

# Supplementary Information: Mechanism of Fibril and Soluble Oligomer Formation in Amyloid Beta and Hen Egg White Lysozyme Proteins

Carlos Perez<sup>1</sup>, Tatiana Miti<sup>1</sup>, Filip Hasecke<sup>2</sup>, Georg Meisl<sup>3</sup>, Wolfgang Hoyer<sup>3,4</sup>, Martin Muschol<sup>1</sup>, and Ghanim Ullah<sup>1,\*</sup>

<sup>1</sup> Department of Physics, University of South Florida, Tampa, FL 33620, USA

<sup>2</sup> Institut für Physikalische Biologie, Heinrich-Heine-Universität, 40204 Düsseldorf, Germany

<sup>3</sup> Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

<sup>4</sup> Institute of Complex Systems (ICS-6), Structural Biochemistry, Research Centre Jülich, Germany

## Modified Powers and Powers Model

The original model used to replicate HewL aggregation was a modified version of the formalism developed by Powers and Powers<sup>2</sup> for nucleated polymerization with off-pathway aggregation (Figure S1). To replicate our observations, we made four key changes to the model by Powers and Powers<sup>1</sup>. (1) We used rate constants that are significantly smaller than those used by Powers and Powers. (2) In the original model, the same on-rate for monomer addition was used along the entire on-pathway, while the off-rate below and above the nucleus size was different. In our model, both the on- and off rates ( $a_1$  versus  $a$  and  $b_1$  versus  $b$  in Figure S1) are different for aggregates that are smaller than the nucleus size and RFs. (3) We included a secondary nucleation mechanism, proposed by Knowles *et al*<sup>3</sup>, where already formed RFs facilitate nucleation of new seeds (blue arrow with rate constant  $k_2$  in Figure S1). This was necessary to replicate the sharp autocatalytic rise in the experimental

ThT fluorescence upon RF nucleation. (4) To incorporate the lack of gOs formation below the COC and the dependence of experimental gOs growth rates on monomer concentration, we multiplied the rate constant ( $\alpha_1$ ) for dimers formation along off-pathway by a sigmoid function. With these changes, the amount of different species in the solution are given by the following rate equations.

$$\begin{aligned} \frac{d[X_1]}{dt} = & -[X_1] \left( 2a_1[X_1] + a_1 \sum_{j=2}^n [Y_j] + a[F^{(0)}] \right) + 2b_1[Y_2] + b_1 \sum_{j=3}^n [Y_j] + b[F^{(0)}] \\ & - [X_1] \left( 2\alpha_1\sigma[X_1] + \alpha \sum_{j=2}^m [Z_j] \right) + 2\beta[Z_2] + \beta \sum_{j=3}^m [Z_j] - nk_2[X_1]^n[F^{(1)}] \end{aligned} \quad (S1)$$

$$\frac{d[Y_2]}{dt} = (a_1[X_1]^2 - b_1[Y_2]) - (a_1[X_1][Y_2] - b_1[Y_3]) \quad (S2)$$

$$\frac{d[Y_j]}{dt} = (a_1[X_1][Y_{j-1}] - b_1[Y_j]) - (a_1[X_1][Y_j] - b_1[Y_{j+1}]), j = 3, 4. \quad (S3)$$

$$\frac{d[Y_n]}{dt} = (a_1[X_1][Y_{n-1}] - b_1[Y_n]) - a[X_1][Y_n] + k_2[X_1]^n[F^{(1)}] \quad (S4)$$

$$\frac{d[Z_2]}{dt} = (\alpha_1[X_1]^2 - \beta[Z_2]) - (\alpha[X_1][Z_2] - \beta[Z_3]) \quad (S5)$$

$$\frac{d[Z_j]}{dt} = (\alpha[X_1][Z_{j-1}] - \beta[Z_j]) - (\alpha[X_1][Z_j] - \beta[Z_{j+1}]), j = 3, 4, \dots, 7. \quad (S6)$$

$$\frac{d[Z_m]}{dt} = \alpha_1[X_1][Z_7] - \beta[Z_m] \quad (S7)$$

$$\frac{d[F^{(0)}]}{dt} = a[X_1][Y_n] \quad (S8)$$

$$\frac{d[F^{(1)}]}{dt} = (n+1)a[X_1][Y_n] + a[X_1][F^{(0)}] - b[F^{(0)}] \quad (S9)$$

where  $[X_1]$ ,  $[Y_j]$ ,  $[Z_j]$ ,  $[F^{(0)}]$ ,  $[F^{(1)}]$ ,  $[Y_n]$ ,  $[Z_m]$  represent the concentration of monomer, oligomers, i-mers along the on-pathway, j-mers along the off-pathway, the RF number concentration, the amount of monomers incorporated in RFs, the on-pathway nucleus of size  $n = 5$ , and the off-pathway final aggregate of size  $m = 8$  respectively in  $\mu$  M. The last term in Eq. S1 and S4 corresponds to secondary nucleation of new fibrils catalyzed by already established ones. To incorporate the transition from sigmoidal growth below COC to biphasic growth above COC, we use the following function.

