Probabilistic Approximations of Bio-Pathways Dynamics

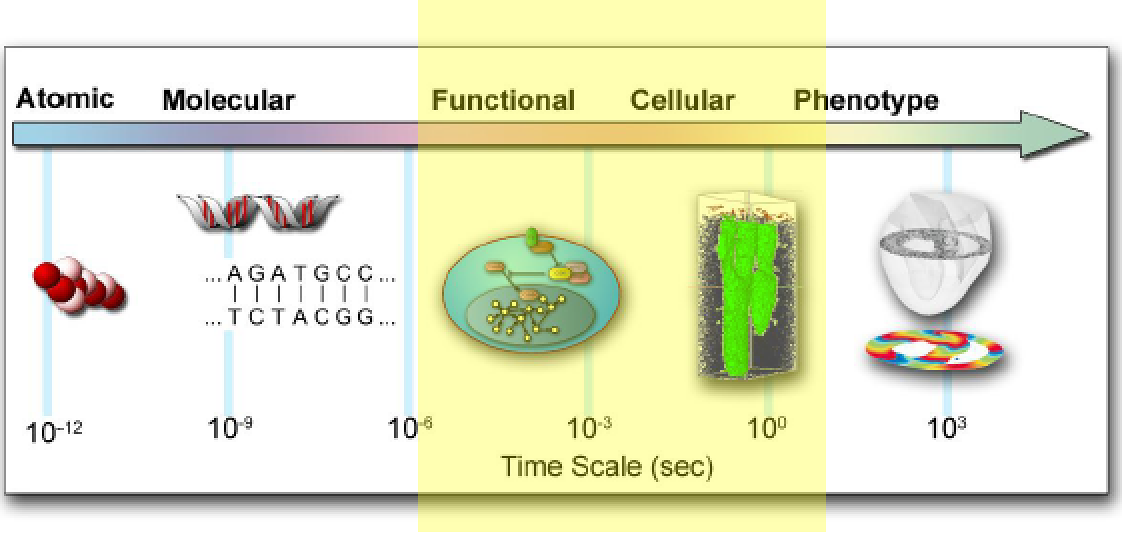
P.S. Thiagarajan

School of Computing, National University of Singapore

Joint Work with: Liu Bing, David Hsu

Granularities

- Several levels of complexity in biological systems



- Our research focuses on the:
  - Intra- Cellular level
  - Bio-pathways

A Common Modeling Approach

- Bio-pathway: A network of bio-chemical reactions
- A system (network) of ODEs
  - One for each molecular species
  - Mass action, Michelis-Menten, Hill, etc.
- Study the ODE system to understand the dynamics of the bio-chemical network

Our work: The main features

- ***ODE systems to describe the dynamics of the bio-chemical networks***
- Decompose → Compute → Compose (ISMB'06, WABI'07)
- Probabilistic models and methods to:
  - update models (RECOMB'10)
  - ***approximate the ODE dynamics***(CMSB'09, TCS'11***)***.

An Example

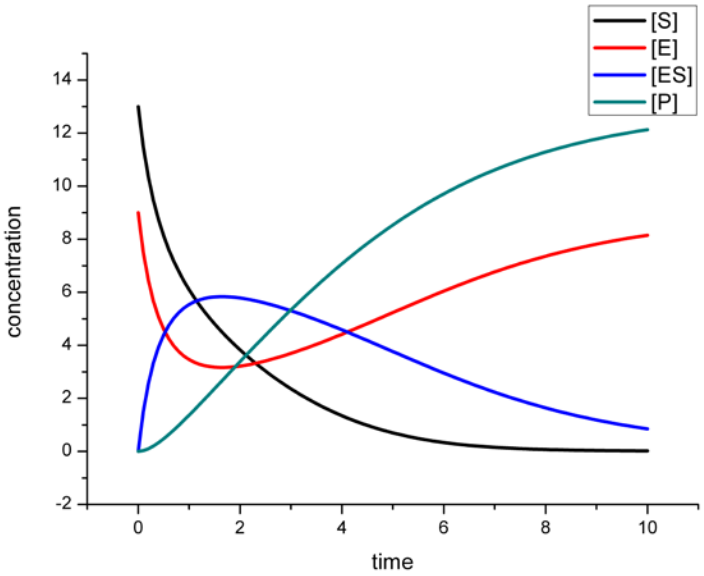
$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P$$

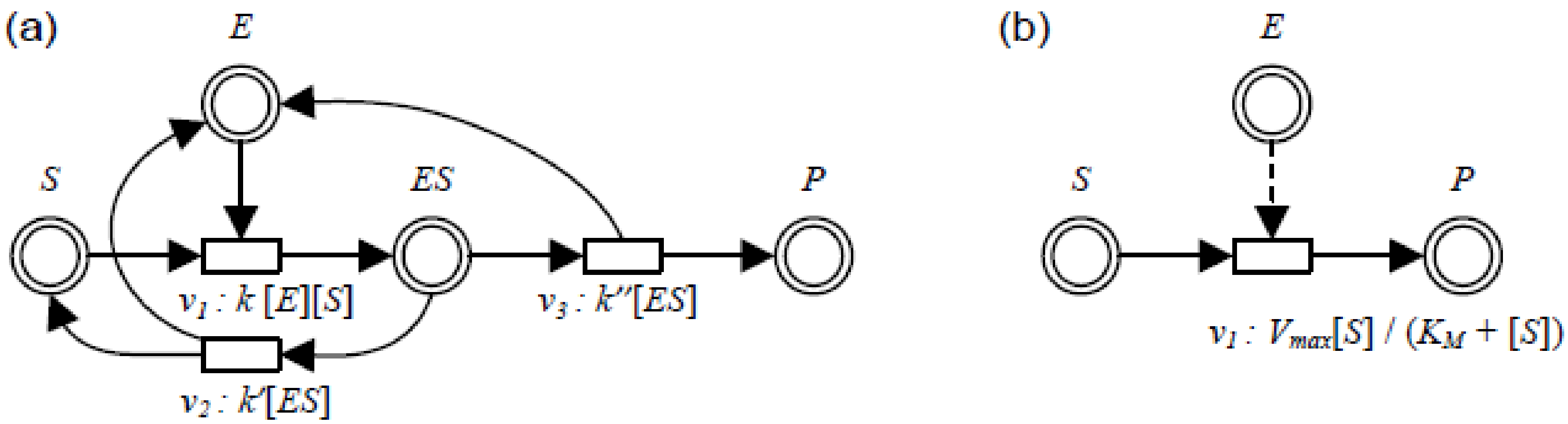
$$\frac{d[S]}{dt} = -k_1[S][E] + k_{-1}[ES]$$

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

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

$$\frac{d[P]}{dt} = k_2[ES]$$




(a)
E
S
ES
P
v1 : k [E][S]
v3 : k''[ES]
v2 : k'[ES]
(b)
E
S
P
v1 : Vmax[S] / (KM + [S])

Many Hurdles

- Many rate constants not known
  - must be estimated
  - noisy data; limited precision; population-based
- High dimensional system
  - closed-form solutions are impossible
  - Must resort to numerical simulations
  - *A set (Interval of values) of initial states must be dealt with;*
  - ***a large number of numerical simulations needed for answering each question***

The "Opinion Poll" Idea

- Discretize the time and value domains.
- Assume a (uniform) distribution of initial states
- Generate a "sufficiently" large number of "typical" trajectories by
  - sampling the initial state and numerical simulations.
- ***Synthesize form this collection of trajectories a dynamic Bayesian network.***
  - **ODEs $\rightarrow$ DBN**

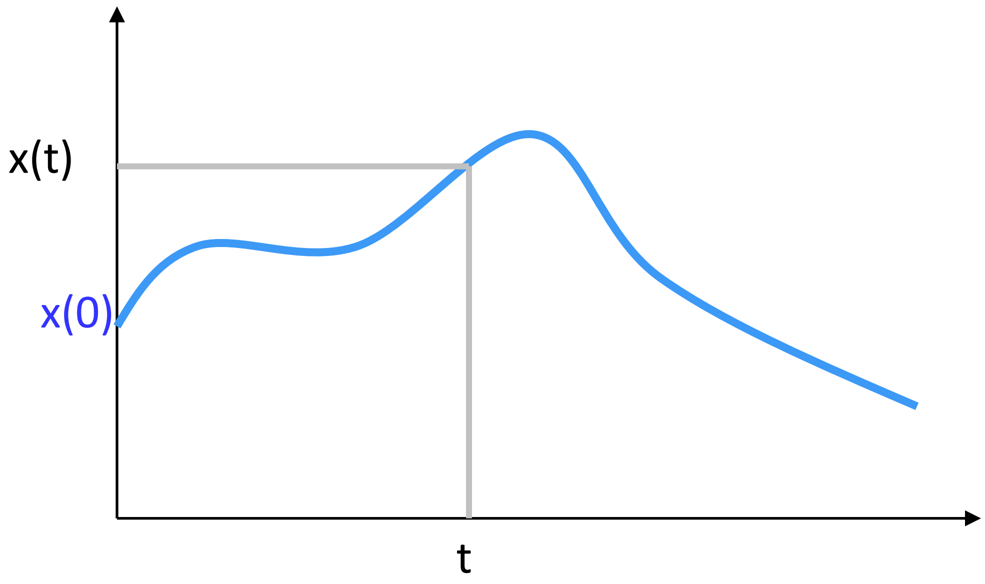
The “opinion poll” Idea

- Pay the one-time cost of constructing the Bayesian network.
- Amortize this cost by performing multiple analysis tasks using the Bayesian network representation.
  - Using inferencing algorithms for DBNs.

