Journal of Theoretical Biology 304 (2012) 219–225

Contents lists available at SciVerse ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Direct Response Analysis in cellular signalling networks

N. Shayeghi [a], T. Ng [b], A.C.C. Coolen [a,b,*]

[a] Department of Mathematics, King's College London, The Strand, London WC2R 2LS, UK
[b] Randall Division of Cell and Molecular Biophysics, King's College London, New Hunts House, London SE1 1UL, UK

ARTICLE INFO

Article history:
Received 9 December 2011
Received in revised form
17 March 2012
Accepted 2 April 2012
Available online 9 April 2012

Keywords:
Systems biology
Signal transduction
Direct Response Analysis

ABSTRACT

Direct Response Analysis is a general computational tool for quantifying direct functional interactions between components in cellular signalling systems from experimental perturbations and measurements alone. This paper aims to reveal the biological meaning of the direct response coefficients obtained upon applying DRA to simple Michaelis–Menten type proteomic and gene regulatory systems. These systems describe dimer formation and dissociation, protein production and decay, and transcription. We derive explicit formulae for the direct response coefficients in terms of biochemical reaction rates, and clarify the potential and limitations of the DRA method. We find that response coefficients are strongly asymmetric, and that they balance persistent characteristics of reactions (e.g. the ratios of on- and off rates) against the time-scales over which these reactions act; fast reactions give stronger response coefficients. The direct interactions between protein species, caused by dimer formation, are effectively negative. We illustrate our results with numerical simulations.

1. Introduction

Macroscopic biological complexity often results from nonlinear interaction among microscopic components, the properties of which appear quite different in context and in isolation. Over the years, many studies have aimed at understanding how components of large and complex biological systems are integrated and coordinated into a functioning macroscopic unit. Nowadays such work is called Systems Biology (Murray, 2001; Citri and Yarden, 2006; Sorkin and Goh, 2008; Ciaccio et al., 2010). The systems approach is particularly fruitful in cellular biology, where the different responses induced in cells (such as proliferation, survival, and motility) are thought to be regulated by signalling via complicated cascades of protein reactions, initiated by ligand binding. Network-level analysis is needed to infer the signal transduction through changes in the activities and concentrations of multiple proteins and their complexes; this information then needs to be mapped to biochemical reaction equations, in order to understand the signalling dynamics quantitatively. Such research typically involves two areas of mathematical analysis (Chen et al., 2009). The first focuses on statistical and topological properties of the interaction network, and inference of biologically relevant modules (Casey et al., 2007). The second on the dynamical analysis of biochemical protein reactions (Murray, 2001; Cornish-Bowden, 2004), usually represented by coupled ordinary differential equations.

If we seek to model mathematically the various cellular biochemical reactions we face serious obstacles. Even if we know the relevant signalling pathways in full, such models typically have many parameters that must be measured or estimated, such as initial protein concentrations and reaction rate constants (e.g. complex formation and enzymatic reactions Cornish-Bowden, 2004). Moreover, often we do not know the pathways in full, and need mathematical analysis in conjunction with biological experiments in order to elucidate the pathway structure of the system. We then require effective means to incorporate experimental data into the model in a manner that reproduces cell specific responses while taking into account the uncertainty in the precise biochemistry of the signalling and the inability of experiments to measure all relevant processes (Chen et al., 2009; Aldridge et al., 2006). The purpose of Direct Response Analysis (DRA) (Kholodenko et al., 1997, 2002; Bruggeman et al., 2002; Sontag et al., 2004; Cho et al., 2005; Andrec et al., 2005; Kholodenko, 2007), proposed by Kholodenko and coworkers, is to achieve this. It is a general protocol with which to extract information on the *direct* pairwise influence which components in a complex dynamical system exert upon each other (as opposed to influence mediated via other components), from only the measurement of responses of these variables to controlled system perturbations. DRA has generally been referred to in literature as Modular Response Analysis, but we feel that this does not capture sufficiently the essence of the method, which is its ability to separate *direct* interactions between components from those that are mediated via third parties. However, it is not yet clear what Kholodenko's response coefficients represent exactly in terms of reaction rates and other biophysical parameters of the cellular signalling system from which they are calculated.

* Corresponding author at: Department of Mathematics, King's College London, The Strand, London WC2R 2LS, UK. Tel.: +44 20 78482235; fax: +44 20 78482017.
E-mail address: ton.coolen@kcl.ac.uk (A.C.C. Coolen).

0022-5193/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jtbi.2012.04.001

In this paper we aim to clarify the meaning of the direct response coefficients that emerge from DRA, by applying DRA to simple models of synthetic proteomic and gene regulation signalling systems. We express the response coefficients explicitly in terms of biochemical parameters (reaction rates), and thereby achieve a better understanding of what these coefficients generally tell us about the underlying biology.

2. Direct Response Analysis

Let us first review briefly the method of Direct Response Analysis (Kholodenko et al., 1997, 2002). We consider a general dynamical system, modelled by a closed set of ordinary differential equations for N dynamical variables $\mathbf{x}=(x_1,\ldots,x_N)$ (such as log-concentrations of proteins and protein complexes, or gene expression levels), and in which K parameters $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_K)$ represent time-independent external perturbations (such as RNA interference):

$$\frac{dx_i}{dt}=f_i(\mathbf{x};\boldsymbol{\theta}),\quad i=1\ldots N \tag{1}$$

The functions f_i are assumed to be smooth, so that their partial derivatives exist. In DRA one assumes also that the system (1) relaxes to a unique equilibrium state[1]; the fixed point is written as $\overline{\mathbf{x}}(\boldsymbol{\theta})=(\overline{x}_1(\boldsymbol{\theta}),\ldots,\overline{x}_N(\boldsymbol{\theta}))$, and is to be solved from the fixed-point equations $f_i(\mathbf{x};\boldsymbol{\theta})=0$ for all i. We next define direct response coefficients r_{ij} for the unperturbed system, and show how these can be calculated upon measuring the system's response to perturbations:

- We first consider the unperturbed system, i.e. $\boldsymbol{\theta}=\mathbf{0}$, and write its stationary state simply as $\overline{\mathbf{x}}(\mathbf{0})=\overline{\mathbf{x}}$. We now imagine a constrained version of (1), where only x_i evolves via $dx_i/dt=f_i(\mathbf{x};\mathbf{0})$, x_j is kept at $x_j=\overline{x}_j+\epsilon$ (with ϵ small) and all other components are held at their equilibrium values $x_k=\overline{x}_k$. The new stationary value for x_i will be of the form $\overline{x}_i'=\overline{x}_i+r_{ij}\epsilon+\mathcal{O}(\epsilon^2)$; the factor r_{ij} measures the direct linear effect of small perturbations in x_j on x_i close to the equilibrium state, and is called the direct response coefficient of the interaction $j\to i$. Its value is determined by the fixed-point equation:

$$f_i(\overline{x}_1,\ldots,\overline{x}_{i-1},\overline{x}_i+r_{ij}\epsilon+\mathcal{O}(\epsilon^2),\overline{x}_{i+1},\ldots,\overline{x}_{j-1},\overline{x}_j+\epsilon,\overline{x}_{j+1},\ldots,\overline{x}_N;\mathbf{0})=0 \tag{2}$$

  Working out this identity, using $f_i(\overline{\mathbf{x}},\mathbf{0})=0$ and the shorthands $J_{ij}=(\partial f_i(\mathbf{x},\mathbf{0})/\partial x_j)\big|_{\mathbf{x}=\overline{\mathbf{x}}}$ for the elements of the Jacobian at the fixed-point, gives $(J_{ii}r_{ij}+J_{ij})\epsilon+\mathcal{O}(\epsilon^2)=0$, and hence

$$r_{ij}=-J_{ij}/J_{ii} \tag{3}$$

- Calculation of (3) requires knowing J_{ij} and J_{ii}, i.e. partial derivatives of the functions $f_i(\mathbf{x},\boldsymbol{\theta})$ in (1). However, we often do not know these functions. The next step in DRA is therefore to construct a protocol for *measuring* all derivatives $\{J_{k\ell}\}$ via perturbations. Differentiation of the general fixed-point $f_i(\mathbf{x}(\boldsymbol{\theta}),\boldsymbol{\theta})=0$ of (1) with respect to the perturbations gives

$$\text{for all } (i,j):\quad \sum_{k=1}^{N}J_{ik}M_{kj}+L_{ij}=0,\quad M_{kj}=\frac{\partial x_k(\boldsymbol{\theta})}{\partial\theta_j},\ L_{ij}=\left.\frac{\partial f_i(\mathbf{x},\boldsymbol{\theta})}{\partial\theta_j}\right|_{\mathbf{x}(\boldsymbol{\theta})} \tag{4}$$

  The entries M_{kj} give the changes in fixed-point values of all components, resulting from small system perturbations; they can in principle be measured or estimated experimentally. The L_{ij} are generally not known; one could limit oneself to combinations (i,j) for which one expects to have $L_{ij}=0$, but this brings in dangerous and uncontrolled assumptions. However, if we limit ourselves to perturbations of production rates,[2] i.e. $f_i(\mathbf{x},\boldsymbol{\theta})=f_i(\mathbf{x})+\theta_i$, then $L_{ij}=\delta_{ij}$, and we can find the $\{J_{k\ell}\}$ by solving

$$\text{for all } (i,j):\quad \sum_{k=1}^{N}J_{ik}M_{kj}=-\delta_{ij} \tag{5}$$

  These are N^2 equations for N^2 unknowns, the solution of which amounts simply to inverting a matrix.

The strength of DRA is that it requires no prior knowledge of the forces $f_i(\mathbf{x},\mathbf{0})$ in the unperturbed system, which allows it to be used for *uncovering* these forces. But DRA also has weaknesses. First, it assumes that our system obeys *closed* equations, and (unless we turn to general perturbations $\boldsymbol{\theta}$, where we need ad hoc assumptions on which of the L_{ij} are zero) it demands that we are able to perturb all production rates of all components independently. In a proteomic system we would then have to perturb production rates of individual complexes independently, for assuming closed equations demands that complexes are included in our component list. This cannot be done without violating chemical conservation laws. Second, DRA requires perturbations that are infinitesimally small; in cellular signalling systems this is experimentally very hard. In most applications of DRA so far the perturbations have been gene knockdowns, so the above formulae do not apply. One could try to remedy the situation somewhat by extending the perturbation theory to higher orders, but this has not yet been done and would still rule out knockdown perturbations. In addition, some of the applicability limitations of DRA can be lifted if it is applied in a modular framework, with dynamical variables representing the activity of a module as a whole, as in e.g. Sontag et al. (2004), Cho et al. (2005), and Yalamanchili et al. (2006).

Here we are concerned with a different question: even if we succeed in measuring the response coefficients (3), how do we interpret the numbers that we find? What do these coefficients mean in terms of biophysical reaction rates, when we apply DRA to cellular signalling systems? Answering these questions would aid our translation of experimental results into practical biological knowledge, which is ultimately the objective of studies that use DRA, such as Santos et al. (2007). To achieve this we calculate the response coefficients (3) for simple mathematical models of proteomic reaction equations and gene regulation.

3. Application to simple proteomic reaction networks

3.1. Derivation of explicit formulae for direct response coefficients

Suppose our cellular signalling system is a simple protein interaction network, described by standard kinetic reaction equations (see e.g. Murray, 2001). If the reactions taking place are hetero-dimer formation and dissociation, production, and decay, and if for simplicity we leave out higher order protein complexes and post-translational modifications, then such a system with n

[1] Evolution to equilibrium is generally thought to be an experimentally reasonable assumption (Murray, 2001), although there are certainly proteomic systems with persistent oscillations (Carlin et al., 2011). The assumption of uniqueness is more suspect.