$$\begin{aligned} \sigma &= \frac{1}{1 + e^{(COC - [X_1])/0.75}} \\ COC &= 3.522e^{-[NaCl]/45.3107} + 45.3107 \end{aligned} \quad (S10)$$

where COC is the critical oligomer concentration. In the model, the primary and the secondary nucleation rate constants  $a_1$  and  $k_2$  are varied over the range  $3.96 \text{ M}^{-1}\text{hr}^{-1}$ - $12.67 \text{ M}^{-1}\text{hr}^{-1}$  and  $7.2 \times 10^{-15} \text{ M}^{-1}\text{hr}^{-1}$  -  $2.25 \times 10^{-5} \text{ M}^{-1}\text{hr}^{-1}$  to obtain the best fit for fibril and aggregation growth with varying initial monomer concentration. Based on our earlier measurements, the ThT response of gOs was taken to be ten-fold weaker than that of RFs.

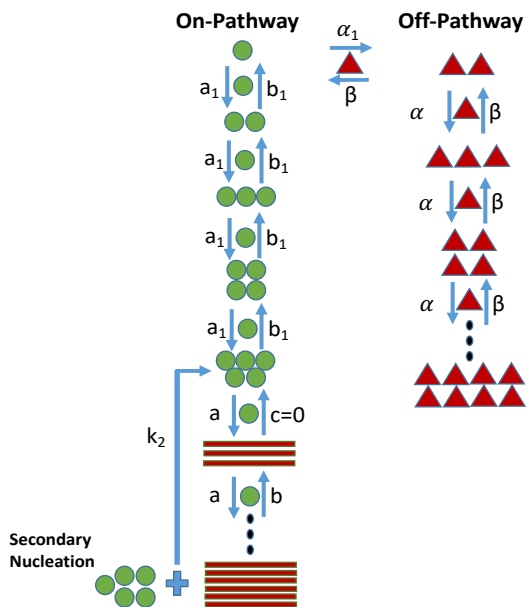


Figure S1: Schematic of the full modified Powers and Powers model. The on-pathway is displayed using green spheres, where in addition to primary nucleation, secondary nucleation contributes to RF seed formation. Within the on-pathway, monomers associate, forming multimers until eventually reaching a nucleus size  $n$  consisting of  $n$  monomers. Beyond this specified nucleus, fibril growth ensues. Already existing RFs (rods) catalyze the formation of new ones through secondary nucleation with rate constant  $k_2$ . On the off-pathway monomers form gOs.

Our full model closely fits all observations about HewL aggregation when the rate constants for both primary and secondary nucleation are allowed to vary (Figure S2). The rate constants for both primary and secondary nucleation decrease by several orders of magnitude as we increase initial monomer concentration.

