Multiscale Models for Synthetic Biology

Yiannis N. Kaznessis

University of Minnesota
Department of Chemical Engineering and Materials Science

Acknowledgements

Jon Tomshine
Vassilis Sotiropoulos
Dan Bolintineanu
John Barrett
Anushree Chaterjee
Kat Volzing
Costas Billliouris
Ben Swiniarski
Dr. Chandrika Mulakala
Dr. Abdallah Sayyed-Ahmad
Dr. Tony Hill
Dr. Poonam Srivastava
Dr. Victor Vivcharuk

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

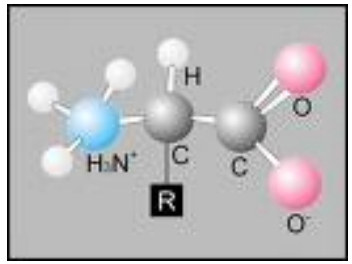
Ted Davis, Wei-Shou Hu, Jennifer Maynard,
Prodromos Daoutidis (CEMS)
George Vasmatzis (Mayo Clinic)
John Lambris (UPenn)
Bob Lehrer, Alan Waring (UCLA)
Sangtae Kim (Purdue)
Eric Fauman, Marc Snow (Pfizer)

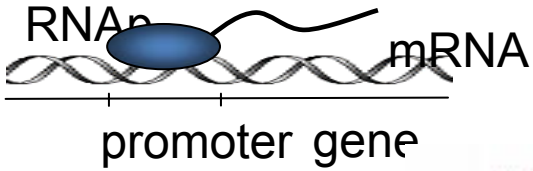
Supported by NIH, NSF CAREER, Army Research Laboratory, 3M, IBM, University of Minnesota Biotechnology Institute

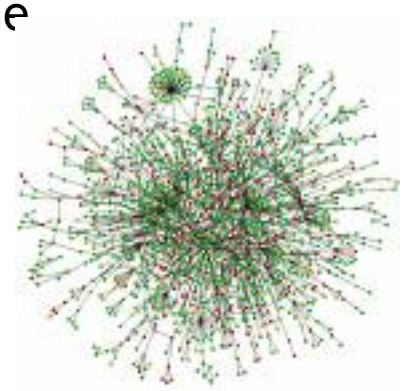
Computational Resources: National Center for Supercomputing Applications and Minnesota Supercomputing Institute.

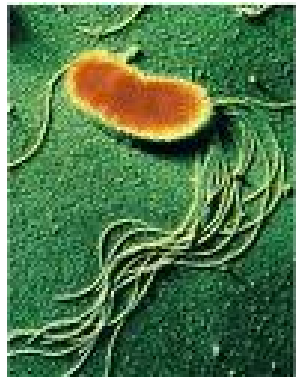
How does life emerge from a soup of chemicals?

- Atoms
- Molecular components (amino acids, nucleotide bases, lipids, metabolites)
- Macromolecular Components (proteins, DNA, RNA, membranes)
- Biomolecular Interactions
- Gene Networks
- Logical and Informational Architectures
- Environmental Context
- Teleonomy – Morphogenesis - Reproduction









Can we develop mathematical representations to describe and predict biological phenotypic complexity?

Can models be based on universal laws of thermodynamics and principles of molecular biology?

- Two major challenges:
  - The complexity is stupefying (number of components, nonlinear interactions, environment and context dependences)
  - Dobzansky's dictum: Biology is a discipline in history

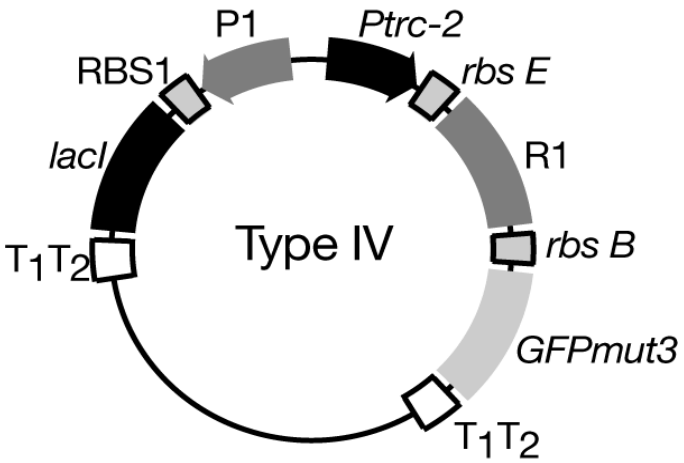
Gene network engineering

- With Genome Projects toolboxes available of
  - Regulatory proteins (activators and repressors)
  - Operator and promoter sites
  - Small inducer molecules

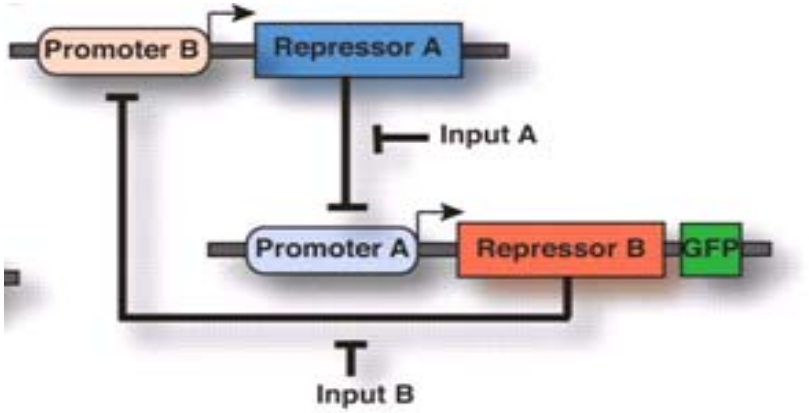
- DNA can be cut and pasted!
- Chemical synthesis of DNA
  - Inexpensive: $0.5/bp
  - Robust: construct 1000-long DNA strands without errors
- Novel gene regulatory networks are at hand.
- Synthetic biology: Forward engineering of biological systems (beyond traditional genetic engineering).

Examples of Synthetic Biology

Bistable switch, Gardner and Collins (2000)

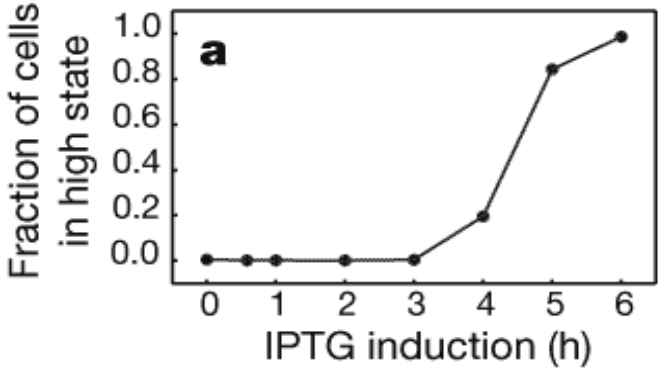


DNA plasmid sequences



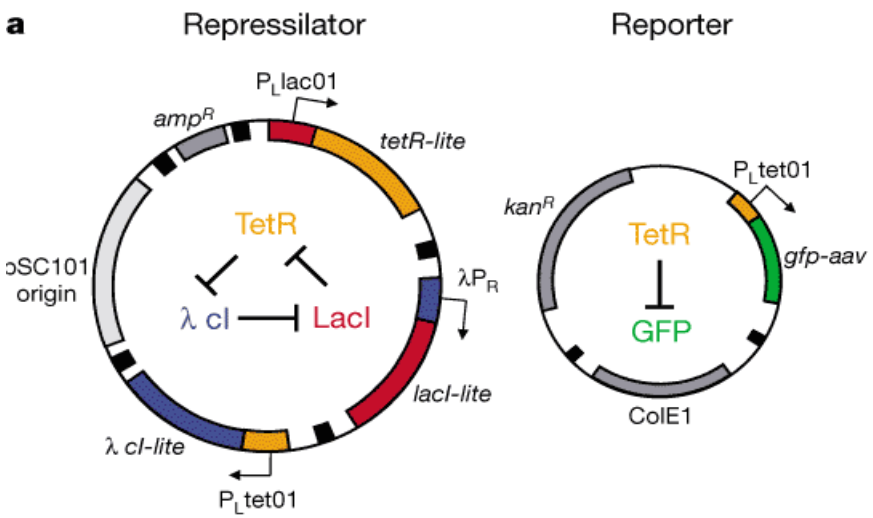
Network topology

Dynamic behavior

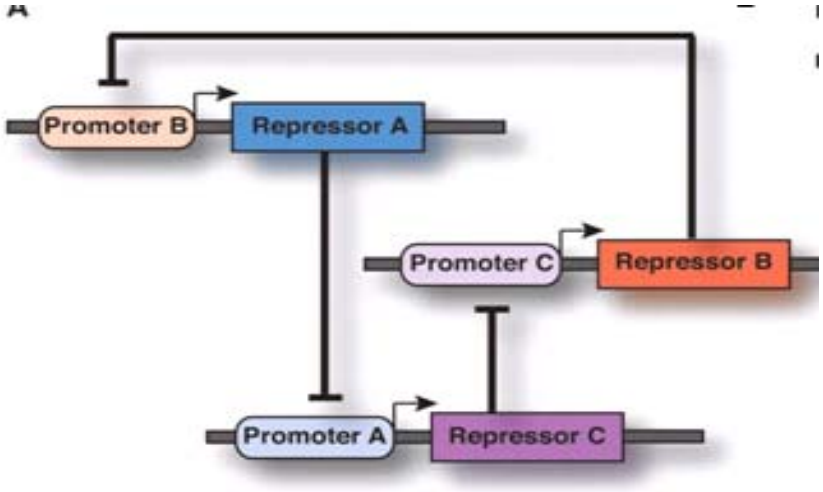


Examples of Synthetic Biology

Repressilator, Elowitz, Leibler (2000)

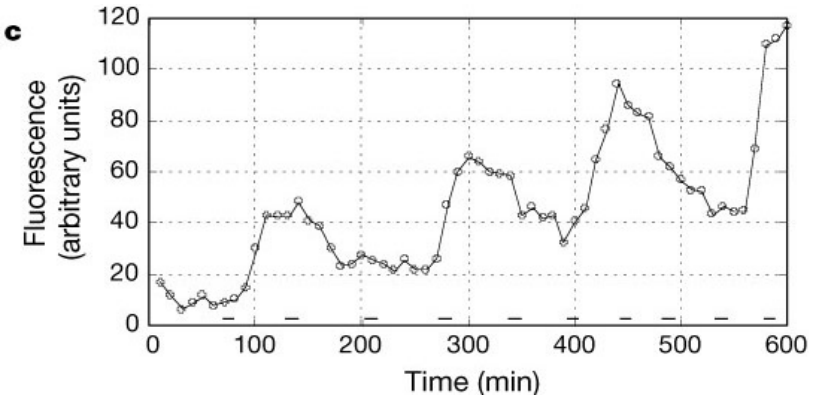


DNA plasmid sequences



Network topology

Dynamic behavior

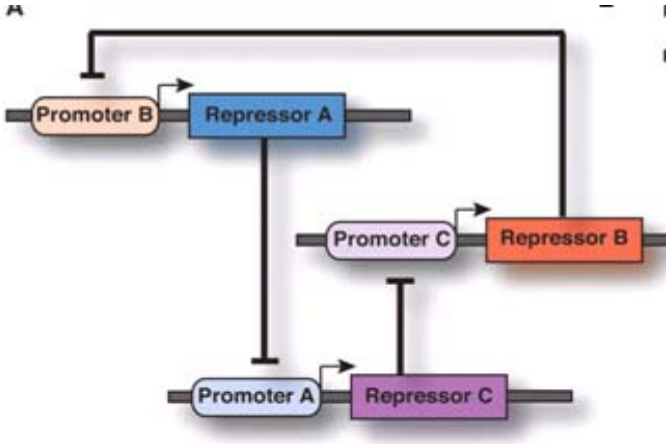


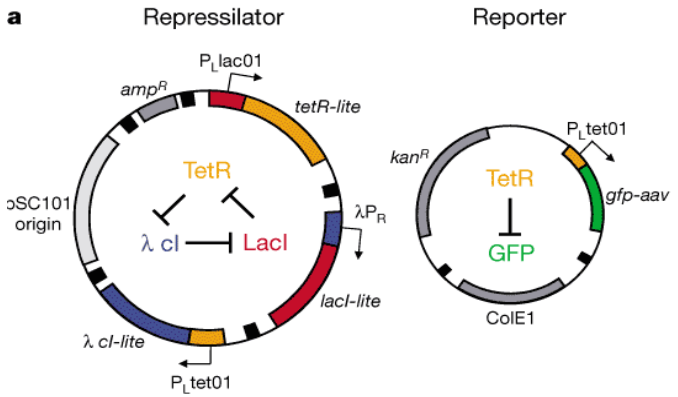
Synthetic Biology

- Controlling the temporal production of protein
  - Switches (e.g. if $C_{inducer}>0$ then turn on production of protein A)
  - Amplitude filters (e.g. if $C_{inducer}>C_{threshold}$ then turn on production of protein A)
  - Logical operators (e.g. if signal molecule 1 AND signal molecule 2 are present, then turn on production of protein A)
  - Numerous applications (e.g. gene therapy, biofuels, sensors, biosynthetic production optimization, biological computing)

