Supplementary Information:

Molecular Mechanisms of Protein Aggregation from Global Fitting of Kinetic Models.

Georg Meisl, Julius B. Kirkegaard, Paolo Arosio, Thomas C. T. Michaels, Michele Vendruscolo, Christopher M. Dobson, Sara Linse and Tuomas P. J. Knowles

1 Data layout

Only a specific format of data can be read in, it is easiest to produce this in a spreadsheet and then copy paste into a plain text file (.txt). The requirements are:

- The data needs to be in a text file.
- Columns correspond to individual experiments (if data is in rows use e.g. Edit > Paste Special > Transpose)
- The first column must be the x values (e.g. time) for the entire dataset. Several columns of y values (e.g. fluorescence intensity) can then follow.
- Columns must be separated by tabs (simply copy pasting from a spreadsheet into a text file should produce tab separated columns, using e.g. Notepad in Windows or TextEditor in Mac), or by comma, if you specify this in the 'Data Format Options'.
- Columns must be of equal length (datasets of different length need to be uploaded as separate files)
- No points can be missing (i.e. no empty cells).
- By default the first row will be assumed to be headers and used to name the datasets. If the first row is also data please tick the relevant box in 'Data Format Options'.

Nature Protocols: doi:10.1038/nprot.2016.010

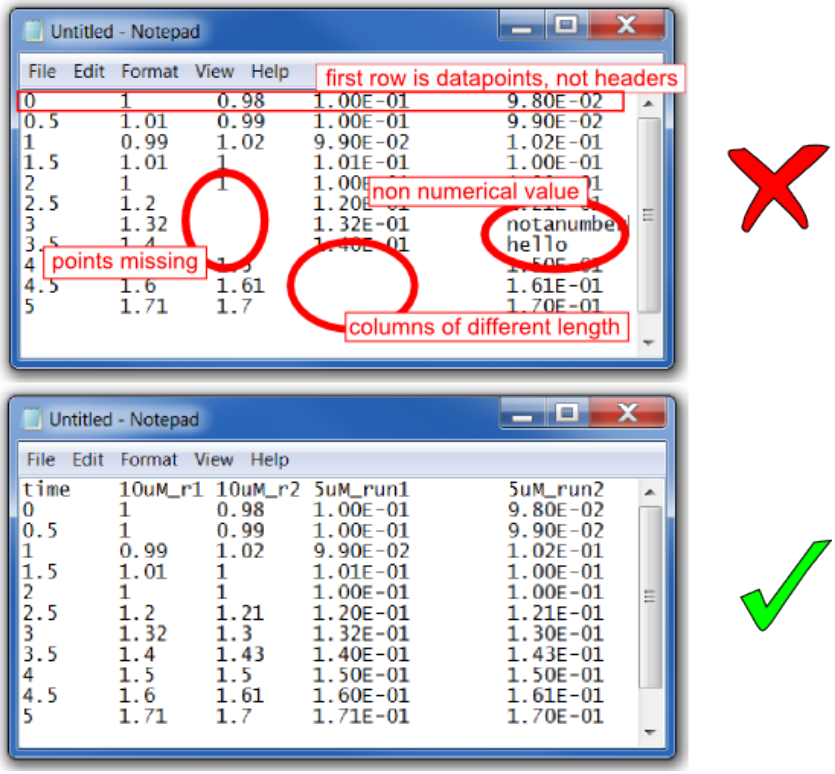

Figure S1: ***Data format for upload.*** *The top dataset shows 4 common problems: (1)The first row should contain headers not data. (2) The second row contains a non-numerical value. Except for the first row the file may contain only numbers. (3) The second column has missing values. (4)The third column is shorter than the rest of the dataset. The bottom dataset is formatted correctly.*

- Units: The units of time in the final output will be the same as the units of the time column in the initial input. The units of protein concentration are irrelevant, unless you decide not to normalise the data.
- A zip file containing several textfiles of this type may also be uploaded.

See also Fig. S1 and the sample data.

