

11

# Applications in bioinformatics, systems biology and artificial life

Cedric Gondro

The Centre for Genetic Analysis and Applications  
University of New England

File Settings Help



## Input Sequences:

laboA	N--LFVALYDFVASGDNTLSITKGEKLRLVLCYMHNGE---	WCEAQTKNGQGWVPSNYITPVN	75.9259259259259	65		
lycsB	KGVIY-ALWDYEPQNDDELPMKEGDCMTI	HREDEDEIEWW--ARLNDKEGYVPRNLLGLYP	75.9259259259259	65		
laboA	NL-FVALYDF-----VASGDNTLSITKGEKLRLVLCYMHNGE	WCEAQTKNGQGW-----PSN--YITP--VN-	60.7843137254902	71		
lpht	GYQYRALYDYKKEREEDIDLHLGD-ILTVNKGS	-SLVALGFS-DGQ--EARP-EEIGWLNGYN	NETTGERGDFPGTYVEYIGRKISP	60.7843137254902	71	
laboA	NLFVALYDFVASGDNTLSITKG-EKLRLVLCYMHNGE	WCEAQTKNGQG--WV--PSNY-ITP-----V-N	61.1111111111111	13		
lihvA	N-F-RVY-Y---RDSRDPVWKGP	AKL--L-----W-----KGEGAVVIQDNSDIKVVP	VPRRKAKIIRD	61.1111111111111	13	
laboA	NLF-V-ALYDFVASGDNTLSITKGEKLRLVLCY---	NHNGEWCEAQTKNGQGW-----V-PSNYITPVN	72.2222222222222	73		
lbb9	-MFKVQAQHDYTATD	TDELQLKAGDVVLVIPFQNPEEQDEGW-LMGVKESD	-WNQHKELEKCRGVFPENFTERVQ	72.2222222222222	73	
<hr/>						
lycsB	KGVIY-ALWDYEPQNDDELPMKEGDCMTI-----I-----	HREDEDEIEWW--ARLNDKEGYVPRNLLGLYP	73.2142857142857	66		
lpht	-GYQYRALYDYKKEREEDIDLHLGD-ILTVNKGS	LVALGFSDGQE	ARPEEIGWLNGYN	NETTGE-RGDFPGTYVEYIGRKIS--P	73.2142857142857	66

## Aligned Sequences:

```

laboA      ----NL-F-V-ALYDFVASGDNTLSITKGEKLRLVLCY----NHNGEWCEAQTKNGQGW-----V-PSNYITPVN-----
lycsB      -----KGVIYALWDYEPQNDDELPMKEGDCMTI--HREDEDEIEWW----UW--ARLNDKEGYVPRNLL-GLYP-----
lpht       -----GYQYRALYDYKKEREEDIDLHLGD-ILTVNKGS
```

```

LVALGFSDGQE
```

```
ARPEEIGWL--NGYN
```

```
ETTGE-RGDFPGTYVEYIGRKIS
```

```
PRRKAKIIRD-----
```

```
-MFKVQAQHDYTATD
```

```
TDELQLKAGDVVLVIPFQNPEEQDEGW-LMGVKESD-WNQHKELEKCRGVFPENFTERVQ-----
```

Number of Sequences: 5

Fitness: 19921.1465972214

# Comparison

	msa-clustal	pre-clustal	msa-pre
better	12	17	8
worse	16	4	20
same	0	7	0
total	28	28	28

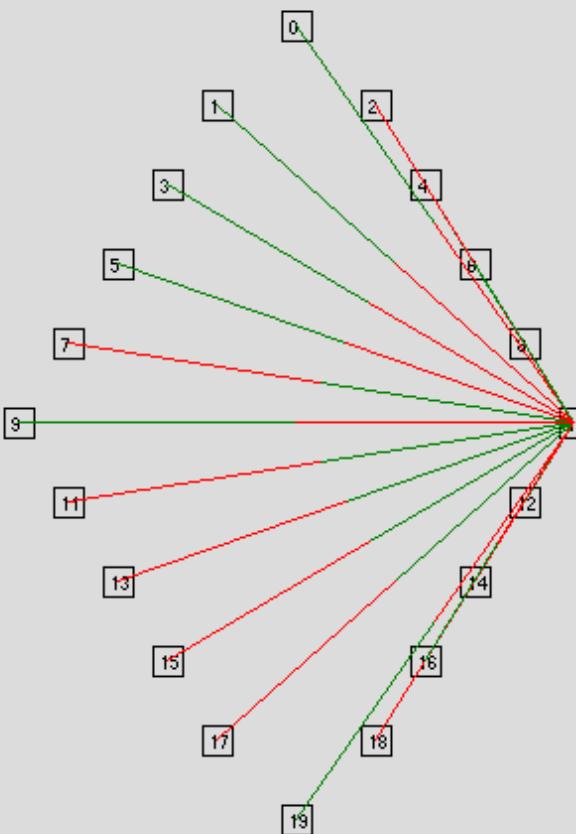
## averages

MSA-GA	<b>0.59042857</b>
MSA-GA w/p	<b>0.65292857</b>
Clustal	<b>0.63978571</b>

Generation: 28719

Organism: 5 Fitness: -36.1

Cy3	0	10	10	10	5	6	10	10	19	4	10	10	12	16	3	10	10	11	10	9	10	1
Cy5	10	18	13	15	10	10	8	17	10	14	10	7	10	10	3	10	2	11	10	9	10	10

**Parameters****Optimization Parameters**

Population size:

