Mathematical Models in Synthetic Biology: From Molecules to Life

Yiannis N. Kaznessis

Department of Chemical Engineering and Materials Science

University of Minnesota

S.V. Sotirchos Memorial Lectureship

8th Panhellenic Scientific Chemical Engineering Congress, Thessaloniki

Mathematical Models

- are at the heart of physical sciences and engineering.

- capture the essential aspects of systems, processes, phenomena.

- are founded on universally accepted laws of physics and chemistry.

- are key to understanding, predicting, designing, optimizing and controlling.

Mathematical models in biology?

- Can biological phenotypes be explained with mathematical models of interacting molecules according to physical laws?

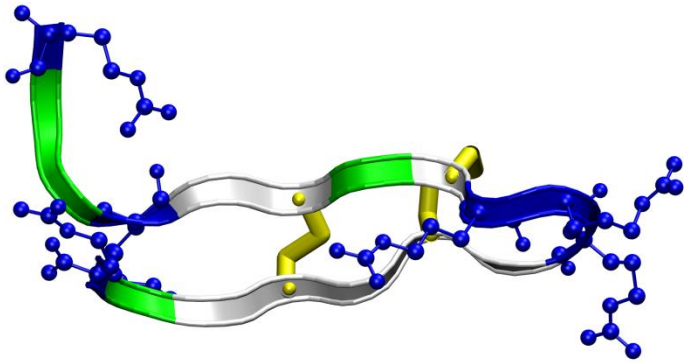
- How does life emerge from a soup of chemicals?

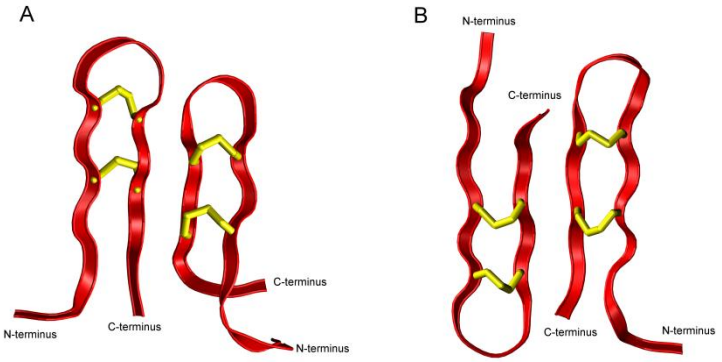
- Two major challenges:

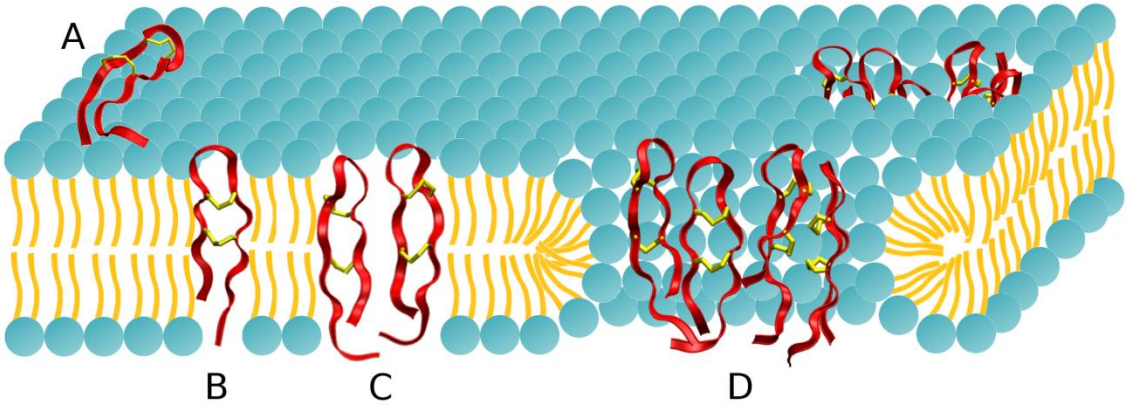
1. Biological systems are not only non-linear and often stochastic; they possess an overwhelming number of variables.
2. Biology is a discipline in history: Dobzhansky's dictum that "Nothing in biology makes sense except in the light of evolution" casts a long shadow on mathematical models of phenotypic complexity.

Antimicrobial Peptides

- Promising alternative to traditional antibiotics: no bacterial resistance.
- How do they work?
- Focus on protegrin-1. Potent against wide spectrum of bacterial organisms.

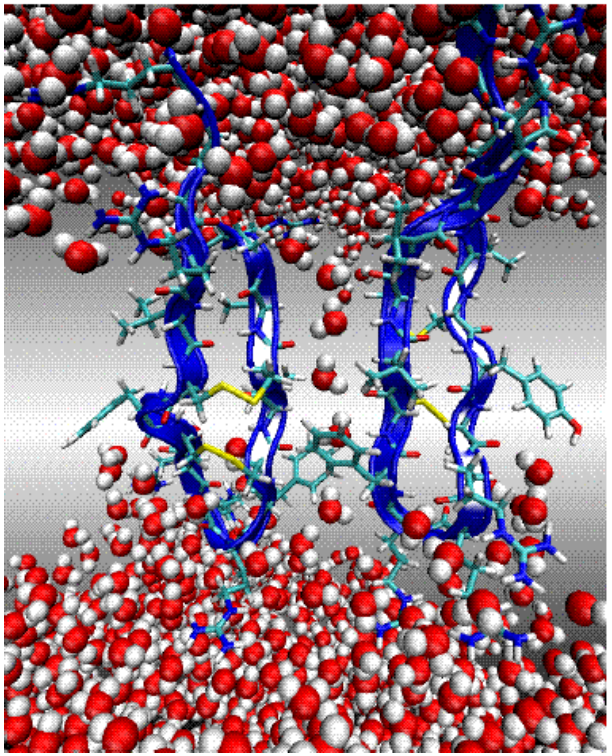






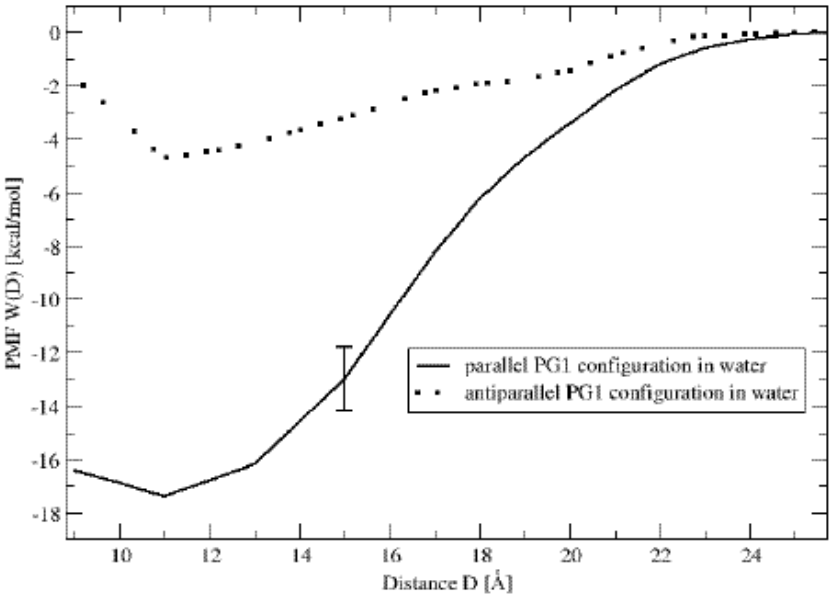
Antimicrobial Peptides

- Calculate the potential of mean force of binding on the membrane, of inserting inside the membrane, of dimerizing, oligomerizing.



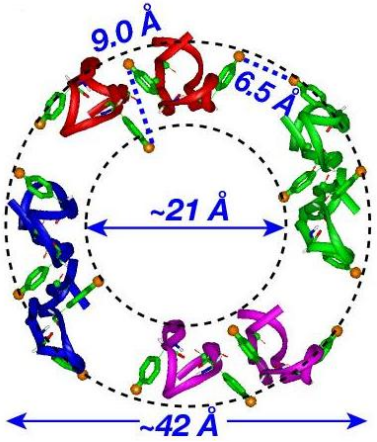
$$W(D) = -\int_{\infty}^{D} \overline{F}(D')dD'$$

$$K = \frac{1}{D_{max} - D_{min}} \int_{D_{min}}^{D_{max}} dD e^{-\beta W(D)} \equiv \langle e^{-\beta W} \rangle_{\Delta l}$$

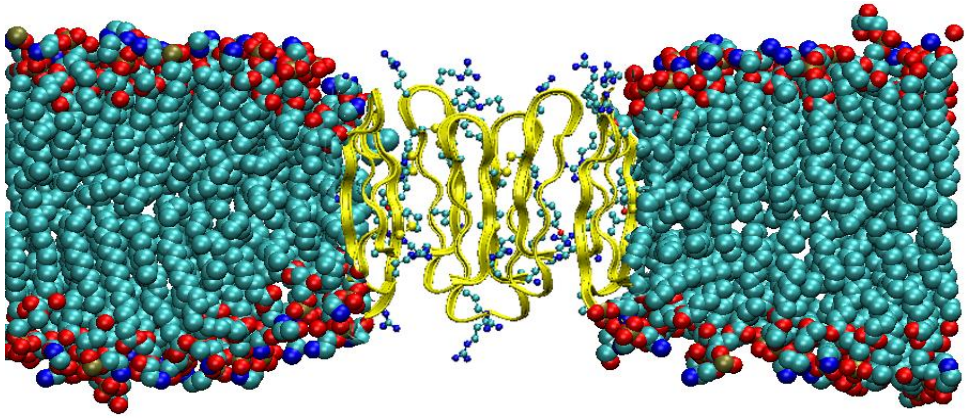


Antimicrobial Peptides

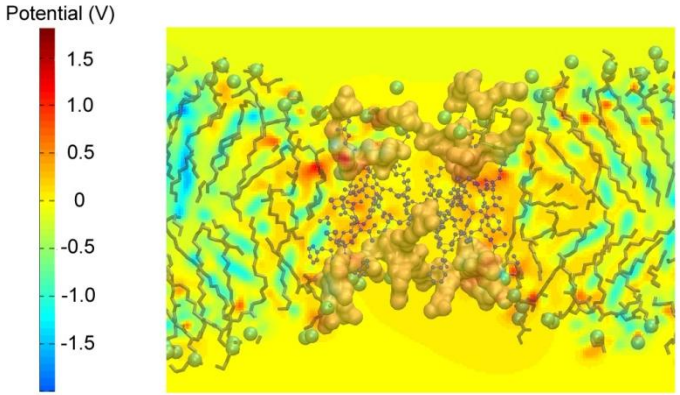
- All-atom molecular dynamics simulations of protegrin pore (98,000 atoms, 150 ns).
- Determine the structure of the pore.
- Determine the electrostatic potential and the dielectric constant.



Mani, R. et al., Biochemistry, **2006**. 45(27): 8341-9.



