Science
AAAS

Supplementary Materials for

Systematic Identification of Signal Activated Stochastic Gene Regulation

Gregor Neuert, Brian Munsky, Rui Zhen Tan,
Leonid Teytelman, Mustafa Khammash, Alexander van Oudenaarden

correspondence to: avano@mit.edu

This PDF file includes:

Materials and Methods–Experimental

Strain and plasmid construction

We used the BY4741 haploid yeast strain with the knockout of the Sho1 scaffold ($sho1\Delta$) to eliminate any crosstalk from other MAP kinase pathways into the HOG pathway and to linearize the pathway (*30, 31*). To visualize the dynamics and activity of the terminal MAP Hog1p kinase, its C-terminus was tagged with a yellow fluorescent protein (YFP) yECitrine using standard PCR-integration (*30*). Experiments in our laboratory demonstrated that the signaling dynamics and the activity of the Hog1p-YFP protein match previously reported data from a Hog1p-GFP fusion protein, which was reported to be fully functional (*32*). For the transcriptional readout of the $STL1$ induction, we used a plasmid in which the $STL1$-promoter was fused to a GFP-protein. The plasmid was then linearized and integrated as a single copy in the endogenous $STL1$-promoter locus. First, 1000bp of the $STL1$-promoter was PCR amplified between -1000bp to 0bp upstream of the 5'-end of the $STL1$-ORF. Second, we ligated the PCR product onto the GFP protein in a plasmid with a $URA3$ selection marker. Third, the plasmid was linearized in the middle of the $STL1$-promotor region to ensure integration upstream of the endogenous $STL1$-locus. A single integration of the plasmid was verified by southern blot and mRNA FISH. For the over expression of the transcription factor Hot1p, we designed primers that bind 1000bp upstream of the 5'-end of the $HOT1$-ORF and 300bp downstream of the 3'-end of the $HOT1$-ORF. This PCR product is then ligated into a plasmid backbone with a MET marker. The plasmid was linearized in the promoter region and integrated into the Hot1p promoter locus using homologous recombination. The plasmid integration allows for the integration of multiple copies of the plasmid in the same genomic location. The fold over expression was measured using $HOT1$ mRNA expression with FISH in over expression strains and compared to wild type strains showing a five fold increase in $HOT1$ mRNA in the Hot1p 5x strain. Knockout of the proteins Arp8p and Gcn5p in strains with a Hog1-YFP tag was generated by crossing the above described Hog1-YFP strain with the BY4742, $GCN5 :: KAN$ ($gcn5\Delta$) or the BY4742, $ARP8 :: NAT$ ($arp8\Delta$) strains.

Yeast growth conditions

The cells are grown on solid minimal media for two days lacking the required amino acids. A single colony is transferred from the plate and inoculated in liquid media to grow for 6-10h. The culture is then diluted to reach a final OD_{600} of 0.1 after 12-16h.

Fluorescence in situ hybridization

For the fluorescent in-situ hybridization experiments, cells have been grown at 30°C. Cells were exposed to a final osmolyte concentration of 0.2M or 0.4M NaCl for 0, 1, 2, 4, 6, 8, 10, 15, 20, 25, 30, 35, 40, 45, 50 and 55 minutes. After the required time, cells were fixed in 4% (v/v) formaldehyde for 30min at room temperature (23±2°C) and then overnight at 4C under constant shaking (see (*33*)). The longer crosslinking is preserving the cell boundary, which is essential for automated image segmentation. The most critical step in doing RNA-FISH in yeast is digestion of the cell wall. This is crucial–if the cell wall is not digested, the number of probes that enter the cell is reduced, and only a fraction of the mRNA spots can be detected. However, if the cells are over-digested, the cells breaks apart and can not be imaged. We quantified the digestion efficiency by analyzing the average mRNA spot intensity in the cell as a function of Zymolase concentration between 10E-3 to 10E2 mg/ml 100T for 30min. In these experiments, cells have been grown to OD600 = 0.1, activated with 0.4M NaCl and fixed after 15min. As shown in fig. S1, the average spot intensity changes three fold and does not change significantly above 2mg/ml Zymolase 100T (US biological). For all the experiments in this study, we choose digestion conditions in this plateau region to ensure maximal cell wall digestion and cell integrity. After cell wall digestion, cells were stored in ethanol at 4°C for a minimum of 12 hours. Cells can be stored for several month in these conditions. Before hybridization, cells were rehydrated in 10% (v/v) formamide and 2x SSC for 5 min. Cells were hybridized in 30μl of hybridization solution containing labeled DNA probes in 10% (v/v) formamide, 2x SSC, 1 mg/ml BSA, 10mM Vanadyl-ribonucleoside complex, 0.5 mg/ml, Escherichia coli tRNA and 0.1 g/ml dextran sulfate overnight at room temperature. Before imaging, cells were washed twice in 10% (v/v) formamide and 2x SSC for 30 min. DAPI was added to the first washing step. The cells were then added to a coverslip and pressed onto a microscope slide. In cases where AF594 and Cy5 labeled probes have been used, we added an anti-oxidation buffer as previously described (*34*) including 2mM Trolox.

Probes for in situ hybridization

The $STL1$, $HOT1$, $CTT1$, and $HSP12$ mRNA were labeled each with a unique set of up to 48 DNA oligonucleotides, each with a length of 20nt. The 3'-end of each probe was modified with an amine group. The DNA-probes have been coupled to tetramethylrhodamine, TMR (Invitrogen), Alexa fluor 594 (Invitrogen) or Cy5 (GE Amersham). After coupling the DNA-probes to the fluorophore, the probes were ethanol precipitated and were purified on an HPLC column to isolate oligonucleotides displaying the highest degree of coupling of the fluorophore to the amine groups.

Image acquisition

Images were taken with a Nikon TI-E inverted fluorescence microscope equipped with a 100x oil-immersion objective and a Princeton Instruments camera using MetaMorph software (Molecular Devices, Downington, PA). Custom filter sets designed to distinguish between the different fluorophores were used. At each time point, Differential Interference Contrast (DIC) and DAPI images were taken. Depending on the strains and mRNA that have been measured, YFP, TMR, AF594 Cy5 images were taken as well. On these fixed yeast cells, z-stacks of images with each separated by 200 nm were taken. The DIC and DAPI images were used to identify single cells. The image of the segmented cells, together with the YFP and the DAPI images, were used to measure the Hog1p localization process into the nucleus (Hog1p nuclear enrichment, Figs. S2 and S3), and to estimate Hog1p kinase activity. Finally, the TMR, AF594 and CY5 3D image stack was used to detect single mRNA transcripts.

Cells for time lapse microscopy have been prepared as described previously (*31*). In brief, a flow chamber was used to monitor the GFP gene expression in single live yeast cells using time-lapse microscopy. The flow chamber consists of a cover glass coated with Concanavalin A (Con A). Yeast cells stick to this coating, while media flows rapidly over it.

Image analysis

(1) A bright field image was chosen in which a clear cell boundary can be observed. The image was then converted into a binary image using automated thresholding. (2) The max projection of a DAPI-image stack was generated and converted into a binary image using a fixed pixel intensity threshold. (3) The binary bright field image was added to the binary DAPI image. (4) The DAPI-nuclei serve as markers for running a marker controlled watershed algorithm over the merged bright-field/DAPI image. Connected regions below or above a specific size were rejected. The cell boundaries were obtained using an edge detection algorithm. Next, a program was run to count the number of RNA molecules in each cell. The general procedure was to run a median filter followed by a Laplacian filter on each optical slice taken. This enhanced particulate signals. A threshold was then selected to pick-up individual spots in each plane. Next, the image with the segmented cells was multiplied with each plane of the TMR, AF594 or Cy5 image stack. Now, in each cell a program counted the total number of isolated signals (i.e. connected components) in three dimensions. The particle count was robust for a range of threshold chosen. Intensity analysis of mRNA spots was performed by integrating the total pixel intensity of a 7x7 pixel square centered on the maximum subtracted by the background.

Results from image analysis

After processing microscopy images, we obtained the number of mRNA per cell from a total of more then 100,000 fixed cells. These data sets include experiments at 0.2M, and 0.4M NaCl in strains expressing wild type levels of the transcription factor Hot1p (Hot1p 1x WT). The experiments in the five fold over expression of Hot1p and in the Arp8p and Gcn5p deletion strains are performed at 0.4M NaCl. Each experimental condition was repeated two times and images where taken at all time points for all experiments. The genes

$STL1$, $CTT1$ and $HSP12$ have been measured together in the same cells. For each experimental condition, the replica experiments are shown in different colors (see figs. S11 for an example). These measurements provide a wealth of quantitative information regarding how mRNA distributions change as functions of time, osmolyte concentration, Hot1p, Arp8p and Gcn5p expression levels.

Hog1p nuclear localization in WT and mutant strains

We generated strains in which the Hog1p nuclear localization dynamics is measured in single cells in WT, Hot1p over-expression and the deletion strains of Gcn5p and Arp8p. The rationale is to distinguish between potential changes in Hog1p signaling dynamics and changes in gene expression due to these mutations. We observed that over expression Hot1p or deleting the Arp8p or Gcn5p chromatin modifiers has a significant affect on the Hog1p nuclear localization and changes the signaling intensity, duration and shape of the Hog1p profile (fig. S4). These Hog1p nuclear localization profiles in the wild-type, Hot1p over-expression and Arp8p/Gcn5p deletion backgrounds are integrated into the modeling as discussed below.

Methods–Computational

Parameterizing the Hog1p signal

Previously measured Hog1p nuclear enrichment curves feature a sharp rise, a saturated level, and then a sharp decline (*30*), and the rate of the decline appears to depend upon the salt level. In order to parameterize this curve in terms of the time and level of osmotic shock, we assume that at $t = 0$ osmotic pressure is applied to the system, and the system evolves according to the diffusion relations:

$$\begin{array}{ll} \frac{d}{dt} Hog1p_{\text{out}}(t) = -\gamma_1 Hog1p_{\text{out}}(t), & Hog1p_{\text{out}}(0) = C_1, \\ \frac{d}{dt} Hog1p_{\text{in}}(t) = C_2 Hog1p_{\text{out}}(t) - \gamma_2 Hog1p_{\text{in}}(t), & Hog1p_{\text{in}}(0) = 0. \end{array} \tag{1}$$

The solution for this simple set of ODES is:

$$Hog1p_{\text{in}}(t) = C_1 C_2 \int_0^t \mathrm{e}^{-\gamma_1 \tau} \mathrm{e}^{-\gamma_2 (t-\tau)} d\tau, \tag{2}$$

$$= C_1 C_2 \mathrm{e}^{-\gamma_2 t} \int_0^t \mathrm{e}^{-(\gamma_1 - \gamma_2)\tau} d\tau, \tag{3}$$

$$= \frac{C_1 C_2}{-(\gamma_1 - \gamma_2)} \mathrm{e}^{-\gamma_2 t} \left(1 - \mathrm{e}^{-(\gamma_1 - \gamma_2)\tau}\right) \tag{4}$$

$$= \frac{C_1 C_2}{-r_1} \mathrm{e}^{-r_2 t} \left(1 - \mathrm{e}^{-r_1 \tau}\right) \tag{5}$$

$$= \hat{C} \left(1 - \mathrm{e}^{-r_1 t}\right) \mathrm{e}^{-r_2 t}, \tag{6}$$

where the rate, $r_1 = \gamma_1 - \gamma_2$, corresponds to the rising behavior of the Hog1p signal (same for all salt levels, which allows us to lump the terms, $\hat{C} = C_1 C_2 / r_1$), and $r_2 = \gamma_2$ corresponds to the decreasing behavior. r_2 is the *only* parameter that varies for different salt levels and it is very well fit with the function

$$r_2 = \frac{\alpha}{\text{[NaCl]}}. \tag{7}$$

In order to capture the saturation level for Hog1p(t) enrichment, we assume a saturation function of the form:

$$Hog1p^{\star}(t) = \left(\frac{Hog1p(t)}{1 + Hog1p(t)/M}\right)^{\eta} = A \left(\frac{\left(1 - \mathrm{e}^{-r_1 t}\right) \mathrm{e}^{-r_2 t}}{1 + \frac{(1 - \mathrm{e}^{-r_1 t}) \mathrm{e}^{-r_2 t}}{M}}\right)^{\eta}, \tag{8}$$

where A and M define the saturation height and midpoint and are the same for all salt levels.

The parameters $\{r_1, \alpha, \eta, A, M\}$ have been fit to match the measured Hog1p nuclear enrichment levels as functions of time, osmolite concentrations, and genetic mutations. The parameters are given in Table S1 and the corresponding fits are shown in fig. S4.