[2] If x_i is a log-concentration, then in terms of actual concentrations $X_i=e^{x_i}$ our equations would be of the form $(d/dt)X_i=F_i(X_1,\ldots,X_N)+\theta_iX_i$, so production rate perturbations would be proportional to the corresponding concentrations.

protein species would be described by equations of the form

$$\frac{\mathrm{d}}{\mathrm{d}t}x_i=\sum_{j=1}^{n}[k_{ij}^{-}x_{ij}-k_{ij}^{+}x_ix_j]+\vartheta_i-\gamma_ix_i,\quad \frac{\mathrm{d}}{\mathrm{d}t}x_{ij}=k_{ij}^{+}x_ix_j-k_{ij}^{-}x_{ij} \tag{6}$$

Here x_i and x_{ij} denote the concentrations of unbound protein i, and the hetero-dimer formed by i and j. The $k_{ij}^{\pm}$ are the on/off rates for the formation and dissociation of the hetero-dimers (if $k_{ij}^{+}=0$ such dimers are not forming); ϑ_i and γ_i represent protein production and decay rates. We consider only hetero-dimers, so $k_{ii}^{+}=x_{ii}=0$ for all i. Our Eq. (6) obey the conservation laws $(\mathrm{d}/\mathrm{d}t)[x_i+\sum_j x_{ij}]=\vartheta_i-\gamma_ix_i$ for all i. We define $x_{ij}=x_{ji}$ and $k_{ij}^{\pm}=k_{ji}^{\pm}$ for all (i,j). It is trivial to determine the stationary solution of (6), giving

$$\overline{x}_i=\frac{\vartheta_i}{\gamma_i},\quad \overline{x}_{ij}=\frac{k_{ij}^{+}}{k_{ij}^{-}}\frac{\vartheta_i\vartheta_j}{\gamma_i\gamma_j} \tag{7}$$

It is also trivial to calculate the entries of the Jacobian matrix. Since our components include two distinct groups of dynamical variables, viz. free protein concentrations labelled by a single index i, and concentrations of complexes labelled by index pairs (i,j), it is natural to write the associated forces as $f_i(\mathbf{x})=\sum_{j=1}^{n}[k_{ij}^{-}x_{ij}-k_{ij}^{+}x_ix_j]+\vartheta_i-\gamma_ix_i$ and $f_{ij}(\mathbf{x})=k_{ij}^{+}x_ix_j-k_{ij}^{-}x_{ij}$, giving a Jacobian matrix with a block structure:

$$J_{ij}=\left.\frac{\partial f_i(\mathbf{x})}{\partial x_j}\right|_{\overline{\mathbf{x}}}=-k_{ij}^{+}\overline{x}_i-\delta_{ij}\left(\sum_{\ell}k_{i\ell}^{+}\overline{x}_\ell+\gamma_i\right),\quad J_{k,ij}=\left.\frac{\partial f_k(\mathbf{x})}{\partial x_{ij}}\right|_{\overline{\mathbf{x}}}=k_{ij}^{-}(\delta_{ki}+\delta_{kj}) \tag{8}$$

$$J_{ij,k}=\left.\frac{\partial f_{ij}(\mathbf{x})}{\partial x_k}\right|_{\overline{\mathbf{x}}}=k_{ij}^{+}(\overline{x}_j\delta_{ki}+\overline{x}_i\delta_{kj}),\quad J_{ij,k\ell}=\left.\frac{\partial f_{ij}(\mathbf{x})}{\partial x_{k\ell}}\right|_{\overline{\mathbf{x}}}=-k_{ij}^{-}(\delta_{ik}\delta_{j\ell}+\delta_{i\ell}\delta_{jk}) \tag{9}$$

From this, via (3), follow our direct response coefficients of DRA. They are only defined for distinct pairs of interacting components, so we have r_{ij} with $i\neq j$, $r_{k,ij}$, $r_{ij,k}$, and $r_{ij,k\ell}$ with $(i,j)\neq(k,\ell),(\ell,k)$:

$$r_{ij}=-\frac{k_{ij}^{+}\overline{x}_i}{\sum_{\ell}k_{i\ell}^{+}\overline{x}_\ell+\gamma_i},\quad r_{k,ij}=\frac{k_{ij}^{-}(\delta_{ki}+\delta_{kj})}{\sum_{\ell}k_{k\ell}^{+}\overline{x}_\ell+\gamma_k},\quad r_{ij,k}=\frac{k_{ij}^{+}}{k_{ij}^{-}}(\overline{x}_j\delta_{ki}+\overline{x}_i\delta_{kj})$$
$$r_{ij,k\ell}=0 \tag{10}$$

Equivalently, with (7) we can eliminate the equilibrium concentrations and express the response coefficients fully in terms of biochemical system parameters:

$$r_{ij}=-\frac{k_{ij}^{+}\vartheta_i/\gamma_i}{\sum_{\ell}k_{i\ell}^{+}\vartheta_\ell/\gamma_\ell+\gamma_i},\quad r_{k,ij}=\frac{k_{ij}^{-}(\delta_{ki}+\delta_{kj})}{\sum_{\ell}k_{k\ell}^{+}\vartheta_\ell/\gamma_\ell+\gamma_k}$$
$$r_{ij,k}=\frac{k_{ij}^{+}}{k_{ij}^{-}}\left(\frac{\vartheta_j}{\gamma_j}\delta_{ki}+\frac{\vartheta_i}{\gamma_i}\delta_{kj}\right),\quad r_{ij,k\ell}=0 \tag{11}$$

3.2. Interpretation and illustration

One understands the signs of the above coefficients. Increasing the amount of an unbound protein ℓ, at equilibrium, will increase the formation rate of all dimers involving ℓ, and hence *reduce* the amount of unbound protein of all its binding partners (so $r_{i\ell}<0$ for all i with $k_{i\ell}^{+}>0$), while *increasing* the dimer concentrations (so $r_{ij,\ell}>0$ for all complexes (i,j) that involve protein ℓ). Similarly, increasing the amount of the (i,j) dimer, in equilibrium, increases the number of (i,j) dissociations and hence the concentrations of its constituent proteins; since it has no direct effect on other dimers we should indeed expect $r_{ij,k\ell}=0$. For further interpretation it is useful to write our equations and results in terms of the following quantities: $\overline{x}_i=\vartheta_i/\gamma_i$ (stationary concentrations of unbound protein), $\Gamma_{ij}=k_{ij}^{+}/k_{ij}^{-}$ (which control the equilibrium balance of bound versus unbound (i,j) dimers), $\tau_{ij}=1/k_{ij}^{-}$ (the characteristic timescale of complex formation/dissociation), and $\tau_i=1/\gamma_i$ (the characteristic timescale of protein decay). This converts (6) into

