The chemical organization of signaling interactions

*Upinder S. Bhalla**

Neurobiology, National Centre for Biological Sciences, GKVK Campus, Bangalore 560065, India

ABSTRACT

Motivation: Cellular chemical signaling pathways form complex networks that are beginning to be studied at the level of chemical kinetics and databases of reactions. Chemical reaction details are traditionally represented as lists of reactions and rates. This does not map readily to the block diagram representation familiar to biologists, and obscures the functional organization of signaling networks. This study examines motifs in signaling chemistry and reports common features that may help to formalize such a mapping between pathway block diagrams and the chemistry. The same motifs may facilitate data representation and provide functional abstraction of the chemistry.

Results: I classified 74 interactions between 25 signaling pathways in terms of shared chemical motifs. All interactions in this dataset consist of a few communicating molecules from one set of pathways, and a replicating set of reactions and molecules from another. Each unique combination of interacting pathways duplicates the chemical reaction scheme of this replicating set, but involves different rate constants. Signaling pathways can therefore be described in an object-oriented manner as sets of core reactions with well-defined interfaces between pathways. This generalization lends itself to designing simulators and databases for signaling networks.

Availability: Software and example models are freely available from http://www.ncbs.res.in/~bhalla/examples/ EGFR_example.html.

Contact: bhalla@ncbs.res.in

INTRODUCTION

The block diagram level representation of signaling pathways is a standard and essential abstraction when dealing with complex signaling networks. This level of description lumps several chemical steps together into blocks that exchange inputs and outputs with other blocks. These blocks usually involve major signaling enzymes or tightly coupled enzyme cascades. Genetic,

pharmacological and molecular biological manipulations have been used to identify upstream and downstream signaling enzymes (Alberts *et al.*, 1994; Lauffenburger and Linderman, 1993), and high-throughput assays promise to greatly expand the coverage of such signaling networks (Tucker *et al.*, 2001; Ideker *et al.*, 2001). Such block level descriptions are qualitative. It is valuable to analyse signaling networks at the level of detailed chemical reactions to understand signaling dynamics and quantitative activation details (Bray, 1995; Bhalla and Iyengar, 1999). The quantitative description of biological signaling relies on an expansion of signaling block diagrams into chemical reaction details. This expansion replaces qualitative information about upstream and downstream interactions with computable rate equations and specific molecular species, whose values can be experimentally measured and used to test signaling models. Unfortunately this results in very complex reaction diagrams which do not readily map onto simple block diagrams, and which are difficult to scale up (Figure 1). This paper reports an analysis of many such reaction schemes and describes a consistent architecture in the chemistry of signaling interactions. This facilitates encapsulation of complex signaling chemistry into self-contained modules with consistent interaction motifs. Other studies have considered motifs at the level of ensembles of signaling pathways (Asthagiri and Lauffenburger, 2000; Bhalla and Iyengar, 1999), but the focus of the current paper is at the level of chemical architecture.

Several studies have been carried out at the 'bottom up' level of chemically detailed models of complex signaling networks. Among these are models of the mitogen associated protein kinase cascade (Huang and Ferrell, 1996; Levchenko *et al.*, 2000; Kholodenko, 2000), phototransduction (Lamb and Pugh, 1992; Blackwell, 2000), chemotaxis (Morton-Firth *et al.*, 1999), genetic circuits (Arkin *et al.*, 1998), and synaptic signaling (Bhalla and Iyengar, 1999; Kuroda *et al.*, 2001). The approach in these cases is to represent each individual chemical reaction separately, providing rate constants and parameters for each reaction along with supplementary information such

