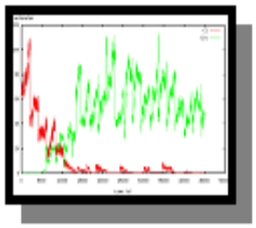


Excursions in Stochastic Dynamics of Complex Biological Systems

Sergey Plyasunov

SPlyasunov@lbl.gov

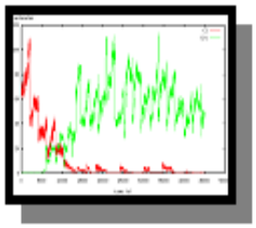
August 31, 2004

Abstract

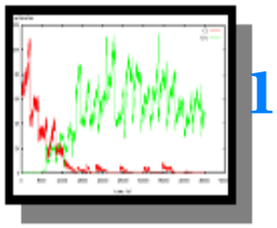
Talk outlines some recent developments in a field of stochastic chemical kinetics and its applications to the models of biological systems.

Topics:

1. New models of transcriptional regulation in λ-phage system,
2. Robustness of lysogenic state
3. Time scales separation in biochemical networks, rare events;
4. Complexity reduction in models with separation of time scale,
5. Model uncertainties and stochastic simulation

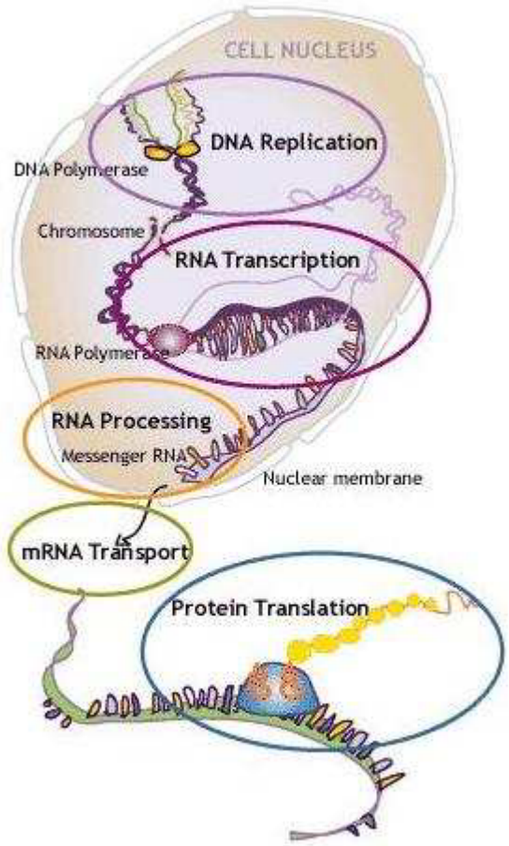


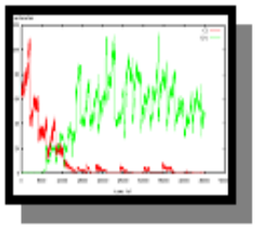
Contents



1. Introduction: Bio-chemical Networks inside the Cell

- **Metabolic (energy, synthesis)**
- **Regulatory (infromation processing : control of gene expression, sensory input signals processing)**



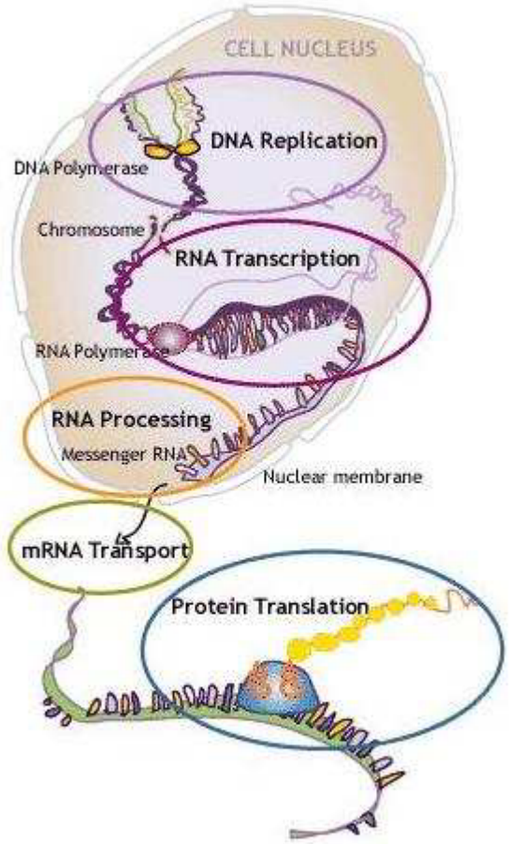


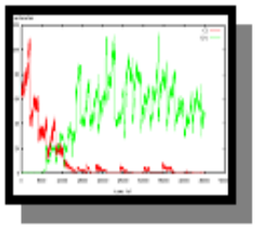
Multiscale Systems

- **System of many interacting components**
- **Multiple Levels of Organization:**
  - **Molecular $\Longleftrightarrow$ Cellular $\Longleftrightarrow$ Population**
  - **Multiple spatial and temporal scales:**

    10^{-8} m 10^{-6} m 10^{-3} m

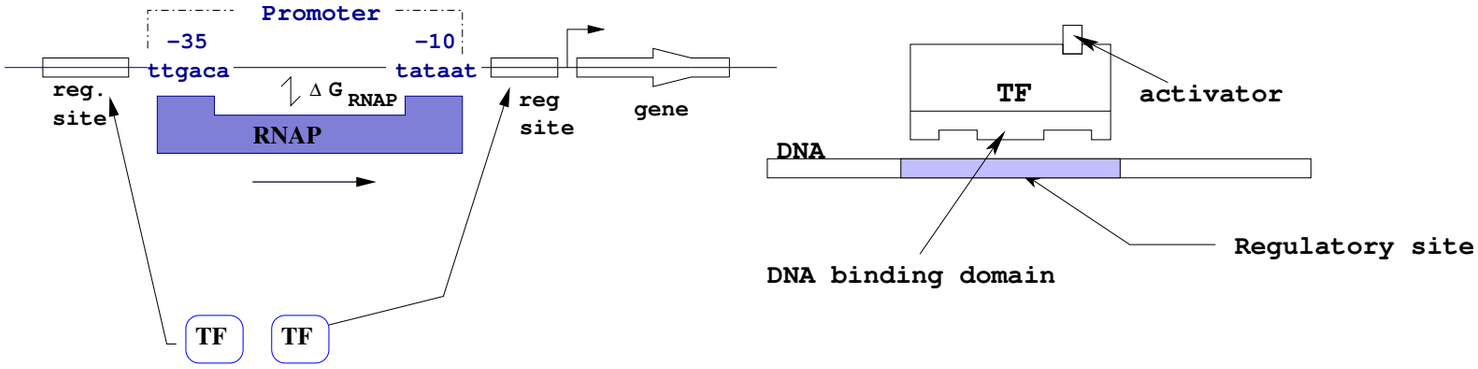
    10^{-7} sec 1 sec 1 hour-1 year
- **Non-Equilibrium Steady State: Conventional Thermodynamics is not applicable**
- **Stochasticity at a basal level of gene expression $\Rightarrow$ Phenotypic variability**





Example:Many-body Interactions and Regulation of Gene Expression

- Gene expression is controlled by binding of transcription factor (TF) proteins to the regulatory sites on DNA, blocking/pushing RNAP from/to the gene.



"Key-lock" principle

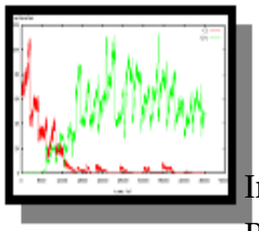
- Specificity and strength varies from promoter to promoter and from TF to TF

Promoter	-35 Region	-10 Region
Consensus	ttgaca	tataat
trp operon	ttgaca	ttaact
rec A	ttgata	tataat

TF protein	reg. sequence
CAP	...aagtga tagctgtc... ...tttgttacctgcctc...
LacI	...aattgtgagcgagtaacaatt... ...aaatgtgagcgagtaacaacc... ...ggcagtgagcgcaacgcaatt...

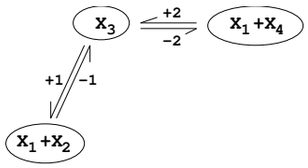
- Activity of the promoter of the $gene_i$ depends on the complicated pattern of regulatory elemens affected by transcription factors $\{\mathbf{TF}\}$:

$$gene_i = F_i(\mathrm{TF}_1, \mathrm{TF}_2, \ldots, \mathrm{TF}_N)$$



Networks of Interacting Species

Integer vector X is a state vector of species numbers (number of proteins , free regulatory sites, RNAP,etc):



$$\sum_{i=1}^{S} \nu_{ir}^{+} X_i \underset{k_{-r}}{\overset{k_{+r}}{\rightleftharpoons}} \sum_{i=1}^{S} \nu_{ir}^{-} X_i \qquad (1)$$

Reaction at each channel changes the state of the system by the vector $\boldsymbol{\nu}_r = (\nu_{1r}^{-} - \nu_{1r}^{+}, \ldots, \nu_{Sr}^{-} - \nu_{Sr}^{+})$:

$$\mathbf{X} \to \mathbf{X} + \boldsymbol{\nu}_r, \;\; \boldsymbol{\nu}_r = \boldsymbol{\nu}_r^{-} - \boldsymbol{\nu}_r^{+}$$

Different classes of problems have different mathematical descriptions:

- Stochastic effects:

$$\frac{\partial P(\mathbf{X},t)}{\partial t} = \sum_r a_r(\mathbf{X} - \boldsymbol{\nu}_r) P(\mathbf{X} - \boldsymbol{\nu}_r, t) - P(\mathbf{X},t) \sum_r a_r(\mathbf{X}), \qquad (2)$$

$$a_{\pm r}(\mathbf{X}) = k_r V \prod_{i=1}^{S} \frac{X_i!}{(X_i - \nu_{ir}^{\pm})! V^{\nu_{ir}^{\pm}}} \qquad (3)$$

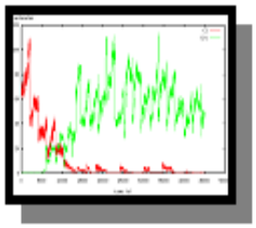
  Eqn. (2) can be solved mostly only by K(inetic)M(onte)C(arlo)(aka Gillespie Algorithm [Bortz et al., 1975, Gillespie, 1977]).

- Add diffusion effects...

- Deterministic mass action kinetics (systems of non-linear/stiff ODE's):

$$\frac{d\mathbf{X}}{dt} = \sum_{r=1}^{R} \boldsymbol{\nu}_r a_r(\mathbf{X}), \qquad (4)$$

$$a_{\pm r}(\mathbf{X}) = k_{\pm r} V \prod_i \frac{X_i^{\nu_{ir}^{\pm}}}{V^{\nu_{ir}^{\pm}}} + O\left(\frac{|\boldsymbol{\nu}_r^{\pm}|}{V}\right) \qquad (5)$$

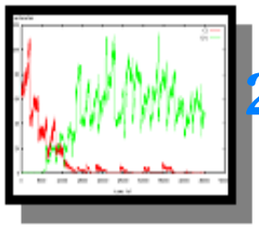


Focus on Stochastic Effects: Why this research is potentially relevant?

The solution of the DNA structure was the results of the integrated theoretical modeling and experimental techniques. Since that time, theory /computation has played probably minor role in biological discoveries.
To overcome this situation traditional simulation techniques must be taken to the new level.

Evangelism:

- General questions about modeling of complex systems: find a biologically/biophysically relevant representation.
- Mathematically rigorous and physically consistent, stochastic algorithms are computationally expensive$\Rightarrow$ corase-graining methods
- How to deal with uncertainties of the model in stochastic/probabilistic setting?
- Problems which are hard to solve with traditional Monte Carlo methods: Large deviations, rare events problems, robustness , stability.



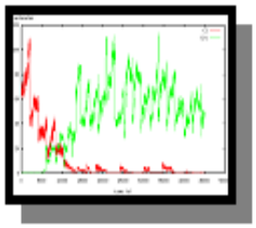
2. Models of transcriptional regulation in bacteriophage λ

S. Plyasunov, R. E. Osterhout, J.W.Little and A.P.Arkin

Abstract

We develop a stochastic model of the bacteriophage-λ lysis/lysogeny switch, taking into account recent experimental evidence demonstrating enhanced cooperativity between the left and right operator regions O_{R} and O_{L} . Model parameters are estimated from available experimental data.

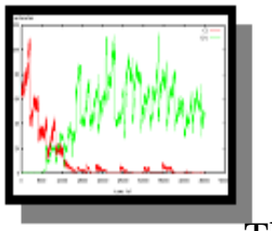
Long distance transcriptional regulation between O_{R} and O_{L} complexes in λ-infected *E. coli* is necessary for efficient repression of λ repressor CI , but its effect on lysogenic stability is unclear. We present a stochastic kinetic model that includes a rigorous mathematical treatment of DNA looping. We use this model to predict the stability of the lysogenic state in wild type and mutant phage, and to investigate the influence of DNA cyclization on the stability of wild type cells and J.W.Little's $O_{\mathrm{R}121}$, $O_{\mathrm{R}323}$ mutants (termed here and after 121 and 323-mutants) [Little et al., 1999].
Keywords: gene regulation,stability, robustness, phage-λ, lysogeny, lysis, DNA cyclization, stochastic model.



Basic facts on *E. coli* λ system

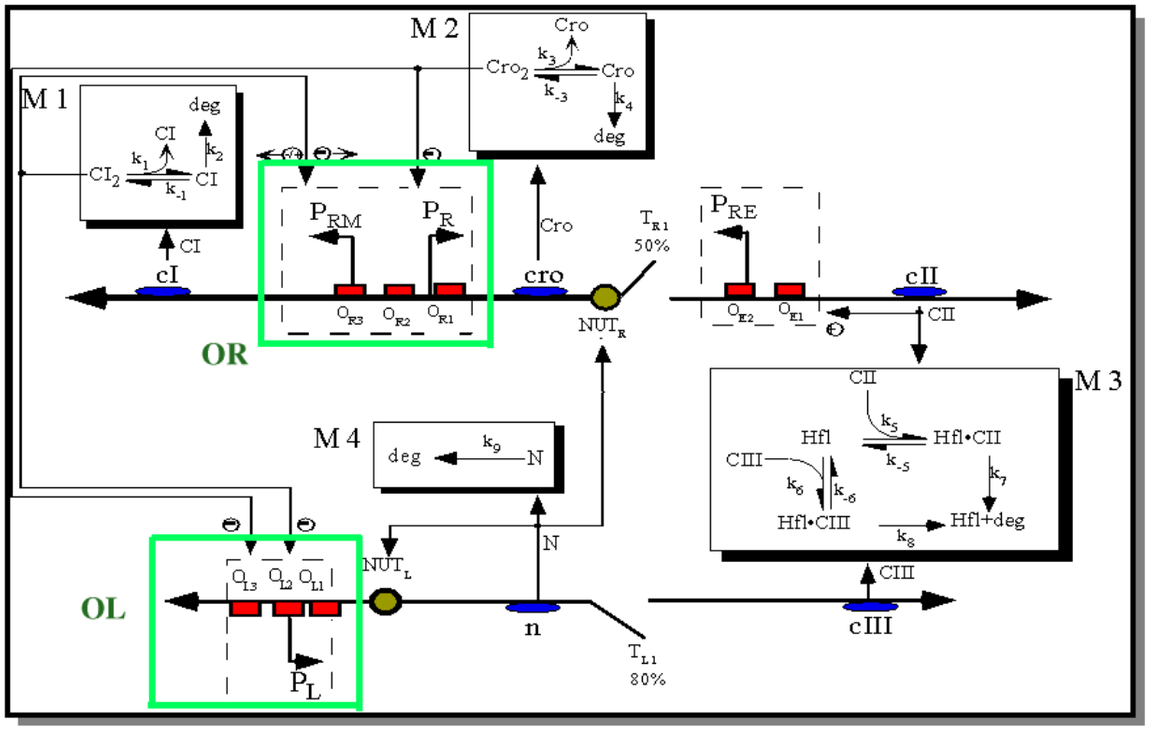
- The genome of *E. coli* consists of a single DNA molecule of 4.6×10^6 bp (length 1.5 mm). It codes for 4226 proteins and number of RNAs.
- Regulatory patterns:
  The "genetic switch" of phage lambda allows a choice between two patterns of gene expression. This switch involves the interplay between two regulatory proteins, CI and Cro which bind to a complex regulatory region termed O_{R} .
- Cooperative interactions of protein binding is important
- These proteins stabilize two mutually exclusive patterns of gene expression. The regulatory circuitry that controls these two alternatives is understood in considerable detail.
- One of the patterns of gene expression (the "lysogenic" state) can be switched to the other (the "lytic" state) by treatments that damage DNA and induce the SOS response. This "genetic switch" has threshold behavior–that is, it occurs above a threshold level of damage, but not below that threshold.
- Relatively well known system