$$\frac{\mathrm{d}}{\mathrm{d}t}x_i=\sum_{j=1}^{n}\frac{1}{\tau_{ij}}(x_{ij}-\Gamma_{ij}x_ix_j)+\frac{1}{\tau_i}(\overline{x}_i-x_i),\quad \frac{\mathrm{d}}{\mathrm{d}t}x_{ij}=\frac{1}{\tau_{ij}}(\Gamma_{ij}x_ix_j-x_{ij}) \tag{12}$$

and gives for the nonzero direct response coefficients:

$$r_{ij}=\frac{-\Gamma_{ij}\overline{x}_i}{\sum_{\ell}\Gamma_{i\ell}\overline{x}_\ell(\tau_{ij}/\tau_{i\ell})+(\tau_{ij}/\tau_i)},\quad r_{k,ij}=\frac{\delta_{ki}+\delta_{kj}}{\sum_{\ell}\Gamma_{k\ell}\overline{x}_\ell(\tau_{ij}/\tau_{k\ell})+(\tau_{ij}/\tau_k)}$$
$$r_{ij,k}=\Gamma_{ij}(\overline{x}_j\delta_{ki}+\overline{x}_i\delta_{kj}) \tag{13}$$

We now see more clearly how direct response coefficients balance the persistent impact of interactions against the timescales over which the interactions act. For instance, if protein decay is much slower than dimer formation and dissociation, i.e. $\tau_{ij}/\tau_k\to 0$ for all (i,j,k), then the above coefficients reduce to

$$r_{ij}=\frac{-\Gamma_{ij}\overline{x}_i}{\sum_{\ell}\Gamma_{i\ell}\overline{x}_\ell(\tau_{ij}/\tau_{i\ell})},\quad r_{k,ij}=\frac{\delta_{ki}+\delta_{kj}}{\sum_{\ell}\Gamma_{k\ell}\overline{x}_\ell(\tau_{ij}/\tau_{k\ell})},\quad r_{ij,k}=\Gamma_{ij}(\overline{x}_j\delta_{ki}+\overline{x}_i\delta_{kj}) \tag{14}$$

whereas if protein decay is much faster than dimer formation and dissociation, i.e. $\tau_{ij}/\tau_k\to\infty$ for all (i,j,k):

$$r_{ij}=0,\quad r_{k,ij}=0,\quad r_{ij,k}=\Gamma_{ij}(\overline{x}_j\delta_{ki}+\overline{x}_i\delta_{kj}) \tag{15}$$

Fig. 1 shows an example of the resulting DRA map for the above system, in which unbound proteins and complexes correspond to circular and square nodes, respectively, and where nonzero direct response coefficients are calculated from (11) and drawn as directed arrows (in shades of red when negative, and shades of black when positive). Reaction rates in this example were generated randomly. The figure emphasises very clearly the general asymmetry of the direct response coefficients, in contrast to more naive methods for quantifying the mutual interactions in interaction networks (such as quantifying interactions by correlation between the values of the dynamical variables, which would always lead to $r_{ij}=r_{ji}$ for all (i,j)).

To gain further intuition on direct response coefficients in proteomic systems we next examine equations (11) numerically for randomly generated network instances, with $N=60$ protein species. We draw the reaction rates for each network from truncated Gaussian distributions (truncated at zero, to have nonnegative values only) and we allow all protein species in principle to form complexes with each other. Our objective is to find out whether there are statistical relations between

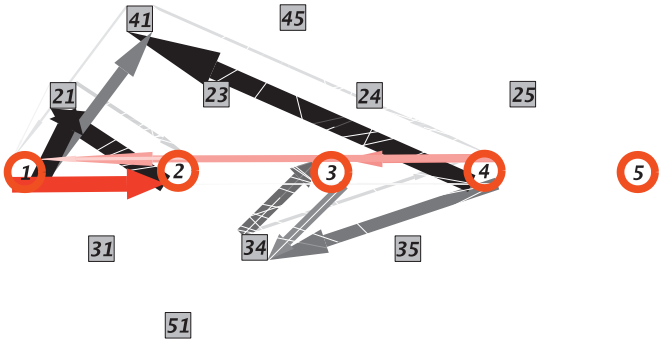

Fig. 1. The DRA map for a simple synthetic protein interaction system with $N=5$ protein species, described by Eq. (6). All rate constants were drawn randomly from a truncated Gaussian distribution. Complexes are represented by grey squares and unbound proteins by red circles. Arrows represent the direct response coefficients (in direction and magnitude) given in (11); interactions between unbound proteins are always negative and shown in red, and interactions between unbound proteins and dimers are always positive and shown in black (thicker and darker arrows indicate stronger direct responses).

