Supplementary Material for: Reverse engineering the human islet amyloid polypeptide aggregation pathway

J Bailey[1], KJ Potter[2], B Verchere[2], L Edelstein-Keshet[1] and D Coombs[1,*]

[1]Department of Mathematics and Institute of Applied Mathematics, University of British Columbia, Vancouver, British Columbia V6T 1Z2, Canada.
[2]Department of Pathology and Laboratory Medicine, University of British Columbia and British Columbia's Childrens Hospital, Vancouver, Canada

E-mail: coombs@math.ubc.ca

1. Analysis of alternative models

Here we analyze the models of Lee et al. [1] and Powers and Powers [2] and rule them out, based on similar methods of analysis and scaling to those shown in the main text.

1.1. The Three-Stage Kinetic Model of Amyloid Fibrillation

Here we show that the model proposed by Lee et al. [1] (abbreviated TSKM in the Supplementary Tables) reduces to our previous NDP model with an arbitrarily assumed value $\gamma = 1$. Supplementary Figure 2 shows a schematic of the proposed model for the species of interest. Here $F(t)$ denotes the fibril number concentration (which is equivalent to the number of fibril tips). It is assumed that fibrils can elongate via addition of any oligomer. This leads to a term of the form $\sum_{j=1}^{k} l_j p_j F$. The new variable $F(t)$ is actually quite similar to the variable $\nu(t)$, i.e. the number of nuclei formed in the previous NDP model. We extend the assumption that negligible monomer mass is contained in oligomers and nuclei to now include the shortest of fibrils, those containing a nucleus and 1 additional monomer. (The need to include these shortest fibers in the negligible mass pool assumption stems from the constraint that a fibril may only form by the addition of a monomer to a nucleus.) Using conventions as in [1], with rates of attachment and detachment for the ith oligomers to fibrils l_i, and s_i respectively leads to the following differential equations

$$\frac{dc}{dt} = -f_{k+1} c F + s_o F, \tag{1}$$

$$\frac{dp_1}{dt} = f_0 c^2 - f_1 p_1 c + b_2 p_2 - b_1 p_1 + s_1 F - l_1 p_1 F, \tag{2}$$

$$\frac{dp_i}{dt} = f_{i-1} p_{i-1} c - f_i p_i c + b_{i+1} p_{i+1} - b_i p_i + s_i F - l_i p_i F \quad (2 \leq i < k), \tag{3}$$

$$\frac{dp_k}{dt} = f_{k-1}p_{k-1}c - f_k p_k c + b_{k+1}F - b_k p_k + s_k F - l_k p_k F, \tag{4}$$

$$\frac{dF}{dt} = f_k p_k c - b_{k+1}F, \tag{5}$$

$$\frac{dM}{dt} = f_{k+1}cF + \sum_{j=1}^{k} l_j p_j F - \sum_{j=0}^{k} s_j F - b_{k+1}F. \tag{6}$$

Observe the new terms $s_i F$ and $l_i p_i F$ in oligomer equations representing the new feature that oligomers of various sizes can attach and detach from fibrils. Further note that this model arbitrarily assumes that $n_0 = 2$ and $n_1 = 1$, whereas we obtain values for these integers directly from the data.

Systematic scaling of this model is accomplished by setting

$$\hat{t} = \frac{t}{\lambda c_o^{-\gamma}}, \quad \hat{c} = \frac{c}{c_0}, \quad \hat{F} = \frac{F}{\mu}, \quad \hat{M} = \frac{M}{c_0}, \quad \text{and } \hat{c}_i = \frac{p_i}{X}, \quad \text{for } 2 \le i \le k$$

where λ is needed to keep the units correct, and X, μ are quantities to be determined by the system. Substituting the above into the model and dropping the hats gives

$$\frac{dc}{dt} = \frac{\lambda c_o^{-\gamma}}{c_0}[-f_{k+1}cc_0F\mu + s_o F\mu], \tag{7}$$

