Stochastic coagulation-fragmentation models for the study of protein aggregation phenomena

Romain Yvinec[1], **Samuel Bernard**[2,3], **Tom Chou**[4], **Julien Deschamps**[5], **Maria R. D'Orsogna**[6], **Erwan Hingant**[7] and **Laurent Pujo-Menjouet**[2,3]

[1]BIOS group INRA Tours, France.

[2]Institut Camille Jordan, Université Lyon 1, Villeurbanne, France.

[3]INRIA Team Dracula, Inria Center Grenoble Rhône-Alpes, France.

[4]Depts. of Biomathematics and Mathematics, UCLA, Los Angeles, USA.

[5]DIMA, Universita degli Studi di Genova, Italy.

[6]Dept. of Mathematics, CSUN, Los Angeles, USA.

[7]Departamento de Matemática, U. Federal de Campina Grande, PB, Brasil.

Amyloid diseases and Becker-Döring model

Numerical results

Coarse-graining

Open problems (for me)

Outline

Amyloid diseases and Becker-Döring model

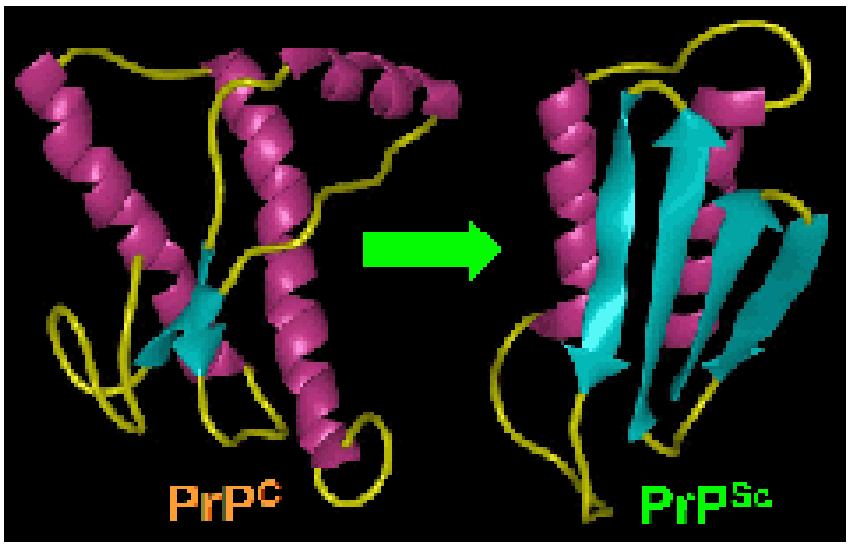
Numerical results

Coarse-graining

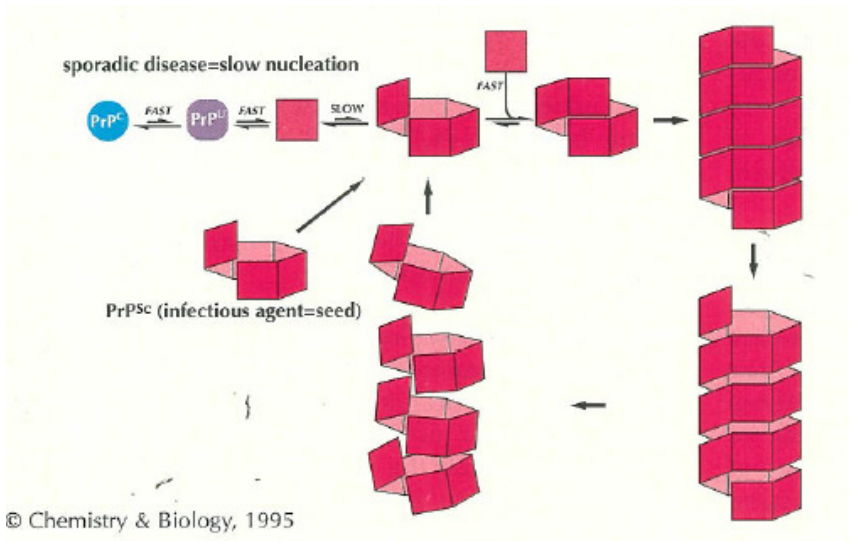
Open problems (for me)

Protein accumulation in amyloid by nucleation-polymerization

Misfolding



Prusiner model for prion



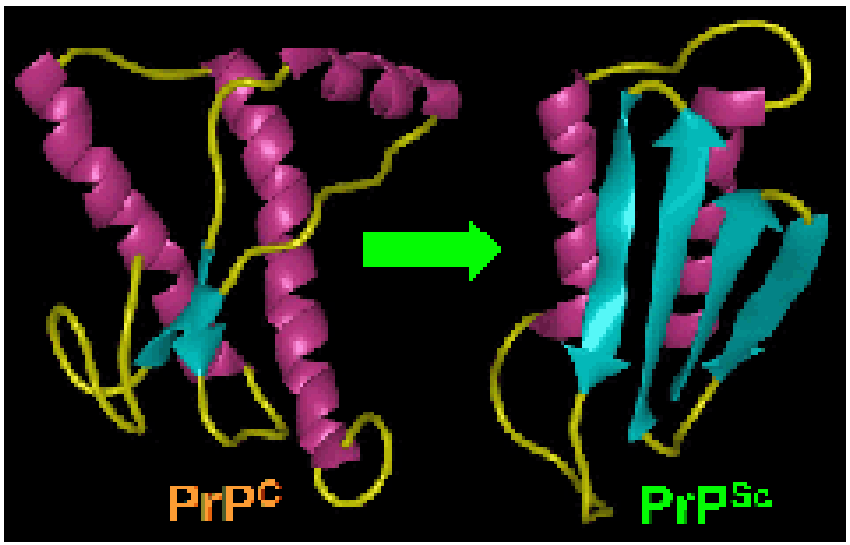
Open access, freely available online PLOS BIOLOGY

Mechanism of Prion Propagation: Amyloid Growth Occurs by Monomer Addition

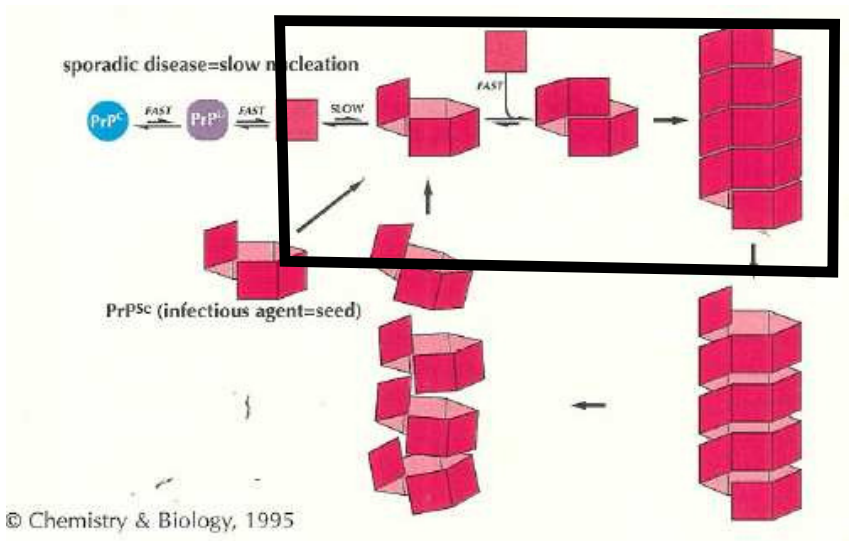
Sean R. Collins, Adam Douglass, Ronald D. Vale, Jonathan S. Weissman*

