## An algebraic-numeric algorithm for the model selection in kinetic networks

## CASC2007, Bonn

Hiroshi Yoshida (Faculty of Math., Kyushu Univ.) Koji Nakagawa (CBRC, AIST) Hirokazu Anai (Fujitsu Lab. LTD./CREST JST.) Katsuhisa Horimoto (CBRC, AIST)

## Contents

- Introduction of kinetic network
- Model and Method
  - (1) Laplace transformation of model formulae and observed data
  - (2) Matching
  - (3) Model consistency estimation
- Result
- Summary
- Announcement of AB2008

## Background

- Metacore: database of networks describing interactions between genes or proteins
- D-sep test: can deal with only directed **acyclic** graph (DAG)



# Aim: selection for the most/more consistent model with the observed data



## More...

- To select the model most/more consistent with the given sampling data
- We have performed model selection over Laplace domain using algebraic equations





$$\frac{d}{dt}SLA(t) = -kSN SLA(t)$$

$$\frac{d}{dt}N(t) = kSN SLA(t) - kNA N(t)$$

$$\frac{d}{dt}A(t) = kNA N(t) - kAG A(t)$$

$$\frac{d}{dt}G(t) = kAG A(t)$$

Assuming a linear relation between the variables, the kinetics of the above network can be described as the system of deferential equations on the left.





## Method

- The kinetics for describing biological phenomena are expressed by a system of differential equations
- 2 The observed data are numerically fitted as a sum of exponentials
- ③ Both the system of differential equations and the sum of exponentials are transformed into the corresponding system of algebraic equations by Laplace transformation
- The two systems of algebraic equations are compared according to *measure*

## Consistency estimation (1)



 $\begin{aligned} \text{Model Consistency Estimation} \\ comp &= \{h(0) - \beta_0, k_1 - \alpha_0, f(0) - (\beta_1 + \beta_2), k_1(f(0) + h(0)) - (\alpha_1\beta_2 + \alpha_2\beta_1), (k_1 + k_2) - (\alpha_1 + \alpha_2), k_1k_2 - \alpha_1\alpha_2 \} \end{aligned}$ 

### Change the system of differential equations into algebraic equations by Laplace transformation

• Ex.) The system of algebraic equations

$$\frac{d}{dt}SLA(t) = -kSN SLA(t)$$
$$\frac{d}{dt}N(t) = kSN SLA(t) - kNA N(t)$$
$$\frac{d}{dt}A(t) = kNA N(t) - kAG A(t)$$
$$\frac{d}{dt}G(t) = kAG A(t)$$

o Into polynomials over Laplace domain

s L[SLA(t)] - SLA(0) = -kSN L[SLA(t)] s L[N(t)] - N(0) = kSN L[SLA(t)] - kNA L[N(t)] s L[A(t)] - A(0) = kNA L[N(t)] - kAG L[A(t)]s L[G(t)] - G(0) = kAG L[A(t)]

## The solution over Laplace domain is:

laplace(SLA(t), t, s) =  $\frac{SLA(0)}{s + kSN}$ 

laplace(A(t), t, s) =

 $\frac{A(0) s^{2} + (kNA N(0) + A(0) kSN + A(0) kNA) s + kNA N(0) kSN + kNA kSN SLA(0) + A(0) kNA kSN}{s^{3} + (kAG + kSN + kNA) s^{2} + (kAG kSN + kAG kNA + kNA kSN) s + kAG kNA kSN}{laplace(N(t), t, s)} = \frac{N(0) s + N(0) kSN + kSN SLA(0)}{s^{2} + (kSN + kNA) s + kNA kSN}$   $laplace(G(t), t, s) = (G(0) s^{3} + (G(0) kAG + G(0) kNA + kAG A(0) + G(0) kSN) s^{2} + (G(0) kAG kNA + kAG kNA N(0) + G(0) kAG kSN + kAG A(0) kNA + kAG A(0) kNA + kAG A(0) kNA kSN) s + kAG kNA N(0) kSN + kAG kNA kSN SLA(0) + G(0) kAG kNA kSN + kAG A(0) kNA kSN) s + kAG kNA N(0) kSN + kAG kNA kSN SLA(0) + G(0) kAG kNA kSN + kAG A(0) kNA kSN) / (s^{4} + (kAG + kSN + kNA) s^{3} + (kAG kSN + kAG kNA + kNA kSN) s^{2} + s kAG kNA kSN)$ 

