A network model of early events in epidermal growth factor receptor signaling that accounts for combinatorial complexity

Michael L. Blinov, James R. Faeder, Byron Goldstein, William S. Hlavacek *

Theoretical Biology and Biophysics Group, Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Received 8 January 2005; received in revised form 6 May 2005; accepted 21 June 2005

Abstract

We consider a model of early events in signaling by the epidermal growth factor (EGF) receptor (EGFR). The model includes EGF, EGFR, the adapter proteins Grb2 and Shc, and the guanine nucleotide exchange factor Sos, which is activated through EGF-induced formation of EGFR–Grb2–Sos and EGFR–Shc–Grb2–Sos assemblies at the plasma membrane. The protein interactions involved in signaling can potentially generate a diversity of protein complexes and phosphoforms; however, this diversity has been largely ignored in models of EGFR signaling. Here, we develop a model that accounts for potential molecular diversity by specifying rules for protein interactions and then using these rules to generate a reaction network that includes all chemical species and reactions implied by the protein interactions. We obtain a model that predicts the dynamics of 356 molecular species, which are connected through 3749 unidirectional reactions. This network model is compared with a previously developed model that includes only 18 chemical species but incorporates the same scope of protein interactions. The predictions of this model are reproduced by the network model, which also yields new predictions. For example, the network model predicts distinct temporal patterns of autophosphorylation for different tyrosine residues of EGFR. A comparison of the two models suggests experiments that could lead to mechanistic insights about competition among adapter proteins for EGFR binding sites and the role of EGFR monomers in signal transduction.

Keywords: Computational systems biology; Receptor tyrosine kinase; Protein complexes; Rule-based modeling; Automatic network generation; BioNetGen

1. Introduction

Processes by which a cell senses and responds to its environment are often marked by combinatorial complexity (Hlavacek et al., 2003). Cellular signaling (Hunter, 2000) generally involves protein–protein interactions and enzymatic activities that imply a diversity of potential protein complexes and phosphoforms, which are difficult to simply enumerate let alone assay or understand. For example, the number of possible phosphoforms of a protein is $2^n$, where $n$ is the number of amino acid residues that are subject to phosphorylation and dephosphorylation by kinases and phosphatases, at least nine tyrosines in the case of epidermal growth factor (EGF) receptor (EGFR) (Jorissen et al., 2003). Additional molecular diversity can arise from the multivalent character of protein–protein interactions. A protein

* Corresponding author. Present address: Mail Stop K710, LANL, Los Alamos, NM 87545, USA. Tel.: +1 505 665 1355; fax: +1 505 665 3493.
E-mail address: wish@lanl.gov (W.S. Hlavacek).

© 2005 Elsevier Ireland Ltd. All rights reserved.
involved in signaling typically consists of multiple protein interaction domains (Pawson and Nash, 2003), such as the Src homology 2 (SH2) and 3 (SH3) domains of the Grb2 adapter protein. Each of these domains can interact with a partner that also contains multiple domains. As a result, proteins can combine in a variety of ways to form a spectrum of heterogeneous complexes. Proteomic studies confirm that diverse molecular species arise during signal transduction (Husi et al., 2000; Bunnell et al., 2002; Blagoev et al., 2003, 2004). Given the protein–protein interactions and enzymatic activities involved in the cellular response to a signal, how do we catalog the potential molecular species implied by these interactions and activities? How do we predict which of the possible molecular species might actually arise during signaling? How do we determine the functional implications of these molecular species or the relative importance of processes that influence them? How can we best use large-scale proteomic measurements to obtain mechanistic insights? These questions are being asked in the emerging field of systems biology, and mathematical models have an important role to play in addressing such questions (Bhalla and Iyengar, 1999; Endy and Brent, 2001; Wiley et al., 2003; Hlavacek et al., 2003; Goldstein et al., 2004). A mathematical model requires an explicit statement of our understanding (or assumptions) about how a signal transduction system operates in a form that allows, through computational analysis, the behavior of the system to be predicted and compared with experimental observations. Here, we provide a demonstration of how a mathematical model, incorporating detail at the level of protein sites and domains, can be used to study signal transduction with a comprehensive treatment of protein complexes and phosphoforms implied by protein interactions.

We develop and analyze a mathematical model for early events in signaling by EGFR, which is a well-studied cell-surface receptor involved in cell proliferation (Schlessinger, 2000; Jorissen et al., 2003). It has been the subject of numerous model-based studies (Wiley et al., 2003). Our model, which we will call the network model, provides a description of EGFR signaling that accounts for the spectrum of molecular species (356) and the reactions among these species (3749) implied by specified interactions and activities of EGFR, EGFR, the adapter proteins Grb2 and Shc, and the guanine nucleotide exchange factor Sos. These interactions and activities are the same as those considered in the seminal model of Kholodenko et al. (1999), which is based on assumptions (simultaneous phosphorylation and dephosphorylation of receptor tyrosines, inability of phosphorylated receptors in a dimer to dissociate, and competition among cytosolic proteins for receptor binding) that significantly limit, a priori, the number of molecular species that can arise during signaling. We will call the model of Kholodenko et al. (1999) the pathway-like model because it represents the signaling system as a set of reaction sequences rather than as a highly branched reaction network.

The rest of this report is organized as follows. First, we describe how the network model is constructed based on the proteins, interactions, and model parameters considered in the work of Kholodenko et al. (1999). Notably, the network model involves no more parameters than the pathway-like model. We then compare the predictions of the two models with the experimental observations of Kholodenko et al. (1999). We find both models are equally consistent. We also present new predictions of the network model and testable predictions that distinguish the two models. A comparison of the models allows us to evaluate the simplifying assumptions of Kholodenko et al. (1999). These assumptions have not been tested so far, even though this model has served as the starting point for a number of modeling studies of EGFR signaling (Schoeberl et al., 2002; Gong and Zhao, 2003; Hatakeyama et al., 2003; Resat et al., 2003; Conzelmann et al., 2004; Liu et al., 2005). We suggest experiments that could lead to insights into the mechanisms of signaling and determine which of the two models better represents signaling. Finally, we use the network model to predict the dynamics of the protein complexes and protein phosphorylation states that are generated during signaling. These predictions provide a picture of molecular diversity that is more detailed than could be currently obtained using the most sophisticated proteomic assays. For example, the model predicts which molecular species containing membrane-proximal Sos are prevalent at different time points. The model could also be used to predict how the population of these species depends on reaction dynamics and concentrations of components. As proteomic technologies mature, testing such predictions will become feasible.

