Computing structural and dynamic properties of biological systems at multiscale

3. Stochastic networks

Jie Liang 梁杰

Dept. of Bioengineering
University of Illinois at Chicago

Outline

1. Chemical master equation and approximations

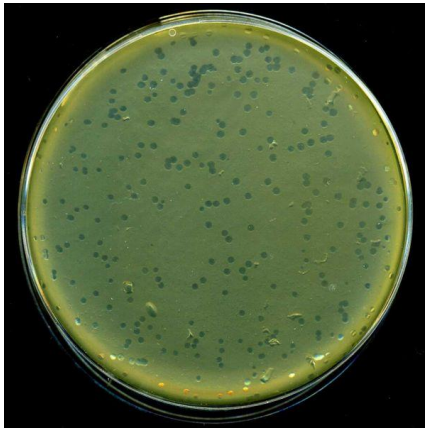
2. Finite buffer algorithm for optimal state enumeration

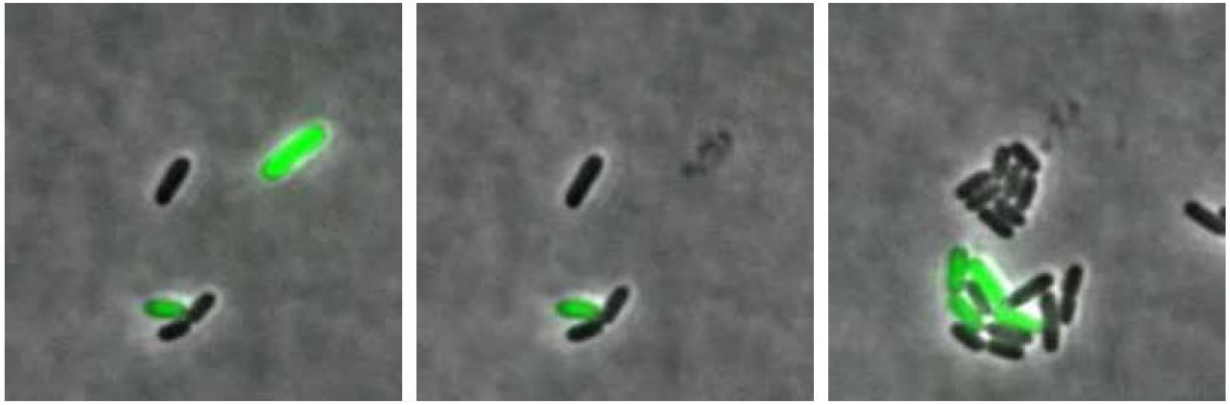
3. How do we know if results are correct and error estimation
   - Probability flux

4. Results of simple and complex stochastic networks
   - Comparison with other method

Stochastic Genetic circuits

- **How is cell fate determined?**





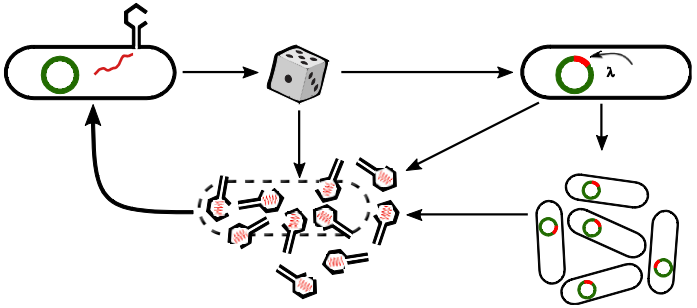
(St-Pierre and Endy, *PNAS*, 2008)

Implication for cancer biology

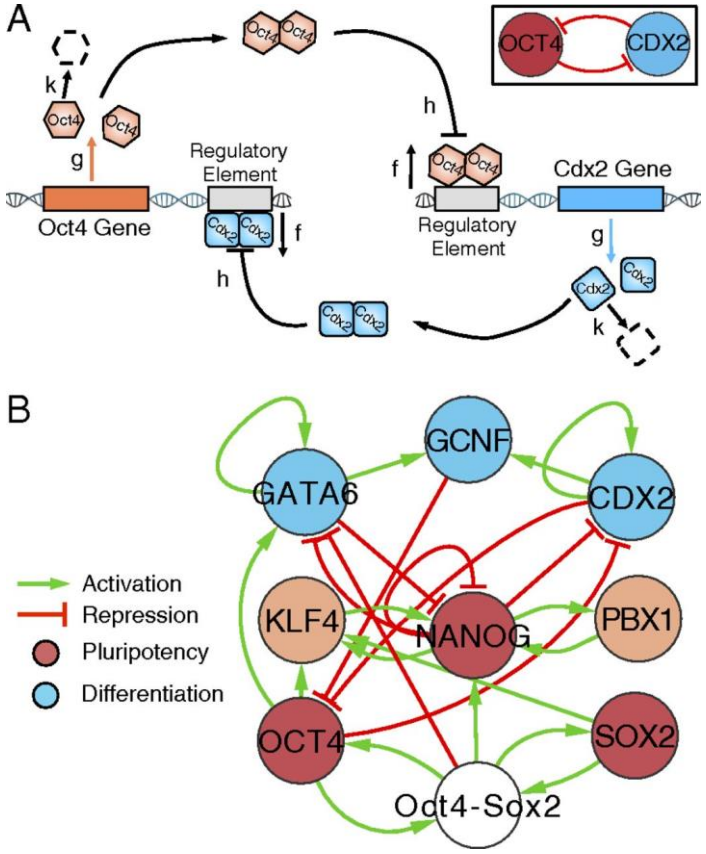
Stochastic network and cellular fate

- Origin of stochasticity:
  - Many molecular networks are of small copy numbers (eg. μM – nM) and are stochastic:
    - transcription regulation, protein synthesis, signal transduction.
  - Slow reactions: multistable systems
  - Burst in transcription and translation
  - Cell-cell variaton
  - External environment

- Epigenetic Landscape of stochastic networks: Broad applications
  - Phage lambda
  - Stem cell differentiation



(Phage work: McAdams and Arkin, 1997; Aurell and Sneppen, 2002; Zhu, Yin, Hood, and Ao, 2004)



(Zhang and Wolyness, 2014)

Chemical Master Equation

- A general framework for studying biological networks
  - Assumption: well-mixed system

Molecular reactions, state, and probability

- Molecular species and reactions:
  - *m* molecular species: $\{X_1, \ldots, X_m\}$
    - X_i: i-th molecular species
  - *n* chemical reactions: $\mathcal{R} = \{R_1, \ldots, R_n\}$

- Microstate:
  - combination of copy numbers at time t:

    $$\mathbf{x}(t) = (x_1(t), \ldots , x_m(t)) \in \mathbb{N}^m$$

  - probability of system in state $\mathbf{x}(t)$:

    $$p(\mathbf{x}(t))$$

State Space and Probability Landscape

- State space $\mathcal{X}$:
  - The set of all possible combination of copy numbers

    $$\mathcal{X} = \{ \mathbf{x}(t) \mid t \in (0, \infty)\},\ \text{discrete}$$
  - Size of state space: $|\mathcal{X}|$, countable

- Probability landscape at time t :

  $$\{p\ (\mathbf{x}(t) \mid (\mathbf{x}(t) \in \mathcal{X}\ \}$$

Chemical Reactions: Stoichiometry and Reaction Rate

- Chemical reaction r_k has the form:

$$c_1(k)\, X_1 + c_2(k)\, X_2 + \ldots + c_m(k)\, X_m \rightarrow c_1'(k)\, X_1 + c_2'(k)\, X_2 + \ldots + c_m'(k)\, X_m.$$

  - Brings system from state $\mathbf{x}_i$ to $\mathbf{x}_j$: A jump process

- Reaction rate: determined by intrinsic rate constant r_k and copy numbers of reactants, which are given by the state $\mathbf{x}_i$

$$A_k(\mathbf{x}_i, \mathbf{x}_j) = A_k(\mathbf{x}_i, \ldots) = A_k(\mathbf{x}_i) = r_k \prod_{l=1}^{m} \binom{x_l}{c_l}$$

  - If a reaction connects $\mathbf{x}_i$ and $\mathbf{x}_j$, $A_k(\mathbf{x}_i, \mathbf{x}_j) > 0$,
    - Else $A_k(\mathbf{x}_i, \mathbf{x}_j) = 0$
  - If more than one reaction connects $\mathbf{x}_i$ and $\mathbf{x}_j$,

$$A(\mathbf{x}_i, \mathbf{x}_j) = \sum A_k(\mathbf{x}_i, \mathbf{x}_j)$$

Discrete Chemical Master Equation

- A foundation framework for studying mesoscopic biological networks

$$d\,p(\mathbf{x}, t)/\,dt = \sum_{x'} [\, A(\mathbf{x}', \mathbf{x})\, p\,(\mathbf{x}', t) - A(\mathbf{x}, \mathbf{x}')\, p\,(\mathbf{x}, t)]$$

  - $p(\mathbf{x}(t))$ is of continuous time
  - **States are discrete**

- Matrix form: let $\mathbf{A} = \{\, \mathbf{A}(x_i, x_j)\}$, $\mathbf{A} \in \mathbb{R}^{|\mathcal{X}| \times |\mathcal{X}|}$

$$d\,p(\mathbf{x}, t)/\,dt = \mathbf{A}\, p(\mathbf{x}, t)$$

Discrete CME

- Full account of probabilities of jumps:
  - Regardless whether copy numbers are small or large, jumps are large or small
  - Full stochasticity
- Can generate trajectories with correct probabilities
  - Based on the network architecture and reaction rates with enumerated states
- Little applications
  - Not feasible beyond very simple systems

Challenge: Requires full description of the discrete state space!

The Holy Grail

- Time-evolving probability landscape p(**x**, t) over all possible states
  - Full stochasticity
  - Can account for nonlinearity in reactions

