

# Moment-based inference predicts bimodality in transient gene expression

Christoph Zechner<sup>a,1</sup>, Jakob Ruess<sup>a,1</sup>, Peter Krenn<sup>a</sup>, Serge Pelet<sup>b</sup>, Matthias Peter<sup>b</sup>, John Lygeros<sup>a</sup>, and Heinz Koeppl<sup>a,2</sup>

<sup>a</sup>Automatic Control Laboratory, and <sup>b</sup>Institute of Biochemistry, ETH Zurich, CH-8092 Zurich, Switzerland

Edited by Peter J. Bickel, University of California, Berkeley, CA, and approved April 3, 2012 (received for review January 5, 2012)

Recent computational studies indicate that the molecular noise of a cellular process may be a rich source of information about process dynamics and parameters. However, accessing this source requires stochastic models that are usually difficult to analyze. Therefore, parameter estimation for stochastic systems using distribution measurements, as provided for instance by flow cytometry, currently remains limited to very small and simple systems. Here we propose a new method that makes use of low-order moments of the measured distribution and thereby keeps the essential parts of the provided information, while still staying applicable to systems of realistic size. We demonstrate how cell-to-cell variability can be incorporated into the analysis obviating the need for the ubiquitous assumption that the measurements stem from a homogeneous cell population. We demonstrate the method for a simple example of gene expression using synthetic data generated by stochastic simulation. Subsequently, we use time-lapsed flow cytometry data for the osmo-stress induced transcriptional response in budding yeast to calibrate a stochastic model, which is then used as a basis for predictions. Our results show that measurements of the mean and the variance can be enough to determine the model parameters, even if the measured distributions are not well-characterized by low-order moments only—e.g., if they are bimodal.

extrinsic variability | high-osmolarity glycerol pathway | moment dynamics | parameter inference | stochastic kinetic models

Building predictive computational models of intracellular reaction kinetics is still a dauntingly ill-posed task (1), characterized by low-dimensional experimental readouts of the hypothesized high-dimensional process. Single-cell technologies hold promise to partly alleviate this ill-posedness by exploiting the observed variability for the calibration of stochastic kinetic models (2, 3). The same technologies, however, also reveal that isogenic cells in a single population exhibit large cell-to-cell variability (4, 5). The variation can be shown to be a convolution of two sources, namely the intrinsic molecular noise and extrinsic factors that render single cells different even in the absence of molecular noise; in many cases, the latter was reported to dominate the former (4, 5). Extrinsic factors comprise difference in cell size, cell-cycle stage, expression capacity, local growth conditions—to name but a few (6, 7). Thus, although single-cell technology offers a way out of the predicament of ill-posedness, it requires new methods to deal properly with intrinsic and extrinsic variability. The effect of extrinsic variability on the dynamics of stochastic models is studied in refs. 7 and 8, whereas first attempts have been made to address the inverse problem of quantifying the extrinsic (9) and any additional intrinsic (10) components from measurements. Because the latter is based on path sampling, its applicability remains limited to small systems. Naturally, extrinsic variability is bypassed when calibrating a stochastic model to one single cell (3, 11), for instance through live-cell imaging data. However, the extent to which such a single path observation of the hypothesized stochastic process—given the notoriously sparse acquisition times—can sufficiently confine unknown process parameters remains questionable.

Stochastic kinetic models, able to capture the intrinsic noise, were proposed for modeling single-cell data and its variability (2,

12, 13). Models that track probabilities over the integer-valued state-space of molecule-counts suffer from the curse of dimensionality and are computationally prohibitive for all but the simplest systems. Similar limitations apply to approximations thereof that retain the discreteness of the state-space (14, 15). While extracting sample paths of such processes is straightforward (16), acquiring their statistics—often necessary for calibration—is hampered by the slow convergence of empirical estimates for high-dimensional models (17). This is particularly challenging because most calibration or inference methods rely on iterative schemes, making it necessary to recompute statistics. Alternative methods set out to reduce the computational burden by tracking only low-order moments instead of the whole probability distribution. A standard scheme in this class is moment closure, which provides a means to capture the stochasticity of reactions while leveraging the scalability of ordinary differential equation models (18–20).

Here we introduce a moment-based inference scheme for calibrating stochastic models with heterogeneous single-cell measurements. We show how by extending the method of moment closure by conditional moment equations one can properly account for extrinsic factors. The proposed method requires no Monte Carlo simulations over extrinsic factors, making this approach very scalable. Moreover, besides parameter estimates and their confidence bounds, the method allows one to quantitatively characterize the cell-to-cell variability, ultimately dissecting the unspecific conglomerate of extrinsic factors (6). Every additionally accounted moment of the stochastic process can make the calibration less ill-posed; in the same way as the mean of the process contains information, so does its variance. Importantly, we show that this also holds true if the process is poorly captured by the accounted moments, for instance, if we just consider first and second order moments of a multimodal process distribution.

We instantiate this computational framework by addressing a widely discussed—and we believe ubiquitous—process motif, namely the transiently induced gene expression (21, 22). Often signaling pathways are activated for a short time window, in which the activated signaling output—such as a mitogen-activated protein kinase (MAPK)—needs to initiate transcription by translocation and interaction with possibly several intermediates. If many intermediates need to be in place, some cells do not manage to transcribe at all, ultimately giving rise to bimodal protein expression profiles. The particular case study we consider is the high-osmolarity glycerol (HOG) pathway in budding yeast (23), where for intermediate induction levels a bimodality in the induced stress proteins was observed (21). We perform time-lapsed flow cytometry measurements to calibrate a stochastic kinetic

Author contributions: C.Z., J.R., M.P., J.L., and H.K. designed research; C.Z., J.R., P.K., S.P., and H.K. performed research; P.K., S.P., and M.P. contributed new reagents/analytic tools; C.Z., J.R., and H.K. analyzed data; and C.Z., J.R., J.L., and H.K. wrote the paper.

The authors declare no conflict of interest.

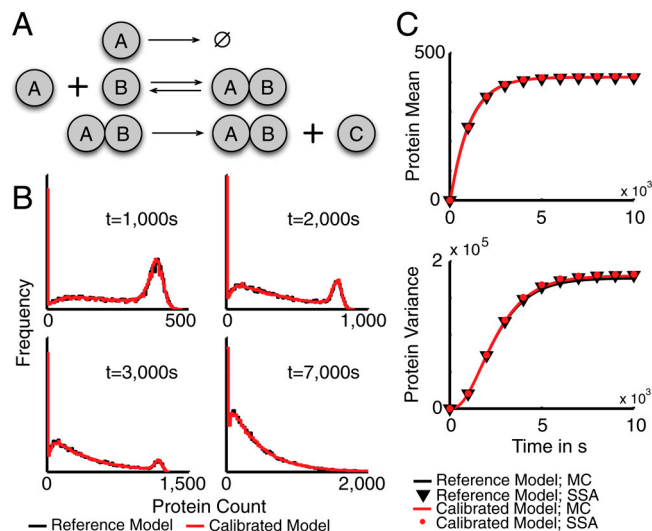
This article is a PNAS Direct Submission.

<sup>1</sup>C.Z. and J.R. contributed equally to this work.

<sup>2</sup>To whom correspondence should be addressed. E-mail: koepplh@ethz.ch.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1200161109/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1200161109/-DCSupplemental).





**Fig. 2.** Moment-based inference using synthetic data. (A) A simple model of transient gene activation: The binding of A to the target gene B aggregates all necessary steps involved in gene activation such as the binding of additional transcription factors, polymerase binding, or chromatin remodeling. Also protein synthesis is reduced to the simplest possible model—i.e., a first order production, abstracting messenger RNA (mRNA) transcription and degradation, translation, and protein folding. (B) The protein distributions predicted by the calibrated model (red) compared to the distributions generated from the reference model (black) at four representative time points (see *SI Appendix, section S.3.2* for further time points). Estimates of the distributions were obtained by stochastic simulation (20,000 runs). (C) The time evolutions of the approximate protein mean and variance obtained from moment closure (MC) differ only little from approximations computed by stochastic simulation (SSA). Therefore, the model parameters can be inferred up to negligibly small deviations.

The role of this kinase is twofold. In the cytoplasm, Hog1 phosphorylates its substrate to increase the internal concentration of glycerol in the cell. In parallel, a large fraction of the active Hog1 translocates to the nucleus where it triggers the activation of a transcriptional program leading to the upregulation of roughly 300 genes (27). Once the internal glycerol concentration allows to balance the external osmotic pressure, the HOG pathway is deactivated, leading to loss of active MAPK and a rapid termination of the transcriptional process.

To quantify the amount of transcription induced by this pathway a fluorescent expression reporter was generated using the promoter pSTL1 (promoter of the sugar transporter-like protein 1), a well-characterized Hog1 expression target driving the expression of a fluorescent protein construct (quadrupleVenus-qV). It was recently shown in ref. 21 that the transient activation of the MAPK Hog1 in conjunction with a slow step in the transcription activation process of the promoter results in a bimodality in the expression profiles of this fluorescent expression reporter.

Nuclear enrichment of Hog1 was measured by microscopy and the pSTL1-qV reporter abundance was quantified by flow cytometry at nine different time points for NaCl concentrations of 0 M, 0.1 M, 0.12 M, and 0.2 M.

**The Model.** Components involved in activation and translocation of Hog1 are present in high abundance (e.g., around 6800 Hog1 molecules per cell) (28). Consequently, intrinsic fluctuations of active Hog1 are relatively small. Experimental results in ref. 29 and our own data support this and also that Hog1 signaling is robust against cell-to-cell variations. Motivated by this, we assume Hog1 signaling to be deterministic rather than stochastic and that the mean dynamics reflect well the signaling behavior (30, 31). Continuous-time functions of nuclear Hog1 were ob-

tained from the experimental data by linear regression with radial basis functions (24) across different NaCl concentrations (see *SI Appendix, section S.4.2*).

Several transcription factors such as Sko1 or Hot1 are under control of active Hog1 once it translocates to the nucleus as shown in ref. 32. This and the experimentally observed switch-like induction of fluorescent reporter expression suggest a high cooperativity in the pSTL1 promoter dynamics. In a purely stochastic mass-action model, one way to model cooperativity is to require multiple Hog1 copies to bind to the promoter before messenger RNAs (mRNA) can be transcribed. However, the previous high copy-number considerations allow us to simplify this step into transforming the fitted Hog1 abundance curves using a Hill-function with tunable parameters (see *SI Appendix, section S.4.2*). The output of this function is then treated as a time-varying kinetic parameter modulating the gene activation rate. Efficient transcription of mRNA requires interaction of the active gene with chromatin remodeling complexes (generic remodeler denoted as CR) (21). Translation is modeled as a one-step linear production event, depending on the number of ribosomes. We assume that extrinsic variability enters the system in the chromatin remodeling (variability in the number of CR) (33) and the translation efficiency (variability in the number of ribosomes) (5). A graphical representation of the model is given in Fig. 3A.

