LETTERS

Unlimited multistability in multisite phosphorylation systems

Matthew Thomson[1] & Jeremy Gunawardena[2]

Reversible phosphorylation on serine, threonine and tyrosine is the most widely studied posttranslational modification of proteins[1,2]. The number of phosphorylated sites on a protein (n) shows a significant increase from prokaryotes, with $n \leq 7$ sites, to eukaryotes, with examples having $n \geq 150$ sites[3]. Multisite phosphorylation has many roles[4,5] and site conservation indicates that increasing numbers of sites cannot be due merely to promiscuous phosphorylation. A substrate with n sites has an exponential number (2^n) of phospho-forms and individual phospho-forms may have distinct biological effects[6,7]. The distribution of these phospho-forms and how this distribution is regulated have remained unknown. Here we show that, when kinase and phosphatase act in opposition on a multisite substrate, the system can exhibit distinct stable phospho-form distributions at steady state and that the maximum number of such distributions increases with n. Whereas some stable distributions are focused on a single phospho-form, others are more diffuse, giving the phosphoproteome the potential to behave as a fluid regulatory network able to encode information and flexibly respond to varying demands. Such plasticity may underlie complex information processing in eukaryotic cells[8] and suggests a functional advantage in having many sites. Our results follow from the unusual geometry of the steady-state phospho-form concentrations, which we show to constitute a rational algebraic curve, irrespective of n. We thereby reduce the complexity of calculating steady states from simulating 3×2^n differential equations to solving two algebraic equations, while treating parameters symbolically. We anticipate that these methods can be extended to systems with multiple substrates and multiple enzymes catalysing different modifications, as found in posttranslational modification 'codes'[9] such as the histone code[10,11]. Whereas simulations struggle with exponentially increasing molecular complexity, mathematical methods of the kind developed here can provide a new language in which to articulate the principles of cellular information processing[12].

A major difficulty in studying multisite phosphorylation from a systems perspective has been the lack of information regarding when different sites on the same protein are simultaneously phosphorylated[13]. Sites may be intricately dependent on each other[14] and both the number and the position of phosphates can affect biological outcome[6,7]. The phospho-form distribution—the relative stoichiometries of each of the 2^n phospho-forms—is thus the appropriate measure of phosphorylation state for a multisite substrate. Such distributions are starting to be measured[15,16], prompted by interest in posttranslational modification (PTM) codes. A theoretical understanding will improve our ability to interpret such data, uncover biological principles and design appropriate experiments. Although we focus on phosphorylation, we hope to lay a foundation for analysing multiple PTMs.

Figure 1 summarizes a general model of multisite phosphorylation. A substrate S, with n sites, is acted on by a kinase E and a phosphatase F. Substrates may have multiple kinases and phosphatases *in vivo* but a single enzyme often addresses multiple sites and we focus here on the minimal enzymatic machinery needed for any n. Each enzyme may act distributively (Fig. 1a) as well as processively (Fig. 1b) using a standard biochemical scheme (Fig. 1c) and arbitrary preferences for site order. These assumptions are more general than in previous models[17–22]: by choosing the details appropriately, any kinase, phosphatase and substrate system can be represented. ATP is assumed to be recharged by some process that is not modelled. It is therefore not treated as a variable but its effect is absorbed into the site-specific parameters (a_u^X, b_u^X, $c_{u,v}^X$ defined in Fig. 1c). Phosphorylation and dephosphorylation are assumed to take place on a fast time scale in comparison to synthesis and degradation of the component proteins. The model is, therefore, effectively closed: there is no flux of material through it and the total amounts of substrate, S_{tot}, and enzymes, E_{tot} and F_{tot}, remain constant at all times. With mass-action kinetics, these assumptions give rise to 3×2^n nonlinear differential equations for the state variables (Fig. 1d).

With limited information on site-specific parameters, numerical simulations can be undertaken for randomly selected parameter values in an attempt to discern typical behaviours. However, doing so in a state space of dimension 3×2^n rapidly becomes intractable as n increases. Here we introduce a new method of analysis, which allows strong conclusions to be drawn about steady states without having to specify parameter values in advance. Experimental evidence indicates that biological systems attain quasi-steady states *in vivo*[23,24], including systems in which multisite phosphorylation has a significant role[25]. While the steady-state assumption must be confirmed in each experimental context, it has been widely used in modelling multisite phosphorylation[17–22].