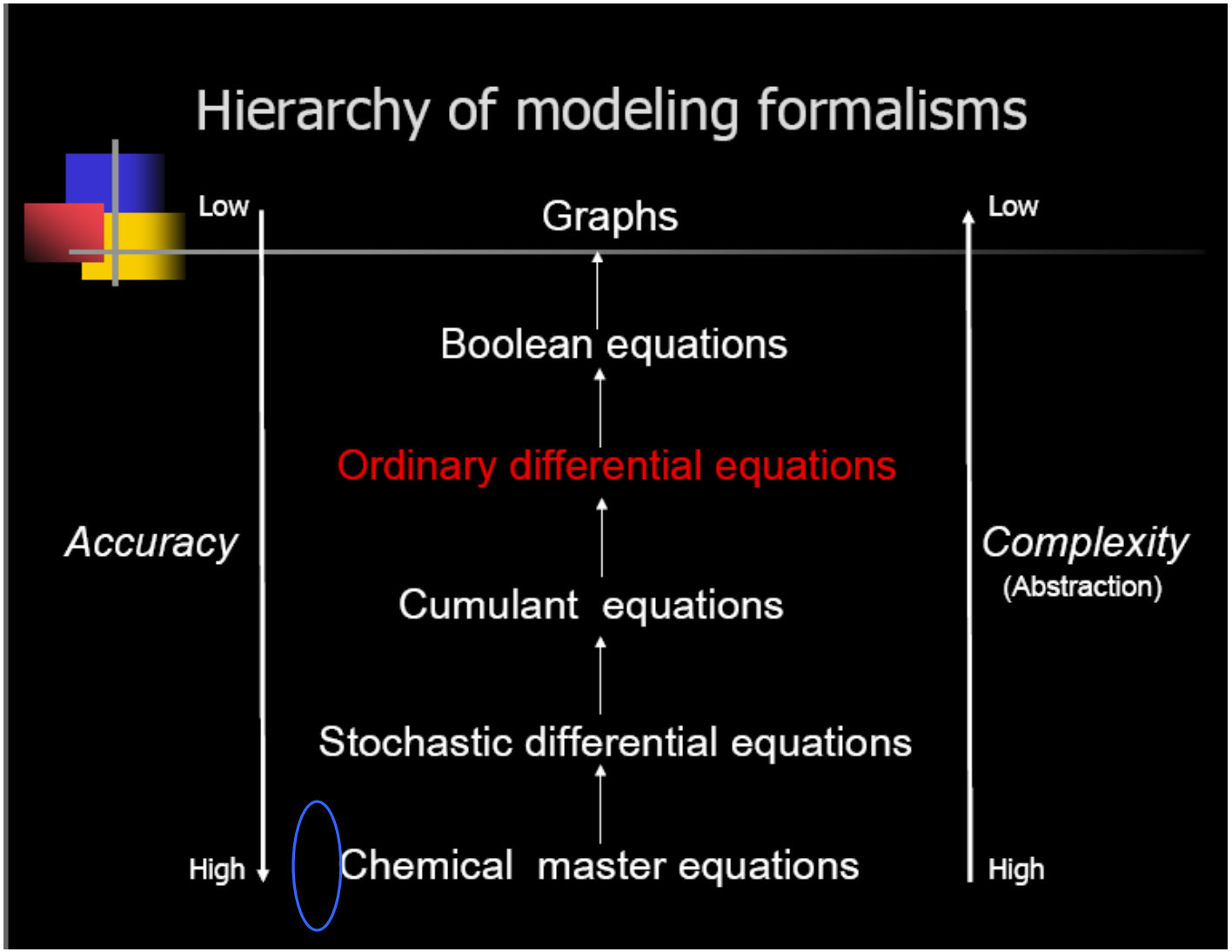
- Vector filed on the same state space
  - Discrete view: same *k*-pointed hats but of very different sizes
  - Continuous view: generic vector field and such structures may be lost.

An Analogy

(see Qian and Beard, 2008)

- Physics and chemistry:
  - Schrodinger Equation
  - Wave function or density function (or partition function in stat mech)
- Exact solution
  - Possible for systems with 100s of atoms
    - e.g. DFT.

- Systems biology:
  - Chemical Master Equation
  - Landscape probability distribution and dynamic changes
- Exact solution
  - Nothing is known about how far we can go!



(Courtesy: Prof Luonan Chen)

Solving Chemical Master Equation

- Analytical solution: not possible in general
  - Except for toy problems or with assumptions
    - e.g., scale separation, fast equilibrium for some reactions
- Direct numerical solution
  - Requires full description of the enormous discrete state space
    - Expecting explosive combinatorial size of space.
  - Not feasible beyond simple systems.

(van Kampen, 1992)

Approximating discrete chemical master equation

- Finite State Projection:
  - Truncating the full state space
  - With error estimation: accuracy certificate

    (Munsky and Khammash, *J Chem Phys*, 2006)
  - Difficulty in choosing specific subspace and limiting its size
  - Need absorption state
  - Cannot compute the steady state probability landscape
    - Leakage accumulates

Stochastic Simulation Algorithm

- Generating reaction trajectories through Monte Carlo
- But follows high probability events (Gillespie, 1977)
  - Very inefficient in sampling rare events.
- Difficult to assess convergency
- Recent work: biased Monte Carlo for rare events:
  - wSSA (Kuwahara & Mura, 2008), dwSSA (Daigle et al, 2011), swSSA (Roh et al, 2011)
  - Adaptively Biased Sequential Importance Sampling (ABSIS)

(Cao and Liang, J Chem Phys, 2013, 139(2):025101)

Approximating dCME with Continuous Chemical Master Equation

- Regarding **state space as continuous**,
- Assumption:
  - Difference in the amount of molecules in neighboring states are infinitesimally small:

$$\partial\, p(\mathbf{x}, t)/\partial\, t = \int_{\mathbf{x}'} [\, A(\mathbf{x}', \mathbf{x})\, p(\mathbf{x}', t) - A(\mathbf{x}, \mathbf{x}')\, p(\mathbf{x}, t)]\, d\,\mathbf{x}'$$

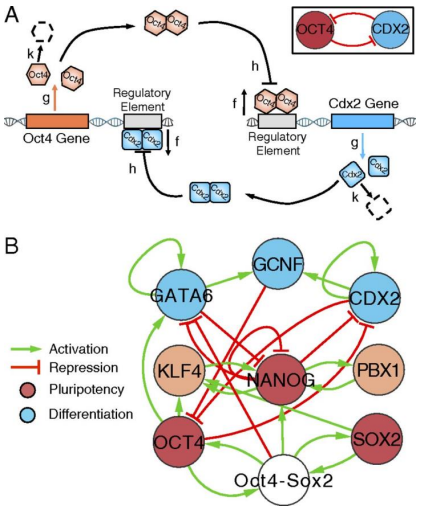
Not valid when only a handful of molecules !

Approximating continuous chemical master equation

- Fokker-Planck Equation: Approximating continuous CME
  - Neighboring states have small difference
  - Further assumptions: (Munsky and Khammash, *J Chem Phys*, 2006)
    - Jumps are small: $|x_i - x_j| < \varepsilon$
    - Transition probability vary slowly: $A(x, y) \approx A(x + \varepsilon_1, y + \varepsilon_2)$
    - Probability vary slowly: $p(\mathbf{x}, t) \approx p(\mathbf{x} + \varepsilon, t)$
  - Replace the jump operator A(x', x) from x' to x:
    - with a differential operator
    - 2nd order Taylor expansion of the differential operator with higher order truncation
- Other Approximation
  - Stochastic differential equation: Langevin equation

Errors due to these approximations are generally not known!

General Approach



- Identify a significant biological problem
- Construct a reasonable stochastic network

- Compute solutions
  - Simulation using Gillespie algorithm
  - Langevin/Fokker-Planck stochastic equation
  - Analysis with Gaussian assumptions
  - Techniques from other fields such as quantum field theory

- But is the network correct? Are we finding the right solutions?

- The truth is unknown except simple cases
  - Experimental results can provide answer but too much degeneracy
    - Not for theory and algorithm development

Can mathematics help?

How to describe the discrete state space for direct solution of dCME?

- Naïve approach (Method 1)
  - B^n, where:
    - B = the maximum possible copy number of each molecular species in the system = "buffer size"
    - n = number of species.
  - e.g. # of states exponentially large:
    - eg, 16 molecular types, 30 total copies: $(30+1)^{16}=7.27\text{x}10^{23}$
  - Many such states will never be visited

- Following trajectory of states from simulation
  - No guarantee all accessible states will be visited

Needs a method that can efficiently describe accessible states

Finite Buffer dCME Method

- Optimal enumeration algorithm of state space
  - Assumes finite number of net molecules synthesized
  - Starts with a given initial condition

- Optimal in memory requirement and time complexity
  - All states reachable from an initial condition will be accounted for
  - No irrelevant states are included
  - All possible transitions will be recorded
  - No infeasible transitions will be attempted

(Cao and Liang, *BMC Systems Biology*, 2008, 2:30;
Cao et al, *PNAS*, 2010, 107(43):18445-50
Cao and Liang, *J Chem Phys*, 2013, ,139(2):025101
Cao, Terebus, and Liang, 2014, manuscript)

Algorithm 1 State Enumerator($\boldsymbol{M}, \boldsymbol{R}, B$)

Network model: $\boldsymbol{N} \leftarrow \{\boldsymbol{M}, \boldsymbol{R}\}$;
Initial condition: $\boldsymbol{s}^{t=0} \leftarrow \{c_1^0, c_2^0, \ldots, c_m^0\}$; Set the value of buffer capacity: $c_{m+1}^0 \leftarrow B$;
Initialize the state space and the set of transitions: $\mathcal{X} \leftarrow \emptyset$; $\boldsymbol{T} \leftarrow \emptyset$;
Stack $ST \leftarrow \emptyset$; Push(ST, $\boldsymbol{s}^{t=0}$); *StateGenerated* ←FALSE
while $ST \neq \emptyset$ **do**
  $\boldsymbol{s}_j \leftarrow$ Pop (ST);
  for $k = 1$ to n **do**
    if reaction R_k occurs under condition $\boldsymbol{s}_j$ **then**
      if reaction R_k is a synthetic reaction and generates u_k new molecules **then**
        $c_{m+1} \leftarrow c_{m+1} - u_k$
        if $c_{m+1} \geq 0$ **then**
          Generate state $\boldsymbol{s}(j, R_k)$ that is reached by following reaction R_k from $\boldsymbol{s}_j$;
          StateGenerated ←TRUE
        end if
      else
        if reaction R_k is a degradation reaction and breaks down u_k molecules **then**
          $c_{m+1} \leftarrow c_{m+1} + u_k$
        end if
        Generate state $\boldsymbol{s}(j, R_k)$ that is reached by following reaction R_k from $\boldsymbol{s}_j$;
        StateGenerated ←TRUE
      end if
      if (*StateGenerated* = TRUE) and ($\boldsymbol{s}(j, R_k) \notin \mathcal{X}$) **then**
        $\mathcal{X} \leftarrow \mathcal{X} \cup \boldsymbol{s}(j, R_k)$;
        Push(ST, $\boldsymbol{s}(j, R_k)$);
        $\boldsymbol{T} \leftarrow \boldsymbol{T} \cup t_{\boldsymbol{s}(j,R_k),\boldsymbol{s}_j}$;
        $a_{i,j} \leftarrow$ Transition Coefficient($\boldsymbol{s}(j, R_k)$, $\boldsymbol{s}_j$, R_k)
      end if
    end if
  end for
end while
Output $\mathcal{X}$, $\boldsymbol{T}$ and $\boldsymbol{A} = \{a_{i,j}\}$.

