Feedback control of stochastic noise

in the yeast galactose utilization pathway

David Orrell, Stephen Ramsey, Marcello Marelli, Jennifer J. Smith,

Timothy W. Petersen, Pedro de Atauri, John D. Aitchison, Hamid Bolouri[*]

Institute for Systems Biology

1441 North 34th Street

Seattle, WA 98103

U.S.A.

Abstract

Genetic networks are often affected by stochastic noise, due to the low number of

molecules taking part in certain reactions. In complex regulatory networks, noise in any

one chemical species may induce noise in the rest of the system. In this paper, we analyse

the sources of stochastic noise in the yeast galactose utilization pathway at the level of

the complete system, by using both computer simulations, and experimental comparisons

between wild-type yeast and a modified strain. A computer model was first used to determine the sources of network noise, as well as mechanisms by which noise is controlled. Results from an estimation tool, confirmed by detailed stochastic simulations, show that noise is caused primarily by fluctuations in mRNA concentrations due to their stochastic creation and decay. The noise is controlled by feedback loops which regulate transcription of certain genes. To test the effect of the feedback loops on the rest of the system, a modified strain of yeast was prepared in which regulation of two key genes is eliminated. As predicted by the model, the mutant strain is induced by galactose in a manner similar to the wild type, but with a higher degree of stochastic noise.

PACS Codes

87.80.Vt; 82.20.Wt; 82.20.Uv; 82.39.Rt;

Keywords

gene expression noise; stochastic kinetics; yeast galactose metabolism

1. Introduction

Genetic networks often contain low copy numbers of many species, including regulatory molecules and DNA. Network simulations using Monte-Carlo techniques, coupled with experimental results [1, 2] have shown that the resulting stochastic effects can lead to significant fluctuations (noise) in network response. Since noise plays a key role in

---

* To whom correspondence should be addressed. hbolouri@systemsbiology.org (email); 206-732-1209 (tel.).; 206-732-1262 (fax)

genetic networks [3], it is interesting to ask whether systems have evolved to control it. For example, negative auto-regulation has been shown to reduce stochastic noise in the production of individual proteins [2]. In this paper, we use estimation techniques to analyse noise, on the scale of the complete regulatory network rather than individual proteins, for galactose utilization in the yeast *Saccharomyces cerevisiae*.

The galactose network has been studied extensively over a number of decades [4-7]. As illustrated in Figure 1, galactose is first transported into the cell through a process which depends on the protein Gal2p. It is then converted into glucose 1-phosphate through the Leloir pathway, which is controlled by the products Gal1p, Gal7p and Gal10p [8, 9]. These proteins are in turn induced by a regulatory apparatus which contains (at least) three additional proteins: Gal3p, Gal4p and Gal80p [7, 10, 11].

The regulatory apparatus operates by a double repression mechanism. The Gal4p protein promotes transcription of the other *GAL* genes by homo-dimerizing and binding with their *cis*-regulatory regions, while the Gal80p protein inhibits transcription by binding with DNA-bound Gal4p. When activated by intra-cellular galactose, Gal3p restores transcription of the *GAL* genes by binding with Gal80p, sequestering it in the cytoplasm, and thus relieving its repressive action. We can therefore say that Gal3p is a positive regulator for the network as a whole, while Gal80p is a negative regulator or inhibitor. An interesting feature of the galactose pathway is that, during activation, both Gal3p and Gal80p are upregulated [12, 13]. Other things being equal, an increase in Gal3p will give a stronger response for a set amount of galactose (i.e. an increase in the level of galactose metabolism, see [14]); but if that is the aim, then why is Gal80p concentration also increased in response to galactose?

In this paper, we analyze sources of noise in the galactose network, and propose reasons for its complex structure, using both computer models and experiments with wild type and modified strains of yeast. We argue two main points: that the major source of noise in the galactose pathway is transcription of mRNA, in particular that for Gal3p and Gal80p; and that noise is controlled by regulatory feedback mechanisms for Gal3p and Gal80p. Section 2 explores the role of the feedback loops by comparing two computer models of the galactose pathway. The first is the wild type model of the regulatory and metabolic systems (see also the works of: [14, 15]); the second is an unregulated version, which is exactly the same in every respect, except that the proteins Gal3p and Gal80p are not upregulated during induction, but are maintained at their low, non-induced levels. The two models appear similar both in equilibrium and transient behavior; however a detailed stochastic simulation shows that the unregulated system is substantially noisier. A noise estimation tool is used in Section 3 to identify both the sources of noise, and mechanisms within the network which help control the major noise sources. It is seen that effectively all system noise can be viewed as a response to mRNA fluctuations caused by the creation and decay of those species. The noise is shown to be controlled by feedback loops which regulate production of mRNA for Gal3p and Gal80p. Section 4 describes experiments using a mutant strain in which feedback regulation of Gal3p and Gal80p has been omitted. A comparison of the mutant and wild-type behaviors is found to be consistent with the theoretical model simulations. Finally, the results are discussed in Section 5.

2. Regulated vs unregulated

We previously presented [16] a computer model for the galactose network which represents the wild-type yeast in which Gal3p and Gal80p are regulated. The model simulates 55 different reaction channels between key species in the galactose network, including regulatory and structural proteins, mRNA, and DNA. The reaction parameters were set to be consistent with existing experimental data from the literature. The equations and details of the model are presented in detail in the above reference, and are summarized briefly in Appendix B herein. To determine the role of Gal3p/Gal80p regulation, we here compare the behavior of that model with an unregulated version, in which the rate of production of mRNA for Gal3p and Gal80p is maintained at the low, non-induced levels. Apart from this change, the structure and parameters of the two models are the same. The difference between the two models therefore illustrates the effect of regulation of these proteins.

The two models were first simulated using an ordinary differential equation (ODE) approach. Figure 2 shows the equilibrium concentrations as a function of external galactose, found by running the ODE simulation until the models achieved steady-state. In either model, the activated form of Gal3p binds with Gal80p after induction to form a complex, and this lowers the quantity of free Gal80p homodimer that is available to inhibit transcription. In the unregulated model, the mRNA for Gal3p and Gal80p are maintained at constant levels, while in the regulated model both are increased after induction. Apart from this difference, however, the two versions show similar behavior. As illustrated in Figure 2, the magnitude and speed of response to galactose availability when both Gal3p and Gal80p are upregulated are essentially the same as in the

unregulated model. Note that this is not the case if only one is upregulated; for example, if Gal3p is upregulated, but Gal80p is not, then the system becomes far more sensitive to external galactose. We note a recent study [14] which demonstrated that positive feedback through Gal3p is responsible for bistability of the system, and that negative feedback through Gal80p is partially responsible for the fluctuations which govern the transition time from the uninduced to induced state. Their work, based on strains in which the GAL3 and GAL80 feedback loops are deleted one at a time, emphasized the separate roles of the Gal80p and Gal3p feedback loops in controlling noise.

There is also little difference in the transient behavior of our two models, for example in the response to induction. The red dashed lines in Figure 3 show an ODE simulation of the two models, initiated in the non-induced state, and run forward in time with external galactose set to a level corresponding to 10% induction, as measured by the concentration of Gal1p (the other structural proteins show a similar induction). This level, marked with a triangle on the horizontal axis in Figure 2, is 0.5 mM for the model with regulation, and is reduced to 0.32 mM for the model with no regulation to give the same mean protein levels. (Another possibility is to change the rate for the activation of Gal3p by internal galactose in the unregulated model; however we found this had no significant effect on the model properties.) The two models also respond in very similar ways to other environmental perturbations, such as a short pulse of external galactose (D.O. unpublished data).

The ODE approach gives an idea of the cell-averaged system behavior, but is only accurate when the number of molecules taking part in all reactions is very high, which is certainly not the case for the galactose model. A more realistic approach is the ensemble

of 30 stochastic simulations shown by the green lines in Figure 3. These were performed using the open-source program Dizzy (http://magnet.systemsbiology.net/dizzy), which incorporates a number of different solvers [17]. The Gillespie solver [18] used here employs a Monte Carlo technique to generate trajectories consistent with the Chemical Master Equation [19]. The green lines can be viewed as corresponding to the behavior of individual cells. The mean of the ensemble of 30 simulations is indicated in the figure by the blue lines. There is considerable variability around the mean, which indicates the range of responses between different cells. However the mean is close to the ODE simulation, so one could say that for this system the ODE adequately captures the average behavior of a large number of cells.