2. The network model

2.1. Basis of the model

The network model (Fig. 1) is based on the same proteins, enzymatic activities and protein–protein interactions considered in the model of Kholodenko et al. (1999). The focus of this model is the cascade of signaling events that lead to recruitment of cytosolic Sos to the inner cell membrane (Fig. 1A and B), which can be described as follows. EGF binds to EGFR, which leads...
Fig. 1. Kinetic scheme for the early events in EGFR signaling. (A) Reactions involving receptors. Only EGFR tyrosine residues that are required for interactions with PLCγ, Grb2, and Shc are included. EGF–EGFR binding and receptor dimerization reactions are illustrated inside Box 1. Although only two reactions are illustrated, EGF may bind a receptor in any cytoplasmic state, and any two EGF-bound receptors can aggregate or dissociate at any time. After receptor aggregation different receptor phosphoforms may be formed as a result of receptor tyrosine kinase activity. Three of the possible phosphoforms are illustrated and named inside Box 2. Cytoplasmic binding reactions related to Sos recruitment and PLCγ activation are shown inside Boxes 3 and 4, respectively. Each reaction is an example of many possible reactions. (B) Cytosolic reactions not involving receptors. In panels A and B, numbers next to each reaction refer to signaling steps and reaction classes described in Table 1. (C) Example of a species omitted from consideration in the model of Khodolenko et al. (1999). (D) Illustration of 20 possible reactions the dimeric species of panel C may undergo in the network model. In this panel, reactions which are relevant only to PLCγ activation are represented inside the dashed boxes and are not included in the model for Sos activation.
to the formation of signaling-competent receptor dimers (Garrett et al., 2002; Ogiso et al., 2002; Ferguson et al., 2003; Burgess et al., 2003). A receptor in a dimer then can be transphosphorylated by the tyrosine kinase domain of the neighboring receptor (Schlessinger, 2000; Jorissen et al., 2003). The cytosolic adapter proteins Grb2 and Shc are recruited to phosphorylated receptor tyrosines Y1068 and Y1148/73 (Batter et al., 1994; Okabayashi et al., 1994). Shc is known to bind to two tyrosine residues, Y1148 via its SH2 domain and Y1173 via its phosphotyrosine binding (PTB) domain (Batter et al., 1994).

When Shc is bound to a receptor, it can be phosphorylated by EGF/R (Pelici et al., 1992). The phosphorylated form of Shc interacts with Grb2 (Lowenstein et al., 1992; Rozakis-Adcock et al., 1992; Sasaoka et al., 1994), which interacts constitutively with Sos (Egan et al., 1993; Rozakis-Adcock et al., 1993; Li et al., 1993).

Translocation of Sos from the cytosol to the membrane is required for activation of the GTPase Ras (Boguski and McCormick, 1993) and downstream signaling events. The model of Khodolenko et al. (1999) accounts for 25 species, including 18 involved in SOS recruitment. Some aspects of the model of Khodolenko et al. (1999) are controversial and might be subject to future investigation and refinement. For example, the model assumes that phosphorylation of Shc leads to a significant reduction in its affinity for EGF/R, which is primarily responsible for the predicted damping of the initial response to EGF. Although recent molecular dynamics simulations support a lower affinity of phosphorylated Shc for EGF/R (Suenga et al., 2004), the implication that Shc recruitment and phosphorylation negatively regulates signaling is problematic in light of earlier experimental work on EGF/R signaling (Sasaoka et al., 1994).

In addition, pre-formed dimers of EGF/R (Jorissen et al., 2003) and other complicating features of ligand-induced receptor dimerization that may influence signalization (Wofsy et al., 1992; Klein et al., 2004) are omitted in the model of Khodolenko et al. (1999) and its extensions. Our main focus here, however, is to evaluate the effects of simplifying assumptions made in developing the pathway-like model, and we therefore keep both the basic reaction processes and their accompanying rate constants in the network model so that we can make a controlled comparison of the two models.

We derive the network model without making the following assumptions, upon which the pathway-like model is based: (1) all receptor tyrosines of both receptors in a dimer are phosphorylated and dephosphorylated simultaneously, (2) dimeric receptors that are phosphorylated or associated with a cytosolic protein cannot dissociate, and (3) binding of cytosolic proteins to dimeric receptors is competitive, with only one cytosolic protein being allowed to associate with a dimer at a time. Without these assumptions, many more molecular species and reactions must be considered. Fig. 1C illustrates one of the molecular species considered only in the network model, and Fig. 1D illustrates the reactions in which this receptor dimer can participate. The network model is derived by applying a rule-based modeling approach (Hlavacek et al., 2003; Goldstein et al., 2004). Rules are specified based on the interactions and activities of protein domains. Each rule defines a reaction class that is composed of multiple reactions that are parameterized by identical rate constants (Tables 1 and 2). This approach to parameterization is based on the assumption that protein domains and sites are modular. In other words, we assume that the activity of, say, the binding site on EGF/R for Grb2 is independent of other sites on EGF/R. A representative of each reaction class is illustrated in Fig. 1A for reactions involving EGF/R and in Fig. 1B for cytosolic reactions. Each representative reaction corresponds with a “step” in the signaling cascade and with an individual reaction considered in the model of Khodolenko et al. (1999).

