INTERFACE

rsif.royalsocietypublishing.org

Research

Cite this article: Birtwistle MR. 2015 Analytical reduction of combinatorial complexity arising from multiple protein modification sites. *J. R. Soc. Interface* **12**: 20141215.
http://dx.doi.org/10.1098/rsif.2014.1215

Received: 3 November 2014
Accepted: 27 November 2014

Subject Areas:
systems biology, biocomplexity, biophysics

Keywords:
kinetic modelling, signal transduction, model reduction, combinatorial complexity

Author for correspondence:
Marc R. Birtwistle
e-mail: marc.birtwistle@mssm.edu

Electronic supplementary material is available at http://dx.doi.org/10.1098/rsif.2014.1215 or via http://rsif.royalsocietypublishing.org.

THE ROYAL SOCIETY PUBLISHING

Analytical reduction of combinatorial complexity arising from multiple protein modification sites

Marc R. Birtwistle

Department of Pharmacology and Systems Therapeutics, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA

Combinatorial complexity is a major obstacle to ordinary differential equation (ODE) modelling of biochemical networks. For example, a protein with 10 sites that can each be unphosphorylated, phosphorylated or bound to adaptor protein requires 3^{10} ODEs. This problem is often dealt with by making *ad hoc* assumptions which have unclear validity and disallow modelling of site-specific dynamics. Such site-specific dynamics, however, are important in many biological systems. We show here that for a common biological situation where adaptors bind modified sites, binding is slow relative to modification/demodification, and binding to one modified site hinders binding to other sites, for a protein with n modification sites and m adaptor proteins the number of ODEs needed to simulate the site-specific dynamics of biologically relevant, lumped bound adaptor states is independent of the number of modification sites and equal to $m + 1$, giving a significant reduction in system size. These considerations can be relaxed considerably while retaining reasonably accurate descriptions of the true system dynamics. We apply the theory to model, using only 11 ODEs, the dynamics of ligand-induced phosphorylation of nine tyrosines on epidermal growth factor receptor (EGFR) and primary recruitment of six signalling proteins (Grb2, PI3K, PLCγ1, SHP2, RasA1 and Shc1). The model quantitatively accounts for experimentally determined site-specific phosphorylation and dephosphorylation rates, differential affinities of binding proteins for the phosphorylated sites and binding protein expression levels. Analysis suggests that local concentration of site-specific phosphatases such as SHP2 in membrane subdomains by a factor of approximately 10^7 is critical for effective site-specific regulation. We further show how our framework can be extended with minimal effort to consider binding cooperativity between Grb2 and c-Cbl, which is important for receptor trafficking. Our theory has potentially broad application to reduce combinatorial complexity and allow practical simulation of a variety ODE models relevant to systems biology and pharmacology applications to allow exploration of key aspects of complexity that control signal flux.

1. Introduction

Biochemical networks are bewilderingly complex. One look at a comprehensive view of metabolic pathways or signalling networks [1,2] leaves one with a sense that predicting their behaviour requires more than intuition and an illustration. Yet, advances in metabolic engineering and systems pharmacology depend on our ability to make such predictions. Thus, many see progression of these and similar fields as depending on our ability to build mathematical, computational representations of relevant biochemical networks, which can subsequently be used to predict their behaviour.

Because biochemical networks, on a basic level, are simply a collection of coupled chemical reactions, one can model their behaviour with differential equation-based chemical kinetics approaches [3,4]. One fundamental but common roadblock to theoretically precise implementation of such approaches is 'combinatorial complexity' [5–7]. Combinatorial complexity arises when

network entities have several *sites* that can each be in several *states*. For example, a protein that contains 10 phosphorylation sites, each of which can be unphosphorylated, phosphorylated or non-competitively bound to a downstream adaptor, leads to 3^{10} (approx. 59 000) unique chemical species. A typical approach to avoid this combinatorial complexity is arbitrarily lumping sites together [3,8,9], but the validity and consequences of such approaches are unclear. Importantly, such reduction approaches result in the inability to track site-specific phenomena, such as differential phosphorylation and dephosphorylation rates on different sites of the same protein. Such site-specific phenomena can have important biological impact. For example, the phosphatase SHP2 dephosphorylates only particular phosphotyrosines on receptor tyrosine kinases that are important for controlling Ras deactivation [10]. This is noteworthy as constitutive Ras activation drives progression of numerous cancer types [11]. On the epidermal growth factor receptor (EGFR), phosphorylation of Y1068 is correlated with longer survival in non-small cell lung cancer patients, whereas phosphorylation of Y1173 is correlated with shorter survival [12]. There are multiple proteins, such as Grb2, PI3K, SHP2, PLCγ1, RasA1 (p120RasGAP) and Shc1, which bind to these and the other approximately 10 tyrosine phosphorylation sites on EGFR in both unique and overlapping manners with varying affinities [13]. The relative recruitment levels of these various proteins determine the strength of downstream signalling to pathways such as Akt, ERK or PKC, which subsequently dictate biological responses such as proliferation, migration, differentiation or apoptosis [14]. However, predicting the recruitment levels of these binding proteins without considering site-specific effects, and therefore combinatorial complexity, is challenging: each EGFR tyrosine is phosphorylated at different rates [15], and the host of binding proteins has distinct affinities for the various EGFR phosphotyrosines [13] as well as different expression levels [16].

Incorporating tens of thousands or even millions of differential equations into a model to account for all these site-specific phenomena in a theoretically precise manner is problematic for several reasons. One is writing the code necessary to simulate so many differential equations. However, rule-based modelling specification languages (e.g. BioNetGen [17,18] and Kappa [19]) allow one to write a small number of reaction rules that are then used to automatically build a set of differential equations, providing the tools necessary to tackle this code-writing problem. Theory that detects independence between sites and states also allows one to reduce combinatorial complexity in a rigorous manner [19–23], but interactions are common and block reduction through such means. Another problem, which rule-based approaches cannot solve, is practical numerical integration of that many differential equations. This is particularly the case considering that the model must be integrated not a single time, but many times over to be useful for routine tasks such as parameter estimation and sensitivity analysis. One potential solution is incorporating kinetic Monte Carlo approaches into a rule-based framework for so-called network-free simulation [24,25]. Such approaches do not enumerate all potential states *a priori*, but rather let initial conditions of the network stochastically propagate into realized states. Yet, such approaches are also computationally intensive for relatively small networks, require specialized software and code (whereas differential equation solvers are widely available and easy to use) and may be difficult to programmatically couple with other models (which are likely to be differential equation-based). Recently, powerful theoretical approaches have been developed that allow one to calculate steady-state behaviour of biochemical networks without having to consider combinatorial complexity [26–29]. Yet, this theory has limitations that preclude its wide application to practical simulation of several systems: (i) enzymes cannot act as substrates, but it is common for signalling network proteins to be both; (ii) there is limited ability to model feedback, but various forms of feedback are central to shaping signalling network control of cell fate [14,30], and (iii) in many situations, signalling dynamics, rather than steady-state behaviour, control biological responses [30,31]. Therefore, a theory that can reduce combinatorial complexity yet retain both the use of differential equations and the ability to describe these important network features may have widespread usefulness.

Here, we provide analytical equations that allow one to reduce combinatorial complexity for a biological scenario that is common to a variety of systems: a protein such as a receptor has multiple modification sites that bind to downstream adaptors, the adaptors sterically hinder each other from simultaneous binding and the modification/demodification kinetics are fast relative to downstream adaptor recruitment. We show that the size of a reduced system which tracks the biologically relevant levels of total recruited adaptor protein and models site-specific phenomena *does not depend on the number of modification sites*, allowing for significant reduction in combinatorial complexity. Predictions of this theory are shown to match well to non-reduced system simulations over a wide range of kinetic parameters, and finally we demonstrate how the theory can be applied to model site-specific EGF-induced signalling through EGFR.

2. Results and discussion

2.1. General modelled system and assumptions

The theory developed in this manuscript considers any protein or stable protein complex having multiple sites that can each be modified and demodified, and then the modified form of this site binds to a downstream protein. Modifications may include, for example, phosphorylation, or any other post-translational modification. The theory allows for bound proteins to alter the modification and demodification rates in a site-specific manner. This overall scenario is common in signalling networks.