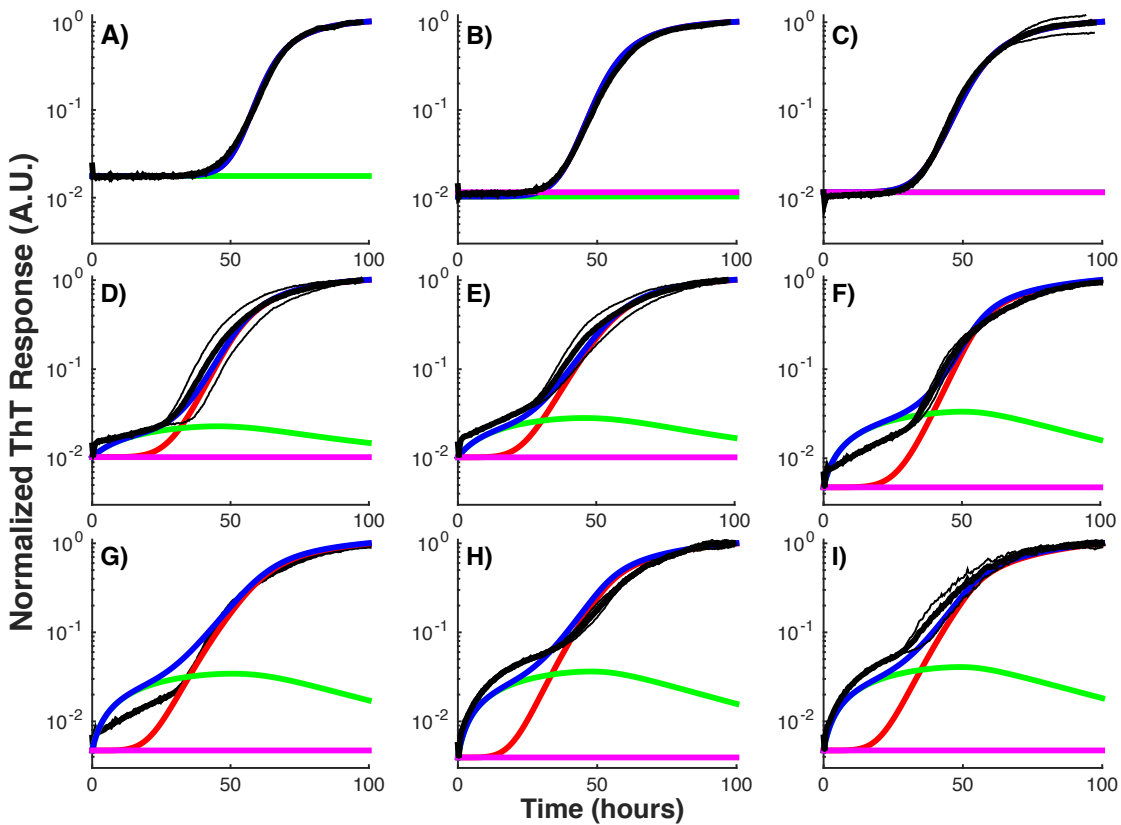


Figure S2: Model fits to experimental data on oligomeric and fibril formation in HewL. Experimentally observed ThT traces (black line where the thick line shows the average of several trials with ThT signal in the range indicated by the thin lines) shown alongside simulated contributions from gOs (green), RFs (red), and all on- and off-pathway species combined (model equivalent of ThT signal) (blue) at initial monomer concentration of (A)  $7\mu\text{M}$ , (B)  $14\mu\text{M}$ , (C)  $24\mu\text{M}$ , (D)  $58\mu\text{M}$ , (E)  $76\mu\text{M}$ , (F)  $126\mu\text{M}$ , (G)  $140\mu\text{M}$ , (H)  $210\mu\text{M}$ , and (I)  $315\mu\text{M}$ .

