Article

Use of Mechanistic Models to Integrate and Analyze Multiple Proteomic Datasets

Edward C. Stites,[1,2,*] Meraj Aziz,[1] Matthew S. Creamer,[1,3] Daniel D. Von Hoff,[1] Richard G. Posner,[1,4,*] and William S. Hlavacek[1,5,*]

[1]Clinical Translational Research Division, Translational Genomics Research Institute, Phoenix, Arizona; [2]Department of Pathology & Immunology, Washington University School of Medicine, St. Louis, Missouri; [3]Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, Connecticut; [4]Department of Biological Sciences, Northern Arizona University, Flagstaff, Arizona; and [5]Theoretical Biology and Biophysics Group, Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico

ABSTRACT Proteins in cell signaling networks tend to interact promiscuously through low-affinity interactions. Consequently, evaluating the physiological importance of mapped interactions can be difficult. Attempts to do so have tended to focus on single, measurable physicochemical factors, such as affinity or abundance. For example, interaction importance has been assessed on the basis of the relative affinities of binding partners for a protein of interest, such as a receptor. However, multiple factors can be expected to simultaneously influence the recruitment of proteins to a receptor (and the potential of these proteins to contribute to receptor signaling), including affinity, abundance, and competition, which is a network property. Here, we demonstrate that measurements of protein copy numbers and binding affinities can be integrated within the framework of a mechanistic, computational model that accounts for mass action and competition. We use cell line-specific models to rank the relative importance of protein-protein interactions in the epidermal growth factor receptor (EGFR) signaling network for 11 different cell lines. Each model accounts for experimentally characterized interactions of six autophosphorylation sites in EGFR with proteins containing a Src homology 2 and/or phosphotyrosine-binding domain. We measure importance as the predicted maximal extent of recruitment of a protein to EGFR following ligand-stimulated activation of EGFR signaling. We find that interactions ranked highly by this metric include experimentally detected interactions. Proteins with high importance rank in multiple cell lines include proteins with recognized, well-characterized roles in EGFR signaling, such as GRB2 and SHC1, as well as a protein with a less well-defined role, YES1. Our results reveal potential cell line-specific differences in recruitment.

INTRODUCTION

Signaling proteins operate within large, complex networks (1,2). One aspect of network complexity is that signaling proteins tend to participate in multiple interactions, with overlapping specificities (3,4). (This molecular promiscuity and polyspecificity is a general feature of proteins.) For example, in a recent study, Jones and co-workers (5) characterized a total of 1405 specific interactions between phosphotyrosine-containing peptides of the four members of the ErbB family of receptor tyrosine kinases (RTKs) and a collection of human Src homology 2 (SH2) and phosphotyrosine-binding (PTB) domains. As is typical of interactions detected and characterized in high-throughput experiments, many of these interactions have affinities that can be considered low, and most do not have known or clear physiological relevance.

In phosphotyrosine signaling (6), although different SH2/PTB domains do recognize a given phosphotyrosine-containing peptide with different affinities (7), affinity differences alone are insufficient to confer a high degree of specificity on phosphotyrosine-SH2/PTB domain interactions (8). It has been speculated that factors beyond affinity, such as compartmentalization, are responsible for conferring specificity (9). However, these possible determinants of specificity are poorly characterized. Moreover, the specificity requirements of cellular information processing are somewhat unclear. It may be that numerous, low-affinity interactions are not only tolerated but are essential for certain information processing functions. Jones (10) has argued that low-affinity interactions are likely to have physiological significance, which has been overlooked mainly because of the difficulty of detecting and studying such interactions.

In any case, for proteins with well-characterized roles in a cell signaling network, it is rare for all of their known interactions to receive attention in an experimental or modeling study. Indeed, there is a stark contrast between the interaction networks considered in typical pathway diagrams (for an example of a large diagram, see (11)) or in typical mathematical/computational models (12) and the much larger interaction networks defined by databases that collect information reported in the primary literature, such as HPRD (13); databases that collect predicted interactions, such as PrePPI (14,15); or databases that collect interactions detected in high-throughput screens, such as BioGRID (16).

Submitted July 10, 2014, and accepted for publication February 24, 2015.

*Correspondence: estites@path.wustl.edu or rposner@tgen.org or wish@lanl.gov

Editor: H. Wiley.

0006-3495/15/04/1819/11 $2.00

http://dx.doi.org/10.1016/j.bpj.2015.02.030

Thus, most interactions are somewhat enigmatic and we lack a sound understanding of the functional consequences of molecular promiscuity and polyspecificity, especially at the systems level.

With the emergence of reliable methods for elucidating key structural details of protein-protein interactions (e.g., the interfaces or the parts/sites of proteins responsible for interactions) (17,18) and techniques that provide comprehensive and high-precision measurements of affinities for site-specific interactions (5,19,20), as well as absolute protein copy numbers (21,22), we now have an opportunity to use this information to evaluate interaction networks in toto and begin to quantitatively assess the relative importance of detected protein-protein interactions. For example, through application of the law of mass action, we can consider how affinity and the abundances of interaction partners combine to influence extent of interaction. We can also consider how extent of interaction is affected by competition, which is a network property that can vary with cell type or changes of the (intra)cellular milieu. However, attempts to interpret interactomic/proteomic data have been limited (23). For example, Jones et al. (24) used affinities alone, together with an arbitrary affinity threshold, to estimate the relative promiscuities of ErbB family RTKs, even though a small K_D for an interaction is meaningless if a protein involved in the interaction is not expressed. Mechanistic modeling, which has been identified as having an important role to play in analyzing cell signaling networks (25), may be able to aid in integrating and interpreting interactomic/proteomic data sets.

Here, building on past modeling efforts and our current quantitative understanding of epidermal growth factor (EGF) receptor (EGFR) signaling (26–32) and using fairly new enabling computational approaches (33–35), we leverage multiple high-throughput proteomic data sets to construct and analyze cell line-specific computational models for early EGFR signaling events. These models break new ground in that each accounts for cell line-specific protein expression and the mapped, direct interactions between six phosphotyrosines in EGFR, which are generated through autophosphorylation, and numerous SH2/PTB domain-containing signaling proteins. In developing these models, we used measurements of affinities made for site-specific interactions (5) to constrain kinetic binding parameters, and we used measurements of protein abundances for 11 different cell lines (22,36) to set the cytoplasmic concentrations of proteins. Thus, our models account for cell line-specific differences in protein copy numbers. (The rate constants for phosphotyrosine-SH2/PTB domain interactions are the same for all models.)

Measurements of binding affinities obtained via a high-throughput technique and relative protein expression levels in multiple cell lines have been used in earlier modeling work focused on EGFR signaling (37), but in a more limited way. For example, in this earlier work, only a few selected direct binding partners of EGFR (GRB2, SHC1, IRS1, and an unspecified protein tyrosine phosphatase) were considered.

We use our cell line-specific models to predict recruitment of signaling proteins to EGFR. We give special attention to HeLa cells, for which copy number measurements are available from two different studies (22,36), and HEK 293 cells. These two cell lines have been used in experimental studies aimed at elucidation of EGFR binding partners (38,39). Our simulation results indicate cell line-specific differences in recruitment of signaling proteins to EGFR, although there are many commonalities across cell lines. Using our models, we measure the importance of an interaction, or set of interactions, by the maximal predicted extent of recruitment facilitated by the interaction(s) of interest. By this metric, many of the well-characterized binding/signaling partners of EGFR, such as SHC1 and GRB2, are identified as important in multiple cell lines. Notably, in addition, we identify the Src-family protein tyrosine kinase (SFK) YES1, which is not widely recognized as playing an important role in EGFR signaling, as one of the more important binding/signaling partners of EGFR. To evaluate these results, we compare importance rankings derived from the predictions of our model, which concern relative levels of recruitment of signaling proteins to activated EGFR, against the rankings obtained from simpler ad hoc approaches, such as affinity-based rankings. We compare the rankings obtained from each metric against experimental measurements of association (38,39).