(Cao and Liang, *BMC Systems Biology*, 2008, 2:30)

- Works for networks of reactions with arbitrary stoichiometry

- Optimal in memory requirement and time complexity
  - All states reachable from an initial condition will be accounted for
  - No irrelevant states are included
  - All possible transitions will be recorded
  - No infeasible transitions will be attempted

Computing Steady State Distribution

- Once states are defined, can compute steady state probability landscape with mild assumptions:

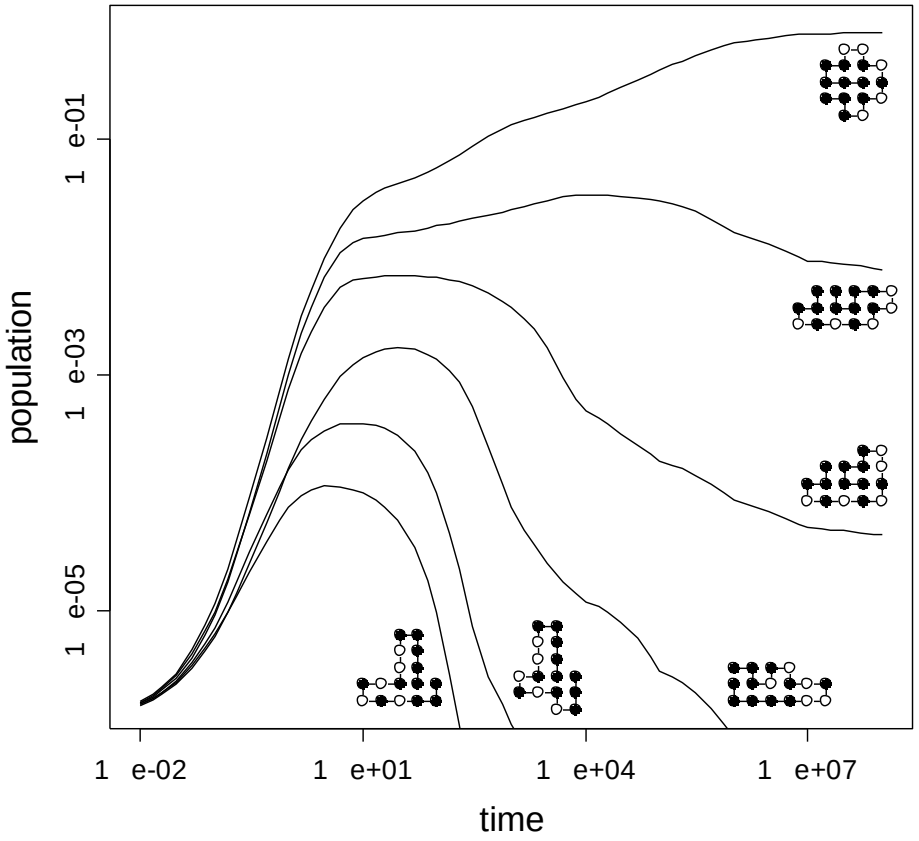
  Let $M = A + I\,\Delta t$, solve $p = M\,p$

  - Does not depend on initial condition
  - It is the eigenvector with eigenvalue = 1
  - e.g. Arnoldi method

(Cao and Liang, *BMC Systems Biology*, 2008, 2:30)

Computing Dynamics

$$d\,p(\mathbf{x}, t)/\,dt = \mathbf{A}\,p(\mathbf{x}, t), \qquad p(\mathbf{x}, t) = p(\mathbf{x}, 0)\,e^{\mathbf{A}t}$$



Can use Krylov subspace method

(Sidje, 2002)

Eg. In a study of model proteins, where time-dependent dynamics changes of probability /concentration of all states/conformations during folding process are computed.

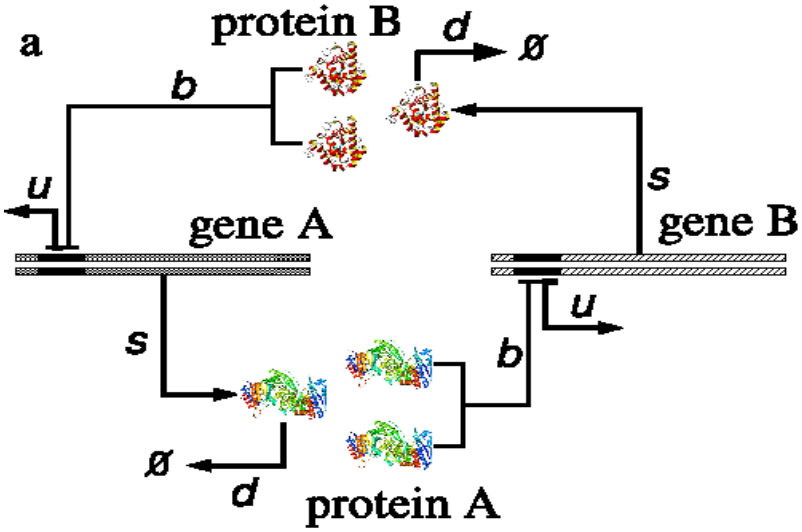
800,000 states

9-orders of magnitude

Kachalo, Lu, and Liang, *Phys Rev Lett*, 2006, 96:058105:1-4

Example: Toggle Switch

- Two genes, A and B.
  - Products repress each other.
  - One of the simplest networks with bistability
  - No general analytical solution.



b

$$Gene A \xrightarrow{s} Protein A$$

$$Gene B \xrightarrow{s} Protein B$$

$$Protein A \xrightarrow{d} \emptyset$$

$$Protein B \xrightarrow{d} \emptyset$$

$$2 \times Protein A + Gene B \xrightarrow{b} Bound Gene B$$

$$2 \times Protein B + Gene A \xrightarrow{b} Bound Gene A$$

$$Bound Gene A \xrightarrow{u} 2 \times Protein B + Gene A$$

$$Bound Gene B \xrightarrow{u} 2 \times Protein A + Gene B$$

J Collins et al, 2000, Nature

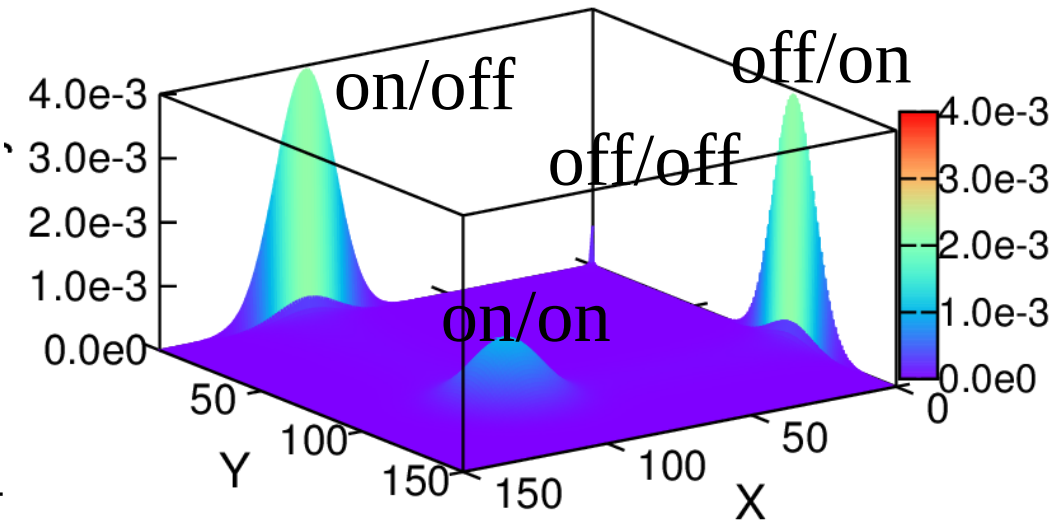
Recent approximation methods:
(Schultz, Onuchic, and Wolyness, 2007, JCP;
Kim and Wang, 2007,)

Toggle-Switch: Our results

- Size of state space:

# Prot A	# B	Size
10	10	764
50	50	19,804
100	100	79,604
150	150	179,404
200	200	319,204
250	250	499,004
300	300	718,804
350	350	978,604
400	400	1,278,404

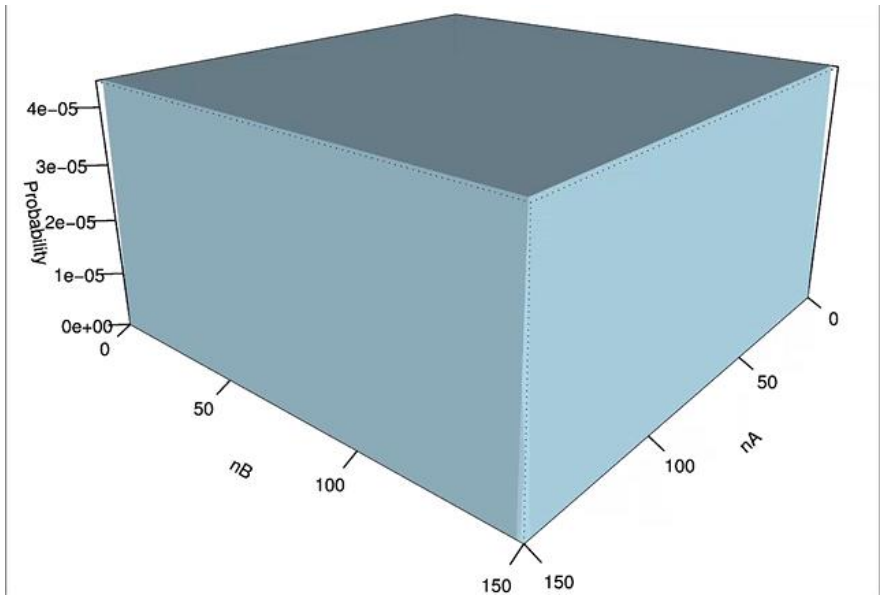
- Exact landscape of steady state probability:



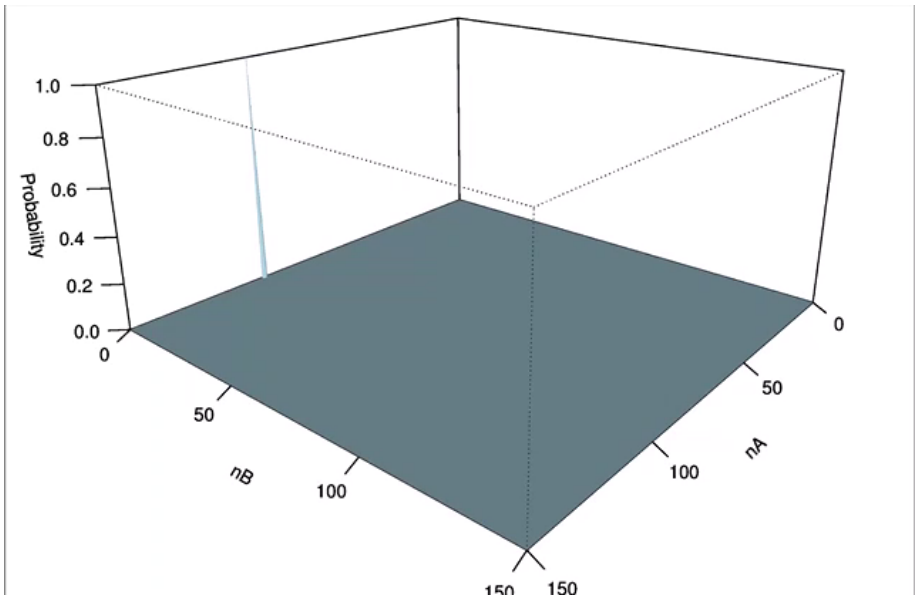
Clear bistability

Time evolving probability landscapes of toggle switch

Starting from uniform distribution



Starting from state (A=100, B=0)



How to go beyond simple motifs?

- Challenging task
  - Most studies have just a few nodes
  - Use approximations: *eg* Langevin, with unknown errors
  - Or run Gillespie algorithm, and hope for the best

Direction Solution: Multi-finite Buffer dCME Algorithm

- How do we know if computed attractive basins are real?
- How to be sure if the computed landscape is not erroneous?
  - How to quantify errors?
- What are the best accuracy we can achieve?

