# Supporting Information for "Growth rates of protein crystals"

Jeremy D. Schmit and Ken Dill

December 13, 2011

## 1 Layer Nucleation and Growth Rate

Here we give a detailed derivation of the growth and nucleation formulas used in the manuscript. These formulas are based on classical nucleation theory. In classical nucleation theory, a droplet is driven to form by a bulk volumetric attraction among particles and is opposed by the creation of surface, which is unfavorable. The latter is expressed as a surface tension $\gamma$. Since we are interested in the formation of two dimensional clusters, $\gamma$ has units of energy/length. When a droplet has a small number of particles, $n < n_0$, the surface tension dominates and the droplet is unstable. However, when the size of the droplet nucleus exceeds the critical size $n > n_0$, the bulk term dominates, and the droplet grows. We now formulate the kinetics of protein-crystal formation in those terms.

The driving force for aggregation is given by $\Delta\mu$, the chemical potential difference between the crystal and solution states. In the solution state the chemical potential is given by $\mu_{\rm sol} = \mu^\circ + \ln c$ where $\mu^\circ$ is the chemical potential at an arbitrary reference concentration. At equilibrium the chemical potentials must be equal between the solution and the crystal phases. This means $\mu_{\rm xtal} = \mu^\circ + \ln c_s$. At concentrations greater than or less than $c_s$ the chemical potentials of the two states are unequal. This provides a net driving force for crystal growth/dissolution which is given by

$$\Delta\mu = \mu_{\rm sol} - \mu_{\rm xtal} \tag{S1}$$

$$= \ln \frac{c}{c_s}. \tag{S2}$$

A circular nucleus containing of radius $r$ has a free energy

$$\Delta F(r) = -\pi(r/d)^2\Delta\mu + 2r\gamma/d. \tag{S3}$$

The first term on the right is the bulk driving force and the second term is the surface tension. Alternatively, we can express this in terms of the number of proteins in the cluster $n = \pi r^2/d^2$

$$F(n) = -\Delta\mu n + \gamma 2\pi\sqrt{nd^2/\pi}. \tag{S4}$$

This free energy has a maxima at the critical size $n_0$ defined by $[dF(n)/dn]_{n_0} = 0$, which gives

$$F^{\ddagger} = F(n_0) = \frac{\pi\gamma_2^2 d^2}{\Delta\mu} \tag{S5}$$

We also require the second derivative with respect to $n$

$$F'' = -\frac{\gamma d}{2}\sqrt{\frac{\pi}{n^3}}. \tag{S6}$$

As shown by Zeldovich,[1] the second derivative describes the random walk of the nucleus size in the neighborhood of the barrier. Briefly, in systems with more sharply peaked free energy barriers (more negative second derivatives), supercritical clusters are less likely to recross the barrier and become subcritical.

The nucleation rate is given by Zeldovich[1–3]

$$J = we^{-F^{\ddagger}/k_{\mathrm{B}}T}\sqrt{\frac{-F''(n_0)}{2\pi k_{\mathrm{B}}T}}, \tag{S7}$$

Here the middle term is an Arrhenius factor describing the probability of a cluster reaching the critical size $n_0$ (given by $F'|_{n_0} = 0$). The final term is the Zeldovich factor $Z = (-F''/2\pi k_{\mathrm{B}}T)^{1/2}$ describing the probability that a supercritical cluster continues to grow without dissolving back to subcritical size. The prefactor $w$ is the rate at which proteins add to the critical nucleus. Within our model this is given by the monomer addition rate multiplied by the number of sites on the cluster perimeter

$$w = r_+\frac{2\pi}{d}\left(\frac{d^2\gamma}{\Delta\mu}\right) \tag{S8}$$

where the factor in parentheses is the critical radius.

