Identifying sources of variation and the flow of information in biochemical networks

Clive G. Bowsher[a,1] **and Peter S. Swain**[b,1]

[a]*School of Mathematics, University of Bristol, Bristol BS8 1TW, United Kingdom; and* [b]*SynthSys—Synthetic and Systems Biology, University of Edinburgh, Edinburgh EH9 3JD, United Kingdom*

AUTHOR SUMMARY

Cells respond to change stochastically because biochemical mechanisms are stochastic: Intermolecular collisions trigger biochemical events at seemingly random times. Yet cells can both control and exploit such stochasticity. To understand how, we must identify the sources of stochasticity and distinguish variation due to information flow, say from the extracellular environment, from that due to confounding "noise". Here, we present a general scheme for dynamic, stochastic systems to decompose variation into as many components as there are potential sources of stochasticity. Further, we identify a particular component that measures information flow and describe conditions that experimental probes, or "reporters," should satisfy to determine this component and all others.

Intuitively, the effect of fluctuations in one stochastic variable on the variation in a variable of interest can be determined by comparing the expected variation in this variable when the first, "explanatory" variable freely fluctuates to that when it is somehow "fixed". Our general decomposition of variance incorporates this idea by mathematically "fixing" the explanatory variable by conditioning on its history (i.e., its current value and its values at all previous times) (1). If we would like to determine how fluctuations of n sources affect the variation of an "output" variable, Z, then the decomposition has $n+1$ terms: one term corresponding to each of the n variables and another term to describe any additional variation in Z that is generated through other sources.

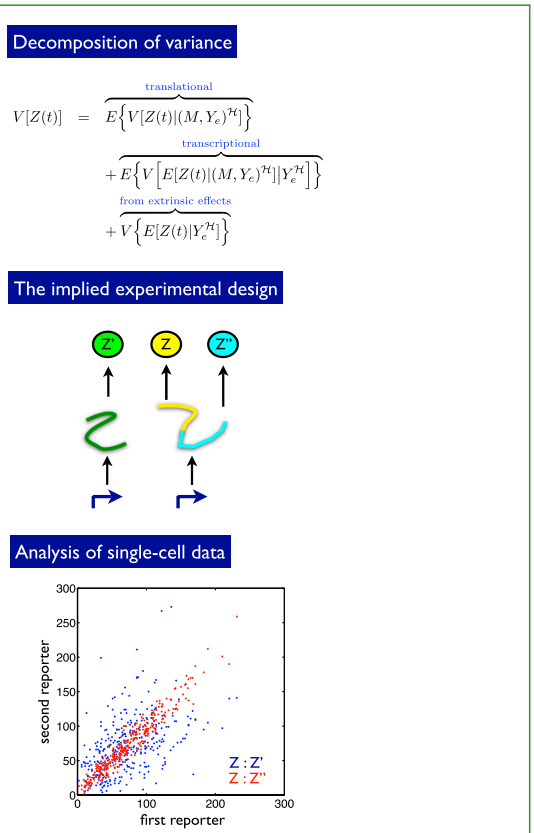

Fig. P1. A simple example: quantifying transcriptional and translational variation. By conditioning on the history of mRNA levels, $M^{\mathcal{H}}$, and of all processes extrinsic to gene expression, $Y_e^{\mathcal{H}}$, we can decompose the variance in the level of the protein, Z, into three components. For example, translational variation is the extra variation generated on average given the joint history of mRNA and Y_e. This decomposition implies that we need a reporter for protein levels (Z), a reporter conjugate given the history of Y_e (Z'), and a reporter conjugate given the joint history of M and Y_e (Z''). Scatter plots (here using simulated data) give a visual representation of the magnitude of the variance components. Each dot represents measurements of two reporters in an individual cell. For Z versus Z', scatter perpendicular to the diagonal $Z=Z'$ gives the sum of transcriptional and translational variation and scatter parallel to the diagonal gives the sum of the variance of Z and extrinsic variation. For Z versus Z'', perpendicular scatter is given by translational variation and parallel scatter is given by the variance of Z and both transcriptional and extrinsic variation. The extrinsic variation is equal to the covariance of Z and Z'.