---

*To whom correspondence should be addressed.
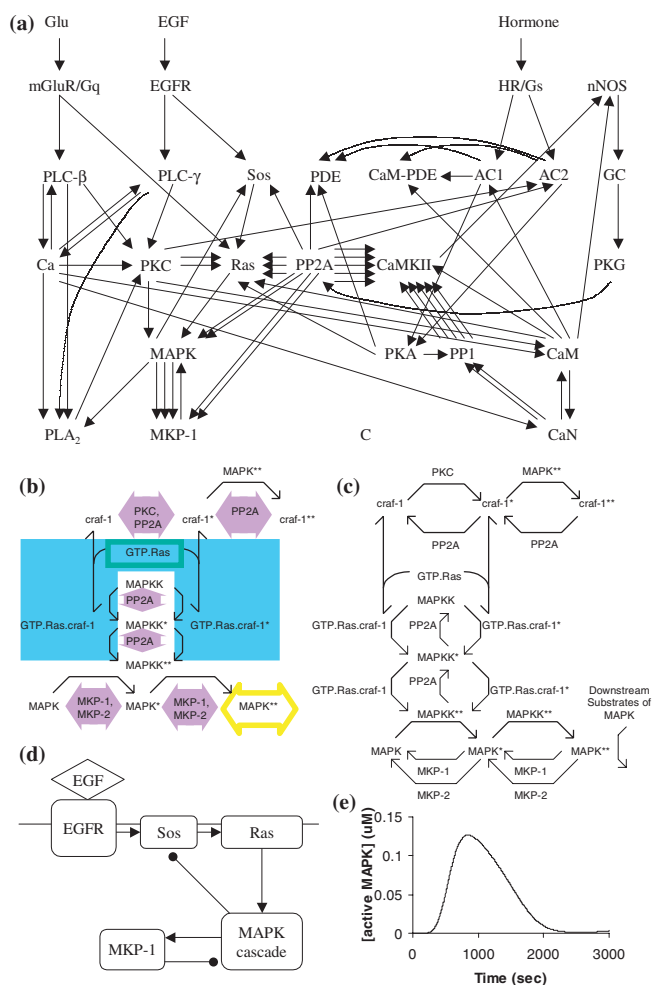
Fig. 1. (a) Block diagram representing the 25 pathway blocks and their 74 interactions from the primary dataset. Arrows represent distinct chemical interactions between pathways but do not distinguish between activation and repression. The number of arrows indicates the number of reaction sites, not the strength of the interaction. For example, each of the 5 arrows from PP2A to CaMKII represents a dephosphorylation of a distinct site or chemical state of CaMKII. The same set of 5 sites are dephosphorylated by PP1. (b) Conversion of MAPK reaction diagram into format based on core reactions and interaction reactions. Pale gray blocks represent reactions where input molecules combine with molecules in the MAPK pathway. The input molecule (GTP.Ras) is enclosed in a thick dark box. Filled arrows represent inputs from enzymatic reactions. In this model PKC, PP2A, MKP1 and MKP2 are inputs at various points. The dark open arrow represents an output reaction, in this case the phosphorylation of substrates by MAP-Kinase. (c) Expanded representation of the MAPK diagram and its inputs and outputs in terms of chemical binding reactions and enzymes. (d) Block-diagram representation of subset of signaling pathways representing EGF stimulation of MAPK cascade. (e) Simulation results for continuous EGF (50 nM) stimulation of MAPK using explicit specification of all reactions, and equivalent cascade connected up using the higher-level specification rules. The two simulations produce identical results.

as data sources. There are several 'top down' projects to handle this sort of complexity at the level of signaling database projects (Hucka *et al.*, 2000; Hedley *et al.*, 2001) and simulators (Stiles and Bartol, 2000; Bhalla, 1998; Schaff *et al.*, 1997; Tomita *et al.*, 1999). In general, these projects adopt the same reaction-by-reaction description supplemented with database, simulator and interface tools. The modular architecture of signaling interactions reported here reconciles the need for complete chemical detail with more analytically and conceptually tractable block diagrams. It is also amenable to an object-oriented representation suitable for simulators and databases.

METHODS AND IMPLEMENTATION

The approach in this paper is to identify common reaction motifs that can be used to classify interactions and provide a higher-level description of complex pathways while retaining quantitative chemical detail. The level of description is that of chemical interactions, and the motifs are therefore common architectural features of the chemistry. The analysis used a primary dataset of 25 signaling enzymes and pathways that are diverse, highly interconnected, and ubiquitous. 74 interactions between these pathways were included in the dataset. These pathways and their interactions were described in terms of reaction schemes based on published models and the literature (Bhalla and Iyengar, 1999; Kuroda *et al.*, 2001; Brondello *et al.*, 1999). The reactions, and rate constants are accessible at http://www.ncbs.res.in/~bhalla/examples/EGFR_example.html. The size of this dataset is limited by the requirement that it include well-defined signaling pathway function and interaction, as well as chemical detail at the level of individual reactions. As a supplementary dataset, other signaling chemistry models including the MAPK cascade (Huang and Ferrell, 1996; Levchenko *et al.*, 2000; Kholodenko, 2000), phototransduction (Lamb and Pugh, 1992; Blackwell, 2000), chemotaxis (Morton-Firth *et al.*, 1999), genetic circuits (Arkin *et al.*, 1998) were also analysed.