MATERIALS AND METHODS

Model development

As described in Text S1 in the Supporting Material, we built a generic (cell type/cell line independent) model for early EGFR signaling events that encompasses the following processes: ligand (EGF) binding to receptor (EGFR); constitutive and ligand-induced receptor dimerization via receptor-receptor interaction; ligand- and dimer-dependent autophosphorylation of tyrosines in EGFR; dephosphorylation of phosphorylated tyrosine residues (when not shielded by a binding partner) through high, constitutive phosphatase activity; endocytosis and recycling and lysosomal degradation of internalized EGF and EGFR; and phosphotyrosine-SH2/PTB domain interactions.

The goal of model development was to build a model capable of 1) incorporating cell line-specific measurements of protein abundances and measurements of binding affinities for phosphotyrosine-SH2/PTB domain interactions and 2) predicting the effects of mass action and competition on recruitment of SH2/PTB domain-containing signaling proteins to ligand-activated EGFR. We focused on using protein abundance measurements of Mann and co-workers (22,36), which characterize 11 different cell lines (including two separate characterizations of protein expression in HeLa cells), and the K_D measurements of Hause et al. (5). Thus, our model omits interactions not characterized in the latter study, such as the interaction of CBL with EGFR and the interactions of EGFR with other ErbB receptor family members. (These interactions are not mediated by an SH2 or PTB domain.) Moreover, our model omits interactions among the direct binding partners of EGFR that are included in the model. Although some of these interactions have been quantitatively characterized

(e.g., the interaction of GRB2 with phosphorylated SHC1), we are not aware of data that could be used to guide the estimation of binding parameters for all mapped interactions. An attempt to include these interactions would entail significant complications of the model and produce unevenness in scope and support (i.e., idiosyncratic inclusion and omission of mechanistic details and varying levels of evidence for different parameters and aspects of model structure). Although the model includes endocytosis, because this process contributes to removal of phosphotyrosines on a fairly fast timescale (40), we do not attempt to account for differences in signaling from different subcellular locations (41).

The generic model is made specific for a particular cell line by parameterizing it using measured protein copy numbers for that cell line. The novelty of the model concerns the treatment of mapped phosphotyrosine-SH2/PTB domain interactions, which is more comprehensive than that of earlier models. Other parts of the model are considered because these parts are necessary to translate a ligand stimulus into phosphorylation of receptor tyrosines and to account for processes that reverse and eliminate receptor phosphorylation. These parts of the model and their associated parameter estimates were taken from the literature. Parameter settings are summarized in Table S1 for the HeLa cell-specific model that was parameterized using the absolute protein copy number measurements of Kulak et al. (22)

Performing simulations

Software S1 consists of a collection of plain-text files that can be processed by BioNetGen (33,42) and NFsim (35). The files with .bngl filename extensions provide executable model specifications, which include parameter settings, and the files with .rnf filename extensions define simulation protocols. Simulations of EGFR signaling in HeLa cells were driven by the file EGFR_K_HeLa.bngl, which defines a HeLa cell-specific version of our generic model for EGFR signaling that is consistent with the copy number measurements of Kulak et al. (22), and the file EGFR_K_HeLa.rnf, which defines a simulation protocol. Similarly, simulations of EGFR signaling in HeLa and HEK 293 cells based on the copy number measurements of Geiger et al. (36) were driven by the files EGFR_G_HeLa.bngl, EGFR_G_HeLa.rnf, EGFR_G_HEK293.bngl, and EGFR_G_HEK293.rnf. Simulations were performed by using BioNetGen (33,42) to process a .bngl file and thereby produce an XML file readable by NFsim (35). This XML file and the appropriate corresponding .rnf file were then provided as inputs to NFsim, which implements a particle-based kinetic Monte Carlo method tailored for stochastic simulation of biomolecular interactions defined in terms of rules (34). Before simulating the response to EGF addition, equilibration simulations, which began at a convenient initial condition, were performed to establish a basal steady state. In other words, we simulated the interactions considered in a cell line-specific model in the absence of EGF until concentrations reached their basal steady-state values. The results of a simulation run are recorded by NFsim in a plain-text file with a .gdat filename extension. Simulation outputs produced by NFsim must be defined; outputs are defined in the observables block of each .bngl file provided in Software S1 using the conventions of BNGL (33).

As discussed in a recent review (43), there are two broad categories of methods available for simulating rule-based models: indirect methods and direct methods. With an indirect method, given a rule-based model, one performs a simulation via a procedure involving the following steps: 1) network generation, in which the rules of a model are translated into a reaction network; 2) specification of a corresponding model in a traditional form, such as that of a coupled system of ordinary differential equations (ODEs); and 3) simulation using a standard method available for the traditional model form found, such as a method for numerical integration of ODEs. Indirect methods break down when the rules of a model imply a reaction network that is very large, which is the case here. In a direct method, the rules of a model are used as event generators in a discrete-event stochastic simulation algorithm, such as that implemented in NFsim (34,35,44). Thus, with use of a direct method, it is not necessary to find the reaction network implied by the rules of a model.

Results from 10 independent (stochastic) simulation runs were averaged to produce smoothed time courses for calculating peak recruitment levels. In these simulations, for reasons of efficiency, we sometimes considered only a fraction of a whole cell, except when considering HeLa cells and the copy number measurements of Kulak et al. (22). In this case, we always considered a whole cell. For HEK 293 cells, we considered 1% of a cell. For other cases, we considered 0.1% of a cell. The cost of stochastic simulation increases with the number of molecules considered; considering a reaction subvolume, thus, decreases computational cost (at the expense of increased fluctuations). A full cell was considered when evaluating stochasticity of recruitment.

The scope of the generic model and each cell line-specific model is limited to six sites of EGFR autophosphorylation, which are the EGFR tyrosines annotated as substrates of the EGFR kinase in the UniProt entry for EGFR (www.uniprot.org/uniprot/P00533).

Ranking interactions and evaluating rankings

We investigated the value of knowing 1) measured protein copy numbers only, 2) measured equilibrium dissociation constants only, 3) the ratios of copy numbers and dissociation constants, and 4) simulated peak association intensities, which are influenced by measured protein copy numbers, measured equilibrium dissociation constants, and competition. We defined importance metrics based on these quantities and compared rankings of phosphotyrosine-SH2/PTB domain interactions based on each metric against available measurements of association. We considered the association data of Schulze et al. (38), who used phosphotyrosine-containing peptides from EGFR in pulldown assays with HeLa cell lysates and a mass spectrometry-based method for binding partner identification. We also considered the association data of Tong et al. (39), who used antibodies against FLAG-tagged EGFR to pull down EGFR and its binding partners in HEK 293 cells and a mass spectrometry-based method for binding partner identification. We limited our analysis to SH2/PTB domain-containing proteins, and further limited ourselves to the list of proteins studied by Hause et al. (5). For a protein that binds EGFR at multiple phosphotyrosines and/or with multiple SH2/PTB domains, we considered only the minimum equilibrium dissociation constant (K_D) among the relevant phosphotyrosines and SH2/PTB domains. For HeLa cells, the copy number of a potential EGFR binding partner was taken to be that reported by Kulak et al. (22). For HEK 293 cells, we used averages of the iBAQ values reported by Geiger et al. (36) to set copy numbers. For simulations, we considered peak association intensity of an EGFR binding partner to be the maximum average number of this protein predicted to be bound to EGFR molecules over repeated simulated time courses, each observed at 1 min intervals, starting with initiation of EGFR signaling by EGF addition at time $t = 0$. We selected an EGF dose comparable to EGFR abundance. We considered the impact of EGF dose on recruitment by considering a significantly lower dose and a significantly higher dose.

