Biophysical Journal Volume 100

Supporting Material

Title: What drives amyloid molecules to assemble into oligomers and fibrils?

Authors: Jeremy D Schmit, Kingshuk Ghosh, and Ken Dill

Supporting Materials for "What drives amyloid molecules to assemble into oligomers and fibrils?"

Jeremy D. Schmit, Kingshuk Ghosh, and Ken Dill

Critical fibril concentration (CFC)

The concentration of peptides in the fibril state is given by the final term in Eq. 7

$$c_{fibril} = \gamma^{\ell p} \sum_{n=n_0}^{\infty} n c_1^n g^{nL} \tag{S1}$$

$$= \gamma^{\ell p} c_1 g^L \frac{d}{d(c_1 g^L)} \sum_{n=n_0}^{\infty} c_1^n g^{nL} \tag{S2}$$

$$= \gamma^{\ell p} \frac{n_0(1-c_1 g^L)(c_1 g^L)^{n_0} + (c_1 g^L)^{n_0+1}}{(1-c_1 g^L)^2}. \tag{S3}$$

The sum in Eq. S1 converges when the argument is less than unity, therefore the monomer concentration must satisfy $c_1 < g^{-L}$ if the bulk concentration is to remain finite. This radius of convergence defines a concentration

$$c_1^{(CFC)} \sim g^{-L}, \tag{S4}$$

which may be interpreted as the CFC for the fibril solution and is plotted as the monomer-fibril boundary in Fig. 4. For c_1 much less than this value the concentration of monomers in the fibril state c_{fibril} is strongly suppressed by the factor $\gamma^{\ell p}(c_1 g^L)^{n_0}$. However, as c_1 approaches g^{-L} Eq. S1 diverges. In this regime we can write $c_1 g^L = 1 - \delta$ and the fibril concentration becomes

$$c_{fibril} \simeq \gamma^{\ell p} \frac{(1-n_0\delta)(1+n_0\delta-\delta)}{\delta^2} \tag{S5}$$

$$= \gamma^{\ell p} \frac{c_1 g^L + \mathcal{O}(\delta^2)}{(1-c_1 g^L)^2}, \tag{S6}$$

which demonstrates that c_{fibril} is insensitive to the lower limit n_0.

Note that our definition for the CFC differs from that given in reference [1].

Critical oligomer concentration (COC)

Here, we calculate the concentration of oligomers both in the presence and absence of fibrils. In the presence of fibrils we can approximate the concentration of peptides in the oligomer state, c_{oligo}, by using the CFC in the second term of Eq. 7 since the monomer concentration varies little in the vicinity of the fibril CFC

$$c_{oligo} \sim N g^{-LN} e^{\chi LN}. \tag{S7}$$

This expression is only valid provided $e^{\chi} < g$. If this condition is not satisfied, then Eq. 2 is a monotonically increasing function of m meaning the oligomer state has a lower free energy per peptide than a fibril of any length. In this case no fibrils will be formed, and the solution will be an equilibrium mixture of oligomers and monomers. We define the COC to be the concentration where the oligomer and monomer states have equal occupancies. Using the appropriate terms from Eq. 7 we find the COC given by

$$c_1^{(COC)} = \left(\frac{e^{-\chi NL}}{N}\right)^{1/(N-1)}. \tag{S8}$$

This expression is plotted as the monomer-oligomer boundary in Fig. 4.

Fibril-oligomer boundary

The boundary between the fibril and oligomer phases is defined, for points suitably removed from the monomer phase, by the condition $c_{fibril} = c_{oligo} \simeq c_0/2$. Using Eqs. S6 and S7 we have

$$c_0/2 = \gamma^{\ell p} \frac{c_1 g^L}{(1 - c_1 g^L)^2} \tag{S9}$$

$$c_0/2 = N c_1^N e^{\chi NL}. \tag{S10}$$

Eq. S9 yields a recursive formula for c_1, which to lowest order gives

$$c_1 \simeq g^{-L}(1 - \sqrt{2\gamma^{\ell p}/c_0}) \tag{S11}$$

which can be combined with Eq. S10 to yield a condition for the phase boundary

$$\frac{\ln g}{\chi} = \frac{NL \ln g}{\ln\left(\frac{c_0}{2Ng^{-L}}\right) - N \ln\left(1 - \sqrt{2\gamma^{\ell p}/c_0}\right)}. \tag{S12}$$

This expression is plotted with long dashes in Fig. 4.