2 Interpreting half times and scalings

By considering how the half times scale with monomer concentration and how this scaling depends on the monomer concentration, one can obtain constraints on possible mechanisms. Fig. S2 guides you to a likely mechanism of aggregation using the information obtained from the double logarithmic plots of half time versus monomer concentration.

Nature Protocols: doi:10.1038/nprot.2016.010

A special note on scaling exponents of $-1/2$, that are constant with monomer concentration, marked by the exclamation mark in Fig. S2: A scaling exponent of $-1/2$, in the absence of saturation effects in elongation, suggests that the secondary processes are monomer independent. This can have 2 possible reasons, either aggregates multiply by fragmentation, or they multiply by fully saturated secondary nucleation (i.e. secondary nucleation where $m_{\mathrm{min}}^{n_2} \gg K_M$). Both mechanisms have a monomer independent secondary nucleation process and in fact are described by exactly the same differential equations for $P(t)$ and $M(t)$. Therefore it is impossible to distinguish between them within the framework of this analysis and different experimental methods need to be employed in order to make the distinction. One possible course of action is to vary the monomer concentrations: if at lower monomer concentrations an increase in scaling is observed the system is likely dominated by secondary nucleation and the initial experiment simply sampled a region of higher monomer where this was fully saturated. By contrast if an increase in scaling is observed at higher concentrations, this suggests that two processes are competing in parallel and the initial, lower monomer, data was dominated by fragmentation, whereas the new, higher monomer data is dominated by secondary nucleation. However, this approach is limited by the monomer concentrations accessible and by the condition that a change in scaling needs to be observed.

2.1 Extracting half times

The algorithm for the extraction of the half times proceeds as follows: first the middle part of the curve is selected, by determining when the average over several points is first above 0.3 and when the average is last below 0.7. The number of points to be averaged over depends on the number of points in the curve. A straight line is then fitted to this middle part of the curve, the point at which it crosses the value of 0.5 is recorded as the half time.

3 Integrated rate laws and approximate scalings

For a general strategy for obtaining the approximate analytical solutions to the rate equations see [4, 3, 2]. Unless stated otherwise the meaning of the parameters is as defined in the main text, m_{tot} is the total protein mass concentration, $m_{\mathrm{tot}} = m(t)+M(t)$.

The scaling exponent as used here is defined as:

$$\gamma = \frac{d}{d(\log(m_0))} \log(t_{1/2}) = m_0 \frac{d}{dm_0} \log(t_{1/2}) \tag{S1}$$

