Bimodal Protein Distributions in Heterogeneous Oscillating Systems

Maciej Dobrzyński, Dirk Fey, Lan K. Nguyen, and Boris N. Kholodenko

Systems Biology Ireland,
University College Dublin, Belfield, Dublin 4, Ireland
{maciej.dobrzynski,dirk.fey,lan.nguyen,boris.kholodenko}@ucd.ie

Abstract. Bimodal distributions of protein activities in signaling systems are often interpreted as indicators of underlying switch-like responses and bistable dynamics. We investigate the emergence of bimodal protein distributions by analyzing a less appreciated mechanism: oscillating signaling systems with varying amplitude, phase and frequency due to cell-to-cell variability. We support our analysis by analytical derivations for basic oscillators and numerical simulations of a signaling cascade, which displays sustained oscillations in protein activities. Importantly, we show that the time to reach the bimodal distribution depends on the magnitude of cell-to-cell variability. We quantify this time using the Kullback-Leibler divergence. The implications of our findings for single-cell experiments are discussed.

Keywords: signaling networks, oscillations, bimodality, stochasticity, protein distributions.

1 Introduction

Protein levels in cellular systems undergo constant changes due to varying extra- and intracellular cues that are dynamically processed by cellular machinery as well as due to thermal noise – an inevitable factor affecting all biochemical reactions. It is because of this variability that cells within a population, be it a bacterial colony or tumor cells, at any given point in time exhibit a distribution of values rather than a precise value of concentrations of its biochemical components, such as proteins or mRNA. Such distributions can be assessed as population snapshots in fluorescence-activated assays using flow cytometry or cell imaging. In both cases the measurement of fluorescence intensity in individual cells correlates with protein abundance. This starkly contrasts to bulk measurements such as Western blots where proteins are detected in cell lysates, which only estimates the average (per-cell) concentration of the entire population.

Of particular interest are bimodal protein distributions that indicate a temporal or steady-state phenotypic division of an isogenic cellular population. Bimodality often reflects the existence of two subpopulations, each capable of performing a different task [2] or having an altered survival rate to stress [3] and

D. Gilbert and M. Heiner (Eds.): CMSB 2012, LNCS 7605, pp. 17–28, 2012.

drug treatment [15]. Bimodal distributions may arise in a number of situations: a purely stochastic genetic switch [1], a bistable system with stochastically induced transitions [11], or noisy networks with sigmoidal response function [8,9]. In this paper we address a much less appreciated mechanism: heterogeneous oscillations. We show that cell-to-cell variability in protein abundances can result in bimodal distributions of concentrations of active (e.g. phosphorylated) protein forms, although individual cells display solely deterministic oscillatory dynamics. We examine analytically and numerically conditions under which this phenomenon occurs.

2 Results

A *single* oscillating cell visits all intermediate levels between the low and the high protein concentrations. A histogram, or a distribution, of concentrations assumed over time can be constructed in the following manner. The range of concentrations between oscillation extrema is divided into infinitesimally small bins and the time the system spends in each of the bins is recorded. For deterministic oscillations, a single period suffices to obtain such a distribution. Depending on the shape of these oscillations, a bimodal single-cell time-averaged histogram of concentrations may arise (Fig. 1). The key question, however, is whether in the presence of cell-to-cell variability which affects the amplitude, phase and frequency of oscillations in individual cells, the described mechanism can also evoke bimodality at the level of a cellular population? The question is equivalent to asking about the ergodicity of such a system: does the distribution of a population coincide with the distribution of an individual measured over time? The disparity of the two has been recently demonstrated experimentally for noisy cellular systems [14]. Protein fluctuations that are high in amplitude and slow compared to cells lifetime may drive a number of cells to a range of concentrations that is only a fraction of the entire concentration spectrum. This condition may persist well over a cells generation thus rendering snapshots of the population incapable of reflecting the underlying network dynamics.

Similar phenomenon may affect a population of oscillating cells. Even though our analysis focuses on oscillations that are deterministic in individual cells, biochemical noise manifests itself in cell-to-cell variability. As a result, oscillations across the population differ in the amplitude, phase and frequency. If this variability is not large enough, a population might not cover the entire concentration spectrum at a given point in time, and a bimodal distribution fails to emerge. An additional condition is required to facilitate this emergence and relates to a so-called mixing time – the time after which all individuals within the population of cells assume all states of the asymptotic (stationary) protein distribution. We therefore set out to answer following questions: (1) under what biochemical circumstances can a heterogeneous population of cells exhibiting oscillatory dynamics give rise to bimodal protein distributions? (2) What is the time after stimulus required to reach a time-independent bimodal distribution?

