Coarse-graining biochemical complexity

Ilya Nemenman
CCS-3/CNLS LANL
with

Nikolai Sinitsyn (CNLS/CCS-3 LANL)

Center for Nonlinear Studies

The First q-bio Conference on Cellular Information Processing

August 8-11, 2007 | Santa Fe, New Mexico, USA

http://cnls.lanl.gov/q-bio
q-bio@cnls.lanl.gov

Speakers Include:

Adam P. Arkin
Lawrence Berkeley National Laboratory

First q-bio Conference on Cellular Information Processi
This conference is intended to advance predictive mo
genetic regulatory systems. The emphasis is
experimentation for the purposes of understanding
particular regulatory systems and of elucidating gene
information processing.

The single-track program will include invited talks
theoretical researchers as well as shorter talks, pos
demonstrations selected from contributed submissio
banquets, six sessions covering a range of topics, a
sessions.

There will be an opportunity for selected participants
presentations made at the conference to a special i
journal indexed by ISI and PubMed.

Lodging is available for participants on the campus of St. John's College, which should facilitate interactions and stimulating informal discussions about quantitative biology. Space is available for 200 participants. In the event that registration demand exceeds capacity, preference will be given to individuals selected to present contributed talks or posters. Abstracts should be submitted for review via the conference web site.

First q-bio Summer School on Cellular Information Processing
This school is designed for researchers new to modeling cellular regulatory systems. It will take place in Los Alamos from July 23 to August 7. Participants will attend daily lectures about signal transduction, gene regulation and stochastic effects in biochemical networks and work in small teams on selected research projects. Tuition includes conference registration.

Deadlines:

Abstract submission	April 15, 2007
Summer School registration	April 15, 2007
Early registration	June 1, 2007

*Travel awards for graduate students and postdocs may be available.
More information and applications are available on the conference website.

Organizing Committee: Jeremy S. Edwards (University of New Mexico); James R. Faeder, William S. Hlavacek, Yi Jiang, Ilya Nemenman, and Michael E. Wall (Los Alamos National Laboratory).

Advisory Committee: William Bialek (Princeton University); Byron Goldstein, John E. Pearson, William H. Press, David H. Sharp, and Pat J. Unkefer (Los Alamos National Lab); Michael A. Savageau (University of California, Davis)

Los Alamos NATIONAL LABORATORY

Center for Nonlinear Studies

Images are of St. John's College, the conference location

Mark Goulian
University of Pennsylvania
Jeff Hasty
University of California, San Diego
Jason M. Haugh
North Carolina State University
Alexander Hoffmann
University of California, San Diego
Ravi Iyengar
Mount Sinai School of Medicine
Boris N. Kholodenko
Thomas Jefferson University
Andre Levchenko
Johns Hopkins University
Wendell A. Lim
University of California, San Francisco
Rune Linding
Samuel Lunenfeld Research Institute
John W. Little
University of Arizona
Leslie M. Loew
University of Connecticut Health Center
Bernhard Ø. Palsson
University of California, San Diego
Johan Paulsson
Harvard Medical School
John Reinitz

Deadlines:

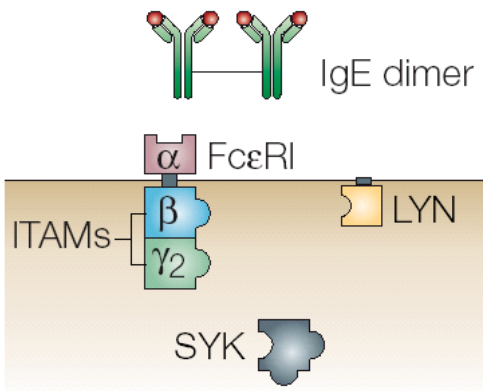
Abstract submission	April 15, 2007
Summer School registration	April 15, 2007
Early registration	June 1, 2007