In Box 1 of Fig. 1A, Step 1 is EGF binding to monomeric EGF/R, and Step 2 is dimerization of EGF-bound receptors. In the network model, we allow modified receptors in dimers to dissociate, and as a result, there are multiple forms of monomeric EGF/R available to bind EGF once signaling begins (but not before EGF stimulation). In contrast, in the pathway-like model, only receptors without modification participate in Steps 1 and 2, at all times. In fact, monomeric EGF/R that is phosphorylated or associated with a cytosolic protein is not considered at all in the pathway-like model. To parameterize ligand-receptor and receptor-receptor interactions in the network model, we assume that the cytoplasmic state of a receptor does not affect ligand-receptor binding or ligand-stimulated receptor dimerization. Thus, we use the same rate constants estimated by Khodolenko et al. (1999) for all reactions involving the various possible states of receptors that can participate in Steps 1 and 2. Similar assumptions are made to parameterize the other steps in the signaling cascade (Tables 1 and 2). This approach to parameterization of the network model is the same approach taken to model early events in sig-
Signaling steps are illustrated in Fig. 1. Initial concentrations are [EGFR]total = 100 nM, [EGF]total = 680 nM, [Grb2]total = 58 nM, [ShcP]1068,1148–1173 = 27 nM, [ShcP]1148–1173 = 150 nM, [ShcP]1148–1173 = 7 nM. Parameter values are taken from Kholodenko et al. (1999). Some of these parameters are based on experimental measurements, but most are estimates obtained through a fitting procedure. Binding and phosphorylation reaction rate constants are denoted as k1 and dissociation and dephosphorylation reaction rate constants are denoted as k−1. First- and second-order rate constants are expressed in units of s−1 and mM−1 s−1, respectively. Modifications of particular rate constants made in this study are marked with footnotes.

This rate constant is that for a first-order reaction with a rate law of the form

\[ \frac{dC}{dt} = -kC \]

This rate constant is that for a second-order reaction with a rate law of the form

\[ \frac{dC}{dt} = -kC^2 \]

This rate constant is scaled by a factor of 0.5 to account for the different stoichiometry of binding sites in the model of Kholodenko et al. (1999).

Table 1

<table>
<thead>
<tr>
<th>Step</th>
<th>Reaction class</th>
<th>Parameter values</th>
<th>Number of reactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ligand–receptor binding</td>
<td>$k_1 = 0.003$, $k_{-1} = 0.06$</td>
<td>48</td>
</tr>
<tr>
<td>2</td>
<td>Receptor dimerization</td>
<td>$k_2 = 0.01$, $k_{-2} = 0.1$</td>
<td>600</td>
</tr>
<tr>
<td>3</td>
<td>Receptor tyrosine phosphorylation</td>
<td>$k_3 = 0.5$, $k_{-3} = 4.505$</td>
<td>$96 \times 144^b$</td>
</tr>
<tr>
<td>4</td>
<td>Receptor tyrosine dephosphorylation</td>
<td>$k_4 = 4.505$</td>
<td>$104 \times 156^b$</td>
</tr>
<tr>
<td>5</td>
<td>Binding of Grb2 to pY1068</td>
<td>$k_5 = 0.0015$, $k_{-5} = 0.05$</td>
<td>$24 \times 2380^b$</td>
</tr>
<tr>
<td>6</td>
<td>Binding of ShcP to pY1148–1173</td>
<td>$k_6 = 0.0001$, $k_{-6} = 0.0015$</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>Binding of ShcP to pY1148–1173</td>
<td>$k_{-7} = 0.003$, $k_{-17} = 0.006$</td>
<td>$16 \times 192^b$</td>
</tr>
<tr>
<td>8</td>
<td>Phosphorylation of ShcP</td>
<td>$k_8 = 3.7 \times 10^{-3}$, $k_{-8} = 0.03$</td>
<td>$8 \times 192^b$</td>
</tr>
<tr>
<td>9</td>
<td>Binding of ShcP to pY1148–1173</td>
<td>$k_{-9} = 0.00045$, $k_{-19} = 0.03$</td>
<td>$16 \times 192^b$</td>
</tr>
<tr>
<td>10</td>
<td>Cytosolic binding of Grb2 and Sos</td>
<td>$k_{10} = 0.0001$, $k_{-10} = 0.0015$</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>Binding of Sos–Grb2 to pY1148–1173</td>
<td>$k_{11} = 0.0001$, $k_{-11} = 0.0015$</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>Cytosolic binding of Grb2 and Sos</td>
<td>$k_{12} = 0.0001$, $k_{-12} = 0.0015$</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>Binding of ShcP to pY1148–1173</td>
<td>$k_{13} = 0.00005$, $k_{-13} = 0.001$</td>
<td>$16 \times 192^b$</td>
</tr>
<tr>
<td>14</td>
<td>Binding of ShcP to pY1148–1173</td>
<td>$k_{14} = 0.00005$, $k_{-14} = 0.001$</td>
<td>$16 \times 192^b$</td>
</tr>
<tr>
<td>15</td>
<td>Binding of ShcP to pY1148–1173</td>
<td>$k_{15} = 0.00005$, $k_{-15} = 0.001$</td>
<td>$16 \times 192^b$</td>
</tr>
<tr>
<td>16</td>
<td>Cytosolic ShcP dephosphorylation</td>
<td>$k_{16} = 0.005$</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>Binding of Grb2 to pY1148–1173–ShcP</td>
<td>$k_{17} = 0.005$, $k_{-17} = 0.001$</td>
<td>$16 \times 192^b$</td>
</tr>
<tr>
<td>18</td>
<td>Binding of ShcP–Grb2 to pY1148–1173–ShcP</td>
<td>$k_{18} = 0.0005$, $k_{-18} = 0.001$</td>
<td>$16 \times 192^b$</td>
</tr>
<tr>
<td>19</td>
<td>Binding of Sos to pY1148–1173–ShcP–Grb2</td>
<td>$k_{19} = 0.01$, $k_{-19} = 0.0124$</td>
<td>$16 \times 192^b$</td>
</tr>
<tr>
<td>20</td>
<td>ShcP–Grb2-Sos binding to pY1148–1173</td>
<td>$k_{20} = 0.0002$, $k_{-20} = 0.12$</td>
<td>$16 \times 192^b$</td>
</tr>
<tr>
<td>21</td>
<td>Cytosolic binding Grb2 to ShcP</td>
<td>$k_{21} = 0.003$, $k_{-21} = 0.001$</td>
<td>2</td>
</tr>
<tr>
<td>22</td>
<td>Cytosolic binding Sos to ShcP–Grb2</td>
<td>$k_{22} = 0.0001$, $k_{-22} = 0.001$</td>
<td>2</td>
</tr>
<tr>
<td>23</td>
<td>Cytosolic binding Grb2–Sos to ShcP</td>
<td>$k_{23} = 0.005$, $k_{-23} = 0.00429$</td>
<td>$16 \times 192^b$</td>
</tr>
<tr>
<td>Total</td>
<td>20 Reaction classes</td>
<td>37 Parameters</td>
<td>3749 Reactions</td>
</tr>
</tbody>
</table>