At steady state, the second equation in Fig. 1d yields an expression for XS_u in terms of X and S_u. Substituting this into the first equation in Fig. 1d makes that equation linear in S_u. The coefficients are algebraic expressions in the site-specific parameters (collectively denoted $\mathbf{a}$) and the auxiliary parameter $t = E/F$, the steady-state ratio of free kinase to free phosphatase. These expressions may be regarded as elements of a set of coefficients, $\mathbb{R}(\mathbf{a}, t)$, in which the $\mathbf{a}$ values and the t have been adjoined to the ordinary numbers, $\mathbb{R}$, as uninterpreted symbols that can be added, subtracted, multiplied and divided as if they were numbers (see the Supplementary Information for more details). The elements of $\mathbb{R}(\mathbf{a}, t)$ correspond to rational functions, or ratios of polynomials, in these parameters (Supplementary Information). By treating the parameters symbolically in this way, they can be used in calculations without their numerical values being known in advance.

The linearized equations for the S_u can be solved by Gaussian elimination, which works as well over the coefficients $\mathbb{R}(\mathbf{a}, t)$ as over $\mathbb{R}$ (Supplementary Information). The steady-state phospho-forms can thereby be shown to satisfy (Supplementary Information)

[1]Biophysics Program, Harvard University, Cambridge, Massachusetts 02138, USA. [2]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115, USA.

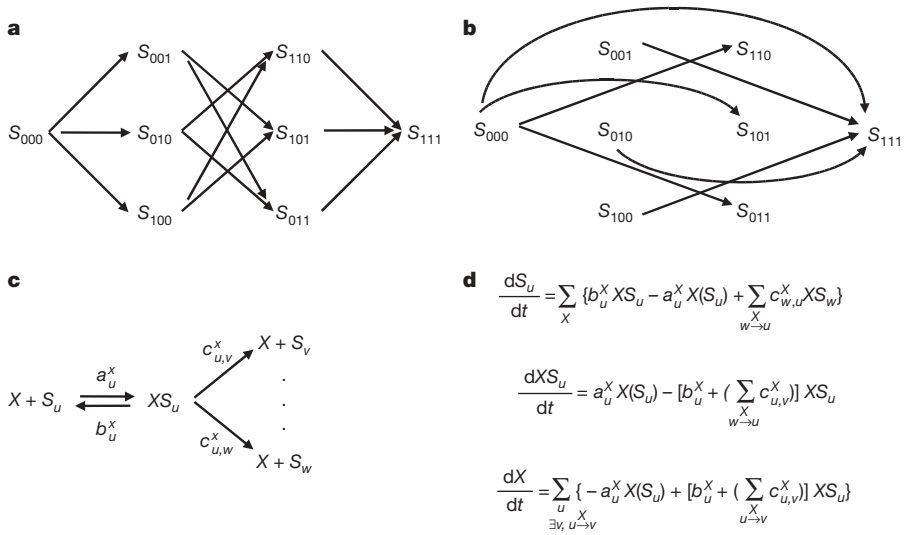

Figure 1 | General model of multisite phosphorylation with substrate *S* having *n* sites, kinase *E* and phosphatase *F*. The phospho-forms are denoted S_u, where u is a bit string indicating the absence/presence (0/1) of phosphate. Kinase reactions for $n = 3$ with one phosphorylation per reaction (**a**, distributivity[29]) and multiple phosphorylations per reaction (**b**, processivity[29]). Phosphatases act similarly in removing phosphates. **c**, Each enzyme ($X = E$ or F) uses a standard biochemical mechanism[30] but may form multiple products, with associated parameters (a_u^X, b_u^X, $c_{u,v}^X$) for mass-action kinetics. ATP is assumed to be held constant and synthesis and degradation are ignored. **d**, The 3×2^n differential equations, where $u \xrightarrow{X} v$ signifies that X converts S_u to S_v. The same symbol is used for a chemical species and for its concentration. Note that $X(S_u)$ here indicates the product of X and S_u.

$$S_u = S_0 r_u(t) \quad (1)$$

where S_0 is the unphosphorylated phospho-form and $r_u(t)$ is a rational function of t with coefficients in $\mathbb{R}(\mathbf{a})$. Although these rational functions are complex, they can be explicitly calculated for any given model (Supplementary Information).