Pathways were initially partitioned in terms of 'core' and 'interaction' reactions by inspection. The core steps are reactions that do not involve any molecules outside the pathway. Interaction steps are those that do. Subsequently the reaction schemes were set up using computer specification of reactions between core and interaction steps to ensure that all reactions were unambiguously categorized (Figure 1). Pathways defined in this manner were interconnected using only the high-level interaction rules. The completeness of the high-level specification of pathways was confirmed by comparing computer-generated reaction schemes with manually specified reaction steps, and comparing computed outcomes of reactions specified at the level of individual reactions with reactions specified at the block level (Figure 1b–e). Computer verification was

performed using developmental version 8 of Kinetikit, the kinetics modeling extension to the simulator GENESIS (Bhalla, 1998). Models, Kinetikit 8, and the GENESIS simulator are available at the site http://www.ncbs.res. in/~bhalla/examples/EGFR_example.html. The goal of this exercise was to test the generality of the above classification of chemical interactions, and to confirm that these rules for defining interactions would indeed completely specify all chemical steps and intermediates. The implementation and simulation of the model was therefore intended as a proof of concept and is not the focus of this paper. A library of pre-existing pathway models (Bhalla and Iyengar, 1999) was used as the starting point for the implementation. This library consists of signaling pathways defined at the level of individual reactions. These models were converted to the modular form discussed below where pathways were defined in terms of core reactions and interaction reactions. The modeling interface in Kinetikit was extended so that dragging from an upstream to a downstream pathway initiated the replication of reactions and interaction molecules according to the rules discussed below. The complete set of 74 pathway interactions were then set up using such drag-and-drop operations. In each case the generated reactions were found to be equivalent to the hand-connected reaction-level chemical scheme of the previous models. This implementation was designed to accommodate pairwise interactions. A serendipitous outcome of the implementation exercise was to uncover the presence of higher-order combinatorial interactions and suggest the utility of constructing higher-order modules. It also played the role of a chemical computer-aided design tool in defining reactions as core reactions and interaction reactions, a process that was tedious and error-prone when done manually.

RESULTS

Classification of interactions

Here I describe how the classification of signaling interactions reduces to a single chemical motif. An initial classification of interactions identified three interaction motifs between pathways: Messengers, upstream replicating reactions, and downstream replicating reactions. As described below, messengers were identified as a special case of downstream replicating reactions. Finally, the chemical architecture of upstream and downstream replicating reactions was seen to be identical although the two cases have opposite directions of information flow. This reduces all three interaction motifs to a single general chemical architecture.

1. Messengers represent the simplest kind of interaction that appears from the above reaction schemes, where a single chemical species communicates be-
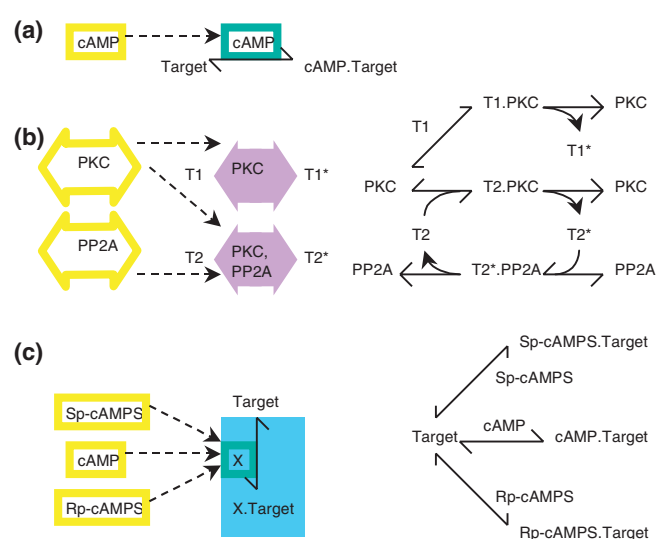
Fig. 2. Motifs for signaling interactions. (a) Messenger interactions. The upstream pathway contributes a molecule which is involved in downstream reactions. (b) Upstream replicating interactions, involving kinases and phosphatases. A modular representation is shown on the left, and the reaction details on the right. The downstream pathway communicates via two molecules, the target and its phosphorylated state. The upstream pathway replicates the enzymatic reaction scheme and distinct enzyme–substrate complexes are formed for each target. (c) Downstream replicating interactions, illustrating the more general case of messenger interactions. Several possible upstream molecules may bind to the downstream target. This leads to formation of multiple bound forms of the target, each with distinct rates.

tween two pathways (Figure 2a). Here the upstream pathway regulates levels of a messenger, which then participates in non-covalent reactions involving the downstream pathway. The conventional second messengers (calcium, cyclic AMP, inositol trisphosphate, diacylglycerol, etc.) fall into this category. Several protein molecules and complexes including $G_{\beta\gamma}$ and calcium-calmodulin interact in an equivalent way.