Table 2

<table>
<thead>
<tr>
<th>Step</th>
<th>Reaction class</th>
<th>Parameter values</th>
<th>Number of reactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ligand–receptor binding</td>
<td>$k_1 = 0.0003$, $k_{-1} = 0.06$</td>
<td>8</td>
</tr>
<tr>
<td>2</td>
<td>Receptor dimerization</td>
<td>$k_2 = 0.0001$, $k_{-2} = 0.001$</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>Receptor tyrosine phosphorylation</td>
<td>$k_3 = 0.5$, $k_{-3} = 4.505$</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>Receptor tyrosine dephosphorylation</td>
<td>$k_4 = 4.505$</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>Binding of PLCγ to Y992</td>
<td>$k_5 = 0.003$, $k_{-5} = 0.02$</td>
<td>12</td>
</tr>
<tr>
<td>6</td>
<td>Transphosphorylation of PLCγ</td>
<td>$k_6 = 0.5$, $k_{-6} = 0.025$</td>
<td>10</td>
</tr>
<tr>
<td>7</td>
<td>Binding of PLCγ to Y992</td>
<td>$k_7 = 0.003$, $k_{-7} = 0.3$</td>
<td>12</td>
</tr>
<tr>
<td>8</td>
<td>Dephosphorylation of PLCγ</td>
<td>$k_8 = 0.001$, $k_{-8} = 0.002$</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>Inactivation of PLCγ</td>
<td>$k_{9} = 0.01$, $k_{-9} = 0.03$</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>9 Reaction classes</td>
<td>15 Parameters</td>
<td>75 Reactions</td>
</tr>
</tbody>
</table>

The number of reactions involving receptor monomers plus the number of reactions involving receptor dimers, respectively.
naling by the immunoreceptor FcRII (Goldstein et al., 2002; Fiedler et al., 2003; Goldstein et al., 2004).

The next steps in the intracellular signaling cascade are transphosphorylation of a receptor in a dimer by the tyrosine kinase domain of the neighboring EGFR (Step 3) and dephosphorylation (Step 4), which is carried out by phosphatases that are considered implicitly, as in the model of Kholodenko et al. (1999). In the pathway-like model, all receptor tyrosines of dimeric EGFR are lumped together, and Step 3 is considered to result in a single form of dimeric EGFR, designated, RP that encompasses, for example, the distinct receptor phosphoforms illustrated in Box 2 of Fig. 1A.

In contrast, we treat the phosphorylation sites of EGFR independently, and assume that there is no interaction among proteins binding to the different sites, which allows us to capture in the network model the full stoichiometric range of EGFR complexes without introducing any new rate parameters. We consider EGFR to contain one docking site for phospholipase C \( \gamma \) (PLC\( \gamma \)), which is localized around Y992 and active when Y992 is phosphorylated, one docking site for Grb2, which is localized around Y1068 and active when Y1068 is phosphorylated, and one docking site for Shc, which is localized around Y1148 and Y1173 and active when tyrosines Y1148 and Y1173, which we lump together, are phosphorylated. These docking sites are assumed to be phosphorylated in separate reactions (via a non-processive mechanism), and association of PLC\( \gamma \), Grb2, and Shc with EGFR is assumed to depend only on the phosphorylation states of Y992, Y1068, and Y1148/73, respectively. Association of a cytosolic protein (e.g., Grb2) with EGFR is assumed not to block association of a second cytosolic protein (e.g., Shc) with the same receptor or with an adjacent receptor in the same dimer. Steps 5–8 and 25 in Fig. 1A and B are related to the dynamics of PLC\( \gamma \), which is considered in the model of Kholodenko et al. (1999) but not required for activation of Shc. Because we assume docking sites on EGFR are independent (see above), interaction of EGFR with PLC\( \gamma \) is unaffected by Grb2, Shc, or Sos. As a result, the models for activation (i.e., phosphorylation) of PLC\( \gamma \) and recruitment of Sos can be derived independently.\(^2\)

Steps 9–24 in Fig. 1A and B are related to the dynamics of Grb2, Shc, and Sos. Each of these steps is described briefly in Table 1. Translocation of Sos from the cytosol to the membrane is accomplished as follows. The cytosolic adapter proteins Grb2 and Shc are recruited to autophosphorylated receptor tyrosines. Grb2 binds at Y1068 via its Src homology 2 (SH2) domain, and Shc binds at Y1148 via is SH2 domain and Y1173 via is phosphotyrosine binding (PTB) domain (Batzer et al., 1994; Okabayashi et al., 1994). Note that, just as we lump tyrosines Y1148 and Y1173 of EGFR together, we lump the SH2 and PTB domains of Shc together, treating Shc as if it has a single EGFR binding domain. When Shc is bound to a receptor, it can be phosphorylated by EGFR (Pellicci et al., 1992). The phosphorylated form of Shc interacts with the SH2 domain of Grb2 (Lowenstein et al., 1992; Rozakis-Adcock et al., 1992; Sasaoka et al., 1994), which interacts constitutively via its two SH3 domains with Sos (Egan et al., 1993; Rozakis-Adcock et al., 1993; Li et al., 1993). The reactions that take place in the cytosol (Steps 12, 16, and 21–23 of Fig. 1B) are the same in the network and pathway-like models. The left-hand side of Box 4 in Fig. 1A illustrates reactions involving EGFR that affect Shc-dependent recruitment of Sos to the membrane, and the right-hand side of Box 4 illustrates reactions involving EGFR that affect Shc-independent recruitment of Sos to the membrane.