## Consistency estimation (2)



 $\begin{aligned} \text{Model Consistency Estimation} \\ comp &= \{h(0) - \beta_0, k_1 - \alpha_0, f(0) - (\beta_1 + \beta_2), k_1(f(0) + h(0)) - (\alpha_1\beta_2 + \alpha_2\beta_1), (k_1 + k_2) - (\alpha_1 + \alpha_2), k_1k_2 - \alpha_1\alpha_2 \} \end{aligned}$ 

## Change the observed data into algebraic equations by Laplace transformation

 The observed data are fitted as a sum of exponentials:

$$\sum_{i=1}^{k} \beta_i \exp(-\alpha_i t)$$

- k is the number of exponentials which can theoretically be determined
- Into functions in s over Laplace domain:  $\sum_{i=1}^{k} \frac{\beta_i}{s + \alpha_i}$

### Previous study: Identifiability problem

**Cobelli, C**., Foster, D. and Toolo, G.: *Tracer Kinetics in Biomedical research: From data to model*, Kluwer Academic/Plenum Publishers, 2000.

$$\frac{dy}{dt} = -k2 \cdot y \qquad \qquad \frac{dy}{dt} = -(k2 + k3)y$$

Both data are fitted as:  $a \exp(-m t)$  k2 = m, uniquely determined k2+k3=m .... unidentifiable globally identifiable

On the assumption of error-free data, ... unrealistic situation

## Our model: we handle noisy data

 We take the position that it is sufficient to fit a noisy observed data as a sum of exponentials:

 $a1e^{-m1t} + a2e^{-m2t} + a3e^{-m3t} + \cdots$ 

# Example: the observed data into Laplace domain



## Consistency estimation procedure (3)



## Comparison of coefficient List in s

 Derive the coefficient List by comparing the algebraic equations of the model and the observed data over Laplace domain:

 $laplace(SLA(t), t, s) = \frac{SLA(0)}{s + KN}$   $laplace(A(t), t, s) = \frac{N(0) s^{2} + (MA N(0) + A(0) KN + A(0) MA) s + MA N(0) KN + MA KN SILA(0) + A(0) MA KN }{s^{3} + (AAG + KN + MA) s^{2} + (AAG KN + AAG MA + MA KN) s + AAG MA KN }{laplace(N(t), t, s) = \frac{N(0) s + N(0) KN + KN SILA(0)}{s^{2} + (KN + MA) s + MA KN }}$   $laplace(G(t), t, s) = (0, 0) s^{3} + (G(0) AAG + G(0) MA + AAG A(0) + G(0) KN) s^{2} + (AAG MA + AAG MA KN) + G(0) AAG KN + AAG A(0) MA + KAG A(0) MA + KAG A(0) MA KN }{s^{4} + (AAG MA N(0) KN) + G^{0} KN + AAG A(0) MA + KAG A(0) MA KN }{s^{4} + (KAG + KN + MA) s^{3} + (KAG KN + KAG MA + MA KN) s^{2} + s AAG MA KN }$   $laplace(CA(t), t, s) = \frac{0.988756614 s + 17.03928819}{s^{2} + 1.102188978 s + 0.1001897313}$   $laplace(CA(t), t, s) = \frac{0.9883934099 s^{3} + 3.059590433 s^{2} + 2.584988429 s + 1.0290575777}{s^{4} + 1.565249130 s^{3} + 0.6365564404 s^{2} + 0.04900672064 s$ 

## **Ex.)** Coefficient List

 Without any error, these polynomials would be zero, but in the case of real noisy observed data, unfortunately not zero... =>

#### Least squares method (LSM)

*kAG kNA kSN*= 0.04900672064

*kSN*=1.00201179553207664

16.98261012 *kSI*V= 17.03928819

20.96775977 kAG kNA kSN= 1.029057577

10.96390626 kAG kNA+3.985149653 kAG kSN+0.99839341 kNA kSN=2.58498843

kAG kSN+ kAG kNA+ kNA kSN= 0.6875282094

*k***A***G* + *k***S***N* + *k***N***A* = 1.704671278

kAG kSN+ kAG kNA+ kNA kSN= 0.6365546405

kAG+ kSN+ kNA = 1.565249130

3.985149653 kAG+0.99839341 kNA+0.99839341 kSN=3.059590448

kNA kSN= 0.1001897313

*kSI*V+ *k*IVA = 1.102188978

kAG kNA kSN= 0.05302211591

9.965512853 *k***N***A*+2.986756243 *kS***N**=4.267391383

19.96936636 kNA kSN= 2.117762332

## Consistency measure:

 The *smallest* sum-square value of the elements in Coefficient List under

*k1 > 0, k2 > 0, ...., kn> 0* (1) Or

k1 >=0, k2 >=0, ..., kn >=0 (2) If e.g. k1 = 0 ... subnetwork



### Concrete procedure

• Coefficient List:

$$l_1(\vec{k}) = r_1, \cdots, l_n(\vec{k}) = r_n$$

The consistency measure can be calculated as the smallest values of *f(k)* among various *ks* : => Least squares method:

$$f(\vec{k}) = \sum_{i=1}^{n} (l_i(\vec{k}) - r_i)^2$$

## Measures (1) and (2)

• Under ki > 0 ... Measure (1)  $\frac{\partial f}{\partial k_1} = 0 \wedge \cdots \wedge \frac{\partial f}{\partial k_n} = 0$ 