Specifically, we assume the following conditions hold

(1) Adaptor binding is slow relative to modification/demodification cycles, such that these cycles reach quasi-equilibrium.
(2) The modification sites are located in close physical proximity to one another relative to the size of the adaptors, such that adaptor binding to one site hinders binding events on other sites.
(3) Binding of adaptor j occurs with rate constants k_{onj} and k_{offj} to one or more modified sites. The assumption simplifies the initial derivation and can be relaxed to allow different on rate constants as indicated later in the manuscript.
(4) Modification and demodification of site i occur as effective first-order processes with rate constants k_{fi} and k_{ri} when adaptor is not bound.

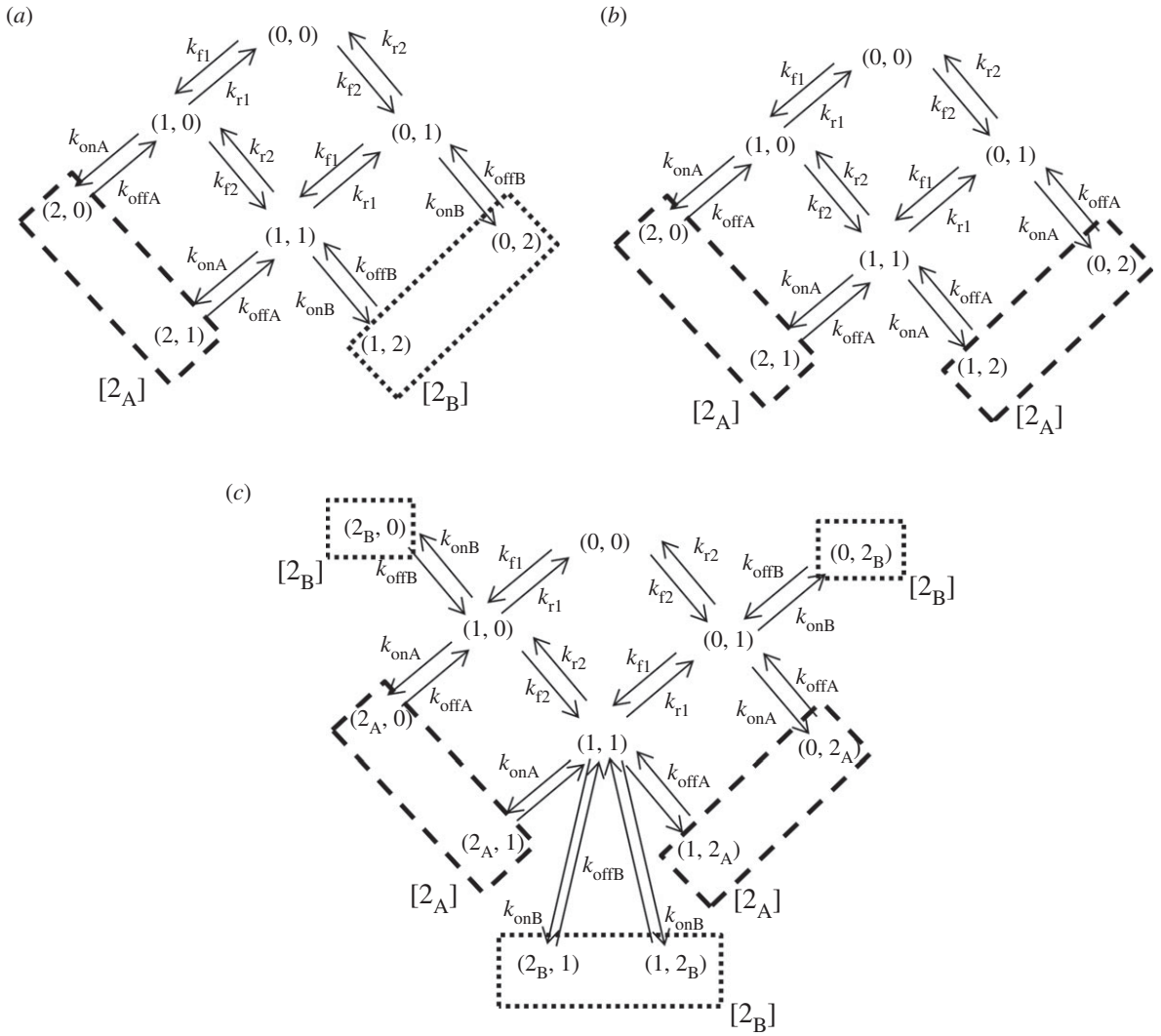

Figure 1. Kinetic schemes for two-site models. For each scheme, protein states are represented as an ordered pair, with the state of site 1 as the first element and the state of site 2 and the second element. 0 denotes unmodified, 1 denotes modified and 2 denotes bound. When different adaptors can bind to the same site (as in *c*), the subscript denotes the bound adaptor. Transitions between states are denoted with arrows, and these transitions occur with the pictured rate constants. States comprising bound adaptor A are denoted as $[2_A]$ with large enclosing dashes, whereas those comprising bound adaptor B are denoted as $[2_B]$ with small enclosing dashes. (*a*) Adaptor A binds to site 1 and adaptor B binds to site 2 in a mutually exclusive fashion. (*b*) Adaptor A binds to site 1 and site 2, but not at the same time. (*c*) Adaptor A and adaptor B compete for binding to site 1 and site 2 in a mutually exclusive fashion.

2.2. A simple two-site scenario

We first consider a simple scenario, where a protein has two sites, site 1 and site 2, which bind adaptors A and B, respectively (figure 1*a*). The states of the protein's sites are described by an ordered pair, where the first position corresponds to the state of site 1 and the second position corresponds to the state of site 2, 0 denotes unmodified, 1 denotes modified and 2 denotes bound to adaptor. Completely unmodified protein (0, 0) can be modified on any site in any order, modified site 1 can bind to adaptor A, and modified site 2 can bind to adaptor B. Because we assume that the sites are in close proximity and bound adaptor sterically hinders access to other sites, we do not consider that enzymes can access any site when adaptor is bound (e.g. (2, 1) cannot go to (2, 0) and vice versa). However, whether these processes are allowed or not has no bearing on the results of the theory derived in this manuscript.

Many systems function by modulating modified site levels to recruit proteins which subsequently turn on signalling cascades to enact biological function. Thus, the total lumped level of recruited protein is often of biological significance. In ErbB receptor signalling, for example, the recruited proteins SOS, PI-3K and PLCγ turn on the pathways for ERK, Akt and PKC, respectively, subsequently affecting cellular processes. We seek to derive expressions for the total, lumped amount of protein recruitment, while retaining the ability to capture, mechanistically, the effects of site-specific phosphorylation and/or dephosphorylation, differential binding affinities and varying protein expression levels on such recruitment. This lumping of recruitment is depicted in figure 1*a* by dashed rectangles, with $[2_A]$ representing the total amount of bound adaptor A, and likewise with $[2_B]$. Thus, overall, we seek a reduced representation for the dynamics of $[2_A]$ and $[2_B]$. These differential equations are given as

$$\frac{d[2_A]}{dt} = k_{onA}[(1,0)+(1,1)][B_A] - k_{offA}[2_A] \qquad (2.1)$$

and

$$\frac{d[2_B]}{dt} = k_{onB}[(0,1)+(1,1)][B_B] - k_{offB}[2_B], \qquad (2.2)$$

where $[B_A]$ and $[B_B]$ represent the bulk concentrations of adaptors A and B, respectively. The quantity $[(1,0)+(1,1)]$

represents the total concentration of modified site 1, and $[(0,1)+(1,1)]$ the same for site 2. From equations (2.1) and (2.2), it is clear that if we can represent $[(1,0)+(1,1)]$ and $[(0,1)+(1,1)]$ in terms of a common variable, then we may be able to close the system and negate the need to track the dynamics of each protein 'microstate'.