Event	gene expressed	Comments
Initial infection	*cro, N*	Only *N, cro* are synthesized until decision point is reached
Lytic pathway	*cro, N, Q*, late genes	*cro* predominates, *N,Q* are anti-terminators
Lysogenuc pathway	*cI, cII, cIII, int*	*cII,cIII* collaborate to establish *cI* synthesis ; after genome integration, only *cI* is expressed during the maintenance of lysogeny

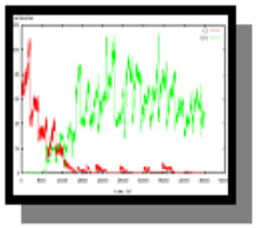


Overview of the Gene Expression Patterns and Genetic Switch

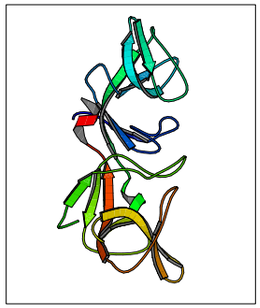
There are two similar complexes in λ-system: O_{R} and O_{L} with similar energetics: O_{R} produces *cI* and *cro*, O_{L} produces transcript of *N*.



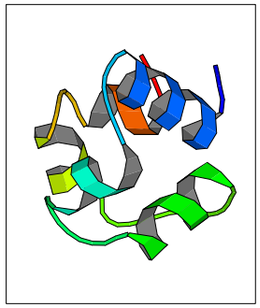
Genetic Switch [Arkin et al., 1998, Ptashne, 1992]



O_{R} and O_{L} complexes

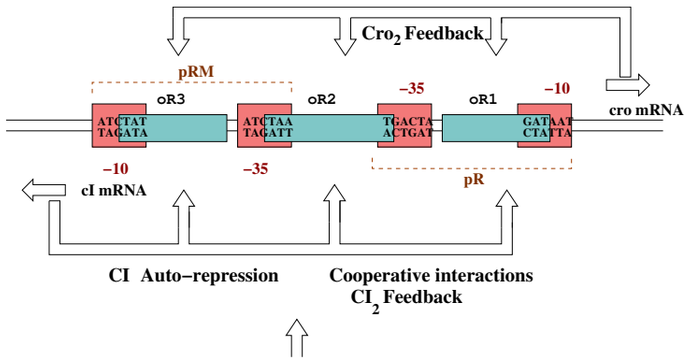


CI protein(λ-repressor)

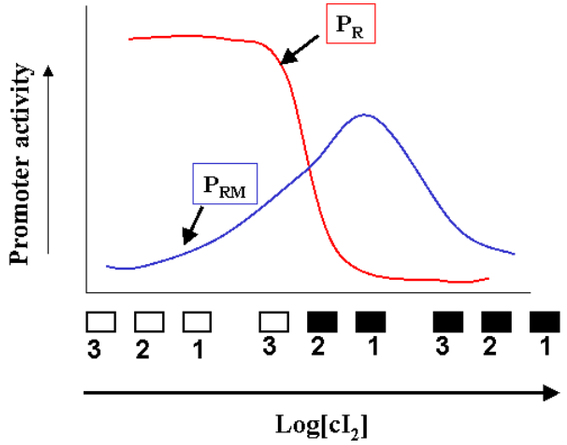


Cro protein

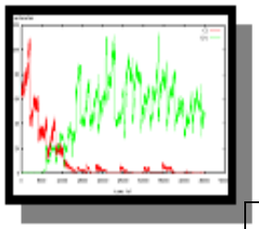
- Only dimers are used for regulation
- Differential binding affinities: Cro_2: $\mathrm{OR}_3 > \mathrm{OR}_2 \approx \mathrm{OR}_1$
  CI_2: $\mathrm{OR}_1 > \mathrm{OR}_2 > \mathrm{OR}_3$
- Both Cro_2 and CI_2 bind to the DNA with helix-turn-helix motif.
- CI has two subunits: cooperativity of interactions is important. Cooperativity of Cro_2 is not important.
- CI can be effectively cleaved by recA protease



Geometric picture of O_{R} sites and pRM/pR promoters inside the O_{R} ;O_{L} is separated by $2.8 \times 10^3 bp$



Promoter activities of O_{R} complex



Kinetic and energetic parameters of O_R complex

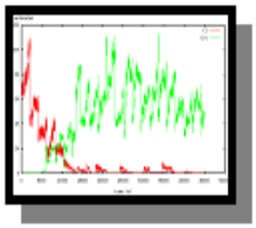
Parameter	Value	Meaning
k_R	$0.013s^{-1}$	pR activity rate
k^u_{RM}	$0.001s^{-1}$	pRM activity rate (basal)
k_{RM}	$0.011s^{-1}$	with CI_2 bound
k_{cro}	$0.00059s^{-1}$	decay/dilution rate
k_{cI}	$0.00034s^{-1}$	decay/dilution rate
RT	0.617 kcal/mol	temperature
$\Delta G_{cI,1,2,3}$	$-12.5, -10.5, -9.5$	independent bindings
$\Delta G_{cro,1,2,3}$	$-12.0, -10.8, -13.4$	independent bindings
$\Delta G_{rnap,32}$	-11.5 kcal/mol	RNAP binding on pRM
$\Delta G_{rnap,1}$	-12.5 kcal/mol	RNAP binding on pR
$\delta G_{cI,12}$	-2.7 kcal/mol	cI_2 cooperativity
$\delta G_{cI,23}$	-2.9 kcal/mol	cI_2 cooperativity
$\delta G_{cro,12}$	-1.0 kcal/mol	Cro_2 cooperativity
$\delta G_{cro,23}$	-0.6 kcal/mol	Cro_2 cooperativity
ΔG_{cro}	-7.0 kcal/mol	Cro dimerization
ΔG_{cI}	-11.1 kcal/mol	CI dimerization
[RNAP]	30nM	RNAp concentration
V	$1.5 \times 10^{-15} l$	*E. coli* volume

- Given the cooperativity and individual binding energies CI_2, Cro_2, and RNAP $\Delta G(\mathbf{s})$ can be calculated for every configuration $\mathbf{s}$ of each different binding state.

Sources: [Shea and Ackers, 1985, Aurell and Sneppen, 2002, Darling et al., 2000]

- CI_2 can block its own production at high concentration
- RNAP forms open complex faster with CI_2 bound at O_R2
- Dimerization reaction: $X_2 \underset{k_{-1}}{\overset{k_{+1}}{\rightleftharpoons}} 2X$, $K_D = k_{+1}/k_{-1}$:

$$[X_2] = \frac{1}{2}[X_{tot}] - \frac{K_D}{8}\left(\sqrt{1 + \frac{8[X_{tot}]}{K_D}} - 1\right) \tag{6}$$



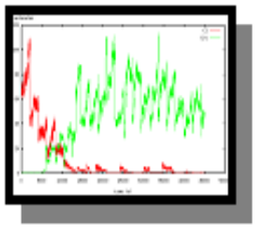
StatMech of transcriptional regulation : QuasiEquilibrium Model

- There are 40 *experimentally distinguishable* states s at O_{R} with CI_2,Cro_2 and RNAP bound in different order.

Label 0 corresponds to the empty site, label 1 corresponds to the CI_2 repressor dimer; 2 corresponds to Cro_2, 3 corresponds to RNAP[Darling et al., 2000]. Total number of protein of each type (monomer units): $N_{\mathrm{Cro}} = 2\mathrm{Cro}_2 + \mathrm{Cro}$,$N_{\mathrm{CI}} = 2\mathrm{CI}_2 + \mathrm{CI}$.

state	O_{R_1}	O_{R_2}	O_{R_3}	$\Delta G(\mathrm{s})$ kcal/mol
0	0	0	0	0.0
1	1	0	0	-12.5
2	0	1	0	-10.5
3	0	0	1	-9.5
4	2	0	0	-12.0
5	0	2	0	-10.8
6	0	0	2	-13.4
7	0	0	3	-11.5
8	3	0	0	-12.5
9	1	1	0	-25.7
10	1	0	1	-22.0
11	0	1	1	-22.9
12	2	2	0	-23.8
13	2	0	2	-25.4
14	0	2	2	-24.8
1515	3	0	3	-24.0
16	1	2	0	-23.3
17	2	1	0	-22.5
18	2	0	1	-21.5
19	1	0	2	-25.9

state	O_{R_1}	O_{R_2}	O_{R_3}	$\Delta G(\mathrm{s})$ kcal/mol
20	0	2	1	-20.3
21	0	1	2	-23.9
22	3	0	1	-22.0
23	0	1	3	-22.0
24	1	0	3	-24.0
25	3	0	2	-25.9
26	0	2	3	-22.3
27	2	0	3	-23.5
32	2	1	1	-34.9
33	2	2	1	-33.3
34	2	1	2	-35.9
35	1	2	2	-37.3
36	1	1	3	-37.2
37	2	2	3	-35.3
38	1	2	3	-34.8
39	2	1	3	-34.0



- Grand-canonical partition function (following [Shea and Ackers, 1985, Aurell and Sneppen, 2002]):

$$Z = \sum_{\mathbf{s}} \mathrm{e}^{-\frac{\Delta G(\mathbf{s})}{RT}} \left(\frac{\mathrm{Cro}_2}{V}\right)^{\sum_i s_{i,1}} \left(\frac{\mathrm{CI}_2}{V}\right)^{\sum_i s_{i,2}} \left(\frac{RNAP}{V}\right)^{\sum_i s_{i,3}}, \tag{7}$$

$$p(\mathbf{s}) = \frac{1}{Z}\mathrm{e}^{-\beta\Delta G(\mathbf{s})} \left(\frac{\mathrm{Cro}_2}{V}\right)^{\sum_i s_{i,1}} \left(\frac{\mathrm{CI}_2}{V}\right)^{\sum_i s_{i,2}} \left(\frac{RNAP}{V}\right)^{\sum_i s_{i,3}} \tag{8}$$

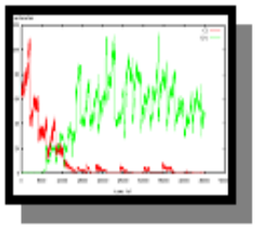
- Activities of promoters (P_{RM} ,P_{R}) are weighted combinations of RNAP-open complex formation rates: CI_2 activity comes from the states where $RNAP$ bound to O_{R_3}:

$$f_1 = f_{\mathrm{CI}}(\mathrm{CI}_2, \mathrm{Cro}_2) = k_{RM} N_{RM}(p_{23} + p_{36} + p_{39}) + \tag{9a}$$
$$+ k^u_{RM} S_{RM}(p_7 + p_{15} + p_{24} + p_{26} + p_{27} + p_{37} + p_{38}) \tag{9b}$$

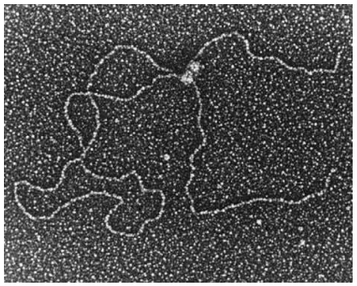
  Cro_2 activity comes from the states where $RNAP$ bound to O_{R_1}:

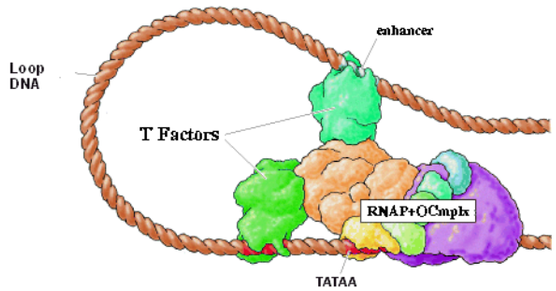
$$f_2 = f_{\mathrm{Cro}}(\mathrm{CI}_2, \mathrm{Cro}_2) = k_R N_R(p_8 + p_{15} + p_{22} + p_{25}) \tag{9c}$$

- "Thermodynamic equilibrium assumption" does not mean that the probabilities $p(\mathbf{s})$ remain constant in time
- How to account for delays due to transcription/translation: $\mathrm{CI}_2(t) \rightarrow CI_2(t-\tau)$?
- In addition there are 30 independent states at O_{L} . For the future: What if they (O_{R} and O_{L}) can interact?