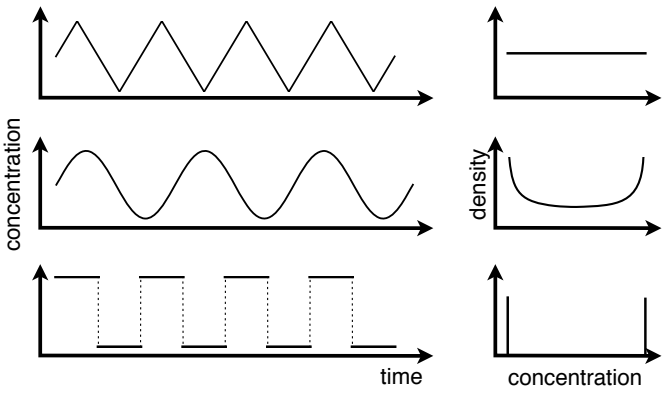

Fig. 1. Protein distributions depend on the functional form of oscillations. Sample time-courses of triangle wave, sinusoidal and step oscillations (left column) along with corresponding time-averaged probability densities ("normalized histograms") of protein concentrations (right).

2.1 Sinusoidal Oscillations Give Rise to Bimodality

We consider an ensemble of cells, each displaying an oscillating level of active protein concentration governed by the intracellular biochemical network dynamics. Cell-to-cell heterogeneity that emerges due to varying gene expression levels induces randomness in the concentration of network components. In an ensemble of oscillating cells, this randomness translates to a distribution of amplitude (A), phase shift (φ), and frequency (ω) of protein activity (y). In order to illustrate the concept of mixing times, we first consider sinusoidal oscillations, $y = A\sin(\omega t+\varphi)$. The three random influences (phase, frequency and amplitude variability) cause qualitatively different behavior with respect to the convergence of the y-distribution in a heterogeneous ensemble of oscillators.

Phase shift variability reflects desynchronization of independent cells within the population and can be quantified in a standard manner, for instance by measuring variance. Narrow distribution of phase shifts compared to the oscillation period, or small desynchronization, implicates that at any point in time the protein levels assumed in the population do not cover the entire range of concentrations. This restricted concentration diversity persists during the time evolution (Fig. 2A, left panel). Stationary distribution emerges instantaneously only when the range of variability uniformly spans the entire oscillation period (Fig. 2A, right panel). In this regime, the probability density function (*pdf*) can be obtained by considering a sine transformation of a uniformly distributed random variable φ restricted to a single oscillation period (cf. Appendix) [6],

$$f(y) = \frac{1}{\pi\sqrt{A^2 - y^2}}\,. \tag{1}$$

The *pdf* is the arcsine distribution (Fig. 2, solid line in *pdf* plots). Notably, it is independent of the time at which the measurement takes place, as well as independent of the frequency of the underlying oscillations.

The variability of frequencies across cellular population stems from intrinsic biochemical noise that affects protein concentrations across the population. Contrary to phase shift variability, an ensemble of sinusoidal oscillators with distributed frequencies reaches the asymptotic stationary distribution regardless of the distribution width; the variance affects only the time to reach it and greater variability accelerates the convergence (Fig. 2B). The asymptotic distribution for uniformly distributed frequencies can be calculated analytically and equals the (previous) result concerning phase-variability (Eq. 1). An intuitive explanation follows from the functional form of the sinusoidal oscillation. The value of random frequency ω is multiplied by time, t. Therefore, regardless of the ω distribution shape, ω is scaled by the increasing time, which accordingly results in the increasing range of frequencies. For large enough t, this range becomes sufficient to facilitate mixing analogous to phase shifts that cover the entire oscillation period.

If cells within the population oscillate with random amplitude only, no stationary distribution can emerge. Since no nonlinear transformation of the random variable takes place, the *pdf* is merely a distribution of the random amplitude A modulated by the sinusoidal wave. The resulting distribution of concentrations cycles over the oscillation period (Fig. 2C).