Synthetic Biology Challenge

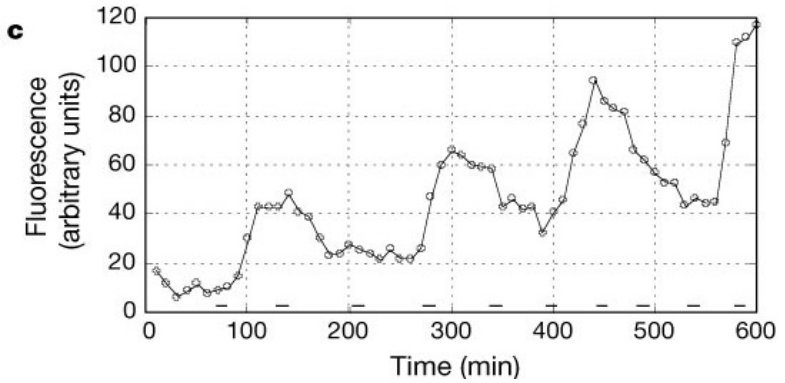
DNA sequences – network topologies





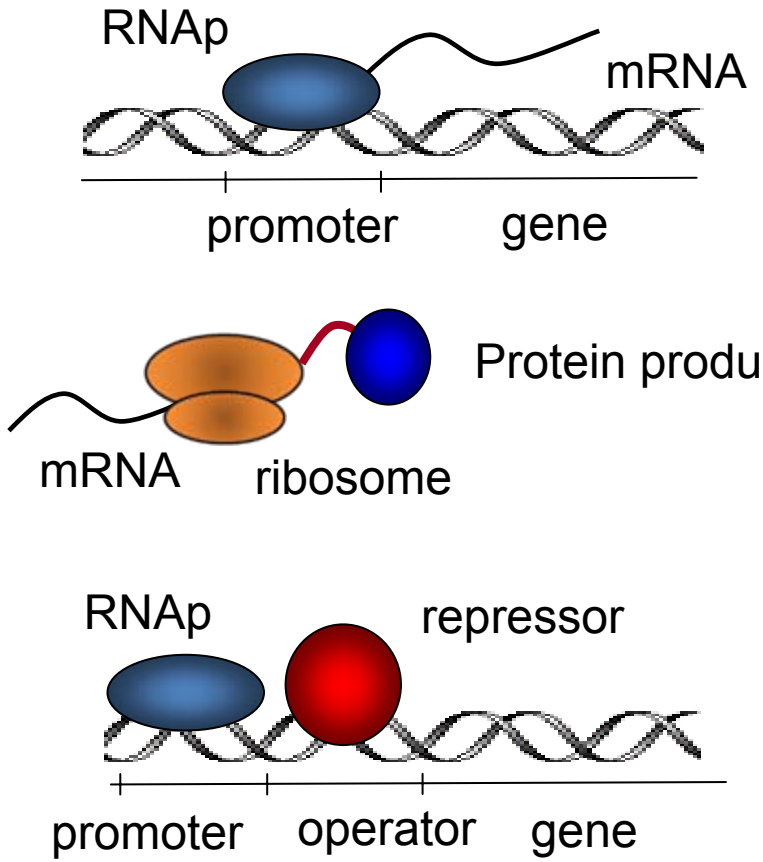
ATGGCATATGGTT

Phenotypic dynamic behavior



Modeling Gene Networks

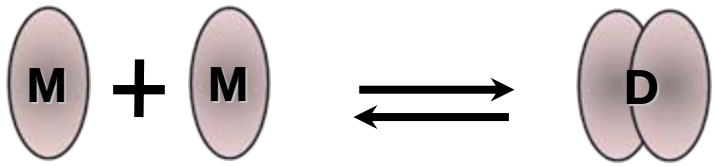
- Adopt Jacob's and Monod's postulate: all biological phenotypic complexity is the result of biomolecules interactions.



Chemical Kinetic Models

- Model all interactions at the molecular level
  - Protein Interactions
  - Transcription
  - Translation
  - Regulation

- Cascades of reactions represent interactions



$$M + M \xrightarrow{k_1} \mathrm{D}$$

$$D \xrightarrow{k_2} M + M$$

Stochastic Kinetics

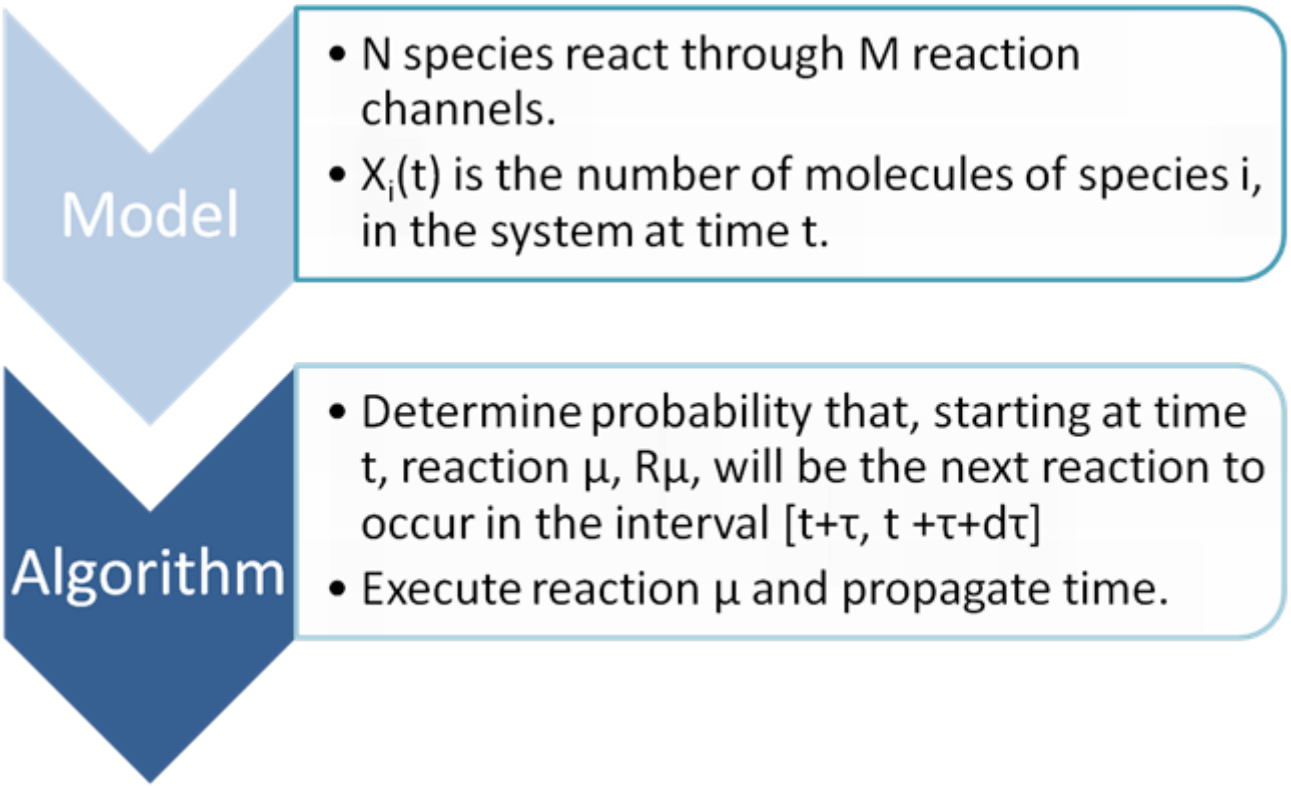
- Modeling cell functions
  - Many rare distinct events
  - Some participating species are sparse and diluted
  - Intrinsic fluctuations important

- Far from the thermodynamic limit: Stochastic chemical kinetics (McQuarrie, 1949; Oppenheim, 1965; Fredrickson, 1963)

- Kinetic Monte Carlo: Stochastic Simulation Algorithm (SSA)
  - Chemical Master equation (CME) instead of ODEs

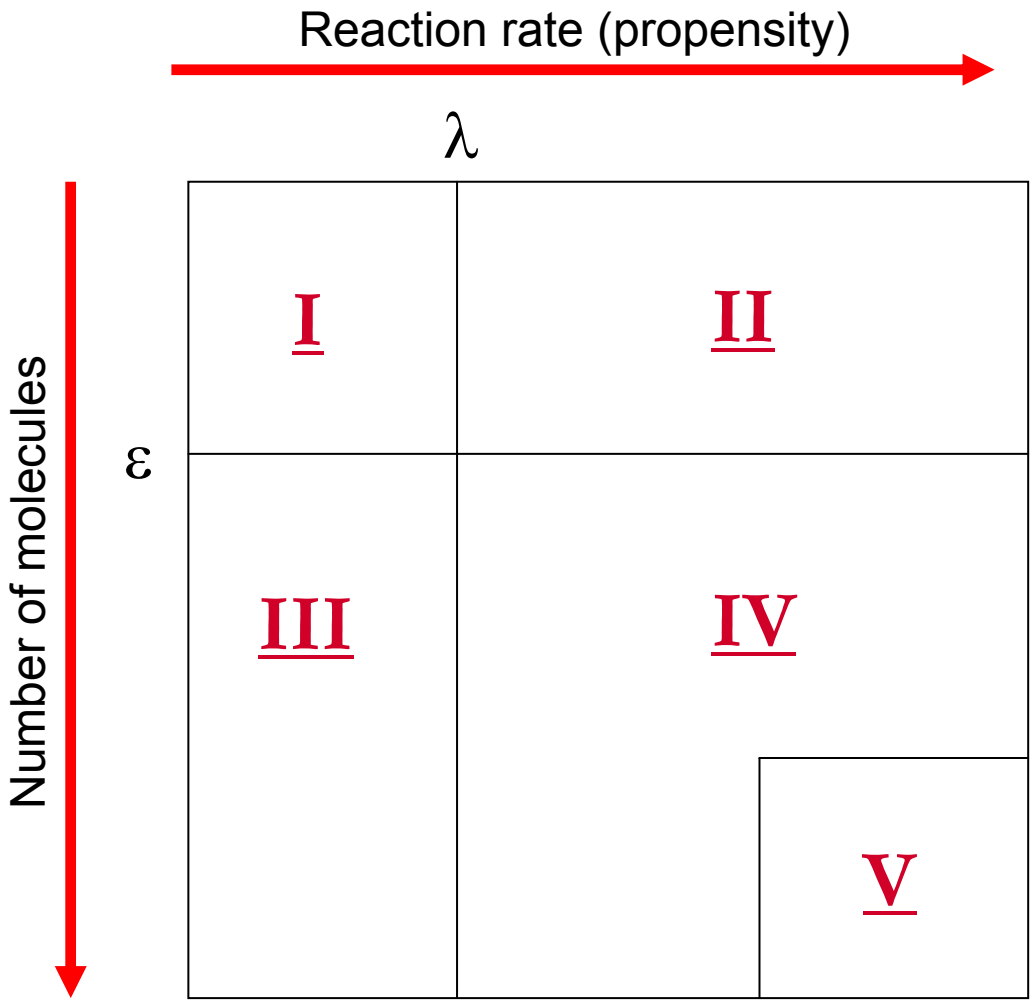
- Stochastic algorithms (Gillespie DT, 1976)

