Evolution of 'design' principles in biochemical networks

P. de Atauri, D. Orrell, S. Ramsey and H. Bolouri

Abstract: Computer modelling and simulation are commonly used to analyse engineered systems. Biological systems differ in that they often cannot be accurately characterised, so simulations are far from exact. Nonetheless, we argue in this paper that evolution results in recurring, dynamic organisational principles in biological systems, and that simulation can help to identify them and analyse their dynamic properties. As a specific example, we present a dynamic model of the galactose utilisation pathway in yeast, and highlight several features of the model that embody such 'design principles'.

1 Introduction

Following the Second World War, possibly spurred by the horror of nuclear warfare, many prominent physicists turned their attention to biology and helped create the fields of structural and molecular biology. Much later, the advent of high throughput sequencing technologies and genome projects, led to a wave of computer scientists joining the fray, and created the field of bioinformatics. Less publicised, but equally successful, was the flow of control engineering ideas into biochemistry which led to the development of metabolic control analysis [1] and biochemical systems theory [2] in the seventies.

Much has been written in recent years of a new wave of migration from the applied quantitative sciences into biology, creating another new discipline: *systems biology*. Curiously, although systems biology is typically defined as the study of the dynamic behaviour of biochemical networks, computer modelling and simulation of network dynamics has so far played a relatively small part within the field compared to the construction and analysis of static large-scale network models from high-throughput data.

Computer modelling, simulation and analysis are of course, widely used in many branches of engineering for a wide variety of purposes. Apart from the common use of simulation to check against unpredicted/undesirable system behaviours, simulation analysis is used to optimise the performance of systems ranging from digital electronics to sailing boats. Additionally, simulation analysis is also a common aspect of safety design: predicting how engineered systems fail under adverse conditions (e.g. cars in collision, power grid failures, and internet attacks). This has been possible because we can characterise the behaviour of engineered components to great accuracy.

By contrast, the components of biological systems are difficult to characterise. The kinetic behaviour of a protein specie may depend on its amount, conformation, cellular location, and the milieu of other molecules present in the cell at the same time. None of these variables can be defined exactly. Many are fundamentally only definable as members of fuzzy sets with intrinsically noisy distribution profiles. Moreover, experimental measurement of molecular concentrations, protein states, interaction kinetics, etc. is inherently inexact. In-vitro measurements often do not reflect conditions inside a cell and can be orders of magnitude different from in-vivo values. In-vivo measurements, on the other hand, can currently only be carried out by proxy and provide very approximate values. For instance, the long half-life of Green Fluorescent Protein (as well as Luciferase and other reporters), means that in-vivo measurements represent the time-average (integral) of an activity, not its instantaneous value.

These, and other concerns that we will outline below, have led many experimental biologists to conclude that simulation and analysis of biochemical pathway kinetics is unlikely to provide predictive insights. Despite the success of biochemical simulation analyses that established the role of positive and negative feedback in providing robust network behaviours such as responses to environmental conditions [3, 4] and development [5–8], many experimentalists argue that cellular pathways are the idiosyncratic result of eons of evolutionary tinkering whose behaviour cannot be understood in terms of engineering principles. The extreme efficiency demonstrated in molecular systems ranging from the ATPase motor to the immune response are assumed to be exceptions, and most cellular pathways are said to resemble Rube-Goldberg [9]/Heath-Robinson [10] 'gadgets' which are neither optimal nor parsimonious. Thus, it is considered unreasonable to ask why a biochemical pathway is organised in a particular manner, not only because our models are inherently inaccurate, but also because there are no organisational (design) principles in Rube-Goldberg machines.

Engineered systems on the other hand, are assumed to be designed rationally and for a given purpose [Note 1]. However, engineered systems are frequently developed

Systems Biology online no. 20045013

doi: 10.1049/sb:20045013

Paper first received 1st April and in revised form 15th May 2004

The authors are with the Institute for Systems Biology, 1441 North 34th Street, Seattle, WA 98103, USA, email: hbolouri@systemsbiology.org

Note 1: It is instructive to note that some languages distinguish between 'why' meaning 'arising from what' and 'why' meaning 'for what purpose' (e.g. 'madua' and 'lama' in Hebrew). Clearly, the former is a reasonable question to ask of the evolution of both biological and engineered systems.

initially for one purpose, but are adapted and adopted for unforeseen new uses that belie the original intention [11]. Furthermore, as the books of Petroski [12] and others amply illustrate, engineered systems fail all too often, and frequently in predictable ways (e.g. the harmonic oscillations that destroyed the Tacoma Narrows Bridge near Seattle in 1940, and later also closed down the London Millennium Bridge in 2001). Numerous additional parallels (e.g. spandrels, vestigial structures, and 'evolutionary ghosts' such as the QWERTY keyboard layout) may be drawn between the evolution of engineered and biological systems. Yet, modelling, simulation, analysis, and even reverse-engineering are part and parcel of modern engineering.

Many cellular pathways are highly efficient and/or are highly conserved across multiple species (e.g. the β-catenin pathway, [13]). Given changing and unpredictable environments, and the competitive pressures of co-evolution, there can be no optimal systems, but rather an ongoing optimisation of trade-offs. Can the mechanisms that underlie biological efficiency be understood in terms of logical organisational principles? Does evolution lead to the emergence of recurring organisational motifs that we may consider analogous to 'design' principles? Or is each biochemical pathway a unique product of bricolage whose organisation cannot throw light on the organisation of other pathways?

Consider the principle of modularity (for a discussion and review see [14]). The modular organisation of stable chemical structures in biology is well established. For example, DNA is organised in a hierarchy of modules ranging from chromosomes, through heterochromatin and euchromatin, to the level of genes which themselves comprise cis-regulatory modules, promoter regions, introns, exons, etc, all the way down to the level of codons and beyond. Likewise, proteins comprise folds, interaction domains, etc., and metazoans comprise body parts, organs, cells and so on. Indeed, the existence of such modules underlies much of bioinformatics (e.g. sequence annotation, transcription factor binding site prediction, etc.) and computational biology (e.g. protein structure/function prediction).

Are inter- and intra-cellular biochemical networks – which are dissipative, far from equilibrium structures – also modularly organised? Certainly, traditional molecular and cell biology has tended to study small parts of molecular interaction networks as though they can be studied in isolation and independently of the context of their cellular interactions. This may however have more to do with the reductionist perspective of these disciplines (emphasising the need to understand the building blocks before attempting to understand a system) than with an explicit assumption of modularity.

We believe that evolution does result in recurring, dynamic organisational principles in biochemical pathways. Moreover, we posit that, in spite of its inherent inaccuracies, computer modelling and simulation can be used to identify and study such 'evolutionary design principles'. To make specific and illustrate our point, in the rest of this paper, we present a model of the yeast galactose utilisation pathway (a metabolic module) built from existing, publicly available data, and highlight several features of the model that embody 'design principles' already predicted theoretically, and/or observed in other biochemical pathways.

2 Overview of galactose uptake in yeast

Galactose utilisation in the yeast *Saccharomyces cerevisiae* has been extensively studied at the genetic and metabolic levels. As such, it provides an attractive case-study and an opportunity to investigate the interactions between these two organisational levels. Figure 1 summarises the pathway. Table 1 summarizes the abbreviations used in Fig. 1 and throughout the text. Galactose uptake begins with the entry of galactose into the cell through a galactose-inducible transport process dependent on the protein gal2p encoded by the gene *GAL2*, followed by the conversion of galactose into glucose 1-phosphate through the Leloir pathway [15, 16]. The metabolic conversions leading to glucose 1-phosphate require the action of three galactose-inducible enzymes, galactokinase, galactose-1-phosphate uridylyltransferase, and UDP-galactose 4-epimerase. These correspond to the products gal1p, gal7p and gal10p encoded by the genes *GAL1*, *GAL7* and *GAL10*, respectively. All of these genes are induced in galactose. For reviews of the regulation of galactose uptake see [17–19].

The *GAL -1*, *-2*, *-7*, and *-10* genes, together with the regulatory *GAL3* and *GAL80* genes, are members of the *GAL* regulon. Throughout this paper, the *GAL 1*, *-2*, *-7*, and *-10* genes are referred to as 'structural genes', and *GAL3* and *GAL80* as 'regulatory genes'. The *GAL4* protein gal4p promotes transcription of all the *GAL* genes by dimerising and binding with their *cis*-regulatory regions. The *GAL80* protein gal80p represses transcription by binding with gal4p. The *GAL3* protein gal3p is thought to be converted into an active form by intra-cellular galactose in an ATP dependent manner. In its active form, gal3p restores transcription of the *GAL* genes by relieving the repressive action of gal80p. Thus, an initial flow of galactose into the cell combines with basal levels of gal3p to switch on transcription of the *GAL* genes. These mechanisms are

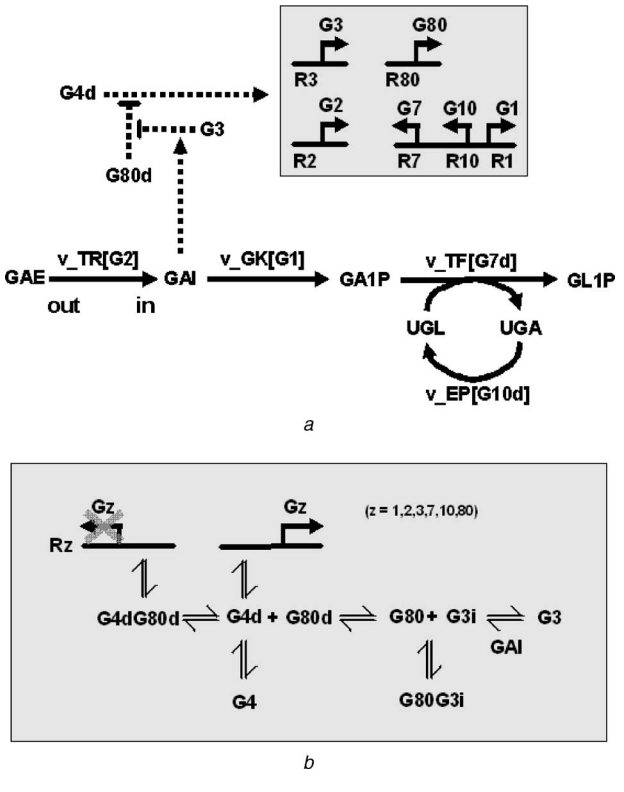

Fig. 1

a Schematic representation of the galactose uptake pathway. Full arrows represent mass transformation; dashed lines represent regulatory interactions, with arrowheads for activation and blunt ends for inhibition
b Detail of the control network with the protein-protein and protein-DNA interactions represented

