

# Supporting Materials for “A model for the stabilities of protein crystals”

## 1 Non-electrostatic terms

We have supposed that protein crystals are stabilized mainly by nonelectrostatic protein-protein contact interactions. Each type of protein in its particular crystal lattice will be modeled by particular values of  $\Delta h_0$  and  $\Delta s_0$ . For our comparison to experiments in the main text, we have simply taken those two quantities as fit parameters. Here we make a few simple estimates to see if those values can be rationalized in terms of hydrophobic and hydrogen bonding interactions.

### 1.1 Hydrophobic interactions

The hydrophobic effect has been studied extensively in proteins for its role in stabilizing the folded state.<sup>1-6</sup> We adopt the usual formulation and express the enthalpy and entropy of hydrophobic burial in terms of the change in heat capacity between the buried and exposed states

$$\Delta h(T) = \Delta c_p(T - T_H) \tag{S1}$$

$$\Delta s(T) = \Delta c_p \ln \frac{T}{T_S}, \tag{S2}$$

where  $\Delta h$  and  $\Delta s$  are the enthalpy and entropy of hydrophobic burial per square Angstrom of solvent-accessible surface area (ASA),  $\Delta c_p$  is the heat capacity change per unit ASA, and  $T_H$  and  $T_S$  are the so-called convergence temperatures where the enthalpy and entropy changes vanish. The specific free energy change is then

$$\Delta g = \Delta c_p(T - T_H) - T\Delta c_p \ln \frac{T}{T_S}. \tag{S3}$$

There are three parameters in Eq. S3: the convergence temperatures,  $T_H$  and  $T_S$ , and the specific heat capacity change,  $\Delta c_p$ . We determine  $T_H$  and  $\Delta c_p$  by fitting Eq. S1 to published values for the enthalpy change.<sup>6</sup> We obtain  $\Delta c_p = -2.0 \text{ JK}^{-1} \text{ mol}^{-1} \text{ \AA}^{-2}$  in reasonable agreement with the work of Privalov and Makhatadze, who found  $-2.3 \text{ JK}^{-1} \text{ mol}^{-1} \text{ \AA}^{-2}$  at  $5^\circ\text{C}$ .<sup>7</sup> For the enthalpy convergence temperature we obtain  $T_H = 359\text{K}$ . The entropy convergence temperature is obtained by fitting Eq. S2 to experimental values for the specific entropy change,<sup>6</sup> using the previously determined heat capacity. We obtain  $T_S = 399\text{K}$ .

The hydrophobic burial term is then

$$\Delta G_{H\phi} = \Delta \text{ASA}_{\text{carbon}} \left[ \Delta c_p (T - T_H) - T \Delta c_p \ln \frac{T}{T_S} \right]. \quad (\text{S4})$$

To find the change in ASA we use a  $1.4 \text{ \AA}$  radius probe to measure the surface area of the crystal forms 193L (tetragonal)<sup>8</sup> and 1AKI (orthorhombic)<sup>9</sup> alone and in the presence of all crystallographic copies that approach the reference molecule within  $4 \text{ \AA}$ . For the tetragonal crystals we find a total  $\Delta \text{ASA}$  of  $1902 \text{ \AA}^2$  of which  $843 \text{ \AA}^2$  is from carbon atoms. For the orthorhombic form the numbers are  $\Delta \text{ASA} = 2344 \text{ \AA}^2$  of which  $\Delta \text{ASA}_{\text{carbon}} = 1042 \text{ \AA}^2$ . We do not distinguish between aliphatic and aromatic carbons because we found that the analysis is not very sensitive to this.

Fig. S1 shows the computed free energy, enthalpy, entropy of hydrophobic burial for tetragonal crystals. By this estimate, on balance, hydrophobic burial helps stabilize these tetragonal crystals of lysozyme.

