# Measurement of in vivo protein binding affinities in a signaling network with mass spectrometry



| | |
|---|---|
| Journal: | *ACS Synthetic Biology* |
| Manuscript ID | sb-2016-00282y.R1 |
| Manuscript Type: | Article |
| Date Submitted by the Author: | 27-Jan-2017 |
| Complete List of Authors: | Gencoglu, Mumun; Universitat Basel Department Biozentrum<br>Schmidt, Alexander; University of Basel, Biozentrum<br>Becskei, Attila; Universität Basel, Biozentrum |
| | |

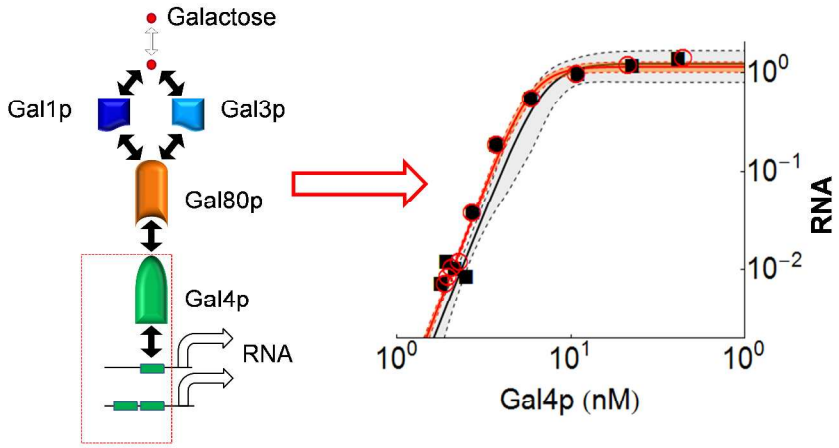
# Measurement of *in vivo* protein binding affinities in a signaling network with mass spectrometry

**Short title: Quantification of protein interactions**

Mumun Gencoglu, Alexander Schmidt & Attila Becskei

Biozentrum, University of Basel, Klingelbergstrasse 50/70, 4056 Basel, Switzerland

Corresponding author: attila.becskei@unibas.ch




Graphical Abstract

**Protein interaction networks play a key role in signal processing. Despite the progress in identifying the interactions, the quantification of their strengths lags behind. Here we present an approach to quantify the *in vivo* binding of proteins to their binding partners in signaling-transcriptional networks, by the pairwise genetic isolation of each interaction and by varying the concentration of the interacting components over time. The absolute quantification of the protein concentrations was performed with targeted mass spectrometry. The strengths of the interactions, as defined by the apparent dissociation constants ranged from subnanomolar to micromolar values in the yeast galactose signaling network. The weak homodimerization of the Gal4 activator amplifies the signal elicited by glucose. Furthermore, combining the binding constants in a feedback loop correctly predicted cellular memory, a characteristic network behavior. Thus, this genetic-proteomic binding assay can be used to faithfully quantify how strongly proteins interact with proteins, DNA and metabolites.**

**KEYWORDS:** *Saccharomyces cerevisiae, GAL1, GAL80, bistability, equilibrium dissociation constant, protein half-life.*

Protein interactions play a major role in shaping network behavior. To understand and predict the behavior of regulatory networks, the strength of the interactions has to be known. Substantial progress has been made to identify protein-protein interactions. Suitable methods are available for specific classes of interactions and experimental purposes[1, 2]. In comparison to the number of identified interactions, only a few of them have been quantified[3]. *In vitro* studies are limited by difficulties to express and purify proteins. Furthermore, binding conditions may be different *in vivo* and *in vitro*[4, 5].

The estimation of binding constants *in vivo* is hampered by the large number of free parameters in the interaction networks and collectively fitting them to experimental data yields large parameter uncertainties[6]. To circumvent this problem, we have broken down the system into smaller parts. We isolated subsystems in the galactose signaling network of the yeast *S. cerevisiae,* the interactions in which have been analyzed in detail and confirmed by multiple studies[7]. Therefore, it is suitable for quantitative studies. The galactose network comprises four proteins involved in signaling. The Gal4p activator is bound to genomic sites both in the presence

and absence of galactose. When Gal80p associates with Gal4p, transcription is inhibited. When galactose binds and activates Gal1p and Gal3p, they sequester Gal80p, which then cannot inhibit Gal4p[7] (Figure 1a). In this way, galactose elicits the expression of Gal4p target genes.

By the systematic variation of the concentrations of interacting proteins and by measuring the expression of target genes in each subsystem, we quantified the strength of interactions of proteins to DNA, proteins and galactose.

## RESULTS

### Design of the subsystems

We have isolated subsystems in the galactose signaling network. To study the interaction in the first subsystem, the Gal4p transcriptional activator – DNA binding, the concentration of the Gal4p was varied and the gene of its inhibitor, *GAL80*, was deleted (Figure 1a). To monitor the effect of Gal4p binding to DNA, we measured the expression of Gal4p target genes. Further subsystems were isolated proceeding from the transcription factor to galactose, which initiates the signaling (Figure 1a). It is important that proteins in each subsystem have single steady-state expression levels. This method takes advantage of the finding that protein-protein interaction networks display a power-law, i.e. most proteins have few, up to 2-3 interactions[8]. Thus, most of the time, few genes have to be deleted. With the availability of widely used genetic methods, genes can be now conveniently deleted and regulated in other organisms, as well[9, 10].

For the mass-spectrometric measurements, proteotypic peptides have to be detected. Since peptides of the Gal4, Gal3 and Gal80 proteins were not detected in wild type cells, we performed a search for peptides in cells with increased expression level (see Selection of Proteotypic Peptides by label-free quantification and directed LC-MS in Methods). Upon identification of the peptides, we aimed to assess the linear quantification range of our stable isotope dilution - selected reaction monitoring mass spectrometry (SID-SRM-MS) approach. For this purpose, we analyzed a dilution series of cellular extracts for each protein (see Methods).

In each subsystem, the protein concentration must be detected linearly. The SID-SRM-MS permitted a quantification with a very broad linear dynamic range, covering over three orders of magnitude (Figures 1b, S1 and S2). The lower limit of quantification was around 50 molecules per cell (Figure 1b), which corresponds to a mean cellular concentration of 2 nM.

In principle, protein concentration can be varied by modulating the activity of a promoter that controls the protein expression. Here, we opted for a different strategy: we shut down the expression and the protein concentration was varying gradually due to the natural decay process (Figures 1c and S3). This approach has two advantages. First, we obtained the rate constants for the protein decay, which is typically slow, with a half-life between 1 and 2 hours (see Supplementary methods, Fitting of parameter values). Second, the decay of the protein concentration has a simpler exponential profile in comparison to the more complex promoter response function, which improves the fitting (Figure 1c).

The Tet-Off system was used to shut down expression (see Methods). We found that the range of expression was higher than optimal for our binding studies, as indicated by the altered cell growth due to the highly expressed transcriptional activator Gal4p or by the too strong basal activation by Gal1p or Gal3p (Figure S4). For this reason, we inserted an RNA stem-loop upstream of the start codon to reduce translation. In this way, suitable ranges of protein concentrations were obtained.

**Weak homodimerization of Gal4p amplifies the glucose signal quadratically**

The mRNA expression of two Gal4p target genes, *GCY1* and *GAL7,* was measured. They have one and two Gal4p binding sites, respectively, and Gal4p binds as a dimer to its recognition sites[11]. Gal4p is the only known regulator of the *GCY1* promoter, while the *GAL7* promoter has a second known regulator, the Nrg1p repressor, which mediates the effect of glucose[12, 13]. After shutting off expression, the Gal4p concentration was declining from the starting concentration of 50 nM. The expression of *GAL7* remained high initially but started to reduce when the Gal4p concentration decreased to below 10 nM, which indicates that the promoter is saturated above 10 nM (Figures 1c, 2a and S5a, b).

Three binding parameters were fitted in this subsystem (Tables 1, S1 and S2, and Subsystem 1 in Supplementary methods): $K_D^{4/4}$ (Gal4 homodimerization), $K_D^{DNA}$ (Gal4p binding to DNA) and $c$ (enhancement factor in cooperative binding). As always, biochemical constants are apparent binding constants since they include the effects of ions and also other solutes in the cell. Gal4p bound the DNA binding sites with an apparent dissociation constant of $K_D$ = 15 nM. When one of the sites is already occupied at the *GAL7* promoter, the binding to the second site is enhanced by a factor of $c$ = 23. In comparison to the relatively strong Gal4p-DNA binding affinity, it was very surprising to find an approximately thousand times weaker binding in the homodimerization reaction, with a $K_D$ = 7.5 μM. The above measurements were performed with

cells grown in glucose. We also performed the same experiments in galactose (Figure S5). The fitted binding constants were very similar (Table 1), indicating that there is no uncharacterized effect of galactose on Gal4p beyond the effect mediated through the Gal1, Gal3 and Gal80 proteins.

The binding constants for dimerization and DNA binding displayed a considerable correlation when sampled from multiple initialization of the genetic algorithm (Figure S5c). This correlation arises because a weaker dimerization results in less dimers, which can be compensated by a stronger binding to the DNA However, the dimerization constant cannot assume a value less than 1 μM without diminishing the goodness of the fit because the nonlinearity is altered, and gene expression becomes less sigmoidal in response to Gal4p (Figure S5d). For the response to be sigmoidal with a quadratic dependence on the amount of a dimeric transcription factor produced, the dissociation constant of the homodimer has to be sufficiently high in relation to the protein concentration[14].

