Probabilistic Approximations of ODEs Based Signaling Pathways Dynamics

P.S. Thiagarajan

School of Computing, National University of Singapore

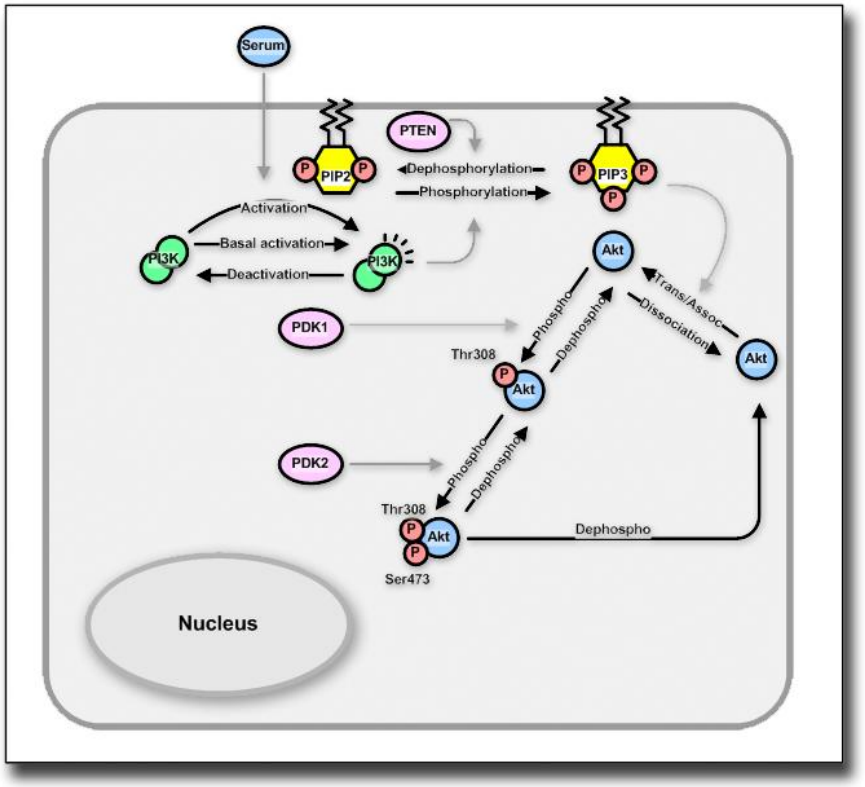
Biopathways

- Biopathways:
  - Metabolic Pathways
  - ***Signaling Pathways***
  - Gene Regulatory Networks

Signaling Pathways

- Chemical reactions in response to external signals (ligands)
- Signals pass into the nucleus through a series of protein modifications

  ‘Data transfer’ mechanism of the cell



A Common Modeling Approach

- View a pathway as a network of bio-chemical reactions
- Model the network as a system of ODEs
  - One for each molecular species
  - Reaction kinetics: Mass action law, Michelis-Menten, Hill, etc.
- Study the ODE system dynamics.

The ODEs model

$$S + E \underset{k_2}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_3} E + P$$

Assume mass law.

$$\frac{dS}{dt} = -k_1 \cdot S \cdot E + k_2 \cdot ES$$

$$\frac{dE}{dt} = -k_1 \cdot S \cdot E + (k_2 + k_3) \cdot ES$$

$$\frac{dES}{dt} = k_1 \cdot S \cdot E - (k_2 + k_3) \cdot ES$$

$$\frac{dP}{dt} = k_3 \cdot ES$$

- Alternative approach:
  - Keep track of exact number of molecules of each type. Simulate the dynamics by executing one reaction at a time stochastically (CTMCs)
  - Stochastic simulations (Gillespie's algorithm)
  - Kappa , BioNetGen, PRISM, Bio-Pepa, ..

ODEs: Major Hurdles

- Many unknown rate constants.
- Must be estimated using limited data:
  - ***Low precision***, population-based, noisy

Major Hurdles

- High dimensional non-linear system
  - no closed-form solutions
  - must resort to numerical simulations
  - *point values of initial states/data will not be available*
  - ***a large number of numerical simulations needed for answering each analysis question***

“Polling” based approximation

- Start with an ODEs system.
- Discretize the time and value domains.
- Assume a (uniform) distribution of initial states
- Generate a “sufficiently” large number of trajectories by
  - Sampling the initial states and numerical simulations.

The “exit poll” Idea

- Encode this collection of discretized trajectories as ***a dynamic Bayesian network.***
- **ODEs $\rightarrow$ DBN**
- Pay the one-time cost of constructing the DBN approximation.
- Do analysis using Bayesian inferencing on the DBN.

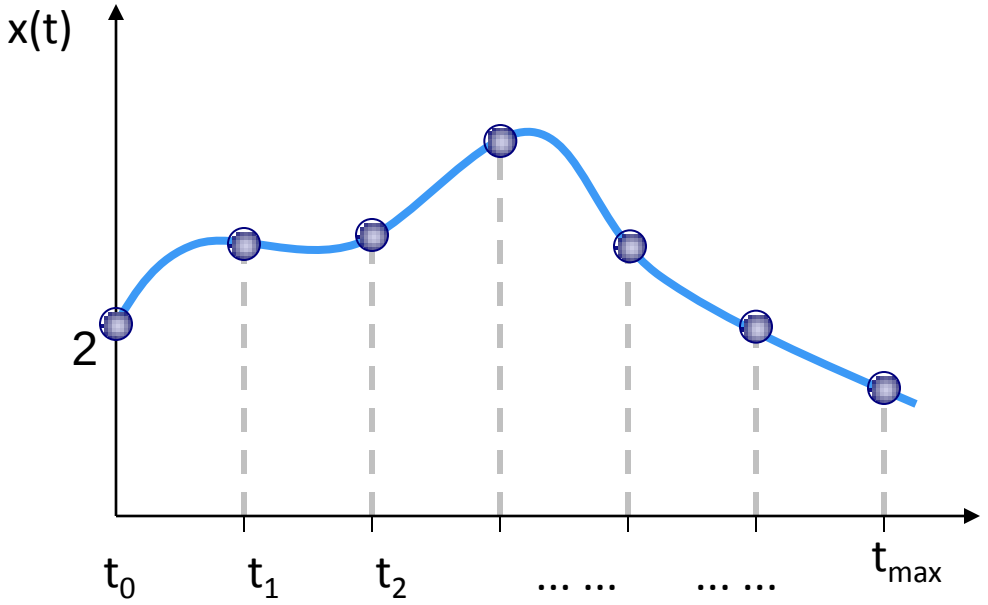
Time Discretization

- Observe the system only at a ***finite*** number of time points.