For biochemical systems, we describe conditions that reporters should satisfy to measure each of the components. A reporter, such as a fluorescently tagged protein, is needed to measure the output of the system. To assess the effect of fluctuations in a variable, Y, on the output, Z, we must construct a second "conjugate" reporter so that its level only correlates with the level of Z because of fluctuations in the levels of Y (the two reporters are conditionally independent, given the history of Y). Then, provided that the reporters have the same conditional means and the same conditional variances, both the covariance of the two reporters and their mean squared difference describe terms in the decomposition. Therefore, for n terms in the decomposition, n reporters are needed.

To illustrate our theory, we define transcriptional and translational variation and show how such variation can be measured experimentally using fluorescent proteins translated from a bicistronic mRNA (Fig. P1). For a standard model of gene expression, we also provide mathematical expressions for these components by

Author contributions: C.G.B. and P.S.S. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. E-mail: c.bowsher@bristol.ac.uk or peter.swain@ed.ac.uk.

See full research article on page E1320 of www.pnas.org.

Cite this Author Summary as: PNAS 10.1073/pnas.1119407109.

including conjugate reporters in the mathematical model we analyze. We expect this analysis technique to be widely applicable for evaluating variance components using stochastic models of biochemical networks.

Much gene expression is initiated by signaling networks, and we go on to explore how fluctuations in the environment and upstream signaling can affect downstream gene expression. Using a three-component decomposition, we reanalyze previous measurements of the response of budding yeast to extracellular pheromones (2) and conclude that, in this example, variation from signal transduction is less substantial than variation arising from other processes extrinsic to gene expression.

Variation in gene expression is also a consequence of information flowing through signaling networks from the extracellular environment. We identify a component in our decomposition whose size relative to the variance of the output, Z, typically indicates how readily a network can "decide" the state of the extracellular environment from the level of Z. For an environmental input, X (for example, the concentrations of a collection of ligands), we express this informational fraction of the output variance as $V\{E[Z|X]\}/V[Z]$, where E denotes expectation and V denotes variance. As the informational fraction tends to its maximum value of one, the distributions of outputs corresponding to each state of the input become more distinct from one another. Consequently, it is easier to unambiguously identify the state of the input: More values of output can be uniquely identified with an input state because the overlap between the (conditional) output distributions decreases.

We apply our results to measurements of sensing of osmotic stress by budding yeast (3) and show that informational variation can be the dominant source of variation. As much as 80% can be informational if we include variation generated in response to changes in the environment's osmolarity. By finding probability distributions of osmotic stress that maximize the informational fraction we can, with some caveats, make predictions about the types of environment most faithfully detected by budding yeast's osmosensing network. These environments have frequent low levels of osmotic stress and infrequent high levels.

A challenge when investigating variation in any system is the influence of the wider stochastic environment in which the system is embedded. Although previous work decomposed variation into intrinsic and extrinsic components (4, 5), there was no means to identify the processes generating variation or to quantify their effects. Our general decomposition and conditions for conjugate reporters provide just such a means. They are essential knowledge for experimental design: Although constructing reporters may sometimes be difficult, these difficulties may not even be apparent, let alone solved, without precisely knowing the properties required of the reporters. Our results hold for any stochastic dynamic system at any point in time and, taken together, thus provide a general mathematical foundation for studies of cellular variation.

1. Bowsher CG (2010) Stochastic kinetic models: Dynamic independence, modularity and graphs. *Ann Stat* 38:2242–2281.
2. Colman-Lerner A, et al. (2005) Regulated cell-to-cell variation in a cell-fate decision system. *Nature* 437:699–706.
3. Pelet S, et al. (2011) Transient activation of the HOG MAPK pathway regulates bimodal gene expression. *Science* 332:732–735.
4. Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. *Science* 297:1183–1186.
5. Swain PS, Elowitz MB, Siggia ED (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci USA* 99:12795–12800.