Figure 2b shows that the expression of *GAL7* increases in response to increasing Gal4p concentrations more steeply than in response to a hypothetical monomeric protein. This quadratic effect of dimerization becomes apparent when the response curves are compared to the gray lines whose unity slope in logarithmic plots indicates linear signal transmission. It can be seen that the monomeric transcription factor generates a higher than unity slope when the relative *GAL4* transcription rate is around 10 (Figure 2b). In this case, the cooperative binding to the promoter accounts for the nonlinearity of signal transmission. Thus, two processes, the weak dimerization and the cooperative binding, introduce nonlinearities into the system.

The effect of these nonlinearities can be tested in cells in which the *GAL4* is controlled by the endogenous promoter since glucose reduces the activity of the *GAL4* promoter[7]. Indeed, the Gal4p concentration decreased from 605 to 183 molecules per cell when the cells were exposed to glucose. This threefold change in Gal4p results in a tenfold change in the *GAL7* expression. A small fraction of this change is due to the direct effect of glucose on gene expression (1.6fold reduction), possibly mediated by the repressor Nrg1. However, much of the change is due to the quadratic effect of dimerization as evidenced by a comparison to a hypothetical monomeric transcription factor (Figure 2c, d). Thus, the dimerization amplifies the glucose signal mediated by Gal4p.

So far, we assumed that the Gal4p binding sites in the *GCY1* and *GAL7* promoters are of identical affinity. When each of these binding sites is inserted in the same promoter, they yield very similar expression levels at maximal induction by galactose[15], which suggests they are bound by Gal4 with similar or equal affinity. However, the concertation of the Gal4p (24 nM, 605 molecules/ cell) is high enough to saturate the high-affinity binding sites. This saturation may hide some of the variations in their affinities. Therefore, we fitted independent dissociation constants for the Gal4p binding sites in the *GCY1* and *GAL7* promoters. The goodness of the fit improved and the fitted binding constant indicated that the binding to a Gal4p recognition site in the *GCY1* promoter is around 3 times stronger in comparison to *GAL7*. We have taken into account this difference when the parameters in upstream subsystems were fitted (see next sections).

**Quantification of interactions between Gal80p, Gal1p, Gal3p and galactose**

To quantify the Gal4p – Gal80p interaction (Figure 1a, 3a), two upstream interactors, *GAL1* and *GAL3* were deleted. Gal80p inhibits Gal4p activity[16]. Upon shutting off the expression of Gal80p, the inhibition of the Gal4p target genes was progressively being relieved (Figure 3b). The values fitted in the first subsystem were fixed and the parameters for the new interactions were fitted (Figure 3a and Subsystem 2 in Supplementary methods). The binding constants were similar in cells grown in glucose and galactose. On the other hand, the interactions were considerably stronger in comparison to the first subsystem: the $K_D$ values for the Gal80 homodimerization and the Gal80-Gal4 heteterodimerization were in the subnanomolar range (Table 1, and Subsystem 2 in Supplementary methods).

To study the Gal1p-Gal80p and Gal3p-Gal80p interactions, *GAL1* or *GAL3* were deleted in the respective subsystems (Figure 1a). Gal1p and Gal3p are homologous proteins. When galactose binds to Gal1p and Gal3p, these activated complexes retain Gal80p in the cytoplasm, which relieves the inhibition of Gal4p (Figure 3a)[7]. We also wanted to quantify the interaction of galactose to Gal1p and Gal3p. To ensure that a single steady-state exists in this subsystem, *GAL2* was deleted because it is under positive feedback control[17]. In the absence of the high affinity transporter Gal2p, the other low-affinity channels facilitate the transport of galactose into the cells.

We used protein expression with and without translational inhibition by stem-loops to span the entire range of concentrations observed in wild-type cells (Figure S4d, e), which

corresponds to three orders of magnitude. As expected, the Gal4p concentration remained approximately constant, while the Gal80p concentration increased slightly at high Gal3p concentrations. This gradual change is due to the negative feedback mediated by a single Gal4p binding site in the *GAL80* promoter. The binding of Gal1p and Gal3p to Gal80p was assessed in the presence of saturating concentration of galactose (0.5%) (Figure 4a, b). The fitting was performed to data representing five system variables (the Gal4, Gal80 and Gal1 proteins, and the *GAL7* and *GCY1* mRNAs). In addition to the binding constant, the production rates of Gal80p and the decay rate of Gal1p were also fitted. The fit was good for the Gal1p-Gal80p interaction and less good for the Gal3p – Gal80p interaction.

Gal3p bound Gal80 around 10 times stronger than Gal1p. The dissociation constant for the Gal1p binding to Gal80 was around 10 nM, with a small error of the fitting (see Subsystem 3 in Supplementary methods). The basal production rate of Gal80p ($\beta_{P80}$) had the largest relative standard error of the fitting.

By adding galactose to the medium at intermediate concentration, we aimed to quantify the galactose binding to Gal1p (see also Subsystem 3 in Supplementary methods). The intracellular galactose reached steady state after around 5 hours and the intra- and extracellular concentration were in equilibrium (Figure 4c, d). The dissociation constant of the galactose binding to Gal1p was in the low milimolar range (Table 1).

**Validation of the parameters by prediction the behavior of a feedback loop**

At this point, we completed the quantification of interactions starting at the signal initiation by galactose and ending at the binding of Gal4p to the DNA. The protein binding constants range from subnanomolar to micromolar values, while the binding to the metabolite is in the low milimolar range. To validate the above measurements, we tested how the apparent binding constants collectively predict a system behavior – cellular memory in the *GAL1* feedback loop (Figure 5a, b). Positive feedback loops have the potential to generate bistability, which permits cells to remember their prior exposure to galactose[18]. While the potential of a system to display bistability is determined by the nonlinearities due to the binding interactions, the translation rate of the *GAL1* mRNA is also required to predict the absolute range of galactose concentration over which bistability is displayed. After fitting this rate (Table S1), the solution of the model revealed a broad range of galactose concentration over which two stable expression states exist, i.e. bistability (at the entire examined galactose concentration above 0.17%, see full

cyan line in Figure 5a). As control of this prediction, we calculated the bistability range when Gal4p dimerizes more strongly. To compensate the higher concentration of the active dimeric Gal4p, its binding affinity to the DNA was reduced. With these parameter values, Gal4p generates a less sigmoidal and hence less nonlinear response (see magenta line in Figure S5d), which is expected to yield a narrower range of bistability. Indeed, bistability was restricted to a narrower range, at around 1% of galactose (Figure 5c).

We also wanted to see how the uncertainty of the fitted parameter values affects the prediction of bistability. The production rate of the inhibitor is known to have a major impact on the bistability range[14], and it was this parameter ($\beta_{P80}$) that had the largest relative uncertainty in the Gal1p-Gal80p subsystem. Therefore, we examined a model with a lowest realistic Gal80p production rate ($\beta_{P80}$) by subtracting the standard error of the fitting from the fitted value. Even in this case bistability was present but had a narrower range (between the dashed cyan lines in Figure 5a).

Bistability is typically evidenced by a bimodal distribution of gene expression, so that most of the cells are expressing either low (OFF-cells) or high (ON-cells) levels of a protein or a reporter gene (Figure S6). A common measure of memory, the memory index, is the difference of the ON cell percentage of cells in cultures that were pre-exposed to media with and without galactose (see Memory experiments in Methods). The degree of bimodality depends on noise and slow transient processes and not only on bistability. Therefore, there is no one-to-one correspondence between the memory index and bistability[19, 20] Generally, the largest memory index is observed between the bistability boundaries because the transitions between the two states are slow. We observed high memory indices at a broad range of galactose concentration 3 days (72 h) after the start of the memory experiment (Figure 5d). Even after 96 hours, a well-defined peak in the memory index was observed. The memory declines faster at the higher galactose concentration because of the faster transitions from the OFF state, which commonly arises in bistable systems due to noise or slow transient processes[19, 20]. At the galactose concentration with peak memory, 50% of cells remembered whether or not they had been exposed to galactose 4 days earlier. Cells undergo around 100 cell divisions in this period.

The bistability range predicted with the "mutant" model, with a strongly dimerizing Gal4p, does not overlap with the range of galactose concentrations with the highest memory indices which is particularly evident at 96 h (magenta line in Figure 5d), underscoring the role of this *in silico* negative control.

The three points with the highest memory indices overlapped with the predicted bistability boundaries independently of which Gal80p production rate was used (cyan full and dashed lines in Figure 5d). The prediction is in good agreement with the measurements especially because this was an absolute prediction without fitting any parameter to the observed range of galactose concentrations in which memory was observed.

## DISCUSSION

To compare our *in vivo* constants to the published *in vitro* binding constants, important details have to be considered. Nearly all studies that quantified interactions with Gal4p use only a fragment of Gal4p, which includes the DNA binding and the dimerization domains. This comprises typically a 100 amino acid long fragment of the 881 amino acid long protein. The difficulty of the expression and purification of the full Gal4p may be related to the fact that overexpression of transcriptional activators is often toxic to the cell, as also evidenced by the fact that we had to use translational inhibition to express Gal4p. The values reported for the Gal4p (fragment) – DNA binding scatter over two orders of magnitude, from 0.5 to 25 nM[11]. We obtained a $K_D$ of 4 to 15 nM (Table 1). Interestingly, the weak homodimerization of Gal4 ($K_D$ = 7.5 to 8.5 μM) is close to the reported value ($K_D$=20 μM*) in vitro*[21].

It is also important to note that binding constants can be fairly compared when the conditions and equations used for the fitting are known. This is true even for the comparison of different *in vitro* measurement. Many transcription factors can bind to the DNA only as dimers; yet, *in vitro* binding constants are often fitted assuming that the total protein can bind to the DNA. This approximation can be a valid assumption if the dimerization is strong and the equilibrium is shifted to a dimeric form or the protein concentration is high enough; thus, the concentration of the dimer is approximately equal to the total protein. However, this assumption is not valid when dimerization is weak and lower protein concentrations are used in the assay. Thus, the reported weak dimerization of Gal4p[21] would require the fitting of equations that include dimerization. However, this was not the case in the relevant *in vitro* studies of Gal4p – DNA binding[11, 22-25].