Consider now assumption 1 above, which states that modification cycles are in quasi-equilibrium. For the first site, this gives

$$\frac{k_{f1}}{k_{r1}} = \frac{(1,0)}{(0,0)} \Rightarrow (1,0) = \frac{k_{f1}}{k_{r1}}(0,0). \tag{2.3}$$

Likewise, we can also obtain

$$(0,1) = \frac{k_{f2}}{k_{r2}}(0,0);\ \frac{k_{f2}}{k_{r2}} = \frac{(1,1)}{(1,0)} \Rightarrow (1,1) = \frac{k_{f1}k_{f2}}{k_{r1}k_{r2}}(0,0). \tag{2.4}$$

Following this approach and substituting into equations (2.1) and (2.2), we obtain

$$\frac{\mathrm{d}[2_A]}{\mathrm{d}t} = k_{onA}\left(\frac{k_{f1}}{k_{r1}} + \frac{k_{f1}k_{f2}}{k_{r1}k_{r2}}\right)(0,0)[B_A] - k_{offA}[2_A] \tag{2.5}$$

and

$$\frac{\mathrm{d}[2_B]}{\mathrm{d}t} = k_{onB}\left(\frac{k_{f2}}{k_{r2}} + \frac{k_{f1}k_{f2}}{k_{r1}k_{r2}}\right)(0,0)[B_B] - k_{offB}[2_B]. \tag{2.6}$$

Now, only five variables appear in these equations: the two lumped states ($[2_A]$ and $[2_B]$), the two free adaptor concentrations and the amount of completely unmodified protein (0, 0). It is trivial to write the differential equations for the free adaptor concentrations based on species balances

$$\frac{\mathrm{d}[B_A]}{\mathrm{d}t} = -\frac{\mathrm{d}[2_A]}{\mathrm{d}t};\quad \frac{\mathrm{d}[B_B]}{\mathrm{d}t} = -\frac{\mathrm{d}[2_B]}{\mathrm{d}t}. \tag{2.7}$$

To close the system and obtain an exact reduced description for the lumped states, we must now represent the differential equation for (0, 0) in terms of only modelled quantities, $[2_A]$ and $[2_B]$. Based on overall protein species balance, we have

$$\frac{-\mathrm{d}[(1,0)+(1,1)+(0,1)+(0,0)]}{\mathrm{d}t} = \frac{\mathrm{d}[2_A]}{\mathrm{d}t} + \frac{\mathrm{d}[2_B]}{\mathrm{d}t}. \tag{2.8}$$

Substituting in quasi-equilibrium relationships (equations (2.3) and (2.4)), we obtain

$$\begin{aligned}
&-\mathrm{d}\left[\frac{k_{f1}}{k_{r1}}(0,0) + \frac{k_{f1}k_{f2}}{k_{r1}k_{r2}}(0,0) + \frac{k_{f2}}{k_{r2}}(0,0) + (0,0)\right]/\mathrm{d}t = \frac{\mathrm{d}[2_A]}{\mathrm{d}t} + \frac{\mathrm{d}[2_B]}{\mathrm{d}t}\\
&\frac{\mathrm{d}(0,0)}{\mathrm{d}t}\left[\frac{k_{f1}}{k_{r1}} + \frac{k_{f1}k_{f2}}{k_{r1}k_{r2}} + \frac{k_{f2}}{k_{r2}} + 1\right] = -\left(\frac{\mathrm{d}[2_A]}{\mathrm{d}t} + \frac{\mathrm{d}[2_B]}{\mathrm{d}t}\right)\\
&\frac{\mathrm{d}(0,0)}{\mathrm{d}t} = -\alpha[k_{onA}\beta_1(0,0)[B_A] - k_{offA}[2_A]\\
&\qquad + k_{onB}\beta_2(0,0)[B_B] - k_{offB}[2_B]],
\end{aligned} \tag{2.9}$$

where

$$\begin{aligned}
&\frac{1}{\alpha} \equiv \left[\frac{k_{f1}}{k_{r1}} + \frac{k_{f1}k_{f2}}{k_{r1}k_{r2}} + \frac{k_{f2}}{k_{r2}} + 1\right];\quad \beta_1 \equiv \frac{k_{f1}}{k_{r1}} + \frac{k_{f1}k_{f2}}{k_{r1}k_{r2}};\\
&\beta_2 \equiv \frac{k_{f2}}{k_{r2}} + \frac{k_{f1}k_{f2}}{k_{r1}k_{r2}}.
\end{aligned} \tag{2.10}$$

Thus, given the above assumptions, the reduced system can be described exactly by a set of three differential equations, none of which depends explicitly on the modification site states.

2.3. Two-sites, one adaptor

Now, consider the case when a single adaptor A binds to both site 1 and site 2 (figure 1*b*). In this case, we have

$$\begin{aligned}
&[2_A] \equiv [(2,0)+(2,1)+(0,2)+(1,2)]\\
&\frac{\mathrm{d}[2_A]}{\mathrm{d}t} = k_{onA}[(0,1)+(1,0)+2^*(1,1)][B_A] - k_{offA}[2_A].
\end{aligned} \tag{2.11}$$

Note here the coefficient of two prior to the (1, 1) state, indicating that adaptor A can bind to either modification site, doubling the concentration-driving force. Similar to above, by substituting in the quasi-equilibrium relationships, we obtain

$$\begin{aligned}
\frac{\mathrm{d}[2_A]}{\mathrm{d}t} &= k_{onA}\left[\frac{k_{f1}}{k_{r1}} + \frac{k_{f2}}{k_{r2}} + 2\frac{k_{f1}k_{f2}}{k_{r1}k_{r2}}\right](0,0)[B_A] - k_{offA}[2_A]\\
&= k_{onA}[\beta_1 + \beta_2](0,0)[B_A] - k_{offA}[2_A]
\end{aligned} \tag{2.12}$$

and

$$\frac{\mathrm{d}(0,0)}{\mathrm{d}t} = -\alpha[k_{onA}[\beta_1 + \beta_2](0,0)[B_A] - k_{offA}[2_A]]. \tag{2.13}$$

Here, β_1 and β_2 are defined as above in equation (2.10). Thus, in this case, this system can be reduced even further, needing only one state for the single lumped adaptor and one for the total amount of unmodified protein.

2.4. Two sites, two adaptors, overlapping binding

Now, we consider a slightly more complex scenario where the two adaptors A and B may both bind to sites 1 and 2, competing with each other (figure 1*c*). Because different adaptors may be bound to the same site, we clarify this ambiguity by denoting adaptor identities in the schematic with a subscript (i.e. 2_A and 2_B). The lumped bound adaptor states and differential equations are given by

$$\begin{aligned}
&[2_A] \equiv [(2_A,0)+(2_A,1)+(0,2_A)+(1,2_A)]\\
&\frac{\mathrm{d}[2_A]}{\mathrm{d}t} = k_{onA}[(0,1)+(1,0)+2^*(1,1)][B_A] - k_{offA}[2_A]
\end{aligned} \tag{2.14}$$

and

$$\begin{aligned}
&[2_B] \equiv [(2_B,0)+(2_B,1)+(0,2_B)+(1,2_B)]\\
&\frac{\mathrm{d}[2_B]}{\mathrm{d}t} = k_{onB}[(0,1)+(1,0)+2^*(1,1)][B_B] - k_{offB}[2_B].
\end{aligned} \tag{2.15}$$

As above, the modified, unbound protein states can be recast through quasi-equilibrium relationships (equations (2.3) and (2.4)) which give

$$\begin{aligned}
\frac{\mathrm{d}[2_A]}{\mathrm{d}t} &= k_{onA}\left[\frac{k_{f1}}{k_{r1}} + \frac{k_{f2}}{k_{r2}} + 2\frac{k_{f1}k_{f2}}{k_{r1}k_{r2}}\right](0,0)[B_A] - k_{offA}[2_A]\\
&= k_{onA}[\beta_1 + \beta_2](0,0)[B_A] - k_{offA}[2_A]c
\end{aligned} \tag{2.16}$$

and

$$\begin{aligned}
\frac{\mathrm{d}[2_B]}{\mathrm{d}t} &= k_{onB}\left[\frac{k_{f1}}{k_{r1}} + \frac{k_{f2}}{k_{r2}} + 2\frac{k_{f1}k_{f2}}{k_{r1}k_{r2}}\right](0,0)[B_B] - k_{offB}[2_B]\\
&= k_{onB}[\beta_1 + \beta_2](0,0)[B_B] - k_{offB}[2_B].
\end{aligned} \tag{2.17}$$

Again, the β factors are as defined in equation (2.10). Following equations (2.8)–(2.10) for the dynamics of the (0, 0) state, we obtain

$$\begin{aligned}
\frac{\mathrm{d}(0,0)}{\mathrm{d}t} &= -\alpha[k_{onA}[\beta_1 + \beta_2](0,0)[B_A] - k_{offA}[2_A]\\
&\quad + k_{onB}[\beta_1 + \beta_2](0,0)[B_B] - k_{offB}[2_B]].
\end{aligned} \tag{2.18}$$

Thus, just as above, the system can be reduced in a similar manner, and the dynamics of lumped adaptor states can be exactly described with only three equations.