2.2 Quantification of the Mixing Time

A real biological oscillatory network exhibits a combination of all three types of variability discussed in the previous section. In a typical experimental scenario, the measurement of oscillations is performed on a population of cells and is preceded by a period of starvation followed by addition of a stimulating agent that evokes oscillations. The procedure corresponds to synchronization of cells such that oscillations begin approximately at the same time. Variability among cells still exists, albeit diminished. The emergence of a stationary population-wide bimodal distribution such as the one depicted in Fig. 2D is therefore delayed. The time to approach it, which we shall call the mixing time, depends on the magnitude of contributions to oscillation variability between cells.

A mixing time larger than zero demonstrates a simple fact that the ensemble of oscillators with small variability of frequencies, amplitudes and phase shifts does not immediately reflect time-averaged statistics. As shown in the section above, the system can reach the stationary distribution, provided variability of frequencies exists.

As a quantification of the mixing time we use the Kullback-Leibler divergence (KL), which, in simple terms, measures the divergence of two distributions [13,7]. Let $P(y,t)$ be the probability density of the y concentration at time point t and let further $Q(y)$ be the asymptotic probability density of y for $t \to \infty$, then the KL(t) is defined as

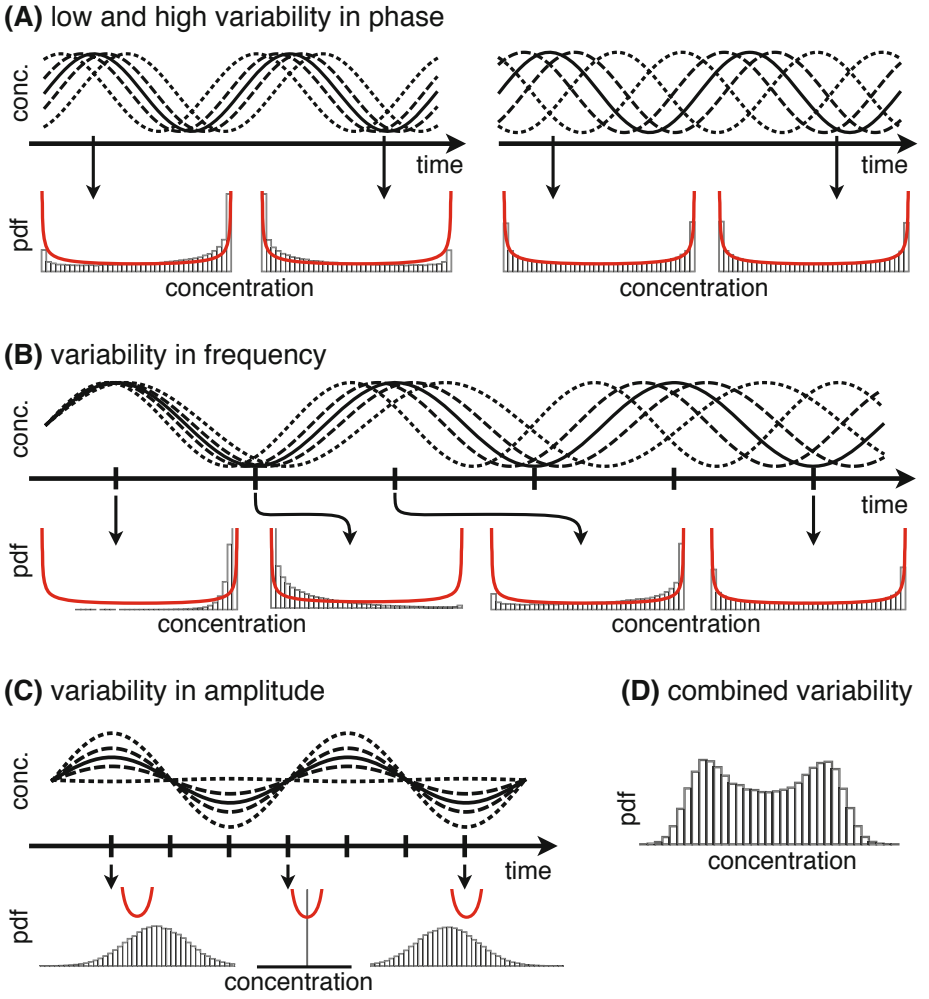

