Supporting Information: Theory of Sequence Effects in Amyloid Aggregation

Caleb Huang, Elaheh Ghanati, and Jeremy D. Schmit*

Department of Physics, Kansas State University, Manhattan, KS 66506, USA

E-mail: schmit@phys.ksu.edu

*To whom correspondence should be addressed

Residence time calculations

In this supporting information, we employ the transfer matrix formalism to solve for the binding lifetimes of sequences with triblock and alternating binding energies. In the zipper model (Fig. 2), where the molecules unbind progressively from one end, the binding lifetimes obey the recursion relationship (Eq. 2)[1]

$$t_R(n) = \tau_R(n)(r_+(n+1)t_R(n+1) + r_-(n)t_R(n-1) + 1). \tag{S1}$$

This relationship says that the system will proceed from state n to state $n+1$ with probability $\tau_R(n)r_+(n+1)$ and to state $n-1$ with probability $\tau_R(n)r_-(n)$. The new random walks starting at these sites will begin after an average waiting time of $\tau_R(n)$ in state n. In the following calculations we drop the R subscripts on t and τ used in the paper and, instead, we employ subscripts that denote the site dependent binding energies.

Triblock sequence

Consider a sequence containing three blocks of amino acids that each have a constant binding energy. At the N-terminal end there is a block of L_1 amino acids with binding energy ε_f ("flanking"), a central block of L_2 amino acids binding energy ε_c ("center"), and a C-terminal block containing the remaining $L - L_1 - L_2$ amino acids that again have binding energy ε_f. These binding energies determine the bond breakage rates $r_-(n)$, which take the values $r_f = r_+e^{-\varepsilon_f}$ and $r_c = r_+e^{-\varepsilon_c}$ for the flanking and center blocks, respectively.

We introduce transformed variables to reduce the recursion relation to a homogenous form

$$\begin{aligned} t(n) &= \theta_f(n) - n/(r_+ - r_f) \qquad 0 < n \leq L_1, \text{ or } L_1 + L_2 < n \leq L \\ t(n) &= \theta_c(n) - n/(r_+ - r_c) \qquad L_1 < n \leq L_1 + L_2 \end{aligned} \tag{S2}$$

The method described in the paper for uniform sequences can be straightforwardly applied to find

the residence time for a molecule bound by the entire first block

$$\begin{aligned}
\mathbf{u}(L_1+1) &= \mathbf{M_f}^{L_1}\mathbf{u}(1) && \text{(S3)}\\
&= \mathbf{T_f}(\mathbf{T_f}^{-1}\mathbf{M_f}\mathbf{T_f})^{L_1}\mathbf{T_f}^{-1}\mathbf{u}(1) && \text{(S4)}\\
&= \begin{pmatrix} \theta_f(1)\frac{\left(\frac{1}{\tau_f r_+}-1\right)^{L_1+1}-1}{\frac{1}{\tau_f r_+}-2} \\ \theta_f(1)\frac{\left(\frac{1}{\tau_f r_+}-1\right)^{L_1}-1}{\frac{1}{\tau_f r_+}-2} \end{pmatrix} && \text{(S5)}
\end{aligned}$$

where the subscripts on M and τ indicate that the flanking binding energy should be used in quantities that depend on r_-.

To determine the residence times for a molecule that forms n bonds where $L_1 < n \leq L_1 + L_2$ we require a second transfer matrix

$$\mathbf{u}(n) = \mathbf{M_c}^{n-L_1-2}\mathbf{u}(L_1+2). \tag{S6}$$

The starting vector for this second round of transfer matrix propagation is given by

$$\begin{aligned}
\mathbf{u}(L_1+2) &= \begin{pmatrix} \theta_c(L_1+2) \\ \theta_c(L_1+1) \end{pmatrix} \\
&= \begin{pmatrix} t(L_1+2)+(L_1+2)/(r_+-r_c) \\ t(L_1+1)+(L_1+1)/(r_+-r_c) \end{pmatrix} && \text{(S7)}
\end{aligned}$$

Eq. (S6) and Eq. (S7) require $t(L_1+2)$ which cannot be obtained directly from the initial transfer matrix. Instead, it can be found by plugging the vector elements of Eq. (S5) into Eq. (S1) which yields

$$t(L_1+2) = \frac{\theta_f(1)\left(-\tau_f r_+\left(\tau_c r_+\left(b_f^{L_1}-1\right)-2b_f^{L_1}\right)-b_f^{L_1}\right)-\tau_f(L_1\tau_c r_+ +1)-2\tau_f r_+\tau_c+\tau_c}{(2\tau_f r_+ -1)\tau_c r_+} \tag{S8}$$

where $b_f = 1/\tau_f r_+ - 1 = e^{\varepsilon_f}$ is the Boltzmann weight for a bond in the flanking region. At this point Eq. (S6) can be used to obtain lifetimes of states L_1+3 through L_1+L_2+1.