$$\frac{dp_1}{dt} = \frac{\lambda c_o^{-\gamma}}{X}[f_0 c^2 c_o^2 - f_1 p_1 X c c_0 + b_2 p_2 X - b_1 p_1 X + s_1 F\mu - l_1 p_1 X F\mu], \tag{8}$$

$$\frac{dp_i}{dt} = \frac{\lambda c_o^{-\gamma}}{X}[f_{i-1}p_{i-1}Xcc_0 - f_i p_i X c c_0 + b_{i+1}p_{i+1}X - b_i p_i X + s_i F\mu - l_i p_i X F\mu] \quad \text{for} \quad 2 \le i \le k, \tag{9}$$

$$\frac{dp_k}{dt} = \frac{\lambda c_o^{-\gamma}}{X}[f_{k-1}p_{k-1}Xcc_0 - f_k p_k X c c_0 + b_{k+1}F\mu - b_k p_k X + s_k F\mu - l_k p_k X F\mu], \tag{10}$$

$$\frac{dF}{dt} = \frac{\lambda c_o^{-\gamma}}{\mu}[f_k p_k X c c_0 - b_{k+1}F\mu], \tag{11}$$

$$\frac{dM}{dt} = \frac{\lambda c_o^{-\gamma}}{c_0}[f_{k+1}cc_0F\mu + \sum_{j=1}^{k} l_j p_j X F\mu - \sum_{j=0}^{k} s_j F\mu - b_{k+1}F\mu]. \tag{12}$$

We now attempt to remove all explicit dependence of the model on c_0. Inspection of the equations for the change in fibril end density and mass imply that the term $f_k p_k X c c_0$ must survive or no fibrils will ever form. Therefore, we set $X = c_o^\gamma$, and $\mu = X$. Other terms must drop out as in the previous example. In the resulting equations, we set $\gamma = 1$ and absorb λ into the reaction rates. The resulting equations are

$$\frac{dc}{dt} = -f_{k+1}cF, \tag{13}$$

$$\frac{dp_1}{dt} = f_0 c^2 - f_1 c p_1 - l_1 p_1 F, \tag{14}$$

$$\frac{dp_i}{dt} = f_{i-1}cp_{i-1} - f_i c p_i - l_i p_i F, \quad \text{for} \quad 2 \le i \le k-1 \tag{15}$$

$$\frac{dp_k}{dt} = f_k p_{k-1} c - f_k p_k c - l_k p_k F, \tag{16}$$
$$\frac{dF}{dt} = f_k p_k c, \tag{17}$$
$$\frac{dM}{dt} = f_{k+1} c F + \sum_{j=1}^{k} l_j p_j F. \tag{18}$$

Elongation kinetics suggest that fibril growth is mediated by monomer addition implying $l_i p_i F = 0$, for $1 \leq i \leq k$. Furthermore, the persistent population of oligomers observed experimentally cannot be established if the elongation terms for oligomeric species are not absent, or trivially small. The model now becomes

$$\frac{dc}{dt} = -f_{k+1} c F, \tag{19}$$
$$\frac{dp_1}{dt} = f_0 c^2 - f_1 c p_1, \tag{20}$$
$$\frac{dp_i}{dt} = f_{i-1} c p_{i-1} - f_i c p_i, \tag{21}$$
$$\frac{dp_k}{dt} = f_{k-1} p_{k-1} c - f_k p_k c, \tag{22}$$
$$\frac{dF}{dt} = f_k p_k c, \tag{23}$$
$$\frac{dM}{dt} = f_{k+1} c F \approx -\frac{dc}{dt}. \tag{24}$$