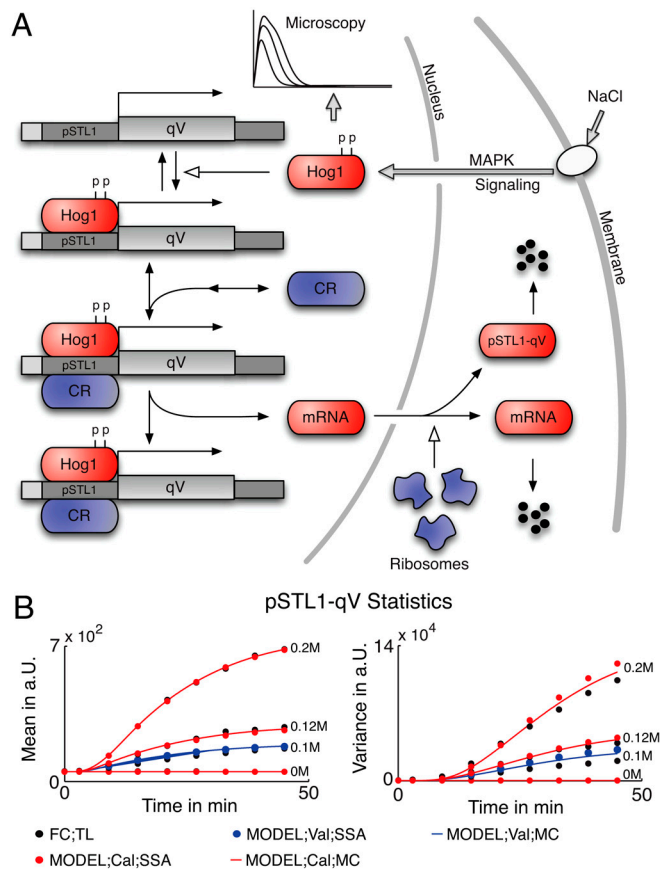
**pSTL1-qV Mean and Variance Predict Transient Bimodality.** The parameters of the model from Fig. 3A were inferred from the time courses of the experimental means and variances (see Fig. 3B) using NaCl concentrations 0 M, 0.12 M, and 0.2 M. We then validated the model by comparing the distributions predicted by the model for 0.1 M NaCl with the experimental results. The pSTL1-qV expression profiles for each measurement time point and NaCl concentration were computed from the calibrated model using stochastic simulation. A comparison between the experimental and the predicted distributions is shown in Fig. 4A. Even though only means and variances were used in the inference, the bimodal distributions are accurately predicted by the model (see also *SI Appendix, section S.4.5*).

We further validated the model using an additional snapshot dataset from ref. 21, where the pSTL1-qV reporter abundance was measured for several other NaCl concentrations between 0 M and 0.3 M, 45 min upon osmotic shock. From the model predictions and the measured distributions, we computed the coefficient of variation (CV) and a dose-response as functions of the NaCl concentration (Fig. 4B). The area around 0.1 M NaCl, where the CV is large and the dose-response curve is rising, indicates the NaCl concentration interval where the expression is in a bimodal regime. Note that also at 0.3 M NaCl, a concentration much larger than the concentrations that were used in the inference, the CV is predicted accurately.

To study the stochastic pSTL1-qV induction, we simulated the model to estimate the average number of cells that (i) never activate the pSTL1 promoter, (ii) activate the promoter at least once, and (iii) induce transcription. Our model predicts that for all NaCl concentrations except 0 M all cells manage to activate the promoter and, therefore, that the bimodality has to be caused by the subsequent—and comparably slow—chromatin remodeling step (see *SI Appendix, section S.4.7*). Further, we performed an *in silico* knock-down of CR by rescaling each cell's amount of CR by a hand-tuned factor, such that the percentage of responding cells saturated around 60% as measured in the experiment (see Fig. 4B). We found that the transition between the non- and all-responding domain is shifted to higher NaCl values and that the slope of the transition edge is decreased.

**In Silico Homogenization of the Cell Population.** After calibrating the model, we switched off extrinsic variability by setting each cell's extrinsic condition to the inferred mean value. We then recom-

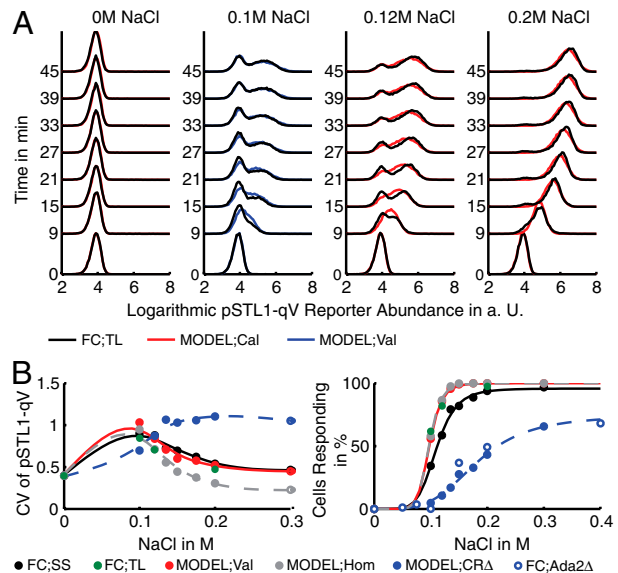




**Fig. 3.** MAPK Hog1 induced pSTL1-qV expression. (A) Osmotic pressure is sensed at the membrane, and results in the activation of the MAPK signaling cascade. Once active, double-phosphorylated MAPK Hog1 translocates to the nucleus, where it can bind via transcription factors to the pSTL1 promoter. Remodeling of the chromatin structure then allows for efficient transcription of the pSTL1-qV, which is exported from the nucleus to yield expression of the fluorescent reporter pSTL1-qV. Blue-shaded entities denote species subject to extrinsic variability. (B) Comparison of pSTL1-qV mean and variance obtained after calibration (Cal) and validation (Val) of the model using moment closure (MC) and 20,000 stochastic simulation runs (SSA) with the experimental estimates obtained from the time-lapsed (TL) flow cytometry (FC) data (approximately 20,000 cells).

puted estimates of the pSTL1-qV distributions using stochastic simulation. The resulting average cell can be interpreted as a homogenized version of the measured population. Again CV and dose response were computed and plotted in Fig. 4B. Interestingly, we find that extrinsic variability does not affect the dose-response behavior in pSTL1-qV induction. In contrast, the homogenized population shows significant differences in the CV. In particular, for larger stress levels the CV is relatively small compared to the heterogeneous counterpart, indicating less variability in pSTL1-qV reporter expression. For intermediate stress levels the homogenized population still shows a bimodal response.

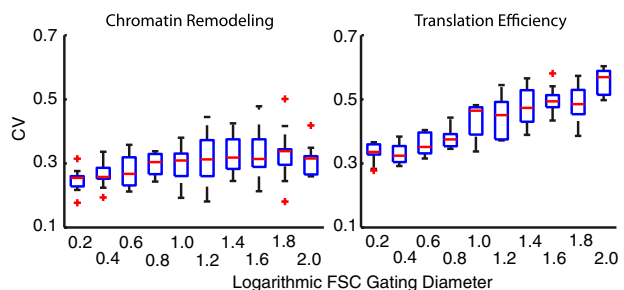
**Cell-Gating Eliminates Only a Fraction of Extrinsic Variability.** To study the extent to which extrinsic variability can be reduced by cell-gating, we reestimated the extrinsic statistics using the time-lapsed flow cytometry dataset for gates of different size, applied on the forward scatter channel (FSC, often used as a proxy for cell volume). We found that the variability in the translation efficiency is significantly reduced for small gating diameters. In contrast, no significant trend was found in the estimated variability in CR (see Fig. 5).



**Fig. 4.** Population mean and variance predict bimodal pSTL1-qV response. (A) Model calibration (Cal) and validation (Val). The parameters were inferred using time-lapsed (TL) flow cytometry (FC) data of the pSTL1-qV reporter at three NaCl concentrations (i.e., 0 M, 0.12 M, and 0.2 M) and used to predict the expression profiles at a NaCl concentration of 0.1 M. (B) Model validation using flow cytometry snapshot data (SS), recorded 45 min after osmotic shock for 0 M, 0.1 M, 0.12 M, 0.135 M, 0.15 M, 0.175 M, 0.2 M, and 0.3 M of NaCl. Left: Coefficient of variation (CV) of pSTL1-qV intensity as a function of NaCl. Right: Dose-response comparison. All curves indicate a Hill-type relation. The calibrated model was homogenized (Hom), giving rise to the average cell's CV and dose-response curves. Additionally, we studied the suppression of the chromatin remodeling *in silico* by reducing the amount of CR, such that the percentage of responding cells saturated around 60% (CRΔ) and compared the model predictions to the results reported in ref. 21, where the authors performed a knock-down of the transcription adapter 2 (Ada2) to demonstrate the impact of chromatin remodeling in pSTL1-qV induction (Ada2Δ).

## Discussion

Studying biological systems with mathematical models requires knowledge of the kinetic rate parameters of the system reactions. These parameters are often hard to measure experimentally and have to be inferred from the measurements that are available. In the simple example of Fig. 24, measurements of the mean dynamics alone did not provide enough information to uniquely identify the parameters (see *SI Appendix, section S.3.3*). This



**Fig. 5.** Cell-gating eliminates only a fraction of extrinsic variability. To study the influence of cell-gating on cell-to-cell variability we fixed all parameters but the extrinsic statistics  $\bar{\alpha}$  to their previously inferred values. The extrinsic statistics were reestimated from the time-lapsed flow cytometry data with increasing FSC gating diameters using MCMC sampling with 100 randomly drawn initial conditions. Statistics were computed and visualized over the 10 runs that achieved the highest posterior values using boxplots with MATLAB's (2009b, Mathworks) default settings. The CV of the translation efficiency increases with the logarithmic gating diameter (right), whereas the CV of the chromatin remodeling remains more or less constant (left).

demonstrates that averaged population data may contain too little information to identify the reaction rates. Contrary to that, additionally measuring the variance in the example of Fig. 2A allows one to uniquely identify all the parameters, even though the measured distributions are bimodal. This implies that the question of whether the measured distributions are well-characterized by low-order moments only is not necessarily of importance. In ref. 2 the authors presented a method that makes use of the information provided by the whole distributions. However, for larger systems, approximation of the probability distribution becomes computationally cumbersome. Focusing the analysis on lower-order moments, as proposed in this paper, means discarding a part of the information but makes the parameter identification feasible for larger systems.

The moment-based inference scheme allowed us to estimate the parameters of a stochastic model of the osmo-stress induced transcriptional activation in budding yeast using distribution measurements of a heterogeneous cell population and thereby enabled us to explain and predict experimental data and to provide computational support for existing biological hypotheses. The inferred model characteristics and parameters agree well with state-of-the-art literature. For instance the predictions, obtained for the *in silico* knock-down of CR, agree well with results in ref. 21, where the authors performed knock-down experiments for different components of the SAGA complex, which is recruited during chromatin remodeling (see Fig. 4B). The pSTL1-qV half-life was estimated to be around 90 min; this value agrees well with ref. 34, where the authors report a high stability of similar fluorescent reporters. The Hill coefficient of the pSTL1 promoter dynamics was estimated to be  $n_H \approx 6$ , indicating high cooperativity in the binding of active Hog1 to the target gene. This seems to be crucial for the cell to achieve the strong switch-like behavior observed in the experimental data and agrees well with previously reported results, where Hog1 dependent transcription factors were shown to have multiple binding sites (35).

The inferred model parameters predict large cell-to-cell variations in the chromatin remodeling as well as in the translation efficiency (CVs around 0.3–0.4). This is in agreement with the experimental results from ref. 33, where the authors found that several chromatin remodeling factors show large variations—e.g., a CV around 0.3 for the transcription adapter 2 (Ada2).

According to our model, variability in the chromatin remodeling is widely independent of the FSC gating radius, indicating that extrinsic noise is suppressed only in the translation efficiency by applying gates on morphological features. In conjunction with the observation that an *in silico* homogenization of the cell population leads to different CVs of the pSTL1-qV distributions (Fig. 4B), this suggests that studies that solely rely on cell gating to eliminate cell-to-cell variability may lead to biased results. Explicitly including extrinsic variability may resolve a systematic mismatch and, additionally, allows to quantify effects of variability on the system. For instance the *in silico* homogenization of the cell population (Fig. 4B) indicates that the dose response is widely insensitive against cell-to-cell variability. This provides computational support for the hypothesis in ref. 21 that the partitioning between reporter inducing and noninducing cells is primarily caused by intrinsic stochasticity.

## Materials and Methods

**Flow Cytometry Measurements.** The yeast cells bearing the integrated pSTL1-quadrupleVenus (21) were cultivated in synthetic (SD) medium (using a yeast nitrogen base w/o folic acid, w/o riboflavin). An overnight saturated culture was diluted and grown in log phase for 24 h (OD<sub>600</sub> nm kept below 0.2 by several dilutions). Hog1 driven gene expression was induced by adding 5 mL (3x) salt solution (SD medium + NaCl) to 10 mL culture containing flasks. At each time point a 0.5 mL aliquot was taken and protein translation was stopped by adding cycloheximide (final CHX concentration: 100 µg/mL). After maturation of the fluorescent reporters (about 2 h), 350 µL cells were added to 400 µL PBS, briefly sonicated and filtered. Finally the fluorescence

was measured using a BD LSR II (excitation: 488 nm, emission: 525/50 nm). If not explicitly stated, we applied cell-gates of log-diameter 0.5 with respect to the forward—and log—diameter 2.5 to the side scatter channel, respectively. Each flow cytometry distribution was obtained from around 55,000 cells, leading to an effective number of cells around 20,000 after applying the gate.