Because next we will derive the results for a system of general size, it is instructive at this point to rehash the general procedure we have used in these different scenarios to derive an exact reduction of the system dynamics in terms of lumped adaptor states:

(1) define lumped, bound states for each adaptor and specify their differential equations;
(2) use quasi-equilibrium relationships to represent the concentration-driving force for adaptor binding to modified proteins in terms of the concentration of completely unmodified protein; and
(3) use overall species balance and quasi-equilibrium relationships to derive the differential equation for completely unmodified protein.

Next, we follow this procedure to derive equations for the general situation.

2.5. A protein with *n* sites and *m* adaptors

Based on the above analysis of various two-site, two-adaptor systems, we now consider the general case when there are n modification sites and m adaptors. First, we present some definitions:

(1) let $[2_j]$ denote the summation over all states that have adaptor j bound, where $j \in \{1, 2, \ldots, m\}$;
(2) let [0] be the state where all n sites are unmodified;
(3) let β_i denote the summation over all states that have site i modified ($i \in \{1, 2, \ldots, n\}$), but are not bound to adaptor on any other site. As in equation (2.10), the βs are cast in terms of quasi-equilibrium relations with [0] factored out;
(4) let d_{ij} be an indicator variable, taking the value of 1 if adaptor j binds to site i, and 0 otherwise; and
(5) let $1/\alpha$ denote the summation over all unbound protein states. As in equation (2.10), it is written in terms of quasi-equilibrium relations with [0] factored out.

Given these definitions, the differential equation for the lumped state for adaptor j $[2_j]$ is

$$\frac{\mathrm{d}[2_j]}{\mathrm{d}t} = k_{\mathrm{on}j}[B_j][0]\sum_{i=1}^{n}\beta_i d_{ij} - k_{\mathrm{off}j}[2_j]. \tag{2.19}$$

The term $[0]\sum_{i=1}^{n}\beta_i d_{ij}$ quantifies the total concentration of modified proteins that adaptor j is able to bind, and assumption 2 above allows us to neglect any bound states from this term. The summation adds up concentrations of only those modified sites where adaptor j is able to bind ('filtered' by the indicator variable). We also note that is possible to relax assumption 3 for on rate constants by incorporating them inside the summation, into the indicator variables. Before writing the expression for the β factors, recall from above (equations (2.3) and (2.4)) that based on quasi-equilibrium relationships, in general, the concentration of a protein with modified sites given by indices in the set X, C_X, is given by

$$C_X = [0]\prod_{i=1;i\in X}^{n}\frac{k_{\mathrm{f}i}}{k_{\mathrm{r}i}}. \tag{2.20}$$

Now, for ease of notation, let $K_i \equiv k_{\mathrm{f}i}/k_{\mathrm{r}i}$. As an example, for a five-site system and $X = [1,3,5]$, we would have $C_x = [0]K_1K_3K_5$. Given equation (2.20), the β factors (see definition 4) are given by the following

$$\beta_i = K_i\left[1 + \sum_{q=1;q\neq i}^{n}\left[K_q + K_q\sum_{r=q+1;r\neq i}^{n}\left[K_r + \cdots + K_y\sum_{z=n;z\neq i}^{n}K_z\right]\right]\right]. \tag{2.21}$$

Equation (2.21) iterates over all potential combinations of modified protein states where site i is modified. The constant factor K_i in front dictates that site i must be modified. The first summation accounts for all doubly modified states, the second summation for all triply modified states, and so on. There are a total of $n-1$ nested summations; the index z on the last summation has a value of n in the first complete loop through all the summations and therefore executes only once, accounting for the state where every site is modified.

The differential equation for the [0] state is, as above, given by overall species balance

$$\frac{\mathrm{d}[0]}{\mathrm{d}t} = -\alpha\sum_{j=1}^{m}\frac{\mathrm{d}[2_j]}{\mathrm{d}t}, \tag{2.22}$$

where the α factor (see Definition 5) is defined by

$$\frac{1}{\alpha} \equiv 1 + \sum_{q=1}^{n}\left[K_q + K_q\sum_{r=q+1}^{n}\left[K_r + K_r\sum_{s=r+1}^{n}\left[K_s + \cdots + K_y\sum_{z=n}^{n}K_z\right]\right]\right]. \tag{2.23}$$

Equation (2.23) gives all potential combinations of modified sites. Here, there are n nested summations. The first summation corresponds to states having only a single site modified, and second summation corresponds to states having any two sites modified and so on. Similar to above, the index z corresponding to the last summation has a value of n in the first loop through, and it therefore executes only once and corresponds to the state where all sites are modified.

From this analysis, it can be seen that when the aforementioned assumptions apply, the number of species needed to describe the lumped dynamics of bound adaptor states exactly is $m+1$ (one for each lumped adaptor state and one for the totally unmodified protein), and does not depend on the number of modification sites n. Thus, by applying these assumptions, one can side step the combinatorial explosion of potential states arising from multiple modification sites, while still retaining the ability to model effects of site-specific phenomena accurately.

2.6. Comparison of theory with simulations

The abovementioned equations dictate the dynamics of a reduced system assuming the specified conditions hold exactly. However, in practice, these conditions do not hold exactly; they only hold to an approximation. Under what conditions do theoretical predictions match exact simulations? Here, we perform simulation studies to answer these questions with a BioNetGen model of a 10-site receptor with four binding proteins that fully considers combinatorial complexity (see Material and methods and figure 2*a*). This model contains 6148 ODEs.

First, we investigated how the main assumption of time-scale separation between modification/demodification rates and binding on rates affected the agreement between

