Appedix A: Supplemental Materials of Amyloid Fibrillation Model

This supplement contains four sections that support the main text including: 1. details of the methods used for parameter estimation, 2. comparisons of the model predictions against a simple empirical model, 3. a sensitivity analysis study that identifies key variables, 4. the MATLAB® [1] computer codes used to implement the parameter estimation and numerical solution of the fibrillation kinetics, and 5. The abstract of the paper entitled "Osmolyte Controlled Fibrillation Kinetics of Insulin" which deals with additive-induced heterogeneous fibrillation. It is the manuscript we wrote together with Nayak and Prof. Belfort from Rensselaer Polytechnic Institute.

A.1 Parameter estimation

Nonlinear least squares regression [3] was used to estimate the model parameters. The mathematical statement of the optimization problem including the physical constraints is of the form:

$$\begin{aligned} &\underset{\underline{\theta}}{Min} && \Phi\left(\underline{\theta}\right) = \sum_{i=1}^{n_{exp}} \left[Y_i - f\left(t_i;\underline{\theta}\right)\right]^2 \\ &Subject\ to && \underline{\theta} \geq \underline{0} \end{aligned} \tag{A.1}$$

where $\Phi(\underline{\theta})$ is the objective function to be minimized and $\underline{\theta}$ is vector of n_p parameters to be estimated from Y_i, the measurements of fibril amount at the n_{exp} time points. The solution to the model built in the main paper describing the kinetics of fibril concentration at the time points is denoted by f. A local minimum for Eq. (A.1) can be found by solving the system of algebraic equations.

$$g_j\left(\underline{\theta}\right) = \frac{\partial \Phi}{\partial \theta_j} = -2\sum_{i=1}^{n_{exp}} \left[Y_i - f\left(t_i;\underline{\theta}\right)\right] \left(\left. \frac{\partial f}{\partial \theta_j} \right|_{t_i} \right) = 0 \quad ; j = 1,...,n_p \tag{A.2}$$

Both the optimization problem and the embedded differential equations, describing the kinetics, were solved with MATLAB® and its library programs **nlinfit**, **nlparci**, and **nlpredci**. (A copy of the MATLAB code is included in Section 4). The algebraic equation solver **nlinfit** uses the Levenberg-Marquardt method to solve the nonlinear system Eq. (A.2) After a limited number of iterations, the calculation converged to the local minimum. Different starting points were used to ensure that the solution to Eq. (A.2) was indeed a global minimum.

Three different metrics were used to assess the goodness-of-fit for the parameters $\underline{\theta}$. One was the coefficient of determination (R^2), which is the ratio of sum of squared errors due to regression to the total sum of square errors. Another was the Root Mean Square Error (RMSE). A third measure of the quality of the statistic fit were the confidence intervals for each of the estimates. Several computational steps were required to calculate the standard error of the confidence intervals. First, the population variance s^2 was estimated from the residual errors,

$$s^2 = \frac{1}{\nu}\sum_{i=1}^{n_{\exp}}\left[Y_i - f(t_i;\underline{\hat{\theta}})\right]^2 \tag{A.3}$$

where ν is the rank of design matrix and equal to the number of experimental points minus the number of parameters ($n_{exp} - n_p$). The half-width confidence interval, given the level of significance, was calculated using

$$\tilde{\theta}_j = \hat{\theta}_j \pm t_{v,\alpha/2}\cdot s\cdot\left[\left(X^T X\right)^{-1}\right]_{jj}^{1/2} \tag{A.4}$$

where $t_{v,\alpha/2}$ is the Student t-distribution with degree of freedom ν and significance level α. The significance level used here was 0.317 which corresponds to one standard deviation. X is the linearized design matrix also known as the Jacobian.

$$X_{ij} = \left.\frac{\partial f}{\partial \theta_j}\right|_{t_i} \tag{A.5}$$

Originally there were two forward and two reverse rate constants to be determined but after exploring parameter variations by sensitivity analysis, we found that k_{nu-} did not influence the response significantly (Figure A.3-2) and so the estimated parameter vector $\underline{\theta}$ has the components $\{k_{nu,1}, k_{fb,1}, k_{fb}\}$

A.2 Comparison of the Kinetic Model against an Empirical Function

In many previous experimental studies regarding fibrillation (see [2] for example), the sigmoidal responses were fitted with an empirical expression of the form.

$$Y = y_i + m_i t + \frac{y_f + m_f t}{1 + e^{-[(t-t_0)/\tau]}} \tag{A.6}$$

where the coefficients (y_i, m_i, y_f, m_f, t_0, τ) were adjusted to fit the time response. Given the often good match to the data it is natural to ask if there is any physical basis for the model.