2. Upstream replicating reactions are exemplified by protein kinase and phosphatase reactions (Figure 2b). Here the upstream pathway enters into a series of reactions (e.g. the Michaelis–Menten enzyme scheme) with the target. There is always at least one reaction intermediate (e.g. the enzyme–substrate complex) which involves both the upstream and downstream molecule. There is also at least one reaction rate that is specific to each upstream–downstream combination. Thus the upstream pathway must define a series of reactions and intermediates that are replicated for every dis-

tinct downstream target. The downstream pathway, however, only needs to provide a few molecules (e.g. the protein and its phosphorylated form) for the upstream pathway to act upon. There is an economy of concepts here when one considers opposing pairs of actions such as kinase and phosphatase, which are represented as an equivalent interaction but with opposite direction.

3. Downstream replicating interactions are a mirror of the second motif, that is, there are replicating reactions that are situated downstream rather than upstream (Figure 2c). Examples include ligand–receptor interactions, the GAP activity of target molecules downstream of activated GTP-Gα and a more flexible representation of messengers. In each of these cases the downstream target may receive inputs from a variety of upstream sources. In the case of ligand–receptor interactions, a receptor may form a number of intermediate states upon binding of a ligand. Each of these intermediate states and the kinetics of their formation are unique to the ligand. Thus the upstream pathway must supply a ligand to enter into this interaction, whereas the downstream pathway must define a series of reaction steps and intermediate states for each ligand. A similar argument applies for GAP activity of target molecules, except that in this case the downstream pathway also effects a transformation of the upstream molecule by converting GTP-Gα to GDP-Gα.

It is useful to identify messenger interactions as a special case of downstream replicating interactions. Most messengers have pharmacologically related chemical species (e.g. Sp-cAMPS for cAMP) which act through similar chemistry but with different rates (Figure 2c). Models of experiments involving radiotracers also require replicated reaction steps to keep track of each radioisotope. Each of these situations involves the messenger molecule generated by the upstream pathway, and a set of downstream reactions which must be replicated for each related messenger. Thus even a single ion messenger such as Ca^{2+} may need to be represented in some cases using replicating downstream reactions.

All the interactions in this dataset fall into one of these three cases. Each of the three cases can be generalized as one pathway presenting a number of communicating molecules, and the other presenting a replicating set of reactions involving those molecules. Thus a single organizational motif underlies all chemical signaling interactions examined in this dataset.

Higher order interactions

Two classes of reaction in this dataset require further generalization. The first is the formation of ternary and higher order complexes. For example, the ligand–receptor-G-protein complex in principle is a combination of five distinct molecules: ligand, receptor and the three G-protein subunits. (Figure 3a, c). Thus the reactions involving this complex must be replicated for every possible combination of the ligand, receptor and G-protein subunits. This situation of combinatorial inputs arises 6 times in a dataset of 25 pathways, and is therefore quite frequent. The second situation arises when the pathway has a replicating input giving rise to multiple active states of an output enzyme, each of which in turn can act on multiple targets. (Figure 3b). This case occurs 5 times in 25 pathways. Both situations involve replicating chemical steps arising in a combinatorial manner rather than just pairwise interactions and can therefore be generalized in the same way.

This generalized organizational principle is that all signaling interactions consist of one or more communicating molecules from one set of pathways, and a replicating set of reactions and molecules from another pathway. The chemical scheme of this replicating set of reactions is repeated for every unique combination of interacting pathways. This motif is a straightforward generalization of the previous one to account for higher-order combinations of interactions.

Higher order modules

Within this framework, multiple isoforms of enzymes sharing common inputs and outputs can be grouped into higher-order modules (Figure 3d). Although the current dataset includes very few isoforms explicitly, almost all modules in the dataset are based on enzymes with multiple isoforms. At least 20 of the 25 pathways would be candidates for higher-order modules, in some cases involving over 8 members (e.g. adenylyl cyclase; Pieroni *et al.*, 1993).

This concept of higher-order modules is also applicable to the frequent cases where there are multiple sites of action of a single enzyme (e.g. phosphatase action on the mitogen activated protein kinase (MAPK) and calcium calmodulin type II kinase (CaMKII) in Figures 1 and 4). This situation occurs between 8 pairs of pathways in the dataset, amounting to 24 interactions out of 74.

Although the formation of higher order modules does not directly modify the concept of a common general motif for signaling interactions, it does broaden the concept of self-contained 'blocks' of signaling pathways. Such blocks are themselves built up from simpler signaling units but present the same interaction motifs to other pathways.