Nature Protocols: doi:10.1038/nprot.2016.010

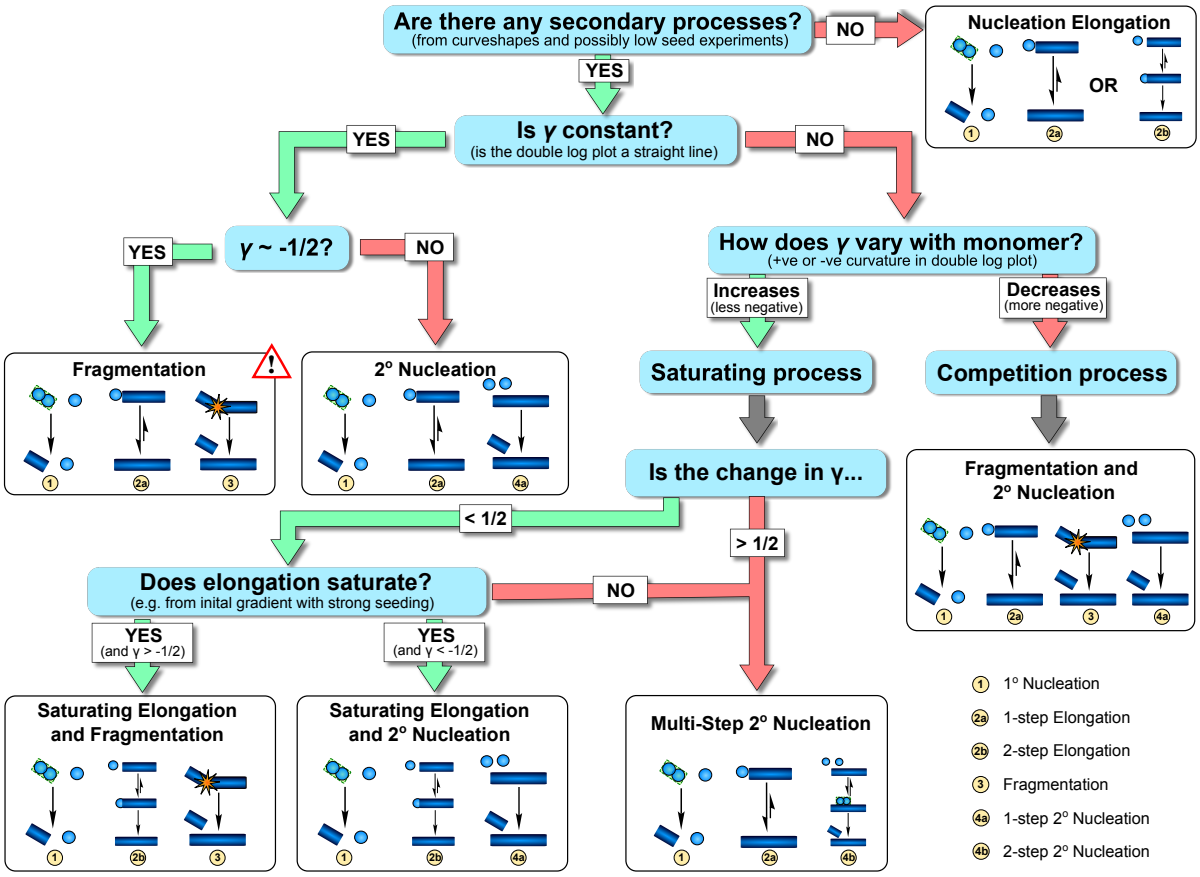

Figure S2: ***From half times to models.*** *The curvature of the double logarithmic plots and the value of their slopes give insights into which aggregation mechanisms are dominant. The flowchart illustrates the decision process to arrive at a likely model, based on the half time plots. The individual processes of each model are shown schematically, for more details see Fig. 6. Care needs to be taken with the interpretation of a scaling exponent of* $-1/2$*, as explained in the text.*

Nature Protocols: doi:10.1038/nprot.2016.010

3.1 Nucleation Elongation

The differential equations describing this system are:

$$\frac{dP}{dt} = k_n m(t)^{n_c} \tag{S2}$$

$$\frac{dM}{dt} = 2(m(t)k_+ - k_{\text{off}})P(t) \tag{S3}$$

For a negligible depolymerisation rate, $k_{\text{off}} \ll k_+ m_0$, the closed for solution, based on Oosawa[6], is:

$$\frac{M}{m_{\text{tot}}} = 1 - \frac{m_0}{m_{\text{tot}}} \left(\frac{1}{\mu} \cosh(\sqrt{\frac{n_c}{2}} \mu \lambda t + \nu) \right)^{-\frac{2}{n_c}} \tag{S4}$$

where the definitions of the parameters are

$$\lambda = \sqrt{2k_+ k_n m_0^{n_c}} \tag{S5}$$

$$\alpha = \sqrt{\frac{k_+ n_c}{k_n m_0^{n_c}}} P_0 \tag{S6}$$

$$\mu = \sqrt{1 + \alpha^2} \tag{S7}$$

$$\nu = \log(\alpha + \mu) \tag{S8}$$

where m_{tot} is the total protein mass concentration.

In the unseeded case this depends only on the combined rate constant $k_+ k_n$, not k_+ and k_n individually.

The approximate scaling exponent is:

$$\gamma \approx -\frac{n_c}{2} \tag{S9}$$

This shows no variation with monomer concentration.

