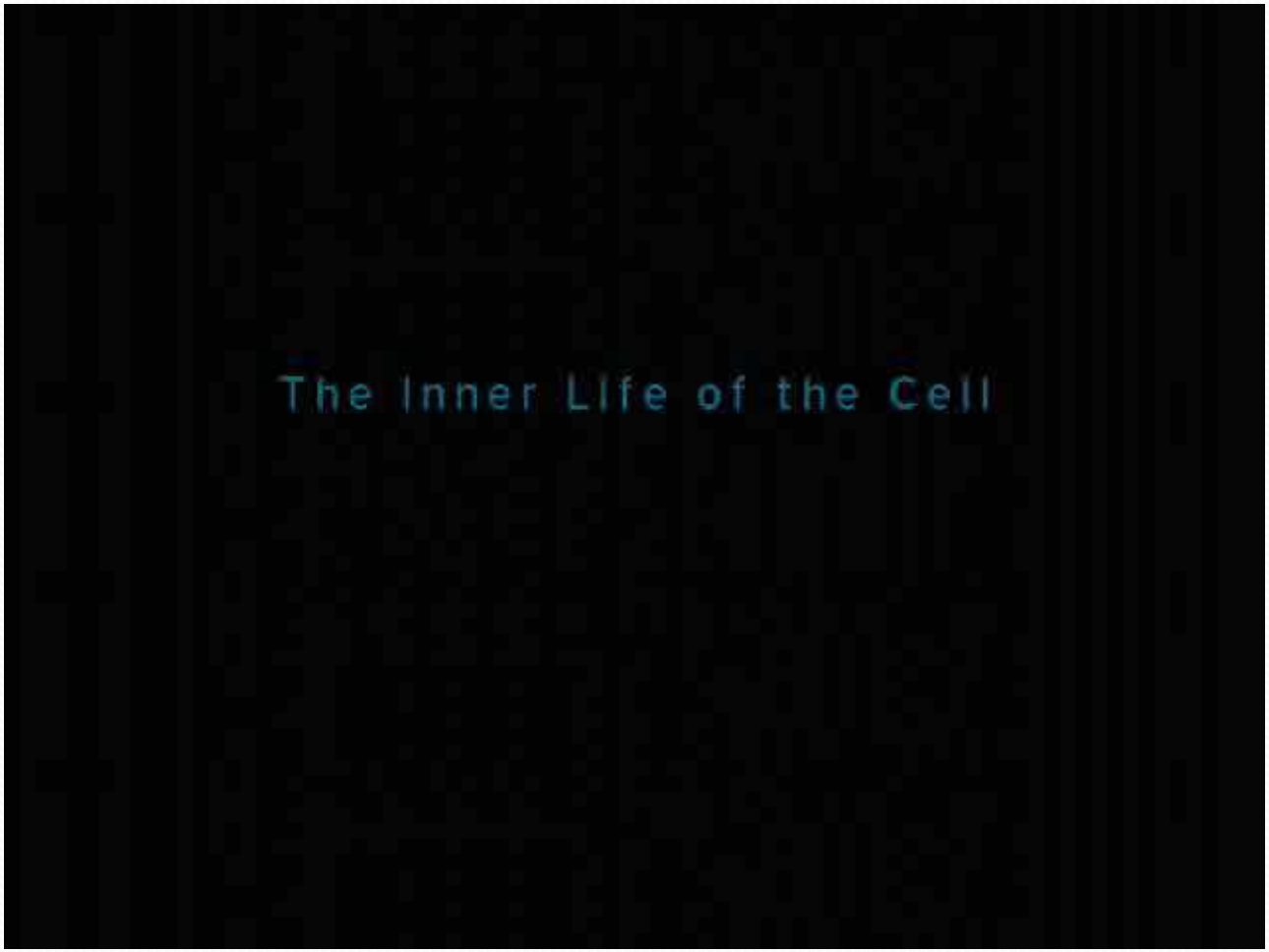
High Performance Computing Challenge 2009

Statistical Analysis of Partially Calibrated Bio-pathway Models

Liu Bing

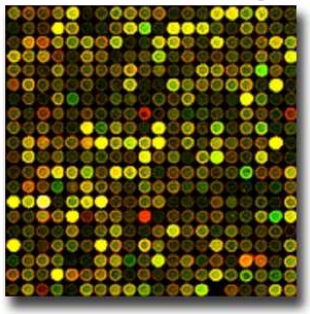
NUS Graduate School for Integrative Sciences and Engineering,
National University of Singapore