100

Number of generations:

100000

Crossover rate:

0.9

Mutation rate:

0.05

**Criterion Weights** Minimize number of arrays:

0.0

 Minimize number of steps:

1

 Weight of contrasts:

0.0

 Balance Dyes:

1

 Variance (A-optimality):

0.0

**Microarray Parameters**

Maximum number of arrays:

4

Number of samples:

4

Maximize Contrasts

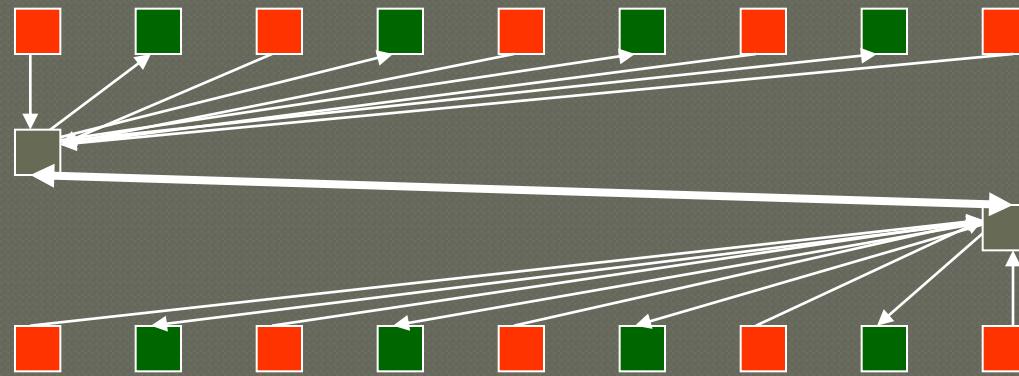
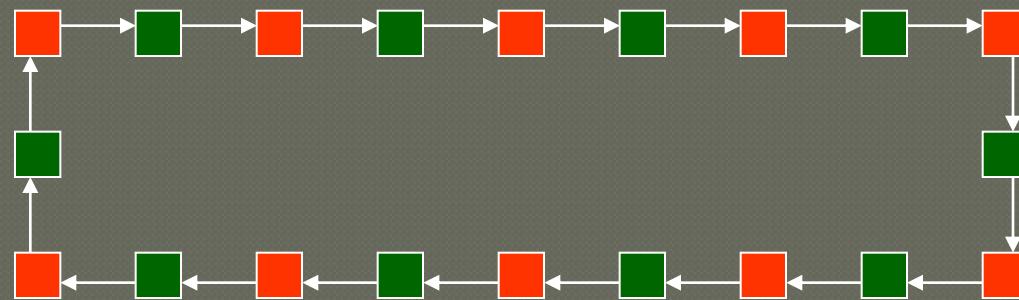
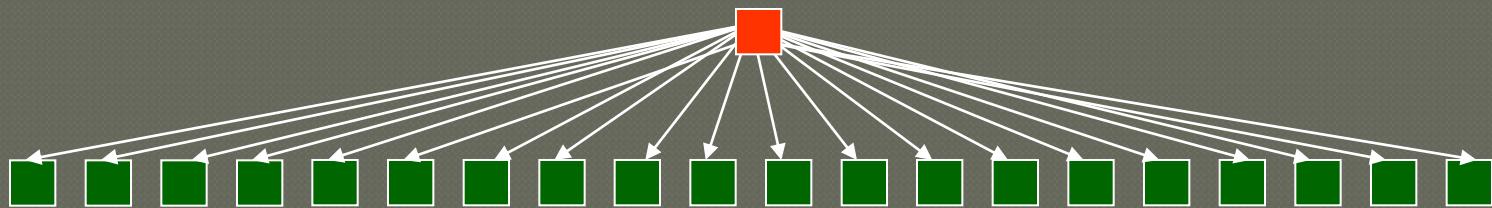


Minimize Contrasts

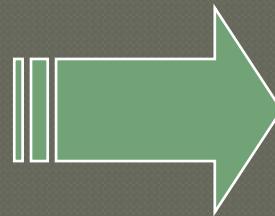
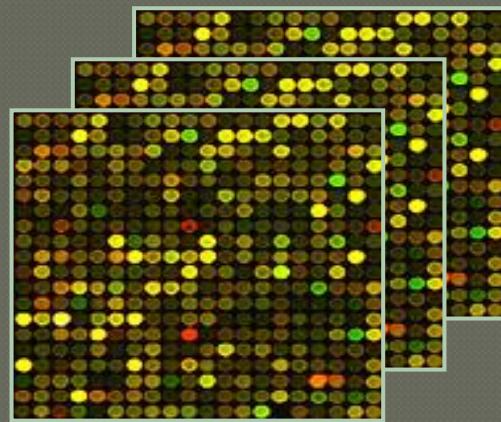