Role of the DNA in Long-Range Interactions

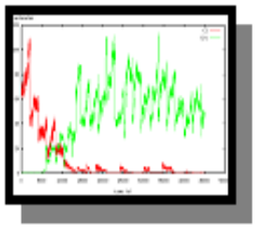




- DNA is a flexible polymer that can adopt a variety of conformations different both in its secondary structure and tertiary structure as determined by intrinsic DNA curvature and DNA super-coiling
- DNA-looping mechanisms are part of networks that regulate all aspects of DNA metabolism, including transcription, replication, and recombination

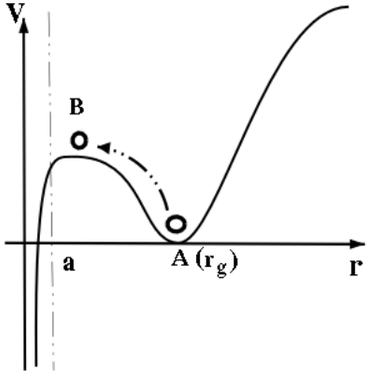
Systems with looping:

- Bacteria: *lac, ara, gal*, distance: $L \approx 100\ bp$
- Viruses: λ-system, distance: $L \approx 60\ bp$ [Ptashne, 1992], $L \approx 2.3 \times 10^3\ bp$ [Dodd et al., 2001] (slow)
- Eukaryots: transcription ($L \approx 5 \times 10^3\ bp$) mating type switching, $L \approx 100 \times 10^3\ bp$
- Multiple looping of DNA reduce the gyration radius $\rightarrow$ easy transfer into cells



DNA looping

- Decrease of the polymer entropy is compensated by the interaction between the segments of the dsDNA: $\Delta G = \Delta G_{TF} - T\Delta S_{loop}$. Polymer cyclization is a very hard computational problem (time scale separation(for $L \gg l_p$) ,non-Markovian process [Szabo et al., 1980, Sokolov, 2003]).
- Huge simplification: Markovian escape problem



Effective potential for the reaction coordinate r for the polymer of length L and Kuhn length l_p. D is the diffusion coefficient of the "monomer" with length l_p.

r is the end-to-end distance: $V(r,L) = -\beta^{-1}\ln[\underbrace{4\pi r^2 G(r,L)}_{\text{radial distr.}}]$

Coordinate r is driven by the white-noise over the barrier $A \to B$

$$\gamma\dot{r} = -\partial_r V(r,L) + \xi(t), \tag{10a}$$

$$\langle \xi(t)\xi(t')\rangle = 2D\delta(t-t'), \gamma^{-1} = \beta D$$

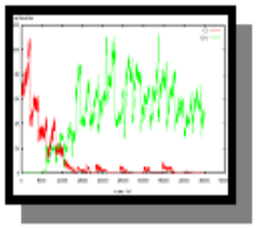
$$\tau_{Kr}^{-1} = \frac{\omega_A\omega_B}{2\pi\gamma}\exp(-\beta\Delta V_{AB}(L)), \tag{10b}$$

$$\omega_{A,B} = l_p^{-1}\sqrt{\partial_{rr}V(r,L)}_{r=A,B}$$

- Kramers escape time for the $G(0,L/l_p) \propto (L/l_p)^{-3/2}$, $L/l_p \gg 1$ [Rippe et al., 1995]:

$$\tau_{Kr} \approx \frac{l_p^2}{D}\left(\frac{l_p}{L}\right)^{\frac{3}{2}}, \tag{11}$$

$$\tau_{Kr} \approx 0.03 - 0.3sec$$



Role of the DNA in Long-Range Interactions

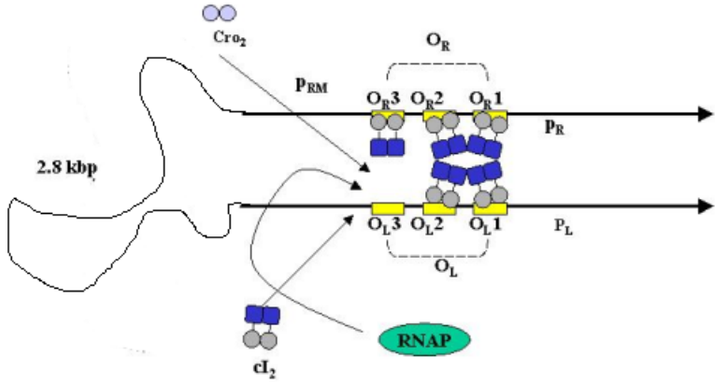
- Long-Range interaction between O_{R} and O_{L} could alter the gene regulation in λ
- Additional change in the Gibbs energy due to the loop formation:

$$\delta\Delta G = -\sum_{ij} \Delta G^{\mathrm{oct}}_{\mathrm{RL}} \underbrace{\sigma_{\mathrm{CI_2\ OR_i}}}_{0,1} \underbrace{\sigma_{\mathrm{CI_2\ OR_{i+1}}}}_{0,1} \underbrace{\sigma_{CI_2\ OR_j}}_{0,1} \underbrace{\sigma_{CI_2\ OR_{j+1}}}_{0,1}, \tag{12a}$$

$$\delta\Delta G = -\Delta G^{\mathrm{tet}}_{\mathrm{RL}} \left[\sigma_{\mathrm{CI_2\ OR_1}}\sigma_{\mathrm{CI_2\ OR_2}}\sigma_{\mathrm{CI_2\ OR_3}}\right]\left[\sigma_{CI_2\ OR_1}\sigma_{CI_2\ OR_2}\sigma_{CI_2\ OR_3}\right], \tag{12b}$$

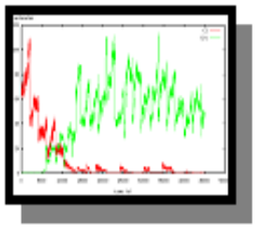
$$\Delta G^{\mathrm{oct}}_{\mathrm{RL}} = -0.5\mathrm{kcal/mol},$$

$$\Delta G^{\mathrm{tet}}_{\mathrm{RL}} = -3.0\mathrm{kcal/mol}$$

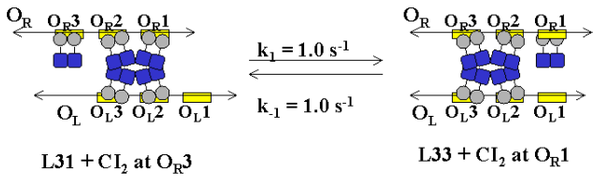


Facts

1. $\mathrm{CI_2}$ can effectively form octamers in solution [Bell and Lewis, 2001]
2. *Repression of* P_{R} *increased* $\times 4$ in the presence of O_{L}
3. *Promoter* P_{RM} *can be also repressed* $\times 1/2.5$ (need site $O_{\mathrm{L}3}$)

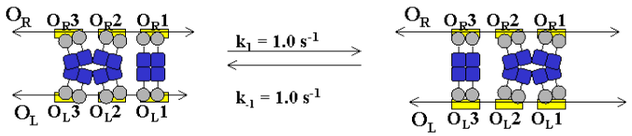


Possible rearrangements of states



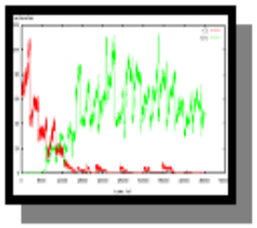
L31 + CI_2 at O_R3

L33 + CI_2 at O_R1



L33 +tetramer

L11 + tetramer



Model Development: Equation-less Modeling

- "On/Off" binding rates:

$$k_{on} = \frac{4\pi D\epsilon}{V}, \varepsilon - \text{target size} 10\,\text{nm}, \tag{13}$$

$$k_{on} \approx 0.1 - 0.05\ s^{-1} \text{ for } D = 5\mu^2 m/sec$$

$$k_{off} = k_{on} V^{\alpha} e^{\beta \Delta G}, \tag{14}$$

Regular non-cooperative binding/release of the transcription factor $\mathrm{X} = \mathrm{CI}_2, \mathrm{Cro}_2$ to the site $O_{\mathrm{R}i}(O_{\mathrm{L}i})$,$i = 1,2,3$ can be expressed as:

$$\mathrm{X} + \mathrm{O}_{\mathrm{Ri}} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \mathrm{XO}_{\mathrm{Ri}} \tag{15a}$$

$$\mathrm{X} + \mathrm{O}_{\mathrm{Li}} \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} \mathrm{XO}_{\mathrm{Li}} \tag{15b}$$

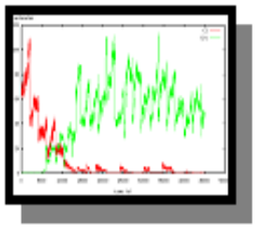
(15c)

Species $O_{\mathrm{R}i}$ and $O_{\mathrm{L}i}$ as well as bound complexes $\mathrm{XO}_{\mathrm{Ri}}$, $\mathrm{XO}_{\mathrm{Li}}$ are essentially binary.

- Cooperativity of binding:

$$\mathrm{X} + \mathrm{OR_i} + \overline{\mathrm{XOR_j}} \rightleftharpoons \mathrm{XOR_i} + \overline{\mathrm{XOR_j}} \tag{16a}$$

$$\mathrm{X} + \mathrm{OR_i} + \mathrm{XOR_j} \rightleftharpoons \mathrm{XOR_i} + \mathrm{XOR_j} \tag{16b}$$



Model Development: Equation-less Modeling (Cont'd)

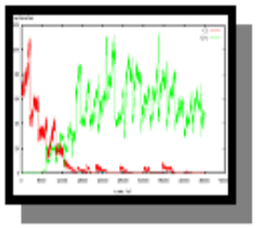
- Dimerization reactions $2X \rightleftharpoons X_2$ take place on the background
- Both dilution and degradation of proteins are accounted.
- Several topological sates of the dsDNA act as pseudospicies

$$L_{00} + O_{Ri} + X_2 \rightleftharpoons L_{00} + X_2O_{Ri} \tag{17}$$
$$L_{00} + \ldots \rightleftharpoons L_{11} + \ldots \tag{18}$$

- Unspecific binding of CI_2 and Cro_2 to the dsDNA is included via simple *projection*:

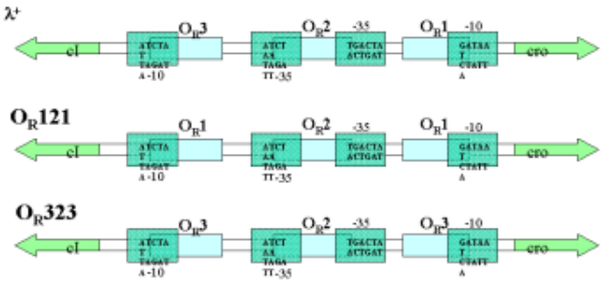
$$X_2(t) = \frac{X_2(t_-) + X_{2DNA}(t_-)}{1 + L_{DNA}/V \exp(-\beta \Delta G_{uX})}, \tag{19a}$$
$$X_{2DNA}(t) = X_2(t_-) + X_{2DNA}(t_-) - X_2(t), \tag{19b}$$
$$X = \{\mathrm{Cro}, \mathrm{CI}\}$$

$L_{DNA} \approx 10^7$ is the number of binding sites on *E. coli* chromosome and $V = 1.2 \times 10^9 \mathrm{M}^{-1}$ is the cell volume.

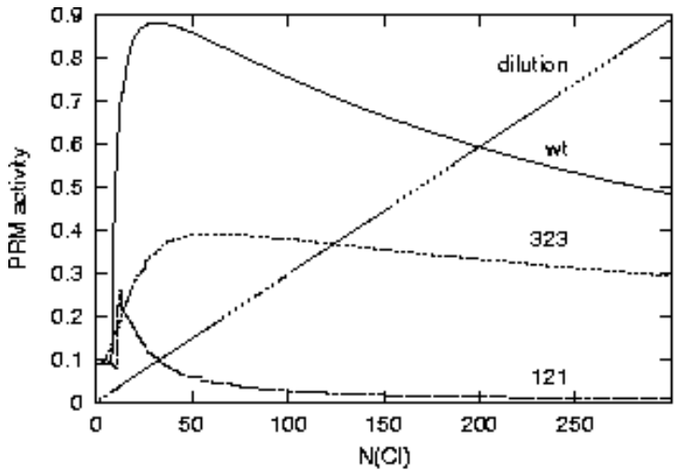


Different Systems: OR_{121} and OR_{323}

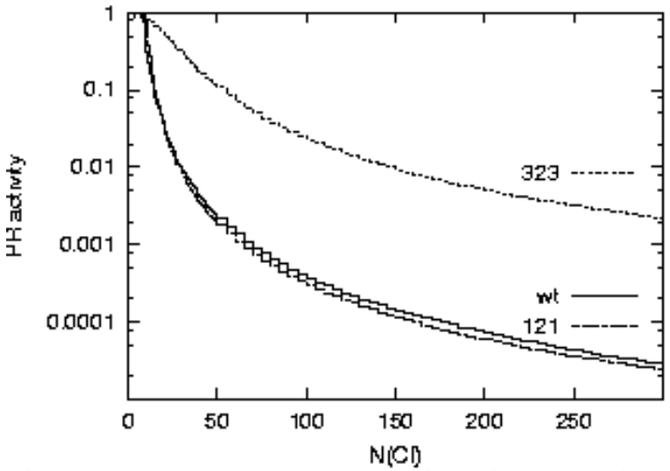
- Structure of the O_{R} can be perturbed [Little et al., 1999]



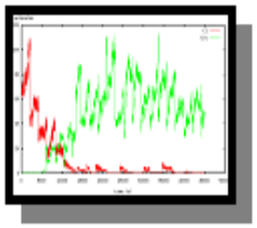
Why lysogenic state is stable overall? What role does O_{L} - O_{R} interaction play?



Plot of $Prob(\mathrm{s} = \mathrm{P_{RM}})$; O_{R121} has reduced activity of P_{RM} [Little et al., 1999, Aurell et al., 2002]

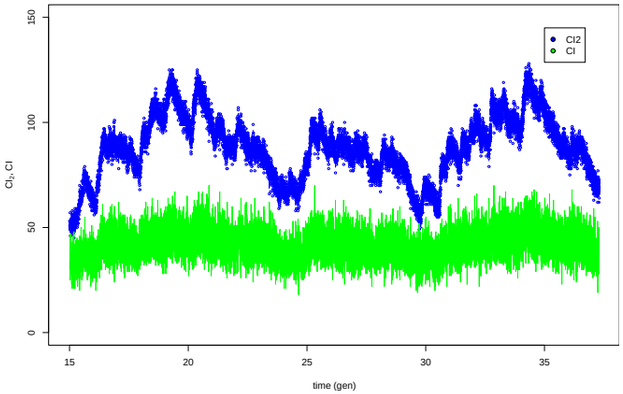


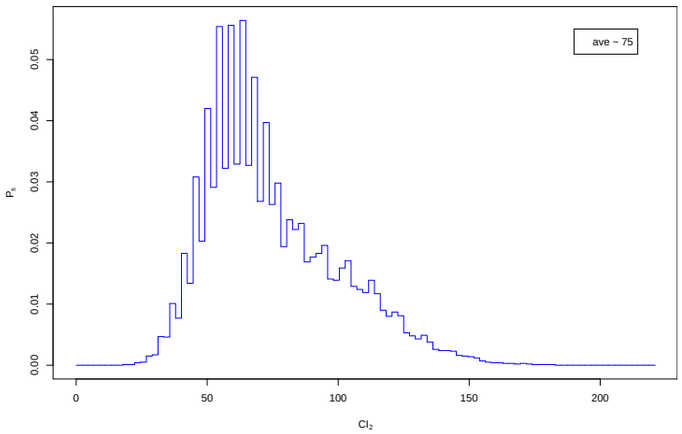
Plot of $Prob(\mathrm{s} = \mathrm{P_{R}})$; O_{R323} has increased activity of P_{R} [Little et al., 1999, Aurell et al., 2002]