Table 1: Abbreviations for chemical species

symbol	description	symbol	description
GAE	external galactose	GL1P	glucose 1-phosphate
GAI	internal galactose	UGL	UDP-glucose
GA1P	galactose 1-phosphate	UGA	UDP-galactose
R1	mRNA for GAL1	R7	mRNA for GAL7
R2	mRNA for GAL2	R10	mRNA for GAL10
R3	mRNA for GAL3	R80	mRNA for GAL80
G1	gal1p	G7d	gal7p dimer
G2	gal2p	G10	gal10p monomer
G3	non-induced form of gal3p	G10d	gal10p dimer
G3i	induced form of gal3p	G80	gal80p monomer
G4	gal4p monomer	G80d	gal80p dimer
G4d	gal4p dimer	G4dG80d	gal4p-gal80p complex
G7	gal7p monomer	G80G3i	gal80p-gal3p complex
v_TR	describes the transport of galactose inside the cell		
v_GK	describes the galactokinase reaction		
v_TF	describes the galactose-1-phosphate uridylyltransferase reaction		
v_EP	describes the UDP-galactose 4-epimerase reaction		

The same symbols are used to refer to species and for concentrations

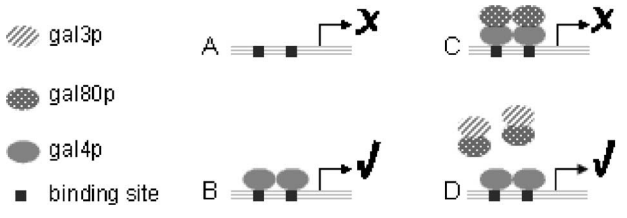

Fig. 2 *Transcriptional regulatory states of the GAL genes. DNA is represented by a thick horizontal line. The bent arrow represents the basal transcription apparatus. The DNA segment upstream (to the left) of this site represents the cis-regulatory region of the gene. Gal4p dimers bind DNA at specific locations with a palindromatic motif comprising three conserved bases at each end and variable base compositions in between. A, there is no induced transcription in the absence of gal4p. B, binding of gal4p dimers induces transcription. C, gal80p binds to gal4p and represses transcriptional activity. D, galactose-activated gal3p molecules bind gal80p and disrupt their repressive function*

illustrated schematically in Fig. 2. While the regulatory genes *GAL3* and *GAL80* each have only one binding site for the gal4p dimer, the structural genes *GAL 1, -2, -7*, and *-10* have multiple (2–5) binding sites.

With respect to galactose, the expression of *GAL1*, *-2*, *-3*, *-7*, *-10*, and *-80* genes falls into one of two states: (1) under non-induced conditions (e.g. in glycerol or raffinose media) *GAL3* and *-80* are expressed at a basal level, while *GAL2, -1*, *-7* and *-10* show no detectable transcriptional activity; (2) with galactose as the carbon source, the *GAL* gene expression is induced to high levels for the transporter and enzyme genes and to moderate levels for *GAL3* and *-80*. In addition, the expression of *GAL* genes is repressed in glucose, and only the (repressor) *GAL80* gene shows significant expression. Finally, *GAL4* expression is repressed in glucose, but appears not to be differentially regulated in the absence of glucose.

3 The galactose utilisation model

Our model is divided into three parts. (1) '**Enzymatic pathway**', refers to the transport of galactose inside the cell and enzymatic transformation of galactose within the cell. (2) '**Transcription and translation of structural genes**', refers to the transcription and translation of *GAL1*, *-2*, *-7*, and *-10*. (3) '**Control network**', refers to the protein-DNA and protein-protein interactions that control transcription of all *GAL* genes and includes the transcription and translation of the *GAL3* and *GAL80* genes. The '**enzymatic pathway**' and '**transcription and translation of structural genes**' models are in direct agreement with data from literature, while optimisation techniques were used to adjust rates where necessary to estimate parameters for the '**control network**'. The model is described in terms of ordinary differential equations in Table 2 (abbreviations listed in Table 1). Full details of the model are given in Appendix 8.1.

The model reproduces the un-induced and induced states, and allows for simulation of partial induction as observed by Biggar and Crabtree [20] who showed that varying external galactose levels produces graded changes in *GAL1* promoter activity. Li *et al.* [21] showed that this graded induction has a plateau reached at galactose concentrations between 1 g/L and 3 g/L (5.5 and 16.7 mM respectively), and maintained up to 20 g/L of applied external galactose. They also showed that at 0.1 g/L (0.55 mM) of external galactose, *GAL* gene expression is only 10% of the maximum. We capture these observations with a model of the fractional saturation of regulatory factors on DNA based on the Arhenius equation [22]. The derivation of this function for the different *GAL* genes, which have different numbers of gal4p dimer binding sites, is given in Appendix 8.2.

Regulation of galactose 1-phosphate concentration is of particular interest because it is known to be toxic, although the mechanism of toxicity remains unclear [23–27]. We optimised the parameters of our model to reproduce a value for galactose 1-phosphate close to that reported by Ostergaard *et al.* [25] for a continuous culture in 0.47 mM of external galactose. We considered two situations, one at 10% of maximal induction with external galactose set to 0.5 mM, and the other at full induction with 111 mM (20 g/L) of external galactose.

Table 2: Model description in terms of ordinary differential equations

Enzymatic part

$$d\,\mathrm{GAl}/dt = \mathrm{v\_TR} - \mathrm{v\_GK} - \mathrm{v\_G3i} \quad (1)$$

$$d\,\mathrm{GA1P}/dt = \mathrm{v\_GK} - \mathrm{v\_TF} \quad (2)$$

$$d\,\mathrm{UGL}/dt = \mathrm{v\_EP} - \mathrm{v\_TF} \quad (3)$$

$$d\,\mathrm{UGA}/dt = \mathrm{v\_TF} - \mathrm{v\_EP} \quad (4)$$

Transcription and translation of structural genes

$$d\,\mathrm{R1}/\mathrm{dt} = \mathrm{kir\_struct} \cdot F(K_P \cdot \mathrm{G4d}, K_Q \cdot K_P \cdot \mathrm{G4d} \cdot \mathrm{G80d}, 4) - \mathrm{kdr\_struct} \cdot \mathrm{R1} \quad (5)$$

$$d\,\mathrm{R2}/\mathrm{dt} = \mathrm{kir\_2} \cdot F(K_P \cdot \mathrm{G4d}, K_Q \cdot K_P \cdot \mathrm{G4d} \cdot \mathrm{G80d}, 5) - \mathrm{kdr\_2} \cdot \mathrm{R2} \quad (6)$$

$$d\,\mathrm{R7}/\mathrm{dt} = \mathrm{kir\_struct} \cdot F(K_P \cdot \mathrm{G4d}, K_Q \cdot K_P \cdot \mathrm{G4d} \cdot \mathrm{G80d}, 2) - \mathrm{kdr\_struct} \cdot \mathrm{R7} \quad (7)$$

$$d\,\mathrm{R10}/\mathrm{dt} = \mathrm{kir\_struct} \cdot F(K_P \cdot \mathrm{G4d}, K_Q \cdot K_P \cdot \mathrm{G4d} \cdot \mathrm{G80d}, 4) - \mathrm{kdr\_struct} \cdot \mathrm{R10} \quad (8)$$

$$d\,\mathrm{G1}/dt = \mathrm{kip\_struct} \cdot \mathrm{R1} - \mathrm{kdp\_struct} \cdot \mathrm{G1} \quad (9)$$

$$d\,\mathrm{G2}/dt = \mathrm{kip\_2} \cdot \mathrm{R2} - \mathrm{kdp\_struct} \cdot \mathrm{G2} \quad (10)$$

$$d\,\mathrm{G7}/\mathrm{dt} = \mathrm{kip\_struct} \cdot \mathrm{R7} - \mathrm{v\_G7d} - \mathrm{kdp\_struct} \cdot \mathrm{G7} \quad (11)$$

$$d\,\mathrm{G7d}/dt = \mathrm{v\_G7d}/2 - \mathrm{kdp\_struct} \cdot \mathrm{G7d} \quad (12)$$

$$d\,\mathrm{G10}/\mathrm{dt} = \mathrm{kip\_struct} \cdot \mathrm{R10} - \mathrm{v\_G10d} - \mathrm{kdp\_struct} \cdot \mathrm{G10} \quad (13)$$

$$d\,\mathrm{G10d}/dt = \mathrm{v\_G10d}/2 - \mathrm{kdp\_struct} \cdot \mathrm{G10d} \quad (14)$$

Control network

$$d\,\mathrm{R3}/dt = \mathrm{kir\_3} \cdot F(K_P \cdot \mathrm{G4d}, K_Q K_P \cdot \mathrm{G4d} \cdot \mathrm{G80d}, 1) - \mathrm{kdr\_reg} \cdot \mathrm{R3} \quad (15)$$

$$d\,\mathrm{R80}/dt = \mathrm{kir\_reg} \cdot F(K_P \cdot \mathrm{G4d}, K_Q K_P \cdot \mathrm{G4d} \cdot \mathrm{G80d}, 1) - \mathrm{kdr\_reg} \cdot \mathrm{R80} \quad (16)$$

$$d\,\mathrm{G4}/dt = \mathrm{kip\_4} - \mathrm{v\_G4d} - \mathrm{kdp\_reg} \cdot \mathrm{G4} \quad (17)$$

$$d\,\mathrm{G4d}/dt = \mathrm{v\_G4d}/2 - \mathrm{v\_G4dG80d} - \mathrm{kdp\_reg} \cdot \mathrm{G4d} \quad (18)$$

$$d\,\mathrm{G3}/dt = \mathrm{kip\_reg} \cdot \mathrm{R3} - \mathrm{v\_G3i} - \mathrm{kdp\_reg} \cdot \mathrm{G3} \quad (19)$$

$$d\,\mathrm{G3i}/dt = \mathrm{v\_G3i} - \mathrm{v\_G80G3i} - \mathrm{kdp\_reg} \cdot \mathrm{G3i} \quad (20)$$

$$d\,\mathrm{G80}/dt = \mathrm{kip\_reg} \cdot \mathrm{R80} - \mathrm{v\_G80d} - \mathrm{v\_G80G3i} - \mathrm{kdp\_reg} \cdot \mathrm{G80} \quad (21)$$

$$d\,\mathrm{G80d}/dt = \mathrm{v\_G80d}/2 - \mathrm{v\_G4dG80d} - \mathrm{kdp\_reg} \cdot \mathrm{G80d} \quad (22)$$

$$d\,\mathrm{G4dG80d}/dt = \mathrm{v\_G4dG80d} - \mathrm{kdp\_reg} \cdot \mathrm{G4dG80d} \quad (23)$$

$$d\,\mathrm{G80G3i}/dt = \mathrm{v\_G80G3i} - \mathrm{kdp\_reg} \cdot \mathrm{G80G3i} \quad (24)$$

v_TR, v_GK, v_TF and v_EP refer to the equations describing the transporter and enzymatic reactions ((1), (2), (3) and (4) in Appendix 8.1); kir_struct, kir_2, kir_3 and kir_reg are the maximum initiation rates for transcription, and F is the fractional saturation ((15) in Appendix 8.2), and depends on the equilibrium constants K_P and K_Q defined in Table 3 ((2) and (6)); kdr_struct, kdr_2, and kdr_reg are the decay rates for mRNA; kip_struct, kip_2, kip_4 and kip_reg are the initiation rates for translation; kdp_struct and kdp_reg are the decay rate for proteins; v_G3i, v_G4d, v_G80d, v_G4dG80d and v_G80G3i account for sugar-protein and protein-protein interactions ((1), (3), (4), (5) and (7) in Table 3); v_G7d and v_G10d refer to the dimerisation of G7 and G10, and use equivalent rate values as for v_80d ((4) in Table 3)

The entire model was implemented and simulated using the Open Source program Dizzy (http://labs.systemsbiology.net/bolouri/Dizzy) which permits stochastic and ODE-based simulations, hierarchical modelling, and model-instantiation and re-use. The model is available as part of the online supplementary materials (available at www.iee.org/sb/) and may be imported into other simulation environments using the ability of Dizzy to read and write the Systems Biology Markup Language (www.sbml.org) model exchange standard. The stochastic simulations are performed using the Accelerated Approximate Stochastic Simulation Algorithm (also called the 'Tau-Leap' method, see [28] for a full description). The computational efficiency of the approximate Tau-Leap method is monitored, and when it becomes inefficient to evaluate the 'leap' time, the method switches to employing Gillespie's well-known discrete-event stochastic simulation technique [29].