Gene regulation models

Now that we have parameterized the Hog1p activation curve, we can introduce that curve as a time varying parameter input into the regulation of downstream genes. We consider model structures with $N = \{2, 3, 4, 5\}$ distinct gene states arranged in a linear chain. For N-states, there are $(N-1)$ forward reactions, $(N-1)$ backward reactions giving a total of $2N-2$ state transition reactions. Each state can also result in mRNA production. Finally mRNA can degrade as a first order process, giving a total of $3N-2$ reactions for an N-state model.

The rate of each state transition $S_i \rightarrow S_j$ may be constant or dependent on the level of Hog1p. These Hog1p dependencies are described with a simple linear form,

$$k_{ij} = \max\{0, a_{ij} + b_{ij}\text{Hog1p}(t)\} \tag{9}$$

The Hog1p effect is chosen to reflect the fact that kinases can have positive ($b_{ij} > 0$) or negative ($b_{ij} < 0$) effects on more complicated processes, including chromatin remodeling, transcription factor binding, polymerase recruitment or elongation, transcription initiation, mRNA degradation, or others. Non-negativity of the transition rate k_{ij} is enforced in the case where $b_{ij} < 0$. Measurements of the mean level of $STL1$ mRNA in the ON cells (*i.e.,* cells with more than a few mRNA) shows that the average amount of mRNA is the same for all experimental conditions (wild type at 0.2M and 0.4M NaCl, as well as the Hot1p 5x, $arp8\Delta$, and $gcn5\Delta$ mutants at 0.4M NaCl (see fig. S5), despite the fact that the Hog1p kinase signal is very different in these strains. This suggests that the transcription rates in active states and the $STL1$ mRNA degradation rates are not affected by Hog1p. To account for the short time between transcript initiation and the actual observation of fully formed mRNA, we add an additional time-lag parameter, t_0, which acts as a delay between the Hog1p signal and downstream events.

Stochastic models of gene regulation

We describe the responses according to the probabilities,

$$P_{i,m} := P(\text{state} = S_i, \text{mRNA} = m). \tag{10}$$

For an N-state model, the index i can take N different values, $i \in [1, 2, \ldots, N]$. For convenience, we enumerate all the possible states into the vector:

$$\begin{bmatrix} \mathbf{P}_0 \\ \\ \mathbf{P}_1 \\ \\ \vdots \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} \mathbf{P}_{1,0} \\ \mathbf{P}_{2,0} \\ \vdots \\ \mathbf{P}_{N,0} \end{bmatrix} \\ \begin{bmatrix} \mathbf{P}_{1,1} \\ \mathbf{P}_{2,1} \\ \vdots \\ \mathbf{P}_{N,1} \end{bmatrix} \\ \vdots \end{bmatrix}, \tag{11}$$

where the vector $\mathbf{P}_m$ corresponds to the probability of the N different gene states with exactly m molecules of mRNA.

Following the methodology in (*35–37*), we formulate the ordinary differential equation that describes the evolution of these probabilities over time. This equation, known as the *Chemical Master Equation* or *Forward Kolmogorov Equation* can be written in vector form as: $\frac{d}{dt}\mathbf{P}(t) = \mathbf{Q}(t)\mathbf{P}(t)$, or more specifically in this instance:

$$\frac{d}{dt}\begin{bmatrix} \mathbf{P}_0 \\ \mathbf{P}_1 \\ \mathbf{P}_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} \mathbf{A}-\mathbf{T} & \mathbf{D} & \mathbf{0} & \cdots \\ \mathbf{T} & \mathbf{A}-\mathbf{T}-\mathbf{D} & 2\mathbf{D} & \ddots \\ \mathbf{0} & \mathbf{T} & \mathbf{A}-\mathbf{T}-2\mathbf{D} & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{bmatrix} \begin{bmatrix} \mathbf{P}_0 \\ \mathbf{P}_1 \\ \mathbf{P}_2 \\ \vdots \end{bmatrix}. \tag{12}$$

In this expression, the matrices for state transitions ($\mathbf{A}$), transcription ($\mathbf{T}$), and degradation ($\mathbf{D}$) are given for the three-state system as:

$$\mathbf{A}(t) = \begin{bmatrix} -k_{12} & k_{21} & 0 \\ k_{12} & -k_{21}-k_{23} & k_{32} \\ 0 & k_{23} & -k_{32} \end{bmatrix}; \quad \mathbf{T}(t) = \begin{bmatrix} k_{r1} & 0 & 0 \\ 0 & k_{r2} & 0 \\ 0 & 0 & k_{r3} \end{bmatrix}; \quad \mathbf{D}(t) = \begin{bmatrix} \delta & 0 & 0 \\ 0 & \delta & 0 \\ 0 & 0 & \delta \end{bmatrix}, \tag{13}$$

where all rates $\{k_{12}, k_{21}, k_{23}, k_{32}, k_{r1}, k_{r2}, k_{r3}, \delta\}$ are potentially functions of Hog1p(t) as given in Eq. S9.

In principle, the number of mRNA can reach any arbitrarily large integer number, $m \in [0, 1, 2, \ldots]$, meaning that the dimension of $\mathbf{P}$ may become arbitrarily large. To address this issue, we use the Finite State Projection Approach to approximate this vector within known error bounds (For more details see (*35–37*)). Typically, this approximation considers larger and larger portions of the vector $\mathbf{P}$ until a specified error tolerance is met. In this case, however, we have experimental data that directly tells us what levels of mRNA numbers to expect. Only one cell in over 60,000 was observed to exceed 100 mRNA, and we have set conservative truncation value of 150 molecules per cell.

After applying the FSP approach, Eq. S12 becomes finite dimensional, and it can be solved in Matlab using the stiff ODE solver, ode15s. For a given set of parameters and a final time of 80 minutes, this solution takes less than one second to complete. This speed is necessary to enable a parameter search to find the maximally likely set of parameters within a large parameter space, allowing a single processor to test over 140,000 model/parameter/experiment combinations per day. Parallel fits were conducted on clusters of 24-64 processors allowing for the consideration of nearly ten million model/parameter/experiment combinations per day. Furthermore, the chosen projection size, guarantees that the approximation errors satisfy a strict criterion of $\sum |\mathbf{P}_{\text{exact}} - \mathbf{P}_{\text{FSP}}| < 10^{-6}$ at all times and all conditions for the final parameter set.

For comparison of the model to the experimental mRNA distributions, it is necessary to convert the $\mathbf{P}$ into a marginal distribution of mRNA numbers. This is relatively easy to do, since

$$\rho_m = P(\text{mRNA} = m) = \sum_{i=1}^{N} P_{i,m}. \tag{14}$$

Now with this definition, we can compare model predictions and experimental data as described below.

Computation of maximum likelihoods

In order to compare a given model and parameter set to the experimental data, it is necessary to determine *'How likely it is that the data could have came from the given model'*. To define this likelihood function, suppose that $n = \{1, 2, \ldots, N\}$ cells were measured various combinations of different experiments (e_n) and different time points (t_n) and each of these cells was found to have exactly m_n copies of a specific type of mRNA. Suppose that a model with parameter set θ, predicts that for each time and experiment, the probability that a given cell has exactly m_n mRNAs in the corresponding conditions is $p(m|\theta, e_n, t_n)$. The total likelihood of all observations, $L(\mathbf{D}|\theta)$, is the product over every cell, or

$$L(\mathbf{D}|\theta) = \prod_{n=1}^{N} p(m_n|\theta, e_n, t_n). \tag{15}$$

Now that we know how likely it is that the data comes from a model and a given set of parameters, θ, our goal is to find the parameter set, θ_{Fit}, which maximizes this likelihood (or equivalently the logarithm of this likelihood):

$$\begin{aligned} \theta_{\text{Fit}} &= \arg\max_{\theta} \left(\log(L(\mathbf{D}|\theta))\right) & (16) \\ &= \arg\max_{\theta} \sum_{n=1}^{N} \log p(m_n|\theta, e_n, t_n). & (17) \end{aligned}$$

Parameter searches

In order to conduct parameter searches, we use iterative combinations of simplex based searches (e.g., Matlab's "fminsearch" function), simulated annealing, and genetic algorithm based searches. The searches are run multiple times from different starting parameter guesses leading to hundreds of thousands of function evaluations per model, for which only the best solution is recorded at the end. We have conducted this search many times for each of the different models under consideration. Parallel fits were conducted on clusters of 24-64 processors allowing for the consideration of nearly ten million model/parameter/experiment combinations per day.

Cross-validation analysis

Because we are considering different model structures of varying complexity, it is necessary to quantify the level of uncertainty for a given model or set of models. This quantification is accomplished through the use of a simple cross-validation analysis. As above, the best fit given all of the data under consideration, $\mathbf{D}$, corresponds to the parameter set θ_{Fit} and likelihood $\log(L_{\text{Fit}})$:

$$\log(L_{\text{Fit}}) = \max_{\theta}\left[\log(L(\mathbf{D}|\theta))\right] \tag{18}$$

$$\theta_{\text{Fit}} = \arg\max_{\theta}\left[\log(L(\mathbf{D}|\theta))\right] \tag{19}$$

In order to cross validate our models, we sample from the full data set to create smaller data sets $\{\mathbf{d}_1, \ldots, \mathbf{d}_n\}$ where each $\mathbf{d}_i \subset \mathbf{D}$. More specifically, we take data from one replica of each experimental condition to form the set $\mathbf{d}_i$. For example, the two replicas of the wild-type experiment at 0.4M NaCl give rise to two data sets $\{\mathbf{d}_1, \mathbf{d}_2\}$. For each data subset, we can do an independent optimization to find the best parameter set given the smaller data set:

$$\theta_i = \arg\max_{\theta}\left[\log(L(\mathbf{d}_i|\theta))\right]. \tag{20}$$

Now for each parameter set, we can compute the likelihood that all of the data in $\mathbf{D}$ matches the model fit only to the subset $\mathbf{d}_i$:

$$\log(L_i) = \log(L(\mathbf{D}|\theta_i)) \leq \log(L_{\text{Fit}}), \tag{21}$$

The cross-validation error is then computed as the average of this quantity:

$$\log(L_{CrossVal}) = \frac{1}{2}\sum_{i=1}^{2}\log(L(\mathbf{D}|\theta_i)) \leq \log(L_{\text{Fit}}). \tag{22}$$

Intuitively, the differences in the parameter sets $\{\theta_i\}$ give a sense of the parameter uncertainty for the corresponding model, and the differences in the cross validation error, $\log(L_{CrossVal})$, give an estimate of the predictive uncertainty for the model, given the current level of data.

Choosing the number of states

We allow each of the $2(N-1)$ state transitions to depend upon the level of Hog1p(t). Using two parameters per Hog1p reaction as in Eq. 9, this requires 4(N-1) parameters. Furthermore, production at each of the N states and degradation and t_0 requires another $N+2$ parameters, giving a total of $5N-2$ parameters for an N-state model. We consider general two-, three-, four-, and five-state model structures of this form, and we fit these to the wild-type data set of $STL1$ mRNA at 0.4M NaCl at all 16 time points.

Fig. S11 shows the comparison between the experimental distributions and the best fit for each of the two-, three-, four-, and five-state model structures. In the figures, the two experimental data sets are shown in red and blue, and the best fit is shown with a black line. From the figure, it is clear that the two-state structure matches the main quantitative features of the experimental data, but fails to capture the quantitative distribution at some of the transient time points (especially at 30 minutes after osmotic shock). The three-state structure does a slightly better job of matching the data, and the four state model provides a near perfect fit. However, the five-state structure provides only a marginal improvement in its match the experimental data. For a comparison of the model complexity, the two-state Hog1p-complete model structure has 8 parameters, the three-state Hog1p-complete model structure has 13 parameters, the four-state Hog1p-complete model structure has 18 parameters, and the five-state model has 23 parameters. For comparison, the best four-state, one kinase model (fig. S11, bottom row) has only 13 parameters.

As expected, the model fits improve with the number of states–the model structure with the fewest states has the worst fit, whereas the model structure with most parameters achieves the best fit. However, as complexity increases, so does the uncertainty and the chance for over-fitting. Since we eventually wish to use the identified models to make predictions of mRNA dynamics, this effect must be avoided. We use a cross validation approach to quantify the parameter uncertainty in the fits, and this uncertainty is plotted as the grey regions in fig. S11 for each of the models. As discussed in the main text and summarized in Fig. 2B, it is clear from fig. S11 that the total uncertainty in the models generally increases with the number of free parameters (or complexity) of the models.

