Supporting information for:

Statistical Mechanics of Globular Oligomer Formation by Protein Molecules

Alexander J. Dear,[†] Anđela Šarić,[‡,¶] Thomas C. T. Michaels,[†,§]

Christopher M. Dobson,[†] and Tuomas P. J. Knowles[*,†,∥]

[†]Centre for Misfolding Diseases, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

[‡]Department of Physics and Astronomy, Institute for the Physics of Living Systems, University College London, Gower Street, London WC1E 6BT, UK

[¶]Corresponding author for simulations. Email: a.saric@ucl.ac.uk

[§]Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA

[∥]Cavendish Laboratory, Department of Physics, University of Cambridge, J J Thomson Avenue, Cambridge CB3 0HE, UK
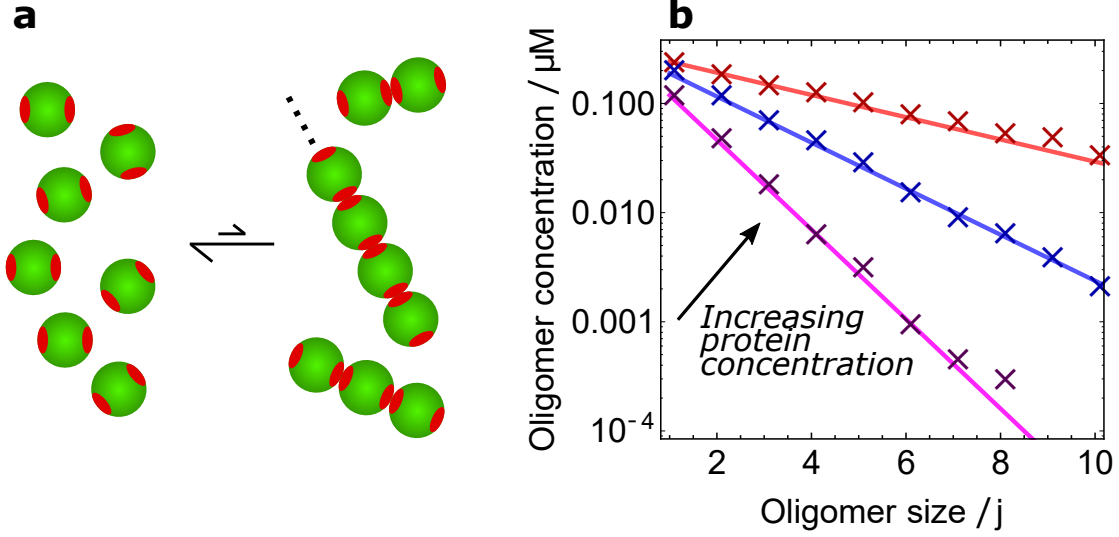
Figure S1: The analytical size distribution model for linear oligomers can be globally fitted to coarse-grained numerical Monte Carlo results (data points) with high accuracy. $\Delta G° = -15kT$; $p = 0.3\mu M$ (purple); $p = 1.3\mu M$ (blue); $p = 5.3\mu M$ (red).

Derivation of general statistical mechanical formulation for oligomers

This section is a reworking of material from ref. S1 and is included here for didactic purposes.

The grand canonical partition function is given by:

$$\Xi(T, V, \mu) = \sum_{N}^{\infty} Q(T, V, N)e^{\beta N \mu} \tag{1}$$

where $Q(T, V, N)$ denotes the canonical partition function of a system that contains N molecules, $\beta = 1/(k_B T)$ is the inverse temperature (k_B is the Boltzmann constant) and μ is the chemical potential of the monomers. We consider the formation of oligomeric aggregates up to an arbitrary maximum size of M monomers. By assuming that the oligomers do not interact, the grand canonical partition function can be expressed in terms of the individual

oligomer partition functions:

$$\Xi(T, V, \mu) = \sum_{N_1=0}^{\infty} \sum_{N_2=0}^{\infty} \cdots \sum_{N_M=0}^{\infty} \frac{\left(q_1(T,V)\,e^{\beta\mu}\right)^{N_1}}{N_1!} \frac{\left(q_2(T,V)\,e^{\beta 2\mu}\right)^{N_2}}{N_2!} \cdots \frac{\left(q_M(T,V)\,e^{\beta M\mu}\right)^{N_M}}{N_M!},$$

$$(2)$$

where $q_j(T,V)$ is the canonical partition function of a j-mer. Eq. (2) can be re-summed to yield