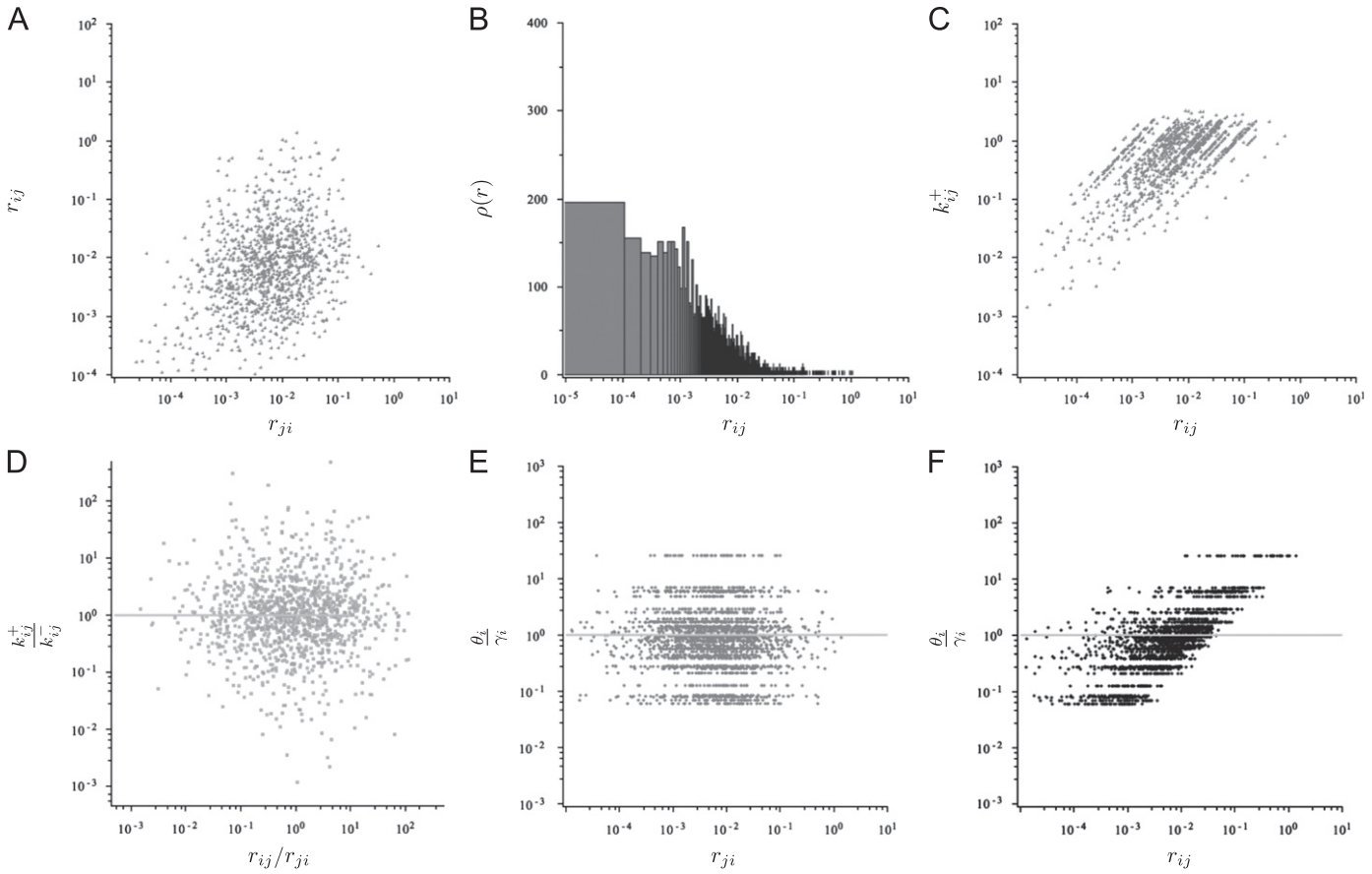

Fig. 2. Values of direct response coefficients between $N=60$ unbound protein species, measured upon generating random protein interaction networks with reaction rates drawn randomly from truncated Gaussian distributions. Top row: A, scatter plot of (r_{ij},r_{ji}) on a log–log scale, plotted for $i<j$ (to prevent artificial symmetries); B, observed distribution $\rho(r)$ of direct response coefficients; C, scatter plot of (r_{ij},k_{ij}^+) on a log–log scale. Bottom row: D, scatter plot of $(r_{ij}/r_{ji},k_{ij}^+/k_{ij}^-)$ plotted for $i<j$ (to prevent artificial symmetries); E, scatter plot of $(r_{ji},\theta_i/\gamma_i)$; F, scatter plot of $(r_{ij},\theta_i/\gamma_i)$. *Note*: for every ratio θ_i/γ_i there are N coefficients r_{ij}, which is why the points in the last two plots show linear sub-structure.

e.g. forward and reverse direct response coefficients r_{ij} and r_{ji}, or between direct response coefficients and the type of reaction rate functions one might expect to dominate their values, and how the coefficients are distributed. The results are shown in Figs. 2 and 3. We observe no significant correlation between r_{ij} and r_{ji} (Pearson coefficient <0.05), see e.g. Fig. 2A. Fig. 2B shows a histogram estimator of the distribution $\rho(r)=N^{-2}\sum_{i,j\le N}\delta[r-r_{ij}]$, which suggests that for the chosen reaction rate statistics the direct response coefficients are, to good approximation, distributed according to a power law. We next plot the coefficients r_{ij} against the associated on-rates k_{ij}^+ in Fig. 2C (showing only a modest positive correlation, as expected from (11)), and the ratios r_{ij}/r_{ji} against the corresponding binding ratios k_{ij}^+/k_{ij}^- in Fig. 2D (where no relation is observed). In Fig. 2E and F we observe a strong correlation between the r_{ij} and the protein synthesis/degradation ratios θ_i/γ_i (Pearson coefficient 0.7), but almost no correlation between the reverse response coefficient r_{ji} and θ_i/γ_i, i.e. the direct response of protein i to perturbation of protein j depends solely on typical concentration of unbound protein i . In Fig. 3 Left and Right we detect a weak inverse linear relation between $r_{ij,k}$ and $r_{jk,i}$, (correlation -0.1). Furthermore, we do observe a roughly linear relation between $r_{i,jk}+r_{jk,i}$ and k_{ij}^+/k_{ij}^-. Overall we can deduce that the most important parameters impacting upon direct response coefficients between unbound proteins are the on-rates for complex formation and their synthesis/degradation ratios. The direct responses of free protein concentrations to perturbations in concentrations of complexes (and vice versa) are influenced strongly by the binding ratios of the complexes. The above regularities are preserved if we use different random initialisations or different distributions of rate constants.

4. Application to a combined proteomic reaction and gene regulation network

4.1. Construction of closed equations for equilibrating gene regulation dynamics

If we wish to apply DRA as a tool for uncovering gene regulation pathways (Yalamanchili et al., 2006; Mettetal et al., 2006; Raingeaud et al., 1996; Williams et al., 2004; Workman et al., 2006), the variables to be perturbed and measured are gene expression levels. Experimental timescales will then have to be sufficiently large for the transcription system to equilibrate; gene regulation is about three orders of magnitude slower than protein reactions. In Eq. (6) the variable ϑ_i can be interpreted as representing the expression level of gene i, so we now extend (6) with further simple equations that describe the evolution of the ϑ_i, giving

$$\frac{\mathrm{d}}{\mathrm{d}t}x_i=\sum_{j=1}^{n}[k_{ij}^-x_{ij}-k_{ij}^+x_ix_j]+\vartheta_i-\gamma_ix_i,\quad \frac{\mathrm{d}}{\mathrm{d}t}x_{ij}=k_{ij}^+x_ix_j-k_{ij}^-x_{ij} \tag{16}$$

$$\tau\frac{\mathrm{d}}{\mathrm{d}t}\vartheta_i=F(A_i(\mathbf{x}),R_i(\mathbf{x}))-\vartheta_i \tag{17}$$

$$A_i(\mathbf{x})=\sum_j J_{ij}^+x_j+\sum_{jk}J_{i,jk}^+x_{jk},\quad R_i(\mathbf{x})=\sum_j J_{ij}^-x_j+\sum_{jk}J_{i,jk}^-x_{jk} \tag{18}$$