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

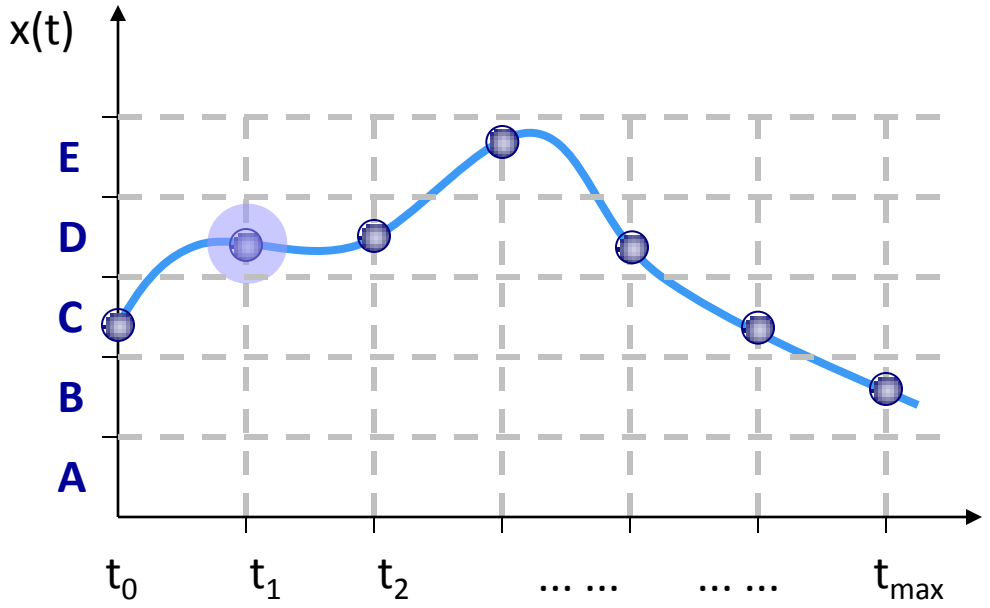
x(0) = 2

$x(t) = t^3 + 4t + 2$



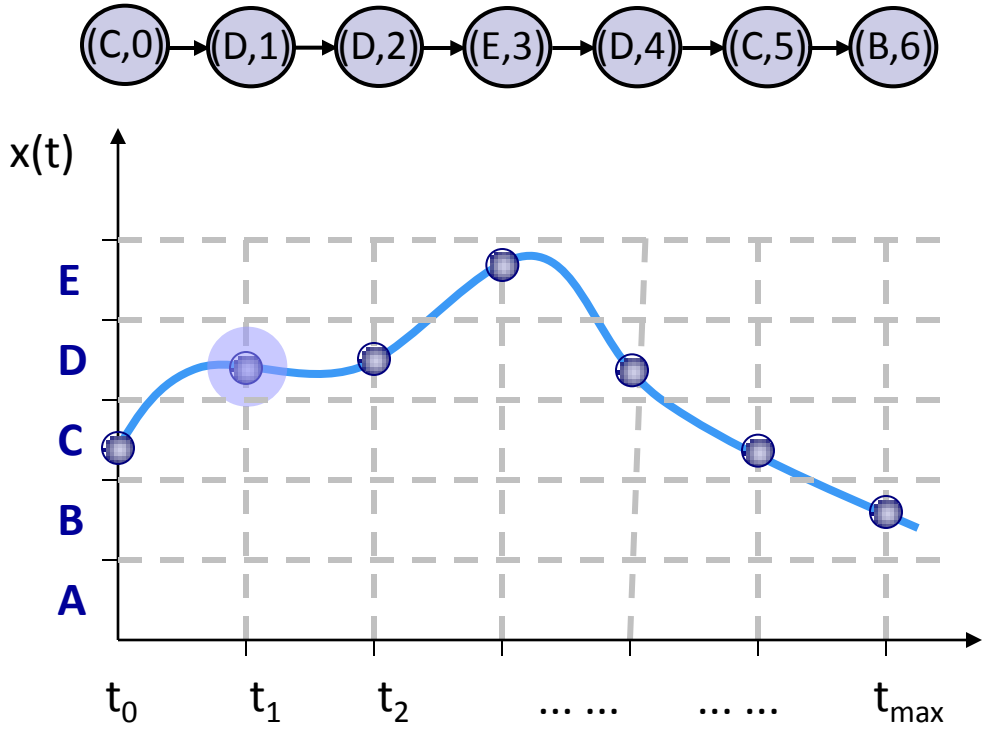
Value Discretization

- Observe only with ***bounded precision***



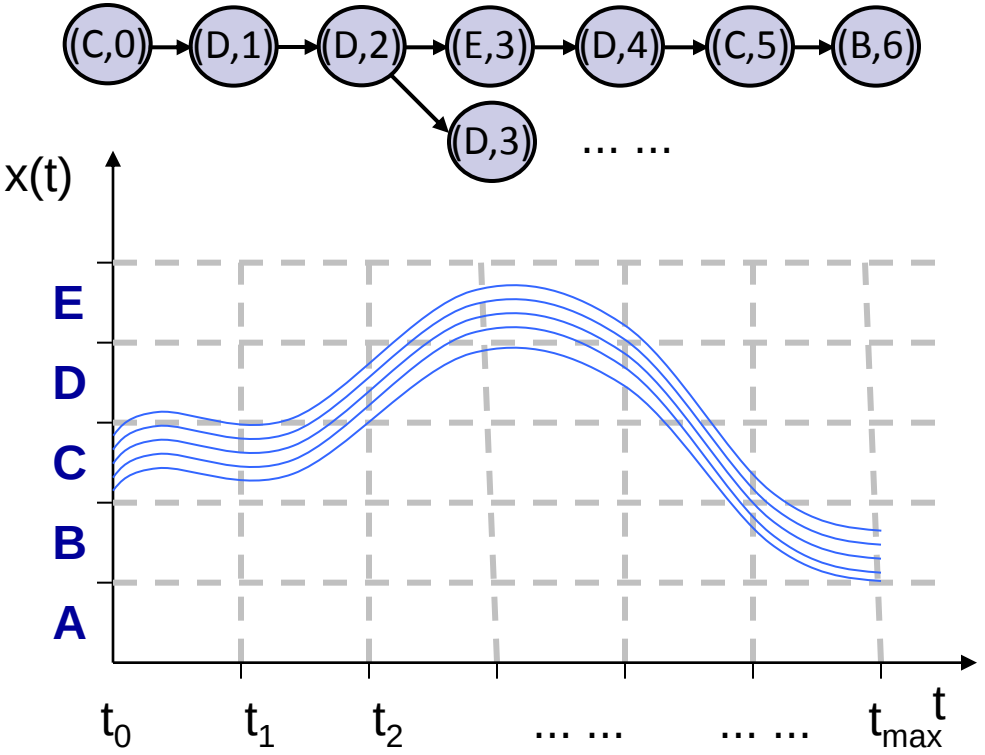
Symbolic trajectories

- A trajectory is recorded as a finite sequence of ***discrete values***.



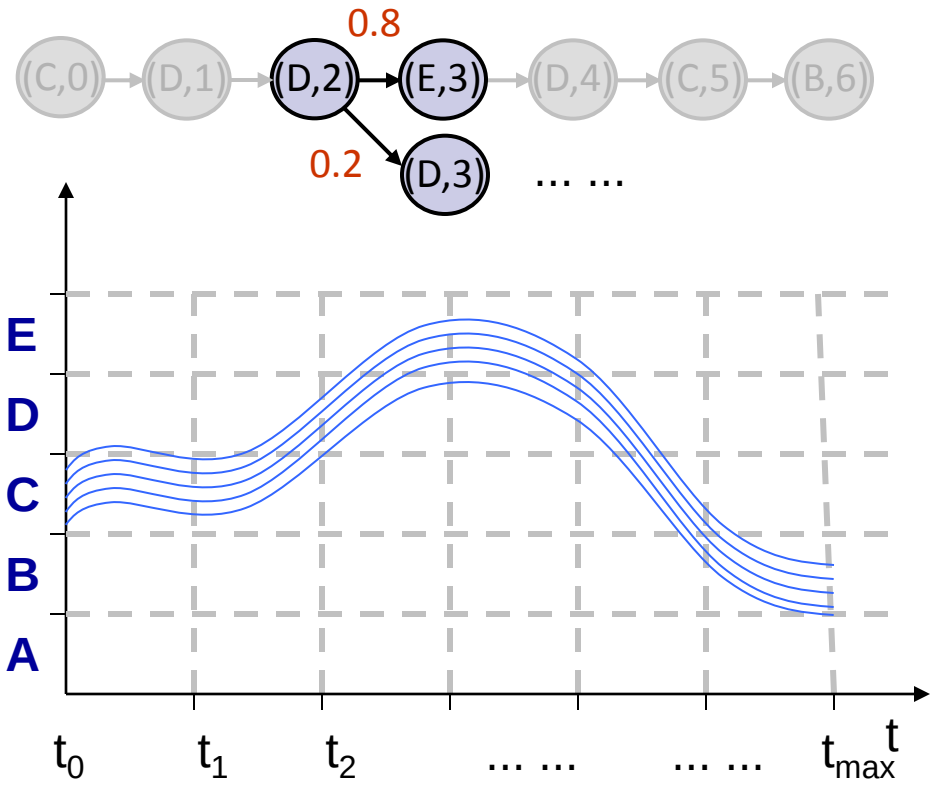
Collection of Trajectories

- Assume a prior distribution of the initial states.
- Uncountably many trajectories. Represented as a set of (timed) finite sequences.



Piecing trajectories together..

- In fact, a probabilistic transition system.
- 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 Justification

- 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 continuously differentiable in C.
  - Mass-law, Michaelis-Menton,…
- Then the solution Φ_t : C $\rightarrow$ C (for each t) exists, is unique, a bijection, continuous and hence ***measurable.***
- ***But the transition probabilities can't be computed.***

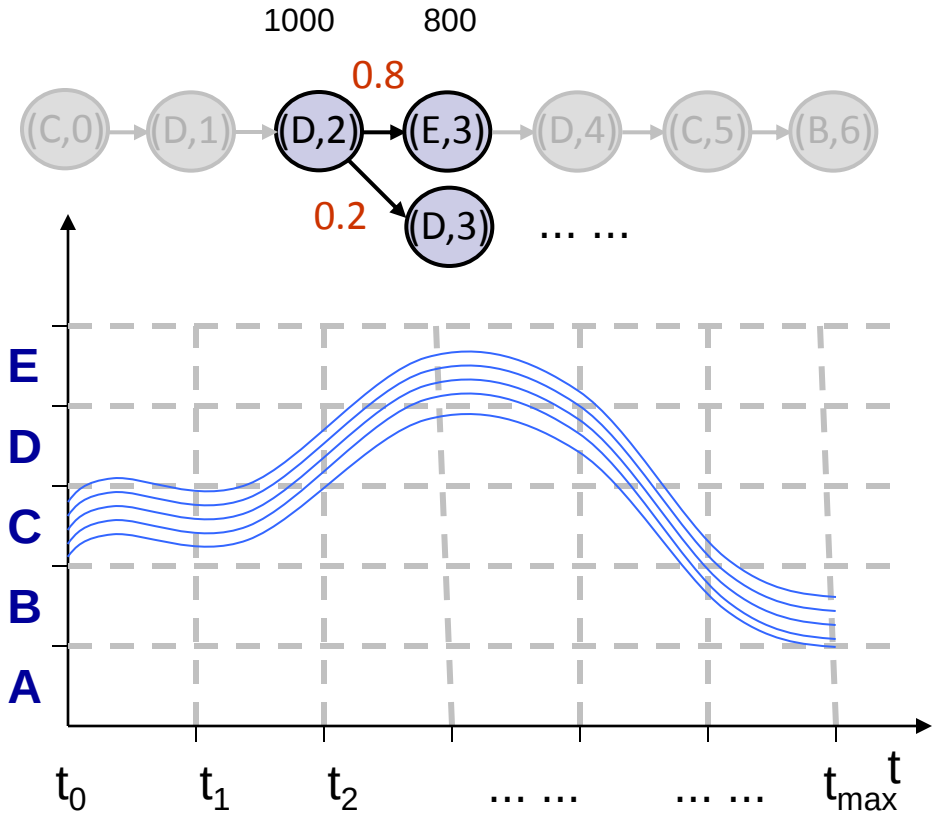
A computational approximation

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



Infeasible Size!

- But the ***transition system*** will be huge.
  - $O(T \cdot k^n)$
  - $k \geq 2$ and n ($\approx$ 50-100).