Figure 3 shows some example simulation results from the model and compares the stochastic and ODE-based simulation results. Note the high degree of variability in

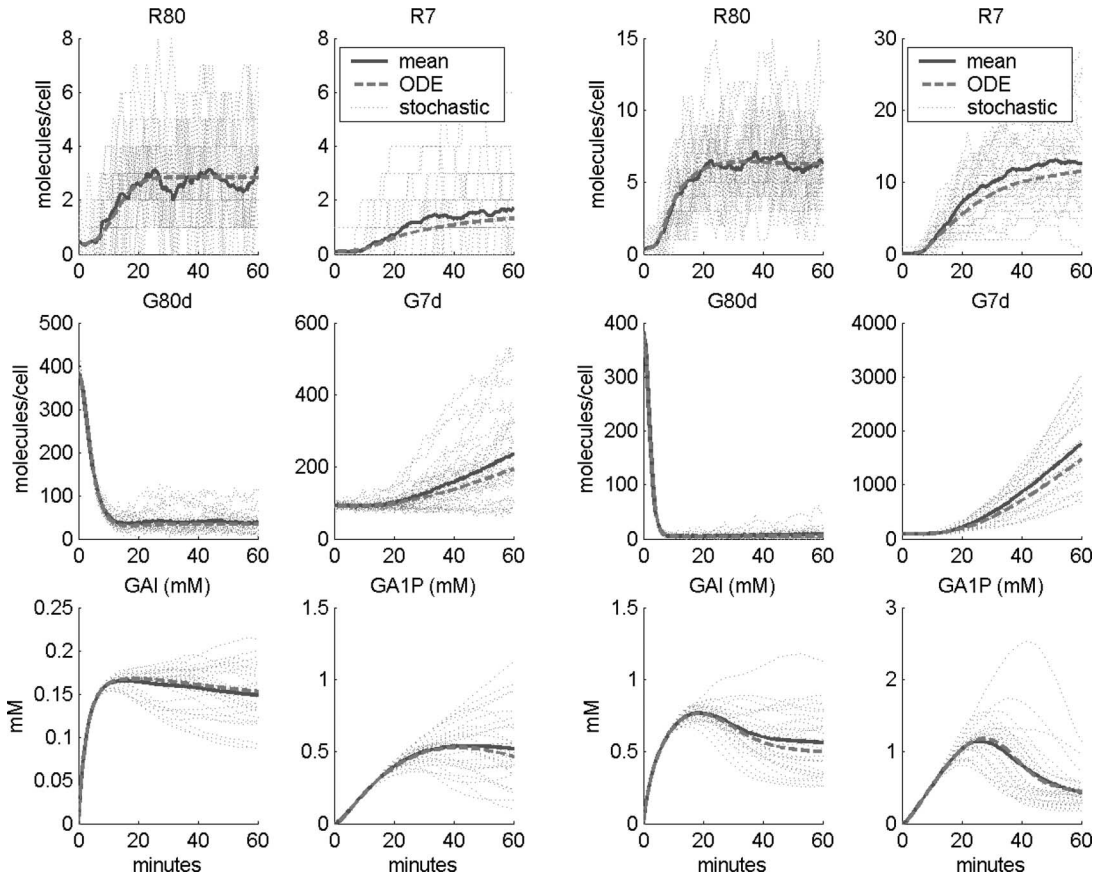

Fig. 3 *Comparison of ODE and stochastic simulations. The left-two columns show the transition from uninduced to low induction (0.5 mM GAE). The right-two columns show the transition to full induction (111.1 mM GAE). Shown are several key species. The regulatory protein gal80p is a repressor, gal7p is a structural protein. Dotted lines show 30 stochastic simulations, solid grey line is the mean of the ensemble, dashed line is the ODE simulation*

mRNA and protein concentrations in individual cells. Note also that the average of several stochastic simulations (here 30 individual simulation runs) approaches the behaviour of the ODE simulations. However, the large deviations from this average for individual stochastic simulations indicate that individual cells may experience conditions very different from the population average (the quantity measured by most experimental assays).

4 Organisational principles in galactose utilisation

In this Section, we use simulation and analysis to show that the organisation of the galactose utilisation pathway confers specific operational advantages to it when compared with alternative implementations. We posit that while such features have arisen by chance, they were selected because of their superior performance. We note that the same, or very similar features have been predicted theoretically and observed in other biochemical pathways (see references in individual Sections below), making these recurring dynamic motifs potential candidates for the role of 'design principles' emerging from evolution.

4.1 Dimerisation reduces intrinsic noise

One example of a dynamic motif is the dimerisation of key proteins. As discussed in the Appendix, the repressor gal80p, and the structural proteins gal7p and gal10p, all form dimers, which has been shown to reduce noise in genetic networks [30]. In order to assess the degree of noise reduction, it is convenient to work in terms of the Fano factor, which is the ratio of variance to mean. In a Poisson process the variance scales with the mean, so the Fano factor is one [31]. Analytical estimates [32] show a significant reduction in the Fano factor for the dimer (compared to the monomer, see Fig. 4). Because protein is produced from mRNA, which itself is subject to stochastic noise, the Fano factor for proteins is typically much higher than one, but dimerisation still reduces the Fano factor by a proportionate amount. For example, at low induction, the gal10p dimer has a Fano factor of about 317, while the total gal10p, which is the number of monomers plus twice the number of dimers, has a Fano factor of about 651, a factor of two greater. Put another way, if gal10p did not dimerise, but was produced to the same level as gal10p dimer, then the expected ratio of standard deviation (square root of the variance) to mean would be higher by about a factor of the square root of two.

4.2 Feedback regulation of gal3p and gal80p reduces sensitivity to transient changes in galactose

A feature of this system, which may initially appear counter-intuitive, is that both gal80p and gal3p proteins are up-regulated when the pathway is activated by galactose [33, 34]. Since gal80p represses transcription of the *GAL* genes, this is analogous to galactose activation trying to turn itself off (negative feedback). *GAL3* also has an autocatalytic effect; however, here the feedback is positive. Galactose-activated gal3p protein molecules result in de-repression of the *GAL3* gene itself. Thus, a basal amount of gal3p protein is necessary in order to lock on the *GAL3* positive feedback loop and activate all the *GAL* genes. Based on simulations of the *GAL3, -4,* and *-80* interactions, Verma *et al.* [35] proposed that the *GAL80* autoregulatory

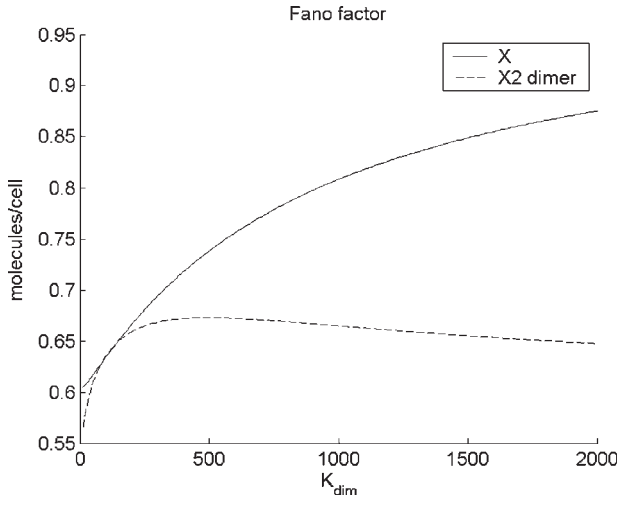

Fig. 4 *Plot of Fano factor (defined as the ratio of variance to mean) for an example scenario. Fano factors for a hypothetical monomer protein X and its dimer X2 are shown [32]. Here K_{dim} is the ratio of reverse to forward dimerisation rates. In this simple example, protein is assumed to be created and destroyed at fixed rates in a simple Poisson process. The plot shows how the dimerisation reaction reduces noise well below what would be expected in a simple Poisson process (i.e. a Fano factor equal to one). Results are consistent with noise reduction for dimers in the galactose model, although in that case, the protein Fano factors are much higher than Poisson because proteins are synthesised from mRNA molecules which are themselves transcribed by a stochastic process (see text)*

feedback has evolved to compensate for the non-inducibility of the galactose pathway if only *GAL3* is autoregulated. Interestingly, *GAL3* is thought to have evolved from a duplication of *GAL1*, but has only one gal4p dimer binding site [35]. In our model, this results in non-zero *GAL3* expression levels when it is either repressed by gal80p or un-induced, which would be sufficient to enable its activation by incoming galactose, independently of the *GAL80* negative feedback.

Negative feedback has been shown to reduce stochastic noise in biochemical networks [36]. To investigate whether the dual-feedback regulation of *GAL3* and *GAL80* plays a similar role in noise reduction, we produced a variant of our model in which neither *GAL3* nor *GAL80* is transcriptionally regulated. Because the up-regulation of the repressor *GAL80* is counteracted by up-regulation of the activator *GAL3*, the performance of the two versions is similar when simulated using the ODE's; however the noise properties are very different. Figure 5 shows stochastic simulations of a key regulatory protein (gal80p dimer, Fig. 5*a*), and a key intermediate product (galactose 1-phosphate) for the normal and unregulated versions of the model in the 10% induced state. The simulation was carried out for a period of 10 000 minutes to get good statistics; equivalently, one could perform simulations of a number of cells over a shorter time period. Figure 6 is a plot of the dimensionless ratio of standard deviation to mean for the full range of species. The model with regulation shows significantly lower noise for almost all species. Analysis using the estimation tool [37] in Dizzy, which estimates the noise as a weighted sum of the reaction rates, shows that the improvement is due to two effects: negative autoregulation of gal80p, which reduces noise in that key protein, and up-regulation of gal3p and gal80p, which increases the number of molecules taking part in reactions and therefore also reduces noise.