Langham A, et al. , JACS, **2008**, 130(13): 4338-4346



Timeline of AMP function

- Ion transport through a protegrin pore using Poisson-Nernst-Planck equations.
- Transient ion transport from bacterial cells. Collapse of transmembrane potential and osmotic swelling.

$$\nabla \cdot \left(D_i \nabla c_i + \frac{D_i q_i F}{kT} c_i \nabla \phi \right) = 0$$

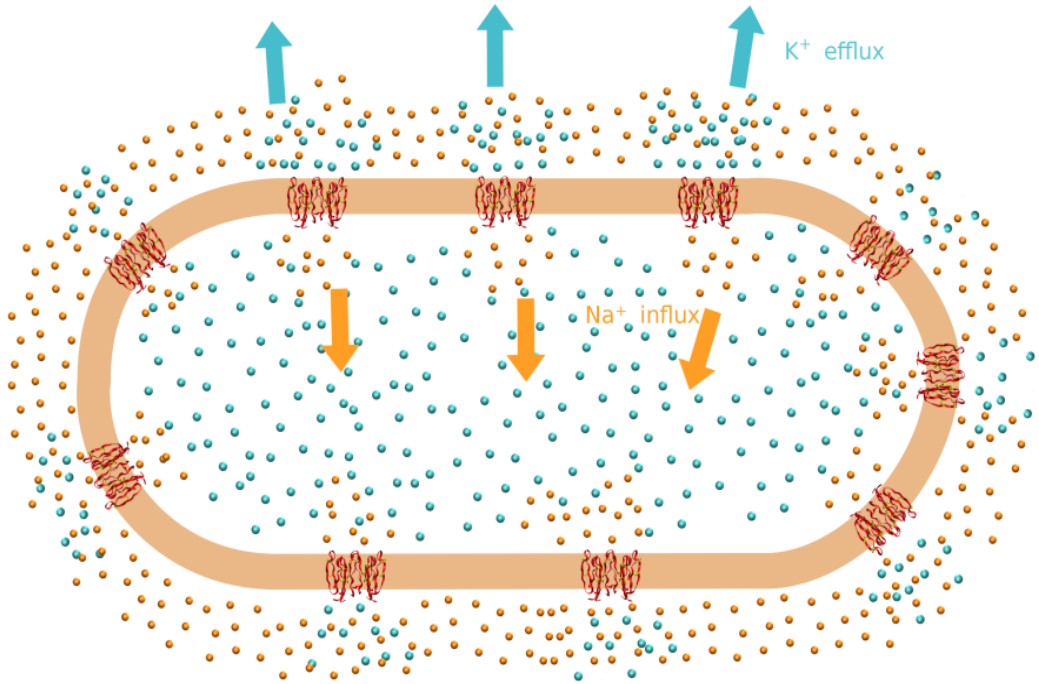
$$-\nabla \cdot (\varepsilon \nabla \phi) = \sum_{i=1}^{n} q_i c_i + \rho_f(\mathbf{r})$$

$$i = 1, \ldots, \text{Number of ionic species}$$

$$\frac{dn_i}{dt} = P_i q_i u \frac{c_i^o - c_i^i e^{q_i u}}{1 - e^{q_i u}} A$$

$$\frac{dV}{dt} = -L_P \left(\Delta P - RT \sum_i (c_i^o - c_i^i) + RT c_f \right) A$$

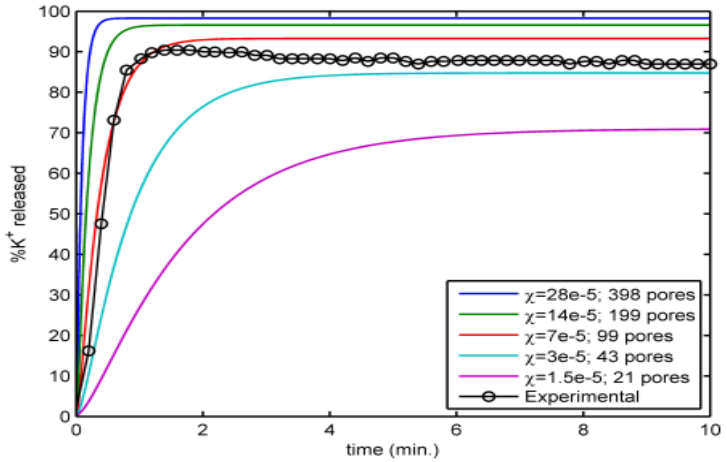
$$\Delta P = \Delta P_0 + m \frac{(V - V_0)}{V_0}; \qquad A = (4\pi)^{1/3} (3V)^{2/3};$$



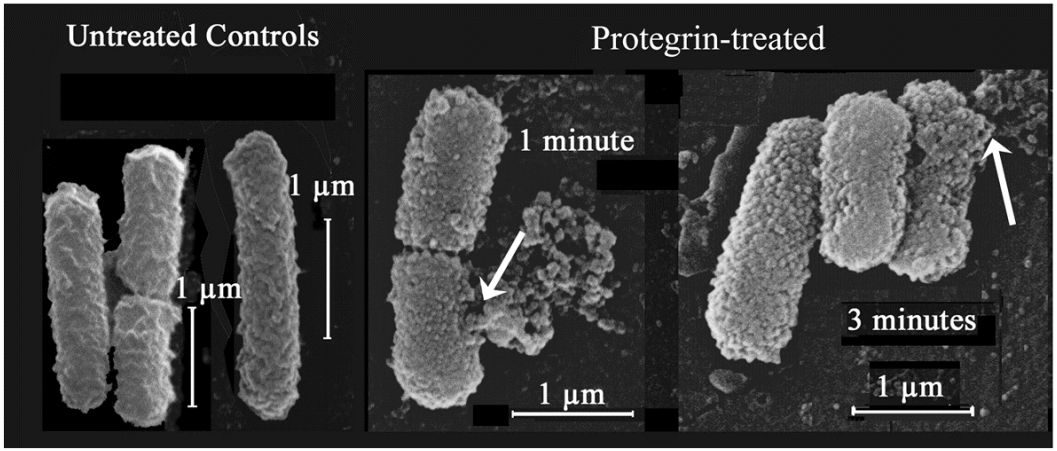
Bolintineanu D, et al. Peptides. **2010**, 31(1):1-8

Antimicrobial Peptides

- Measured potassium release matched by models.

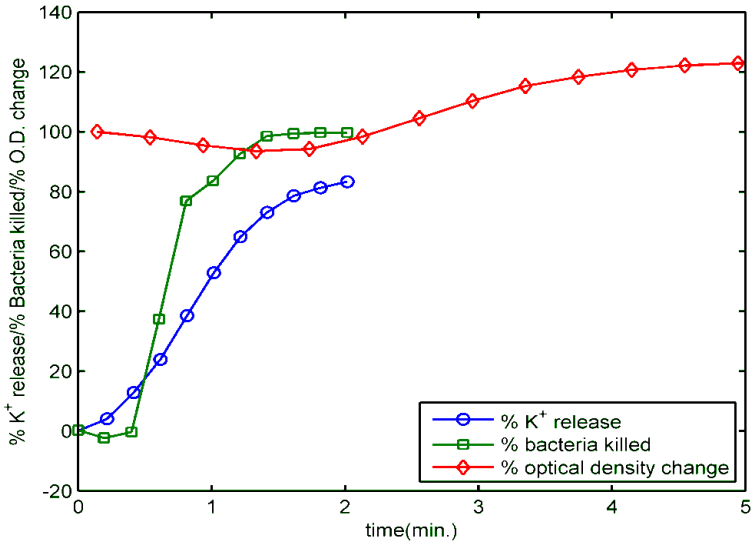


- Osmotic swelling observed by SEM.



Bacterial killing kinetics

- Mathematical models can provide insight into the mechanism of action of antimicrobial peptides. Multiscale models capture all the interactions that are relevant to activity and provide a quantitative narrative that is useful in designing new peptides.



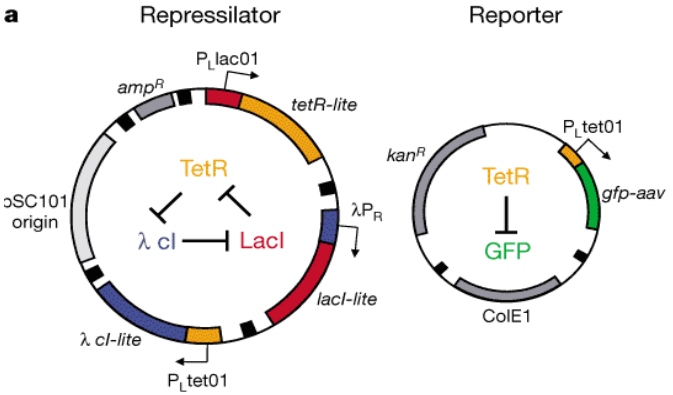
- % bacteria killed and % K+ release data were measured in the same experiment, while optical density data are from a different experiment.

- K^+ release kinetics apparently concurrent with bacterial killing kinetics.
- Bacterial death occurs faster than osmotic swelling → this may be an overkill mechanism, and K^+ release/T.M. potential decay the leading cause of death.

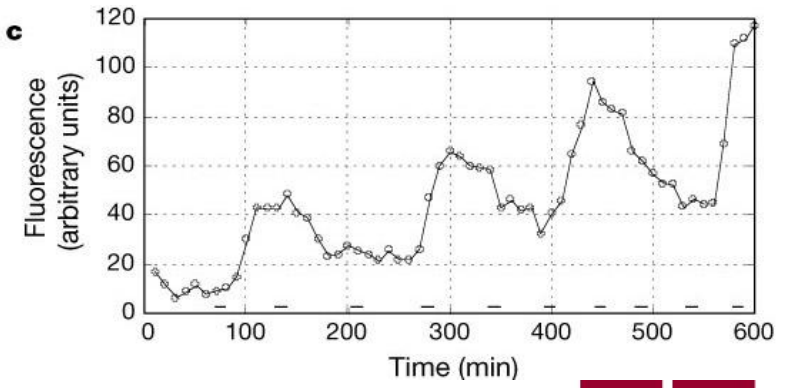
Bolintineanu D, et al. Peptides. **2010**, 31(1):1-8

Synthetic biology may usher a new era for modeling in biology

- Synthetic biology: Forward engineering of biological *systems* (beyond traditional genetic engineering).
- Chemical synthesis of DNA
  - Inexpensive: $0.5/bp
- DNA can be cut and pasted, assembled to generate new functions.
- With Genome Projects toolboxes available of
  - Regulatory proteins (activators and repressors)
  - Operator and promoter sites
  - Small inducer molecules
- Novel gene regulatory networks are at hand.

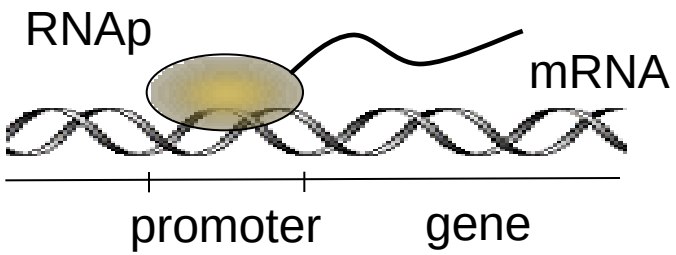


Elowitz, Leibler, Nature, (2000)

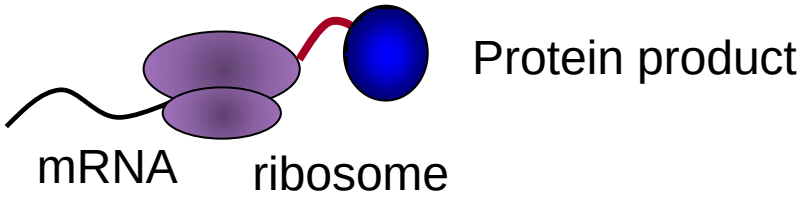


Mathematical Models of Synthetic Gene Networks

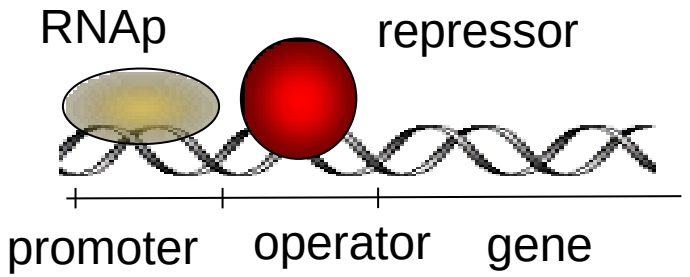
- Synthetic biosystems may be small enough and sufficiently context-independent for us to test universal mathematical models based on first principles.