Compact Representation

- Exploit the network structure (additional independence assumptions) to construct a DBN instead.
- The DBN is a ***factored*** form of the probabilistic transition system.

The DBN representation

$$S + E \underset{k_2}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_3} E + P$$

Assume mass law.

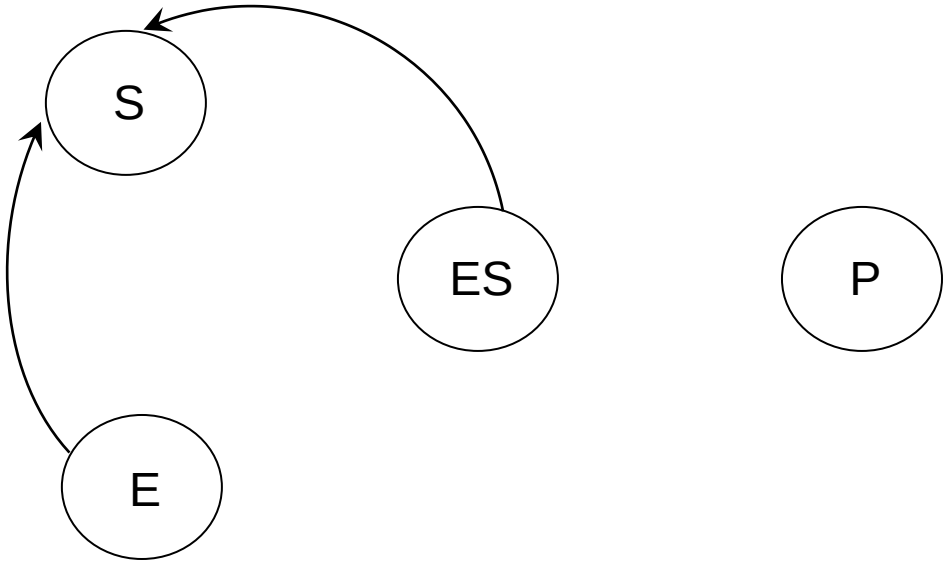
$$\frac{dS}{dt} = -k_1 \cdot S \cdot E + k_2 \cdot ES$$

$$\frac{dE}{dt} = -k_1 \cdot S \cdot E + (k_2 + k_3) \cdot ES$$

$$\frac{dES}{dt} = k_1 \cdot S \cdot E - (k_2 + k_3) \cdot ES$$

$$\frac{dP}{dt} = k_3 \cdot ES$$

$$S + E \underset{k_2}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_3} E + P$$



Dependency diagram

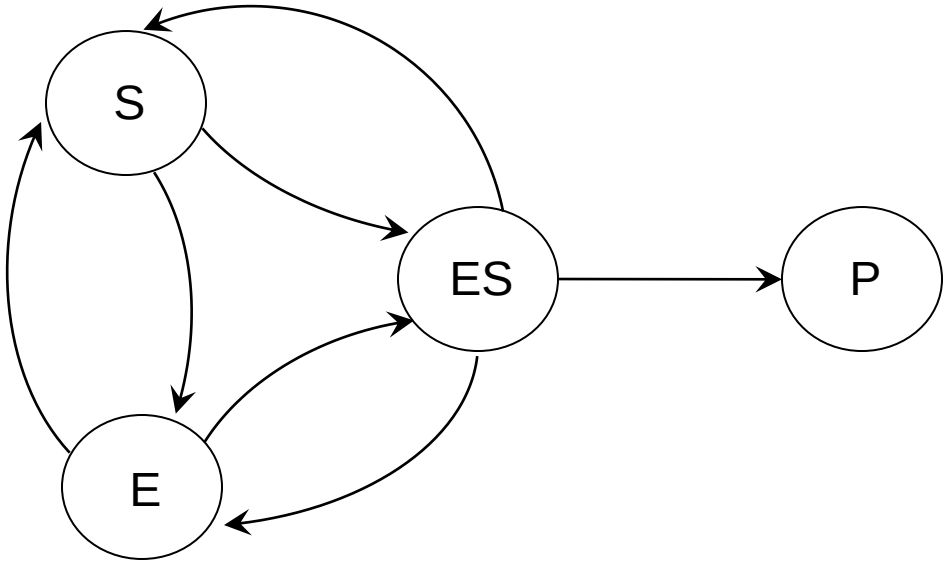
$$\frac{dS}{dt} = -k_1 \cdot S \cdot E + k_2 \cdot ES$$

$$\frac{dE}{dt} = -k_1 \cdot S \cdot E + (k_2 + k_3) \cdot ES$$

$$\frac{dES}{dt} = k_1 \cdot S \cdot E - (k_2 + k_3) \cdot ES$$

$$\frac{dP}{dt} = k_3 \cdot ES$$

$$S + E \underset{k_2}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_3} E + P$$



Dependency diagram

$$\frac{dS}{dt} = -k_1 \cdot S \cdot E + k_2 \cdot ES$$

$$\frac{dE}{dt} = -k_1 \cdot S \cdot E + (k_2 + k_3) \cdot ES$$

$$\frac{dES}{dt} = k_1 \cdot S \cdot E - (k_2 + k_3) \cdot ES$$

$$\frac{dP}{dt} = k_3 \cdot ES$$

The DBN Representation

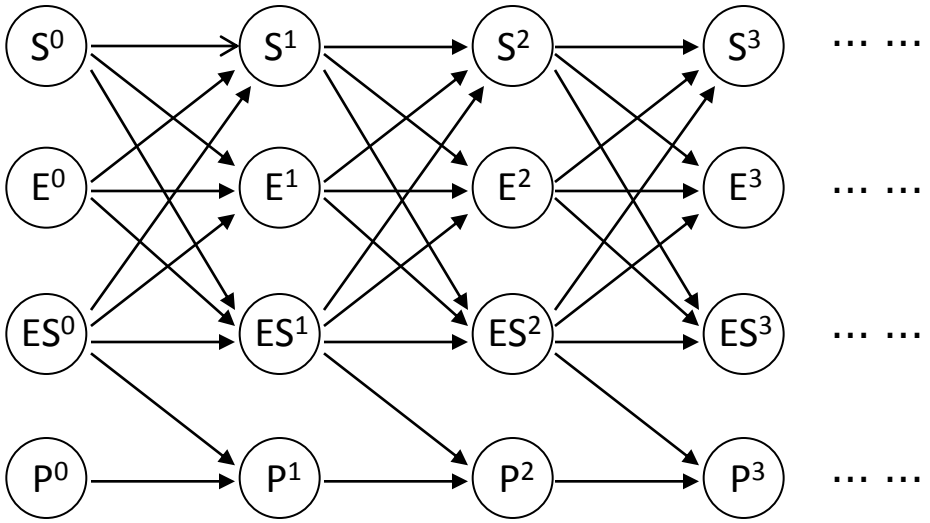
$$S + E \underset{k_2}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_3} E + P$$

$$\frac{dS}{dt} = -k_1 \cdot S \cdot E + k_2 \cdot ES$$

$$\frac{dE}{dt} = -k_1 \cdot S \cdot E + (k_2 + k_3) \cdot ES$$

$$\frac{dES}{dt} = k_1 \cdot S \cdot E - (k_2 + k_3) \cdot ES$$

$$\frac{dP}{dt} = k_3 \cdot ES$$

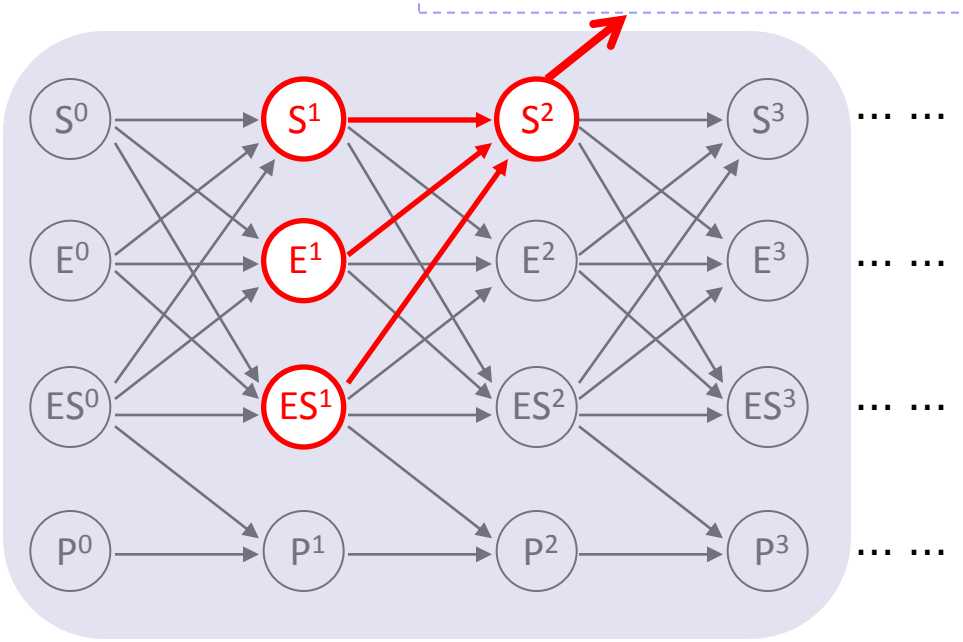


- Each node has a CPT associated with it.
- This specifies the local (probabilistic) dynamics.