OK

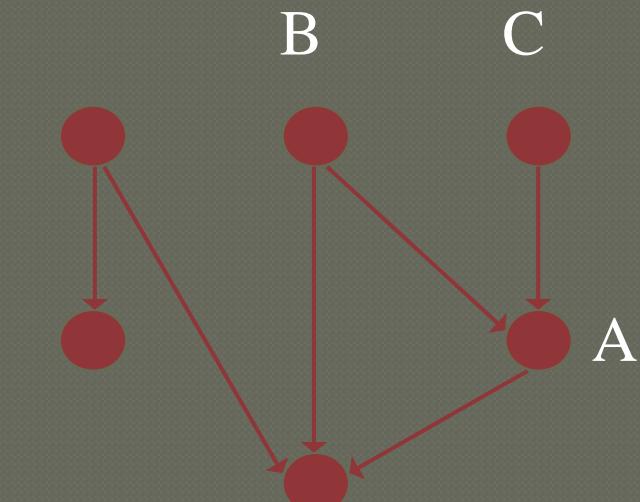
Cancel



# Biological Process Reconstruction



Microarray gene expression  
data



Genetic network

# Objectives

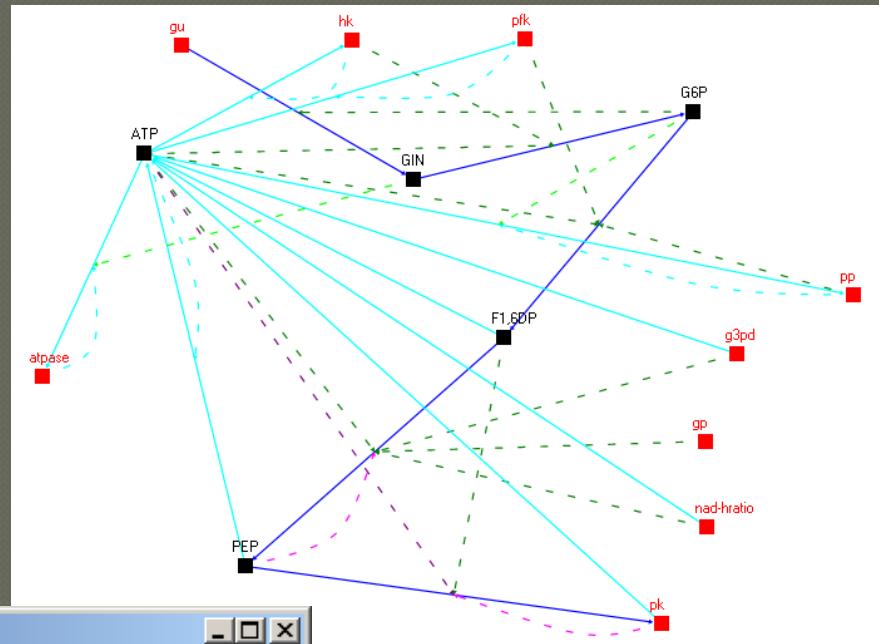
---

- To find a mathematical model that:
  - predicts how a biological process changes over time
  - predicts measurable outputs of the process
  - Helps explain/understand how the process operates
- Parameterize the model with realistic biological parameters
  - Estimate how parameter changes affect the process

# Parameterization of S-Systems Equations

## Fermentation pathway in *Saccharomyces cerevisiae*

$$\dot{X}_i = \alpha_i \prod_{j=1}^{n+m} X_j^{g_{ij}} - \beta_i \prod_{j=1}^{n+m} X_j^{h_{ij}}$$