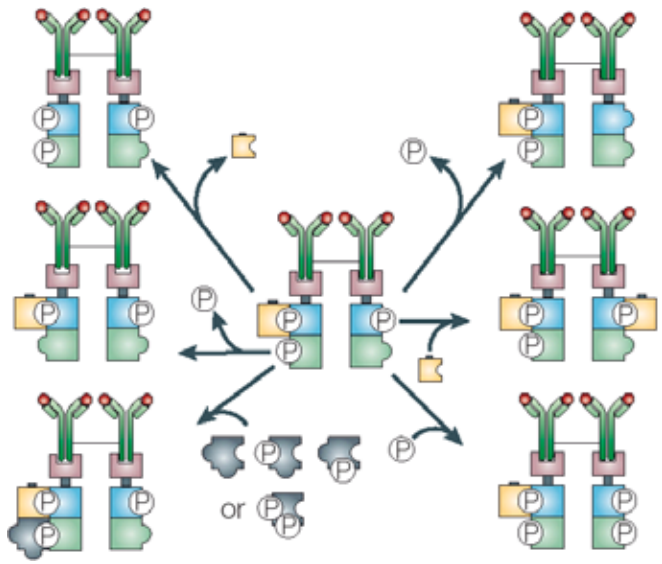
*Travel awards for graduate students and postdocs may be available.
More information and applications are available on the conference website.

Topic

- Animal learning, multiscale or power-law memory: *world is complex*
- *Predictability, complexity, and learning.* NeCo 2001: *what is complexity (of a time series)?*
- Coarse graining biochemical networks: *how to deal with complexity?*

Biochemical complexity:
Example - IgE receptor (From Faeder, Hlavacek, et al.)

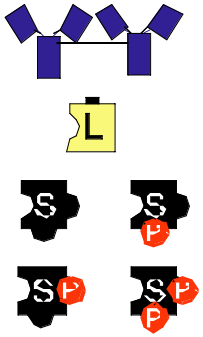




354 species / 3680 reactions

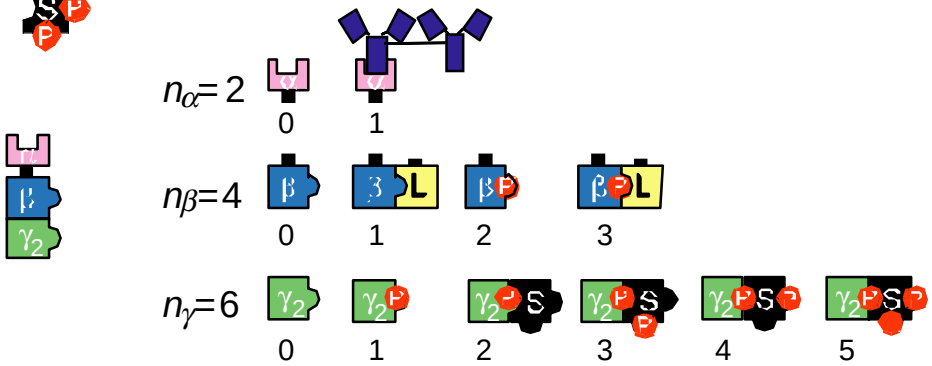
Why such complexity?

6 free states



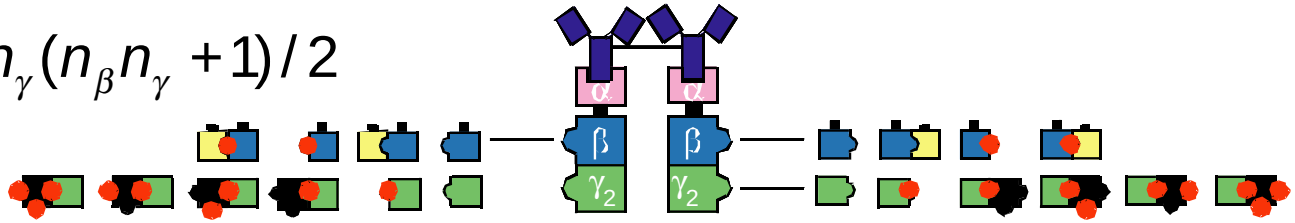
48 monomer states

$$n_\alpha n_\beta n_\gamma$$



300 dimer states

$$n_\beta n_\gamma (n_\beta n_\gamma + 1)/2$$



354 chemical species (*2954 for trimers*)