$P(S^2=C \mid S^1=B, E^1=C, ES^1=B) = 0.2$

$P(S^2=C \mid S^1=B, E^1=C, ES^1=C) = 0.1$

$P(S^2=A \mid S^1=A, E^1=A, ES^1=C) = 0.05$



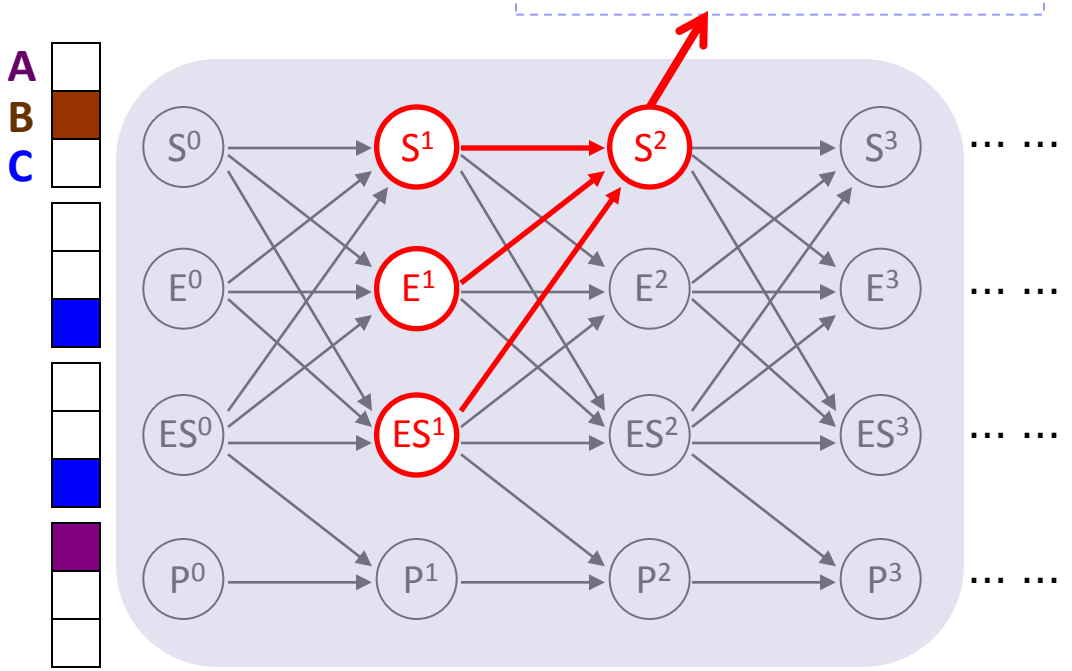
A
B
C

- Fill up the entries in the CPTs by sampling, simulations and counting

$P(S^2=C|S^1=B,E^1=C,ES^1=B)= 0.2$

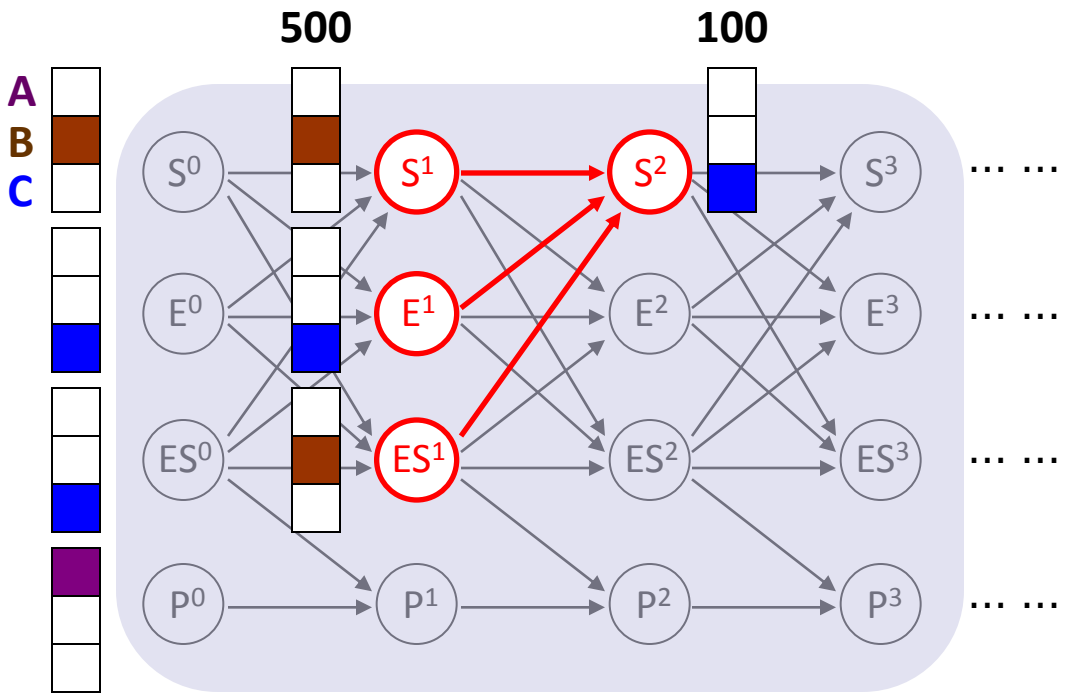
$P(S^2=C|S^1=B,E^1=C,ES^1=C)= 0.1$

$P(S^2=A|S^1=A,E^1=A,ES^1=C)= 0.05$



Computational Approximation

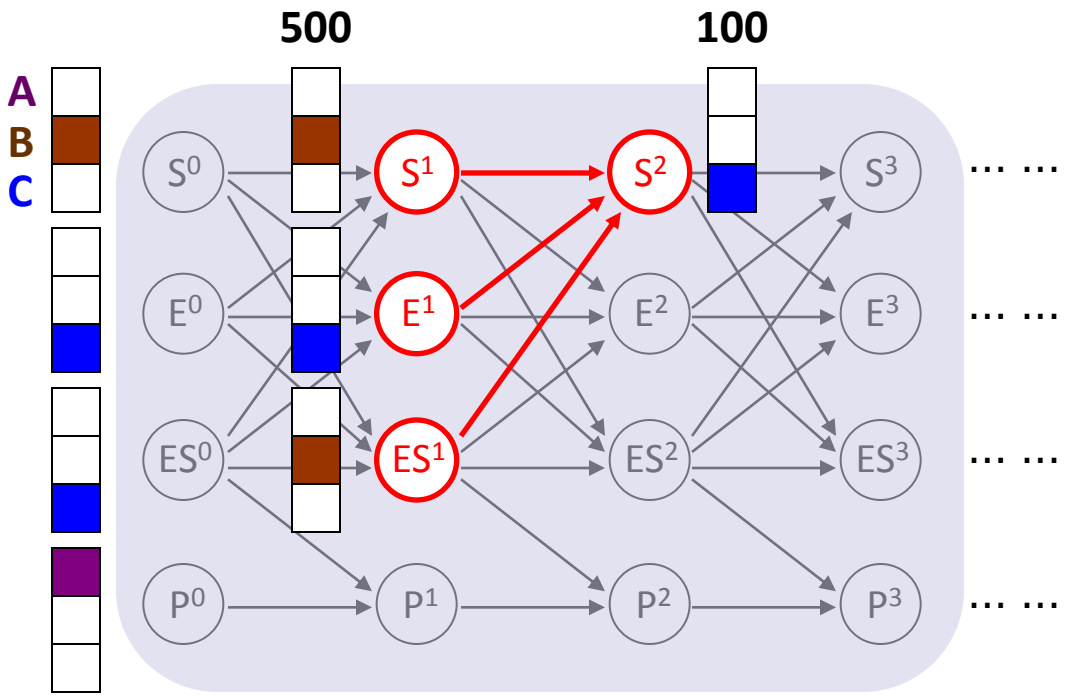
- Fill up the entries in the CPTs by sampling, simulations and counting



The Technique

$P(S^2=\mathbf{C} \mid S^1=\mathbf{B}, E^1=\mathbf{C}, ES^1=\mathbf{B}) = 100/500 = 0.2$

- Fill up the entries in the CPTs by sampling, simulations and counting

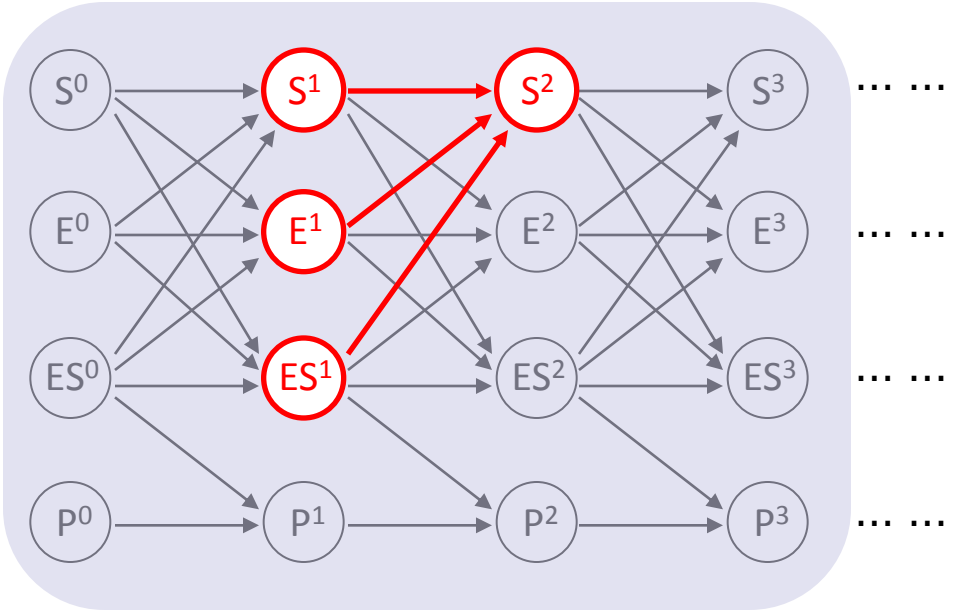


The Technique

The size of the DBN is:

$O(T \cdot n \cdot k^d)$

d will be usually much smaller than n.