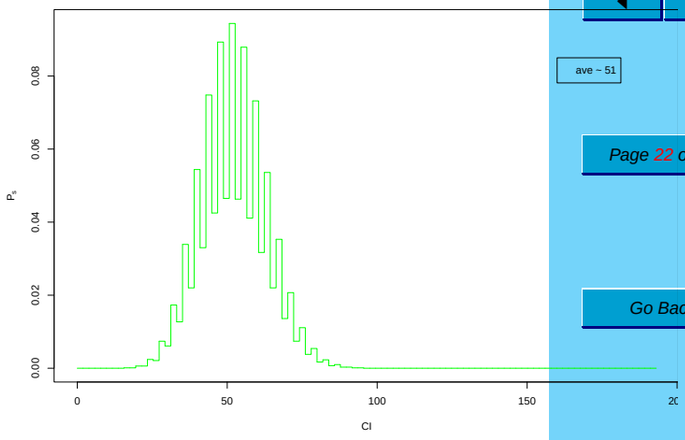


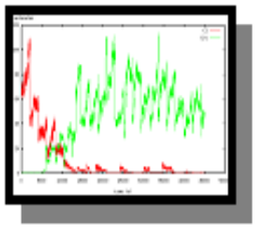
Results

λ^+-system



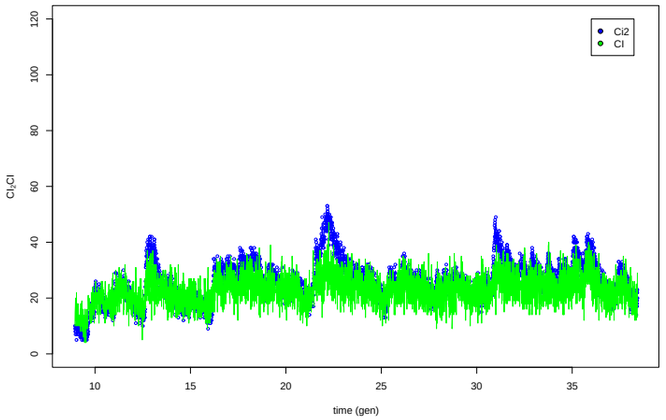


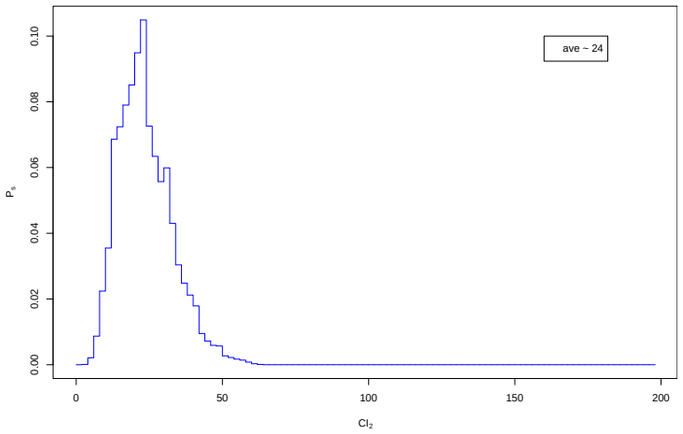


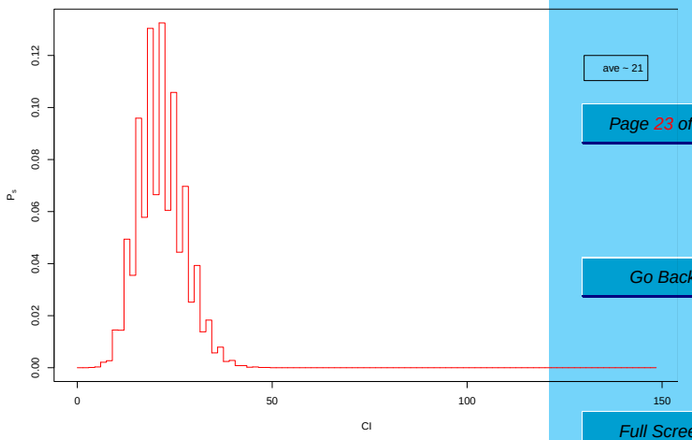


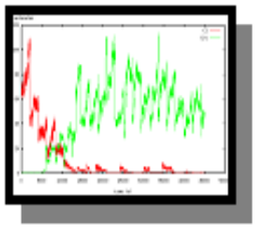
Results

CI $\lambda-\mathrm{OR}_{121}$-system

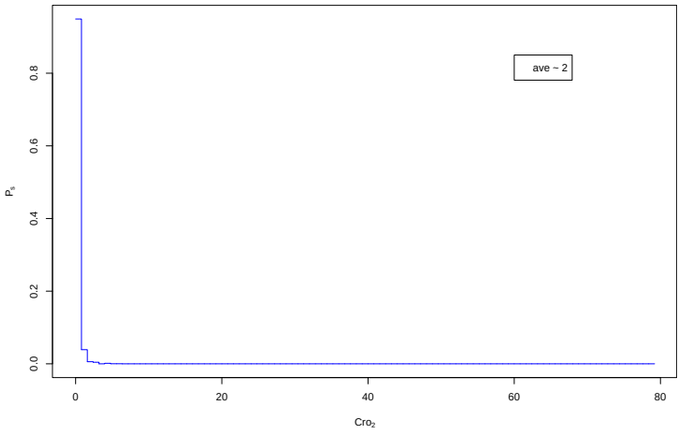


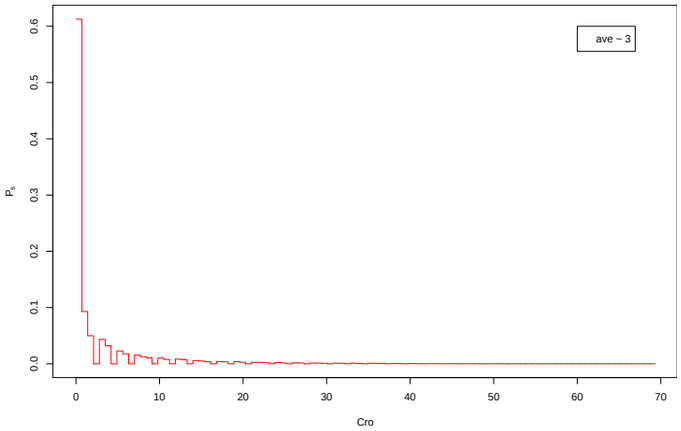


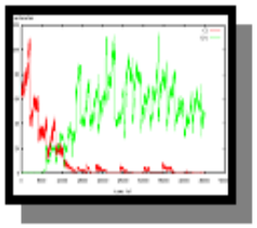




Cro $\lambda-\mathrm{OR}_{121}$-system

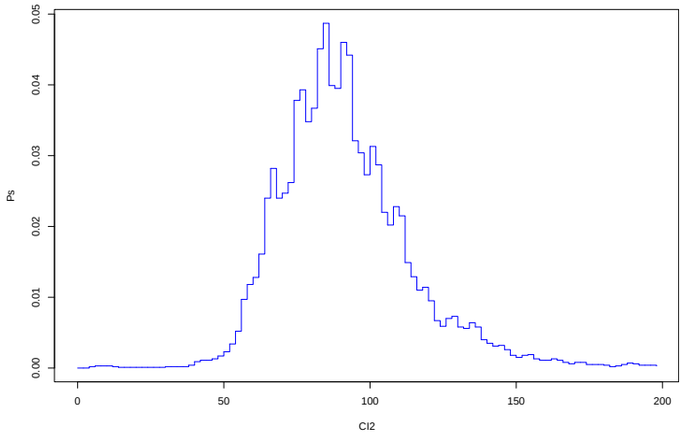


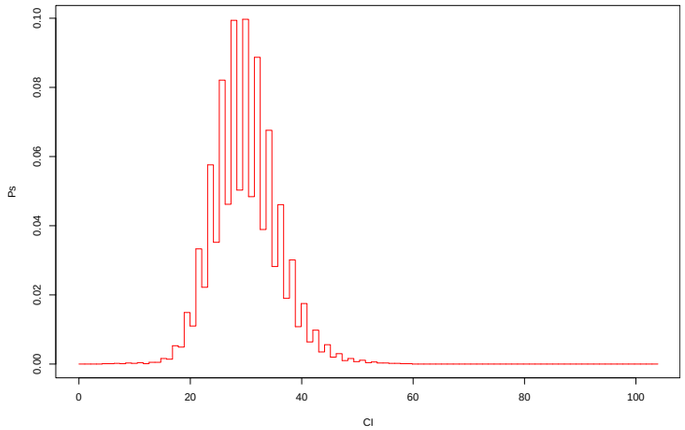


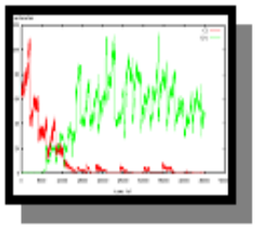


Results

CI $\lambda-\mathrm{OR}_{323}$-system

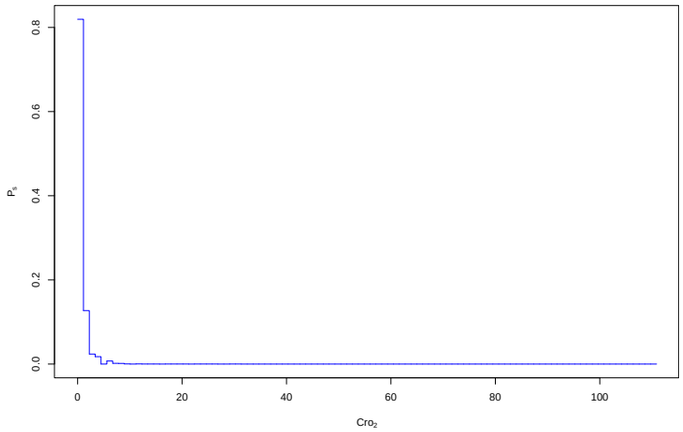


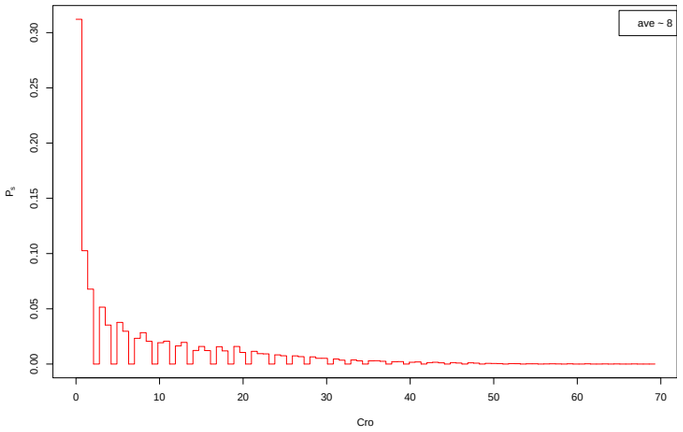


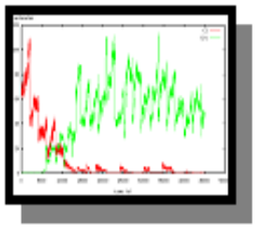


Results

Cro $\lambda-\mathrm{OR}_{323}$-system

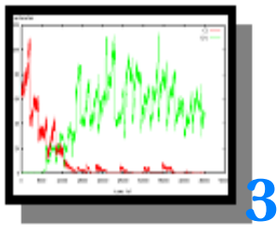






Results (contd)

- O_{R} -CI - O_{L} interaction may lead to stability
- O_{R121} may not be stable at very strong cyclizatin rates

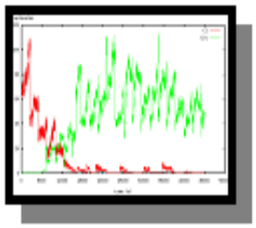


3. Rare events in transcriptional regulation of λ-phage infected *E. coli* cells.

Sergey Plyasunov

Abstract

We examine the statistical picture of transition pathways that describe the decay from a meta-stable lysogeny state in λ-phage infected *E. coli* cells, which is known to have an exponentially large stability under normal immune conditions. We present results on identification of the transition pathways and computation of the effective rate of the transition *lysogeny* $\rightarrow$ *lysis*. This formalism defines the quantitative measure of the robustness of epigenetic states.



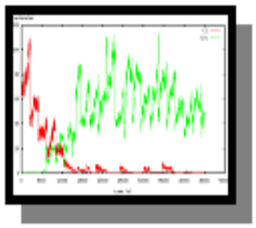
Transition lysogeny $\rightarrow$ lysis in λ phage.

- In the absence of *recA* -mediated cleavage of the repressor (so called *recA*$^-$ system) λ^+ system is exceptionally stable ($5-7$ years; compare to $30min \approx 1gen$). Experiments of [Toman et al., 1985] show possibility of switching back to lysogenic state from anti-immune state in a defect λ-phage that can not escape the *E. coli* chromosome. In this case system switches back to lysogenic state with high Cro numbers with rate $10^{-2}-10^{-3}$ per generation and per cell:

$$\underbrace{A}_{lysogeny} \overset{k_{AB}}{\underset{k_{BA}}{\rightleftharpoons}} \underbrace{B}_{lysis} \tag{20}$$

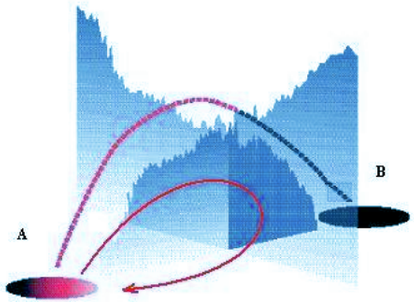
(21)

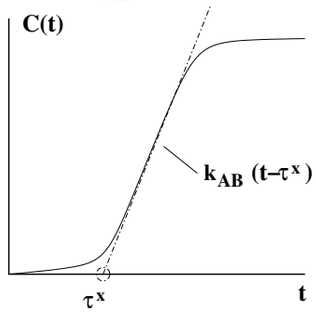
- Quasi-stationary state A for the wild-type system corresponding to the total number of CI ≈ 200 (≈ 100 CI_2) and Cro ≈ 0 . In the lytic state (B) CI ≈ 0 and Cro $\approx 40-80$ molecules in total.
- How one can predict "macroscopic" rates k_{AB}, k_{BA} from "microscopic" parameters (kinetic rates, binding energies, etc.)??



Kinetic Rates and Rare Events

What is kinetic rate k_{AB}?



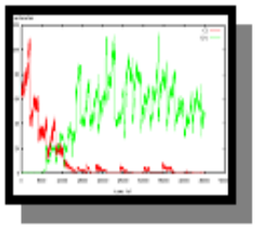


$$C(t) = \frac{\langle 1_A(\mathbf{X}(0)) 1_B(\mathbf{X}(t)) \rangle}{\langle 1_A(\mathbf{X}(0)) \rangle}, \quad C(t) \approx \begin{cases} 0, \ t \leq \tau^X \\ k_{AB} \times (t - \tau^x), \ \tau^X < t < k_{AB}^{-1} \\ \frac{k_{AB}}{k_{AB}+k_{BA}} \times e^{-(k_{AB}+k_{BA})t}, \ t > k_{AB}^{-1} \end{cases} \tag{22}$$

Kinetic rate can be found as a slope of the correlation function $C(t)$.
But straightforward approach:to follow the time evolution of the system with molecular dynamics simulations until a reasonable number of events has been observed will fail.
Examples:

- Chemical kinetics ([Kramers, 1940],[Hänggi et al., 1990])
- Protein folding
- Complex database query (e.g. statistics of alignment scores)
- Communication networks failures
- etc

Use of traditional Monte Carlo methods is “prohibited” even for the “simple” chemical systems: (e.g. proton transfer in H_2O: $\tau_{\text{dwell } H_2O} = 1 \ hour$,$\tau_{\text{vib}} = 10^{-15} \ sec$) or more complex (hydrophobic polymer collapse[tenWolde and Chandler, 2002], DNA polymerase β closing [Radhakrishnan and Schlick, 2004]).

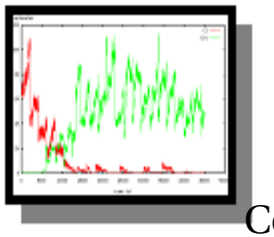


Breaking the Barrier of Rare Events: Study of the rare events/large deviations

1. H. Kramers (1940)[Kramers, 1940] and his early theory of chemical reaction rates as a diffusion over the simple barrier (Kramers Theory).
2. Transition State Theory (TST) Requires the identification of the potential barrier and *transition state*:

$$k_{AB} = \omega_A \exp(-\Delta G^{\ddagger}_{AB}/RT)$$

   Equilibrium systems only.
3. Large Deviation Theory in dynamical systems (small noise limit) [Freidlin and Wentzel, 1984]. Applicable for non-equilibrium systems [Aurell and Sneppen, 2002].
4. Transition Path Sampling (TPS)[Pratt, 1986], [Dellago et al., 1998, Berne et al., 1997, tenWolde and Chandler, 2002, Dellago et al., 2002, Hagan et al., 2003]. Statistical mechanics of transition pathways connecting meta-stable states of the equilibrium system. Crucial point: need "seed" pathway and efficient sampling in pathway-space.
5. Multilevel methods (e.g. Transition Interface Method (TIS) [van Erp et al., 2003]). Diffusive transition with multiple re-crossings [Bolhuis, 2003].