These simulations again show no great differences between the regulated and unregulated models. However, one expected property of the galactose network [20], and indeed other such networks, is that the structural proteins Gal1p, Gal2p, Gal7p and Gal10p have slow decay times, so perturbations due to stochastic effects lead to low-frequency fluctuations. Therefore while Figure 3 gives an impression of how individual cells react to induction, the time scale is not sufficiently long to accurately measure the noise properties. To get reliable results, the stochastic effects must either be computed for many cells, or alternatively for a single cell over a long simulation time (we established empirically that 10000 minutes was adequate for the results to converge). Figure 4 compares long simulations, using Dizzy's implementation of the Tau-Leap algorithm [21-23], for the two versions of the model in the 10% induction state. The models have the same mean concentrations (except for internal galactose, denoted GAI), but the unregulated model in the right panels appears noisier. This is confirmed by the top panel

of Figure 5, which shows the dimensionless ratio of standard deviation to mean for the range of species. Noise is higher for the unregulated model almost across the board. Simulations were performed at partial induction since the impact of noise is greater at low species concentrations [24, 25], and the computation time is also faster.

Of particular interest is the metabolic intermediary galactose 1-phosphate (denoted GA1P). Regulation of Gal3p and Gal80p affects noise not just in the regulatory network, but also in the downstream metabolic products. Galactose 1-phosphate is thought to be toxic to the cell at high concentrations [26-29]. As shown in the simulation results in Figure 4, greater noise in the unregulated model can result in 'spikes' of GA1P more than twice the physiological concentration. Regulation of noise could improve the viability of the organism, by controlling the maximum concentration of GA1P experienced, and the magnitude of peak-to-peak fluctuations. From these simulations, it therefore appears that the regulatory feedback loops are not necessary for induction, but reduce system noise. The next section looks more closely at the specific sources of noise, and how the feedback loops function to reduce it.

3. Sources of noise

Since GA1P is part of an interconnected system, fluctuations in its concentration depend on variations in all the other species in the network. In order to understand how noise in GA1P arises, and how the network might control it, it is necessary to identify the sources of noise at a system level. For example, one reaction in the network might be extremely noisy, but have little effect on the rest of the system; while another might generate less

noise, but have a magnified effect downstream. To this end, we used the noise estimation
tool in Dizzy [30], which estimates the total variance as a weighted sum of the reaction
rates. For a system with N species and M reactions, the expression for the total variance
in the different species is the N-dimensional vector:

$$\mathbf{v} = \frac{1}{2} \sum_{j=1}^{M} s_j \mathbf{w}_j \qquad (1)$$

where the s_j are reaction rates, and the $\mathbf{w}_j$ are weighting factors, which are derived from
the system equations. The derivation is given in [30], and a detailed definition of the
terms is given in Appendix A.

Eq. (1) provides a general framework for the analysis of noise in biochemical
networks. The equation is obtained by considering the stochastic noise as a perturbation
to the steady state solution of the ODEs, akin to a stochastic model error. The deviation
of the stochastic trajectory from the ODE steady state can then be estimated using error
growth methods [31, 32]. The variance $\mathbf{v}$ represents a balance between two effects. The
first is the size of the stochastic perturbations in the various reactions, which depends on
the reaction rates s_j. The second is the system stability as measured by the negative real
part of the eigenvalues of the system Jacobian, which appear in the weighting factors $\mathbf{w}_j$.
The $\mathbf{w}_j$ have units of time, and are analogous to decay times for the perturbations arising
from the different reactions.

The resulting estimates for system noise are shown in the lower panel of Figure 5.
The estimation tool captures the differences between the two models, and takes only

seconds, instead of days, to compute. Its main advantage is that it can be used to determine the primary causes of stochasticity in the system. Each term in the summation Eq. (1) corresponds to the variability which is introduced by a particular reaction channel. Referring to that equation, if the j'th reaction were somehow made non-stochastic, the total variance in GA1P would decrease by an amount about equal to the corresponding term $\frac{1}{2}s_j\mathbf{w}_j$ in the sum. (There is also a secondary effect on the weighting factors for the other terms, as evident from the equations in Appendix A, but it is quite small. See for example the discussion of Figure 6, panel C below.)

Panels A and B of Figure 6 show the largest contributions to GA1P noise in the regulated and unregulated models respectively. They are expressed in terms of the ratio of standard deviation to the mean: $\frac{1}{GA1P}\sqrt{\frac{1}{2}s_j\mathbf{w}_j}$, where $GA1P$ is the mean GA1P concentration. In either model, the reactions which contribute the most are those which create (or destroy) mRNA. In fact, these reactions account for about 98% of the total estimated noise, while all other reactions contribute only about 2%. This makes sense since the mRNA only exist in low copy numbers, tend to be highly stochastic, and drive the formation of protein. Because the reactions of creation and destruction are in equilibrium, the two contribute an equal amount.

Since mRNA creation/decay reactions account for most of the system noise, it follows that fluctuations in downstream proteins are primarily a reaction to changing mRNA levels. This is graphically demonstrated in Figure 7, which compares a simulation of the full model (blue line, compare left panels of Figure 4), with a simulation in which the time course mRNA concentrations from the stochastic simulation were a fixed input,

but the concentrations of all other species were determined from that using ODEs (red lines). Any change in protein levels is therefore due not to protein reactions themselves (which are simulated with ODEs) but are reactions to the changes in mRNA. The close agreement implies that protein fluctuations are almost entirely a result of mRNA noise.

Another way to look at this is that, if fluctuations in mRNA levels could be removed completely, then the rest of the system would experience hardly any noise. Reactions such as translation of proteins, or interactions between proteins, would result in only mild stochasticity. As an illustration, the green lines in Figure 7 show a stochastic simulation in which all mRNAs are kept artificially fixed at their equilibrium levels, but the other reactions proceed in a stochastic fashion as before. It therefore indicates the noise which would result in the rest of the system if the creation and decay of mRNA were somehow to be completely noise-free. The extremely low noise level confirms that system noise is due primarily to mRNA creation/decay.

Using the estimation tool, we can go further and determine the dependence of system noise on particular creation/decay reactions. For example, the second highest contribution in the unregulated model is from the mRNA for Gal80p. If the mRNA for Gal80p only is kept at a fixed level without formation or decay, then the noise due to its formation and decay reactions will be muted. As shown in Figure 6, panel C, the contribution from those reactions drops out and the total variability is lowered; however the contributions to variability from the other reactions are little changed.

It is of course impossible to completely eliminate reaction noise in this way. It can be reduced through negative autoregulation or feedback [33, 34], and in the regulated model the protein Gal80p does reduce its own formation in this way. At partial induction,

however, there is also positive feedback, through Gal3p. Increasing Gal3p tends to increase the level of activation and so produce more Gal3p, which in turn increases Gal80p. Just as negative feedback reduces noise, so positive feedback increases it; and the total system includes both.

In fact, the noise reduction comes about because the net effect of the feedback loops is to increase the rate of production of both Gal80p and Gal3p. The amount of these proteins, in monomer and dimer form, as well as the various complexes they form, all increase, so the entire control network grows in size. Since the degree of stochastic variability tends to decrease with the number of molecules, the network is less noisy. This mechanism was demonstrated by estimating the noise for another variant of the model in which Gal3p and Gal80p are again not regulated; however rather than being maintained at the non-induced level, they are maintained at the higher level of 10% induction. The estimated noise for this unregulated model, shown in panel D, is close to that of the full regulated model. The result was confirmed by running a full stochastic simulation; the ratio of standard deviation to noise for GA1P was 0.22, in near agreement with the regulated model at 0.23.