All parameters were fixed based upon their fits to the 0.4M osmotic shock, and the resulting models were used to predict the distributions of $STL1$ mRNA at a lower osmotic shock of 0.2M NaCl. These predictions and their uncertainty are shown in fig. S12. From the figure it is clear that as the number of states increases from two to four, the predictions first improve, but are subject to greater uncertainty. The four-state, one-kinase model provides the best quantitative prediction of the distributions.

Multi-objective fits

Having determined that the chosen four-state, one-kinase model provides the best combination of fit and prediction accuracy of the considered models, we test the possibility of fitting that same model structure to several additional mutants and genes. These include a five-fold Hot1p over-expression strain and gene knockouts of the chromatin modifiers $ARP8$ or $GCN5$ as well as two additional stress response genes $CTT1$ and $HSP12$. For this, we use a multiple-objective fitting criterion which includes $STL1$ expression in four strains (wild-type, Hot1p 5x, $arp8\Delta$, and $gcn5\Delta$) simultaneously with $CTT1$ and $HSP12$ in wild-type strains.

As before, our goal is to maximize the likelihood that the observed data came from the considered model. This objective can be extended as:

$$\max_{\{\theta_e\}} \sum_e \log(L(\mathbf{D}_e|\theta_e)) + \lambda(\{\theta_e\}), \tag{23}$$

where $e \in \{STL1_{WT}, STL1_{Hot1p5x}, STL1_{arp8\Delta}, STL1_{gcn5\Delta}, CTT1_{WT}, HSP12_{WT}\}$ refers to the six different mRNA distribution series, θ_e refers to the parameter set for that experimental condition, and $\log(L(\mathbf{D}_e|\theta_e))$ is the log likelihood that the measured mRNA distribution series corresponding, $\mathbf{D}_e$, comes from the model with the parameters θ_e. The function $\lambda(\{\theta_e\})$ is used to penalize differences in parameters between the different genes, experimental conditions and genetic mutations.

We consider two different choices for the parameter constraining function $\lambda(\{\theta_e\})$. For the different genes in the wild-type strain ($STL1$, $CTT1$, and $HSP12$), we allow all parameters to be completely independent, $\lambda(\{\theta_e\}) = 0$. This allows us to fit each condition separately to get the best possible fit to the mRNA distributions. Moreover, we fit and conduct the cross validation analysis for these distributions. Fig. S14 shows the aggregated fits and cross-validation results with this approach versus increasing model complexity. As was the case for the wild-type $STL1$, the four-state, one-kinase model remains the best choice when compared on all three wild-type genes.

Next we apply a non-trivial constraint function on the parameters:

$$\lambda(\{\theta_e\}) = \sum_{e,i} \left| \log\left(\frac{\theta_{e,i}}{\theta_{STL1-WT,i}}\right)\right|. \tag{24}$$

Essentially, we are penalizing any fold change in the parameters in relationship to that corresponding to the $STL1$ dynamics in the wild-type model. The summation over the index i corresponds to summing over all of the individual parameters. By adding this constraint on the parameters set and forcing a single model to fit all six sets of mRNA distributions, the fit to the wild-type data is slightly diminished (compare fig. S14, red bar to the red line). However, the ability of the model to predict new data is substantially improved (compare fig. S14, green bar to the green line).

Supplementary Text

Importance of using full probability distributions

Several recent studies have shown that transcriptional models can be identified from information about the means and variance alone (*38, 39*). With this in mind, we also identified all parameters of the final model from the means and variances at all time points for the wild type cells at 0.4M NaCl. It is interesting to compare the fits and predictions of the model when identified from the two different data sets: means and variances or from full distributions. As expected, when the full distribution is ignored, one can achieve better fit for the means and variances. However, when these two different fits are used to predict the dynamics of the means and variances at another experimental condition (0.2M NaCl), we find that the model identified from the distributions does slightly better (compare gold and red bars in fig. S13A, left). However, a much larger difference is seen in the predictions of the full distributions at 0.2M NaCl. The prediction data at 0.2M NaCl is e^{1566} times more likely to have come from the model fit to the full distributions (compare gold and red lines in fig. S13A, right). The reason for this extreme difference is that the fit with only the means and variances fails to capture the bimodality in the experimental data (fig. S13B, left), whereas the model fit to the full data set can fit this bimodality at 0.4M NaCl and predict it at 0.2M NaCl (fig. S13B, right). Thus, although means and variances alone can sufficiently constrain models and correctly predict bimodality in other cases (*38*), quantification and analysis of full distributions is crucial for this system and these data.

Parameters for the final model

The final parameters of the model after multi-objective fitting are given in Table S2. This final parameter set has been fit to the mRNA distributions at 16 transient time points following osmotic shock for the $STL1$ mRNA in four different mutant strains as well as for the $CTT1$ and $HSP12$ mRNA in the wild-type strain. These full fits are shown in fig. S15. Table S2 shows which fit parameters are most sensitive to the genetic mutations, which parameters change the most from one gene to another, and which are held constant in all conditions. In particular, deletion of $ARP8$ increases the transition rate from S1 to S2, which results in faster activation of $STL1$ expression. Conversely, deletion of $GCN5$ decreases the transition rate from S1 to S2, which results in slower activation. Both $GCN5$ and $ARP8$ decrease the rate of transition from S2 to S1, which result in more prolonged activation times. Furthermore, $CTT1$ and $HSP12$ exhibit slower mRNA degradation rates and lower mRNA expression rates in the S4 state, which prolong their expression in relationship to $STL1$ while maintaining a similar maximum level of mRNA expression.

Predictions and comparisons to new experiments

Due to space constraints, the main text shows only a small portion of the predictions of the final model. This section documents some other predictions that have been made and validated with experimental measurements. Predictions of the distributions for $STL1$, $CTT1$ and $HSP12$ in the wild-type strain subjected to a 0.2M NaCl osmotic shock are shown in fig. S16. Predictions for the transient distributions of $CTT1$ and $HSP12$ genes in response to 0.4M NaCl osmotic shock under the effect of Hot1p over-expression or $ARP8$/$GCN5$ deletions are shown in fig. S17. Fig. S18 shows the measured and predicted fractions of cells with more than a single mRNA as a function of time for the different experimental conditions. Fig. S19 shows the measured and predicted mean level of mRNA in each cell as a function of time for the different experimental conditions.

References

30. Muzzey, D., C. Gomez-Uribe, J. Mettetal, and A. van Oudenaarden. 2009. A systems-level analysis of perfect adaptation in yeast osmoregulation. *Cell*. 138:160–171.

31. Mettetal, J. T., D. Muzzey, C. Gmez-Uribe, and A. van Oudenaarden. 2008. The Frequency Dependence of Osmo-Adaptation in Saccharomyces cerevisiae. *Science*. 319:482–484.

32. Ferrigno, P., F. Posas, D. Koepp, H. Saito, , and P. Silver. 1998. Regulated nucleo/cytoplasmic exchange of hog1 mapk requires the importin bold beta homologs nmd5 and xpo1. *The EMBO Journal*. 17:5606 – 5614.

33. Bumgarner, S., G. Neuert, B. Voight, A. Symbor-Nagrabska, P. Grisafi, A. van Oudenaarden, and G. Fink. 2012. Single-cell analysis reveals that noncoding rnas contribute to clonal heterogeneity by modulating transcription factor recruitment. *Molecular Cell*. 45:470–482.

34. Raj, A., P. van den Bogaard, S. Rifkin, A. van Oudenaarden, and S. Tyagi. 2008. Imaging individual mrna molecules using multiple singly labeled probes. *Nature Methods*. 5:877–887.

35. Munsky, B., and M. Khammash. 2006. The finite state projection algorithm for the solution of the chemical master equation. *J. Chem. Phys.* 124.

36. Munsky, B., and M. Khammash. 2008. The finite state projection approach for the analysis of stochastic noise in gene networks. *IEEE Trans. Automat. Contr./IEEE Trans. Circuits and Systems: Part 1*. 52:201–214.

37. Munsky, B. 2008. The finite state projection approach for the solution of the chemical master equation and its application to stochastic gene regulatory networks. Ph.D. thesis, Univ. of California at Santa Barbara, Santa Barbara.

38. Zechner, C., J. Ruessa, P. Krenna, S. Pelet, M. Peter, J. Lygeros, and H. Koeppl. 2012. Moment-based inference predicts bimodality in transient gene expression. *Proc. Nat. Acad. Sci. USA*. 109:8340–8345.

39. Singh, A., B. Razooky, R. Dar, and L. Weinberger. 2010. Dynamics of protein noise can distinguish between alternate sources of gene-expression variability. *Mol Syst Biol*. 8.

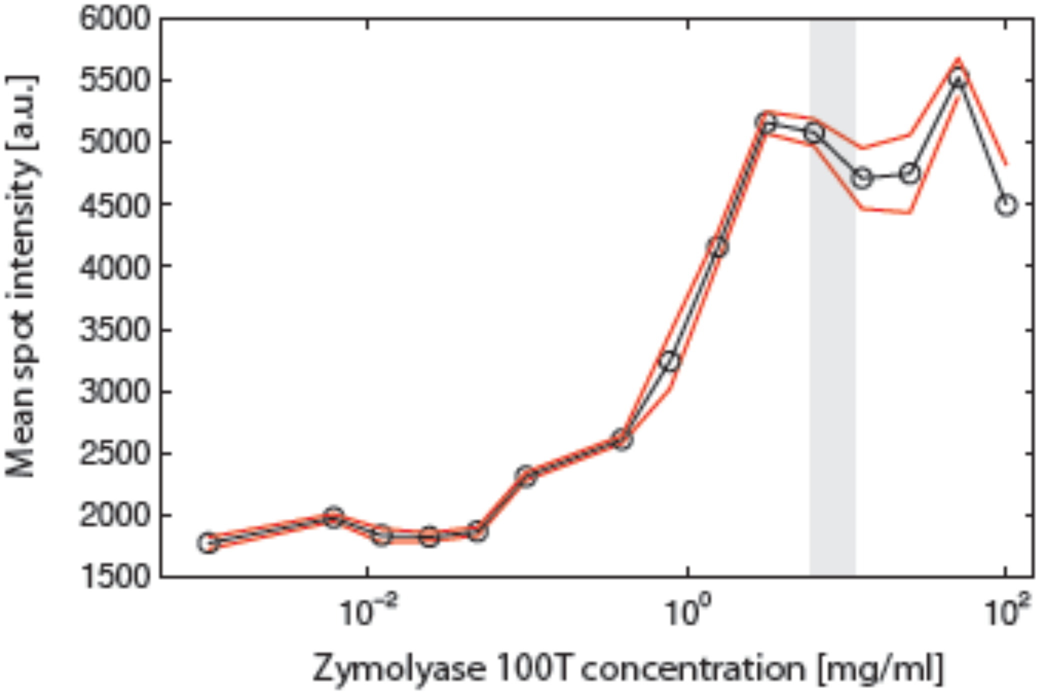

Figure S1: Average spot intensity versus Zymolase concentration.