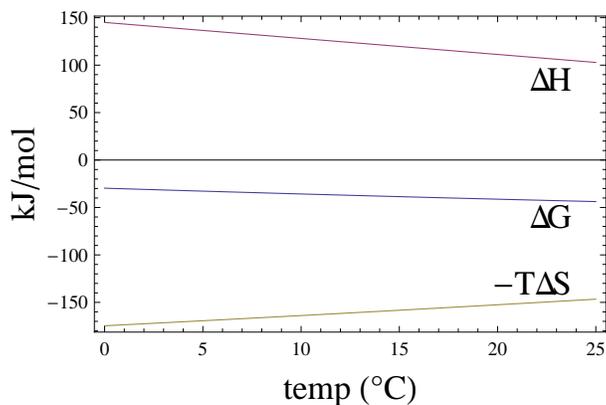


Figure S1: Enthalpy, entropy, and free energy of hydrophobic burial in tetragonal lysozyme crystals.

## 1.2 Hydrogen Bonds

There is considerable debate as to the role of H-bonds in the stabilization of proteins and protein-protein complexes,<sup>10</sup> undoubtedly at least partly because the strength of an H-bond is context dependent.<sup>11,12</sup> We make a simple estimate here.

For the temperature dependence of H-bonding, we use a functional form similar to that for the hydrophobic interactions,

$$\Delta G^{HB} = \Delta C_p^{HB}(T - T_H^{HB}) - T\Delta C_p^{HB} \ln \frac{T}{T_S^{HB}}, \quad (\text{S5})$$

where  $\Delta G^{HB}$  and  $\Delta C_p^{HB}$  are the free energy and heat capacity change per H-bond.<sup>11</sup> For two reasons, we must make choices here: (1) the experimental data<sup>11</sup> gives a wide range of values ( $66 < \Delta C_p^{HB} < 90 \text{ J K}^{-1} \text{ mol}^{-1}$ ,  $380 < T_H^{HB} < 410 \text{ K}$ , and  $360 < T_S^{HB} < 466 \text{ K}$ ), and (2) the authors have studied cyclodextrin, which may differ from peptide bonds. Our best guess is to adopt the convergence temperatures of their "A" data set,  $T_H^{HB} = 380 \text{ K}$  and  $T_S^{HB} = 360 \text{ K}$  (although similar result may be obtained with the other sets) and scale the overall magnitude of the H-bond term so that the enthalpy contribution from the hydrophobic and H-bond terms sum to  $\Delta h_0$ . This is equivalent to using  $\Delta C_p^{HB}$  as a fitting parameter.

The total H-bond free energy is  $n_H \Delta G^{HB}$  where  $n_H$  is the number of H-bonds observed in the crystal structure. The crystal structure 193L<sup>8</sup> shows 15 H-bonds between the reference molecule and its six nearest neighbors, so we set  $n_H = 15$ . Of these, only one is between two charged groups, although we will not discriminate between charge-charge, charge-polar, or polar-polar contacts in our present analysis. The orthorhombic crystal form, 1AKI,<sup>9</sup> has  $n_H = 14$  H-bonds.

In Fig. S2 we plot the enthalpy, entropy, and free energy of the H-bonds in tetragonal crystals. Note that the enthalpy and entropy have opposite slopes compared to the hydrophobic interactions (Fig. S1). This follows from the fact that H-bonds and hydrophobic interactions have heat capacity changes with differing signs.<sup>13</sup> This leads to the fortuitous cancellation of the temperature dependence that allows the temperature independent approximation of Eqs. 5 and 6.

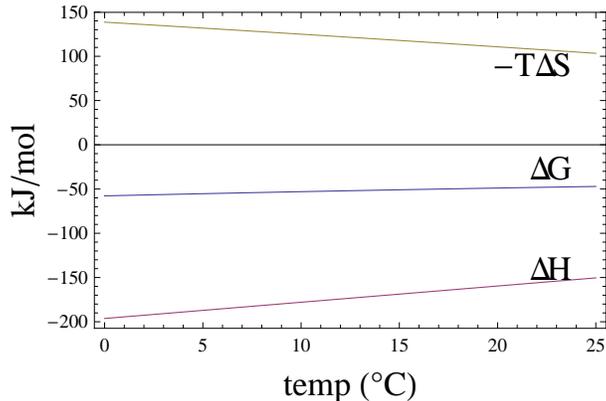


Figure S2: Enthalpy, entropy, and free energy of the H-bonding term in tetragonal lysozyme crystals.