Fig. 2. Approach to an asymptotic distribution. Time courses of oscillations mark the 25^{th} and 75^{th} percentile (dotted), 40^{th} and 60^{th} percentile (dashed), and 50^{th} percentile (solid) of the corresponding parameter distribution. Protein probability density functions (*pdf*) are evaluated numerically at points indicated by arrows. Asymptotic solution, Eq. 1, is marked by the solid line. (A) Phase shifts follow Gaussian distribution with zero mean and standard deviation $\sigma = \pi/2$ (left) and π (right). For large σ, the *pdf* is time-independent and equals the asymptotic *pdf*. (B) Frequency follows log-normal distribution with median 1 and $\sigma = 0.2$. (C) Amplitude follows Gaussian distribution with mean 1 and $\sigma = 1$. The distribution cycles over the oscillation period. (D) Sample stationary protein distribution when all three variability influences are combined.

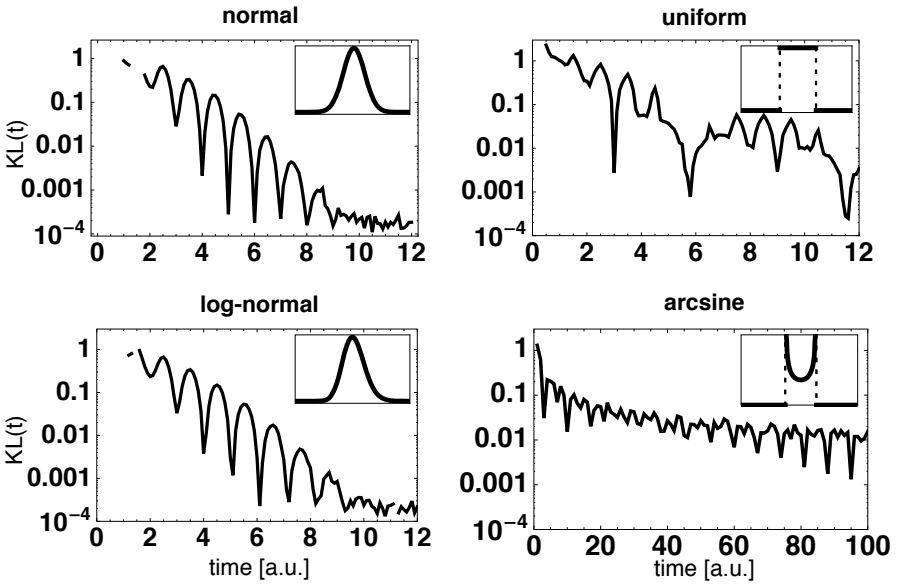

Fig. 3. Kullback-Leibler divergence for a population of sinusoidal oscillators with random frequencies drawn from normal, log-normal, uniform and arcsine distributions (insets). All frequency distributions have the same mean, $\mu = \pi$, and standard deviation, $\sigma = \pi/10$. We calculate KL between numerically sampled protein distributions (based on 100'000 points) at time t and the asymptotic distribution from Eq. 1.

$$KL(t) = \int_{-\infty}^{+\infty} P(y,t) \ln \frac{P(y,t)}{Q(y)} dy \ . \tag{2}$$

Here, $KL(t)$ measures the divergence rather than distance of the snapshot at time t from the asymptotic true snapshot distribution. It is worth emphasizing that KL is always non-negative but it is not a metric in the mathematical sense for it is asymmetric and it does not satisfy triangle inequality.

Temporal behavior of $KL(t)$ is shown in Fig. 3 where we measure the divergence between the protein distribution in an ensemble of oscillators with random frequencies and the asymptotic arcsine distribution. Regardless of the type of frequency distribution, the KL divergence decays at an exponential rate as the oscillating ensemble evolves in time.