The surface growth rate is given by the time $t$ required to complete a layer. Upon the nucleation of a cluster, the perimeter advances at a rate $v_{\rm step} = (r_+ - r_-)d$, so after time $\tau$ the cluster has an area $\pi(v_{\rm step}\tau)^2$. Since nuclei are continuously forming, a completed layer is a collection of clusters ranging in age from $0$ to $t$. If the crystal has a surface area $A$, the cluster nucleation rate is $r_{\rm nuc} = JA/d^2$. Then the area of the clusters that nucleated between time $\tau$ and $\tau + d\tau$ prior to the completion of the layer is $r_{\rm nuc}\pi(v_{\rm step}\tau)^2 d\tau$. Adding up the areas of clusters of all ages we have

$$A = \int_0^t J\frac{A}{d^2}\pi(v_{\rm step}\tau)^2 d\tau, \tag{S9}$$

which yields a layer completion time that is independent of $A$

$$t = \left(\frac{3d^2}{\pi J v_{\rm step}^2}\right)^{1/3} \tag{S10}$$

$$= \left(\frac{3}{\pi J(r_+ - r_-)^2}\right)^{1/3}. \tag{S11}$$

Combining Eqs. S7 and S11 we arrive at an expression for the growth rate (in nm/sec)[2]

$$V = \frac{d}{t} \tag{S12}$$

$$= r_+ d\left(\frac{\pi}{3}\right)^{1/3}\left(1 - \frac{r_-}{r_+}\right)^{2/3}\left(\frac{\Delta\mu}{k_{\rm B}T}\right)^{1/6} e^{-\frac{\pi\gamma_2^2 d^2}{3\Delta\mu k_{\rm B}T}}, \tag{S13}$$

which is Eq. 2 in the manuscript.

## 2 Fitting procedure and results

The parameters $\gamma$, $d$, and $c_s$, were obtained from the literature. $d = 3.8$nm for the dimension of lysozyme perpendicular to the (110) crystal face,[4] and $c_s$ was computed using the method in.[5] Analysis of the (101) face growth rates[6] yields similar trends and comparable parameters (data not shown). The surface tension is more conveniently expressed as the product $d\gamma$, the interface energy per protein, given in[7] to be $7.2$ kJ/mol.

The fitting parameters $f_b$ and $k$ were obtained from a two-parameter fit to the data in Fig 2b. Growth rates at 3% NaCl (Figs 2a) were obtained using the same value of $f_b$, however, it was necessary to to re-fit $k$, a fact that can be justified on the grounds that the expanded Debye layer

would hinder binding attempts. In total, Figs. 2 and 3 feature 17 experimental curves that we model using three free parameters.

## 3 Effect of salt and pH on attempt rate

A curious result of the fitting procedure is the finding that the attempt rate $k$ depends strongly on salt concentration, but only weakly on pH. We can understand this result using a crude model for the distance dependence of the potential energy of a protein approaching the crystal surface. We begin by expressing the potential as the sum of an attractive term and a repulsive term

$$U(d) = U_{\rm att}(d) + U_{\rm rep}(d) \tag{S14}$$

where $d$ is the closest approach between the protein and the crystal. The repulsive term has a maximum value at $d = 0$ and decays on a length scale corresponding to the Debye length. Thus, we write

$$U_{\rm rep}(d) = \Delta f_{\rm es} e^{-\kappa_D d} \tag{S15}$$

where $\kappa_D^{-1}$ is the Debye screening length and $\Delta f_{\rm es}$ is given by Eq. 7. We assume that the attractive term arises primarily from hydrophobic interactions and desolvation effects. The distance dependence of these forces is not clear, so we employ two different approximations; a square well

$$U_{\rm att}^{\rm SW}(d) = \begin{array}{ll} -f_b & d \le d_w \\ 0 & d > d_w, \end{array} \tag{S16}$$

and an exponential decay

$$U_{\rm att}^{\rm exp}(d) = -f_d e^{-d/d_w}. \tag{S17}$$

In both cases we choose the range of the interaction to be comparable to the diameter of a water molecule $d_w = 0.2$nm reflecting the dominant role of solvent degrees of freedom in mediating these forces. In Fig. S1 we plot $U(d)$ for each of these choices of the attractive interaction. The square well model (Fig. S1a) confirms our finding that the barrier depends weakly on pH ($\sim 30\%$) and strongly on salt (by a factor of 2). The exponential attraction also show the effect of pH to be less important than salt, but here the changes in the barrier height are factors of 2 and 4 for pH

and salt, respectively.