If numerical values are to be given to the site-specific parameters, it is important to know that the rational functions $r_u(t)$ remain well defined. For instance, the rational function $c/(1-c)$ becomes undefined when $c = 1$. We show that, for any positive site-specific parameter values and positive t, $r_u(t)$ is always well defined and positive ('positivity', Supplementary Information), ruling out such problems.

Equation (1) implies that the steady-state phospho-forms can be described through a single auxiliary variable, t, so that they form a one-dimensional geometric object, or 'curve'. Despite increasing numbers of sites, and the exponentially increasing size of the model, the steady-state phospho-forms always remain a curve, providing the basis for an exponential reduction in complexity. What changes with n is the extent to which the curve undulates, which determines how many steady states can co-exist for given amounts of substrate and enzymes (see below). Not all curves can be described by rational functions; those that can are of considerable geometric interest, as explored in an earlier paper[26] and discussed further in the Supplementary Information.

The upshot of equation (1) is that, at steady state, the 3×2^n state variables are determined by S_0, E and F. Because the substrate and enzyme totals remain constant, we can formulate three equations for these three unknowns. The equation for S_0 can be solved directly in terms of S_{tot}, leaving a pair of equations (defined in the Supplementary Information):

$$\Phi_1(E, F) = E_{\mathrm{tot}} \quad \text{and} \quad \Phi_2(E, F) = F_{\mathrm{tot}} \quad (2)$$

which determine the steady-state E and F values corresponding to any given substrate and enzyme totals. Equation (2) exactly characterizes the steady states of the model (Supplementary Information). To find steady states, it is no longer necessary to numerically simulate 3×2^n differential equations; this can be done by only solving two algebraic equations. The complexity arising from the dynamics has been distilled away. This exponential reduction of complexity is the key to what follows.

Figure 2a shows an example with four sites and five steady states. (We assume sequentiality, as in Fig. 3a, but merely for convenience.) Only stable states are detected experimentally, or found by numerical simulation, and this example is tristable; the corresponding stable phospho-form distributions are markedly distinct. Whereas distributions 1 and 3 are each focused on a single phospho-form, distribution 2 is broader. With multiple stable states a system can encode many outcomes, or several bits of information, enabling complex information encoding and processing[8]. Such multistability is believed to underlie cellular differentiation and other decisions[8,27] but experimental examples have, so far, only demonstrated bistability[23–25]. Bistability in two-site phosphorylation was previously shown by modelling[20].

It remained unknown how the number of stable states depends on n. We found that multistability tends to occur when substrate is in excess over enzymes (as in Fig. 2a). In this approximation, the two equations for E and F in equation (2) can be reduced to one polynomial equation for $t = E/F$ (Supplementary Information):

$$P(t) = \alpha_N t^N + \alpha_{N-1} t^{N-1} + \cdots + \alpha_1 t + \alpha_0 = 0 \quad (3)$$

where $\alpha_i \in \mathbb{R}(\mathbf{a})$ and N lies between $n + 1$ and 2^n depending on the model. Positivity of the $r_u(t)$ is essential here. Positive solutions of $P(t) = 0$ correspond approximately to steady states. We show that the discrepancy between the exact steady states found by equation (2) and the approximate ones found by equation (3) can be made as small as desired by increasing the excess of S_{tot} over E_{tot} and F_{tot} (Supplementary Information).

The advantage of equation (3) over equation (2) is that we can readily construct solutions of the former. If n is even and we choose any $n + 1$ distinct positive numbers, then we can always find a model for which the corresponding $P(t)$ has these numbers as solutions (Supplementary Information). Moreover, S_{tot} can be chosen arbitrarily, so that the approximation of equation (2) by equation (3) is as accurate as desired. It follows that the corresponding model has $n + 1$ steady states. The example in Fig. 2a meets this bound for $n = 4$. If n is odd the same can be achieved for any set of n distinct positive numbers (Supplementary Information). Note that odd numbers of steady states are constructed in both cases.