Protein accumulation in amyloid by nucleation-polymerization

Misfolding



Prusiner model for prion

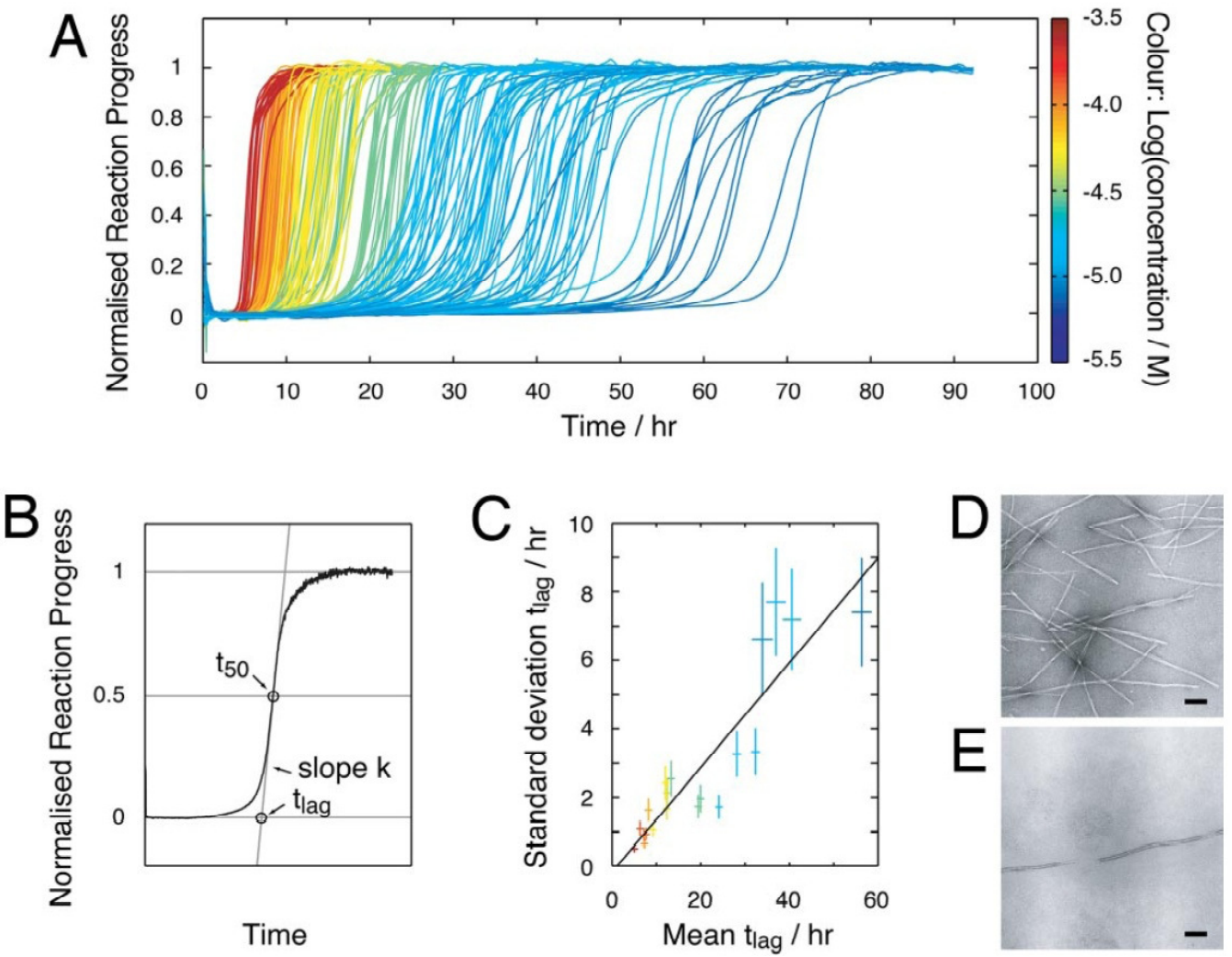


Open access, freely available online PLOS BIOLOGY

Mechanism of Prion Propagation: Amyloid Growth Occurs by Monomer Addition

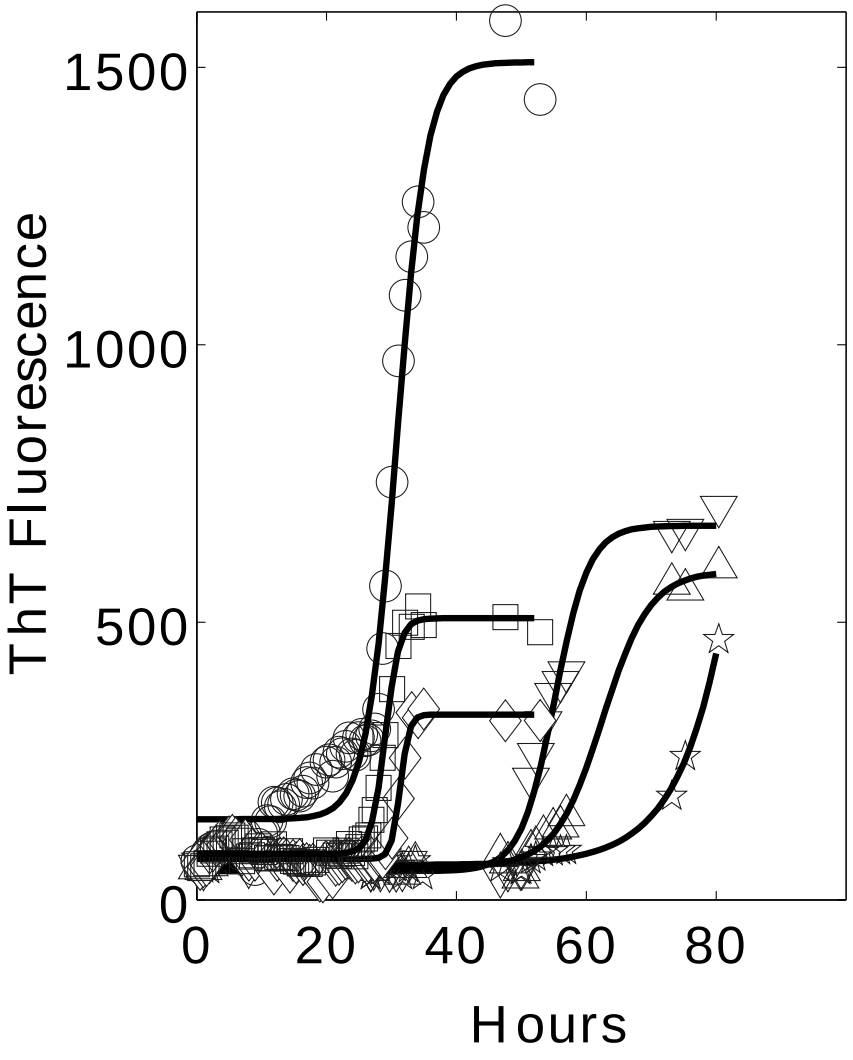
Sean R. Collins, Adam Douglass, Ronald D. Vale, Jonathan S. Weissman*

Times series of *in-vitro* spontaneous polymerization

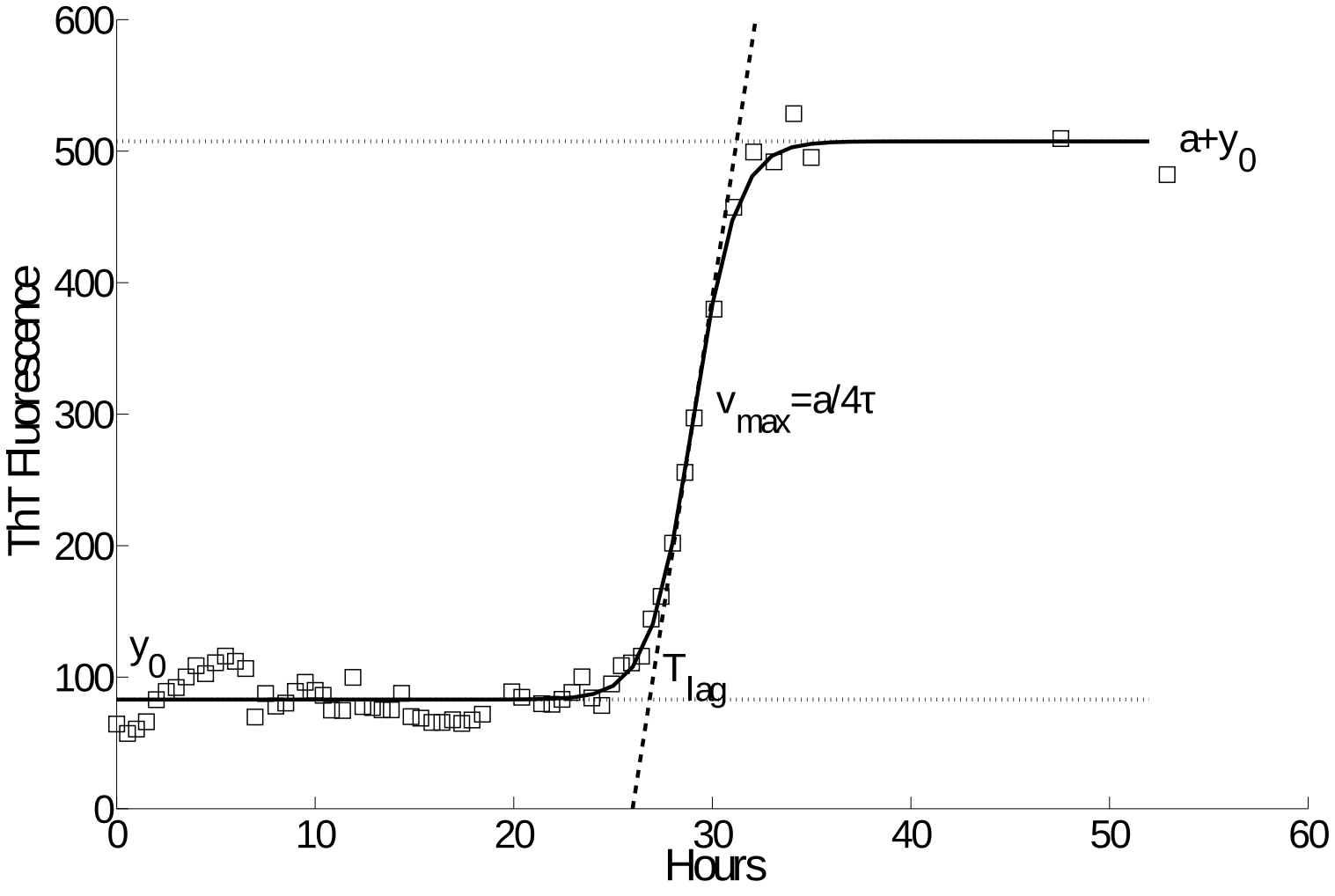


Xue et al. PNAS (2008)
Eugene el al. hal-01205549 (2015)