Stochastic Simulation Algorithm



- 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

Modeling Regimes

Reaction rate (propensity)

λ

es

$$t \neg \ln\left(\frac{1}{\mathcal{U}(0,1)}\right) / a_s$$

II

Integrate stochastic-discrete and stochastic-continuous models
Hy3S, Salis and Kaznessis, J.Chem. Phys. 2005

Nur

$$dX_i = \sum_{j=1}^{M} \nu_{ji}\alpha_j\left(\underline{X}(t)\right)dt + \sum_{j=1}^{M} \nu_{ji}\sqrt{\alpha_j\left(\underline{X}(t)\right)}dW_j, \qquad j = 1,\ldots,M$$

V

Hybrid Equations

- Partition into slow/discrete and fast/continuous reactions

- The effects of the fast/continuous reactions are described by Itô SDEs, called the chemical Langevin equation

$$dX_i = \sum_{j=1}^{M} \nu_{ji} \alpha_j \left(\underline{X}(t)\right) dt + \sum_{j=1}^{M} \nu_{ji} \sqrt{\alpha_j \left(\underline{X}(t)\right)} dW_j, \qquad j = 1,\ldots,M^{\,fast}$$

- The times of the slow/discrete reaction events are governed by a system of differential Jump equations, describing the time evolution of the reaction residuals, Rj
    - When Rj(t) = 0, then the jth reaction has occurred at time t.
    - These are also Itô SDEs, but without a Wiener process, W

$$dR_j = \alpha_j^{slow}\left(\underline{X}(t)\right) dt, \quad R_j(t_o) = \log(URN_j), \quad j = 1,\ldots,M^{\,slow}$$

SDE integration

- Euler-Maruyama Scheme, numerical error α $\sqrt{(\Delta t)}$

$$X_i(t+Dt) = X_i(t) + \sum_{j=1}^{M^{fast}} v_{ji} a_j^f(\underline{X}(t))Dt + \sum_{j=1}^{M^{fast}} v_{ji}\sqrt{a_j^f(\underline{X}(t))}DW_j$$

$$R_j(t+\Delta t) = R_j + a^s(\underline{X}(t))\,\Delta t$$

- Milstein Scheme, numerical error $O(\Delta t)$

$$X_i(t+\Delta t) = X_i(t) + \sum_{j=1}^{M^{fast}} v_{ji} a_j^f(\underline{X}(t))\Delta t + \sum_{j=1}^{M^{fast}} v_{ji}\sqrt{a_j^f(\underline{X}(t))}\Delta W_j$$

$$+\frac{1}{2}\sum_{j_1,j_2=1}^{M^{fast}}\sum_{n=1}^{N} v_{j_1 n} v_{j_2 i}\sqrt{\frac{a_{j_1}(\underline{X}(t))}{a_{j_2}(\underline{X}(t))}}\frac{\partial a_{j_1}}{\partial X_n} I(j_1, j_2)$$

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^{ANRH}
100	9.64
1000	116.1
10,000	1198.2
100,000	**20535**