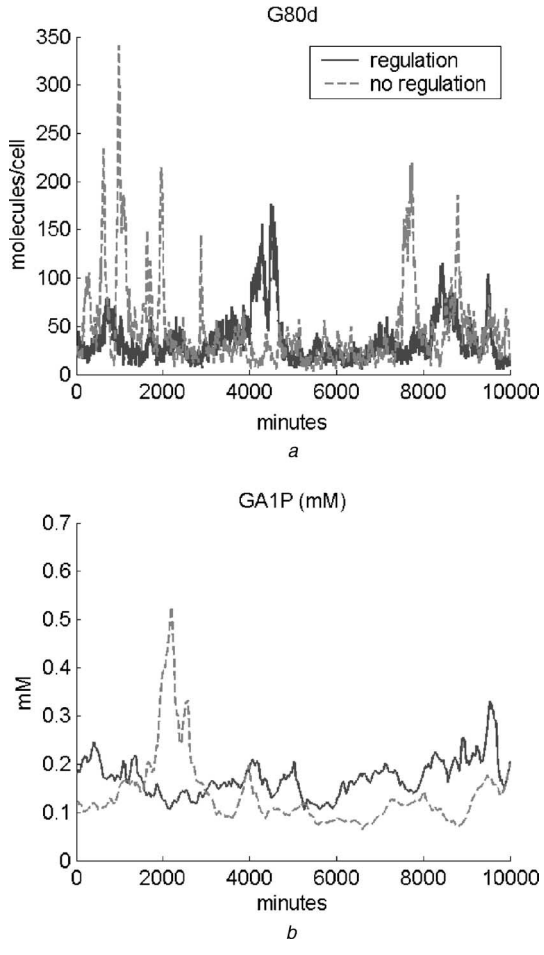

Fig. 5 *Plot of gal80p dimer (a) and galactose 1-phosphate (b) for the regulated and unregulated cases at 10% induction. The unregulated case (dashed line) has higher stochastic noise in both species*

4.3 Multi-site modulation avoids toxic build up of intermediary products

It has been shown that a coordinated and simultaneous increase in the activity of all of the enzymes of a pathway allows an efficient increase of the steady state flux without affecting the level of intermediates, and other connected pathways [38, 39]. This form of flux control is referred to as Multisite modulation [39]. For our simple pathway, this principle predicts that the proportional increase of all enzyme activities produces a proportional increase of flux, with no changes in the concentrations of the intermediaries. An application of this principle to the galactose pathway was reported by Ostergaard *et al.* [40] who obtained an increase in the flux by eliminating the three negative regulators of the *GAL* system (gal6p, gal80p, and mig1p).

Figure 7 shows simulations of the concentrations of internal galactose and galactose 1-phosphate when the system is induced with a large concentration of external galactose (20 g/L). As a result of induction by galactose, transporter and enzyme concentrations increase slowly during the entire simulation time in a coordinate manner. The coordinated-induction case is compared with a hypothetical case where the galactose transporter (gal2p) concentration is fixed to a high level during the entire time-course of the simulation. The effect of fixing gal2p is a change in the balance in supply and demand, with very high levels of supply, and consequently high accumulation of both intermediaries [41]. It is interesting that perfect regulation of internal concentrations is not achieved even for

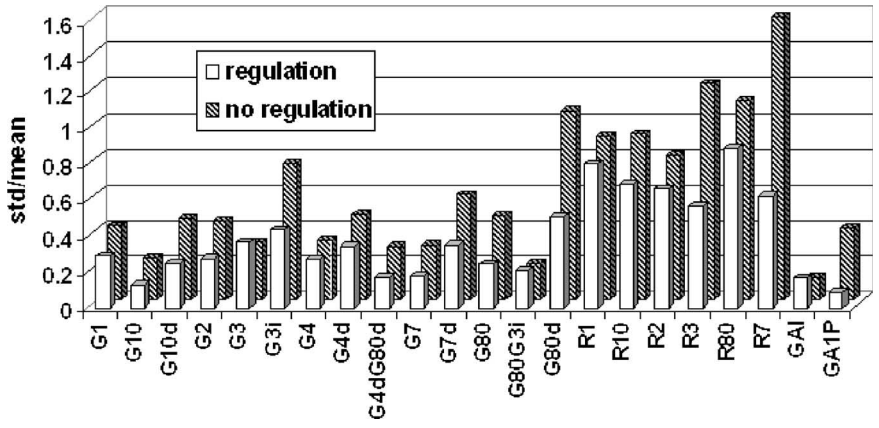

Fig. 6 *Comparison of stochastic variability for the regulated and unregulated versions of the model in the 10% induction state. Shown, is the dimensionless ratio of standard deviation to mean. The system without regulation has significantly higher noise*

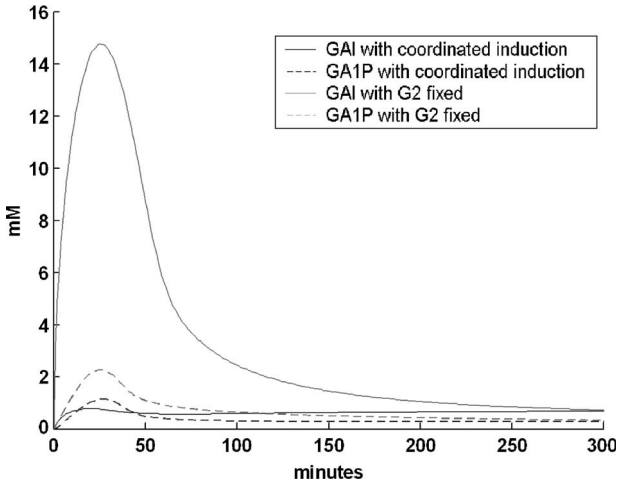

Fig. 7 *Time-course simulation of full induction. Dark grey lines show the complete model, with balanced induction. Light grey lines show the model with gal2p value held constant. This level was fixed to be the same as that reached by simulations of the full coordinate induction model at 500 minutes. Solid lines are internal galactose, dashed lines galactose 1-phosphate*

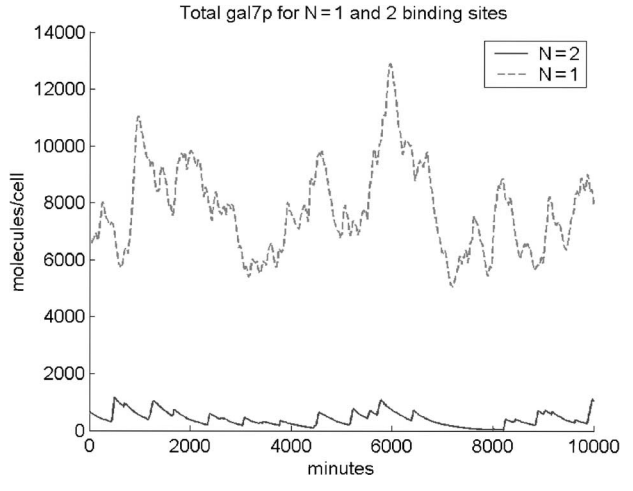

Fig. 8 *Plot of total gal7p protein produced, for $N = 2$ binding sites (usual model), and a modified version with $N = 1$ binding site. The basal output is much higher in the modified version (dashed line). The plot shows how cooperativity in gal80p binding keeps basal output of structural proteins low*

coordinate control, and a low transient accumulation is observed. This may be an example of an evolutionary compromise between multiple system trade-offs. It should be noted that there is a very small constitutive level of galactose transport which is not considered in our model (see Appendix 8.1). The true situation will be more pronounced than that shown for coordinate induction in Fig. 7.

4.4 The number of binding sites per gene reflects the need for silenced vs. basal transcription levels

Multiple binding sites for a single transcription factor on a given target gene are common in higher eukaryotes [42]. An interesting feature of this system is that, while the regulatory genes have only one gal4p dimer binding site, the structural genes have between two and five [18]. There is a clear difference in behaviour between genes with a single site and those with two or more, since synergistic binding of gal80p means that repression is much more effective in the latter case [43]. The regulatory genes therefore have a fairly large basal transcription. This is necessary since both gal3p and gal80p have to be present in the un-induced state in order for the system to sense and respond to external galactose. The structural genes, on the other hand, have a more switch-like behaviour, with only minimal basal transcription in the un-induced state. It should be noted that even with multiple binding sites, the basal expression levels are not absolute zero. Indeed, in our model, the non-zero basal expression of *GAL2* is in fact necessary to initiate induction.

To illustrate this effect, we produced a modified version of the model in which the structural gene *GAL7*p has only one binding site. Figure 8 compares the levels of protein for the two cases, in the non-induced state. The basal output is much higher when the gene has only a single binding site. However, even with multiple binding sites our model predicts significant basal activity, which may be responsible for the transient increase in galactose 1-phosphate during induction (see Section 4.3).

4.5 Other features of the system requiring further study

The transient accumulation in galactose 1-phosphate during induction (see bottom-right panel in Fig. 3 and also Fig. 7), although moderate, may indicate that the model is not complete. Interestingly Greger and Proudfoot [44] showed that *GAL7* is induced about five minutes before *GAL1* and *GAL10*. This could ensure an increase in the consumption of galactose 1-phosphate and a decrease in its concentration at the start of induction. Therefore, this may be a fruitful area for further investigation.

5 Conclusions

Selective evolutionary pressure results in ongoing performance optimisation and favours the emergence of dynamic structures that support complex behaviours. We argued that this evolutionary process results in recurring organisational motifs that may be viewed as evolutionary building blocks and 'design principles'. Computer modelling and simulation offers a well-established framework for the analysis of the dynamic behaviour of systems.

We noted that, based on experimental evidence published over the past 37 years, the yeast galactose utilisation pathway includes a number of organisational motifs, including negative feedback, dimerisation, multi-site modulation, and the use of multiple transcription factor binding sites. We constructed a fairly detailed biochemical model of this relatively well-studied pathway. Our model closely reproduces experimental observations. Simulations of the model showed that these motifs ensure robust system behaviour in spite of random variations in a cell's environment, as well as intrinsic noise in gene expression. Each of the motifs considered have also been observed in other biochemical pathways. The functional advantages we highlighted have also been noted in these pathways. The observation that these organisational motifs recur and confer performance advantages in different pathways, suggests that their repeated occurrence may not be the result of random chance, but rather the recurrent selection of a good 'design principle'.

6 Acknowledgments

We would like to thank Lee Hood and Andrea Weston for sharing unpublished data with us, and also for much useful guidance. We thank Reiko Tanaka for discussions on and help with parameter fitting of the control network to experimental data, Alpana Waghmare for help with data curation and literature searching, Daehee Hwang and Alistair Rust for advice on network structure and *GAL* cis-regulatory analysis, and Lesley Wilkerson for help with the manuscript.