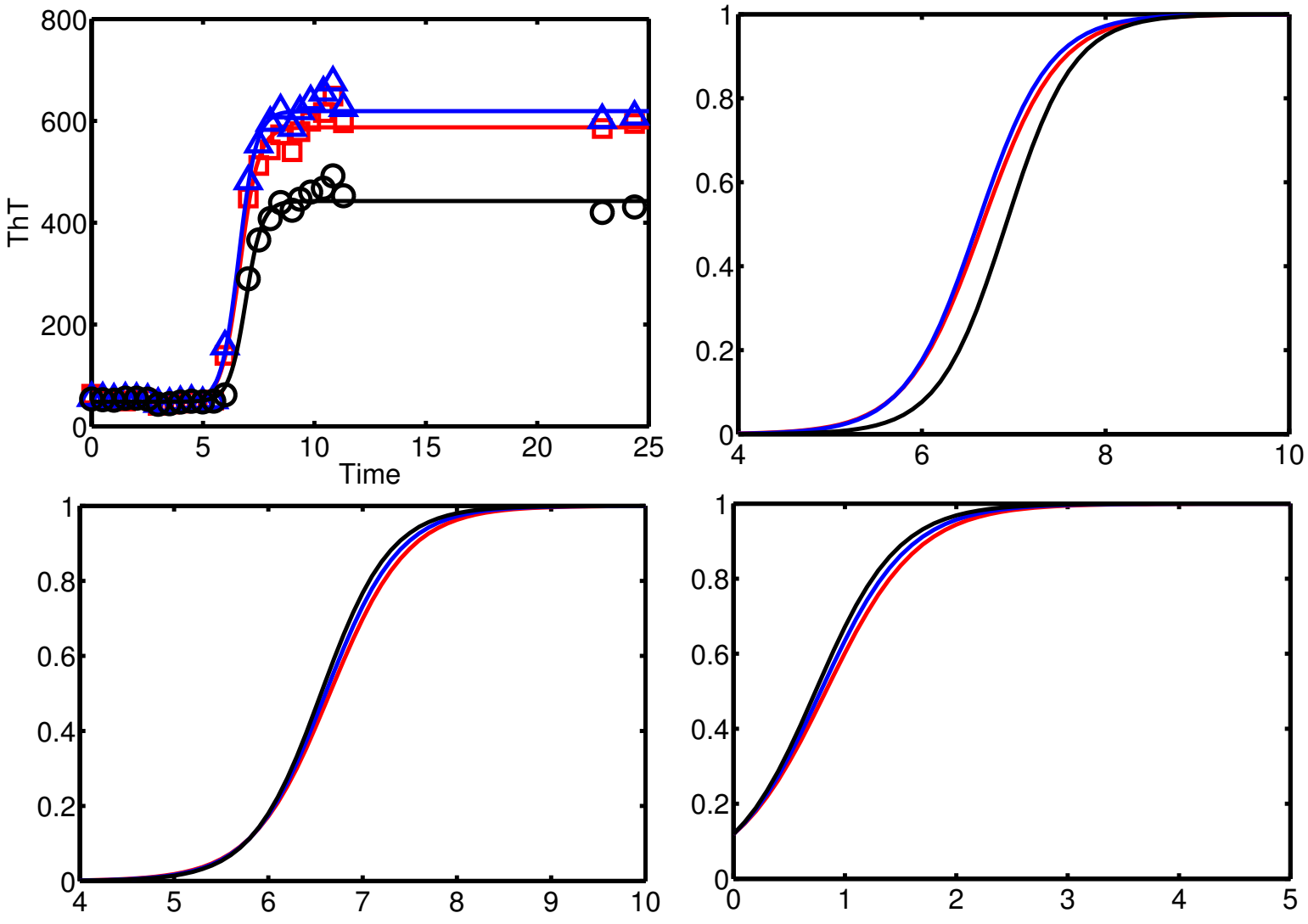
Times series of *in-vitro* spontaneous polymerization



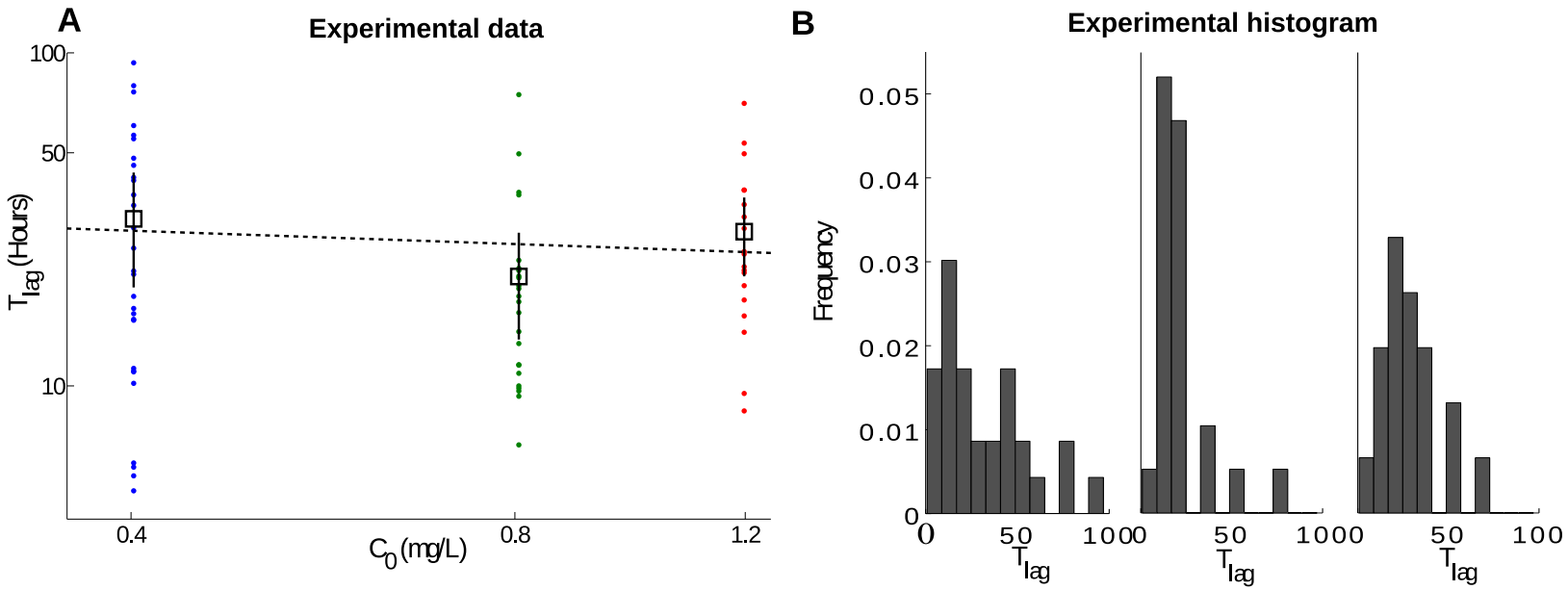
Quantification of experiment



Quantification of experiment



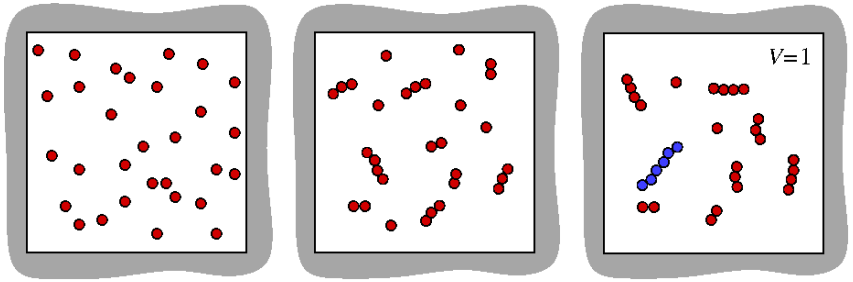
Statistics of lag time



Nucleation time in the Stochastic Becker-Döring model

Reversible one-step agregation

$$C_i + C_1 \underset{q_{i+1}}{\overset{p_i}{\rightleftharpoons}} C_{i+1} \qquad (1)$$



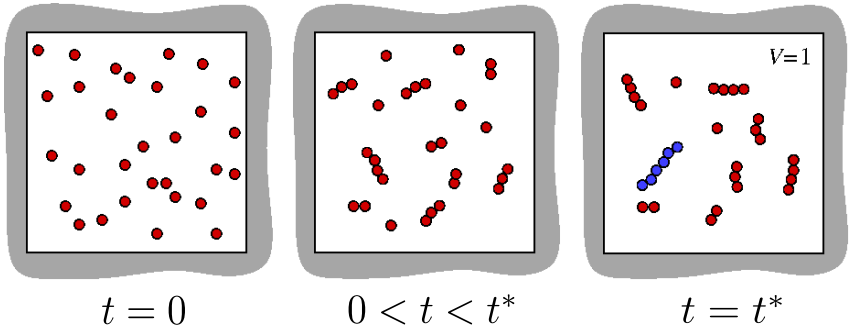
The **nucleation time** is given by the following First Passage Time,

$$T^{N,M} := \inf\{t \geqslant 0 : C_N(t) = 1 \mid C_i(t = 0) = M\delta_{i=1}\}. \qquad (2)$$

Nucleation time in the Stochastic Becker-Döring model

Reversible one-step agregation

$$C_i + C_1 \underset{q_{i+1}}{\overset{p_i}{\rightleftharpoons}} C_{i+1} \qquad (1)$$



The **nucleation time** is given by the following First Passage Time,

$$T^{N,M} := \inf\{t \geqslant 0 : C_N(t) = 1 \mid C_i(t = 0) = M\delta_{i=1}\}. \qquad (2)$$

- What are the dependencies of the nucleation time with respect to the model parameters ?