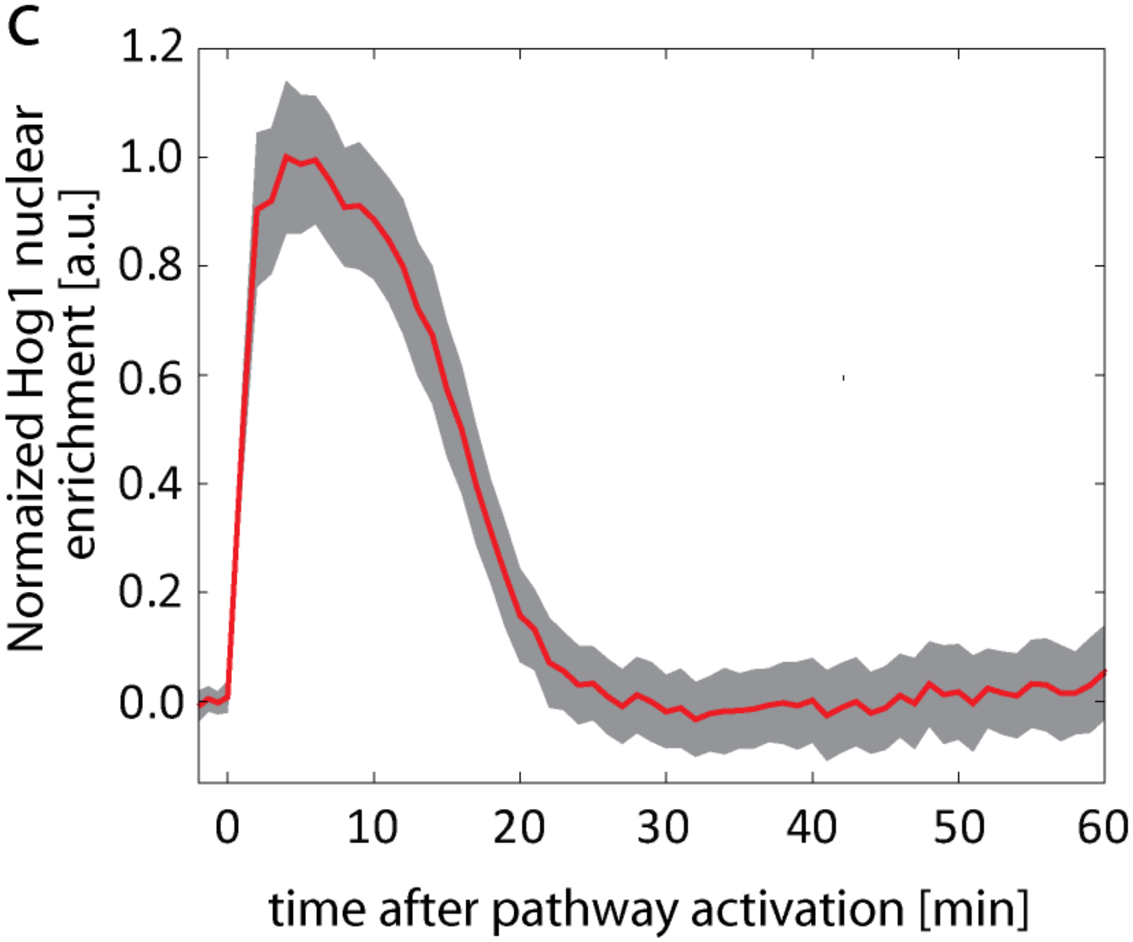

Figure S2: Normalized Hog1p nuclear enrichment measured at 0.4M NaCl. Mean (solid lines) and standard deviation (grey shade) from ∼300 single cells (adapted from (*30*)). Variation in signaling dynamics is primarily an artifact of the image analysis in the experimental approach.

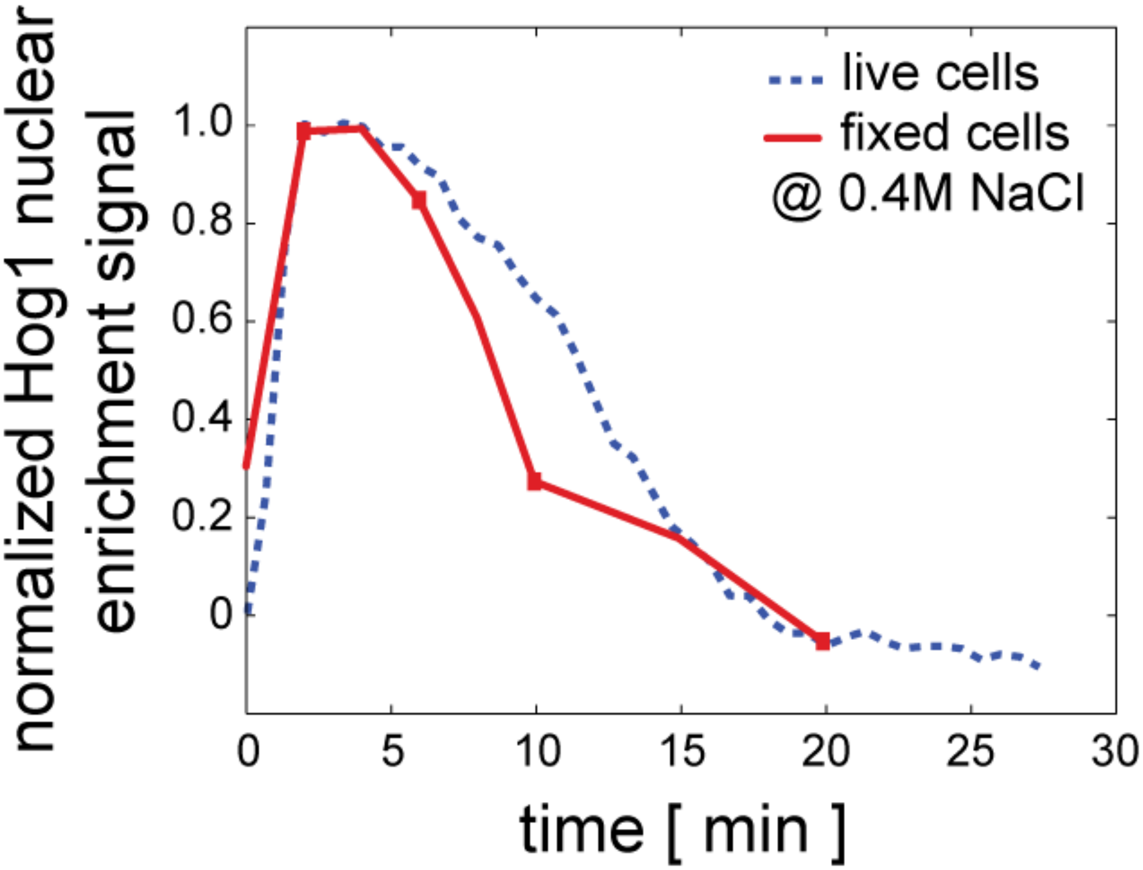

Figure S3: Comparison of normalized nuclear localization between live (blue dash line) and fixed cells (redline).

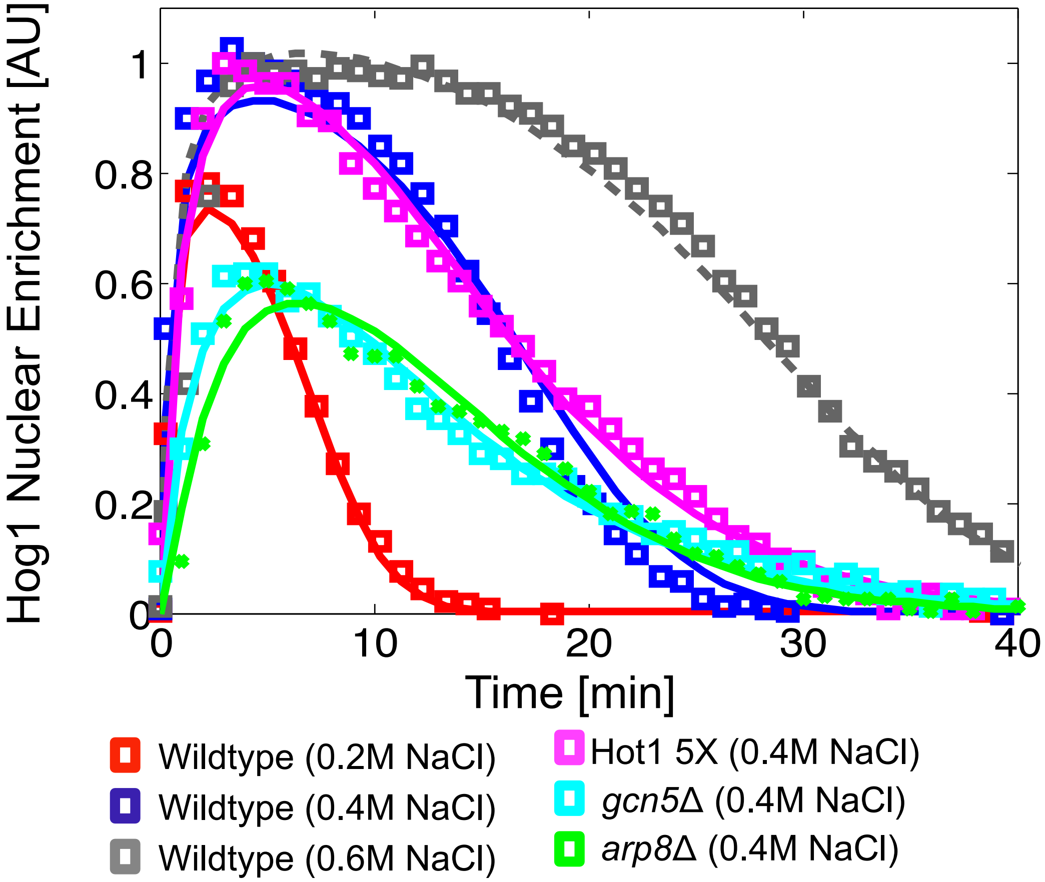

Figure S4: Kinase nuclear enrichment, Hog1p*, after activation of the HOG-pathway with different osmolyte concentrations and different genetic modifications. Color bands denote the mean measurement plus or minus one standard deviation derived from three independent experiments (adapted from (*30*)). Dashed lines correspond to the best fit for the signal at 0.2, 0.4N and 0.6M NaCl for the wild-type cells and for the *arp8*Δ, *gcn5*Δ and $Hot1p\ 5x$ mutants at 0.4M NaCl.

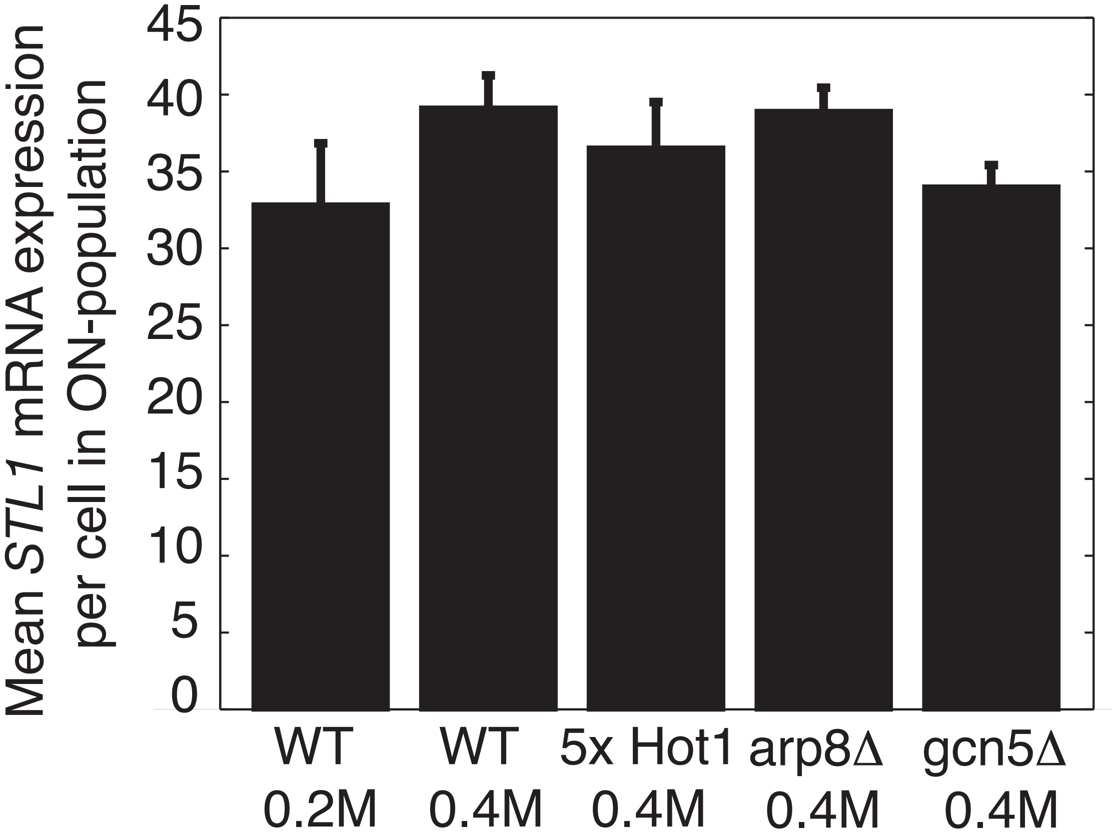

Figure S5: Mean expression level for $STL1$ in ON cells for different conditions.