The Technique

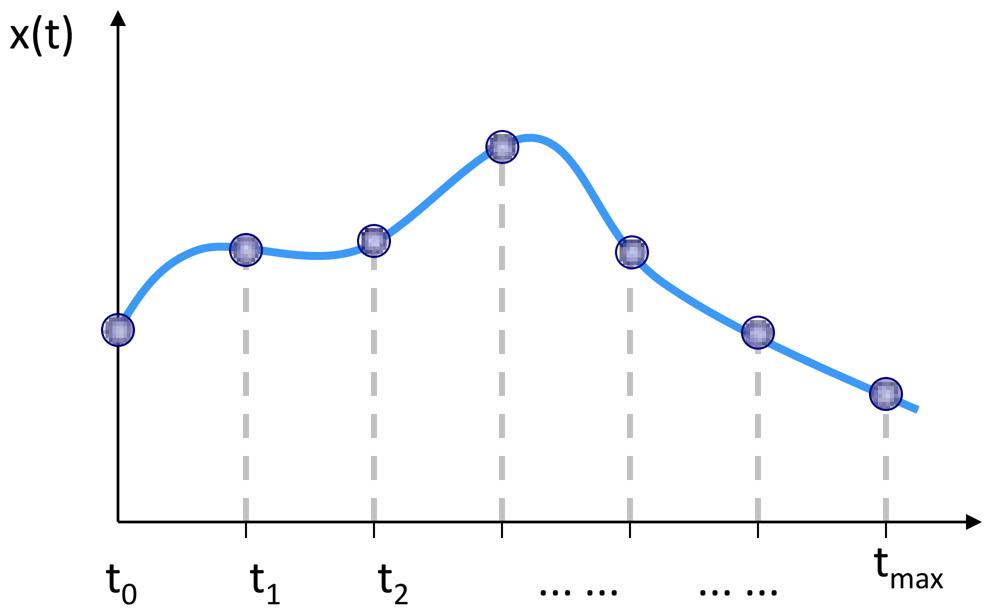
$$\frac{dx}{dt} = 3t^2 + 4$$

- We want to study x(t) the "solution" to the above equation. [$x(t) = t^3 + 4t + x(0)$]



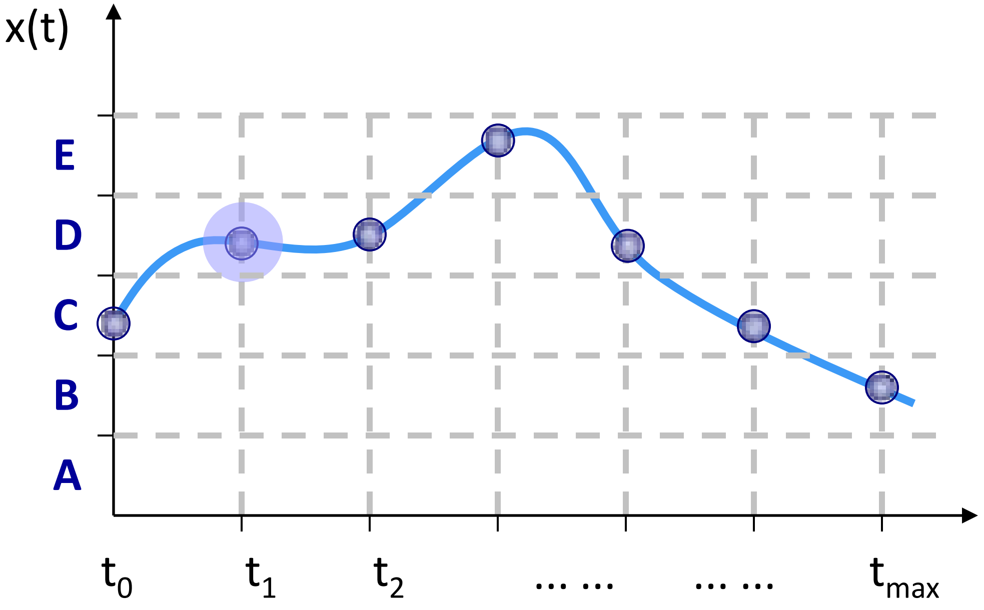
The Technique

- Observe the system only at discrete time points; in fact, only at finitely many time points