- Probability flux, boundaries, and truncation error of finite state space
  - Boundary probability and convergency theory
  - Can answer all above questions

- Mb-dCME algorithm and optimized finite state space
  - Decomposition of reaction graphs
  - Independent Birth-Death components
  - Multiple buffer queues to control truncation errors

(Cao, Terebus, and Liang, 2014, manuscript)

Independent Birth-Death Components of Reactions

- iBDs: Decompose reactions into independent components
- Each iBD is of infinite size, except *j*-th iBD
- Study the steady state distribution

$$\bar{\pi}_N^{(N)} \equiv \pi_{\partial,\, B_j}^{(\mathcal{I}_j)} = \sum_{\boldsymbol{x}\in\partial,\, B_j} \pi^{(\mathcal{I}_j)}(\boldsymbol{x}) = \sum_{\boldsymbol{x}\in\partial,\, B_j} p^{(\mathcal{I}_j)}(\boldsymbol{x},\, t=\infty)$$

Analysis of Truncation of Infinite Rate Matrix

- Truncation of infinite spaces, where

$$\mathcal{I} = (\infty, \cdots, \infty) \qquad \mathcal{I}_i = (\infty, \cdots, B_j, \cdots, \infty)$$
$$\mathcal{I}_{i,j} = (\infty, \cdots, B_i, \cdots, B_j, \cdots, \infty)$$

where $\Omega^{\mathcal{B}} \subseteq \Omega^{\mathcal{I}_{i,j}} \subseteq \Omega^{\mathcal{I}_j} \subseteq \Omega^{\mathcal{I}}$.

Lemma 5. *At steady state,* $\boldsymbol{\pi}^{\mathcal{I}_{i,j}}(\boldsymbol{x}) \geq \boldsymbol{\pi}^{\mathcal{I}_j}(\boldsymbol{x})$ *and* $\boldsymbol{\pi}^{\mathcal{I}_{i,j}}(\boldsymbol{x}) \to \boldsymbol{\pi}^{\mathcal{I}_j}(\boldsymbol{x})$ *componentwise for any* $\boldsymbol{x} \in \Omega^{\mathcal{I}_{i,j}}$ *when buffer capacity* $B_i \to \infty$.

- Can determine minimum buffer required for errors to be smaller than a pre-defined tolerance

- Optimal memory allocation so errors are minimized

Direction Solution: Multi-finite Buffer dCME Algorithm

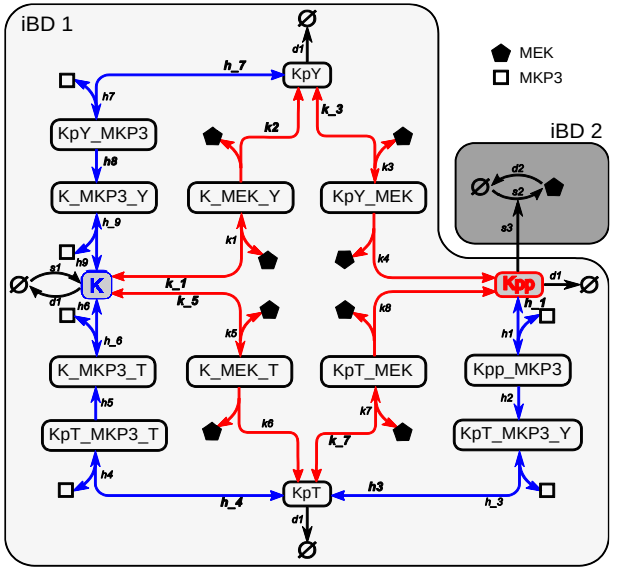
- How do we know if computed attractive basins are real?
- How to be sure if the computed landscape is not erroneous?
  - How to quantify errors?
- What are the best accuracy we can achieve?

- Yes to all with mf-dCME method!

(Cao, Terebus, and Liang, 2014, manuscript)

MAPK cascade

- Closed system, 16 molecular species (n=16), 35 reactions.
- Reduction factor of *ca*. 10^9



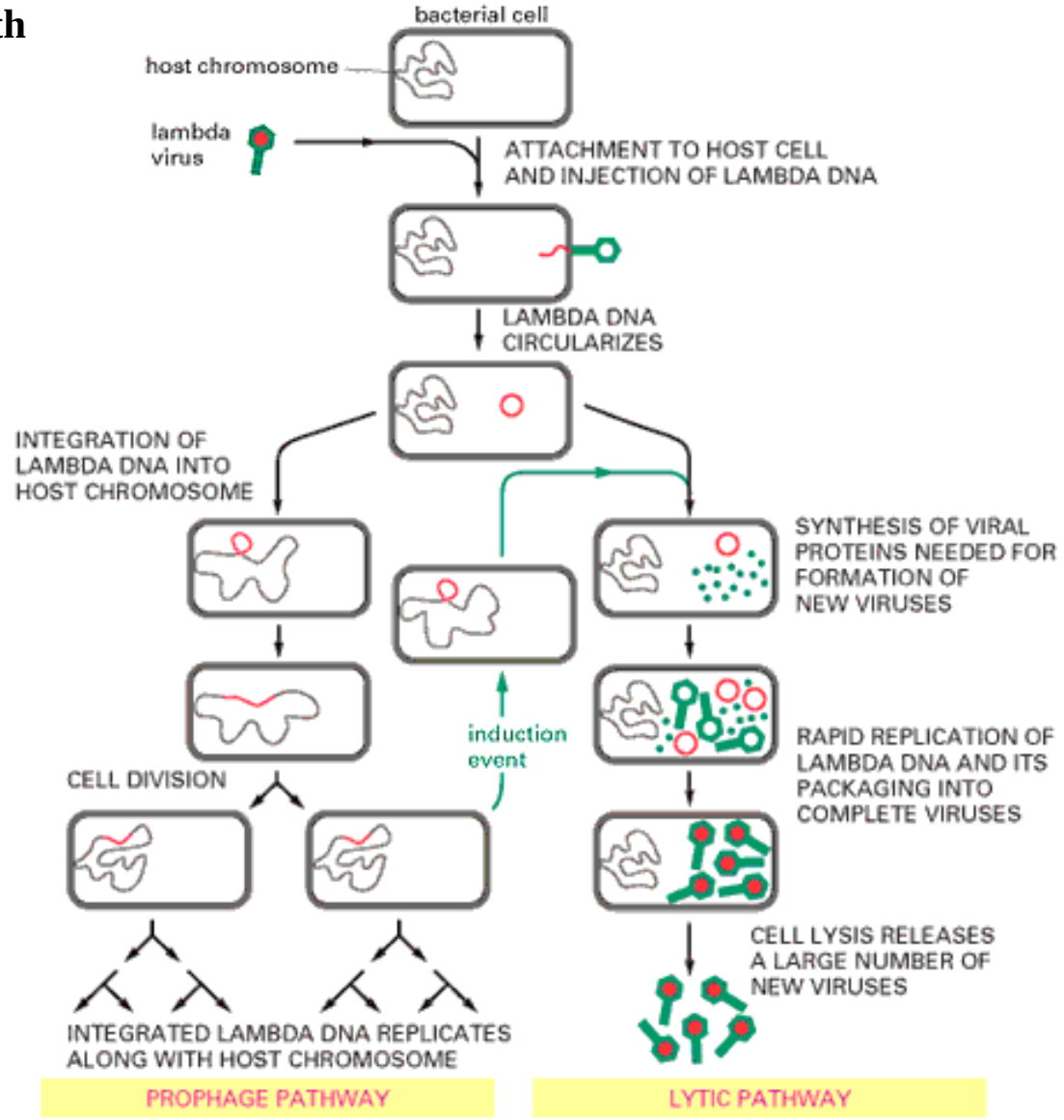
Sizes of Buffer Queues	Mb-dCME Finite Buffer	Hypercube Method	Reduction Factor
3, 3	2, 176	4.3×10^{9}	2.0×10^{6}
6, 6	209,304	3.3×10^{13}	1.6×10^{8}
9, 9	6,210,644	1.0×10^{16}	1.6×10^{9}
14, 6	2,706,935	1.1×10^{11}	4.1×10^{4}

- All with errors quantified and ensured that they are small (Cao, Terebus, and Liang, 2014, manuscript)

Example: Phage Lambda's Life Cycle:

A choice between lysogenic and lytic growth

- Two pathways:
  - Prophage/ Lysogenic pathway
    - Integrate into host genome
  - Lytic pathway:
    - genome replication, protein sysnthesis, assembly of viral particles, cell lysis, escape of mature phage
- Epigenetic memory
  - Passed along



(Arkin, Ross, McAdam, 1998)

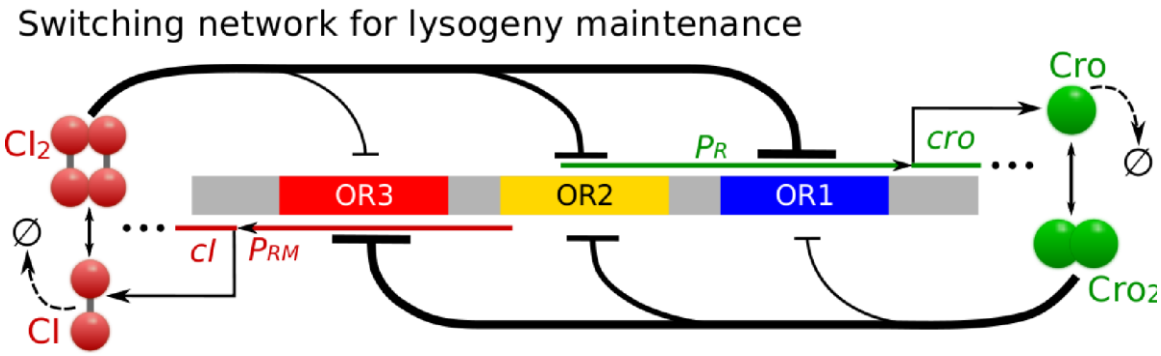
(Essential Cell Biology, Second Edition, Garland Science, 2004)

How does it work?

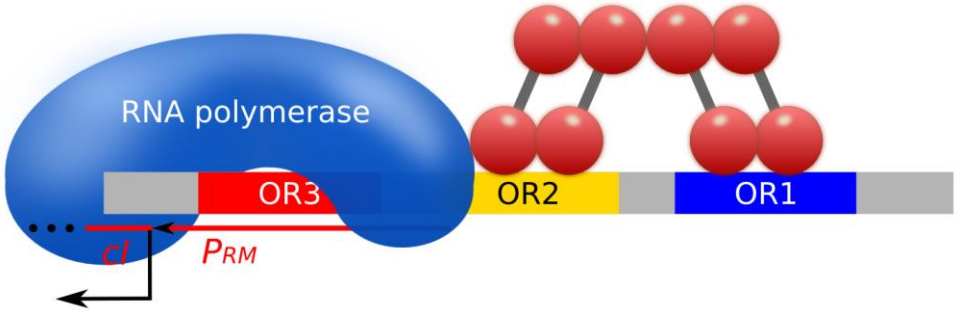
- Systems stability against perturbation
- Robustness against genetic mutation
- Regulation mechanism of the switch
- Heritable epigenetic state
  - DNA damage due to UV

Our model

- 13 molecular species
- 51 reactions
- 1.7M microstates



- Cooperativities
  - CI_2, Cro_2
  - Neighboring sites
- Implicit OR-OL looping effect
  - Stabilized CI_2 binding to OR_2, with 10 × higher CI synthesis rate with CI_2 bound OR_2
  - P_{RM} suppressed when CI_2 bound to OR_3
    - Only possible with looping