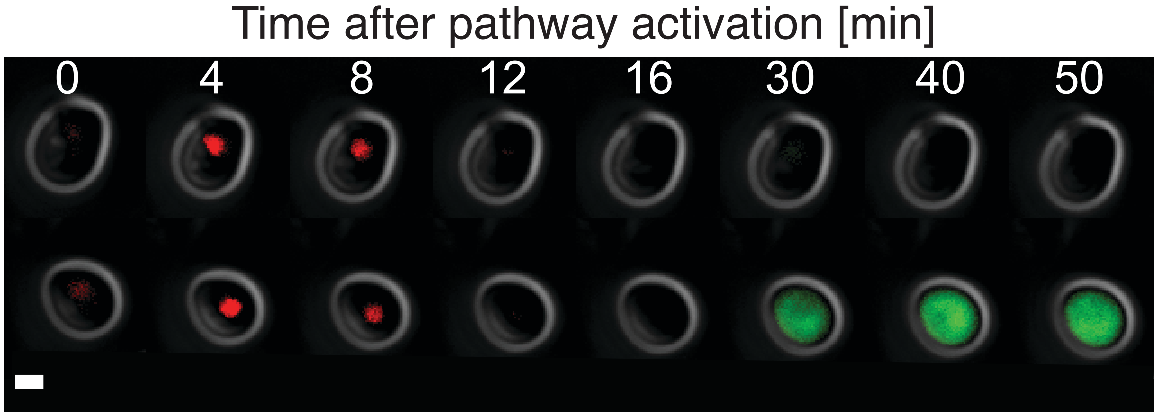

Figure S6: Time lapse microscopy of yeast cells (grey) showing nuclear localization of the yellow fluorescent protein (YFP) tagged Hog1p mitogen-activated protein kinase (MAPK, red) and stochastic and bimodal activation of the osmo-sensitive STL1 gene, as measured with a GFP fusion reporter (green, scale bar: 2 μm).

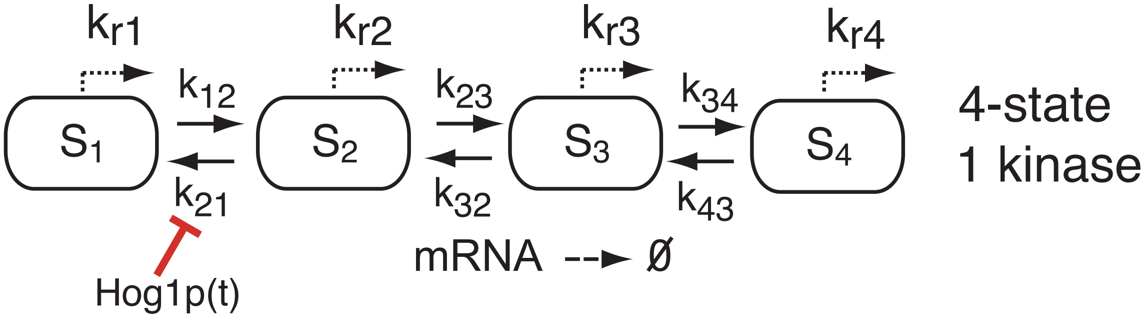

Figure S7: Schematic diagram of the best four-state Hog1p-model structure with one kinase-dependent rate (Hog1p interrupts the $S_2 \rightarrow S_1$ reaction).

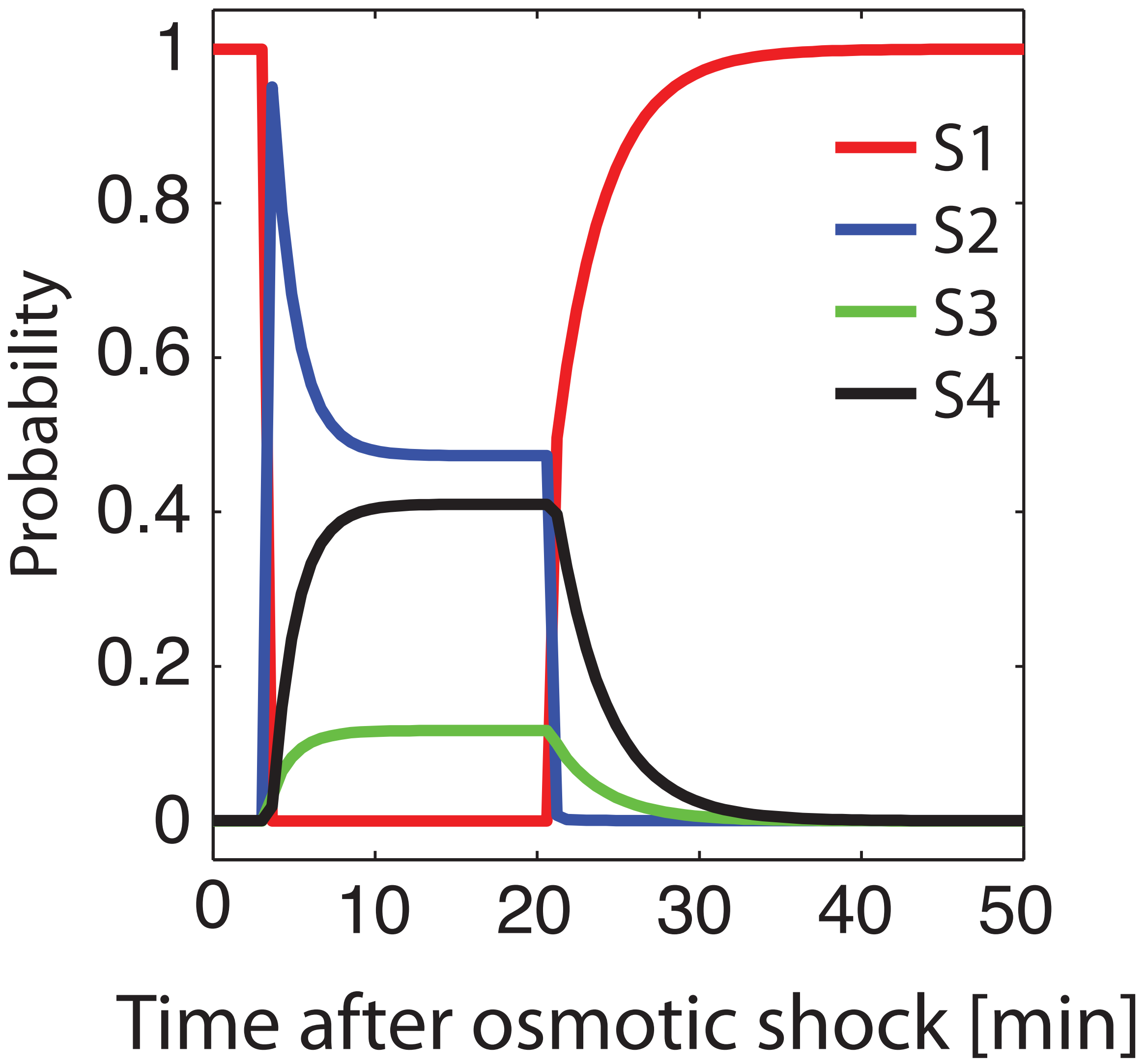

Figure S8: When the level of Hog1p exceeds a fixed threshold, it activates a rapid switch out of the OFF state S_1. When Hog1p falls below that threshold, the system slowly relaxes back to S_1. The state S_3 is a short-lived intermediate state.

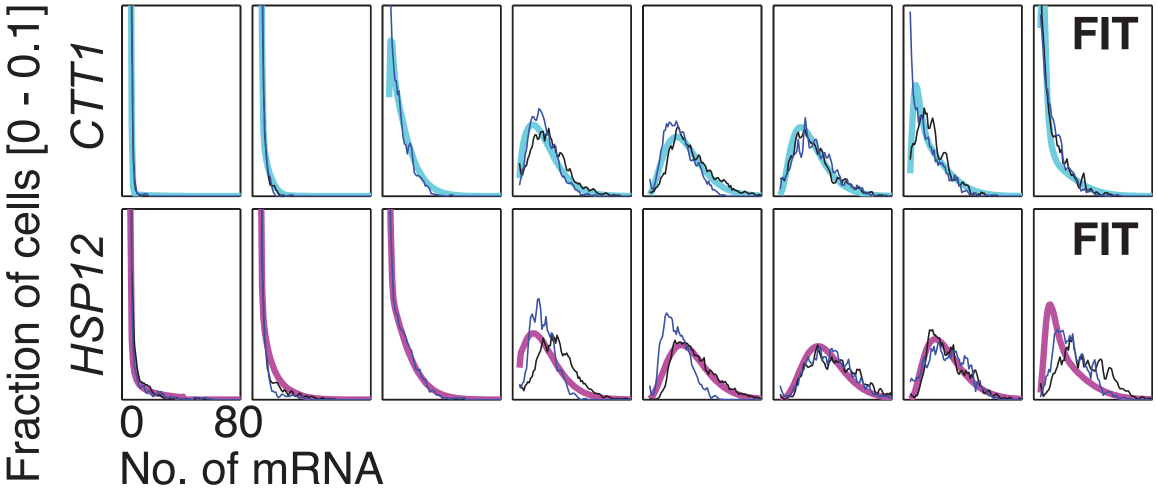

Figure S9: Combined fit of the model structure identified in Fig. 2 to Hog1p-activated mRNA expression of *CTT1* (cyan) and *HSP12* (magenta) at 0.4 M NaCl. Experimental replicas are shown with thin blue and black lines.

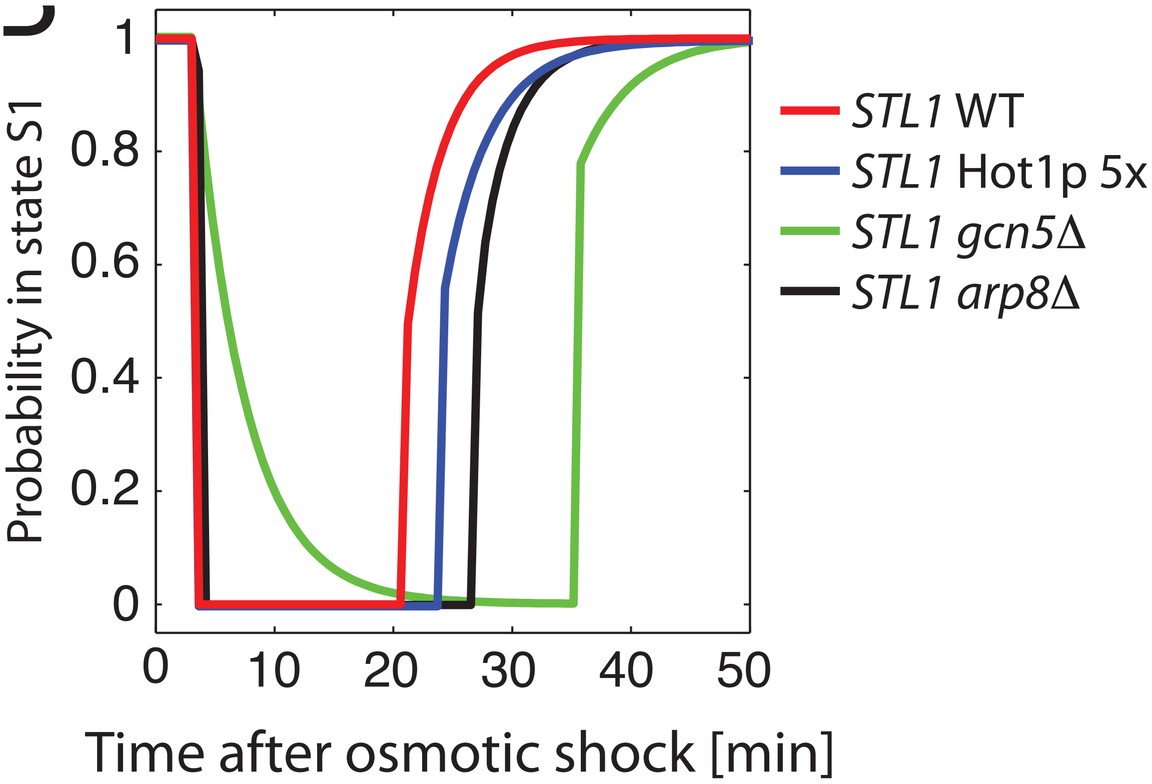

Figure S10: Probabilities in state S_1 for cells expressing *STL1* for different mutant strains.