**Equations: Anaerobic Fermentation**

File Options

Equations:

```

GIN = 0.8122*gu^1*G6P^-0.2344 - 2.8632*GIN^0.7464*ATP^0.0243*hk^1

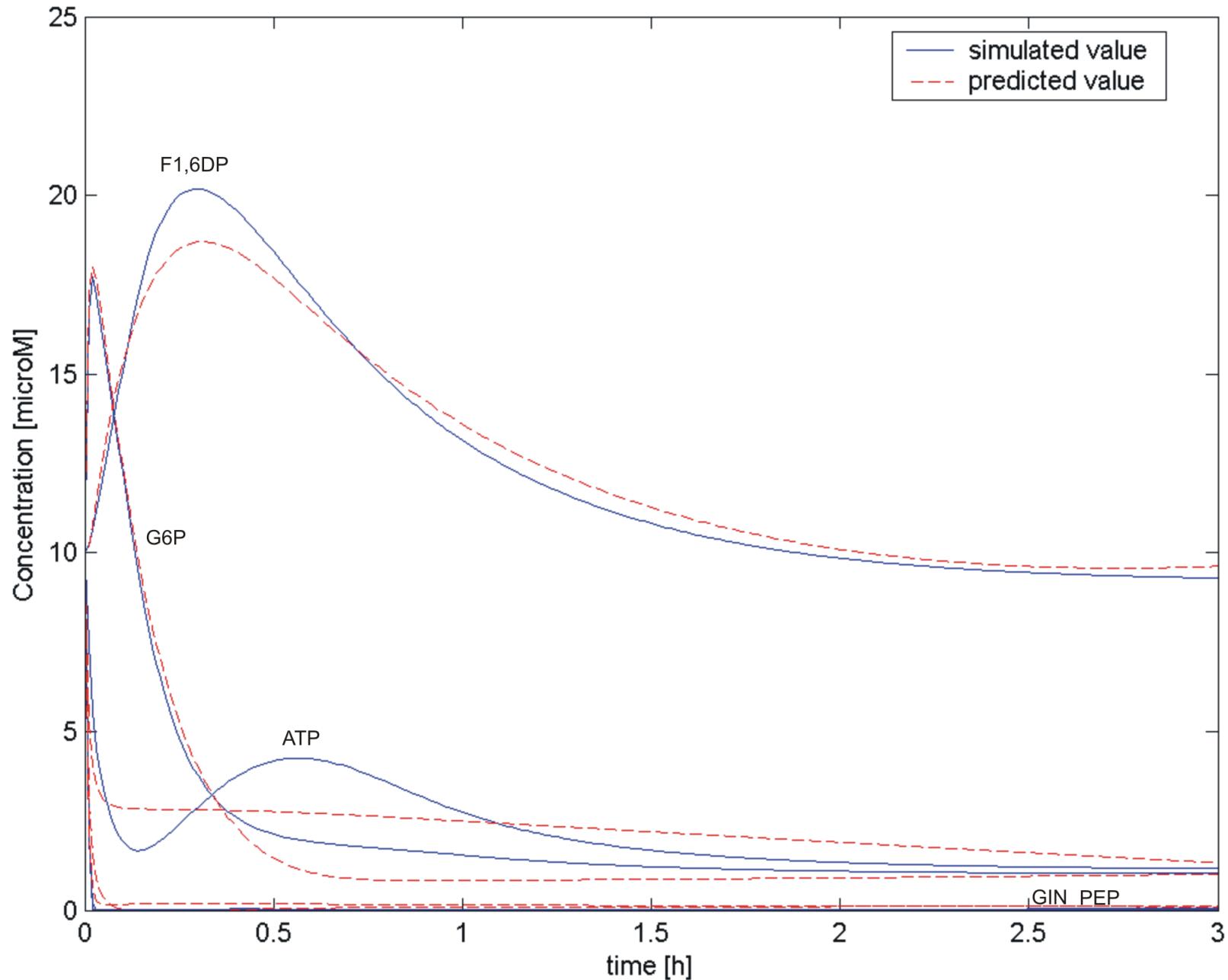
G6P = 2.8632*GIN^0.7464*ATP^0.0243*hk^1 - 0.5239*G6P^0.735*ATP^-0.394*pfk^0.999*pp^0.001

F1,6DP = 0.5232*G6P^0.7318*ATP^-0.3941*pfk^1*pp^0 -
0.0148*F1,6DP^0.584*ATP^0.119*g3pd^0.944*gp^0.056*nad-hratio^-0.575*PEP^0.03

PEP = 0.022*F1,6DP^0.6159*ATP^0.1308*g3pd^1*gp^0*nad-hratio^-0.6088*PEP^0 -
0.0945*PEP^0.533*F1,6DP^0.05*ATP^-0.0822*pk^1

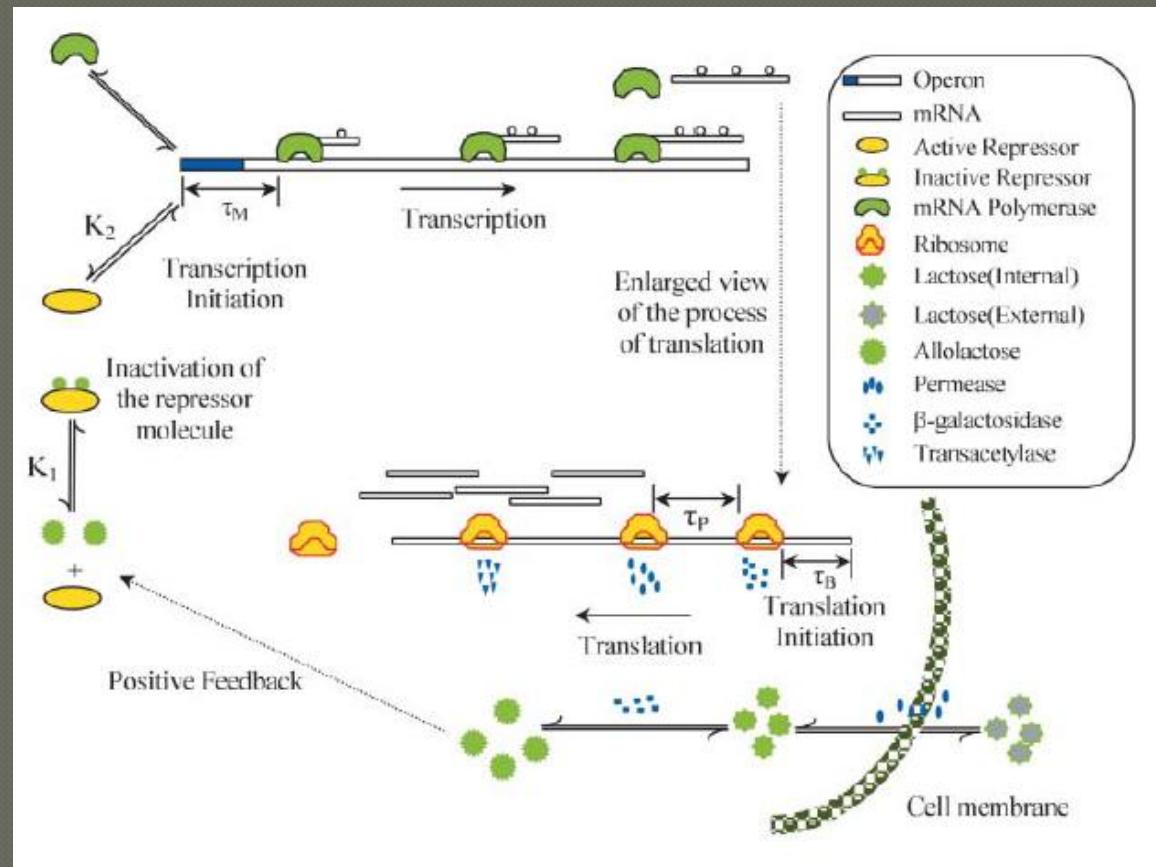
ATP = 0.0913*F1,6DP^0.333*PEP^0.266*g3pd^0.5*pk^0.5*nad-hratio^-0.304*ATP^0.024 -
3.2097*ATP^0.372*pp^0.0002*atpase^0.47*pfk^0.265*hk^0.265*GIN^0.198*G6P^0.196