Furthermore, two or more binding constants can be correlated and have a relatively broad distribution of realistic values even when fitted to data obtained from *in vitro* measurements of simple binding reactions[26].

Our Gal4 –DNA binding constant is defined for the binding between the Gal4 dimer and the DNA. Thus, the value of this constant (4 - 15 nM) is likely to reflect weaker binding than the weakest reported *in vitro* binding (20 nM) because the *in vitro* constant is defined for the total Gal4p. There are multiple reasons for differences between *in vivo* and *in vitro* binding constants. Solutes, ions and macromolecular crowding *in vivo* are typically different from the *in vitro* conditions. Protein association in cells can occur both in mature form and in partially folded form during translation. Binding of transcription factors is measured with naked DNA in vitro while DNA is wrapped around histones in the cell. For example, the nucleosomes can accelerate the dissociation of the proteins bound to DNA[27].

For the Gal4p-Gal80p interaction, reported *in vitro* values are also scattered over two orders of magnitudes[16] (and references therein). Our *in vivo* value is closest to the strongest reported *in vitro* value ($K_D$ = 0.3 nM)[16], which is also similar to the value of the *K. lactis* Gal80p ($K_D$ = 1 nM)[28].

A recent *in vitro* study showed that the binding of Gal1p to Gal80p is ten times weaker than the binding of Gal3p to Gal80p[29], which is very similar to our findings *in vivo* (Table S1). However, the absolute values differ since the binding *in vivo* is 40 times stronger than *in vitro*. It is important to note that different binding mechanisms and equations were used to fit the binding constants in our and their studies. In our study we allowed the binding of Gal1 and Gal3 only to the monomeric form of Gal80p, and not to the dimeric form. On the other hand, only a single form of Gal80p was considered in the in vitro study (i.e. the total protein amount)[29]. *In vitro* studies suggests that Gal1p/Gal3p can form only 1:1 (i.e. monomer:monomer) complexes with Gal80[30, 31]. It remains to be determined if these heterodimeric complexes can also form as a result of dissociation of the Gal80p dimer when Gal1p/Gal3p binds to it.

The fitting of the parameters in the Gal1p/Gal3p -Gal80p subsystem depends on the mechanisms in the downstream subsystems. The downstream subsystems (Gal4p – DNA and Gal4p – Gal80p) are more directly read out by the system output, gene expression (i.e. the mRNA); therefore, the parameter values fitted in these subsystems can be more easily compared to the *in vitro* values.

We confirmed the fitted parameter values with two experiments in this study: the amplification of the glucose signal in the Gal4p-DNA subsystem, and bistability based on the parameters fitted in all three subsystems.

Protein affinities play an important role in shaping the network behavior[32]. Our work revealed a nonlinear effect of binding reactions, as evidenced by the signal amplifying effect of weak Gal4p homodimerization. A majority of proteins, particularly transcription factors, form dimers[33]. Therefore, we expect that similar effects of homodimerization will play an important role of network behavior. With the availability of detailed metabolic measurements, it will be possible to understand and model the behavior of the galactose network[34, 35].

Mass spectrometry has been used to quantify interactions between proteins and small molecule inhibitors in cell extracts[36]. With the genetic – proteomic method presented in this study, interactions in transcriptional – signaling networks can be quantified *in vivo*. We expect that optimal assays will be employed for specific experimental purposes and interaction types, as it has been done to identify protein-protein interactions. Tagging of proteins with fluorescent proteins has been used to identify interactions and progress has been made to use this method to quantify protein - protein interactions *in vivo*[37, 38]. There are several advantages of our genetic-proteomic method. It requires no tagging and interactions between native proteins can be quantified. Secondly, the power of mass spectrometry permits detection of weakly expressed proteins, which are below the florescence detection limit (e.g. Gal4)[39]. Furthermore, multiple proteins can be detected simultaneously (e.g. Figure 3b, c), with an upper limit of around 50 proteins[40]. Importantly, the interactions can be studied between protein homodimers and heterodimers, protein-DNA pairs and protein-metabolite pairs. This is crucial since quadratic signal amplification and other nonlinear effects of protein interactions arise due to the joint effect of protein homo- and hetero-dimerization. Thus, we expect that our genetic – proteomic approach will be particularly advantageous to quantify interactions with strong nonlinear effects on network behavior.

## METHODS

See the Supplementary methods for detailed protocols for the mass-spectrometry (Sample Preparation for LC-MS Analysis, SRM-Assay Development and Protein Quantification), RNA isolation and quantitation by quantitative real-time PCR (qPCR) and the flow cytometric measurements of fluorescent proteins ($P_{GAL1}$::GFP reporter).

**Selection of Proteotypic Peptides by label-free quantification and directed LC-MS.**

To detect efficiently peptides, we obtained extracts from cells in which the expression of GAL genes was controlled by the Tet-off system (see Construction of strains for controllable protein expression). The concentration of the Gal4, Gal3 and Gal80 proteins in these strains was higher than in the wild type strains. To set up highly sensitive and specific SRM assays for protein quantification, we selected five proteotypic peptides (PTPs) for each protein as described [41]. In brief, we excluded peptides which sequence matched to multiple proteins in the database and peptide containing missed cleavages, glutamine at the n-termini, more than 20 amino acids and, if at least five PTPs were found, methionine. We then ranked the filtered peptides according to their precursor ion MS-intensity determined in a label-free quantification (LFQ) experiment from whole cell extracts and selected the five peptides with the highest MS-response per protein. See Supplementary methods for details.

Since we were not able, due to sensitivity issues, to identify five PTPs for the proteins Gal80, Gal4 and Gal3 using LFQ, we additionally carried out an inclusion mass list (INL) experiment to direct MS-sequencing to the missing tryptic peptides of these three proteins[42]. Therefore, we generated a list of precursor ion masses for all missing peptides (Data S2), imported it into the MS and re-analyzed the same peptide samples as described above with the following changed parameters: The preview scan option was disabled, the resolution for MS1 scans was reduced to 30,000 FWHM and MS/MS scans were also acquired in the orbitrap at a resolution of 7,500 FWHM using a fill time of 100 ms. The acquired raw-files were converted to the mascot generic file (mgf) format using the msconvert tool (part of ProteoWizard, version 3.0.4624 (2013-6-3)). With the MASCOT algorithm (Matrix Science, Version 2.4.0), the mgf files were searched using the same settings as for LFQ above, only the fragment ion tolerance was set to 0.02 Da. Next, the database search results were imported to the Scaffold software (version 4.3.2, Proteome Software Inc., Portland, OR) and the protein false identification rate was set to 1% based on the number of decoy hits. Specifically, peptide identifications were accepted if they could achieve an FDR less than 1.0% by the scaffold local FDR algorithm generating a list of confidently identified peptides (Data S3). We then ordered the identified peptides by decreasing number of identified spectra and selected the highest scoring peptides. By this, we were able to identify 5 PTPs per protein with the exception of Gal4 (only 4 peptides were identified). Since this protein is a very low abundant and crucial for the analysis, we selected 13 additional peptides (Data S4) that were predicted to be highly amenable for LC-MS analysis according to the final suitability score provided by

PeptideAtlas[43]. In total, we ordered 37 synthetic heavy reference peptides (JPT Peptide Technologies GmbH, Berlin, Germany) for SRM assays development and selected 11 of these for absolute quantification of the 5 proteins of interest (Data S5).

**Determination of absolute molecule numbers.**

All raw-files were imported into Skyline for protein / peptide quantification[44] following best practices for confident peptide quantification[45]. Herein, integrated peak areas of the 5 most intense transitions exceeding the precursor ion mass associated to the reference (heavy) and endogenous (light) peptide were summed, respectively. From the obtained light-to-heavy peptide ratios, the absolute endogenous peptide concentration was determined in fmol/µg. Based on the number of cells counted for each sample by flow cytometry and assuming complete protein extraction and digestion efficiency, absolute abundances for the selected proteins (in copies/cell) were calculated (see Equation (1)):

$$\text{Protein copy number [molecule /cell]} = \frac{\frac{N_A n(H) PA(L)}{PA(H)} Nf}{N(cell)} \quad (1)$$

$N_A$ is the Avogadro number, *N(cell)* is the number of cells in the sample for the measurement, $n(H)$ [mole] is the amount of the heavy peptide added to the digested sample, PA is the peak area measured for each peptide in the MS/MS. *H* and *L* refer to the synthesized heavy peptide and the light peptide derived from the endogenous protein, respectively. *Nf* is the normalization factor to account for variability in the sample injected into the MS/MS. The determination of the normalization factor is described in the next section.

**Determination of the normalization factor.**

Variability in the sample injected into the MS/MS can arise due to the following steps during the sample preparation: (1) cell lysis, (2) protein digestion and peptide recovery and (3) purification of peptides with C-18 column. To identify the causes of peptide loss, we prepared a dilution series. For this purpose, cells expressing high levels of Gal1p were mixed with Δ*gal1* cells to obtain a dilution series and the Gal1 protein amount was measured (Figure S2a).

Firstly, the efficiency of cell lysis was assessed by examining cells under microscope before and after lyses. It was always above 95%. Secondly, we assessed the loss of peptide during C-18

column purification. For this purpose, heavy Gal1p peptides were added to each sample in equal amount before C-18 purification. Their peak areas should be equal in each sample (Figure S2b). Indeed, there was a very small experimental variation in the detected amount of the heavy peptide. Thus, the variation must arise between cell lysis and C-18 column purification, during protein digestion and peptide recovery ($2^{nd}$ stage).