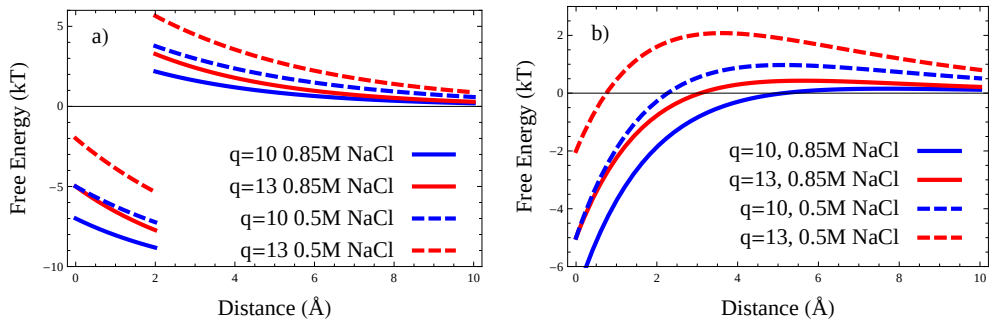


Figure S1: Plot of Eq. S14 as a function of distance $d$ for two choices of the attractive interaction: a) square well, and b) exponential decay.

## 4 Cooperative NP binding

The model presented in the manuscript fails to quantitatively capture the crystal growth rates at high concentrations. This discrepancy is shown in Fig. S3a, which shows the growth rates abruptly plateau at a critical concentration. This is not captured by our model. We attribute this to a failure of the approximation in Eq. 3 that the potential binding sites are all independent. As the occupancy of NP proteins approaches unity this approximation becomes increasingly poor. We can correct for this by accounting for the free energy of interaction between NP proteins on adjacent binding sites

$$f_{\text{int}}(i) = \frac{1}{2}\sum_{\{j\}} \delta n_j \tag{S18}$$

where $\delta$ is the free energy of interaction between two NP proteins, the summation is over the binding sites adjacent to site $i$, and $n_j$ takes the value 1 if site $j$ is occupied by an NP protein and zero otherwise. For a growing crystal the potential binding sites are arranged on the perimeter of a growing 2D layer on the crystal surface, see Fig. S2. This means that the potential binding sites form a 1D array and the occupancy of NP proteins can be solved with a transfer matrix formalism. If there are $N$ binding sites on the perimeter of the cluster the partition function is $Q = \text{Tr}(M^N)$

where

$$M = \begin{pmatrix} 1 & s \\ 1 & s\sigma \end{pmatrix} \tag{S19}$$

and $s = e^{-f/k_BT}$ and $\sigma = e^{-\delta/k_BT}$. Using the cyclical property of the trace we can evaluate the partition function by diagonalizing $M$

$$Q = \lambda_+^N + \lambda_-^N \tag{S20}$$

where the eigenvalues are

$$\lambda_\pm = \frac{1}{2}(s\sigma + 1 \pm \sqrt{(s\sigma+1)^2 - 4s(\sigma-1)}). \tag{S21}$$

For large $N$ we can neglect the smaller eigenvalue and the blocking probability is

$$1 - P_{\text{free}} = \langle n \rangle \simeq \frac{d \ln \lambda_+}{d \ln s} \tag{S22}$$

$$= \frac{s}{2\lambda_+}\left(\sigma + \frac{\sigma(s\sigma+1) - 2(\sigma-1)}{\sqrt{(s\sigma+1)^2 - 4s(\sigma-1)}}\right). \tag{S23}$$

The monomer binding rate is then

$$r_+ = kcP_{\text{free}} \tag{S24}$$

with $P_{\text{free}}$ given by Eq. S23.