If this unregulated model succeeds in reproducing the noise characteristics of the regulated model, one might then ask why the organism doesn't simply maintain a higher concentration of Gal3p and Gal80p at all times. One way to do this is to slow the decay rates for the regulatory mRNA, which in the model decay relatively rapidly compared to structural mRNA [20]. (Another way is to simply speed production, but this might be metabolically constrained.) The results are shown in panel E of Figure 6 for a tenfold decrease in the decay rate of the regulatory mRNA, which brings the protein

concentrations up to the levels of panel B. There is now a much larger contribution to noise from the formation of regulatory mRNA. The reason is that, with a slower decay time, the number of proteins produced per mRNA molecule is increased. As discussed, and demonstrated experimentally [2], this leads to more noise being transmitted to the regulatory proteins, and, it follows, the rest of the network.

What counts is therefore not just the level of regulatory proteins, but the rate at which the mRNAs are degraded. But if the regulatory mRNA must be degraded quickly to avoid the transmission of noise, it means that many must be produced to maintain a given level, which is metabolically expensive. The cell therefore appears to adopt a "just-in-time" approach, ramping up the control network when necessary, but maintaining it at a low level when used only in a sensing capacity. This structure is consistent with the aims of reducing noise, while avoiding the metabolic cost of producing excess mRNA and protein. Indeed, if the metabolic cost were not a significant constraint, then presumably noise would rarely be an issue in genetic networks, since all regulatory species and mRNA would be maintained at high levels (except where randomizing noise was desired).

We therefore conclude from the theoretical results that most noise is due to transcription of mRNA; it is controlled by Gal3p and Gal80p feedback mechanisms; and while the feedbacks reduce noise, they do not appear to be necessary otherwise. In the next section, we demonstrate experimentally the role of the feedback loops, by omitting them from a mutant strain of yeast.

4. Experimental results

In order to test the effect of regulation on network noise, a mutant strain of yeast was prepared in which Gal3p and Gal80p were expressed at low, constant levels. The wild-type promoters of GAL3 and GAL80 were replaced with the minimal promoter of the cytochrome-c isoform 1 (CYC1) gene, with no upstream activating sites. The green fluorescent protein gene (GAL1:GFP) was also inserted into the yeast genome to act as a reporter [35, 36]. This strain therefore corresponds to the unregulated version of the model, with the reporter protein indicating the level of system activation. Its behavior was compared with a second strain which was wild-type except for the addition of the GFP reporter.

Using flow cytometry, we measured the fluorescence of cells in the mutant and wild-type at low concentrations of galactose. From the histogram of fluorescences, we calculated the cell-to-cell variation in the fluorescence intensity. The results for 0.05% galactose are shown in Figure 8. This corresponds to approximately 50% induction. The noise coefficients, defined as the ratio of standard deviation to mean, are 0.066 for the wild-type, and 0.126 for the mutant. (There is a slight difference in the mean, but by whatever measure the unregulated case is noisier.) Some of this noise will be due to effects such as variation in cell size, stage in the cell cycle, and so on. As discussed in Appendix B, however, this additional noise is decreased somewhat by the experimental procedure, which sorts for cells of a similar size. The effect of removing the regulation is to increase the system noise by a factor of about 1.91.

The experimental results can be compared with Figure 9, which was obtained by running an ensemble of 100 stochastic simulations. The noise coefficients here are 0.113

for the wild-type, and 0.224 for the mutant. These are in approximate agreement with estimated values, from the noise estimation tool in Dizzy, of 0.105 and 0.266 respectively. See Table 1 for a summary of the stochastic and experimental results. While the noise is slightly higher in the computer simulations than in the experimental results, the model correctly predicts two things. Firstly, the mutant strain can respond to galactose. The mean level of induction in either case is quite similar. Thus, if one were only to measure the average induction over a large number of cells, it would be hard to understand the role of regulation. Secondly, the mutant strain is considerably noisier than the wild-type version. Again, this can only be determined experimentally using measurements of individual cells. The main effect of regulation therefore appears to be to control system noise.

5. Discussion

The above computer and experimental results show how analysis of the sources of noise in the galactose network allows us to "reverse engineer" some of its features. If an engineer were to design this system, they might apply the following logic. Effectively all noise in the system is a response to fluctuations in mRNA; so to control noise, the main priority is to reduce the relative size of these fluctuations, and reduce their transmission to the rest of the system. The structural mRNAs, such as the GAL7 transcript, are major contributors to system noise, but are already upregulated to high levels during induction. The regulatory mRNAs are not required to produce large numbers of proteins, but should have both a reasonably short decay time, and a high concentration after induction, to

prevent noise. Since this combination is metabolically expensive, the solution is to regulate these mRNA, so their rate of production is increased when necessary, but kept low when used only in a sensing capacity.

Of course, the system was not designed according to some logical plan; but through understanding the sources of noise, we can gain some insight into the role of features which otherwise seem unnecessarily complicated and metabolically expensive. Up-regulation of Gal3p and Gal80p requires the formation of more mRNA and protein and adds complexity. For this feature to have been selected by evolution, it is reasonable to suppose that it confers some advantage, and the control of noise is a likely contributor.

In any case, it is clear that, in analyzing the properties of a system, ODE simulations of computer models, or experimental results averaged over a large number of cells, may disguise important properties of the system. The computer models with and without regulation only reveal their differences through detailed stochastic simulations. Similarly, single-cell experimental analysis was required to show that, while lack of regulation has only a small effect on the average level of induction, it does result in greater stochastic noise.

Materials and Methods

The model used is as described in our previous study [38], except as indicated here. For the mutant strain, we modified the fractional saturation for *GAL3* and *GAL80* to a fixed value of 0.04. This fractional saturation value is consistent with the data in Table 3 of [39], which shows the ratio of basal expression, to UAS-activated expression,

of a *CYC1-lacZ* reporter construct. We added a reporter gene driven by the

GAL4/GAL80/GAL3 regulatory mechanism, with one binding site for *GAL4*. The half-

life of the reporter product was defined to be ~350 minutes, which is consistent with *in*

vivo measurements of GFP degradation [40]. The maximal rate of initiation of

transcription, the maximal rate of initiation of translation per transcript, and the half-life

of the transcript for the reporter gene were chosen to be identical to the other structural

GAL genes in the model.

All simulations of the *GAL* model were performed using the Dizzy software

program [17] version 1.9.1, on a 46-node Intel dual-processor cluster computer (IBM)

running Linux kernel version 2.4.22 (Red Hat). The Java virtual machine used is the Java

2 Runtime Environment, version 1.4.2 (IBM).

The dynamics of the model was solved in two steps. First, the steady-state was

obtained by numerically solving, out to a long time scale, the deterministic chemical

kinetics using ordinary differential equations (ODEs). The steady-state concentrations of

the species provided the initial conditions for the stochastic simulations. Second, the

steady-state stochastic dynamics were solved using a Monte Carlo technique in which the

results were averaged over an ensemble of realizations of a Markov process representing

the stochastic kinetics. The initial conditions for the first (ODE) step were the species

concentrations at steady-state in the uninduced model (0% galactose, corresponding to

growth in non-*GAL*-repressing media).

Deterministic simulations were conducted using the fifth-order Dormand-Prince

ODE solver with a fourth-order error estimation formula and adaptive step-size control.

The ODE solver used is the `odeToJava` solver (Raymond Spiteri & Murray Patterson, `www.netlib.org/odeToJava`), available as the simulator option `odeToJava-dopr54-adaptive` within the Dizzy software program. The relative and absolute error tolerances were set to 0.0001. The initial step size was set to 0.4. The stop time was set to 4000 minutes, which was sufficient for the wild-type and mutant models to reach steady-state, for the galactose concentrations used.

Stochastic simulations were conducted using the Gillespie Tau-Leap Algorithm [22] which is available within Dizzy as simulator option `tauleap-complex`. A stop time of 1440 minutes was used, with 400 time samples per simulation. The Tau-Leap relative error threshold was set to 0.0005. The number of separate simulations conducted was 200 (for mutant and wild-type models separately), yielding an overall effective ensemble size of 40,000.