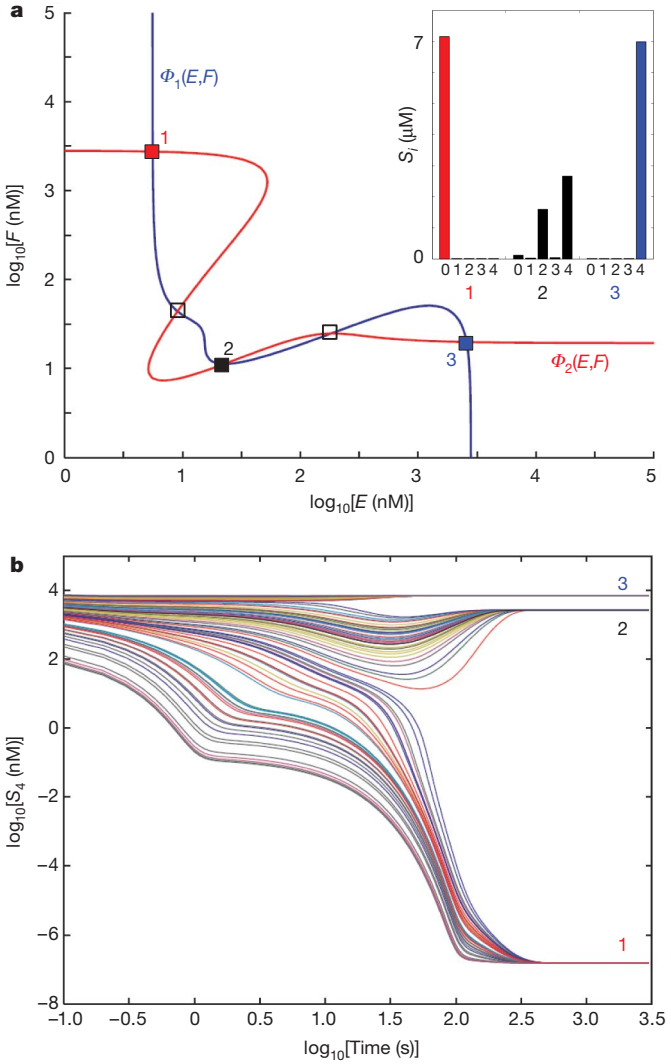

Figure 2 | Multistability for an $n = 4$ distributive, sequential system, as in Fig. 3a. For parameter values, see Supplementary Table 1. **a**, Plots of equation (2) for $S_{tot} = 10\,\mu M$ and $E_{tot} = F_{tot} = 2.8\,\mu M$. The intersections correspond to the steady states. Filled squares are stable: 1 (red), 2 (black) and 3 (blue); open squares are unstable. Stability was determined by standard methods (Supplementary Information). The inset shows the corresponding phospho-form distributions, following the notation in Fig. 3a. **b**, Time courses of S_4 reaching its three stable values from initial conditions $S_0 = \alpha S_{tot}$, $S_4 = (1-\alpha)S_{tot}$ and $X = 0$ for all other variables, with α chosen randomly from the uniform distribution on [0, 1] (100 samples), determined by simulation (Supplementary Information).