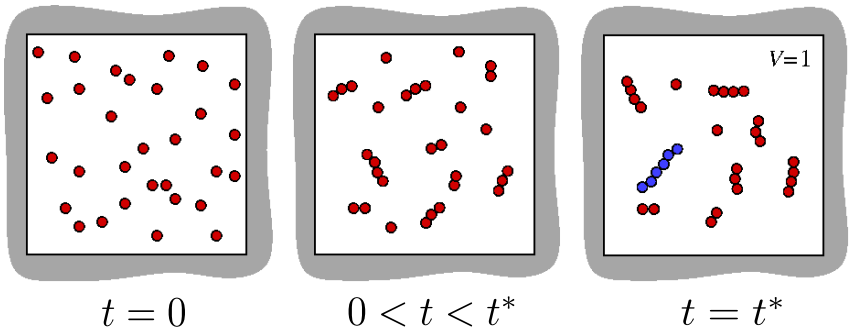
total mass : M ; nucleus size : N

aggregation rates : p_i , $i \geqslant 1$ fragmentation rates : q_i , $i \geqslant 2$

Nucleation time in the Stochastic Becker-Döring model

Reversible one-step agregation

$$C_i + C_1 \underset{q_{i+1}}{\overset{p_i}{\rightleftharpoons}} C_{i+1} \qquad (1)$$



The **nucleation time** is given by the following First Passage Time,

$$T^{N,M} := \inf\{t \geqslant 0 : C_N(t) = 1 \mid C_i(t = 0) = M\delta_{i=1}\}. \qquad (2)$$

- What is the nucleation time for very large initial quantity M and nucleus size N ?

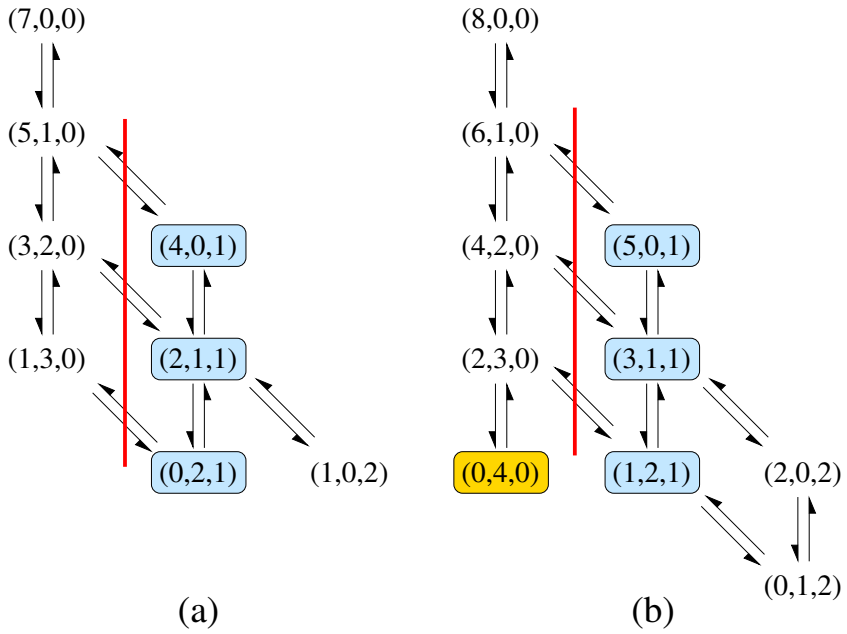
$$\lim_{M,N\to\infty} T^{N,M}$$

In data, $M \approx 10^{10} - 10^{15}$, Size of (observed) polymers $\approx 10^3 - 10^6$, $N =$?.

Nucleation time in the Stochastic Becker-Döring model

Reversible one-step agregation

$$C_i + C_1 \underset{q_{i+1}}{\overset{p_i}{\rightleftharpoons}} C_{i+1}$$



The **nucleation time** is given by the following First Passage Time,

$$T^{N,M} := \inf\{t \geqslant 0 : C_N(t) = 1 \mid C_i(t=0) = M\delta_{i=1}\}\,.$$

Outline

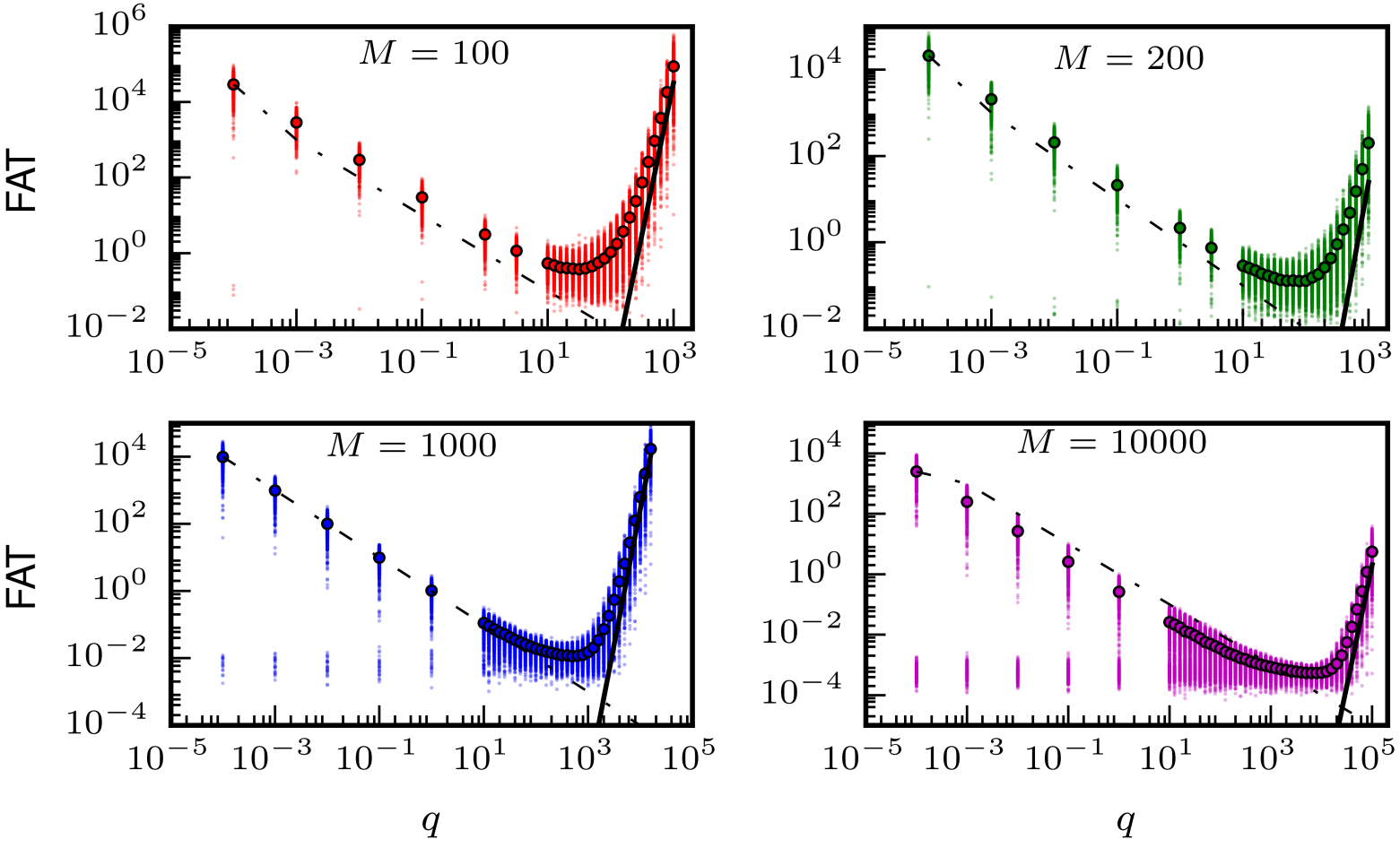
Amyloid diseases and Becker-Döring model

Numerical results

Coarse-graining

Open problems (for me)