2.2. Automatic generation of the reaction network

One can now enumerate the 356 molecular species considered in the network model for EGFR-mediated activation of Sos. The extracellular domain of EGFR can be free or bound to EGF. The Grb2 docking site of EGFR can be free, phosphorylated, bound to Grb2, or bound to Grb2 and Sos in complex. The Shc docking site of EGFR can be free, phosphorylated, bound to Shc, bound to phosphorylated Shc, bound to Shc and Grb2 in complex, or bound to Shc, Grb2, and Sos in complex. Thus, there are \( 2 \times 4 \times 6 = 48 \) forms of monomeric EGFR, 24 forms of symmetric receptor dimers (each receptor in a dimer is

\[ \text{symmetric receptor dimers (each receptor in a dimer is bound to EGF),} \]

\[ \begin{align*}
\text{and } \left\{ \begin{array}{l}
\text{4 forms of monomeric EGFR, 4 possible forms of}
\text{symmetric receptor dimers (each receptor in a dimer is bound to EGF),}
\end{array}\right.
\end{align*} \]

\[ \text{of either Grb2 or Shc, and 24 forms of complex dimers.} \]

\[ \text{of either Grb2 or Shc, and 24 forms of complex dimers.} \]

\[ \text{of symmetric receptor dimers (each receptor in a dimer is bound to EGF),} \]

\[ \text{and } \left\{ \begin{array}{l}
\text{4 forms of monomeric EGFR, 4 possible forms of}
\text{symmetric receptor dimers (each receptor in a dimer is bound to EGF),}
\end{array}\right.
\end{align*} \]

\[ \text{of either Grb2 or Shc, and 24 forms of complex dimers.} \]
bound to EGF), and $\binom{24}{2} = 276$ forms of asymmetric receptor dimers. There are eight additional species associated with EGFR: (1) free EGF, (2) cytosolic Grb2, (3) cytosolic Sos, (4) cytosolic Grb2 in complex with Sos (Grb2–Sos), (5) cytosolic Shc, (6) phosphorylated Shc in the cytosol (ShcP), (7) cytosolic Shc in complex with Grb2 (ShcP–Grb2), and (8) cytosolic Shc in complex with Grb2 and Sos (ShcP–Grb2–Sos).

Fig. 1D illustrates 20 possible reactions a particular dimer can undergo: it can break up into two receptor-monomers; unprotected phosphotyrosines can be dephosphorylated; phosphotyrosines can bind different cytosolic protein complexes; proteins can bind to proteins already associated with a receptor; some protein complexes may dissociate from receptor or from receptor-bound proteins. This example is typical. Table 1 gives the number of reactions of each reaction type included in the model. For example, a ligand can potentially bind to any receptor-monomer without a ligand, which gives rise to 48 unidirectional reactions; any two ligand-associated receptors can aggregate into a dimer, which gives rise to 600 potential dimerization reactions, etc. In the end, we consider 3749 unidirectional reactions among the 356 species in the network model of Sos activation.

A computational model that describes interactions among these species typically includes an ordinary differential equation (ODE) for each chemical species. Each equation contains a term on the right-hand side for every reaction that influences the concentration of the species that corresponds to the ODE. To account for hundreds to thousands of species, one must abandon the approach of writing equations manually. An automatic procedure and software called BioNetGen (Blinov et al., 2004; Fae der et al., 2005a) was used to generate the chemical reaction network (and corresponding system of coupled ODEs) based on the reaction classes defined in Tables 1 and 2 and illustrated in Fig. 1. The software and models used in our studies are available at our website (http://cellsignaling.lanl.gov). Given the relatively large size of the network model compared to most models of signal-transduction systems reported in the literature, it is reasonable to be concerned about the computational costs of a rule-based approach to modeling. The computational cost of integrating a system of 356 ODEs is manageable: computations using a standard method appropriate for stiff systems take at most a few minutes on a single processor. On the other hand, for larger systems, such as some of the simple extensions of the network model we have considered, which have tens of thousands of ODEs (unpublished material), computational efficiency is likely to be a major concern, and novel methods of simulation, in which rule evaluation is embedded in the method of simulation, may be needed (Lok and Brent, 2005; Fae der et al., 2005a).

Values of parameters in the network models for Sos and PLCγ activation (i.e., rate constants for each of the reaction classes and concentrations for EGFR, EGFR, Grb2, Shc, Sos, and PLCγ) are summarized in Tables 1 and 2. These values correspond to parameter estimates of Kholodenko et al. (1999) with the following exceptions. Rate constants for reactions in which a cytosolic protein binds a docking site on EGFR have each been scaled by a factor of 0.5. This adjustment was made to account for the different stoichiometry of the two models: the pathway-like model accounts for only one docking site per receptor dimer, whereas the network model accounts for one docking site per receptor, or two docking sites per receptor dimer. Also, as a simplification, we have replaced Michaelis–Menten rate laws for phosphatase-catalyzed reactions with rate laws of the form $kC$, where $k$ is a rate constant and $C$ is the concentration of a phosphotyrosine-containing molecule. In each case, we have set the value of $k$ equal to the value of the corresponding ratio $V_M/K_M$, which can be obtained from the estimates of Kholodenko et al. (1999). Note that $V_M$ is the maximum velocity of a reaction, and $K_M$ is the Michaelis–Menten constant. This simplification, which we have made in all the models that we consider, has an insignificant effect on results (results not shown). Schoeberl et al. (2002) also made this simplification.