A corollary of this architecture of pathways is that the mechanistic details of any pathway can change without impinging on the reaction schemes for other pathways. Such changes can apply both to the core reactions and the
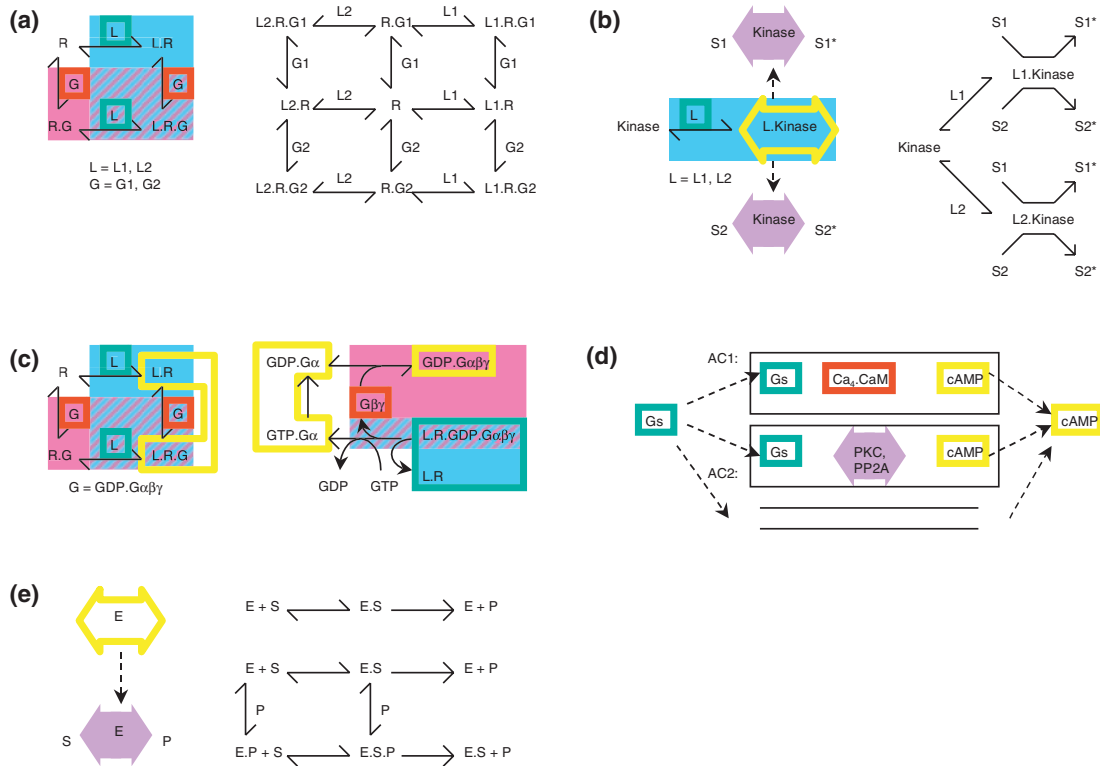
Fig. 3. Combinatorial interactions. (a) Combinations of receptor, ligand and G-protein in block and reaction diagram forms. Two ligands and two G-proteins are illustrated. Stripes indicate higher-order combinations. In the expanded reaction scheme on the right, there are four combinations of ligand and G protein. (b) Combinatorial reactions arising from two inputs activating a kinase, and two substrates. There are four combinations of ligand and substrate. (c) Partitioning a ligand–receptor–G-protein complex between receptor/ligand reactions and G-protein interactions. The reactions are tightly coupled. Three sets of molecules must communicate, and both upstream and downstream molecules must be replicated for each combination of signals. This set of reactions is therefore more suitable for representation as a single block. (d) A composite module where multiple adenylyl cyclase isoforms share Gs inputs and cAMP outputs. Only the possible inputs and outputs for this composite module are displayed. (e) Modularity in interactions. A simple enzyme reaction scheme is replaced by a more complex reaction scheme including product inhibition. The downstream pathway in either case needs only to provide the communicating molecules S and P, and the details of the reaction are specified by the replicating part of the upstream pathway.

reactions involved in interactions, provided the number of communicating molecules does not change (Figure 3e). This permits individual pathway specifications to be upgraded to incorporate more detail without disturbing any other part of the signaling network.

Rules for constructing modules

The partitioning of complex signaling networks into pathways and their interactions is a three-step process:

(1) Identify interactions in the form of messengers, upstream replicating, or downstream replicating motifs. This sets up the basic interaction motifs.

(2) Identify instances of higher order interactions. This determines cases where interactions need to be treated in a combinatorial manner.

(3) Determine whether closely linked modules can be

encapsulated as higher-order modules. This makes the granularity of the description somewhat coarser and simplifies the overall network description.

It is probable that this process could be automated, but the identification of interactions is a trivial addition to the labor-intensive process of development of chemically detailed signaling models. Furthermore, the concept of replicating interactions is itself a useful organizing principle for developing such models since it partitions the problem into the specification of core reactions and interactions. It is therefore unclear that such an algorithm would have practical utility.