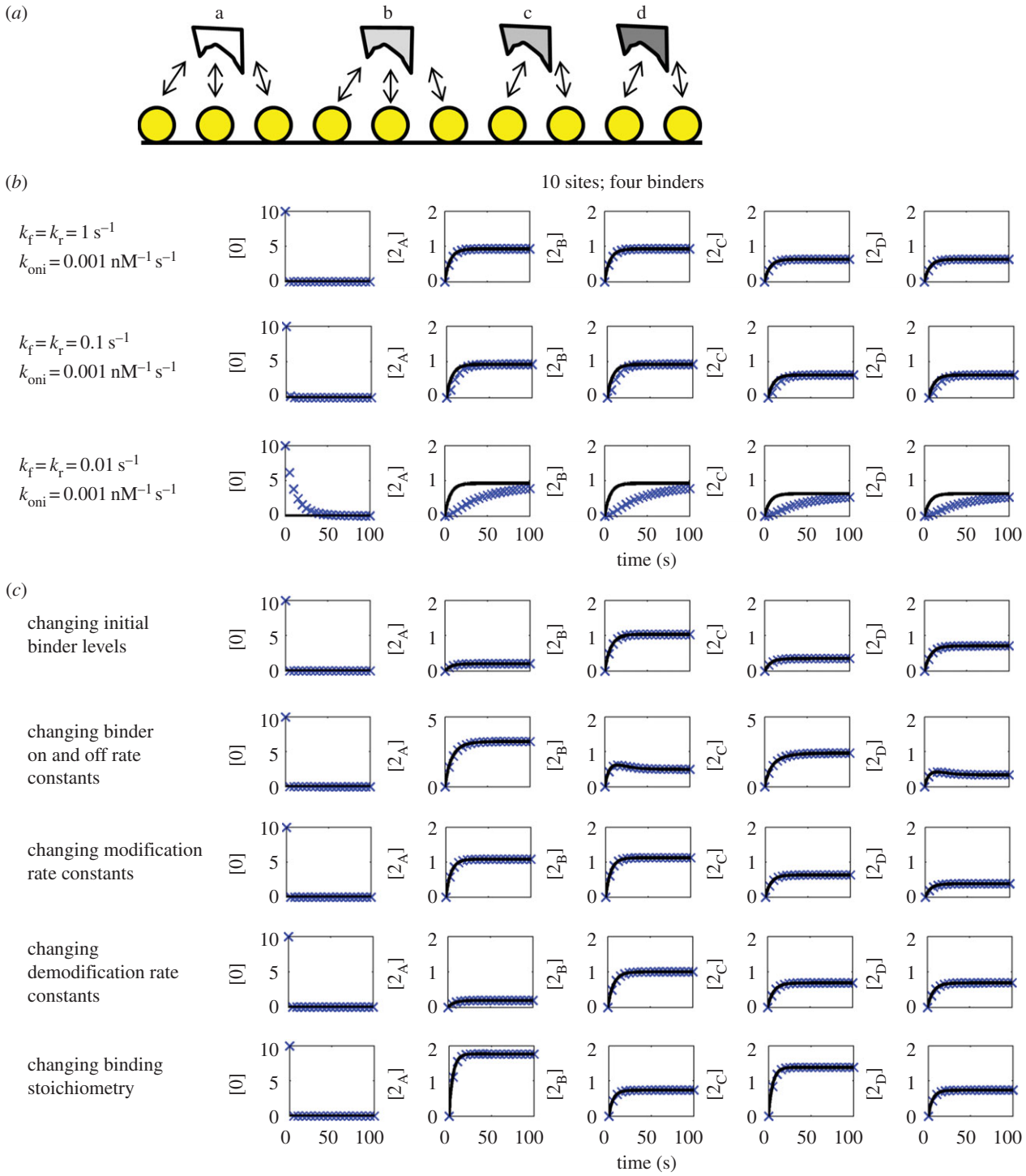

Figure 2. Comparing theory to a combinatorially complex 10-site model. (*a*) Simplified cartoon schematic of the 10-site model. Ten modification sites (denoted by yellow circles) can bind to four different proteins (labelled a–d). Arrows to sites denote allowed binding events. (*b*) Comparison of BioNetGen-based simulations of the full model (blue *x*'s) to theoretical predictions (black lines) for various time-scale differences between modification and binding on rate constants. Parameter values are indicated to the left of the plots and in materials and methods. (*c*) Comparison for a host of parameter variations, as indicated by text to the left of the plots. In the first row, A levels were reduced by 80% and C levels were reduced by 50%. In the second row, on rate constants for A and C were multiplied by 3, and off rate constants for A and C were divided by 3. In the third row, modification rate constants were multiplied by 4, 7 and 0.1 for the first, fourth, and 10th sites, respectively. In the fourth row, demodification rate constants were multiplied by 10 for sites 1, 2 and 3. In the fifth row, the **d** matrix was randomly populated with 0 s and 1 s, which corresponded to: A binding sites 3, 4, 5, 6, 7, 8, 9 and 10; B binding sites 3, 9 and 10; C binding sites 1, 2, 3, 4, 6 and 9; and D binding sites 4, 8 and 10. (Online version in colour.)

BioNetGen simulations and our theory (figure 2*b*—first three rows). Here, binder levels and affinities are taken as equal for clarity of presentation (they are altered directly below). When there was a 100-fold or greater difference between modification rate constants and the binding on-rate, the agreement between our theory (black lines) and BioNetGen-based ODE simulations (blue *x*'s) was excellent. When the difference was 10-fold, discrepancies between theory and simulation were observed, although they disappeared as the system reached steady-state. Thus, our theory may also be applicable in such scenarios so long as discrepancies occurring on the order of 1 min are not a concern. Many meaningful biological changes occur over

much longer timescales (hours), so such discrepancies may in fact be quite tolerable, and worth the trade-off for model reduction capacity provided by our theory.

We further challenged our theory to account for a number of biologically plausible scenarios (figure 2*c*): changing binder concentrations (first row); changing on and off rate constants for binders (second row); changing site-specific modification rate constants (third and fourth row) and changing the binding stoichiometry (fifth row—**d** matrix). In each case, predictions from our theory match well to simulations from the full combinatorial complexity model. Thus, over a wide range of conditions, our theory was able to reduce the dimension of the system from 6148 to five independent ODEs, over a 10^3-fold reduction. Further reduction is possible, if a larger combinatorially complex model is considered.

2.7. Applying the theory to model ligand-induced epidermal growth factor receptor activation

Here, we show how our theory can be applied to a practical scenario where EGF binds to the EGFR, causing signalling through asymmetric dimers (figure 3*a*) [32]. Before proceeding, however, we must evaluate whether the four assumptions underlying our theory may reasonably apply to the EGFR system. This analysis of assumptions is presented in the electronic supplementary material. Although there are certainly some potential deviations from the assumptions in a strict sense, in many respects, we conclude that they may reasonably apply to the EGFR system. One significant deviation is the cooperative recruitment of the E3 ubiquitin ligase c-Cbl to the EGFR, depending on the presence of Grb2 [33]. The theory as developed above does not allow for such a possibility, but we derive a special case here that allows us to introduce this feature without considering the full array combinatorial complexity. The derivation is also shown in the electronic supplementary material, and our simulation analysis of this scenario is presented below.

2.7.1. Choosing tyrosine phosphorylation sites, binders and rate constant parameters

There are multiple potential tyrosine phosphorylation sites on EGFR and many different proteins which can bind to their phosphorylated form. We limited ourselves to EGFR tyrosines that are that are C-terminal to the kinase domain or reasonably close to it, eliminating tyrosines N-terminal of Y920, yielding nine sites (table 1). To reduce the number of potential binding proteins, we relied on two sets of experimental data. First, we required that binding proteins have less than a 3 µM K_d to any EGFR tyrosine phosphorylation site [13]. Second, we required that binding proteins have detectable concentrations as measured in NIH3T3 cells [16]. This resulted in six binding proteins—Grb2, PI3K (PIK3R1—we assume the SH2 domain containing regulatory subunit controls levels of the complete PI3K complex), PLCγ1, SHP2 (PTPN11), RasA1 (p120 RasGTPase activating protein) and Shc1, each with known expression levels and unique dissociation constants for the considered phosphotyrosines (table 1). Taking $k_{off} = 0.1\ s^{-1}$ (see the electronic supplementary material) gives the k_{on}s. Site-specific kinase catalytic constants were previously measured [34], and we assumed $k_f = 0.17\ s^{-1}$ if no site-specific data were available (table 1). First-order phosphatase rate constants were taken as $0.1\ s^{-1}$ [35].

2.7.2. Simulating epidermal growth factor-induced primary binding responses

With the model specified based on a variety of experimental data as detailed above, we could now perform simulations of EGF-induced binding of these six proteins to the variety of EGFR phosphotyrosines. We compared the predictions of our reduced model theory, which consists of only 11 independent ODEs, to simulations using NFsim accounting for combinatorial complexity [24] (our computers ran out of memory when trying to generate the network with BioNetGen and perform ODE simulations). Our reduced model predictions for the dynamics of primary binder recruitment to EGF match well to NFsim results (figure 3*b*—first two rows). There is some discrepancy between our theory and NFsim results for the completely dephosphorylated receptor state at short times. We suspect this discrepancy is mainly due to the time required for newly formed receptor dimers to reach a phosphorylation equilibrium, which our theory treats as instantaneous. Grb2 and Shc1 are predicted to be the predominant primary binding proteins, with the others having recruitment levels approximately an order of magnitude lower. Interestingly, the mechanism by which Grb2 and Shc1 are the predominant binders are predicted to be different based on table 1. Shc1 has moderate abundance but high affinity for many sites, whereas Grb2 has high abundance which makes up for its low affinity to a single site.

2.7.3. Analysis of SHP2

Previous models have elucidated mechanisms of site-specific signalling by SHP2 in a variety of systems [36–38]. It has been shown that SHP2 has activity specifically towards Y992 on EGFR [39,40]. We simulated this situation by allowing SHP2-bound states to exhibit phosphatase activity towards pY992 (figure 3*b*, third row), which predicted that recruitment of RasA1, a RasGAP that turns off the ERK pathway, would be inhibited. This is functionally similar to simulated effects of a Y992F mutation (figure 3*b*—fourth row). These simulation results are consistent with previous experimental studies on both scenarios, in that SHP2 is needed for activation of the ERK pathway [39,41,42], and that expression of the Y992F EGFR mutant leads to sustained EGF-induced ERK activity [43].