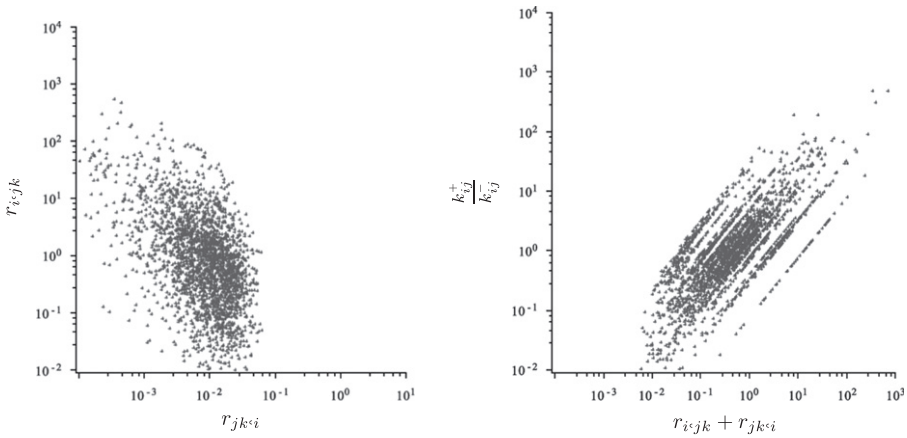

Fig. 3. Values of direct response coefficients between $N=60$ unbound protein species and their binary complexes, measured upon generating synthetic protein interaction networks with reaction rates drawn randomly from truncated Gaussian distributions. Left: scatter plot of $(r_{i,jk}, r_{jk,i})$ on a log-log scale. Right, scatter plot of $(r_{i,jk}+r_{jk,i}, k_{ij}^+/k_{ij}^-)$ on a log–log scale. *Note*: for every ratio k_{ij}^+/k_{ij}^- there are N combinations of $r_{i,jk}+r_{jk,i}$, which explains the linear sub-structure on the right.

Here $A_i(\mathbf{x})$ and $R_i(\mathbf{x})$ represent formulae for the cumulative activation and repression of gene i, in which nonnegative parameters $\{J_{ij}^{\pm}\, J_{i,jk}^{\pm}\}$ specify the efficacies of the various proteins j and dimers (i,j) as activators $(+)$ or repressors $(-)$ of gene i, and with $J_{i,jk}^{\pm}=J_{i,kj}^{\pm}$ for all (i,j,k). If protein j does not transcribe gene i then $J_{ij}^{+}=J_{ij}^{-}=0$, etc. $F(A,R)$ is a nonnegative nonlinear Hill-type function, with $\partial F/\partial A \geq 0$, $\partial F/\partial R \leq 0$, such as $F(A,R)=\mathrm{Erf}(A-R)$ or $F(A,R)=\frac{1}{2}+\frac{1}{2}\tanh(A-R)$. The parameter τ gives the characteristic timescale for gene regulation dynamics.

Since the proteomic dynamics is much faster than gene regulation, we may assume that (16) reaches equilibrium very fast, and replace the instantaneous values $\mathbf{x}$ in (18) by their equilibrated expressions $\mathbf{x}(\boldsymbol{\vartheta})$ given in (7). We also choose the simple form $F(A,R)=F(A-R)$ for our transcription function. All this reduces our equations to the following simple set that involves expression levels only:

$$\tau\frac{\mathrm{d}}{\mathrm{d}t}\vartheta_i = F(U_i(\boldsymbol{\vartheta}))-\vartheta_i \tag{19}$$

$$U_i(\boldsymbol{\vartheta})=\sum_j U_{ij}\vartheta_j+\sum_{jk}U_{i,jk}\vartheta_j\vartheta_k,\quad U_{ij}=\frac{J_{ij}^{+}-J_{ij}^{-}}{\gamma_j},\quad U_{i,jk}=\frac{k_{jk}^{+}J_{i,jk}^{+}-J_{i,jk}^{-}}{k_{jk}^{-}\gamma_j\gamma_k} \tag{20}$$

The DRA formalism demands that (19) evolves to a unique fixed-point. Numerical examination, however, reveals that this is not always the case; the system (19) can evolve into oscillatory trajectories. To overcome this problem and get meaningful results from DRA, we have to postulate that, whenever the fixed-points of (19) are locally unstable, asymptotically the system (19) will exhibit persistent oscillations *close to* these fixed-points. This would allow us to apply DRA after all, in reasonable approximation. In order to find the unstable fixed points of (19) numerically, we modify (19) into

$$\frac{\mathrm{d}}{\mathrm{d}t}\vartheta_i=\mathcal{G}_i(\boldsymbol{\vartheta})=F(U_i(\boldsymbol{\vartheta}))-\vartheta_i-\sum_\ell(F(U_\ell(\boldsymbol{\vartheta}))-\vartheta_\ell)\frac{\partial}{\partial\vartheta_i}F(U_\ell(\boldsymbol{\vartheta})) \tag{21}$$

All stationary solutions of (19) are also stationary solutions of (21). However, in contrast to (19), the modified process (21) is guaranteed to evolve to a fixed point, i.e. a solution of

$$\forall i=1\ldots N:\quad \sum_\ell(F(U_\ell(\boldsymbol{\vartheta}))-\vartheta_\ell)\frac{\partial}{\partial\vartheta_i}F(U_\ell(\boldsymbol{\vartheta}))=F(U_i(\boldsymbol{\vartheta}))-\vartheta_i \tag{22}$$

since it defines gradient descent on the Lyapunov function

$$\mathcal{L}(\boldsymbol{\vartheta})=\frac{1}{2}\sum_\ell(\vartheta_\ell-F(U_\ell(\boldsymbol{\vartheta})))^2 \tag{23}$$

Moreover, the global minima of (23), i.e. the dominant solutions of (22), are indeed the fixed-points of (19).