A mutant strain was constructed containing *GAL80* and *GAL3* behind the *CYC1* minimal promoter, by replacing the endogenous promoters of each gene with the *CYC1* minimal promoter. Briefly, *URA3* was fused up stream of the *CYC1* promoter, comprising nucleotides –138 to –1 (where +1 is the A of the *CYC*1 start codon), by sequential PCR with flanking oligomers containing ~60 bp arms homologous to either *GAL3* or *GAL80*. This construct was integrated into BY4742 or BY4741 [41] replacing nucleotides –1 to –899 of *GAL3* or nucleotides –1 to –249 of *GAL80*, respectively. The individual *CYC1:GAL3* and *CYC1:GAL80* strains were crossed and diploids sporulated to obtain haploid strains containing both modifications.

The reporter gene encoding GFP+ under the control of the *GAL1* promoter was generated by PCR. To do this, the gene encoding GFP+ followed by the

Schizosaccharomyces pombe gene *HIS5*, was amplified from pGFP+/HIS5, generated by replacing EGFP in pGFP/HIS5 [42] with GFP+ [43]. The *GAL1* promoter consisting of nucleotides -12 to –690 was amplified from genomic DNA and the two PCR products were fused by PCR using flanking oligomers containing ~60 bp sequences homologous to *HIS3*. The gene fusion was integrated into BY4741 and into an isogenic strain containing both *GAL3* and *GAL80* modifications, replacing +1 to +400 of the *his3Δ*1 locus. Strains with galactose-dependent green fluorescence were selected by fluorescence microscopy.

All yeast strains were grown overnight in YEP (1% yeast extract, 2% peptone) supplemented with 2% glucose (YEPD), washed and transferred to YEP containing 2% raffinose (YEPR), and grown for 16 hours at 30 ºC. Cells were collected and induced in YEPR or YEP supplemented with galactose (YEPG) at the concentrations indicated in the main text, for 6 hours. Induced cells were suspended at $1-10 \times 10^6$/ml (approximately 0.6 OD_{600}) in the induction media and measured on a high-speed inFlux cell sorter (Cytopeia). Cells were analyzed based on scattered laser light and fluorescence from a focused 488-nm argon ion laser operating at 300 mW. The data for each event consisted of an ordered tuple of the forward light scatter (FSC), perpendicular (side) light scatter (SSC), pulse width of the forward scattered light, and the integrated fluorescence from 510 nm to 550 nm to quantify the green fluorescent protein concentration. Data from a single run of the flow cytometer was stored in a Flow Cytometry Standard (FCS) version 3.0 format file incorporating 100,000 events. Initial visualization of the flow cytometry data was performed using Summit Offline (DakoCytomation) version 3.1. The correlation between fluorescence and cell size was corrected by fitting a linear function to

the fluorescence vs. forward-scatter data. The first-order term was subtracted from the ordinate (fluorescence) values, and the ordinates additively rescaled so that the mean adjusted fluorescence is the same as the mean raw fluorescence. This adjustment removed the cell-size bias in the fluorescence values. A subset of events within a window of forward-scatter was used to generate the final fluorescence histogram. The window had a width of 1 standard deviation (of the forward scattering) and was centered at the median forward scattering value. The purpose of this window selection was to ensure approximate homogeneity of analyzed cell volumes, and to exclude multiple-cell events.

Appendix A: Noise estimation formula

We present here the definition of the terms in Eq. (1) of the main text. See [37] for the derivation. We assume a biochemical network consisting of N molecular species with concentrations x_i molecules/cell, and M reactions.

$\mathbf{r}_j$ is first defined to be the N-dimensional unit column vector which describes the direction of the j'th reaction. If for example the reaction is $x_a + x_b \xrightarrow{\ k_j\ } x_c$, then species x_a and x_b lose one molecule, and species x_c gains one molecule, so $\mathbf{r}_j$ is the column vector with the value $\dfrac{-1}{\sqrt{3}}$ in rows a and b, and $\dfrac{1}{\sqrt{3}}$ in row c.

s_j is the reaction rate. For the above example, it is $s_j = k_j x_a x_b$. This term appears directly in the summation in Eq. (1).

J is next defined as the Jacobian of the ODE system equations. If the system is at a stable equilibrium, the eigenvalues u_j of **J** have negative real part.

Q is the matrix $\mathbf{Q} = \mathbf{PTP}^{-1}$ where **P** is the matrix formed from the column eigenvectors of **J**, and **T** is the diagonal matrix with entries $\mathbf{T}(j,j) = \dfrac{1}{\sqrt{-\text{Re}(u_j)}}$.

$\mathbf{w}_j$ is a vector with entries equal to the square of those in $\mathbf{Qr}_j$. These are the weighting factors which reflect the system's ability to damp out internal fluctuations.

The column vector giving the variance of each species is then given by $\mathbf{v} = \dfrac{1}{2}\displaystyle\sum_{j=1}^{M} s_j \mathbf{w}_j$, which is Eq. (1).

This algorithm has been tested against stochastic simulations, and incorporated into the program Dizzy [17].

References

1. McAdams, H.H. and A. Arkin, *It's a noisy business! Genetic regulation at the nanomolar scale.* Trends in genetics, 1999. **15**(2): p. 65-69.
2. Thattai, M. and A. van Oudenaarden, *Intrinsic noise in gene regulatory networks.* Proc Natl Acad Sci U S A, 2001. **98**(15): p. 8614-9.
3. Arkin, A., J. Ross, and H.H. McAdams, *Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected Escherichia coli cells.* Genetics, 1998. **149**(4): p. 1633-48.
4. Johnston, M., *A model fungal gene regulatory mechanism: the GAL genes of Saccharomyces cerevisiae.* Microbiol. Rev., 1987. **51**: p. 458-476.
5. Peng, G. and J. Hopper, *Evidence for Gal30'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**: p. 5140-5148.
6. Peng, G. and J. Hopper, *Gene activation by interaction of an inhibitor with a cytoplasmic signaling protein.* Proceedings of the National Academy of Science, 2002. **99**: p. 8548-8553.

7. Lohr, D., P. Venkov, and J. Zlatanova, *Transcriptional regulation in the yeast GAL gene family: a complex genetic network.* FASEB J, 1995. **9**: p. 777-787.

8. 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**: p. 441-470.

9. Holden, H.M., I. Rayment, and J.B. Thoden, *Structure and function of enzymes of the Leloir pathway for galactose metabolism.* J. Biol. Chem., 2003. **278**: p. 43885-43888.

10. Melcher, K., *Galactose metabolism in Saccharomyces cerevisiae: a paradigm for eukaryotic gene regulation*, in *Yeast Sugar Metabolism*, F.K. Zimmermann and K.-D. Entian, Editors. 1997, Technomic Publishing Inc.: Lancaster, PA.

11. Bhat, P.J. and T.V. Murthy, *Transcriptional control of the GAL/MEL regulation of yeast Saccharomyces cerevisiae: mechanism of galactose-mediated signal transduction.* Molecular Microbiology, 2001. **40**(5): p. 1059-1066.

12. Shimada, H. and T. Fukasawa, *Controlled transcription of the yeast regulatory gene GAL80.* Gene, 1985. **39**: p. 1-9.

13. Bajwa, W., T.E. Torchia, and J.E. Hopper, *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**: p. 3439-47.

14. Acar, M., A. Becskei, and A. van Oudenaarden, *Enhancement of cellular memory by reducing stochastic transitions.* Nature, 2005. **435**(7039): p. 228-32.

15. Verma, M., P.J. Bhat, and K.V. Venkatesh, *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.

16. de Atauri, P., et al., *Evolution of "design" principles in biochemical networks.* IEE Systems Biology, 2004. **1**(1): p. 28-40.

17. Ramsey, S., D. Orrell, and H. Bolouri, *DIZZY: Stochastic Simulation of Large-Scale Genetic Regulatory Networks.* Journal of Bioinformatics and Computational Biology, 2005. **3**(2): p. 1-21.

18. Gillespie, D.T., *A general method for numerically simulating the stochastic time evolution of coupled chemical reactions.* Journal of computational physics, 1976. **22**: p. 435-450.