Trajectory of the Markov Process

Consider the Markov process $\{X_t\}_{0\leq t\leq T}$.
Assumption: ergodisity w.r.t. some invariant measure (Gibbs measure for "classical" MD).
Types of dynamics:

- Langevin dynamics:

$$\dot{X} = p, \tag{23a}$$

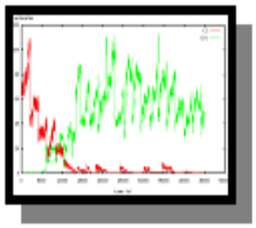
$$\dot{p} = -\nabla_X U(X) - \gamma p + \underbrace{\sigma\xi}_{white\ \ noise} \tag{23b}$$

- Overdumped-"Chemical" Langevin equation:

$$\dot{X}_t = a(X_t) + \underbrace{\sigma(X_t)\xi}_{white\ noise} \tag{23c}$$

- Jump-process:

$$dX_t = \sum_r \nu_r \underbrace{dN_r(dt|X_t)}_{state\ dep.\ Poisson\ noise} \tag{23d}$$



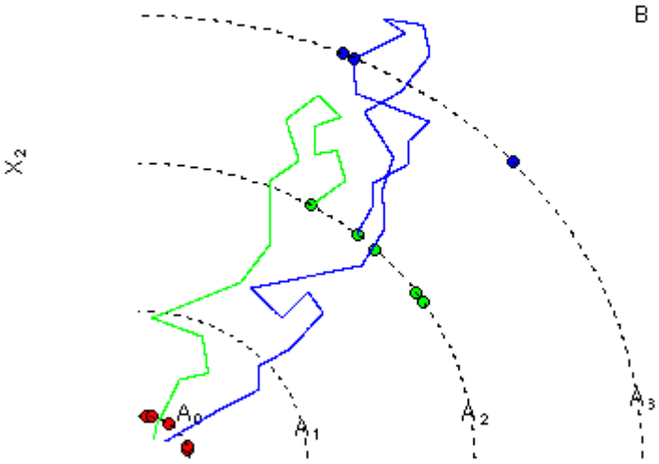
Computational framework for the calculation of k_{AB}.

Approach:

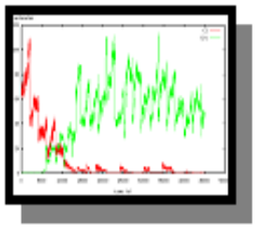
- Introduce interfaces $A_i : A_0 = A, A_1, A_2, \ldots A_n = B$
- Random crossing time(s): $\tau_B = \inf\{0 \leq t \leq \infty : \mathbf{X}_t \in B\}$ τ_A (first return back to A): $\tau_i = \inf\{0 < t \leq \infty : \mathbf{X}_t \in A_i\}$.
- Transition rate in diffusive limit:

$$k_{AB} = \nu_{A,0}\mathbb{P}(\tau_B < \tau_A) \tag{24a}$$

$$\mathbb{P}(\tau_B < \tau_A) = \mathbb{P}(\tau_1 < \tau_A)\prod^{N}\mathbb{P}(\tau_i < \tau_A | \tau_{i-1} < \tau_A) \tag{24b}$$



Stochastic trajectory starting at A_0 and labeled as a corresponds to the event $\{\tau_1 > \tau_0\}$ while pathway labeled as a' corresponds to the event $\{\tau_1 < \tau_0\}$. Similar, trajectory b corresponds to the event $\{\tau_2 < \tau_0\}$ took place conditional on event $\{\tau_1 < \tau_0\}$, while b' corresponds to the event $\{\tau_2 > \tau_0\}$ conditional on event $\{\tau_1 < \tau_0\}$.
$\nu_{A,0}$-frequency of crossing events through the A_0.



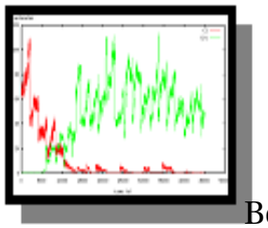
Computational framework for the rate calculation

- At every interface i one runs n_i replications of the trajectory from the point $\mathbf{X}_{0,i}$. Trajectory is stopped either when it reaches the interface of the level $i+1$ or return back to the original state $A_0 = A$.

$$\mathbb{P}(\tau_{i+1} < \tau_A | \tau_i < \tau_A) = p_i \approx \frac{n_{i \to i+1}}{n_i} \tag{25}$$

- New starting position $\mathbf{X}_{0,i+1}$ is the average $\mathbf{X}_{\tau_{i+1}}$: $\mathbf{X}_{0,i+1} = \frac{1}{n_{i \to i+1}} \sum_{j=1}^{n_{i \to i+1}} \mathbf{X}_{\tau_{i+i}^j}$
- Estimator for $\mathbb{P}$ is unbiased but has a variance:

$$\sqrt{\text{var}\{p_i\}} = \frac{\sqrt{(1-p_i)p_i}}{n_i} \tag{26}$$



Transition pathway

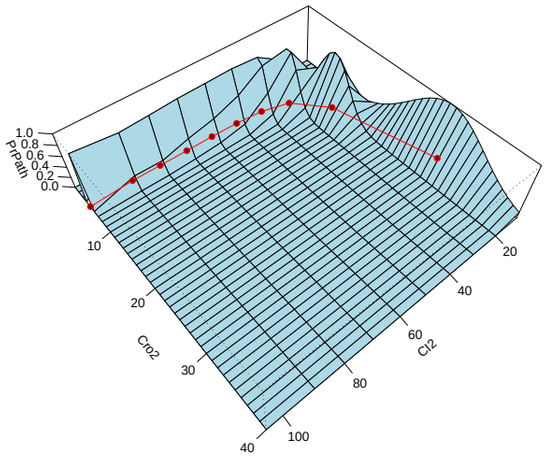
Between interfaces A_i trajectories are simulated with ME:

$$\frac{\partial P(X_1,X_2,t)}{\partial t} = f_1(X_1-1,X_2)P(X_1-1,X_2,t) + f_2(X_1,X_2-1)P(X_1,X_2-1,t)+ \quad (27a)$$

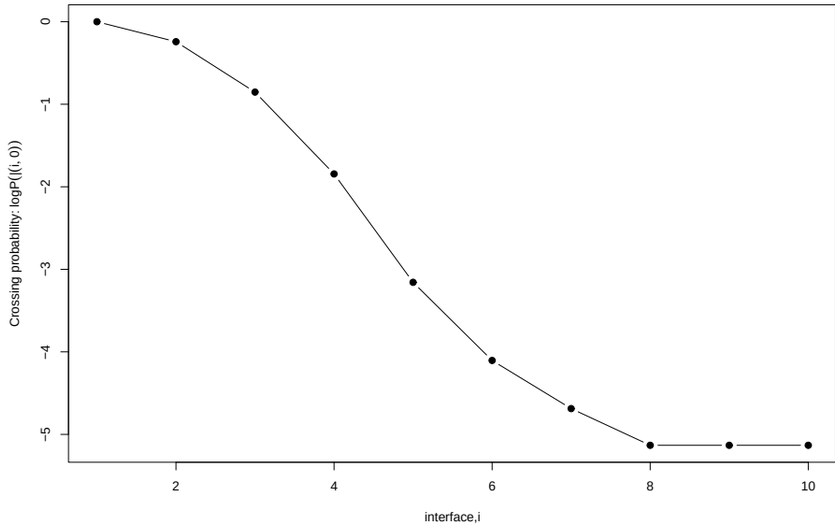
$$+k_1(X_1+1)P(X_1+1,X_2,t) + k_2(X_2+1)P(X_1,X_2+1,t)- \quad (27b)$$

$$-(f_1(X_1,X_2)+f_2(X_1,X_2)+k_1X_1+k_2X_2)P(X_1,X_2,t) \quad (27c)$$

variables $X_1 = \mathrm{CI}_2$ and $X_2 = \mathrm{Cro}_2$.

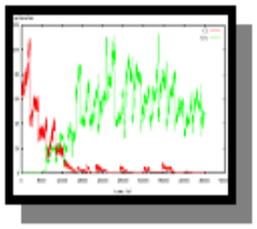


Resulting transition pathway for the lysogeny-lysis transition corresponding to the w.t. parameters. $n_i = 10^4$ trajectories are used at every interface. Interfaces $i = 0 \ldots 10$ are located at CI_2=const

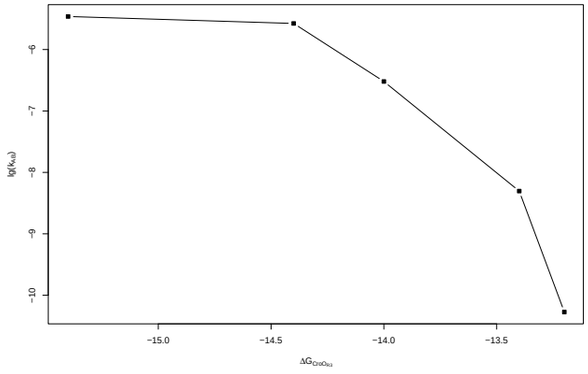


Cumulative crossing probability $\lg \mathbb{P}(0 \to i)$ at the different interfaces i for the lysogeny-lysis transition corresponding to w.t. λ phage.

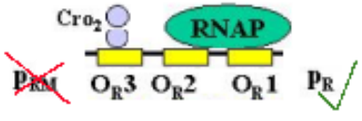
Slope of $\lg \mathbb{P}$ is maximal at $i = 4-6$ corresponding to CI_2=50-60 and $\mathrm{Cro}_2 \approx 8$. (“TS”)



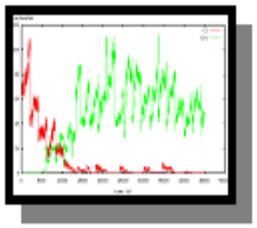
Robustness



Dependence of the transition rate k_{AB} on the Gibbs energy of the Cro_2 dimer binding to the operator site O_{R3}. One can see that k_{AB} increases almost 3 orders of magnitude when $\Delta G_{\mathrm{Cro}O_{\mathrm{R}}3}$ is decreased, but it still stays at very low values ($\propto 10^{-6} s^{-1}$, compare with the time scales of one generation of *E. coli* cells $\approx 2\ 10^{3} sec$) and lysogenic states remains robust under large perturbations in $\Delta G_{\mathrm{Cro}O_{\mathrm{R}}3}$

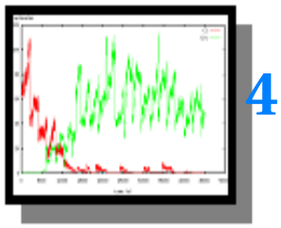
- 

Gibbs energy $\Delta G_{\mathrm{Cro}O_{\mathrm{R}}3}$ (kcal/mol)	$k_{AB}(sec^{-1})$
−13.10	$5.33\ 10^{-11}$
−13.40 [Darling et al., 2000]	$4.96\ 10^{-9}$
−14.00	$3.02\ 10^{-7}$
−14.40	$2.66\ 10^{-6}$
−15.40	$3.45\ 10^{-6}$



Conclusions

- Concept of robustness is introduced
- Algorithm is presented and used to study the transition rates of the lysogeny $\rightarrow$ lysis
- Stability of the λ phage is investigated in response to the change in $\Delta G_{\mathrm{Cro}O_{\mathrm{R}}3}$

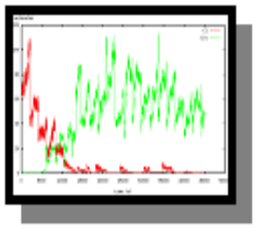


4. Coarse-grained Kinetic Monte Carlo simulations: separation of time scales and renormalization of transition rates.

Sergey Plyasunov

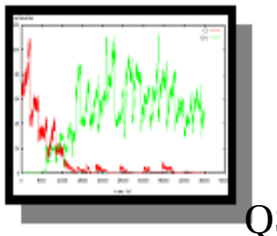
Abstract

This work addresses the theoretical framework and numerical methods for performing stochastic simulations of reaction dynamics in chemical networks with time-scales separation. This technique is based on application of the projection technique and cumulant expansion to the chemical Master Equation. We present a general and systematic procedure for the elimination of the fast irrelevant variables and present a new form of the chemical master equation which involves only relevant species with the ratio of time-scales serving as a small perturbation parameter. Accuracy of the perturbation expansion is analyzed. This approach is applicable to a wide range of problems including typical modeling framework of biochemical/genetic networks.



Coarse-Graining

- In many cases separation of time scales is very well developed (example: binding/dissociation events of TF or change of DNA topology v.s. gene translation/transcription): fast and slow manifolds. Many other examples can be given across different scientific disciplines.
- System takes the “closure” on the slow manifold.
- Fast reactions are becoming the computational bottleneck of KMC $\Rightarrow$ Need for computational techniques which are able to “coarse-grain” on irrelevant features of the system (think of Claude Monet or Renoir)
- “Coarse-graining” has to be done in stochastic framework (Reason: irrelevant species may have low copy number [Kepler and Elston, 2001, Bundschuh et al., 2003, Rao and Arkin, 2003, Shibata, 2003]).
- Maintain accuracy and achieve speed up.



Deterministic QSSA

QSSA provides the dimensionality reduction for deterministic systems with separation of time scales: Examples

1. Enzymatic Networks:

$$\mathrm{X} + \mathrm{E} \underset{k_{-1}}{\overset{k_{+1}}{\rightleftharpoons}} \mathrm{EX} \xrightarrow{k_2} \mathrm{E} + \mathrm{X}^* \tag{28a}$$

$$dEX/dt \to 0, \;\; EX = E_0 X/(X + K_M) + O(\epsilon) \tag{28b}$$

$$\varepsilon = E_0/(K_M + X_0), \tag{28c}$$

$$E_0 = E + EC, \;\; K_M = (k_{-1} + k_2)/k_1 \tag{28d}$$

2. These two networks are dynamically equivalent (Brusselator[Nicolis and Prigogine, 1977], non-linear chemical "oscillator"):

$$\mathrm{A} \to \mathrm{X} \tag{29a}$$

$$\mathrm{X} + 2\mathrm{Y} \xrightarrow{\tilde{\mathrm{k}}} 3\mathrm{Y} \tag{29b}$$

$$\mathrm{Y} \to \mathrm{B} \tag{29c}$$

$$\mathrm{A} \to \mathrm{X} \tag{30a}$$

$$2\mathrm{Y} \underset{k_{-1}}{\overset{k_{+1}}{\rightleftharpoons}} \mathrm{Z} \tag{30b}$$

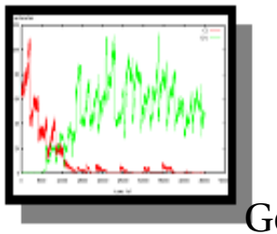
$$\mathrm{X} + \mathrm{Z} \xrightarrow{k_2} \mathrm{Y} + \mathrm{Z} \tag{30c}$$

$$\mathrm{Y} \to \mathrm{B} \tag{30d}$$

$k_1 \ll k_{-1}, \;\; k_2 = O(k_{-1}), \;\; \varepsilon = 1/k_{-1}.$
If $\tilde{y} = y + 2z$,then:

$$z = k_1 \tilde{y}^2 \varepsilon + O(\varepsilon^2) \tag{31}$$

$$\tilde{k} = \frac{k_1 k_2}{k_{-1}} \tag{32}$$



Fast and slow reactions

Goal: exploit separation of time scales to simplify the ME:

$$\frac{\partial p(\mathbf{S},t)}{\partial t} = \mathbb{L}p(\mathbf{S},t), \qquad (33)$$

linear operator $\mathbb{L}$ for the pure jump Markov process:

$$\mathbb{L}\ldots = \sum_{r=1}^{R} a_r(\mathbf{S}-\boldsymbol{\nu}_r)\ldots - \sum_{r=1}^{R} a_r(\mathbf{S})\,. \qquad (34)$$

set of reactions

$$\mathcal{R} = \mathcal{R}_0 \cup \mathcal{R}_! \{\text{fast}, \text{slow}\}, \qquad (35)$$

$$\mathbf{S} = (\mathbf{Y}, \mathbf{X}) = (\text{fast}, \text{slow}), \qquad (36)$$

$$\epsilon = \tau_Y/\tau_X \ll 1 \qquad (37)$$

New Chemical Master Equation:

$$\frac{\partial p(\mathbf{X},t)}{\partial t} = \sum_{r\in\mathcal{R}_1} \tilde{a}_r(\mathbf{X}-\boldsymbol{\nu}_{rX},t)p(\mathbf{X}-\boldsymbol{\nu}_{rX},t) - p(\mathbf{X},t)\sum_{r\in\mathcal{R}_1}\tilde{a}_r(\mathbf{X},t). \qquad (38)$$

Assumptions:

- Conditional on the slow species, fast should reach a stable distribution quick: $p(\mathbf{Y},t|\mathbf{X}) \to \hat{p}(\mathbf{Y}|\mathbf{X})$ on the time scale $\tau_Y \ll \tau_X$.
- Cumulants:

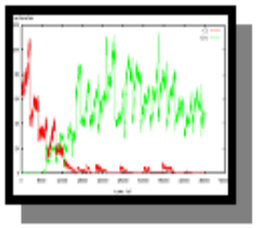
$$C_1(t;\mathbf{X}) = \langle \mathbf{Y} \rangle \qquad (39)$$

$$\mathbb{C}_2(t_1,t_2;\mathbf{X}) = \langle\langle \mathbf{Y}(t_1)\mathbf{Y}^T(t_2)\rangle\rangle = \langle \mathbf{Y}(t_1)\mathbf{Y}^T(t_2)\rangle - \langle \mathbf{Y}(t_1)\rangle\langle \mathbf{Y}^T(t_2)\rangle \qquad (40)$$

$$\ldots \qquad (41)$$

  Computed over the $\hat{p}(\mathbf{Y}|\mathbf{X})$ must exist and be finite.