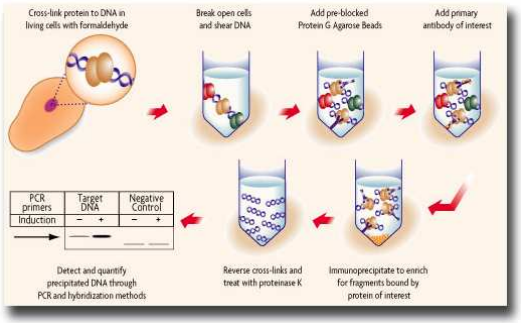
The Inner life of the Cell, Harvard University

Introduction

Microarray



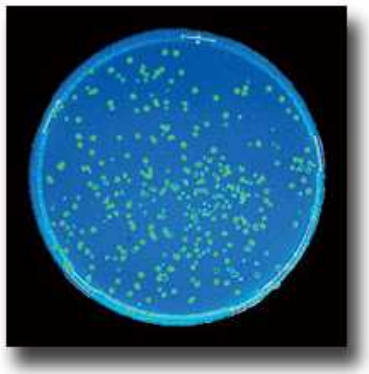
ChiP



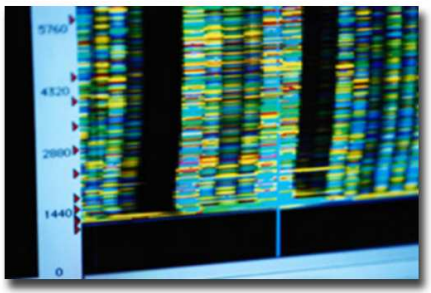
Western Blotting



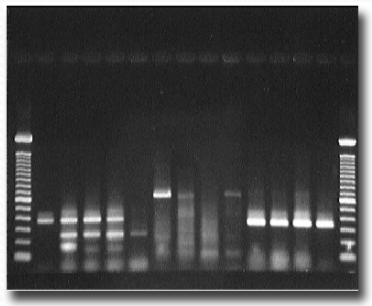
Two-Hybrid Assay



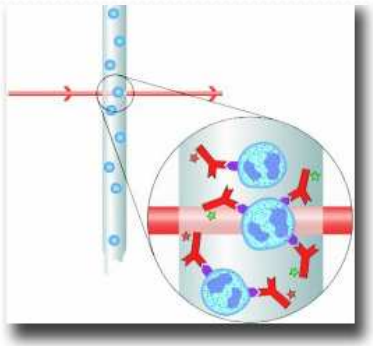
Gene Sequencing



PCR

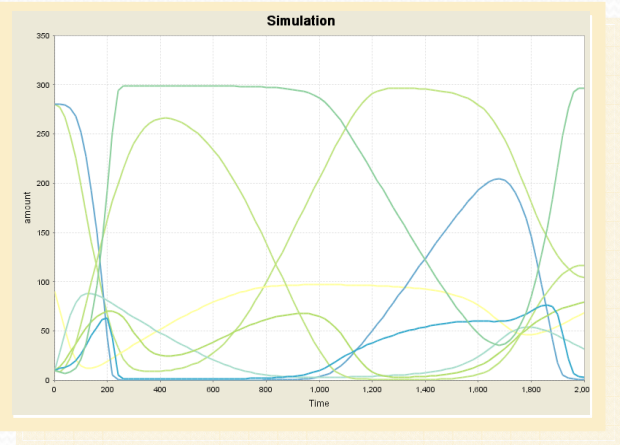


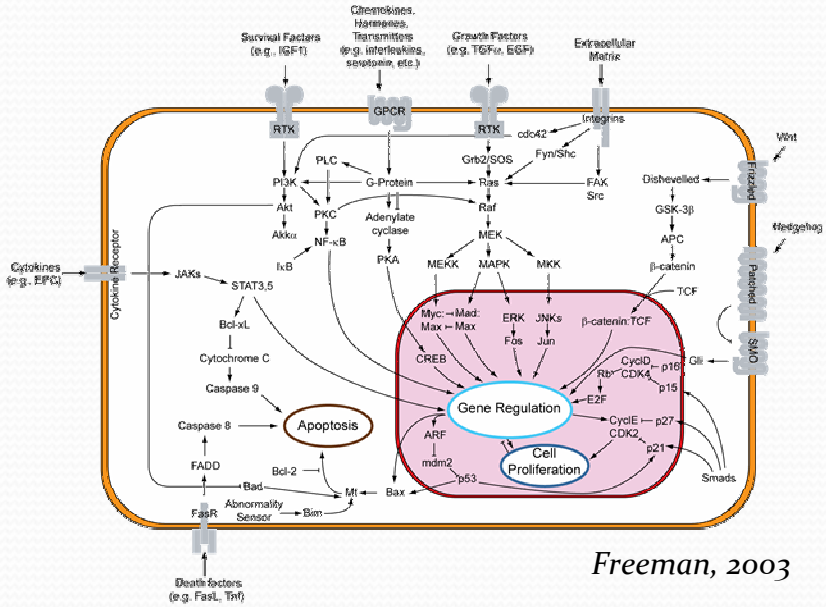
Flow cytometry



Introduction

- Previous research in biology focused on identifying:
  - Components
  - Interactions
- How about ***dynamics***?
  - Bistability, Oscillation, …