It is possible to contrive subdivisions of pathways that require a dual up–downstream replicating set of reactions, but these require partitioning of pathways at the point where reactions are dense (Figure 3c). The classification adopted in this analysis avoids this situation.
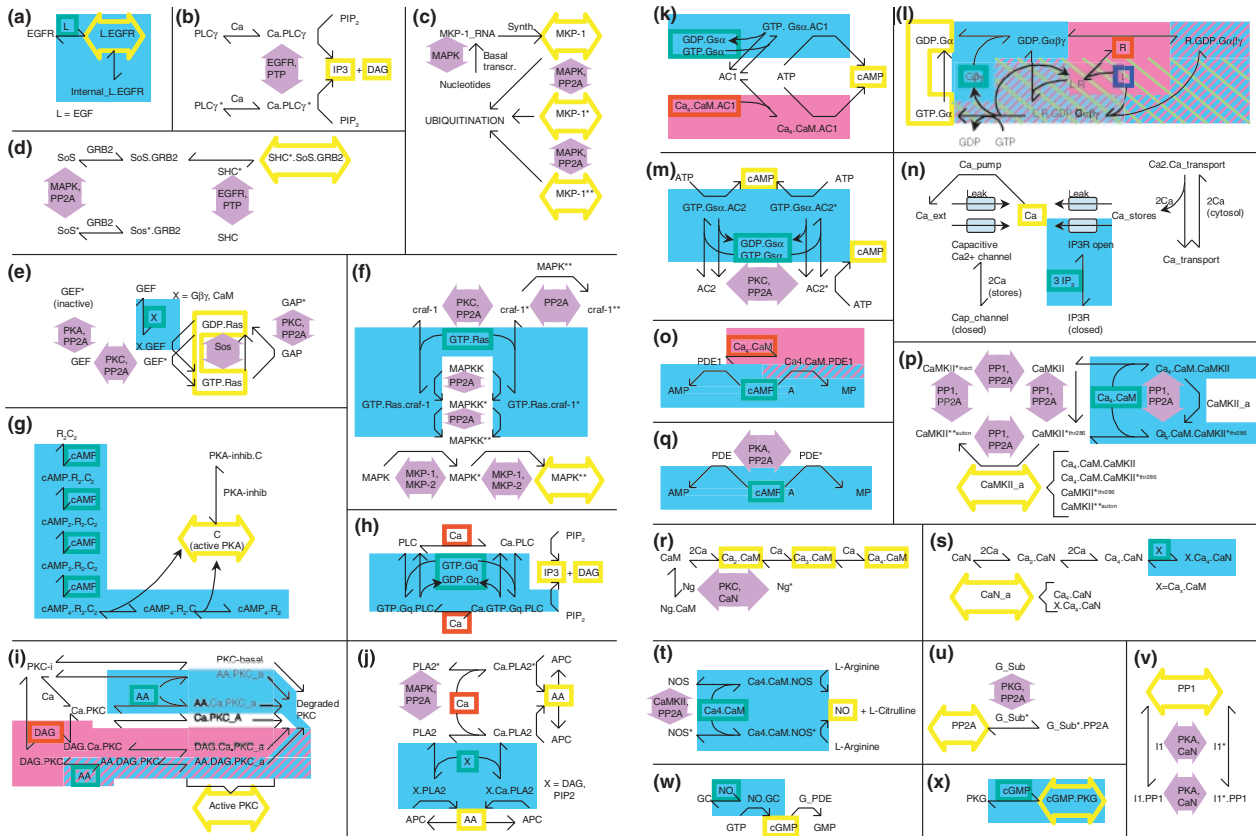
Fig. 4. aset of reaction mechanisms. Solid light gray blocks represent replicating reaction steps, with the input molecule(s) enclosed in a thick dark gray box. Additional replicating inputs occur in a few cases and are represented in other shades. Combinatorial replication of reactions is represented by slanted shaded lines. Pale gray boxes represent output molecules. Pale gray arrows represent output replicating reactions, typically Michaelis–Menten enzymes. Filled gray arrows represent inputs from replicating reactions. (a) Epidermal Growth Factor Receptor. A distinct complex for active enzyme is formed for each ligand, and the enzyme–substrate complex is unique for each substrate, giving a combinatorial number of reactions. (b) phospholipase C gamma. This has two output molecules, inositol trisphosphate (IP3) and diacylglycerol (DAG). (c) MAP kinase phosphatase-1 (MKP-1). MAPK enzyme activity is represented as stimulating transcription, and synthesis as a zeroth order reaction. There are three output enzymes for each of the phosphorylation states of MKP-1. (d) SoS. (e) Ras. (f) Mitogen activated protein kinase (MAPK) regulation. (g) Protein kinase A. (h) Phospholipase C-β. G-proteins communicate by the GTP— as well as GDP-bound alpha subunits, to accommodate GTPase-activating (GAP) activity of the target enzyme. The replicating input handles multiple isoforms of G proteins. (i) Protein kinase C (PKC). Ca, DAG, and AA are the inputs. DAG and AA are both replicating inputs. Active PKC is the sum of activities of six activated forms. (j) Phospholipase A2 (PLA2). It is activated by phosphorylation by MAPK, by Ca, and by replicating reactions involving inputs from DAG and PIP2. The output is AA. (k) AC1 activation by Gsα and calcium–calmodulin (Ca4.CaM). These are treated as independent replicating sets of reactions. (l) Ligand–Receptor–G-protein complex. This is in principle a 5th order complex involving combinations of ligand, receptor, and the three G-protein subunits. Here we consider only the receptor, ligand and G$\beta\gamma$ in combination. The ligand replicating interactions are represented by a dark shaded box and pale slanting lines over the background. As the reaction mechanism is general it can represent various Gα subunits, in this case Gs and Gq. (m) Adenylyl cyclase type 2. (n) Calcium regulation. (o) Calmodulin-dependent phosphodiesterase. (p) Calcium-calmodulin dependent protein kinase type II (CaMKII). The total kinase activity is the combination of four active forms of the kinase. (q) cAMP phosphodiesterase. (r) Calmodulin (CaM). (s) Calcineurin (CaN). The total activity is the sum of Ca4.CaN and its binding to various Ca-bound states of CaM. (t) Nitric oxide synthase. (u) Protein phosphatase 2A. (v) Protein phosphatase 1. (w) Soluble guanylyl cyclase. (x) Protein Kinase G.