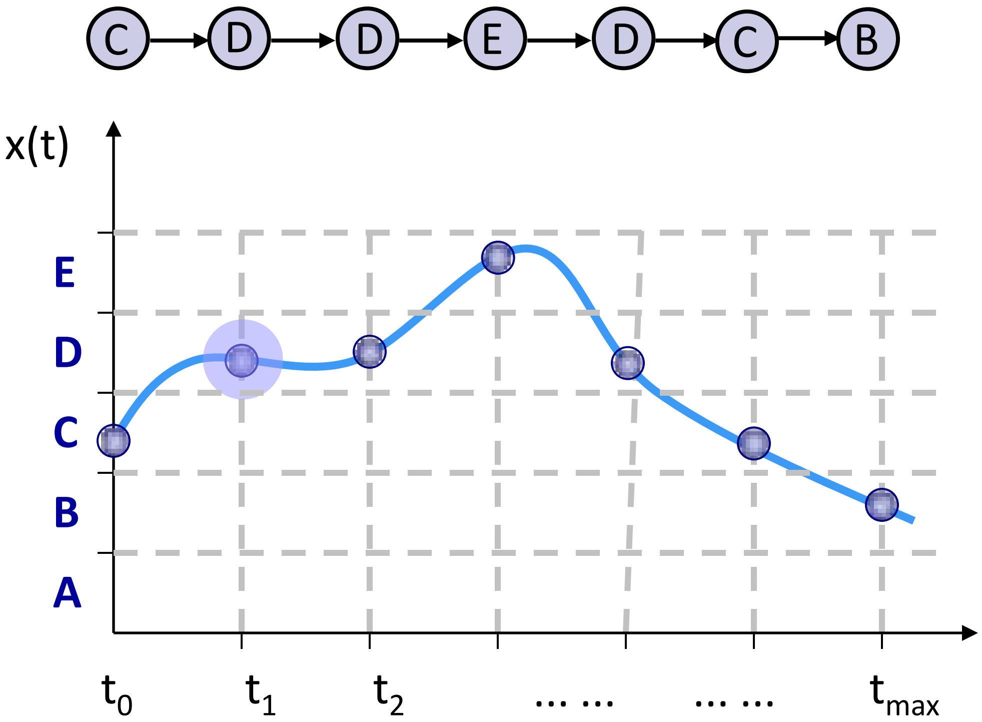
The Technique

- Observe only with bounded precision



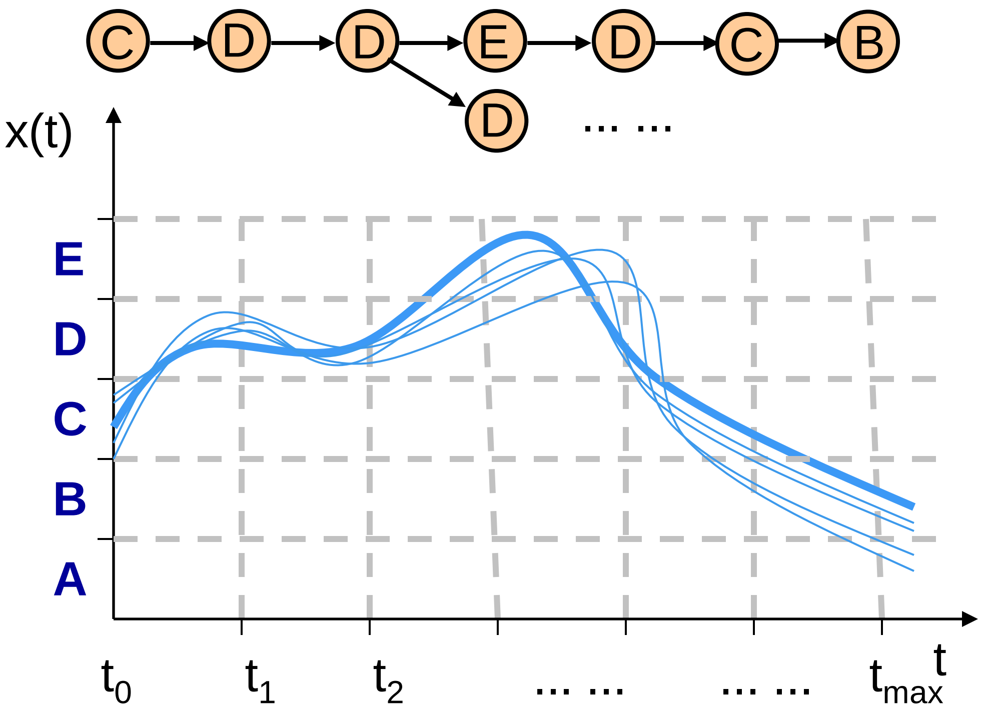
The Technique

- A trajectory now is sequence of *discrete values*



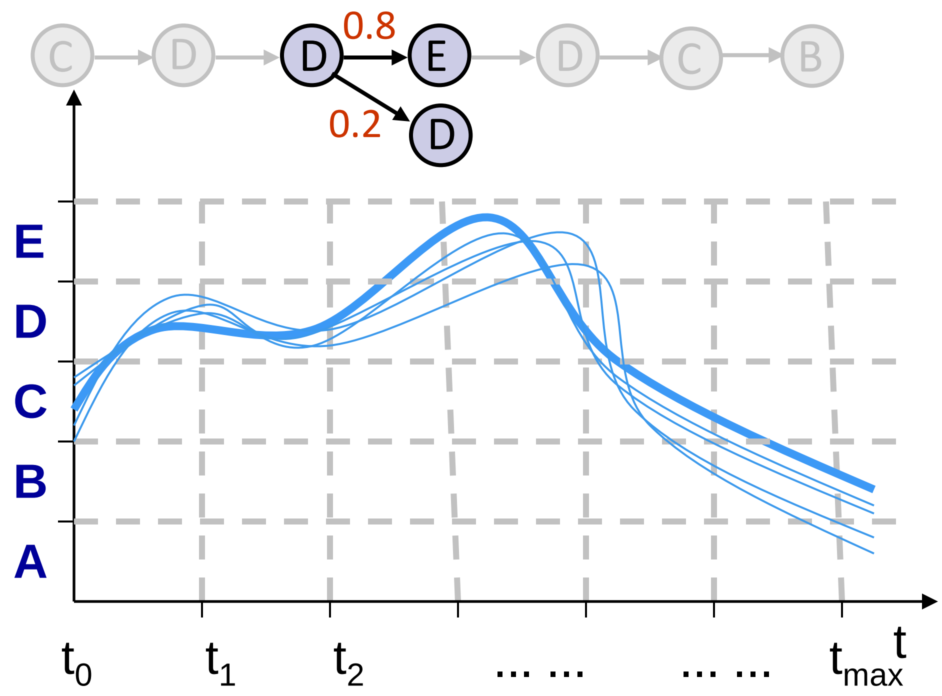
Main Idea

- With a prior distribution of the initial states
- Many sequences of ***discrete values***!



The Technique

- In fact, a Markov chain.
- Pr((D, 2) $\longrightarrow$ (E, 3)) is the “fraction” of the trajectories residing in D at t = 2 that land in E at t = 3.



The Technique

- The value space of the variables is assumed to be a compact subset C of $\mathcal{R}^n$
- In **Z'** = F(Z), F is assumed to be differentiable everywhere in C.
  - Mass-law, Michaelis-Menton,…
- Then the solution (flow) Φ_t : C $\rightarrow$ C (for each t) exists, is unique, is a bijection, is continuous and hence ***measurable.***

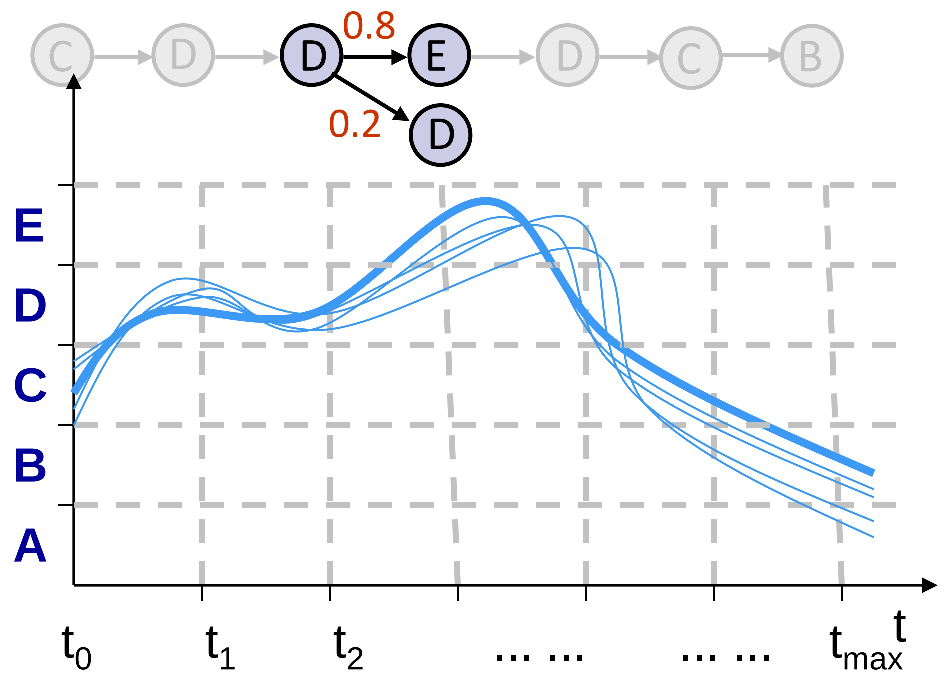
(s, i) – States;
(s, i) $\rightarrow$ (s', i+1) -- Transitions

Sample, say, 1000 times the initial states.

Through numerical simulation, generate 1000 trajectories.

Pr(s, i) is the fraction of the 1000 trajectories that are in the discrete state s at t_i.

Pr((s, i) $\rightarrow$ (s' i+1)) is the fraction of the trajectories that are in s at t_i which land in s' at t_{i+1}.



The Technique

- But the ***Markov chain*** will be huge.
  - $O(T \cdot k^n)$
  - 50 variables (n) , each with two (k) intervals, T time points
  - $T \times 2^{50}$ states !

The Technique

- Exploit the network structure (additional independence assumptions) to construct a dynamic Bayesian network instead.
  - Sampling + numerical + simulations + counting
- The DBN is a ***factored*** form of a Markov chain.