### 1.3 Internal Entropy

One other contributor to the nonelectrostatic contact interactions is the internal entropy, accounting for the restriction of side chains at the crystal contacts,<sup>14</sup> release of bound waters,<sup>15</sup> and the loss of rotational entropy.<sup>16</sup> The latter contribution is independent of the crystal form, while the former two will be roughly proportional to the area buried at the crystal contacts. Since the difference in  $\Delta\text{ASA}$  between the two crystal forms is only  $\sim 20\%$  and water release and side chain restriction are expected to affect the entropy in opposite directions, we expect that  $\Delta S_{int}$  will be similar for the two crystal forms.

### 1.4 Enthalpy and entropy budget

We combine these non-electrostatic contributions to get

$$\begin{aligned}
 H_{tot} &= H_{H\phi} + H_{HB} \\
 S_{tot} &= S_{H\phi} + S_{HB} + S_{int}.
 \end{aligned}
 \tag{S6}$$

For the heat capacities we obtain  $\Delta C_p^{HB} = 122$  and  $110$  J/K/mol, which corresponds to average H-bond energies of  $-3.1$  and  $-2.8$  kJ/mol for tetragonal and orthorhombic crystals, respectively. The internal entropies are  $-235$  and  $-248$  J/K/mol, respectively.

	Tetragonal			Orthorhombic		
	Enthalpy	Entropy	$H - TS$	Enthalpy	Entropy	$H - TS$
Hydrophobic	123.9	564.3	-37.2	111.5	556.4	-58.5
H-bond	-173.3*	-425.3	-51.9	-115.1*	-253.7	-37.6
$S_{int}^{**}$	$\emptyset$	-234.7	67.0	$\emptyset$	-248.4	75.9
Total	-49.4	-95.7	-22.1	-3.6	54.3	-20.2

Table 1: Decomposition of  $\Delta h_0$  and  $\Delta s_0$  into the contributions discussed in the text. Enthalpy and free energy values are given in kJ/mol, entropy values are kJ/K/mol. Values are computed at  $T = 12.5^\circ\text{C}$  for tetragonal crystals and  $T = 32.5^\circ\text{C}$  for orthorhombic.

\* Fit by adjusting  $\Delta C_p$

\*\* Fitting parameter

In Table 1 we show each of the terms in Eqs. S6 after optimizing  $\Delta C_p^{HB}$  and  $S_{int}$  so that  $H_{tot} = \Delta h_0$  and  $S_{tot} = \Delta s_0$ . They show our estimate that the crystals are stabilized by both hydrophobic and hydrogen bonding interactions, roughly at equal magnitudes, and with overall very little net temperature dependence of the contact free energy. Table 1 shows these various contributions for lysozyme in both crystal forms.

We note that the approach presented in this appendix is functionally equivalent to that we presented in the main text. Here the free parameters are the H-bond strength and the internal entropy. In fact, fitting the solubility using these parameters instead of  $\Delta h_0$  and  $\Delta s_0$  reproduces the data with a fidelity comparable to that shown in Figs. 2 and 3. This insensitivity to the details of the model is a consequence of the temperature independence of the non-electrostatic enthalpy and entropy. This temperature independence, and also the wide range of  $\Delta h_0$  and  $\Delta s_0$  values we observe, follow directly from the thermodynamic properties of hydrophobic interactions and H-bonds.

## 2 Linearized approximation

Here we develop an analytic expression for the electrostatic enthalpy change. To do this we employ a linearized (Debye-Huckel) approximation to Eq. 7. While this approximation is quantitatively poor due to the large potentials in the crystal, the qualitative trends are illuminating and it may be useful for weakly charged proteins.

Upon the linearization of Eq. 7 we have

$$\nabla^2\psi(\vec{x}) = \kappa^2\psi(\vec{x}), \quad (\text{S7})$$

where  $\kappa^{-1} = (\epsilon_w\epsilon_0k_B T/2e^2N_{ACbulk})^{1/2}$  is the Debye screening length. The radial solution to Eq. S7 for the geometry shown in Fig. 1 is

$$\psi(r) = A\frac{e^{-\kappa r}}{r} + B\frac{e^{\kappa r}}{r}, \quad (\text{S8})$$

where the constants  $A$  and  $B$  are determined by the boundary conditions. Applying the boundary conditions discussed in the text

$$-\psi'(r = a) = \frac{Ze}{4\pi\epsilon_w\epsilon_0 a^2} \quad (\text{S9})$$

$$-\psi'(r = b) = 0, \quad (\text{S10})$$

we arrive at an expression for the potential

$$\psi^{cryst}(r) = \frac{Ze}{4\pi\epsilon_w\epsilon_0 r} \frac{e^{\kappa(b-r)}(\kappa b - 1) + e^{-\kappa(b-r)}(\kappa b + 1)}{[e^{\kappa(b-a)}(\kappa b - 1)(\kappa a + 1) - e^{-\kappa(b-a)}(\kappa b + 1)(\kappa a - 1)]}. \quad (\text{S11})$$