2.3 Oscillations and Population Snapshots in a Two-Layer GTPase System

To analyze how oscillations mix in a biologically realistic scenario, we numerically study a model of a generic two-layered GTPase system. Small GTPases can cycle between an inactive GDP-bound state (G) and an active GTP-bound state

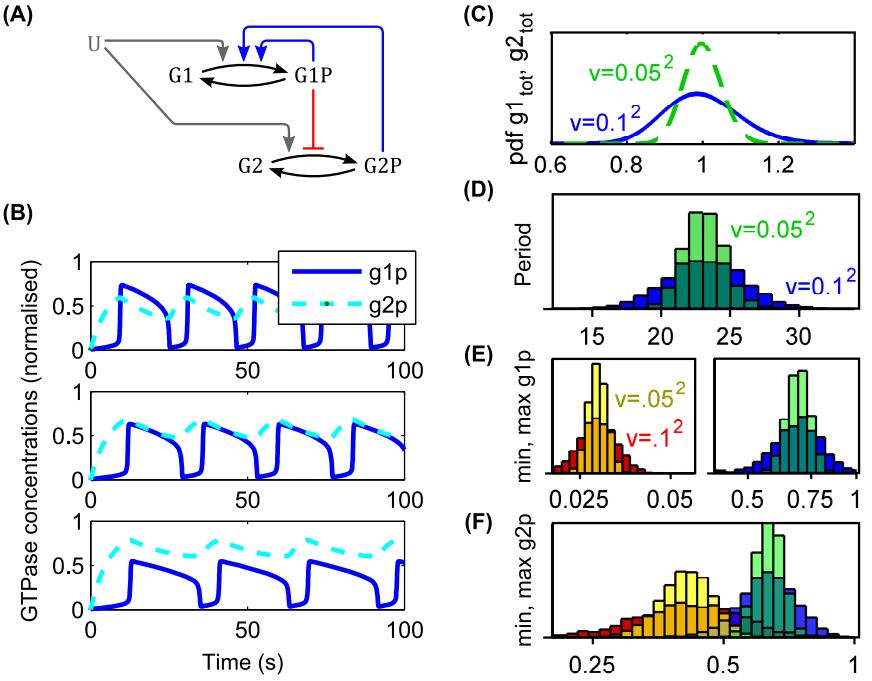

Fig. 4. Dynamic model of GTPase cascade. (A) Interaction scheme. (B) Simulated trajectories of three random cells. (C) Distribution of total GTPase concentrations used for ensemble simulations. The resulting distributions of periods (D), extrema for concentrations of g_1p (E) and g_2p (F). Note the logarithmic x-axis in panels E and F. Parameters used in the simulation: maximal rates $r_1 = 10$, $r_2 = 6.5$, $r_3 = 1$, $r_4 = 0.55$; half-activation constants $m_1 = 25$, $m_2 = 0.09$, $m_3 = 5$, $m_4 = 14$; positive interaction $G1 \to G1$, $a_{11} = 200$, $m_{11} = 10$; negative interaction $G1 \dashv G2$, $a_{12} = 0.005$, $m_{13} = 0.05$; positive interaction $G2 \to G1$, $a_{21} = 80$, $m_{21} = 20$.

(GP). They are important transducers of cell signaling that regulate a wide range of biological processes, for instance cell proliferation, cell morphology as well as nuclear and vesicle transport. Individual GTPase are often interlinked, thereby generating positive and negative feedback systems that are theoretically capable of exhibiting rich dynamics including oscillations. Indeed oscillations have been observed experimentally for several GTPases. For example, the small Rho-GTPase cdc42 regulates polarized growth in fission yeast using oscillating activity arising from both positive and negative feedback [4].

We consider the GTPase cascade depicted in Fig. 4A and let G1, G2 and G1P, G2P denote the inactive and active form of the corresponding GTPase, respectively. The system features a positive auto-regulatory loop in which G1P enhances its own activation and a negative feedback loop in which G1P inhibits

the activation of G2 and in turn G2P activates G1. The following ordinary differential equations in normalized coordinates represent the system [16],

$$\frac{d}{dt}g_1p = \alpha_{11}\alpha_{21}\frac{r_1\, g_1}{m_1 + g_1} - \frac{r_2\, g_1p}{m_2 + g_1p}\,, \qquad \alpha_{11} = \frac{m_{11} + a_{11}\, g_1p}{m_{11} + g_1p}\,,$$
$$\alpha_{21} = \frac{m_{21} + a_{21}\, g_2p}{m_{21} + g_2p}\,,$$
$$\frac{d}{dt}g_2p = \alpha_{13}\frac{r_3\, g_2}{m_3 + g_2} - \frac{r_4\, g_2p}{m_4 + g_2p}\,, \qquad \alpha_{13} = \frac{m_{13} + a_{13}\, g_1p}{m_{13} + g_1p}\,, \qquad (3)$$