7 References

1 Fell, D.A.: 'Understanding the Control of Metabolism' (Portland Press, London, 1997)
2 Voit, E.O.: 'Computational Analysis of Biochemical Systems. A Practical Guide for Biochemists and Molecular Biologists' (Cambridge University Press, Cambridge, UK, 2000)
3 Barkai, N., and Leibler, S.: 'Robustness in simple biochemical networks', *Nature*, 1997, **387**, (6636), pp. 913–917
4 El-Samad, H., Khammash, M., Kurata, H., and Doyle, J.C.: 'Feedback regulation of the heat shock response in *E. coli*', *Lect. Notes Control Inf. Sci.*, 2003, **289**, pp. 115–128
5 Von Dassow, G., Meir, E., Munro, E.M., and Odell, G.M.: 'The segment polarity network is a robust developmental module', *Nature*, 2000, **406**, (6792), pp. 188–192
6 Von Dassow, G., and Odell, G.M.: 'Design and constraints of the Drosophila segment polarity module: robust spatial patterning emerges from intertwined cell state switches', *J. Exp. Zool.*, 2002, **294**, (3), pp. 179–215
7 Davidson, E.H., Rast, J.P., Oliveri, P., Ransick, A., Calestani, C., Yuh, C.H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C.T., Livi, C.B., Lee, P.Y., Revilla, R., Rust, A.G., Pan, Z., Schilstra, M.J., Clarke, P.J., Arnone, M.I., Rowen, L., Cameron, R.A., McClay, D.R., Hood, L., and Bolouri, H.: 'A genomic regulatory network for development', *Science*, 2002, **295**, (5560), pp. 1669–1678
8 Raya, A., Kawakami, Y., Rodriguez-Esteban, C., Ibanes, M., Rasskin-Gutman, D., Rodriguez-Leon, J., Buscher, D., Feijo, J.A., and Izpisua Belmonte, J.C.: 'Notch activity acts as a sensor for extracellular calcium during vertebrate left-right determination', *Nature*, 2004, **427**, (6970), pp. 121–128
9 Wolfe, M.F.: 'Rube Goldberg: Inventions!' (Simon & Schuster, London, 2000)
10 Beare, G.: 'Heath Robinson: Machines and Inventions' (Chris Beetles Ltd., London, 1992)
11 Lienhard, J.H.: 'The engines of our ingenuity: an engineer looks at technology and culture' (Oxford University Press, Oxford, New York, 2000)
12 Petroski, H.: 'Design paradigms: case histories of error and judgment in engineering' (Cambridge University Press, Cambridge [England], New York, NY, 1994)
13 Nelson, W.J., and Nusse, R.: 'Convergence of Wnt, beta-catenin, and cadherin pathways', *Science*, 2004, **303**, (5663), pp. 1483–1487
14 Hartwell, L.H., Hopfield, J.J., Leibler, S., and Murray, W.M.: 'From molecular to modular cell biology', *Nature*, 1999, **402**, (supp. 2), pp. C47–C52
15 Frey, P.A.: 'The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose', *FASEB J.*, 1996, **10**, (4), pp. 461–470
16 Holden, H.M., Rayment, I., and Thoden, J.B.: 'Structure and function of enzymes of the Leloir pathway for galactose metabolism', *J. Biol. Chem.*, 2003, **278**, (45), pp. 43885–43888
17 Lohr, D., Venkov, P., and Zlatanova, J.: 'Transcriptional regulation in the yeast GAL gene family: a complex genetic network', *FASEB J.*, 1995, **9**, (9), pp. 777–787
18 Melcher, K.: 'Galactose metabolism in Saccharomyces cerevisiae: a paradigm for eukaryotic gene regulation', in Zimmermann, F.K., and Entian, K.D. (Eds.): 'Yeast Sugar Metabolism' (Technomic Publishing Inc., Lancaster, PA, 1997), pp. 235–269
19 Bhat, P.J., and Murthy, T.V.: 'Transcriptional control of the GAL/MEL regulon of yeast Saccharomyces cerevisiae: mechanism of galactose-mediated signal transduction', *Mol. Microbiol.*, 2001, **40**, (5), pp. 1059–1066
20 Biggar, S.R., and Crabtree, G.R.: 'Cell signaling can direct either binary or graded transcriptional responses', *EMBO J.*, 2001, **20**, (12), pp. 3167–3176
21 Li, J., Wang, S., Vandusen, W.J., Schultz, L.D., George, H.A., Herber, W.K., Chae, H.J., Bentley, W.E., and Rao, G.: 'Green fluorescent protein in Saccharomyces cerevisiae: real-time studies of the GAL1 promoter', *Biotechnol. Bioeng.*, 2000, **70**, (2), pp. 187–196
22 Gibson, M.A., and Mjolsness, E.: 'Modeling the activity of single genes', in Bower, J.M., and Bolouri, H. (Eds.): 'Computational modeling of genetic and biochemical networks' (MIT Press, Cambridge, MA, 2001), Chapter 1, pp. 3–48
23 Mehta, D.V., Kabir, A., and Bhat, P.J.: 'Expression of human inositol monophosphatase suppresses galactose toxicity in Saccharomyces cerevisiae: possible implications in galactosemia', *Biochim. Biophys. Acta*, 1999, **1454**, (3), pp. 217–226
24 Lai, K., and Elsas, L.J.: 'Overexpression of human UDP-glucose pyrophosphorylase rescues galactose-1-phosphate uridyltransferase-deficient yeast', *Biochem. Biophys. Res. Commun.*, 2000, **271**, (2), pp. 392–400
25 Ostergaard, S., Olsson, L., and Nielsen, J.: 'In vivo dynamics of galactose metabolism in Saccharomyces cerevisiae: metabolic fluxes and metabolite levels', *Biotechnol. Bioeng.*, 2001, **73**, (5), pp. 412–425
26 Lai, K., Langley, S.D., Khwaja, F.W., Schmitt, E.W., and Elsas, L.J.: 'GALT deficiency causes UDP-hexose deficit in human galactosemic cells', *Glycobiology*, 2003, **13**, (4), pp. 285–294
27 Bhat, P.J.: 'Galactose-1-phosphate is a regulator of inositol monophosphatase: a fact or a fiction?', *Med. Hypotheses*, 2003, **60**, (1), pp. 123–128
28 Gillespie, D.T., and Petzold, L.R.: 'Improved leap-size selection for accelerated stochastic simulation', *J. Chem. Phys.*, 2003, **119**, (16), pp. 8229–8234
29 Gillespie, D.T.: 'A general method for numerically simulating the stochastic time evolution of coupled chemical reactions', *J. Comp. Phys.*, 1976, **22**, (4), pp. 403–434
30 Bundschuh, R., Hayot, F., and Jayaprakash, C.: 'The role of dimerization in noise reduction of simple genetic networks', *J. Theor. Biol.*, 2003, **220**, (2), pp. 261–269
31 Van Kampen, N.G.: 'Stochastic processes in physics and chemistry', (North-Holland, Amsterdam, 1992)
32 Orrell, D., and Bolouri, H.: 'Control of internal and external noise in genetic regulatory networks', *J. Theor. Biol.*, in press
33 Shimada, H., and Fukasawa, T.: 'Controlled transcription of the yeast regulatory gene GAL80', *Gene*, 1985, **39**, (1), pp. 1–9
34 Bajwa, W., Torchia, T.E., and Hopper, J.E.: 'Yeast regulatory gene GAL3: carbon regulation; UASGal elements in common with GAL1, GAL2, GAL7, GAL10, GAL80, and MEL1; encoded protein strikingly similar to yeast and Escherichia coli galactokinases', *Mol. Cell. Biol.*, 1988, **8**, (8), pp. 3439–3447
35 Verma, M., Bhat, P.J., and Venkatesh, K.V.: 'Quantitative analysis of GAL genetic switch of Saccharomyces cerevisiae reveals that nucleocytoplasmic shuttling of Gal80p results in a highly sensitive response to galactose', *J. Biol. Chem.*, 2003, **278**, (49), pp. 48764–48769
36 Becskei, A., and Serrano, L.: 'Engineering stability in gene networks by autoregulation', *Nature*, 2000, **405**, (6786), pp. 590–593
37 Orrell, D., Ramsey, S., de Atauri, P., and Bolouri, H.: 'A method for estimating stochastic noise in large genetic regulatory networks', *Bioinformatics*, Submitted
38 Kacser, H., and Acerenza, L.: 'A universal method for achieving increases in metabolite production', *Eur. J. Biochem.*, 1993, **216**, (2), pp. 361–367
39 Fell, D.A., and Thomas, S.: 'Physiological control of metabolic flux: the requirement for multisite modulation', *Biochem. J.*, 1995, **311**, (Pt 1), pp. 35–39
40 Ostergaard, S., Olsson, L., Johnston, M., and Nielsen, J.: 'Increasing galactose consumption by Saccharomyces cerevisiae through metabolic