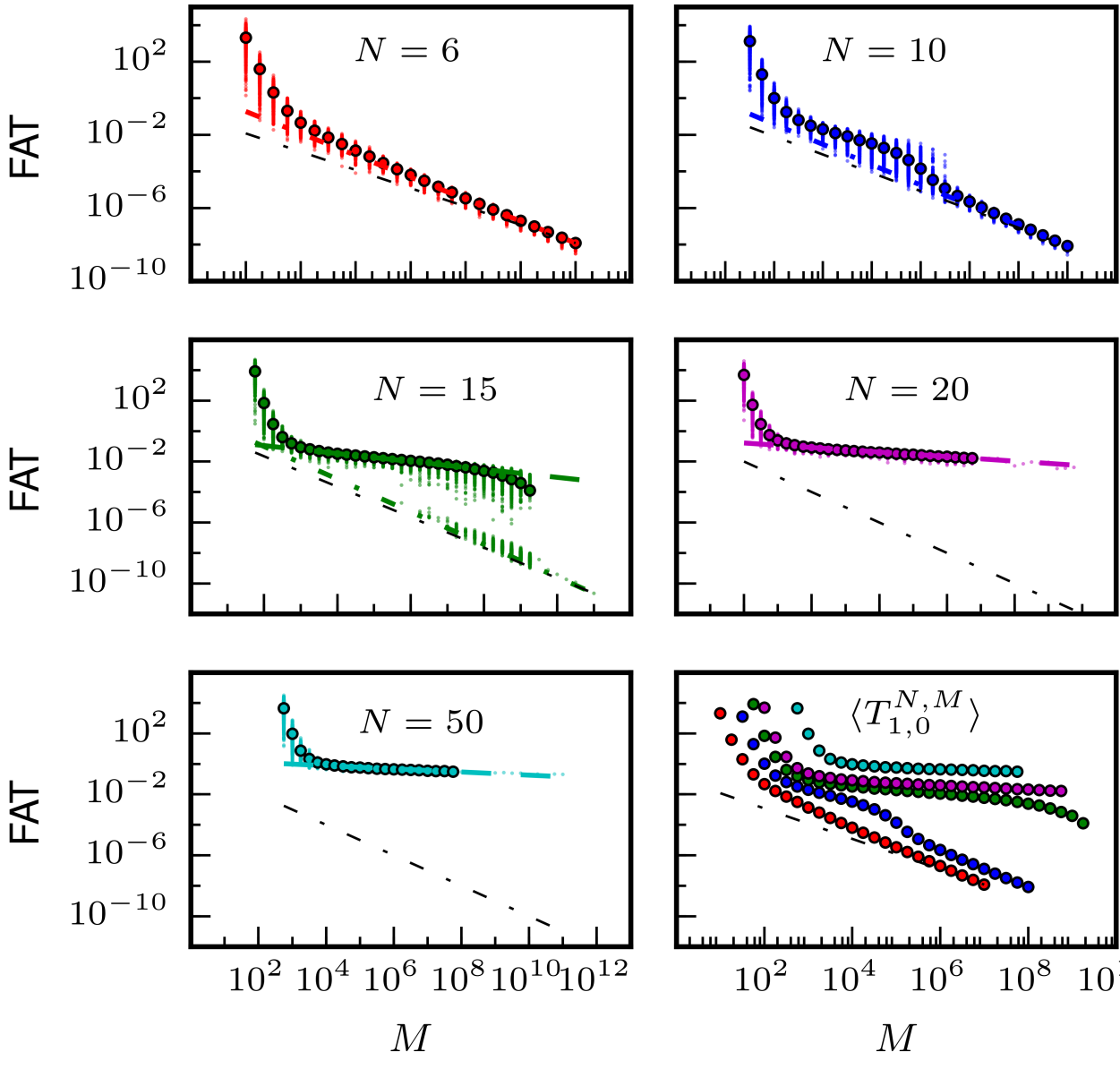
- Non-monotonous w.r.t reaction rate
- Bimodal for 'small' fragmentation rate



N=10, $p_1 = 0.5$, $p_k = 1$ and $q_k \equiv q$ for $k \geqslant 2$,.

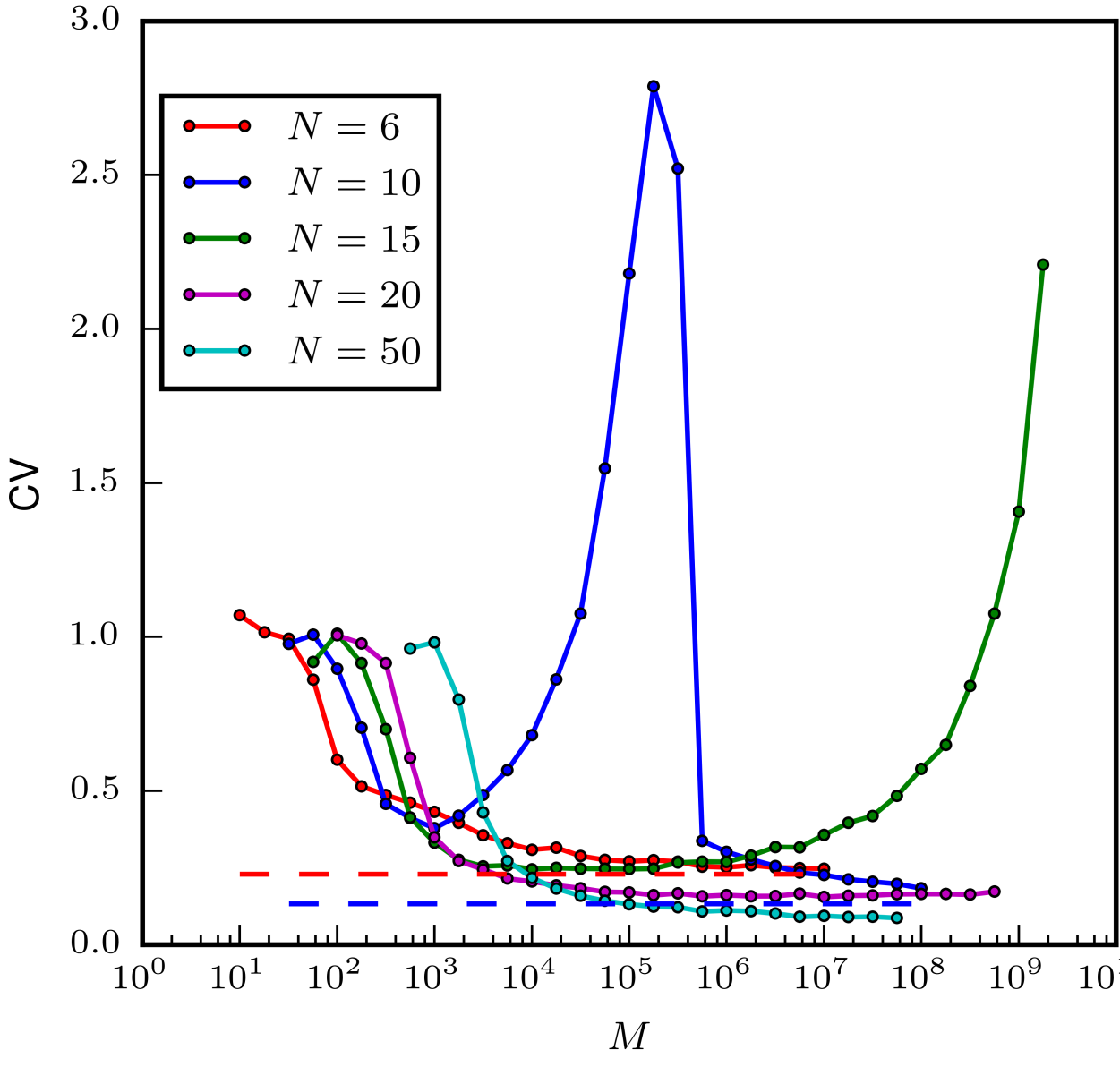
- 'Weak' dependency w.r.t. total monomer number M

$p_1 = 0.5$, $p_k = 1$ and $q_k \equiv 100$ for $k \geqslant 2$,.