```



# Lac Operon – A Well Characterized System

- Classic model of gene regulation
- Few elements
- Well studied and parameterized
- Biological measurements available
- Various mathematical models



# A System of Delay Differential Equations for the lac operon in *E. coli* – I

---

- 4 time delay differential equations
- 8 parameters
- 6 functions:
  - Plus
  - Minus
  - Power
  - Division
  - Multiplication
  - Time delay

$$\frac{dM}{dt} = \frac{1 + k_1 y_4^\rho}{1 + y_4^\rho} - b_1 y_1$$

$$\frac{dP}{dt} = y_1 - b_2 y_2$$

$$\frac{dB}{dt} = r_3 y_1 - b_3 y_3$$

$$\frac{dL}{dt} = S y_2 - y_3 y_4$$

# A System of Delay Differential Equations for the lac operon in *E. coli* – II

---

- 3 time delay differential equations

- 17 parameters

- 6 functions:

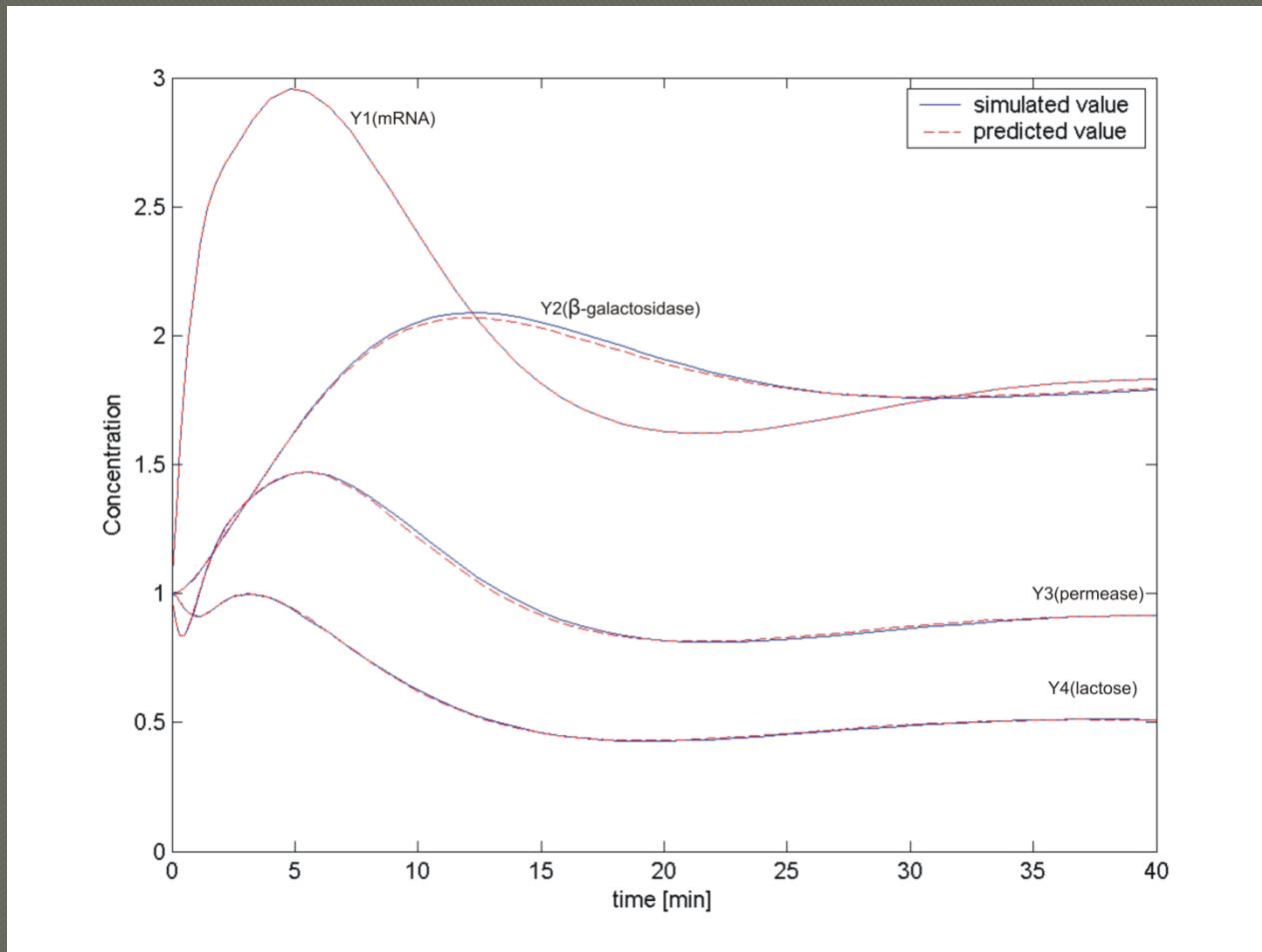
- Plus
- Minus
- Power
- Division
- Multiplication
- Time delay

$$\frac{dM}{dt} = \alpha_M \frac{1 + k_1(e^{-\mu\tau_M} A_{\tau_M})^n}{k + k_1(e^{-\mu\tau_M} A_{\tau_M})^n} - \tilde{Y}_M M$$