3. Results

3.1. Similar predictions of the pathway-like and network models

The pathway-like model was parameterized so that its predictions would match certain dynamic responses to EGF that were assayed by Kholodenko et al. (1999). In Fig. 2A–E, the experimental results of Kholodenko et al. (1999) are compared with predictions of the network model and two forms of the pathway-like model: one with PLCγ included, as in the original model, and a reduced form with PLCγ omitted. We consider these two forms of the pathway-like model to illustrate that competitive binding of PLCγ, which is omitted in the network model, has a minimal effect on activation of Sos. As can be seen, for the quantities measured by Kholodenko et al. (1999), the network and two pathway-like models make similar predictions, especially at steady-state. The different models are more-or-less equally consistent with...
the measured time courses of direct Grb2 association with EGFR (Fig. 2A), PLCγ phosphorylation (Fig. 2B), indirect Grb2 association with EGFR via Shc (Fig. 2C), Shc phosphorylation (Fig. 2D), and receptor phosphorylation (Fig. 2E). The differences at early times in Fig. 2A–D are explained by the assumption that cytosolic proteins compete for receptor binding in both forms of the pathway-like model, and the differences of Fig. 2E are explained by the assumption of the pathway-like model that all receptor phosphorytrosines are protected from phosphatase activity if a single cytosolic protein is associated with a receptor dimer. After the transient, the models agree, largely because Shc and PLCγ are found mostly in their phosphorylated cytosolic forms at steady-state (not shown), and these forms have low affinity for docking sites on EGFR (Kholodenko et al., 1999, Tables 1 and 2). Also, the dynamics of cytosolic reactions are identical in the network and pathway-like models. Because binding of cytosolic proteins to EGFR is non-competitive in the network model, the predictions of the network model shown in Fig. 2 are independent of whether PLCγ is considered or omitted.

3.2. New predictions of the network model

The network model, because it incorporates more molecular details than the pathway-like model, can predict the results of experiments that are beyond the scope of the pathway-like model (Fig. 3). Fig. 3A shows that individual tyrosines of EGFR are predicted by the network model to display distinct dynamics during a response to EGF, even though the rise and fall of total receptor phosphorylation is predicted by the network and pathway-like models alike. Comparison of Fig. 3A with Fig. 2E reveals that the pattern of total receptor phosphorylation can be attributed to phosphorylation of Y1148/73, binding of Shc at pY1148/73, where it protects pY1148/73 from phosphatase activity, phosphorylation of Shc, and return of phosphorylated Shc to the cytosol, where it concentrates in its phosphorylated form.
The remaining panels of Fig. 3 illustrate that the network model predicts greater molecular diversity than the pathway-like model. Fig. 3B shows that a substantial fraction of receptor dimers are predicted by the network model to associate transiently with two molecules of phosphorylated Shc during the response to EGFR. At the peak of the time course of Fig. 3B, 26% of total phosphorylated Shc is found in complexes containing two molecules of EGFR and two molecules of Shc. Fig. 3C shows that a substantial fraction of monomeric receptors are also predicted by the network model to be found in association with Sos. The slightly delayed association of Sos with monomeric receptors compared to association of Sos with dimeric receptors reflects the time required for receptors in dimers to become phosphorylated and dissociate from each other. As can be seen, the fraction of membrane-associated Sos in complexes with monomeric receptors is small at short times, but becomes more significant as the steady-state is approached. At steady-state, 36% of total Sos is associated with monomeric receptors. The assumptions of the pathway-like model preclude the formation of these complexes and many others, as indicated in Fig. 3D. The plots of Fig. 3D represent the numbers of molecular species containing 98% of receptors as a function of time after EGF stimulation according to the network and pathway-like models. As can be seen, at the peak of the time course predicted by the network model, 98% of receptors are distributed among more than 100 molecular species. The diversity of molecular species containing 98% of receptors falls as the steady-state is approached to about one third of its peak transient value. However, at all times, the distribution of receptors among molecular species predicted by the network model is greater than that predicted by the pathway-like model. The molecular species (and reactions) that are prevalent at any given time are determined by the dynamics of the signaling cascade and the connectivity of the reaction network (Faeder et al., 2005b).

3.3. Distinguishing predictions of the two models

The two models make distinct predictions, which could help us determine if assumptions restricting the range of species are valid (Fig. 4). Fig. 4A illustrates the predicted effects of two mutations: (1) knocking out the EGFR tyrosines that bind Shc, and (2) knocking out Shc tyrosines that bind Grb2. The first mutation blocks Shc recruitment to receptor. The second mutation blocks Grb2 binding to Shc, but allows Shc to bind EGFR. Both
mutations eliminate Shc-mediated recruitment of Sos and thus might be expected to downregulate Sos activation. Fig. 4A demonstrates that both mutations do indeed lower Sos recruitment compared to the wild type (WT) during transient activation, but, counter to intuition, produce a much higher level of steady-state Sos activation.

For the first mutation, both the network and pathway-like models predict the same steady-state levels because in the absence of Shc binding, Grb2 recruits Sos at the same rate in both models via direct binding of Grb2 to EGF.

For the second mutation, the predictions of the two models differ because unphosphorylated Shc can bind EGF, and this binding competes with Grb2-EGFR binding in the pathway-like model but not in the network model. This competition results in partial inhibition of Sos activation in the pathway-like model, although the level of steady-state activation is still higher in the mutant than in WT. In contrast, the level of steady-state activation predicted by the network model is the same for both mutations. If the levels of steady-state Sos activation for the two mutants could be measured experimentally, the results presented above suggest that independence of the Grb2 and Shc binding sites in EGFR could be established or refuted.