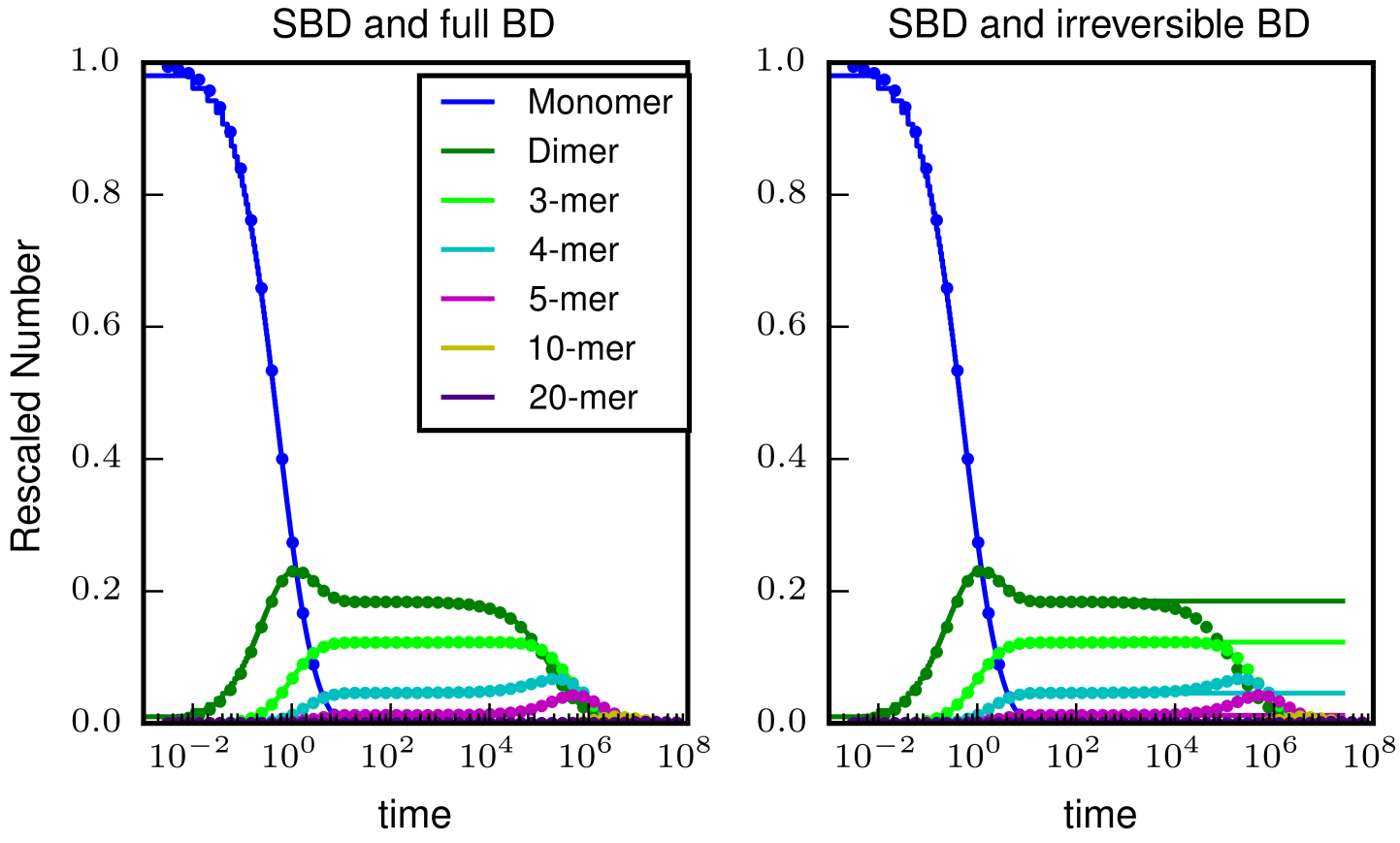


- Normalized variance non-monotonous.
- Normalized variance non zero pour $M \to \infty$.

$p_1 = 0.5$, $p_k = 1$ and $q_k \equiv 100$ for $k \geqslant 2$,.



- For $M \to \infty$: deterministic trajectory.
- Metastable behavior : 'pure-aggregation'.
- Medium-large polymer formed only after a longer time



$p_k = 1$ and $q_k \equiv 1$ for $k \geqslant 2$, $M = 10^5$ (we plot $M^{-1} C_k(tM^{-1})$).
Detailed Analysis : cf [Wattis, J. Phys. A : Math., 35 (2002)]

Outline

Amyloid diseases and Becker-Döring model

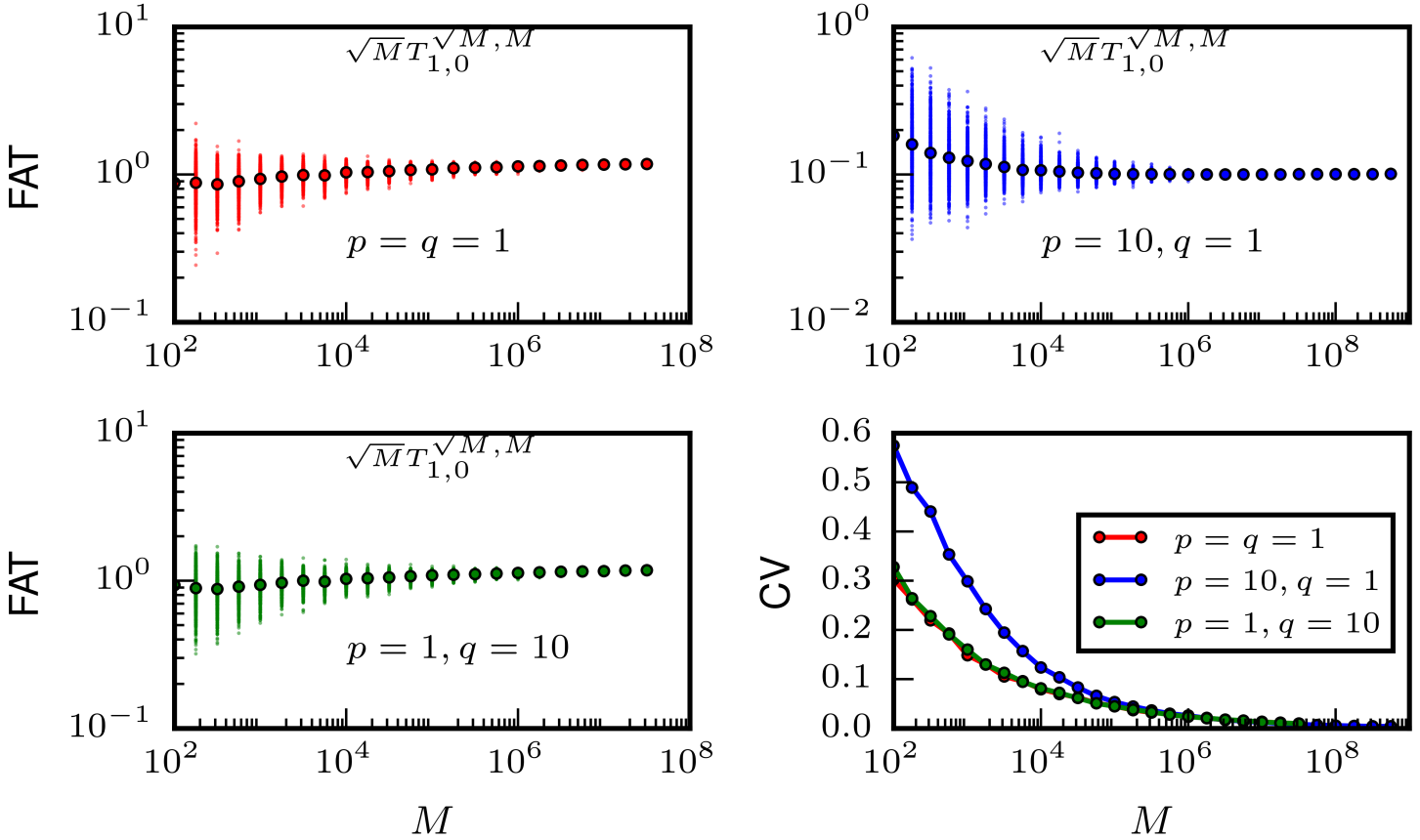
Numerical results

Coarse-graining

Open problems (for me)

Large nucleus $N \sim \sqrt{M}$

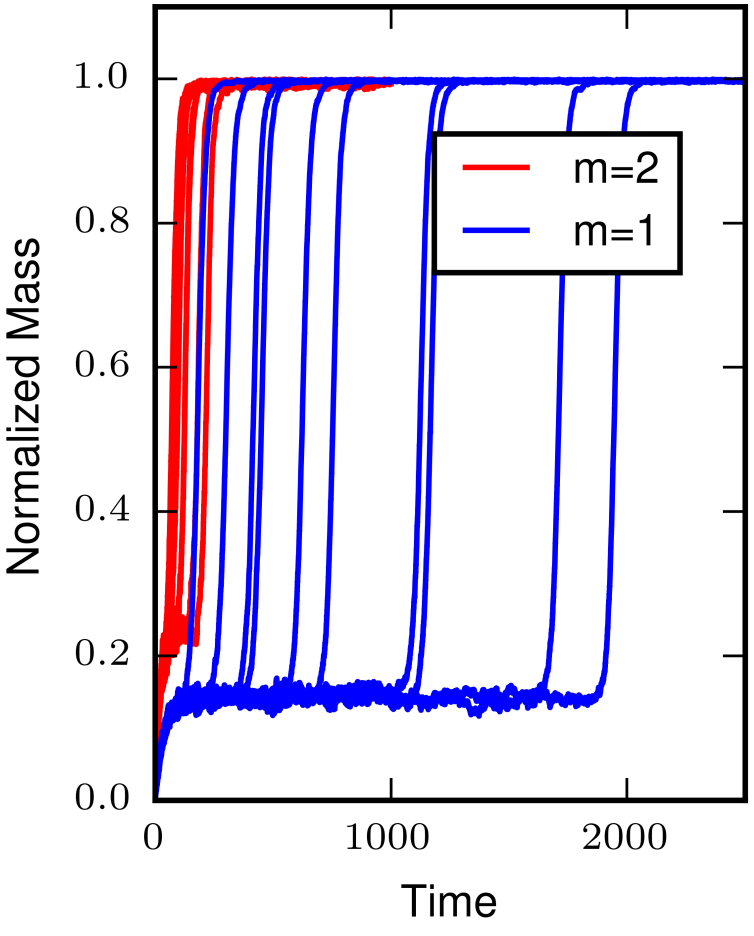
- First case $(p(0)m > q(0))$: Convergence towards a deterministic value.

