Available online at www.sciencedirect.com
ScienceDirect
BioSystems 91 (2008) 231–244
BioSystems
www.elsevier.com/locate/biosystems

Evolutionary modelling of feed forward loops in gene regulatory networks

Max B. Cooper, Matthew Loose, John F.Y. Brookfield *

Institute of Genetics, School of Biology, University of Nottingham, Queens Medical Centre, Nottingham NG7 2UH, United Kingdom

Received 16 February 2007; received in revised form 16 July 2007; accepted 18 September 2007

Abstract

Feed forward loops (FFLs) are gene regulatory network motifs. They exist in different types, defined by the signs of the effects of genes in the motif on one another. We examine 36 feed forward loops in *Escherichia coli*, using evolutionary simulations to predict the forms of FFL expected to evolve to generate the pattern of expression of the output gene. These predictions are tested using likelihood ratios, comparing likelihoods of the observed FFL structures with their likelihoods under null models. The very high likelihood ratios generated, of over 10^{11}, suggest that evolutionary simulation is a valuable component in the explanation of FFL structure.

Keywords: Feed forward loops; Evolutionary simulation; Regulatory networks; Motifs; Likelihood ratios

1. Introduction

Gene regulatory networks are patterns of directed interactions between transcription factors and their target genes. These networks are complex, and, within them, smaller subsets of transcription factor–gene interactions can be described, called "network motifs". The feed forward loop (FFL) is a 3-gene network motif that is over-represented in real networks when compared with randomised networks with the same degree of connectivity (Conant and Wagner, 2003; Milo et al., 2004, 2002). A FFL comprises three genes, X, Y and Z, where X regulates Y and Z, and Y regulates Z (see Fig. 1). A FFL can be controlled by upstream transcription factors or cofactors, which regulate gene expression or control transcription factor activity by cofactor binding or covalent modification. Any factor that regulates X is described as an input on X, symbolised by Sx. Correspondingly, there can be an input on Y, or Sy, acting in addition to the regulation of Y by X (Mangan and Alon, 2003).

Fig. 1 shows the eight possible FFL types (coherent 1–4 and incoherent 1–4), defined by the signs of the effects of the genes' transcription factor products on their targets. In addition, many FFLs show autoregulations, which can be either positive or negative in sign.

Many authors have examined the role of FFLs, and why these motifs should be over-represented in gene networks (e.g. Ishihara et al., 2005; Mangan and Alon, 2003; Wall et al., 2005). One potential explanation for FFLs is to hypothesise that the FFL has evolved to convert a particular pattern of Sx input into a particular pattern of expression of Z. Here, we test this idea with an evolutionary model, which is designed and tested using *Escherichia coli* gene expression and regulatory network

* Corresponding author.
E-mail address: john.brookfield@nottingham.ac.uk (J.F.Y. Brookfield).

0303-2647/$ – see front matter © 2007 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.biosystems.2007.09.004

data as described below. The aims of this study are to test the validity of the evolutionary model in predicting FFL network structures, to increase our understanding of, and make predictions about, FFL gene expression and evolution, and to establish a method for quantifying the evolutionary importance of a gene expression function, in this case, the response of FFL gene Z to the input Sx.

For each of 42 FFLs identified in *E. coli* by Shen-Orr et al. (2002), we identify the relationship between the pattern of Sx expression and the pattern of Z expression that it creates. For example, we might see a steady increase in Sx being converted to a sudden increase in Z as the Sx level passes a threshold. We would describe this relationship as an activation threshold. For each FFL we wish to model, we propose, from published data, a relationship between Sx and Z. We can also observe whether any autoregulatory loops (Fig. 1, lowest row) exist in the FFL.