Average fibril length

The critical concentrations for fibril and oligomer formation, $c_1^{(CFC)}$ and $c_1^{(COC)}$ are notably lacking a dependence on the nucleation parameter γ. While this parameter has little effect on the relative stability of the fibril and oligomer phases, we expect that it will play a large role in determining the equilibrium lengths of mature fibrils. To see this we consider a system that is deep within the regime where fibrils are the dominant species so that $c_0 \simeq c_{fibril}$. Using Eq. S6 we find

$$c_1 \simeq g^{-L}(1 - \sqrt{\gamma^{\ell p}/c_0}). \tag{S13}$$

The jth moment of the fibril length distribution is given by

$$\langle n^j \rangle = \frac{\gamma^{\ell p} \sum_n n^j c_1^n g^{nL}}{\gamma^{\ell p} \sum_n c_1^n g^{nL}} \tag{S14}$$

$$= \left(\frac{(c_1 g^L)^{n_0}}{1 - c_1 g^L}\right)^{-1} \left(c_1 g^L \frac{d}{d(c_1 g^L)}\right)^j \left(\frac{(c_1 g^L)^{n_0}}{1 - c_1 g^L}\right). \tag{S15}$$

The average length is given by the first moment $j = 1$

$$\langle \ell \rangle = n_0 + \frac{c_1 g^L}{1 - c_1 g^L} \tag{S16}$$

$$\sim c_0^{1/2} \gamma^{-\ell p/2} + const, \tag{S17}$$

where Eq. S13 has been used to extract the scaling behavior.

The lack of a dependence on g in this scaling relation is a result of an approximation based on $c_0 \gg c_1^{(CFC)}$. When this assumption is satisfied the large majority of protein is in the fibril state, and the problem of determining the fibril lengths is reduced to a question of fibril breakage statistics. Since each breakage incurs a statistical penalty γ^{ℓ_p}, the functional form of Eq. S17 is not surprising (the numerical factor in the exponent is a result of the degeneracy of breakage points). For systems near the onset of fibrillization $c_0 \simeq c_1^{(CFC)}$ the fibril lengths depend sensitively on g and Eq. S14 must be used to model the lengths.

Denaturants

Denaturants destabilize the folded states of proteins by weakening hydrophobic interactions and weakening peptide-peptide H-bonds relative to peptide-solvent H-bonds [2]. To capture the effect of denaturants on the interaction free energy quantities g and γ, we use

$$\begin{aligned} kT \ln g_c &= kT(\ln g - a_0 c_d) \\ kT \ln \gamma_c &= kT(\ln \gamma + a_0 c_d), \end{aligned} \tag{S18}$$

where g_c and γ_c are the propagation parameters in the presence of denaturant, c_d is the denaturant concentration and a_0 is a constant describing the destabilizing effect of the osmolyte on the fibril. This form reflects the fact that the denaturant weakens the H-bonds captured in g, but also reduces the fibril end free energy $-\ln\gamma$, which arises largely from unsatisfied H-bonds.

Using Eqs. S4, S8, and S18 we can compute how the critical concentrations will shift as a function of denaturant concentration. At the onset of fibrillization we have $c_0 = c_1^{(CFC)}$, so from Eqs. S4 and S18 we have

$$\ln c_1 = -L \ln g_c \tag{S19}$$

$$= -L(\ln g - a_0 c_d), \tag{S20}$$

which can be solved for c_d to give the phase boundary

$$c_d = \frac{\ln c_0 + L \ln g}{L a_0}. \tag{S21}$$

Similarly, for the oligomer state we write

$$\chi_c = \chi - a_1 c_d \tag{S22}$$

where a_1 has been introduced to reflect the fact that since the oligomers are more dependent on hydrophobic interactions and less dependent on H-bonds for stability, and therefore, the destabilization coefficient will, in general, be different. The onset of oligomerization may be determined from Eq. S8

$$c_0 = \left(\frac{e^{-NL(\chi - a_1 c_d)}}{N} \right)^{1/(N-1)}, \tag{S23}$$

which can be rearranged to yield

$$c_d = \frac{1}{a_1} \left(\chi + \frac{(N-1)\ln c_0 + \ln N}{NL} \right). \tag{S24}$$

The fibril-oligomer boundary can be derived from Eqs. S9 and S10

$$\ln c_0/2N = NL(\chi - a_1 c_d) - NL(\ln g - a_0 c_d), \tag{S25}$$

where we have dropped the correction term in Eq. S9. In Fig. 8 we take $a_1 = a_0/3$ reflecting our expectation that the hydrophobic interactions stabilizing the oligomer are less affected by the presence of denaturant than the H-bonds stabilizing the fibril [3, 4]. However, the particular choice of $a_0/3$ is for illustration purposes. With this approximation for a_1 Eq. S25 becomes

$$c_d = \frac{3}{2a_0 NL}\left(\ln\frac{c_0}{2N} - NL(\chi + \ln g)\right). \tag{S26}$$

Based on denaturation studies, we expect that $a_0 = 0.022M^{-1}$ for urea and 0.042 for guanidinium [4]. Using our estimates of $L\ln g = 13.1$ and $\ell p \ln\gamma = -15.5$ (in the absence of urea) for Aβ from our previous analysis of fraction fibril as a function of concentration (Fig. 5) we can predict the fibril fraction as a function of urea using Eq. S18 without a fit parameter. Our prediction is compared with the experimental data in Table 1.

Electrostatics

To compute F_{es} we approximate the fibril as a smooth cylinder of radius R and uniform charge density. The linear charge density may be computed by noting that the average charge per β-strand is q/n_s and there are $2p$ strands per layer in the fibril.