engineering of the GAL gene regulatory network', *Nat. Biotechnol.*, 2000, **18**, (12), pp. 1283–1286
41 Hofmeyr, J.S., and Cornish-Bowden, A.: 'Regulating the cellular economy of supply and demand', *FEBS Lett.*, 2000, **476**, (1-2), pp. 47–51
42 Davidson, E.H.: 'Genomic Regulatory Systems: Development and Evolution' (Academic, San Diego, CA, 2001)
43 Melcher, K., and Xu, H.E.: 'Gal80-Gal80 interaction on adjacent Gal4p binding sites is required for complete GAL gene repression', *EMBO J.*, 2001, **20**, (4), pp. 841–851
44 Greger, I.H., and Proudfoot, N.J.: 'Poly(A) signals control both transcriptional termination and initiation between the tandem GAL10 and GAL7 genes of Saccharomyces cerevisiae', *EMBO J.*, 1998, **17**, (16), pp. 4771–4779
45 Guthrie, C., and Fink, G.R. (Eds.): 'Guide to Yeast Genetics and Molecular Biology' (Academic Press, Inc., San Diego, CA, 1991), p. 17
46 Ramos, J., Szkutnicka, K., and Cirillo, V.P.: 'Characteristics of galactose transport in Saccharomyces cerevisiae cells and reconstituted lipid vesicles', *J. Bacteriol.*, 1989, **171**, (6), pp. 3539–3544
47 Reifenberger, E., Boles, E., and Ciriacy, M.: 'Kinetic characterization of individual hexose transporters of Saccharomyces cerevisiae and their relation to the triggering mechanisms of glucose repression', *Eur. J. Biochem.*, 1997, **245**, (2), pp. 324–333
48 Teusink, B., Diderich, J.A., Westerhoff, H.V., Van Dam, K., and Walsh, M.C.: 'Intracellular glucose concentration in derepressed yeast cells consuming glucose is high enough to reduce the glucose transport rate by 50%', *J. Bacteriol.*, 1998, **180**, (3), pp. 556–562
49 Kotyk, A.: 'Mobility of the free and of the loaded monosaccharide carrier in Saccharomyces cerevisiae', *Biochim. Biophys. Acta*, 1967, **135**, (1), pp. 112–119
50 Schell, M.A., and Wilson, D.B.: 'Purification and properties of galactokinase from Saccharomyces cerevisiae', *J. Biol. Chem.*, 1977, **252**, (4), pp. 1162–1166
51 Segawa, T., and Fukasawa, T.: 'The enzymes of the galactose cluster in Saccharomyces cerevisiae. Purification and characterization of galactose-1-phosphate uridylyltransferase', *J. Biol. Chem.*, 1979, **254**, (21), pp. 10707–10709
52 Fukasawa, T., Obonai, K., Segawa, T., and Nogi, Y.: 'The enzymes of the galactose cluster in Saccharomyces cerevisiae. II. Purification and characterization of uridine diphosphoglucose 4-epimerase', *J. Biol. Chem.*, 1980, **255**, (7), pp. 2705–2707
53 Dey, P.M.: 'UDP-galactose 4′-epimerase from Vicia faba seeds', *Phytochemistry*, 1984, **23**, (4), pp. 729–732
54 Majumdar, S., Ghatak, J., Mukherji, S., Bhattacharjee, H., and Bhaduri, A.: 'UDPgalactose 4-epimerase from Saccharomyces cerevisiae. A bifunctional enzyme with aldose 1-epimerase activity', *Eur. J. Biochem.*, 2004, **271**, (4), pp. 753–759
55 Christacos, N.C., Marson, M.J., Wells, L., Riehman, K., and Fridovich-Keil, J.L.: 'Subcellular localization of galactose-1-phosphate uridylyltransferase in the yeast Saccharomyces cerevisiae', *Mol. Genet. Metab.*, 2000, **70**, (4), pp. 272–280
56 Timson, D.J., and Reece, R.J.: 'Kinetic analysis of yeast galactokinase: implications for transcriptional activation of the GAL genes', *Biochimie*, 2002, **84**, (4), pp. 265–272
57 Wang, Y., Liu, C.L., Storey, J.D., Tibshirani, R.J., Herschlag, D., and Brown, P.O.: 'Precision and functional specificity in mRNA decay', *Proc. Natl. Acad. Sci. USA*, 2002, **99**, (9), pp. 5860–5865
58 Iyer, V., and Struhl, K.: 'Absolute mRNA levels and transcriptional initiation rates in Saccharomyces cerevisiae', *Proc. Natl. Acad. Sci. USA*, 1996, **93**, (11), pp. 5208–5212
59 Lashkari, D.A., Derisi, J.L., Mccusker, J.H., Namath, A.F., Gentile, C., Hwang, S.Y., Brown, P.O., and Davis, R.W.: 'Yeast microarrays for genome wide parallel genetic and gene expression analysis', *Proc. Natl. Acad. Sci. USA*, 1997, **94**, (24), pp. 13057–13062
60 Bolouri, H., and Davidson, E.H.: 'Transcriptional regulatory cascades in development: initial rates, not steady state, determine network kinetics', *Proc. Natl. Acad. Sci. USA*, 2003, **100**, (16), pp. 9371–9376
61 Pratt, J.M., Petty, J., Riba-Garcia, I., Robertson, D.H., Gaskell, S.J., Oliver, S.G., and Beynon, R.J.: 'Dynamics of protein turnover, a missing dimension in proteomics', *Mol. Cell. Proteomics*, 2002, **1**, (8), pp. 579–591
62 Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S.: 'Global analysis of protein expression in yeast', *Nature*, 2003, **425**, (6959), pp. 737–741
63 Carey, M., Kakidani, H., Leatherwood, J., Mostashari, F., and Ptashne, M.: 'An amino-terminal fragment of GAL4 binds DNA as a dimer', *J. Mol. Biol.*, 1989, **209**, (3), pp. 423–432
64 Rodgers, K.K., and Coleman, J.E.: 'DNA binding and bending by the transcription factors GAL4(62*) and GAL4(149*)', *Protein Sci.*, 1994, **3**, (4), pp. 608–619
65 Kang, T., Martins, T., and Sadowski, I.: 'Wild type GAL4 binds cooperatively to the GAL1-10 UASG in vitro', *J. Biol. Chem.*, 1993, **268**, (13), pp. 9629–9635
66 Giniger, E., and Ptashne, M.: 'Cooperative DNA binding of the yeast transcriptional activator GAL4', *Proc. Natl. Acad. Sci. USA*, 1988, **85**, (2), pp. 382–386
67 Xu, H.E., Kodadek, T., and Johnston, S.A.: 'A single GAL4 dimer can maximally activate transcription under physiological conditions', *Proc. Natl. Acad. Sci. USA*, 1995, **92**, (17), pp. 7677–7680
68 Lohr, D., and Lopez, J.: 'GAL4/GAL80-dependent nucleosome disruption/deposition on the upstream regions of the yeast GAL1-10 and GAL80 genes', *J. Biol. Chem.*, 1995, **270**, (46), pp. 27671–27678
69 Timson, D.J., Ross, H.C., and Reece, R.J.: 'Gal3p and Gal1p interact with the transcriptional repressor Gal80p to form a complex of 1:1 stoichiometry', *Biochem. J.*, 2002, **363**, (Pt 3), pp. 515–520
70 Leuther, K.K., and Johnston, S.A.: 'Nondissociation of GAL4 and GAL80 in vivo after galactose induction', *Science*, 1992, **256**, (5061), pp. 1333–1335
71 Parthun, M.R., and Jaehning, J.A.: 'A transcriptionally active form of GAL4 is phosphorylated and associated with GAL80', *Mol. Cell. Biol.*, 1992, **12**, (11), pp. 4981–4987
72 Platt, A., and Reece, R.J.: 'The yeast galactose genetic switch is mediated by the formation of a Gal4p-Gal80p-Gal3p complex', *EMBO J.*, 1998, **17**, (14), pp. 4086–4091
73 Peng, G., and Hopper, J.E.: 'Evidence for Gal3p's cytoplasmic location and Gal80p's dual cytoplasmic-nuclear location implicates new mechanisms for controlling Gal4p activity in Saccharomyces cerevisiae', *Mol. Cell. Biol.*, 2000, **20**, (14), pp. 5140–5148
74 Peng, G., and Hopper, J.E.: 'Gene activation by interaction of an inhibitor with a cytoplasmic signaling protein', *Proc. Natl. Acad. Sci. USA*, 2002, **99**, (13), pp. 8548–8553
75 Venkatesh, K.V., Bhat, P.J., Kumar, R.A., and Doshi, P.: 'Quantitative model for Gal4p-mediated expression of the galactose/melibiose regulon in Saccharomyces cerevisiae'', *Biotechnol. Prog.*, 1999, **15**, (1), pp. 51–57
76 Lue, N.F., Chasman, D.I., Buchman, A.R., and Kornberg, R.D.: 'Interaction of GAL4 and GAL80 gene regulatory proteins in vitro', *Mol. Cell. Biol.*, 1987, **7**, (10), pp. 3446–3451
77 Vollenbroich, V., Meyer, J., Engels, R., Cardinali, G., Menezes, R.A., and Hollenberg, C.P.: 'Galactose induction in yeast involves association of Gal80p with Gal1p or Gal3p', *Mol. Gen. Genet.*, 1999, **261**, (3), pp. 495–507
78 Platt, A., Ross, H.C., Hankin, S., and Reece, R.J.: 'The insertion of two amino acids into a transcriptional inducer converts it into a galactokinase', *Proc. Natl. Acad. Sci. USA*, 2000, **97**, (7), pp. 3154–3159
79 Gancedo, J.M., Lopez, S., and Ballesteros, F.: 'Calculation of half-lives of proteins in vivo. Heterogeneity in the rate of degradation of yeast proteins', *Mol. Cell. Biochem.*, 1982, **43**, (2), pp. 89–95
80 Ideker, T., Thorsson, V., Ranish, J.A., Christmas, R., Buhler, J., Eng, J.K., Bumgarner, R., Goodlett, D.R., Aebersold, R., and Hood, L.: 'Integrated genomic and proteomic analyses of a systematically perturbed metabolic network', *Science*, 2001, **292**, (5518), pp. 929–934
81 Laughon, A., and Gesteland, R.F.: 'Isolation and preliminary characterization of the GAL4 gene, a positive regulator of transcription in yeast', *Proc. Natl. Acad. Sci. USA*, 1982, **79**, (22), pp. 6827–6831

8 Appendix

8.1 Details of the model

The three parts of the model, '**Enzymatic pathway**', '**Transcription and translation of structural genes**', and '**Control network**', were modelled separately and then joined. Although some parameters are introduced using mM as units of concentration, the complete model is computed in 'molecules per cell' (m/c). The conversion between m/c and mM is done by assuming 2.38 mL of cell volume per gram of cell dry weight [25], and a cell dry weight of 15×10^{-12} g per haploid cell [45].

8.1.1 Enzymatic pathway

Transport: Galactose transport can be divided in three components: two galactose inducible transport processes, a high-affinity process and a low-affinity process, both dependent of gal2p; one residual constitutive low affinity transport, independent of gal2p [46, 47]. The mechanism of the inducible two-component transport is unknown. Only the high-affinity component is assumed in this model and an equation describing transport as a symmetric facilitated diffusion process is applied [48]:

$$\begin{aligned} \mathrm{v\_TR} = \mathrm{k\_TR} \cdot \mathrm{G2} \Big((\mathrm{GAE}/\mathrm{Km\_TR} - \mathrm{GAI}/\mathrm{Km\_TR}) \Big/ \\ (1 + \mathrm{GAE}/\mathrm{Km\_TR} + \mathrm{GAI}/\mathrm{Km\_TR} \\ + \alpha\mathrm{\_TR} \cdot \mathrm{GAE} \cdot \mathrm{GAI}/\mathrm{Km\_TR}^2) \Big) \end{aligned} \quad (1)$$

Km_TR is 1 mM for the inducible high affinity transport [47]; α_TR is the *interactive constant* [48], with a value of 1 [49]; k_TR was adjusted to 4350 min^{-1}. This adjustment

provided at GAE 0.5 mM a value of GA1P in the range of the values provided by Ostergaard *et al.* [25].

Galactokinase: Galactokinase is a monomeric protein [50]. The reaction catalysed by it depends on ATP, but because ATP is not a dependent variable in the model, we assume that the galactokinase reaction is only dependent on GAI:

$$v\_GK = kcat\_GK \cdot G1 \cdot \frac{GAI}{Km\_GK + GAI} \quad (2)$$

kcat_GK is 3350 min^{-1} and Km_GK is 0.6 mM [50]. The kinetic measurements correspond to apparent values with ATP concentration fixed.

Galactose-1-phosphate uridylyltransferase: The reaction catalysed by galactose-1-phosphate uridylyltransferase, depends on a dimeric protein [51], and has been described to follow a ping-pong mechanism [15]. We describe it with the following equation:

$$v\_TF = kcat\_TF \cdot G7d \cdot \frac{GA1P \cdot UGL}{Km\_ga1p\_TF \cdot UGL + Km\_ugl\_TF \cdot GA1P + GA1P \cdot UGL} \quad (3)$$

kcat_TF is 59 200 min^{-1} (per enzyme molecule; dimer), Km_ga1p_TF (Km for GA1P) is 4.0 mM, and Km_ugl_TF (Km for UGL) is 0.26 mM [51].