19. McQuarrie, D.A., *Stochastic approach to chemical kinetics.* J. Appl. Prob., 1967. **4**: p. 413-478.

20. Wang, Y., et al., *Precision and functional specificity in mRNA decay.* Proc Natl Acad Sci U S A, 2002. **99**(9): p. 5860-5.

21. Gibson, M.A. and J. Bruck, *Efficient exact stochastic simulation of chemical species with many species and many channels.* J. Phys. Chem., 2000. **104**: p. 1876-1889.

22. Gillespie, D.T., *Approximate accelerated stochastic simulation of chemically reacting systems.* Journal of Chemical Physics, 2001. **115**(4): p. 1716-1733.

23. Gillespie, D.T. and L.R. Petzold, *Improved leap-size selection for accelerated stochastic simulation.* Journal of chemical physics, 2003. **119**: p. 8229-8234.

24. Elowitz, M.B., et al., *Stochastic gene expression in a single cell.* Science, 2002. **297**(5584): p. 1183-6.

25. Ozbudak, E.M., et al., *Regulation of noise in the expression of a single gene.* Nat Genet, 2002. **31**(1): p. 69-73.

26. Deng, M., et al., *Inferring domain-domain interactions from protein-protein interactions.* Genome Res, 2002. **12**(10): p. 1540-8.

27. Lai, K. and L.J. Elsas, *Overexpression of human UDP-glucose pyrophosphorylase rescues galactose-1-phosphate uridyltransferase-deficient yeast.* Biochem. Biophys. Res. Commun., 2000. **271**: p. 392-400.

28. Ostergaard, S., L. Olsson, and J. Nielson, *In Vivo Dynamics of Galactose Metabolism in Saccharomyces cerevisiae: Metabolic Fluxes and Metabolite Levels.* Biotechnology and Bioengineering, 2001. **73**(5): p. 412-425.

29. Lai, K., et al., *GALT deficiency causes UDP-hexose deficit in human galactosemic cells.* Glycobiology, 2003. **13**: p. 285-294.

30. Orrell, D., et al., *A method for estimating stochastic noise in large genetic regulatory networks.* Bioinformatics, 2005. **21**(2): p. 208-17.

31. Orrell, D., *Estimating error growth and shadow behavior in nonlinear dynamical systems.* Int. J. Bifurcat. Chaos, in press.

32. Orrell, D., *Role of the metric in forecast error growth: how chaotic is the weather?* Tellus, 2002. **54A**: p. 350-362.

33. Orrell, D. and H. Bolouri, *Control of internal and external noise in genetic regulatory networks.* J Theor Biol, 2004. **230**(3): p. 301-12.

34. Becskei, A., B. Seraphin, and L. Serrano, *Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion.* Embo J, 2001. **20**(10): p. 2528-35.

35. Li, J., et al., *Green fluorescent protein in Saccharomyces cerevisiae: real-time studies of the GAL1 promoter.* Biotechnol. Bioeng., 2000. **70**: p. 187-196.

36. Elowitz, M.B., et al., *Stochastic gene expression in a single cell.* Science, 2002. **297**: p. 1183-1186.

37. Orrell, D., et al., *A method for estimating stochastic noise in large genetic regulatory networks.* Bioinformatics, 2004.

38. de Atauri, P., et al., *Evolution of "design" principles in biochemical networks.* IEE Journal of Systems Biology, 2004. **1**(1): p. 22-40.

39. Blair, W.S. and B.R. Cullen, *A yeast TATA-binding protein mutant that selectively enhances gene expression from the weak RNA polymerase II promoters.* Mol Cell Biol, 1997. **17**: p. 2888-2896.

40. Corish, P. and C. Tyler-Smith, *Attenuation of green fuorescent protein half-life in mammalian.* Protein Engin, 1999. **12**: p. 1035-1040.

41. Brachmann, C.B., et al., *Designer deletion strains derived from Saccharomyces cerevisiae S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications reference.* Yeast, 1998. **14**: p. 115-132.

42. Dilworth, D.J., et al., *Nup2p dynamically associates with the distal regions of the yeast nuclear pore complex.* J Cell Biol, 2001. **153**(7): p. 1465-78.

43. Scholz, O., et al., *Quantitative analysis of gene expression with an improved green fluorescent protein.* Eur J Biochem, 2000. **267**: p. 1565-1570.
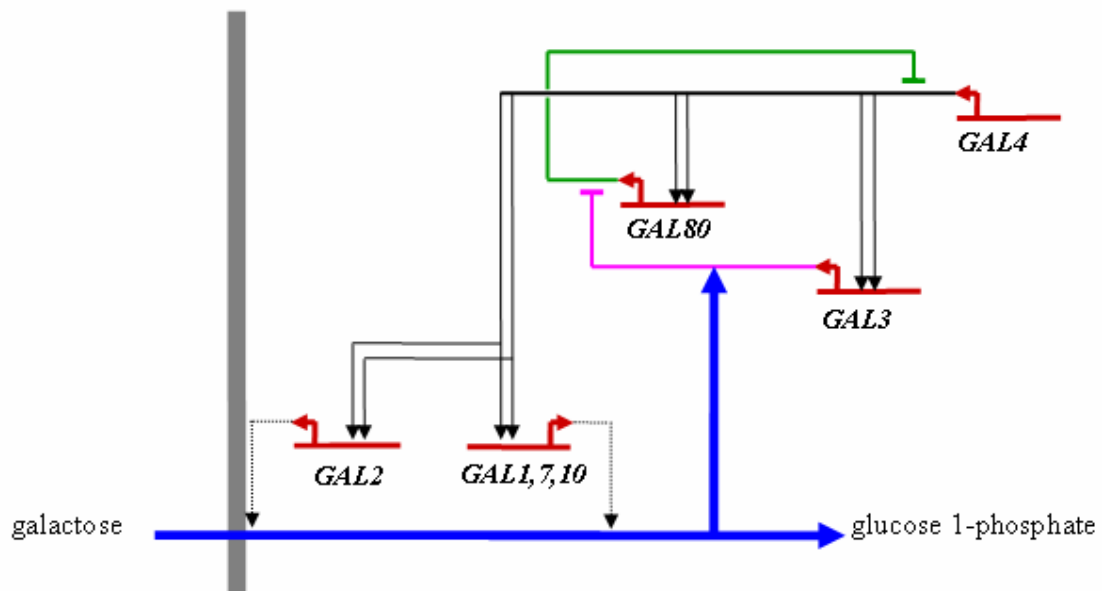
FIGURES