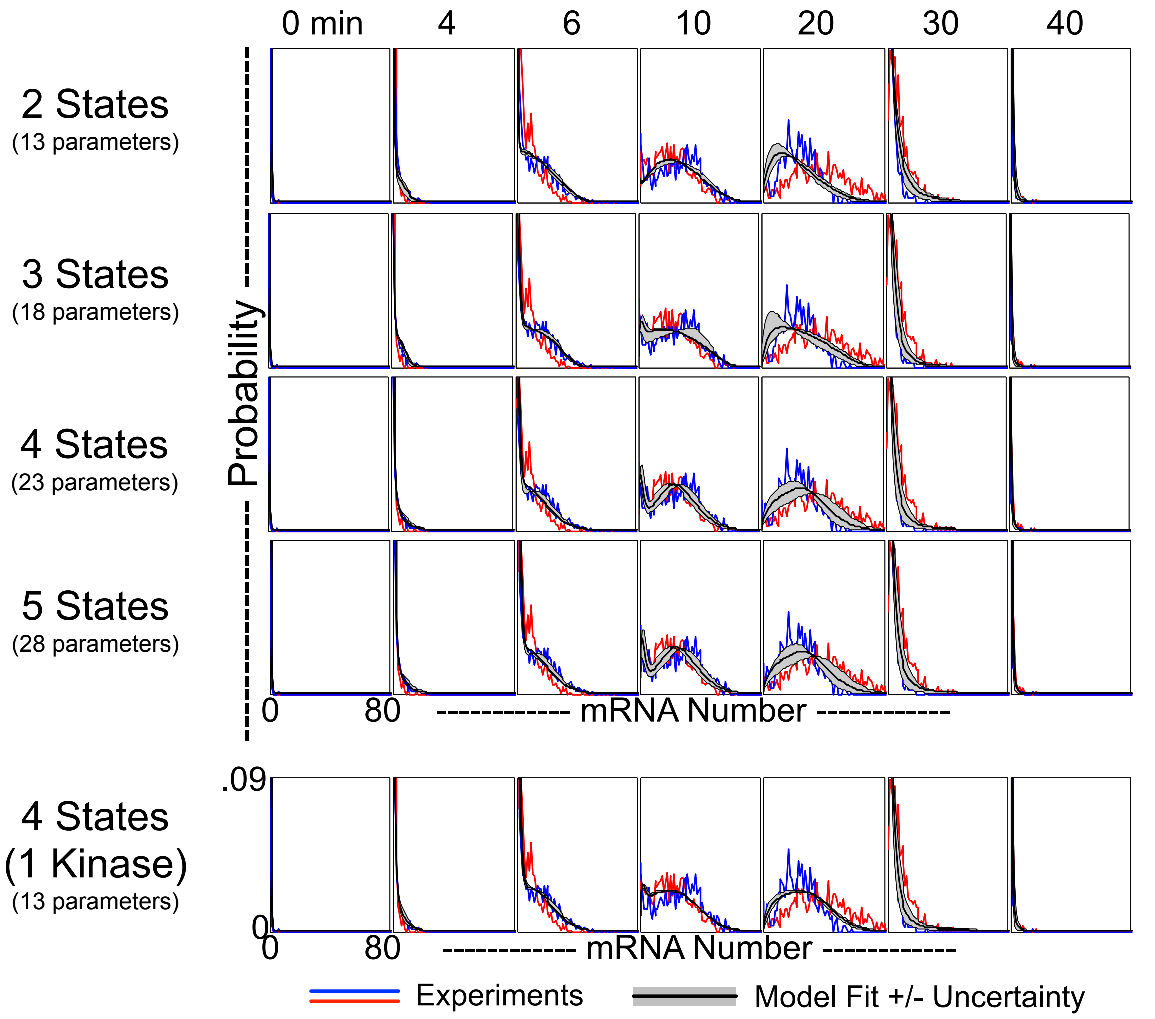

Figure S11: Measured and fitted probability distributions for the number of mRNA's versus time for an osmotic shock of 0.4M NaCl in wild-type cells. Experimental data are shown in red and blue, the best model fit is shown with black lines, and the model uncertainty is shown with grey shading. In general, as the number of parameters increases, the fits improve, but the amount of uncertainty increases.

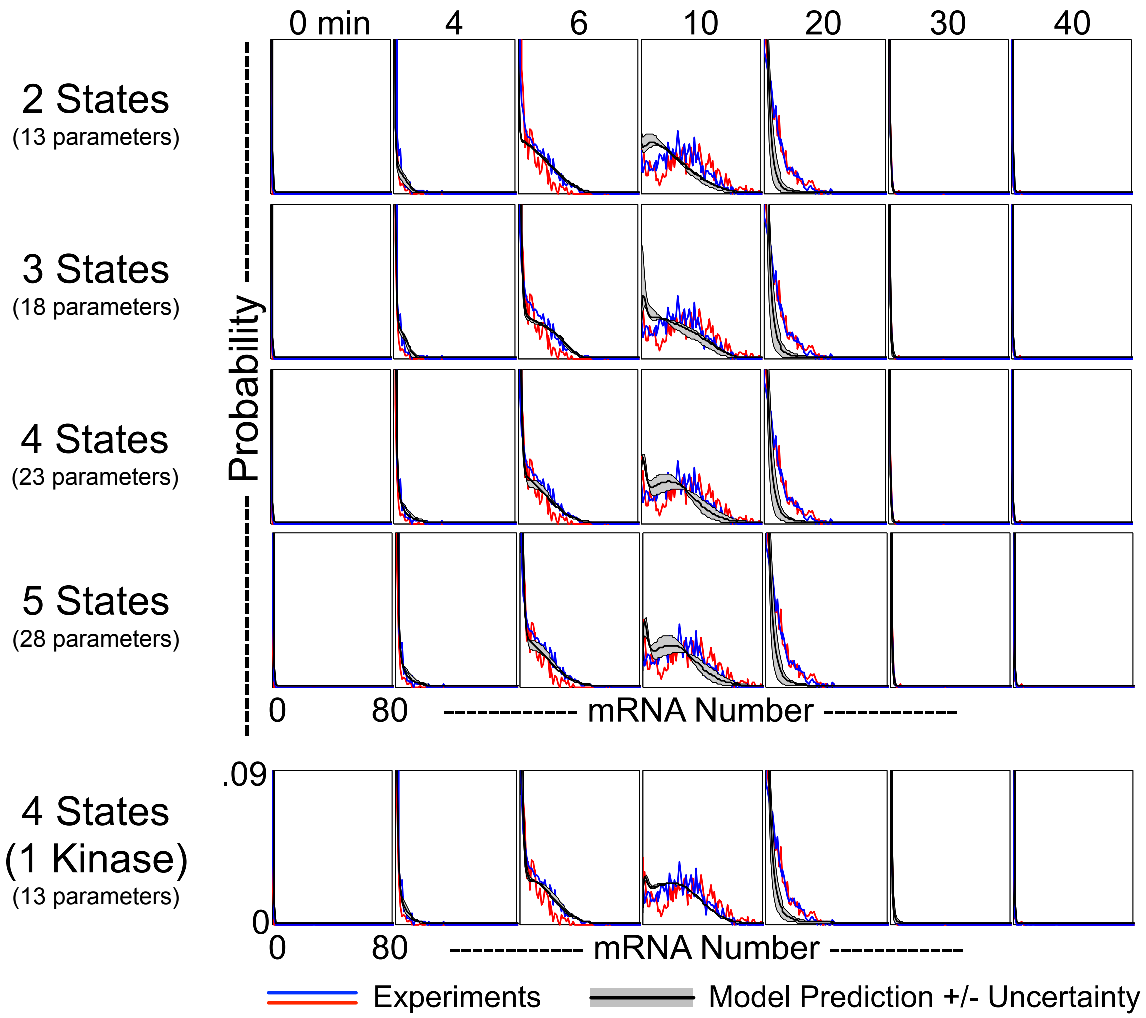

Figure S12: Predicted and measured probability distributions for the number of mRNA's versus time for an osmotic shock of 0.2M NaCl in wild-type cells. Model predictions are shown with black lines, and model uncertainty is shown with grey shading. Experimental data from two replicates are shown in red and blue. As the number of parameters increases, the fits first improve, but then worsen due to too much uncertainty. The four-state, one-kinase model provides the best quantitative prediction of the distributions.

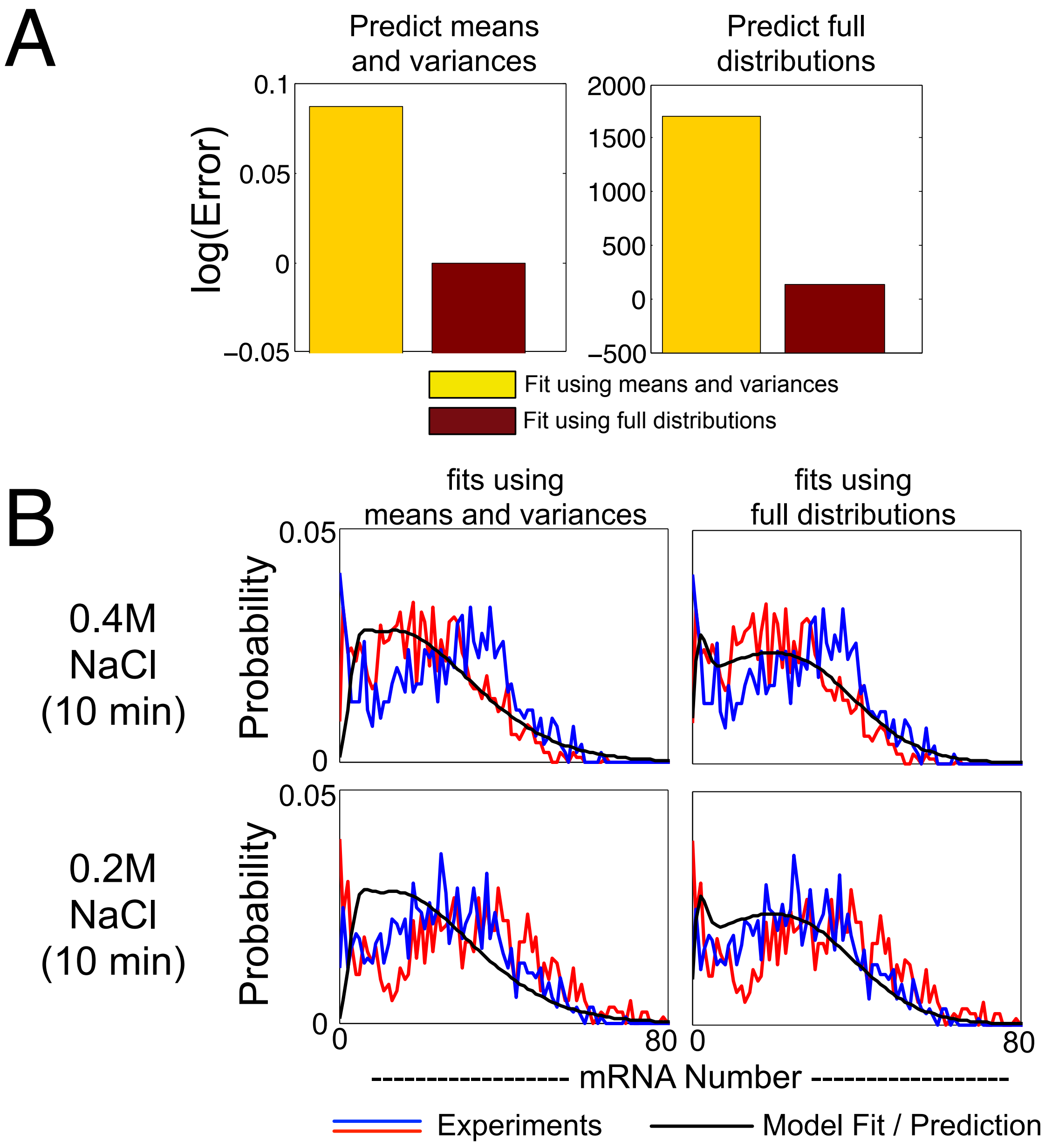

Figure S13: Comparison between model fits and predictions, when the model is identified from the first two moments (*i.e.*, means and variances) or from the full probability distributions. A) Errors in the first two moments (left) or in the full distributions (right). Gold: prediction errors at 0.2M NaCl, when means and variances are used in the fitting. Red: prediction errors at 0.2M NaCl, when full distributions are used in the fitting. B) Fits at 0.4M NaCl (top) and predictions at 0.2M NaCl (bottom), when the model has been identified using the first two moments (left) or with the full distributions (right). Experimental data sets are shown in red and blue and the model fit (top) and predictions (bottom) are shown in black.