The two models make different predictions about EGFR dimerization during signaling, as shown in Fig. 4B. The pathway-like model predicts a slower approach to steady-state and a larger number of receptors in dimers at steady-state than the network model. The pathway-like model also predicts a transient overshoot, whereas the network model does not. These predictions of the pathway-like model arise, because in this model signaling affects receptor dimerization. Recall that phosphorylated receptors in a dimer are not allowed to dissociate. In contrast, in the network model, receptor dimerization is independent of the signaling events under consideration. Thus, monitoring receptor dimerization as parameters of signaling are varied (e.g., protein expression levels of Grb2 or Shc) could help determine if signaling affects dimerization. The predicted effect of varying Grb2 concentration is shown in Fig. 4B. The number of receptor dimers depends on the total cytosolic Grb2 concentration in the pathway-like model but not in the network model.

The two models also make different predictions about the outcome of the following hypothetical experiment, which depends on whether or not receptor monomers are involved in signaling. Let us consider a panel of bivalent antibodies that crosslink EGFR with different kinetics. Fig. 4C illustrates the predicted effects at steady-state of crosslinking receptors with an antibody as a function of dimer dissociation rate. The network model predicts that Shc recruitment is maximal at an optimal dimer lifetime, whereas the pathway-like model predicts a low level of Shc recruitment that is relatively insensitive to the lifetime of dimers. In the pathway-like model only dimers...
can recruit Shc, whereas in the network model phosphorylated receptor-monomers can be associated with Shc. Model parameters are such that unphosphorylated Shc has high affinity for EGFR and remains associated with receptor for a long time, which explains the build-up of the receptor-associated pool of Shc proteins as dimer lifetime decreases. A shorter lifetime allows more receptors to traffic through the dimeric state become phosphorylated and recruit Shc. This effect is one of serial engagement (Goldstein et al., 2004). As the lifetime becomes shorter still, Shc recruitment falls, because dimers break apart before receptors can be phosphorylated, an effect known as kinetic proofreading (Hlavacek et al., 2002; Goldstein et al., 2004).

3.4. Molecular diversity during signaling

In Fig. 3D, we pointed out that only a relatively small fraction of the possible molecular species is populated at steady-state, but many more species are populated during the transient peak. This behavior is described in greater detail in Fig. 5, which helps to clarify the differences and similarities between the network and pathway-like models. The distribution of Sos in different species is shown at short times in panel A and at long times in panel B. Fig. 5B demonstrates that at steady-state only two chemical species contain 72% of all Sos recruited through ShcP. These species are a dimer and a monomer that each contain one phosphorylated tyrosine. This kind of species is accounted for in the model of Kholodenko et al. (1999). In contrast with the narrow distribution at steady-state, the transient distribution is much broader. The two most prevalent steady-state species contain only 47% of Sos recruited through ShcP (Fig. 5A), with Sos distributed among a variety of other receptor species: 26 species account for 95% of Sos recruited through ShcP. The next two most prevalent receptor species during the transient contain 18% of Sos at 10 s. These species contain receptor dimers with each receptor bound to an adapter; this kind of species is unaccounted for in the pathway-like model. Similar distributions are observed for Sos bound to receptor via Grb2 recruited at Y1068 (Fig. 5C and D).

These observations provide an explanation for the match in quantitative predictions between the two models at steady-state. According to the network model, under steady-state conditions, only one phosphorylated tyrosine in a dimer is engaged in protein-binding activity, as assumed in the pathway-like model of Kholodenko et al. (1999). Complexes in which more than one protein is bound to a receptor at the same time are rare at steady-state, but common during the transient, which explains the differences between the two models at short times.

3.5. Distinct reaction sequences that lead to Sos recruitment

Another important question is what reaction paths are involved in recruitment of Sos to the membrane. To
answer this question we use the method of path analysis, described by Faeder et al. (2005b). This method determines the prevalence of sequences of reactions by which a molecular component is transformed from one state into another one. Here, we focus on pathways that transform an unmodified receptor monomer into a receptor associated with Sos under steady-state conditions.

There are nine distinct reaction paths for Sos recruitment in the model of Kholodenko et al. (1999). Fig. 6A illustrates the top three, which account for 70% of Sos recruited at steady-state. In contrast, there are a variety of paths in the network model, with the 50 most prevalent paths accounting only for 45% of Sos recruitment. All of the paths in the network model can be grouped into nine classes of reaction paths, each of which corresponds to an individual path in the model of Kholodenko et al. (1999). Fig. 6B illustrates the most prevalent paths within one such class, in which recruitment of Grb2 and Sos is sequential.

According to both models, at steady-state, about a third of Sos is activated via sequential recruitment of Grb2 and Sos. Parameters of the models are such that the amount of free Grb2–Sos complex in the cytosol is less than the amount of free cytosolic Grb2, which explains why simultaneous recruitment of Grb2 and Sos in complex, the second most prevalent route to Sos activation,
is less common than sequential recruitment of Grb2 and Sos to Y1068. These two classes of paths are responsible for about 50% of the Sos recruited at steady-state. The third most prevalent class of paths is binding of the complex Shc-P-Grb2-Sos to Y1148/73. Although the amount of Shc-P-Grb2-Sos complexes is high at steady-state, the prevalence of this pathway is suppressed by the low affinity of the complex for EGFR.

As illustrated in Fig. 6B, within each class of Sos activation pathways there is molecular diversity. The network model allows one to identify the individual chemical species involved in these reaction paths. As can be seen, sequences that recruit Sos may involve receptor monomers.