We caution that our simulation-based rankings of interactions are based on the specific, technical definition of importance given previously, which measures recruitment. With this approach, the recruitment of any one signaling protein to EGFR is equally important regardless of the identity of the protein. We do not attempt to account for the signaling activity of a recruited protein, nor do we attempt to account for any dependence of signaling activity on subcellular location, such as plasma membrane-associated versus endosomal.

RESULTS AND DISCUSSION

Development of cell line-specific models for EGFR signaling based on K_D and protein copy number measurements

To demonstrate how proteomic data sets can be integrated and leveraged within the formal framework of a model to rank the

importance of protein-protein interactions, we formulated a physicochemical, mechanistic model for early events in EGFR signaling (see Materials and Methods section, and Text S1, Table S1, and Software S1 in the Supporting Material). The model was formulated for a generic/hypothetical cell line taken to express all of the SH2/PTB domain-containing EGFR binding partners characterized in the study of Hause et al. (5). A total of 12 cell line-specific versions of the generic model were derived by setting protein copy numbers on the basis of protein abundance measurements made by Mann and co-workers for 11 mammalian cell lines (22,36). There are two models for HeLa cells, one parameterized on the basis of data from the study of Kulak et al. (22) and the other parameterized on the basis of data from the study of Geiger et al. (36). The other 10 models, which are parameterized on the basis of data from the study of Geiger et al. (36), represent EGFR signaling in HEK 293, A549, GAMG, HepG2, Jurkat, K562, LnCap, MCF7, RKO, and U2OS cells (36). Geiger et al. (36) did not report an EGFR copy number for Jurkat, K562, or RKO cells, presumably because these cells do not express EGFR. For these cells, we considered the case of exogenous expression of EGFR at a level of 50,000 copies per cell (i.e., we set the EGFR copy number at 50,000 per cell to mimic stable transfection of EGFR). Interactions considered in the models were formalized in terms of rules (43). Rules for interactions involving EGFR phosphotyrosines and SH2/PTB domain-containing binding partners of EGFR were assigned rate constants consistent with the K_D measurements of Hause et al. (5). Use of rules allowed us to concisely capture in our models site-specific details, such as the phosphorylation and occupancy levels of individual SH2/PTB domain docking sites in EGFR (Fig. 1). Accounting for site-specific details is difficult or impossible when attempting to specify a model using a traditional modeling approach (43,45,46), such as writing ODEs.

The structure of the generic model is based on mechanisms considered in earlier modeling studies but is more comprehensive in its consideration of EGFR binding partners. The cell line-specific versions of the model account for protein expression levels in various commonly used mammalian cell lines. The parameter values of a HeLa cell-specific version of the generic model are summarized in Table S1; except for protein copy numbers, the other cell line-specific models share the same parameter values. Parameter values were set in accordance with earlier estimates reported in the literature, except for protein copy numbers and binding parameters for interactions between phosphotyrosines in EGFR and SH2/PTB domain-containing proteins, which were set in accordance with high-throughput proteomic data. For example, for the HeLa cell-specific model of Table S1, copy numbers were assigned based on absolute protein abundance measurements made for HeLa cells by Kulak et al. (22). Similarly, for the HEK 293 cell-specific model, copy numbers were assigned based on relative protein abundance measurements

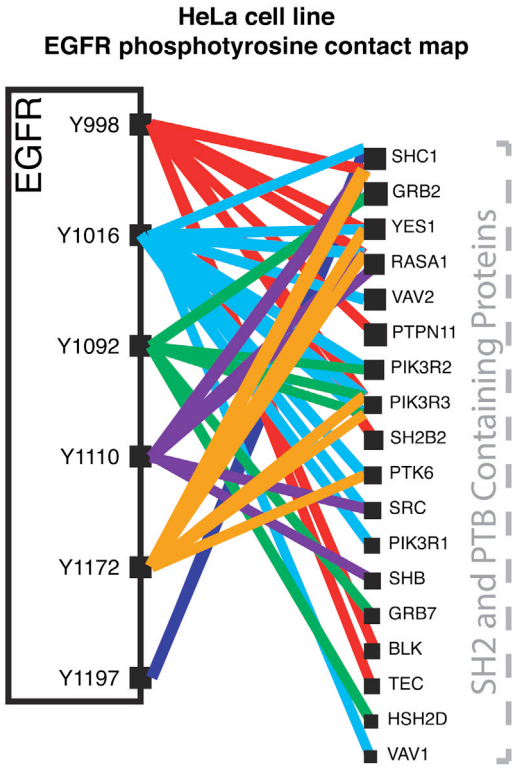

FIGURE 1 Interactions considered in a HeLa cell-specific version of a model for early events in EGFR signaling. Squares at left represent sites of EGFR autophosphorylation, squares at right represent SH2/PTB domain-containing proteins expressed in HeLa cells, and lines connecting squares represent interactions mapped and characterized experimentally. The squares at right are drawn such that area is proportional to the logarithm of measured protein copy number, and lines are drawn such that thickness is proportional to the logarithm of $1/K_D$. The squares at right are ordered from top to bottom according to predicted peak level of association with EGFR after addition of EGF. The most highly recruited protein, SHC1, is represented at the top. To see this figure in color, go online.

made for HEK 293 cells by Geiger et al. (36). As noted earlier, except for differences in protein copy numbers, the cell line-specific versions of the generic model are otherwise identical. Forward rate constants for EGFR phosphotyrosine interactions with SH2/PTB domain-containing proteins were all assigned a typical value for a protein-protein interaction. Reverse rate constants were assigned values derived from the forward rate constant and the affinity measurements of Hause et al. (5), which characterize interactions between phosphopeptides from EGFR and a set of human SH2 and PTB domains. Fig. 1 illustrates the phosphotyrosine-SH2/PTB domain interactions considered in the HeLa cell version of the generic model that was parameterized using data from the study of Kulak et al. (22).

Simulation of single-cell behavior

The (generic) model was formulated using BNGL, a rule-based model-specification language (33), for the scenario

of a single cell, and simulations were performed using the BNGL-compatible software tools BioNetGen (33,42) and NFsim (35), together with plain-text files that define cell-specific versions of the generic model and corresponding simulation protocols (Software S1). Simulations yielded site-specific time courses of phosphorylation and recruitment of signaling proteins to EGFR via their SH2/PTB domain-mediated interactions with EGFR phosphotyrosines. Representative simulation results, obtained from the HeLa cell-specific version of the model that was parameterized using data from the study of Kulak et al. (22), are shown in Fig. 2. Simulation data for all 12 cell line-specific models considered are given in Table S2. Simulations were performed using a stochastic method, which generates single-cell trajectories and accounts for the stochastic nature of chemical reaction kinetics. (We obtained population-averaged quantities by averaging the results from multiple simulation runs.)