$$A \xrightarrow{k_1} B$$

$$B \xrightarrow{k_2} C$$

$$C \xrightarrow{k_3} D$$

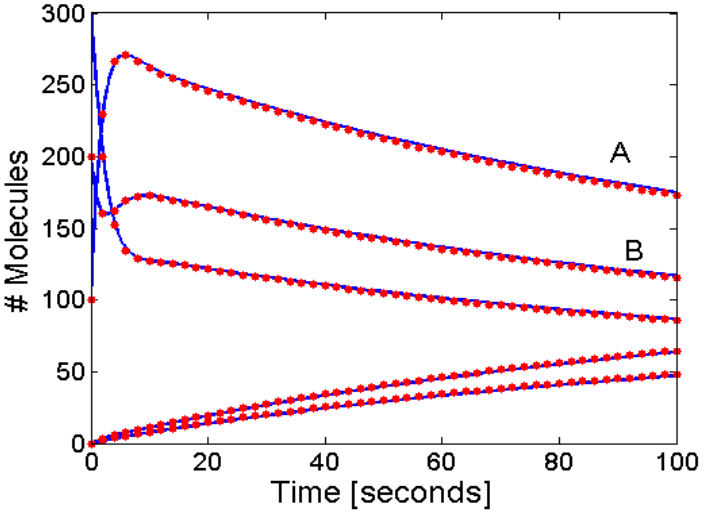
$$A + C \xrightarrow{k_4} D$$

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

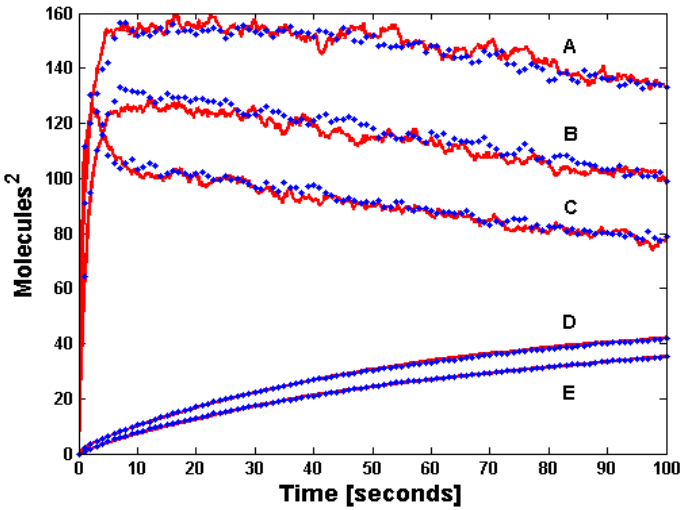
$$k_1, k_2, k_3 << k_4, k_5$$

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

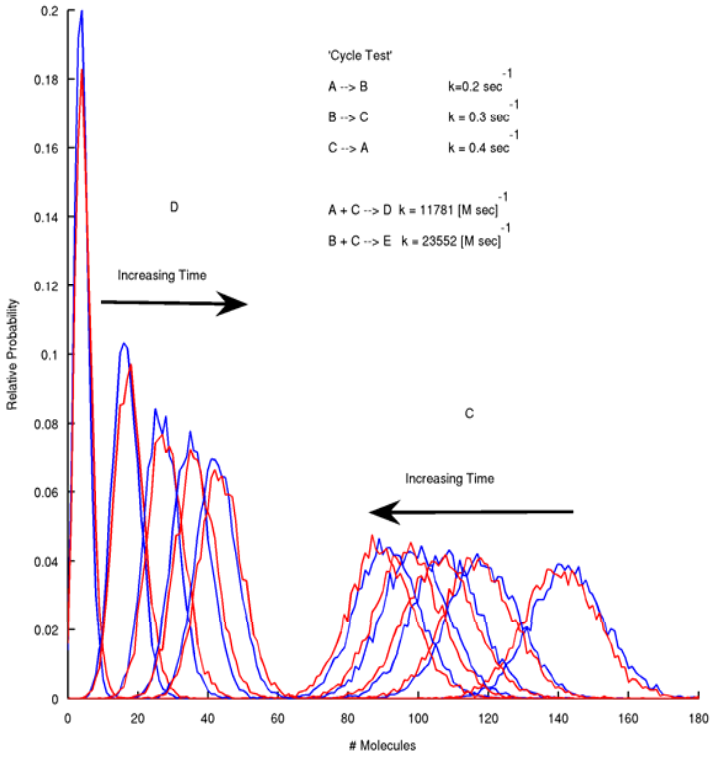
Accuracy: A Cycle Test



Mean

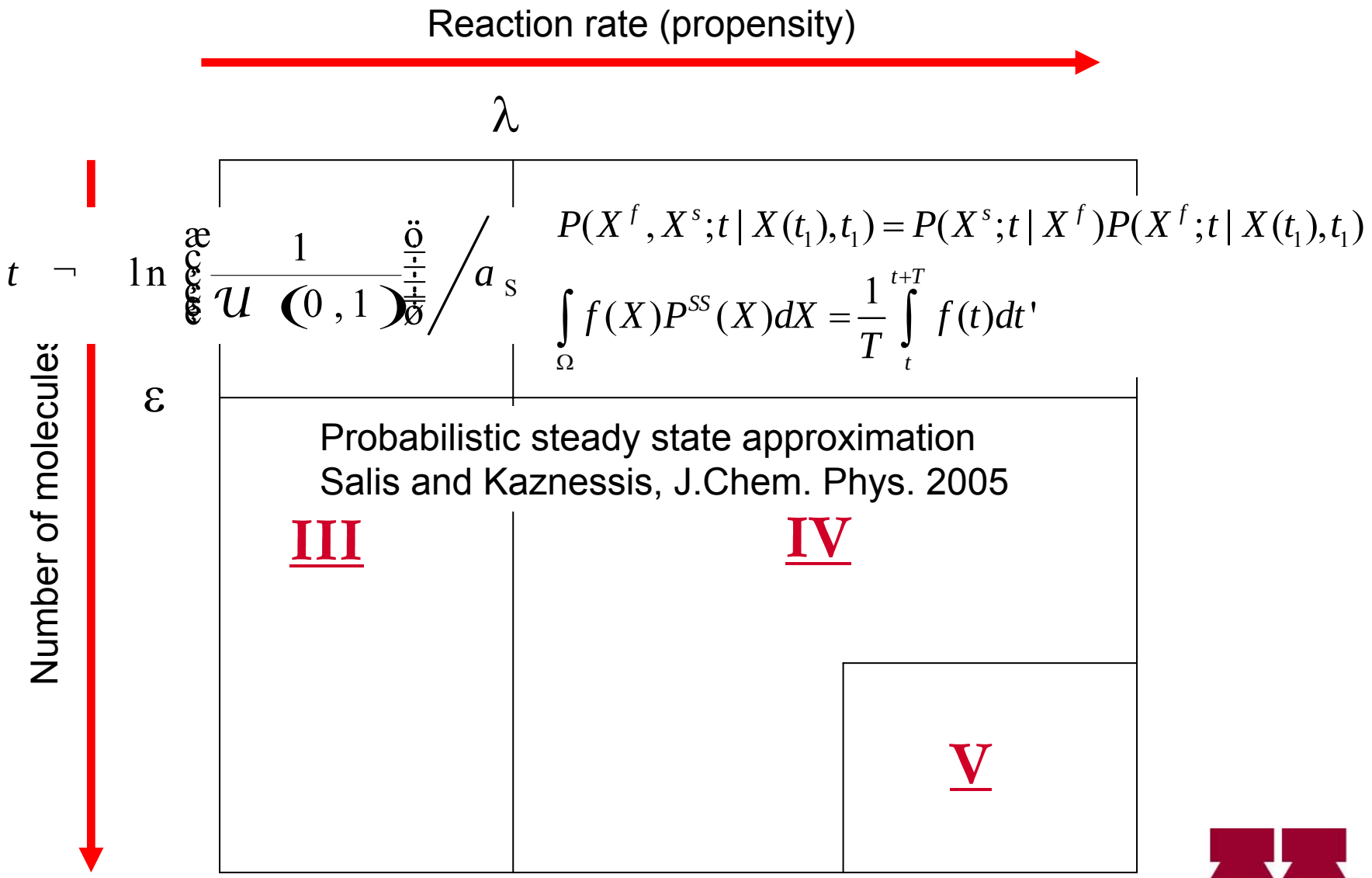


Probability Distribution

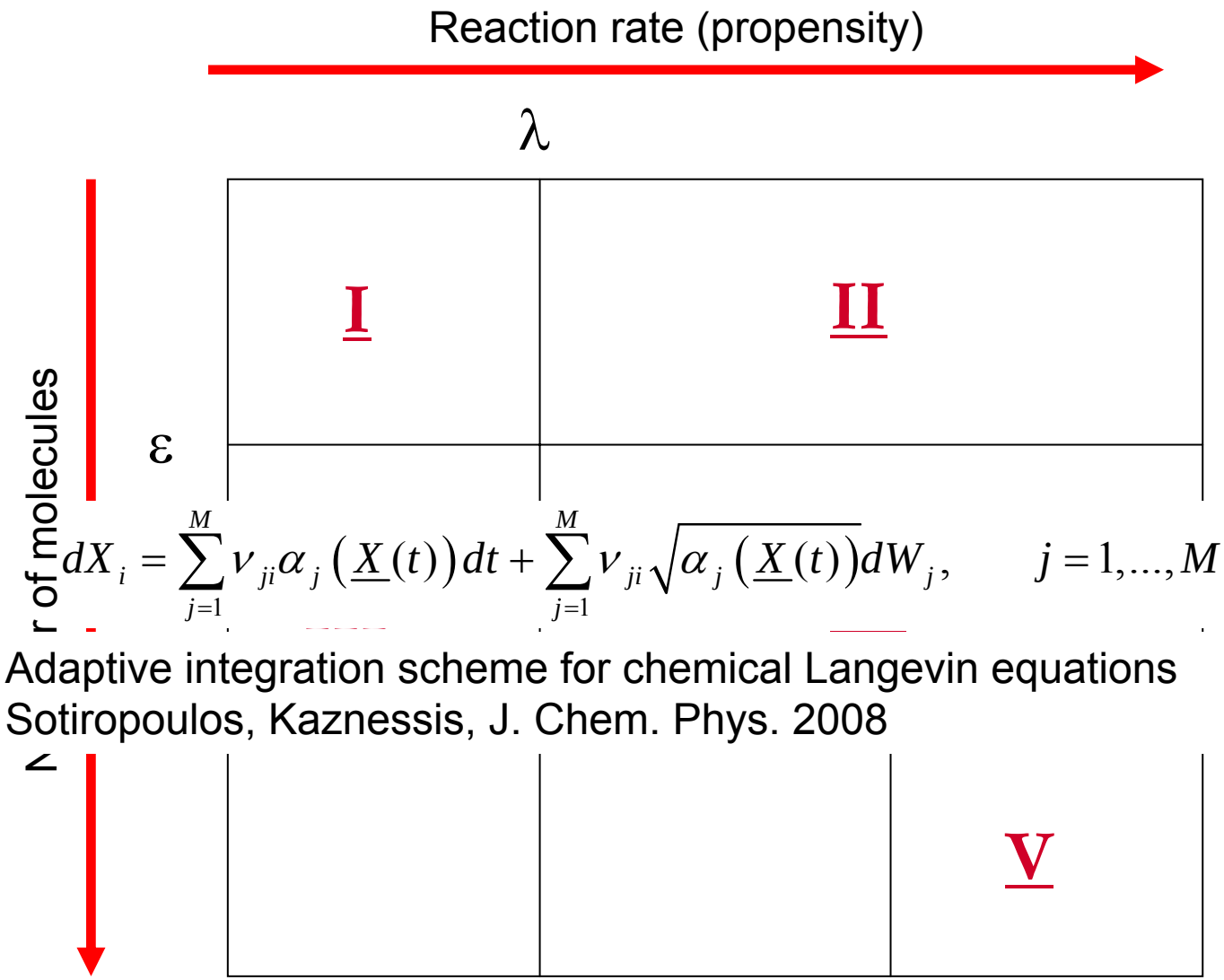


Variance

Modeling Regimes



Modeling Regimes



$$dX_i = \sum_{j=1}^{M} \nu_{ji} \alpha_j \left(\underline{X}(t)\right) dt + \sum_{j=1}^{M} \nu_{ji} \sqrt{\alpha_j \left(\underline{X}(t)\right)} dW_j, \qquad j = 1, \ldots, M$$

Adaptive integration scheme for chemical Langevin equations
Sotiropoulos, Kaznessis, J. Chem. Phys. 2008

Computational Synthetic Biology

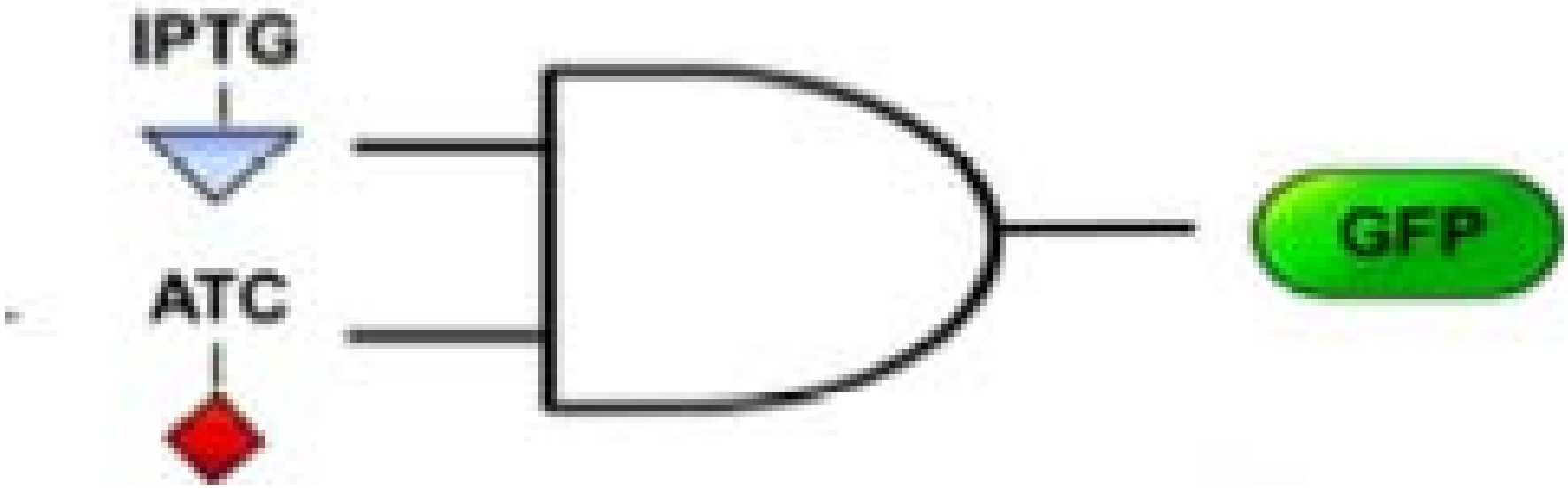
- Multi-scale stochastic-discrete and stochastic-continuous algorithm enables simulations of hundreds of species involved in thousands of reactions with disparate kinetic constants.

- Model gene networks with all the known molecular components.

- Generate detailed design principles.

Tuttle, Salis, Tomshine, Kaznessis, Biophys. J. (2005)
Salis, Kaznessis, Phys. Biol. (2007)
Tomshine and Kaznessis, Biophys. J. (2006)
Sotiropoulos, Kaznessis, BMC Systems Biology, (2007)
Kaznessis, BMC Systems Biology, (2007)

Synthesis of a Bio-Logical AND Gate



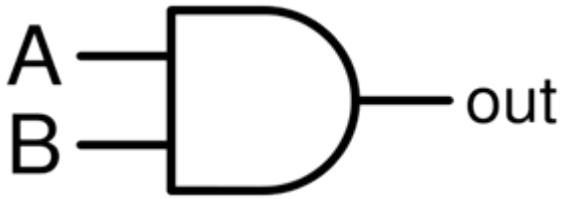
Our Molecular Toolbox

LacI and TetR repressors

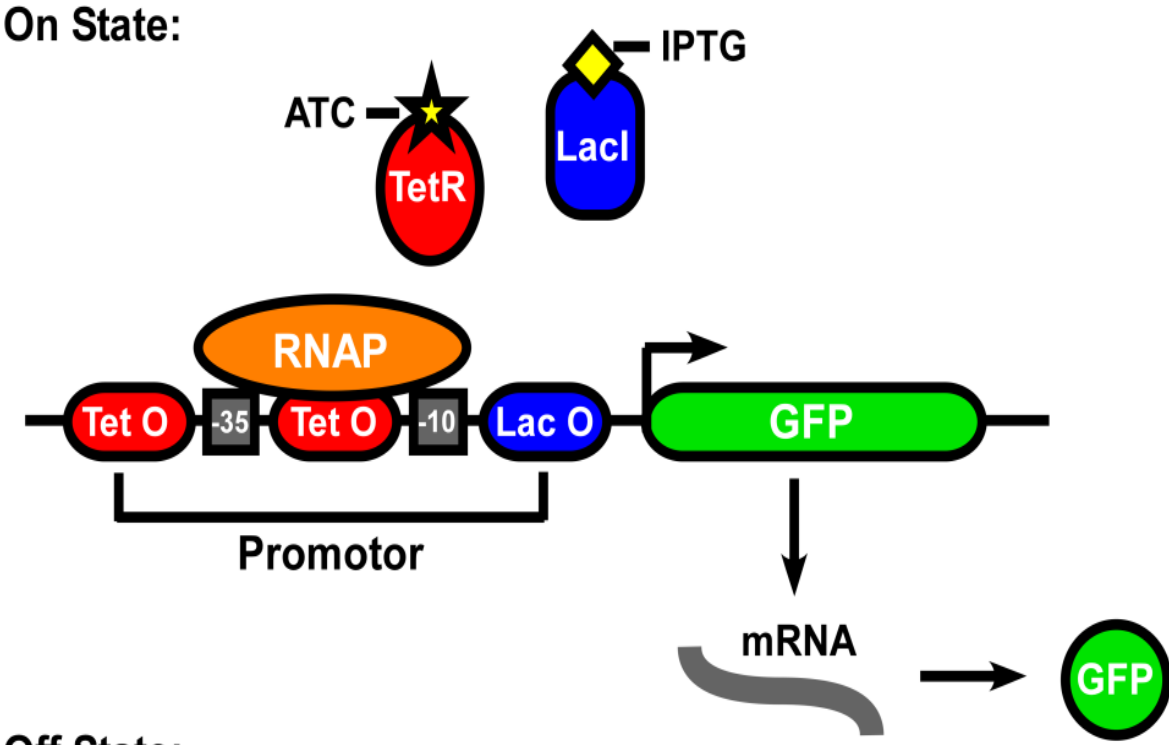
DNA sites: *lac* operators (lacO1, lacO2, lacO3), *tet* operators (tetO1, tetO2)
Promoter sequences (-35 and -10 σ70 dependent hexamers)
RBS sequences (hairpin secondary structures, RNAse binding sites)