- Adopt Jacob's and Monod's postulate: biological phenotypes, however complex, can be explained in terms of cascades of biomolecular interactions.

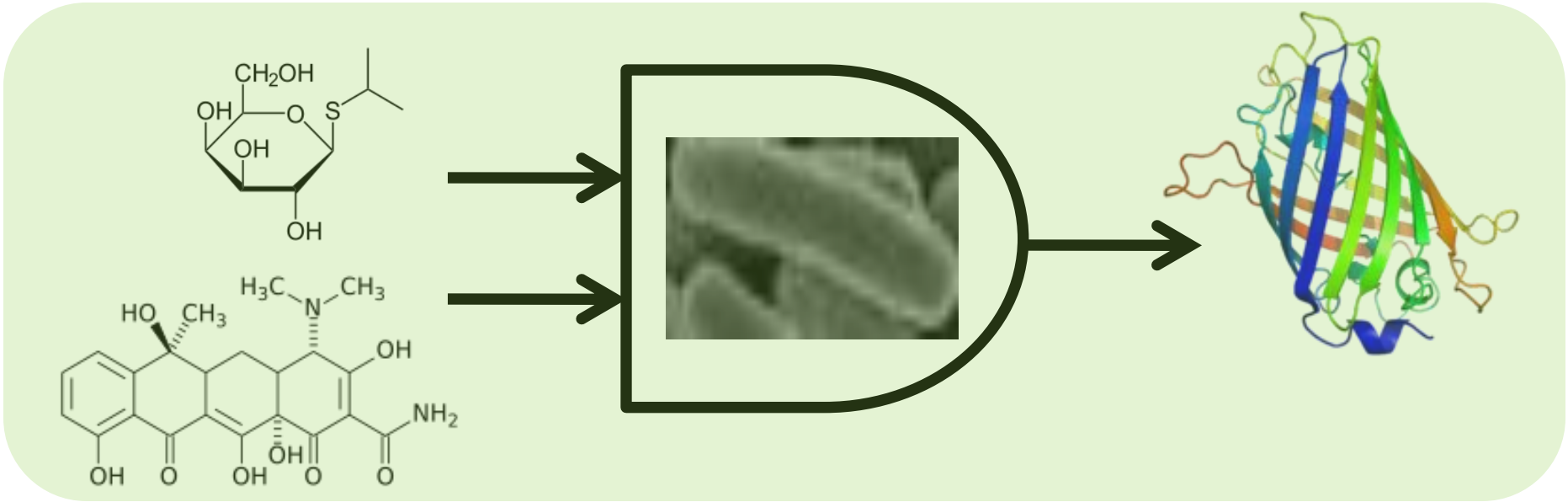


- All interactions dictated by thermodynamics and kinetics.
  - Protein Interactions
  - Transcription
  - Translation
  - Regulation



Synthetic Bio-Logical AND Gate

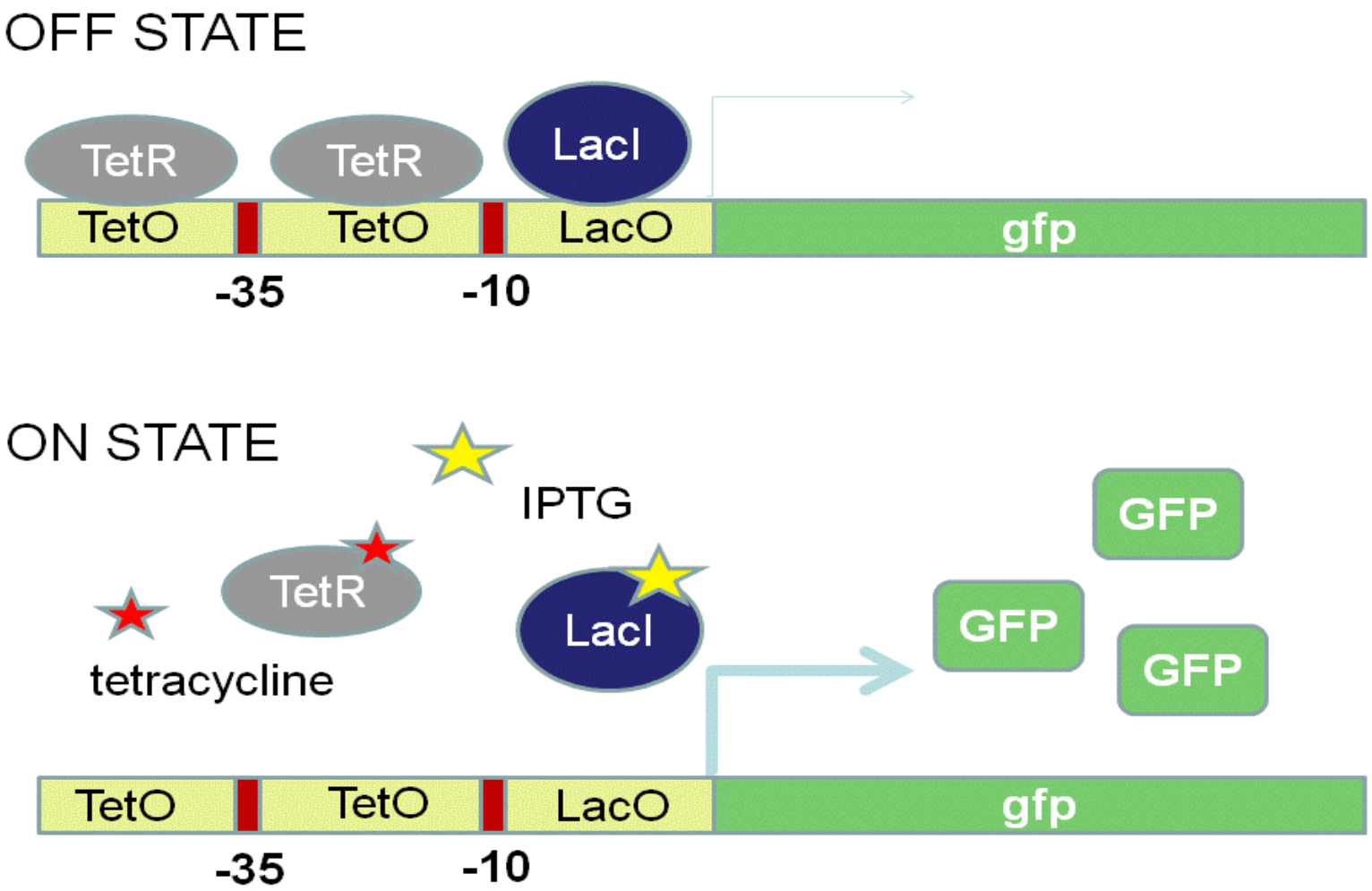
We engineered E. coli to produce green fluorescence protein if IPTG AND tetracyline are added in the bacterial culture.



Ramalingam, et al., Biochem. Eng. J. (2009)