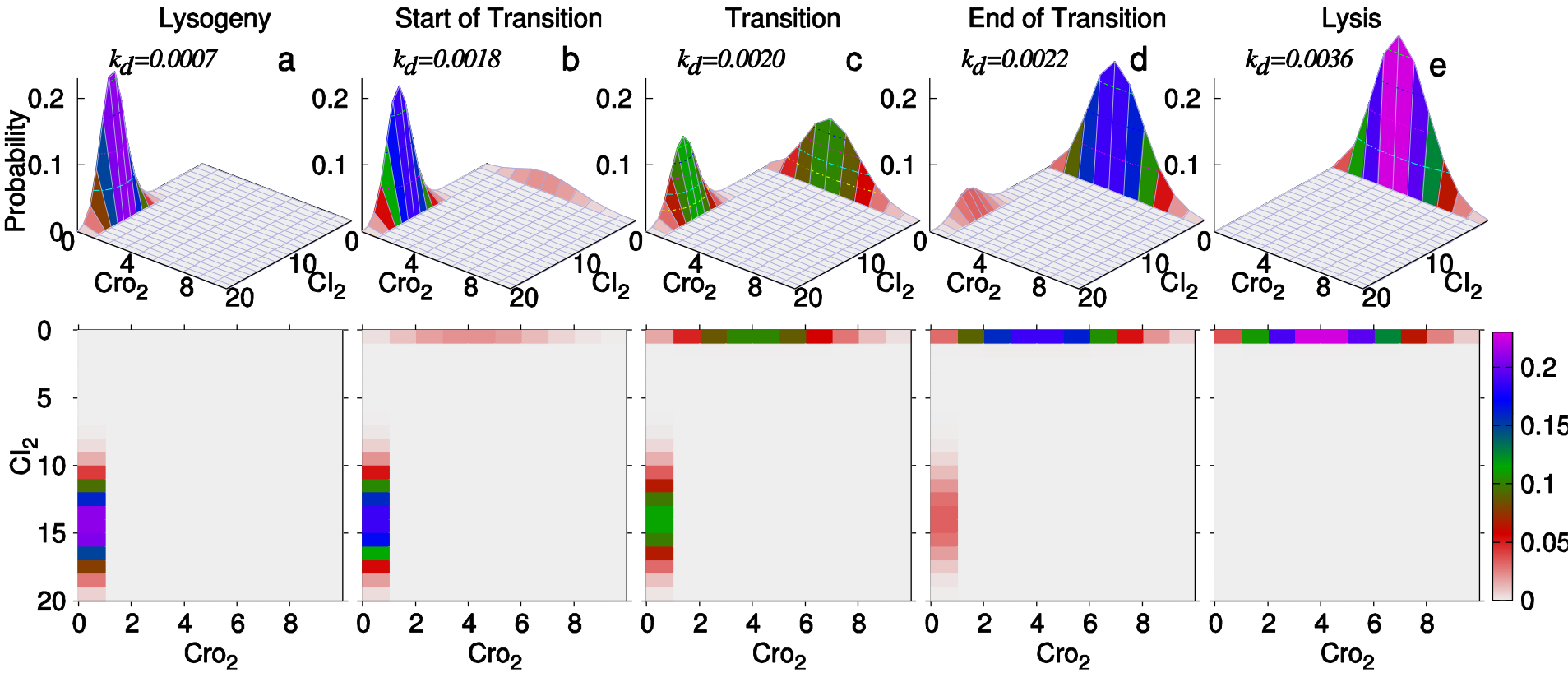


(Cao, Lu, and Liang, *Proc Natl Acad Sci USA*, 2010)

Steady State Probablistic Landscape

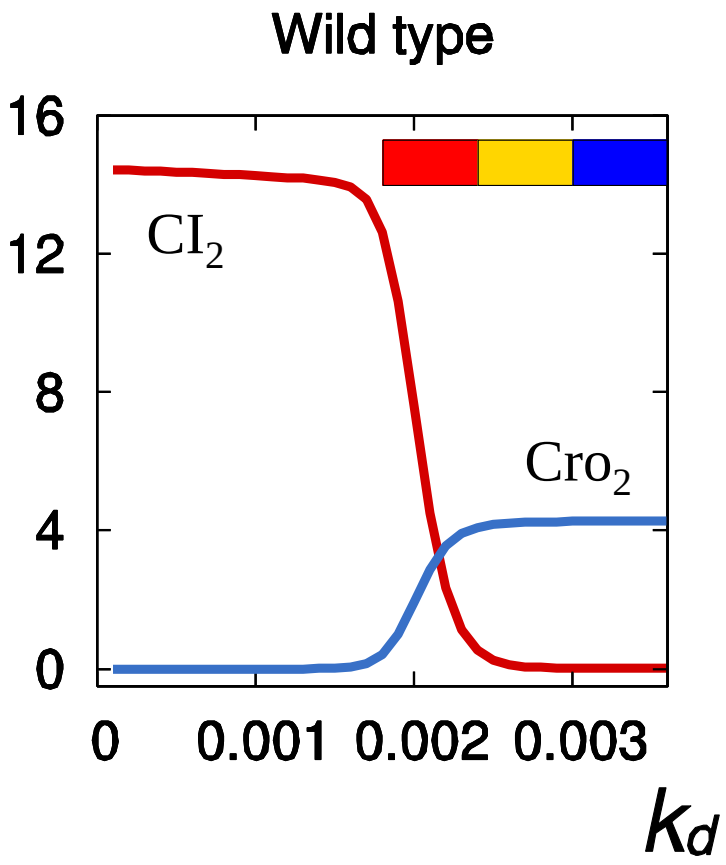
• Lysogenic induction (phage induction):
-- Switching from lysogeny to lytic development

- Different UV dosage: varying CI degradation rate
  - Projection of 13D landscape to CI_2-Cro_2 subspace



Titration Curve: A Mechanistic Picture

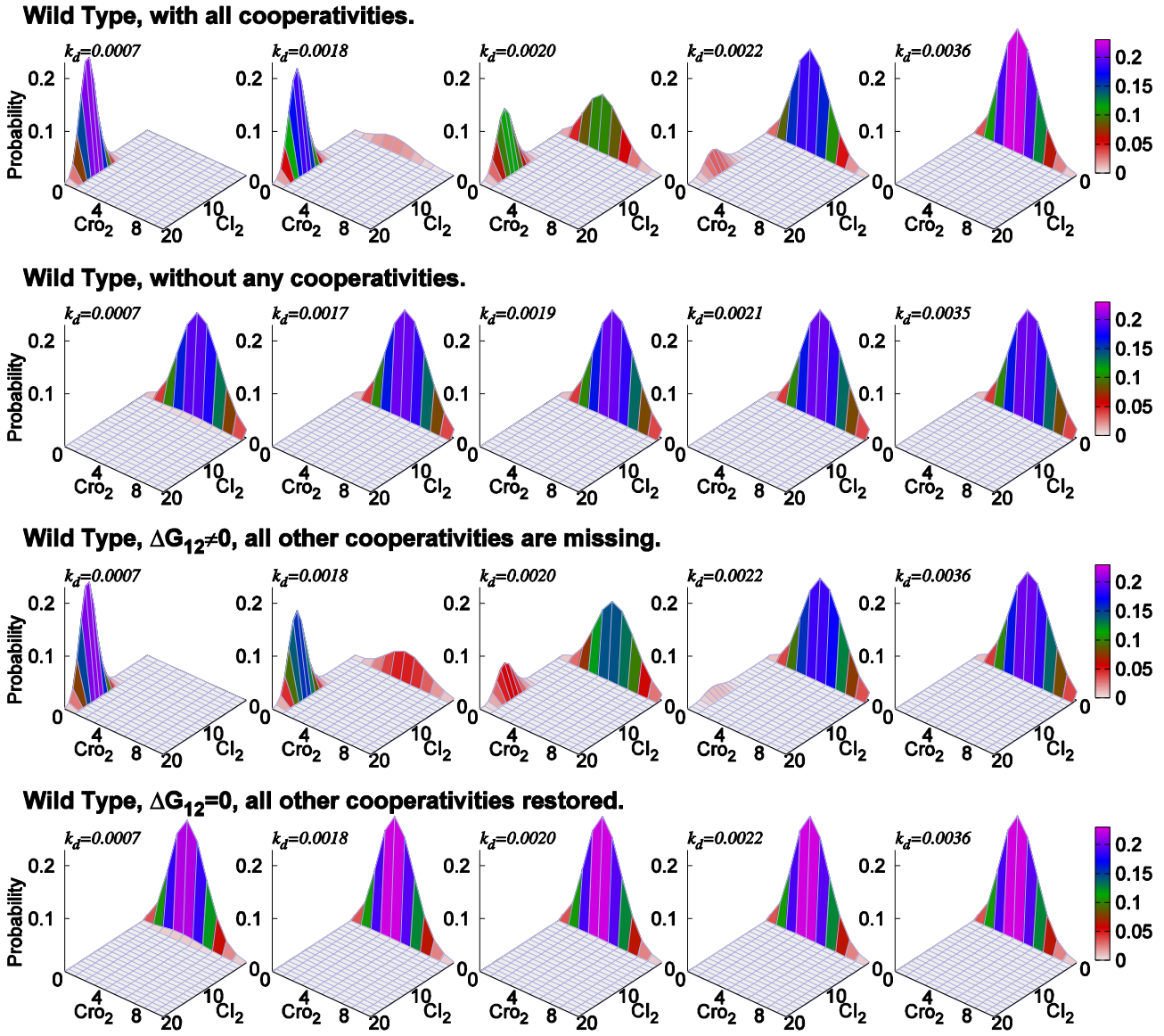
- **CI_2 and Cro_2 level:**
  - Integrate CI_2 and Cro_2 over landscape
- **Wild type: Deep Threshold**
  - Stable to UV irradiation
    - CI degradation rate can fluctuate over a wide region
    - CI level changes little and Cro suppressed
  - Efficient switch over a narrow region:
    - Ultransensitivity for true signal after set point



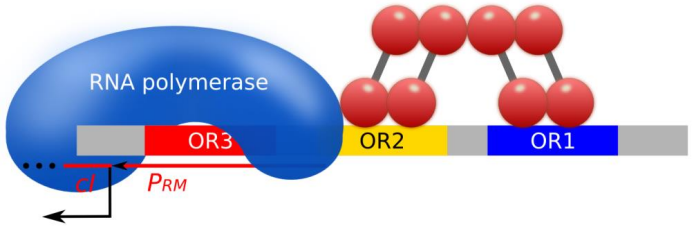
- Maintenance of epigenetic state
  - Same lysogeny upon cell growth and cell division

Mechanism:

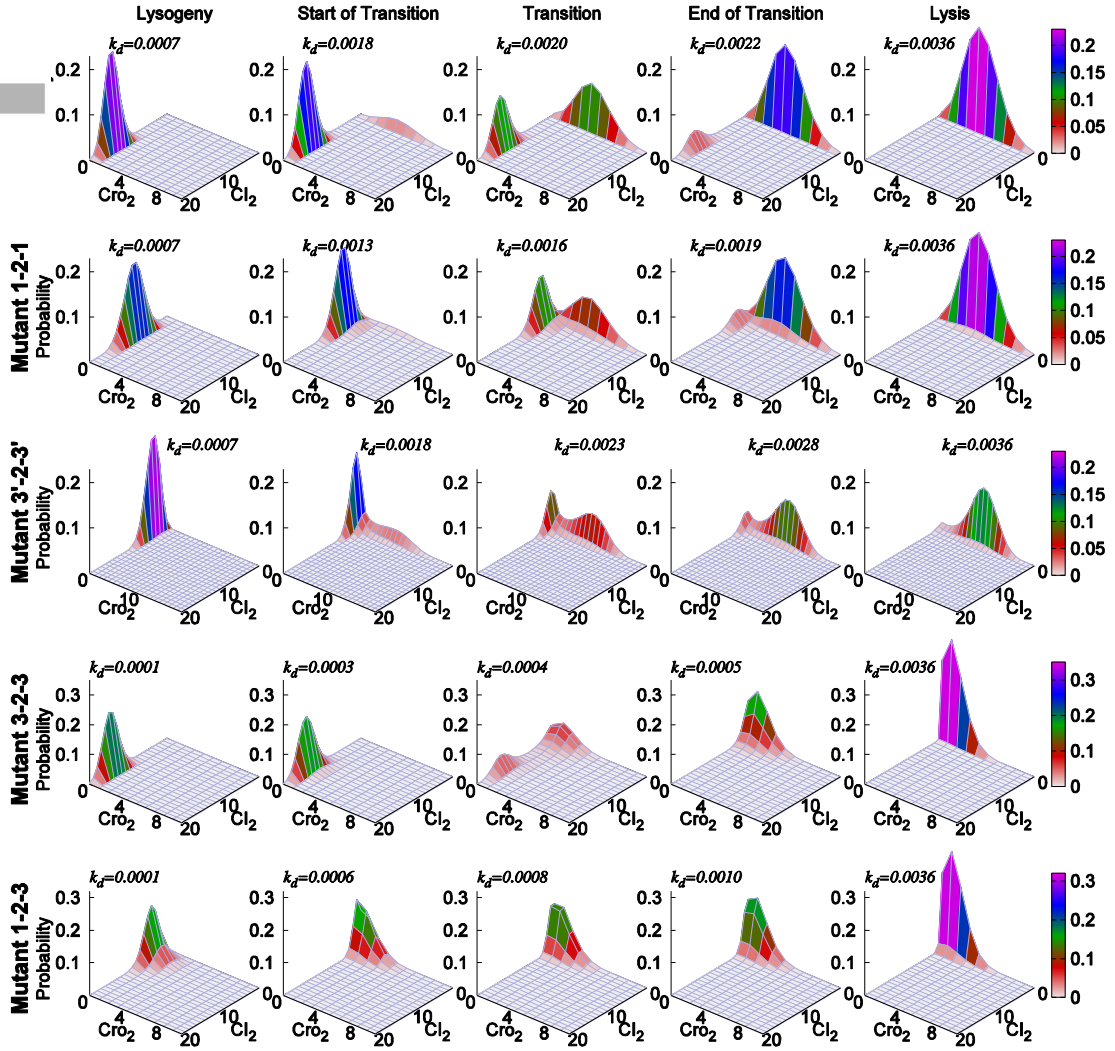
- Cooperativity is essential.
  - Otherwise, no lysogeny.



Effects of Altered Operators



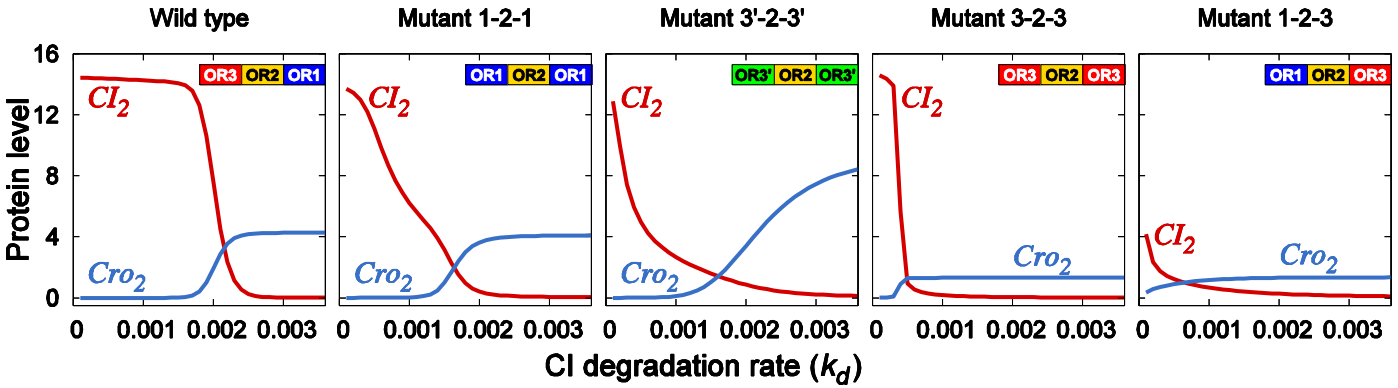
- Wild Type
- Mutants:
  - 1-2-1
  - 3'-2-3'
  - 3-2-3
  - 1-2-3



(Little et al, 2003, *EMBO J*)

Titration Curves: A Mechanistic Picture of Little Study

- All mutants have lower threshold in lysogenic-lytic state transition
  - And also leaky: switching not efficient
- Mutants can be induced to lytic state with lower UV dosage
  - Consistent with Little's "hair trigger" mechanism
- Some cannot lysogenize
  - 1-2-3



Comparison with Stochastic Simulation Algorithm

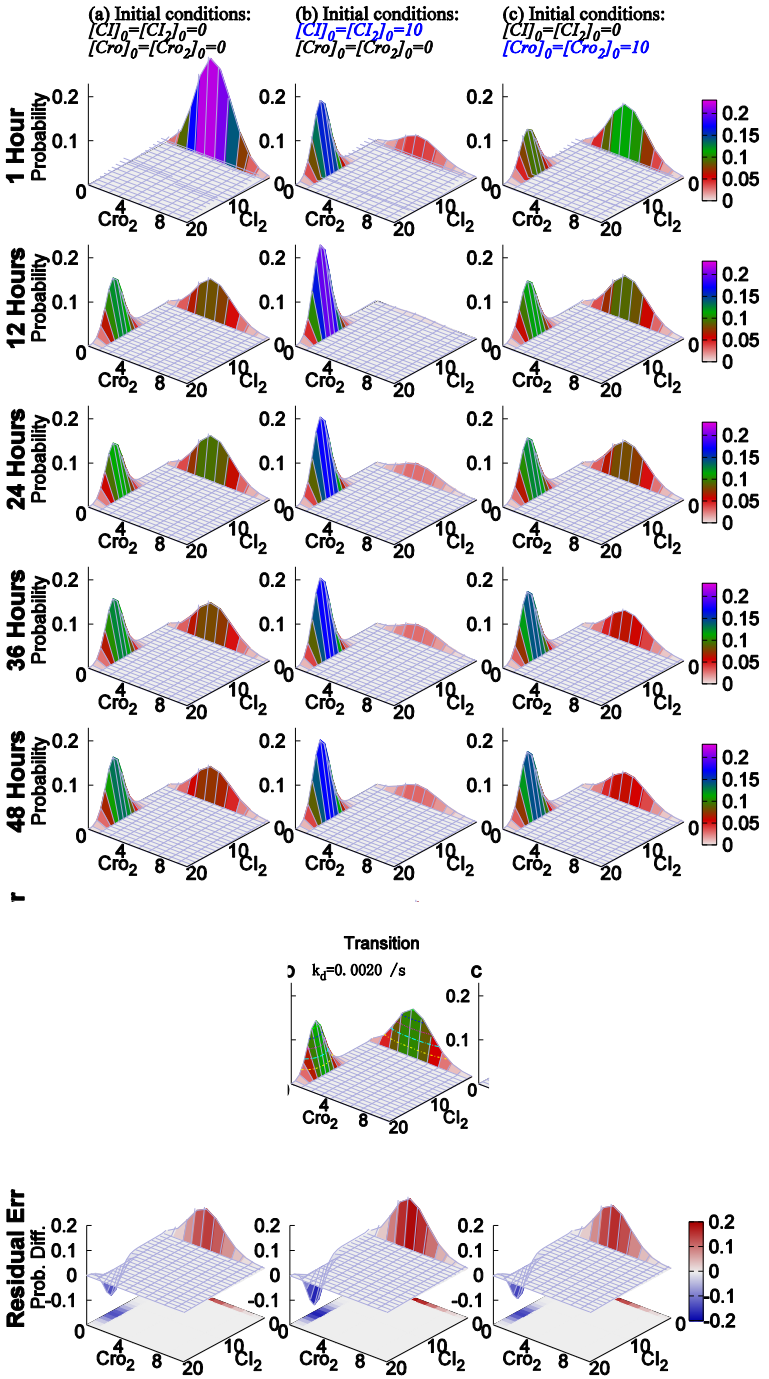
- Improved Gillespie algorithm: StochKit
  - Same model parameters, with CI k_d = 0.002/s
  - Three different initial conditions

(Li, Cao, Petzhold, and Gillespie, 2008)

- Time:
  - dCME: 8 hours, 2GHz quad core, AMD
  - SSA: not yet converged after 48 hours

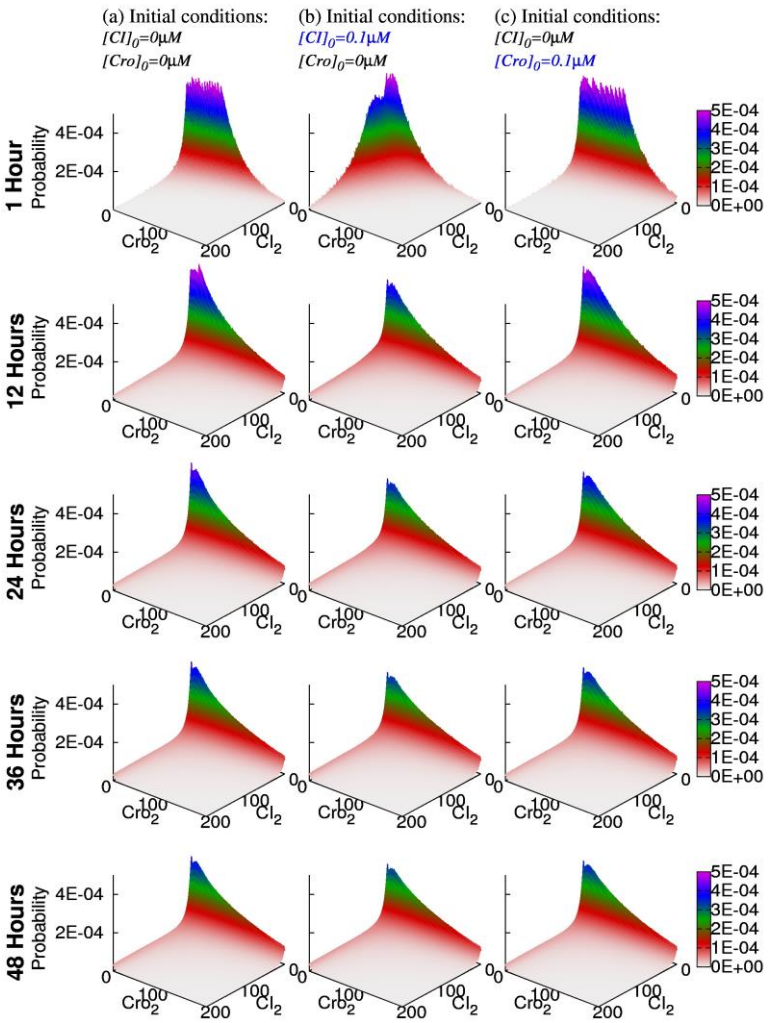
- Wrong conclusion

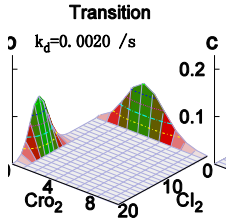
(Cao, Lu, and Liang, *Proc Natl Acad Sci USA*, 2010)



Comparison with Langevin SDE Method

- Langevin SDE
  - Same model parameters, with CI kd = 0.0021/s
  - Three different initial conditions
- Time:
  - dCME: 8 hours,
  - Langevin SDE: not yet converged after 48 hours
- Wrong conclusion
  - Transition phase vs lytic state
- It is unclear how much improvement can be achieved if different formulation of SDE and different stochastic parameters are chosen.