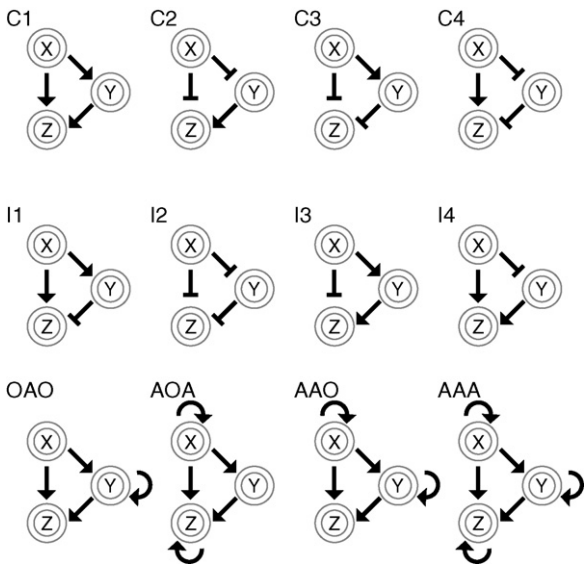

Fig. 1. Feed forward loops. The figure shows FFL types and some patterns of FFL autoregulation. Circular nodes represent genes, and black lines represent directed regulatory effects (arrowheads are activating, flat ends repress). Feed forward loops are termed coherent ("C1–4": top row) or incoherent ("I1–4": second row), depending on whether there are complementary or antagonistic effects of X on Z directly and indirectly, via Y. FFLs have been numbered according to their relative abundance in *Escherichia coli* and *Saccharomyces cerevisiae*, with the more common forms given lower numbers (Ishihara et al., 2005; Mangan and Alon, 2003; Shen-Orr et al., 2002; Wall et al., 2005). The third row shows examples of FFL autoregulation along with the terminology used to describe it: The letters "A" or "O", indicate the position of autoregulating genes in the order X, Y, then Z in the FFL, where A represents autoregulation, which can be positive, as shown here, or negative.

We seek to explain, at least partially, why a particular feed forward loop has the structure that it has, why it is a C1, or an I3, for example, and why the signs of any autoregulations are as they are. However, the type of a FFL, plus its autoregulatory interactions, does not uniquely specify the relationship between the input Sx and the output Z. A given type of FFL can create many different relationships between Sx and Z, depending on the quantitative details of the interactions between the genes. Equally, for a given relationship between Sx and Z, many FFL types, shown in the first two rows of Fig. 1, will be capable of creating this relationship. How, then, can we predict which types of FFL are likely to be seen in living organisms? Our approach is explicitly evolutionary. We note that, while many different FFL types might be equally capable of converting a given input Sx to a given output Z, these FFL types may well have unequal probabilities of evolving.

Therefore, for each FFL, we describe the interactions between the components of the FFL in terms of a gene expression model, which creates a pattern of expression of gene Z with respect to time, as a consequence of a temporally changing input Sx (see Fig. 2A). Parameters describing the interactions between gene products and their targets evolve in a simulation model which randomly mutates these parameters, and tests the effects on Z gene expression. A parameter change which improves the fit between the gene expression model and the relationship between Sx and Z observed in nature (which forms the aim of the simulation) is retained. After 1000 mutations have been tested, the gene expression pattern of the final outcome of the evolutionary simulation is tested against its aim, and classified as to whether it constitutes a successful solution to the gene expression task. Thus, for each evolutionary simulation, a simulated FFL is generated, which is classified in terms of which of the eight types of FFL it belongs to (themselves defined by the signs of the three between-gene effects in the FFL), and in terms of the signs of any autoregulations. We examine in detail the subset of these FFLs that constitute successful solutions to the gene expression task that has been set.

These evolutionary simulations, therefore, create, for each *E. coli* FFL, a frequency distribution of types of FFL among simulated FFLs that solve the gene expression task performed by that *E. coli* FFL. This allows testing of the model, and indication of whether our proposed expression patterns are likely to have shaped *E. coli* FFLs over evolutionary time. To explain FFL structures, we have to show that, among FFL types that could, in theory, be used to produce the observed relationship between Sx and Z, those that tend to be used, in real-