**Moment Closure.** Equations that describe the time evolution of all the moments of the distribution can be derived from the CME (36). Extracting the equations for the moments up to some order  $n$  leads to a finite open system that possibly depends on higher order moments. Approximating the higher order moments by some nonlinear function  $f$  of the lower-order moments (see ref. 19 and *SI Appendix, sections S.3.1 and S.4.1*) leads to a closed system of the form

$$\frac{d}{dt}\tilde{\mu} = A(\theta)\tilde{\mu} + B(\theta)f(\tilde{\mu}),$$

where  $\tilde{\mu}$  is a vector containing the moments of order up to  $n$  and  $A(\theta)$  and  $B(\theta)$  are determined by the model structure.

**Moment Uncertainties and Data Modeling.** Asymptotically unbiased estimates for central moments of order  $k$  at time  $t_i$  were computed from  $M$  samples as

$$\hat{\mu}_k(t_i) = \begin{cases} \frac{1}{M} \sum_{i=1}^M x_i(t_i) & k = 1 \\ \frac{1}{M} \sum_{i=1}^M (x_i(t_i) - \hat{\mu}_1(t_i))^k & k > 1. \end{cases}$$

The central limit theorem implies that for large  $M$  (i.e., around 20,000 within our experiments) the moment estimates are approximately normally distributed—i.e.,  $\hat{\mu}_k(t_i) \sim \mathcal{N}(\mu_k(t_i), \sigma_k^2(t_i))$ —with  $\mu_k(t_i)$  as the true  $k$ -th moment. We further validated this assumption for both case studies by comparing bootstrapped distributions of the empirical moments to normal distributions using probability-probability (P-P) and quantile-quantile (Q-Q) plots. Additionally, we used a Kolmogorov–Smirnov–Lilliefors test to assess normality (see *SI Appendix, sections S.3.2 and S.4.5*). For  $k = 1$  and  $k = 2$ , the estimators variance can be estimated as

$$\sigma_1^2(t_i) = \frac{1}{M} \hat{\mu}_2^2(t_i) \quad \text{and} \quad \sigma_2^2(t_i) = \frac{1}{M} \left( \hat{\mu}_4(t_i) - \frac{M-3}{M-1} \hat{\mu}_2^2(t_i) \right),$$

respectively.

**Modeling Fluorescence Intensities.** We assumed that the measured fluorescence intensity for a given cell is proportional to the number of fluorescent proteins (33)—i.e.,  $I_{\text{tot}}(t_i) = \epsilon I_{\text{tot}}(t_i)$  with scaling parameter  $\epsilon$ . Due to the nonidentifiability in the translation step (see *SI Appendix, section S.4.5*), only the product of the translation rate and  $\epsilon$  can be determined. Furthermore, we assumed that the reporter abundance  $I_R(t_i)$  is corrupted by autofluorescence and measurement artifacts, modeled as an additive random variable  $I_{\text{AF}}(t_i)$ , independent of the reporter abundance—i.e.,  $I_{\text{tot}}(t_i) = I_R(t_i) + I_{\text{AF}}(t_i)$ . Mean and variance of  $I_{\text{AF}}(t_i)$  were estimated from the flow cytometry data for 0 M NaCl, collected over the measurement time points. As this allows very accurate estimates ( $M$  in the order of hundreds of thousands), the uncertainty of those estimates can be well neglected. The experimental means and variances of pSTL1-qV abundance at a given measurement time point were calculated as  $\hat{\mu}_R^k(t_i) = \hat{\mu}_{\text{tot}}^k(t_i) - \mu_{\text{AF}}^k(t_i)$  for  $k \in \{1, 2\}$ . Note that moment-based inference and analysis of the model can be carried out without any assumptions on the autofluorescence distribution. In order to compare protein distributions from the model with experimentally obtained distributions, we sampled autofluorescence values from the measured flow cytometry distribution for 0 M NaCl.

**Model Calibration.** For all experiments, we assumed flat prior distributions over parameters  $\gamma_j \in \gamma$  (with zero probability for negative values). In the M-H MCMC scheme, for each of the  $J$  parameters in  $\gamma$ , we used independent log-normal proposal distributions such that  $q(\gamma^{\text{new}} | \gamma^{\text{old}}) = \prod_{j=1}^J q(\gamma_j^{\text{new}} | \gamma_j^{\text{old}})$  with  $q(\gamma_j^{\text{new}} | \gamma_j^{\text{old}}) = \mathcal{LN}(\ln \gamma_j^{\text{old}}, v_j^2)$ . A detailed configuration can be found in *SI Appendix, sections S.3.2 and S.4.5*. Proposed parameter samples are accepted with probability

$$a = \min \left\{ 1, \frac{p(\gamma^{\text{new}} | \hat{\mu}_1, \dots, \hat{\mu}_n) q(\gamma^{\text{old}} | \gamma^{\text{new}})}{p(\gamma^{\text{old}} | \hat{\mu}_1, \dots, \hat{\mu}_n) q(\gamma^{\text{new}} | \gamma^{\text{old}})} \right\}.$$

From the resulting Markov chain we extracted the parameter configuration which maximized the posterior density. The distance between predicted and measured protein distributions was quantified using the Kolmogorov metric (see *SI Appendix, section S.4.5*).

**ACKNOWLEDGMENTS.** We are grateful to J. Paulsson, A. Hilfinger, and J. Hasenauer for interesting discussions. C.Z. and H.K. acknowledge the sup-

port from the Swiss National Science Foundation, Grant PP00P2 128503. The work of J.R. and J.L. was supported in part by SystemsX.ch under the project YeastX and by the European Commission under the project MoVeS. S.P. and M.P. acknowledge the support from the European project UNICELLSYS, the European Research Council, the SystemsX.ch organization (LiverX), the Competence Centre for Systems Physiology and Metabolic Disease, the Swiss National Science Foundation, and the ETH Zurich.

- Hengl S, Kreutz C, Timmer J, Maiwald T (2007) Data-based identifiability analysis of non-linear dynamical models. *Bioinformatics* 23:2612–2618.
- Munsky B, Trinh B, Khammash M (2009) Listening to the noise: Random fluctuations reveal gene network parameters. *Mol Syst Biol* 5:318.
- Boys RJ, Wilkinson DJ, Kirkwood TBL (2007) Bayesian inference for a discretely observed stochastic kinetic model. *Stat Comput* 18:125–135.
- Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. *Science* 297:1183–1186.
- Colman-Lerner A, et al. (2005) Regulated cell-to-cell variation in a cell-fate decision system. *Nature* 437:699–706.
- Snijder B, Pelkmans L (2011) Origins of regulated cell-to-cell variability. *Nat Rev Mol Cell Biol* 12:119–125.
- Shahrezaei V, Ollivier JF, Swain PS (2008) Colored extrinsic fluctuations and stochastic gene expression. *Mol Syst Biol* 4:196.
- Hilfinger A, Paulsson J (2011) Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *Proc Natl Acad Sci USA* 108:12167–12172.
- Hasenauer J, et al. (2011) Identification of models of heterogeneous cell populations from population snapshot data. *BMC Bioinformatics* 12:125.
- Koepl H, Zechner C, Ganguly A, Pelet S, Peter M (2012) Accounting for extrinsic variability in the estimation of stochastic rate constants. *Int J Robust Nonlin*, 10.1002/rnc.2804.
- Golightly A, Wilkinson DJ (2005) Bayesian inference for stochastic kinetic models using a diffusion approximation. *Biometrics* 61:781–788.
- Mettetal JT, Muzzey D, Pedraza JM, Ozbudak EM, van Oudenaarden A (2006) Predicting stochastic gene expression dynamics in single cells. *Proc Natl Acad Sci USA* 103:7304–7309.
- Guido NJ, et al. (2006) A bottom-up approach to gene regulation. *Nature* 439:856–860.
- Munsky B, Khammash M (2006) The finite state projection algorithm for the solution of the chemical master equation. *J Chem Phys* 124:044104.
- Wolf V, Goel R, Mateescu M, Henzinger TA (2010) Solving the chemical master equation using sliding windows. *BMC Syst Biol* 4:42.
- Gillespie D (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J Comput Phys* 22:403–434.
- Fishman GS (1996) Monte Carlo: Concepts, algorithms, and applications. (Springer, Berlin).
- Whittle P (1957) On the use of the normal approximation in the treatment of stochastic processes. *J Roy Stat Soc B Met* 19:268–281.
- Hespanha J (2008) Moment closure for biochemical networks. *Proceedings of the 3rd International Symposium on Communications, Control and Signal Processing (IEEE)* pp 142–147.
- Ruess J, Miliias-Argeitis A, Summers S, Lygeros J (2011) Moment estimation for chemically reacting systems by extended Kalman filtering. *J Chem Phys* 135:165102.
- Pelet S, et al. (2011) Transient activation of the HOG MAPK pathway regulates bimodal gene expression. *Science* 332:732–735.
- Shalem O (2008) Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. *Mol Syst Biol* 4:4.
- Hohmann S (2002) Osmotic stress signaling and osmoadaptation in yeasts. *Microbiol Mol Biol Rev* 66:300–372.
- Bishop CM (2007) Pattern Recognition and Machine Learning. New York (Springer).
- Bellman R, Astrom K (1970) On structural identifiability. *Math Biosci* 7:329–339.
- Gilks WR, Richardson S, Spiegelhalter DJ (1996) Markov chain Monte Carlo in practice. (Chapman & Hall, New York).
- Gasch AP, et al. (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell* 11:4241–4257.
- Ghaemmaghami S, et al. (2003) Global analysis of protein expression in yeast. *Nature* 425:737–741.
- Muzzey D, Gómez-Urbe Ca, Mettetal JT, van Oudenaarden A (2009) A systems-level analysis of perfect adaptation in yeast osmoregulation. *Cell* 138:160–171.
- Zi Z, Liebermeister W, Klipp E (2010) A quantitative study of the Hog1 MAPK response to fluctuating osmotic stress in *Saccharomyces cerevisiae*. *PLoS One* 5:e9522.
- Macia J, et al. (2009) Dynamic signaling in the Hog1 MAPK pathway relies on high basal signal transduction. *Sci Signal* 2:ra13.
- Capaldi AP, et al. (2008) Structure and function of a transcriptional network activated by the MAPK Hog1. *Nat Genet* 40:1300–1306.
- Newman JRS, et al. (2006) Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* 441:840–846.
- Mateus C, Avery SV (2000) Destabilized green fluorescent protein for monitoring dynamic changes in yeast gene expression with flow cytometry. *Yeast* 16:1313–1323.
- Proft M, Gibbons F, Copeland M, Roth F, Struhl K (2005) Genomewide identification of Sko1 target promoters reveals a regulatory network that operates in response to osmotic stress in *Saccharomyces cerevisiae*. *Eukaryot Cell* 4:1343–1352.
- Hespanha J (2006) Modeling and analysis of stochastic hybrid systems. *IEEE Proceedings Control Theory And Applications* 153:520–535.

# Moment-Based Inference Predicts Bimodality in Transient Gene Expression

C. Zechner, J. Ruess, P. Krenn, S. Pelet, M. Peter, J. Lygeros, H. Koepl

## S.1 Population Dynamics and Extrinsic Variability

Consider a chemical reaction system of  $m$  different species in a constant random environment, which is given by a  $\mathcal{D}$ -valued random variable  $Z$  that is distributed according to the probability measure  $P_Z$ , where  $\mathcal{D}$  is some set. Denote by  $X(t) = (X_1(t), \dots, X_m(t))$  the stochastic process that describes the time evolution of the number of molecules of the  $m$  species and by  $p(x, t|z)$  the conditional probability  $P(X(t) = x|Z = z)$ , where  $x = (x_1, \dots, x_m) \in \mathcal{X}$  and  $\mathcal{X} \subset \mathbb{N}^m$  is the set of all reachable states. The time evolution of the conditional probability distribution of this system then follows a conditional chemical master equation:

$$\frac{d}{dt}p(x, t|z) = \sum_{k=1}^K -p(x, t|z)a_k(x, z) + p(x - \nu_k, t|z)a_k(x - \nu_k, z),$$

where  $a_k(x, z)$ ,  $k = 1, \dots, K$ , are the propensity functions and  $\nu_k$ ,  $k = 1, \dots, K$ , the stoichiometric transition vectors of the  $K$  reactions of the system.