• Under *ki >= 0* ... Measure (2)

$$\frac{\partial f}{\partial k_1} = 0 \wedge \dots \wedge \frac{\partial f}{\partial k_n} = 0$$

Including subnetworks

# Algorithm to compute the measure (2) (Recursive procedure)

MinimizePositive.. Compute the measure (1) MinimizePositive(f): INPUT: Polynomial f with variables  $k_1, \dots, k_n$ OUTPUT: Tuple  $(k_1, \dots, k_n)$  satisfying  $k_1 \ge 0, \dots, k_n \ge 0$  which minimizes f BEGIN if n = 0 then Return({})  $M \leftarrow \{(k_1, \cdots, k_n) \mid k_1 \ge 0 \land \cdots \land k_n \ge 0 \land \partial f / \partial k_1 = 0 \land \cdots \land \partial f / \partial k_n = 0\}$  $m_1 \leftarrow \text{MinimizePositive}(f(0, \cdots, k_n))$  $m_2 \leftarrow \text{MinimizePositive}(f(k_1, 0, \cdots, k_n))$ *MinimizePositive(f(k1, ..., 0, .... kn))* . . .  $m_n \leftarrow \text{MinimizePositive}(f(k_1, \cdots, 0))$  $mres \leftarrow (j_1, \cdots, j_n)$ s.t.  $f(k_1, \dots, k_n) \ge f(j_1, \dots, j_n)$  for all  $(k_1, \dots, k_n) \in M \cup m_1 \cup \dots \cup m_n$ 

Return({mres}) END

## Result

#### We used the following five models

Model A:





Model C:

Model D:









## Data generation for simulation

We have generated the time series of data for the consistent molecules for the simulation study, before the model consistency estimation.



The given and estimated parameters

are as follows: S LA; 1, 1 (given) and 1.00 (estimated); S LA; 1, 10 and 10.0; N; 1, 1/10 and 0.100; N; 2, 1 and 1.00; N; 1, 163/9 and 18.1; N; 2, 100/9 and 11.1; A; 1, 1/10 and 0.100; A; 2, 1/2 and 0.500; A; 3, 1 and 1.00; A; 1, 163/36 and 4.53; A; 2,  $\Box$ 15=4 and  $\Box$ 3:75; A; 3, 20/9 and 2.22; G; 1, 1/10 and 0.100; G; 2, 1/2 and 0.500; G; 3, 1 and 1.00; G; 1,  $\Box$ 815=36 and  $\Box$ 22:6; G; 2, 15/4 and 3.75; G; 3,  $\Box$ 10=9 and  $\Box$ 1:11; G; 4, 21 and 21.0. Each figure corresponds to the four variables (molecules) in the model: (a) S LA, (b) N, (c) A, (d) G.

#### Table of Consistency measure

Measure (1)