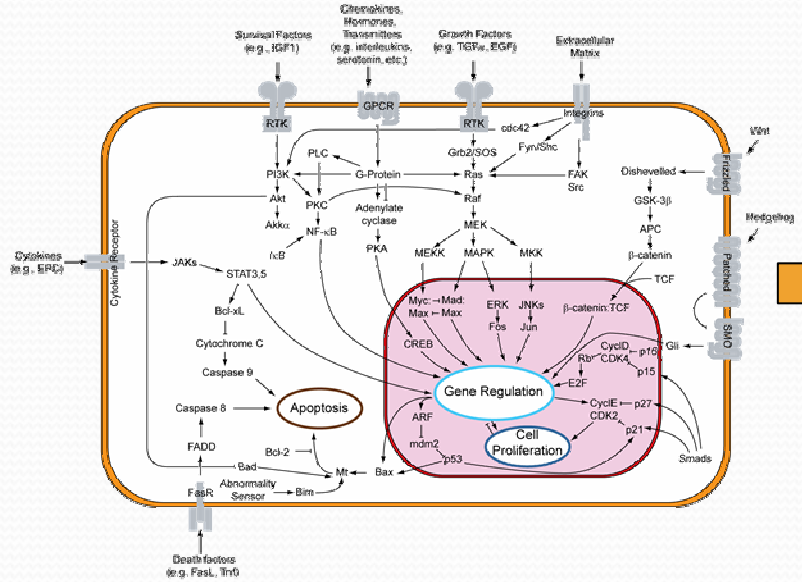


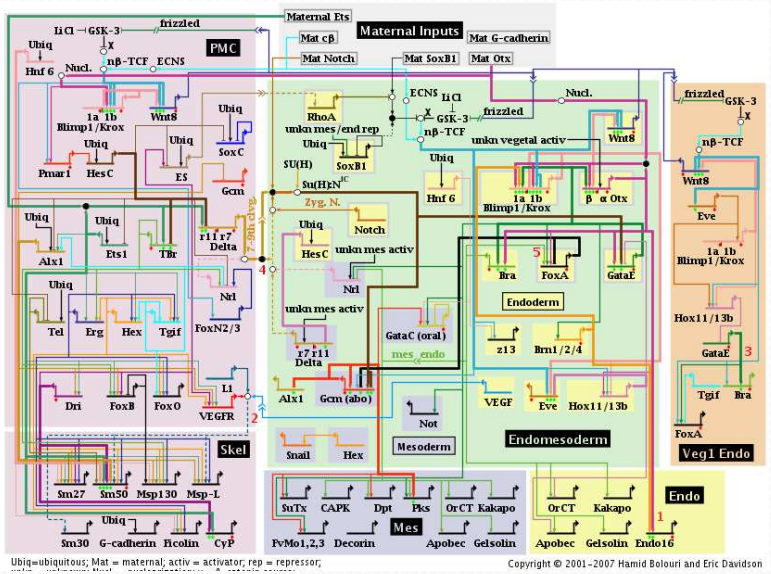


Freeman, 2003

Introduction

- The emerging of ***Computational Systems Biology***
- Long-term vision — ***Virtual Cell***