In order to correct for incomplete protein digestion and peptide loss during processing, we measured the endogenous actin amount (Figure S2a, b). Indeed, it displayed large variations similar to Gal1p. To correct for the variations, we prepared an ideally processed sample by using excess trypsin and used a macro C18 column to avoid loss of the peptide and saturation of the column. The amount of Act1p measured from this sample is the expected (ideal) actin amount (Figure S2b)**.** Thus, the normalization factor is given by:

$$Nf = \frac{PA(Act, exp)}{PA(Act, obs)}$$

The expected and observed PA for the actin peptides are denoted by PA(Act, exp) and PA(Act, obs). Using this normalization factor increased considerably the linearity of the dilution curve (Figures 1b and S2c).

**Construction of deletion strains.**

All yeast strains are derivatives of *S. cerevisiae* BY4741 (*MAT*a his3Δ1; leu2Δ0; met15Δ0, ura3Δ0) and BY4742 (*MAT*α, his3Δ1; leu2Δ0; lys2Δ0; ura3Δ0) (EUROSCARF, Frankfurt). For most experiments, strains were required, in which two genes were deleted. To obtain these strains, we mated two haploid strains with single gene deletions, each of them deleted with *kanMX*. The resulting diploid strains were sporulated randomly and haploid strains with a *MAT*a his3Δ1; leu2Δ0; lys2Δ0; ura3Δ0 genotype were selected, using alpha-factor to determine the mating type. The markers of the gene deletion were confirmed with PCR. The resulting background strains (YmmnH02 (*MAT*a his3Δ1; leu2Δ0; lys2Δ0; ura3Δ0, Δ*gal1::kanMX* Δ*gal3::kanMX),* YmmnH03 (*MAT*a his3Δ1; leu2Δ0; lys2Δ0; ura3Δ0, Δ*gal2::kanMX* Δ*gal3::kanMX)*, YmmnH04 (*MAT*a his3Δ1; leu2Δ0; lys2Δ0; ura3Δ0, Δ*gal1::kanMX* Δ*gal2::kanMX* )) were used for the further strain constructions (Table S3).

**Construction of strains for controllable protein expression.**

First, we transformed the background strains with the pRS305::$P_{CLN3}$::tTA plasmid. Second, the pRS303::$P_{GAL1}$::GFP was integrated, which serves as a reporter to read out the activity of the GAL network. Last, pRS303::P*[tetO2]2CYC1* – GAL gene constructs were integrated, which together with tTA constitute the Tet-Off system to control the expression of the GAL genes[46]. Around 300 bp of the 5' fragment of the GAL genes was cloned into the plasmids to provide a homologous sequence long enough for chromosomal integration. Strains with different protein expression levels were screened to find the optimal expression range spanned by the expression with and without doxycycline (Figure S4). Protein expression in different strains was altered with stem-loops inserted upstream of the GAL genes, which reduce the translational efficiency[47]. A U3 intron was inserted downstream of the start codon to provide the option to measure the mature mRNA by qPCR. For the expression of *GAL1*, an expression cassette with a P*[tetO2]4inGAL1* promoter was also used[15].

**Growth conditions.**

All cultures were grown at 30°C in a synthetic raffinose medium containing yeast nitrogen base, 2% filter-sterilized raffinose and 0.005% glucose as carbon source and the SC-Leu/-His/-Ura drop-out supplement, unless otherwise indicated. This basic raffinose medium was supplemented with inducers, glucose or galactose as indicated at the specific experiments. Cultures were grown either in 96 well plates or Erlenmeyer flasks. Both the growth rate and induction time of GFP were similar in these conditions.

**Memory experiments.**

Cellular memory was assayed by hysteresis experiment[18]. First, cells were pre-incubated in 0 or 2% galactose for 5 days to obtain un-induced and fully induced cells. These cells were then inoculated into media containing a range of galactose concentrations so that cells with different pre-incubation histories were grown in identical conditions. The cultures were diluted three times a day during the memory experiment to keep the $OD_{600}$ below 1.0. At the time of measurement by flow cytometry, the cell density ($OD_{600}$) was between 0.2 and 0.6. The histograms of the fluorescent gene reporter were plotted and the OFF and ON cells were identified (Figure S8). The difference between the ON cell percentages between the cultures with the two different preincubation histories defines the memory index: ON% (0% Gal pre-incubation) – ON% (2% Gal pre-incubation)[15].

**Shutting off gene expression to vary the protein concentration.**

The cells were cultured overnight in media supplemented with 0.5 % galactose or 0.2% glucose. They were transferred to a refreshment medium, starting at an $OD_{600}$ ~ 0.1 in 60 ml conical flasks. When the cell density reached $OD_{600}$~0.5, transcription was shut off by adding 10 µg/µl of doxycycline. 5-5 ml cultures were collected at each indicated time point (1.5, 3, 4. 5, 6, 7.5, 9, 10.5, 12, 13.5 and 24 hours) for RNA and protein measurements. The cell density was kept below an $OD_{600}$ of 1.0 throughout the course of the experiment.

**Measurement of intracellular galactose concentration.**

Cells were grown in large volumes (50 ml) in order to get a pellet of 10-20 µl since the pellet volume is around 0.1% of the total culture at $OD_{600}$ = 1. We extracted intracellular galactose with a Boiling Ethanol extraction method as described [48] with minor modifications. Briefly, the culture flasks were transferred to 50% methanol (kept on dry ice) and were centrifuged for 5 min at -20°C. The resulting pellet was dissolved in 2 ml boiling 75% ethanol, i.e. preheated to 95°C. The Eppendorf tubes were kept at 95°C for 5 min. The tubes were transferred to heated vacuum and rotated until the debris became dry. The debris was dissolved in 100 $\mu$l of water. To measure the intracellular galactose concentration, we used the Raffinose/D-Galactose assay kit (Megazyme).

**Data fitting.**

First, the RNA half-life was fitted by simple exponential models. After having confirmed the short half-life, the protein decay rate constant and the interaction parameters were fitted with MathWorks® product SimBiology in MATLAB. The best fits of the parameter values were found using a genetic algorithm (GA, in the Global Optimization Toolbox). The algorithm repeatedly modifies a population of individual solutions. At each step, GA randomly selects individuals from the current population and uses them as parents to produce the children for the next generation. Over successive generations, the population "evolves" toward an optimal solution. The retrieved parameter values were typically distributed over a narrow range. We selected the one with the lowest error of the fitting. The binding constants fit in a given subsystem were fixed in the subsequent fittings of the upstream subsystems.

**Supporting Information:** Supplementary methods, Figures S1-S6, Tables S1-S3, Data files S1-S6.

**Author contributions:** A.B and M.G. designed the study, analyzed the data and wrote the paper. MG performed the experiments and data fitting. AS provided mass spectrometry expertise and contributed to the experimental design.

## ACKNOWLEDGEMENTS

We thank Matthias Heinemann, Amirhossein Hajihosseini and Takeo Wada for helpful discussions, Sylvia Voegeli and Vincent Jaquet for the PGAL1-YFP plasmid and experimental help. M. G. was supported by a SystemsX IPhD fellowship. This work was supported by the Systems X StoNets and SNSF grants.

## REFERENCES

[1] Snider, J., Kotlyar, M., Saraon, P., Yao, Z., Jurisica, I., and Stagljar, I. (2015) Fundamentals of protein interaction network mapping, *Molecular systems biology 11*, 848.

[2] Vidal, M., and Fields, S. (2014) The yeast two-hybrid assay: still finding connections after 25 years, *Nat Methods 11*, 1203-1206.

[3] Kastritis, P. L., and Bonvin, A. M. (2013) On the binding affinity of macromolecular interactions: daring to ask why proteins interact, *J R Soc Interface 10*, 20120835.

[4] Oda, M., and Nakamura, H. (2000) Thermodynamic and kinetic analyses for understanding sequence-specific DNA recognition, *Genes Cells 5*, 319-326.

[5] Garcia-Contreras, R., Vos, P., Westerhoff, H. V., and Boogerd, F. C. (2012) Why in vivo may not equal in vitro - new effectors revealed by measurement of enzymatic activities under the same in vivo-like assay conditions, *The FEBS journal 279*, 4145-4159.

[6] Gutenkunst, R. N., Waterfall, J. J., Casey, F. P., Brown, K. S., Myers, C. R., and Sethna, J. P. (2007) Universally sloppy parameter sensitivities in systems biology models, *PLoS computational biology 3*, 1871-1878.

[7] Bhat, P. J. (2008) *Galactose Regulon of Yeast. From Genetics to Systems Biology.*, Springer, Heidelberg.

[8] Janjic, V., Sharan, R., and Przulj, N. (2014) Modelling the yeast interactome, *Sci Rep 4*, 4273.

[9] Dominguez, A. A., Lim, W. A., and Qi, L. S. (2016) Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation, *Nat Rev Mol Cell Biol 17*, 5-15.

[10] Bashor, C. J., Horwitz, A. A., Peisajovich, S. G., and Lim, W. A. (2010) Rewiring cells: synthetic biology as a tool to interrogate the organizational principles of living systems, *Annu Rev Biophys 39*, 515-537.

[11] Hong, M. Q., Fitzgerald, M. X., Harper, S., Luo, C., Speicher, D. W., and Marmorstein, R. (2008) Structural basis for dimerization in DNA recognition by Gal4, *Structure 16*, 1019-1026.

[12] MacIsaac, K. D., Wang, T., Gordon, D. B., Gifford, D. K., Stormo, G. D., and Fraenkel, E. (2006) An improved map of conserved regulatory sites for Saccharomyces cerevisiae, *BMC Bioinformatics 7*, 113.

[13] Zhou, H., and Winston, F. (2001) NRG1 is required for glucose repression of the SUC2 and GAL genes of Saccharomyces cerevisiae, *BMC Genet 2*, 5.