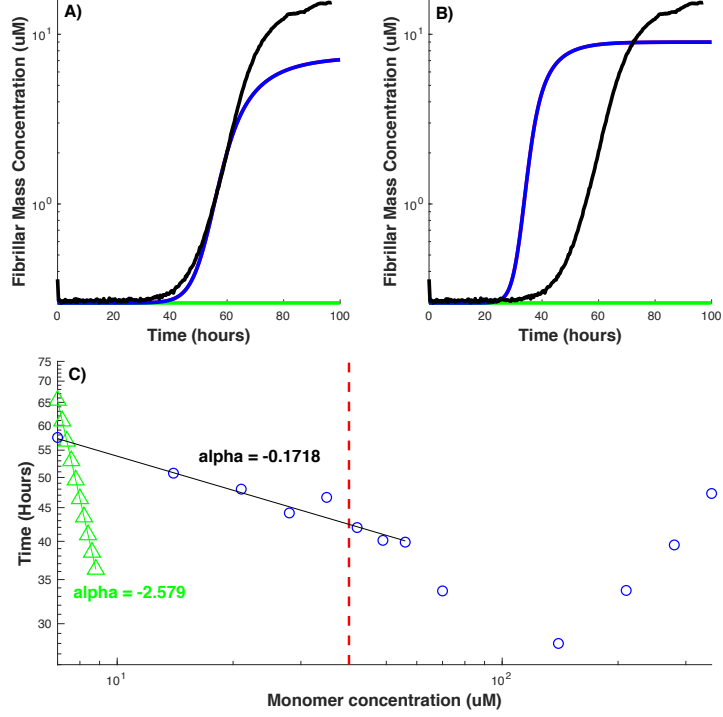


Figure S3: With fixed nucleation rate constants, the lag period drops significantly faster than the experimental values as we increase initial monomer concentration. (A) Simulated fibril formation at  $7.0\mu\text{M}$  initial monomer concentration (blue) shown together with experimental trace (black). (B) Increasing the initial monomer concentration by only  $2\mu\text{M}$  results in a drastic change in the lag period in the model. Lag period as a function of initial monomer concentration given by the model at fixed nucleation rate constants (green symbols) (C) and observed experimentally (blue symbols), where the vertical dashed red line represents the COC. The theoretical and experimental lag periods are fitted with a power law with exponents  $-2.579$  (green line) and  $-0.1718$  (black line).

## Reduced Model

To get rid of the sigmoid function, we consider cooperative aggregation of multiple monomers so that the nucleus along on-pathway and gOs along off-pathway are formed in a single reaction step each. After this change, the model is given by the following equations.

$$\frac{d[X_1]}{dt} = -5a_1[X_1]^5 - a[X_1]F^{(0)} + 5b_1[Y_5] + b[F^{(0)}] - n\alpha_1[X_1]^n + n\beta[Z_n] - 5k_2[X_1]^5[F^{(1)}] \quad (\text{S11})$$

$$\frac{d[Y_5]}{dt} = a_1[X_1]^5 - b_1[Y_5] - a[X_1][Y_5] + k_2[X_1]^5[F^{(1)}] \quad (\text{S12})$$

$$\frac{d[Z_n]}{dt} = \alpha[X_1]^n - \beta[Z_n] \quad (\text{S13})$$

$$\frac{d[F^{(0)}]}{dt} = a[X_1][Y_5] \quad (\text{S14})$$

$$\frac{d[F^{(1)}]}{dt} = 6a[X_1][Y_5] + a[X_1][F^{(0)}] - b[F^{(0)}] \quad (\text{S15})$$

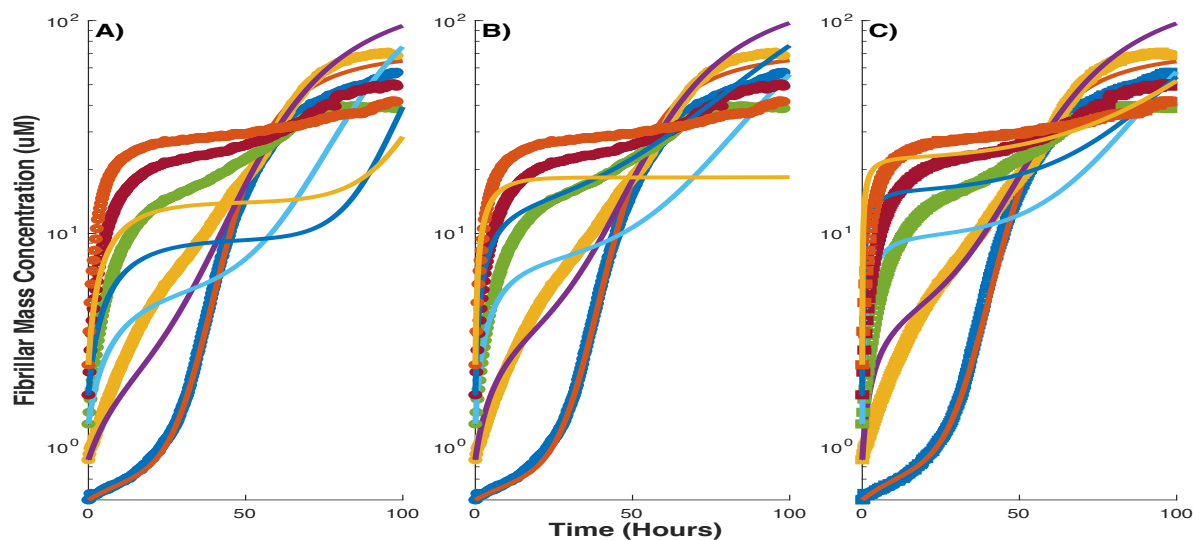


Figure S4: Fitting the reduced model (Eqs. S11-S15) to ThT traces representing HewL fibril formation when a final aggregate size of 3 (A), 4 (B), and 5 (C) along off-pathway is used in the model at HewL concentration of 70  $\mu\text{M}$ , 140  $\mu\text{M}$ , 210  $\mu\text{M}$ , 280  $\mu\text{M}$ , and 350  $\mu\text{M}$  (350, 280, 210, 140, and 70  $\mu\text{M}$  (bottom to up)). Theoretical and experimental results are represented by solid lines and circles respectively.