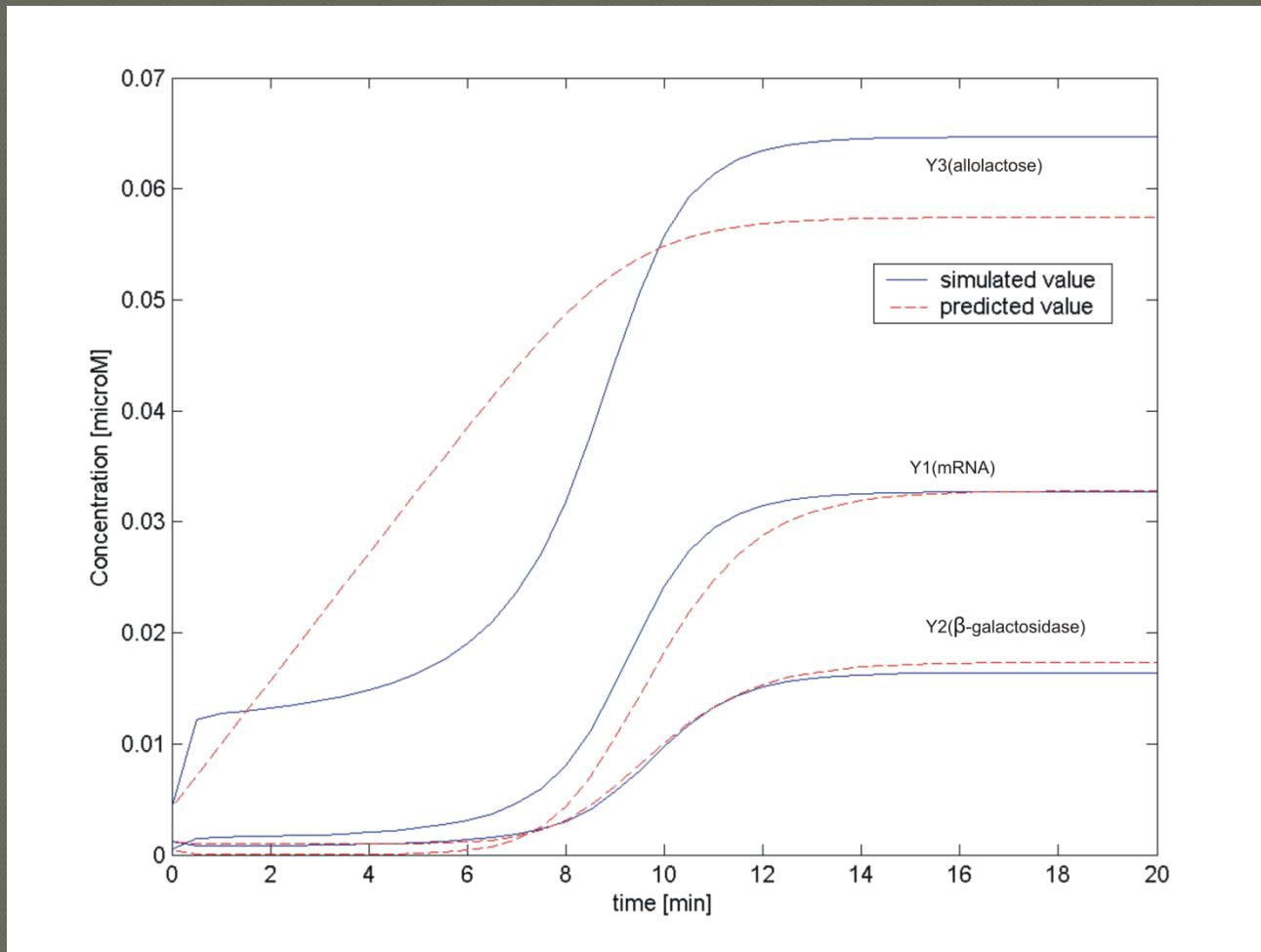
$$\frac{dB}{dt} = \alpha_B e^{-\mu\tau_B} M_{\tau_B} - \tilde{Y}_B B$$

$$\frac{dA}{dt} = \alpha_A B \frac{L}{K_L + L} - \beta_A B \frac{A}{K_A + A} - \tilde{Y}_A A$$

# Goodness of Fit Model I



## Goodness of Fit Model II



## Evolved solutions

---

$$y1 = \frac{4.9987 y_4^{t-0.64} y_4^{t-0.64} + 1}{y_4^{t-0.64} y_4^{t-0.64} + 1} - y_1$$