Images of IPTG, TC and GFP from Wikipedia

Synthetic Bio-Logical AND Gate

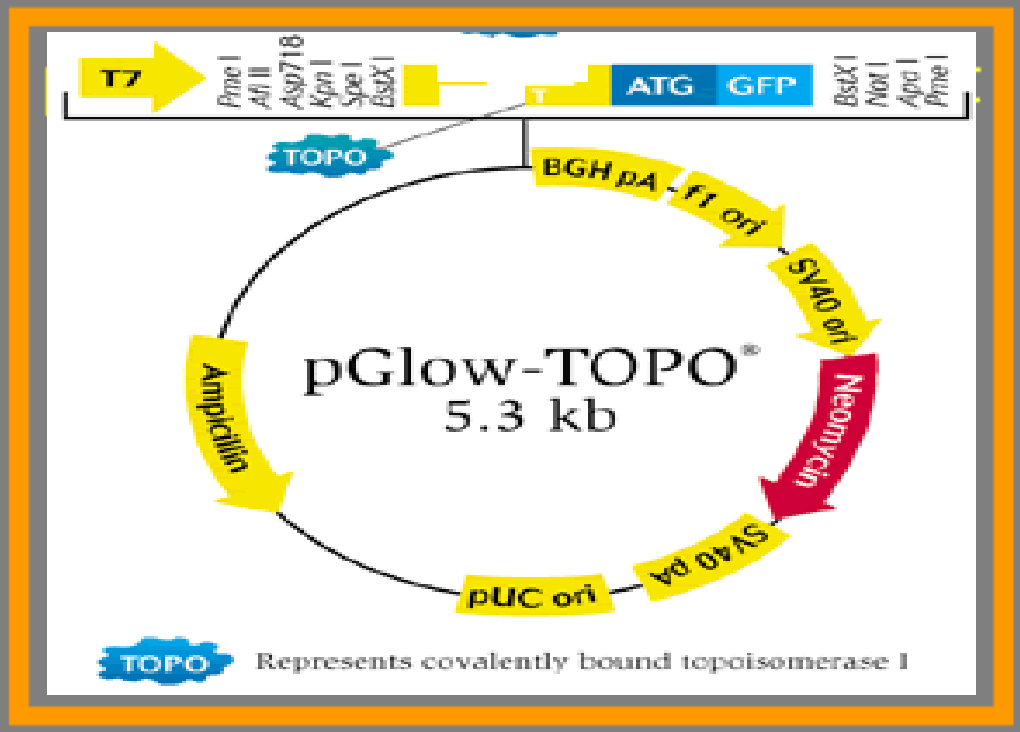


Experimental Construction of a Lac/Tet AND Gate

E. coli DH5α Pro cells
with lacI and tetR

aTC inducer

IPTG inducer



GFP

P lac/2tet

Lac Repressors

Tet Repressors

cycle3 gfp

AAATTGTGAGCGGATAACAA TTGACA TCCCTATCAGTGATAGA GATACT ATCCCTATCAGTGATAGAGA AGGAAACCGGTTC ATG

-35

-10

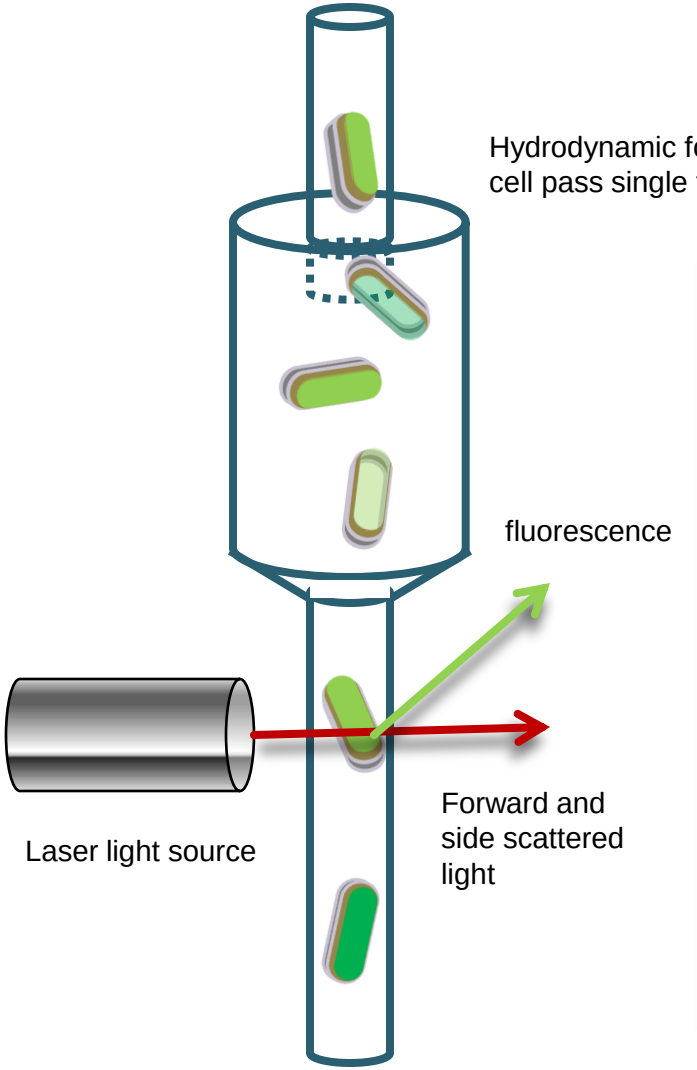
lacO1

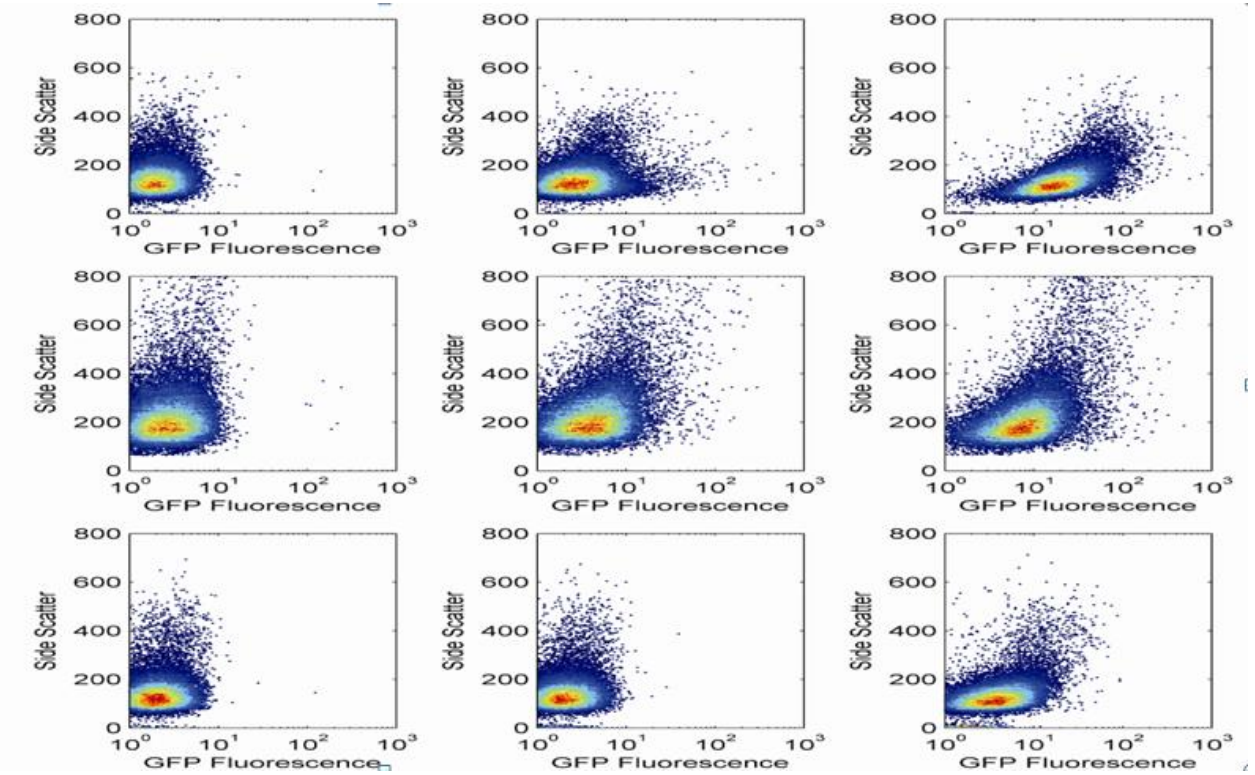
tetO2

tetO2

RBS

Flow cytometry





Reaction Network

2 TetR1 → TetR2
TetR2 → 2 TetR1

TetR2 + tetO2 → TetR2:tetO2
TetR2:tetO2 → TetR2 + tetO2

RNAp + tetP → RNAp:tetP
RNAp:tetP → RNAp + tetP

RNAp:tetP → RNAp:tetP*
tetO2 + RNAp:tetP* → tetP + RNAp:tetO2

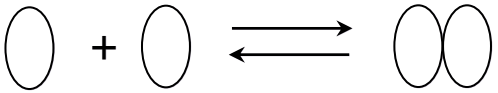
RNAp:tetO2 → RNAp + tetO2
RNAp:tetO2 → tetO2 + RNAp:DNA_lac
RNAp:DNA_lac → RNAp + mRNA_lac

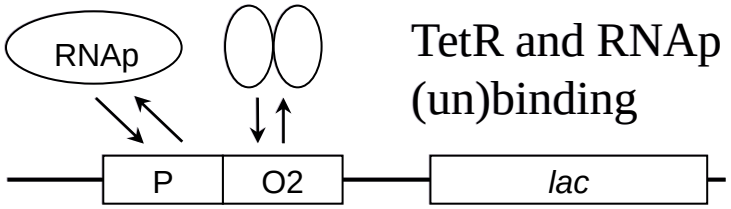
rib + mRNA_lac → rib:mRNA_lac
rib:mRNA_lac → mRNA_lac + rib:mRNA_lac1
rib:mRNA_lac1 → LacR1 + rib + mRNA_lac1

RNase + mRNA_lac → RNase:mRNA_lac
RNase:mRNA_lac → RNase

DTet→
TetR1 →
TetR2 →

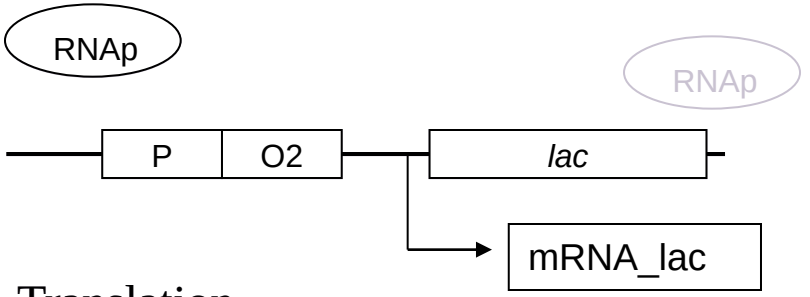
dimerization



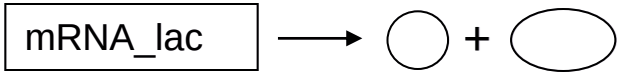


TetR and RNAp (un)binding

Opening of DNA, Transcription



Translation



Degradation of proteins and mRNA

Synthetic Reaction Network

Number	General Transcription & Translation Reactions	k	Source
1	RNAp + lacP + lacO1 + tetO1 + tetO2 → RNAp:lacP	1.00E+07	30
2	(see below)		
3	(see below)		
4	RNAp:lacP → RNAp:lacP*	0.01	31
5	RNAp:lacP → RNAp + lacP + lacO1 + tetO1 + tetO2	1	30
6	RNAp:lacP* → lacP + lacO1 + tetO1 + tetO2 + RNAp:DNAgfp	30	32
7	RNAp:DNAgfp → RNAp + gfp_mRNA	30	32 §
8	gfp_mRNA + rib → rib:gfp_mRNA	100000	¶
9	rib:gfp_mRNA → rib:gfp_mRNA_1 + gfp_mRNA	33	32
10	rib:gfp_mRNA_1 → rib + gfp	33	32 §
	LacI Repression at Lac Operator		
11	lacI4 + lacO1 → lacI4:lacO1	2E+09	27
12	lacI4:lacO1 → lacI4 + lacO1	4.00E-04	27
13	lacI4 + IPTG → lacI4:IPTG	4.60E+06	27
14	lacI4:IPTG → lacI4 + IPTG	0.2	27
15	lacI4:lacO1 + IPTG → lacI4:lacO1:IPTG	1.00E+06	27
16	lacI4:lacO1:IPTG → lacI4:lacO1 + IPTG	0.8	27
17	lacI4:IPTG + lacO1 → lacI4:lacO1:IPTG	2E+09	27
18	lacI4:lacO1:IPTG → lacI4:IPTG + lacO1	0.4	27
	TetR Repression, 1st Tet Operator		
19	tetR2 + aTc → tetR2:aTc	100000000	28 *
20	tetR2:aTc → tetR2 + aTc	0.001	28 *
21	tetR2:aTc + aTc → tetR2:aTc2	100000000	28 *
22	tetR2:aTc2 → tetR2:aTc + aTc	0.001	28 *
23	tetR2 + tetO1 → tetR2:tetO1	100000000	29 *
24	tetR2:tetO1 → tetR2 + tetO1	0.001	29 *
25	tetR2:aTc + tetO1 → tetR2:tetO1:aTc	100000000	28 *
26	tetR2:tetO1:aTc → tetR2:aTc + tetO1	1	28 *,†
27	tetR2:aTc2 + tetO1 → tetR2:tetO1:aTc2	100000000	28 *
28	tetR2:tetO1:aTc2 → tetR2:aTc2 + tetO1	100000	28 *,†
29	tetR2:tetO1 + aTc → tetR2:tetO1:aTc	100000000	28 *
30	tetR2:tetO1:aTc → tetR2:tetO1 + aTc	0.001	28 *
31	tetR2:tetO1:aTc + aTc → tetR2:tetO1:aTc2	100000000	28 *
32	tetR2:tetO1:aTc2 → tetR2:tetO1:aTc + aTc	0.001	28 *
	LacI / lacO Leakiness Reactions		
2	RNAp + lacP + lacI4:lacO1 + tetO1 + tetO2 → RNAp:lacP + lacI4	6.23E+05	‖
3	RNAp + lacP + lacI4:lacO1:IPTG + tetO1 + tetO2 → RNAp:lacP + lacI4:IPTG	6.23E+05	‖