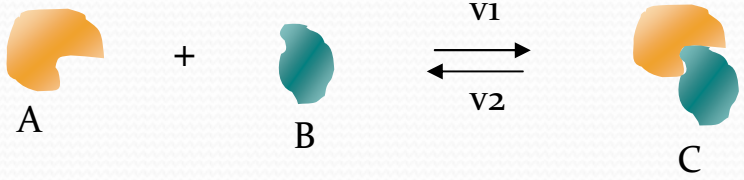




Longabaugh et al. 2008

Introduction

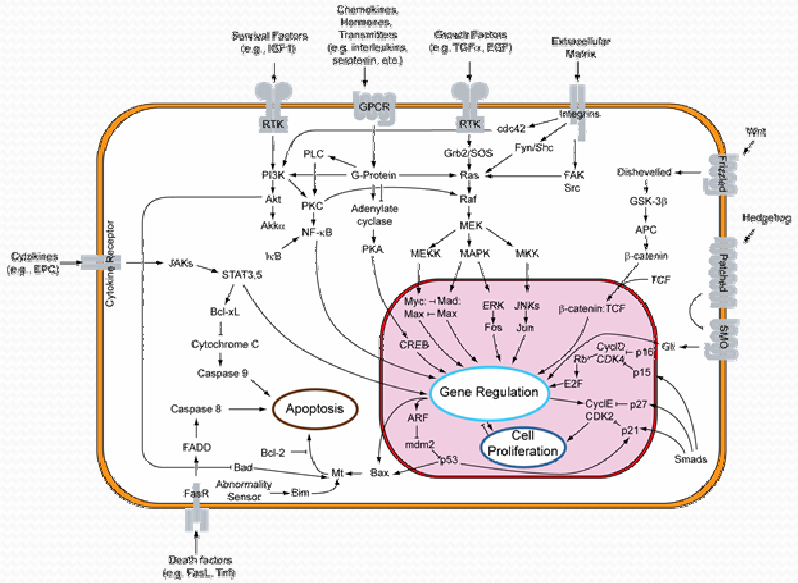
- How can we describe the signaling pathways dynamics?
- Modeling formulism
  - Hybrid automata, **Ordinary Differential Equations**…
  - Stochastic approaches
- ODE Example:



$$v_1 = k_1[A][B]$$

$$v_2 = k_2[C]$$

$$d[B]/dt = -v_1 + v_2 = k_1[A][B] - k_2[C]$$

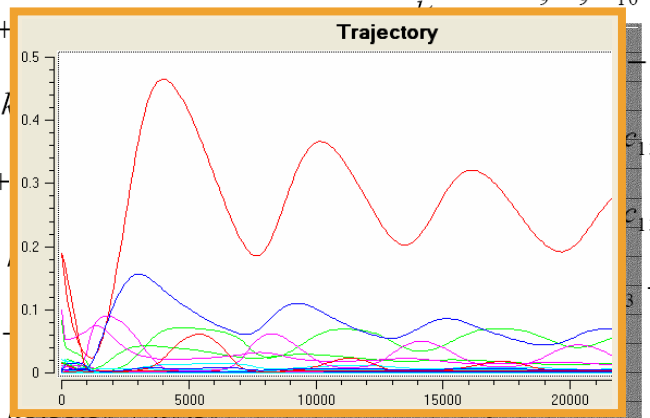


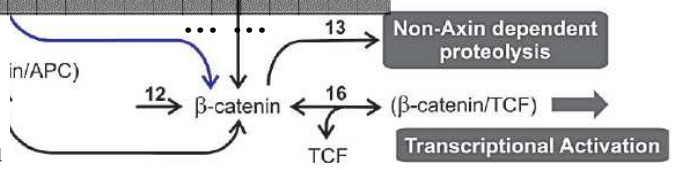
Introduction

- Modeling signaling pathways with ODEs

$$\frac{dc_1}{dt} = -k_1.c_1.c_2 + k_2.c_3$$
$$\frac{dc_2}{dt} = -k_1.c_1.c_2 + k_2.c_3 +$$
$$\frac{dc_3}{dt} = k_1.c_1.c_2 - k_2.c_3 - k$$
$$\frac{dc_4}{dt} = -k_3.c_3.c_4 + k_4 c_5 +$$
$$\frac{dc_5}{dt} = k_3.c_3.c_4 - k_4.c_5 -$$
$$\frac{dc_6}{dt} = -k_5.c_5.c_6 + k_6.c_7 -$$
$$\frac{dc_7}{dt} = k_5.c_5.c_6 - k_6.c_7 -$$
$$\frac{dc_8}{dt} = -k_7.c_7.c_8 + k_8.c_9 + k_{11}.c_{11} + k_{20}.c_{21}$$
$$\frac{dc_9}{dt} = k_7.c_7.c_8 - k_8.c_9 - k_9.c_9.c_{10} + k_{10}.c_{11}$$
$$-k_{15}.c_9.c_{17} + k_{16}.c_{18}$$

$$\frac{dc_{10}}{dt} = -k_9.c_9.c_{10} + k_{10}.c_{11} + k_{14}.c_{14}$$
$$- k_{10}.c_{11} - k_{11}.c_{11}$$
$$c_{13} + k_{13}.c_{14} + k_{11}.c_{11}$$
$$c_{13} + k_{13}.c_{14} + k_{29}.c_{16}.c_{31}$$
$$- k_{13}.c_{14} - k_{14}.c_{14}$$