Noisy recruitment of signaling proteins to EGFR

Several of the time courses shown in Fig. 2 exhibit fluctuations, such as those for BLK, TEC, and SHB, which have reported abundances of 242, 424, and 859 copies per cell (22), respectively. Noisy recruitment is even discernible for SH2B2, which has a reported abundance of 7693 copies per cell (22). The fluctuations reflect noisy, stochastic recruitment of signaling proteins to EGFR because of a combination of low copy number, low affinity, and high competition. It is not clear whether intermittent recruitment of signaling proteins to an activated receptor has any physiological implications, but it is interesting to note that the molecular circuitry of RAS-mediated control of ERK activation, which is downstream of EGFR, has been shown to filter transient input signals (47). This type of filtering may be important for suppressing noise arising from promiscuous/polyspecific interactions of signaling proteins.

Predicting recruitment of signaling proteins to EGFR on the basis of affinity, abundance, and competition

Based on physicochemical principles, we can expect that recruitment of a signaling protein to an activated receptor will be influenced by multiple factors. A low abundance protein may be strongly recruited if the protein's affinity for the receptor is strong. Similarly, a low affinity protein may be strongly recruited if the protein's abundance is high. A low affinity, low abundance protein may be strongly recruited if it has no competition for binding to the receptor. Thus, assessment of recruitment, which is likely a key aspect of interaction importance, can benefit from an integration of different types of data in a model. Simulations indicate that recruitment of signaling proteins to EGFR indeed depends on multiple factors. As can be seen in Fig. 3, there is a correlation between predicted recruitment and single factors but the correlation is imperfect.

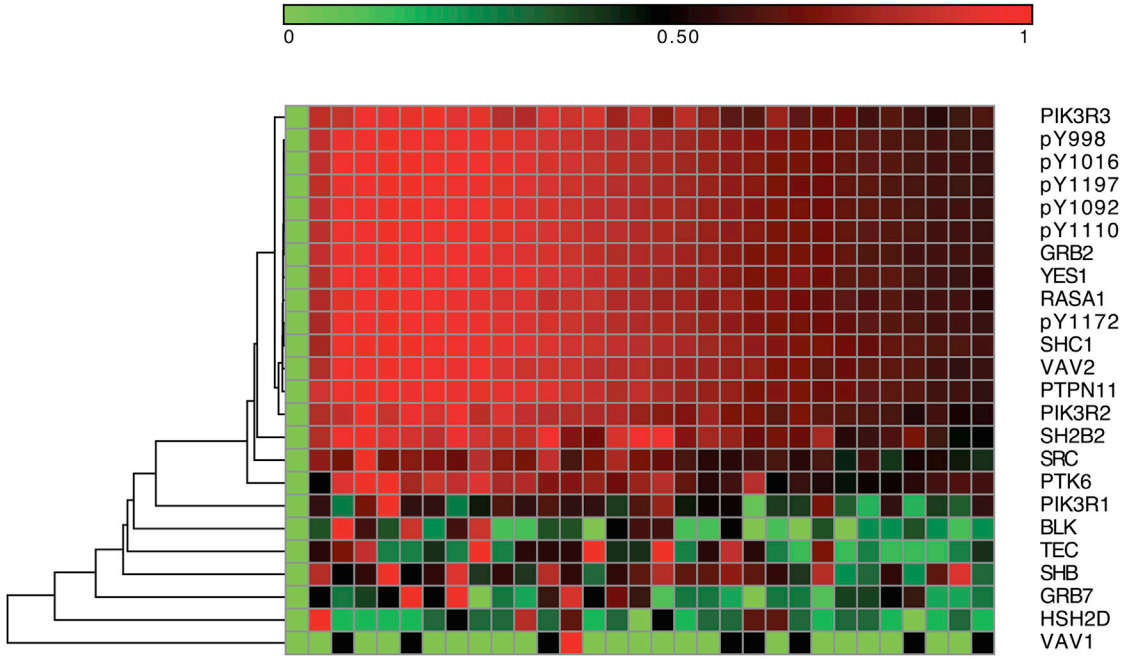

FIGURE 2 Heatmap summarizing predicted time courses of protein recruitment for a HeLa cell-specific version of a model for early events in EGFR signaling. Simulation data have been scaled such that the minimum and maximum values for each time course are 0 and 1, respectively. The results shown were obtained from a stochastic simulation of EGFR signaling in a single cell. To see this figure in color, go online.

Single factors are not expected to always be predictive of recruitment. For example, in the middle panel of Fig. 3, it can be seen that four proteins (RASA1, YES1, VAV2, and PTPN11) are predicted to be more or less equally recruited to EGFR despite having rather different copy numbers, which range from ~26,000 copies per cell for RASA1 to ~300,000 copies per cell for PTPN11 (22). Equal levels of recruitment are predicted for these proteins because of compensating differences in their affinities and network connections (Table S1). Thus, we expect that accurate rankings of interaction importance will require measurements of the multiple quantitative factors that influence interaction strength and the integration of these measurements within appropriate, physicochemical models. It should be noted that single factors are not equally predictive of recruitment. Protein copy number seems to be most predictive; network connectivity seems to be least predictive (Fig. 3).

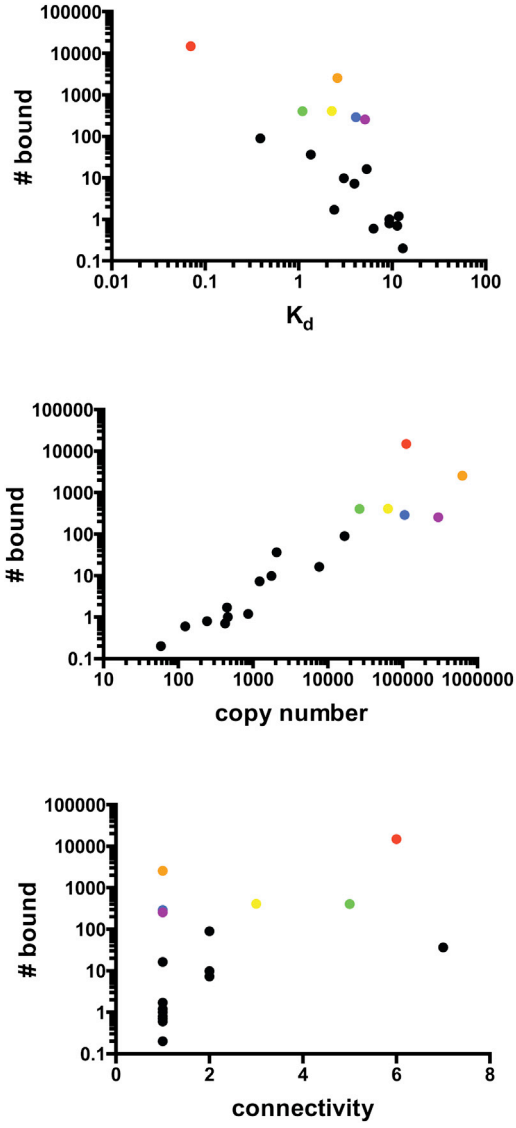

FIGURE 3 The predicted peak level of recruitment of a protein to EGFR in HeLa cells correlates with the protein's highest affinity interaction with EGFR (*top*), its copy number (*middle*), and the number of phosphotyrosines in EGFR with which it interacts (*bottom*). However, the correlations are imperfect. The colored points correspond to SHC1 (*red*), GRB2 (*orange*), YES1 (*yellow*), RASA1 (*green*), VAV2 (*blue*), and PTPN11 (*purple*). To see this figure in color, go online.