Therefore, the model as proposed in [1] results in a special case of the generic NDP model from the view of collapsible data and persistent oligomer populations, or elongation kinetics. Removing the condition that only one monomer may join other monomers or oligomers to form the next oligomeric species, thus allowing the formation of each oligomeric species through any number of additional monomer units results in the NDP model

$$\frac{dp_1}{dt} = f_0 c^{2\gamma} - f_1 c^{\gamma} p_1, \tag{25}$$
$$\frac{dp_i}{dt} = f_{i-1} c^{\gamma} p_{i-1} - f_i c^{\gamma} p_i, \tag{26}$$
$$\frac{dp_k}{dt} = f_{k-1} p_{k-1} c^{\gamma} - f_k p_k c, \tag{27}$$
$$\frac{dF}{dt} = f_k p_k c^{\gamma}, \tag{28}$$
$$\frac{dM}{dt} = f_{k+1} c F = -\frac{dc}{dt}. \tag{29}$$

Thus, the "Three-Stage Kinetic Model of Amyloid Fibrillation" reduces to the NDP model, under the systematic scaling and elongation kinetics implied by our data.

1.2. Off-Pathway Kinetics for Aggregate Formation

Powers and Powers [2] proposed an NDP model with a competing off-pathway branch that produces other types of aggregates (represented by the variables z_i). We abbreviate

this model by the initials OPK in the Supplementary Tables. The assumptions made for the basic NDP model apply here but the authors assumed what amounts to an arbitrary scaling $t_0 \propto A_\infty^{-\gamma}$, for $\gamma = 1$.

With terminology summarized in Supplementary Tables 2 and 3, the model equations are:

$$\frac{dp_1}{dt} = f_0 c^2 - f_1 c p_1 + b_2 p_2 - b_1 p_1, \tag{30}$$

$$\frac{dp_i}{dt} = f_{i-1} c p_{i-1} - f_i c p_i - b_i p_i + b_{i+1} p_{i+1}, \quad \text{for } 2 \le i \le k, \tag{31}$$

$$\frac{dz_1}{dt} = \alpha_0 c^2 - \alpha_1 z_1 c + \beta_2 z_2 - \beta_1 z_1, \tag{32}$$

$$\frac{dz_i}{dt} = \alpha_{i-1} z_{i-1} c - \alpha_i z_i c - \beta_i z_i + \beta_{i+1} z_{i+1}, \quad \text{for } 2 \le i \le m-1, \tag{33}$$

$$\frac{dz_m}{dt} = \alpha_{m-1} z_{m-1} c - \beta_m z_m, \tag{34}$$

$$\frac{d\nu}{dt} = f_k c^{n_k} p_k - b_\nu \nu, \tag{35}$$

$$\frac{dM}{dt} = f_{k+1} c \nu - b_M \nu. \tag{36}$$

We now introduce the additional scaled variable $\hat{z}_i = z_i / Y$ for $1 \le i \le m$, where Y is to be determined. Substituting the scaling variables into the model and dropping the hats gives

$$\frac{dp_1}{dt} = \lambda[f_0 c^2 - f_1 p_1 c], \tag{37}$$

$$\frac{dp_i}{dt} = \lambda[f_{i-1} p_{i-1} c - f_i p_i c c], \quad \text{for } 2 \le i \le k, \tag{38}$$

$$\frac{d\nu}{dt} = \lambda[f_k p_k c], \tag{39}$$

$$\frac{dM}{dt} = \lambda[f_{k+1} \nu c], \tag{40}$$

$$\frac{dz_1}{dt} = \lambda[\alpha_0 c^2 - \alpha_1 z_1 c], \tag{41}$$

$$\frac{dz_i}{dt} = \lambda[\alpha_{i-1} z_{i-1} c - \alpha_i z_i c] \quad \text{for } 2 \le i \le m-1, \tag{42}$$

$$\frac{dz_m}{dt} = \lambda[\alpha_{m-1} z_{m-1} c]. \tag{43}$$

We find, given our data constraints, that this model reduces to the NDP model where aggregates of the same monomeric size as oligomers are produced. Furthermore, these aggregate populations are ever increasing, and given that no "bottleneck" of nucleation exist along the aggregate pathway, it is likely that such aggregation events would effectively compete for monomers, especially during early stages of fibrillogenesis. The result is a notable mass of monomers being sequestered into aggregates (a prediction not supported by our experimental data). This would contradict our assumption that most of the mass is concentrated in either the fibril or the monomer pools during the

fibril formation process. This model may be easily modified in such a way that allows for any scaling of γ. However, the observation that the off-pathway aggregates will be of the same size in monomer units as the persistent oligomer population will hold, except perhaps for the smallest aggregate species.