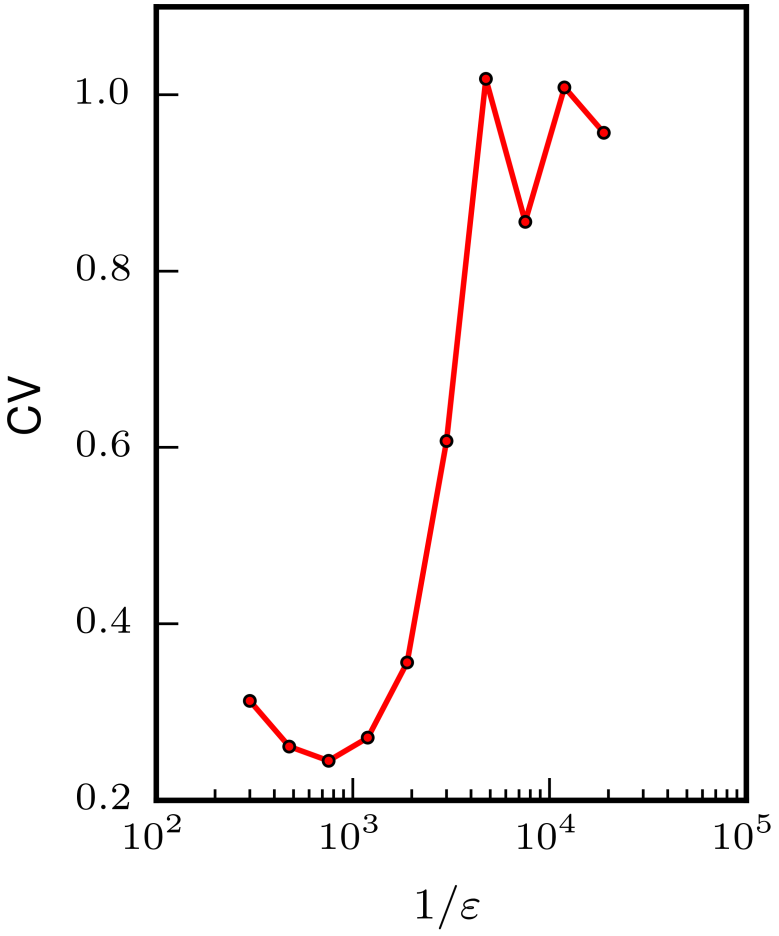


- case A
- case B

Large nucleus $N \sim \sqrt{M}$

- Second case ($p(0)M < q(0)$) : Exponentially large time and 'translated' trajectory.





Outline

Amyloid diseases and Becker-Döring model

Numerical results

Coarse-graining

Open problems (for me)

Quantifying the large deviation in the SBD model

A much simpler version of this model consider that a **single** aggregate may be formed at a time :

$$k \underset{q_{k+1}}{\overset{p_k(m-k\varepsilon)}{\rightleftharpoons}} k+1\,,$$

which converges (with time rescaling) to

$$\frac{dx}{dt} = p(x)(m-x) - q(x)$$

FPT are given by

$$\mathbb{E}\big[T^N_{1,0}\big] = \sum_{i=1}^{N-1}\sum_{j=1}^{i} \frac{\prod_{k=j+1}^{i} q_k}{\prod_{k=j}^{i} p_k(m-\varepsilon k)}\,.$$

which converges (with time rescaling) to

$$\int_0^1 \int_0^y . \exp\left[\varepsilon^{-1}\int_z^y \ln\left(\frac{q(x)}{p(x)(m-x)}\right) dx\right].$$

Quantifying the large deviation in the SBD model

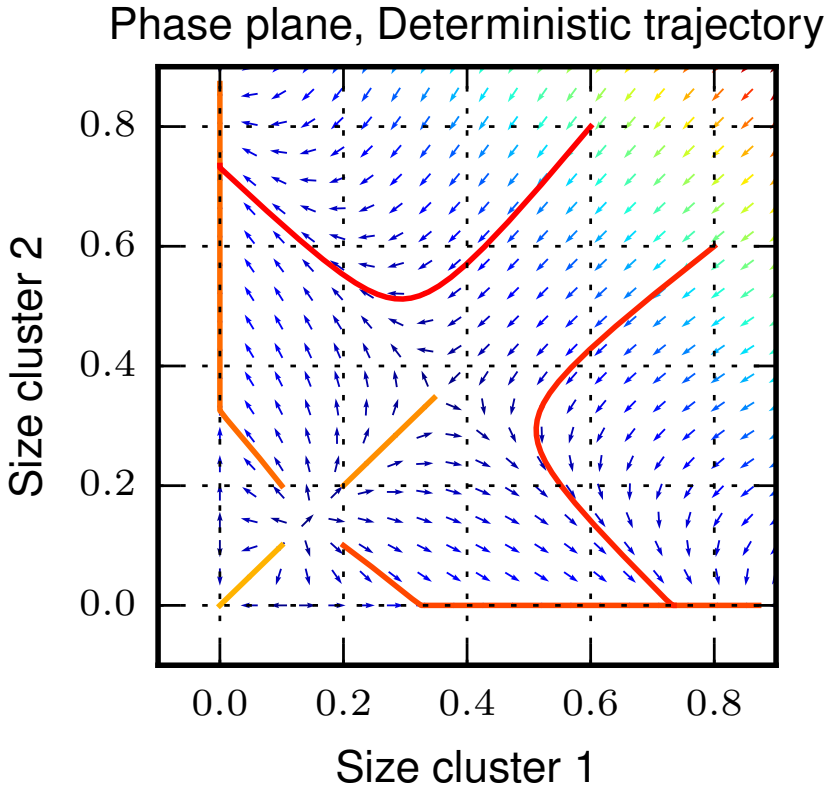
Can we perform similar calculations with n clusters ?

$$(k_0, k_1) \underset{q_{k_0+1}}{\overset{p_{k_0}(m-(k_0+k_1)\varepsilon)}{\rightleftharpoons}} (k_0+1, k_1),$$

$$(k_0, k_1) \underset{q_{k_1+1}}{\overset{p_{k_1}(m-(k_0+k_1)\varepsilon)}{\rightleftharpoons}} (k_0, k_1+1),$$

which converges (with time rescaling) to

$$\frac{dx}{dt} = p(x)(m-x-y) - q(x)$$
$$\frac{dy}{dt} = p(y)(m-x-y) - q(y)$$



Quantifying the large deviation in the SBD model

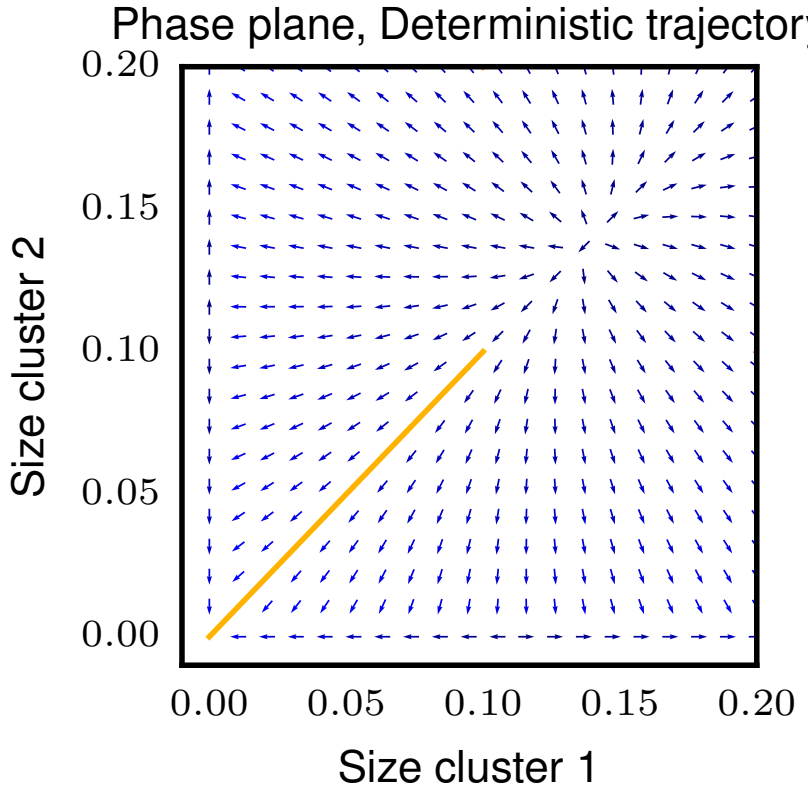
Can we perform similar calculations with n clusters ?

$$(k_0, k_1) \underset{q_{k_0+1}}{\overset{p_{k_0}(m-(k_0+k_1)\varepsilon)}{\rightleftharpoons}} (k_0+1, k_1),$$

$$(k_0, k_1) \underset{q_{k_1+1}}{\overset{p_{k_1}(m-(k_0+k_1)\varepsilon)}{\rightleftharpoons}} (k_0, k_1+1),$$

which converges (with time rescaling) to

$$\frac{dx}{dt} = p(x)(m-x-y) - q(x)$$
$$\frac{dy}{dt} = p(y)(m-x-y) - q(y)$$



Quantifying the large deviation in the SBD model

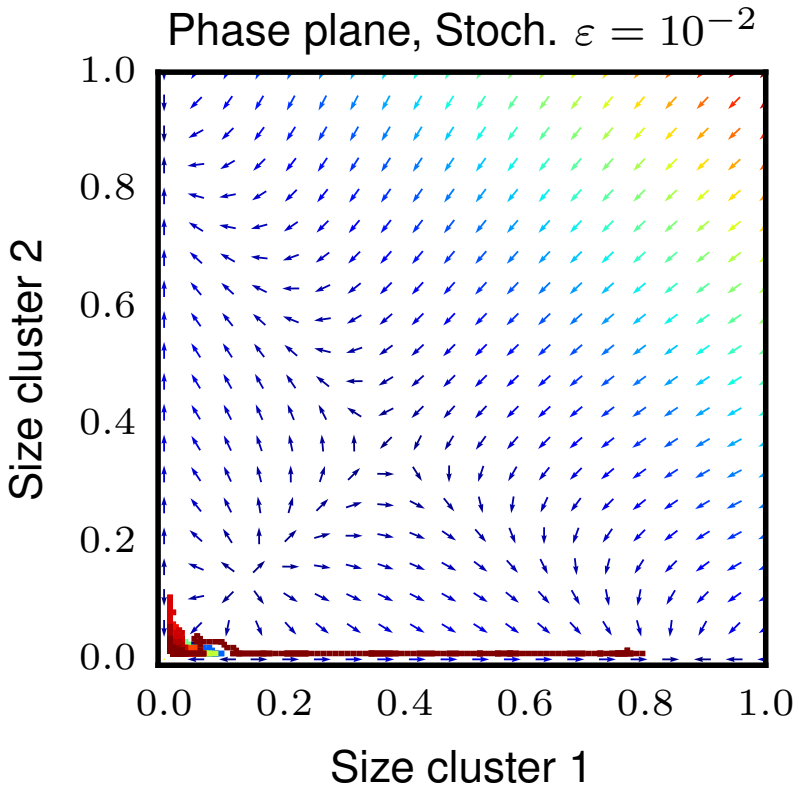
Can we perform similar calculations with n clusters ?