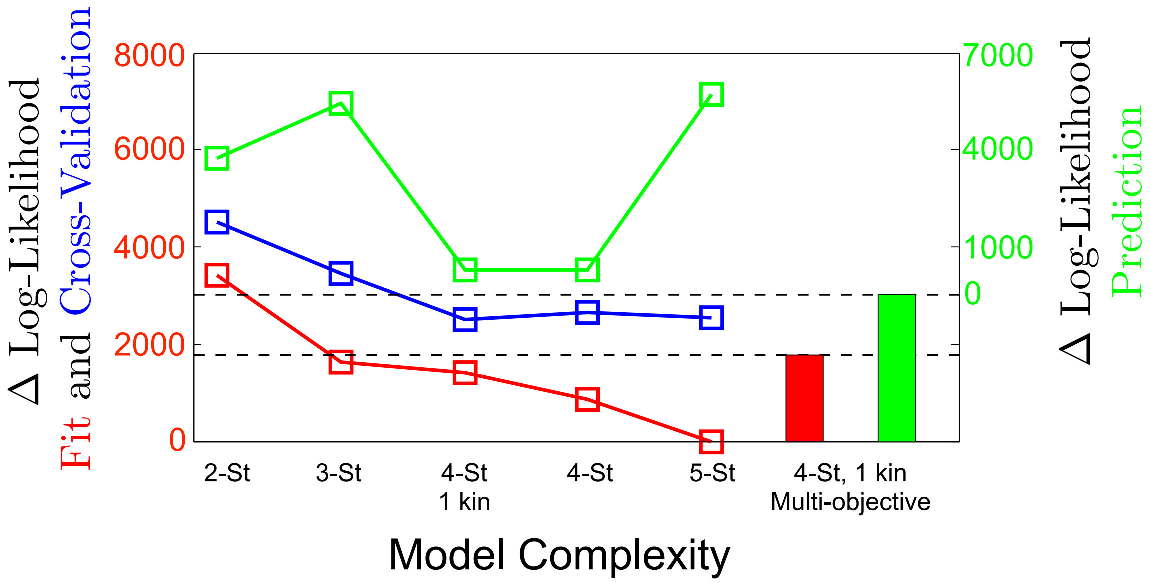

Figure S14: Relative likelihoods of best fit for different model structures at 0.4M NaCl (red, left axis) and the resulting predictions at 0.2M NaCl (green, right axis) for the three genes: $STL1$, $CTT1$, and $HSP12$. Cross-validation at 0.4M NaCl is used to quantify predictive uncertainty and yields excellent a priori knowledge of predictive power (compare blue and green lines). The lines correspond to the case where the three wild-type genes ($STL1$, $CTT1$, and $HSP12$) are fit independently. The bars on the right correspond to the fit and prediction quality when the three genes ($STL1$, $CTT1$, and $HSP12$) and the three mutants (Hot1p 5x, $arp8\Delta$, and $gcn5\Delta$) are fit simultaneously but compared only in the context of the $STL1$, $CTT1$, and $HSP12$ mRNA distributions.

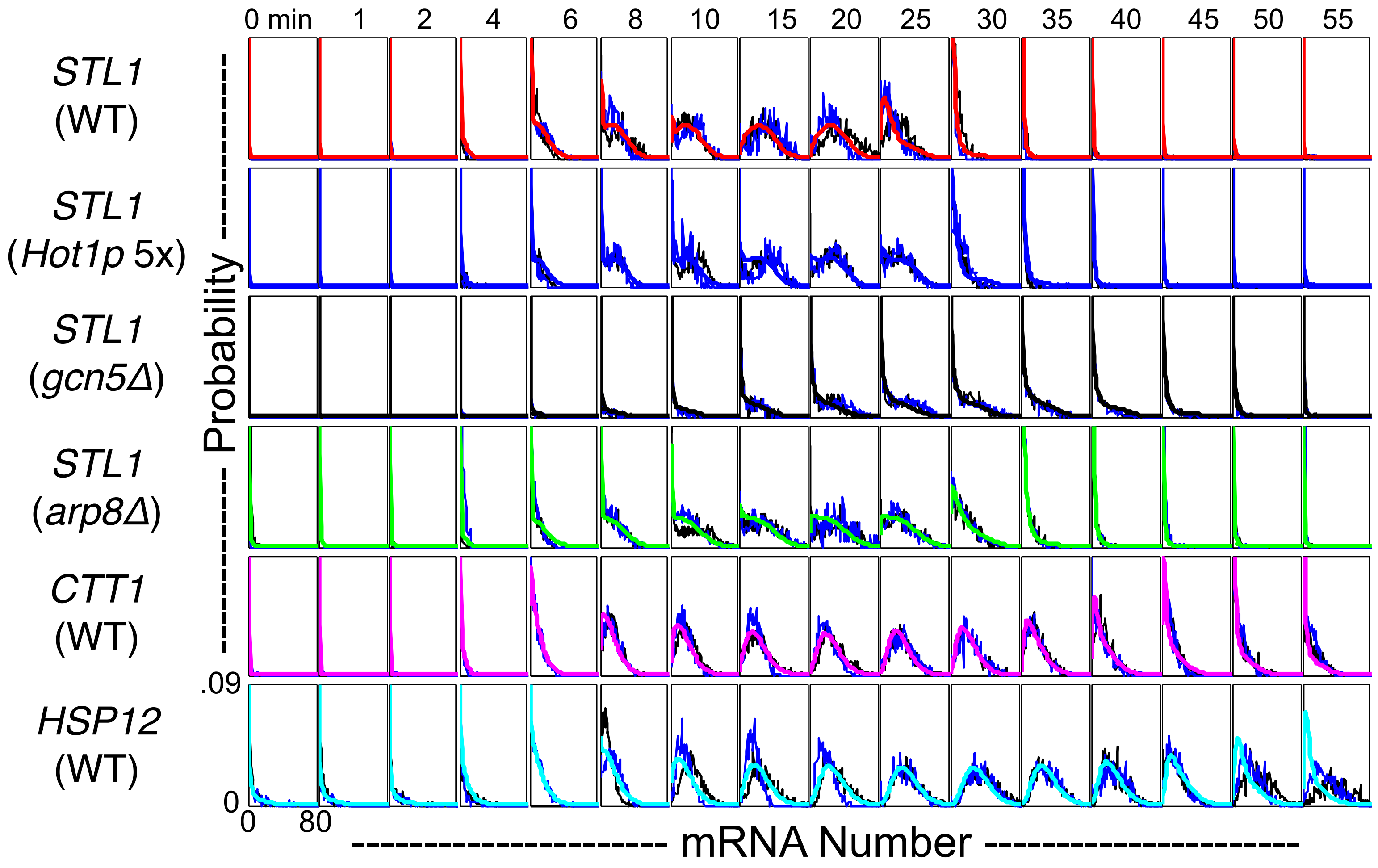

Figure S15: Model fits for the final model (structure and parameters) at the time points {0,1,2,4,6,8,10,15,20,25,30,35,40,45,50,55} minutes after osmolite induction for each of the six genes and mutations: $STL1$ wild-type (red), $STL1$ 5x Hot1p expression (blue), $STL1$ $gcn5\Delta$ mutant (black), $STL1$ $arp8\Delta$ mutant (green), $CTT1$ wild-type (magenta), and $HSP12$ wild-type (cyan). In each condition, there are two experimental replicas shown in blue and black for an osmotic shock of 0.4M NaCl. The final model parameters are shown in Table S2.

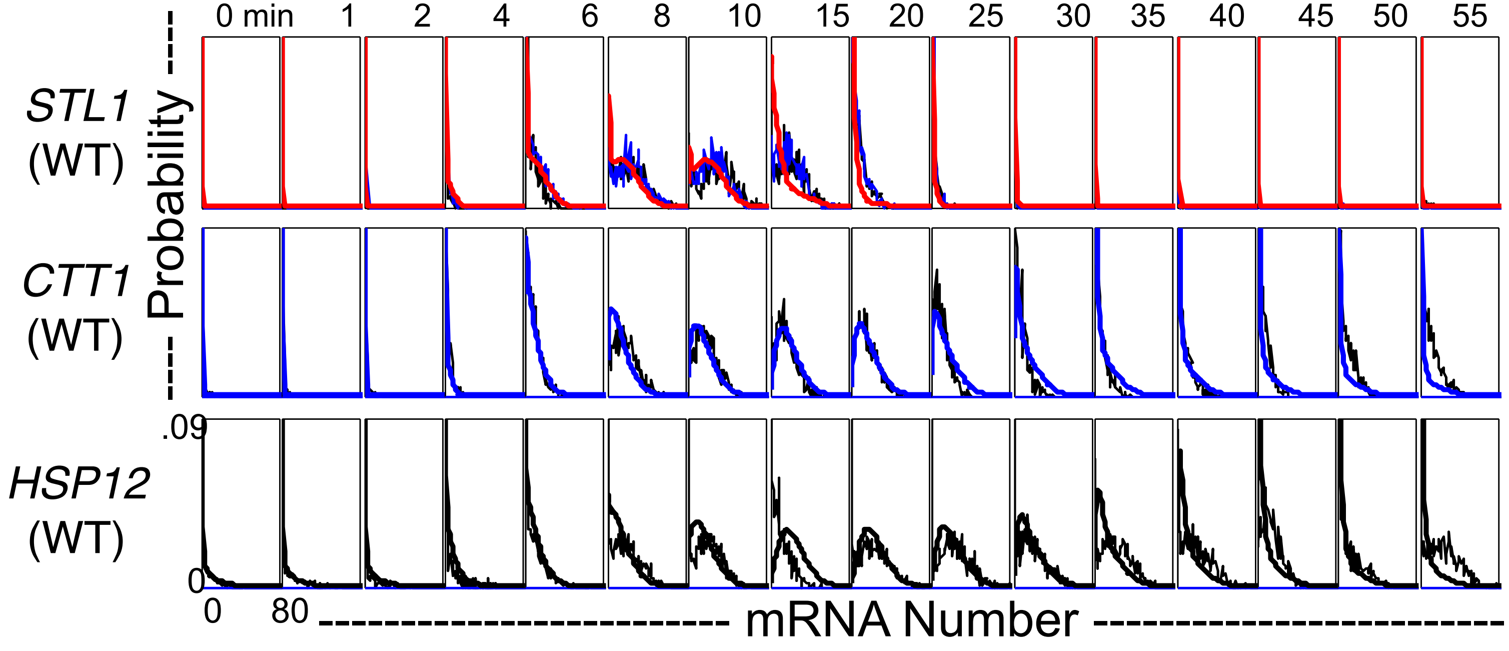

Figure S16: Final model predictions at the time points {0,1,2,4,6,8,10,15,20,25,30,35,40,45,50,55} minutes after a 0.2M NaCl osmolite induction $STL1$ (red), $CTT1$ (blue) and $HSP12$ (black). Parameters were taken from the response of these genes under 0.4M NaCl osmotic shock (i.e., there are no free parameters).