10+ reactions/species

(an example with a relatively short RHS)

- XDOT(1)= (1.0*km1*X(8)*X(0)-2.0*kp1*X(7)*X(1))/1.0
  +(1.0*km1*X(10)*X(0)-2.0*kp1*X(9)*X(1))/1.0
  +(1.0*km1*X(28)*X(0)-2.0*kp1*X(33)*X(1))/1.0
  +(1.0*km1*X(35)*X(0)-2.0*kp1*X(17)*X(1))/1.0
  +(1.0*km1*X(40)*X(0)-2.0*kp1*X(36)*X(1))/1.0
  +(1.0*km1*X(43)*X(0)-2.0*kp1*X(37)*X(1))/1.0
  +(1.0*km1*X(46)*X(0)-2.0*kp1*X(38)*X(1))/1.0
  +(1.0*km1*X(49)*X(0)-2.0*kp1*X(39)*X(1))/1.0
  +(1.0*km1*X(56)*X(0)-2.0*kp1*X(55)*X(1))/1.0
  +(1.0*km1*X(60)*X(0)-2.0*kp1*X(117)*X(1))/1.0
  +(1.0*km1*X(66)*X(0)-2.0*kp1*X(24)*X(1))/1.0
  +(1.0*km1*X(67)*X(0)-2.0*kp1*X(77)*X(1))/1.0
  +(1.0*km1*X(68)*X(0)-2.0*kp1*X(72)*X(1))/1.0
  +(1.0*km1*X(69)*X(0)-2.0*kp1*X(78)*X(1))/1.0
  +(1.0*km1*X(70)*X(0)-2.0*kp1*X(75)*X(1))/1.0
  …

And, on top of this, everything is stochastic and dynamic!

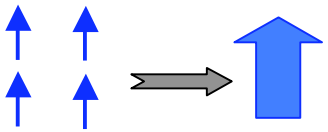
What to do?

- Coarse graining! *out = f(in)*; $x_{last} = f(x_{first})$
- Already are doing this (in deterministic context)

$$\frac{d[\mathrm{lacZ}]}{dt} = \frac{V_0}{K_0^2 + [\mathrm{lacI}]^2}$$

- Is this legitimate?
  - Is the functional form correct?
  - Are these events Poisson?
- How can simulations be done?
  - Simple SSA-Gillespie won't work (though recall Golding's talk)

Which coarse-graining method to use?



- Combining nodes
  - how?
- Fast rates vs. slow rates
  - Rates concentration dependent
  - May couple very different species types
- Momentum space RG
  - Does not decrease # of nodes
- Fast nodes vs. slow nodes
  - All couples, all same speed
- High abundance (relatively slow) vs. Low abundance (relatively fast): adiabatic approximation
  - That's what biochemists have been using
  - Stochasticity?

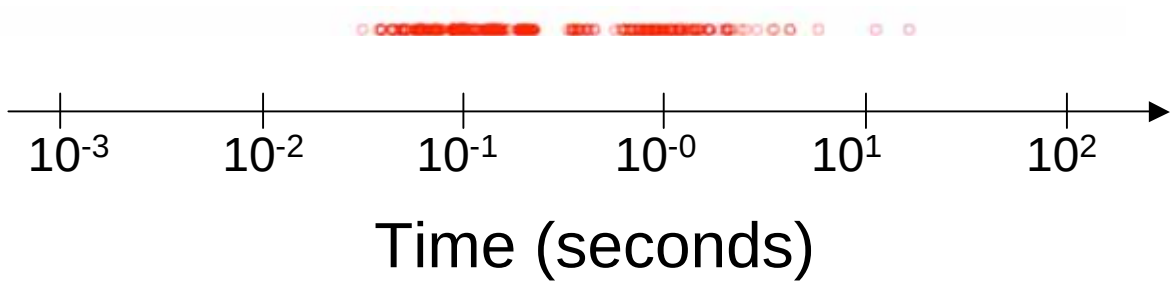
Why adiabaticity?
(Kozdon, Faeder)