$$(k_0, k_1) \underset{q_{k_0+1}}{\overset{p_{k_0}(m-(k_0+k_1)\varepsilon)}{\rightleftharpoons}} (k_0+1, k_1),$$

$$(k_0, k_1) \underset{q_{k_1+1}}{\overset{p_{k_1}(m-(k_0+k_1)\varepsilon)}{\rightleftharpoons}} (k_0, k_1+1),$$

which converges (with time rescaling) to

$$\frac{dx}{dt} = p(x)(m-x-y) - q(x)$$
$$\frac{dy}{dt} = p(y)(m-x-y) - q(y)$$



Quantifying the large deviation in the SBD model

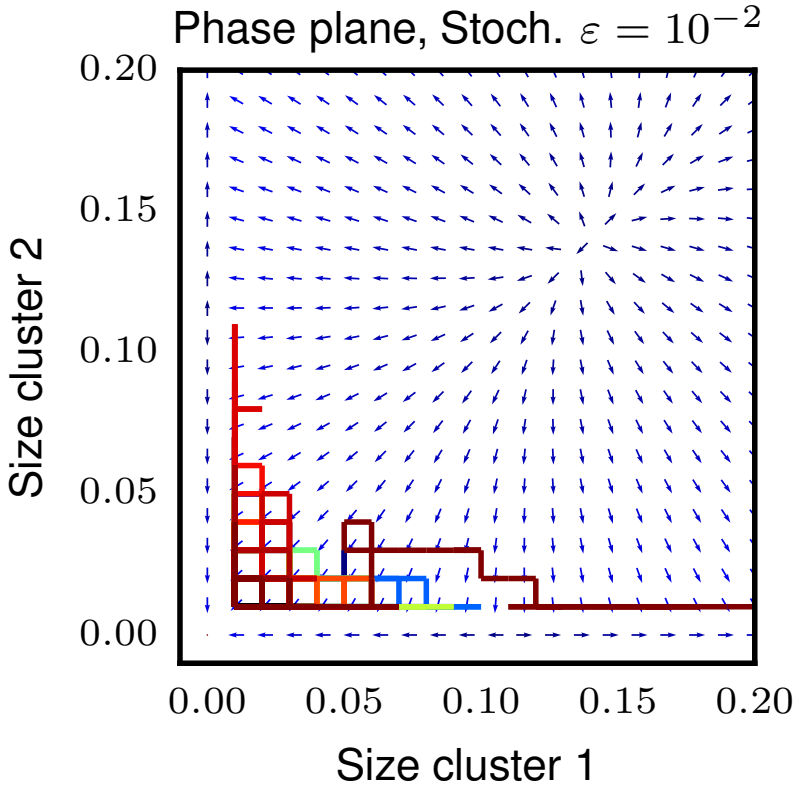
Can we perform similar calculations with n clusters ?

$$(k_0, k_1) \underset{q_{k_0+1}}{\overset{p_{k_0}(m-(k_0+k_1)\varepsilon)}{\rightleftharpoons}} (k_0+1, k_1),$$

$$(k_0, k_1) \underset{q_{k_1+1}}{\overset{p_{k_1}(m-(k_0+k_1)\varepsilon)}{\rightleftharpoons}} (k_0, k_1+1),$$

which converges (with time rescaling) to

$$\frac{dx}{dt} = p(x)(m-x-y) - q(x)$$
$$\frac{dy}{dt} = p(y)(m-x-y) - q(y)$$



Quantifying the large deviation in the SBD model

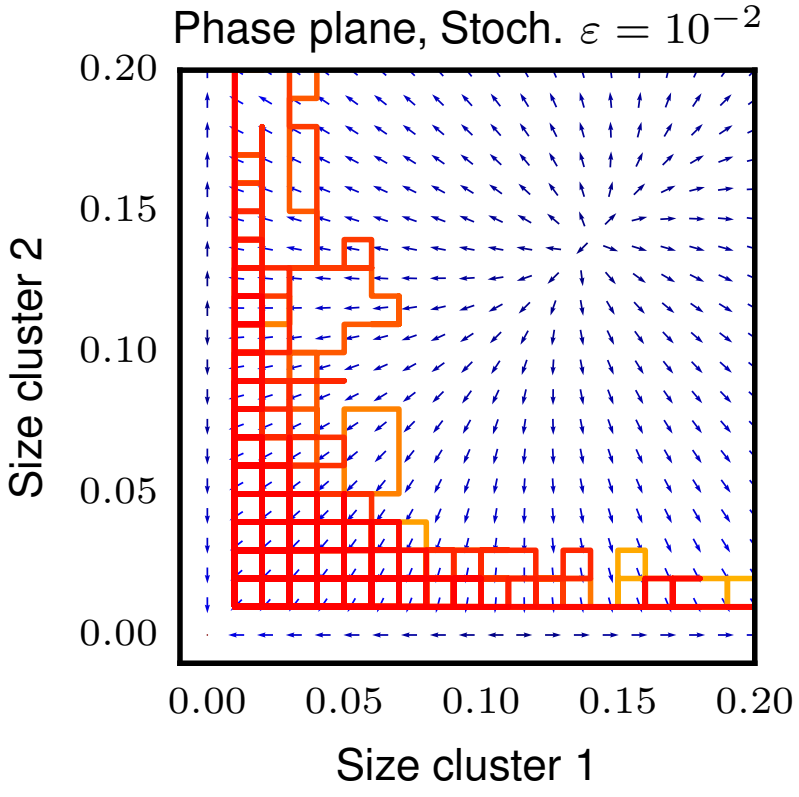
Can we perform similar calculations with n clusters ?

$$(k_0, k_1) \underset{q_{k_0+1}}{\overset{p_{k_0}(m-(k_0+k_1)\varepsilon)}{\rightleftharpoons}} (k_0+1, k_1),$$

$$(k_0, k_1) \underset{q_{k_1+1}}{\overset{p_{k_1}(m-(k_0+k_1)\varepsilon)}{\rightleftharpoons}} (k_0, k_1+1),$$

which converges (with time rescaling) to

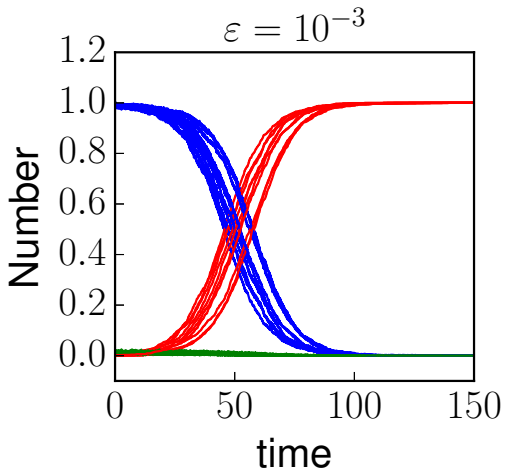
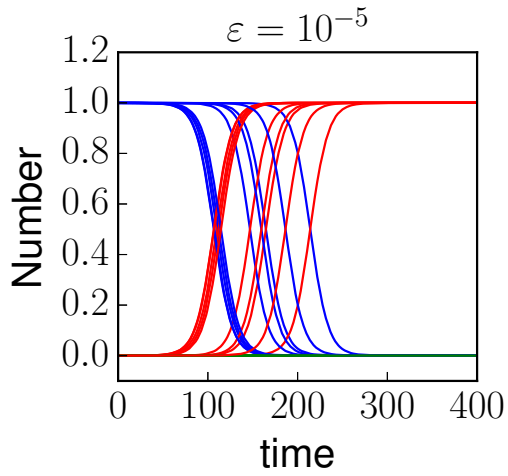
$$\frac{dx}{dt} = p(x)(m-x-y) - q(x)$$
$$\frac{dy}{dt} = p(y)(m-x-y) - q(y)$$



- **Extension to spatial models (diffusion)** ?
- **Data fitting with** $10^{10} - 10^{15}$ **proteins** ?

Toy model with time-scale separation

$$\begin{cases} X & \underset{\gamma^*/\varepsilon^r}{\overset{\gamma}{\rightleftharpoons}} & X^*\,, \\ X^* + X^* & \xrightarrow{\varepsilon\alpha} & 2Y\,, \\ X + Y & \xrightarrow{\varepsilon\beta} & 2Y\,, \end{cases}$$

with initial condition

$$Z^\varepsilon(0) = (\lfloor x_0/\varepsilon \rfloor, 0, 0)\,.$$