[14] Majer, I., Hajihosseini, A., and Becskei, A. (2015) Identification of optimal parameter combinations for the emergence of bistability, *Physical Biology 12*, 066011.
[15] Hsu, C., Scherrer, S., Buetti-Dinh, A., Ratna, P., Pizzolato, J., Jaquet, V., and Becskei, A. (2012) Stochastic signalling rewires the interaction map of a multiple feedback network during yeast evolution, *Nat Commun 3*, 682.
[16] Melcher, K., and Xu, H. E. (2001) Gal80-Gal80 interaction on adjacent Gal4p binding sites is required for complete GAL gene repression, *Embo J 20*, 841-851.
[17] Hawkins, K. M., and Smolke, C. D. (2006) The regulatory roles of the galactose permease and kinase in the induction response of the GAL network in Saccharomyces cerevisiae, *J Biol Chem 281*, 13485-13492.
[18] Acar, M., Becskei, A., and van Oudenaarden, A. (2005) Enhancement of cellular memory by reducing stochastic transitions, *Nature 435*, 228-232.
[19] Maleki, F., and Becskei, A. (2016) An open-loop approach to calculate noise-induced transitions, *J Theor Biol 415*, 145-157.
[20] Hsu, C., Jaquet, V., Maleki, F., and Becskei, A. (2016) Contribution of Bistability and Noise to Cell Fate Transitions Determined by Feedback Opening, *Journal of molecular biology 428*, 4115-4128.
[21] Gadhavi, P., Morgan, P. J., Alefounder, P., and Harding, S. E. (1996) A physico-chemical investigation of the self-association of the DNA binding domain of the yeast transcriptional activator GAL4, *Eur Biophys J 24*, 405-412.
[22] Liang, S. D., Marmorstein, R., Harrison, S. C., and Ptashne, M. (1996) DNA sequence preferences of GAL4 and PPR1: how a subset of Zn2 Cys6 binuclear cluster proteins recognizes DNA, *Molecular and cellular biology 16*, 3773-3780.
[23] Kang, T., Martins, T., and Sadowski, I. (1993) Wild type GAL4 binds cooperatively to the GAL1-10 UASG in vitro, *The Journal of biological chemistry 268*, 9629-9635.
[24] Vashee, S., Xu, H., Johnston, S. A., and Kodadek, T. (1993) How do "Zn2 cys6" proteins distinguish between similar upstream activation sites? Comparison of the DNA-binding specificity of the GAL4 protein in vitro and in vivo, *The Journal of biological chemistry 268*, 24699-24706.
[25] Parthun, M. R., and Jaehning, J. A. (1990) Purification and characterization of the yeast transcriptional activator GAL4, *The Journal of biological chemistry 265*, 209-213.
[26] Tochtrop, G. P., Bruns, J. L., Tang, C., Covey, D. F., and Cistola, D. P. (2003) Steroid ring hydroxylation patterns govern cooperativity in human bile acid binding protein, *Biochemistry 42*, 11561-11567.
[27] Chen, C., and Bundschuh, R. (2014) Quantitative models for accelerated protein dissociation from nucleosomal DNA, *Nucleic acids research 42*, 9753-9760.
[28] Anders, A., Lilie, H., Franke, K., Kapp, L., Stelling, J., Gilles, E. D., and Breunig, K. D. (2006) The galactose switch in Kluyveromyces lactis depends on nuclear competition between Gal4 and Gal1 for Gal80 binding, *J Biol Chem 281*, 29337-29348.
[29] Lavy, T., Yanagida, H., and Tawfik, D. S. (2016) Gal3 Binds Gal80 Tighter than Gal1 Indicating Adaptive Protein Changes Following Duplication, *Mol Biol Evol 33*, 472-477.
[30] Timson, D. J., Ross, H. C., and Reece, R. J. (2002) Gal3p and Gal1p interact with the transcriptional repressor Gal80p to form a complex of 1:1 stoichiometry, *Biochem J 363*, 515-520.
[31] Egriboz, O., Goswami, S., Tao, X., Dotts, K., Schaeffer, C., Pilauri, V., and Hopper, J. E. (2013) Self-association of the Gal4 inhibitor protein Gal80 is impaired by Gal3: evidence for a new mechanism in the GAL gene switch, *Molecular and cellular biology 33*, 3667-3674.
[32] Marles, J. A., Dahesh, S., Haynes, J., Andrews, B. J., and Davidson, A. R. (2004) Protein-protein interaction affinity plays a crucial role in controlling the Sho1p-mediated signal transduction pathway in yeast, *Molecular cell 14*, 813-823.
[33] Marsh, J. A., and Teichmann, S. A. (2015) Structure, dynamics, assembly, and evolution of protein complexes, *Annu Rev Biochem 84*, 551-575.

[34] Peng, W., Liu, P., Xue, Y., and Acar, M. (2015) Evolution of gene network activity by tuning the strength of negative-feedback regulation, *Nat Commun 6*, 6226.

[35] Zamboni, N., Saghatelian, A., and Patti, G. J. (2015) Defining the metabolome: size, flux, and regulation, *Molecular cell 58*, 699-706.

[36] Bantscheff, M., Eberhard, D., Abraham, Y., Bastuck, S., Boesche, M., Hobson, S., Mathieson, T., Perrin, J., Raida, M., Rau, C., Reader, V., Sweetman, G., Bauer, A., Bouwmeester, T., Hopf, C., Kruse, U., Neubauer, G., Ramsden, N., Rick, J., Kuster, B., and Drewes, G. (2007) Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors, *Nat Biotechnol 25*, 1035-1044.

[37] Maeder, C. I., Hink, M. A., Kinkhabwala, A., Mayr, R., Bastiaens, P. I., and Knop, M. (2007) Spatial regulation of Fus3 MAP kinase activity through a reaction-diffusion mechanism in yeast pheromone signalling, *Nature cell biology 9*, 1319-1326.

[38] Liao, J. Y., Song, Y., and Liu, Y. (2015) A new trend to determine biochemical parameters by quantitative FRET assays, *Acta Pharmacol Sin 36*, 1408-1415.

[39] Huh, W. K., Falvo, J. V., Gerke, L. C., Carroll, A. S., Howson, R. W., Weissman, J. S., and O'Shea, E. K. (2003) Global analysis of protein localization in budding yeast, *Nature 425*, 686-691.

[40] Bourmaud, A., Gallien, S., and Domon, B. (2016) Parallel reaction monitoring using quadrupole-orbitrap mass spectrometer: Principle and applications, *Proteomics*.

[41] Glatter, T., Ludwig, C., Ahrne, E., Aebersold, R., Heck, A. J., and Schmidt, A. (2012) Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion, *J Proteome Res 11*, 5145-5156.

[42] Schmidt, A., Gehlenborg, N., Bodenmiller, B., Mueller, L. N., Campbell, D., Mueller, M., Aebersold, R., and Domon, B. (2008) An integrated, directed mass spectrometric approach for in-depth characterization of complex peptide mixtures, *Mol Cell Proteomics 7*, 2138-2150.

[43] Deutsch, E. W., Lam, H., and Aebersold, R. (2008) PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows, *EMBO Rep 9*, 429-434.

[44] MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., Kern, R., Tabb, D. L., Liebler, D. C., and MacCoss, M. J. (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments, *Bioinformatics 26*, 966-968.

[45] Carr, S. A., Abbatiello, S. E., Ackermann, B. L., Borchers, C., Domon, B., Deutsch, E. W., Grant, R. P., Hoofnagle, A. N., Huttenhain, R., Koomen, J. M., Liebler, D. C., Liu, T., MacLean, B., Mani, D. R., Mansfield, E., Neubert, H., Paulovich, A. G., Reiter, L., Vitek, O., Aebersold, R., Anderson, L., Bethem, R., Blonder, J., Boja, E., Botelho, J., Boyne, M., Bradshaw, R. A., Burlingame, A. L., Chan, D., Keshishian, H., Kuhn, E., Kinsinger, C., Lee, J. S., Lee, S. W., Moritz, R., Oses-Prieto, J., Rifai, N., Ritchie, J., Rodriguez, H., Srinivas, P. R., Townsend, R. R., Van Eyk, J., Whiteley, G., Wiita, A., and Weintraub, S. (2014) Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach, *Mol Cell Proteomics 13*, 907-917.

[46] Belli, G., Gari, E., Piedrafita, L., Aldea, M., and Herrero, E. (1998) An activator/repressor dual system allows tight tetracycline-regulated gene expression in budding yeast, *Nucleic acids research 26*, 942-947.

[47] Hsu, C., Jaquet, V., Gencoglu, M., and Becskei, A. (2016) Protein Dimerization Generates Bistability in Positive Feedback Loops, *Cell Rep 16*, 1204-1210.

[48] Canelas, A. B., ten Pierick, A., Ras, C., Seifar, R. M., van Dam, J. C., van Gulik, W. M., and Heijnen, J. J. (2009) Quantitative evaluation of intracellular metabolite extraction techniques for yeast metabolomics, *Anal Chem 81*, 7379-7389.

## Figure legends

**Figure 1.** Design of the subsystems and measurement of target response to varying protein concentrations. (a) The galactose signaling network (left). Positive and negative feedback loops are indicated by full and dashed gray lines, respectively. The green squares denote the Gal4p binding sites within the promoters. The subsystems (red dashed squares) to study a specific interaction were obtained by deleting the regulators (double red lines) of the protein whose concentration is varied under the control of the Tet-Off system (magenta rotary switch) (see also Table S3). (b) Linearity of the detection of protein concentration by SID-SRM-MS. Cells with maximal expression of the specified gene (0 dox) were mixed with cells in which the specified gene is deleted to obtain a dilution series. Regression lines are shown in the form $y = ax$. (c) The decay of Gal4p (Gal4p-DNA subsystem) is shown after addition of 1 μg / ml doxycycline. The corresponding changes in *GAL7* expression are shown in the right panel. The black squares denote the measured data, while the empty circles denote the expected Gal4p concentration at specific time points based on the exponential fit. The *GAL7* data align more tightly when the fitted promoter response (Hill) function is plotted with respect to the expected Gal4p concentration, indicated by the narrower confidence interval (68%).