4.2. Derivation of explicit formulae for direct response coefficients

Since stationary can now be relied upon, we can calculate from (21) after equilibration the direct response coefficient R_{ij} that quantifies the impact of expression level j on expression level i:

$$R_{ij}=-\frac{\partial\mathcal{G}_i(\boldsymbol{\vartheta})/\partial\vartheta_j}{\partial\mathcal{G}_i(\boldsymbol{\vartheta})/\partial\vartheta_i} \tag{24}$$

with

$$\begin{aligned}\frac{\partial}{\partial\vartheta_j}\mathcal{G}_i(\boldsymbol{\vartheta})&=\frac{\partial}{\partial\vartheta_j}\sum_\ell(F(U_\ell\boldsymbol{\vartheta}))-\vartheta_\ell)\left(\delta_{i\ell}-\frac{\partial}{\partial\vartheta_i}F(U_\ell(\boldsymbol{\vartheta}))\right)\\&=\sum_\ell\left\{\left(\frac{\partial F(U_\ell(\boldsymbol{\vartheta}))}{\partial\vartheta_j}-\delta_{j\ell}\right)\left(\delta_{i\ell}-\frac{\partial F(U_\ell(\boldsymbol{\vartheta}))}{\partial\vartheta_i}\right)\right.\\&\quad\left.-(F(U_\ell(\boldsymbol{\vartheta}))-\vartheta_\ell)\frac{\partial^2F(U_\ell(\boldsymbol{\vartheta}))}{\partial\vartheta_i\partial\vartheta_j}\right)\Big\}\end{aligned} \tag{25}$$

Upon choosing the simple sigmoidal function $F(U)=\frac{1}{2}+\frac{1}{2}\tanh(U)$, with $F'(U)=\frac{1}{2}1-\tanh^2(U)]$, the remaining partial derivatives come out to be

$$\frac{\partial F(U_\ell(\boldsymbol{\vartheta}))}{\partial\vartheta_i}=\frac{1}{2}[1-\tanh^2(U_\ell(\boldsymbol{\vartheta}))]\frac{\partial U_\ell(\boldsymbol{\vartheta})}{\partial\vartheta_i} \tag{26}$$

$$\begin{aligned}\frac{\partial^2F(U_\ell(\boldsymbol{\vartheta}))}{\partial\vartheta_i\partial\vartheta_j}&=[1-\tanh^2(U_\ell(\boldsymbol{\vartheta}))]\left\{\frac{1}{2}\frac{\partial^2U_\ell(\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}\right.\\&\quad\left.-\tanh(U_\ell(\boldsymbol{\vartheta})\frac{\partial U_\ell(\boldsymbol{\vartheta})}{\partial\vartheta_i}\frac{\partial U_\ell(\boldsymbol{\vartheta})}{\partial\vartheta_j}\right\}\end{aligned} \tag{27}$$

in which we have

$$\frac{\partial U_\ell(\boldsymbol{\vartheta})}{\partial\vartheta_i}=U_{\ell i}+2\sum_k U_{\ell,ik}\vartheta_k,\quad \frac{\partial^2U_\ell(\boldsymbol{\vartheta})}{\partial\vartheta_i\partial\vartheta_j}=2U_{\ell,ij} \tag{28}$$

We should bear in mind that Eq. (21) will generally have more than one fixed-point solution; for each such solution we can calculate corresponding direct response coefficients R_{ij}, and these will generally differ between fixed-points. Fig. 4 shows an example of the resulting DRA map for the (modified) gene regulation system (21), in which genes correspond to circular nodes, and where nonzero direct response coefficients are calculated from (24) and drawn as directed arrows (in shades of red when negative, and shades of black when positive). Transcription rates in this example were generated randomly. Again we observe clearly the general asymmetry of the direct response coefficients.

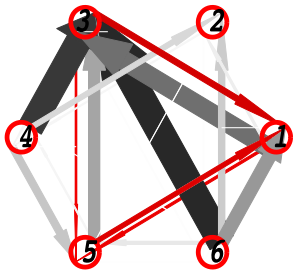

Fig. 4. The DRA map for a simple gene regulation system of the type (21) with $N=6$ genes. All transcription rate constants were drawn randomly from a truncated Gaussian distribution. Individual genes are shown as red circles. The dynamics (21) is run following random initialisation with expression levels drawn from a truncated Gaussian distribution. Arrows represent the direct response coefficients R_{ij} (in direction and magnitude) given in (24); negative coefficients are shown in red, positive ones in black (thicker and darker arrows indicate stronger direct responses). We emphasise that this example serves only as an illustration; the direct response coefficients found (and hence the DRA diagram) obviously depend on the chosen rate constants.

5. Discussion

Direct Response Analysis (DRA) is a versatile computational tool proposed by Kholodenko (2007, and see references therein) for quantifying functional interactions between components in cellular signalling systems from experimental perturbations and measurements alone. In this paper we sought to understand in greater depth the potential and limitations of this method when applied to cellular signalling systems, and the biological meaning of the direct response coefficients that would result from application of DRA to simple Michaelis–Menten type proteomic and gene regulatory systems by deriving explicit formulae for the direct response coefficients in terms of biochemical reaction rates.

As with any mathematical formalism, the validity of DRA rests on various assumptions and conditions. It is hard to see how these can be satisfied in the context of proteomic signalling systems. For instance, DRA demands that one perturbs and measures independently all dynamical variables of a *closed* dynamical system. In proteomic pathways this implies that one has to include the concentrations of all complexes that can be formed by the proteins under study. But independent perturbation of concentrations of individual complexes without perturbing the concentrations of the constituent proteins is impossible in practice, as it would violate biochemical conservation laws. Second, DRA is a linear perturbation theory, and thus demands that all perturbations are infinitesimally small; yet in practice it is often applied in a context where perturbations take the form of gene knockdowns—violating the basis of the method. It is not clear to what extent the outcome of DRA can be trusted when applied to a situation that fails to meet its conditions for validity. If DRA is applied to gene regulation pathways a further complication arises: the typical equations describing the dynamics of expression levels in gene regulation networks do not always evolve to equilibrium, in conflict with another assumption of DRA. One can circumvent the problem via a slightly modified dynamical process that does go to stationary states, but this requires further assumption that the gene regulation dynamics stabilises into oscillatory trajectories *close to* fixed-points.