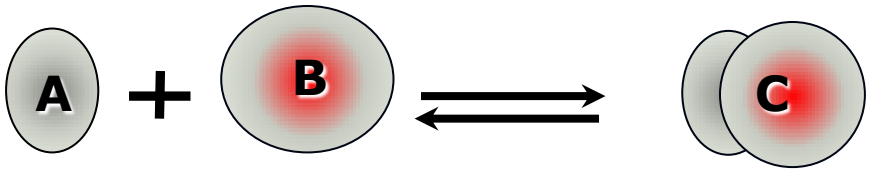
Number	TetR Repression, 2nd Tet Operator	k	Source
33	tetR2 + tetO2 → tetR2:tetO2	100000000	29 *
34	tetR2:tetO2 → tetR2 + tetO2	0.001	29 *
35	tetR2:aTc + tetO2 → tetR2:tetO2:aTc	100000000	28 *
36	tetR2:tetO2:aTc → tetR2:aTc + tetO2	1	28 *,†
37	tetR2:aTc2 + tetO2 → tetR2:tetO2:aTc2	100000000	28 *
38	tetR2:tetO2:aTc2 → tetR2:aTc2 + tetO2	100000	28 *,†
39	tetR2:tetO2 + aTc → tetR2:tetO2:aTc	100000000	28 *
40	tetR2:tetO2:aTc → tetR2:tetO2 + aTc	0.001	28 *
41	tetR2:tetO2:aTc + aTc → tetR2:tetO2:aTc2	100000000	28 *
42	tetR2:tetO2:aTc2 → tetR2:tetO2:aTc + aTc	0.001	28 *
	Nonspecific DNA Interactions		
43	lacI4 + nsDNA → lacI4:nsDNA	1000	33 *
44	lacI4:nsDNA → lacI4 + nsDNA	0.0041667	33 *
45	lacI4:IPTG + nsDNA → lacI4:IPTG:nsDNA	1000	33 *
46	lacI4:IPTG:nsDNA → lacI4:IPTG + nsDNA	0.0041667	33 *
47	tetR2 + nsDNA → tetR2:nsDNA	1000	33 *
48	tetR2:nsDNA → tetR2 + nsDNA	3.2409	33 *
49	tetR2:aTc + nsDNA → tetR2:aTc:nsDNA	1000	33 *
50	tetR2:aTc:nsDNA → tetR2:aTc + nsDNA	3.2409	33 *
	Degradation and Dilution Reactions		
51	→ tetR2	1.00E-11	‖
52	tetR2 →	2.89E-04	‖
53	tetR2:aTc → aTc	2.89E-04	‖
54	tetR2:aTc2 → 2 aTc	2.89E-04	‖
55	→ lacI4	1.00E-09	‖
56	lacI4 →	2.89E-04	‖
57	lacI4:IPTG → IPTG	2.89E-04	‖
58	gfp_mRNA →	1.16E-03	¶
59	gfp →	3.21E-05	‡
60	lacI4:nsDNA → nsDNA	1.93E-04	**
61	lacI4:IPTG:nsDNA → nsDNA + IPTG	1.93E-04	**
62	tetR2:aTc:nsDNA → nsDNA + aTc	1.93E-04	**
63	tetR2:nsDNA → nsDNA	1.93E-04	**

Reference Key:

A. Levandowski *et. al.*, 1996
B. Gilbert and Müller-Hill, 1970
C. Vogel and Jensen, 1994
D. Sorensen and Pedersen, 1991
E. Elowitz and Leibler, 2000
F. Kędracka-Krok, 1999
G. Bertrand-Burggraf *et. al.*, 1984
H. Stickle *et. al.*., 1994

Chemical Kinetics Models

- Represent interactions with chemical reactions.



$$A + B \xrightarrow{k_1} C$$

$$C \xrightarrow{k_2} A + B$$

- Reaction rate laws.
- Ordinary differential equations.

$$\frac{dC_A}{dt} = -k_1 C_A C_B + k_2 C_C$$

$$\frac{dC_B}{dt} = -k_1 C_A C_B + k_2 C_C$$

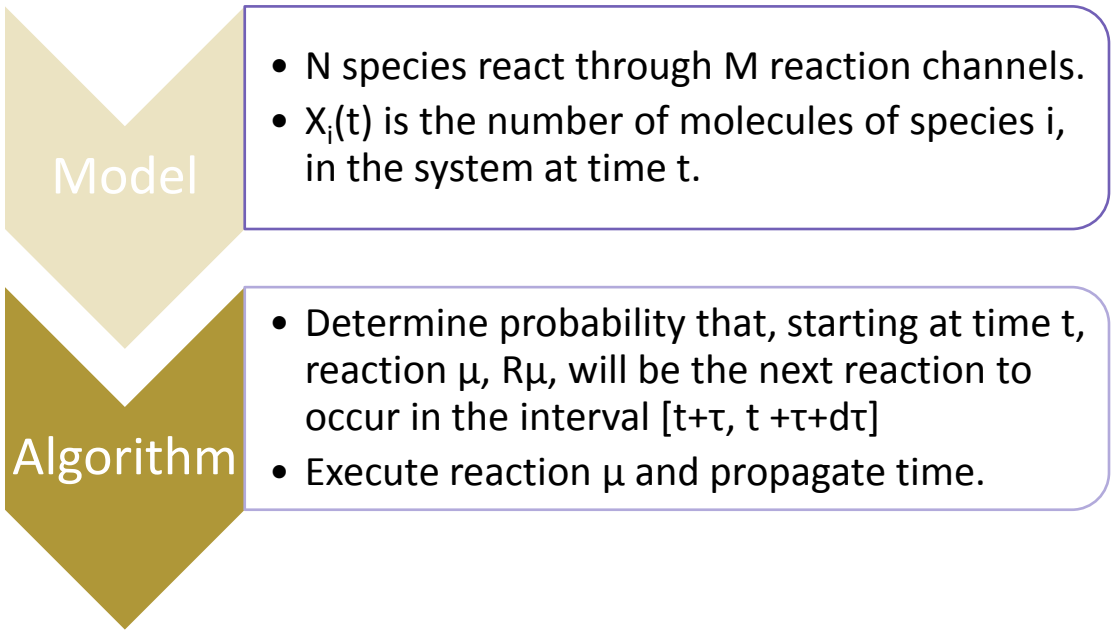
$$\frac{dC_C}{dt} = k_1 C_A C_B - k_2 C_C$$

- Far from the thermodynamic limit. Stochastic chemical kinetics
- Master equation formalism (McQuarrie, 1949; Oppenheim, 1965; Fredrickson, 1963).

$$\frac{\partial P(\underline{X})}{\partial t} = \int P(\underline{X}')\, W(\underline{X}' \mid \underline{X}) - P(\underline{X})\, W(\underline{X} \mid \underline{X}')\; d\underline{X}'$$

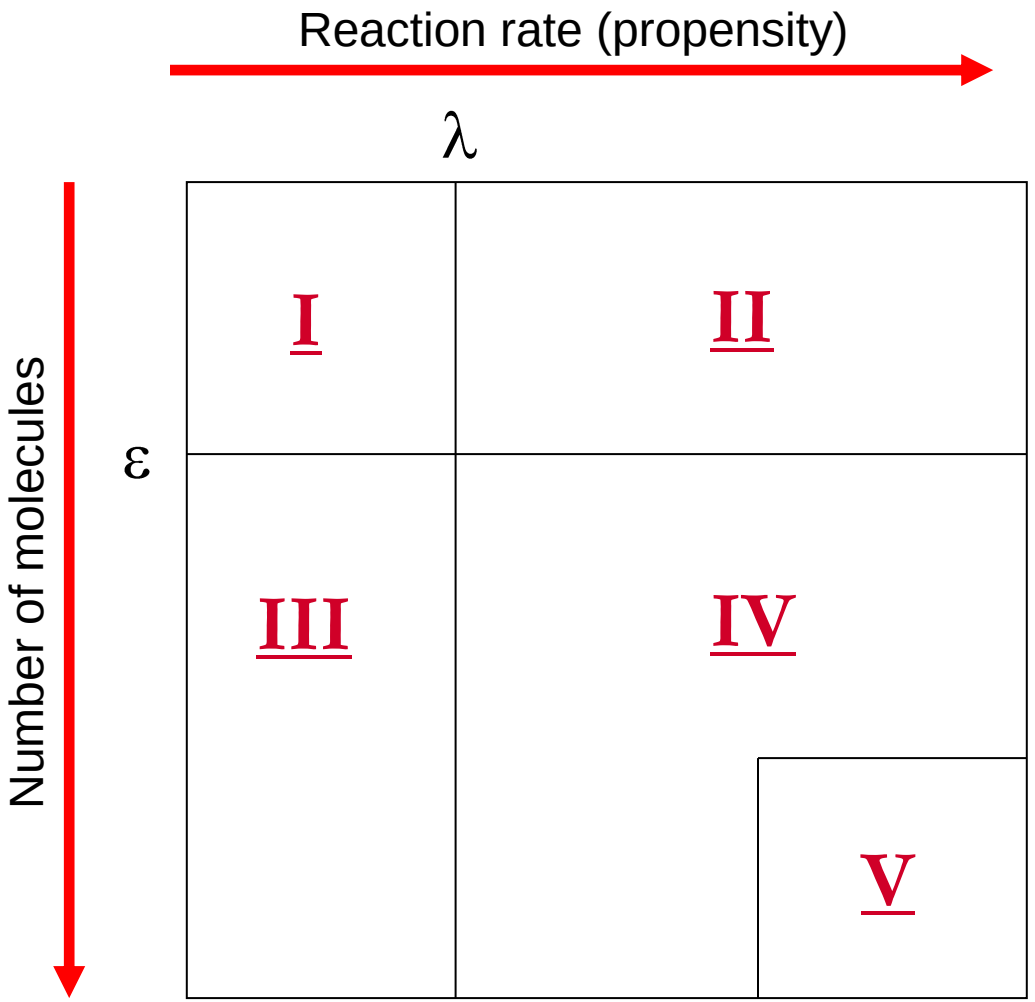
Stochastic Simulation Algorithm

- Stochastic simulation algorithm (Gillespie, 1976) samples the probability distribution.



- The system may contain rare, discrete, but critical events *and* continuously occurring deterministic or stochastic transitions.
- Simulation using the SSA will be *very* slow. Computational time scales with the number of reaction occurrences .