Interaction statistics

The generality of the above conceptual framework was empirically supported by using it to represent 74 interactions between 25 pathways (Figures 1 and 4). To the extent that this list includes types of interactions seen in other signaling pathways, the current organizational framework can be applied to additional networks. The 74 interactions come from 29 output interaction points and lead into 54 inputs on the 25 pathways. On average, therefore, each pathway output connects to 2.5 other pathways. The strength

of interaction is not necessarily related to the number of interactions between modules. For example, the single-connection interaction from EGF to the EGFR represents a nanomolar-affinity interaction, whereas the multiple arrows from PP1 to CaMKII represent phosphatase action on multiple states of the enzyme, each with a K_m of 5 μM.

16 of the output types are single-molecule or messengers, and the remainder are mostly phosphatases or kinases. 26 of the inputs are single-molecule or messenger inputs, the remainder are mostly (de)phosphorylations. 6 pathways have combinatorial (greater than pairwise) replicating interactions so this is a significant fraction of the total. Ras is the only pathway in this dataset with more than 4 distinct inputs, but there is growing evidence that many other pathways may also be regulated in several ways (e.g. phospholipase C β is regulated by G$\beta\gamma$, PIP2 levels and phosphorylation in addition to Gqα (Singer *et al.*, 1997; Ryu *et al.*, 1990)).

Applicability of the classification beyond point mass-action kinetics

The current analysis is based on chemical reaction architecture rather than details of the kinetics involved. Thus it can be applied to situations other than point mass-action kinetics. Stochastic chemical kinetics utilize equivalent reaction schemes and therefore also fit in this framework (Gillespie, 1977). Genetic circuits can be represented as chemical networks, (Arkin *et al.*, 1998) and these published networks can also be described in terms of replicating interactions where at any given instant one of the possible DNA/promoter combinations is stochastically selected. Thus stochastic and genetic networks fit directly into the framework for signaling interactions as described in this paper. Spatial and structural aspects of signaling add further dimensions to the chemistry (Stiles and Bartol, 2000), but this analysis remains applicable to the definition of the underlying reaction systems.

DISCUSSION

At the conceptual level, this study suggests that all signaling interactions may be described by a single chemical motif wherein some pathways present a number of communicating molecules, and a further pathway includes sets of reactions and molecules which are replicated for every combination of the interacting pathways. This generalization is an unexpected outcome of a preliminary taxonomic classification of signaling interactions based on chemical motifs. Indeed, from the viewpoint of designing databases and interfaces, these results largely obviate the need for such a taxonomy. The number of interactions examined in this study is limited by the need for a complete chemical specification of interacting pathways, and a rather small number of examples have been described at this level

of detail. Nevertheless, many of the known classes of signaling pathways have been represented. The selected pathways are ubiquitous, and are likely to be substantially equivalent in many tissues and species. I performed a further analysis of other published chemical signaling models including the MAPK cascade (Huang and Ferrell, 1996; Levchenko *et al.*, 2000; Kholodenko, 2000), phototransduction (Lamb and Pugh, 1992; Blackwell, 2000), chemotaxis (Morton-Firth *et al.*, 1999; Shimizu *et al.*, 2000), genetic circuits (Arkin *et al.*, 1998). In addition, several variations on the signaling chemistry within the current dataset were also considered from a database of pathways at http://doqcs.ncbs.res.in. Together these data sources double the number of signaling pathways and their various implementations available for analysis, to approximately 50. In all cases inspection of the reaction schemes confirmed the applicability of the above procedure for classifying interactions.