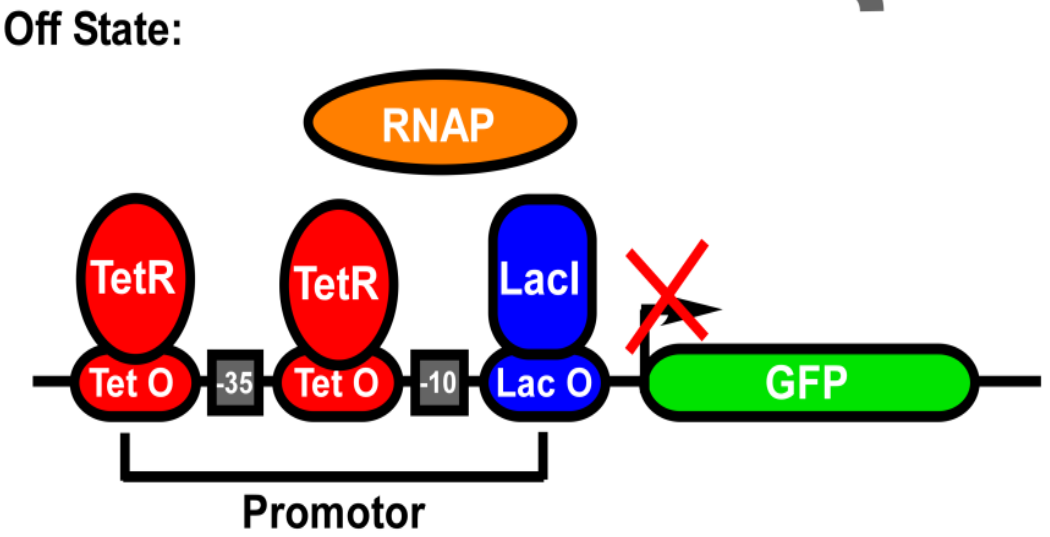
AND Logic Gate

- Express a gene when two signals are present:
  - aTc
  - IPTG

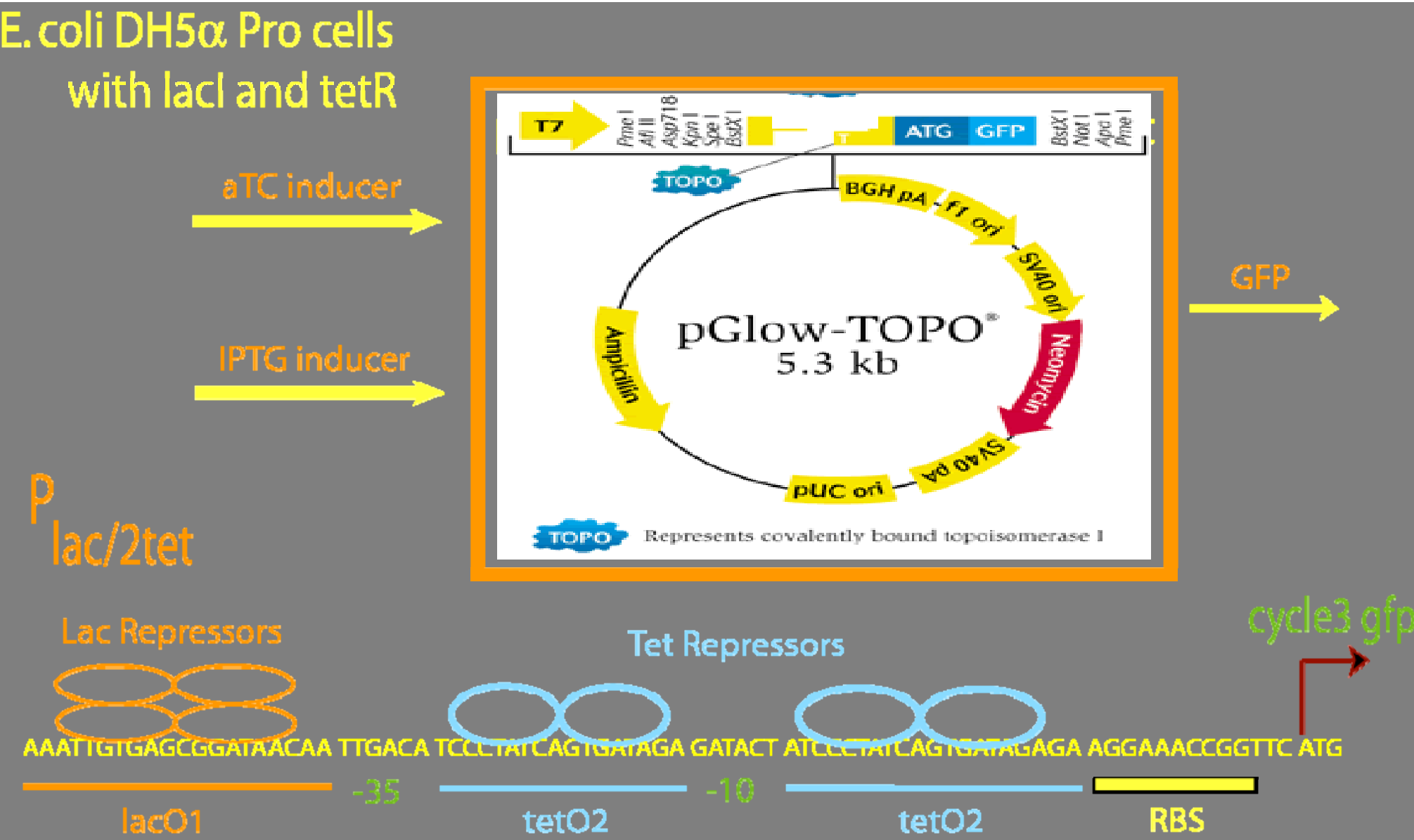


- Six possible designs

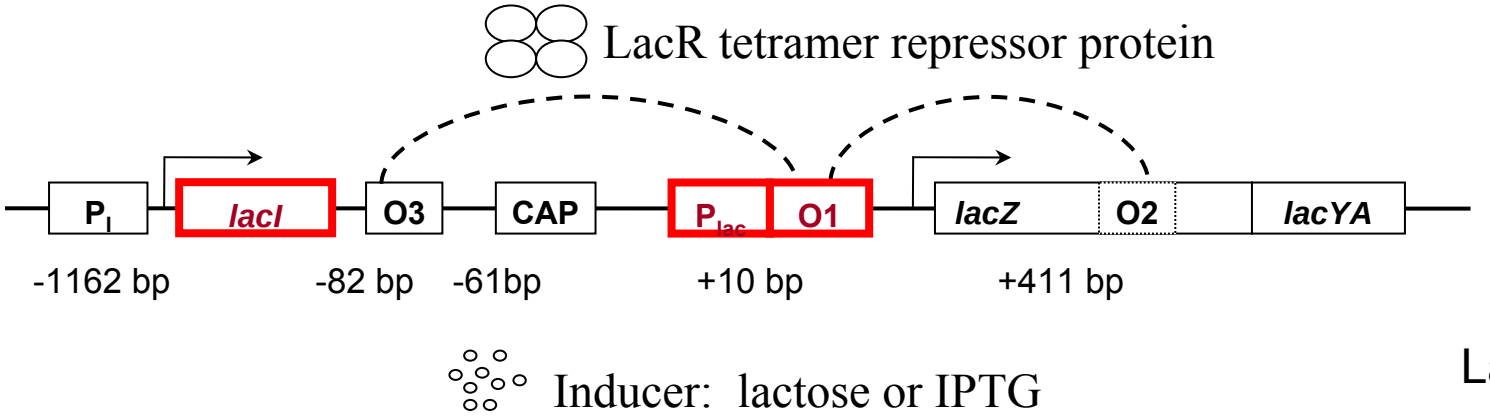




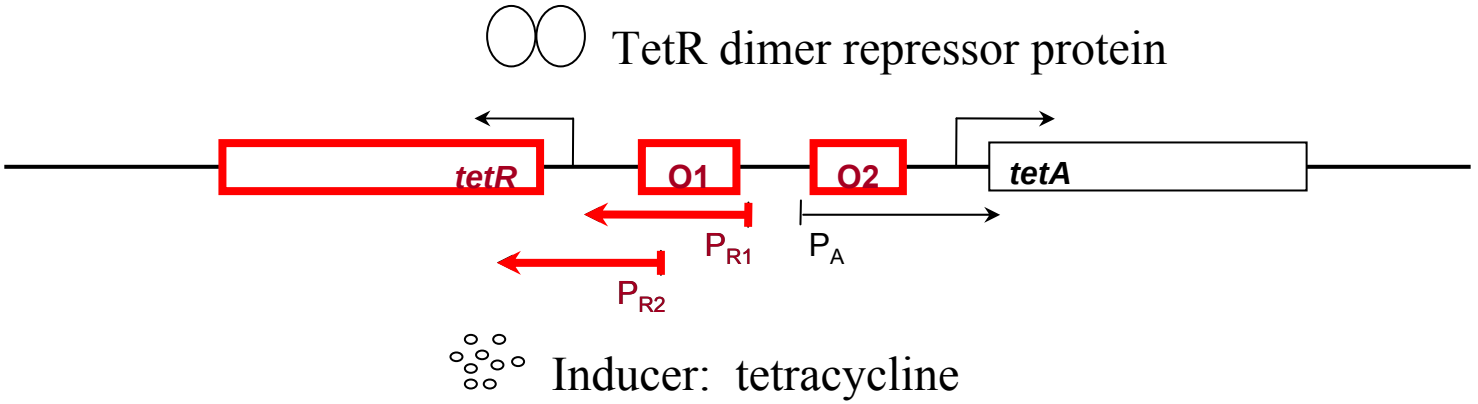
Experimental Construction of a Lac/Tet AND Gate



Models/Experiments Based on Real Biomolecular Components



Lac operon



Tet operon

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

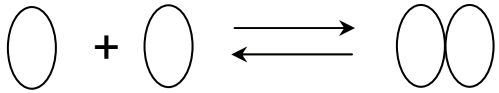
RNAp:tetO2 → RNAp + tetO2
RNAp:tetO2 → tetO2 + RNAp:DNA_lac
RNAp:DNA_lac → RNAp + mRNA_lac

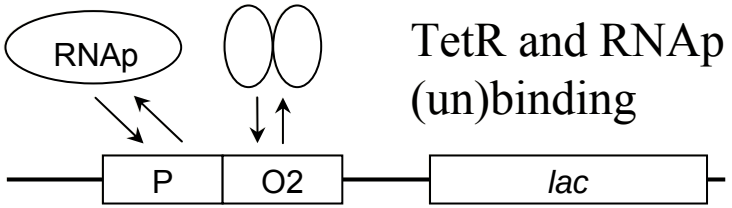
rib + mRNA_lac → rib:mRNA_lac
rib:mRNA_lac → mRNA_lac + rib:mRNA_lac1
rib:mRNA_lac1 → LacR1 + rib + mRNA_lac1

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

DTet→
TetR1 →
TetR2 →

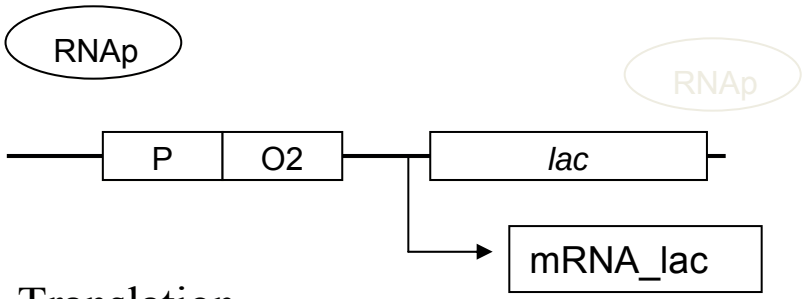
dimerization



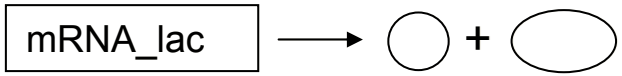


TetR and RNAp (un)binding

Opening of DNA, Transcription



Translation



Degradation of proteins and mRNA

AND Gate Simulations

- Network with 63 reactions.

- 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.

- 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).

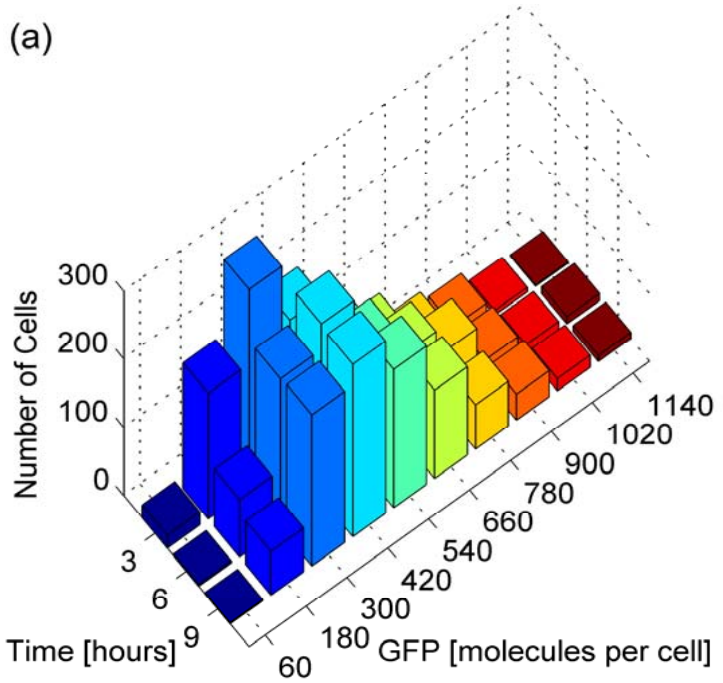
- Measure GFP number of molecules for 216,000 trajectories (36,000 CPU hours).