Measure (2).. Including subnetworks

query	CM1, under the constraint (2.4)									CM2, under the constraint (2.6)							
	estimated	smallest	$k_{SN}$	$k_{NA}$	$k_{AG}$	$k_{NG}$	$k_{SA}$	$k_{SS}$	estimated	smallest	k <sub>sn</sub>	k <sub>NA</sub>	$k_{AG}$	$k_{NG}$	$k_{SA}$	k <sub>SS</sub>	
А	A	1.34×10 <sup>-11</sup>	1.00	0.100	0.500	-			С	7.66×10 <sup>-12</sup>	1.00	0.100	0.500	0*	0*	-	
	E	1.36×10 <sup>-11</sup>	1.00	0.100	0.500	-	-	$1.40 \times 10^{-6}$	Α	1.34×10 <sup>-11</sup>	1.00	0.100	0.500	-	-	-	
	D	×	×	-	×	-	×	-	Е	1.36×10 <sup>-11</sup>	1.00	0.100	0.500	-	-	$1.40 \times 10^{-6}$	
	В	×	×	×	×	-	×	-	В	$1.35 \times 10^{-10}$	1.00	0.100	0.500	-	0*	-	
		×	×	×	×	×	×	-	D	1.68	0*	-	0.435	-	0.0439	-	
В	В	4.20×10 <sup>-11</sup>	1.00	0.100	0.500	-	0.400	-	В	4.20×10 <sup>-11</sup>	1.00	0.100	0.500	-	0.400	-	
	Α	20.1	1.19	0.167	0.637	-	-	-	С	6.44×10 <sup>-11</sup>	1.00	0.100	0.500	0*	0.400	-	
	D	×	×	-	×	-	×	-	E	20.1	1.19	0.167	0.637	-	-	0*	
	E	×	×	×	×	-	-	×	Α	20.1	1.19	0.167	0.637	-	-	-	
	$\sim$	×	×	×	×	×	×	-	D	1050	0*	-	0.0351	-	1.97	-	
С	C	$2.78 \times 10^{-9}$	1.00	0.100	0.500	0.200	0.400	-	С	2.78×10 <sup>-9</sup>	1.00	0.100	0.500	0.200	0.400	-	
	В	0.558	0.994	0.160	0.913	-	0.408	-	В	0.558	0.994	0.160	0.913	-	0.408	-	
	Α	28.0	1.19	0.213	1.17	-	-	-	D	23.9	0*	-	1.22	-	0.418	-	
	D	×	×	-	×	-	×	-	E	28.0	1.19	0.213	1.17	-	-	0*	
	F	×	×	×	×	-	-	×	Α	28.0	1.19	0.213	1.17	-	-	-	
D	D	1.83×10 <sup>-14</sup>	1.00	-	0.500	-	0.400	-	D	$1.83 \times 10^{-14}$	1.00	-	0.500	-	0.400	-	
	А	576	1.02	3.98	0.623	-	-	-	E	358.	1.13	3.63	0.285	-	-	0*	
	E	×	×	×	×	-	-	×	В	399.	1.10	3.74	0.395	-	0*	-	
	В	×	×	×	×	-	×	-	C	434.	1.18	3.43	0.454	0.528	0*	-	
	C	×	×	×	×	×	×	-	A	576.	1.02	3.98	0.623	-	-	-	
E	E	9.26×10 <sup>-11</sup>	1.00	0.100	0.500	-	-	0.700	E	9.26×10 <sup>-11</sup>	1.00	0.100	0.500	-	-	0.700	
	A	1.46	0.702	0.0564	0.367	-	-	-	С	1.46	0.702	0.0564	0.367	0*	0*	-	
	D	×	×	-	×	-	×	-	Α	1.46	0.702	0.0564	0.367	-	-	-	
	В	×	×	×	×	-	×	-	В	1.46	0.702	0.0564	0.367	-	0*	-	
	С	×	×	×	×	×	×	-	D	2.57	0*	-	0.258	-	0.0284	-	

0\*: exact zero value.

-: no corresponding parameters.

×: no real positive solutions.

#### The verification of our method, Measure 1



## When *kSS* is small, Model E is almost equal to model A

Model E



#### The verification of our method, Measure 2





## Summary

- We have proposed a method to select a model which is more/most consistent with the time series of observed data.
- We have verified our method, using generated data, handling a cyclic relationship hitherto unavailable in previous methods.

## Future works ... scalability

- Focusing on a local network within a large-scale network.
- Easy elimination of the unnecessary variables in virtue of algebraic equations Eliminating C(t),



$$k_{2}C_{1}(t) = k_{1}C_{2}(t) - k_{1}k_{d}\int_{0}^{t} e^{-k_{d}(\tau-t)}C_{2}(\tau)d\tau$$
  
-(sk\_{2} + k\_{2}k\_{d})L[C\_{1}(t)](s) + sk\_{1}L[C\_{2}(t)](s) = 0

## International Conference on Algebraic Biology

2005 (1st) November 28-30 FUJITSU SOLUTION SQURE Tokyo, Japan organized by H. Anai, H. Hong, K. Horimoto



from Universal Academy Press

2007 July 2-4
RISC, Johannes Kepler University
Linz, Austria

organized by B. Buchberger, H. Hong, K. Horimoto



LNCS 4545 from *Springer* 

July 31-August 2 RISC, Johannes Kepler University Linz, Austria organized by B. Buchberger, K. Horimoto, R. Laubenbacher, B. Mishra

Paper deadline is Jan.-14, 2008



http://www.risc.uni-linz.ac.at/about/conferences/ab2008/

2009 (4th)

## SAMSI, NC, USA organized by R. Laubenbacher