The Technique

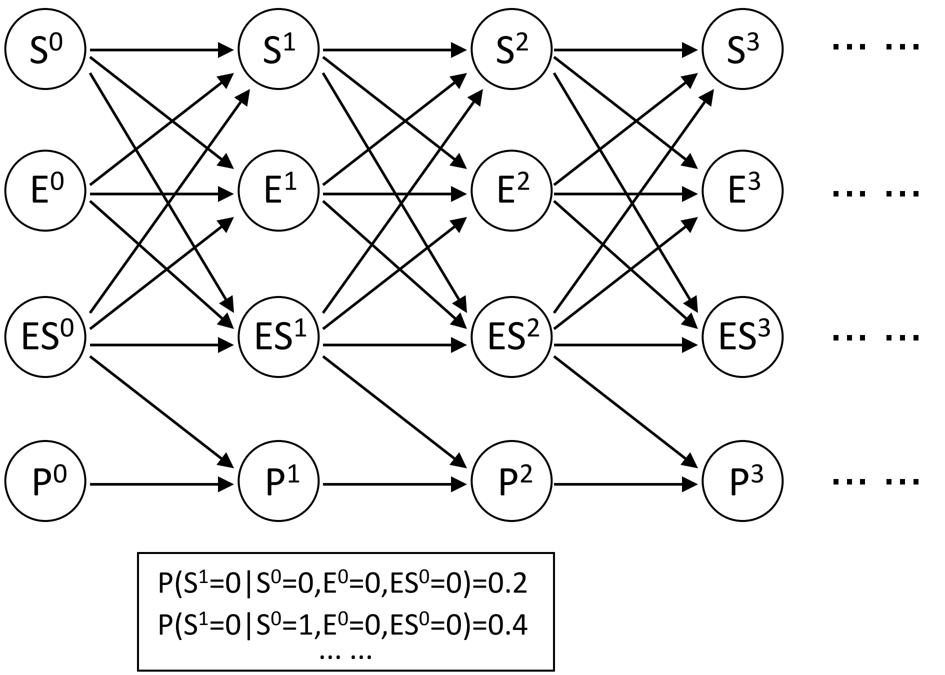
$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \overset{k_2}{\rightarrow} E + P$$

$$\frac{d[S]}{dt} = -k_1[S][E] + k_{-1}[ES]$$

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

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

$$\frac{d[P]}{dt} = k_2[ES]$$

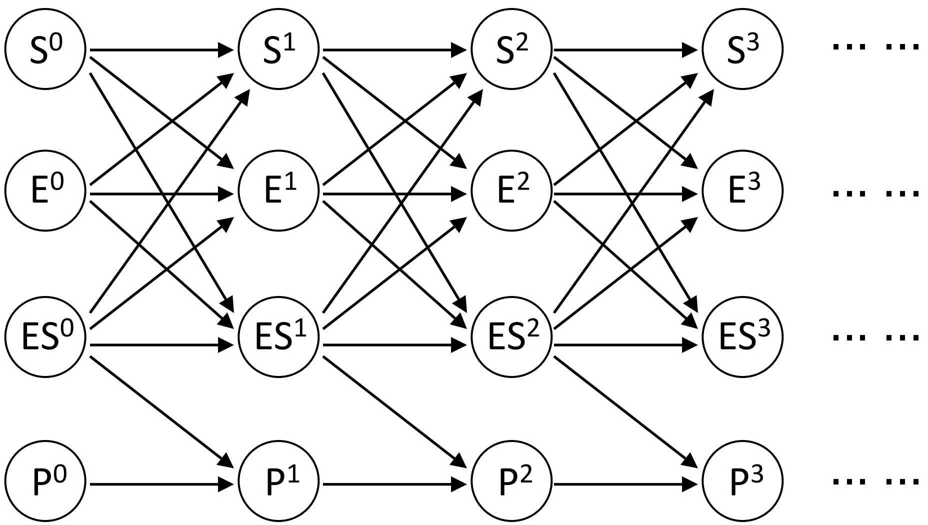


The Technique

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \overset{k_2}{\rightarrow} E + P$$

Fill up the entries in the CPTs by sampling, simulations and counting as before.

O(T . n . k^d)



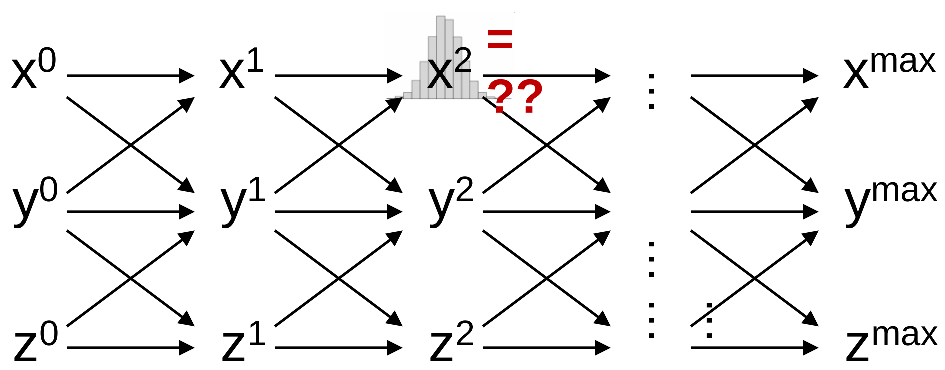
$P(S^1=0 | S^0=0, E^0=0, ES^0=0)=0.2$

$P(S^1=0 | S^0=1, E^0=0, ES^0=0)=0.4$

… …

Method

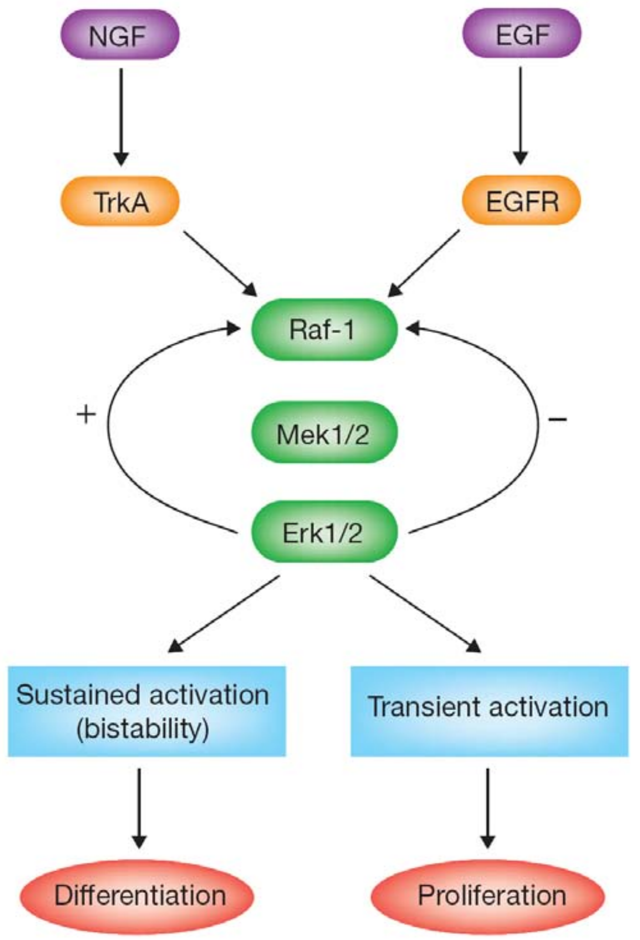
- Model analysis → Bayesian inference
- Approximate inference algorithm (Factored Frontier algorithm) (Murphy &Weiss)
- $O(T \,.\, n \,.\, k^d)$



$$Pr(x^{i+1}) = \sum_{x^i,y^i} Pr(x^{i+1}|x^i,y^i)\, Pr(x^i)\, Pr(y^i)$$

Two Synthetic Models

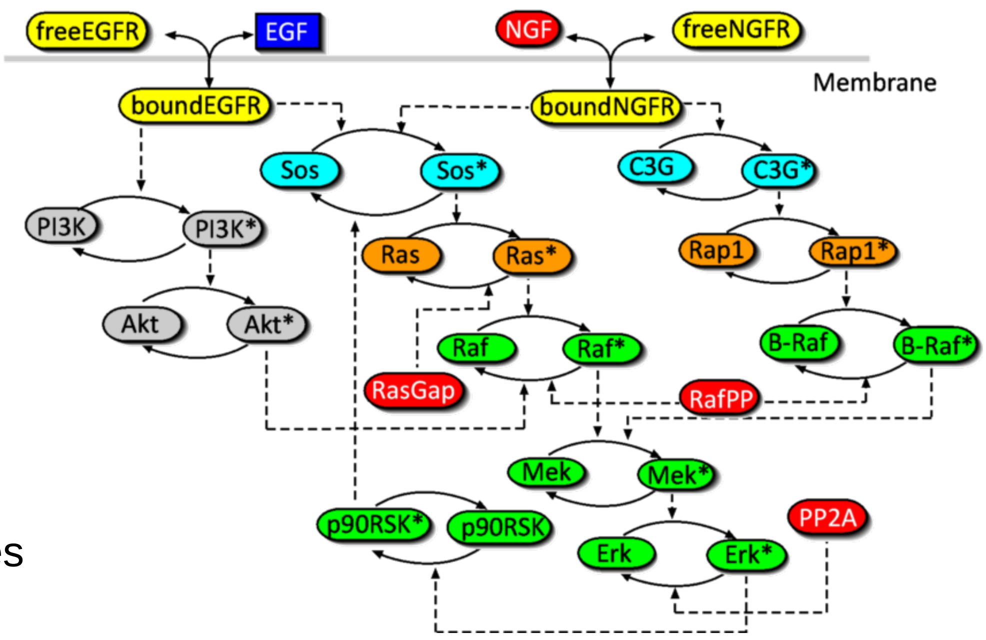
- PC12 cells
- The EGF-NGF signaling pathway
- Distinct signals dictate different cellular outcomes by activating the same signaling cascade