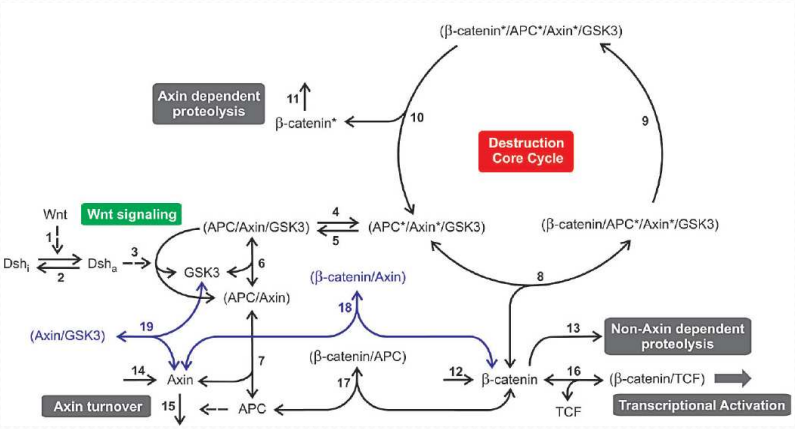


What can we do next

- Enrich our understanding
  - Why is the dynamics like that?
  - How is the dynamics affected by mutations?
  - Which species plays a critical role in shaping signaling dynamics?
  - What is the role of a crosstalk?
  - … …
- Many applications
  - Setting experiments
  - Simulation of clinical responses
  - Optimization of therapy
  - … …

Hurdle

- Many unknown parameters (rate constants)



$$\frac{dc_{10}}{dt} = -k_9.c_9.c_{10} + k_{10}.c_{11} + k_{14}.c_{14}$$

$$\frac{dc_{11}}{dt} = k_9.c_9.c_{10} - k_{10}.c_{11} - k_{11}.c_{11}$$

$$\frac{dc_{12}}{dt} = -k_{12}.c_{12}.c_{13} + k_{13}.c_{14} + k_{11}.c_{11}$$

$$\frac{dc_{13}}{dt} = -k_{12}.c_{12}.c_{13} + k_{13}.c_{14} + k_{29}.c_{16}.c_{31}$$

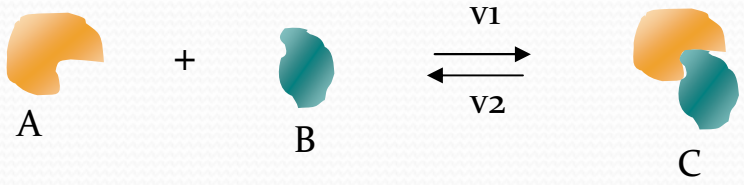
$$\frac{dc_{14}}{dt} = k_{12}.c_{12}.c_{13} - k_{13}.c_{14} - k_{14}.c_{14}$$

$$\frac{dc_{15}}{dt} = k_{14}.c_{14}$$

… … … …

Hurdle

- Many unknown parameters (rate constants)
- Parameter estimation
  - Guess → Evaluate → Guess → Evaluate …

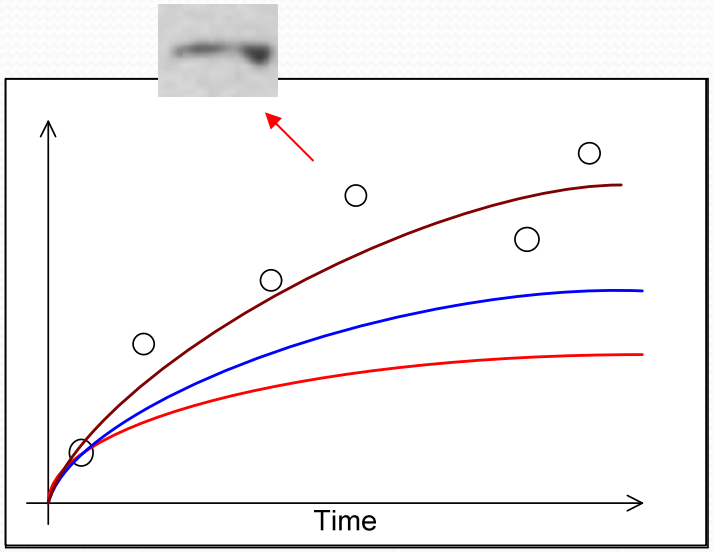


$$v_1 = k_1[A][B]$$

$$v_2 = k_2[C]$$

$$d[B]/dt = -v_1 + v_2 = k_1[A][B] - k_2[C]$$

?? ??