3.2 Secondary Nucleation

The differential equations describing this system are:

$$\frac{dP}{dt} = k_n m(t)^{n_c} + k_2 m(t)^{n_2} M(t) \tag{S10}$$

$$\frac{dM}{dt} = 2m(t)k_+ P(t) \tag{S11}$$

Nature Protocols: doi:10.1038/nprot.2016.010

The approximate analytical solution as obtained in [1] is:

$$\begin{aligned}\frac{M}{M_\infty} &= 1-\left(1-\frac{M_0}{M_\infty}\right)e^{-k_\infty t} \\ &\quad\cdot\left(\frac{B_-+C_+e^{\kappa t}}{B_++C_+e^{\kappa t}}\cdot\frac{B_++C_+}{B_-+C_+}\right)^{\frac{k_\infty}{\kappa k_\infty}}\end{aligned} \tag{S12}$$

where the definitions of the parameters are

$$\kappa = \sqrt{2m_0k_+m_0^{n_2}k_2} \tag{S13}$$

$$\lambda = \sqrt{2k_+k_nm_0^{n_c}} \tag{S14}$$

$$C_\pm = \frac{k_+P_0}{\kappa}\pm\frac{k_+M_0}{2m_0k_+}\pm\frac{\lambda^2}{2\kappa^2} \tag{S15}$$

$$k_\infty = \sqrt{(2k_+P(0))^2+\frac{4k_+k_nm_0^{n_c}}{n_c}+\frac{4k_+k_2m_{\mathrm{tot}}m_0^{n_2}}{n_2}+\frac{4k_+k_2m_0^{n_2+1}}{n_2+1}} \tag{S16}$$

$$\bar{k}_\infty = \sqrt{k_\infty^2-2C_+C_-\kappa^2} \tag{S17}$$

$$B_\pm = \frac{k_\infty\pm\bar{k}_\infty}{2\kappa} \tag{S18}$$

where the mass at long times, M_∞, is just the total protein mass concentration, m_{tot}. This solution is more accurate than the ones involving fragmentation, i.e. very close to the numerical integration of the differential equations, however it only applies for negligible off rates $k_{\mathrm{off}} \ll k_+m_0$.

In the unseeded case this depends only on the combined rate constants k_+k_n and k_+k_2, not k_+, k_n and k_2 individually.

The approximate scaling exponent is:

$$\gamma \approx -\frac{n_2+1}{2} \tag{S19}$$

This shows no variation with monomer concentration.

3.3 Fragmentation

The differential equations describing this system are:

$$\frac{dP}{dt} = k_nm(t)^{n_c}+k_-M(t) \tag{S20}$$

$$\frac{dM}{dt} = 2(m(t)k_+-k_{\mathrm{off}})P(t) \tag{S21}$$

The linearised solution as obtained in [4] is:

Nature Protocols: doi:10.1038/nprot.2016.010

$$
\begin{aligned}
M(t) &= M_\infty + \mathrm{Exp}\left[-\frac{k_+(4c\kappa\mathrm{Cosh}(\kappa t) + 4P_0\kappa^2\mathrm{Sinh}(\kappa t))}{2\kappa^3}\right] \\
&\quad \left((M_0 - M_\infty)e^{\frac{2k_+c}{\kappa^2}}\right) \qquad \text{(S22)}
\end{aligned}
$$

where

$$
\begin{aligned}
c &= k_n m_0^{n_c} + k_- M_0 \\
\kappa &= \sqrt{2(k_+ m_0 - k_{\mathrm{off}})k_-} \\
M_\infty &= m_{\mathrm{tot}} - k_{\mathrm{off}}/k_+ \qquad \text{(S23)}
\end{aligned}
$$

where m_{tot} is the total protein mass concentration.

In the unseeded case, for negligible k_{off}, this depends only on the combined rate constants k_+k_n and k_+k_-, not k_+, k_n and k_- individually.

The approximate scaling exponent is:

$$
\gamma = -\frac{1}{2} \qquad \text{(S24)}
$$

This shows no variation with monomer concentration.