Kholodenko 2007

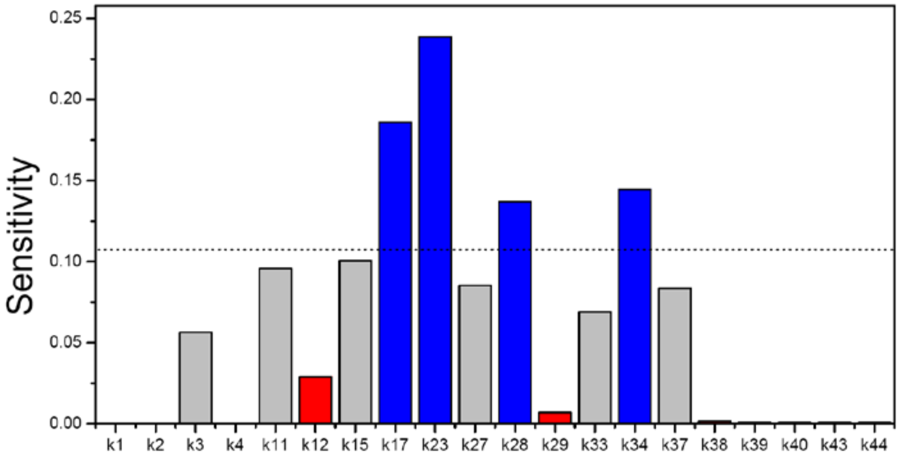
EGF-NGF Pathway

- ODE model (*Brown et al. 2004*)
  - 32 species
  - 48 parameters
  - Features:
    - Good size
    - Feedback loops
- DBN construction
  - Settings
    - 5 intervals
    - 1 min time-step, 100 minutes
    - 3 x 10^6 samples
  - Runtime
    - 4 hours on a cluster of 10 PCs



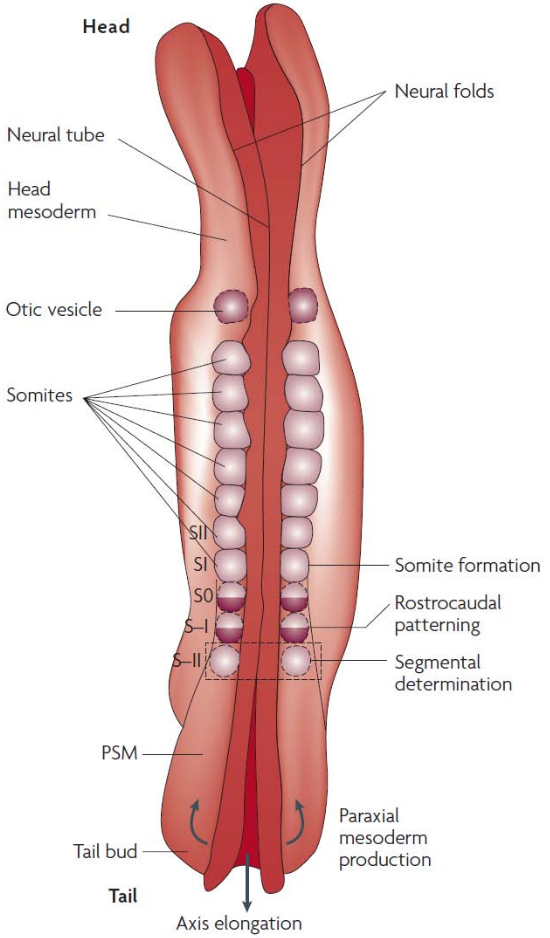
Global Sensitivity Analysis

- Running time
  - ODE based: 22 hours
  - DBN based: 34 minutes

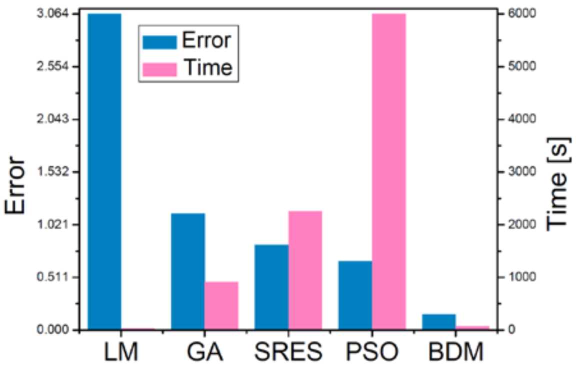


Segmentation Clock Network

- A nice example of biological rhythms
- Governs the periodic formation process of somites during embryogenesis
- The underlying signaling network couples three oscillating pathways consisting of the FGF, Wnt and Notch signaling pathways