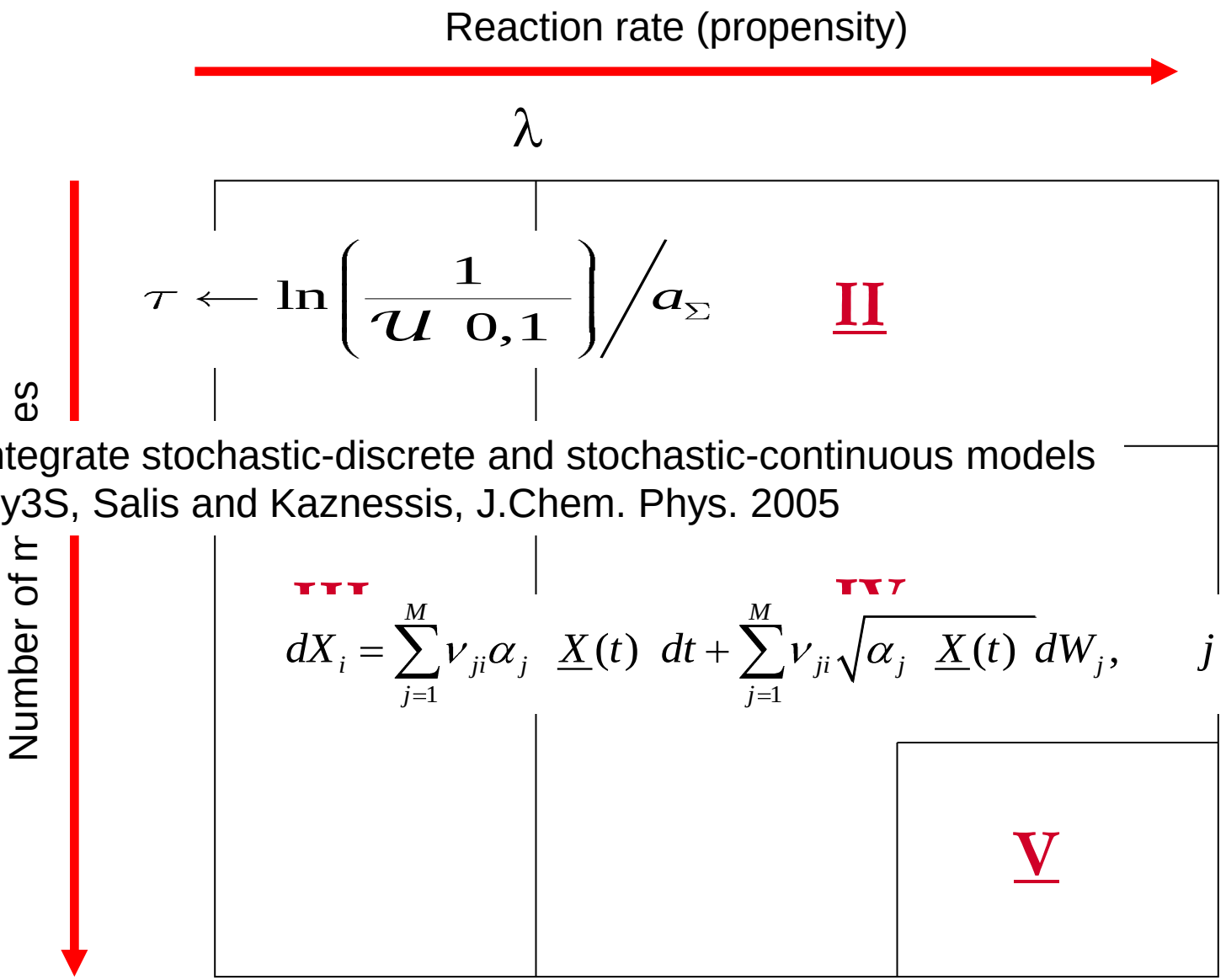
Multiscale Modeling Framework



- I: Discrete / Stochastic
  - Jump Markov process
  - Stochastic simulation algorithm (Gibson and Bruck, 2000)
- II: Discrete / Stochastic
  - Tau-Leaping (Cao, Petzold, Gillespie, 2005)
  - Probabilistic steady state
- III-IV: Continuous / Stochastic
  - Valid continuous Markov process
  - Chemical Langevin equation
  - (Gillespie, 2001; Haseltine and Rawlings, 2002)
- V: Continuous / Deterministic
  - Valid ordinary differential equations (Amundson, 1966)

Petzold, Gillespie, Cao, Vlachos, Kevrekidis, Vanden-Eijnden, Arkin, Khammash

Multiscale Modeling Regimes



Speed Comparisons with SSA

The Cycle Test

System Size proportional to the number of reactant molecules of fast reactions

Ratios of Computational Run Times	
System Size	T^{SSA}/T^{NRH}
100	7.94
1000	95.59
10,000	986.8
100,000	**16912**