## Fitting to off-pathway dynamics with varying aggregate sizes in the presence of on-pathway

To test whether the presence of on-pathway while fitting the off-pathway rate equations to the initial segment of ThT signal, representing gOs growth at monomer concentration above COC, we repeated the fits searching for the optimal values for  $k$  and  $m$  in the presence of the on-pathway. We found that the inclusion of on-pathway does not change our conclusion (compare Figure 4 of the main text to Figure S5). We also note that the  $\chi^2$  values obtains for the initial segments of the traces are almost the same as when monomers' access to the on-pathway is restricted.

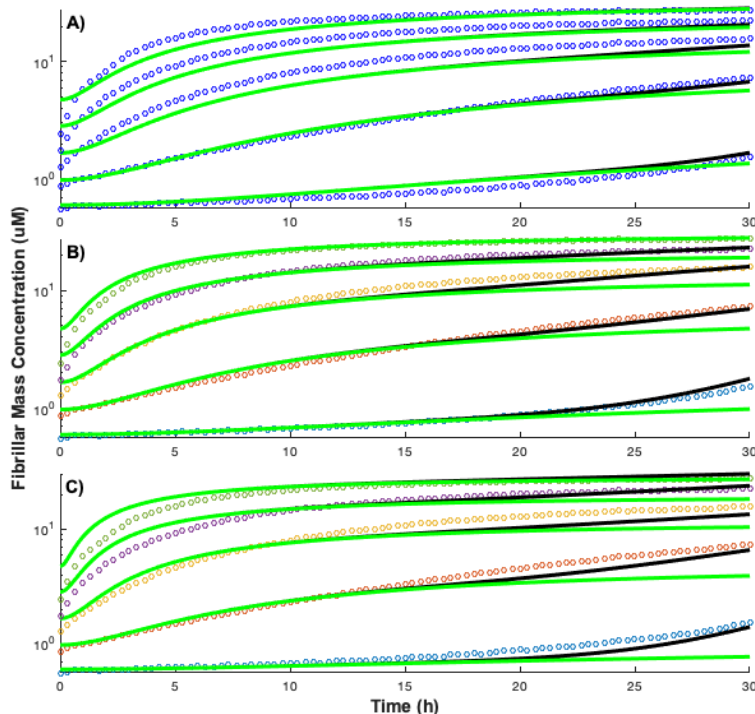


Figure S5: Fitting the model to the kinetics along off-pathway with monomers having access to both off- and on-pathway at HewL concentration of  $70 \mu\text{M}$ ,  $140 \mu\text{M}$ ,  $210 \mu\text{M}$ ,  $280 \mu\text{M}$ , and  $350 \mu\text{M}$  (bottom to top). Fits with different combinations of intermediate ( $Z_k$ ) and final ( $Z_m$ ) aggregates with different sizes ( $k$  and  $m$  respectively). Model fits with combination (A)  $Z_2Z_4$ ,  $\chi^2=96.17$ , (B)  $Z_2Z_5$ ,  $\chi^2=53.46$ , and (C)  $Z_2Z_6$ ,  $\chi^2=103.15$ . Circles and lines represent experimental data, and model fits respectively. The green and black lines represent the gOs and all species (gOs and RFs) given by the model fits to the initial part of the ThT signal.

## Fits with the model by Knowles and collaborators

To show that the variation in the primary and secondary nucleation rate constants within the on-pathway is not limited to one model, we simulated ThT traces at different monomer concentrations below COC using the model developed by Knowles and collaborators (see<sup>4</sup> for details about the model). Like our model, the formalism by Knowles and collaborators uses both primary and secondary nucleation to model the kinetics of amyloid beta aggregation. This version of their model does not include fragmentation of existing RFs and uses the primary nucleation, secondary nucleation, and elongation rate constants as the fitting parameters. The model fits our experimental data below COC very closely when the traces are fitted individually (Figure S6A-I). In agreement with our model, the primary and secondary nucleation rate constants decrease significantly as we increase initial monomer concentration (Figure S6J&K). To see if it is possible to perform a global fit, we attempted to

fit the model simultaneously to as few as two data sets recorded at different initial monomer concentrations by minimizing Pearson's chi-squared function. Despite using only two traces recorded at initial monomer concentration less than  $2 \mu\text{M}$  apart, the model fails to produce close fits using the same set of forward rate constants for primary and secondary nucleation. This further supports our initial hypothesis that the observed changes in the primary and secondary nucleation rate constants are intrinsic to the system and is not an artifact of our model.