Figure 1. Schematic diagram of the regulatory portion of the galactose network in yeast. When galactose enters the cell, in a process involving Gal2p, it causes Gal3p to bind with Gal80p, thus activating the production of structural proteins Gal1p, Gal7p, and Gal10p. These convert the galactose to glucose 1-phosphate, with galactose 1-phosphate (not shown) as one intermediary in the process. See text for discussion.
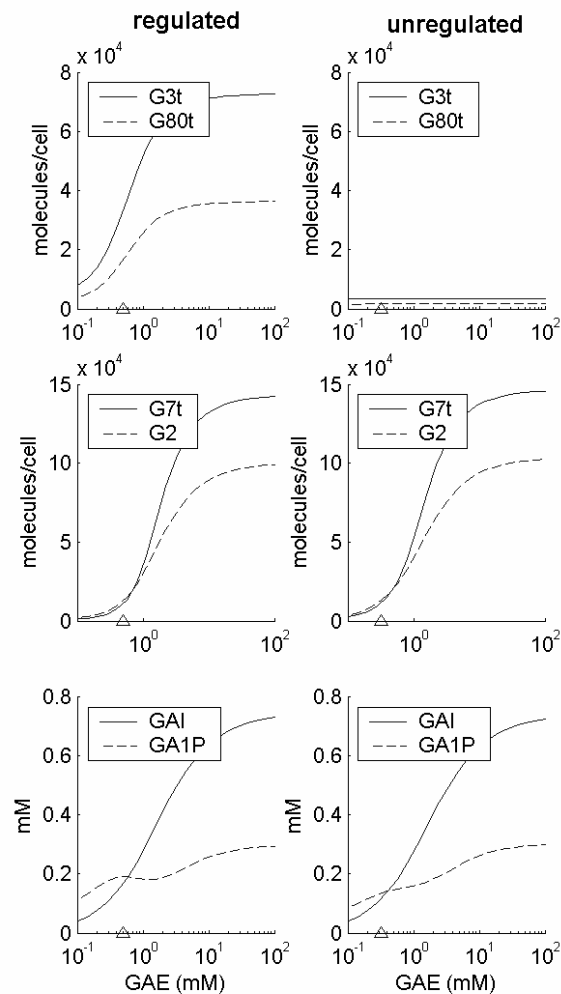
Figure 2. Comparison of equilibrium behavior as a function of external galactose (GAE), for the galactose pathway model with regulation (left panels), and a version in which Gal3p and Gal80p are not upregulated (right panels). Shown from top to bottom are regulatory protein; structural protein; internal galactose GAI and galactose 1-phosphate GA1P (see [16] for details). G80 refers to Gal80p protein, G80t is total Gal80p in monomer or dimer form, and likewise for the other symbols. The two models show similar behavior for the structural proteins and metabolic quantities. The triangle indicates the 10% induction level.
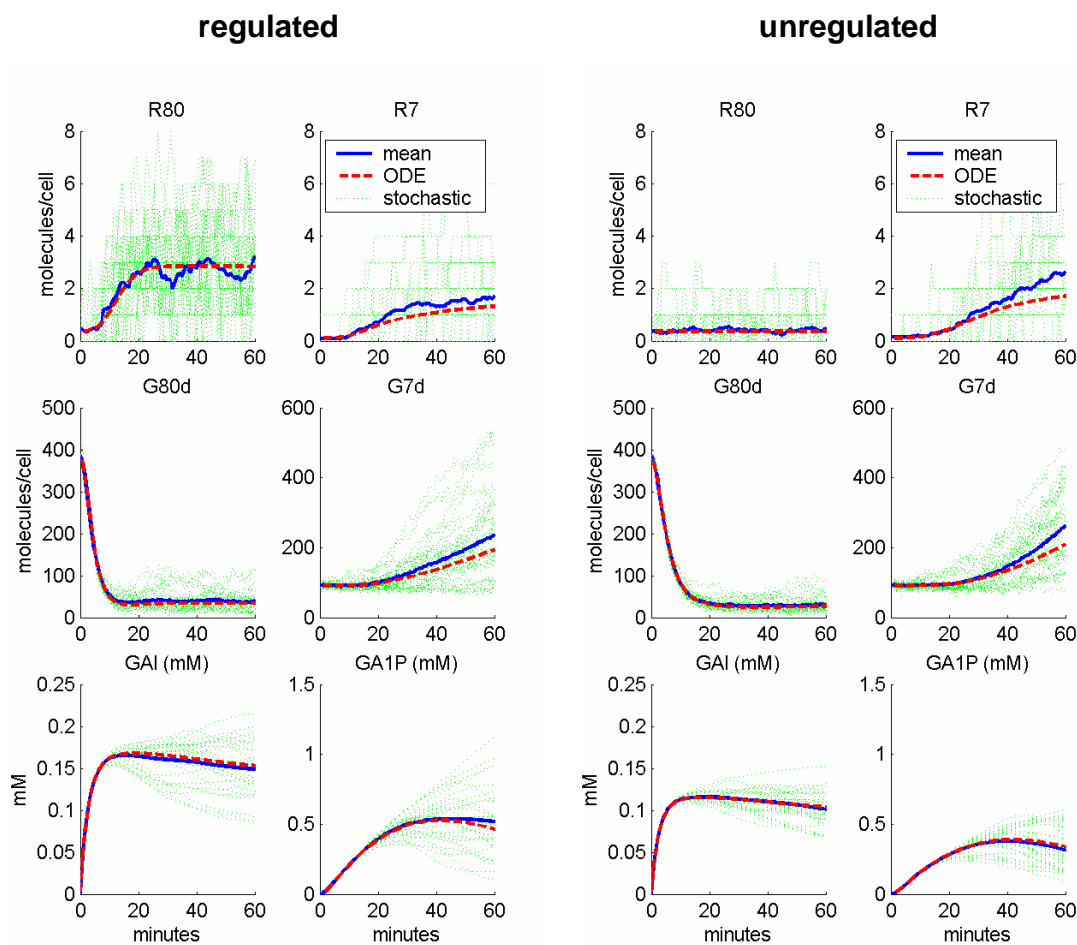
Figure 3. Comparison of the models with regulation (left panels) and without regulation (right panels) during induction. Top panels show mRNA for Gal80p and Gal7p, denoted R80 and R7; middle panels show protein dimers, denoted G80d and G7d; lower panels show metabolic quantities. Red dashed lines are from an ODE simulation, green lines are an ensemble of 30 stochastic simulations, blue lines are the mean of the stochastic simulations. The model is initiated in the non-induced condition, then external galactose is added corresponding to 10% induction (0.5 mM for the normal model, 0.32 for the version without regulation). The two versions of the model appear quite similar in behavior, except for R80 which is held fixed in the unregulated model.
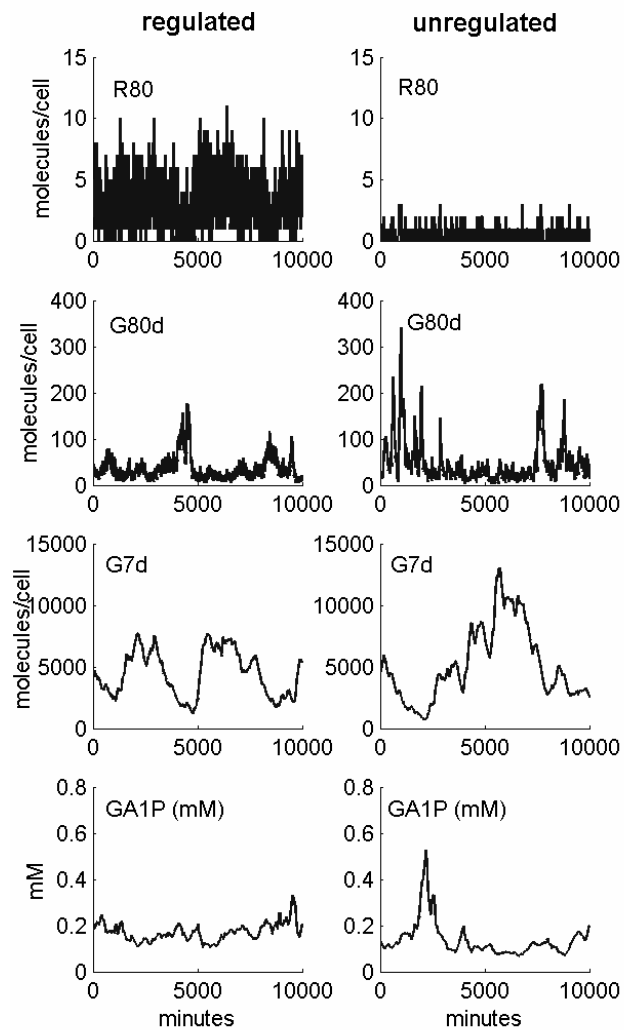