Results



LM: Levenberg–Marquardt

GA: Genetic Algorithm

SRES: Evolutionary Strategy

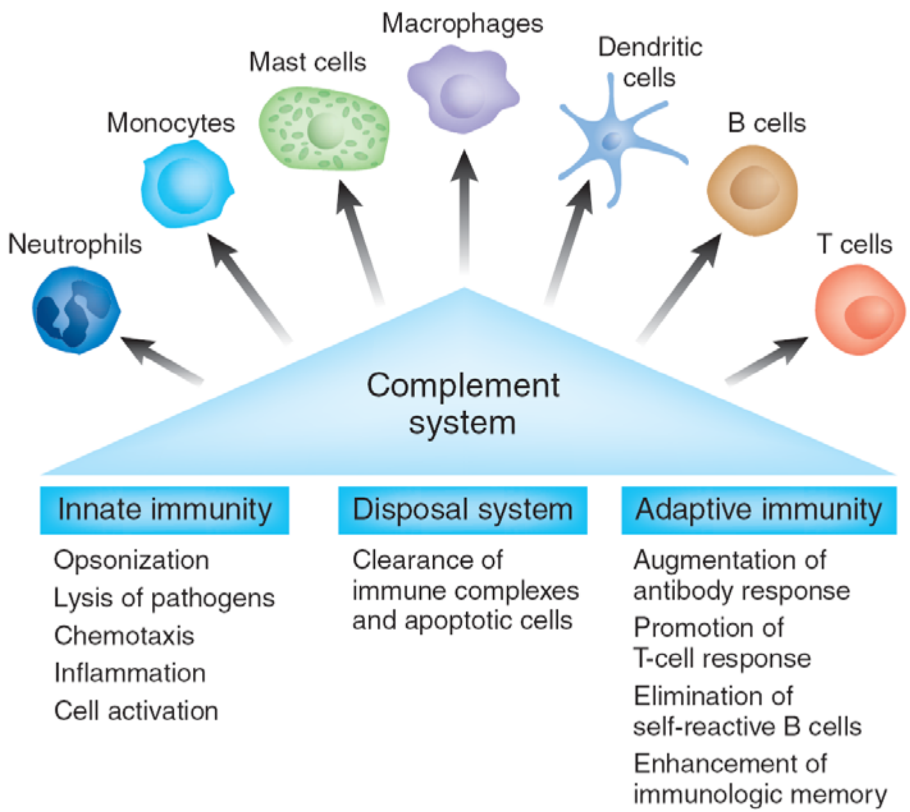
PSO: Particle Swarm Optimization

BDM: our method

- Global sensitivity analysis
  - ODE based: 81 hours
  - DBN based: 3 hours

Complement System

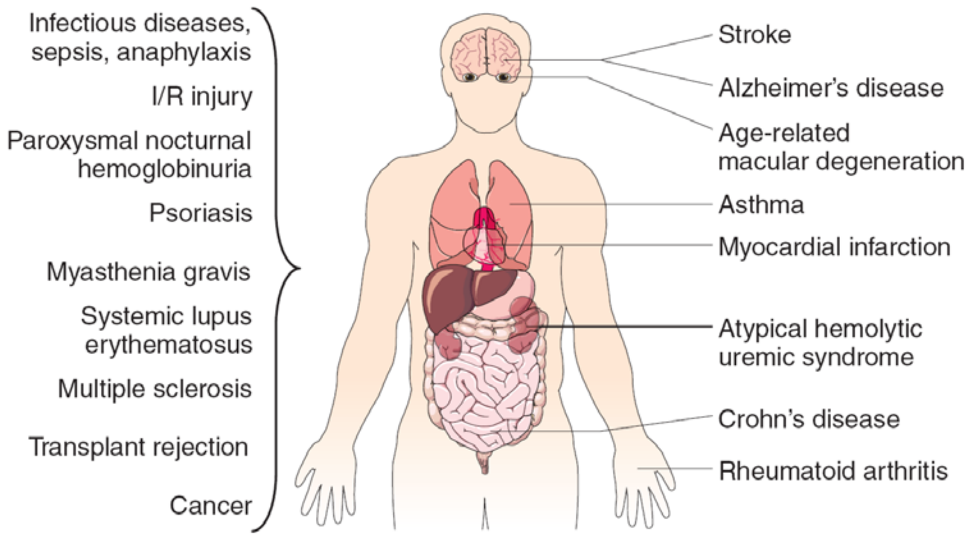
- Complement system is a critical part of the immune system



Ricklin et al. 2007

Complement System

- Adequate complement activation is necessary
- Excessive complement activation is harmful
- Dysregulation of the balance between activation and inhibition can lead to many diseases



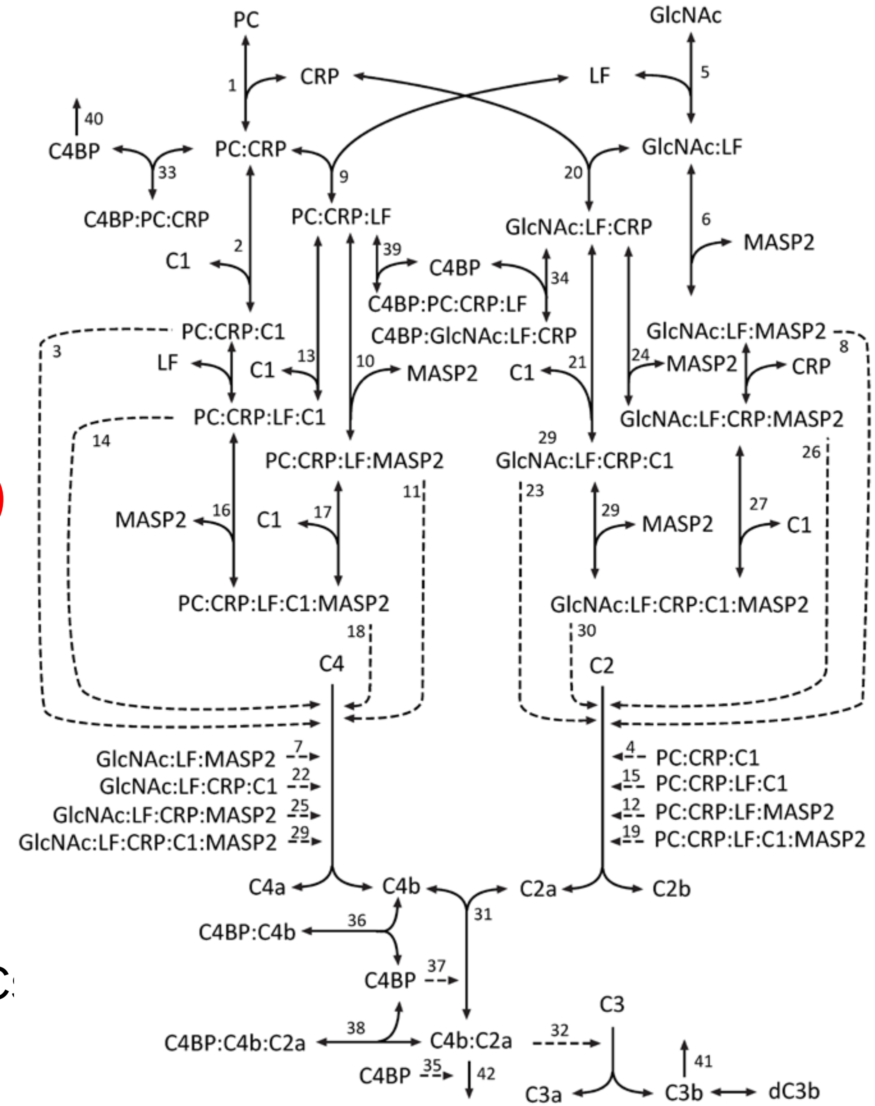
Ricklin et al. 2007

Motivation

- Quantitatively understand the regulatory mechanisms of complement system
  - How is the complement activity enhanced under inflammation condition?
  - How is the excessive response of the complement avoided?
- The model:
  - Classical pathway + the lectin pathway
    - enhancement
  - Inhibitory mechanisms
    - C4BP

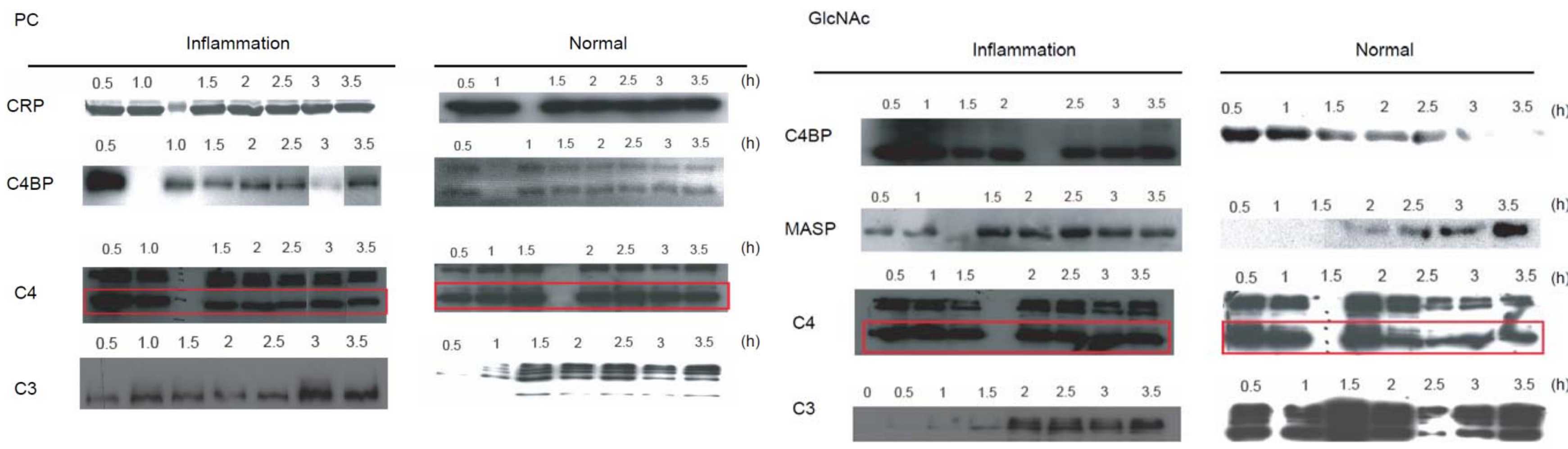
Complement System

- ODE Model
  - 42 Species
  - 45 Reactions
    - Mass law
    - Michaelis-Menten kinetics
  - 92 Parameters ***(71 unknown!)***
- DBN Construction
  - Settings
    - 6 intervals
    - 100s time-step, 12600s
    - 1.2 x 10^6 samples
  - Runtime
    - 12 hours on a cluster of 20 PC