We found that the catalytic efficiency (k_{cat}/K_m) for SHP2 had to be on the order of $100\ s^{-1}\ nM^{-1}$ to have an appreciable effect on site-specific phosphorylation levels (results in figure 3*b* are for $1000\ s^{-1}\ nM^{-1}$). This is a remarkable seven orders of magnitude higher than reported values of $10^{-5}\ s^{-1}\ nM^{-1}$ [44]. Perhaps, *in vivo* phosphatase activation processes play a role, but because the reported k_{cat}s for these phosphatases are already quite high (approx. $10\ s^{-1}$), such mechanisms are unlikely to explain the entire effect. Another possibility is local concentration of the receptor-phosphatase complex with its phosphorylated substrates. If one simply considers concentration of the phosphatase at the plasma membrane and 100 nm of cytoplasm associated with it, one obtains a 60-fold increase in effective concentration (assuming a 20 µm cell diameter and 2000 μm^3 cell volume). Because this factor would apply to both phosphatase and its substrate, these factors multiply to give a 3600-fold increase (60×60) in reaction rate. Even further concentration within membrane microdomains such as clathrin-coated pits is known to occur for many RTK systems [45]. It is estimated that clathrin-coated pits account for approximately 2% of the plasma membrane area [46].

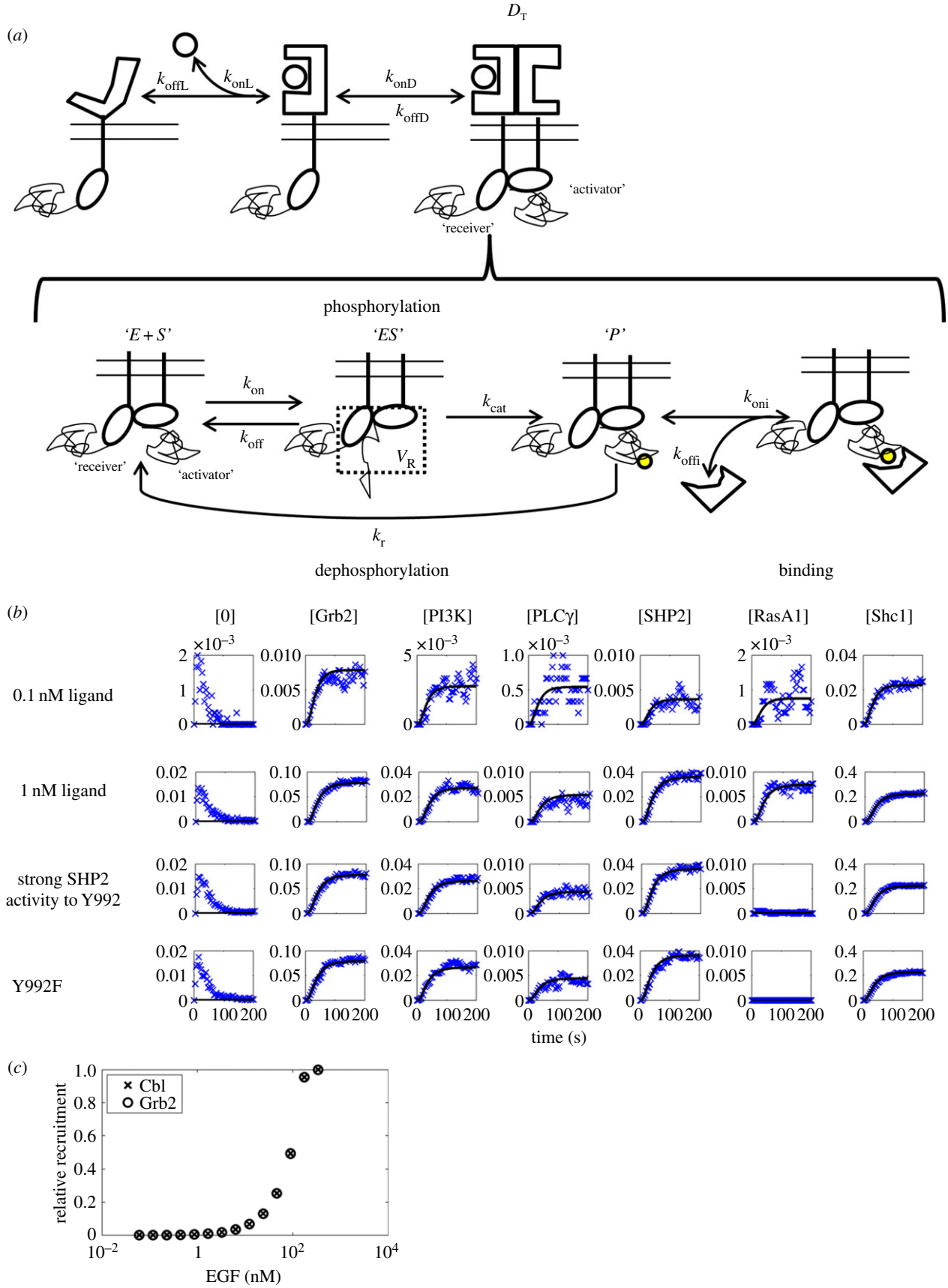

Figure 3. Modelling site-specific EGFR signalling without combinatorial complexity. (*a*) The ligand (small circle) binds to the receptor, stabilizing a conformation that can dimerize with another receptor, forming an asymmetric dimer. These dimers, referred to as D_T, are catalytically active. Phosphorylation of receptor-tethered substrate is accomplished via the 'trans' mechanism, whereby the 'receiver' kinase domain phosphorylates tyrosines on the C-terminus of the 'activator' receptor (phosphorylation denoted by the yellow circle). The effective volume of the phosphorylation reaction, V_R, is denoted by the dotted square. Species analogous to a classical enzyme kinetics scheme for phosphorylation are indicated in quotes above their cartoon representation. Phosphorylated receptor can be dephosphorylated, or can be bound by cytoplasmic adaptor protein. (*b*) Comparing NFsim simulation results to those predicted by our theory. Columns correspond to different species as indicated, whereas rows are simulations under different conditions (indicated by text to the left of the plots). All *y*-axes are concentrations in nM. First and second rows are responses to 0.1 and 1 nM ligand without SHP2 activity. All other rows consider SHP2 activity with $k_{dcat}/K_{dm} = 1000\ \text{nM}^{-1}\ \text{s}^{-1}$ for Y992 only. (*c*) Simulated dose–response of Grb2 and cCbl recruitment to EGFR given high-affinity binding of Grb2–cCbl complexes to dually phosphorylated Y1045/Y1068. (Online version in colour.)

Table 1. Parameters for the EGFR model. Abundances for each protein in the model are listed at the bottom, and phosphorylation rate constant parameters for each site are listed on the left. Site-specific parameters for protein binding to phosphotyrosine are given in the centre ($K_d = k_{off}/k_{on}$).

k_{cat} (min^{-1})	EGFR	Grb2	PI3K	PLCγ1	SHP2	RasA1	Shc1
	tyrosine sites			K_d (μM)			
10	920		0.07	1.41	0.69		2.31
10	954			0.51			
10	974						2.84
14.4	992		2.4	1.93		1.1	
17.4	1068	2.6					
17.2	1086						0.41
12.9	1114						
13.1	1148						0.07
15.1	1173						1.42
abundance (nM)[a]	7.05	4.58	4.97	4.43	66.6	19.4	31.3

[a]Based on a 2000 μm^3 cell.

Taking this further concentration factor into account for phosphatase and substrate (50×50) yields the approximately 10^7-fold increase in effective reaction rate needed for site-specific dephosphorylation to have appreciable effects on phosphorylated receptor levels. Thus, these modelling results suggest that local concentration of phosphatase with phosphorylated receptors in plasma membrane subdomains is a key feature of site-specific dephosphorylation.