To determine the peptide charge as a function of pH, we use

$$q = \sum_i^{\text{acidic residues}} -\frac{10^{pH-pKa_i}}{1+10^{pH-pKa_i}} + \sum_i^{\text{basic residues}} +\frac{10^{pKa_i-pH}}{1+10^{pKa_i-pH}}, \tag{S27}$$

where the pKas of the amino acids are taken from Ref. [5]. So, the charge density on each peptide-molecule cylinder is $\rho = 2qp/n_s a$, where $a = 4.7$ Å is the spacing between β-strands.

We then solve for the electrostatic potential ψ using the Poisson-Boltzmann equation

$$\epsilon\frac{1}{r}\frac{\partial}{\partial r}r\frac{\partial\psi}{\partial r} = -e(c_+ - c_-), \tag{S28}$$

where the ion concentrations are $c_\pm = c_s e^{\mp e\psi/kT}$. In the linearized (Debye-Huckel) approximation, Eq. S28 has the solution

$$\psi(r) = \frac{2q}{2\pi R a\epsilon\kappa \mathrm{K}_1(\kappa R)}\mathrm{K}_0(\kappa r), \tag{S29}$$

however, we use the numerical solution of the nonlinear Eq. S28 as the dimensionless potential $e\psi/kT$ can reach values in excess of unity at the low salt concentrations we consider. Here κ^{-1} is the Debye length, defined via $\kappa^2 \equiv 2e^2 c_s/(\epsilon k_{\mathrm{B}}T)$. At infinite dilution the appropriate boundary conditions are $\psi'(R) = -\rho/2\pi\epsilon R$ and $\psi(\infty) = 0$, but for the purposes of the numerical solver we employ the outer boundary condition $\psi'(d) = 0$ corresponding to a solution of fibrils separated by an average distance $2d$. We take $d = R + 5\kappa^{-1}$ (see Fig. S1). For $d \gg \kappa^{-1}$ the influence of the outer boundary condition will be minimal.

Once we have computed ψ, we get the electrostatic free energy density of the peptide cylinder as [6]

$$f = \frac{\epsilon}{2}\left(\frac{d\psi}{dr}\right)^2 + kT(c_+\ln(c_+/c_s) + c_-\ln(c_-/c_s) - c_+ - c_- + 2c_s). \tag{S30}$$

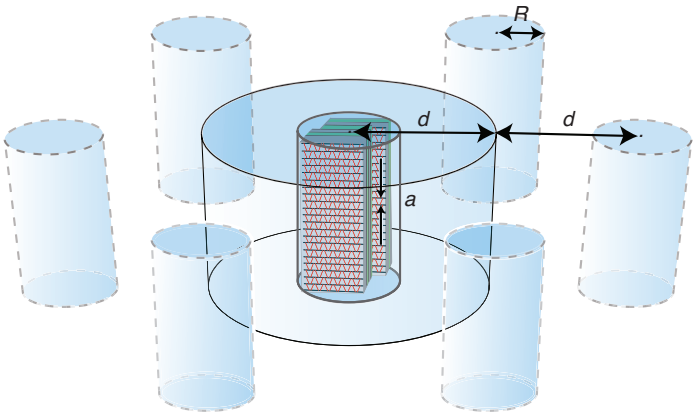

Figure S1: Cylindrical fibril geometry used to solve Eq. S28. Fibrils are taken as cylinders of radius R separated by a distance $2d$. Eq. S28 is solved from surface of the reference fibril (center), out to a distance d (outer cylinder) with the boundary condition $\psi'(d) = 0$ reflecting the symmetry of the electric potential between the cylinders.

The first term in Eq. S30 is the electrostatic energy stored in the electric field, and the remaining terms account for the translational entropy of the ions in the screening layer. We then compute the free energy per peptide using

$$\Delta F_{es} = 2\pi \frac{n_s a}{2p} \int_R^\infty f(r) r dr \tag{S31}$$

where r is the radial coordinate perpendicular to the axis of the cylinder.

References

[1] Lee, C., 2009. Self-assembly of protein amyloids: A competition between amorphous and ordered aggregation. *Phys. Rev. E* 80:031922.

[2] Auton, M., L. M. F. Holthauzen, and D. W. Bolen, 2007. Anatomy of energetic changes accompanying urea-induced protein denaturation. *Proc. Nat. Acad. Sci.* 104:15317–15322.

[3] Bolen, D., and G. Rose, 2008. Structure and energetics of the hydrogen-bonded backbone in protein folding. *Ann. Rev. of Biochem* 77:339–362.

[4] Ghosh, K., and K. Dill, 2009. Computing protein stabilities from their chain lengths. *Proc. Nat. Acad. Sci.* 106:10649–10654.

[5] Sillero, A., and J. Ribeiro, 1989. Isoelectric Points of Proteins: Theoretical Determination. *Analytical Biochemistry* 179:319–325.

[6] Andelman, D., 2006. Proceedings of the Nato ASI & SUSSP on ”soft condensed matter physics in molecular and cell biology”, Taylor & Francis, New York, chapter Introduction to electrostatics in soft and biological matter, 97–122.