**Figure 2.** Determination of the binding constants in the Gal4p-DNA subsystem and detection of signal amplification by the Gal4p homodimerization. (a) Expression of the *GAL7* and *GCY1* genes as a function of the Gal4p concentration in glucose. $n = 3$ (biological replicates). Error bars denote standard deviations. The data were replotted from time series data as in Figure 1c. (b) *GAL7* expression, as the response to varying *GAL4* production (transcription) rate ($\beta_{P4}$), with the fitted parameters (full line). The dashed line represents a hypothetical system in which the Gal4p is a monomeric protein. The binding of the hypothetical monomeric transcription factor (TF) to the DNA was defined to have a $K_d$ = 354 nM, at which the fractional saturation of the promoter equals the fractional saturation of the promoter bound by the Gal4p in the wild-type cell in galactose (see intersection of dashed and full lines). The gray lines represent linear signal transmission: the relative change in the system output is equal to the relative change in the input. (c) Processes in the first subsystem, including the regulation by glucose. (d) The response of target genes in Δ*gal80* cells as a function of mean cellular concentration of Gal4p in cells grown in the presence of glucose and galactose. The direct effect of glucose on the expression of the

*GAL7* gene is indicated by gray lines. The direct effect was quantified by calculating the ratio of the maximal transcription rate ($V_{max}$) measured in galactose to that in glucose (Table S1) since the direct glucose effect can be discerned when the promoter is saturated by Gal4p. The additional change due to binding of the Gal4p dimer and a hypothetical monomeric TF (as defined in (b)) are indicated by full and dashed lines when the production rate of the Gal4 ($\beta_{P4}$) is decreased from 175 to 53 (as in (b)). These values of $\beta_{P4}$ were calculated from the mean Gal4p concentration in the cells grown in galactose (24.2 nM) and glucose (7.32 nM).

**Figure 3.** Determination of the binding constants in the Gal80p-Gal4p subsystem. (a) Diagram of the binding reactions and constants in the Gal80p – Gal4p subsystem. The dashed circle denotes the nucleus. The circles and triangles denote the Gal4p and Gal80p, respectively. (b) For the Gal80p – Gal4p subsystem, the binding was examined in cells grown in the presence of 0.5% galactose (top) and 0.2% glucose (bottom). The data were replotted from time series data as in Figure 1c. *n = 3* (biological replicates). Error bars denote standard deviations. The gray lines represent the equivalence between the independent and dependent variables for the protein concentration

**Figure 4.** Determination of the binding constants in the Gal1p-Gal80p and Gal3p-Gal80p subsystems and assessment of the system response to galactose. (a, b) Cells were grown in the presence of 0.5% galactose. (b) To determine the response to galactose in the Gal1p-Gal80p subsystem the protein concentration was adjusted by varying the concentration of doxycycline. Cells were incubated for 24 hours in the presence of 0.02% galactose to reach steady-state. The culture was refreshed twice to keep galactose concentrations steady and to prevent the culture from exiting the logarithmic growth phase. The *GAL7* expression was measured (red diamonds). (c) Intracellular accumulation of galactose in Δgal1 Δgal2 cells after addition of galactose to the medium at the indicated concentrations. The first measurement t=0 was performed immediately after addition of galactose. The low level of intracellular galactose at this time point indicates that the galactose in the medium is efficiently washed out during the extraction of galactose from the cells. The curves denote fits with the Widda's formula for reversible transport. The steady-state intracellular concentration closely matches the extracellular one. (d) The galactose concentrations have been determined using the Raffinose/D-Galactose enzymatic assay. The linearity of the assay was assessed by supplementing media with galactose at the indicated galactose

concentrations. Linear response was observed in the concentration range from 0.001 to 0.1% galactose. Thus, media with concentrations above 0.1% where diluted to obtain a linear response.

**Figure 5.** Validation of system parameters by prediction of network response in the Gal1p mediated feedback loop. (a) Bistability predicted for the *GAL1* feedback loop (in Δ*gal2*, Δ*gal3* background) based on the original parameters fitted in the three subsystems and the *GAL1* transcription and translation rates. The black and red lines denote the lower and upper stable *GAL1* mRNA expression levels. The predictions with the lowest predicted Gal80p production rate (predicted value minus standard error of the fitting: $\beta_{P80}$ = 180.1 – 31.2 = 148.9 are also shown (orange and gray dashed lines). (b) Molecular interactions in the *GAL1* loop in the Δ*gal2*, Δ*gal3* cell background. (c) Prediction of bistability with altered parameters reflecting strong Gal4p dimerization. The parameters are identical to original parameter values used in the model (a), with the following exceptions: $K_D^{4/4}$ = 100 nM and $K_D^{DNA}$ = 1039 nM. (d) Comparison of the bistability range and cellular memory. The memory index was measured 72 and 96 h after pre-exposing the cells to 0 and 2% galactose. Error bars denote standard deviations (*n = 3*). The full cyan line represents the predicted bistability ranges with the wild type parameters, while the cyan dashed line indicates the predicted bistability with $\beta_{P80}$ = 148.9. The magenta lines denote the bistability range calculated with the parameters reflecting the strong Gal4p dimerization (c).

**Table 1**. Binding constants. The values of the apparent equilibrium dissociation constants are expressed in nM, μM and mM. The numbers in the parenthesis in the second column denote the subsystem, in which the parameter was fitted (1: Gal4p – DNA, 2: Gal80p – Gal4p, 3a: Gal1p – Gal80p, 3b: Gal3p – Gal80p).

| Binding process | Fitted in sub-system | Parameter | Cells grown in the presence of | |
|---|---|---|---|---|
| | | | Galactose | Glucose |
| Gal4p – Gal4p | (1) | $K_D^{4/4}$ | 8.46 μM | 7.49 μM |
| Gal4p – DNA | (1) | $K_D^{DNA}$ | 14.3 nM | 15.0 nM |
| Gal4p – DNA (single binding site in promoter) | (1) | $K_D^{DNA1}$ | 4.8 nM | 5.0 nM |
| Cooperative Binding Gal4p –DNA (GAL7) | (1) | c | 24.4 | 23.9 |
| Gal80p – Gal80p | (2) | $K_D^{80/80}$ | 0.645 nM | 0.681 nM |
| Gal80p – Gal4p | (2) | $K_D^{4/80}$ | 0.039 nM | 0.041 nM |
| Gal1p (complexed with galactose) – Gal80p | (3a) | $K_D^{1G/80}$ | 10.5 nM | - |
| Gal1p – Galactose | (3a) | $K_D^{1/G}$ | 4.44 mM | - |
| Gal3p (complexed with galactose) – Gal80p | (3b) | $K_D^{3G/80}$ | 1.2 nM | - |