In Fig. S3b we compare experimental growth rates at 5% NaCl and $14^\circ$C to the theory given by Eqs. S13, and S23 with $f_b$, $k$, and $\delta$ as fitting parameters. The theory now accurately reproduces the plateau at high concentrations, although the pH dependence is somewhat too strong. This is to be expected due to two approximations in our electrostatic model: 1) our neglect of the favorable interaction with the protein upon counterion condensation,[5] and 2) $v_w$ is likely an underestimate of the volume accessible to ions associated with NP proteins. Fig. S3c compares the temperature dependence of the growth rate to the theoretical prediction. This plot uses the parameters of Fig. S3b with no further fitting. The growth rates show a non-monotonic behavior with respect to temperature which can be understood as follows. At high temperatures, lowering the temperature increases the growth rate by increasing the supersaturation. However, lowering the temperature further eventually leads to a reduction in the growth rate due to increasing populations of NP

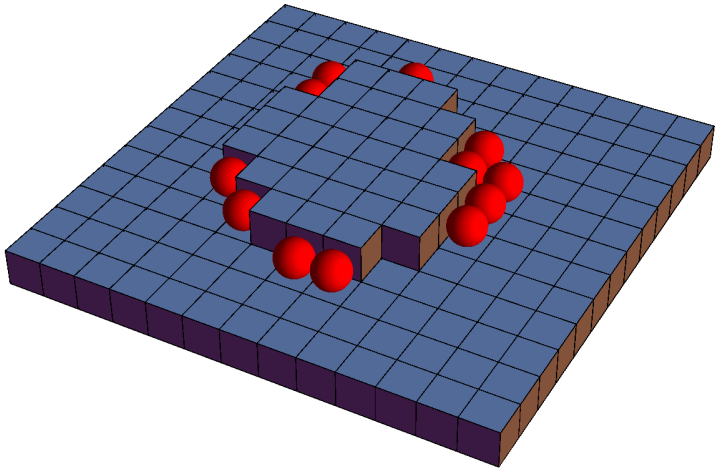

Figure S2: NP proteins on the perimeter of a growing 2D cluster on the surface of a crystal. Red spheres represent the ensemble of disordered NP states. Neighboring spheres have a favorable binding interaction $\delta$.

proteins. This non-monotonicity is not seen at low protein concentrations due to the lower population of NP proteins. Our cooperative binding model correctly captures these trends without fitting.

## 5 Tetramer growth units

It has been suggested that the growth of lysozyme crystals proceeds via the addition of oligomeric growth units.[4,8,9] In particular, a tetramer growth unit is consistent with AFM observations of the growth process.[10] However, this idea has been very controversial.[11–13] To assess this mechanism we can generalize Eq. 3 to allow for the addition of tetramer growth units.

The concentration of tetramers will be

$$c_4 = c_1^4 e^{-f_4/k_{\mathrm{B}}T}, \tag{S25}$$

where $c_1$ is the monomer concentration and $f_4$ is the free energy of the tetramer. Then we would expect the addition rate to be proportional to $k_4 c^4 c^{-f_4/k_{\mathrm{B}}T} = k_4' c^4$ where $k_4$ is the tetramer attempt rate. To find the success rate we need to know the occupancy of the NP state. To do this, we require the relative concentrations of monomers and tetramers in the solution.