$$\Xi(T, V, \mu) = \exp\left(\sum_{j=1}^{M} q_j(T,V)\,e^{\beta j\mu}\right).$$

$$(3)$$

From Eq. (3), the concentration of oligomers of size j, $f(j) \equiv N_j(T,V)/(N_A V)$, is given by

$$f(j) = \frac{1}{N_A V}\,q(j)\,e^{\beta j\mu},$$

$$(4)$$

where N_A is Avogadro's number and we have used the short hand notation $q(j) \equiv q_j(T,V)$. Note that the chemical potential μ is set implicitly by letting $M \to \infty$ in Eq. (3) and imposing the conservation-of-mass condition, $p = \sum_{j=1}^{\infty} j f(j)$, which can be written more conveniently as

$$p = -\frac{1}{N_A V}\frac{k_B T}{\Xi}\frac{\partial\Xi}{\partial\mu}$$

$$(5)$$

We must now address the partition function $q(j)$. This is given in general by a product of independent contributions from translational and internal degrees of freedom: $q(j) = q_{\text{trans}}(j)\,q_{\text{int}}(j)$. In particular, the translational partition function $q_{\text{trans}}(j)$ is proportional to the system volume, and so can be written as $q_{\text{trans}} = V/v_0(j)$, where $v_0(j)$ is a fundamental volume (in the gaseous phase, it is given by the cube of the thermal wavelength). Making this substitution, we arrive at Eq. 1 in the main text.

Calculating the critical monomer concentration m^*

The maximum of the length distribution is found by solving $df(j)/dj = 0$:

$$\frac{df(j)}{dj} = \frac{1}{N_A v_0} \frac{d}{dj} e^{-\beta(\Delta G(j) - j\mu)} = f(j) \cdot -\beta \left(\frac{d\Delta G(j)}{dj} - \mu \right) = 0 \tag{6a}$$

$$\implies \frac{d\Delta G(j)}{dj} - \mu = \frac{5}{3} \frac{h}{\beta} j^{2/3} + \left(\frac{z_\infty}{2} G_b - \mu \right) - \frac{2}{3} j^{-1/3} \frac{z_\infty}{2} \alpha' G_b = 0. \tag{6b}$$

By making the substitution $x = j^{1/3}$, this can be transformed into the simpler problem $\mathcal{F}(x) = 0$, where:

$$\mathcal{F}(x) = ax^3 + cx + d \tag{7a}$$

$$= \frac{5}{3} \frac{h}{\beta} x^3 + \left(\frac{z_\infty}{2} G_b - \mu \right) x - \frac{z_\infty}{3} \alpha' G_b. \tag{7b}$$

We in fact need both a maximum and a minimum, so we need at least 2 real solutions to this equation. The cubic function must then cross the x axis three times. For this to happen it must possess two real stationary points $x_\pm$:

$$\mathcal{F}'(x) = 3ax^2 + c = 0 \implies x_\pm = \pm \sqrt{\frac{-c}{3a}}. \tag{8}$$

We thus need $c = \frac{z_\infty}{2} G_b - \mu < 0$ as a necessary but not sufficient condition. This translates to a fairly tiny minimum monomer concentration so we look for a further constraint. We

need that the first of these stationary points gives a value of $\mathcal{F} > 0$, and the second $\mathcal{F} < 0$:

$$\mathcal{F}(x_\pm) = \pm a \left(\frac{-c}{3a}\right)^{3/2} \pm c \left(\frac{-c}{3a}\right)^{1/2} + d \tag{9a}$$

$$= \pm \left(\frac{(-c)^3}{27a}\right)^{1/2} \mp (-c)^{2/2} \left(\frac{-c}{3a}\right)^{1/2} + d \tag{9b}$$

$$= \pm \frac{1}{3} \left(\frac{(-c)^3}{3a}\right)^{1/2} \mp \left(\frac{(-c)^3}{3a}\right)^{1/2} + d \tag{9c}$$

$$\therefore \mathcal{F}(x_-) = \frac{2}{3} \left(\frac{(-c)^3}{3a}\right)^{1/2} + d, \quad \mathcal{F}(x_+) = -\frac{2}{3} \left(\frac{(-c)^3}{3a}\right)^{1/2} + d. \tag{9d}$$

Given that $d > 0$, the first condition is always satisfied. To satisfy the second, we require (using $z_\infty = 12$):

$$\frac{2}{3} \left(\frac{(-c)^3}{3a}\right)^{1/2} > d \tag{10a}$$

$$(-c)^3 > 3a \frac{9d^2}{4} = 5 \frac{h}{\beta} \frac{(-z_\infty \alpha' G_b)^2}{4} \tag{10b}$$

$$\mu - 6G_b > \left(\frac{5h}{\beta}\right)^{1/3} (-6\alpha' G_b)^{2/3}. \tag{10c}$$