Robustly recruited signaling proteins

According to the HeLa cell-specific model parameterized using the data of Kulak et al. (22) (i.e., the model of Table S1), the four signaling proteins most strongly recruited to EGFR in HeLa cells are SHC1, GRB2, YES1, and RASA1, as measured by peak association with EGFR after EGF stimulation (Figs. 1 and 3). Given what we know about EGFR signaling, it is not surprising that SHC1, GRB2, and RASA1 should be among the most strongly recruited proteins; however, strong recruitment of YES1 is arguably surprising. The role of YES1 in EGFR signaling has not been extensively studied, but available results suggest that YES1 contributes to important processes, such as EGFR-mediated regulation of tight junction formation (48). YES1 has also recently been found to play a role in nuclear translocation of EGFR, which promotes resistance to the EGFR inhibitor cetuximab (49).

Relative levels of recruitment predicted by each of our 12 cell line-specific models are summarized in Table S3. Notably, the results obtained from the two HeLa cell-specific models are very similar, which indicates a degree of consistency between the characterizations of HeLa cell protein abundances reported by Geiger et al. (36) and Kulak et al. (22). The results summarized in Table S3 collectively indicate that five well-characterized EGFR binding partners are robustly recruited to EGFR in multiple cell lines: SHC1, GRB2, SRC, RASA1, and PTPN11. For 6 out of 11 cell lines, YES1 is predicted to be more strongly recruited to EGFR than at least one of these five proteins. However, YES1 is predicted to be more strongly recruited than SRC only according to the HeLa cell-specific model parameterized using data from the study of Kulak et al. (22). SRC appears to be the more important Src-family member. In several cell lines, SRC and YES1 are predicted to be co-recruited, which suggests that these two kinases may have nonredundant functions. YES1 may be a more important player in EGFR signaling than is currently appreciated.

Effect of EGF dose on recruitment of signaling proteins to EGFR

Using the HeLa cell-specific model parameterized using data from the study of Kulak et al. (22) (i.e., the model of Table S1), we investigated the effect of EGF dose on predictions of relative levels of protein recruitment. We considered a dose of EGF 10-fold lower and 10-fold higher than the dose used in other calculations (namely, the calculations

of Figs. 1 and 3). The results, which are summarized in Table S4, indicate that relative recruitment levels are insensitive to EGF dose over a wide range of EGF concentrations.

Differences and commonalities in recruitment across cell lines

A notable result of our simulations is prediction of cell line-specific differences in relative recruitment of signaling proteins to EGFR. For example, PLCG1 is predicted to be an important interaction partner of EGFR in HEK 293 cells (Table S3) but not in HeLa cells (Figs. 1 and 3, Table S3). Other signaling proteins that are predicted to be robustly recruited to EGFR only in particular cell lines include SH2B3 in HEK 293 cells and SOCS6 in A549 cells (Table S3). These peculiarities represent potentially testable predictions. Although cell line-specific differences are predicted, there are many commonalities. For example, according to our simulations, SHC1 and GRB2 are among the three proteins most highly recruited to EGFR in both HeLa and HEK 293 cells (Table S3). Signaling proteins that are predicted to be robustly recruited to EGFR in multiple cell lines include SHC1, GRB2, SRC, RASA1, YES1, and PTPN11 (Table S3).

A proposed metric for ranking the relative importance of mapped interactions in a cell signaling network

One would like to have the ability to meaningfully rank the mapped interactions in a cell signaling network, so that characterization of newly discovered (or documented but under studied) interactions can be focused productively on the interactions most likely to have physiological relevance. To rank the relative importance of EGFR interactions with intracellular signaling proteins, various measures of interaction importance could be considered, such as maximal or average extent of association between binding partners over a period of interest. Lifetime of interaction is another potential measure of importance. Here, we consider use of maximal extent of recruitment of a cytoplasmic protein to EGFR after stimulation of signaling by EGF to measure the importance of an interaction between EGFR and a binding partner. Although extent of recruitment is not necessarily indicative of function, the simple colocalization of two signaling proteins is known to have functional significance in specific cases and is believed to be a widely used means for achieving specificity in cell signaling networks (50). The rankings provided in Table S3 are based on the proposed simulation metric, i.e., this table reports relative maximal levels of recruitment. For each of the 11 cell lines of interest, the highly ranked proteins (according to the simulation metric of importance) include well-characterized players in EGFR signaling (Table S3), such as the adaptor proteins SHC1 and GRB2.

Assessment of importance rankings

To evaluate the simulation metric, we considered other simpler metrics. In total, we considered four different metrics for interaction importance: the proposed metric based on simulation results (predicted peak level of recruitment), two single-factor metrics (protein copy number and equilibrium dissociation constant K_D), and a metric based on the ratio of copy number to K_D. We assessed these metrics by comparing predictions of interaction importance (as defined in Materials and Methods) against measurements of EGFR association made for HeLa (38) and HEK 293 (39) cells (Tables 1 and 2). In our analysis of the importance rankings given in Tables 1 and 2, we asked if highly ranked interactions include interactions that are detected experimentally.

Let us first consider the HeLa cell-specific association data of Schulze et al. (38) (Table 1). The measurements of Schulze et al. (38) identified various SH2/PTB domain-containing binding partners for different phosphotyrosine-containing peptides from EGFR. Five of these detected binding partners (SHC1, GRB2, PTPN11, SRC, and CRK) were evaluated for EGFR binding capacity in the study of Hause et al. (5). Hause et al. (5) did not report a K_D for

TABLE 1 Important EGFR binding partners in HeLa cells according to each of four different metrics

Simulation	Copy number (#)	K_D	Ratio (#/K_D)
SHC1	TXN	**SHC1**	**SHC1**
GRB2	**GRB2**	SHC2	**GRB2**
YES1	**PTPN11**	PIK3R2	**PTPN11**
RASA1	CRKL	SOCS6	PIK3R2
VAV2	**CRK**	RASA1	YES1
PTPN11	**SHC1**	PIK3R3	VAV2
PIK3R2	VAV2	PLCG1	RASA1
PIK3R3	SUPT6H	GRAP2	PIK3R3
SH2B2	INPPL1	SH2B3	SH2B2
PTK6	LYN	YES1	PTK6
SRC	YES1	SHD	**SRC**
PIK3R1	NCK1	PIK3R1	PIK3R1
SHB	RASA1	**GRB2**	SHB
GRB7	TNS3	LCK	GRB7
BLK	GRB10	PTK6	TEC
TEC	TNS1	SHC4	SYK
HSH2D	FER	**SRC**	BLK
VAV1	PIK3R2	SH2D1B	HSH2D

The 18 proteins considered in the first column are those expressed in HeLa cells according to Kulak et al. (22) and that also have a measured K_D for at least one EGFR phosphotyrosine, as determined in the study of Hause et al. (5). Bold text is used to indicate proteins detected to bind EGFR (or more precisely, to bind at least one phosphotyrosine-containing peptide derived from EGFR) in the study of Schulze et al. (38). In each column, EGFR binding partners are ranked according to the indicated importance metric: the simulation, copy number (#), K_D or ratio (#/K_D) metric. The most important protein is listed at the top of each column. Except for the third column, rankings are based on the protein copy measurements of Kulak et al. (22). The rankings of the third column are based on the K_D measurements of Hause et al. (5): the ranking of a protein is determined by the minimum K_D among all interactions of the protein with phosphotyrosines in EGFR. The protein with highest affinity is at the top.