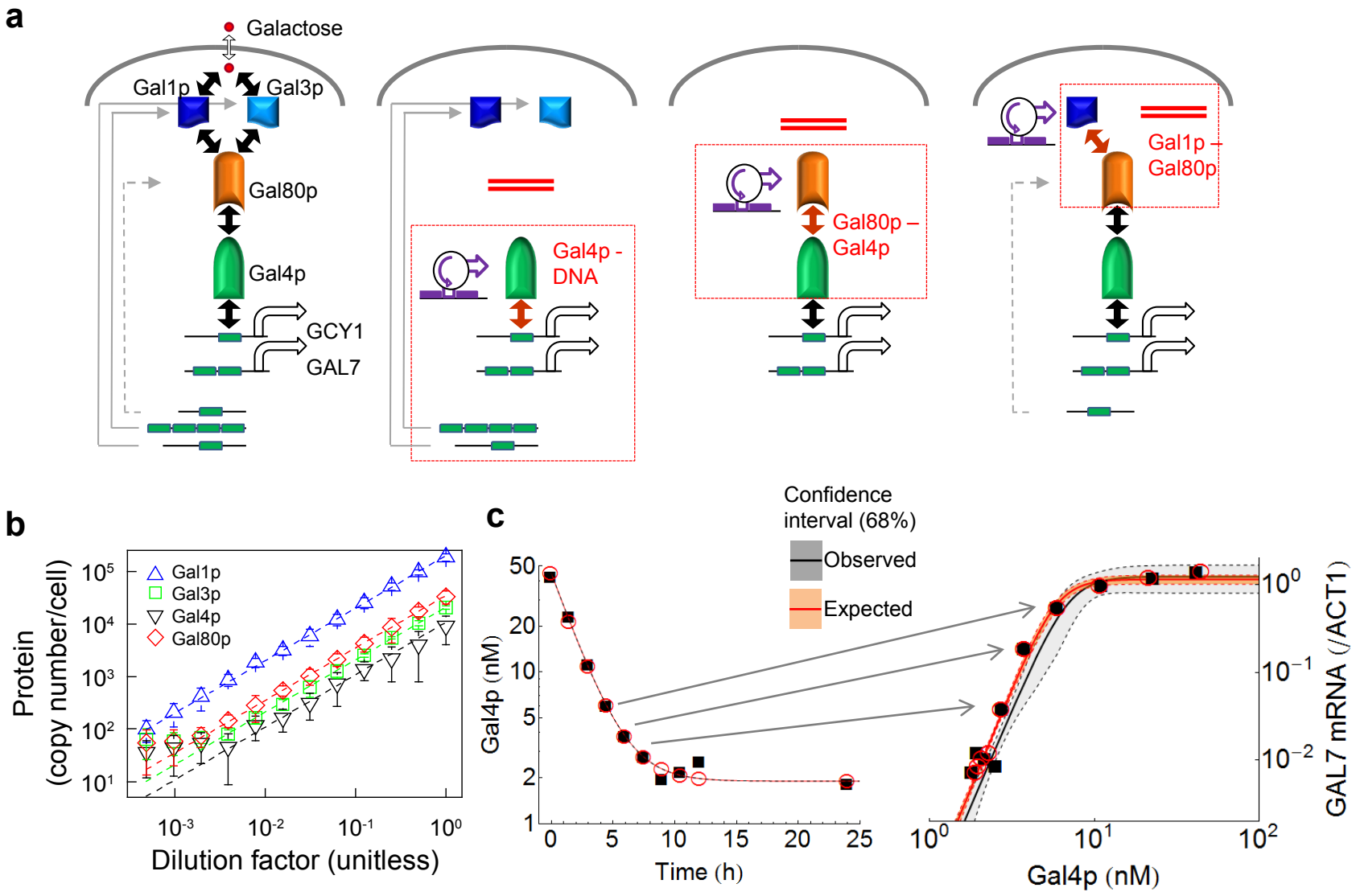

a
Galactose
Gal1p
Gal3p
Gal80p
Gal4p
GCY1
GAL7
Gal4p - DNA
Gal80p – Gal4p
Gal1p – Gal80p
b
Protein (copy number/cell)
Gal1p
Gal3p
Gal4p
Gal80p
Dilution factor (unitless)
c
Confidence interval (68%)
Observed
Expected
Gal4p (nM)
Time (h)
GAL7 mRNA (/ACT1)



Figure 1.

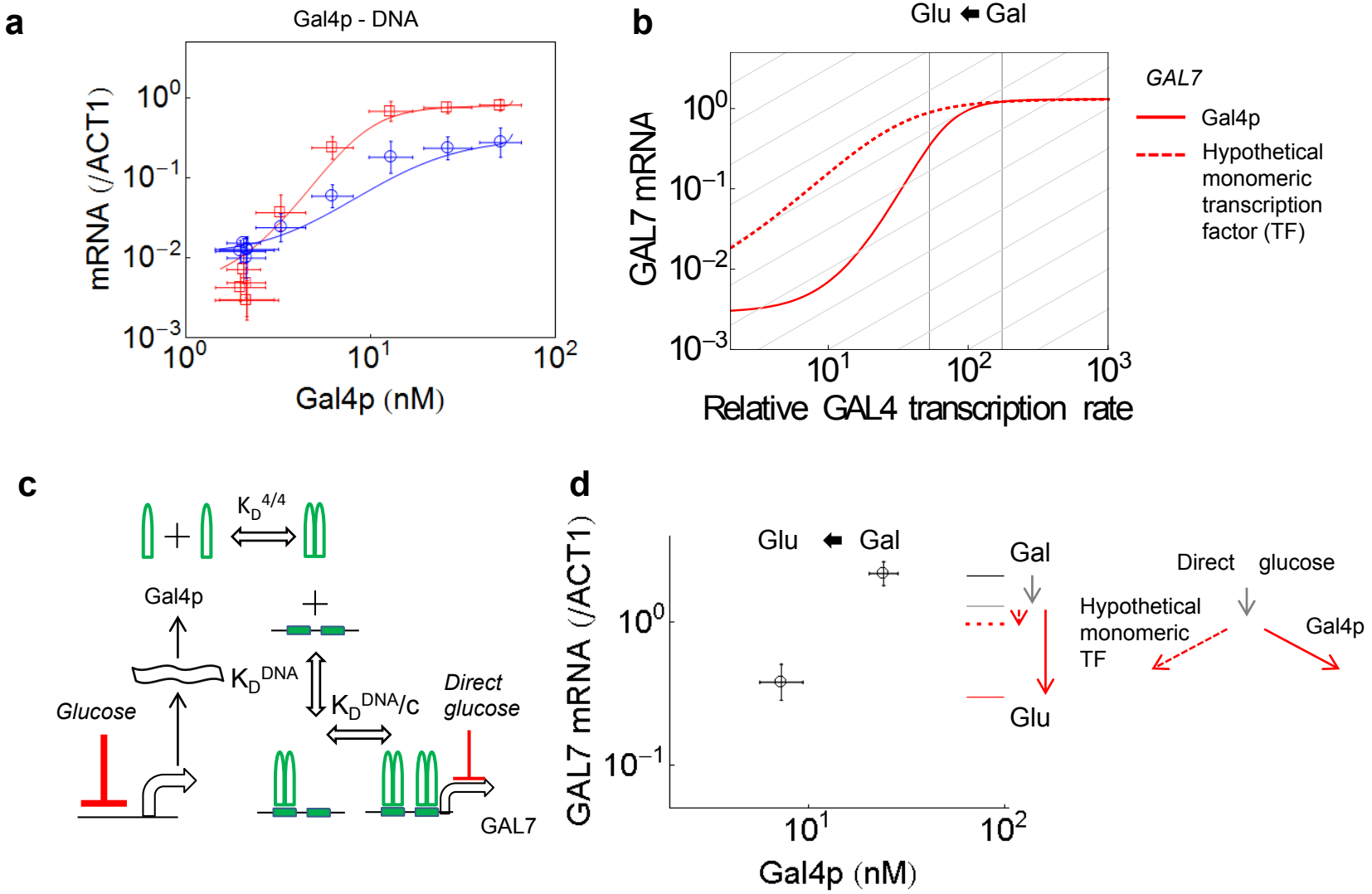

a
Gal4p - DNA
mRNA (/ACT1)
Gal4p (nM)
b
Glu ⬅ Gal
GAL7 mRNA
Relative GAL4 transcription rate
GAL7
Gal4p
Hypothetical monomeric transcription factor (TF)
c
$K_D^{4/4}$
Gal4p
$K_D^{DNA}$
$K_D^{DNA}/c$
Glucose
Direct glucose
GAL7
d
Glu ⬅ Gal
GAL7 mRNA (/ACT1)
Gal4p (nM)
Gal
Glu
Direct glucose
Hypothetical monomeric TF
Gal4p



Figure 2.

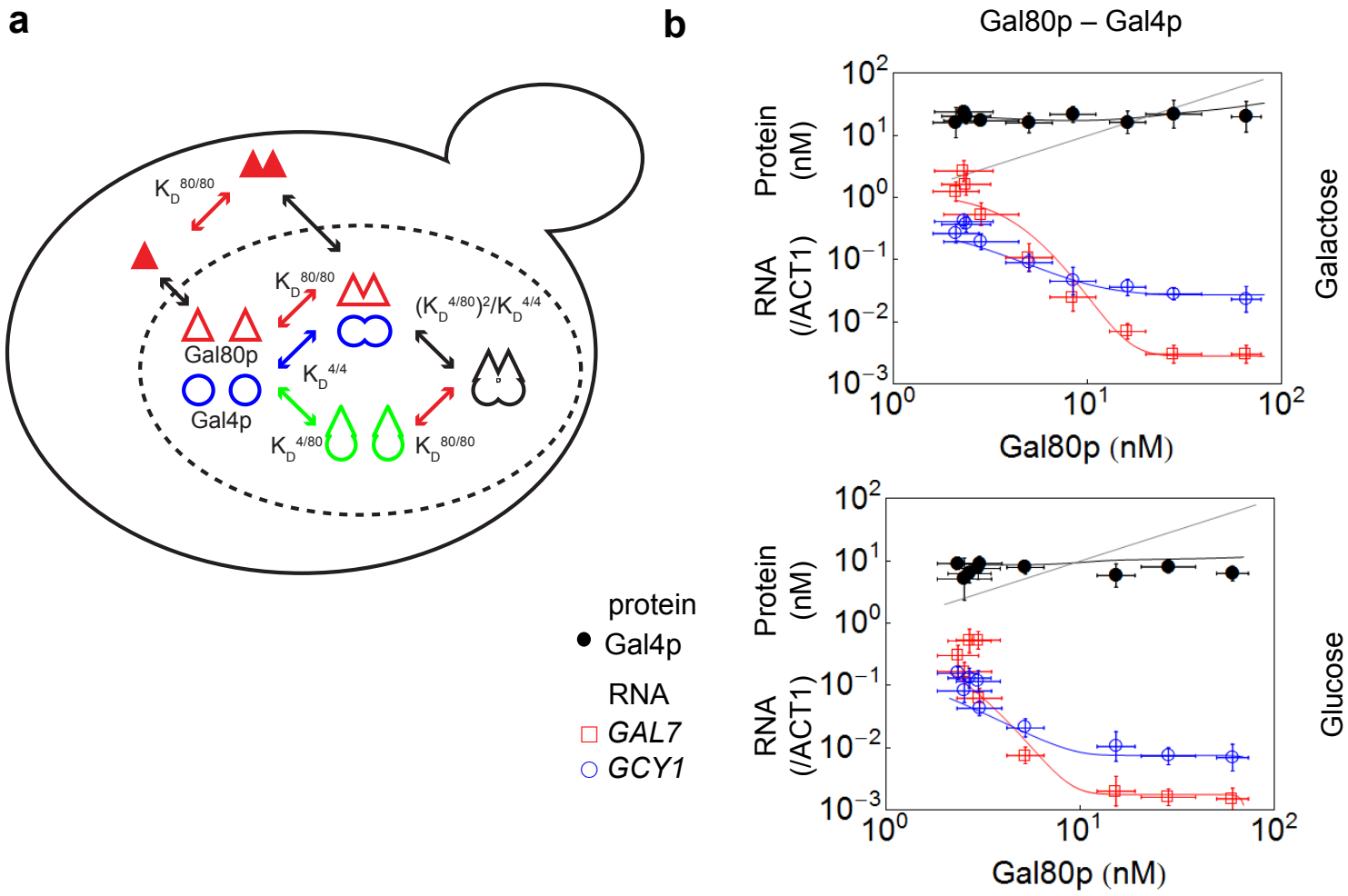

a
$K_D^{80/80}$
$K_D^{80/80}$
$(K_D^{4/80})^2/K_D^{4/4}$
Gal80p
$K_D^{4/4}$
Gal4p
$K_D^{4/80}$
$K_D^{80/80}$
protein
Gal4p
RNA
GAL7
GCY1
b
Gal80p – Gal4p
Protein (nM)
RNA (/ACT1)
Galactose
Gal80p (nM)
Protein (nM)
RNA (/ACT1)
Glucose
Gal80p (nM)



Figure 3.