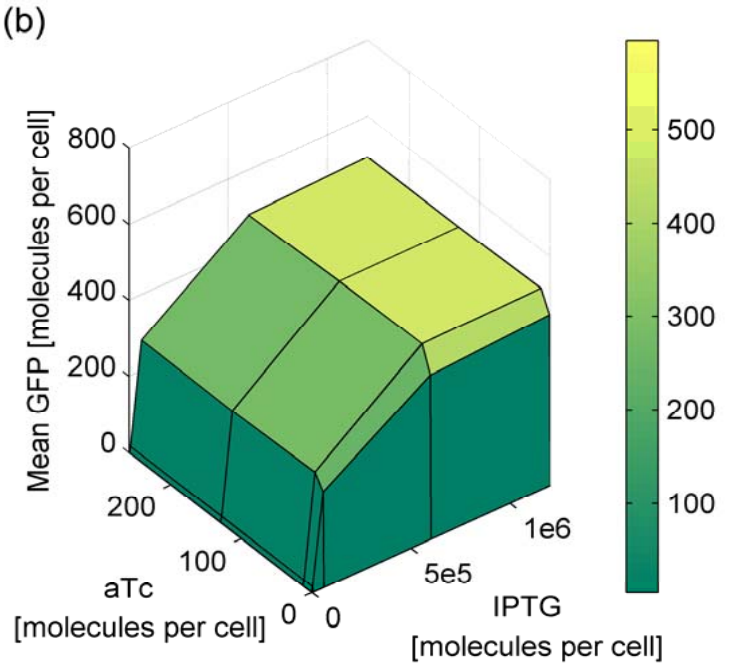
Stochastic simulations

LTT

TLT

TTL



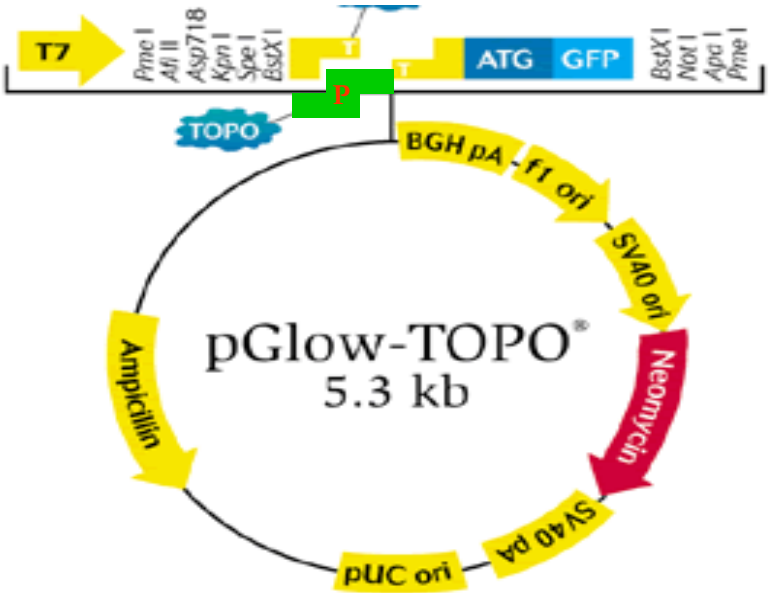


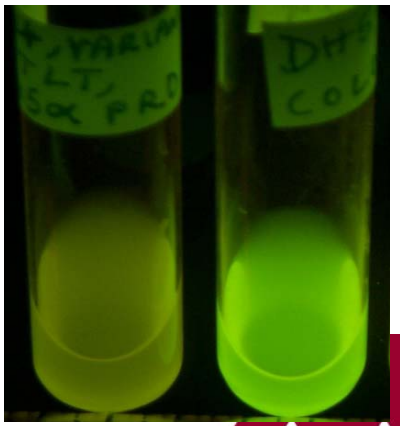
Synthetic Promoter Designs

LTT

TLT

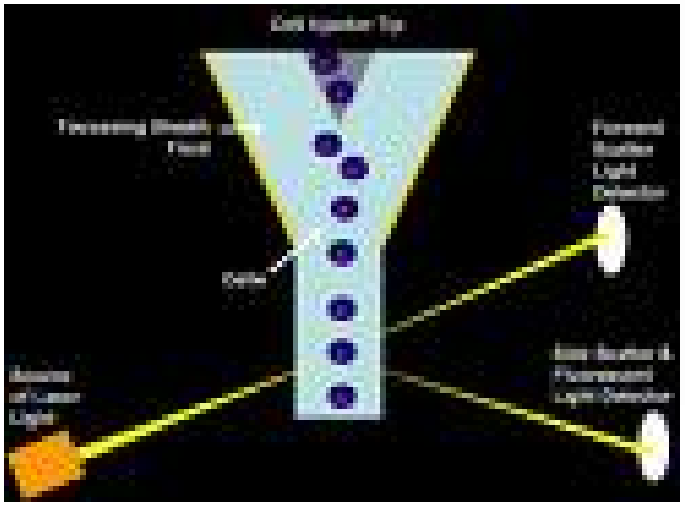
TTL

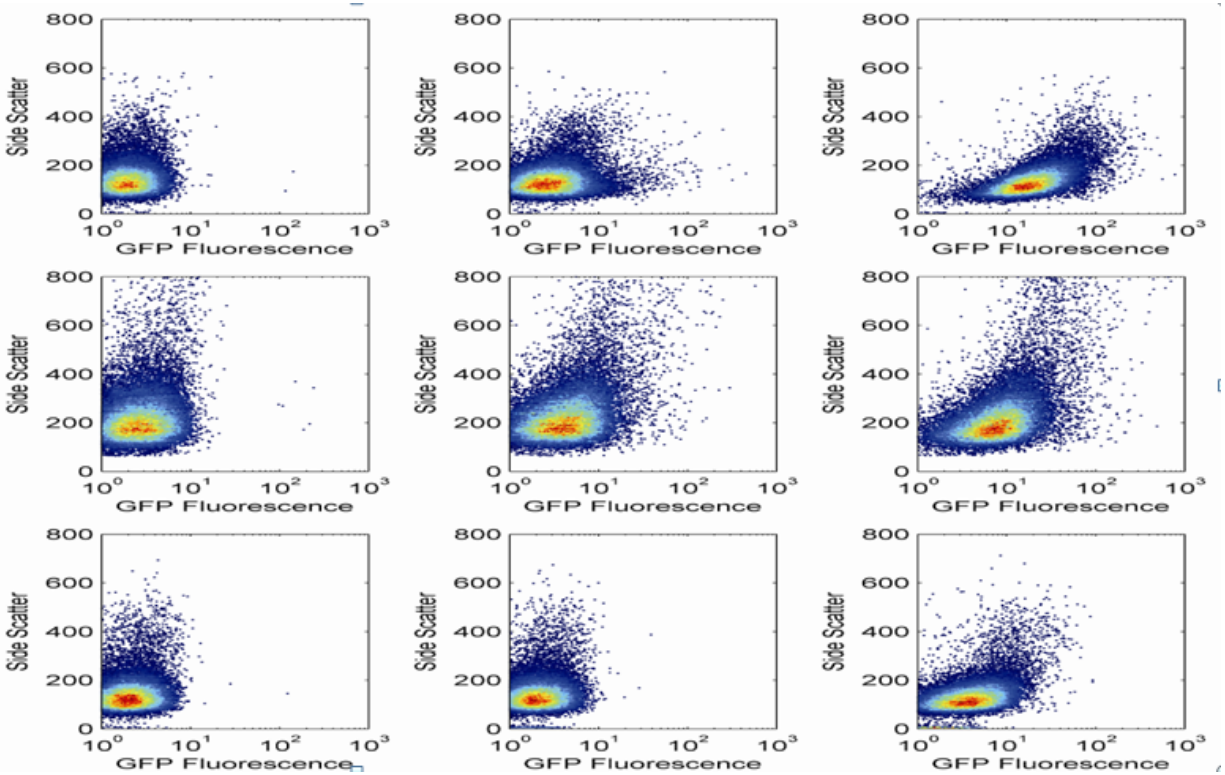




TOPO Represents covalently bound topoisomerase I

Flow cytometry





Computer-Aided Design of Bio-Logical AND Gates

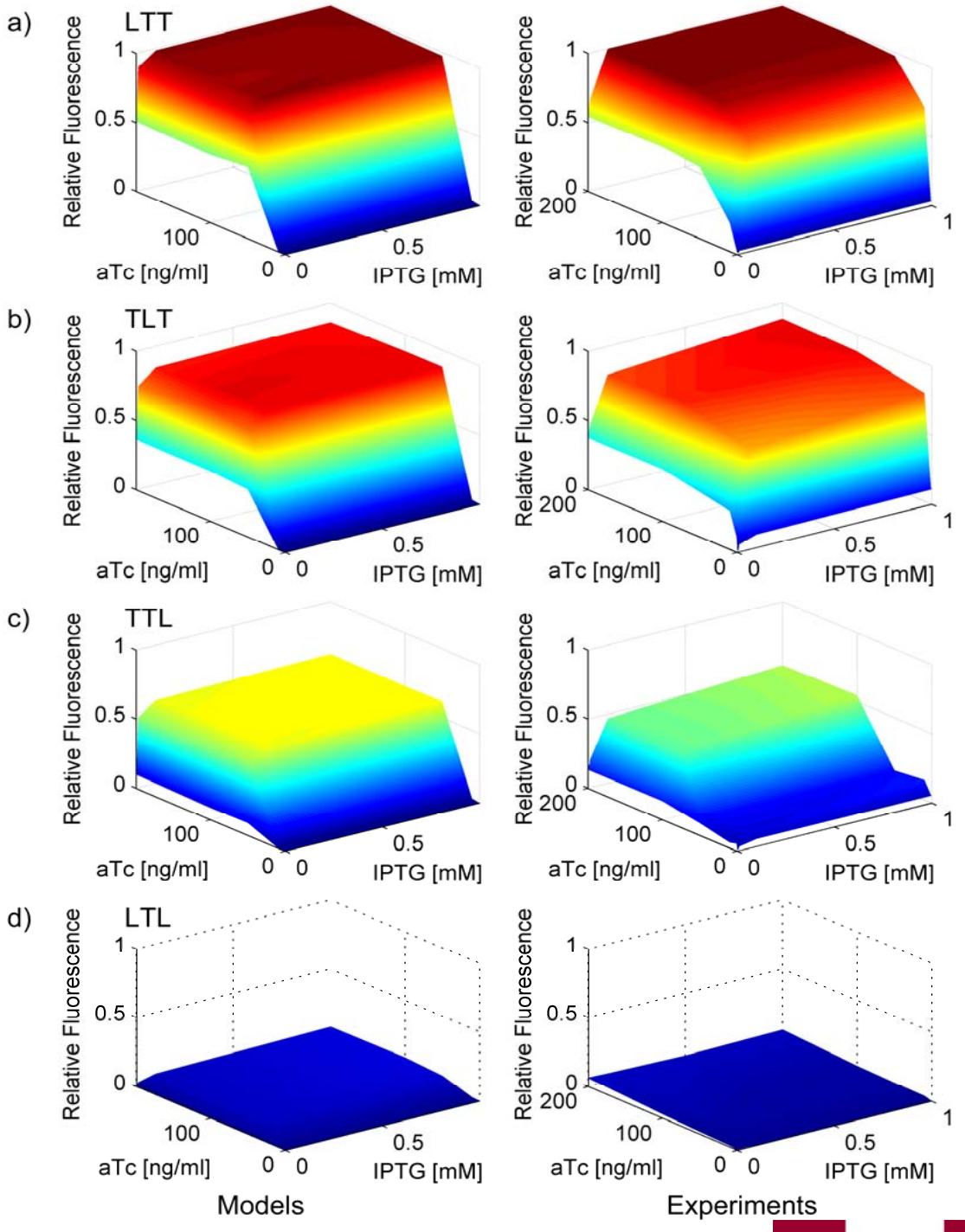
•TTL is the highest-fidelity AND gate.

• Leakage of lacO can explain the variable phenotypic behavior.

•Biological insight: leakage as a function of promoter topology.