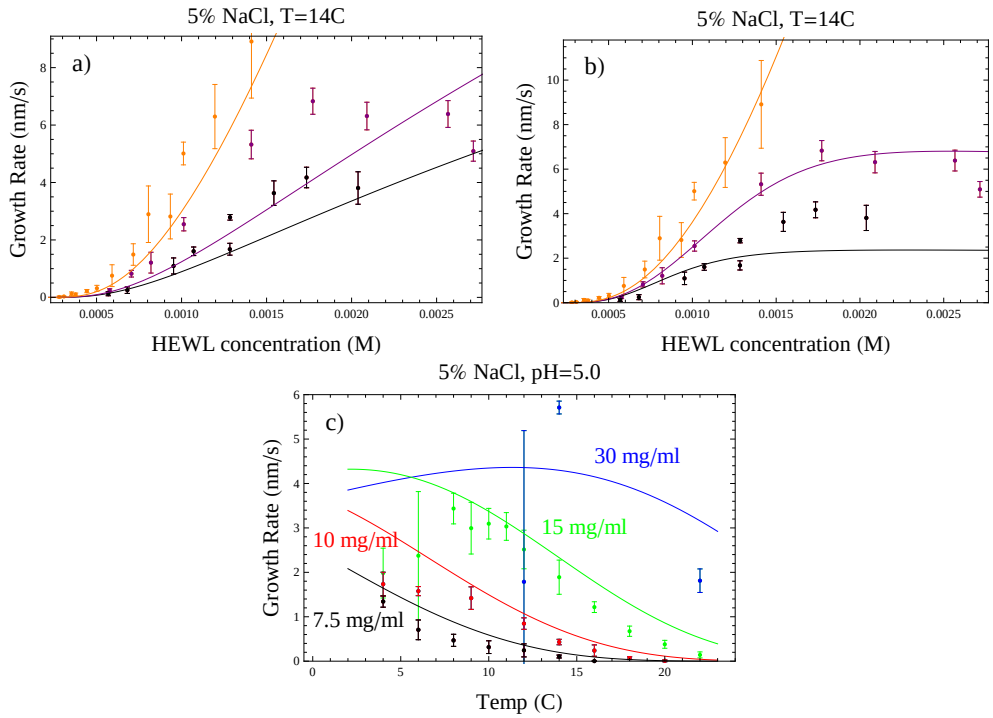


Figure S3: Growth rates of the (110) face of tetragonal lysozyme crystals compared to Eq. S13 without (a) and with (b,c) accounting for interactions between NP proteins. Colors in panels (a,b) indicate different pH conditions as follows: 4.0 (orange), 4.8 (purple), 5.4 (black). $k = 6.8 \times 10^4\ \mathrm{sec}^{-1}\mathrm{Mol}^{-1}$, $f_b = -19$ kJ/mol, $\delta = -5.0$ kJ/mol

To estimate $f_4$ we note that the tetramer contains four protein-protein contacts[9] compared to six for a protein in the crystal. Therefore, we generously estimate the free energy of the tetramer using the binding free energy of a protein in the crystal $\sim 10k_{\mathrm{B}}T$.[5] Plugging this value into Eq. S25 we find that the concentrations of monomers and tetramers will be equal when $c_1 \simeq 0.1$M. Since this is two orders of magnitude larger than the concentration range of interest, we conclude that the majority of proteins in solution will be monomers, and likewise we expect the majority of NP proteins will be monomers as well. Neglecting the possibility of NP tetramers, we can write

$$P_{4free} = \frac{1}{(1 + ce^{-f/k_BT})^4}. \tag{S26}$$

Eq. S26 says that all four binding sites must be free of NP proteins in order for a tetramer to bind. So, the on rate for tetramer binding is