FcεRI (trimer)
2954 states

Relaxation time scales of different species

FcεRI (dimer)
354 states

EGFR
356 states



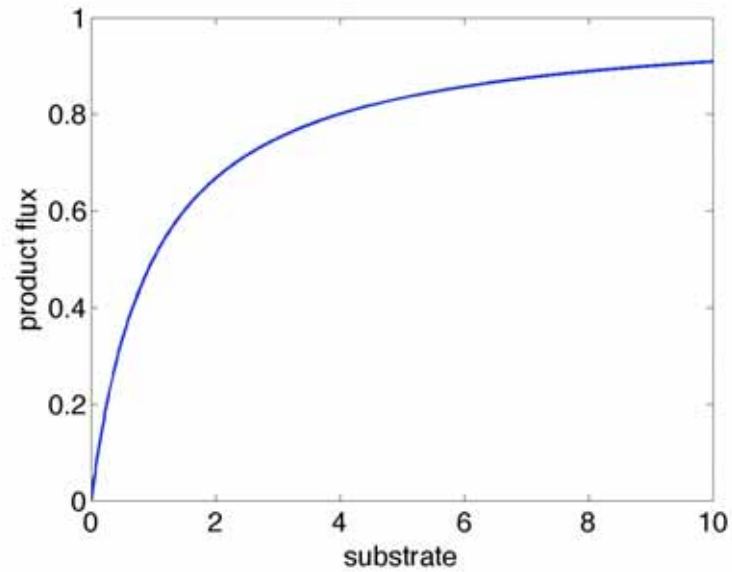
Michaelis-Menten reaction: Deterministic coarse-graining

Slow modulation

$$S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} SE \xrightarrow{k_2} P$$

$$\frac{d[SE]}{dt} = k_1[S][E] - (k_{-1} + k_2)[SE] = 0$$

$$\frac{dP}{dt} = \frac{k_1 k_2 [E][S]}{k_2 + k_{-1} + k_1[S]} = J_{cl}$$

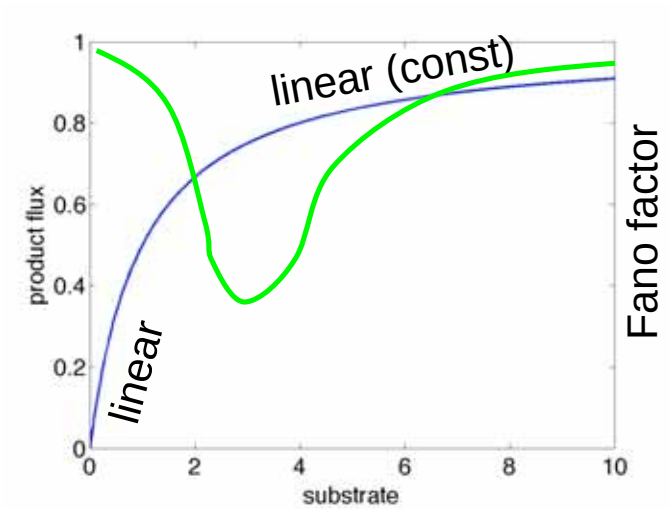


- Adiabatic approximation
  - Many enzyme turnovers for small fractional change in [*P*], [*S*]
- How to do coarse-graining with fluctuations?

MM with fluctuations

(Hwa, Bundschuh, Vanden-Eijnden, Ehrenberg, Szabo, Arkin, et al.)