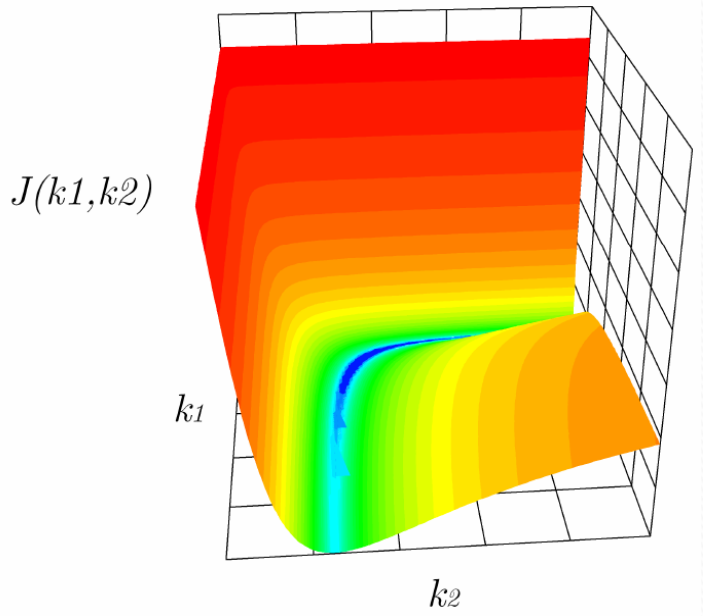


Parameter Estimation

- Optimization problem
- Methods
  - Levenberg-Marquardt
  - Stimulated Annealing
  - Evolution Strategy
  - Particale Swarm Optimization
  - …

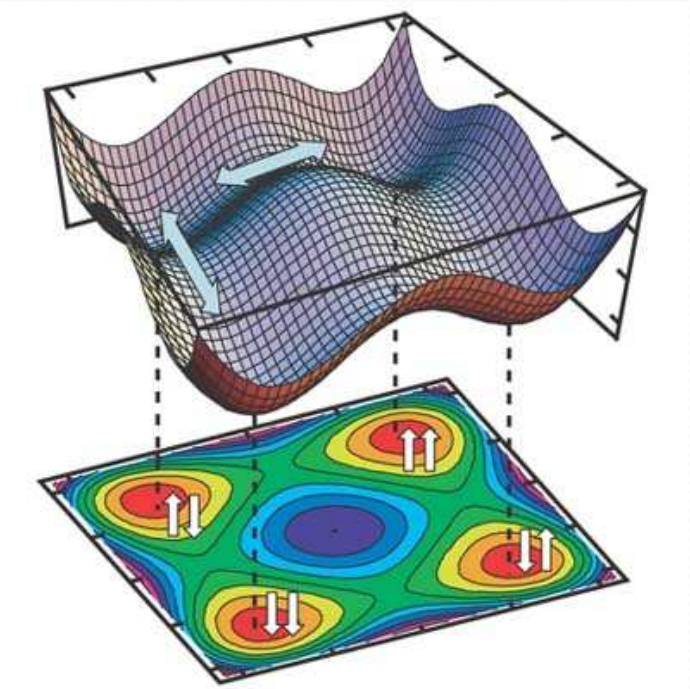
Given data $\tilde{\mathbf{x}}(t_j)$, find $\mathbf{k}$ to

minimize $J(\mathbf{k}) = \sum_j ||\mathbf{x}(t_j;\mathbf{k}) - \tilde{\mathbf{x}}(t_j)||^2$



Difficulties

- High dimension system
  - Many local minima
- Limited and noisy data
  - Insufficient to constrain all parameters to have a unique solution