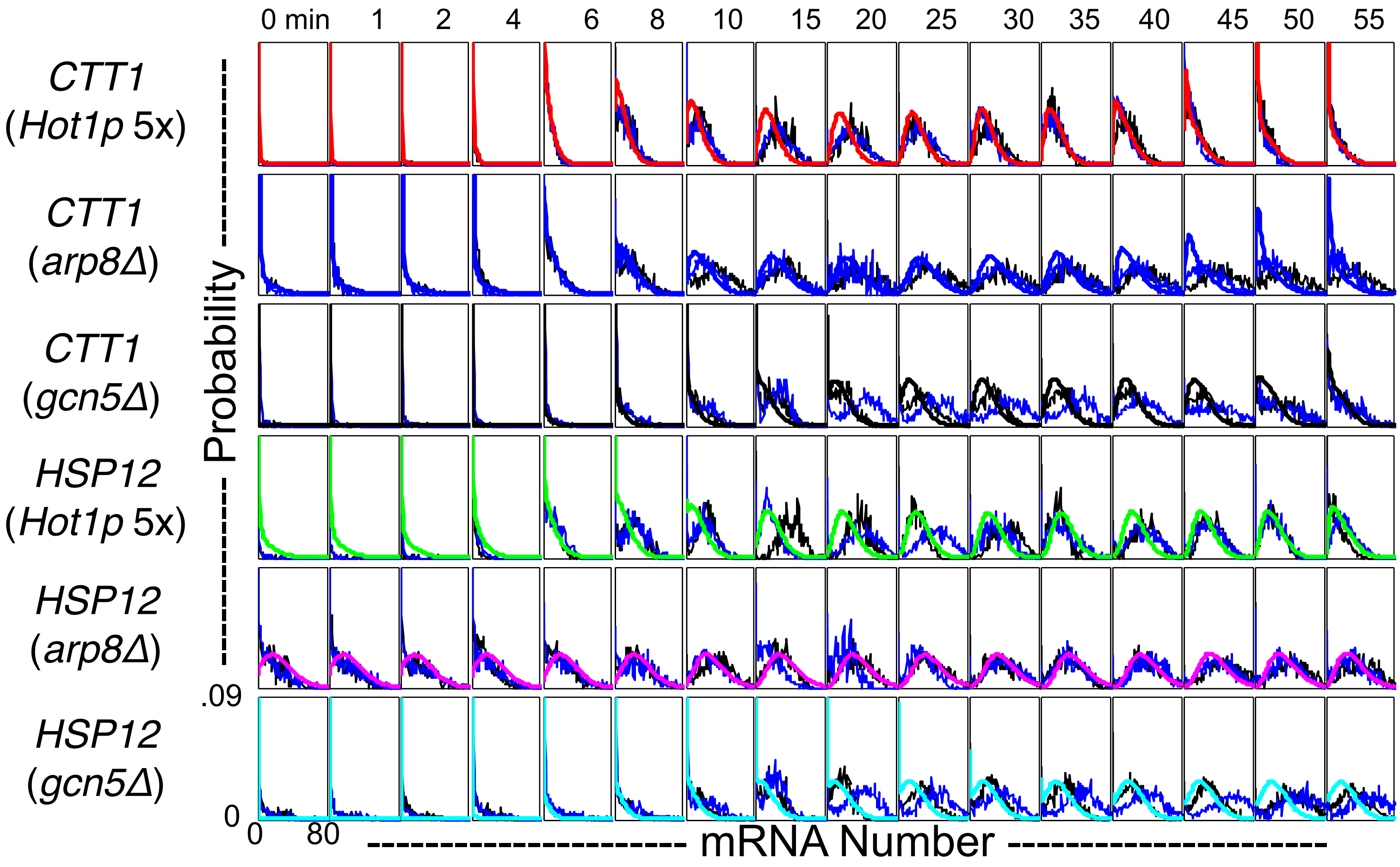

Figure S17: Final model predictions at the time points $\{0,1,2,4,6,8,10,15,20,25,30,35,40,45,50,55\}$ minutes after a 0.4M NaCl osmolite induction for $CTT1$ and $HSP12$ mRNA distributions under three different mutants: 5x Hot1p expression, $arp8\Delta$, and $gcn5\Delta$. The parameters used in these predictions are determined from the mutational effects on the $STL1$ transcriptional dynamics and the differences between $STL1$ and $CTT1/HSP12$ in the wild-type strain (i.e., there are no free parameters).

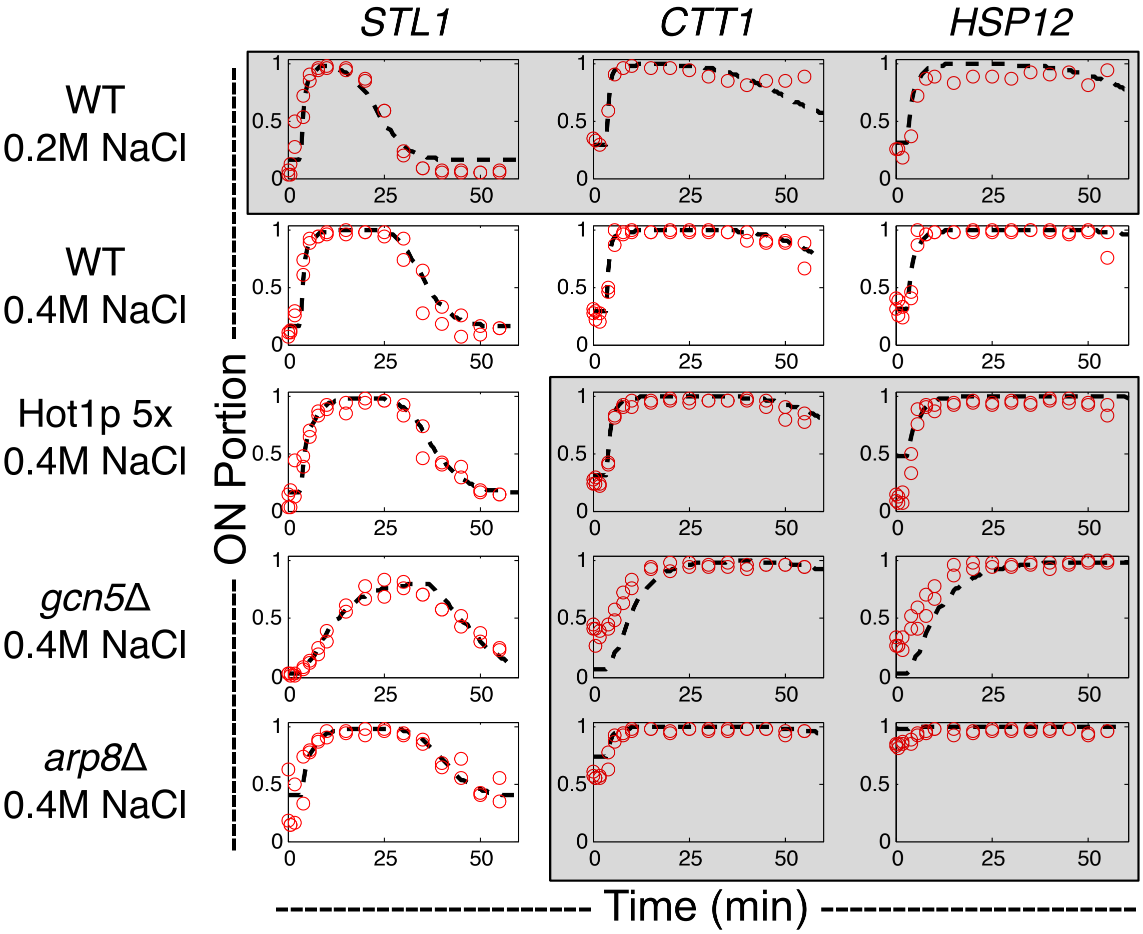

Figure S18: ON-fraction predictions for the final model versus time after osmolite induction. Predictions are made for each of the three genes ($STL1$, $CTT1$, and $HSP12$) in each of five genetic and environmental conditions (wild-type at 0.4M NaCl, wild-type at 0.2M NaCl, 5x Hot1p over expression at 0.4M NaCl, $arp8\Delta$ mutation and $gcn5\Delta$ mutation). The genes/conditions combinations in the shaded region were predicted with no free parameters.

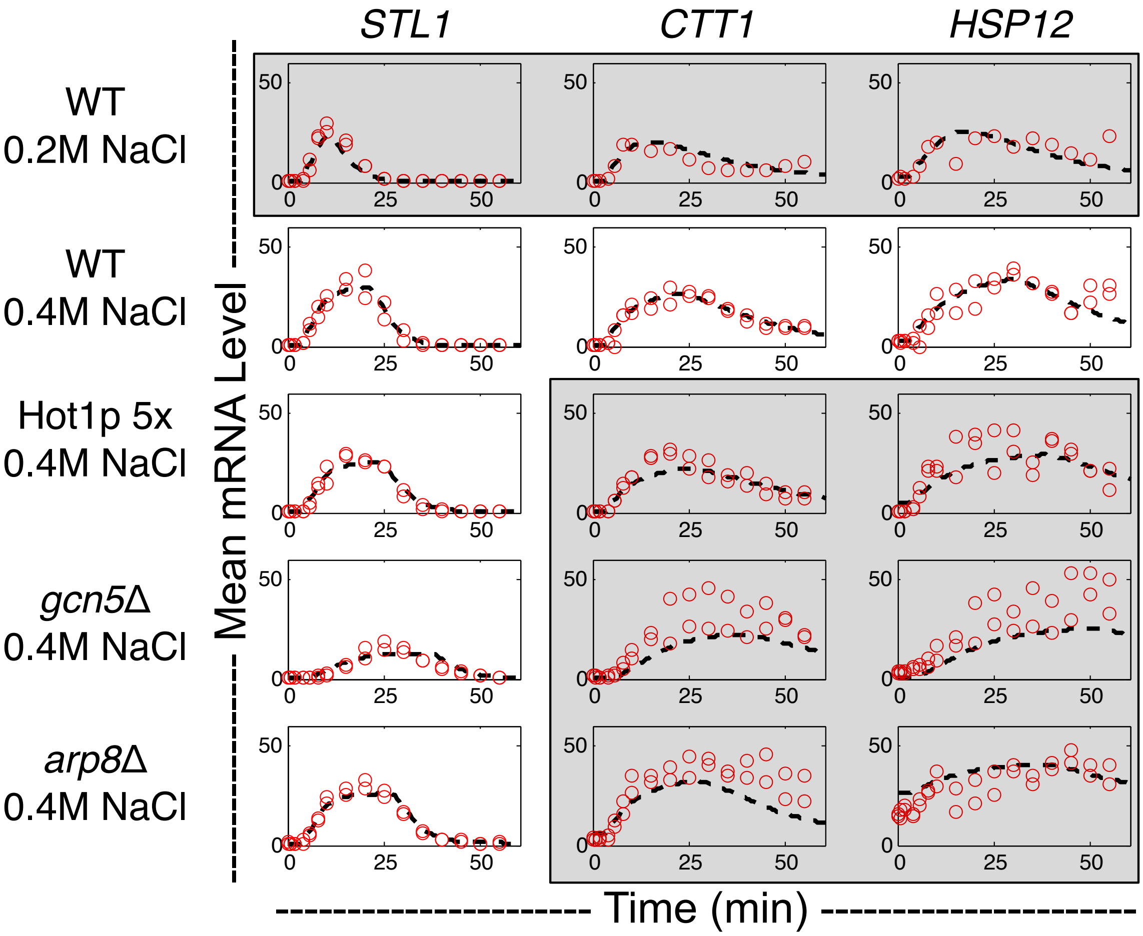

Figure S19: Mean mRNA level predictions for the final model versus time after osmolite induction. Predictions are made for each of the three genes ($STL1$, $CTT1$, and $HSP12$) in each of five genetic and environmental conditions (wild-type at 0.4M NaCl, wild-type at 0.2M NaCl, 5x Hot1p over expression at 0.4M NaCl, $arp8\Delta$ mutation and $gcn5\Delta$ mutation). The genes/conditions combinations in the shaded region were predicted with no free parameters.

Mutant Strain	Parameter Values				
	r_1 (s^{-1})	r_2 (s^{-1})	η -	A -	M -
Wildtype (0.2M NaCl)	6.9×10^{-5}	7.1×10^{-3}	3.1	9.3×10^{9}	6.4×10^{-4}
Wildtype (0.4M NaCl)	6.9×10^{-5}	3.6×10^{-3}	3.1	9.3×10^{9}	6.4×10^{-4}
Wildtype (0.6M NaCl)	6.9×10^{-5}	2.4×10^{-3}	3.1	9.3×10^{9}	6.4×10^{-4}
$arp8\Delta$ (0.4M NaCl)	2.2×10^{-9}	2.6×10^{-3}	1.6	9.3×10^{9}	4.1×10^{-7}
$gcn5\Delta$ (0.4M NaCl)	4.4×10^{-15}	3.2×10^{-3}	0.85	9.3×10^{9}	1.7×10^{-12}
$Hot1p\ 5x$ (0.4M NaCl)	2.9×10^{-14}	3.4×10^{-3}	0.87	9.3×10^{9}	3.2×10^{-12}

Table S1: Parameterization of the Hog1p nuclear enrichment signal for the wild-type strain at different levels of osmotic shock and for the mutant strains at 0.4M NaCl osmotic shock.

Mutant Strain	**Final Model Parameter Values**											
	k_{12}	k_{21}	k_{23}	k_{32}	k_{34}	k_{43}	k_{r1}	k_{r2}	k_{r3}	k_{r4}	δ	t_0
$STL1$ (WT)	1.3	$3200(1 - 2.4Hog1p)$	0.0067	0.027	0.13	0.038	0.00078	0.012	0.99	0.054	0.0049	190
$STL1$ (Hot1p 5x)	1.3	$1100(1 - 2.4Hog1p)$	0.0045	0.024	0.15	0.038	0.00080	0.0052	0.90	0.082	0.0049	190
$STL1$ ($arp8\Delta$)	0.0039	$120(1 - 2.4Hog1p)$	0.0013	0.021	0.16	0.038	0.00017	0.0020	1.0	0.054	0.0049	190
$STL1$ ($gcn5\Delta$)	5.8	$420(1 - 2.4Hog1p)$	0.0048	0.027	0.11	0.038	0.0021	0.0054	1.0	0.054	0.0047	180
$CTT1$ (WT)	1.3	$3200(1 - 2.4Hog1p)$	0.019	0.018	0.13	0.0083	0.00062	0.0098	1.0	0.0016	0.0020	200
$HSP12$ (WT)	4.7	$490(1 - 2.4Hog1p)$	0.0096	0.024	0.13	0.0083	0.00013	0.0023	1.0	0.0022	0.0014	190

Table S2: Parameter values for the final model for each of the different genes and mutant strains.