This gives the critical monomer concentration:

$$m^* = \frac{1}{N_A v_0} e^{\beta\mu} = \frac{1}{N_A v_0} \exp\left(6\beta G_b + (5h)^{1/3} \left(-6\alpha'\beta G_b\right)^{2/3}\right). \tag{11}$$

The value x^* for which the roots merge, or for which $\Delta = 0$, is given by $9d/2\Delta_0 = -3d/2ac$, which gives the location of the new maximum:

$$j^* = \frac{d}{2} = -\frac{6\alpha'\beta G_b}{5h}. \tag{12}$$

Head group sterics

We expect that head groups have some translational, rotational and conformational degrees of freedom, and that these are restricted through steric interactions upon addition of amphiphiles to a micelle. Indeed, it is known that head group size can influence micelle geometry as a result of these interactions.[S2] This free energy penalty for formation of larger micelles was neglected by Chandler *et al*, presumably because it is small relative to the connectivity-enforcing energy penalty.

We here verify this hypothesis by calculating the steric penalty explicitly. In previous MT approaches, this has been dealt with using a test particle approach,[S3] yielding the expression:

$$G_{\text{st}} = -jkT \ln\left[1 - \frac{a_h}{a}\right], \tag{13}$$

where a_h is the average cross-sectional area of the head group (in our Monte Carlo simulations $\pi\sigma^2/4$), and a is the core surface area per monomer. We first calculate a for spherical micelles:

$$V = jv = \frac{4}{3}\pi r^3 \implies r = \left(\frac{3jv}{4}\right)^{1/3} \tag{14a}$$

$$a = \frac{S}{j} = \frac{Sv}{V} = \frac{3v}{r} \tag{14b}$$

$$\therefore a = 3v\left(\frac{4}{3jv}\right)^{1/3} = \left(\frac{36v^2}{j}\right)^{1/3}. \tag{14c}$$

Now since v scales with σ^3, we have the scaling behaviour for G_{st}:

$$G_{\text{st}} \simeq -jkT \ln\left[1 - Cj^{\frac{1}{3}}\right], \tag{15}$$

with constant C. Thus, G_{st} has lower scaling with j than the connectivity-enforcing free energy penalty ($j^{5/3}$). Furthermore, C is given approximately by:

$$C < \frac{\pi\sigma^2/4}{(36\sigma^6)^{1/3}} \simeq 0.24. \tag{16}$$

We see, therefore, that not only does the head-group steric repulsion free energy term scale less strongly with j than the connectivity-enforcing term, but that also its coefficient is less than half of the magnitude of the connectivity-enforcing term's coefficient (found from fitting to be $h = 0.58$). Thus it is reasonable to neglect the steric contribution for modelling the Monte Carlo simulations, and oligomers in general.

Computing h analytically

The term h in the model can be computed from a knowledge of the geometry of amphiphiles within micelles. Specifically, h is given approximately by the formula:

$$h \simeq 0.75(\xi/\delta)^{4/3}, \tag{17}$$

where δ is the "length" of the hydrophobic segment, or its greatest distance from the hydrophilic segment, and ξ is the "width" of the hydrophobic segment, or its average linear dimension perpendicular to its length. In the case of our simulations we may identify $\xi = \sigma$. Moreover, we may calculate the length of the hydrophobic segment as the how far we may move one monomer along the length of another whilst maintaining their interaction. This gives $\delta = 1.33\sigma$. Feeding these numbers into Eq. (17) gives $h \simeq 0.51$. This is in reasonably good agreement with our fitted value of $h = 0.58$. The agreement can be improved by explicitly including the steric term computed in the previous section in the model when fitting; doing so yields a fitted value of $h = 0.49$.

References

(S1) Iljina, M.; Garcia, G. A.; Dear, A. J.; Flint, J.; Narayan, P.; Michaels, T. C. T.; Dobson, C. M.; Frenkel, D.; Knowles, T. P. J.; Klenerman, D. Quantitative analysis of co-oligomer formation by amyloid-β peptide isoforms. *Sci. Rep.* **2016**, *6*, 28658.

(S2) Tanford, C. Theory of micelle formation in aqueous solutions. *J. Phys. Chem.* **1974**, *78*, 2469–2479.

(S3) Puvvada, S.; Blankschtein, D. Molecular-thermodynamic approach to predict micellization, phase behavior and phase separation of micellar solutions. I. Application to nonionic surfactants. *J. Chem. Phys.* **1990**, *92*, 3710–3724.