Figure 4. Comparison of the two versions of the model in the 10% induction state for a long time scale (the statistics are equivalent to shorter simulations over a number of cells). The model without regulation (panels on right) has increased noise in regulatory and structural proteins, as well as the intermediary GA1P (galactose 1-phosphate). See also Figure 9 for GA1P results over many simulations.
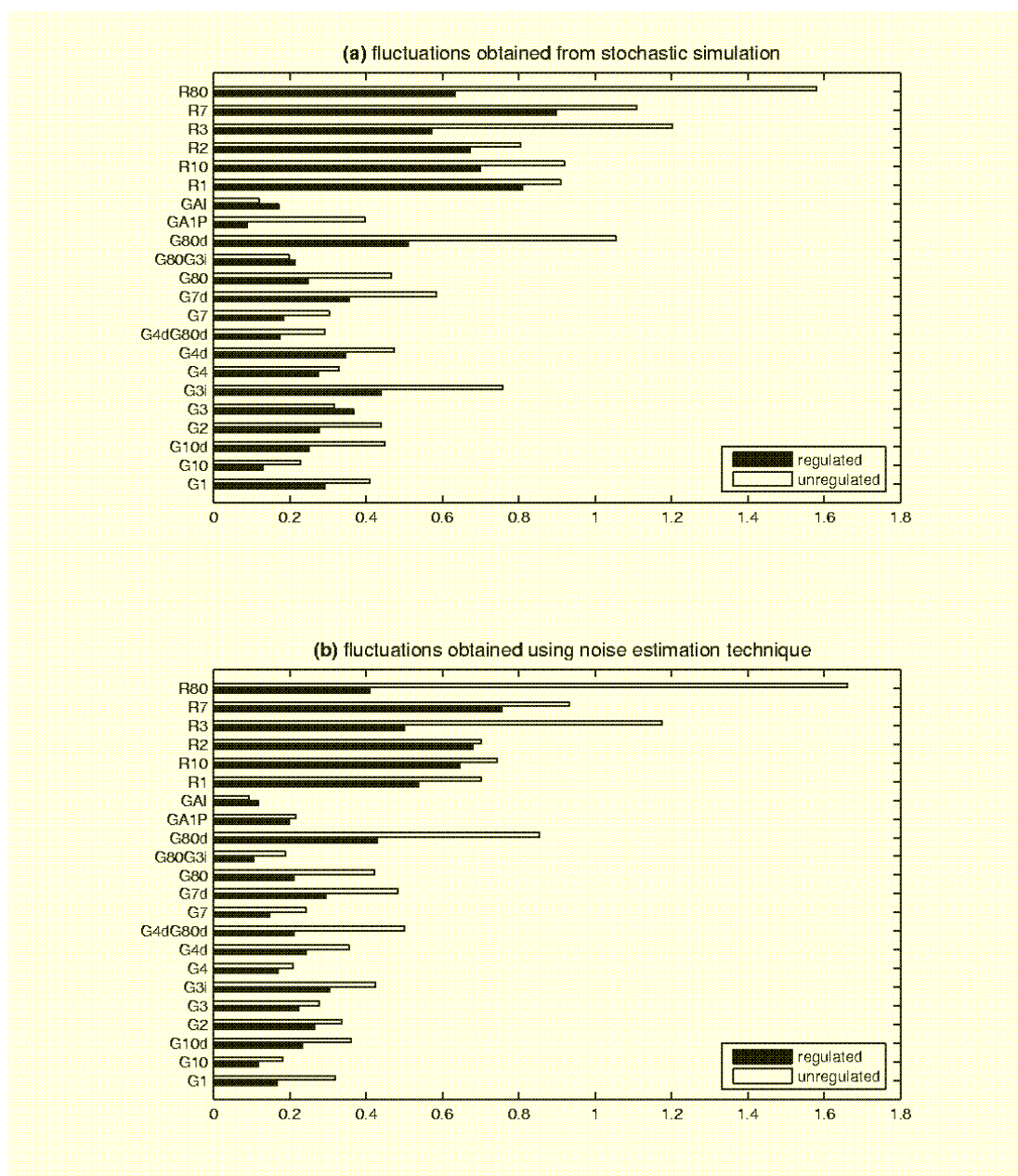
Figure 5. Comparison of stochastic variability for the two versions of the model in the 10% induction state. Shown is the dimensionless ratio of standard deviation to mean, calculated both from the stochastic simulation using the Tau-Leap solver in Dizzy (upper panel) and the estimation technique (lower panel). Both methods are in reasonable agreement, and indicate that the system without regulation has higher noise. Symbols for species is as in previous figures, but G3i refers to the activated form of Gal3p, and G3 to the inactive form.
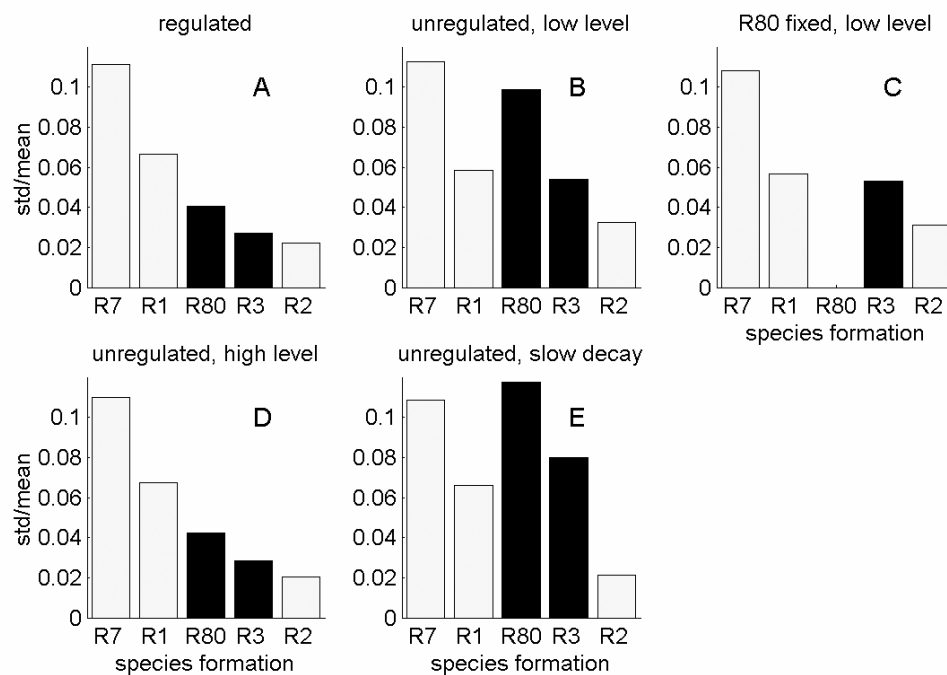
Figure 6. The reactions which contribute most to noise in GA1P, at the 10% induction level, are the formation and destruction of the mRNA in the horizontal axis of panel A (both contribute an equal amount). Noise is higher in the unregulated model (panel B) with Gal3p and Gal80p at low levels. If the mRNA for Gal80p is kept at a fixed level (no formation or decay reactions) in this unregulated model, then the contribution of those reactions is zero, but the contributions of the other reactions are little changed (panel C). If the unregulated model is adjusted so that Gal3p and Gal80p are kept at the high level of 10% induction (panel D), then the noise reduction is almost the same as in the regulated case. If the same effect is achieved by slowing the decay rate of regulatory mRNA, noise is enhanced (panel E).
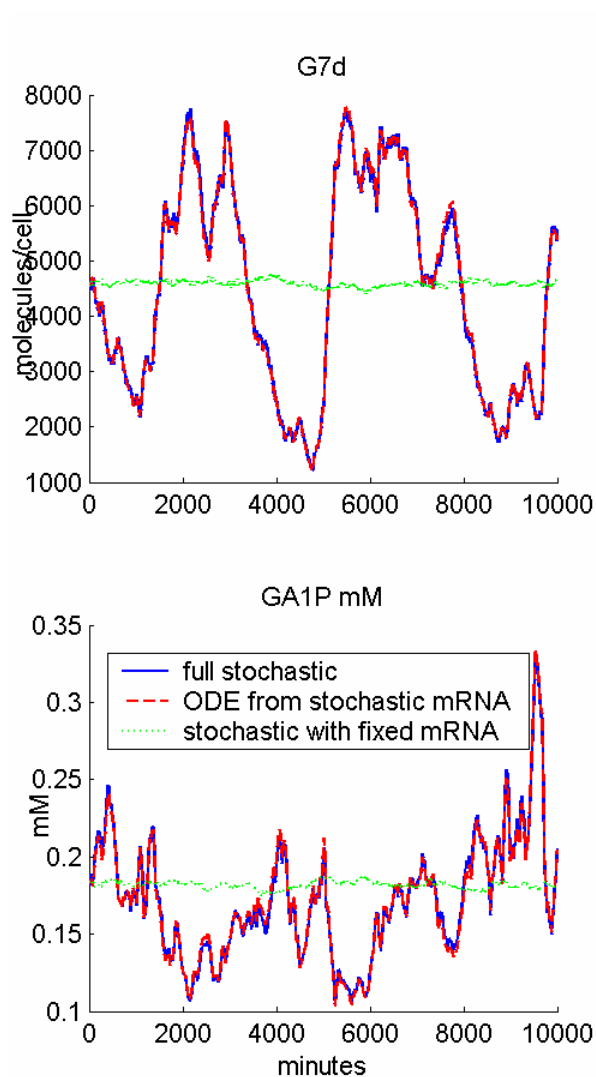
Figure 7. Plot illustrating that noise is due to mRNA fluctuations. Shown are the gal7p dimer (upper) and galactose 1-phosphate (lower) concentrations. Red dashed line shows an ODE simulation in which the stochastic mRNA levels are used as input, so noise in protein is solely a response to the changing mRNA levels. It is almost identical to the full simulation shown by the blue line (from Figure 5), indicating that system noise arises primarily from mRNA fluctuations. Green line shows a stochastic simulation in which mRNA levels are kept artificially fixed at their equilibrium values, thus eliminating their contribution to noise; the noise in the rest of the system is now negligible.