Unknown rate constants

$$S + E \underset{k_2 = 0.2}{\overset{k_1 = 0.1}{\rightleftharpoons}} ES \xrightarrow{k_3} E + P$$

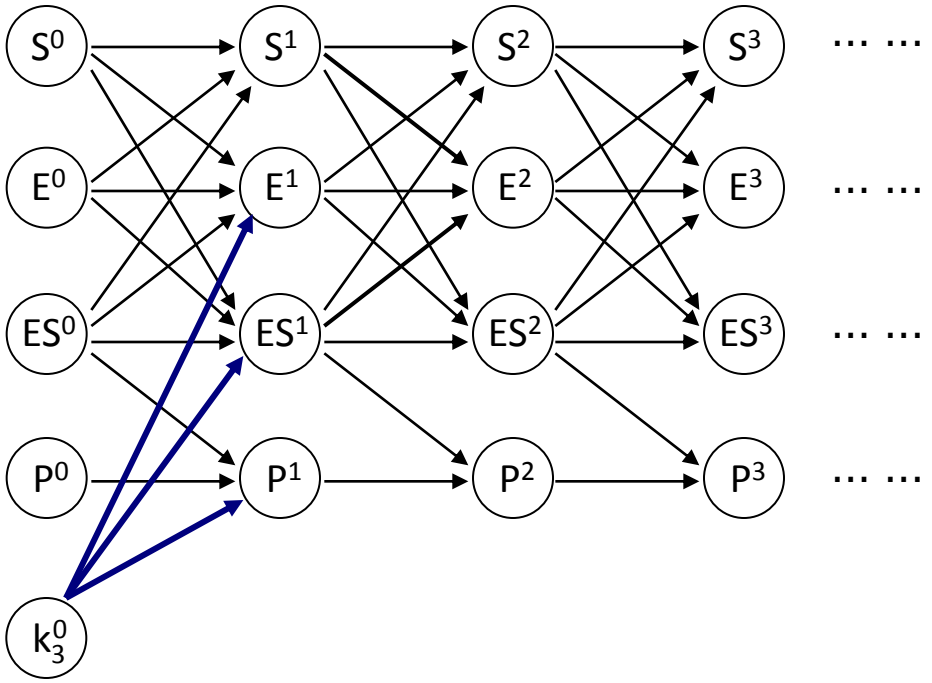
$$\frac{dS}{dt} = -0.1 \cdot S \cdot E + 0.2 \cdot ES$$

$$\frac{dE}{dt} = -0.1 \cdot S \cdot E + (0.2 + k_3) \cdot ES$$

$$\frac{dES}{dt} = 0.1 \cdot S \cdot E - (0.2 + k_3) \cdot ES$$

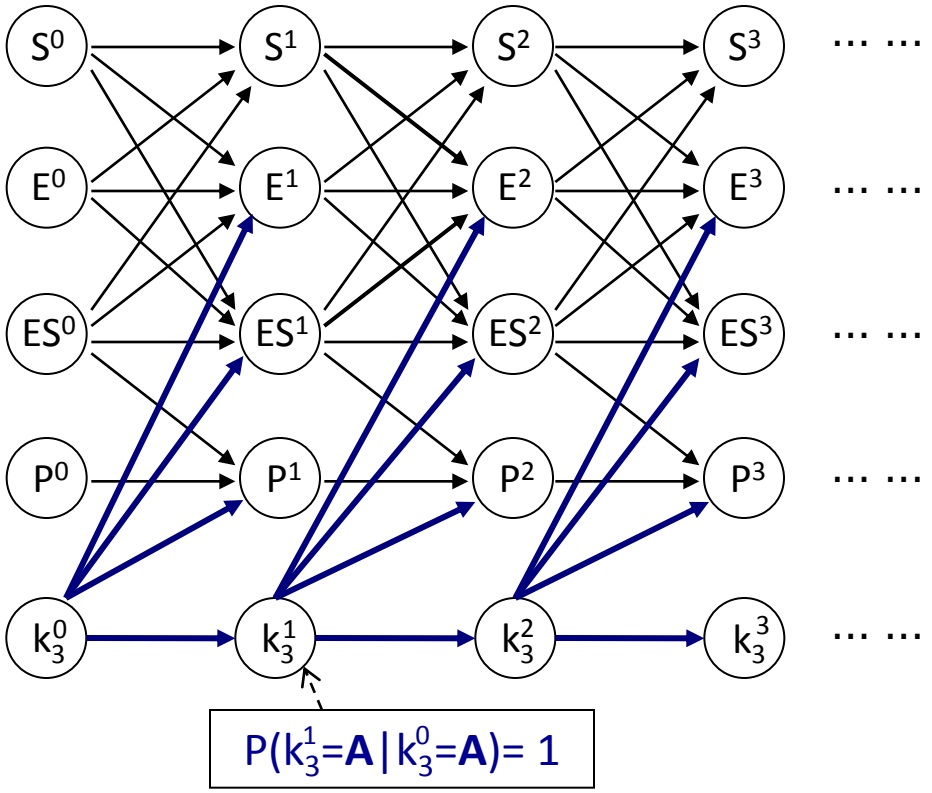
$$\frac{dP}{dt} = k_3 \cdot ES$$

$$\frac{dk_3}{dt} = 0$$



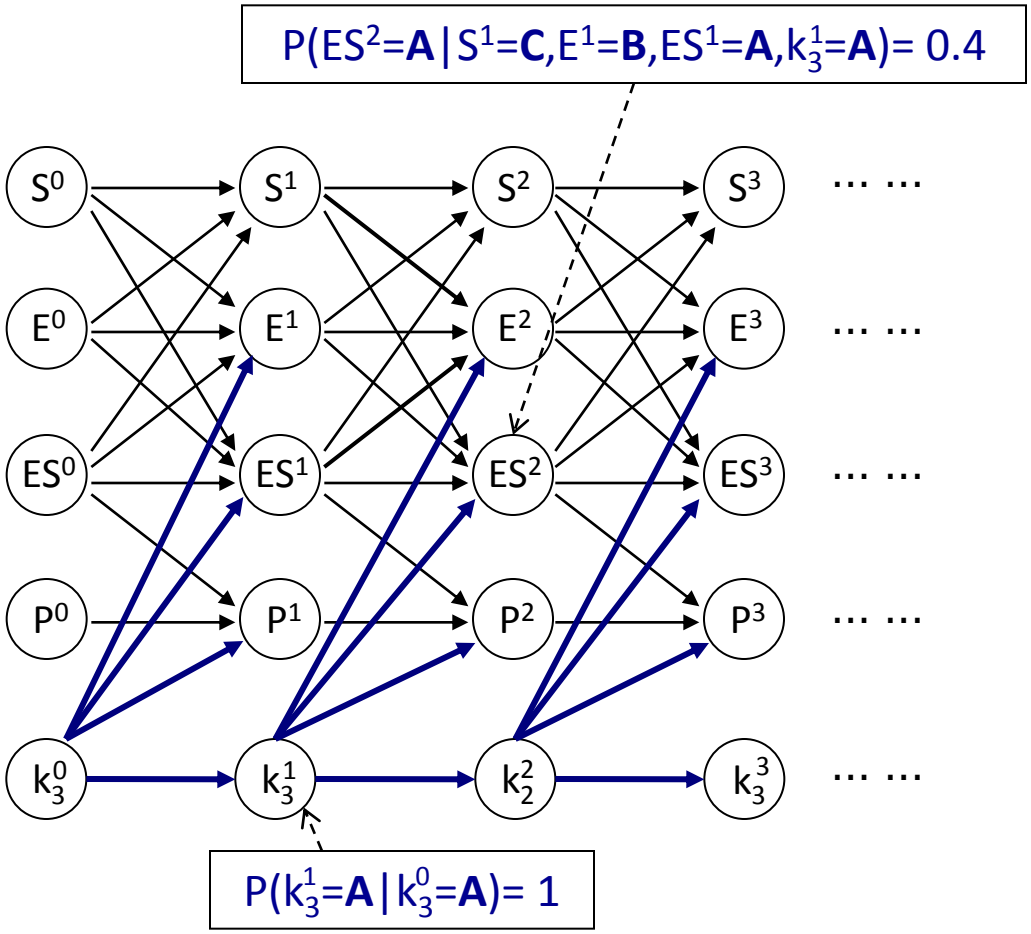
Unknown rate constants

During the numerical generation of a trajectory, the value of k_3 does not change after sampling.