Multiplying both sides of the conditional CME by the  $i$ -th component  $x_i$  of the state vector  $x$ , summing over all states  $x \in \mathcal{X}$  and integrating over all possible values of  $Z$  with respect to the probability measure  $P_Z$  yields:

$$\int_{\mathcal{D}} \sum_{x \in \mathcal{X}} \frac{d}{dt}p(x, t|z)x_i dP_Z = \int_{\mathcal{D}} \sum_{x \in \mathcal{X}} \sum_{k=1}^K [-p(x, t|z)a_k(x, z)x_i + p(x - \nu_k, t|z)a_k(x - \nu_k, z)x_i] dP_Z$$

The left hand side is then the time derivative of the marginal mean of  $X_i$ , whereas the right hand side can be simplified by a change of variables as follows:

$$\begin{aligned} \frac{d}{dt}\mathbb{E}[X_i(t)] &= \int_{\mathcal{D}} \sum_{k=1}^K \sum_{x \in \mathcal{X}} [-p(x, t|z)a_k(x, z)x_i + p(x, t|z)a_k(x, z)(x_i + \nu_{k_i})] dP_Z \\ &= \sum_{k=1}^K \int_{\mathcal{D}} \sum_{x \in \mathcal{X}} p(x, t|z)a_k(x, z)(-x_i + x_i + \nu_{k_i}) dP_Z \\ &= \sum_{k=1}^K \nu_{k_i} \int_{\mathcal{D}} \sum_{x \in \mathcal{X}} p(x, t|z)a_k(x, z) dP_Z \\ &= \sum_{k=1}^K \nu_{k_i} \int_{\mathcal{D}} \mathbb{E}[a_k(X(t), z)|z] dP_Z \\ &= \sum_{k=1}^K \nu_{k_i} \mathbb{E}[a_k(X(t), Z)], \end{aligned}$$

where  $\nu_{k_i}$  is the  $i$ -th component of  $\nu_k$ .

Multiplying both sides with  $x_i x_j$  instead of  $x_i$  and performing the same calculations yields the time derivatives of the second order moments:

$$\begin{aligned}
\frac{d}{dt} \mathbb{E}[X_i(t) X_j(t)] &= \int_{\mathcal{D}} \sum_{k=1}^K \sum_{x \in \mathcal{X}} [-p(x, t|z) a_k(x, z) x_i x_j + p(x, t|z) a_k(x, z) (x_i + \nu_{k_i})(x_j + \nu_{k_j})] dP_Z \\
&= \sum_{k=1}^K \int_{\mathcal{D}} \sum_{x \in \mathcal{X}} p(x, t|z) a_k(x, z) (-x_i x_j + x_i x_j + x_i \nu_{k_j} + x_j \nu_{k_i} + \nu_{k_i} \nu_{k_j}) dP_Z \\
&= \sum_{k=1}^K (\nu_{k_i} \mathbb{E}[X_j(t) a_k(X(t), Z)] + \nu_{k_j} \mathbb{E}[X_i(t) a_k(X(t), Z)] + \nu_{k_i} \nu_{k_j} \mathbb{E}[a_k(X(t), Z)])
\end{aligned}$$

Through the terms  $\mathbb{E}[a_k(X(t), Z)]$  and  $\mathbb{E}[X_i(t) a_k(X(t), Z)]$  cross moments of the species and the extrinsic variable influence the time evolution of the species moments. Even if the extrinsic variable  $Z$  is independent of the species initial conditions, over time it will become correlated with  $X(t)$ . This requires computing the time evolution of these correlations along with the species moments.

The time derivatives of the cross moments  $\mathbb{E}[X_i(t) Z]$  can be obtained with the same calculations as above by multiplying both sides by  $x_i z$ :

$$\begin{aligned}
\frac{d}{dt} \mathbb{E}[X_i(t) Z] &= \int_{\mathcal{D}} \sum_{k=1}^K \sum_{x \in \mathcal{X}} [-p(x, t|z) a_k(x, z) x_i z + p(x, t|z) a_k(x, z) (x_i + \nu_{k_i}) z] dP_Z \\
&= \sum_{k=1}^K \int_{\mathcal{D}} \sum_{x \in \mathcal{X}} p(x, t|z) a_k(x, z) (-x_i z + x_i z + \nu_{k_i} z) dP_Z \\
&= \sum_{k=1}^K \nu_{k_i} \mathbb{E}[Z a_k(X(t), Z)]
\end{aligned}$$

If we close the moment equations and replace all the higher order moments of the species and all the higher order cross moments of the species and the extrinsic variable by functions of the lower order moments, we obtain a dynamical system for the moments and cross moments of order up to two that depends only on some lower order moments of the distribution of the extrinsic variable. If this distribution is unknown, these moments can be included as unknown parameters in the parameter search. The specific moment equations and closure functions for the examples considered in this work are given in Section S.3.1 and Section S.4.1.

## S.2 Separating extrinsic from intrinsic noise in a birth-death process with variable birth rate

To demonstrate that, in principle, the extrinsic statistics can be identified from measurements of species moments we consider a simple birth-death process, initialized at zero, with death rate  $b$ , where the birth rate  $Z$  is a random variable distributed according to some probability measure  $P_Z$ .

$$\emptyset \xrightarrow{Z} X \xrightarrow{b} \emptyset$$



Using the derivation of the previous section the time evolution of the moments of this system can be computed as

$$\begin{aligned}\frac{d}{dt}\mathbb{E}[X] &= \mathbb{E}[Z] - b\mathbb{E}[X], \\ \frac{d}{dt}\mathbb{E}[X^2] &= \mathbb{E}[Z] + 2\mathbb{E}[XZ] - 2b\mathbb{E}[X^2] + b\mathbb{E}[X], \\ \frac{d}{dt}\mathbb{E}[XZ] &= \mathbb{E}[Z^2] - b\mathbb{E}[XZ].\end{aligned}$$

This is a closed linear system and can be solved analytically. Denote by  $\mathbb{E}[X]^{tr}$  and  $\mathbb{E}[X]^{ss}$  two observations of the mean of  $X$ , taken at a transient time point  $t_{tr}$  and at stationarity, respectively. These two observations allow to compute the death rate and the mean of  $P_Z$  as

$$b = -\frac{1}{t_{tr}} \log \left( \frac{\mathbb{E}[X]^{ss} - \mathbb{E}[X]^{tr}}{\mathbb{E}[X]^{ss}} \right) \quad \text{and} \quad \mathbb{E}[Z] = b\mathbb{E}[X]^{ss}.$$

If also an observation  $\mathbb{E}[X^2]^{ss}$  of the stationary second moment is available, the second moment of  $P_Z$  can be computed as

$$\mathbb{E}[Z^2] = b^2 \left( \mathbb{E}[X^2]^{ss} - \mathbb{E}[X]^{ss} \right)$$

Therefore, for this process, mean and variance of the extrinsic variable can be computed from measurements of the mean at one transient time point and mean and second moment at stationarity.

### S.3 A Simple Model of Transient Gene Activation

The simple mass-action model under consideration is defined by the four reactions



and the corresponding stochastic rate constants given in Tab. S.1.

#### S.3.1 Moment Dynamics

Denote by  $\mu_A^1, \mu_B^1, \mu_D^1$  and  $\mu_C^1$  the means of  $A, B, AB$  and  $C$ , respectively. Let  $\mu_A^2, \mu_{AB}^2, \mu_{AD}^2, \mu_{AC}^2, \mu_B^2, \mu_{BD}^2, \mu_{BC}^2, \mu_D^2, \mu_{DC}^2, \mu_C^2$  and  $\mu_{ABD}^3, \mu_{A2B}^3, \mu_{ABC}^3, \mu_{AB2}^3$  be the second and third order moments, respectively. Since we assumed that this system is unaffected by extrinsic variability, the moment dynamics are not affected by moments of an extrinsic variable. Several tools are available, e.g. [7], that are able to automatically compute moment dynamics and closure functions from the reaction system and allow fast implementation of our identification method. For the system in question the moment equations become

$$\begin{aligned}
\frac{d}{dt}\mu_A^1 &= c_3 \cdot \mu_D^1 - c_2 \cdot \mu_{AB}^2 - c_1 \cdot \mu_A^1 \\
\frac{d}{dt}\mu_B^1 &= c_3 \cdot \mu_D^1 - c_2 \cdot \mu_{AB}^2 \\
\frac{d}{dt}\mu_D^1 &= c_2 \cdot \mu_{AB}^2 - c_3 \cdot \mu_D^1 \\
\frac{d}{dt}\mu_C^1 &= c_4 \cdot \mu_D^1 \\
\frac{d}{dt}\mu_A^2 &= c_1 \cdot \mu_A^1 - 2 \cdot c_1 \cdot \mu_A^2 + c_2 \cdot \mu_{AB}^2 - 2 \cdot c_2 \cdot \mu_{A2B}^3 + c_3 \cdot \mu_D^1 + 2 \cdot c_3 \cdot \mu_{AD}^2 \\
\frac{d}{dt}\mu_{AB}^2 &= c_3 \cdot \mu_D^1 - c_2 \cdot \mu_{AB2}^3 - c_2 \cdot \mu_{A2B}^3 + c_3 \cdot \mu_{AD}^2 + c_3 \cdot \mu_{BD}^2 - (c_1 - c_2) \cdot \mu_{AB}^2 \\
\frac{d}{dt}\mu_{AD}^2 &= c_2 \cdot \mu_{A2B}^3 - c_2 \cdot \mu_{AB}^2 - c_2 \cdot \mu_{ABD}^3 - c_3 \cdot \mu_D^1 + c_3 \cdot \mu_D^2 - (c_1 + c_3) \cdot \mu_{AD}^2 \\
\frac{d}{dt}\mu_{AC}^2 &= c_3 \cdot \mu_{DC}^2 - c_2 \cdot \mu_{ABC}^3 - c_1 \cdot \mu_{AC}^2 + c_4 \cdot \mu_{AD}^2 \\
\frac{d}{dt}\mu_B^2 &= c_2 \cdot \mu_{AB}^2 - 2 \cdot c_2 \cdot \mu_{AB2}^3 + c_3 \cdot \mu_D^1 + 2 \cdot c_3 \cdot \mu_{BD}^2 \\
\frac{d}{dt}\mu_{BD}^2 &= c_2 \cdot \mu_{AB2}^3 - c_2 \cdot \mu_{AB}^2 - c_2 \cdot \mu_{ABD}^3 - c_3 \cdot \mu_D^1 + c_3 \cdot \mu_D^2 - c_3 \cdot \mu_{BD}^2 \\
\frac{d}{dt}\mu_{BC}^2 &= c_3 \cdot \mu_{DC}^2 - c_2 \cdot \mu_{ABC}^3 + c_4 \cdot \mu_{BD}^2 \\
\frac{d}{dt}\mu_D^2 &= c_2 \cdot \mu_{AB}^2 + 2 \cdot c_2 \cdot \mu_{ABD}^3 + c_3 \cdot \mu_D^1 - 2 \cdot c_3 \cdot \mu_D^2 \\
\frac{d}{dt}\mu_{DC}^2 &= c_2 \cdot \mu_{ABC}^3 - c_3 \cdot \mu_{DC}^2 + c_4 \cdot \mu_D^2 \\
\frac{d}{dt}\mu_C^2 &= c_4 \cdot \mu_D^1 + 2 \cdot c_4 \cdot \mu_{DC}^2
\end{aligned}$$

A closed system is obtained by replacing the third order cumulants by zero, which is equivalent to replacing the third order moments by functions of the lower order moments as follows:

$$\begin{aligned}
\mu_{ABD}^3 &= \mu_A^1 \cdot \mu_{BD}^2 + \mu_B^1 \cdot \mu_{AD}^2 + \mu_D^1 \cdot \mu_{AB}^2 - 2 \cdot \mu_A^1 \cdot \mu_B^1 \cdot \mu_D^1, \\
\mu_{A2B}^3 &= -2 \cdot \mu_B^1 \cdot \mu_A^{1^2} + 2 \cdot \mu_{AB}^2 \cdot \mu_A^1 + \mu_B^1 \cdot \mu_A^2, \\
\mu_{ABC}^3 &= \mu_A^1 \cdot \mu_{BC}^2 + \mu_B^1 \cdot \mu_{AC}^2 + \mu_C^1 \cdot \mu_{AB}^2 - 2 \cdot \mu_A^1 \cdot \mu_B^1 \cdot \mu_C^1, \\
\mu_{AB2}^3 &= -2 \cdot \mu_A^1 \cdot \mu_B^{1^2} + 2 \cdot \mu_{AB}^2 \cdot \mu_B^1 + \mu_A^1 \cdot \mu_B^2
\end{aligned}$$

Note that the accuracy of moment closure methods depends not only on the network structure, but also on the unknown parameters. Therefore, it is a priori unclear if a certain closure method leads to accurate approximations of the moments. However, for a given set of parameters accuracy of a moment closure method can be checked by comparing the approximations with estimates computed from a large number of stochastic simulations. Therefore we arbitrarily picked a moment closure method, performed the parameter search with it and checked the accuracy of the approximations locally for the found parameters. Second order zero cumulant moment closure led to good agreement of the approximations with estimates obtained from stochastically simulating the system with the estimated parameters and was therefore sufficient for our model.