3.4 Fragmentation and Secondary Nucleation

The differential equations describing this system are:

$$
\begin{aligned}
\frac{dP}{dt} &= k_n m(t)^{n_c} + k_- M(t) + k_2 m(t)^{n_2} M(t) \qquad \text{(S25)} \\
\frac{dM}{dt} &= 2(m(t)k_+ - k_{\mathrm{off}})P(t) \qquad \text{(S26)}
\end{aligned}
$$

The linearised solution is obtained by setting $m(t) = m_0$. This is then used in a fixed point iteration to give the first order fixed point solution:

$$
\begin{aligned}
M(t) &= M_\infty + \mathrm{Exp}\left[-\frac{k_+(4c\kappa\mathrm{Cosh}(\kappa t) + 4P_0\kappa^2\mathrm{Sinh}(\kappa t))}{2\kappa^3}\right] \\
&\quad \left((M_0 - M_\infty)e^{\frac{2k_+c}{\kappa^2}}\right) \qquad \text{(S27)}
\end{aligned}
$$

where

$$
\begin{aligned}
a &= k_2 m_0^{n_2} + k_- \\
c &= k_n m_0^{n_c} + aM_0 \\
\kappa &= \sqrt{2(k_+ m_0 - k_{\mathrm{off}})(k_2 m_0^{n_2} + k_-)} \\
M_\infty &= m_{\mathrm{tot}} - k_{\mathrm{off}}/k_+ \qquad \text{(S28)}
\end{aligned}
$$

Nature Protocols: doi:10.1038/nprot.2016.010

In the unseeded case, for negligible k_{off}, this depends only on the combined rate constants k_+k_n, k_+k_2 and k_+k_-, not k_+, k_n, k_2 and k_- individually.

The approximate scaling exponent is:

$$\gamma = \frac{d\log(t_{1/2})}{d\log(m(0))} \approx -\frac{1}{2}\left(\frac{n_2}{1+K/m(0)^{n_2}}+1\right) \tag{S29}$$

where $K = k_-/k_2$. In the limit of low and high monomer concentration this becomes $\gamma = -1/2$ and $\gamma = -(n_2+1)/2$ respectively, i.e. the scaling exponent decreases with increasing monomer and there is negative curvature in the double logarithmic plots.

3.5 Multi-Step Secondary Nucleation

The differential equations describing this system are:

$$\frac{dP}{dt} = k_n m(t)^{n_c} + k_2 \frac{m(t)^{n_2}}{1+m(t)^{n_2}/K_M} M(t) \tag{S30}$$

$$\frac{dM}{dt} = 2m(t)k_+P(t) \tag{S31}$$

The approximate analytical solution as obtained in [5] is:

$$\begin{aligned}\frac{M}{M_\infty} &= 1-\left(1-\frac{M_0}{M_\infty}\right)e^{-k_\infty t} \\ &\quad \cdot \left(\frac{B_- + C_+e^{\kappa t}}{B_+ + C_+e^{\kappa t}} \cdot \frac{B_+ + C_+}{B_- + C_+}\right)^{\frac{k_\infty}{\kappa k_\infty}}\end{aligned} \tag{S32}$$

where the definitions of the parameters are

$$\kappa = \sqrt{2m_0k_+\frac{m_0^{n_2}k_2}{1+m_0^{n_2}/K_M}} \tag{S33}$$

$$\lambda = \sqrt{2k_+k_nm_0^{n_c}} \tag{S34}$$

$$C_\pm = \frac{k_+P_0}{\kappa} \pm \frac{k_+M_0}{2m_0k_+} \pm \frac{\lambda^2}{2\kappa^2} \tag{S35}$$

$$k_\infty = \sqrt{(2k_+P(0))^2 - 2A - 4k_+k_2m_{\text{tot}}K_M\frac{\log\left[K_M\right]}{n_2}} \tag{S36}$$

$$\begin{aligned}A &= -\frac{2k_+k_nm_0^{n_c}}{n_c} - 2k_+k_2m_{\text{tot}}K_M\frac{\log\left[K_M+m_0^{n_2}\right]}{n_2} \\ &\quad - 2k_+k_2K_Mm_0\left({}_2F_1\left[\frac{1}{n_2},1,1+\frac{1}{n_2},-\frac{m_0^{n_2}}{K_M}\right]-1\right)\end{aligned} \tag{S37}$$

$$\bar{k}_\infty = \sqrt{k_\infty^2 - 2C_+C_-\kappa^2} \tag{S38}$$

$$B_\pm = \frac{k_\infty \pm \bar{k}_\infty}{2\kappa} \tag{S39}$$

Nature Protocols: doi:10.1038/nprot.2016.010

where the mass at long times, M_∞, is just the total protein mass concentration, $m_{\rm tot}$. This solution is more accurate than the ones involving fragmentation, i.e. very close to the numerical integration of the differential equations, however it only applies for negligible off rates $k_{\rm off} \ll k_+ m_0$.