In the limit  $b \rightarrow \infty$  we recover the Debye-Huckel result for an ion in solution

$$\psi^{sol}(r) = \frac{Ze}{4\pi\epsilon_w\epsilon_0 r} \frac{e^{-\kappa(r-a)}}{(\kappa a + 1)}, \quad (\text{S12})$$

which, at the protein surface, can be written as

$$\psi^{sol}(a) = \frac{Ze}{4\pi\epsilon_w\epsilon_0} \left( \frac{1}{a} - \frac{\kappa}{\kappa a + 1} \right). \quad (\text{S13})$$

In Eq. S13 the first term is instantly recognized as the potential due to the protein macroion,  $\psi_p(r) = Ze/4\pi\epsilon_0\epsilon_w r$ , and therefore the second term is the potential due to the associated salt ions. The interaction energy between the soluble protein and the salt ions is then

$$E_{sol} = -\frac{Z^2 e^2 \kappa}{4\pi\epsilon_w\epsilon_0(\kappa a + 1)}. \quad (\text{S14})$$

In the crystal state the situation is somewhat more complex. Here the electrostatic potential can

be decomposed as in Eq. 13

$$\psi^{crys}(r) = \psi_p(r) + \psi_s^{crys}(r) + \psi_0^{crys}, \quad (S15)$$

where the three terms on the right represent the potential due to the protein, salt, and crystal surface charges, respectively. The electrostatic potential energy of the protein interacting with its ion cloud is  $Ze\psi_s(a)$ , or

$$E_{crys} = Ze(\psi^{crys}(a) - \psi^{crys}(b) - \psi_p(a)). \quad (S16)$$

The change in the interaction energy between the crystal and solution states is

$$\Delta h_{ES} = E_{crys} - E_{soln} \quad (S17)$$

$$= Ze(\psi^{crys}(a) - \psi^{crys}(b) - \psi^{sol}(a)) \quad (S18)$$

$$= \frac{Z^2 e^2}{4\pi\epsilon_w\epsilon_0} \frac{2\kappa[e^{-\kappa(b-a)}(\kappa b + 1) - (\kappa a + 1)]}{(\kappa a + 1)[e^{\kappa(b-a)}(\kappa b - 1)(\kappa a + 1) - e^{-\kappa(b-a)}(\kappa b + 1)(\kappa a - 1)]}. \quad (S19)$$

### 3 Fitting procedure

To determine the free parameters  $\Delta h_0$  and  $\Delta s_0$  we define the sum of square errors

$$SSE = \sum_i (\ln(c_i/M_w) - \mu(pH_i, [NaCl]_i, T_i, \Delta h_0, \Delta s_0))^2 \quad (S20)$$

where  $c_i$  is the protein solubility in mg/ml,  $M_w = 14,400$  is the molecular weight,  $\mu$  is given by Eq. 27, and the sum is over the 598 data points shown in Fig. 2. A similar procedure was followed for orthorhombic crystals, but since the experimental data was presented in the form of quadratic fits, the sum over  $i$  was done at 3K increments at each pH and salt concentration. The  $SSE$  is plotted in Fig. S3 as a function of  $\Delta h_0$  and  $\Delta s_0$ . It is readily evident that the  $SSE$  has a prominent valley. The valley floor is inclined at an angle  $\Delta\Delta h_0/\Delta\Delta s_0 = 285K$  corresponding to the mid point of the temperature range studied. For convenience we rescale the  $\Delta s_0$  axis by this value so that both axes have the same units. The second derivatives at the minimum  $SSE = 25.0$  are then  $5.48 \times 10^{-8}(\text{kJ/Mole})^{-2}$  along the valley floor and  $4.25 \times 10^{-4}(\text{kJ/Mole})^{-2}$  in the perpendicular direction.