[1] Chuang-Chung Lee, Arpan Nayak, Ananthakrishnan Sethuraman, Georges Belfort, and Gregory J McRae. A three-stage kinetic model of amyloid fibrillation. *Biophys J*, 92(10):3448–3458, May 2007.

[2] Evan T Powers and David L Powers. Mechanisms of protein fibril formation: nucleated polymerization with competing off-pathway aggregation. *Biophys J*, 94(2):379–391, Jan 2008.

2. Supplementary Tables

Supplementary Table 1. Estimating the number of steps in the nucleation pathway. As described in the main text, we generated log-log plots of the fluorescence $A(t)$ versus time, for times between 5% and 30% of the asymptotic fluorescence. The slope of the best-fit linear function is our estimate of $k + 2$ where k is the number of steps in the pathway to nucleation. Here we give the estimates of k for each experiment with at least 4 data points in the time range considered, and the R^2 (coefficient of determination), a measure of the goodness of the linear fit.

Initial concentration (μM)	Estimated k	No. of data points	R^2 statistic
30	10.4	8	0.88
35	8.9	11	0.99
40	6.3	17	0.98
45	7.9	9	0.98
50	7.3	10	0.93
55	8.2	8	0.90
60	8.2	8	0.90
65	7.0	7	0.91
70	9.7	4	0.99
75	7.7	5	0.99
80	7.7	5	0.99

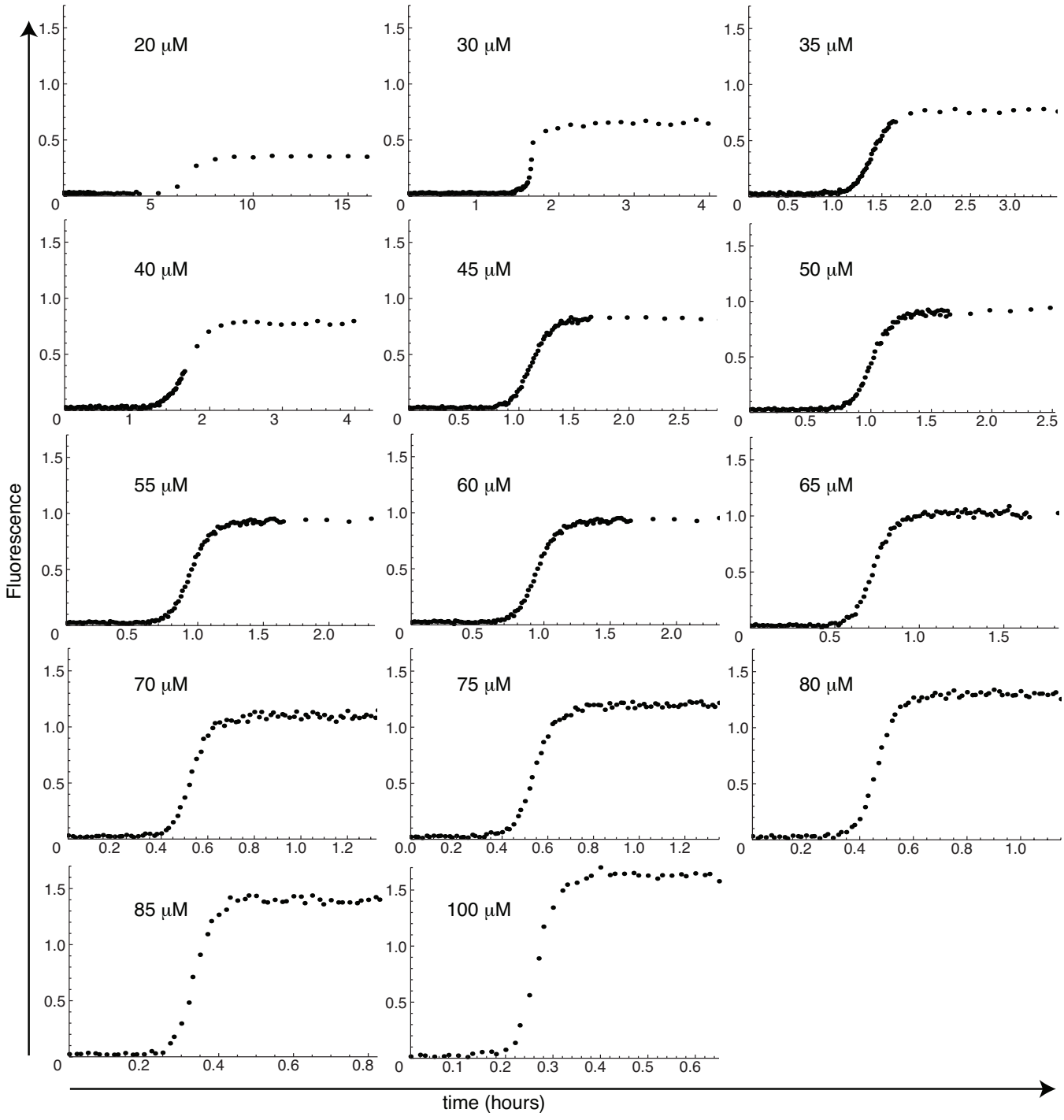
Supplementary Table 2. Variables and their meanings in the various models. Abbreviations: NDP, nucleation dependent polymerization; FDSN, fibril-dependent secondary nucleation; TSKM, three-stage kinetic model; OPK, nucleation dependent model with off-pathway kinetics.