Remark

- dCME method can compute exact probabilit landscape of stochastic network
  - Can solve problems that Gillespie algorithm is challenged
- Study critical rare events
- Mechanistic understanding

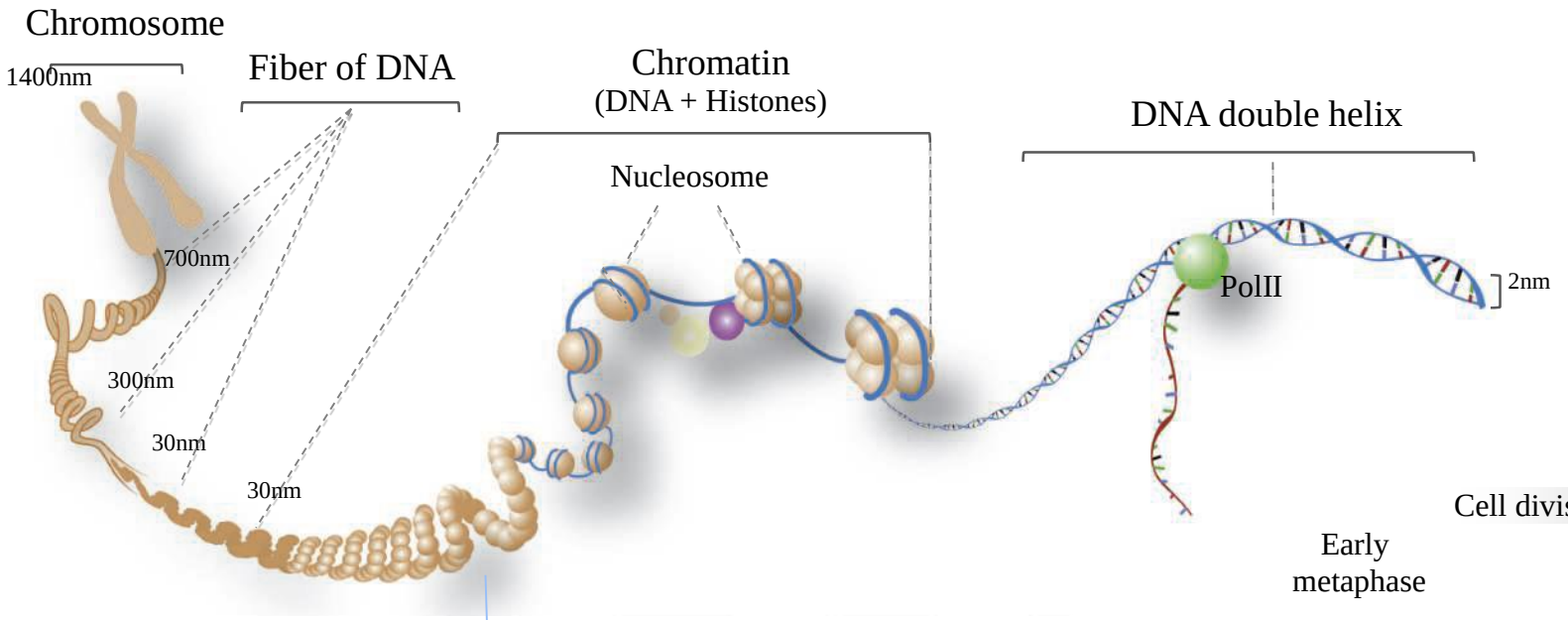
3. Chromosome Folding: Epigenetic States of Cell

- How do distant genetic elements cooperate dynamically?

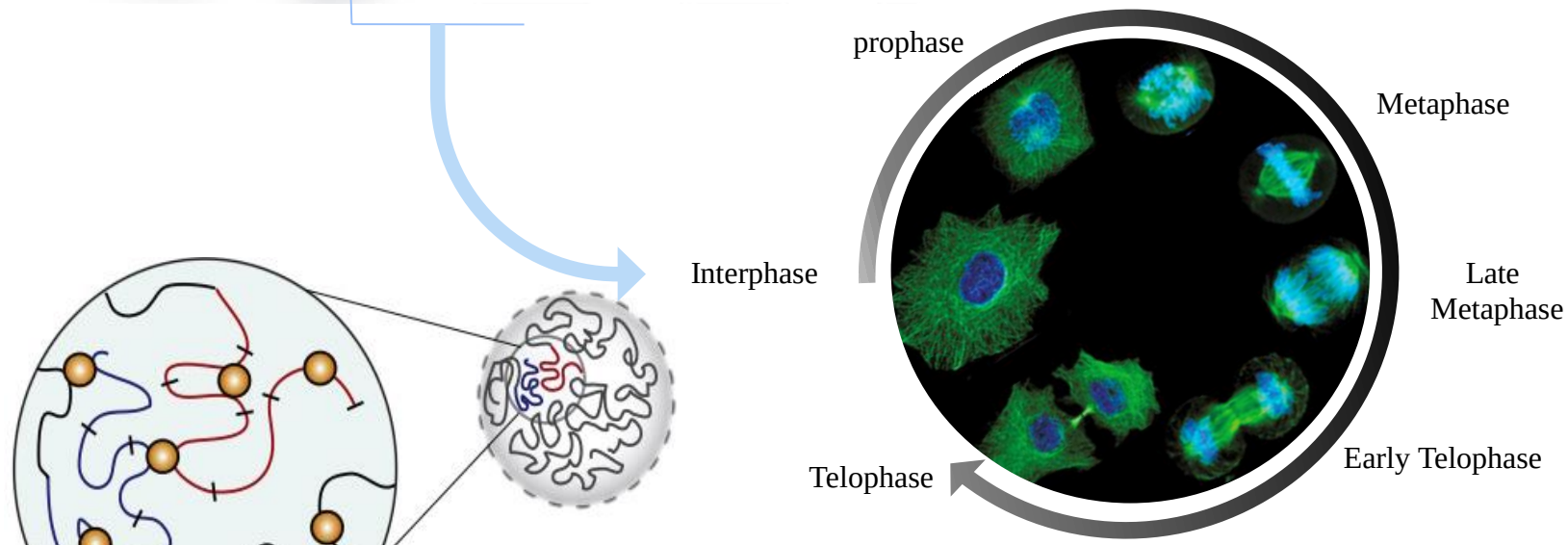
- Higher-order folding behavior of chromosome

- **Noise removal: exclude nonspecific interactions**
- Reconstructing 3D structures of chromosome

(with Gamze Gürsoy, Yun Xu, and Prof. Amy Kenter, manuscript)

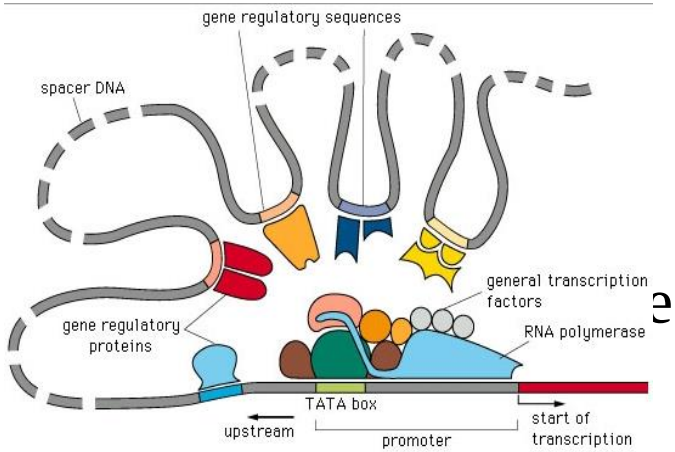


Cell division / cell cycle

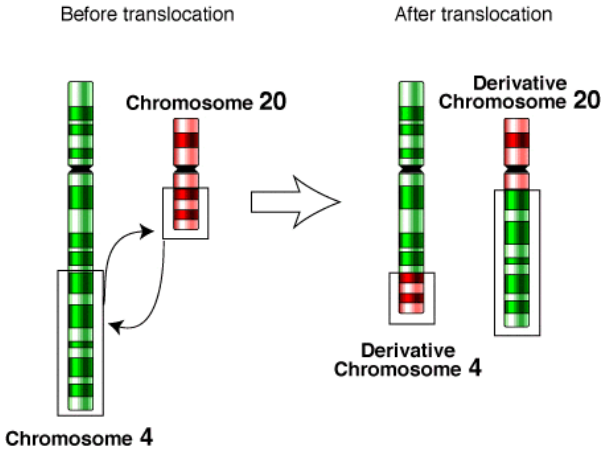


Modified from ELIZABETH PENNISI. **SCIENCE** VOL 337 7 SEPTEMBER 2012

- Chromatin looping and organization: key to cell functions
  - Gene regulation
  - Translocation

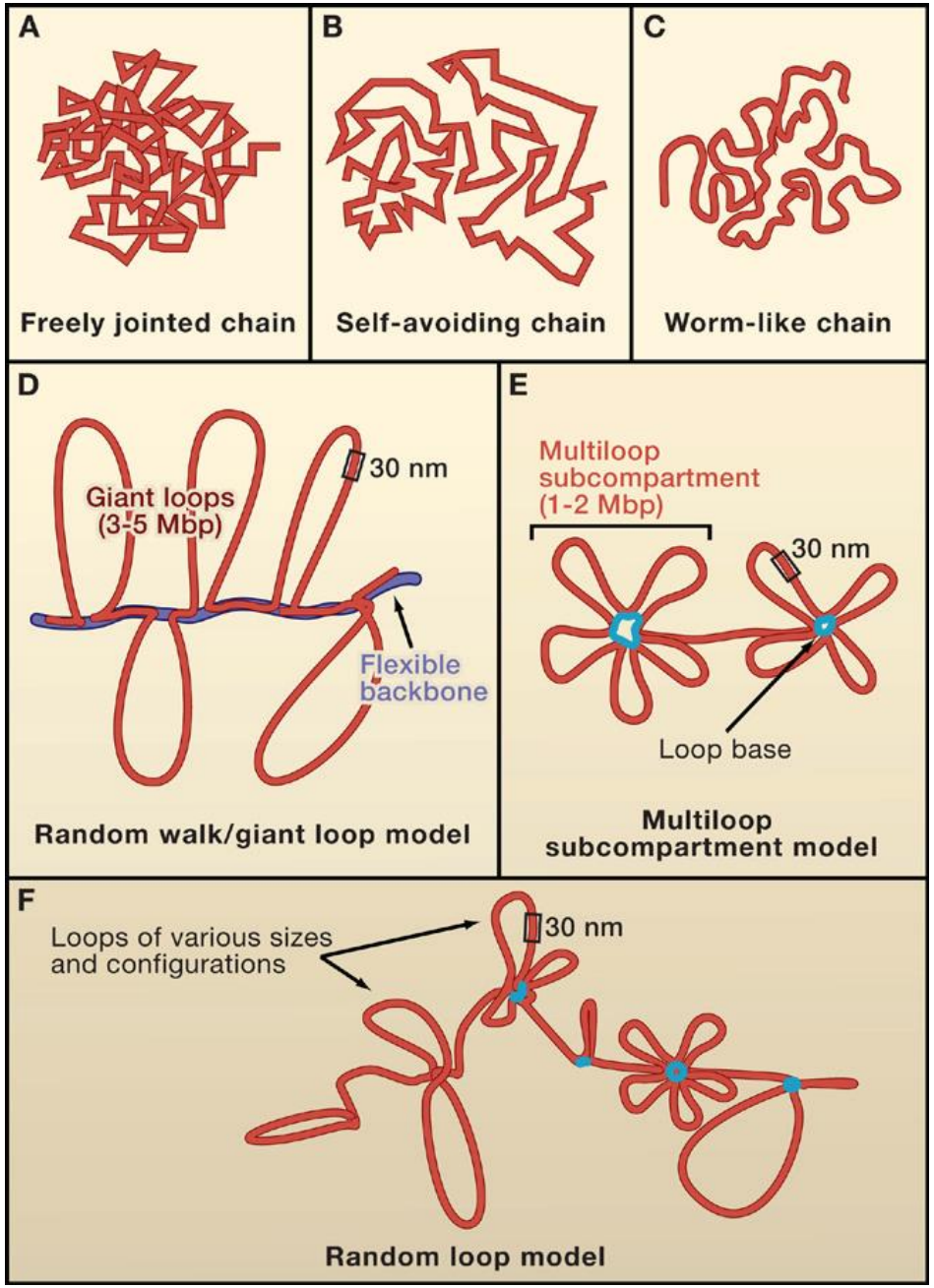


es?