Gong and Zhao (2003) have also analyzed the relative contributions of the Shc-dependent and Shc-independent pathways for Sos recruitment and subsequent Ras activation using the model of Schoerber et al. (2002), and they find that over the full time course of receptor stimulation there is a significant preference for the Shc-dependent pathway. We find for both the pathway-like and the network models that the two pathways make nearly equal contributions to Sos activation at steady-state. The differences between our results and those of Gong and Zhao (2003) might arise from the differences between the model of Khodolenko et al. (1999) and that of Schoerber et al. (2002), such as the assumption that GAP stabilizes the formation of dimers in the latter model. In any event, the results of Section 3.3 demonstrate that pathway analysis is not always a reliable indicator of the effect of blocking the activity of a signaling protein, which in the case of Shc results in a large increase in the level of steady-state Sos activation. Thus, when Shc is present it may carry a major portion of the Sos activation flux, but its removal, at least in the case of Shc results in a large increase in the level of steady-state Sos activation. Thus, when Shc is present it may carry a major portion of the Sos activation flux, but its removal, at least in the case of Shc results in a large increase in the level of steady-state Sos activation. Thus, when Shc is present it may carry a major portion of the Sos activation flux, but its removal, at least in the case of Shc results in a large increase in the level of steady-state Sos activation. Thus, when Shc is present it may carry a major portion of the Sos activation flux, but its removal, at least in the case of Shc results in a large increase in the level of steady-state Sos activation. Thus, when Shc is present it may carry a major portion of the Sos activation flux, but its removal, at least in the case of Shc results in a large increase in the level of steady-state Sos activation. Thus, when Shc is present it may carry a major portion of the Sos activation flux, but its removal, at least in the case of Shc results in a large increase in the level of steady-state Sos activation. Thus, when Shc is present it may carry a major portion of the Sos activation flux, but its removal, at least in the case of Shc results in a large increase in the level of steady-state Sos activation. Thus, when Shc is present it may carry a major portion of the Sos activation flux.
receptor dimers, and competitive binding of adapter proteins to a receptor dimer. These assumptions may or may not be appropriate under all experimental conditions. We would like to reduce such limiting assumptions to a minimum even though we recognize that not all possible species are important under all conditions. The reason is that we cannot identify the critical species through intuition, as the importance of species depends on network dynamics and connectivity. Input to a model in the form of known molecular interactions and activities seems preferable to input in the form of an intuitively selected set of chemical species and reactions.

The network model, without adjustment of parameter values, reproduces the predictions of the pathway-like model that were compared with experimental data by Khodolenko et al. (1999) (Fig. 2). However, other predictions of the two models differ, and the network model yields new predictions that are unavailable from the pathway-like model. For example, in the network model, individual tyrosines of EGFR are considered. Thus, this model can be used to predict the dynamics of phosphorylation for each tyrosine, and we find that different tyrosines have different temporal patterns of phosphorylation (Fig. 3A). The results of Fig. 4 suggest three types of experiments that could be performed to determine which model better represents early events in EGFR signaling and obtain mechanistic insights. For example, the pathway-like model predicts that mutations of EGFR and Shc tyrosines should have different effects, whereas the network model predicts that these mutations should have identical effects (Fig. 4A). Interestingly, both models make the surprising prediction that elimination of Shc-mediated recruitment of Sos will have a positive effect on overall Sos recruitment at steady-state. This result is obtained because phosphorylated Shc in the models, consistent with molecular dynamics simulations (Suemaga et al., 2004), binds EGFR with low affinity and sequesters Sos in the cytosol, inhibiting Sos activation. This finding appears to contradict earlier experimental results (Sasaoka et al., 1994) and probably merits further experimental investigation.

Kinetic rate constants in the network model are the same as those in the pathway-like model with the exception of a scale factor, which was introduced for stoichiometric reasons (cf. Tables 1 and 2 and Khodolenko et al., 1999). For example, the rate constant for Grb2 binding to phosphorylated receptor dimers in the pathway-like model is twice the value of the rate constant for binding of the SH2 domain of Grb2 to receptor phosphorytrosine Y1068 in the network model. To avoid introducing additional rate parameters when constructing the network model, we made the assumption that proteins bind to the phosphorylation sites of EGFR independently, which creates large classes of reactions that can be characterized by the same rate constant. This lumping of rate constants in the network model is motivated by the modularity of protein domains (see Hlavacek et al., 2003), and it can be tested and refined as more experimental data becomes available about the effects of protein-protein interactions. There is, however, already some evidence that a large number of proteins can bind independently to the relatively small cytoplasmic portion of a membrane-bound receptor, such as EGFR (Jiang and Sorkin, 2002) or CD19 (Brooks et al., 2004).

The simple structure of the pathway-like model, on the other hand, comes at the cost of introducing complex and often hidden correlations among model parameters and variables. For example, the assumption that only unmodified dimers of EGFR can dissociate leads to a correlation between adapter protein expression levels and ligand-receptor binding (Fig. 4B), and masks the effects of serial engagement and kinetic proofreading on EGFR-Shc association predicted by the network model (Fig. 4C).

The pathway-like and network models provide different microscopic pictures of EGFR signaling. As might be expected, the network model predicts greater molecular diversity (Figs. 3B–D, 5, and 6). For example, multiple adapter proteins are predicted to associate with the same receptor dimer, which is not allowed in the pathway-like model but consistent with the results of Jiang and Sorkin (2002). Also, receptor monomers are predicted to contribute to Sos recruitment, which again is not allowed in the pathway-like model. Spatial spread of receptor phosphorylation in response to localized EGF stimulation (Verveer et al., 2000; Reynolds et al., 2003) might be explained by the involvement of receptor monomers in signaling. Also consistent with predictions of the network model are proteomic assays indicating that EGFR is a member of a large number of heterogeneous protein complexes that arise during signaling (Blagoev et al., 2003, 2004). At this time, predictions about molecular diversity are difficult to test, but such predictions will become more important with advances in methods for monitoring protein modifications and interactions ( Aebersold and Mann, 2003; Mann and Jensen, 2003; Meyer and Tersel, 2003).

Acknowledgements

This work was supported by DOE through contract W-7405-ENG-36 and by NIH through grants GM35556, AI28433 and RR18754. We thank Ed Stites for stimulating discussions.
References