In the unseeded case this depends only on the combined rate constants k_+k_n and k_+k_2, not k_+, k_n and k_2 individually.

The approximate scaling exponent is:

$$\gamma \approx -\frac{1}{2}\left(\frac{n_2}{1+m(0)^{n_2}/K_M}+1\right) \tag{S40}$$

In the limit of low and high monomer concentration this becomes $\gamma = -(n_2+1)/2$ and $\gamma = -1/2$respectively, i.e. the scaling exponent increases with increasing monomer and there is positive curvature in the double logarithmic plots.

3.6 Saturating Elongation

The differential equations describing this system are:

$$\frac{dP}{dt} = k_n m(t)^{n_c} \tag{S41}$$

$$\frac{dM}{dt} = 2k_+ P(t)\frac{m(t)}{1+m(t)/K_E} \tag{S42}$$

$$\begin{aligned}\frac{M}{m_{\rm tot}} = 1 - \frac{K_E}{m_{\rm tot}}\mathrm{W}\Bigg[&\frac{m_{\rm tot}-M_0}{K_E} \\ &\exp\left(\frac{m_{\rm tot}-M_0-K_E k_+ t(2P_0+k_n m_0^{n_c} t)}{K_E}\right)\Bigg]\end{aligned} \tag{S43}$$

where $W(x)$ is the Lambert W function.

In the unseeded case this depends only on the combined rate constants k_+k_n and k_+k_2, not k_+, k_n and k_2 individually.

The scaling exponent is:

$$\gamma = -\frac{nc}{2} + \frac{1}{2}\frac{1}{1+\log(4)K_E/m_0} \tag{S44}$$

In the limit of low and high monomer concentration this becomes $\gamma = -(n_c)/2$ and $\gamma = -(n_c-1)/2$ respectively, i.e. the scaling exponent increases by 0.5 with increasing monomer and there is positive curvature in the double logarithmic plots.

Nature Protocols: doi:10.1038/nprot.2016.010

3.7 Saturating Elongation and Secondary Nucleation

The differential equations describing this system are:

$$\frac{dP}{dt} = k_n m(t)^{n_c} + k_2 m(t)^{n_2} M(t) \tag{S45}$$

$$\frac{dM}{dt} = 2k_+ P(t) \frac{m(t)}{1 + m(t)/K_E} \tag{S46}$$

The approximate analytical solution is:

$$\frac{M}{m_{\text{tot}}} = 1 - \frac{K_E}{m_{\text{tot}}} \text{W} \left[\frac{m_{\text{tot}} - M_0}{K_E} \exp\left(\frac{(m_{\text{tot}} - M0)\kappa^2 - 2K_E k_+ (P_0 \kappa \sinh(\kappa t) + \alpha(\cosh(\kappa t) - 1))}{K_E \kappa^2} \right) \right] \tag{S47}$$

where $W(x)$ is the Lambert W function and the definitions of the parameters are

$$\kappa = \sqrt{2 \frac{m_0 k_+}{1 + m_0/K_E} m_0^{n_2} k_2} \tag{S48}$$

$$\alpha = k_n m_0^n c + k_2 m_0^n 2M0 \tag{S49}$$

In the unseeded case this depends only on the combined rate constants $k_+ k_n$ and $k_+ k_2$, not k_+, k_n and k_2 individually.

The approximate scaling exponent is:

$$\gamma \approx -\frac{n_2}{2} - \frac{1}{2(1 + m(0)/K_E)} \tag{S50}$$

In the limit of low and high monomer concentration the scaling exponent becomes $\gamma = -(n_2 + 1)/2$ and $\gamma = -(n_2)/2$ respectively, i.e. the scaling exponent increases by 0.5 with increasing monomer and there is positive curvature in the double logarithmic plots.