- Comment: "Slow" species participate only in "slow" reactions (i.e. reactions with small $a_r(\mathbf{X},\mathbf{Y}),\ r\in\mathcal{R}_1$)



Basics of the Kinetic Monte Carlo

- Survival/waiting probability:

$$Q(t|\mathbf{X},\mathbf{Y}) = \exp(-t\sum_r a_r(\mathbf{X},\mathbf{Y})) = \prod_r \exp(-t a_r(\mathbf{X},\mathbf{Y})) \equiv \prod_r Q_r(t|\mathbf{X},\mathbf{Y}), \tag{42}$$

$$p_r(t|\mathbf{X},\mathbf{Y}) = -\frac{\partial}{\partial t} Q_r(t|\mathbf{X},\mathbf{Y}) \tag{43}$$

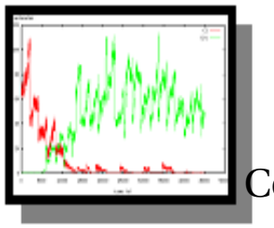
- Time steps τ of the reactions are sampled from $Q_r(t|\mathbf{X},\mathbf{Y})$ and smallest is chosen.
- Update time-step:

$$\begin{aligned} \tau_1 &\propto Q_1(t|\mathbf{X}), \\ \tau_2 &\propto Q_2(t|\mathbf{X}), && (44) \\ &\dots && (45) \\ \tau = \min(\tau_1,\tau_2,\dots) &= \tau_{\mathrm{r}^*},\ t \leftarrow t+\tau_{\mathrm{r}^*}, && (46) \end{aligned}$$

- Update species:

$$(\mathbf{X},\mathbf{Y}) = (\mathbf{X},\mathbf{Y}) + \boldsymbol{\nu}_{\mathrm{r}^*} \tag{47}$$

- How does distribution $Q_r(t|\mathbf{X},\mathbf{Y})$ looks like for the slow reactions? How strong non-Markovian effects?
- When it's possible to introduce the *effective transition rate*?



Example

Consider the system:

$$\mathsf{X} + \mathsf{X} \underset{k_{-1}}{\overset{k_{+1}}{\rightleftarrows}} \mathsf{X}_2$$

$$\mathsf{E} + \mathsf{X}_2 \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} \mathsf{EX}_2$$

$$\mathsf{EX}_2 \xrightarrow{k_3} \mathsf{EX}_2 + \mathsf{P}(*)$$

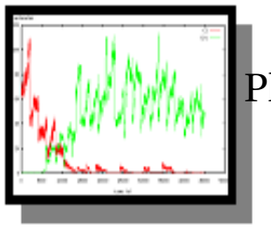
$$\mathsf{P} \xrightarrow{k_4} \emptyset$$

$$N = E + EX_2, \ \varepsilon = \frac{k_3}{k_{-2} + k_2 X_2} \leq 1.0$$

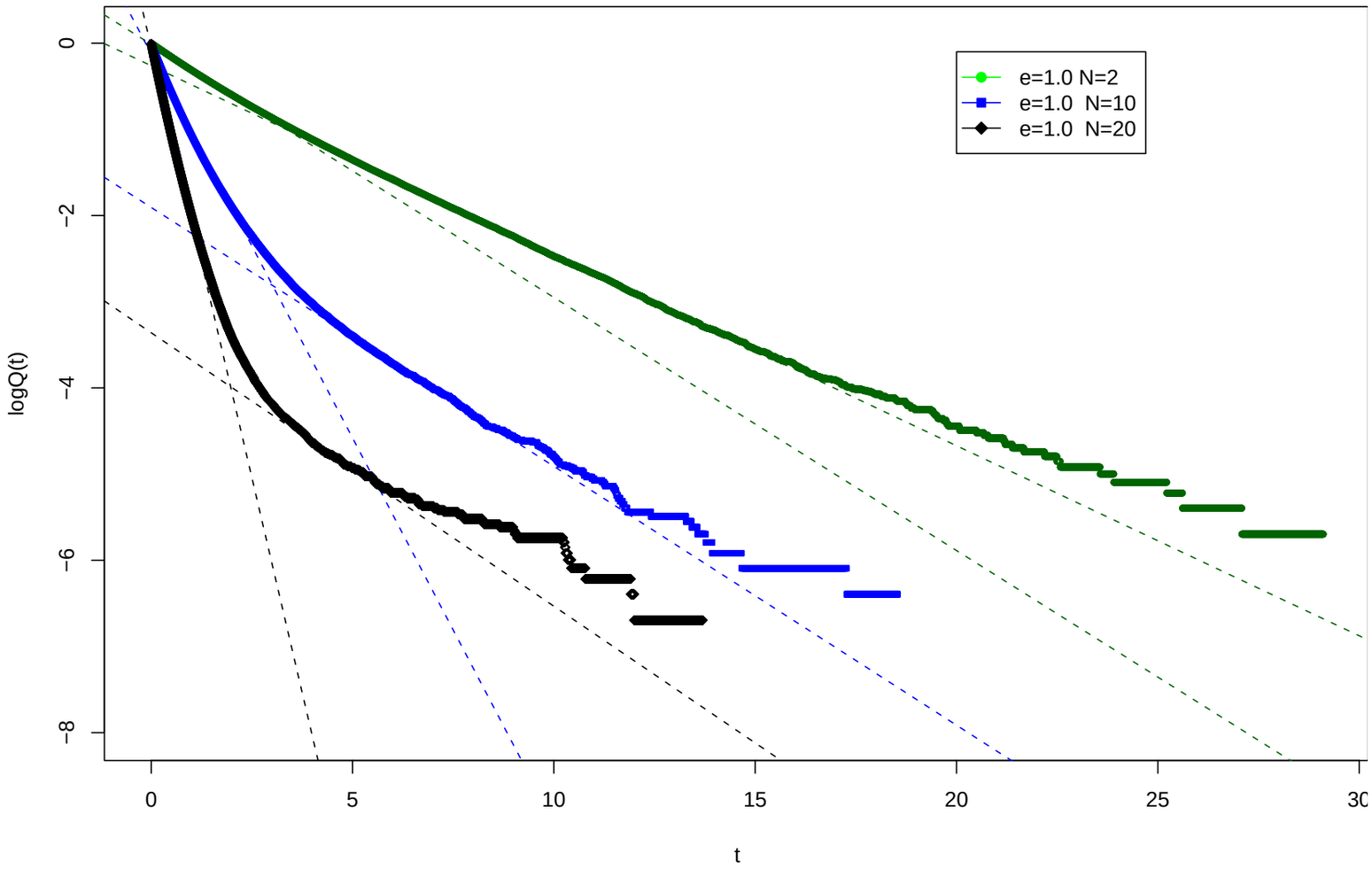
$$\mathbb{E}\{EX_2 | X_2\} = \frac{N k_2 X_2}{k_{-2} + k_2 X_2}$$

$$\langle\langle EX_2(t) EX_2(0) \rangle\rangle = \langle \Delta EX_2^2 \rangle e^{-k_{-2} + k_2 X_2 t},$$

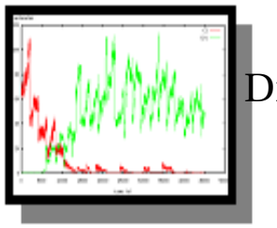
- Investigate $Q_*(\tau | \mathbf{X})$ at different ϵ, N
- Can the distribution be fitted to a set of lines?



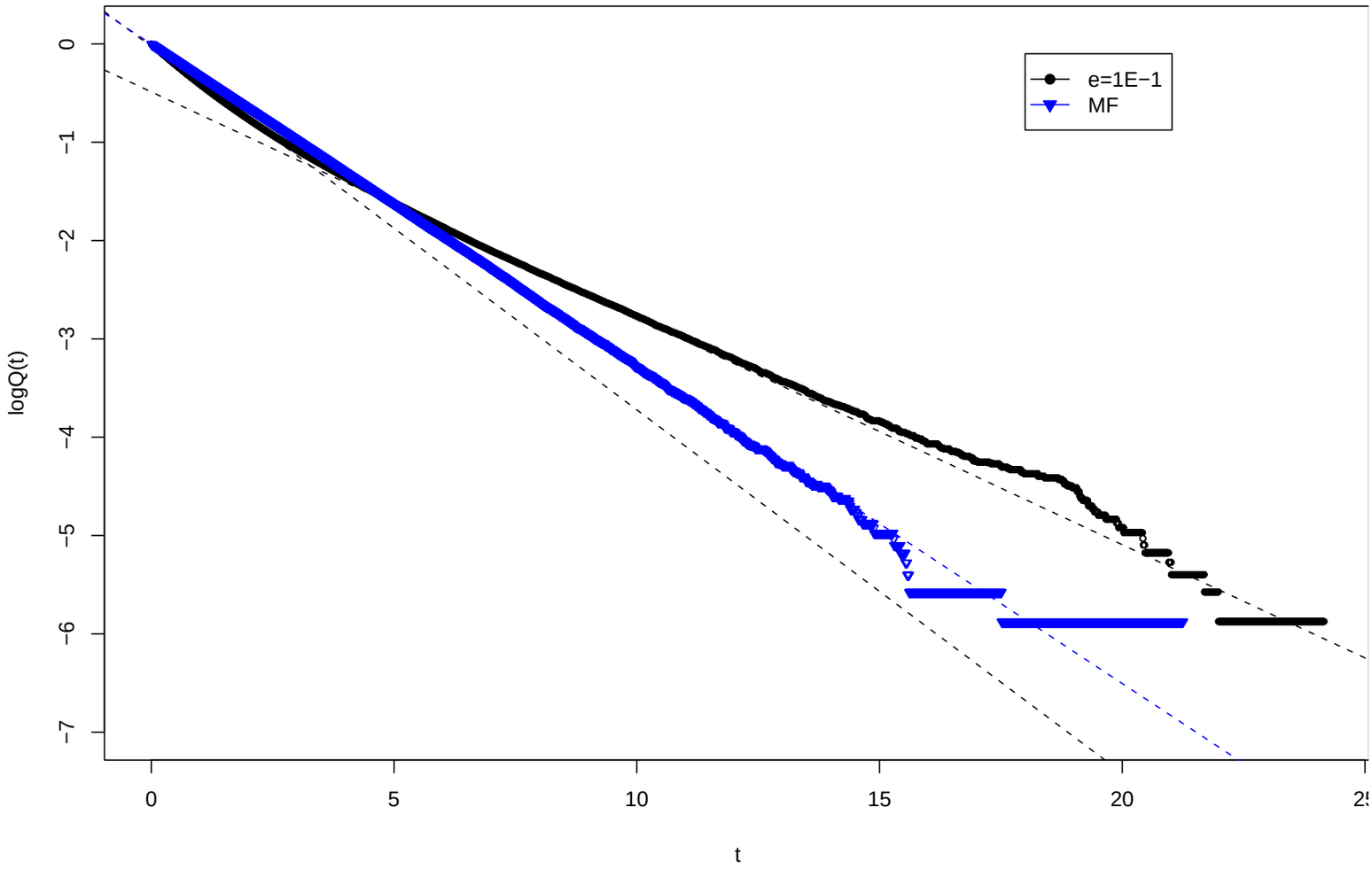
Plots of $\log\ Q_*(\tau|\mathbf{X})$ for different ϵ and N(Strong non-Markovian effects):



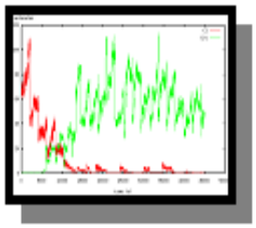
- Two asymptotic for the kinetic rate $-\frac{\partial \log Q_*}{\partial t}$
- The bigger number of states (N) in the fast manifold the stronger non-Markovian effects
- at large t $\log Q_*(t)$ is a straight line again (intermitancy dies out)



Distribution $Q_*(t)$ for $\epsilon = 0.1$ and $N = 3$



- Mean-field rate $\sum_Y a_*(\mathbf{X}, \mathbf{Y})\hat{p}(\mathbf{Y}|\mathbf{X})$ goes in between of the asymptotic of $-\frac{\partial \log Q_*(t)}{\partial t}$

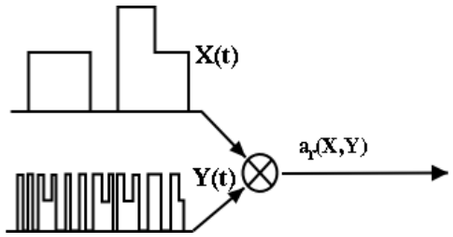


Statistics of waiting times and Renormalization of Rates

General approach for the effective transition rates.

Consider one particular "slow" reaction r:

$$\nu^{+}_{X1r}X_1 + \cdots + \nu^{+}_{Y1r}Y_1 + \ldots \xrightarrow{k_r} \nu^{-}_{X1r}X_1 + \cdots + \nu^{-}_{Y1r}Y_1 + \ldots \tag{49}$$



which involves species from both subsets $\mathbf{X}$ and $\mathbf{Y}$.

- Averaged survival probability:

$$\tilde{Q}_r(t|\mathbf{X}) \equiv \left\langle \exp\left(-\int_0^t ds a_r(\mathbf{X}, \mathbf{Y}_s)\right)\right\rangle \tag{50}$$

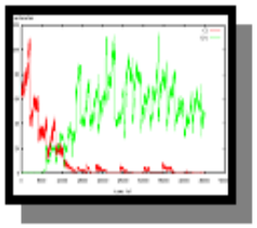
  Average $\langle\cdot\rangle$ is taken over the realizations of the jump process $\mathbf{Y}_s, [0 \leq s \leq t]$ with probability density $\hat{p}(\mathbf{Y}, t|\mathbf{X})$.