with $g_1 = g_1^{tot} - g_1p$, $g_2 = g_2^{tot} - g_2p$, and where g_1, g_2 and g_1p, g_2p denote the concentrations of inactive and active GTPases, respectively, and r_i, m_i, a_{ij}, m_{ij} are kinetic parameters. The factors α_{11}, α_{21}, α_{13} model the described interactions with the parameters $a_{11} > 1$, $a_{21} > 1$ (positive interactions) and $0 < a_{13} < 1$ (negative interaction).

In accordance with the literature [12], we model a population of cells as an ensemble [10,5] of single cells in which the total concentrations of both GTPases are log-normally distributed with mean one and standard deviations consistent with experimentally reported values ranging from 0.12 to 0.28 in human cells [14] (Fig. 4C).

Fig. 4B illustrates the dynamics of the model with representative responses of three random cells to a step input. The model exhibits switch-like G1P oscillations and triangle wave-like G2P oscillations, thus providing a convenient tool to investigate how differentially shaped oscillations manifest in the distribution of a population snapshots taken at a particular time point. Figs. 4D-F demonstrate how the periods and the minima and maxima of the oscillations are distributed in the population. Decreasing the variability of the total GTPase distribution (from $\sigma = 0.1$ to 0.05) yields more narrowly distributed periods and extrema while their means remain unchanged.

The distribution of $g_1p(t)$ and $g_2p(t)$ concentrations, in the following referred to as a distribution snapshot, changes over time. For $t < 0$ the entire population is synchronized; the phase of all oscillations is zero and the first peak occurs roughly at the same time; at the initial time all cells exhibit zero GTPase activity, while after 15 seconds most cells have progressed to the first peak. Over time, the cell-to-cell variability has an increasing effect on the population snapshots; different periods shift the phases of subsequent peaks until the phases are uniformly distributed. During this transition period, the snapshot distribution dynamically changes (Fig. 5). The evolution of the distribution crucially depends on the shape of the underlying oscillations. For example, switch-like G1P oscillations result in uni-modal ($t = 5\,\mathrm{s}$), bi-modal ($t = 9\,\mathrm{s}$) and even tri-modal ($t = 78\,\mathrm{s}$) distributions. In contrast, the triangle wave G2P oscillations yield uni-modal snapshot distributions at all times (Fig. 5B).

Next we sought to assess how quickly the snapshot distribution converges to the asymptotic one using Kullback-Leibler divergence and asked whether it is possible to find the time point at which the oscillations are well mixed. The results are shown in Fig. 5C and D. The snapshot distribution converges exponentially

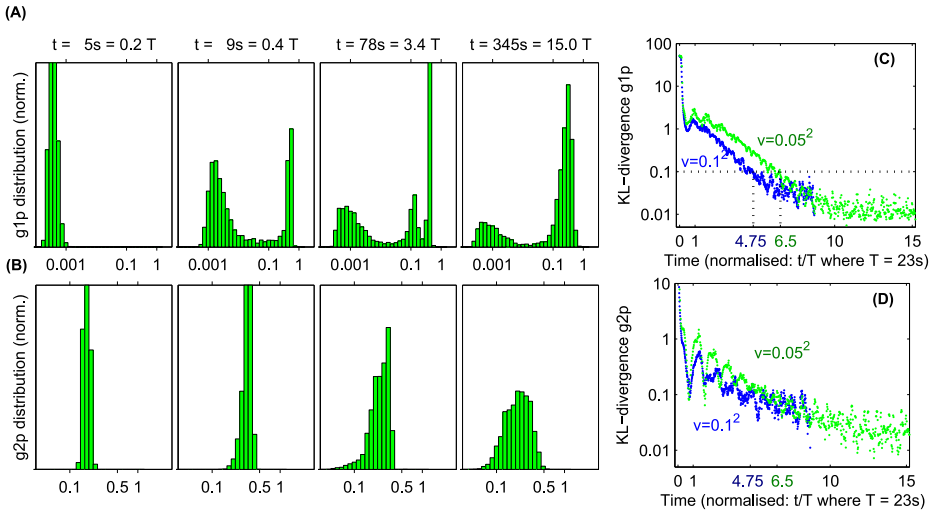

Fig. 5. Numerical simulations of time evolution of g_1p (A) and g_2p (B) protein distribution for indicated time points. The distribution of total G1 and G2 concentrations is normal with mean 1 and standard deviation 0.05. Panels C and D show a comparison of Kullback-Leibler divergence for total G1 and G2 distributions with standard deviation 0.1 and 0.05. Simulation parameters same as described in caption of Fig. 4.