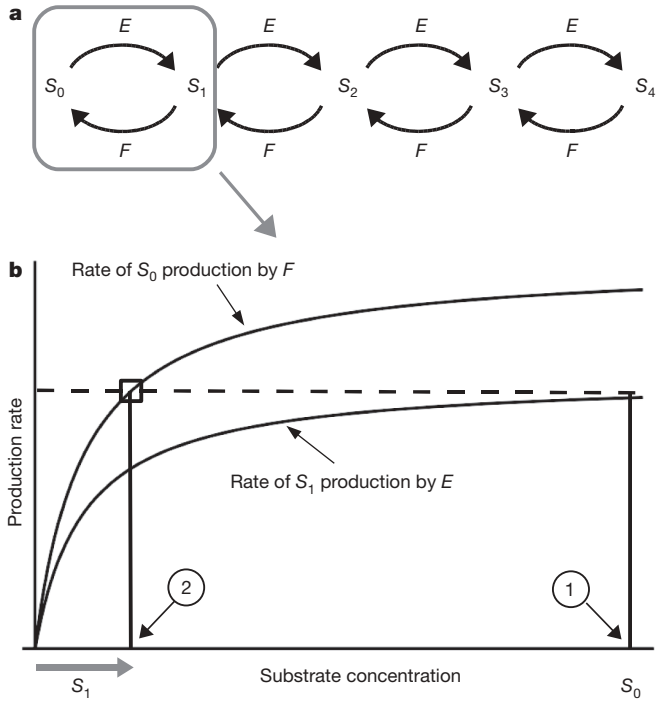

Figure 3 | Multistability by kinetic trapping. **a**, A distributive, sequential system with $n = 4$; E phosphorylates in order, F dephosphorylates in reverse order. The phospho-forms are denoted S_i, $i = 0,\ldots, 4$, where i is the number of phosphates. **b**, Rate functions for production of S_1 from S_0 by E and of S_0 from S_1 by F are approximated as Michaelis–Menten hyperbolas[30], with the phosphatase curve to the left and above the kinase curve. If the system is initiated with substrate entirely in S_0 and in excess over both enzymes and saturating them, S_0 will sequester E and produce S_1 at nearly maximal rate (point 1). F, however, will be unoccupied, so that as S_1 increases (grey arrow), phosphorylation and dephosphorylation will balance (point 2). Any leak of S_1 into S_2 can be opposed by F, which is not sequestered. The system will hence reach steady state with substrate remaining predominantly unphosphorylated. If similar conditions are applied to S_3 and S_4, but reversed with respect to E and F, then S_4 can be similarly trapped. The parameter values required for this argument are discussed in the Supplementary Information.

This is also what we find in simulations. Furthermore, if the steady states are ordered by their corresponding E/F values then unstable states always occur between stable ones (Supplementary Information), as in Fig. 2a. Hence, if there are $2k + 1$ steady states, $k + 1$ of them are stable. It follows that a model with n sites can have as many as $\lfloor (n+2)/2 \rfloor$ stable states ($\lfloor x \rfloor$ being the greatest integer not greater than x). We see that the tristability in Fig. 2a is only the tip of the iceberg: the maximum number of stable states increases with increasing numbers of sites.

Experimental detection of stable states requires an understanding of how they arise dynamically. Figure 3b considers the inter-conversion of S_0 and S_1, for the sequential system in Fig. 3a. An informal argument based on the Michaelis–Menten approximation suggests that, for suitable parameter values, if the system is started entirely in S_0 then substrate can remain trapped predominantly in that state, as in distribution 1 in Fig. 2a. A similar argument applies to trapping of S_4, as in distribution 3. Simulations show that all three stable distributions can be reached by starting from suitable mixtures of S_0 and S_4 (Fig. 2b).

In vivo, the most likely way that a phosphorylation system is regulated is by modulating its enzymes. Because the enzymes have been assumed to be in their active states, this corresponds to altering E_{tot} or F_{tot}. Figure 4a shows, for the tristable system in Fig. 2, that changes in E_{tot} can switch the system between stable states 1 and 3. It is possible that more complex modulations, involving both E_{tot} and F_{tot}, could access all three stable states or that different parameter values could facilitate additional switching capability. Figure 4b shows another option, in which changes in E_{tot} can switch between three stable states even though there is only a narrow window of E_{tot} values for which three stable states coexist. In other words, the system may not need to exhibit robust tristability to have access to three stable states.

DNA provides a static, structural mechanism for encoding information at a capacity of 2 bits per base pair. The idea that PTMs provide a dynamic mechanism for information encoding has been broadly influential[10,11] but the mechanistic details remain a matter of debate[9,28] and no estimate of information capacities has emerged for any such code. Our result provides the first demonstration of a PTM mechanism that can, in principle, encode an arbitrary amount of information, along with an estimate of its information capacity. If natural selection has found such a capability useful, that may help account for the emergence of large numbers of phosphorylation sites.