- Mean = deterministic
- Var = mean for linear regimes (one step dominated)

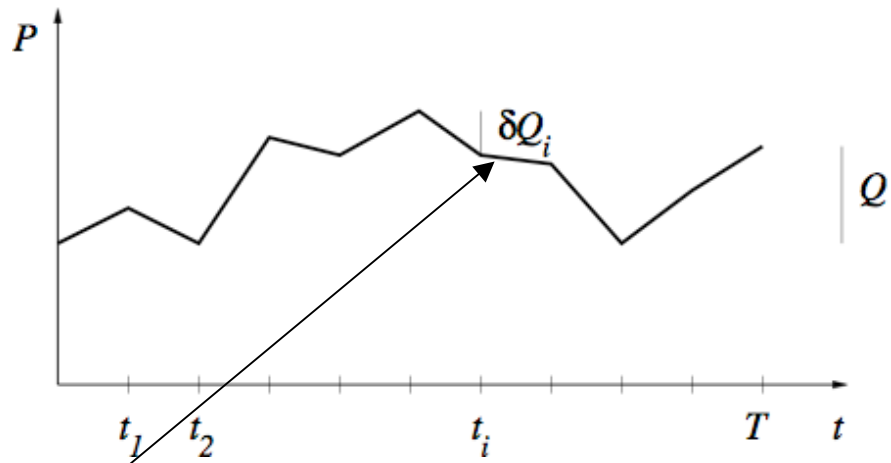


- Is first statement correct? What about the bend area for the second?

Michaelis-Menten reaction (or a pore): Stochastic coarse-graining

$$S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} SE \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} P$$

$$Q = P(T) - P(0)$$



4 Poisson processes
with (almost) constant rates k_i

$$P(Q) = \int d\delta Q_1 \int d\delta Q_2 \ldots \int d\delta Q_{T/\delta t} \prod P(\delta Q_i)\,\delta\left(Q - \sum \delta Q_i\right)$$

Functional integral over all paths - can get full MGF

(Simper version of Sinitsyn and Nemenman, 2007)

Adiabatic approximation

$$P(Q) \sim \int \prod [dN][d\chi] \exp\left[i\chi \dot{N} + N_E H \right]$$

occupied enzymes

Lagrange multiplier

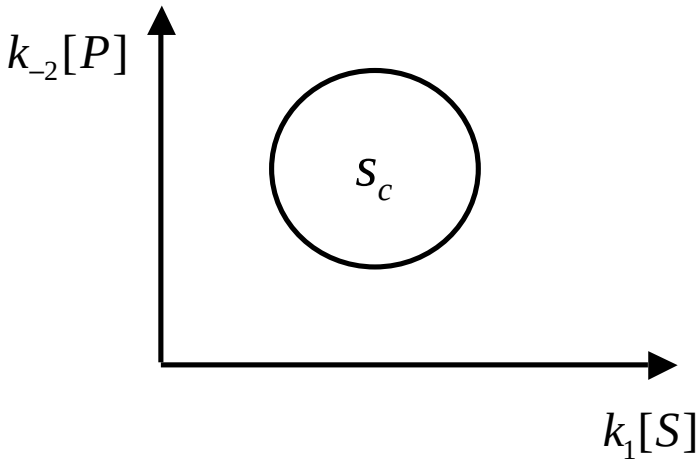
enzymes

Saddle point solution (exact due to linearity of *S*)

$$\frac{\partial H(N,\chi)}{\partial \chi_{cl}} = -i\dot{N}_{cl} = 0$$

Adiabatic solution

Michaelis-Menten reaction: Periodic modulation of two rates

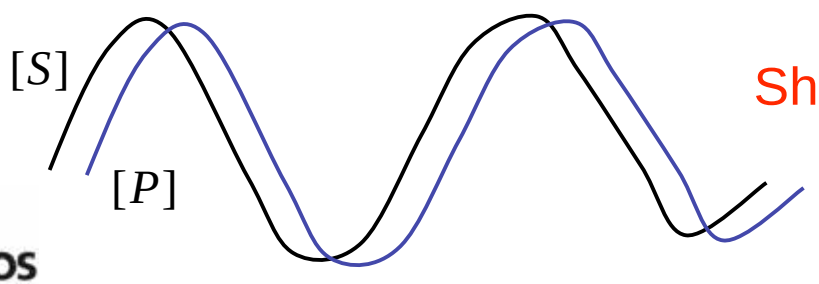


$$\mathrm{MGF} = \mathrm{MGF}_{\mathrm{cl}} + \iint_{s_c} dS dP \; F$$

Berry curvature
As in adiabatic QM

$$J = J_{cl} + J_{pump} = J_{cl} + \iint_{s_c} d^2k \frac{k_2 + k_{-1}}{T_0 \left(\sum k_i\right)^3}$$