### S.3.2 Model Calibration

We first validated the normality assumption of the empirical moment estimates as described in the Materials and Methods section in the main text. As graphical method we applied standard probability-probability ( $P - P$ ) and quantile-quantile ( $Q - Q$ ) plots, whereas to quantitatively assess normality we performed the Kolmogorov-Smirnov-Lilliefors test [4]. Results and  $p$ -values for the tests are demonstrated for the protein distribution at 10000s in Fig. S.1.

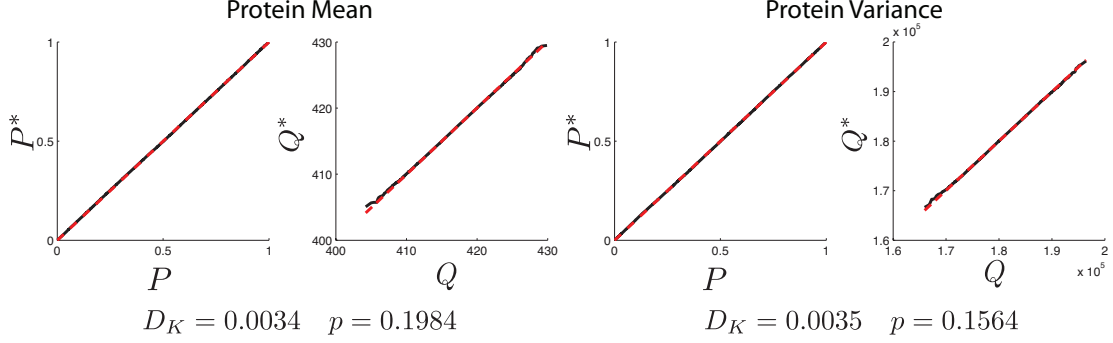


Figure S.1: Normality of empirical moments. Distributions of empirical moments were computed using bootstrapping. In particular, we randomly subsampled 10000 abundance values from the simulated target distribution and computed the sample mean and variance. This was repeated 50000 times to obtain approximations of the distributions of the moment estimates. For comparison, analytical normal distributions were fitted to the bootstrapped data and compared against each other using standard probability-probability ( $P - P$ ) and quantile-quantile ( $Q - Q$ ) plots [4]. In all four plots, the x-axes correspond to the analytical normal distributions, whereas the y-axes correspond to the bootstrapped empirical distributions. The low Kolmogorov-Smirnov test statistics  $D_K$  in conjunction with the large  $p$ -values indicate that the distribution of the empirical moment estimates can be well approximated by a normal distribution.

In order to calibrate the model to the reference data, we estimated the kinetic parameters from Tab. S.1. For simplicity, we assumed that the initial conditions of the species are known. If in some application the initial conditions are unknown, they can be included as unknown parameters in the parameter search and estimated along with the other parameters. For species  $A$  we set the initial amount to 50 molecules,  $B$  was initialized at 1 (corresponding to the gene being initially in the inactive state), whereas all other species were initialized at 0 molecules. The reference data was generated by stochastic simulation using  $M = 20000$  sample paths of length  $T = 10000s$ . We computed first and second order moments of species  $C$  and their corresponding uncertainties (see main text, Materials and Methods) each 2000s.

The obtained data was treated as measurements and used to calibrate the model by computing MAP estimates using a standard Metropolis-Hastings (M-H) sampler (see main text, Materials and Methods). The scaling parameters of the log-normal proposal densities were set to  $v_j = 0.01$ . After burn-in we recorded around 10000 samples of the M-H algorithm and determined MAP values. This inference was performed multiple times using random initial parameter values, drawn from a log-normal distribution  $\mathcal{LN}(\ln 0.002, 2^2)$ . Each time, the inference scheme ended up with equivalent MAP estimates (up to small random deviations, introduced by the randomized parameter search).

The inferred parameter set used for further analysis is given in Tab. S.1. The calibrated model was validated by comparing the distributions of species  $C$  to the distributions obtained from the reference model at the measurement time points (for better accuracy the comparison was based on 50000 sample paths for both models), as shown in Fig. S.2.

Table S.1: Inferred model parameters and Metropolis-Hastings setup. The reference and inferred MAP estimates are denoted as  $\gamma_j$  and  $\gamma_{j,MAP}$ , respectively.

Parameter	$c_1$	$c_2$	$c_3$	$c_4$
$\gamma_j$	$1.500 \cdot 10^{-2}$	$8.000 \cdot 10^{-4}$	$1.000 \cdot 10^{-3}$	$4.000 \cdot 10^{-1}$
$\gamma_{j,MAP}$	$1.380 \cdot 10^{-2}$	$7.050 \cdot 10^{-4}$	$9.865 \cdot 10^{-4}$	$3.988 \cdot 10^{-1}$
Unit	$s^{-1}$	$s^{-1}$	$s^{-1}$	$s^{-1}$

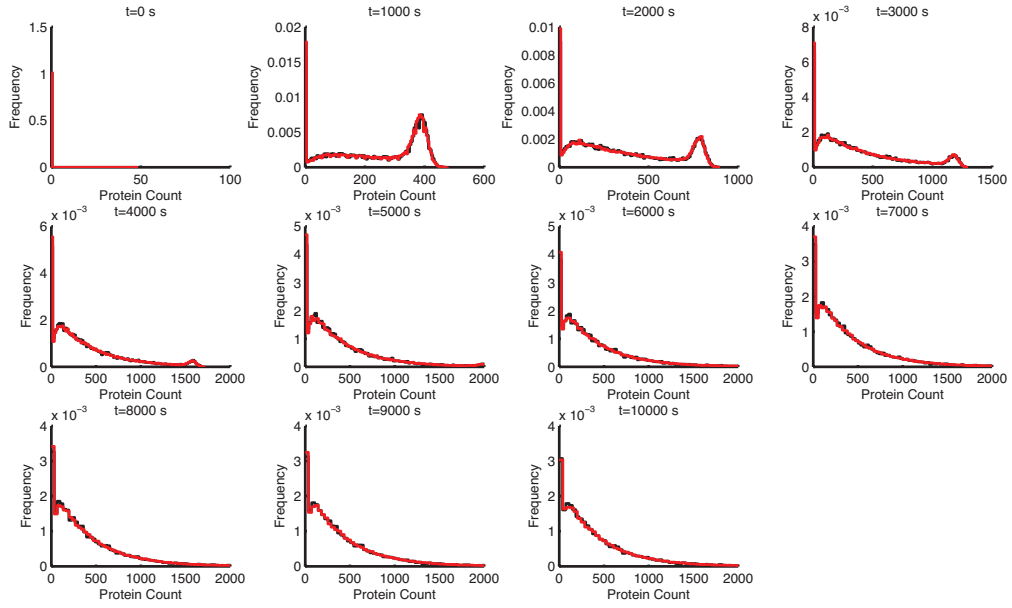


Figure S.2: Protein distributions for each measurement time point. Red: Calibrated model; Black: Reference model.



### S.3.3 Second Order Moment Resolves Non-Identifiability

For the simple model of transient gene activation, the population mean alone does not provide enough information to uniquely determine the four model parameters. This is demonstrated in Fig. S.3, where we compared two different model configurations, each of them found by running the same MCMC algorithm with different initial conditions. In both cases, the estimated means fit well the reference mean. In contrast, the variances significantly differ from each other. Neither parameter set can reproduce the underlying distribution. For the second parameter configuration, the distribution is even unimodal (see Fig S.4).

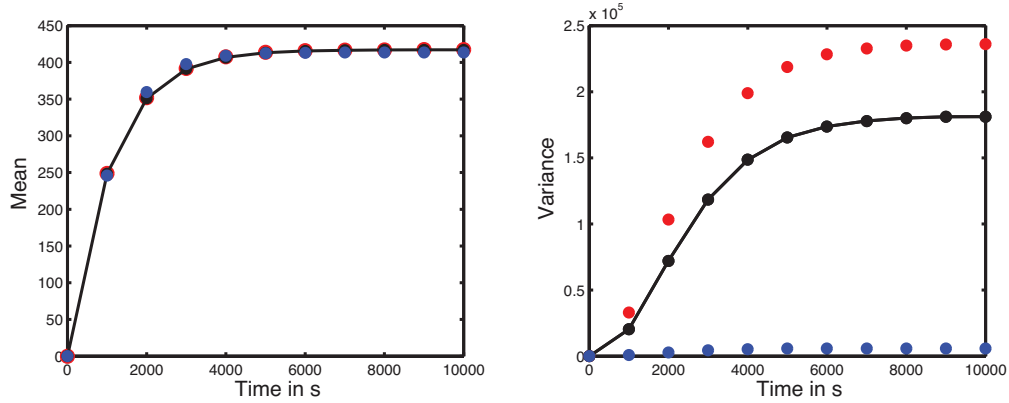


Figure S.3: The plot shows results for two parameter configurations that achieved almost equivalent mean values but strongly differ in the variance. Red, Blue: Calibrated models; Black: Reference model

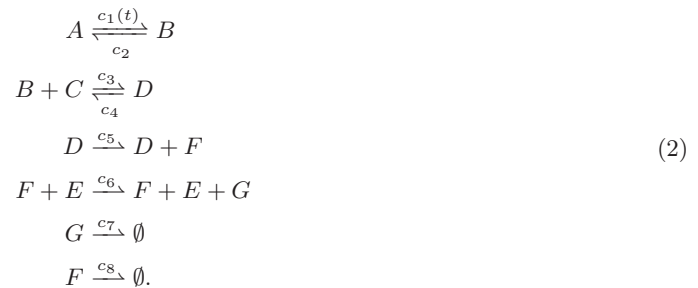
### S.4 Osmo-Stress induced MAPK Hog1 Activity and pSTL1-qV Protein Expression

To keep the notation simple, we use the following acronyms for the model species:

Table S.2: Acronyms for the chemical species.