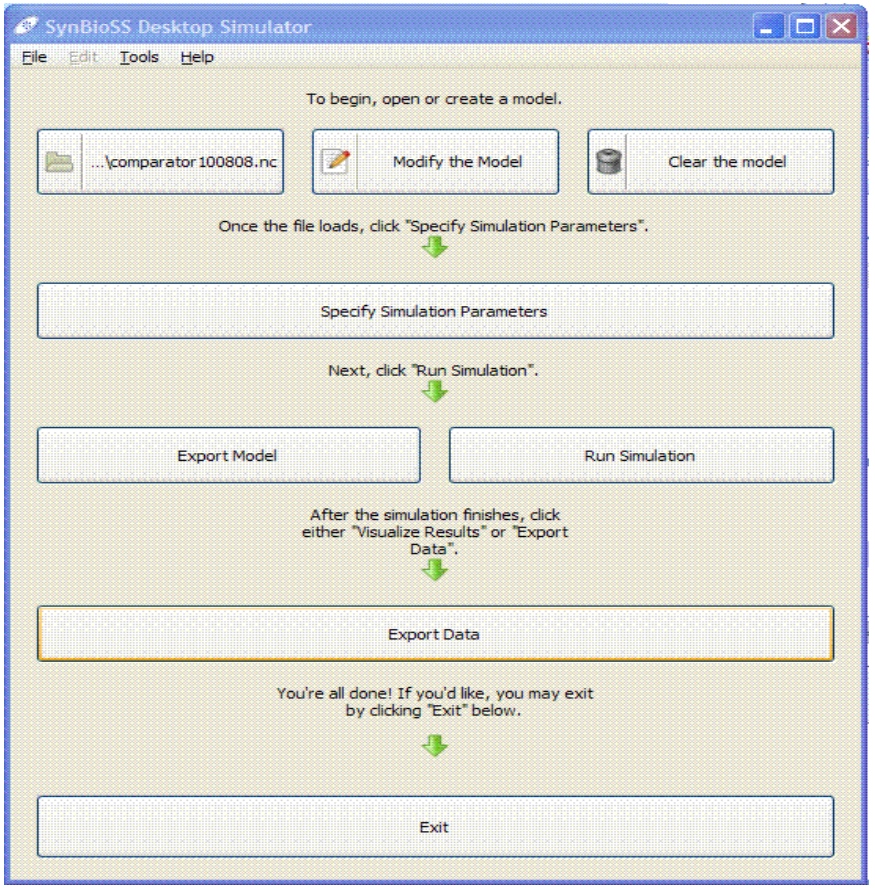
•Double-L systems not expressing enough GFP. Too much LacI in *E.coli* strain.

• Models capture experimental phenotype.



Synthetic Biology Software Suite
synbioss.sourceforge.net

- Numerical simulations are conducted with SynBioSS Desktop.



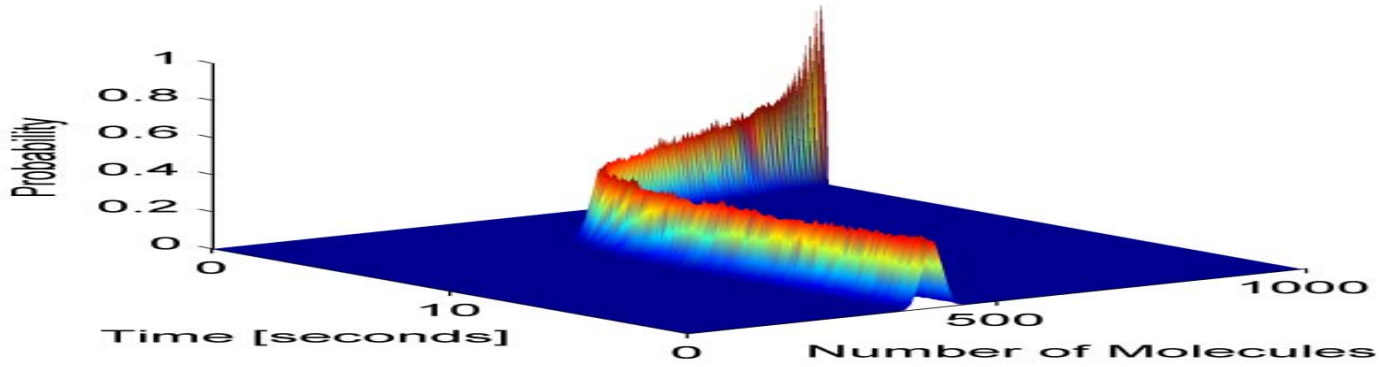
- Windows installation executable
- Codes for UNIX/Linux available
- Working on MacOS executable

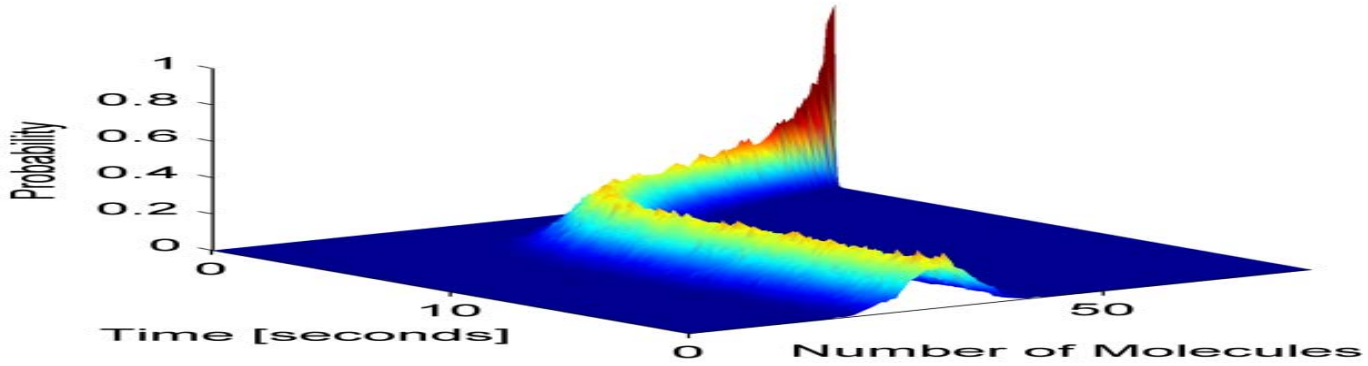
- Graphical User Interface:
  - Upload models
  - Build new models
  - Change existing models
  - Set simulation parameter

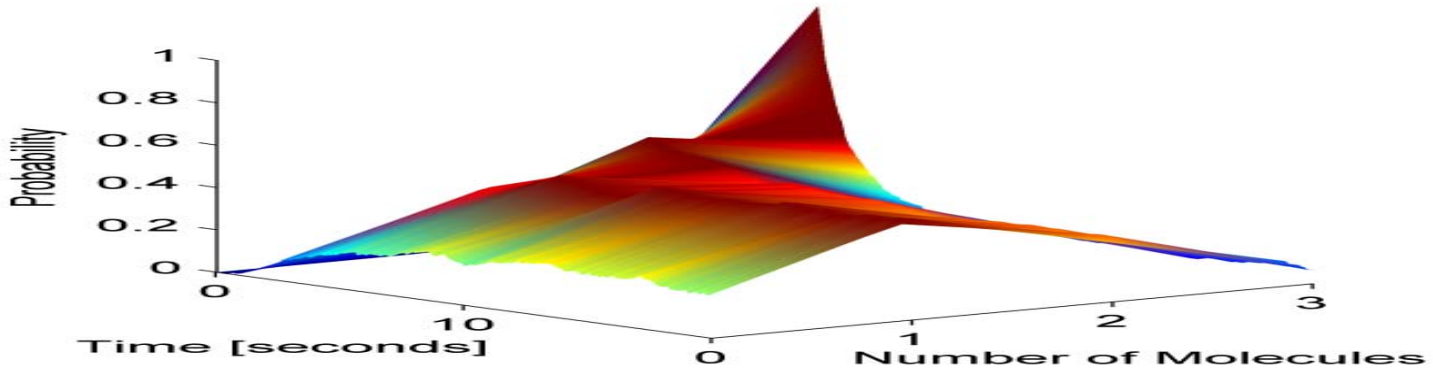
- Conduct Simulations
  - Multiscale algorithms
  - Stochastic algorithms
  - Generate probability distributions

- Output
  - NetCDF/SBML
  - CSV for analysis of probability distributions with Excel

$A \leftrightarrow B$

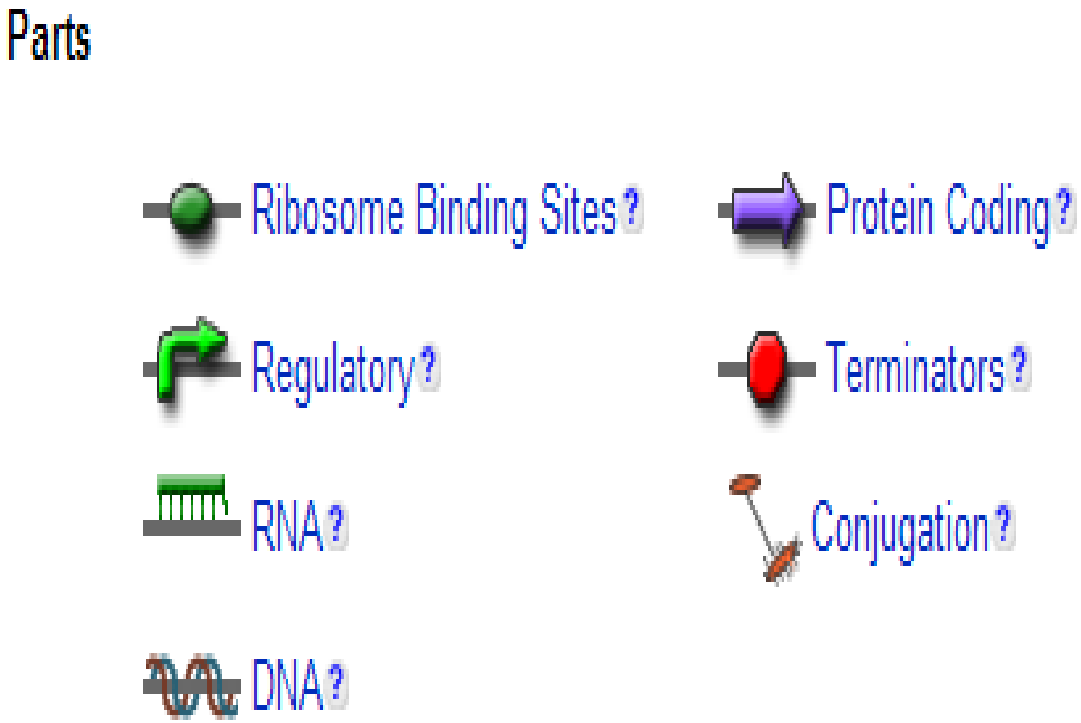


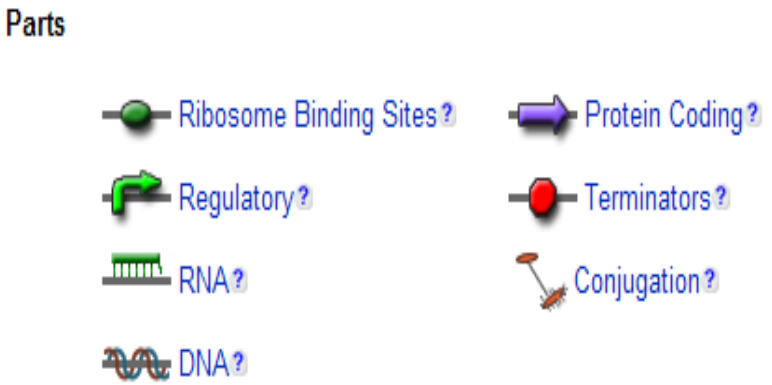




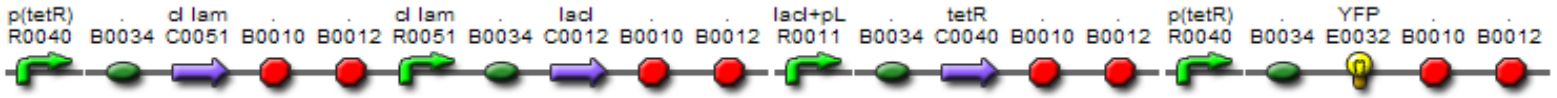
BioBricks (http://partsregistry.org)

- **BioBrick** standard biological parts are DNA sequences of defined structure and function.
- designed to be composed and incorporated into living cells such as *E. coli* to construct new biological systems.