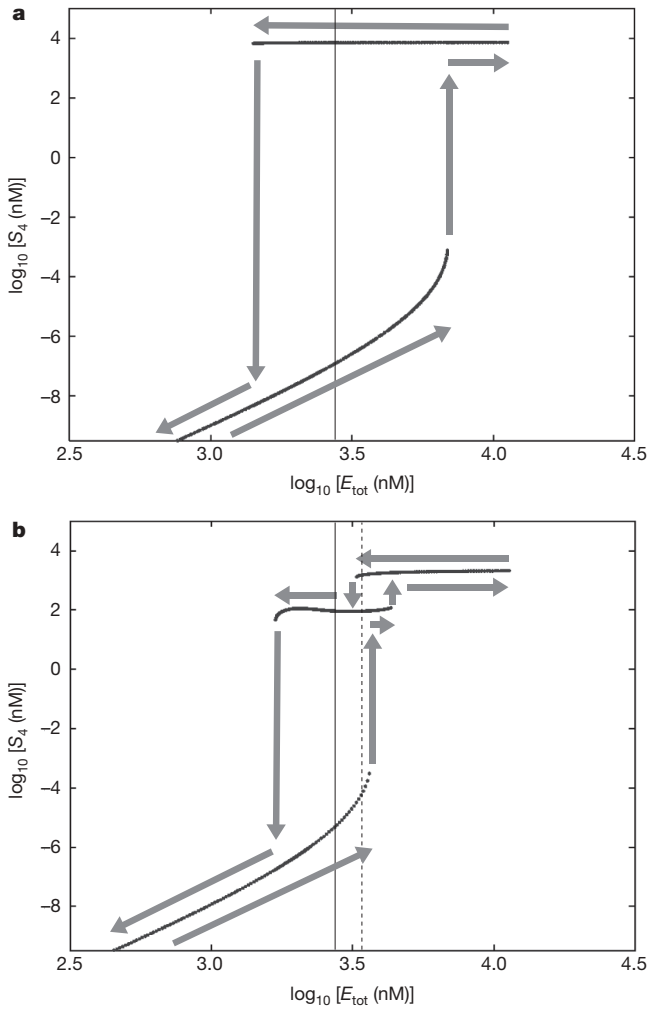

Figure 4 | Switching between stable states. **a**, The system in Fig. 2a with $E_{tot} = F_{tot} = 2.8\ \mu M$ and $S_{tot} = 10\ \mu M$ is taken in a cycle (grey arrows) using simulation (Supplementary Information). The free kinase is repeatedly changed by a small amount and the system allowed to relax back to a stable state. Starting on the lower branch, corresponding to distribution 1 in Fig. 2a, the system switches abruptly to the higher branch (distribution 3), remains on that branch as E_{tot} is lowered, and then switches abruptly back down to the lower branch (hysteresis). **b**, The system with $E_{tot} = F_{tot} = 2.8\ \mu M$, $S_{tot} = 5\ \mu M$ has only two stable states (not shown) but occupies three when E_{tot} is cycled. Three stable states only coexist in a narrow window around the dotted line. The solid lines mark $E_{tot} = 2.8\ \mu M$.

Received 10 November 2008; accepted 24 April 2009.
Published online 17 June 2009.