$$r_{4+} = \frac{k'c^4}{(1 + ce^{-f/k_BT})^4}, \tag{S27}$$

and the detachment rate of the four protein growth unit is

$$r_{4-} = \frac{k' c_s^4}{(1 + c_s e^{-f/k_B T})^4}. \tag{S28}$$

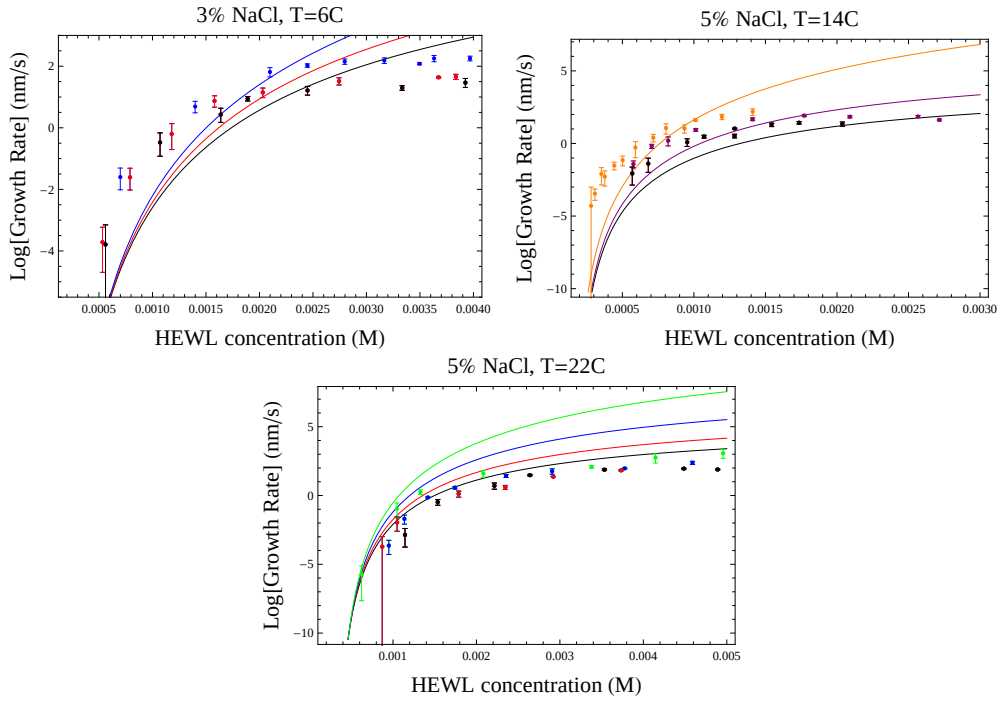


Figure S4: Growth rates of the (110) face of tetragonal lysozyme crystals compared to Eq. S13 for tetramer addition. Colors indicate different pH conditions as follows: 4.0 (orange), 4.2 (green), 4.6 (blue), 4.8 (purple), 5.0 (red), 5.4 (black).

The theoretical predictions of Eq. S27 with Eq. S13 are shown in Fig. S5. The tetramer model fails to capture the extent of the plateau behavior at high concentrations because, due to the requirement that all four sites must be free for a tetramer to bind, $P_{\text{free}}$ is of order $10^{-4}$ and varies weakly over the concentration range of interest. These plateaus were previously attributed to a change in the growth mechanism from the nucleation of 2D clusters to random addition of proteins on the crystal surface.[6,7,14,15] We favor the monomer addition model with the NP-induced plateau for two reasons. First, the layer nucleation mechanism is expected to fail when the critical size of the nucleus reaches the size of single growth unit. For a tetramer this critical value $n_0 = 4$ coincides with the sharp increase in the growth rates (0.0005-0.001M) rather than the onset of the plateau (0.002-0.004M). Therefore, a tetramer growth model would be expected to obey random deposition growth at all concentrations. Secondly, a random deposition growth model would be

expected to give a linearly increasing growth rate with increasing concentrations. Instead, the growth rates appear to plateau and, in some cases, decline with increasing concentrations. This decline is more easily explained by the cooperative deposition of NP proteins presented in the previous section. Our model does not explain the AFM experiments showing that the growth layer expands by two proteins at a time.[10] However, these experiments were done at low concentrations where the density of NP proteins is negligible, and since the our fit attempt rate $kc$ is comparable to the 50msec temporal resolution of the experiment, only a minor asymmetry in the on rates of the two layers is necessary to explain the observation.

We conclude that the growth of tetragonal lysozyme crystals is most likely via the addition of monomers, as only this mechanism is able to capture both the low and high concentration growth behavior in a single model.