http://pbil.univ-lyon1.fr/baobab/article.php3?id_article=278
http://upload.wikimedia.org/wikipedia/commons/c/cb/Translocation-4-20.png

(Work of Kuhn, Flory, de Gennes, Heerman, Langowski)



- Chain polymer models: very useful
- Challenging
  - Excluded volume
  - Physical properties
    - 30 nm fibre
    - Persistence length
  - Extremely challenging in sampling:
    - **Severe nuclear confinement**

      \+

      Experimental constraints

(from Jhunjhunwala, et al, *Cell*, 2009)

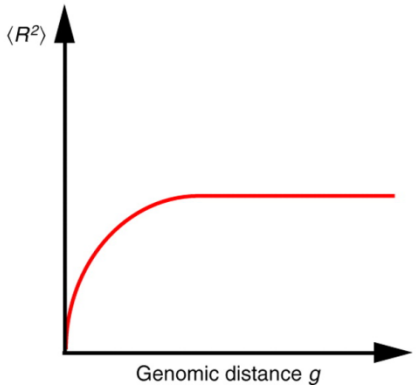
Higher Order Architecture of Chromatin Folding

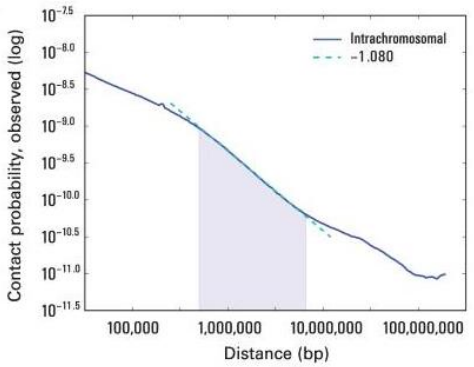
$R^2(g)$ scaling: Spatial vs genomic distances
(FISH studies)

- Short genomic distances

$$R^2(g) \sim g^{2\nu}, \text{with } \nu \sim 0.3$$

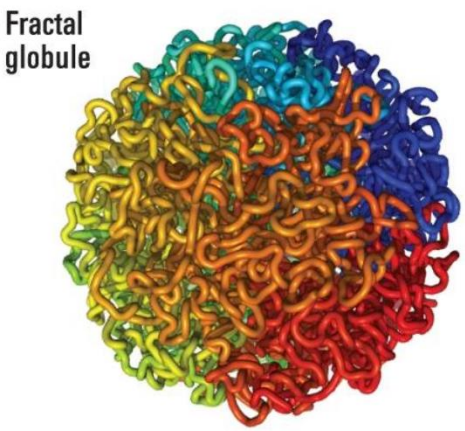
- Longer genomic distances: levels off

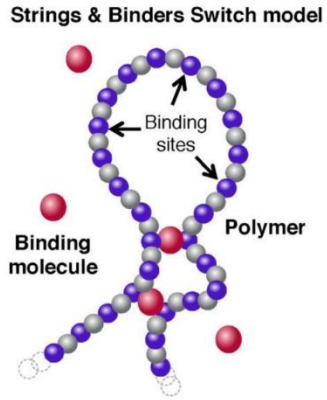




$P_c(g)$ scaling: Looping probability vs genomic distances
(HiC studies) $P_c(g) \sim g^{-\alpha}$

- Diverse, but on average: $\alpha \sim 1.08$





Fractal Globule (FG) model:

- Explains α and ν at short genomic distance
- Not: leveling-off effects and the diverse α exponents

Strings and Binders Switch (SBS) model:

- Average scaling of α and ν
- But, significant parameter tuning
- Results are mutually exclusive

(Lieberman-Eiden et al, 2010, *Science*,)

(Barbieri et al, 2012, *PNAS*)

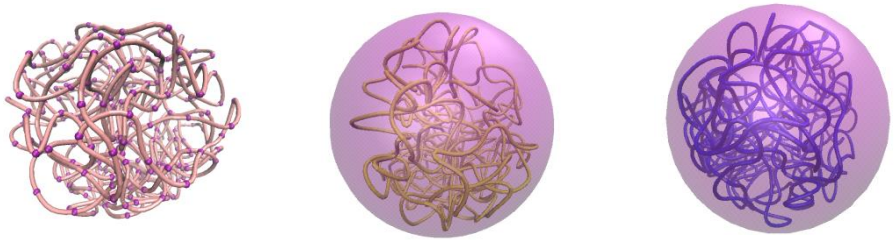
Our Approach: C-SAC Model

Constrained Self-Avoiding Chromatin model

- Physical model of chromatin: with appropriate properties
  - Polymer beads
  - *D*: 30 nm chromatin fiber
  - Bond length: persistence length L_p =150 nm
  - Self-avoiding with excluded volume
  - 1,000 units ~ 5,000 beads ~ 15 Mbp long chain

- **Spatial confinement explicitly modeled !**
  - Not possible with MD or Metropolis Monte Carlo

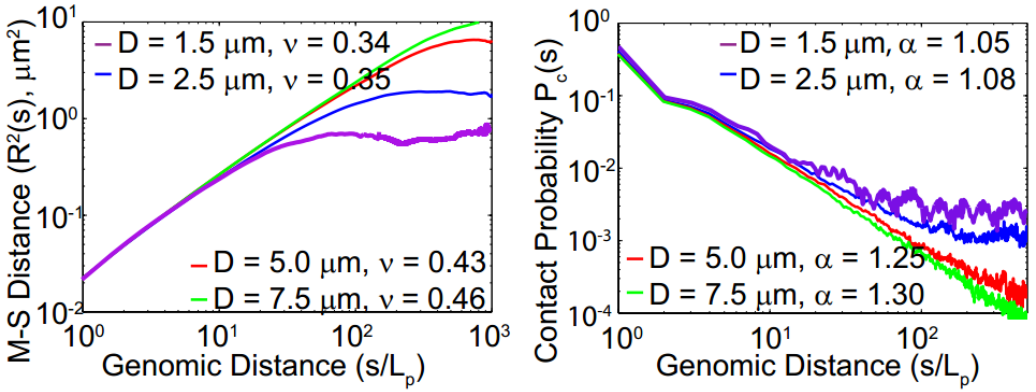
- 10,000 C-SAC chains
  - Properly weighted!



(Gürsoy, Xu, Kenter, and Liang, *Nucleic Acid Res*, 2014)

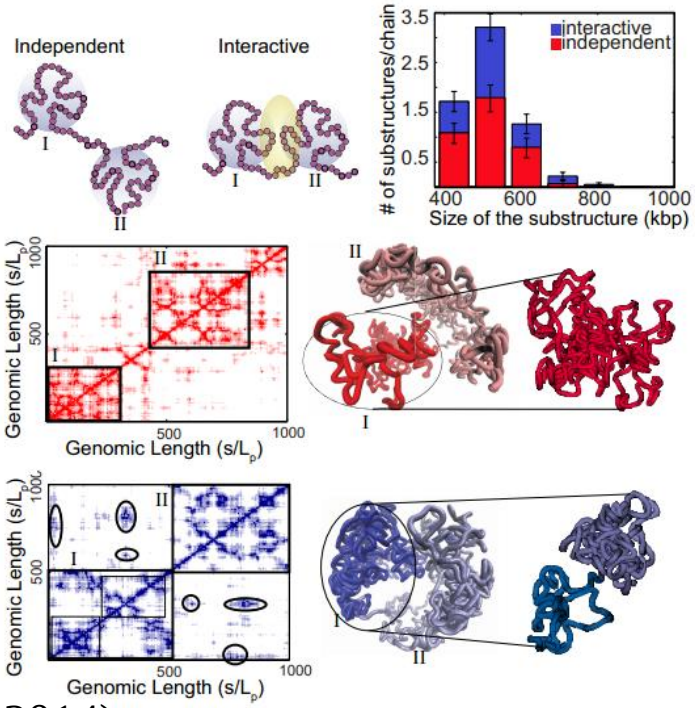
Spatial Confinement Important for Chromosome Folding

- Reproduces $R^2(g)$ and $P_c(g)$ scaling
  - Currently the only polymer model account for all facts
  - Without tuning parameter



- Drives formations of topological domains

- **Nucleus size regulates chromosome structure**
  - induces different folding landscapes
    - e.g., Stem cell with nucleus occupying 95% of volume



(Gürsoy, Xu, Kenter, and Liang, *Nucleic Acid Res*, 2014)

Remark

- Chromatin Chain Polymer Model and Sampling Techniques
  - In confinement and with experimental constraints
- **Removal of noise due to non-specific interactions**

Summary

- Stochastic genetic circuits and cellular state
  - Foundational principles and algorithms
  - Cellular fate

- Chromosome folding and nuclear confinement: 3D structure predictions
  - Removal of noise due to non-specific interactions.
  - 3D physical basis of gene activation and cellular programming

→ Acknowledgement!

Collaborators

- **Youfang Cao (**UIC)
- **Dennis Gessmann (UIC)**
- **Gamze Gursoy** (UIC)
- Sema Kachalo (now Intel graphics unit)
- Ronald Jackups, Jr (Wash U, Assist Prof)
- Yingzi Li (SJTU)
- **Meishan Lin** (UIC)
- **Hammad Naveed** (UIC/now Toyota Institute)
- **Ke Tang** (UIC)
- **Anna Terebus** (UIC)
- Wei Tian (UIC)
- **Yun Xu** (UIC)
- **Jieling Zhao** (UIC)
- Jinfeng Zhang (FSU, Associate Prof)

- David Stone (UIC)
- **Luisa Di Pietro (UIC)**
- **Linda Kenney (UIC)**
- **Amy Kenter (UIC)**
- John Marko (NWU)
- Lisa Xu (SJTU)

Papers: http://gila.bioe.uic.edu/lab/ (left column)

Acknowledgement

- NSF DBI
- NIH NIGMS
- CBC Catalyst