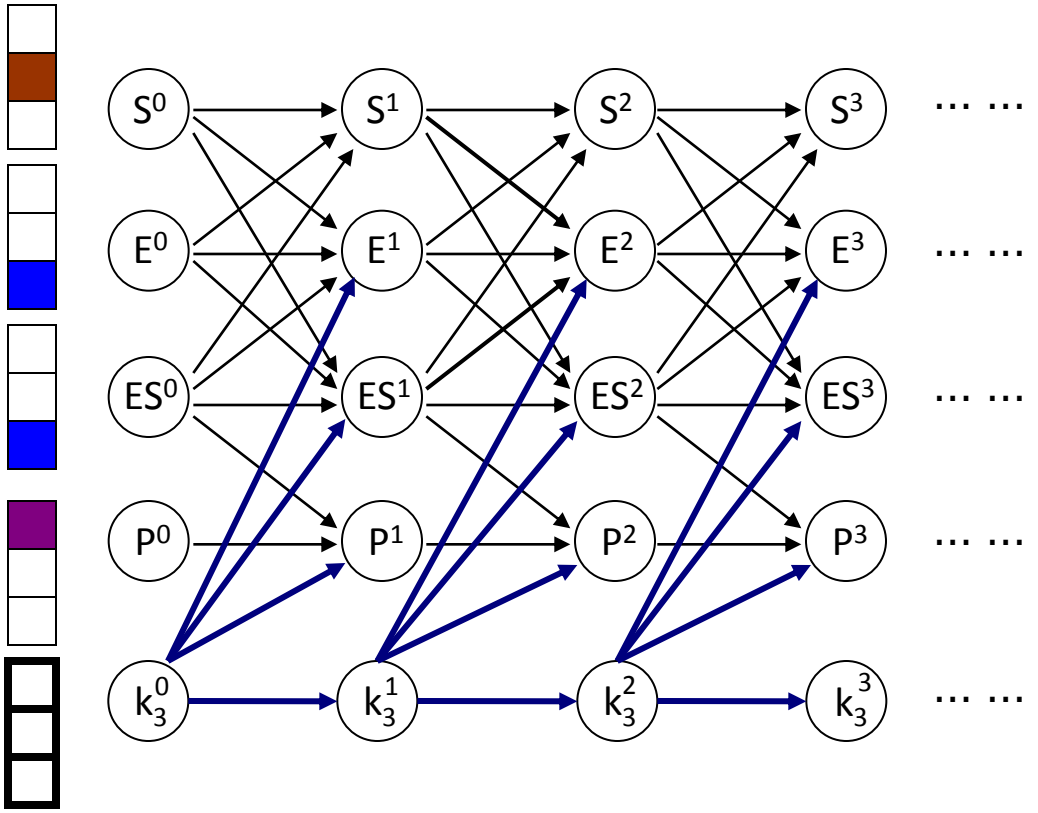
Unknown rate constants

During the numerical generation of a trajectory, the value of k_3 does not change after sampling.



Unknown rate constants

Sample uniformly across ***all*** the Intervals.



DBN based Analysis

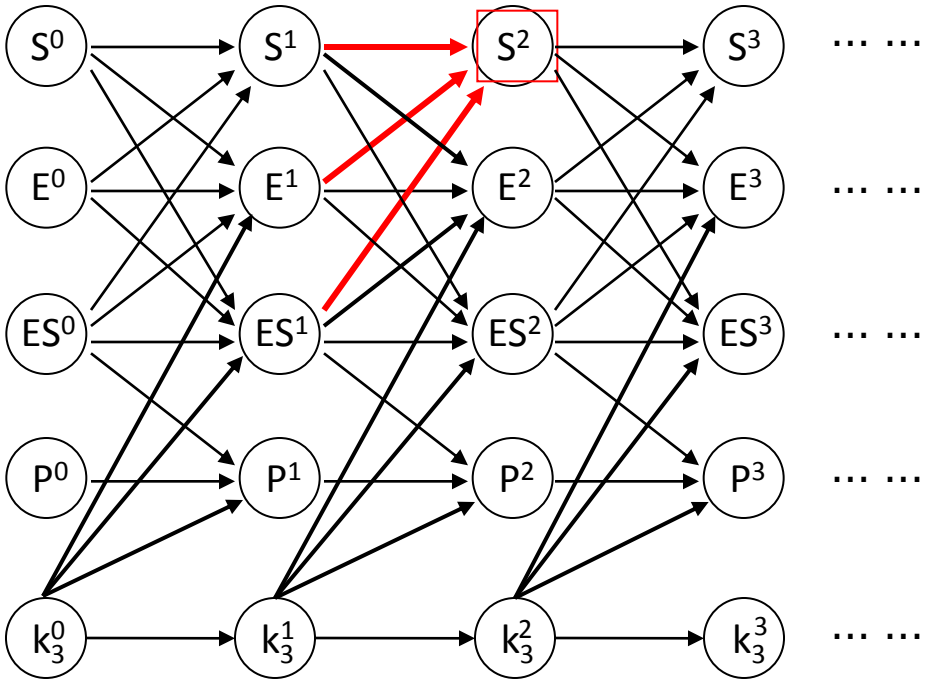
- Use Bayesian inferencing to do parameter estimation, sensitivity analysis, probabilistic model checking …
- Exact inferencing is not feasible for large models.
- We do approximate inferencing.
  - ***Factored Frontier algorithm***.

Parameter Estimation

1. For each choice of (interval) values for unknown parameters, run FF, compare with experimental data and assign a score using FF.

2. Return parameter estimates as maximal likelihoods.

3. FF can be then used on the calibrated model to do sensitivity analysis, probabilistic verification etc.

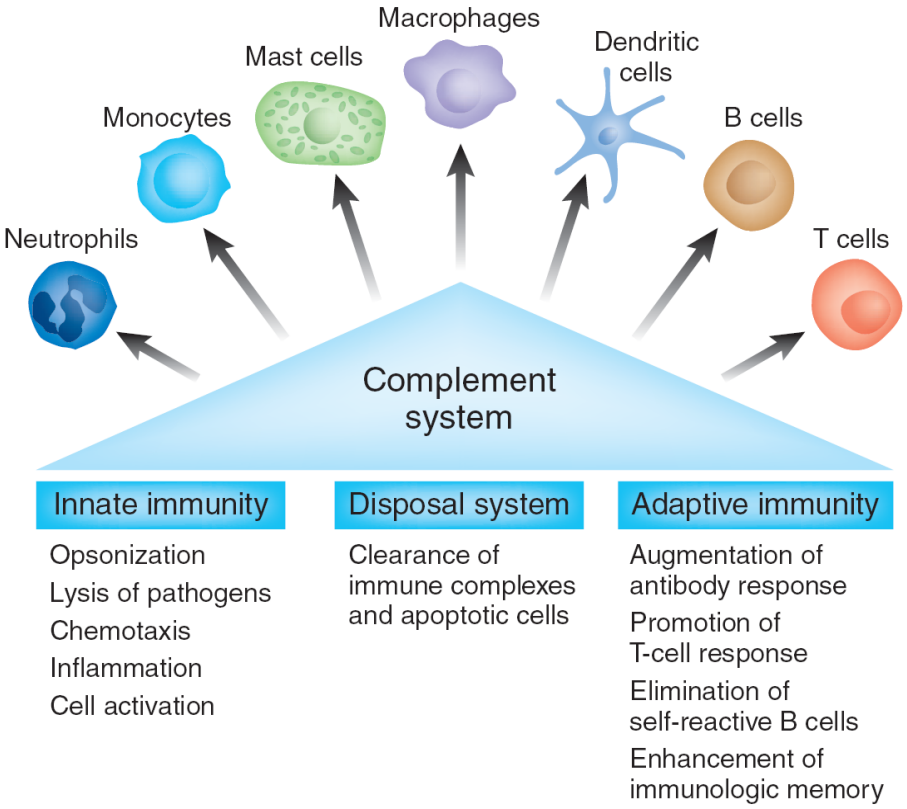


DBN based Analysis

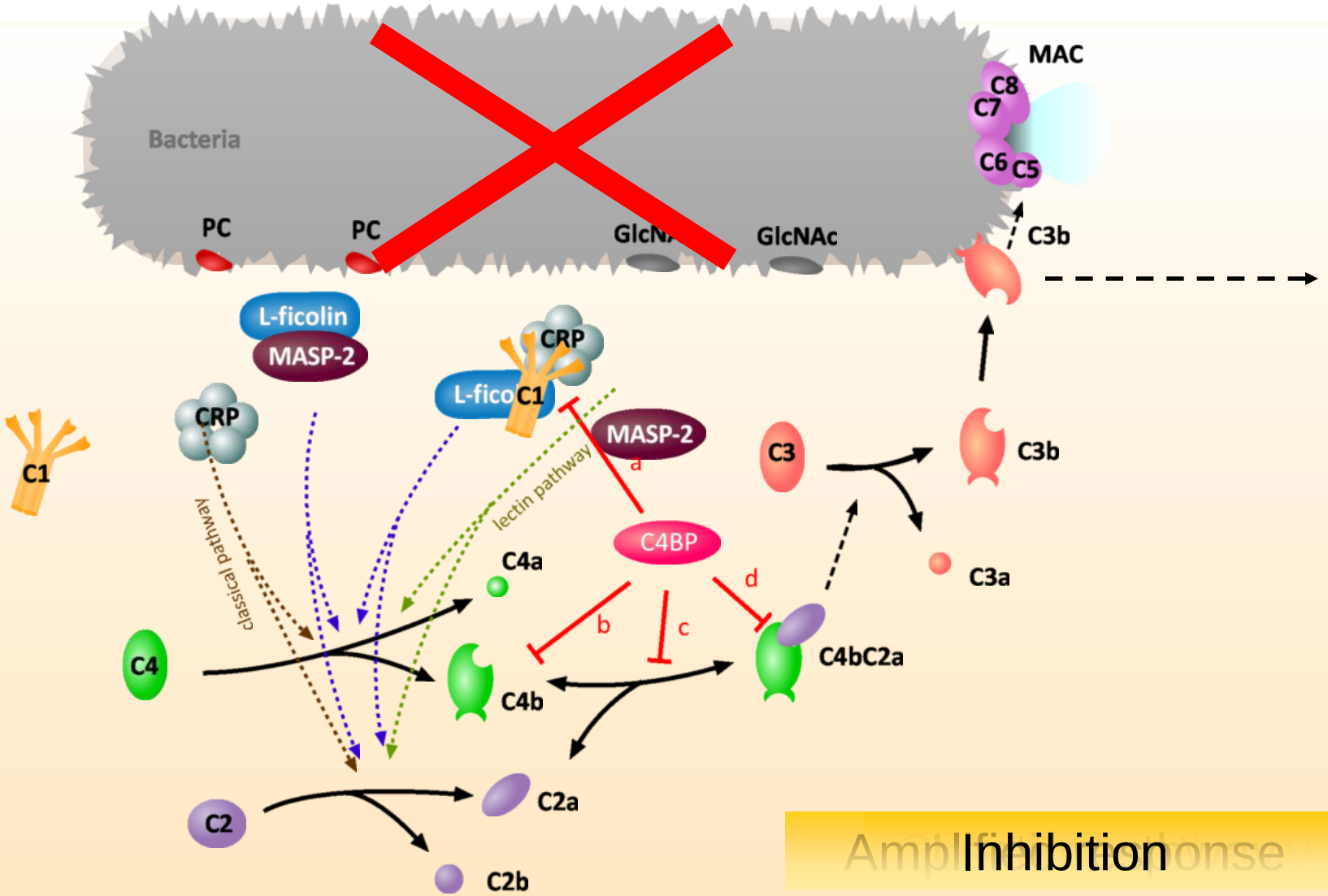
- Our experiments with signaling pathways models (taken from the ***BioModels data base***) show:
  - The one-time cost of constructing the DBN can be easily amortized by using it to do parameter estimation and sensitivity analysis.
  - Good compromise between efficiency and accuracy.

