Statistical Properties of Evolved Gene Regulatory Networks-Feed Forward Loops in *Escherichia coli*

Max B. Cooper*, John F.Y. Brookfield, Matthew Loose
School of Biology, University of Nottingham
*COMPLEX, University College, London

ESEB Turin, 25th August 2009
Gene Regulation Networks

Genes Control the Expression of Other Genes

Boolean or Quantitative Effects
Feed Forward Loops

C1  X  Y  Z
I1

C2
I2

C3
I3

C4
I4

OAO
AOA
AAO
AAA
Equations for Gene Expression

\[ P_{ij} = \frac{Q_{ij}[i]}{(Q_{ij}[i] + 1)} \]

\[ T_j = \sum_i P_{ij} E_{ij} \]

\[ E_{ij} < 0 = \text{Repression} \]
\[ E_{ij} > 0 = \text{Activation} \]

\[ [j]_{t+1} = 0.9[j]_t + T_j \]
Escherichia coli FFLs

Data from Shen-Orr et al. (2002) Nature Genetics 31:64

36 Feed Forward Loops where we can:

Classify in terms of 6 Possible Expression Patterns for Gene X

Classify in terms of whether input on gene X (Sx) is a linear increase or sigmoidal growth curve

Classify in terms of whether there are autoregulations on genes X, Y or Z

Classify in terms of whether binding of gene X and gene Y products on gene Z is independent or competitive
The Structure of the Model

E.coli

Input

Output

Model

?=?

X

Y

Z

?=?
Evolutionary Change

The genotype is defined by $Q_{ij}$ and $E_{ij}$ values for all interactions in the network.

Initially, $\log_{10}Q_{ij}$ is randomly distributed from -2 to 2, and $\log_{10}E_{ij}$ randomly distributed from -1 to 2.

Mutation: Make random changes to a random $Q_{ij}$ or $E_{ij}$ (in range $0.5=>2$), randomly reallocate sign of $E_{ij}$. 
Assessment of Fitness

3300 time steps - Assess fitness every 300 time steps:
1 - (summed absolute deviations of [Z] from target) Retain mutations where:
(summed absolute deviations of [Z] from target) < previous value x 1.000001
Successful run has < 20% summed absolute deviations of [Z] from target after 1000 mutational steps
Define FFL Model

Define Sx Expression Pattern
Allocate autoregulation on X,Y&Z
Indepenent/Competitive Binding Sites
Constitutive inputs on Y/Z

Simulate Model

Randomize starting parameters, Qij and Eij

Run developmental simulation
(3300 rounds, sample every 300 cycles)

Estimate model fitness

From literature:
Threshold/Curve/Gradient
Activation/Repression

Estimate Z Expression

Predicted Z Expression

Significantly deleterious
Beneficial or weakly deleterious

Reject
Accept

Randomly mutate one parameter

Repeat with new start conditions
Activation Threshold for Z-Coherent 4 FFL

Sx → X

Y → Z

[Graph showing concentration over time for Sx, X, Y, and Z]
Activation Threshold for Z-Incoherent 3

\[ S_x \rightarrow X \]
\[ \downarrow \]
\[ Y \]
\[ \rightarrow Z \]

![Graph showing concentration over time](image)
For each *E. coli* FFL

Examine successful simulations for that FFL:

Classify each successful simulation in terms of the sign of the interactions in the FFL \((E_{ij} \text{ values})\)

What proportion of successful simulations match the signs of the interaction in the real FFL?
Likelihood Ratio

Numerator: Proportion of simulated FFLs that match the signs of real FFL

Denominator: Reciprocal of the number of FFLs that could theoretically yield the observed Z output pattern
Likelihood Ratio: an example

X=fnr
Y=arcA
Z=icdA

Coherent 3 : Negative Autoregulation on X
Proportion of successful solutions with correct signs =41.1%
Number of possible FFLs=16
LR=16x0.411 =6.581
Likelihood Ratio

Product of Likelihood Ratios across 36 FFLs in the data set:

\[2.27 \times 10^{11}\]
Biosystems (2008) 91: 231-244

Thanks to BBSRC for studentship for Max Cooper