$$A \xrightarrow{k_1} B$$

$$B \xrightarrow{k_2} C$$

$$C \xrightarrow{k_3} D$$

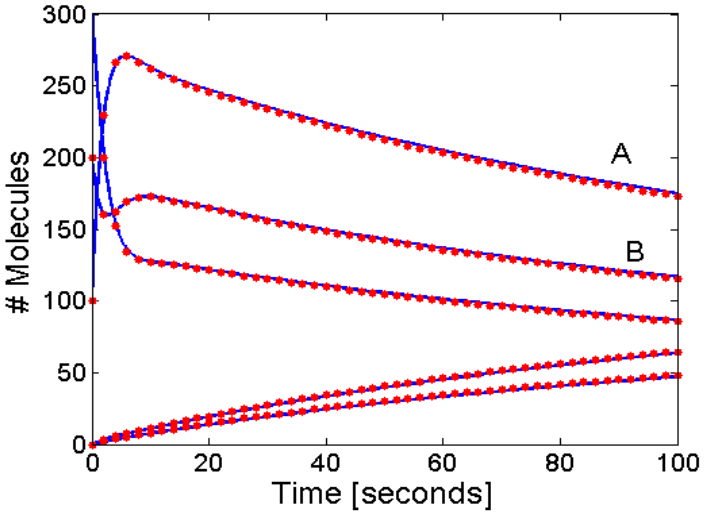
$$A + C \xrightarrow{k_4} D$$

$$B + C \xrightarrow{k_5} E$$

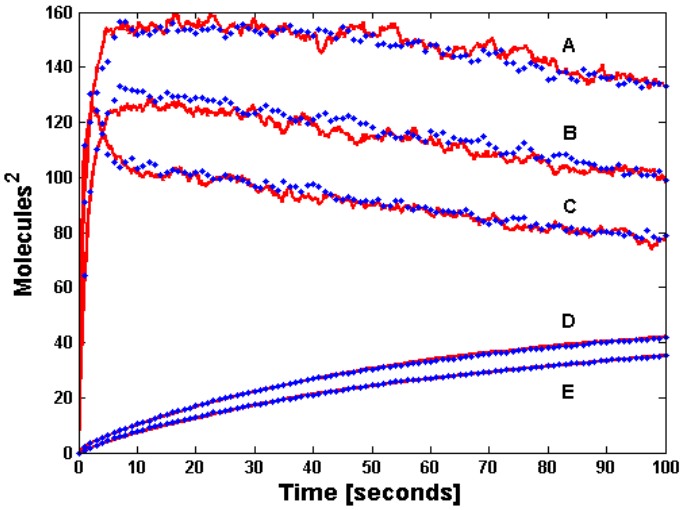
$$k_1, k_2, k_3 << k_4, k_5$$

Large scale benchmark in Salis and Kaznessis J. Chem. Phys. 2005a

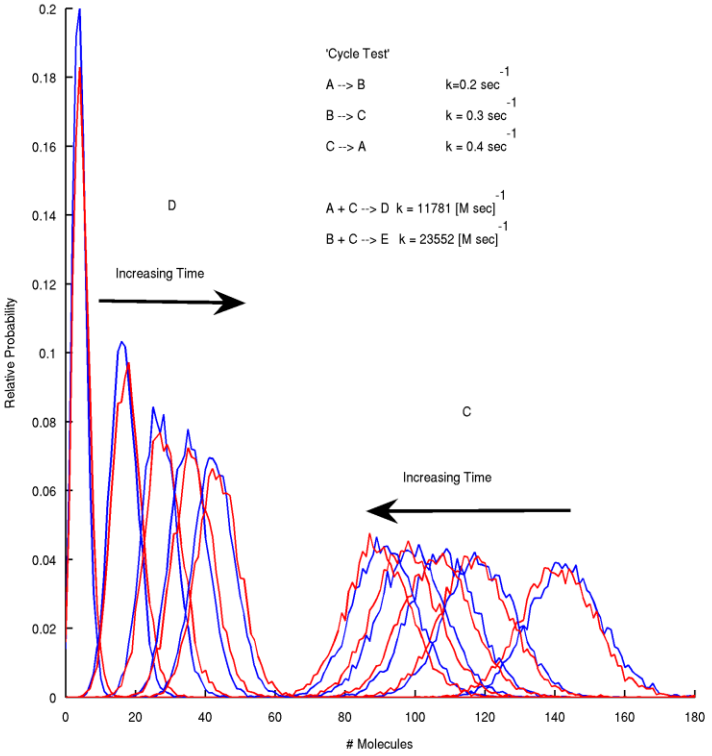
Accuracy: A Cycle Test



Mean

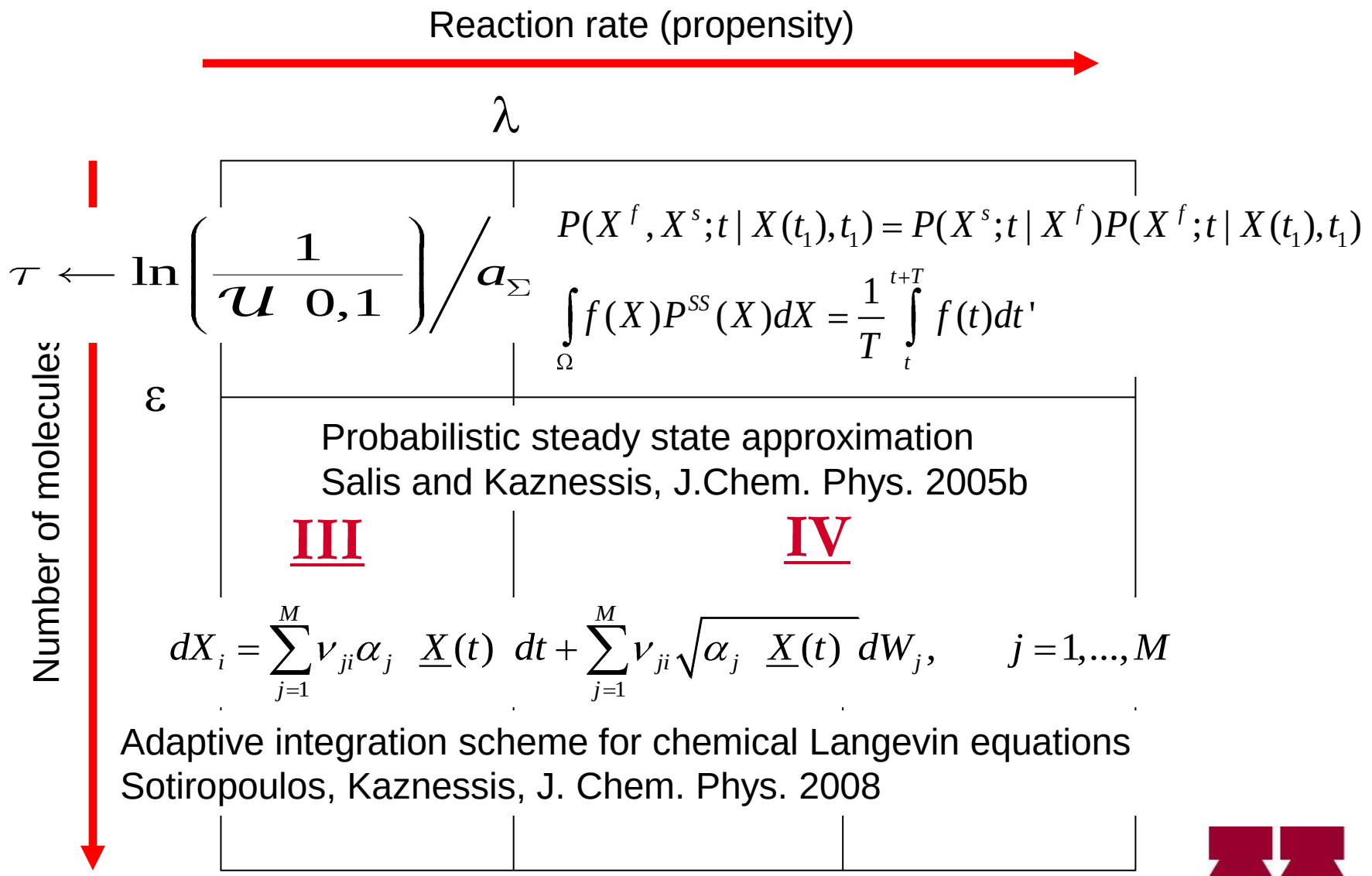


Probability Distribution



Variance

Modeling Regimes



Logical AND Gate Simulations

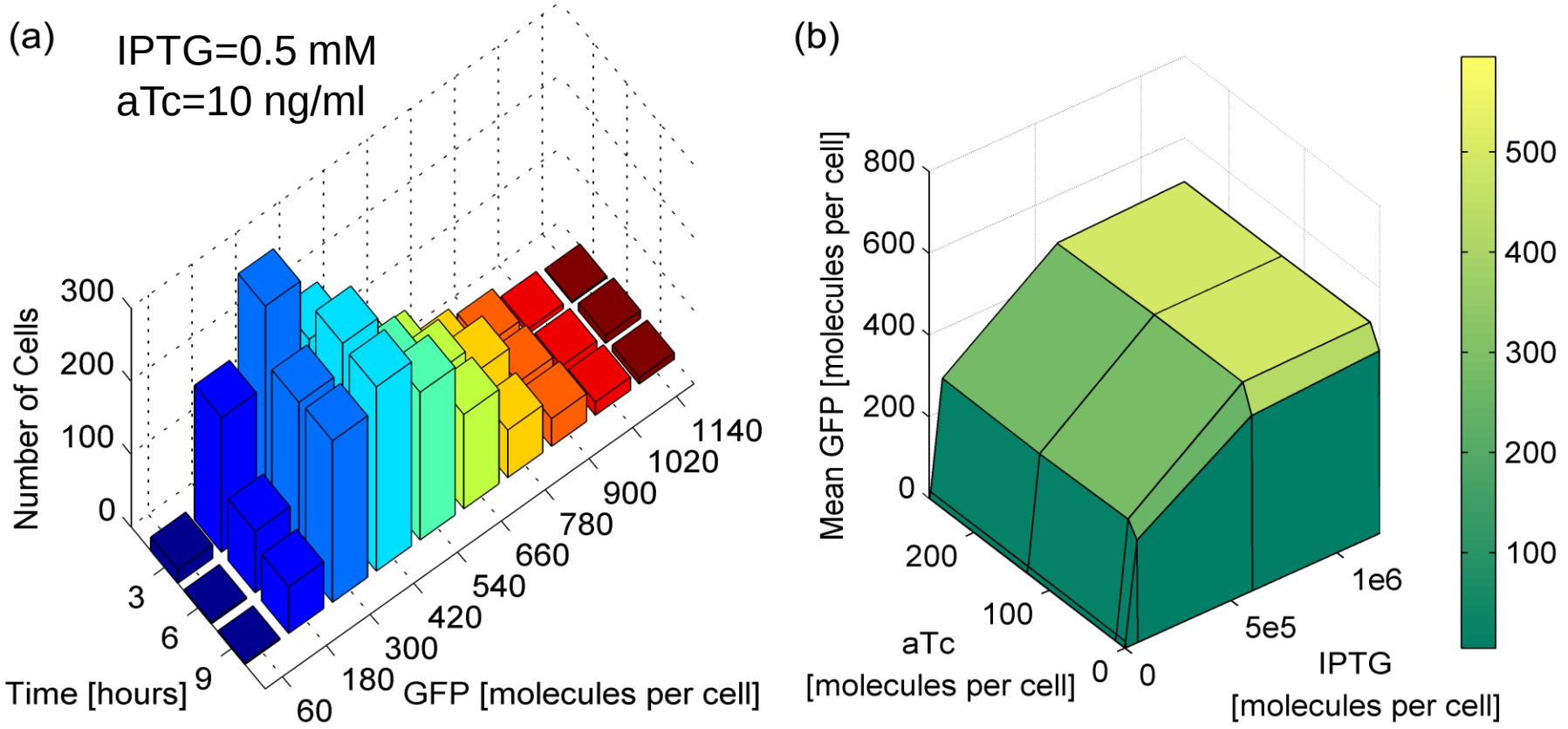
• Network with 63 reactions. Simulate a grid of 6x6 aTc-IPTG pair concentrations (0-200 ng/ml and 0-2mM). Simulate 1,000 trajectories for each pair. Simulate six designs (LLT, TTL).

•Species are uniformly distributed in the cell. Initial cell volume is 10^{-15} L. Cell division occurs every 30±5 minutes: the volume doubles exponentially and then halves.

• Multiscale simulations including stochasticity.

•Measure GFP number of molecules for 216,000 trajectories (36,000 CPU hours).

Stochastic simulations

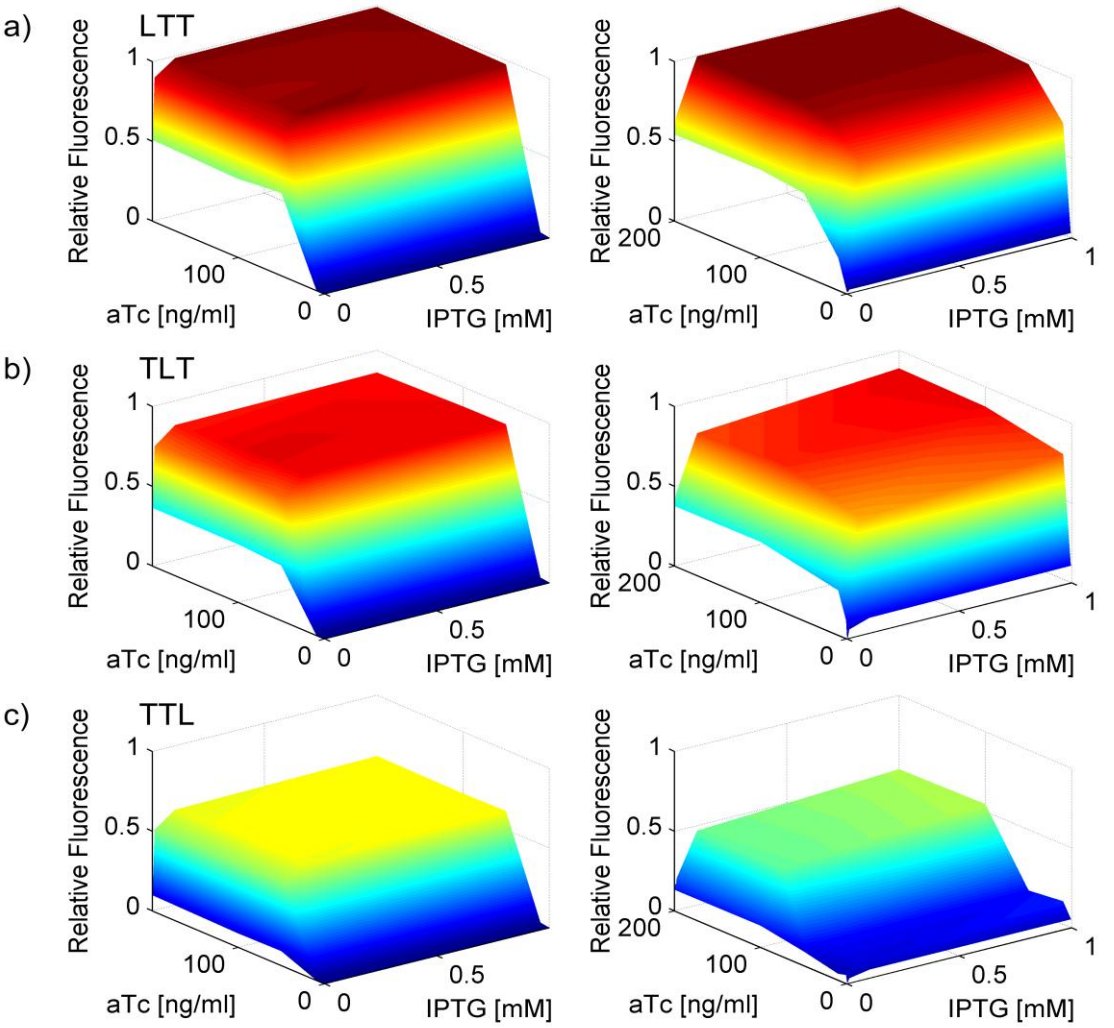


Computer-Aided Design of Bio-Logical AND Gates

•TTL is the highest-fidelity AND gate.

• Leakage of lacO can explain the variable phenotypic behavior. Biological insight.

• Mathematical models capture experimental phenotype in a way fit for analysis and design.



Ramalingam, Tomshine, Maynard, Kaznessis, Biochem. Eng. J. (2009)

Computational Synthetic Biology -

www.SynBioSS.org

- Synthetic biological systems confer advantages:
  1. They are small and well-defined to be captured by universal yet tractable models.
  2. They are modular, allowing us to build complex logical and informational architectures.
  3. They are our designs, not nature's, letting us avoid some of the historical difficulties.

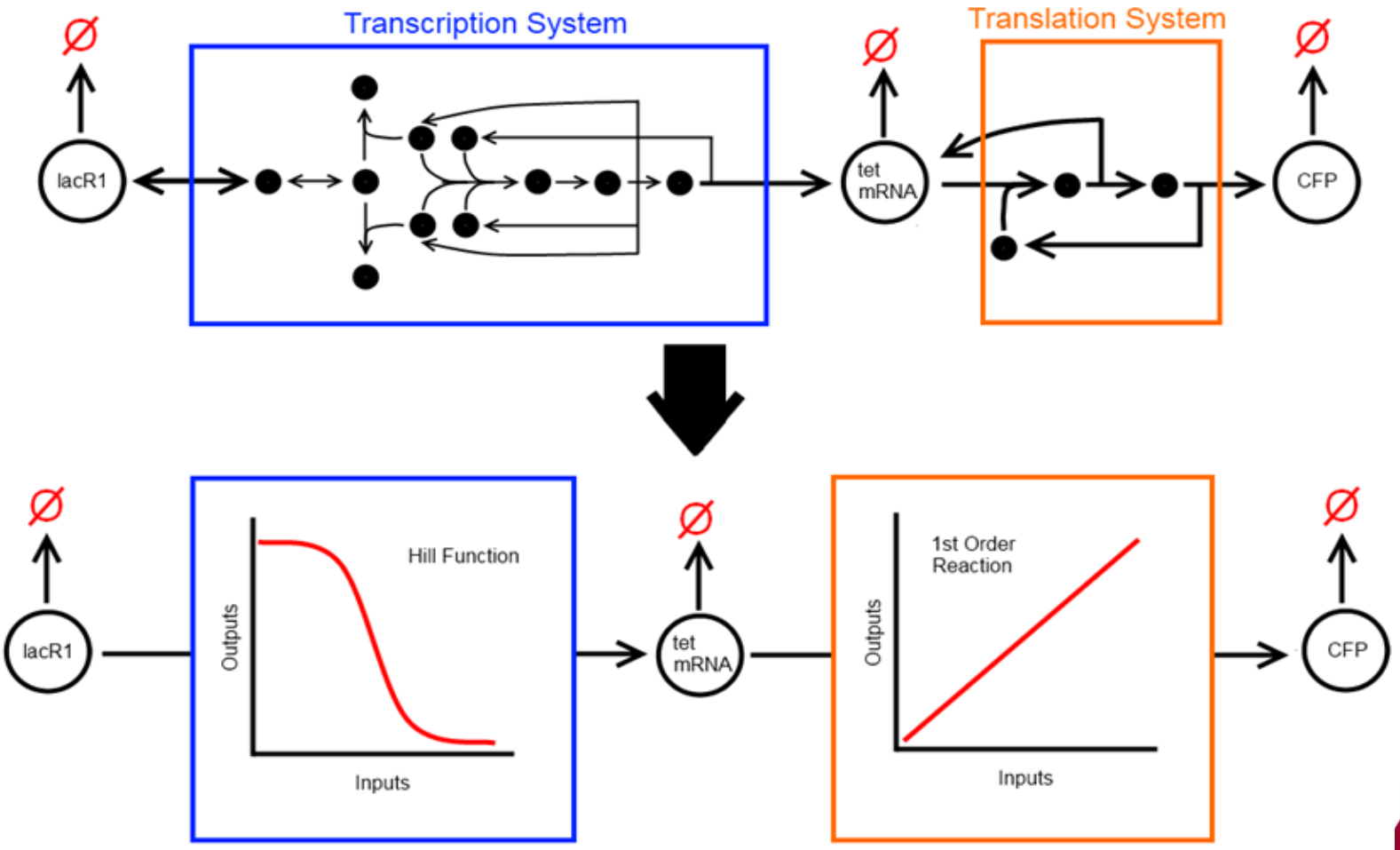
*Tuttle, et al., Biophys. J. (**2005**)*
*Salis, Kaznessis, Phys. Biol. (**2007**)*
*Salis, et al., BMC Systems Biology, (**2007**)*
*Tomshine, Kaznessis, Biophys. J. (**2006**)*
*Sotiropoulos, Kaznessis, BMC Systems Biology, (**2007**)*
*Kaznessis, Biotechnology Journal, (**2009**)*
*Weeding, et al., Briefings in Bioinformatics (**2010**)*

- Challenges: constants not known; molecular biology not known; assumption of context free system not valid.
- Can we start with a full model and reduce it to a few equations?