In their recent work, Knowles and collaborators listed several different versions of their model, incorporating different mechanisms involved in fibril assembly along on-pathway<sup>5</sup>. In addition to the primary and secondary nucleation, one version of the model includes the fragmentation of existing RFs and the saturation of their elongation. In other words, the monomer dependence of RF elongation decreases with increasing monomer concentration and saturates at high monomer concentrations. Using this version of the model (Eqs. S51-S56 in Ref.<sup>5</sup>) with the nucleus size ( $n_c$ ), the primary and secondary nucleation rate constants, the elongation rate constant, and the Michaelis constant for elongation (the parameter that gives the monomer concentration at which elongation is half saturated - i.e., the elongation step proceeds at half its maximum speed) as fit parameters significantly improves the global fit to the initial part of the HewL data (Figure S7). The global fit results in the primary nucleus size of 0.65. Meisl et al.<sup>5</sup> justify the non-integer value of  $n_c$  by arguing that interpreting  $n_c$  as a nucleus size is only valid if the reaction is a simple single-step process. Instead, they term this parameter as the reaction order of the primary nucleation. Regardless of this, there is considerable room for improvement in the fits to the lag period as a function of initial monomer concentration and the later part of the experiment.



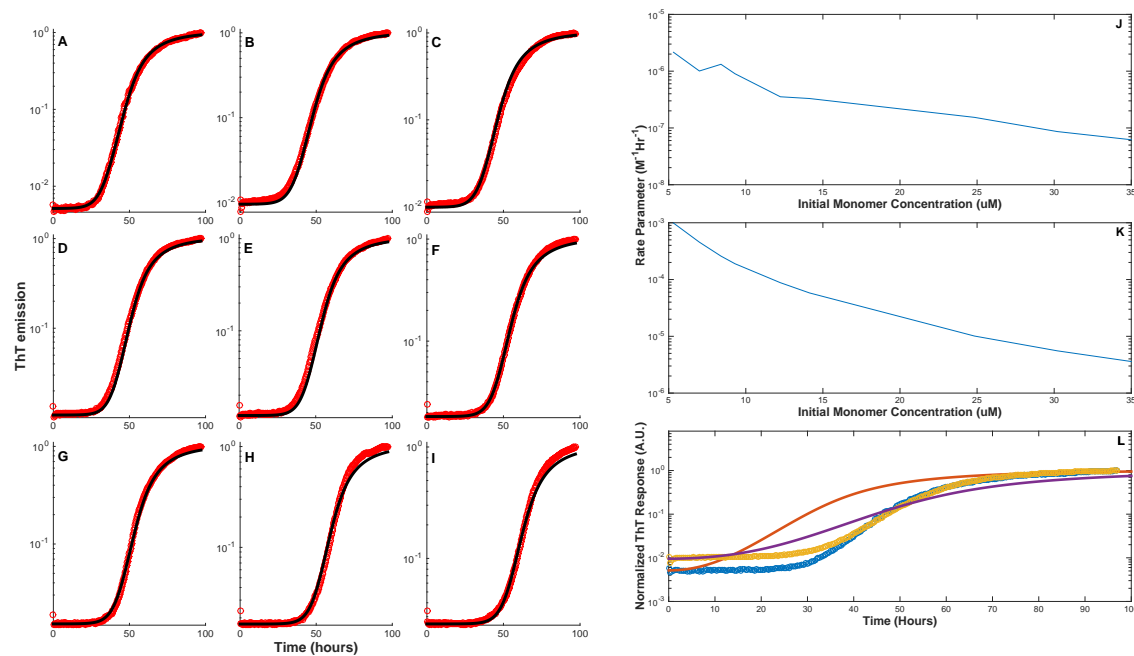


Figure S6: Fitting the model by Knowles and collaborators to observed ThT traces representing fibril assembly in HewL. Observed ThT traces (red) plotted on top simulated results (black) at initial monomer concentration of (A)  $5.31\mu\text{M}$ , (B)  $6.99\mu\text{M}$ , (C)  $8.39\mu\text{M}$ , (D)  $9.3\mu\text{M}$ , (E)  $12.24\mu\text{M}$ , (F)  $14.13\mu\text{M}$ , (G)  $24.83\mu\text{M}$ , (H)  $30.21\mu\text{M}$ , and (I)  $35\mu\text{M}$ . Rate constants for primary (J) and secondary (K) nucleation used to reproduce the fits to experimental results as functions of initial monomer concentrations. (L) Global fits to two traces recorded at two different initial monomer concentrations less than  $2\mu\text{M}$  apart, where experimental and theoretical results represented by circles and lines respectively. The ThT traces are normalized to the peak intensity in these experiments, following the approach used by Knowles and collaborators.