- Eqn. (50) can be represented as a sum over all possible cumulants of the process $\mathbf{Y}$:

$$\tilde{Q}_r(t|\mathbf{X}) = \exp\left[\sum_{m=1}^{\infty} \frac{(-1)^m}{m!} \int_0^t dt_1 \ldots \int_0^t dt_m C_r^{(m)}(\mathbf{X}, t_1, \ldots, t_m)\right] \tag{51}$$

- Effective transition rates $\tilde{a}_r$:

$$\tilde{a}_r(\mathbf{X}, t) = -\frac{\partial}{\partial t} \ln \tilde{Q}_r(t|\mathbf{X}) \tag{52}$$



Effective Rates

- In the first order approximation [Rao and Arkin, 2003] (mean-field):

$$\tilde{a}_r(\mathbf{X},t) = C_r^{(1)}(\mathbf{X},t) = \langle a_r(\mathbf{X},\mathbf{Y})\rangle,\ r \in \mathcal{R}_1 \tag{53}$$

This gives Michaelis-Menton, Hill transition rates in deterministic framework.

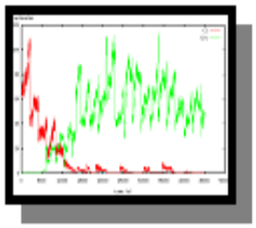
- The difference:

$$\Delta a_r(\mathbf{X},t) = \tilde{a}_r(\mathbf{X},t) - \langle a_r(\mathbf{X},\mathbf{Y})\rangle = \frac{\partial}{\partial t}[\sum_{m=2}^{\infty}\frac{(-1)^{m-1}}{m!}\int_0^t \dots \int_0^t C_r^{(m)}(\mathbf{X},t_1,\dots,t_m)dt_1\dots dt_m] \tag{54}$$

expresses the contribution of the fluctuations of the eliminated fast variables to the effective rate.

$$C_r^{(m)}(t_1,\dots,t_m) = A_r^{(m)}(\epsilon)\prod_u e^{-\mu_r(\epsilon)|t_u - t_m^*|}, \tag{55}$$

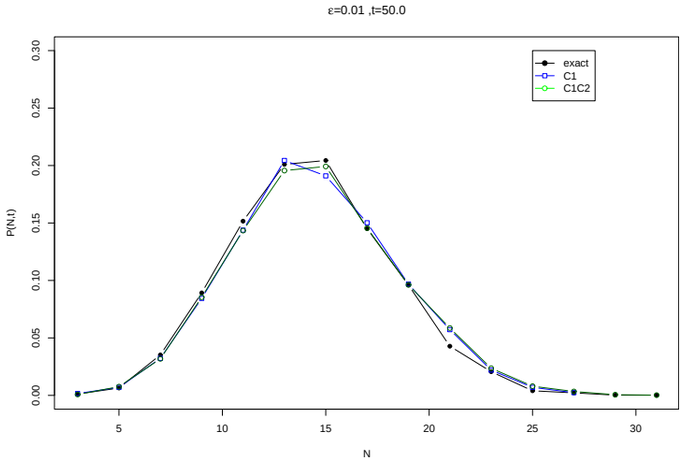
$t_m^* = \min\{t_1,\dots,t_m\}$ and $\mu_r(\epsilon) \propto \epsilon^{-1}$.



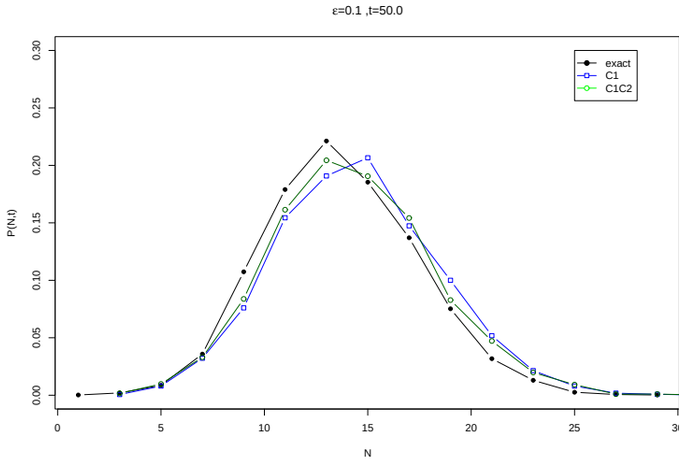
- Regime of the short memory the effective rates are independent of the time:

$$\tilde{a}_r(\mathbf{X}, t) = \text{independent of } t \text{ as } t \to \infty$$

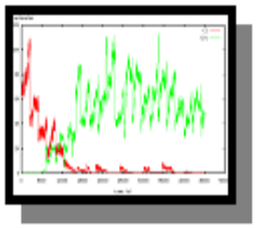
and description becomes Markovian at the time scales larger then the correlation length of the fast species $\tau_Y \Rightarrow$ Regular kinetic Monte Carlo schemes (Bortz et al. [1975], Gillespie [1977]) for stochastic simulation of contracted system i.e. sampling trajectories $\mathbf{X}_t$.



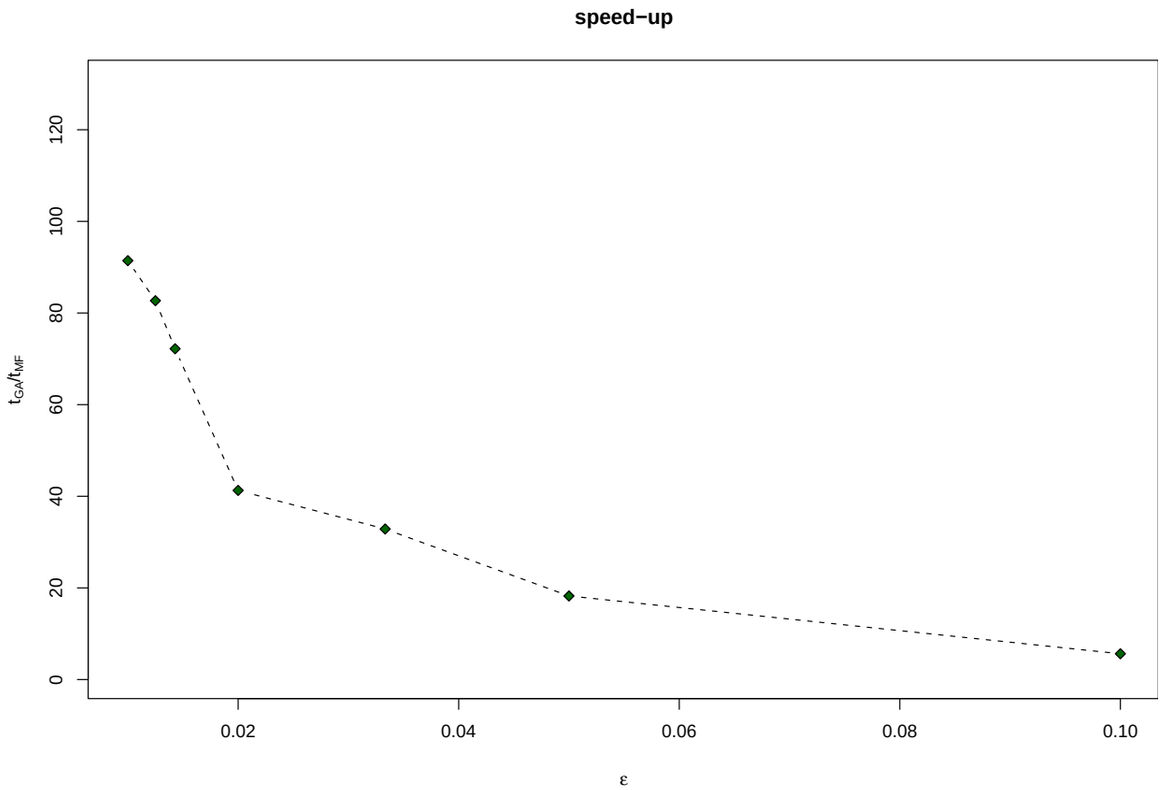
Histograms of the number of reaction events in the reaction channel (*) obtained by exact kinetic Monte Carlo (Gillespie method), by mean-filed reaction and second order correlation correction for $\epsilon = 0.01$.

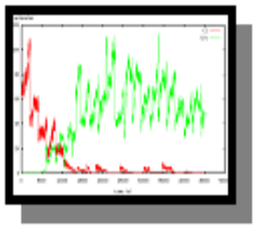


Distributions of the number of reaction events in the reaction channel (*) obtained by exact kinetic Monte Carlo, by mean-filed reaction and second order correlation correction for $\epsilon = 0.1$



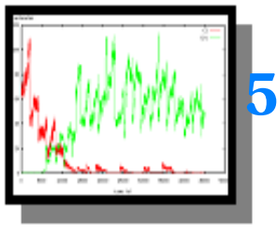
Speed-Up





Conclusions

- Approach is applicable for the reaction systems which display well-developed separation of time scales between relevant and irrelevant species.
- Effective kinetic rates can be identified through the averaging of the transition rates over the statistics of $\mathbf{Y}|\mathbf{X}$ (using mean and correlation functions of the conditional process $\mathbf{Y}|\mathbf{X}$) leading to the KMC for the coarse-grained model.



5. Uncertainty Propagation in Models of multi-Variable Chemical Reaction Networks: Separation of State Variables and Parameters.

Sergey Plyasunov

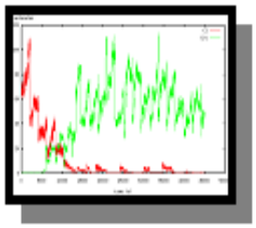
Abstract

Tentative

Uncertainty propagation scheme is presented for the stochastic system described by the chemical master equation. Method relies on Poisson mapping technique and use of Polynomial Chaos Expansion (PCE) for the propagation the uncertain structure of parameters.

Coefficients of the expansion are computed through the Galerkin procedure. The convergence of the solution with respect to the resolution level is investigated.

This computational approach can be useful for the purposes of the parameter estimation since it provides with efficient computational schemes for the evaluation of the sensitivities with respect to the kinetic rates.



Propagation of Uncertainties

- Models are always uncertain
- Sensitivity of the (stochastic) non-linear dynamics with respect to the values of parameters are crucial for design and identification
- Uncertainty can be modeled as "disorder" of the parameters, i.e. dependence of the parameters on *random variable(s)* of some type (Poissonian, Gaussian, Uniform, etc.) or even *stochastic processes* of some type (white noise).
- Even in linear systems with simple distributions of parameters (i.e. Gaussian type) resulting uncertainties in state space are usually more complicated:

$$dx/dt = L(k(\xi), x) = \pm k(\xi)x, \;\; k(\xi) = k_0(1+\sigma\xi), \;\; x \propto \mathcal{N}(0,1), \tag{56a}$$

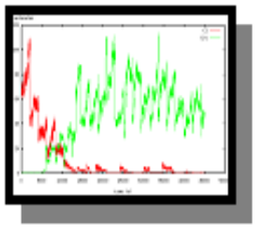
$$dP(k) = \frac{dk}{\sqrt{2\pi}k_0\sigma}\exp(-\frac{(k-k_0)^2}{2k_0^2\sigma^2}), \tag{56b}$$

$$\Downarrow$$

$$dP(x|t) = \frac{dx}{k_0\sigma\sqrt{2\pi}}\exp(-\frac{(\ln(x/x(0)) \pm k_0 t)^2}{k_0\sigma t}) \tag{56c}$$

- Small perturbations/multiple shooting with stiff ODEs/Monte Carlo; group theoretical analysis - might be too "complex" for complex systems.

One has to take uncertainty directly into the modeling approach



Polynomial Chaos Expansion

- Any "signal" $x(t) \in L_2([0,T])$ can be decomposed into the frequency spectrum:

$$\mathbf{x}(t) = \sum_n \mathbf{x}_n(\omega) e^{i\omega_n t}, \tag{57a}$$

$$||\mathbf{x}||^2 = \sum_n |\mathbf{x}_n|^2 \tag{57b}$$

- Linear system with the input $\mathbf{u}(\omega)$ is related to the response $\mathbf{x}(\omega)$:

$$\mathbf{x}(\omega) = H(\omega)\mathbf{u}(\omega)$$

- Similar to that[Cameron and Martin, 1947], any function of the random variable $\xi(\omega)$ with measure $d\mu(\xi)$ can be considered as a map the space $(\Omega, \mathcal{F}_\Omega, \mathcal{P})$ to $\mathbb{R}^n$ and can be expanded in basis of orthogonal polynomials $\{H_n(\cdot)\}$:

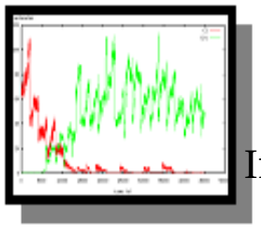
$$\mathbf{x}(t, \xi(\omega)) = \sum_{n=0}^{\infty} \mathbf{x}_n H_n(\xi(\omega)), \tag{58a}$$

$$\langle H_n, H_m \rangle = \int d\mu(\xi) H_n(\xi) H_m(\xi) = \delta_{mn}, \tag{58b}$$

$$||\mathbf{x}||^2 = \sum_{n=0}^{\infty} ||\mathbf{x}_n||^2, (\text{a wonder!}) \tag{58c}$$

- Functions $\mathbf{x}_n$ *spectral modes* representing propagation of disorder from parameters k into state variables X.
- For the nonlinear system:

$$d\mathbf{x}/dt = L(\mathbf{x}, \mathbf{k}) \rightarrow d\mathbf{x}_n/dt = \langle H_n, L(\sum_p \mathbf{k}_p H_p, \sum_m \mathbf{x}_m H_m) \rangle \tag{59}$$



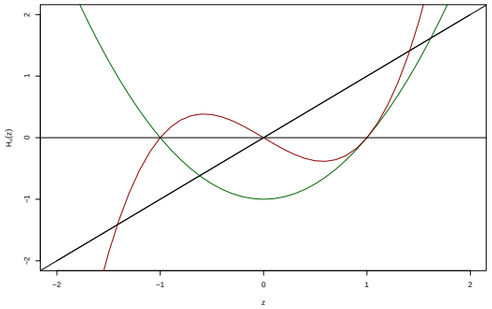
Polynomial Chaos Expansion (Cont'd)

If $d\mu(\xi) = \frac{d\xi}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}}$ then most suitable basis are

$$H_n(z) = (-1)^n e^{z^2/2} \frac{d^n}{dz^n} e^{-z^2/2}, \quad H_0 = 1, \; H_1(z) = z, \; H_2(z) = z^2 - 1, \; H_3(z) = z^3 - z$$

Then:

$$\mathbb{E}_\xi\{x(t)\} = x_0(t), \tag{60}$$
$$\mathbb{V}\mathrm{ar}_\xi\{x(t)\} = x_1(t)^2 + 2x_2^2(t) + 6x_3^2(t), \ldots \tag{61}$$



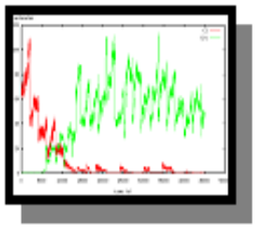
"Wiener Chaos" [Wiener, 1938],[Chorin, 1974],[Ghanem and Spanos, 1996]

- Different types of polynomials may be efficently use for different types of "disorder"