In general, the set of differential equations defined in our model need to be solved numerically except for some special cases. One is when the natural insulin hexamer is assumed to dissociate instantaneously and the critical size *n* is as small as two. Under these assumptions there are two chemical species left: soluble monomers and insoluble fibrils. Since the kinetic model now consists of only two distinct stages (Eq. (A.7)): the slow nucleation process, and the fast fibrillation stage, the nucleation and the fibril elongation reactions both follow second order kinetics.

$$\begin{aligned} A_1 + A_1 &\xrightarrow{k_{nu,1}} F \\ A_1 + F &\underset{k_{fb-}}{\overset{k_{fb,1}}{\rightleftharpoons}} F \end{aligned} \tag{A.7}$$

In this model fibril dissociation is first order with respect to fibril concentration.

$$\frac{dF}{dt} = k_{nu,1}A_1^{\,2} + k_{fb,1}FA_1 - k_{fb-}F \tag{A.8}$$

If the fibrils were on average composed of *N* monomers then the time derivative of monomer concentration should have been *N* times as large as that of fibril concentration.

$$\frac{dA_1}{dt} = -N\frac{dF}{dt} \tag{A.9}$$

Under these conditions the monomer concentration, A_1 could be expressed in terms of initial concentrations (A_0) and *F* by a mass balance; thus, the time derivative of fibril now becomes a single variable differential equation.

$$\begin{aligned} \frac{dF}{dt} &= k_{nu,1}(A_0 - NF)^2 + k_{fb,1}F(A_0 - NF) - k_{fb-}F \\ &= (N^2k_{nu,1} - Nk_{fb,1})F^2 + (k_{fb,1}A_0 - 2Nk_{nu,1} - k_{fb-})F + k_{nu,1}A_0^{\,2} \end{aligned} \tag{A.10}$$

Since, at steady state, the right hand side of Eq. 10 was equal to zero, it is a second order algebraic equation in *F* and we know that the system had at least one stable steady state, so there must be two real roots (r_1, r_2) [4].

$$r_1, r_2 = \frac{-(k_{fb,1}A_0 - 2Nk_{nu,1} - k_{fb-}) \pm \sqrt{(k_{fb,1}A_0 - 2Nk_{nu,1} - k_{fb-})^2 - 4(N^2k_{nu,1} - Nk_{fb,1})k_{nu,1}A_0^{\,2}}}{2(N^2k_{nu,1} - Nk_{fb,1})} \tag{A.11}$$

The set of differential equations were simplified into the form explicitly showing the loci of two roots; one the attractor and the other the repeller (say for $r_1 < r_2$) as in Figure A.2-1.

$$\frac{dF}{dt} = (N^2 k_{nu,1} - N k_{fb,1})(F - r_1)(F - r_2) \tag{A.12}$$

$$\frac{dA_1}{dt} = (k_{fb,1} - N k_{nu,1})(A_1 - (A_0 - N r_1))(A_1 - (A_0 - N r_2)) \tag{A.13}$$

Each time derivative is a parabolic function of its specie concentration where one is concave and the other is convex in schematic Figure A.2-1. Referring to Figure A.2-1 (left), we can track the fibrillation response from a different perspective. At first, the fibril concentration was zero and *dF/dt* was positive. This gave the accelerated concentration until it reached the stationary point which corresponded to the inflection point of $F(t)$. Then, the fibril concentration kept on increasing but in a decelerated way until it reached the stable steady state (saturation condition). A similar argument can be made with respect to monomer concentration by using Figure A.2-1 (right).

Eq. 13 has a similar form to the Verhurst equation [5] and can be solved analytically to give

$$F(t) = r_1 + \frac{r_2 - r_1}{1 + e^{-(t - t_0)/\tau}} \tag{A.14}$$

where the constants are given by

$$\frac{1}{\tau} = (N k_{fb,1} - N^2 k_{nu,1})(r_2 - r_1) \tag{A.15}$$

$$t_0 = \tau \ln(-r_2 / r_1) \tag{A.16}$$

Given that there were no fibrils present at the beginning of the experiment, the analytic solution to the equation shares the same functional form as Eq. (A.6) when m_i and m_f are equal to zero. The models Eq. (A.6) and Eq. (A.14) can now be seem to be related to several experimentally measurable parameters. First, $1/\tau$ is the apparent growth rate constant k_{app}. Second, t_0 is the inflection point where the second order derivative of $F(t)$ equal to zero. Finally, the delay time can be calculated as $t_0 - 2\tau$.