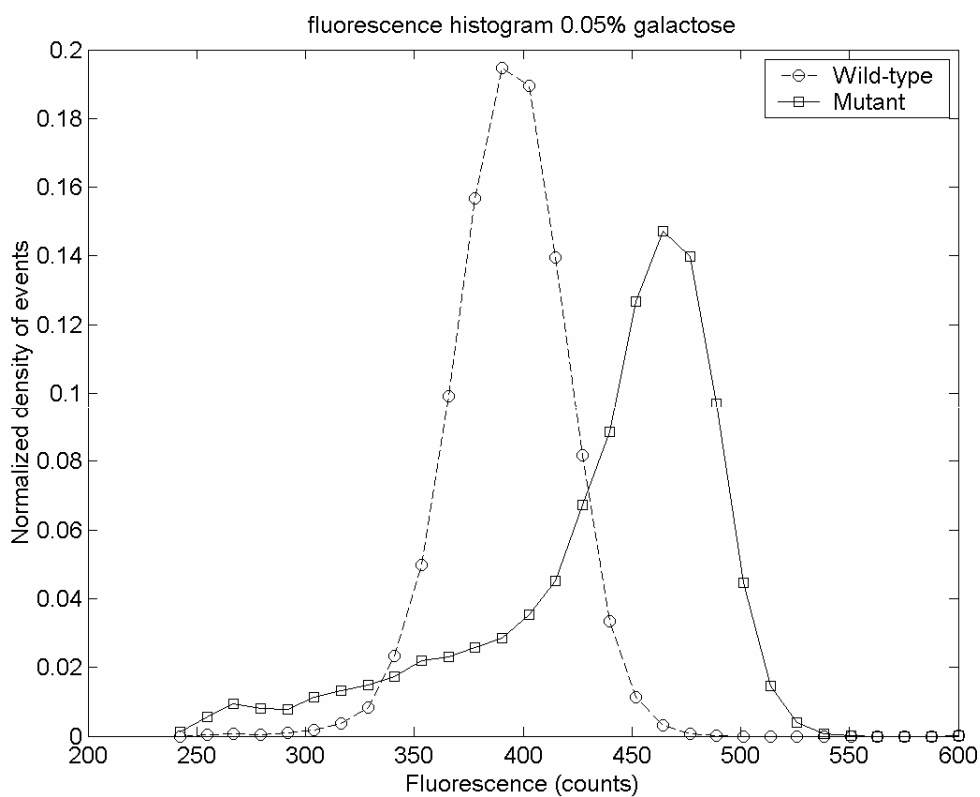
Figure 8. Histogram of adjusted fluorescence values for the selected flow cytometry events for the wild-type and mutant stains cultured in 0.05% galactose, which corresponds to approximately 50% induction. The histograms are normalized to have unit area, over the fluorescence range of the data points shown. The data was acquired six hours after the introduction of galactose. See Appendix B for details. The noise coefficients are 0.066 for the wild-type, and 0.126 for the mutant. The means are 393.1981 (wild-type) and 437.3525 (mutant).
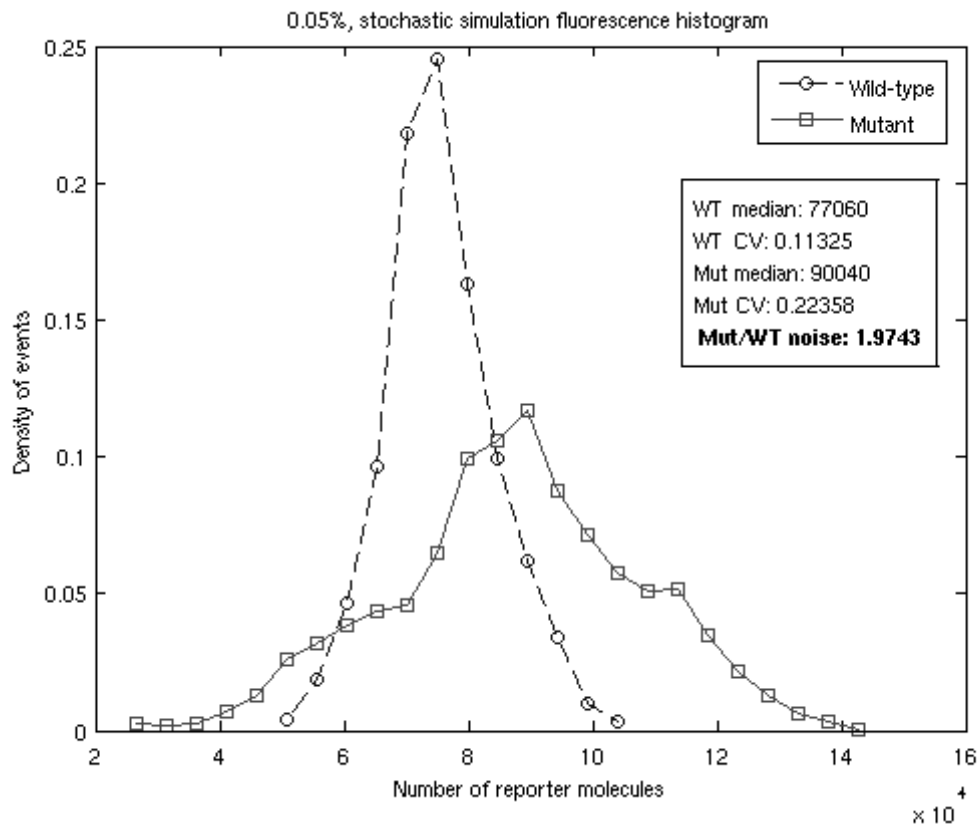
Figure 9. Histogram of the number of reporter molecules, from stochastic simulations of the two versions of the GAL model at steady-state. The noise coefficients are 0.113 for the wild-type, and 0.224 for the mutant. These compare with estimated values, from the noise estimation tool in Dizzy, of 0.105 and 0.266 respectively. The means are 7.54E4 (wild-type) and 8.73E4 (mutant). Stochastic results are from a 100-member ensemble of 1440 minute simulations, with 400 samples taken per simulation, performed using the Tau-Leap ("tauleap-complex") simulator in Dizzy version 2.1.1 with the relative error tolerance set to 0.0005.

TABLES

	Estimate	Stochastic	Experimental
Wild-type	0.105	0.113	0.066
Mutant	0.266	0.224	0.126
Ratio	2.53	1.974	1.91

Table 1. Coefficients of variation, defined as the standard deviation divided by the mean, for wild-type and mutant strains of yeast. Stochastic calculations were performed using the estimation tool in Dizzy, and the Tau-Leap simulator data of Figure 9. Experimental values are from the fluorescence data of Figure 8. The ratio between mutant and wild-type shows the noise enhancement due to elimination of the feedback loops.