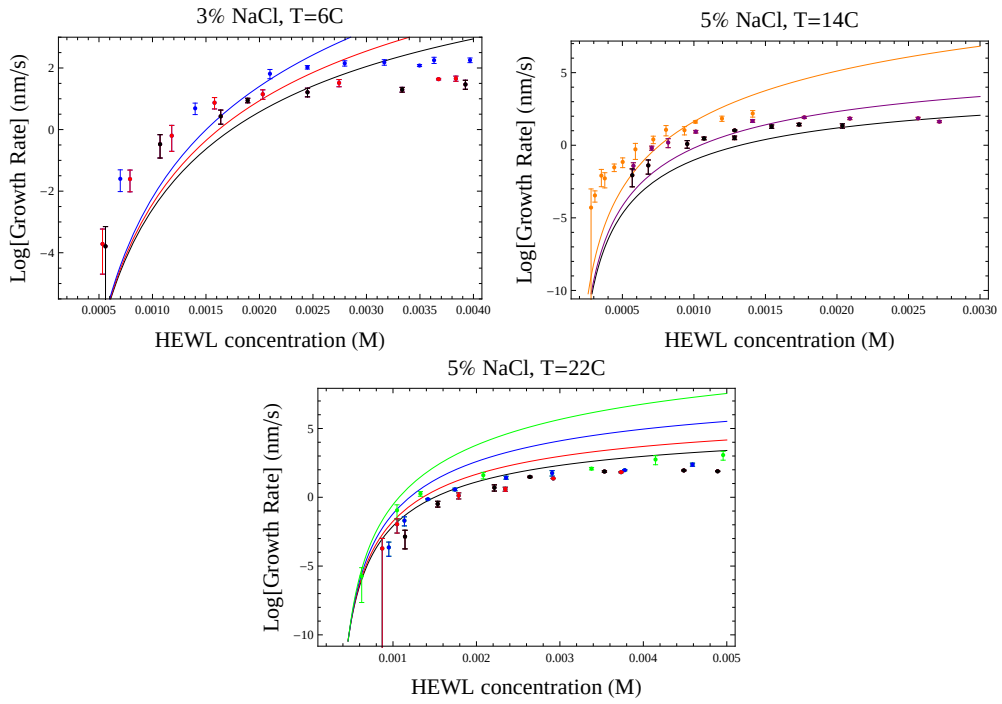


Figure S5: Growth rates of the (110) face of tetragonal lysozyme crystals compared to Eq. S13 for tetramer addition. Colors indicate different pH conditions as follows: 4.0 (orange), 4.2 (green), 4.6 (blue), 4.8 (purple), 5.0 (red), 5.4 (black).

## References

[1] Zeldovich, J. *JETP* **1942,** *12,* 525 (English translation: Acta Physicochim. URRS 18, 1 (1943)).

[2] Saito, Y. *Statistical physics of crystal growth;* World Scientific: Singapore, 1996.

[3] Kashchiev, D.; van Rosmalen, G. *Cryst. Res. Technol.* **2003,** *38,*.

[4] Nadarajah, A.; Pusey, M. *Acta Cryst.* **1996,** 983-996.

[5] Schmit, J.; Dill, K. *J. Phys. Chem. B* **2010,** *114,* 4020.

[6] Gorti, S.; Forsythe, E.; Pusey, M. *Cryst. Growth Des.* **2005,** *5,* 473 - 482.

[7] Gorti, S.; Konnert, J.; Forsythe, E.; Pusey, M. *Cryst. Growth Des.* **2005,** *5,* 535 - 545.

[8] Li, M.; Nadarajah, A.; Pusey, M. *J. Cryst. Growth* **1995,** *156,* 121-132.

[9] Pusey, M.; Nadarajah, A. *Cryst. Growth Des.* **2002,** *2,* 475-483.

[10] Li, H.; Nadarajah, A.; Pusey, M. *Acta. Cryst.* **1999,** 1036-1045.

[11] Muschol, M.; Rosenberger, F. *J. Cryst. Growth* **1995,** *167,* 738-747.

[12] Grimsley, G.; Scholtz, J.; Pace, C. *J. Cryst. Growth* **1997,** *178,* 575-584.

[13] Vekilov, P.; Rosenberger, F. *J. Cryst. Growth* **1996,** *158,* 540-551.

[14] Gorti, S.; Forsythe, E.; Pusey, M. *Cryst. Growth Des.* **2004,** *4,* 691 - 699.

[15] Gorti, S.; Forsythe, E.; Pusey, M. *Acta. Cryst.* **2005,** 837-843.