A B

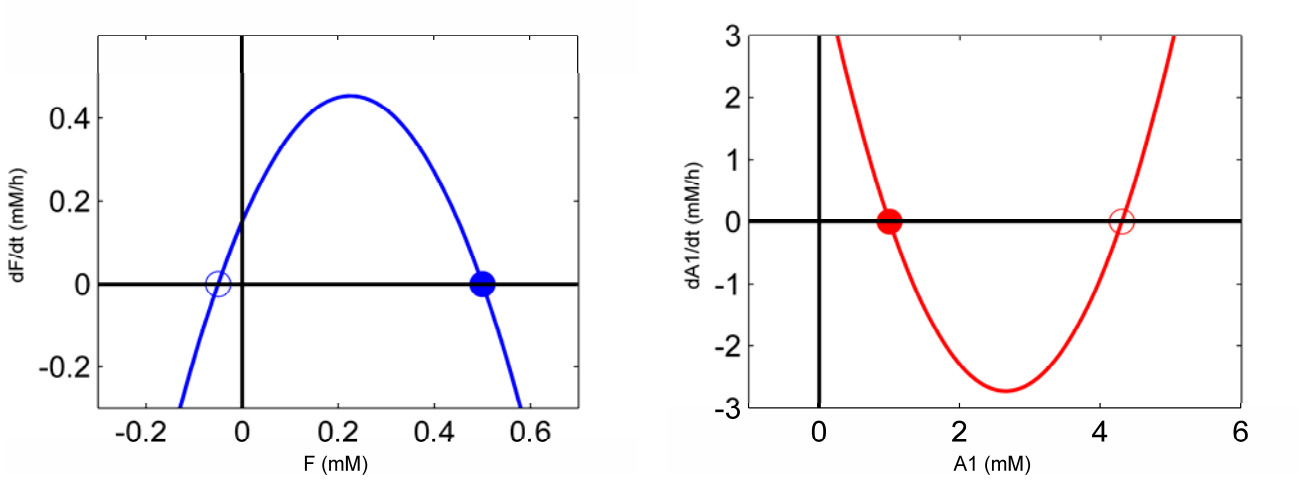

Figure A.2-1 (A) Time derivative of the fibril concentration as a function of the fibril concentration. (B) Time derivative of the monomer concentration as a function of the monomer concentration. Solid circles are attractors (stable steady states); open circles are repellers (unstable steady states).

A.3 Sensitivity analysis

Sensitivity analysis was conducted to evaluate the impact of each of the four parameters. The partial derivative of the model response with respect to the j-th parameter was computed by a finite difference method.

$$\frac{\partial f(\underline{\theta})}{\partial \theta_j} \cong \frac{f(\theta_j + \Delta\theta_j) - f(\theta_j)}{\Delta\theta_j} \tag{A.17}$$

To obtain a good estimate of the first derivate as well as to minimize the round-off error, the optimal value of $\Delta\theta_j$ was chosen to be 10^{-4} of θ_j. Due to the fact that rate constants estimated in this study had different units and their values span several orders of magnitude, the logarithmic from of Eq. 17 provided a better way to compare the relative contributions,

$$\frac{\partial \ln f(\underline{\theta})}{\partial \ln \theta_j} = \frac{\theta_j}{f} \frac{\partial f(\underline{\theta})}{\partial \theta_j} \tag{A.18}$$

Figure A.3-2 shows the results of logarithmic sensitivity analysis as a function of time demonstrating the impact of each parameter at different stages of fibrillation. Notice that changing the value of k_{nu-} several fold did not affect the Y response significantly compared with perturbing the other parameters so k_{nu-} was not considered in the parameter estimation. The rate constant $k_{nu,1}$, followed by k_{fb-} was the most significant parameter influencing the earlier stage of fibrillation; thus, the delay time is determined

by both $k_{nu,1}$ and k_{fb-}. Finally, the equilibrium gain of the signal roughly depended on the products of $k_{nu,1}$ and k_{fb-} divided by $k_{fb,1}$.

