order epistasis

In this, we consider mutations at three sites as an extension of the two-site case. Below is a table describing the three-mutation case, directly paralleling the two-site case in Table 1.

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

We can solve for each coefficient in the epistatic model in thermodynamic terms:

$$\alpha_{abc} = \Delta G_i^{abc} - \langle \Delta G_{j,k}^{abc} \rangle$$

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

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

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

$$\beta_{AB} = (\langle \Delta G_{j,k}^{Abc} \rangle - \langle \Delta G_{j,k}^{abc} \rangle) + (\langle \Delta G_{j,k}^{aBc} \rangle - \langle \Delta G_{j,k}^{abc} \rangle) - (\langle \Delta G_{j,k}^{ABc} \rangle - \langle \Delta G_{j,k}^{abc} \rangle)$$

$$\beta_{AC} = (\langle \Delta G_{j,k}^{Abc} \rangle - \langle \Delta G_{j,k}^{abc} \rangle) + (\langle \Delta G_{j,k}^{abC} \rangle - \langle \Delta G_{j,k}^{abc} \rangle) - (\langle \Delta G_{j,k}^{AbC} \rangle - \langle \Delta G_{j,k}^{abc} \rangle)$$

$$\beta_{BC} = (\langle \Delta G_{j,k}^{aBc} \rangle - \langle \Delta G_{j,k}^{abc} \rangle) + (\langle \Delta G_{j,k}^{abC} \rangle - \langle \Delta G_{j,k}^{abc} \rangle) - (\langle \Delta G_{j,k}^{aBC} \rangle - \langle \Delta G_{j,k}^{abc} \rangle)$$

$$\beta_{ABC} = - \left[(\langle \Delta G_{j,k}^{Abc} \rangle - \langle \Delta G_{j,k}^{abc} \rangle) + (\langle \Delta G_{j,k}^{aBc} \rangle - \langle \Delta G_{j,k}^{abc} \rangle) + (\langle \Delta G_{j,k}^{abC} \rangle - \langle \Delta G_{j,k}^{abc} \rangle) \right] + \left[(\langle \Delta G_{j,k}^{ABc} \rangle - \langle \Delta G_{j,k}^{abc} \rangle) + (\langle \Delta G_{j,k}^{AbC} \rangle - \langle \Delta G_{j,k}^{abc} \rangle) + (\langle \Delta G_{j,k}^{aBC} \rangle - \langle \Delta G_{j,k}^{abc} \rangle) \right] - \left[(\langle \Delta G_{j,k}^{ABC} \rangle - \langle \Delta G_{j,k}^{abc} \rangle) \right]$$

This shows, directly analogous to the pairwise epistatic case, that nonlinear perturbations of mutations to unobserved structures j and k lead to a potentially non-zero three-way interaction term β_{ABC} .

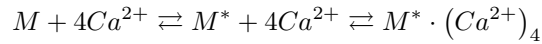
52 4 Ensemble Modeling

53 4.1 Deriving model

54 S100A4 populates both a closed conformation (M) and an open conformation (M^*), differentiated by exposure
55 of a hydrophobic cleft by rotation of two helices. In the absence of Ca^{2+} , M is favored over M^* . Ca^{2+} binds
56 cooperatively to four sites in the M^* state [Garrett et al., 2008]. The $M^* \cdot (Ca^{2+})_4$ and M species correspond
57 to the “*ca*” and “*apo*” species from the main text. Finally, *peptide* binds preferentially to the M^* state. To
58 model the system, we make the following assumptions:

- 59 1. M is strongly favored over M^* in the absence of Ca^{2+} .
- 60 2. Ca^{2+} binds cooperatively at four equivalent sites on M^* .
- 61 3. Ca^{2+} binds much more tightly to M^* than M , allowing us to neglect the $M \cdot Ca_4^{2+}$ state.
- 62 4. *peptide* binds much more tightly to M^* than M , allowing us to neglect any $M \cdot peptide$ states.

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



64

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

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

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

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

66 Substitute the equilibrium constants and simplify:

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

67

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

68

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

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

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

70

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

71

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

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

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

73 ΔG_* is the stability of M^* relative to M in the absence of Ca^{2+} . ΔG_C describes the affinity of the M^* state74 for Ca^{2+} . The terms ΔG_* and ΔG_C , together, describe the intrinsic stability of the active, metal-bound75 “ ca ” complex at a reference $[Ca^{2+}]$. We therefore define a new constant:

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

76 The final expression for ΔG is:

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

77 The microscopic free energy of the *apo* (M) state does not depend on the concentration of Ca^{2+} ; therefore,78 ΔG_{apo} is a constant:

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

79 **4.2 Setting arbitrary offset**80 We do not know ΔG_{ca}° or ΔG_{apo}° . We do know, however, that at a low calcium concentration $\Delta G_{apo}(\mu_{Ca^{2+}}) \ll$ 81 $\Delta G_{ca}(\mu_{Ca^{2+}})$ (meaning, the M form is favored over M^* at low calcium). We also know that $\Delta G_{ca}(\mu_{Ca^{2+}})$ 82 will increase linearly relative to ΔG_{apo}° as a function of $\mu_{Ca^{2+}}$. If we do not care about the absolute value83 of $[Ca^{2+}]$ at which the system transitions between favoring *apo* and *pep*, we can choose arbitrary values for84 ΔG_{ca}° and ΔG_{apo}° and then still calculate how epistasis should change as a function of $\mu_{Ca^{2+}}$ for the protein.85 For convenience, we set $\Delta G_{apo}^\circ = 0$ and $\Delta G_{ca}^\circ = 10$ at $\mu_{Ca^{2+}} = 0$.

86 **4.3 Modeling mutant cycles**

87 *ab* genotype:

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

88

$$\Delta G_{apo}^{ab} = \Delta G_{apo}^{\circ}$$

89

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

90 *Ab* genotype:

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

91

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

92

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

93 *aB* genotype:

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

94

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

95

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

96 *AB* genotype:

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

97

$$\Delta G_{AB,apo} = \Delta G_{apo}^{\circ} + \Delta \Delta G_{A,apo} + \Delta \Delta G_{B,apo}$$

98

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

99 Final expression for $\mu_{Ca^{2+}}$ -dependence of β_{AB} :

$$\beta_{AB}(\mu_{Ca^{2+}}) = (\langle \Delta G_{ca,apo}^{Ab} \rangle - \langle \Delta G_{ca,apo}^{ab} \rangle) + (\langle \Delta G_{ca,apo}^{aB} \rangle - \langle \Delta G_{ca,apo}^{ab} \rangle) - (\langle \Delta G_{ca,apo}^{AB} \rangle - \langle \Delta G_{ca,apo}^{ab} \rangle)$$

100 References

- 101 Sarah C. Garrett, Louis Hodgson, Andrew Rybin, Alexei Touthkine, Klaus M. Hahn, David S. Lawrence,
102 and Anne R. Bresnick. A Biosensor of S100A4 Metastasis Factor Activation: Inhibitor Screening and
103 Cellular Activation Dynamics. *Biochemistry*, 47(3):986–996, January 2008. ISSN 0006-2960. doi:
104 10.1021/bi7021624. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3227476/>.