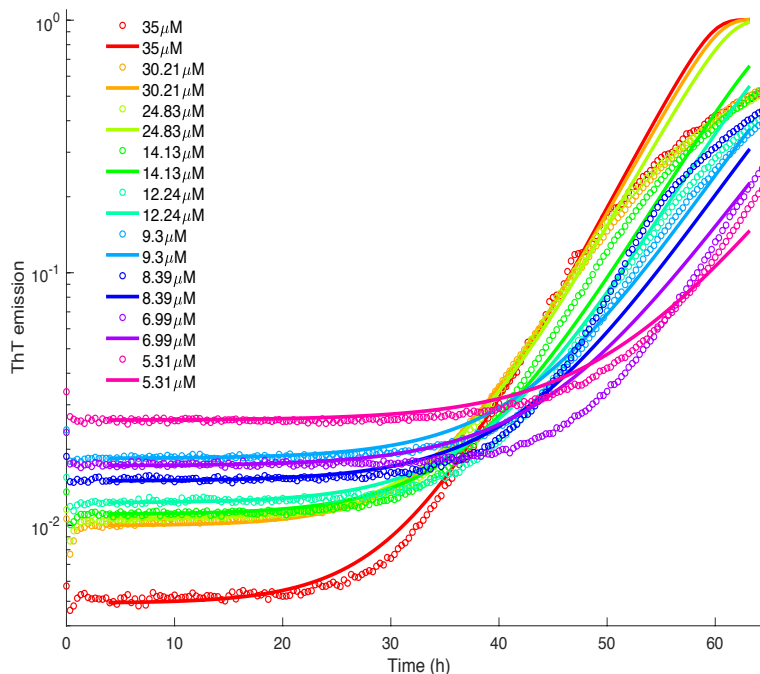


Figure S7: Fitting the model by Knowles and collaborators incorporating fragmentation and elongation saturation of RFs to observed ThT traces representing fibril assembly in HewL. Observed ThT traces (circles) plotted with simulated results (lines) at initial monomer concentrations as indicated. The ThT traces are normalized to the peak intensity in these experiments, following the approach used by Knowles and collaborators.

## Fits with Eden *et al.* model

In addition to the model developed by Knowles and collaborators, we also simulated our experiments with monomer concentrations below COC using the model put forward by Eden *et al.* (see<sup>6</sup> for details about the model). The model reproduces the initial phase of the fibril assembly very well but plateaus relatively faster than experimental results (Figure S8A-I). Like our and Knowles's models, both primary and secondary nucleation rate constants decrease as we increase initial monomer concentration (Figure S8J & K). Also like the model by Knowles and collaborators, global fit to two traces recorded at two different monomer concentrations fails to give satisfactory results. These findings further support the hypothesis that fixed nucleation rate constants can not be used to simulate the observed traces at all monomer concentrations and must be allowed to vary.