2.7.4. Refining the model for analysis of Grb2–cCbl cooperativity

A recent study has shown that cCbl is cooperatively recruited to EGFR as EGF dose is increased, and that this cooperativity is dependent on Grb2 and the presence of the recruitment sites for both Grb2 (Y1068) and cCbl (Y1045) [33]. It was suggested that this effect may be due to cooperativity on the level of the phosphorylation sites themselves, implying that a Grb2–cCbl complex may bind cooperativity to EGFR, because Y1045 and Y1068 are more likely to be phosphorylated together. We extended our theory to accommodate such cooperative behaviour by adding Y1045, cCbl, and the Grb2–cCbl complex to the model (see the electronic supplementary material), which resulted in only a slight increase of model complexity (21 states). We assumed that cCbl alone could bind to phoshphorylated Y1045 with similar affinity as Grb2 for its site, and the Grb2–cCbl complex could bind to receptors having both Y1045 and Y1068 phosphorylated, with 100-fold greater affinity than either protein alone. To accomplish this, we created a new site that takes into account the simultaneous phosphorylation states of Y1045 and Y1068 while still using our model reduction assumptions (see the electronic supplementary material). Using this extended model, we performed simulations to explore whether these assumptions could reproduce the cooperative behaviour in cCbl binding to EGFR observed [33]. Surprisingly, we found that there was no predicted difference in recruitment of Grb2 or cCbl as a function of EGF dose (figure 3*c*), despite the higher affinity of the Grb2–cCbl complex and the reliance on two sites to be simultaneously phosphorylated. Thus, more complex mechanisms, such as ordered cooperative binding of Grb2 and cCbl to pY1045/pY1068, or increased phosphorylation of one of these sites as a result of phosphorylation of the other, may be needed to explain the observed cCbl dose responses.

3. Conclusion

Combinatorial complexity can cripple mechanistic modelling of signal transduction. Here, we show that given assumptions which may be biologically plausible, the dynamics of biologically relevant lumped adaptor states do not depend on the number of modification sites, eliminating combinatorial complexity when such assumptions are satisfied, while still describing the effects of site-specific phenomena. It may apply to, for example, RTK and histone modification systems. Yet, it is important to note that regulatory complexity, which is defined as the large number of parameters required to model such systems, remains a problem that is not addressed by the current theory but rather by other recently published approaches [47]. We demonstrate numerically that our theory for describing a reduced system of lumped adaptor states provides predictions which match well to the true system dynamics. This correspondence is maintained over a wide range of parameter values, the edge of which comes remarkably close to violating the theory's main assumptions, but without significant deviation between reduced and full model simulation results. Importantly, existing data suggest that our theory is applicable to the EGFR system, and we demonstrate how to apply it, even when the non-cooperative binding assumption must be relaxed for Grb2–cCbl interactions. Analysis of our EGFR model suggests that site-specific tyrosine dephosphorylation may only be possible by concentrating phosphatases with their substrates by a surprising 10^7-fold, which may be possible with membrane microdomain signalling. Wide application of this theory may allow for practical simulation of large biochemical networks that otherwise would not be feasible to explore key aspects of receptor signalling complexity in a succinct manner.

4. Material and methods

4.1. Simulation of the full 10-site model

BioNetGen (v. 2.2.2) was used to generate and simulate the ODEs for the full 10 site model described in figure 2, and MATLAB (R2013a, The Mathworks, Natick, MA) with `ode15s` was used to apply our theory. Unless otherwise specified as in figure 2 legend, initial concentrations of receptor and free binding proteins were 10 nM, modification rate constants were $1\ s^{-1}$, and

binder on and off rates were 0.001 s^{-1} nM^{-1} and 0.1 s^{-1}, respectively. The BNGL file associated with this model is given in the electronic supplementary mateial, text 1, and MATLAB code is provided in the electronic supplementary material.

4.2. Simulation of the ligand-induced signalling model

BioNetGen (v. 2.2.2) was used to specify the reaction rules for the ligand-induced EGFR model and NFsim (v. 1.11) was used for simulation. Ligand binding on and off rates was both 0.01 (s^{-1} nM^{-1} and s^{-1}) and dimerization on and off rates were 0.01 s^{-1} nM^{-1} and 0.001 s^{-1}. The BNGL file associated with this model is given in the electronic supplementary material, text 2, and MATLAB code is given in the electronic supplementary material. We assumed that newly formed receptor dimers would be completely unphosphorylated, but according to assumption 1, be rapidly converted to the equilibrium between phosphorylated and unphosphorylated receptor states. Moreover, in equation (2.10), the first-order kinase and phosphatase rate constants were time-invariant; however, here they may change over time. Given these considerations, we have

$$\frac{\mathrm{d}[0]}{\mathrm{d}t} = -\alpha\left[\sum_{j=1}^{m}\frac{\mathrm{d}[2_j]}{\mathrm{d}t} - \frac{\mathrm{d}D_\mathrm{T}}{\mathrm{d}t}\right] - [0]\frac{\mathrm{d}\ln(1/\alpha)}{\mathrm{d}t}, \qquad (4.1)$$

where D_T is the total concentration of active receptor dimers (figure 3*a*), which changes as a result of ligand binding and receptor dimerization processes. Because we had to track dynamics of the parameter α, we used `ode45` in MATLAB (R2013a) to integrate the ODEs. This allowed us to keep track of α-values between time steps numerically in a straightforward manner. Such tracking could potentially be accomplished with `ode15s` but would require more complex coding, because it is a variable-time step, iterative solver.

Acknowledgements. We acknowledge Avner Schlessinger and Joseph Schlessinger for helpful discussions.

Funding statement. Funding from the Icahn School of Medicine at Mount Sinai and the NIH grants P50 GM071558 (Systems Biology Center New York), R01GM104184 and U54HG008098 (LINCS Center).

References

1. Oda K, Matsuoka Y, Funahashi A, Kitano H. 2005 A comprehensive pathway map of epidermal growth factor receptor signaling. *Mol. Syst. Biol.* **1**, 20050010. (doi:10.1038/msb4100014)
2. Caron E, Ghosh S, Matsuoka Y, Ashton-Beaucage D, Therrien M, Lemieux S, Perreault C, Roux PP, Kitano H. 2010 A comprehensive map of the mTOR signaling network. *Mol. Syst. Biol.* **6**, 453. (doi:10.1038/msb.2010.108)
3. Kholodenko BN, Demin OV, Moehren G, Hoek JB. 1999 Quantification of short term signaling by the epidermal growth factor receptor. *J. Biol. Chem.* **274**, 30 169–30 181. (doi:10.1074/jbc.274.42.30169)
4. Bhalla US, Iyengar R. 1999 Emergent properties of networks of biological signaling pathways. *Science* **283**, 381–387. (doi:10.1126/science.283.5400.381)
5. Blinov ML, Faeder JR, Goldstein B, Hlavacek WS. 2006 A network model of early events in epidermal growth factor receptor signaling that accounts for combinatorial complexity. *Biosystems* **83**, 136–151. (doi:10.1016/j.biosystems.2005.06.014)
6. Faeder JR, Blinov ML, Goldstein B, Hlavacek WS. 2005 Combinatorial complexity and dynamical restriction of network flows in signal transduction. *Syst. Biol. (Stevenage)* **2**, 5–15. (doi:10.1049/sb:20045031)
7. Creamer MS *et al*. 2012 Specification, annotation, visualization and simulation of a large rule-based model for ERBB receptor signaling. *BMC Syst. Biol.* **6**, 107. (doi:10.1186/1752-0509-6-107)
8. Birtwistle MR, Hatakeyama M, Yumoto N, Ogunnaike BA, Hoek JB, Kholodenko BN. 2007 Ligand-dependent responses of the ErbB signaling network: experimental and modeling analyses. *Mol. Syst. Biol.* **3**, 144. (doi:10.1038/msb4100188)
9. Schoeberl B, Eichler-Jonsson C, Gilles ED, Muller G. 2002 Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat. Biotechnol.* **20**, 370–375. (doi:10.1038/nbt0402-370)
10. Montagner A, Yart A, Dance M, Perret B, Salles JP, Raynal P. 2005 A novel role for Gab1 and SHP2 in epidermal growth factor-induced Ras activation. *J. Biol. Chem.* **280**, 5350–5360. (doi:10.1074/jbc.M410012200)
11. Pylayeva-Gupta Y, Grabocka E, Bar-Sagi D. 2011 RAS oncogenes: weaving a tumorigenic web. *Nat. Rev. Cancer* **11**, 761–774. (doi:10.1038/nrc3106)
12. Wang F, Wang S, Wang Z, Duan J, An T, Zhao J, Bai H, Wang J. 2012 Phosphorylated EGFR expression may predict outcome of EGFR-TKIs therapy for the advanced NSCLC patients with wild-type EGFR. *J. Exp. Clin. Cancer Res.* **31**, 65. (doi:10.1186/1756-9966-31-65)
13. Hause Jr RJ, Leung KK, Barkinge JL, Ciaccio MF, Chuu CP, Jones RB. 2012 Comprehensive binary interaction mapping of SH2 domains via fluorescence polarization reveals novel functional diversification of ErbB receptors. *PLoS ONE* **7**, e44471. (doi:10.1371/journal.pone.0044471)
14. Lemmon MA, Schlessinger J. 2010 Cell signaling by receptor tyrosine kinases. *Cell* **141**, 1117–1134. (doi:10.1016/j.cell.2010.06.011)
15. Fan YX, Wong L, Johnson GR. 2005 EGFR kinase possesses a broad specificity for ErbB phosphorylation sites, and ligand increases catalytic-centre activity without affecting substrate binding affinity. *Biochem. J.* **392**, 417–423. (doi:10.1042/BJ20051122)
16. Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M. 2011 Global quantification of mammalian gene expression control. *Nature* **473**, 337–342. (doi:10.1038/nature10098)
17. Faeder JR, Blinov ML, Hlavacek WS. 2009 Rule-based modeling of biochemical systems with BioNetGen. *Methods Mol. Biol.* **500**, 113–167. (doi:10.1007/978-1-59745-525-1_5)
18. Blinov ML, Faeder JR, Goldstein B, Hlavacek WS. 2004 BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics* **20**, 3289–3291. (doi:10.1093/bioinformatics/bth378)
19. Feret J, Danos V, Krivine J, Harmer R, Fontana W. 2009 Internal coarse-graining of molecular systems. *Proc. Natl Acad. Sci. USA* **106**, 6453–6458. (doi:10.1073/pnas.0809908106)
20. Borisov NM, Markevich NI, Hoek JB, Kholodenko BN. 2006 Trading the micro-world of combinatorial complexity for the macro-world of protein interaction domains. *Biosystems* **83**, 152–166. (doi:10.1016/j.biosystems.2005.03.006)
21. Borisov NM, Markevich NI, Hoek JB, Kholodenko BN. 2005 Signaling through receptors and scaffolds: independent interactions reduce combinatorial complexity. *Biophys. J.* **89**, 951–966. (doi:10.1529/biophysj.105.060533)
22. Conzelmann H, Fey D, Gilles ED. 2008 Exact model reduction of combinatorial reaction networks. *BMC Syst. Biol.* **2**, 78. (doi:10.1186/1752-0509-2-78)
23. Conzelmann H, Saez-Rodriguez J, Sauter T, Kholodenko BN, Gilles ED. 2006 A domain-oriented approach to the reduction of combinatorial complexity in signal transduction networks. *BMC Bioinformatics* **7**, 34. (doi:10.1186/1471-2105-7-34)
24. Sneddon MW, Faeder JR, Emonet T. 2011 Efficient modeling, simulation and coarse-graining of biological complexity with NFsim. *Nat. Methods* **8**, 177–183. (doi:10.1038/nmeth.1546)
25. Colvin J, Monine MI, Gutenkunst RN, Hlavacek WS, Von Hoff DD, Posner RG. 2010 RuleMonkey: software for stochastic simulation of rule-based models. *BMC Bioinformatics* **11**, 404. (doi:10.1186/1471-2105-11-404)