Kim et al. 2007

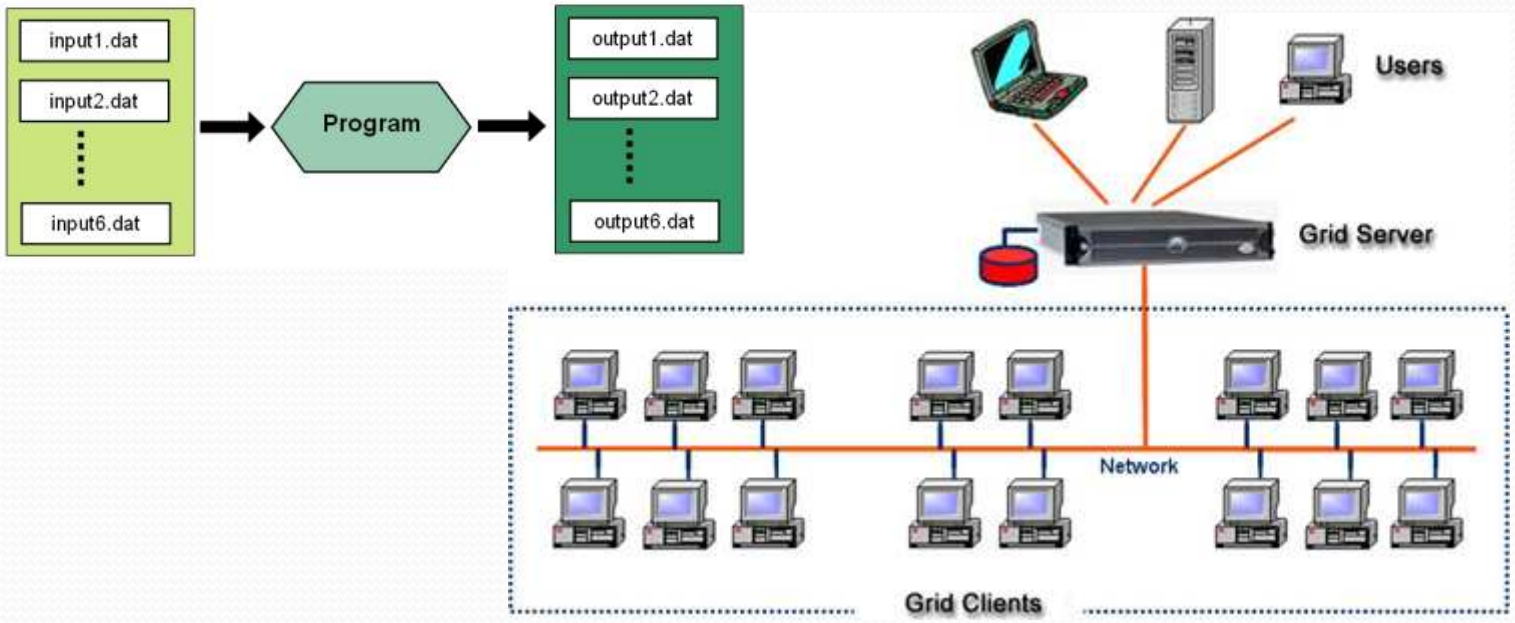
What we want to do

- Perform parameter estimation many times
- Statistically analyze the estimated parameters