TABLE 2 Important EGFR binding partners in HEK 293 cells according to each of four different metrics

Simulation	Copy number (#)	K_D	Ratio (#/K_D)
SHC1	TXN	**SHC1**	**SHC1**
PLCG1	CRKL	SHC2	**GRB2**
GRB2	**GRB2**	PIK3R2	SH2B3
SH2B3	CRK	SOCS6	**PLCG1**
SRC	PTPN11	RASA1	PTPN11
RASA1	SH2B3	PIK3R3	PIK3R2
PTPN11	SUPT6H	**PLCG1**	RASA1
YES1	**PLCG1**	GRAP2	SRC
PIK3R2	NCK1	SH2B3	**YES1**
PIK3R3	SRC	**YES1**	PIK3R1
VAV2	LYN	SHD	**VAV2**
PIK3R1	**YES1**	PIK3R1	SYK
VAV3	**SHC1**	**GRB2**	PIK3R3
SHB	RASA1	LCK	SLA2
TEC	SYK	PTK6	VAV3

The 15 proteins considered in the first column are those expressed in HEK 293 cells according to Geiger et al. (36) and that also have a measured K_D for at least one EGFR phosphotyrosine, as determined in the study of Hause et al. (5). Bold text is used to indicate proteins detected to bind EGFR (or more precisely, to coimmunoprecipitate with EGFR) in the study of Tong et al. (39). It should be noted that copy numbers determined by Tong et al. (39) were not used in our analysis. In each column, EGFR binding partners are ranked according to the indicated importance metric: the simulation, copy number (#), K_D or ratio (#/K_D) metric. The most important protein is listed at the top of each column. Except for the third column, rankings are based on the protein copy measurements of Geiger et al. (36). The rankings of the third column are based on the K_D measurements of Hause et al. (5): the ranking of a protein is determined by the minimum K_D among all interactions of the protein with phosphotyrosines in EGFR. The protein with highest affinity is at the top.

CRK interaction with any of the six EGFR phosphotyrosines of interest. Thus, only the copy number metric provides a rank for CRK. As indicated in Table 1, the simulation and ratio metrics provide high ranks for more detected binding partners than the other two metrics: four detected EGFR binding partners (SHC1, GRB2, PTPN11, and SRC) are among the top 11 in each case. Although the simulation and ratio metrics provide similar rankings for these four EGFR binding partners, the complete rankings produced by the two metrics are distinct, which indicates that importance, as defined by the simulation metric, is affected by competition. The ratio metric does not consider competition. We caution that Schulze et al. (38) did not detect all known EGFR binding partners. Several well-established binding partners, such as RASA1, were not detected. We also caution that the results of Schulze et al. (38), which are based on using bait peptides to capture SH2/PTB domain-containing proteins from whole cell lysates, may not perfectly reflect physiological interactions.

Let us now consider the HEK 293 cell-specific association data of Tong et al. (39) (Table 2). Because of the approach used in the study Tong et al. (39), the experimental results, arguably, better reflect physiological interactions and better correspond to predicted quantities than the experimental results of Schulze et al. (38). Tong et al. (39) detected five EGFR binding partners (SHC1, GRB2, PLCG1, YES1, and VAV2) that contain at least one SH2 or PTB domain among the SH2/PTB domains evaluated for EGFR binding capacity in the study of Hause et al. (5). The simulation and ratio metrics provide similar ranks for these EGFR binding partners: the five EGFR binding partners are among the top 11 according to each metric. As before (in the analysis of Table 1), the rankings provided by the two metrics are distinct. In contrast, the copy number and K_D metrics did not provide similarly high ranks for all of the five detected EGFR binding partners.

The simulation and ratio metrics appear to perform better than the single-factor metrics at assigning high importance ranks to detected EGFR binding partners (Tables 1 and 2). The superior performance of the simulation metric suggests that models that can account for physicochemical and network factors influencing protein-protein interactions (e.g., mass action and competition) may provide more useful measures of interaction importance than rankings based on copy numbers or K_D's alone. Although developing models for cell signaling networks via traditional approaches (e.g., ODE modeling) can be problematic because of the significant combinatorial complexity that arises when site-specific details are considered (43,46,51–53), the rule-based modeling approach used here is relatively straightforward. Thus, we expect that rule-based models will be increasingly used in the future for integrating and analyzing multiple high-throughput proteomic data sets, such as those considered in this study.

Discrimination between the two best-performing metrics considered here, the simulation and ratio metrics, seems not to be possible because of the limited amount of association data available for testing predictions of interaction importance (Tables 1 and 2). Thus, evaluation of importance ranking approaches would benefit from generation of more sensitive and comprehensive assays of association. Notably, the simulation and ratio metrics produce distinct rankings, and with new data, it could potentially be established that one of the two metrics performs better at accurately identifying interactions responsible for recruitment of signaling proteins to EGFR. In any case, it appears that competition, which is considered in the simulation metric (but not in any of the simpler metrics), may influence recruitment of signaling proteins to EGFR.

CONCLUSIONS

There has been much discussion about the promise of combining modeling and proteomics (54), but these two fields have not so far become tightly integrated. One reason is the difficulty of building and analyzing a model with the scope and mechanistic detail required to connect to proteomic data. Such models must encompass a larger number of proteins and interactions than are usually considered in models to take full advantage of proteomic data (e.g., near

comprehensive measurements of absolute protein copy numbers). The technical challenges of model construction and parameterization have meant that most models developed to date only connect to a handful of biochemical readouts. With the development of scalable, rule-based modeling and simulation approaches, there are now exciting opportunities to broaden the scope of our modeling efforts. As demonstrated here, it is possible to use a mechanistic model based on the law of mass action to integrate measurements of multiple quantitative factors that influence interactions among signaling proteins, allowing the combined effects of copy number, affinity, and competition on recruitment to be elucidated. We note that attempts to integrate multiple data sets that characterize cell signaling networks, as in the study of EGFR signaling by Waters et al. (55), typically do not involve the use a mechanistic model (i.e., a model that synthesizes and formalizes available insights about molecular mechanisms of signaling).

By using mechanistic, cell line-specific models to integrate high-throughput K_D and protein copy number measurements, we were able to assess the effects of affinity, abundance, and competition on recruitment of signaling proteins to EGFR in particular cellular contexts. Our findings include 1) the possibility of noisy recruitment, even when an EGFR binding partner is present at thousands of copies per cell; 2) the identification of YES1, which is not widely viewed as an important player in EGFR signaling, as a protein that is robustly recruited to EGFR in multiple cell lines; and 3) the elucidation of several nonobvious cell line-specific differences in recruitment of signaling proteins to EGFR (e.g., strong recruitment of PLCG1 to EGFR in HEK 293 cells but not in other cell lines considered). Moreover, we evaluated a proposed metric of interaction importance, which labels well-known players in EGFR signaling as important and provides rankings of relative importance that agree well with experimental measurements of association. Finally, it should be noted that our models yield cell line-specific predictions about the kinetics of recruitment, which could potentially be tested experimentally.