towards its asymptotic distribution. Further, the rate of convergence depends on the cell-to-cell variability in the population; lower variability of the total GTPase concentrations causes lower variability of periods and results in longer mixing times.

3 Discussion

A stationary bimodal protein distribution may arise in a heterogeneous population of independent cells with sustained deterministic oscillations of active protein levels. The emergence of population-level bimodality is inevitable in the presence of cell-to-cell variability that affects oscillation frequency. Importantly, the type of the frequency distribution across the population is irrelevant for the emergence of bimodality; only the time of convergence is affected (Fig. 3).

Detecting the oscillatory nature of signaling networks with bulk measurements (e.g. immunoblotting) is only possible for synchronized cells. This synchronization, for instance, occurs at the point of stimulation preceded by a period of starvation. If this condition is not satisfied, cell-to-cell variability introduces phase desynchronization and diversity in frequencies, which averages out the oscillations. On top of that, sampling frequency in the experiment should be sufficiently higher than frequency of the oscillations. Otherwise the measurement captures only the population mean, which does not oscillate. In this regime, single-cell measurement methods can give an additional insight, for they record the amount of protein in individual cells.

A population snapshot obtained with flow cytometry or imaging is equivalent to the probability density functions discussed throughout the paper. The time dependence of such a distribution may become very intricate for realistic systems, which we demonstrated in Fig. 5. Nonetheless, there exists a time scale – the mixing time, which we estimated using Kullback-Leibler divergence – when the stationary distribution emerges. If amplitude variability is smaller than the oscillation amplitude itself (cf. g_1p and g_2p oscillations in Fig. 4B), the distribution may become bimodal. The time to converge to this distribution is independent of the population size: as long as mixing of frequencies and phases in a population has not commenced, the protein distribution remains different from the asymptotic one. This behavior contrasts with other mechanisms that generate bimodality where increasing the number of independent individuals results in a better indication of the population-wide asymptotic distribution. Experimental verification of sources of bimodality might benefit from this feature.

The second implication of our finding relates to mechanisms that preserve synchronized population-wide oscillations generated by biochemical networks. In the presence of frequency heterogeneity the mixing and eventual convergence to the asymptotic distribution is only a matter of time. If oscillations are a physiologically relevant trait, as is the case of circadian rhythms, the convergence is undesirable because it would indicate that oscillations are out of sync and each cell within an organ, for instance, has its own day and night pattern. This could explain why biochemical coupling is present in such systems in order to facilitate spontaneous synchronization across the population.

Acknowledgments. MD, LKN and BNK were supported by Science Foundation Ireland under Grant No. 06/CE/B1129. DF received funding from the European Union Seventh Framework Programme (FP7/2007-2013) ASSET project under grant agreement number FP7-HEALTH-2010-259348.

References

1. Acar, M., Mettetal, J.T., van Oudenaarden, A.: Stochastic switching as a survival strategy in fluctuating environments. Nat. Genet. 40(4), 471–475 (2008)
2. Ackermann, M., Stecher, B., Freed, N.E., Songhet, P., Hardt, W.D., Doebeli, M.: Self-destructive cooperation mediated by phenotypic noise. Nature 454(7207), 987–990 (2008)
3. Blake, W.J., Balázsi, G., Kohanski, M.A., Isaacs, F.J., Murphy, K.F., Kuang, Y., Cantor, C.R., Walt, D.R., Collins, J.J.: Phenotypic consequences of promoter-mediated transcriptional noise. Molecular Cell 24(6), 853–865 (2006)
4. Das, M., Drake, T., Wiley, D.J., Buchwald, P., Vavylonis, D., Verde, F.: Oscillatory dynamics of Cdc42 GTPase in the control of polarized growth. Science 337(6091), 239–243 (2012)
5. Kuepfer, L., Peter, M., Sauer, U., Stelling, J.: Ensemble modeling for analysis of cell signaling dynamics. Nat. Biotechnol. 25(9), 1001–1006 (2007)
6. Miller, S., Childers, D.: Probability and Random Processes. With Applications to Signal Processing and Communications. Academic Press (January 2012)