*Sotiropoulos, et al., IEEE/ACM Trans Comput Biol Bioinform. (**2009**)*

Reduced Models for Synthetic Biology

By reducing the model a relatively simple model (15 reactions, 9 components) can be determined with more complex reaction rate equations. Use in microbial community models

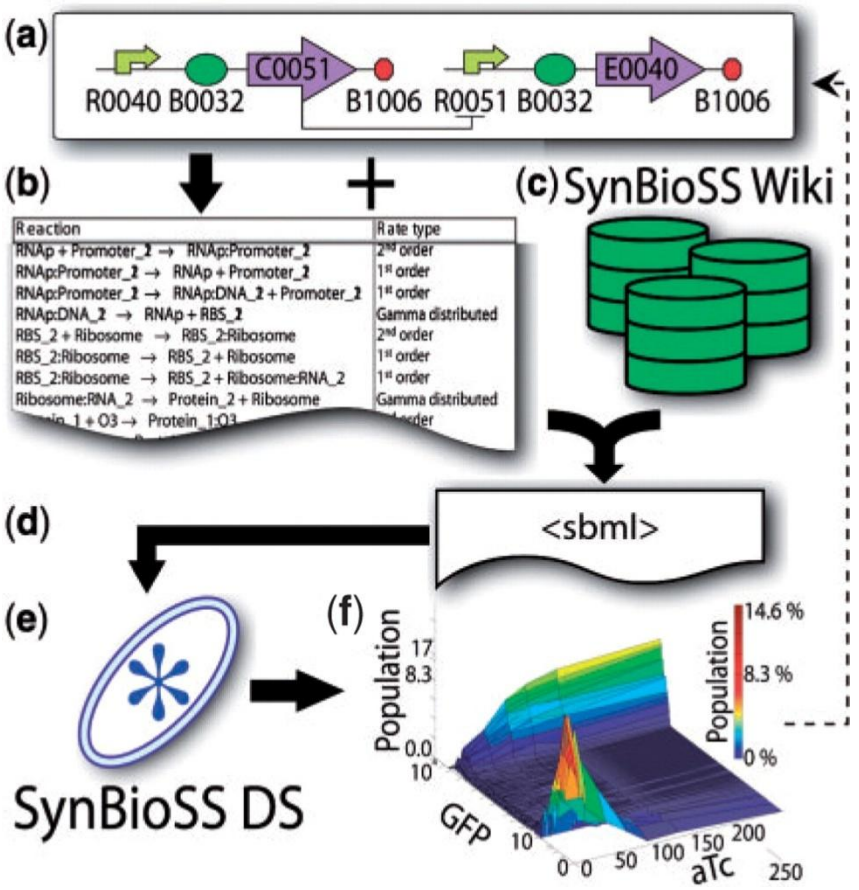


Hysss - SynBioSS

hysss.sourceforge.net

www.SynBioSS.org

- Tool for generation, curation and simulation of synthetic biological networks. Three components:
  - Designer: Reaction network generation for arbitrary synthetic construct
  - Wiki: Kinetic data storage/retrieval. Community driven effort
  - Desktop Simulator: Numerical simulation with multiscale algorithms
- Goal
  - **Directly connect DNA sequences with dynamic phenotypes**

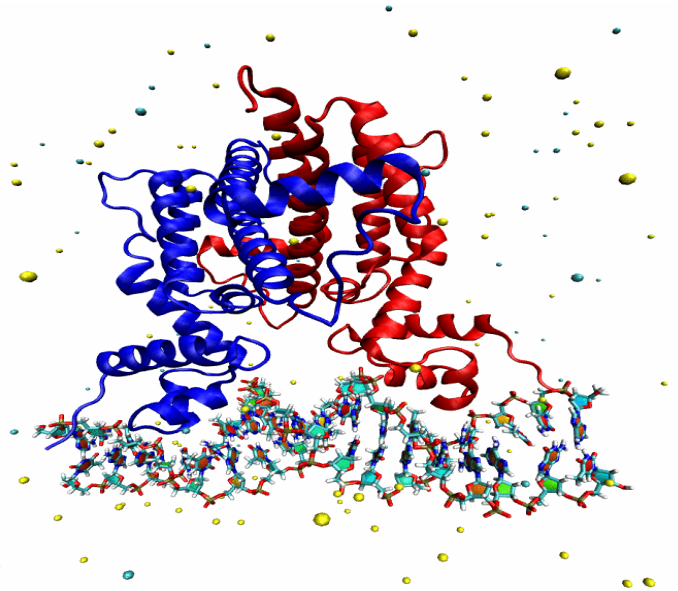


Hill, et al. Bioinformatics, (**2008**)
Salis, et al. BMC Systems Biology, (**2007**)

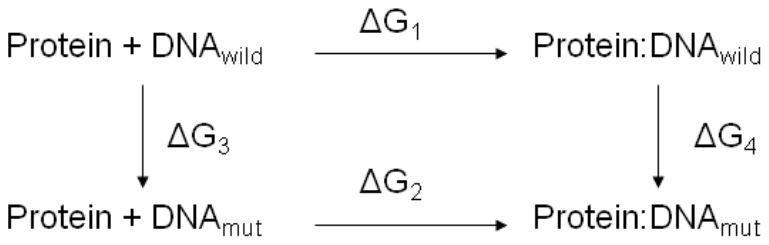
Molecular Models: TetR-TetO binding

- Determine the free energy of binding between the repressor and its cognate DNA sequence.
- Determine the change of the binding strength upon mutation.

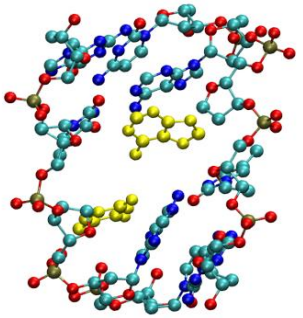
Name	Sequence	Reporter activity* (%)	$\Delta\Delta G=\Delta G_4-\Delta G_3$ kcal/mol
Wild type	TCCCTATCAGTGATAGAGA	0	**0**
M3	TCCCTA**A**CAGTG**T**TAGAGA	52	**-7.76**
M4	TCCCT**T**TCAGTGA**A**AGAGA	2	**-2.91**
M6	TCC**G**TATCAGTGATA**C**AGA	20	**-6.6**



Zwanzig, R. W. J. Chem. Phys. (1954)

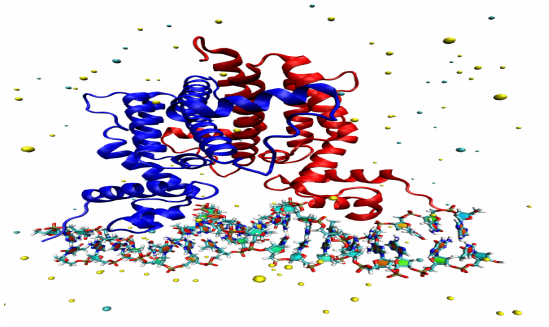


$$\Delta G_{A\to B} = -k_B T \sum_{i=1}^{N} \ln \left\langle \exp\left(-\frac{E_{\lambda_{i+1}} - E_{\lambda_i}}{k_B T} \right) \right\rangle_{\lambda_i}$$

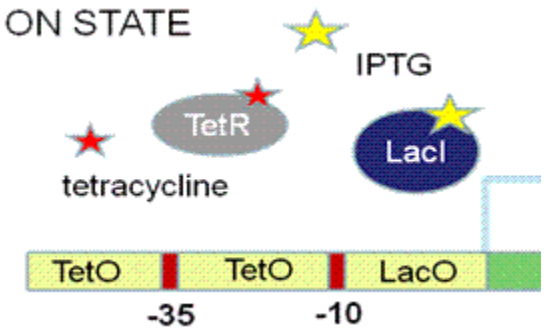


How does life emerge from a soup of chemicals?

- Atoms, molecules

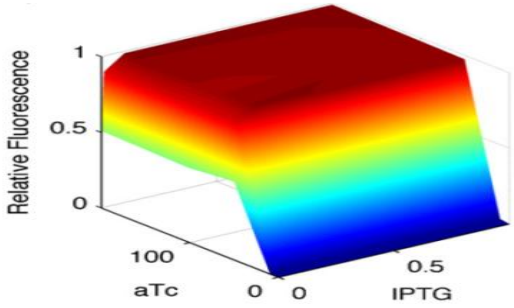


- Pairwise interactions



- Networks of biomolecular interactions

- Informational and logical architectures
  (Paul Nurse, Nobel Prize Medicine, 2001)



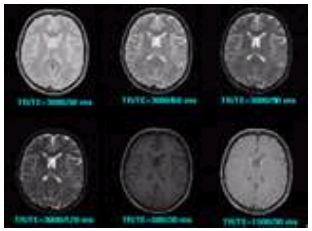
- Development, differentiation, metabolism, immune responses, etc.

- Life (biological systems) defined by three major features
  (Jacques Monod, Nobel Prize Medicine, 1976):
  1. Reproductive invariance
  2. Autonomous morphogenesis
  3. Teleonomy

The role of high-performance computing

Computational mathematics enhance the unaided human brain in two major ways:

1. Extrapolation, related to innate human computing capacities.



2. Augmentation, past innate human computing abilities. Augmentation can be considered as an important shift in human abilities.

With extrapolation and augmentation, high-performance computing provides access to tractable mathematics and solutions previously inaccessible to the human brain..

With results attainable only with computer simulations, the foundation can be solidly laid for mathematical models that capture and explain biological phenomena (Humphreys: “Extending Ourselves”.

Images from http://www.stanleylondon.com/tele13prem.htm and http://iacl.ece.jhu.edu/projects/pdf/

Acknowledgements

Kostantinos Biliouris
Jon Tomshine (tools)
Vassilis Sotiropoulos (theory)
Dan Bolintineanu (simulations)
John Barrett
Anushree Chaterjee
Kat Volzing (experiments)
Dr. Chandrika Mulakala
Dr. Abdallah Sayyed-Ahmad
Dr. Tony Hill (tools)
Dr. Poonam Srivastava (experiments)
Dr. Victor Vivcharuk

Dr. Himanshu Khandelia
Dr. Howard Salis (theory)
Dr. Spyros Vicatos
Dr. Allison Langham
Dr. Vijay Reddy
Dr. Yuhua Duan
Dr. Ting-Lan Chiu
Dr. Kavita Iyer (experiments)
Dr. Hairong Wei

Supported by:
NIH (R01GM070989 and R01GM086865),
NSF CAREER
Army Research Laboratory
3M
IBM
University of Minnesota Biotechnology Institute

Computational Resources:
National Center for Supercomputing Applications
Minnesota Supercomputing Institute.