Species	$pSTL1^{off}$	$pSTL1^{on}$	$CR$	$pSTL1^{on} \cdot CR$	$RIB$	$mRNA$	$pSTL1 - qV$
Acronym	$A$	$B$	$C$	$D$	$E$	$F$	$G$

The Hog1 induced gene expression model is given by the reaction network



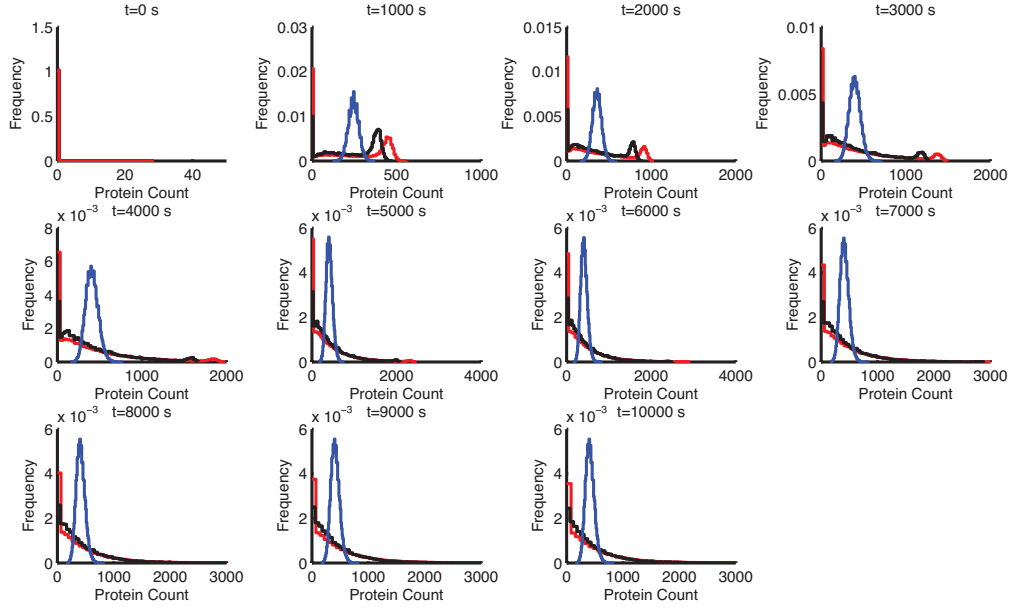


Figure S.4: Protein distributions for two parameter configurations. As the blue distribution is unimodal for all time points, it follows that the multi-modality cannot be predicted from the protein mean alone. Red, Blue: Calibrated models; Black: Reference model.

Note that  $c_1(t)$  denotes a time-varying kinetic parameter as described in Section S.4.2. Note that the synthesis reaction for protein  $G$  can equivalently be written as the standard first-order reaction  $F \xrightarrow{c_6 E} F + G$ .

### S.4.1 Moment Dynamics

Denote by  $\mu_A^1, \mu_B^1, \mu_C^1, \mu_D^1, \mu_E^1, \mu_F^1$ , and  $\mu_G^1$  the means of  $A, B, C, D, E, F$  and  $G$ , respectively. Let  $\mu_A^2, \mu_{AB}^2, \mu_{AC}^2, \mu_{AD}^2, \mu_{AE}^2, \mu_{AF}^2, \mu_{AG}^2, \mu_B^2, \mu_{BC}^2, \mu_{BD}^2, \mu_{BE}^2, \mu_{BF}^2, \mu_{BG}^2, \mu_C^2, \mu_{CD}^2, \mu_{CE}^2, \mu_{CF}^2, \mu_{CG}^2, \mu_D^2, \mu_{DE}^2, \mu_{DF}^2, \mu_{DG}^2, \mu_E^2, \mu_{EF}^2, \mu_{EG}^2, \mu_F^2, \mu_{FG}^2, \mu_G^2$  and  $\mu_{ABC}^3, \mu_{AEF}^3, \mu_{B2C}^3, \mu_{BCD}^3, \mu_{BCE}^3, \mu_{BC2}^3, \mu_{BCF}^3, \mu_{BCG}^3, \mu_{E2F}^3, \mu_{DEF}^3, \mu_{EFG}^3, \mu_{EF2}^3, \mu_{CEF}^3, \mu_{BEF}^3$  be the second and third order moments, respectively. Then the moment equations for this system are given by:

$$\begin{aligned}
\frac{d}{dt}\mu_A^1 &= c_2 \cdot \mu_B^1 - c_1 \cdot \mu_A^1 \\
\frac{d}{dt}\mu_B^1 &= c_1 \cdot \mu_A^1 - c_2 \cdot \mu_B^1 - c_3 \cdot \mu_{BC}^2 + c_4 \cdot \mu_D^1 \\
\frac{d}{dt}\mu_C^1 &= c_4 \cdot \mu_D^1 - c_3 \cdot \mu_{BC}^2 \\
\frac{d}{dt}\mu_D^1 &= c_3 \cdot \mu_{BC}^2 - c_4 \cdot \mu_D^1 \\
\frac{d}{dt}\mu_E^1 &= 0 \\
\frac{d}{dt}\mu_F^1 &= c_5 \cdot \mu_D^1 - c_8 \cdot \mu_F^1 \\
\frac{d}{dt}\mu_G^1 &= c_6 \cdot \mu_{EF}^2 - c_7 \cdot \mu_G^1 \\
\frac{d}{dt}\mu_A^2 &= c_1 \cdot \mu_A^1 - 2 \cdot c_1 \cdot \mu_A^2 + c_2 \cdot \mu_B^1 + 2 \cdot c_2 \cdot \mu_{AB}^2 \\
\frac{d}{dt}\mu_{AB}^2 &= c_1 \cdot \mu_A^2 - c_1 \cdot \mu_A^1 - c_2 \cdot \mu_B^1 + c_2 \cdot \mu_B^2 - c_3 \cdot \mu_{ABC}^3 + c_4 \cdot \mu_{AD}^2 - (c_1 + c_2) \cdot \mu_{AB}^2 \\
\frac{d}{dt}\mu_{AC}^2 &= c_2 \cdot \mu_{BC}^2 - c_1 \cdot \mu_{AC}^2 - c_3 \cdot \mu_{ABC}^3 + c_4 \cdot \mu_{AD}^2 \\
\frac{d}{dt}\mu_{AD}^2 &= c_2 \cdot \mu_{BD}^2 + c_3 \cdot \mu_{ABC}^3 - (c_1 + c_4) \cdot \mu_{AD}^2 \\
\frac{d}{dt}\mu_{AE}^2 &= c_2 \cdot \mu_{BE}^2 - c_1 \cdot \mu_{AE}^2 \\
\frac{d}{dt}\mu_{AF}^2 &= c_2 \cdot \mu_{BF}^2 + c_5 \cdot \mu_{AD}^2 - (c_1 + c_8) \cdot \mu_{AF}^2 \\
\frac{d}{dt}\mu_{AG}^2 &= c_2 \cdot \mu_{BG}^2 + c_6 \cdot \mu_{AEF}^3 - (c_1 + c_7) \cdot \mu_{AG}^2 \\
\frac{d}{dt}\mu_B^2 &= c_1 \cdot \mu_A^1 + 2 \cdot c_1 \cdot \mu_{AB}^2 + c_2 \cdot \mu_B^1 - 2 \cdot c_2 \cdot \mu_B^2 + c_3 \cdot \mu_{BC}^2 - 2 \cdot c_3 \cdot \mu_{B2C}^3 + c_4 \cdot \mu_D^1 + 2 \cdot c_4 \cdot \mu_{BD}^2 \\
\frac{d}{dt}\mu_{BC}^2 &= c_1 \cdot \mu_{AC}^2 - c_3 \cdot \mu_{B2C}^3 - c_3 \cdot \mu_{BC2}^3 + c_4 \cdot \mu_D^1 + c_4 \cdot \mu_{CD}^2 + c_4 \cdot \mu_{BD}^2 - (c_2 - c_3) \cdot \mu_{BC}^2 \\
\frac{d}{dt}\mu_{BD}^2 &= c_1 \cdot \mu_{AD}^2 - c_3 \cdot \mu_{BC}^2 + c_3 \cdot \mu_{B2C}^3 - c_3 \cdot \mu_{BCD}^3 - c_4 \cdot \mu_D^1 + c_4 \cdot \mu_D^2 - (c_2 + c_4) \cdot \mu_{BD}^2 \\
\frac{d}{dt}\mu_{BE}^2 &= c_1 \cdot \mu_{AE}^2 - c_2 \cdot \mu_{BE}^2 - c_3 \cdot \mu_{BCE}^3 + c_4 \cdot \mu_{DE}^2 \\
\frac{d}{dt}\mu_{BF}^2 &= c_1 \cdot \mu_{AF}^2 - c_3 \cdot \mu_{BCF}^3 + c_4 \cdot \mu_{DF}^2 + c_5 \cdot \mu_{BD}^2 - (c_2 + c_8) \cdot \mu_{BF}^2 \\
\frac{d}{dt}\mu_{BG}^2 &= c_1 \cdot \mu_{AG}^2 - c_3 \cdot \mu_{BCG}^3 + c_4 \cdot \mu_{DG}^2 + c_6 \cdot \mu_{BEF}^3 - (c_2 + c_7) \cdot \mu_{BG}^2 \\
\frac{d}{dt}\mu_C^2 &= c_3 \cdot \mu_{BC}^2 - 2 \cdot c_3 \cdot \mu_{BC2}^3 + c_4 \cdot \mu_D^1 + 2 \cdot c_4 \cdot \mu_{CD}^2
\end{aligned}$$

$$\begin{aligned}
\frac{d}{dt}\mu_{CD}^2 &= c_3 \cdot \mu_{BC}^3 - c_3 \cdot \mu_{BC}^2 - c_3 \cdot \mu_{BCD}^3 - c_4 \cdot \mu_D^1 + c_4 \cdot \mu_D^2 - c_4 \cdot \mu_{CD}^2 \\
\frac{d}{dt}\mu_{CE}^2 &= c_4 \cdot \mu_{DE}^2 - c_3 \cdot \mu_{BCE}^3 \\
\frac{d}{dt}\mu_{CF}^2 &= c_4 \cdot \mu_{DF}^2 - c_3 \cdot \mu_{BCF}^3 + c_5 \cdot \mu_{CD}^2 - c_8 \cdot \mu_{CF}^2 \\
\frac{d}{dt}\mu_{CG}^2 &= c_4 \cdot \mu_{DG}^2 - c_3 \cdot \mu_{BCG}^3 + c_6 \cdot \mu_{CEF}^3 - c_7 \cdot \mu_{CG}^2 \\
\frac{d}{dt}\mu_D^2 &= c_3 \cdot \mu_{BC}^2 + 2 \cdot c_3 \cdot \mu_{BCD}^3 + c_4 \cdot \mu_D^1 - 2 \cdot c_4 \cdot \mu_D^2 \\
\frac{d}{dt}\mu_{DE}^2 &= c_3 \cdot \mu_{BCE}^3 - c_4 \cdot \mu_{DE}^2 \\
\frac{d}{dt}\mu_{DF}^2 &= c_3 \cdot \mu_{BCF}^3 + c_5 \cdot \mu_D^2 - (c_4 + c_8) \cdot \mu_{DF}^2 \\
\frac{d}{dt}\mu_{DG}^2 &= c_3 \cdot \mu_{BCG}^3 + c_6 \cdot \mu_{DEF}^3 - (c_4 + c_7) \cdot \mu_{DG}^2 \\
\frac{d}{dt}\mu_E^2 &= 0 \\
\frac{d}{dt}\mu_{EF}^2 &= c_5 \cdot \mu_{DE}^2 - c_8 \cdot \mu_{EF}^2 \\
\frac{d}{dt}\mu_{EG}^2 &= c_6 \cdot \mu_{E2F}^3 - c_7 \cdot \mu_{EG}^2 \\
\frac{d}{dt}\mu_F^2 &= c_5 \cdot \mu_D^1 + 2 \cdot c_5 \cdot \mu_{DF}^2 + c_8 \cdot \mu_F^1 - 2 \cdot c_8 \cdot \mu_F^2 \\
\frac{d}{dt}\mu_{FG}^2 &= c_5 \cdot \mu_{DG}^2 - (c_8 + c_7) \cdot \mu_{FG}^2 + c_6 \cdot \mu_{EF2}^3 \\
\frac{d}{dt}\mu_G^2 &= c_6 \cdot \mu_{EF}^2 + 2 \cdot c_6 \cdot \mu_{EFG}^3 + c_7 \cdot \mu_G^1 - 2 \cdot c_7 \cdot \mu_G^2
\end{aligned}$$