Although our generic model captures EGFR phosphotyrosine interactions with SH2/PTB domain-containing proteins more comprehensively than earlier models for EGFR signaling, we caution that this model does not represent a comprehensive synthesis of the mechanistic knowledge available about EGFR signaling. Indeed, parts of it represent important processes (e.g., endocytosis) only very coarsely, and the treatment of phosphotyrosine-SH2/PTB domain interactions is simplified, as a matter of necessity. For example, cooperative binding, about which little is known, is not considered. In the future, it will be important to find a means to consider the effects of negative and positive binding cooperativity, which are likely at play. Negative cooperativity can arise from steric clashes, which have been studied computationally in the context of EGFR signaling on the basis of protein structures (56). Positive cooperativity is known to be involved in the interaction of EGFR with CBL, which depends on GRB2 (57,58). (It should be noted that CBL, which recognizes phosphotyrosines using an assembly of conserved domains, is not included in the models discussed here.) It will also be important to consider other complications, including site-specific differences in phosphorylation and dephosphorylation kinetics, heterodimers of ErbB family members (which may recruit signaling proteins differently than EGFR homodimers), and the effects of subcellular location of the receptor and other signaling proteins on receptor signaling.

Despite the usual uncertainties surrounding computational models of cellular regulatory systems, we are optimistic that analysis of a model incorporating mechanistic understanding and connecting to multiple proteomic data sets, as demonstrated here, has the potential to become a routinely applied and fairly reliable means for obtaining insights into system-level behaviors that would not otherwise be discernible.

SUPPORTING MATERIAL

Supporting text, four tables, and Supporting Software are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(15)00231-3.

ACKNOWLEDGMENTS

This work was supported by grants from the Arizona Biomedical Research Commission (0806) and the National Institute of General Medical Sciences of the National Institutes of Health (P50GM085273 and R01GM111510). E.C.S. and W.S.H. acknowledge support from the Randy Pausch Scholars Program, which is sponsored by the TGen Foundation, Howard Young, and the Global Cure National Advisory Council. M.S.C. acknowledges support from the Helios Foundation.

SUPPORTING CITATIONS

References (59–67) appear in the Supporting Material.

REFERENCES

1. Kandasamy, K., S. S. Mohan, …, A. Pandey. 2010. NetPath: a public resource of curated signal transduction pathways. *Genome Biol.* 11:R3.
2. Csermely, P., T. Korcsmáros, …, R. Nussinov. 2013. Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol. Ther.* 138:333–408.
3. Nobeli, I., A. D. Favia, and J. M. Thornton. 2009. Protein promiscuity and its implications for biotechnology. *Nat. Biotechnol.* 27:157–167.
4. Schreiber, G., and A. E. Keating. 2011. Protein binding specificity versus promiscuity. *Curr. Opin. Struct. Biol.* 21:50–61.
5. Hause, Jr., R. J., K. K. Leung, …, R. B. Jones. 2012. Comprehensive binary interaction mapping of SH2 domains via fluorescence polarization reveals novel functional diversification of ErbB receptors. *PLoS ONE.* 7:e44471.
6. Lim, W. A., and T. Pawson. 2010. Phosphotyrosine signaling: evolving a new cellular communication system. *Cell.* 142:661–667.
7. Panayotou, G., G. Gish, …, M. D. Waterfield. 1993. Interactions between SH2 domains and tyrosine-phosphorylated platelet-derived

growth factor beta-receptor sequences: analysis of kinetic parameters by a novel biosensor-based approach. *Mol. Cell. Biol.* 13:3567–3576.

8. Ladbury, J. E., and S. Arold. 2000. Searching for specificity in SH domains. *Chem. Biol.* 7:R3–R8.
9. O'Rourke, L., and J. E. Ladbury. 2003. Specificity is complex and time consuming: mutual exclusivity in tyrosine kinase-mediated signaling. *Acc. Chem. Res.* 36:410–416.
10. Jones, R. B. 2013. Do low-affinity ErbB receptor protein interactions represent the base of a cell signaling iceberg? *Expert Rev. Proteomics.* 10:115–118.
11. Oda, K., Y. Matsuoka, …, H. Kitano. 2005. A comprehensive pathway map of epidermal growth factor receptor signaling. *Mol. Syst. Biol.* 1: 2005.0010.
12. Chelliah, V., C. Laibe, and N. Le Novère. 2013. BioModels Database: a repository of mathematical models of biological processes. *Methods Mol. Biol.* 1021:189–199.
13. Keshava Prasad, T. S., R. Goel, …, A. Pandey. 2009. Human Protein Reference Database—2009 update. *Nucleic Acids Res.* 37:D767–D772.
14. Zhang, Q. C., D. Petrey, …, B. Honig. 2012. Structure-based prediction of protein-protein interactions on a genome-wide scale. *Nature.* 490:556–560.
15. Zhang, Q. C., D. Petrey, …, B. Honig. 2013. PrePPI: a structure-informed database of protein-protein interactions. *Nucleic Acids Res.* 41:D828–D833.
16. Chatr-Aryamontri, A., B. J. Breitkreutz, …, M. Tyers. 2013. The BioGRID interaction database: 2013 update. *Nucleic Acids Res.* 41:D816–D823.
17. Kuzu, G., O. Keskin, …, R. Nussinov. 2012. Constructing structural networks of signaling pathways on the proteome scale. *Curr. Opin. Struct. Biol.* 22:367–377.
18. Petrey, D., and B. Honig. 2014. Structural bioinformatics of the interactome. *Annu. Rev. Biophys.* 43:193–210.
19. Koytiger, G., A. Kaushansky, …, G. MacBeath. 2013. Phosphotyrosine signaling proteins that drive oncogenesis tend to be highly interconnected. *Mol. Cell. Proteomics.* 12:1204–1213.
20. Tinti, M., L. Kiemer, …, G. Cesareni. 2013. The SH2 domain interaction landscape. *Cell Reports.* 3:1293–1305.
21. Mann, M., N. A. Kulak, …, J. Cox. 2013. The coming age of complete, accurate, and ubiquitous proteomes. *Mol. Cell.* 49:583–590.
22. Kulak, N. A., G. Pichler, …, M. Mann. 2014. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods.* 11:319–324.
23. Johnsson, N. 2014. Analyzing protein-protein interactions in the post-interactomic era. Are we ready for the endgame? *Biochem. Biophys. Res. Commun.* 445:739–745.
24. Jones, R. B., A. Gordus, …, G. MacBeath. 2006. A quantitative protein interaction network for the ErbB receptors using protein microarrays. *Nature.* 439:168–174.
25. Kholodenko, B., M. B. Yaffe, and W. Kolch. 2012. Computational approaches for analyzing information flow in biological networks. *Sci. Signal.* 5:re1.
26. Kholodenko, B. N., O. V. Demin, …, J. B. Hoek. 1999. Quantification of short term signaling by the epidermal growth factor receptor. *J. Biol. Chem.* 274:30169–30181.
27. Macdonald, J. L., and L. J. Pike. 2008. Heterogeneity in EGF-binding affinities arises from negative cooperativity in an aggregating system. *Proc. Natl. Acad. Sci. USA.* 105:112–117.
28. Chen, W. W., B. Schoeberl, …, P. K. Sorger. 2009. Input-output behavior of ErbB signaling pathways as revealed by a mass action model trained against dynamic data. *Mol. Syst. Biol.* 5:239.
29. Schoeberl, B., E. A. Pace, …, U. B. Nielsen. 2009. Therapeutically targeting ErbB3: a key node in ligand-induced activation of the ErbB receptor-PI3K axis. *Sci. Signal.* 2:ra31.
30. Kleiman, L. B., T. Maiwald, …, P. K. Sorger. 2011. Rapid phosphotransfer by receptor tyrosine kinases impacts downstream signaling and drug binding. *Mol. Cell.* 43:723–737.
31. Creamer, M. S., E. C. Stites, …, R. G. Posner. 2012. Specification, annotation, visualization and simulation of a large rule-based model for ERBB receptor signaling. *BMC Syst. Biol.* 6:107.
32. Volinsky, N., and B. N. Kholodenko. 2013. Complexity of receptor tyrosine kinase signal processing. *Cold Spring Harb. Perspect. Biol.* 5:a009043.
33. Faeder, J. R., M. L. Blinov, and W. S. Hlavacek. 2009. Rule-based modeling of biochemical systems with BioNetGen. *Methods Mol. Biol.* 500:113–167.
34. Yang, J., M. I. Monine, …, W. S. Hlavacek. 2008. Kinetic Monte Carlo method for rule-based modeling of biochemical networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 78:031910.
35. Sneddon, M. W., J. R. Faeder, and T. Emonet. 2011. Efficient modeling, simulation and coarse-graining of biological complexity with NFsim. *Nat. Methods.* 8:177–183.
36. Geiger, T., A. Wehner, and M. Mann. 2012. Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. *Mol. Cell. Proteomics.* 11: M111.014050.
37. Klinke, 2nd, D. J. 2010. Signal transduction networks in cancer: quantitative parameters influence network topology. *Cancer Res.* 70:1773–1782.
38. Schulze, W. X., L. Deng, and M. Mann. 2005. Phosphotyrosine interactome of the ErbB-receptor kinase family. *Mol. Syst. Biol.* 1: 2005. 0008.
39. Tong, J., P. Taylor, and M. F. Moran. 2014. Proteomic analysis of the epidermal growth factor receptor (EGFR) interactome and post-translational modifications associated with receptor endocytosis in response to EGF and stress. *Mol. Cell. Proteomics.* 13:1644–1658.
40. Goh, L. K., and A. Sorkin. 2013. Endocytosis of receptor tyrosine kinases. *Cold Spring Harb. Perspect. Biol.* 5:a017459.
41. Haugh, J. M., K. Schooler, …, D. A. Lauffenburger. 1999. Effect of epidermal growth factor receptor internalization on regulation of the phospholipase C-gamma1 signaling pathway. *J. Biol. Chem.* 274:8958–8965.
42. Blinov, M. L., J. R. Faeder, …, W. S. Hlavacek. 2004. BioNetGen: software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics.* 20:3289–3291.
43. Chylek, L. A., L. A. Harris, …, W. S. Hlavacek. 2014. Rule-based modeling: a computational approach for studying biomolecular site dynamics in cell signaling systems. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 6:13–36.
44. Yang, J., and W. S. Hlavacek. 2011. The efficiency of reactant site sampling in network-free simulation of rule-based models for biochemical systems. *Phys. Biol.* 8:055009.
45. Blinov, M. L., J. R. Faeder, …, W. S. Hlavacek. 2006. A network model of early events in epidermal growth factor receptor signaling that accounts for combinatorial complexity. *Biosystems.* 83:136–151.
46. Stefan, M. I., T. M. Bartol, …, M. B. Kennedy. 2014. Multi-state modeling of biomolecules. *PLOS Comput. Biol.* 10:e1003844.
47. Toettcher, J. E., O. D. Weiner, and W. A. Lim. 2013. Using optogenetics to interrogate the dynamic control of signal transmission by the Ras/Erk module. *Cell.* 155:1422–1434.
48. Wang, Y., D. Du, …, Z. Chen. 2006. Tyrosine phosphorylated Par3 regulates epithelial tight junction assembly promoted by EGFR signaling. *EMBO J.* 25:5058–5070.
49. Iida, M., T. M. Brand, …, D. L. Wheeler. 2013. Yes and Lyn play a role in nuclear translocation of the epidermal growth factor receptor. *Oncogene.* 32:759–767.
50. Park, S. H., A. Zarrinpar, and W. A. Lim. 2003. Rewiring MAP kinase pathways using alternative scaffold assembly mechanisms. *Science.* 299:1061–1064.
51. Hlavacek, W. S., J. R. Faeder, …, B. Goldstein. 2003. The complexity of complexes in signal transduction. *Biotechnol. Bioeng.* 84:783–794.