To solve for lifetimes within the third block, we repeat the boundary matching process described above to obtain $\mathbf{u}(L_1+L_2+2)$. At that point, a final round of transfer matrix propagation can be used for the final block

$$\mathbf{u}(n) = \mathbf{M_f}^{n-L_1-L_2-2}\mathbf{u}(L_1+L_2+2). \tag{S9}$$

After a tedious calculation (performed with the aid of Mathematica), the final boundary condition $r_+(L) = 0$ can be applied yielding

$$\theta_R(1) = \frac{b_c^{-L_2} b_f^{-L-L_1}}{r_- \tau_f r_+ c_f c_c} \times \tag{S10}$$
$$\left(r_- \left(c_c \tau_f \left(b_f^{L_1+L_2} - \tau_f r_+ d\right) - \tau_f r_+ c_f \tau_c b_f^{\ell_R}\left(b_c^{L_2}-1\right)\right) - \left(2(\tau_f r_+)^2 - 3\tau_f r_+ + 1\right) c_c b_f^{L_1+L_2}\right)$$

$$\theta_R(L) = -\frac{r_+\tau_f}{\left(1-2r_+\tau_c\right)^2\left(1-2r_+\tau_f\right)^2}\left(\left(-b_f^{L-L_1-L_2-2}+b_f\right)A + \frac{\left(b_f^{L-L_1-L_2-2}-1\right)B}{r_+\tau_f}\right) \tag{S11}$$

where $c_i = (2\tau_i r_+ - 1)$, $d = -b_f^{\ell_R} b_c^{L_2} + b_f^{\ell_R} + b_f^{L_1+L_2}$,

$$\begin{aligned} A &= -\theta_1 c_c \left(r_+\tau_c\left(b_c^{L_2}+1\right) - b_c^{L_2}\right) b_f^{L_1} \\ &\quad + r_+\tau_f\left(\theta_1 c_c\left(-2b_c^{L_2}b_f^{L_1} + 2r_+\tau_c\left(b_c^{L_2}b_f^{L_1}+1\right) + b_f^{L_1} - 1\right) + 2\tau_c\left(-c_c L_2 + \left(r_+\tau_c - 1\right)\left(b_c^{L_2}-1\right)\right)\right) \\ &\quad - \left(\tau_f c_c + \tau_c\right)\left(L_2\left(1-2r_+\tau_c\right) + \left(r_+\tau_c - 1\right)\left(b_c^{L_2}-1\right)\right), \end{aligned} \tag{S12}$$

$$\begin{aligned} B &= r_+^2\tau_f^2\left(\theta_1\left(1-2r_+\tau_c\right)\left(2r_+\tau_c\left(b_c^{L_2}b_f^{L_1}-1\right) - b_f^{L_1}+1\right) - 2\tau_c\left(r_+\tau_c\left(b_c^{L_2}-1\right) + L_2 c_c\right)\right) \\ &\quad + r_+\tau_f\left(\theta_1 c_c\left(r_+\tau_c\left(5b_c^{L_2}-1\right) - 2b_c^{L_2}\right)b_f^{L_1} + \tau_c f\left(b_c^{L_2}-1\right) + L_2 c_c\left(\tau_f c_c + \tau_c\right)\right) \\ &\quad - c_c\left(\theta_1 c_c b_c^{L_2} b_f^{L_1} + \left(\tau_f c_c + \tau_c\right)\left(b_c^{L_2}-1\right)\right), \end{aligned} \tag{S13}$$

and $f = \left(2r_+^2\tau_c\tau_f + r_+\left(5\tau_c - \tau_f\right) - 2\right)$.

Alternating sequence

For the alternating sequence, Eq. (S1) is generalized to a pair of recursion relationships

$$t_A(n) = p_{A+}t_B(n+1) + p_{A-}t_B(n-1) + \tau_A \tag{S14}$$

$$t_B(n) = p_{B+}t_A(n+1) + p_{B-}t_A(n-1) + \tau_B \tag{S15}$$

where the "A" subscript refers to odd numbered sites and the "B" subscript refers to even sites. The forward and backward probabilities are given by

$$p_{A+} = \frac{r_+}{r_+ + r_{A-}} \qquad p_{A-} = \frac{r_{A-}}{r_+ + r_{A-}} \tag{S16}$$