A closed system is then obtained by replacing the third order cumulants by zero, which is equivalent to replacing the third order moments by functions of the lower order moments as follows:



$$\begin{aligned}
\mu_{ABC}^3 &= \mu_A^1 \cdot \mu_{BC}^2 + \mu_B^1 \cdot \mu_{AC}^2 + \mu_C^1 \cdot \mu_{AB}^2 - 2 \cdot \mu_A^1 \cdot \mu_B^1 \cdot \mu_C^1, \\
\mu_{AEF}^3 &= \mu_A^1 \cdot \mu_{EF}^2 + \mu_E^1 \cdot \mu_{AF}^2 + \mu_F^1 \cdot \mu_{AE}^2 - 2 \cdot \mu_A^1 \cdot \mu_E^1 \cdot \mu_F^1, \\
\mu_{B2C}^3 &= -2 \cdot \mu_C^1 \cdot \mu_B^{1^2} + 2 \cdot \mu_{BC}^2 \cdot \mu_B^1 + \mu_C^1 \cdot \mu_B^2, \\
\mu_{BCD}^3 &= \mu_B^1 \cdot \mu_{CD}^2 + \mu_C^1 \cdot \mu_{BD}^2 + \mu_D^1 \cdot \mu_{BC}^2 - 2 \cdot \mu_B^1 \cdot \mu_C^1 \cdot \mu_D^1, \\
\mu_{BCE}^3 &= \mu_B^1 \cdot \mu_{CE}^2 + \mu_C^1 \cdot \mu_{BE}^2 + \mu_E^1 \cdot \mu_{BC}^2 - 2 \cdot \mu_B^1 \cdot \mu_C^1 \cdot \mu_E^1, \\
\mu_{BC2}^3 &= -2 \cdot \mu_B^1 \cdot \mu_C^{1^2} + 2 \cdot \mu_{BC}^2 \cdot \mu_C^1 + \mu_B^1 \cdot \mu_C^2, \\
\mu_{BCF}^3 &= \mu_B^1 \cdot \mu_{CF}^2 + \mu_C^1 \cdot \mu_{BF}^2 + \mu_F^1 \cdot \mu_{BC}^2 - 2 \cdot \mu_B^1 \cdot \mu_C^1 \cdot \mu_F^1, \\
\mu_{BCG}^3 &= \mu_B^1 \cdot \mu_{CG}^2 + \mu_C^1 \cdot \mu_{BG}^2 + \mu_G^1 \cdot \mu_{BC}^2 - 2 \cdot \mu_B^1 \cdot \mu_C^1 \cdot \mu_G^1, \\
\mu_{E2F}^3 &= -2 \cdot \mu_F^1 \cdot \mu_E^{1^2} + 2 \cdot \mu_{EF}^2 \cdot \mu_E^1 + \mu_F^1 \cdot \mu_E^2, \\
\mu_{DEF}^3 &= \mu_D^1 \cdot \mu_{EF}^2 + \mu_E^1 \cdot \mu_{DF}^2 + \mu_F^1 \cdot \mu_{DE}^2 - 2 \cdot \mu_D^1 \cdot \mu_E^1 \cdot \mu_F^1, \\
\mu_{EFG}^3 &= \mu_E^1 \cdot \mu_{FG}^2 + \mu_F^1 \cdot \mu_{EG}^2 + \mu_G^1 \cdot \mu_{EF}^2 - 2 \cdot \mu_E^1 \cdot \mu_F^1 \cdot \mu_G^1, \\
\mu_{EF2}^3 &= -2 \cdot \mu_E^1 \cdot \mu_F^{1^2} + 2 \cdot \mu_{EF}^2 \cdot \mu_F^1 + \mu_E^1 \cdot \mu_F^2, \\
\mu_{CEF}^3 &= \mu_C^1 \cdot \mu_{EF}^2 + \mu_E^1 \cdot \mu_{CF}^2 + \mu_F^1 \cdot \mu_{CE}^2 - 2 \cdot \mu_C^1 \cdot \mu_E^1 \cdot \mu_F^1, \\
\mu_{BEF}^3 &= \mu_B^1 \cdot \mu_{EF}^2 + \mu_E^1 \cdot \mu_{BF}^2 + \mu_F^1 \cdot \mu_{BE}^2 - 2 \cdot \mu_B^1 \cdot \mu_E^1 \cdot \mu_F^1.
\end{aligned}$$

As in the example of Section S.3, second order zero cumulant moment closure led to good agreement of the approximations with estimates obtained from stochastically simulating the system with the found parameters and was therefore sufficient for this model.

#### S.4.2 Nuclear Hog1 Activity as Model Input

We used fluorescence microscopy to measure nuclear enrichment of MAPK Hog1. The data set was recorded by co-authors, originally published in [9]. To extract the time evolution from the single-cell measurements (around 50-120 cells), we first removed outliers manually and normalized the data. In particular, we divided nuclear Hog1 (obtained by image segmentation) by the total Hog1 in the cell to reduce photo-bleaching artifacts. Note that the relocation data measures active and also inactive Hog1, whereas the latter cannot bind to the target gene. Thus, we also subtracted the mean basal levels, calculated from time-points before osmotic stress. The basal active Hog1 level - denoted  $u_0$  is then inferred during model calibration, together with the other parameters. Note that this procedure is based on the assumption that export of nuclear Hog1 is fast once it has been deactivated. The resulting data was then used to perform a linear regression with radial basis functions (RBFs). The base functions were centered around the measurement time points and the width parameter was adjusted according to the increasing measurement intervals. This procedure was repeated for each salt concentration. Note that to properly capture the stress-dependent protein expression, we additionally measured the STL1 reporter for intermediate salt concentrations in the flow cytometry experiments. The corresponding deterministic Hog1 functions were obtained by performing a linear interpolation along the NaCl-time plane (see Fig S.5).

In the following, the extracted nuclear Hog1 function for a given salt concentration is denoted as  $u(t)$ . To obtain the corresponding time-varying gene activation intensity, we transformed the Hog1 abundance using a Hill-function, i.e.,

$$c_1(t) = \frac{V_{max}(u(t) + u_0)^{n_H}}{K_d^{n_H} + (u(t) + u_0)^{n_H}}, \quad (3)$$

with  $u_0$ ,  $n_H$ ,  $V_{max}$  and  $K_d$  as unknown parameters, estimated during model calibration.

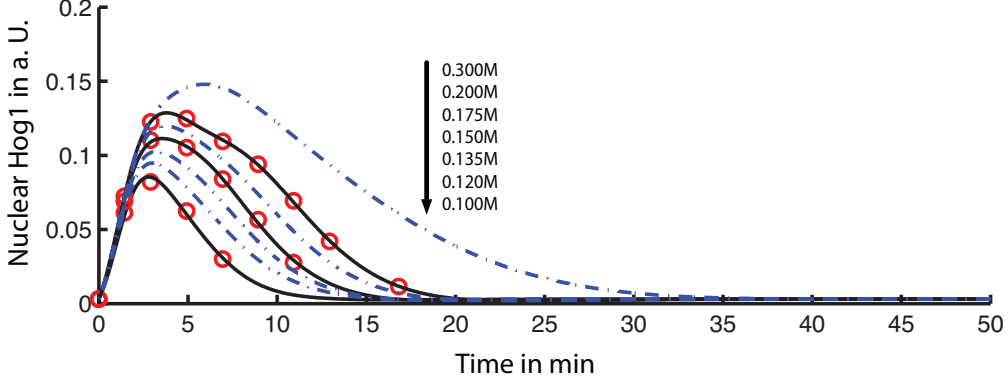


Figure S.5: Extracted nuclear Hog1 enrichment over time. Black: Fitted using linear RBF regression; blue: interpolated; red: mean abundance obtained from microscopy.

### S.4.3 Extrinsic Variability

We model extrinsic variability at two different stages of the proposed model. First, we assume variability in chromatin remodeling, because it depends on a variety of different complexes which might be subject to cell-to-cell variations (such as RSC or the SAGA complex, [8]). Chromatin remodeling is modeled by recruitment of species  $C$  by the active gene  $B$ . Thus, variability in the total number of CR ( $Z_1 = [C] + [D]$ ) leads to variability in the remodeling efficiency. Furthermore, we assume heterogeneity in the translation efficiency, which is reflected by a variability in the proxy species  $Z_2 = [E]$ . Mean and covariance matrix of the random vector  $Z = [Z_1, Z_2]^T$  are defined as

$$\mathbb{E}[Z] = \begin{pmatrix} \alpha_1^1 \\ \alpha_2^1 \end{pmatrix} \quad (4)$$

and

$$\mathbb{E}[(Z - \mathbb{E}[Z])(Z - \mathbb{E}[Z])^T] = \begin{pmatrix} \alpha_{11}^2 & \alpha_{12}^2 \\ \alpha_{12}^2 & \alpha_{22}^2 \end{pmatrix} \quad (5)$$

The extrinsic statistics then enter the moment equations in terms of the initial conditions for the means  $\mathbb{E}[C + D](0) = \mathbb{E}[C](0) = \alpha_1^1$  (as initially all CR is unbound),  $\mathbb{E}[E](0) = \alpha_2^1$ , the variances  $Var(C + D)(0) = Var(C)(0) = \alpha_{11}^2$ ,  $Var(E)(0) = \alpha_{22}^2$  and the covariance  $Cov(C + D, E)(0) = Cov(C, E)(0) = \alpha_{12}^2$  and remain constant for all times, since  $[C] + [D]$  and  $[E]$  are conserved. The extrinsic statistics were assumed to be unknown and inferred from the measurements as explained in Section S.4.5. Note that within a moment-based approach, no assumptions on the distribution  $P_Z$  are required. However, once a comparison between the protein distributions is desired, realizations  $z$  have to be drawn from  $P_Z$  for each SSA run. As  $P_Z$  is not fully characterized by mean and variance only, further assumptions need to be made. Here, we restrict the shape of  $P_Z$  to be log-normal [3] and compute its parameters from the inferred extrinsic statistics (i.e., first and second order moments).

### S.4.4 Flow Cytometry Data for pSTL1-qV Reporter

The time-lapse pSTL1-qV distribution dataset (FC;TL) was measured using flow cytometry at different time points between zero and 48 minutes in six-minute intervals. Salt was added to the media straight after time  $t = 0min$  by hand, which yielded a slightly delayed rather than an immediate exposition to the desired salt level. We accounted for this by correcting the later time points by  $-3min$ , and

consequently, we obtained the modified measurement time points  $0min, 3min, 9min, \dots, 45min$ . We observed a strong bimodality in the logarithmic forward scatter (FSC) and side scatter (SSC) plots, indicating the presence of two phenotypic subpopulations. Consequently, we applied a weak cell gate on the FSC and SSC channel to exclude one of the two subpopulations. In analogy, this procedure was performed on the additional snapshot dataset (FC;SN).

#### S.4.5 Model Calibration

Analogous to Section S.3.2, we first validated the normality assumption of the empirical moment as described in the Materials and Methods section in the main text. As graphical method we applied standard probability-probability ( $P-P$ ) and quantile-quantile ( $Q-Q$ ) plots whereas to quantitatively assess normality we performed the Kolmogorov-Smirnov-Lilliefors test. The results and the corresponding  $p$ -values of the normality tests are shown in Fig. S.6. The described analysis was performed on the pSTL1-qV distribution for  $0.1M$  NaCl at time  $45min$ , where the measured distribution is bimodal.

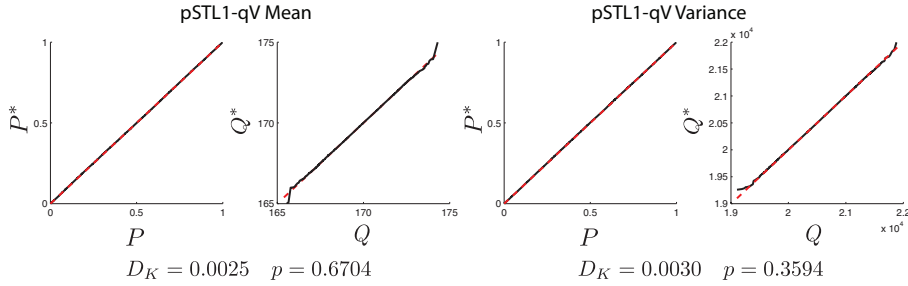


Figure S.6: Normality of empirical moments. Distributions of empirical moments were computed using bootstrapping. In particular, we randomly subsampled 10000 abundance values from the pSTL1-qV distribution and computed the sample mean and variance. This was repeated 50000 times to obtain approximations of the distributions of the moment estimates. For comparison, analytical normal distributions were fitted to the bootstrapped data and compared against each other using standard probability-probability ( $P-P$ ) and quantile-quantile ( $Q-Q$ ) plots. In all four plots, the x-axes correspond to the analytical normal distributions, whereas the y-axes correspond to the bootstrapped empirical distributions. The low Kolmogorov-Smirnov test statistics  $D_K$  in conjunction with the large  $p$ -values indicate that the distribution of the empirical moment estimates can be well approximated by a normal distribution.