Complement System

- Complement system is a critical part of the immune system



Ricklin et al. 2007

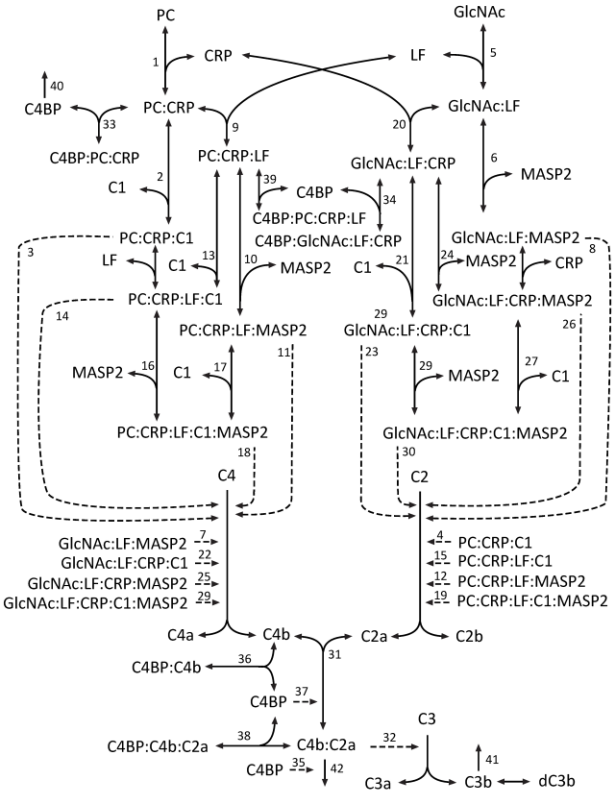

Bacteria
PC
PC
GlcNAc
GlcNAc
MAC
C8
C7
C6
C5
C3b
L-ficolin
MASP-2
CRP
L-fico
C1
CRP
C1
MASP-2
a
lectin pathway
classical pathway
C3
C3b
C3a
C4BP
C4a
d
b
c
C4
C4b
C4bC2a
C2
C2a
C2b
Amplification
Inhibition

Goals

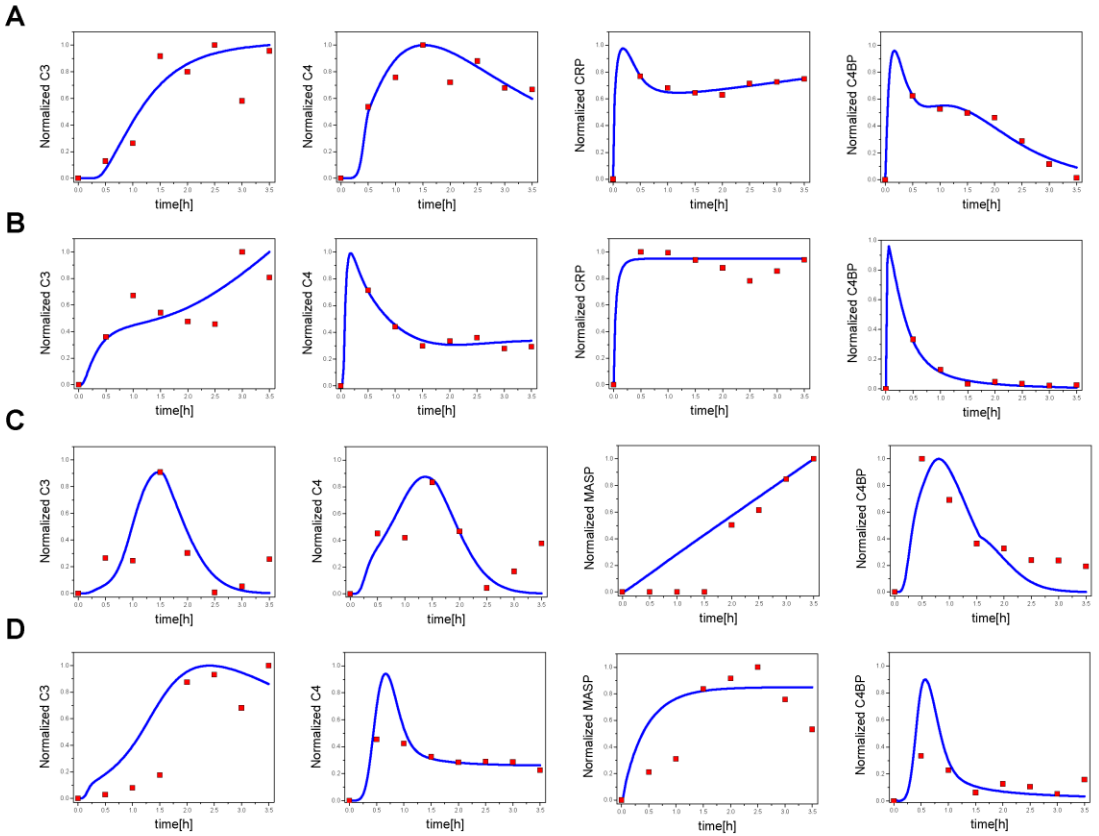
- Quantitatively understand the regulatory mechanisms of complement system
  - How is the excessive response of the complement avoided?
- The model:
  - Classical pathway + the lectin pathway
  - Inhibitory mechanism
    - ✓ C4BP

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
    - ✓ 2.4 x 10^6 samples
  - Runtime
    - ✓ 12 hours on a cluster of 20 PCs
- Model Calibration:
  - Training data: 4 proteins, 7 time points, 4 experimental conditions
  - Test data: *Zhang et al, PLoS Pathogens, 2009*

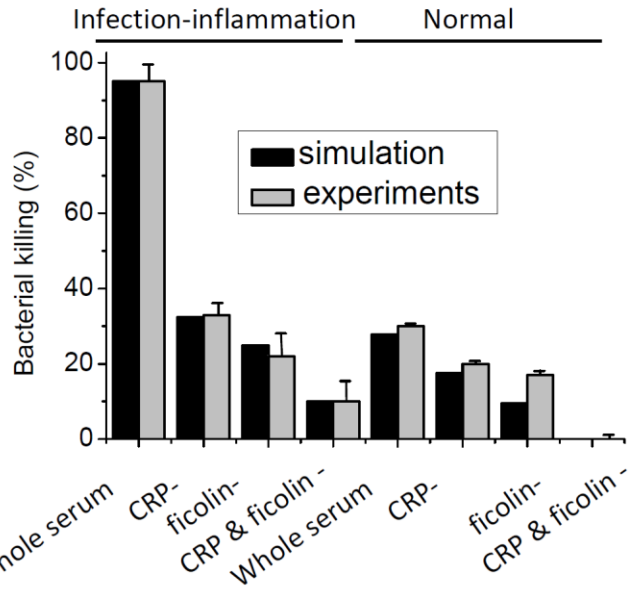


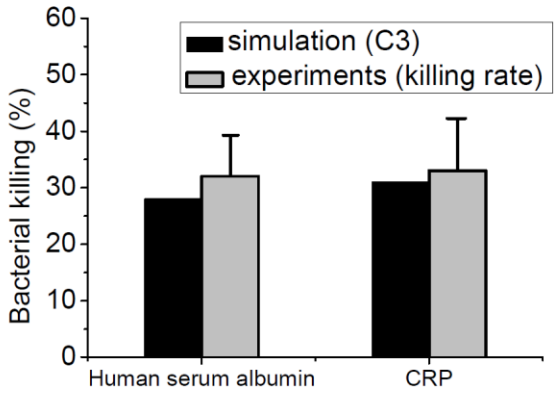
Model Calibration (parameter estimation)