$$y2 = y_1 - (y_2 + y_2)$$

$$y3 = y_1 / 10.1023 - y_3 / 10.1023$$

$$y4 = y_2 - y_3 y_4$$

$$\frac{dM}{dt} = \frac{1 + k_1 y_4^\rho}{1 + y_4^\rho} - b_1 y_1$$

$$\frac{dP}{dt} = y_1 - b_2 y_2$$

$$\frac{dB}{dt} = r_3 y_1 - b_3 y_3$$

$$\frac{dL}{dt} = S y_2 - y_3 y_4$$

---


$$y1 = 0.123003^{8.1379 - (\frac{y3}{y1 + y1 + 8.7614})} - y1$$

$$y2 = y1 - y1^{t-0.53614}$$

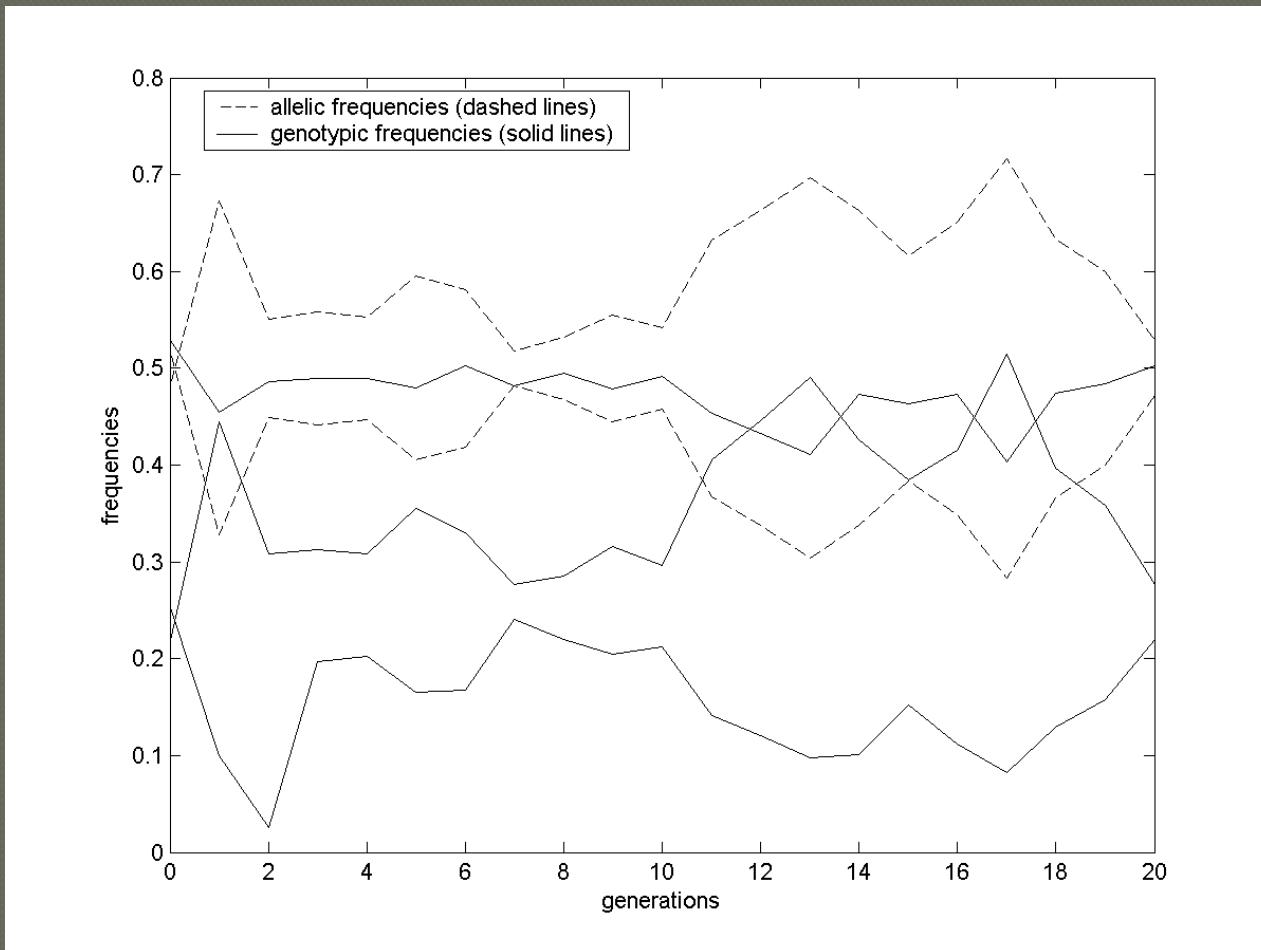
$$y3 = 0.12303 - (y2^{0.516968})$$

$$\frac{dM}{dt} = \alpha_M \frac{1 + k_1 (e^{-\mu \tau_M} A_{\tau_M})^n}{k + k_1 (e^{-\mu \tau_M} A_{\tau_M})^n} - \tilde{Y}_M M$$

$$\frac{dB}{dt} = \alpha_B e^{-\mu \tau_B} M_{\tau_B} - \tilde{Y}_B B$$

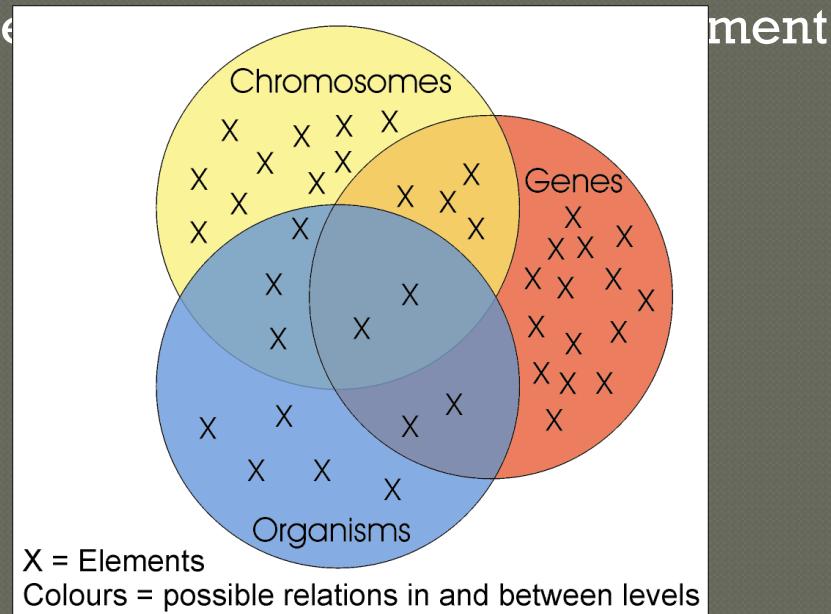
$$\frac{dA}{dt} = \alpha_A B \frac{L}{K_L + L} - \beta_A B \frac{A}{K_A + A} - \tilde{Y}_A A$$