UDP-galactose 4-epimerase: The reaction catalsed by UDP-galactose 4-epimerase, depends on a dimeric protein [52], which transforms UGA in UGL in a reversible manner. We describe it with the following equation, which models the reversibility of the reaction:

$$v\_EP = kcat\_EP \cdot G10d \cdot \frac{(1/Km\_uga\_EP) \cdot (UGA - UGL/Keq\_EP)}{1 + UGA/Km\_uga\_EP + UGL/Km\_ugl\_EP} \quad (4)$$

kcat_EP is 3890 min^{-1} (per enzyme molecule; dimer), Km_uga_EP (Km for UGA) is 0.22 mM, and Keq_EP is 3.5 [52]. Keq_EP refers to the ratio UGL/UGA [52]. Km_ugl_EP (Km for UGL) is 0.25 mM, corresponding to the value in *Vicia faba* [53].

The value for the conservation UGL + UGA is set to that reported by Lai *et al.* [26] in fibroblasts. kcat_EP was increased ten-fold (kcat_EP = 38 900 min^{-1}) in order to change the ratio UGL/UGA to a value closer to that reported by Lai *et al.* [26].

The dimerisation rates for G7 and G10 were assumed to be equal to that of G80 ((4) in Table 3).

Our description of the galactose uptake at the level of enzymes does not include the following points: 1) GL1P is transformed in glucose 6-phosphate by the action of phosphoglucomutase and with UTP is transformed to UGL and PPi by the action of UDP-glucose pyrophosphorylase [15]; 2) UDP-galactose 4-epimerase has been shown to be a bi-functional enzyme with aldolase 1-epimerase (mutarotase) activity [54]; 3) Christacos *et al.* [55] have suggested substrate/product channelling or other interactions that could alter the kinetics of the enzymes, based on the experimental evidence of gal7p subcellular localisation, which is dependent upon coexpression of gal1p and gal10p; 4) The equation describing the galactokinase catalysed reaction does not allow the use of the model to simulate the expected high accumulation of GA1P in galactose-1-phosphate uridylyltransferase deficient strains. The reasons for this are: a) GA1P is a product inhibitor at very high concentrations [50, 56]; b) UDP-glucose pyrophosphorylase has been shown also to be a multi-functional enzyme that catalyses the conversion of GA1P to UGA [24, 26], and although its activity is very low compared with this for galactose-1-phosphate uridylyltransferase [26], we can not evaluate this effect on GA1P under galactose-1-phosphate uridylyltransferase deficiency.

8.1.2 Transcription and translation of structural genes:

The transcription of the structural genes is described in Table 2 ((5) to (8)). The degradation rates kdr_struct and kdr_2 are equal to the summation of two components, one intrinsic degradation rate of the RNA and the other the dilution rate that accounts for cell growth. A dilution rate of 0.097 $hour^{-1}$ is used, and corresponds to that for the chemostat continuous culture used by [25] to study the steady state at GAE 0.5 mM. The intrinsic degradation rate components are deduced from the poly(A) half-life estimations from Wang *et al.* [57]. From their values we assume the same half-life of 22 min for *GAL1*, *-7*, and *-10*, and 49 min for *GAL2*. The maximum initiation rate of transcription, kir_struct and kir_2, are

Table 3: Optimised association and dissociation rate constants

Interaction	k_f	k_r	Equations	
Sugar-Protein				
$G3 + GAI \underset{k_r}{\overset{k_f}{\rightleftharpoons}} G3i$	4×10^{-7}	890	$v\_G3i = k_f \cdot G3 \cdot GAI - k_r \cdot G3i$	(1)
Protein-DNA				
$DNA + G4d \underset{k_r}{\overset{k_f}{\rightleftharpoons}} DNA\text{-}G4d$	0.1	1.1	$K_P = k_f/k_r$	(2)
Protein-Protein				
$G4 + G4 \underset{k_r}{\overset{k_f}{\rightleftharpoons}} G4d$	0.1	1	$v\_G4d = 2 \cdot k_f \cdot G4 \cdot G4 - 2 \cdot k_r \cdot G4d$	(3)
$G80 + G80 \underset{k_r}{\overset{k_f}{\rightleftharpoons}} G80d$	0.1	170	$v\_G80d = 2 \cdot k_f \cdot G80 \cdot G80 - 2 \cdot k_r \cdot G80d$	(4)
$G4d + G80d \underset{k_r}{\overset{k_f}{\rightleftharpoons}} G4dG80d$	0.1	1.8	$v\_G4dG80d = k_f \cdot G4d \cdot G80d - k_r \cdot G4dG80d$	(5)
			$K_Q = k_f/k_r$	(6)
$G80 + G3i \underset{k_r}{\overset{k_f}{\rightleftharpoons}} G80G3i$	0.1	0.03	$v\_G80G3i = k_f \cdot G80 \cdot G3i - k_r \cdot G80G3i$	(7)

k_f units are $(molecules/cell)^{-1}$ min^{-1}, and k_r units are min^{-1}

adjusted to approximate reported steady-state values of R1, R2, R7 and R10. The model is fit to provide for structural genes a value of around 30 m/c of mRNA under full induction conditions and a very low basal expression level under non-induction conditions. Iyer and Struhl [58] have reported a value R1 of 33 m/c, under conditions of high galactose concentration, which we assume as full induction. This level of induction should be similar for the other R2, R7 and R10, as deduced from the similar induction observed when comparing results from microarray experiments of steady-state induction (galactose) *versus* repression (glucose) [59].

The level of transcriptional induction is a function of galactose induction, but also a function of the number of binding sites for gal4p. Regulatory genes in the galactose pathway have only one binding site, but structural genes have multiple binding sites [43], with the exception of the structural gene *MEL1*, which is also a gal4p regulated gene. Our model assumes two binding sites for *GAL7*, four for *GAL1* and *GAL10* (these four binding sites are in a shared promoter region), and five for *GAL2* [18]. Synergistic or stabilising effects, depending on the number of binding sites, have been suggested [43] for the binding of the repressor gal80p. Such stabilisation could be due to gal80p dimerisation and gal80p dimer-dimer interactions, which could in turn stabilise gal4p-gal80p interactions. This has been proposed to explain the low basal expression of both *GAL3* and *GAL80* [43], both of which have a single gal4p binding site, and the more efficient repression of *GAL* genes with multiple gal4p binding sites. Melcher and Xu [43] discount cooperative binding of gal80p dimers, and favour a mechanism in which gal80p multimerisation reduces accessibility of the gal4p-gal80p complexes to the transcriptional machinery. Our model emulates this stabilisation effect by changing the association rate of the binding of gal80p dimers to the gal4p-DNA complex by an arbitrarily large but plausible cooperativity factor of 30 (see [60], and Appendix 8.2).

The protein translation process is described in Table 2 ((9), (10), (11), and (13)). The degradation rates, as for transcription, are equal to the summation of two components, the intrinsic degradation rate of the proteins and the dilution rate. The intrinsic degradation rate component is assumed to be 0.022 $hour^{-1}$ for all structural genes, which is the measured average for 50 proteins in yeast [61]. The initiation rates are adjusted to provide specific steady-state values for the protein-mRNA ratios (G/R). Recently, Ghaemmaghami *et al.* [62], through a global analysis of proteins in yeast, have shown a significant relationship between mRNA levels and protein levels, although individual genes with equivalent mRNA levels can have large differences in protein abundances. From the supplementary material that they provide it is possible to estimate a global G2/R2 of $\sim$3500 for gal2p, as this is the mean value for the functional category of 'Transport facilitation'. They provide a value of G/R of $\sim$5000 for the functional category of 'Metabolism', which we apply to gal1p, gal7p and gal10p. It should be noted that our protein degradation rates and the ratio of Protein/DNA are based on general measures of *Saccharomyces cerevisiae*, not specifically for *GAL* products.

8.1.3 Control network

Simplified picture of protein-DNA and protein-protein interactions: The simplified picture of protein-DNA and protein-protein interactions that we provide is based in the following observation: (1) gal4p dimerises and binds DNA as a dimer (G4d); (2) gal80p dimerises (G80d) and forms a complex with the gal4p dimer; (3) gal3p and gal80p form a complex containing one molecule of gal3p (G3) and one molecule of gal80p (G80). (4) gal3p is activated (G3i) by interacting with the internal galactose (GAI). These assumptions and other general considerations are discussed below.

The nature of G4 interactions: Gal4p binds DNA as a dimer and dimerises in solution [63, 64]. The association and dissociation rate constants for DNA binding and dimerisation are in Table 3 ((2) and (3)). However, some points are not reflected in the model: 1) affinity changes depending on the binding site, as Kang *et al.* [65] have shown for the shared *GAL1-GAL10* promoter; 2) a phenomenon of cooperativity of the binding of gal4p in adjacent sites has been described [65–67], but shown to depend on the affinity of the binding site [66, 67]; 3) the role of chromatin remodelling is not included in our model, but it has been shown that the *TATA* or *transcription start sites* for *GAL1, -10*, *-80* genes are nucleosome blocked, depending on gal4p for disruption and gal80p for nucleosome reorganisation [68].

Direct or indirect induction? Gal3p and gal80p, in the presence of galactose and ATP, form a complex containing one molecule of gal3p and one molecule of gal80p [69]. However, the exact nature of such an interaction *in vivo* remains unclear. On the one hand it has been proposed that the gal4p-gal80p complex associated with the DNA does not dissociate after galactose induction [70–72]. Platt and Reece [72] showed that *in vitro* induction involves the formation of a ternary protein complex composed of gal80p, gal4p, and gal3p, which in turn is proposed to activate transcription. A different proposal is that gal3p induces transcription indirectly, by binding with the repressor gal80p, which dissociates from the gal80p-gal4p complex, releasing active gal4p [73, 74]. The hypothesis of a ternary complex implies that gal3p and gal80p are located in the nucleus. However, Peng and Hopper [73, 74] have suggested, based on *in vivo* localisation experiments, that gal80p can be localised to both the nucleus and cytoplasm, but gal3p is located in the cytoplasm and there it cannot interact with the gal80p-gal4p-DNA complex, but rather initiates induction via interaction with gal80p in the cytoplasm, modulating the gal80p–gal4p interaction.

Both the ternary and non-ternary scenarios have been previously modelled [35, 75] by considering their behaviour at equilibrium. Our model assumes the indirect and non-ternary hypothesis, although we do not model the transport of gal80p between the cytoplasm and nucleus. Peng and Hopper [73] have suggested a rapid and efficient nuclear import of gal80p and Verma *et al.* [35] have shown, through mathematical modelling, that the shuttling of gal80p is a key step for a highly sensitive response to the inducer. While such effects could in principle be modelled, at the dynamical level the complexity of the model rapidly exceeds our understanding of the mechanism, and the result will be extremely dependent on the assumptions. In our approach, the simple way is chosen, and all interactions are modelled as occurring in a single volume. Reaction rates were based on bibliographical values; optimisation techniques were used to adjust rates where necessary, and within acceptable limits, to give properties which agree with the observed system. The resulting association and dissociation rate constants affecting all protein-protein interactions, and also DNA-protein, are provided in Table 3 ((2), (3), (4) (5) and (7)). The dissociation constant K_D for the complex G4dG80d is equivalent to 0.85×10^{-9} M, close to the 0.3×10^{-9} M reported by Melcher & Xu [43], or the 5×10^{-9} M reported by Lue *et al.* [76]. However, the

affinity of G80 proteins to produce a dimer is much lower (7.9×10^{-8} M) than the reported by Melcher and Xu [43] ($1 - 3 \times 10^{-10}$ M). This discrepancy reflects the fact that our model is a simple representation of a complex phenomenon, where we try to keep only the essential features and where the reproduction of system behaviour is essentially a phenomenological approximation of a simple and poorly understood process. Finally, regarding the interaction of gal4p with its promoter binding site, our computed dissociation constant K_D is 0.53×10^{-9} M, close of the 1.3×10^{-9} M reported by Melcher and Xu [43].