Model validation

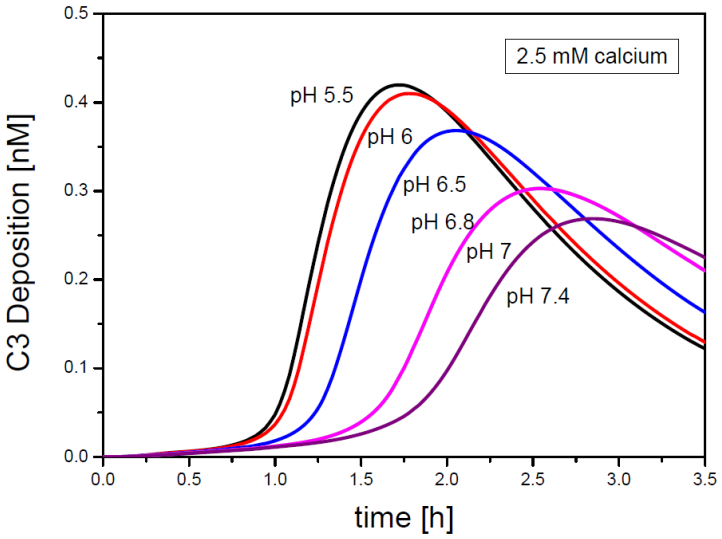
- Validated the model using previous published data (*Zhang et al 2009)*

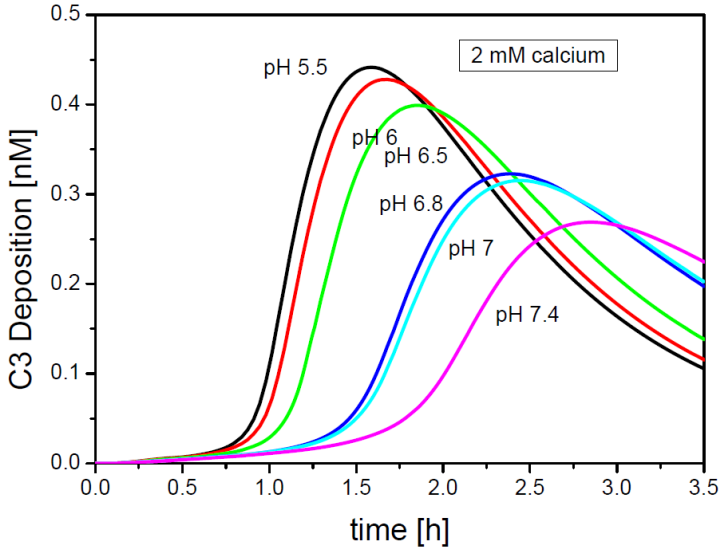




Enhancement mechanism

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



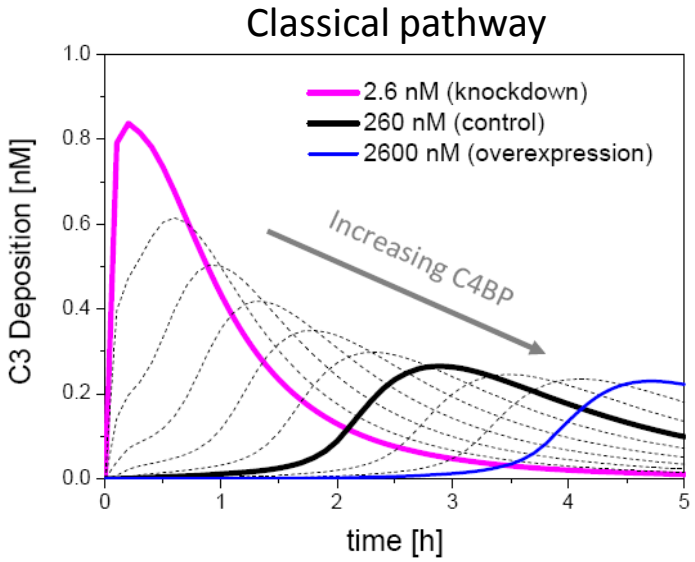


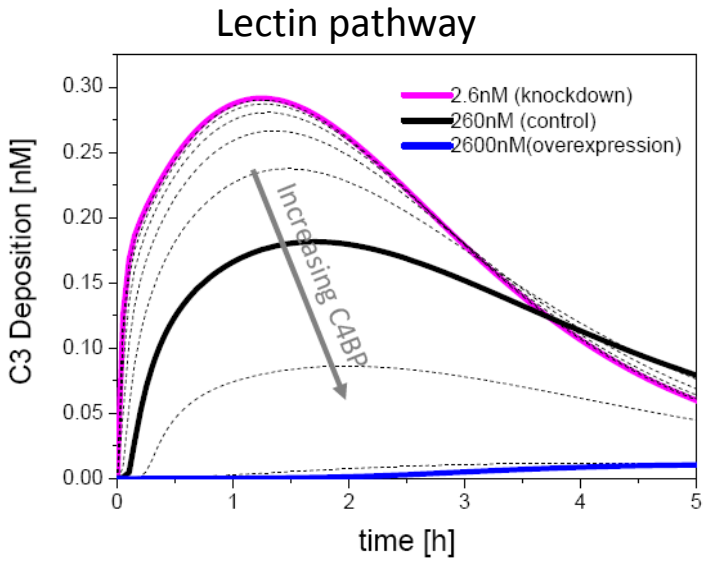
Analysis.

- (Local and global) sensitivity analysis.
- *in silico* experiments.

Model predictions: The regulatory effect of C4BP

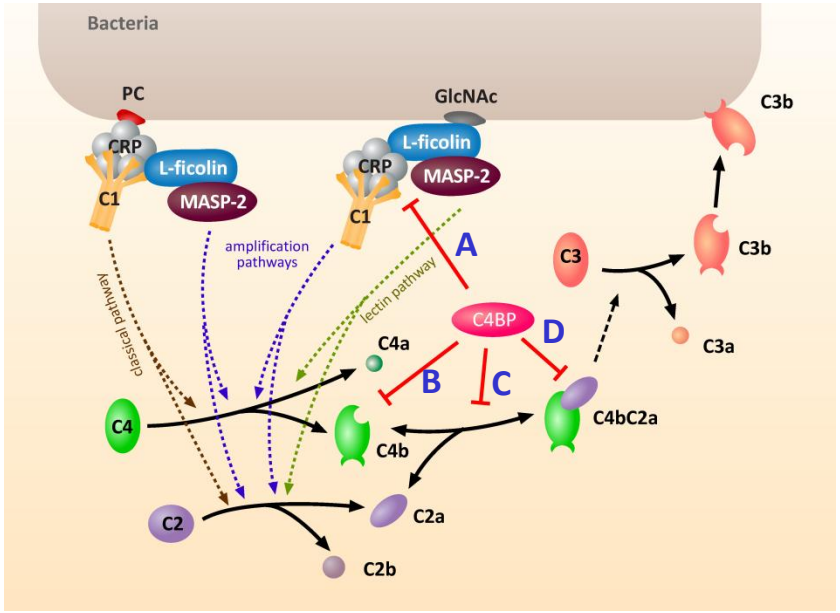
- C4BP maintains classical complement activation but delays the maximal response time
- But attenuates the lectin pathway activation

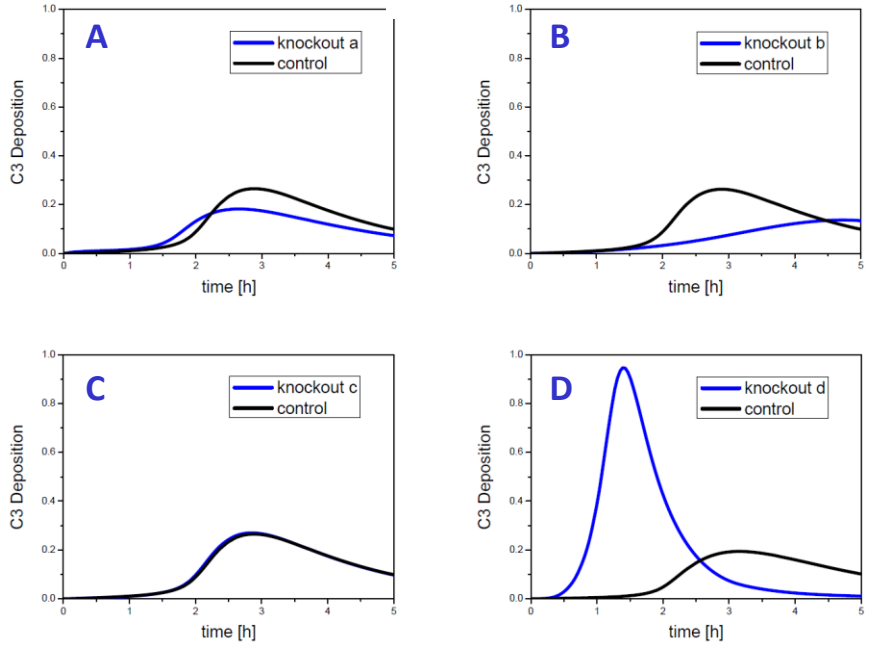




The regulatory mechanism of C4BP

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





Results

- Both predictions concerning C4BP were experimentally verified.

[PLoS Comp.Biol (2011)] ***[***BioModels database (***303.Liu***)]

Some extensions

- Parametrized version of FF
  - Reduce errors by investing more computational time

  [CMSB'11, TCBB 2012]
- GPU implementation:
  - Significant increase in performance and scalability
  - Thrombin-dependent MLC p-pathway
    - **105 ODEs**; 197 rate constants ; **164 "unknown" rate constants.**
    - (FF based approximate) probabilistic verification method *[Bioinformatics 2012]*

Current Collaborations

Ding Jeak Ling

Immune system signaling during Multiple infections

Marie-Veronique Clement

DNA damage/response pathways

G V Shivashankar

Chromosome co-localizations and co-regulations

Conclusion

- The DBN approximation method is useful and efficient.
- When does it (not) work?
- How to relate ODEs based dynamical properties to the DBN based ones?
- How to extend the approximation method to multi-mode signaling pathways?

Acknowledgements:

Liu Bing

Suchee Palaniappan

Benjamin Gyori

Gireedhar Venkatachalam

Wang Junjie

Blaise Genest

David Hsu