26. Karp RL, Perez Millan M, Dasgupta T, Dickenstein A, Gunawardena J. 2012 Complex-linear invariants of biochemical networks. *J. Theor. Biol.* **311**, 130–138. (doi:10.1016/j.jtbi.2012.07.004)
27. Gunawardena J. 2012 A linear framework for time-scale separation in nonlinear biochemical systems. *PLoS ONE* **7**, e36321. (doi:10.1371/journal.pone.0036321)
28. Thomson M, Gunawardena J. 2009 The rational parameterization theorem for multisite post-translational modification systems. *J. Theor. Biol.* **261**, 626–636. (doi:10.1016/j.jtbi.2009.09.003)
29. Thomson M, Gunawardena J. 2009 Unlimited multistability in multisite phosphorylation systems. *Nature* **460**, 274–277. (doi:10.1038/nature08102)
30. Nakakuki T *et al*. 2010 Ligand-specific c-Fos expression emerges from the spatiotemporal control of ErbB network dynamics. *Cell* **141**, 884–896. (doi:10.1016/j.cell.2010.03.054)
31. Purvis JE, Karhohs KW, Mock C, Batchelor E, Loewer A, Lahav G. 2012 p53 dynamics control cell fate. *Science* **336**, 1440–1444. (doi:10.1126/science.1218351)
32. Liu P, Bouyain S, Byrne PO, Longo PA, Leahy DJ. 2012 A single ligand is sufficient to activate EGFR dimers. *Proc. Natl Acad. Sci. USA* **109**, 10 861–10 866. (doi:10.1073/pnas.1201114109)
33. Sigismund S *et al*. 2013 Threshold-controlled ubiquitination of the EGFR directs receptor fate. *EMBO J.* **32**, 2140–2157. (doi:10.1038/emboj.2013.149)
34. Brignola PS *et al*. 2002 Comparison of the biochemical and kinetic properties of the type 1 receptor tyrosine kinase intracellular domains. Demonstration of differential sensitivity to kinase inhibitors. *J. Biol. Chem.* **277**, 1576–1585. (doi:10.1074/jbc.M105907200)
35. Kleiman LB, Maiwald T, Conzelmann H, Lauffenburger DA, Sorger PK. 2011 Rapid phospho-turnover by receptor tyrosine kinases impacts downstream signaling and drug binding. *Mol. Cell* **43**, 723–737. (doi:10.1016/j.molcel.2011.07.014)
36. Barua D, Faeder JR, Haugh JM. 2007 Structure-based kinetic models of modular signaling protein function: focus on Shp2. *Biophys. J.* **92**, 2290–2300. (doi:10.1529/biophysj.106.093484)
37. Gaffney EA, Heath JK, Kwiatkowska MZ. 2008 A mass action model of a fibroblast growth factor signaling pathway and its simplification. *Bull. Math. Biol.* **70**, 2229–2263. (doi:10.1007/s11538-008-9342-1)
38. Tan WH, Popel AS, Mac Gabhann F. 2013 Computational model of Gab1/2-dependent VEGFR2 pathway to Akt activation. *PLoS ONE* **8**, e67438. (doi:10.1371/journal.pone.0067438)
39. Agazie YM, Hayman MJ. 2003 Molecular mechanism for a role of SHP2 in epidermal growth factor receptor signaling. *Mol. Cell Biol.* **23**, 7875–7886. (doi:10.1128/MCB.23.21.7875-7886.2003)
40. Sugimoto S, Lechleider RJ, Shoelson SE, Neel BG, Walsh CT. 1993 Expression, purification, and characterization of SH2-containing protein tyrosine phosphatase, SH-PTP2. *J. Biol. Chem.* **268**, 22 771–22 776.
41. Noguchi T, Matozaki T, Horita K, Fujioka Y, Kasuga M. 1994 Role of SH-PTP2, a protein-tyrosine phosphatase with Src homology 2 domains, in insulin-stimulated Ras activation. *Mol. Cell Biol.* **14**, 6674–6682.
42. Yart A *et al*. 2001 A critical role for phosphoinositide 3-kinase upstream of Gab1 and SHP2 in the activation of ras and mitogen-activated protein kinases by epidermal growth factor. *J. Biol. Chem.* **276**, 8856–8864. (doi:10.1074/jbc.M006966200)
43. Tasaki S *et al*. 2010 Phosphoproteomics-based modeling defines the regulatory mechanism underlying aberrant EGFR signaling. *PLoS ONE* **5**, e13926. (doi:10.1371/journal.pone.0013926)
44. Peters GH, Branner S, Moller KB, Andersen JN, Moller NP. 2003 Enzyme kinetic characterization of protein tyrosine phosphatases. *Biochimie* **85**, 527–534. (doi:10.1016/S0300-9084(03)00036-1)
45. Sorkin A, Goh LK. 2009 Endocytosis and intracellular trafficking of ErbBs. *Exp. Cell Res.* **315**, 683–696. (doi:10.1016/j.yexcr.2008.07.029)
46. Lodish HB, Zipursky A, Matsudaira L, Baltimore P, Darnell J. 2000 *Molecular cell biology*, 4th edn. New York, NY: W. H. Freeman.
47. Ollivier JF, Shahrezaei V, Swain PS. 2010 Scalable rule-based modelling of allosteric proteins and biochemical networks. *PLoS Comput. Biol.* **6**, e1000975. (doi:10.1371/journal.pcbi.1000975)