# Alife – stealing it back



# A conceptual model of Mendelian populations

- Empirical model / echo model – simplified abstraction of the main mechanistic properties of biological populations
  - Universe – environment and population
  - Organisms – chromosomes, genes and alleles
  - Phenotypic correlation of the environment and other organisms



# Population genetics and evolution simulation

**Sigex AD: Example**

File Analysis Help

**Evolution**

Select the frequency:  Genotype  Allele

Care  Detect  Vision Generation 0 to 10  
 Reproduction  Partition  Fight OK  
 Death  Intelligence  Neutral  Sex Ratio

**Frequencies**

Select the frequency:  Genotype  Allele Generation << 10 >> OK

Care  Detect  Vision  
 Reproduction  Partition  Fight  
 Death  Intelligence  Neutral

Trait	Genotype	Allele
Care	0.17391	0.82609
Death	0.16500	0.83500
Intelligence	0.07000	0.36500
Neutral	0.21000	0.35500

**Initial Population Definition**

Male  Female  
 Maternal Inheritance  
 X Chromosome Detect Gene D d  
 Autosomal chromosomes Vision Gene V1 V2 V3 V4  
 Reproduction Gene R r  
 Partition Gene P p  
 Fight Gene L l  
 Death Gene M m  
 Intelligence Gene I1 I2 I3 I4  
 Neutral Gene N n

Add Individual Amount: 1

Id	Sex	Egg	Detect	Care	Vision	Repr	
00001	M	A	D	C	V1	V1	R
00002	M	A	D	C	V1	V1	R
00003	M	A	D	C	V1	V1	R

**HARDY-WEBERG**

Select the genes: Gen. Obs. Exp.  $\chi^2$

Gene	Obs.	Exp.	$\chi^2$
Vision	0	1.3225	1.3225
V1V1:	0	1.3225	2.1278
V2V2:	3	1.3225	0.0000
V3V3:	26	26.0100	0.0000
V4V4:	6	6.7600	0.0854
V1V2:	7	2.6450	7.1705
V1V3:	13	11.7300	0.1375
V1V4:	3	5.9800	1.4850
V2V3:	5	11.7300	3.8613
V2V4:	5	5.9800	0.1606
V3V4:	32	26.5200	1.1324
Total chi-square		17.4830	
Partition			
pp:	95	94.0900	0.0088
Pp:	4	5.8200	0.5691

Chi-Square Generation << 10 >> OK

**SIGEX: Untitled**

Simulation Configuration Help

**Simulation Parameters**

Initial Population:  Random 100  with Defined  
 Defined ...

Food: Initial amount 20 Reposition rate (%) 100 Energetic value 200 to 200

Generation: Overlapping Qty. 200

Recombination rate ... Advanced configuration ...

Individual(s): 3

Animation Interval 1 cycles Show energy  Cycle real time 0 ms

Generate Trace file

Start Gon... Sig... Cor... SIG... Sig... Cor... EN 5:33 PM

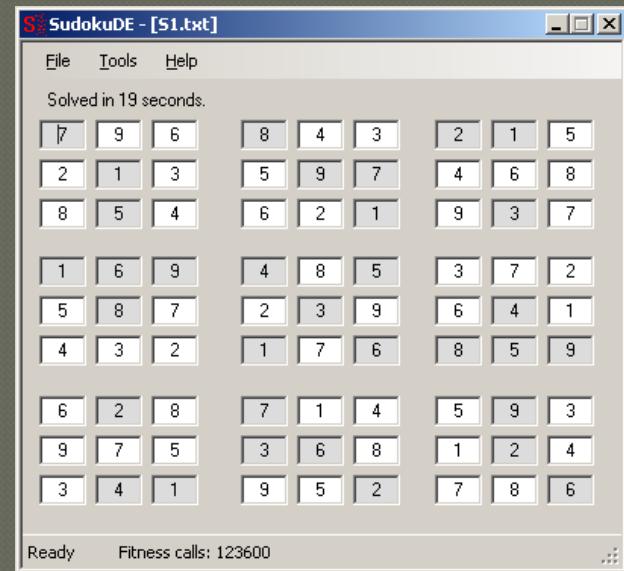
## Metaheuristics can solve sudoku puzzles

Rhyd Lewis

Received: 7 July 2005 / Revised: 15 March 2006 /  
Accepted: 6 June 2006 / Published online: 1 May 2007  
© Springer Science+Business Media, LLC 2007

**Abstract** In this paper we present, to our knowledge, the first application of a metaheuristic technique to the very popular and NP-complete puzzle known as ‘sudoku’. We see that this stochastic search-based algorithm, which uses simulated annealing, is able to complete logic-solvable puzzle-instances that feature daily in many of the UK’s national newspapers. We also introduce a new method for producing sudoku problem instances (that are not necessarily logic-solvable) and use this together with the proposed SA algorithm to try and discover for what types of instances this algorithm is best suited. Consequently we notice the presence of an ‘easy-hard-easy’ style phase-transition similar to other problems encountered in operational research.

**Keywords** Metaheuristics · Sudoku · Puzzles · Phase-transition



# University of New England