52. Hlavacek, W. S., J. R. Faeder, …, W. Fontana. 2006. Rules for modeling signal-transduction systems. *Sci. STKE.* 2006:re6.

53. Mayer, B. J., M. L. Blinov, and L. M. Loew. 2009. Molecular machines or pleiomorphic ensembles: signaling complexes revisited. *J. Biol.* 8:81.

54. Cox, J., and M. Mann. 2011. Quantitative, high-resolution proteomics for data-driven systems biology. *Annu. Rev. Biochem.* 80:273–299.

55. Waters, K. M., T. Liu, …, B. D. Thrall. 2012. Network analysis of epidermal growth factor signaling using integrated genomic, proteomic and phosphorylation data. *PLoS ONE.* 7:e34515.

56. Hsieh, M. Y., S. Yang, …, B. S. Wilson. 2010. Spatio-temporal modeling of signaling protein recruitment to EGFR. *BMC Syst. Biol.* 4:57.

57. Jiang, X., F. Huang, …, A. Sorkin. 2003. Grb2 regulates internalization of EGF receptors through clathrin-coated pits. *Mol. Biol. Cell.* 14:858–870.

58. Sigismund, S., V. Algisi, …, P. P. Di Fiore. 2013. Threshold-controlled ubiquitination of the EGFR directs receptor fate. *EMBO J.* 32:2140–2157.

59. Kozer, N., D. Barua, …, A. H. Clayton. 2013. Exploring higher-order EGFR oligomerization and phosphorylation—a combined experimental and theoretical approach. *Mol. Biosyst.* 9:1849–1863.

60. Kozer, N., D. Barua, …, A. H. Clayton. 2014. Recruitment of the adaptor protein Grb2 to EGFR tetramers. *Biochemistry.* 53:2594–2604.

61. Shankaran, H., Y. Zhang, …, H. Resat. 2012. Integrated experimental and model-based analysis reveals the spatial aspects of EGFR activation dynamics. *Mol. Biosyst.* 8:2868–2882.

62. Moran, U., R. Phillips, and R. Milo. 2010. SnapShot: key numbers in biology. *Cell.* 141:1262–1262.e1.

63. Wofsy, C., B. Goldstein, …, H. S. Wiley. 1992. Implications of epidermal growth factor (EGF) induced egf receptor aggregation. *Biophys. J.* 63:98–110.

64. Zhou, M., S. Felder, …, J. Schlessinger. 1993. Real-time measurements of kinetics of EGF binding to soluble EGF receptor monomers and dimers support the dimerization model for receptor activation. *Biochemistry.* 32:8193–8198.

65. Low-Nam, S. T., K. A. Lidke, …, D. S. Lidke. 2011. ErbB1 dimerization is promoted by domain co-confinement and stabilized by ligand binding. *Nat. Struct. Mol. Biol.* 18:1244–1249.

66. Fan, Y. X., L. Wong, …, G. R. Johnson. 2004. Ligand regulates epidermal growth factor receptor kinase specificity: activation increases preference for GAB1 and SHC versus autophosphorylation sites. *J. Biol. Chem.* 279:38143–38150.

67. Endres, N. F., R. Das, …, J. Kuriyan. 2013. Conformational coupling across the plasma membrane in activation of the EGF receptor. *Cell.* 152:543–556.