Pump, ratchet

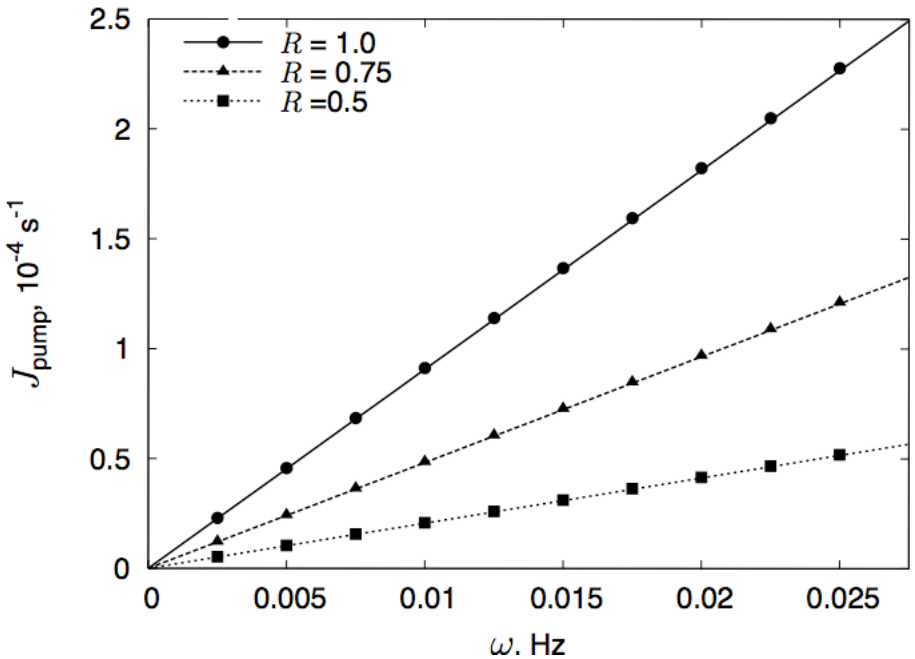


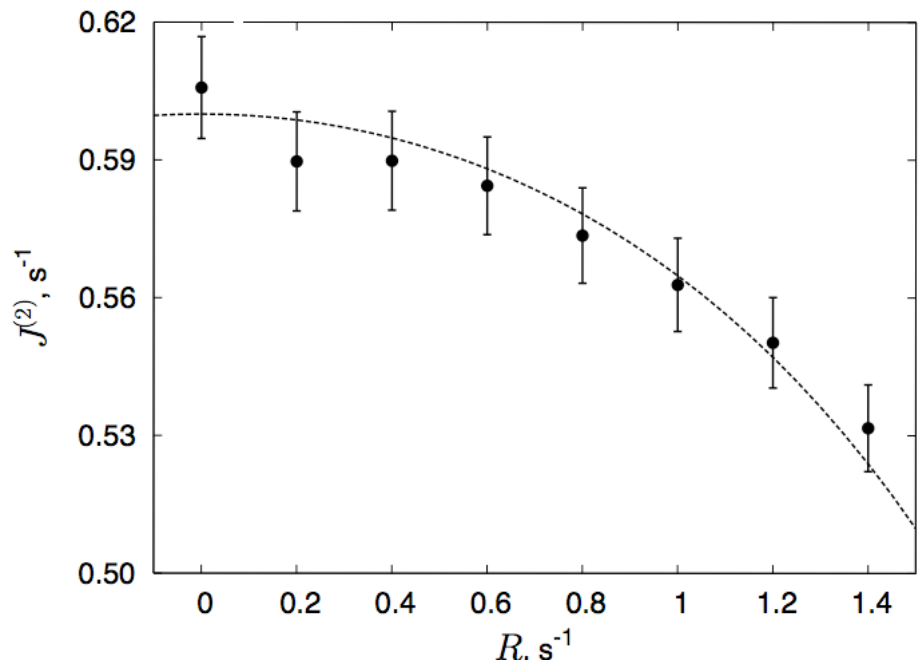
Shielding

Example 1: Bulk fluxes

$$k_1 = 1.5 + R\cos\omega t; \qquad k_{-2} = 1.5 + R\sin\omega t; \qquad k_{-1} = k_2 = 1$$

equilibrium, on average: $J_{cl} = 0$



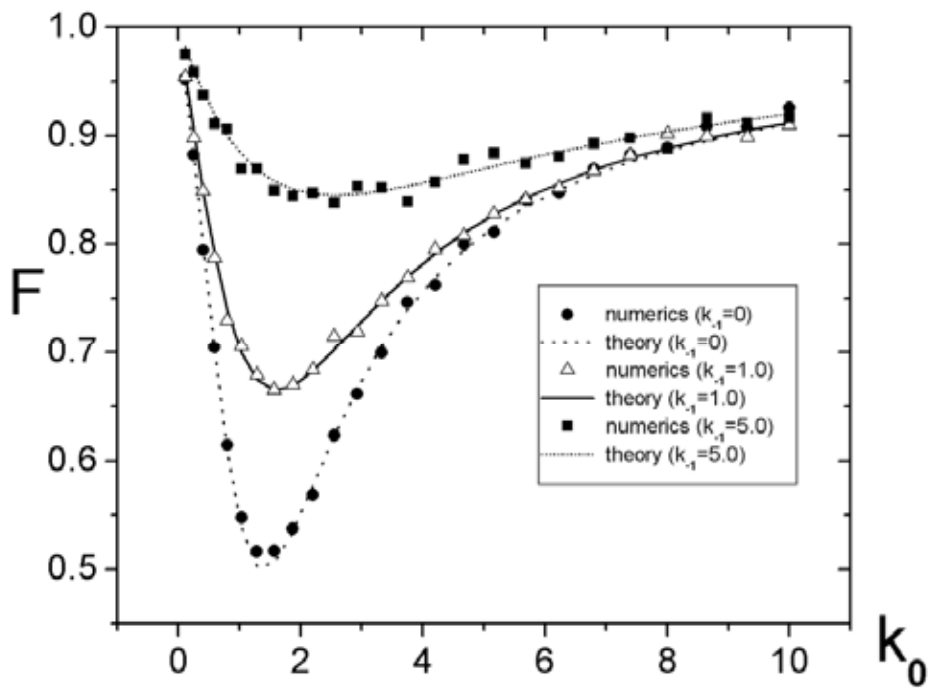


Pump current up to 10% for realistic enzymes

Example 2:
Noise in single molecule experiments

$$\varnothing \underset{q}{\overset{k_0}{\rightleftharpoons}} S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} SE \xrightarrow{k_2} P$$

Xie et al.
Bezrukov et al.



Example 3:
Nonperiodic correction to MM rate

$$S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} SE \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} P$$

$$\dot{N}_p = J^{cl}_{S\to P} - J^{cl}_{P\to S} + \boxed{\left(k_2 + k_{-1}\right)\frac{\left(k_1\dot{N}_S + k_{-2}\dot{N}_P\right)\left(k_2 + k_{-2}N_P\right)}{\left(k_1 N_S + k_{-1} + k_2 + k_{-2}N_P\right)^3}}$$

Conclusions

- Adiabatic coarse-graining of stochastic biochemical networks
- Nonzero mean corrections (pump effects) -- geometric nature
- Nonpoisson statistics
- Developing symbolic package for coarse graining (to be built into BioNetGen -- network simulation package from LANL, NAU, and now Pitt)