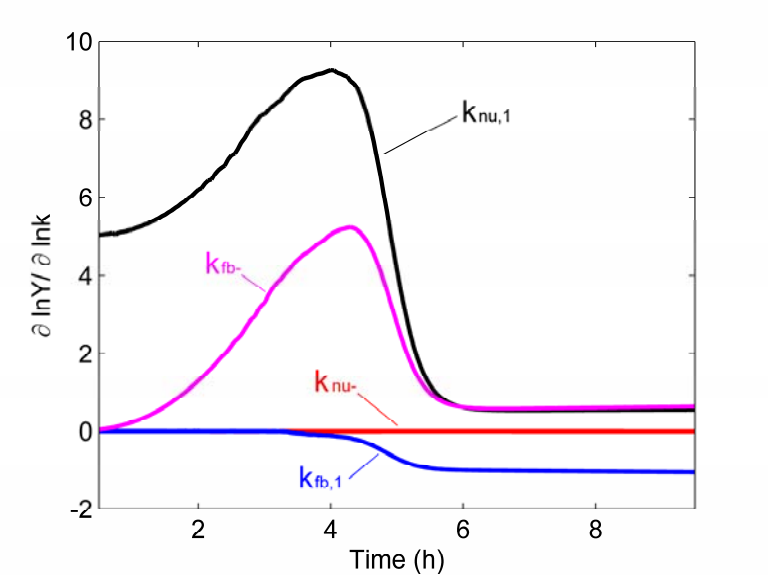

Figure A.3-2 *∂lnY/∂lnk* represents the logarithmic sensitivity analysis of the rate constant influence on the fibrillation response around the standard condition. Note that *Y* is ThT fluorescence intensity and rate constants include $k_{nu,1}$ (black), k_{nu-} (red), $k_{fb,1}$ (blue), and k_{fb-} (magenta).