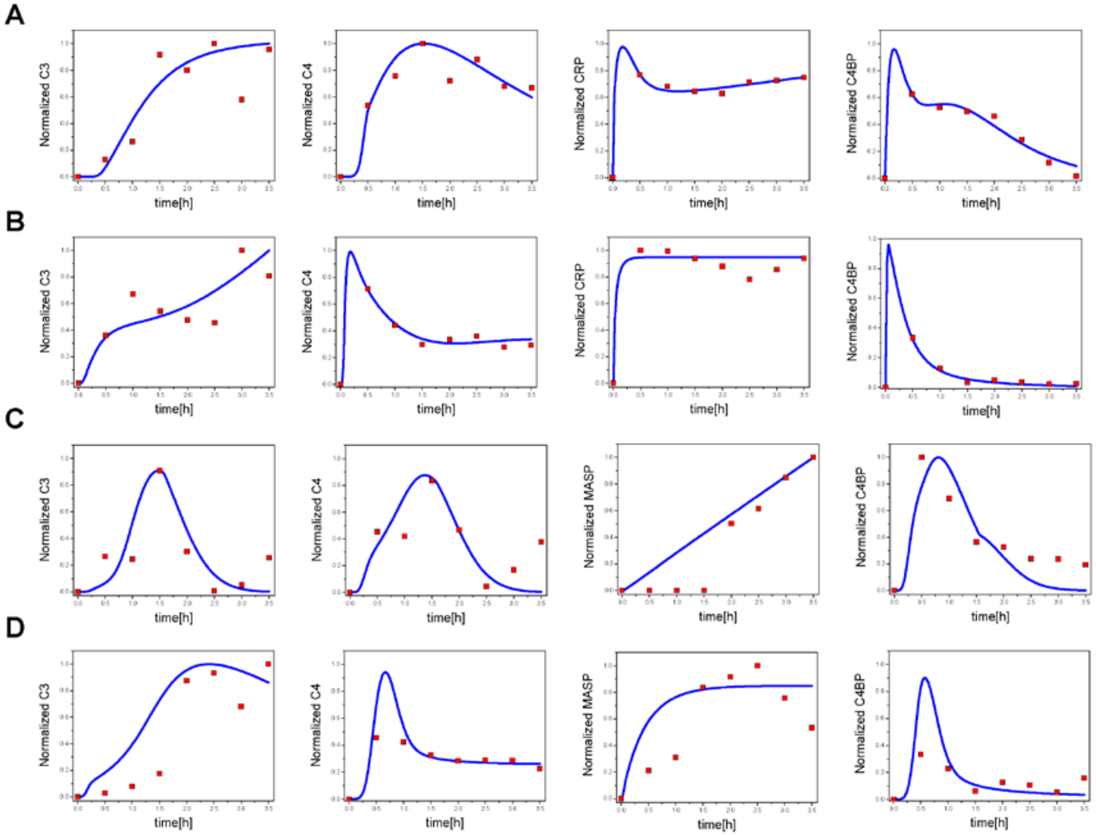
Parameter estimation

- Training data
  - 4 proteins, 7 time points, 4 conditions



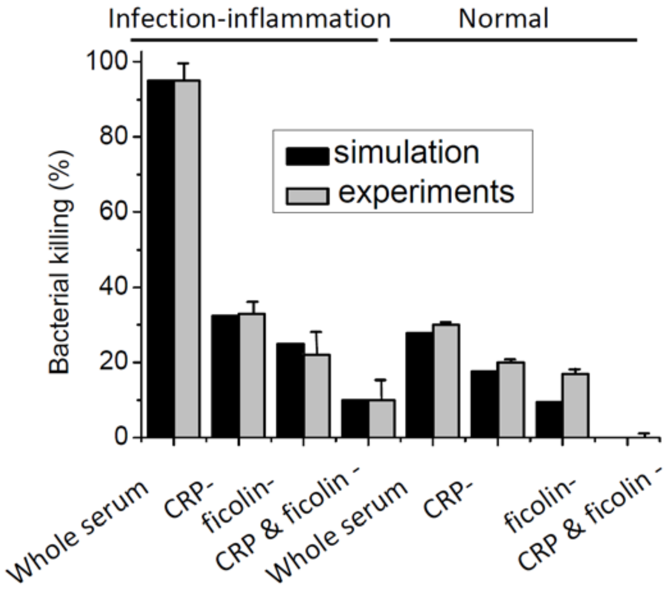
* inflammation condition: pH 6.5 calcium 2 nM

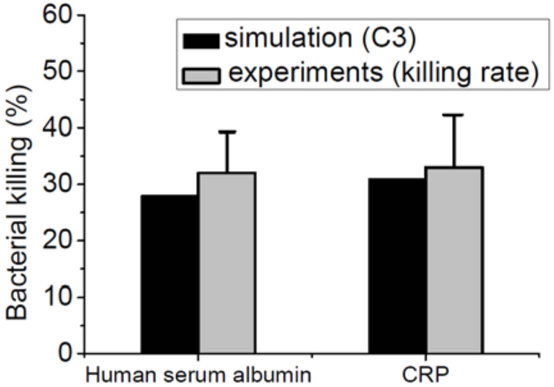
Model Calibration



Model validation

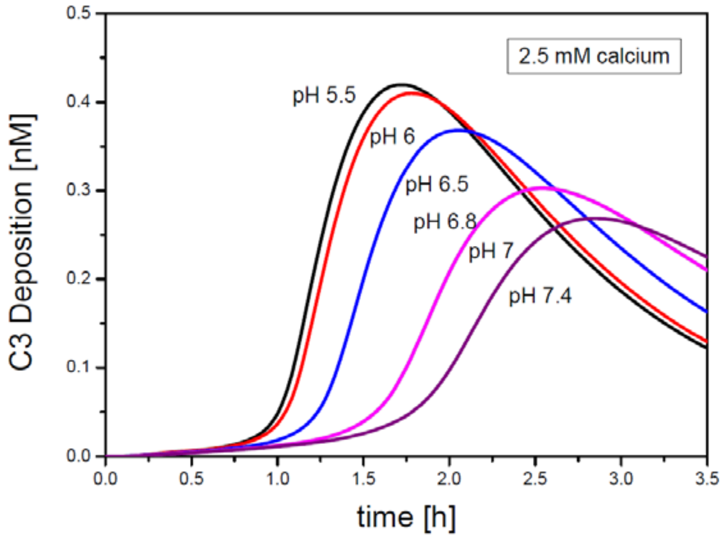
- Validated the model using previous published data (*Zhang et al 2009)*