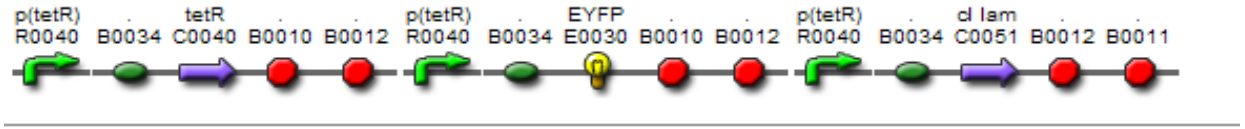




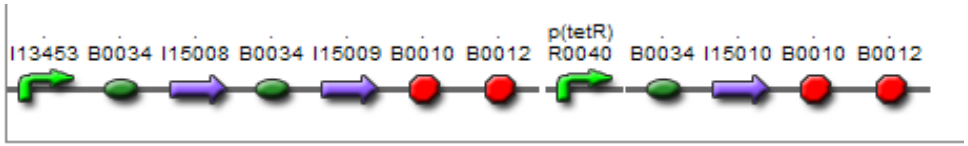
Light responsive system, dual regulation



Repressilator



Toggle Switch



Synthetic Biology Software Suite

- **SynBioSS.sourceforge.net**

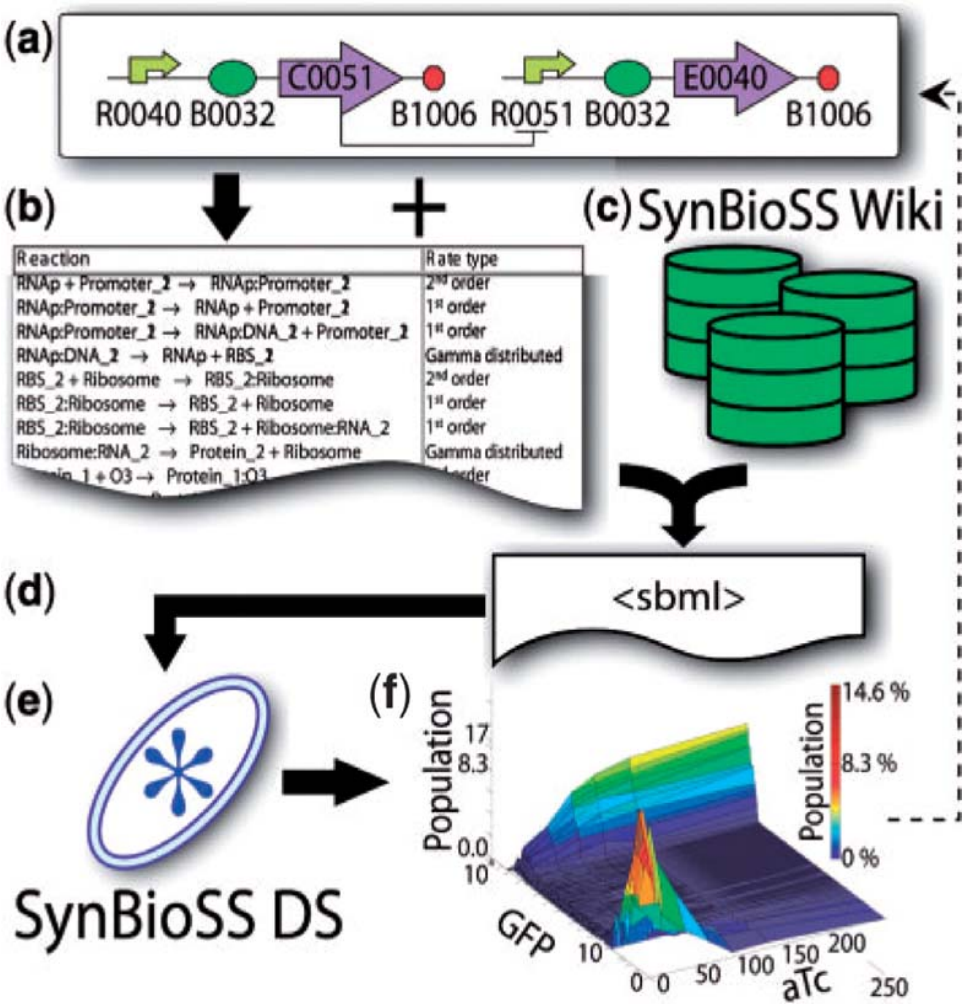
- Complete tool for generation, curation and simulation of synthetic biological networks.
- Three components:
  - Designer: Reaction network generation for arbitrary synthetic construct
  - Wiki: Kinetic data storage/retrieval. Community driven effort
  - Desktop Simulator: Numerical simulation with multiscale algorithms
- Designer and Wiki developed for iGEM
- Goals
  - Use accurate, fast and detailed quantitative models
  - Make model creation simple, faster
  - **Directly connect DNA sequences with dynamic phenotype**

SynBioSS: users can quickly build models of arbitrary synthetic gene regulatory networks (a-b), store and retrieve quantitative information (c), conduct numerical simulations (d-e), compare results with targeted synthetic phenotype and, if necessary, go back to re-design (f).

SynBioSS connects DNA sequences to targeted phenotypes. We believe this can rationalize synthetic biology.

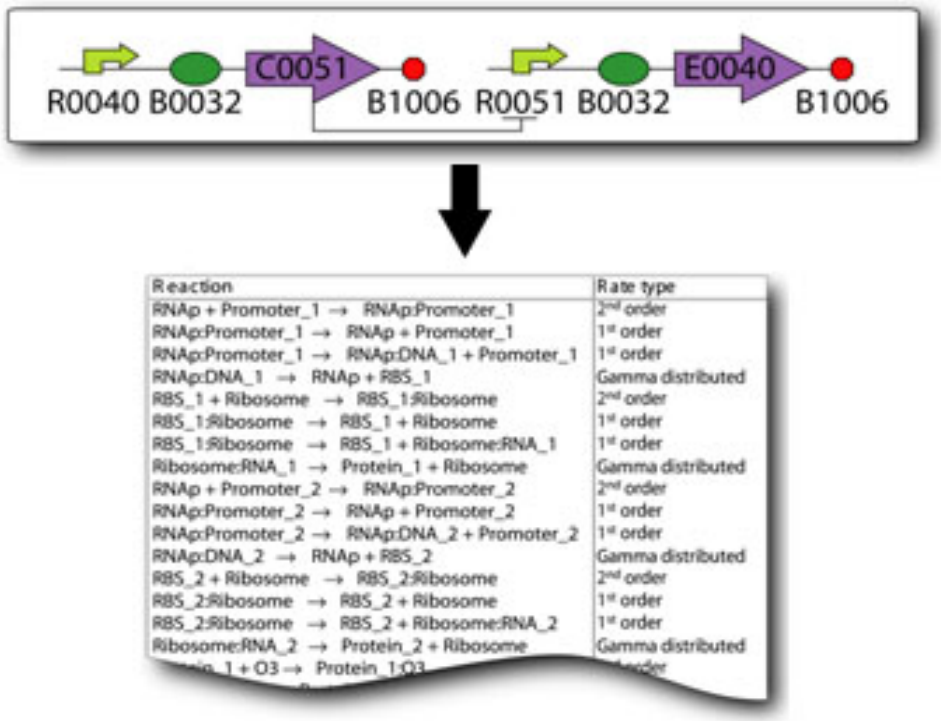
Everything is available at synbioss.sourceforge.net

Hill et al. Bioinformatics 2008 24(21):2551



SynBioSS Designer

- Automated generation of kinetic models.
- Molecular biology dogma.
- Web interface tool.
- Input
  - BioBrick components and relations
- Output
  - NetCDF
  - SBML
  - Anyone can build a model in 15 minutes



Model Structure: Network

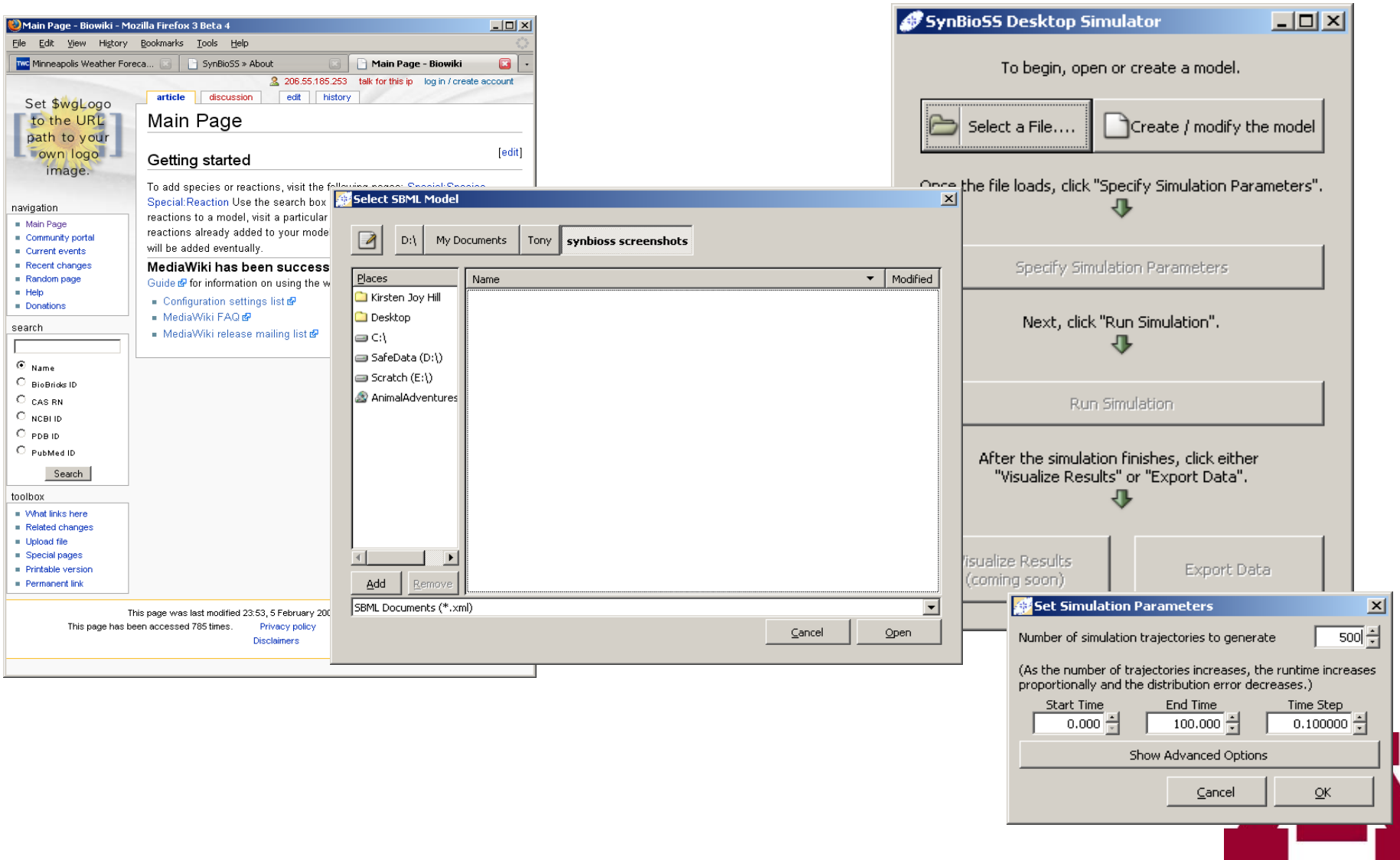
Number	General Transcription & Translation Reactions	k	Source
1	RNAp + lacP + lacO1 + tetO1 + tetO2 → RNAp:lacP	1.00E+07	30
2	(see below)		
3	(see below)		
4	RNAp:lacP → RNAp:lacP*	0.01	31
5	RNAp:lacP → RNAp + lacP + lacO1 + tetO1 + tetO2	1	30
6	RNAp:lacP* → lacP + lacO1 + tetO1 + tetO2 + RNAp:DNAgfp	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		
3	RNAp + lacP + lacI4:lacO1:IPTG + tetO1 + tetO2 → RNAp:lacP + lacI4:IPTG		

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	**

SynBioSS Wiki

- Kinetic data repository
  - Formatted & searchable
  - Includes references
- Curated by users
- Manual network creation
  - SBML output
- Connections to Designer
  - Data retrieval
  - Network input
- Have entered data for tetracyline, lactose and arabinose operons.

http://synbioss.sourceforge.net/



Summary

- Available toolbox of DNA sequences and regulatory proteins. Design novel gene networks to control protein production.
- Hybrid stochastic-discrete and stochastic-continuous network simulations tackle multiple scales.
- Software tool available to the synthetic and systems biology community (SynBioSS).
- http://synbioss.sourceforge.net/
- Computer-assisted design of a synthetic Bio-Logical AND-gate.
- Can reductionism be validated?