A.4 MATLAB® Codes of the Model

The results shown in Fig. 2A and 2B in the main paper were obtained using MATLAB® [1] and two m-file functions**: `solveode_fibril`** and **`dA_dt=ode_fibril(t,A,n,theta)`.** The function **`solveode_fibril`,** sets up the parameters for the kinetic model and then solves the resulting set of differential equations using the MATLAB® function **`ode23s`**. The outputs include the fibrillation response and oligomer concentrations as functions of time. The right hand side of the differential equations are evaluated using the function **`dA_dt = ode_fibril(t,A,n,theta)`**.

```
function solveode_fibril;
% % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % %
% This function solves the set of differential equations defined @   %
% ode_fibril, and plots sigmoidal curves of fibril and oligomer conc.%
% % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % %

% The following is the list of parameters and initial conditions
n=6; % Critical number of monomers to form a neucleus
b=6e5; % The proportional constant of signal response
theta=[3.54e-2,2.73e6,1.94e3]; % The rate constants [knu1, kfb1, kfb_]
hex0_mg=2; % Initial insulin wt conc. of insulin hexamers (mg/mL)
insulin_kD=6; % The molecular weight of insulin monomer (kD)
```

```
hex0_mM=hex0_mg/insulin_kD/6; % Convert unit from mass to molar conc(mM)
Y0=zeros(1,n+1); % The initial concentration of insulin hexamers (mM)
Y0(n+1)=hex0_mM;

% Run the simulation and plot the results
close all
slots=201; % Specify number of simulation points
t_range=linspace(0,8,slots); % Specify the time scale vector
[t_val,Y_val]=ode23s(@ode_fibril,t_range,Y0,[],n,theta);
signal=Y_val(:,n)*b; % Fluores Signal is proportional to fibril conc.

% Plot the data of UV-vis signal and simulation results
plot(t_val, signal,'b-','LineWidth',1.5);
xlabel('Time (h)','FontSize', 14);
ylabel('UV-vis absorbance @ 600 nm','FontSize', 14);

% Plot the concentration of natural insulin hexamer, monomer, and dimer
figure
plot(t_val, Y_val(:,n+1)*6,'r:',t_val, Y_val(:,1),'b-',t_val, ...
Y_val(:,2),'k-.','LineWidth',1.5);
legend('natural hexamer','monomer','dimer');
xlabel('Time (h)','FontSize', 14);
ylabel ('Cluster Concentration (mM)','FontSize', 14)

% Plot the concentration of 3-mers, 4-mers, and 5-mers
figure
plot(t_val, Y_val(:,3),'r:',t_val, Y_val(:,4),'b-',t_val,
Y_val(:,5), ...
'k-.','LineWidth',1.5);
legend('3-mer','4-mer','5-mer','location','southeast');
xlabel('Time (h)','FontSize', 14);
ylabel ('Cluster Concentration (mM)','FontSize', 14)


function dA_dt=ode_fibril(t,A,n,theta)
% % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % %
% Defines the set of ODEs to be solved to simulate fibrillation      %
% A(1~n-1) is the vector of i-mer concentrations (i=1~n-1)           %
% A(n) is fibril conc. and A(n+1) is natural hexamer conc.           %
% t is time and dA_dt is the first order derivatives of A vector     %
% n is the critical size of clusters and is assigned to be 6         %
% theta vector is the set of rate constants [knu1, kfb1, kfb_]       %
% % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % % %

dA_dt=zeros(size(A)); % First order derivatives of i-mer concentrations
Jnu=zeros(size(A)); % Flux of i-th nucleation reaction
Jfb=zeros(size(A)); % Flux of i-th fibrillation reaction

% The following is the list of rate constants
khex=3; % Forward rate constant of insulin hexamer dissociation
knu=ones(n,1)*theta(1); % First foward nucleation rate constants
for i=1:n-1
   knu(i)=theta(1)/2*(1+i^(-1/3)); % Correct knu(i) by Stokes-Einstein
Eq.
end
knu_=ones(n,1)*1e-3; % Reverse nucleation constants
```

```
kfb=ones(n,1)*theta(2); % First forward fibrillation rate constant
for i=1:n-1
   kfb(i)=theta(2)*i^(-1/3); % Correct kfb(i) by Stokes-Einstein Eq.
end
kfb_=ones(n,1)*theta(3); % Reverse fibrillation rate constant

% Definitions of reaction fluxes Jhex, Jnu, and Jfb
Jhex=khex*A(n+1); % The flux of hexamer decomposition reaction
for i=1:n-1
   Jnu(i)=knu(i)*A(1)*A(i)-knu_(i)*A(i+1); % The flux of nucleation rxn
   Jfb(i)=kfb(i)*A(n)*A(i)-kfb_(i)*A(n); % The flux of i-mer elongation
end

% There are n+1 equations representing the conc. change of n+1 species
dA_dt(1)=6*Jhex-sum(Jnu(1:n-1))-Jnu(1)-Jfb(1); % Derivative of [monomer]
for i=2:n-1 % from dimer to (n-1)-mer
   dA_dt(i)=-Jnu(i)+Jnu(i-1)-Jfb(i); % Derivatives of [oligomer]
end
dA_dt(n)=Jnu(n-1); % Derivative of fibril concentration
dA_dt(n+1)=-Jhex; % Derivative of insulin hexamer concentration
```

A.5 Abstract of osmolyte controlled fibrillation kinetics of insulin

How proteins, implicated in amyloid diseases, impart toxicity is unknown. What is known is that they are converted from their native-fold to long 鶸-sheet-rich fibrils in a typical sigmoidal time-dependent curve. This reaction process from monomer or dimer to oligomer to nuclei and then to fibrils is the subject of intense study. Here, we probe this reaction process using a model protein, human insulin, through the addition of a comprehensive series of stabilizing and destabilizing osmolytes and quantify the analysis using our new kinetic rate model. The results are then collated into a cogent explanation using the preferential exclusion and accumulation of osmolytes away from and at the protein surface during nucleation, respectively. Both the heat of solution and the neutral molecular surface area of the osmolytes correlate linearly with two fitting parameters of the kinetic rate model, i.e. the lag-time and the nucleation rate prior to fibril formation. Also, the action of osmolytes on the reaction appears to be independent and additive. These kinetic and thermodynamic results support the preferential exclusion model and the existence of oligomers including nuclei and larger structures that could induce toxicity.

1. Matlab. The MathWorks, Natick, MA
2. Nielsen L, Khurana R, Coats A, Frokjaer S, Brange J, Vyas S, Uversky VN, Fink AL (2001) Effect of environmental factors on the kinetics of insulin fibril formation: elucidation of the molecular mechanism. Biochemistry 40:6036-6046
3. Seber GAF, Wild CJ (2003) Nonlinear Regression. John Wiley and Sons, New York City, NY
4. Strogatz SH (1994) Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering. Perseus Books, Cambridge, MA

5. Verhulst PF (1844) Recherches mathématiques sur la loi d'accroissement de la population. Mem Acad Roy, Brussels, Belgium