$$p_{B+} = \frac{r_+}{r_+ + r_{B-}} \qquad p_{B-} = \frac{r_{B-}}{r_+ + r_{B-}} \tag{S17}$$

and the site waiting times are

$$\tau_A = \frac{1}{r_+ + r_{A-}} \tag{S18}$$

$$\tau_B = \frac{1}{r_+ + r_{B-}} \tag{S19}$$

where the bond breakage and formation rates are related by the local binding energy $r_{i-} = r_+ e^{-\varepsilon_i}$. Using Eq. (S14) to eliminate the A terms from Eq. (S15) we obtain

$$t_B(n) = p_{2+}t_B(n+2) + p_{2-}t_B(n-2) + \tau_2 \tag{S20}$$

where

$$p_{2+} = \frac{p_{A+}p_{B+}}{1 - p_{A+}p_{B-} - p_{A-}p_{B+}} \tag{S21}$$

$$p_{2-} = \frac{p_{A-}p_{B-}}{1 - p_{A+}p_{B-} - p_{A-}p_{B+}} \tag{S22}$$

$$\tau_2 = \frac{\tau_A + \tau_B}{1 - p_{A+}p_{B-} - p_{A-}p_{B+}} \tag{S23}$$

Eq. (S20) has a straightforward interpretation. It says that a random walk at a given B site will remain confined to that B site and the two adjacent A sites for an average time τ_2. When it exits that cluster of three sites, the next B site it will arrive at will be the one to the right with probability p_{2+} and the one to the left with probability p_{2-}.

The similarity of Eq. (S20) to Eq. (S1) means that we can straightforwardly apply the same matrix formalism. First, we convert the Eq. (S20) to a homogenous form using

$$t_B(n) = \theta_B(n) - \frac{n}{2}\frac{\tau_2}{p_{2+} - p_{2-}} \tag{S24}$$

The resulting homogenous equation can be re-written in the matrix form $\mathbf{v}(n+2) = \mathbf{M_2}\mathbf{v}(n)$ where

$$\mathbf{v}(n) = \begin{pmatrix} \theta_B(n) \\ \theta_B(n-2) \end{pmatrix} \tag{S25}$$

$$\mathbf{M_2} = \begin{pmatrix} \frac{1}{p_{2+}} & -\frac{p_{2-}}{p_{2+}} \\ 1 & 0 \end{pmatrix} \tag{S26}$$

By comparison to Eq. (S5) we can construct $\mathbf{v}(n)$ for all B sites

$$\mathbf{v}(n) = \begin{pmatrix} \frac{\theta_B(2)\left(\left(\frac{p_{2-}}{p_{2+}}\right)^{n/2} - 1\right)}{\frac{p_{2-}}{p_{2+}} - 1} \\ \frac{\theta_B(2)\left(\left(\frac{p_{2-}}{p_{2+}}\right)^{\frac{n}{2}-1} - 1\right)}{\frac{p_{2-}}{p_{2+}} - 1} \end{pmatrix} \tag{S27}$$

where we have used $\theta_B(0) = 0$. The two elements from this vector can be used with Eq. (S14) and Eq. (S24) to find the residence times for all A sites

$$t_A(n) = -\frac{\tau_2\left(p_{A+} - p_{A-} + n\right)}{2\left(p_{2+} - p_{2-}\right)} + \tau_A + \frac{\theta_B(2)\left(\left(1 - p_{A+}\right)\left(\frac{p_{2-}}{p_{2+}}\right)^{\frac{n-1}{2}} + p_{A+}\left(\frac{p_{2-}}{p_{2+}}\right)^{\frac{n+1}{2}} - 1\right)}{\frac{p_{2-}}{p_{2+}} - 1} \tag{S28}$$

In particular, we are interested in the $n = 1$ site

$$t_A(1) = \theta_B(2) p_{A+} - \frac{\tau_2 p_{A+}}{p_{2+} - p_{2-}} + \tau_A \tag{S29}$$

The unknown constant, $\theta_B(2)$, is determined from the right boundary condition $t_B(L) = t_A(L-1) + 1/r_{B-}$. Using Eq. (S27), Eq. (S24), and Eq. (S28) we obtain

$$\theta_B(2) = \left(\frac{p_{2-}}{p_{2+}}\right)^{1-\frac{L}{2}} \left(\frac{\tau_A}{1-p_{A+}} + \frac{1}{(1-p_{A+})\, r_{B-}} + \frac{\tau_2}{p_{2+} - p_{2-}}\right) \tag{S30}$$

References

(1) Redner, S. *A Guide to First-Passage Processes*; Cambridge University Press, 2007 .