1. Walsh, C. T. *Posttranslational Modification of Proteins* (Roberts and Company, 2006).
2. Cohen, P. The role of reversible protein phosphorylation in health and disease. *Eur. J. Biochem.* **268,** 5001–5010 (2001).
3. Gnad, F. *et al.* PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. *Genome Biol.* **8,** R250 (2007).
4. Cohen, P. The regulation of protein function by multisite phosphorylation — a 25 year update. *Trends Biochem. Sci.* **25,** 596–601 (2000).
5. Holmberg, C. I., Tran, S. E. F., Eriksson, J. E. & Sistonen, L. Multisite phosphorylation provides sophisticated regulation of transcription factors. *Trends Biochem. Sci.* **27,** 619–627 (2002).
6. Wu, R. C. *et al.* Selective phosphorylations of the SRC-3/AIB1 coactivator integrate genomic responses to multiple cellular signaling pathways. *Mol. Cell* **15,** 937–949 (2004).
7. Park, K.-S., Mohapatra, D. P., Misonou, H. & Trimmer, J. S. Graded regulation of the Kv2.1 potassium channel by variable phosphorylation. *Science* **313,** 976–979 (2006).
8. Nurse, P. Life, logic and information. *Nature* **454,** 424–426 (2008).
9. Sims, R. J. & Reinberg, D. Is there a code embedded in proteins that is based on posttranslational modification? *Nature Rev. Mol. Cell Biol.* **9,** 815–820 (2008).
10. Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293,** 1074–1080 (2001).
11. Turner, B. Cellular memory and the histone code. *Cell* **111,** 285–291 (2002).
12. Cohen, J. E. Mathematics is biology's next microscope, only better; biology is mathematics' next physics, only better. *PLoS Biol.* **2,** e439 (2004).
13. Hunter, T. The age of crosstalk: phosphorylation, ubiquitination and beyond. *Mol. Cell* **28,** 730–738 (2007).
14. Ferrarese, A. *et al.* Chemical dissection of the APC repeat 3 multistep phosphorylation by the concerted action of protein kinases CK1 and GSK3. *Biochemistry* **46,** 11902–11910 (2007).
15. Phanstiel, D. *et al.* Mass spectrometry identifies and quantifies 74 unique histone H4 isoforms in differentiating human embryonic stem cells. *Proc. Natl Acad. Sci. USA* **105,** 4093–4098 (2008).
16. Pesavento, J. J., Bullock, C. R., LeDuc, R. D., Mizzen, C. A. & Kelleher, N. L. Combinatorial modification of human histone H4 quantitated by two-dimensional liquid chromatography coupled with top down mass spectrometry. *J. Biol. Chem.* **283,** 14927–14937 (2008).
17. Goldbeter, A. & Koshland, D. E. An amplified sensitivity arising from covalent modification in biological systems. *Proc. Natl Acad. Sci. USA* **78,** 6840–6844 (1981).
18. Lisman, J. E. A mechanism for memory storage insensitive to molecular turnover: a bistable autophosphorylating kinase. *Proc. Natl Acad. Sci. USA* **82,** 3055–3057 (1985).
19. Salazar, C. & Höfer, T. Allosteric regulation of the transcription factor NFAT1 by multiple phosphorylation sites: a mathematical analysis. *J. Mol. Biol.* **327,** 31–45 (2003).
20. Markevich, N. I., Hoek, J. B. & Kholodenko, B. N. Signalling switches and bistability arising from multisite phosphorylation in protein kinase cascades. *J. Cell Biol.* **164,** 353–359 (2004).
21. Gunawardena, J. Multisite protein phosphorylation makes a good threshold but can be a poor switch. *Proc. Natl Acad. Sci. USA* **102,** 14617–14622 (2005).
22. Kim, S. Y. & Ferrell, J. E. Substrate competition as a source of ultrasensitivity in the inactivation of Wee1. *Cell* **128,** 1133–1145 (2007).
23. Ozbudak, E. M., Thattai, M., Lim, H. N., Shraiman, B. I. & van Oudenaarden, A. Multistability in the lactose utilization network of *Escherichia coli*. *Nature* **427,** 737–740 (2004).
24. Sha, W. *et al.* Hysteresis drives cell-cycle transitions in *Xenopus laevis* egg extracts. *Proc. Natl Acad. Sci. USA* **100,** 975–980 (2003).
25. Ferrell, J. E. Jr & Machleder, E. M. The biochemical basis of an all-or-none cell fate switch in *Xenopus* oocytes. *Science* **280,** 895–898 (1998).
26. Manrai, A. & Gunawardena, J. The geometry of multisite phosphorylation. *Biophys. J.* **95,** 5533–5543 (2008).
27. Monod, J. & Jacob, F. General conclusions: teleonomic mechanisms in cellular metabolism, growth and differentiation. *Cold Spring Harb. Symp. Quant. Biol.* **26,** 389–401 (1961).
28. Berger, S. L. The complex language of chromatin regulation during transcription. *Nature* **447,** 407–411 (2007).
29. Huang, C.-Y. F. & Ferrell, J. E. Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc. Natl Acad. Sci. USA* **93,** 10078–10083 (1996).
30. Cornish-Bowden, A. *Fundamentals of Enzyme Kinetics* 2nd edn 23–28 (Portland Press, 1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported in part by the NIH under grant R01-GM081578. We thank A. Manrai for scientific discussions, R. Ward for editorial help and HMS RITG for support with cluster computing. We acknowledge the encouragement of the late Stephen Thomson (1946–2006) and Charles Gunawardena (1929–2007).

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.G. (jeremy@hms.harvard.edu).