- Challenge: ***it require a huge amount of time!***

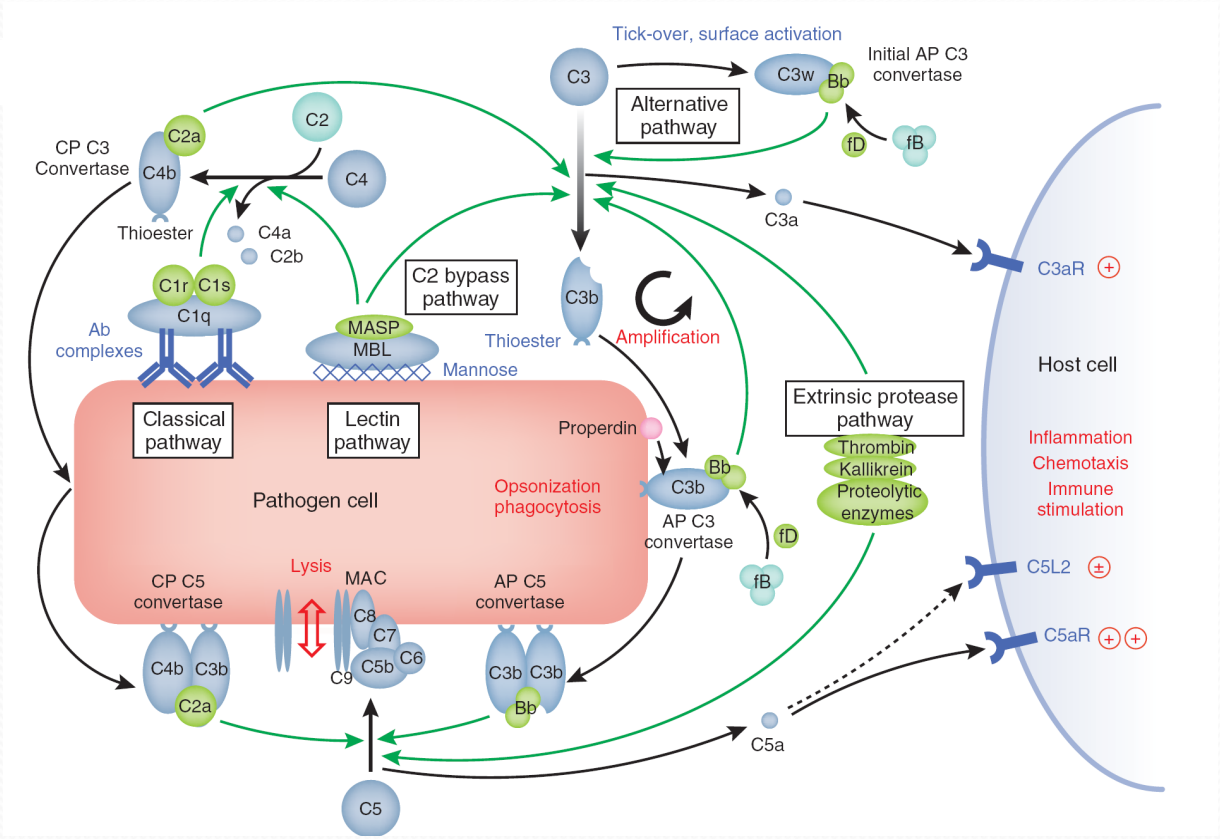
Grid Computing

- Tera-Scale Campus Grid at NUS (TSC@NUS)
- Parametric Study Program



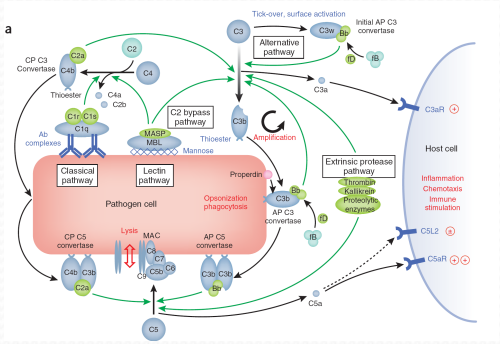
Case study

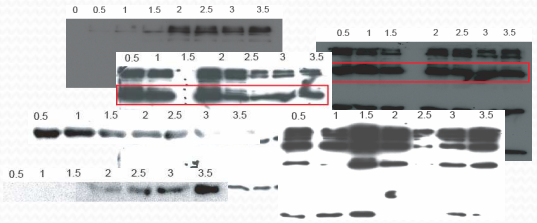
- Complement system

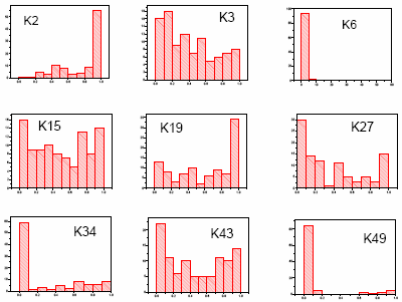


Ricklin et al. 2007

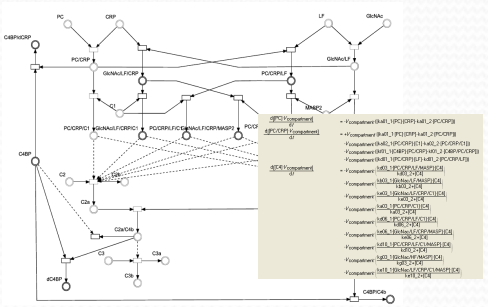
Modeling complement system*



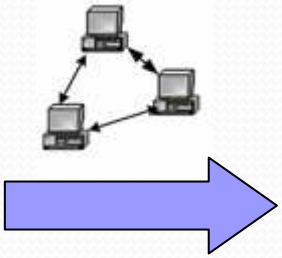




ODE Model



Parameter estimation on TSC@NUS



Statistical analysis

Para Parameter set

Parameter set

ter set

* This is an on-going joint project with Prof. Ding Jeak Ling's group from DBS, NUS

Experiments

- 42 differential equations
- 89 unknown parameters
- Parameter estimation method
  - Stochastic ranking evolutionary strategy (SRES)
- 3 experiments:

Job ID	μ	λ	Gen	No. of Samples
9944	30	200	6000	99
10055	40	300	10000	99
10282	30	200	500	2000

Table 1. Experiments

Results

Job ID	No. of Workunits	Running Time
9944	1	481 min
10055	1	1208 min
10282	1	42 min

Table 2. Performance on single PC

Job ID	No. of Workunits	No. of Nodes	Total CPU Time	Running Time	Speedup
9944	99	96	44928 min	729 min	61.6
10055	99	100	108780 min	3245 min	33.5
10282	2000	91	67815 min	826 min	82.1

Table 3. Performance on the Grid

Conclusion

- By enabling our application on TCG@NUS, we were able to achieve up to ***82.1 speedup***.
- Currently, we are investigating more properties of Complement system using the parameters estimated.

Acknowledgements

- Computer Center staffs:
  - Mr. Wang Junhong
  - Ms. Grace Foo
- Supervisors and Collaborator:
  - Prof Thiagarajan
  - A/P David Hsu
  - Prof Ding Jeak Ling