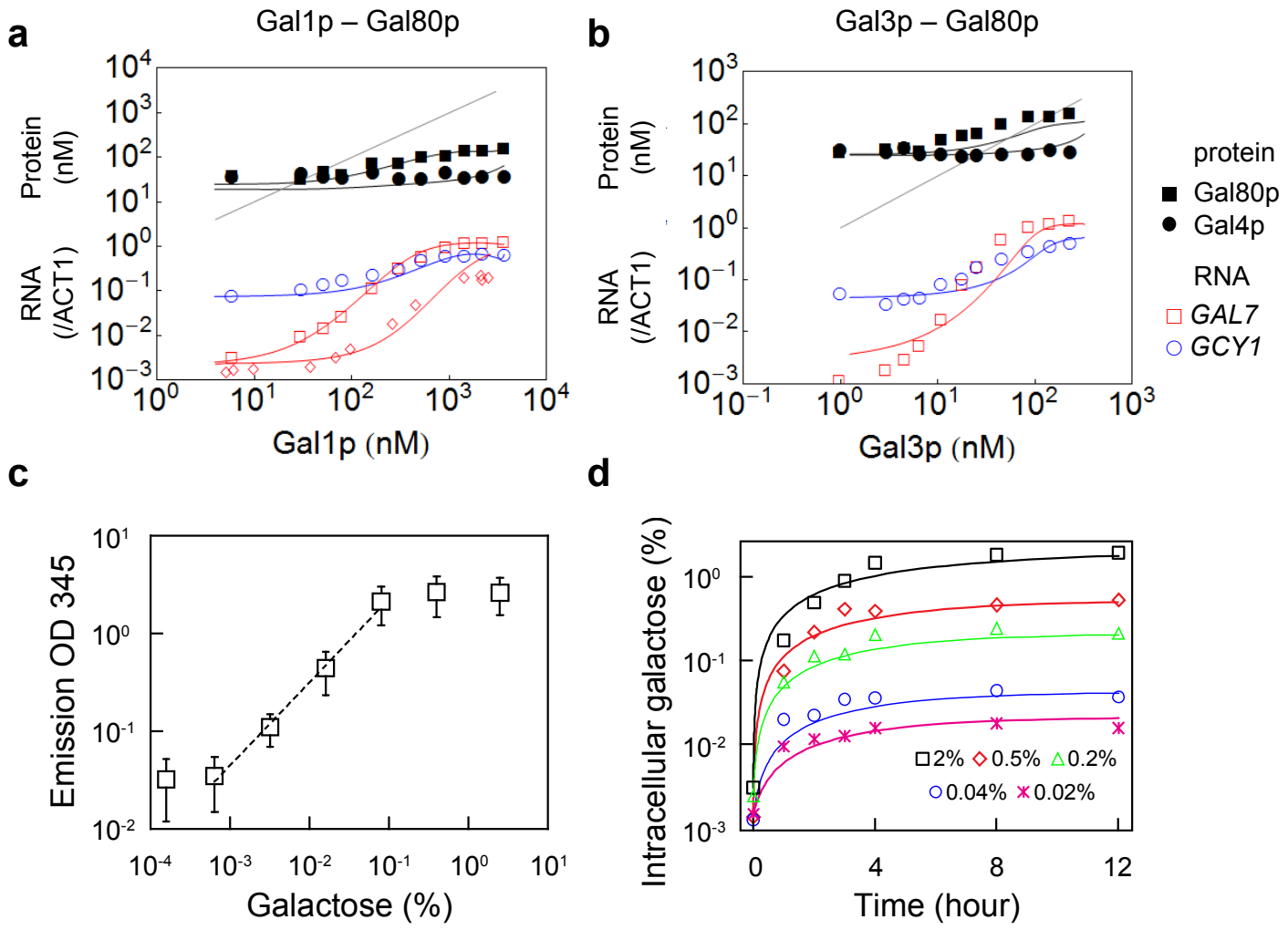

a
Gal1p – Gal80p
Protein (nM)
RNA (/ACT1)
Gal1p (nM)
b
Gal3p – Gal80p
Protein (nM)
RNA (/ACT1)
Gal3p (nM)
protein
Gal80p
Gal4p
RNA
GAL7
GCY1
c
Emission OD 345
Galactose (%)
d
Intracellular galactose (%)
Time (hour)
2%
0.5%
0.2%
0.04%
0.02%

Figure 4.

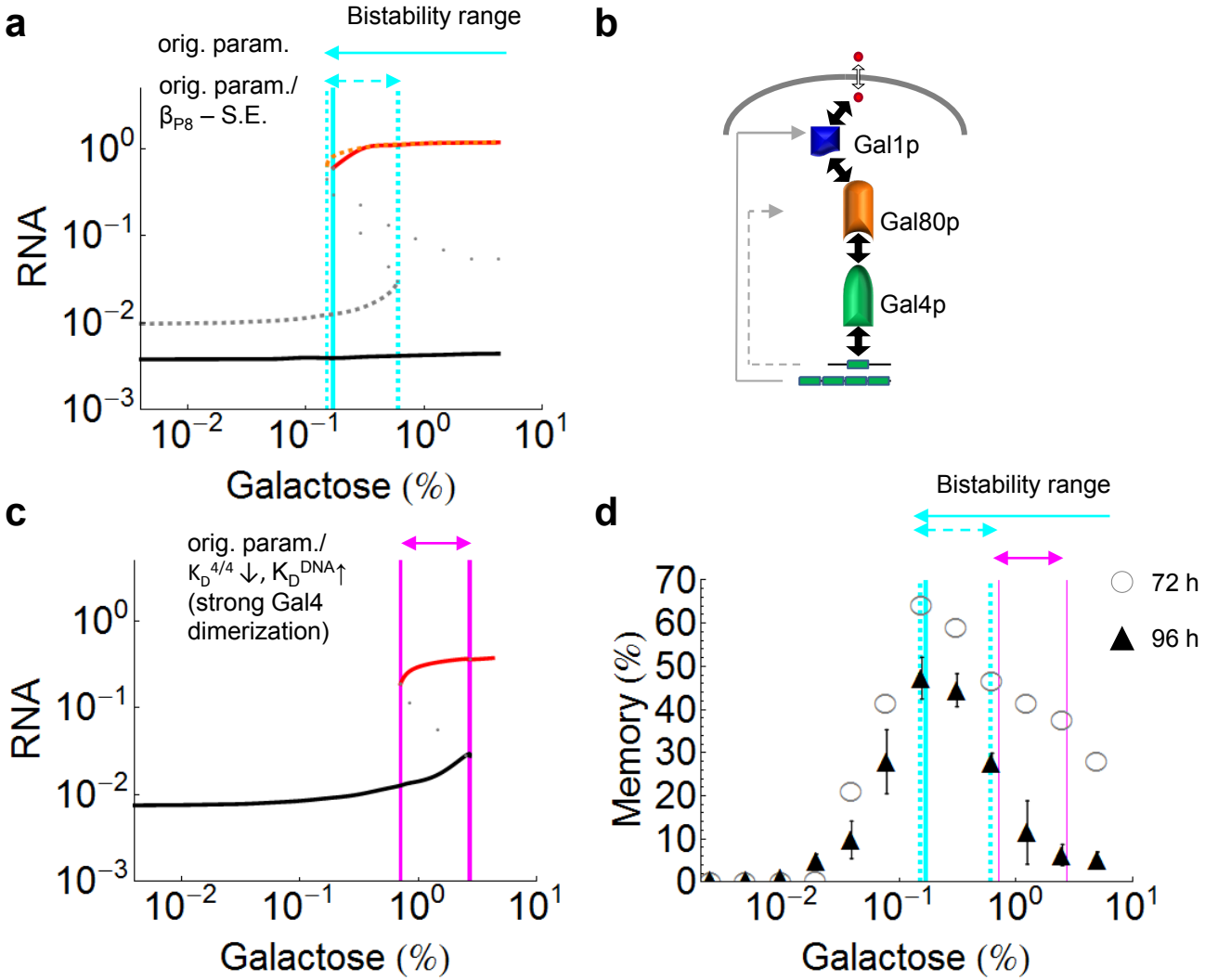

a
Bistability range
orig. param.
orig. param./
$\beta_{P8}$ – S.E.
RNA
$10^0$
$10^{-1}$
$10^{-2}$
$10^{-3}$
$10^{-2}$ $10^{-1}$ $10^0$ $10^1$
Galactose (%)
b
Gal1p
Gal80p
Gal4p
c
orig. param./
$K_D^{4/4}$ ↓, $K_D^{DNA}$↑
(strong Gal4
dimerization)
RNA
$10^0$
$10^{-1}$
$10^{-2}$
$10^{-3}$
$10^{-2}$ $10^{-1}$ $10^0$ $10^1$
Galactose (%)
d
Bistability range
Memory (%)
70
60
50
40
30
20
10
0
72 h
96 h
$10^{-2}$ $10^{-1}$ $10^0$ $10^1$
Galactose (%)



Figure 5.