We have expressed the direct response coefficients of DRA explicitly in terms of biochemical rate constants, for simple models of proteomic and gene regulation systems. This led to valuable insight into what exactly is measured by these coefficients. For instance:

1. Response coefficients are strongly asymmetric (i.e. directional). The direct effect of perturbing the concentration of a protein A on that of another protein B with which it interacts is not invariant under exchanging the roles of A and B. The same is true for perturbations of expression levels.
2. Response coefficients balance persistent characteristics of reactions (e.g. the ratios of on- and off rates) against the time-scales over which these reactions act. The direct response mediated by a biochemical reaction increases with the impact of that reaction on steady-state concentrations, but decreases with the characteristic time scale of this reaction (fast reactions give stronger response coefficients).
3. The direct interactions between protein species, caused by dimer formation, are negative. The intuition is that increasing the concentration of a protein leads to increased formation of dimers with its reaction partners; this reduces the amount of free protein of its partners. This is why oscillatory protein concentration dynamics can be linked to short loops of odd length in protein interaction networks (Carlin et al., 2011).

The main reasons for using DRA in the study of real biological systems have traditionally been to understand the system responses that are observed experimentally following specific biological perturbations such as gene knockdown and cell adhesion, and to have a systematic tool for mapping unknown cellular signalling pathways. Our present work might add to these the potential for calculating estimates of actual kinetic parameters from experimental perturbations, either by explicit inversion of identities such as (11) and (24), or by using these identities in a Bayesian estimation framework. The main limitation of our analysis is that it is based on a simplified mathematical representation of protein–protein interaction and gene regulation, as will be any quantitative study. Our results are thus valid to the extent that this representation captures the dominant signalling events in the cell. More specifically: our model does not include molecular complexes of order higher than two, post-translational modifications, stochasticity, or spatial variation of molecular concentrations, and involves simplified equations for transcription.

However, we hope that this paper may aid experimentalists who use the DRA method for uncovering cellular signalling pathways, in exposing the dangers caused by violation of the assumptions underlying DRA, and in providing a better understanding of the biochemical meaning of the direct response coefficients which they would be measuring.

Acknowledgments

This study was supported by the Biotechnology and Biological Sciences Research Council (BBSRC, grant BB/G007160/1) of the United Kingdom.

References

Aldridge, B.B., Burke, J.M., Lauffenburger, D.A., Sorger, P.K., 2006. Nat. Cell Biol. 8, 1195–1203.
Andrec, M., Kholodenko, B.N., Levy, R.M., Sontag, E., 2005. J. Theor. Biol. 232, 427–441.
Bruggeman, F.J., Westerhoff, H.V., Hoek, J.B., Kholodenko, B.N., 2002. J. Theor. Biol. 218, 507–520.
Carlin, L.M., Milewicz, H., Matthews, D.R., Fernandes, L., Perani, M., Levitt, J., Keppler, M., Coolen, A.C.C., Barber, P., Vojnovic, B., Suhling, K., Fraternali, F., Ameer-Beg, S., Parker, P.J., Thomas, N.S.B., Ng, T., 2011. Sci. Signal. 4, 1–11.
Casey, F.P., Baird, D., Feng, Q., Gutenkunst, R.N., Waterfall, J.J., Myers, C.R., Brown, K.S., Cerione, R.A., Sethna, J.P., 2007. Syst. Biol. IET 1 (3), 190–202.
Chen, W.W., Schoeberl, B., Jasper, P.J., Niepel, M., Nielsen, U.B., Lauffenburger, D.A., Sorger, P.K., 2009. Mol. Syst. Biol. 5, 239.
Cho, K.-H., Choo, S.-M., Wellstead, P., Wolkenhauer, O., 2005. FEBS Lett. 579, 4520–4528.

Ciaccio, M.F., Wagner, J.P., Chuu, C., Lauffenburger, D.A., Jones, R.B., 2010. Nat. Methods 7, 148–155.
Citri, A., Yarden, Y., 2006. Nat. Rev. 7, 505–516.
Cornish-Bowden, A., 2004. Fundamentals of Enzyme Kinetics, 3rd edition. Portland Press.
Kholodenko, B.N., 2007. Nat. Cell. Biol. 9 (3), 247–249.
Kholodenko, B.N., Hoek, J.B., Westerhoff, H.V., Brown, G.C., 1997. FEBS Lett. 414 (2), 430–434.
Kholodenko, B.N., Kiyatkin, A., Bruggeman, F.J., Sontag, E., Westerhoff, H.V., Hoek, J.B., 2002. Cell. Biol. 99 (20), 12841–12846.
Mettetal, J.T., Muzzey, D., Pedraza, J.M., Ozbudak, E.M., van Oudenaarden, A., 2006. Proc. Natl. Acad. Sci. USA 103 (19), 7304–7309.
Murray, J.D., 2001. Mathematical Biology I. An Introduction. Springer, New York.
Raingeaud, J., Whitmarsh, A.J., Barrett, T., Derijard, B., Davis, R.J., 1996. Mol. Cell. Biol. 16, 1247–1255.
Santos, S.D.M., Verveer, P.J., Bastiaens, P.I.H., 2007. Nat. Cell. Biol. 9, 324–330.
Sontag, E., Kyatkin, A., Kholodenko, B.N., 2004. Bioinformatics 20, 1877–1886.
Sorkin, A., Goh, L.K., 2008. Exp. Cell. Res. 315 (4), 683–696.
Williams, C.C., Allison, J.G., Vidal, G.A., Burow, M.E., Beckman, B.S., Marrero, L., Jones, F.E., 2004. J. Cell. Biol. 167, 469–478.
Workman, C.T., Mak, H.C., McCuine, S., Bosco Tagne, J., Agarwal, M., Ozier, O., Begley, T.J., Samson, L.D., Ideker, T., 2006. Science 312 (5776), 1054–1059.
Yalamanchili, N., Zak, D.E., Ogunnaike, B.A., Schwaber, J.S., Kriete, A., Kholodenko, B.N., 2006. Syst. Biol. 153, 236–246.