Variable	meaning	model
$c(t)$	IAPP monomer concentration	all
$p_i(t)$	number concentration of the ith oligomer ($i = 1 \dots k$)	all
$\nu(t)$	number concentration of nuclei	NDP
$M(t)$	monomer mass in fibril, excluding monomers in oligomer and nuclei	NDP
$A(t)$	the measured fluorescence, assumed to be a measure of M	NDP
$F(t)$	fibril concentration (similar to $\nu(t)$)	TSKM
$z_i(t)$	number concentration of ith aggregate (excluding oligomers)	OPK

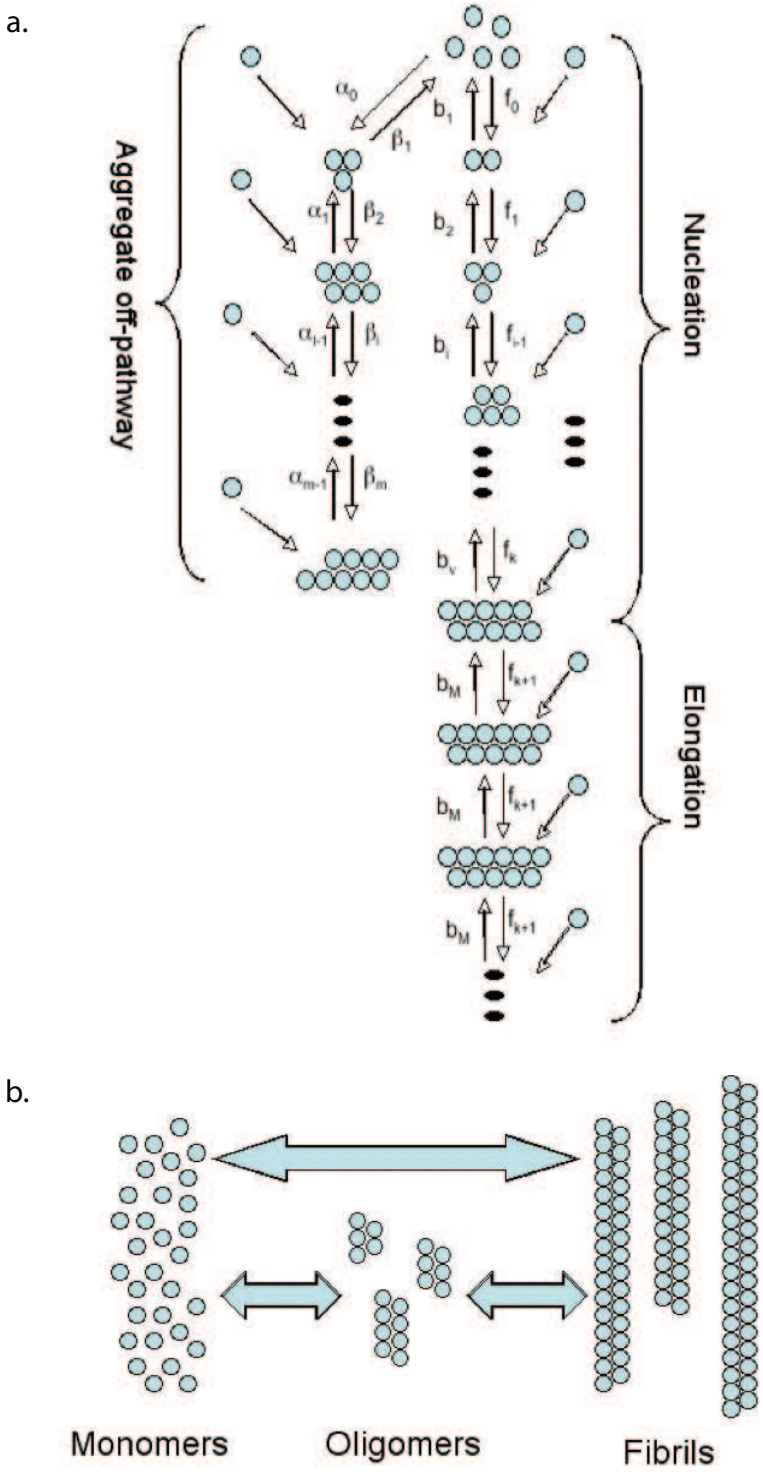
Supplementary Table 3. Parameters and their meanings in the various models. Abbreviations: NDP, nucleation dependent polymerization; FDSN, fibril-dependent secondary nucleation; TSKM, three-stage kinetic model; OPK, nucleation dependent model with off-pathway kinetics.

Parameter	meaning	model
c_0	the initial IAPP monomer concentration	NDP
n_i	number of monomers added to p_i to form p_{i+1}	NDP
k	number of oligomer species	NDP
f_i	forward rate constant for the ith oligomer species	NDP
f_k	forward rate constant for nuclei formation	NDP
f_{k+1}	forward rate constant for polymer elongation	NDP
b_i	reverse rate constant for the ith oligomer species	NDP
d_i	disintegration rate constant for the ith oligomer species	NDP
δ	rate of secondary nucleation	FDSN
l_i	attachment rate of ith oligomer to fibrils	TSKM
s_i	detachment rate of ith oligomer from fibrils	TSKM
m	number of aggregate species (other than oligomers)	OPK
α_i	forward rate constant for the ith aggregate species	OPK
β_i	reverse rate constant for the ith aggregate species	OPK
b_ν	reverse rate constant for nuclei	OPK
b_M	reverse rate constant for fibrils	OPK

3. Supplementary Figures



Supplementary Figure 1. All 14 sets of fluorescence data for hIAPP polymerization using the thioflavin-T assay. We collected 20 hours of data in each experiment but the range of data shown has been restricted in each case, for easier viewing.



Supplementary Figure 2. Schematics of alternative models for amyloid nucleation and polymerization. a. the model proposed by Powers and Powers. b. the model proposed by Lee et al.