7. Mirsky, H.P., Taylor, S.R., Harvey, R.A., Stelling, J., Doyle, F.J.: Distribution-based sensitivity metric for highly variable biochemical systems. IET Syst. Biol. 5(1), 50 (2011)
8. Niepel, M., Spencer, S.L., Sorger, P.K.: Non-genetic cell-to-cell variability and the consequences for pharmacology. Curr. Opin. Chem. Biol. 13(5-6), 556–561 (2009)
9. Ochab-Marcinek, A., Tabaka, M.: Bimodal gene expression in noncooperative regulatory systems. PNAS 107(51), 22096–22101 (2010)
10. Ogunnaike, B.A.: Elucidating the digital control mechanism for DNA damage repair with the p53-Mdm2 system: single cell data analysis and ensemble modelling. J. R. Soc. Interface 3(6), 175–184 (2006)
11. Samoilov, M., Plyasunov, S., Arkin, A.P.: Stochastic amplification and signaling in enzymatic futile cycles through noise-induced bistability with oscillations. PNAS 102(7), 2310–2315 (2005)
12. Schliemann, M., Bullinger, E., Borchers, S., Allgöwer, F., Findeisen, R., Scheurich, P.: Heterogeneity reduces sensitivity of cell death for TNF-stimuli. BMC Syst. Biol. 5, 204 (2011)
13. Shahrezaei, V., Swain, P.S.: Analytical distributions for stochastic gene expression. PNAS 105(45), 17256–17261 (2008)
14. Sigal, A., Milo, R., Cohen, A., Geva-Zatorsky, N., Klein, Y., Liron, Y., Rosenfeld, N., Danon, T., Perzov, N., Alon, U.: Variability and memory of protein levels in human cells. Nature 444(7119), 643–646 (2006)
15. Spencer, S.L., Gaudet, S., Albeck, J.G., Burke, J.M., Sorger, P.K.: Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. Nature 459(7245), 428–432 (2009)
16. Tsyganov, M.A., Kolch, W., Kholodenko, B.N.: The topology design principles that determine the spatiotemporal dynamics of G-protein cascades. Mol. Biosyst. 8(3), 730–743 (2012)

Appendix

Asymptotic Protein Distribution

Consider an ensemble of cells with sinusoidal oscillations of protein levels. Variability of phases φ is accounted for by a random variable Φ uniformly distributed on the range of an oscillation period, 2π. We set out to obtain a distribution of protein levels y denoted by a random variable Y, which is the result of a nonlinear transformation $Y = A\sin(\omega t + \Phi)$. Since the transformation is periodic, without loss of generality we first set $t = 0$ and focus on the shorter range, $-\pi/2 < \Phi < \pi/2$, where sine function is monotonically increasing.

The cumulative distribution function (CDF) of Y is simply the probability that the random variable Y takes on a value less than or equal y, $F_Y(y) = Pr(Y \leq y)$. From this we obtain,

$$F_Y(y) = Pr\left[A\sin(\Phi) \leq y\right] = Pr\left[\Phi \leq \arcsin\left(\frac{y}{A}\right)\right], \text{ and } |y| \leq A\,. \tag{4}$$

The CDF of Y can be therefore expressed in terms of the CDF of Φ. The probability density function (*pdf*), denoted by $f_Y(y)$, is CDF's first derivative,

$$\begin{aligned} F_Y(y) &= F_\Phi\left(\arcsin\left(\frac{y}{A}\right)\right)\,, \\ \frac{d}{dy}F_Y(y) &= f_Y(y) = f_\Phi\left(\arcsin\left(\frac{y}{A}\right)\right)\frac{1}{\sqrt{A^2+y^2}} \\ &= \frac{1}{\pi}\,\frac{1}{\sqrt{A^2+y^2}}\,. \end{aligned} \tag{5}$$

The procedure can be repeated to yield the same result, the arcsine distribution, for every range of length π, where the transformation is monotonic.