3.8 Saturating Elongation and Fragmentation

The differential equations describing this system are:

$$\frac{dP}{dt} = k_n m(t)^{n_c} + k_- M(t) \tag{S51}$$

$$\frac{dM}{dt} = 2k_+ P(t) \frac{m(t)}{1 + m(t)/K_E} \tag{S52}$$

Nature Protocols: doi:10.1038/nprot.2016.010

The approximate analytical solution is:

$$\frac{M}{m_{\rm tot}} = 1 - \frac{K_E}{m_{\rm tot}} \mathrm{W}\left[\frac{m_{\rm tot} - M_0}{K_E} \exp\left(\frac{(m_{\rm tot} - M0)\kappa^2 - 2K_E k_+(P_0\kappa\sinh(\kappa t) + \alpha(\cosh(\kappa t) - 1))}{K_E\kappa^2}\right)\right] \quad \text{(S53)}$$

where $W(x)$ is the Lambert W function and the definitions of the parameters are

$$\kappa = \sqrt{2\frac{m_0 k_+}{1 + m_0/K_E}k_-} \quad \text{(S54)}$$

$$\alpha = k_n m_0^n c + k_- M0 \quad \text{(S55)}$$

In the unseeded case this depends only on the combined rate constants k_+k_n and k_+k_-, not k_+, k_n and k_- individually.

The approximate scaling exponent is:

$$\gamma \approx -\frac{1}{2(1 + m(0)/K_E)} \quad \text{(S56)}$$

In the limit of low and high monomer concentration the scaling exponent becomes $\gamma = -1/2$ and $\gamma = 0$ respectively, i.e. the scaling exponent increases by 0.5 with increasing monomer and there is positive curvature in the double logarithmic plots.

References

[1] Samuel I. A. Cohen, Sara Linse, Leila M. Luheshi, Erik Hellstrand, Duncan A. White, Luke Rajah, Daniel E. Otzen, Michele Vendruscolo, Christopher M. Dobson, and Tuomas P. J. Knowles. Proliferation of amyloid-beta42 aggregates occurs through a secondary nucleation mechanism. *Proceedings of the National Academy of Sciences*, 110:9758–9763, 2013.

[2] Samuel I A Cohen, Michele Vendruscolo, Christopher M Dobson, and Tuomas P J Knowles. Nucleated polymerization with secondary pathways. ii. determination of self-consistent solutions to growth processes described by non-linear master equations. *J Chem Phys*, 135(6):065106, Aug 2011.

[3] Samuel I A Cohen, Michele Vendruscolo, Mark E Welland, Christopher M Dobson, Eugene M Terentjev, and Tuomas P J Knowles. Nucleated polymerization with secondary pathways. i. time evolution of the principal moments. *J Chem Phys*, 135(6):065105, Aug 2011.

Nature Protocols: doi:10.1038/nprot.2016.010

[4] Tuomas P J Knowles, Christopher A Waudby, Glyn L Devlin, Samuel I A Cohen, Adriano Aguzzi, Michele Vendruscolo, Eugene M Terentjev, Mark E Welland, and Christopher M Dobson. An analytical solution to the kinetics of breakable filament assembly. *Science*, 326(5959):1533–1537, Dec 2009.

[5] Georg Meisl, Xiaoting Yang, Erik Hellstrand, Birgitta Frohm, Julius B. Kirkegaard, Samuel I. A. Cohen, Christopher M. Dobson, Sara Linse, and Tuomas P. J. Knowles. Differences in nucleation behavior underlie the contrasting aggregation kinetics of the aβ40 and aβ42 peptides. *Proceedings of the National Academy of Sciences*, 111:9384–9389, 2014.

[6] Fumio Oosawa and Sho Asakura. *Thermodynamics of the Polymerization of Protein*. Academic Press, 1975.

Nature Protocols: doi:10.1038/nprot.2016.010