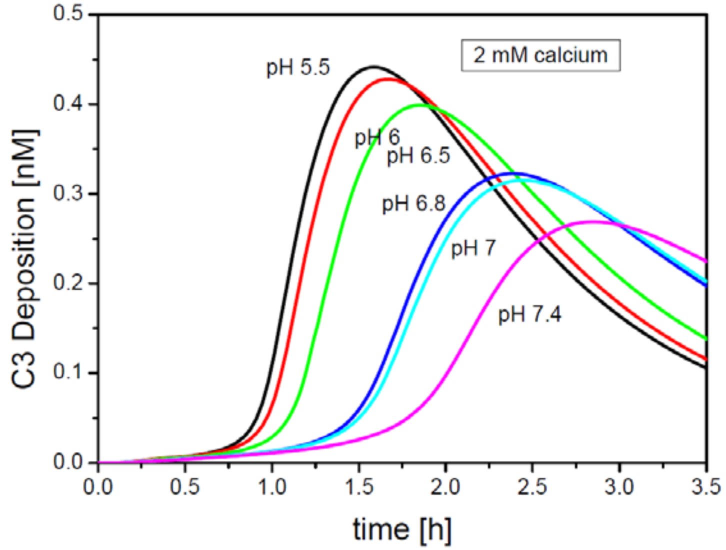




Enhancement mechanism

- The antimicrobial response is sensitive to the pH and calcium level

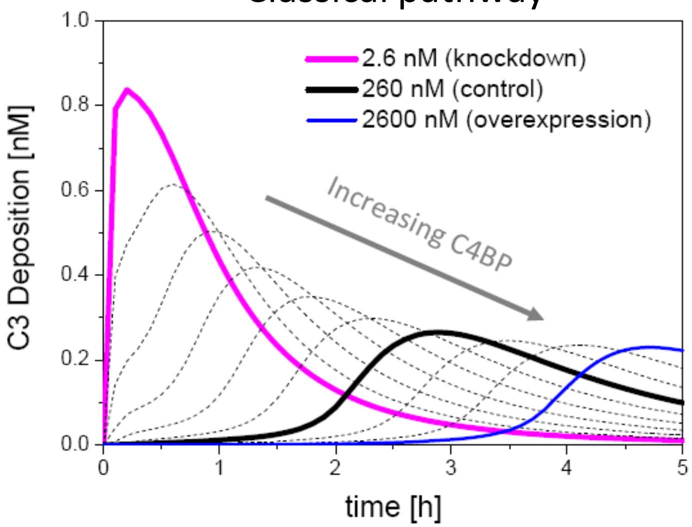




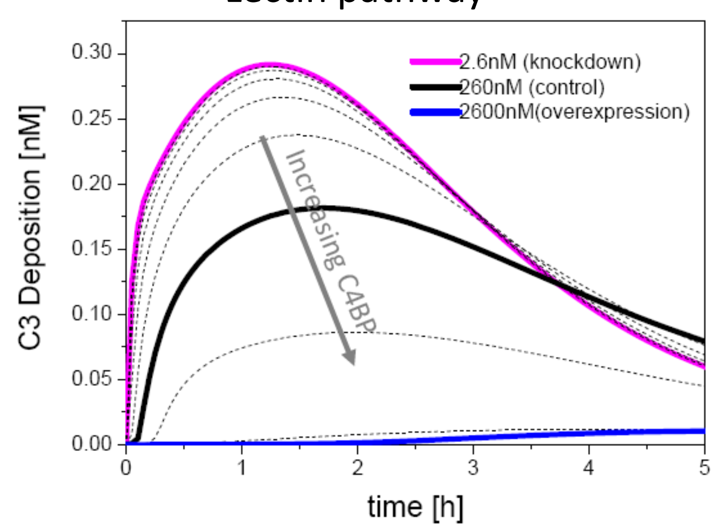
Model prediction: The regulatory effect of C4BP

- C4BP maintains classical complement activation but delays the maximal response time.
- But attenuates the lectin pathway activation.

Classical pathway

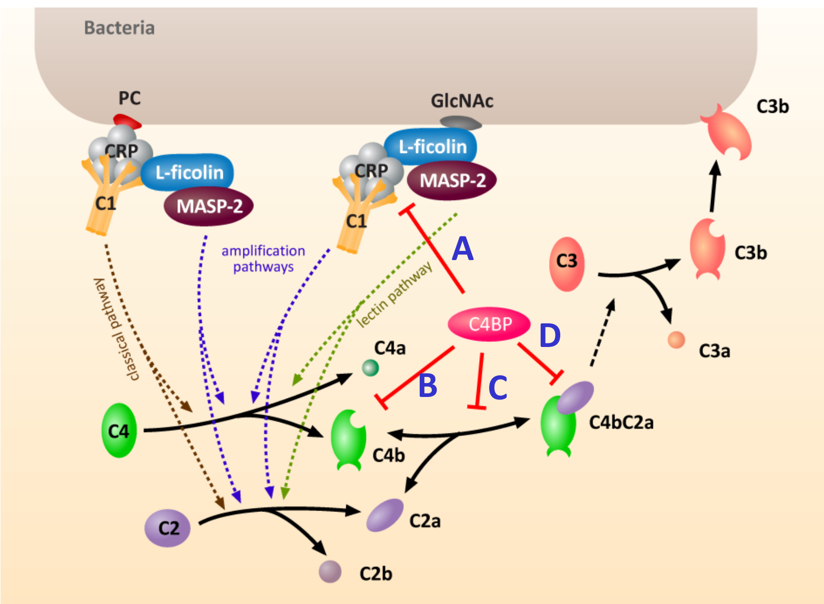


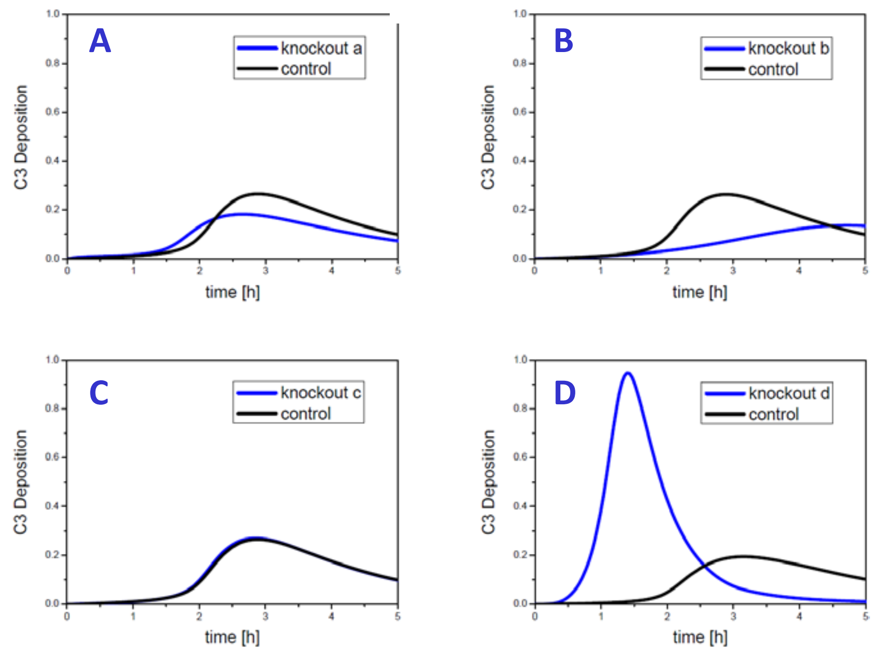
Lectin pathway



The regulatory mechanism of C4BP

- The major inhibitory role of C4BP is to facilitate the decay of C3 convertase





Results

- Both predictions concerning C4BP were experimentally verified.
- *“A Computational and Experimental Study of the Regulatory Mechanisms of the Complement System”*
  - **Bing Liu, Jing Zhang, Pei Yi Tan, David Hsu, Anna M. Blom, Benjamin Leong, Sunil Sethi, Bow Ho, Jeak Ling Ding, P.S. Thiagarajan**

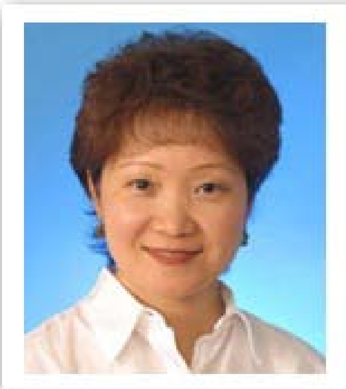
PLoS Computational Biology Vol. 7, Issue.1 (2011)

Conclusion

- The DBN approximation technique is useful and realistic.
- (Approximate) probabilistic verification methods
- Implementation on a GPU platform.
- Error analysis.
- A general abstraction scheme?
  - When is it effective?

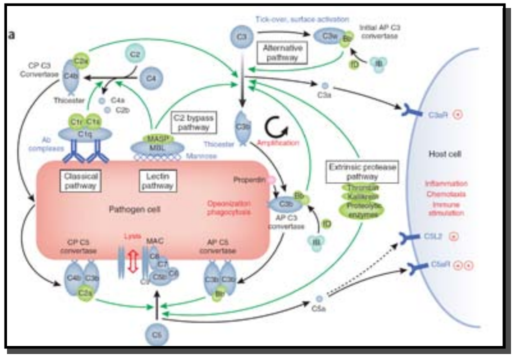
Current Collaborators

Ding Jeak Ling

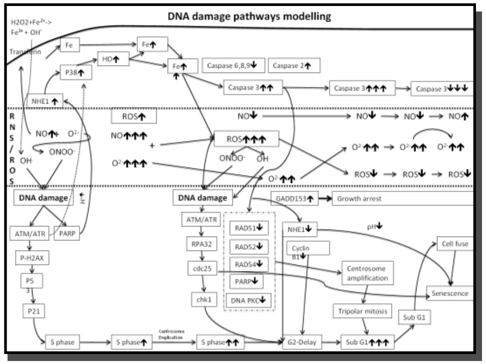


Marie-Veronique Clement

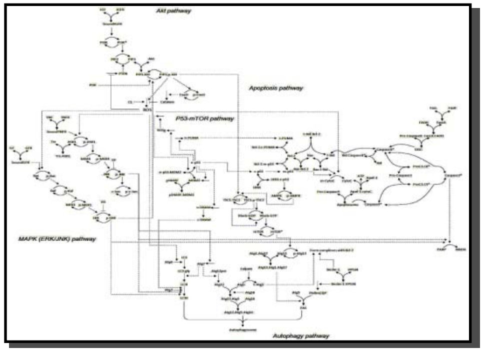
Shazib Pervaiz



Complement pathway



DNA damage pathway



Apoptosis-Autophage- pathway

Our group:

S. Akshay

Liu Bing

Abhinav Dubey

Blaise Genest

Benjamin Gyori

David Hsu

Suchee Palaniappan

P.S. Thiagarajan

Wang Junjie

Gireedhar Venkatachalam