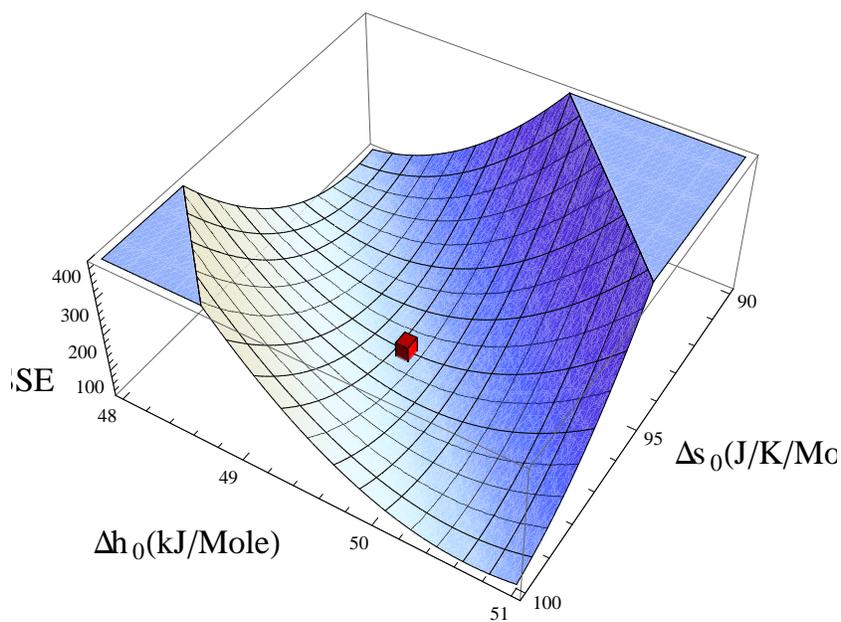


Figure S3: Plot of the sum of square errors vs. the fitting parameters  $\Delta h_0$  and  $\Delta s_0$ . The minimum is indicated by the red square.

## References

- [1] Livingstone, J.; Spolar, R.; Record, M. **1991**, *30*, 4237 - 4244.
- [2] Spolar, R.; Ha, J.; Record, M. *Proc. Nat. Acad. Sci.* **1989**, *86*, 8382 - 8385.
- [3] Spolar, R.; Livingstone, J.; Record, M. **1992**, *31*, 3947 - 3955.
- [4] Makhatadze, G.; Privalov, P. *J. Mol. Biol.* **1993**, *232*, 639 - 659.
- [5] Privalov, P.; Makhatadze, G. *J. Mol. Biol.* **1993**, *232*, 660 - 679.
- [6] Makhatadze, G.; Privalov, P. *Biophys. Chem.* **1994**, *51*, 291 - 309.
- [7] Privalov, P.; Makhatadze, G. *FASEB J.* **1992**, *6*, A344 - A344.
- [8] Vaney, M.; Maignan, S.; Ries-Kautt, M.; Ducriux, A. *Acta Crystallogr.* **1996**, *52*, 505-517.
- [9] Artymiuk, P.; Blake, C.; Rice, D.; Wilson, K. *Acta Crystallogr.* **1982**, *38*, 778.

- [10] Baldwin, R. *J. Mol. Biol.* **2007**, *371*, 283-301.
- [11] Ross, P.; Rekharsky, M. *Biophys. J.* **1996**, *71*, 2144-2154.
- [12] Gao, J.; Bosco, D.; Powers, E.; Kelly, J. *Nat. Struct. Mol. Biol.* **2009**, *16*, 684 - 690.
- [13] Richardson, J.; Makhatadze, G. *J. Mol. Biol.* **2004**, *335*, 1029-1037.
- [14] Wang, J.; Szewczuk, Z.; Yue, S.; Tsuda, Y.; Konishi, Y.; Purisima, E. *J. Mol. Biol.* **1995**, *253*, 473-492.
- [15] Vekilov, P.; Feeling-Taylor, A.; Yau, S.-T.; Petsiv, D. *Acta. Cryst.* **2002**, 1611 - 1616.
- [16] Irudayam, S.; Henchman, R. *J. Phys. Chem. B* **2009**, DOI: 10.1021/jp809968p.