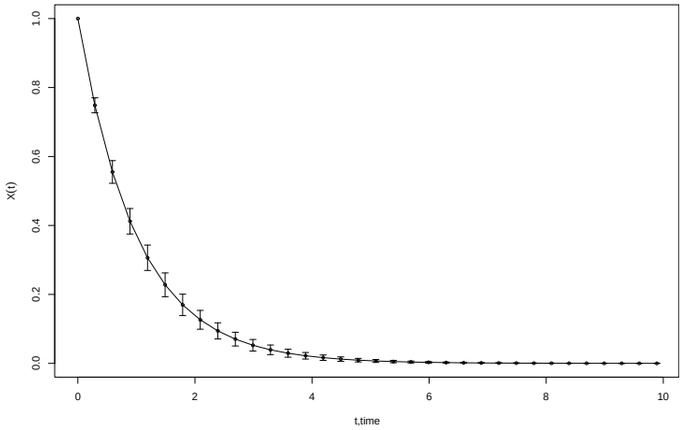
Simple Example

Consider: $\mathsf{X} \xrightarrow{k} \emptyset$ or $dx/dt = -kx$.
Gaussian disorder: $k(\xi) = k_0 H_0(\xi) + k_0 \sigma H_1(\xi), \quad x(t,\xi) \approx \sum_{n=0}^{N} x_n H_n(\xi)$
Results in the coupled chain of equations (all $x_n = 0$ for $n > N$):

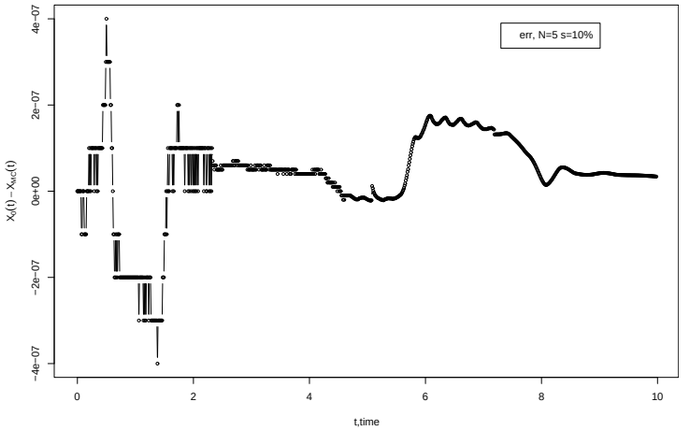
$$\dot{x}_0 = -k_0 x_0 - k_0 \sigma x_1 \tag{62a}$$
$$\dot{x}_1 = -k_0 x_1 - k_0 \sigma x_0 - 2\sigma x_2 \tag{62b}$$
$$\dot{x}_2 = -k_0 x_2 - k_0 \sigma x_1 - 3\sigma x_3 \tag{62c}$$
$$\ldots \tag{62d}$$



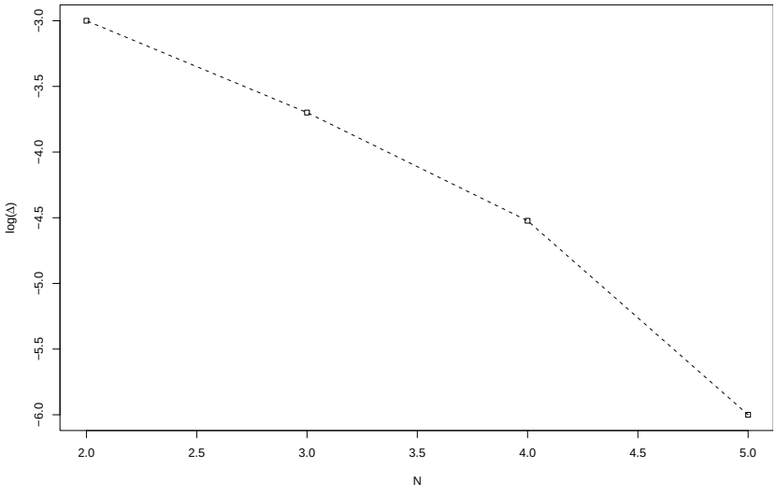
Example Cont'd



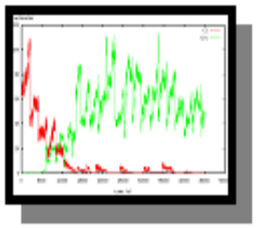
Trajectory $X_0(t)$ and uncertainty $varX(t)$



Error between MC result and $x_0(t)$



Log-Error $log(\Delta)$ vs expansion order N.



Stochastic Setting

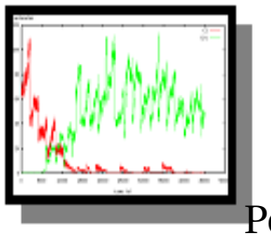
- In stochastic setting $\mathbf{X}(t)$ is a random variable for every moment of time t, hence joint pdf $P(\mathbf{X}, t; \mathbf{k})$ must be investigated on parameter sensitivity

$$\sum_{i=1}^{S} \nu_{ir}^{+} X_i \underset{k_{-r}}{\overset{k_{+r}}{\rightleftharpoons}} \sum_{i=1}^{S} \nu_{ir}^{-} X_i \tag{63a}$$

$$\frac{\partial P(\mathbf{X}, t)}{\partial t} = \sum_{r} a_r(\mathbf{X} - \boldsymbol{\nu}_r) P(\mathbf{X} - \boldsymbol{\nu}_r, t) - \tag{63b}$$

$$- P(\mathbf{X}, t) \sum_{r} a_r(\mathbf{X}) \tag{63c}$$

- Development PCE schemes directly for the Chem. Master Equation could be problematic: considering ME as a linear system is hindered by huge amount of states.



Stochastic Setting

Possible solution:

- Use some parametrization of $P(\mathbf{X},t)$ e.g. for the system near non-equilibrium steady state characterized by the drift matrix $\mathbf{A}$ and diffusion matrix $\mathbf{D}$ (Linear Noise approximation):

$$P(\mathbf{X},t) \propto e^{-\frac{1}{2}(\mathbf{X}-\boldsymbol{\mu}(t))^T\boldsymbol{\Sigma}^{-1}(t)(\mathbf{X}-\boldsymbol{\mu}(t))}, \tag{64}$$

$$\dot{\boldsymbol{\mu}} = -\mathbf{A}\boldsymbol{\mu}(t), \tag{65}$$

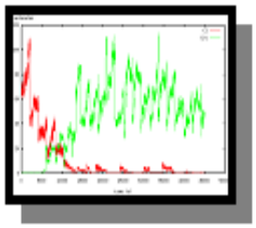
$$\dot{\boldsymbol{\Sigma}}(t) = -\mathbf{A}\boldsymbol{\Sigma}(t) - \boldsymbol{\Sigma}(t)\mathbf{A}^T + \mathbf{D}. \tag{66}$$

- Now PCE may be applied as an expansion of $\boldsymbol{\mu}(t)$ and $\boldsymbol{\Sigma}(t)$:

$$\mathbf{k}(\xi) = \mathbf{k}_0 + \mathbf{k}_1 H_1(\xi), \tag{67}$$

$$\boldsymbol{\mu}(t,\xi) \approx \sum_{n=1}^{N} \boldsymbol{\mu}_n(t) H_n(\xi) \tag{68}$$

$$\boldsymbol{\Sigma}(t,\xi) \approx \sum_{n=1}^{N} \boldsymbol{\Sigma}_n(t) H_n(\xi) \tag{69}$$



Uncertanty Propogation: Stochastic Setting

- General parametrization of $P(\mathbf{X},t)$ e.g. [Chaturvedi and Gardiner, 1979, Gilchrist et al., 1997]:

$$P(\mathbf{X},t) = \int \prod D\mu(\mathbf{q}) p(\mathbf{q},t) \underbrace{\frac{q_i^{X_i} e^{-q_i}}{X_i!}}_{\text{independent Poisson pdf}} \tag{70}$$

- Master equation can be mapped to PDE

$$\frac{\partial p(\mathbf{q},t)}{\partial t} = L(\mathbf{q}, \frac{\partial}{\partial \mathbf{q}}) p(\mathbf{q},t) \tag{71}$$

- Small noise problems/Rare Events: $p(\mathbf{q},t) \rightarrow \exp(-W(\mathbf{q}))$.
  $\mathbf{p} = \frac{\partial W}{\partial \mathbf{q}}$,

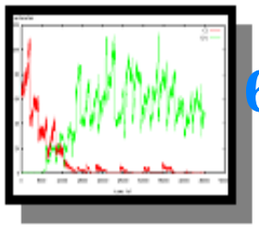
$$\dot{\mathbf{q}} = \frac{\partial H(\mathbf{q},\mathbf{p})}{\partial \mathbf{p}}, \tag{72a}$$

$$\dot{\mathbf{p}} = -\frac{\partial H(\mathbf{q},\mathbf{p})}{\partial \mathbf{q}}, \tag{72b}$$

$$H(\mathbf{q},\mathbf{p}) = \sum_r (\exp(\nu_r \mathbf{p}) - 1) a_r(\mathbf{q}) \tag{72c}$$

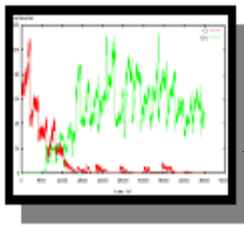
- PCE can be used to study sensitivity of the trajectories $(q(t), p(t))$.

More ot come...



6. Acknowledgments

- Prof. A. Arkin
- Prof. C. Rao
- R. Osterhout
- A. Rizvi
- Dr. M. Samoilov
- Dr. E. Alm
- T. Altman
- Flaherty, Patrick J.
- Prof. J. W. Little, U. of Arizona, Tuscon
- Prof. D. Chandler, Center for Theoretical Chemistry, Chem. Department
- Dr. M.O. Vlad, Stanford University



References

A.P. Arkin, John Ross, and H. H. McAdams. Stochastic kinetic analysis of pathway bifurcation in phage lambda-infected Escherichia coli cells. *Genetics*, 1998.

E. Aurell and K. Sneppen. "Epigenetics as a first exit problem". *Phys. Rev. Lett.*, 88, 2002.

Erik Aurell, Stanley Brown, Johan Johanson, and Kim Sneppen. "Stability puzzles in phage λ". *Physical Review E*, 65, 2002.

C.E. Bell and M. Lewis. Crystal Structure of the λ repressor C-terminal domain. *J. Mol. Biol.*, 314:1127–1136, 2001.

Bruce J. Berne, G. Ciccotti, and David F. Coker, editors. *A story of Rare Events: From barriers, to electrons and unknown pathways*, 1997.

Peter G. Bolhuis. Transition-path sampling of β-hairpin folding. *PNAS*, 100(21):12129–12134, 2003.

A.B. Bortz, M.H. Kalos, and J.L. Lebowitz. New algorithm for Monte-Carlo simulation of Ising spin systems. *J. Comput. Phys.*, 17:10, 1975.

R. Bundschuh, F. Hayot, and C. Jayaprakash. Fluctuations and Slow Variables in Genetic Networks. *Biophysical Journal*, 84:1606–1615, 2003.

R. Cameron and W. Martin. The orthogonal development of non-linear functionals in series of fourier-hermit functionals. *Ann. of Math.*, 48, 1947.

S. Chaturvedi and C. W. Gardiner. Time-convolutionless projection operator formalism for elemination of fast variables.Application to the Brownian motion. *J. Stat. Phys.*, 18, 1979.

A. Chorin. "gaussian fields and random flow". *J. Fluid Mech.*, 85:325–347, 1974.

Paul J. Darling, Jo M. Holt, and Gary K. Ackers. "Coupled energetics of the λ cro repressor self-assembly and site-specific dna operator binding: Cooperative interaction of cro dimers". *JMB*, 302:625–638, 2000.

C. Dellago, P. Bolhuis, F. S. Csajka, and D. Chandler. Transition path sampling and the calculation of rate constants. *J. Chem. Phys.*, 108:1964, 1998.

C. Dellago, P.G. Bolhuis, and P.L. Geissler. Transition path sampling. *Adv. Chem. Phys.*, 123(1), 2002.

I.B. Dodd, A.J. Perkins, D. Tsemitsidis, and J.B. Egan. "Octamerization of lambda CI repressor is needed for effective repression of p(rm) and efficient switching from lysogeny.". *Genes Dev.*, pages 3013–22, Nov. 2001.

M.I. Freidlin and A.D. Wentzel. *Random Perturbations of Dynamical Systems*. Springer,New York, 1984.

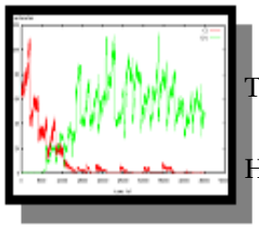
Roger G. Ghanem and Pol D. Spanos. *Stochastic finite elements: a spectral approach*. Springer-Verlag New York, Inc., New York, NY, 1996.

A. Gilchrist, C.W. Gardiner, and P.D. Drummond. Positive P-representation: Application and Validity. *Physical Review A*, pages 3014–3032, 1997.

D.T. Gillespie. Exact Stochastic Simulation of Coupled Chemical Reactions. *J. Phys. Chem.*, 81:2340–2361, 1977.

Michael F. Hagan, Aaron R. Dinner, David Chandler, and Arup K. Chakraborty. Atomistic understanding of kinetic pathways for single base-pair binding and unbinding in DNA. *PNAS*, 100:13922–13927, 2003.

P. Hänggi, P. Talkner, and M. Borkovec. Reaction rate theory: Fifty years after Kramers. *Rev. Mod. Phys.*, 62:251–342, 1990.



Thomas B. Kepler and Timothy C. Elston. Stochasticity in Transcriptional Regulation: Origins, Consequences and mathematical representations. *Biophys J.*, 81(6):3116–3136, 2001.

H. Kramers. Brownian motion in the field force and the diffusion model of chemical reactions. *Physica*, 7:284, 1940.

John W. Little, Donald P. Shepley, and David W. Wert. "Robustness of a gene regulatory circuit". *The EMBO Journal*, 18(15):4299–4307, 1999.

G. Nicolis and I. Prigogine. *Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order through Fluctuations*. Wiley Ltd., 1977.

Lawrence R. Pratt. A statistical method for identifying transition states in high dimensional problems. *J. Chem. Phys.*, 85:5045, 1986.

M. Ptashne. *"A Genetic Switch: Phage λ and Higher Organisms"*. Blackwell Scientific Publications and Cell Press, 1992.

Ravi Radhakrishnan and Tamar Schlick. Orchestration of cooperative events in dna synthesis and repair mechanism unraveled by transition path sampling of dna polymerase β's closing. *PNAS*, 101(16):5970–5975, 2004.

C. V. Rao and A. P. Arkin. Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the gillespie algorithm. *Journal of Chemical Physics*, 118(11):4999–5010, 2003.

K. Rippe, P. von Hippel, and J. Langowski. Action in the distance: Dna-looping and initiation of transcription. *Trends in Biochem. Sciences*, 20:500–506, 1995.

M.A. Shea and G.K. Ackers. "The O_R control system of bacteriophage lambda: A physical-chemical model for gene regulation". *J. Mol. Biol.*, 181:211–230, 1985.

T. Shibata. Fluctuating Reaction Rates and their Application to the problems of Gene Expression. *Phys. Rev. E*, 67: 061906, 2003.

I. M. Sokolov. Cyclization of a Polymer: First-Passage Problem for a non-Markovian Process. *Phys. Rev. Lett*, 90: 080601–080605, 2003.

A. Szabo, K. Schulten, and Z. Schulten. *J. Chem. Phys.*, 72:4350, 1980.

P. Rein tenWolde and David Chandler. Drying-induced hydrophobic polymer collapse. *PNAS*, pages 6539–6543, 2002.

Z. Toman, C. Dambly-Chaudire, L. Tenenbaum, and V. Radman. "a system for detection of genetic and epigenetic alterations in *escherichia coli* induced by dna-damaging agents". *J. Mol. Biology*, 186:97, 1985.

Titus S. van Erp, Daniele Moroni, and Peter G. Bolhuis. A novel path sampling method for the calculation of rate constants. *J. Chem. Phys.*, 118(17), 2003.

N. Wiener. "the homogeneous chaos". *Am. J. Math.*, 60:897–936, 1938.