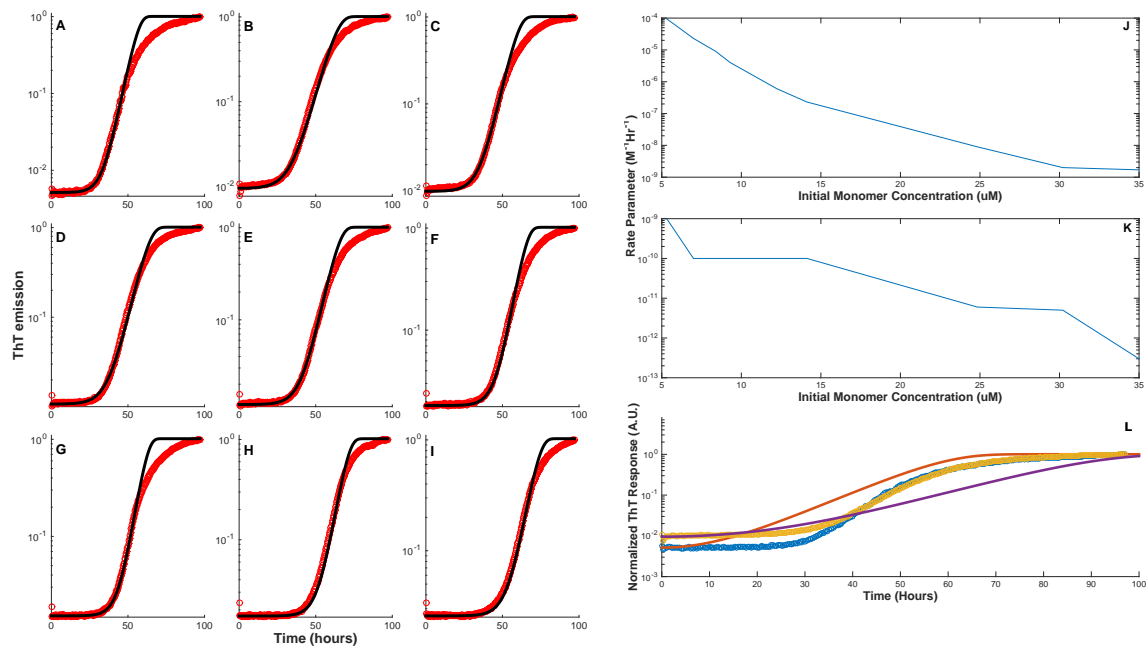


Figure S8: Fitting the model by Eden *et al* to observed ThT traces representing fibril assembly in HewL. Observed ThT traces (red) plotted on top simulated results (black) at initial monomer concentration of (A)  $5.31\mu\text{M}$ , (B)  $6.99\mu\text{M}$ , (C)  $8.39\mu\text{M}$ , (D)  $9.3\mu\text{M}$ , (E)  $12.24\mu\text{M}$ , (F)  $14.13\mu\text{M}$ , (G)  $24.83\mu\text{M}$ , (H)  $30.21\mu\text{M}$ , and (I)  $35\mu\text{M}$ . Rate constants for primary (J) and secondary (K) nucleation used to reproduce the fits to experimental results as functions of initial monomer concentrations. (L) Global fits to two traces recorded at two different initial monomer concentrations less than  $2\mu\text{M}$  apart, where experimental and theoretical results represented by circles and lines respectively. For consistency with the approach adopted by Eden *et al*, ThT traces are normalized to the peak intensity in these experiments,.

## References

- [1] Hasecke, F.; Miti, T.; Perez, C.; Barton, J.; Schölzel, D.; Gremer, L.; Grüning, C. S.; Matthews, G.; Meisl, G.; Knowles, T., *et al.* Origin of Metastable Oligomers and their Effects on Amyloid Fibril Self-assembly. *Chem. Sci.* **2018**, *9*, 5937–5948.
- [2] Powers, E. T.; Powers, D. L. Mechanisms of Protein Fibril Formation: Nucleated Polymerization with Competing Off-pathway Aggregation. *Biophys. J.* **2008**, *94*, 379–391.
- [3] Knowles, T. P.; Waudby, C. A.; Devlin, G. L.; Cohen, S. I.; Aguzzi, A.; Vendruscolo, M.; Terentjev, E. M.; Welland, M. E.; Dobson, C. M. An Analytical Solution to the Kinetics of Breakable Filament Assembly. *Science* **2009**, *326*, 1533–1537.
- [4] Cohen, S. I.; Linse, S.; Luheshi, L. M.; Hellstrand, E.; White, D. A.; Rajah, L.; Otzen, D. E.; Vendruscolo, M.; Dobson, C. M.; Knowles, T. P. Proliferation of Amyloid- $\beta$ 42 Aggregates Occurs Through a Secondary Nucleation Mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 9758–9763.
- [5] Meisl, G.; Kirkegaard, J. B.; Arosio, P.; Michaels, T. C. T.; Vendruscolo, M.; Dobson, C. M.; Linse, K. T. P. J., Sara. Molecular Mechanisms of Protein Aggregation from Global Fitting of Kinetic Models. *Nat. Protoc.* **2016**, *11*, 252–272.
- [6] Eden, K.; Morris, R.; Gillam, J.; MacPhee, C. E.; Allen, R. J. Competition between Primary Nucleation and Autocatalysis in Amyloid Fibril Self-assembly. *Biophys. J.* **2015**, *108*, 632–643.