Sugar-protein interaction? The closely related yeast *Kluyveromyces lactis* lacks a gal3p homolog and contains a single galactokinase-like molecule that functions both as a galactokinase and as transcriptional inducer [77]. In *Saccharomyces cerevisiae*, the protein responsible for galactokinase activity, gal1p, is a bi-functional protein that also can induce *GAL* gene expression like gal3p, although the inducer role of gal1p is neglected in the model, because approximately 40-fold more gal1p than gal3p is required to activate the *GAL* genetic switch *in vitro* [72] and the level of approximation of our model can not account for this detail. It has been suggested [56, 78] that the binding of galactose, and also ATP, to gal3p and gal1p induces a conformational change in the proteins and this conformational change is required to promote association with gal80p. Platt *et al.* [78] have shown that the insertion of just two amino acids from gal1p into the corresponding region of gal3p confers galactokinase activity onto the resultant protein, as gal1p. This observation supports the hypothesis of a sugar-protein interaction as the phenomenon that induces gal3p. We assume a simple equilibrium among galactose and gal3p reproducing a sugar-protein interaction. The association and dissociation rate constants have been selected ((1) in Table 3) to allow a 10% induction at GAE 0.5 mM and full induction at 111 mM, with around 80% at 16.7 mM.

Transcription and translation: The transcription and translation for *GAL3* and -80 are described in Table 2 ((15), (16), (19) and (21)). We assume that *GAL3* and *GAL80* as autoregulated genes have a high turnover. This assumption is based on the observations of Gancedo *et al.* [79], which measured protein decay in *Saccharomyces cerevisiae* and showed that proteins can be divided into two groups of 'high' and 'low' turnover. The kinetic parameters are adjusted as: degradation rate kdr_reg $0.16\,\mathrm{min}^{-1}$; maximum initiation rates kir_3 $2.9\,m/c\,min^{-1}$ and kir_reg $1.4\,m/c\,min^{-1}$; degradation rate kdp_reg $0.0037\,\mathrm{min}^{-1}$; and initiation rate kip_reg $18\,\mathrm{min}^{-1}$. The higher maximum initiation rate for *GAL3* (kir_3) satisfies the observation that gal3p is more abundant than gal80p in the cell [74]. In agreement with microarray experiments comparing un-induced (raffinose) and induced (galactose) states [80], and other observations introduced above, the parameters provided satisfy a basal expression for these two regulatory genes under non-induction/non-repression, compared with the very low expression for the structural genes, and a moderate expression under induction with respect to structural genes. Also, the timescale of several minutes for increase of mRNA levels after induction (see examples in Fig. 3) approximates experimental time-course observations (Hood and Weston; personal communication). Finally, the ratio Protein/RNA achieved at steady state is 4800 for both *GAL3* and *GAL80*, corresponding to the measured mean ratio of protein to mRNA molecules [62].

The expression of *GAL4* is not regulated by itself and was estimated by Laughon and Gesteland [81] to be almost identical in galactose-induced and non-induced cells. The translation of gal4p is introduced ((17) in Table 2), with the same degradation rate kdp_reg as *GAL3* and *-80*, and an initiation rate kip_4 of $0.86\,\mathrm{m/c\,min}^{-1}$, adjusted to provide a total gal4p value of 230 m/c [35].

8.2 Derivation of the parameterised fractional saturation function for genes with multiple gal4p dimer binding sites

Below are the fractional saturation functions for different cases of promotion/repression. In the galactose model, P represents the gal4p dimer, and Q represents the gal80p dimer. Reaction rates and species number are in terms of molecules per cell. It is initially assumed that there is no cooperativity, so binding of one protein does not affect the binding rates of subsequent proteins. The case with cooperative binding is discussed at the end.

Case 1: Single binding of P with gene G leads to promotion of transcription. This would be the case where gal4p dimer promotes transcription without repression, and is given as an illustrative example only.

We treat each binding site as a separate species G_n for n = 1 to N, and assume that P binds G_n with an equilibrium constant K_P (ratio of the forward reaction rate to the backwards rate). The binding reactions can be written

$$G_n + P \overset{K_P}{\longleftrightarrow} G_nP \tag{5}$$

so at equilibrium $G_nP = K_P \cdot P \cdot G_n$. If there is only one copy of the binding site, we have

$$1 = G_n + G_nP = (1 + K_P \cdot P)G_n \tag{6}$$

Setting p equal to the dimensionless quantity $K_P \cdot P$ gives $G_n = 1/(1+p)$, $G_nP = p/(1+p)$.

The probability of all sites being unoccupied (no transcription) is $\left(1/(1+p)\right)^N$. Expanding the denominator $D(p,N) = (1+p)^N$ gives

$$D(p,N) = \sum_{h=0}^{N} \binom{N}{h} p^h \tag{7}$$

The terms in the polynomial correspond to the relative probability of the various possible configurations of the gene's N binding sites. For example, if N = 2, then

$$D(p,2) = \sum_{h=0}^{2} \binom{2}{h} p^h = p^2 + 2p + 1 \tag{8}$$

The first term corresponds to the situation where each binding site is occupied, which can occur in $\binom{2}{0} = 1$ different ways. The second term corresponds to the case where only one binding site is occupied, which can occur in $\binom{2}{1} = 2$ ways (i.e. one site or the other), while the last term corresponds to the case where no binding site is occupied, which again can only occur in one way. The function D(p, N) therefore accounts for all the possible configurations. To determine the probability of transcription, we find the relative probability of only those states which lead to transcription, and divide by D(p, N). Therefore the fractional saturation function is

$$F(p,N) = \frac{\sum_{h=1}^{N} \binom{N}{h} p^h}{\sum_{h=0}^{N} \binom{N}{h} p^h} \tag{9}$$

where the numerator omits the term with $h = 0$ which corresponds to the case where no binding site is occupied.

Note that the fractional saturation function can also be written as $F(p, N) = 1 - (1/(1 + p))^N$, which is 1 minus the probability that no binding site is occupied. The main advantage of the summation notation is that it is useful in accounting for cooperative binding, as shown later.

Case 2: Single binding of P with gene G leads to promotion of transcription; binding of Q with GP represses transcription. This is the case when gal80p represses by binding with the gal4p dimer/DNA complex. If Q forms a complex with a species R, then the system will be activated by increasing R.

Proceeding as above, we have

$$
\begin{aligned}
&G_n + P \overset{K_P}{\longleftrightarrow} G_nP\\
&G_nP + Q \overset{K_Q}{\longleftrightarrow} G_nPQ\\
&\dot{G}_n = 0 \Rightarrow G_nP = K_P \cdot P \cdot G_n\\
&\dot{G}_nP = 0 \Rightarrow G_nPQ = K_Q \cdot G_nP \cdot Q\\
&\therefore G_nPQ = K_QK_P \cdot P \cdot Q \cdot G_n\\
&1 = G_n + G_nP + G_nPQ = (1 + K_P \cdot P + K_QK_P \cdot P \cdot Q)G_n\\
&\text{Set } p = K_P \cdot P,\ pq = K_QK_P \cdot P \cdot Q\\
&G_n = \frac{1}{1 + p + pq},\ G_nP = \frac{p}{1 + p + pq},\ G_nPQ = \frac{pq}{1 + p + pq}
\end{aligned}
\tag{10}
$$

Expanding the denominator raised to the power *N* as before gives

$$
\begin{aligned}
D(p, pq, N) &= (1 + p + pq)^N\\
&= \sum_{i=0}^{N} \binom{N}{i} pq^i \sum_{h=0}^{N-i} \binom{N-i}{h} p^h
\end{aligned}
\tag{11}
$$

The fractional saturation function is

$$
F(p, pq, N) = \frac{\sum_{i=0}^{N} \binom{N}{i} pq^i \sum_{h=0}^{N-i} \binom{N-i}{h} T(h)\, p^h}{\sum_{i=0}^{N} \binom{N}{i} pq^i \sum_{h=0}^{N-i} \binom{N-i}{h} p^h}
\tag{12}
$$

where

$$
T(h) = \begin{cases} 1 & \text{if } h > 0 \\ 0 & \text{otherwise} \end{cases}
\tag{13}
$$

in the numerator picks out those terms corresponding to a transcription state. Alternatively this can be written as $F(p, pq, N) = 1 - ((1 + pq)/(1 + p + pq))^N$.

Now, suppose that Q forms a complex with R, so that only a portion Q_B is available to bind with DNA, and we ignore other reactions such as decay of the proteins or complexes with other proteins. At equilibrium we have

$$
\begin{aligned}
&Q_B + R \overset{K_c}{\longleftrightarrow} QR\\
&\dot{Q}_B = 0 \Rightarrow QR = K_c \cdot Q_B \cdot R\\
&Q = Q_B + QR = (1 + K_c \cdot R) \cdot Q_B
\end{aligned}
\tag{14}
$$

Setting $pq = K_QK_P \cdot P \cdot Q_B = K_QK_P \cdot P \cdot (Q/(1 + K_c \cdot R))$ in Equation (12) means that the repressive term pq will decrease with an increase in R, so the system is activated via double repression. Note that the function F(p, pq, N) has a maximum value of $1 - (1/(1 + p))^N$, which is less than unity, even if all the repressor Q is removed.

8.2.1 Cooperativity:

When the fractional saturation functions are expressed (as here) by a series of summations, it is easy to modify them to account for cooperative binding. Suppose for case 2 above that if one molecule of *P* is bound, the binding coefficient for subsequent molecules increases to C_PK_P, and similarly for *Q*. In the summation, the index h counts the number of sites bound with P, and i counts the sites bound with *PQ*. Therefore, the total number of molecules of P bound is given by $h + i$. Since all but the first of these binding reactions has a binding coefficient of C_PK_P, the corresponding term in the expansion must be multiplied by a factor C_P^{h+i-1}. A similar argument for *Q* gives a fractional saturation function

$$
F(p, pq, N) = \frac{\sum_{i=0}^{N} \binom{N}{i} pq^i \sum_{h=0}^{N-i} \binom{N-i}{h} T(h)\, C(h, i)\, p^h}{\sum_{i=0}^{N} \binom{N}{i} pq^i \sum_{h=0}^{N-i} \binom{N-i}{h} C(h, i)\, p^h}
\tag{15}
$$

Here

$$
C(h, i) = C_P^{h+i-1} C_Q^{i-1}
\tag{16}
$$

where negative exponents are set to zero (cooperativity only has an effect if at least one molecule is bound). The galactose model assumes cooperativity in the binding of gal80p dimer with $C_Q = 30$, but no cooperativity in gal4p binding.