The model comprises the parameters  $u_0, n_H, K_d, V_{max}, c_2, c_3, c_4, c_5, c_6, c_7, c_8, \alpha_1^1, \alpha_2^1, \alpha_{11}^2, \alpha_{22}^2$  and  $\alpha_{12}^2$ . Note that the propensity of a translation event is proportional to the product  $c_6 \cdot [E]$ . Thus, the parameters  $c_6, \alpha_2^1$  and  $\alpha_{22}^2$  are structurally unidentifiable [1, 6]; hence we estimated the products  $c_6\alpha_2^1$  and  $c_6\alpha_{22}^2$ . Given those products, statistics of the number of ribosomes could for instance be estimated by setting  $c_6$  to values from literature and quantifying the remaining part.

Due to the high-dimensional state and parameter space and the fact that our flow cytometry experiments only captured distributions of a single protein, we expected the inference problem to be non-convex (i.e., characterized by a multi-modal posterior distribution). For this reason, a M-H scheme analogous to Section S.3.2 was performed 50 times, each time with randomly drawn initial parameter configurations. Note that in general, this is likely to give parameter configurations for which the nonlinear moment system is numerically unstable and thus, we first selected a stable parameter configuration  $\gamma_s$  (see Tab. S.3). Then, each parameter value was randomly initialized around those initial parameter values, i.e.,  $\gamma_j^0 \sim \mathcal{LN}(\ln \gamma_{j,s}, 0.5^2)$ . The scaling parameters  $v_j$  of the log-normal proposal densities are given in Tab. S.3. We then sorted the 50 parameter sets according to their maximum a-posteriori probabilities and selected the best five parameter sets for further inspection.

To quantitatively assess the goodness-of-fit we made use of a distance measure between probability distributions. Among the various metrics proposed in literature (see [5] for an overview) we chose the uniform or Kolmogorov metric (measuring the maximum deviation between the cumulative distributions). As it is scale-invariant and bounded by one it appears more amenable to interpretation than others distance measures. Moreover, it was already used in the context of stochastic simulations of chemical kinetics [2]. Thus, for each of the above parameter sets we computed the Kolmogorov distance for all NaCl concentrations and time-points between the empirical and the predicted pSTL1-qV distributions. The final parameter set was selected such as to minimize the total distance, i.e., the sum of individual Kolmogorov distances over all time-points and the three concentrations used for fitting (i.e., 0M, 0.12M and 0.2M of NaCl). Fig. S.7 depicts the Kolmogorov distance for the best performing parameter set which is given in Tab. S.3. The results suggest that the model provides a reasonable approximation of the experimental data; indeed we are not aware of any study in the literature reporting comparable accuracy in predicting time lapsed distribution data. We believe that a further decrease in distance  $D_K$  may be achievable by adding complexity to the transcriptional and translational steps. However, in the absence of any further experimental readout, such extensions are likely to introduce new structural identifiability problems.

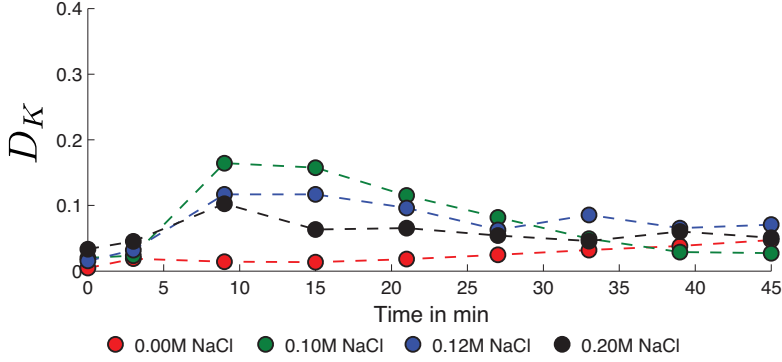


Figure S.7: Distance in distribution between model predictions and data. The Kolmogorov distance  $D_K$  was computed between the empirical and the predicted pSTL1-qV distribution for each time point and concentration. Note that the distance  $D_K$  is bounded by one.

#### S.4.6 Model Validation

Besides validating the predicted distributions with the time-lapsed dataset (FC;TL) for an intermediate NaCl concentration (i.e., 0.1M), we further validated the inferred model using the additional snapshot dataset (FC;SN). In particular, we computed the coefficient of variation and the percentage of protein producing cells at  $t = 45min$  for different salt concentrations between 0M and 0.3M. The former was computed as

$$CV = \frac{\hat{\sigma}_F}{\hat{\mu}_F}, \quad (6)$$

where  $\hat{\mu}_F$  and  $\hat{\sigma}_F$  are the empirical means and standard deviations obtained from the protein distributions. For comparison with the experimental data, we added the estimated mean and variance of the autofluorescence intensity to the means and variances of the simulated distributions. To estimate the number of responding cells for the different concentrations (dose-response) from the experimental data, we removed the autofluorescence by performing a deconvolution between the measured flow cytometry distributions and the autofluorescence distribution (obtained by flow cytometry for 0M NaCl, see main text, Materials and Methods). Subsequently, the percentage of non-expressing cells was given by the



Table S.3: Inferred model parameters and Metropolis-Hastings setup. The  $\gamma_{j,s}$  denote the initial parameter values, chosen such as to obtain stable moment dynamics. The inferred MAP estimates are denoted as  $\gamma_{j,MAP}$ . The scaling parameters of the log-normal proposal distributions used in the Metropolis-Hastings algorithm are denoted as  $v_j$ .

Parameter	$\gamma_{j,s}$	$\gamma_{j,MAP}$	Unit	$v_j$
$u_0$	$3.000 \cdot 10^{-2}$	$1.581 \cdot 10^{-2}$	$a.U.$	0.01
$n_H$	3.000	6.130	1	0.01
$K_d$	$2 \cdot 10^{-1}$	$1.418 \cdot 10^{-1}$	$a.U.$	0.01
$V_{max}$	1.000	1.025	$s^{-1}$	0.01
$c_2$	1.000	1.384	$s^{-1}$	0.02
$c_3$	$4.000 \cdot 10^{-4}$	$6.669 \cdot 10^{-4}$	$s^{-1}$	0.02
$c_4$	$1.000 \cdot 10^{-3}$	$1.469 \cdot 10^{-2}$	$s^{-1}$	0.02
$c_5$	1.000	$2.825 \cdot 10^{-1}$	$s^{-1}$	0.02
$c_7$	$1.000 \cdot 10^{-3}$	$5.476 \cdot 10^{-4}$	$s^{-1}$	0.02
$c_8$	$1.000 \cdot 10^{-4}$	$1.283 \cdot 10^{-4}$	$s^{-1}$	0.02
$\alpha_1^1$	$3.300 \cdot 10^1$	$2.250 \cdot 10^2$	1	0.01
$c_6 \alpha_2^1$	$3.300 \cdot 10^{-2}$	$5.663 \cdot 10^{-3}$	$s^{-1}$	0.01
$\alpha_{11}^2$	$1.900 \cdot 10^3$	$7.809 \cdot 10^3$	1	0.02
$c_6 \alpha_{22}^2$	$1.900 \cdot 10^{-3}$	$3.098 \cdot 10^{-6}$	$s^{-2}$	0.02
$c_6 \alpha_{12}^2$	$1.100 \cdot 10^{-1}$	$8.935 \cdot 10^{-2}$	$s^{-1}$	0.01

histogram value at bin zero. To interpolate between the measurement time points we fitted curves for the CV and the dose-response using polynomial base functions and a Hill-function, respectively. In both cases, the parameters were found by minimizing the  $L_2$  - norm between the parametric model output and the CV and dose-response values.

#### S.4.7 Slow Transcriptional Activation Causes Transient Bimodality

The authors of [9] hypothesize that the induced pSTL1-qV bimodality originates from slow stochastic events in the transcriptional activation (i.e., interaction of the active gene with the chromatin remodeling complex, mRNA Polymerase II binding, etc.). Depending on the kinetic parameters, our model can induce bimodality either in the gene activation - or the subsequent CR binding step. To computationally verify the hypothesis from [9], we performed a statistical analysis of those events using stochastic simulation with the inferred model parameters from Tab.S.3. In particular, we found that for intermediate salt concentrations, all cells have active Hog1 bound to the promoter, whereas CR binding occurs in only a fraction of the cells (see Fig. S.8). This observation provides strong computational support for the hypothesis that chromatin remodeling is the key factor in bimodal pSTL1-qV induction and furthermore that the induction is governed by intrinsic stochasticity.

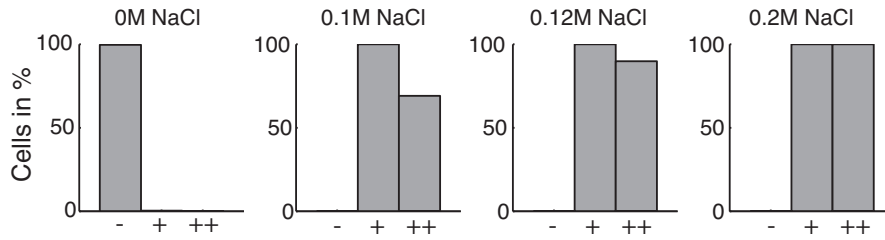


Figure S.8: Statistical analysis of the transcriptional activation in pSTL1-qV expression. Bars indicate the percentage of cells that never activated the gene (-), that activated the gene at least once (+) and cells that initiated transcription (++). Statistics were computed from 1000 traces obtained using stochastic simulation with the inferred parameters.

## References

- [1] R. Bellman and K. Astrom. On structural identifiability. *Mathematical Biosciences*, 7(3-4):329–339, 1970.
- [2] Y. Cao and L. Petzold. Accuracy limitations and the measurement of errors in the stochastic simulation of chemically reacting systems. *Journal of Computational Physics*, 212(1):6–24, Feb. 2006.
- [3] C. Furusawa, T. Suzuki, A. Kashiwagi, T. Yomo, and K. Kaneko. Ubiquity of log-normal distributions in intra-cellular reaction dynamics. *Biophysics*, 1:25–31, 2005.
- [4] J. D. Gibbons and S. Chakraborti. *Nonparametric statistical inference*. Marcel-Dekker, New York, 4 edition, 2003.
- [5] A. L. Gibbs and F. E. Su. On choosing and bounding probability metrics. *International Statistical Review*, 70(3):419–435, 2002.
- [6] S. Hengl, C. Kreutz, J. Timmer, and T. Maiwald. Data-based identifiability analysis of non-linear dynamical models. *Bioinformatics*, 23(19):2612–2618, 2007.
- [7] J. P. Hespanha. **StochDynTools** — a MATLAB toolbox to compute moment dynamics for stochastic networks of bio-chemical reactions. Available at: <http://www.ece.ucsb.edu/~hespanha>, May 2007.
- [8] F. C. Holstege, E. G. Jennings, J. J. Wyrick, T. I. Lee, C. J. Hengartner, M. R. Green, T. R. Golub, E. S. Lander, and R. a. Young. Dissecting the regulatory circuitry of a eukaryotic genome. *Cell*, 95(5):717–28, Nov. 1998.
- [9] S. Pelet, F. Rudolf, M. Nadal-Ribelles, E. de Nadal, F. Posas, and M. Peter. Transient activation of the HOG MAPK pathway regulates bimodal gene expression. *Science (New York, N.Y.)*, 332(6030):732–5, May 2011.