This study provides an empirical definition of pathway blocks and interaction arrows that underlie most contemporary descriptions of signaling networks. The classification scheme discussed in this paper leads to finer blocks than commonly used to express signaling pathways, but tightly coupled sets of molecules such as receptor–ligand–G-protein complexes are lumped together. Despite this small difference in granularity, there is a satisfying degree of overlap between this chemically defined block diagram and the traditional description. This overlap is strengthened since higher-order modules are nearly identical to classical representations of pathways. A further conceptual point of the paper is the separation between reactions comprising the core pathway and interactions respectively. This subdivision distinguishes between the intrinsic properties of the pathway, and those that derive their identity from combinations of interacting pathways.

Comparison with enzyme nomenclature

An existing standard for classifying reactions is the nomenclature for enzymatic reactions. There are interesting parallels between this nomenclature and the current analysis of signaling chemistry. To recapitulate, an enzyme is defined primarily by the reactions it catalyzes (IUBMB Nomenclature Committee, 1992). A signaling 'interaction' is where one pathway presents a number of communicating molecules, and the other presents a replicating set of reactions involving those molecules. The first parallel between the two cases is that an enzyme is identified by the reactions it catalyzes, and an 'interaction' by the chemical steps it is involved in. The second similarity is that the basis for enzyme nomenclature is common chemical function, for example, transfer of a chemical group from one molecule to another. This is analogous to the notion of upstream replicating reactions,

since the same chemical operation may be carried out on many different substrates. The presence of these parallels is not surprising, since a large subset of interactions are indeed enzymatic reactions.

Differences between the classifications arise because enzyme classification focuses on chemistry, whereas signaling motifs involve information flow as well as chemistry. Three main differences arise: (1) pathway interactions need not involve any covalent modifications. Binding reactions are quite common; (2) enzymatic reactions are the defining feature of an enzyme, but in the case of this analysis the interactions are just one of several features of a pathway. Pathways can include several interactions and also a set of core reactions; and (3) a single, specific interaction can represent inter-conversions involving completely different enzymatic steps. For example, the interaction between PKC/PP2A (upstream) and cRaf (downstream) for phosphorylat-ing/dephosphorylating cRaf involves a phosphorylation step and ATP for one direction, and a dephosphorylation step catalyzed by a different enzyme for the reverse. Nevertheless, in terms of this analysis, the interaction has the same chemical architecture involving the phosphorylated and unphosphorylated cRaf.

Designing databases of signaling kinetics

This study provides an empirical framework for designing databases for the chemistry of signaling pathways. Current databases and XML extensions designed for signaling represent the chemistry primarily through listings of chemical steps and rates (Hucka *et al.*, 2000; Hedley *et al.*, 2001). The CellML specification (Hedley *et al.*, 2001http://www. Cellml.org) also allows for encapsulation of reactions and molecules. Interactions between pathways involve the definition of explicit reactions between the pathways. The current analysis suggests a more structured organization. First, each pathway specification should include the chemical reaction mechanisms of the core pathway as well as replicating reactions for the interactions. Further, it should include rate constants and concentrations for the core reactions, and a set of values as defaults for the interactions. The second part of the database would specify interaction parameters including kinetic constants and the identity of the interacting pathways. The reaction mechanisms for interactions would already be defined as part of the pathway definition. In principle the number of interactions could scale combinatorially with N pathways. This analysis, which involves a highly interconnected set of pathways, suggests that the actual number is much smaller and may possibly be more like 3 to 4 N. Indirect support for this comes from an analysis of the *S. cerevisiae* protein–protein interaction network, which suggests that about 93% of proteins interact with five or fewer other proteins (Jeong *et al.*, 2001).

Object-oriented specification of signaling pathways

There has hitherto been a divide between the block-diagram representation of signaling interactions, and simulators and databases that must explicitly represent each chemical reaction. Through this analysis it becomes possible to modularize signaling pathways in a manner reminiscent of object-oriented programming. First, the chemistry within a pathway does not depend on other pathways. This means that the details of the implementation are hidden or encapsulated. Second, interactions between pathways are conducted through well-defined interfaces consisting of the replicating interactions. Because the interaction interfaces are consistent for all pathways, they provide for data abstraction: all pathways can be regarded as based on the same class. Third, new pathways can be derived from existing ones to account for isoform or species specialization, or 'upgraded' with improved data. This corresponds to class derivation and inheritance of properties of existing pathways. Fourth, composite pathways can embed existing ones into larger modules. In object-oriented terms, these are nested classes. These properties facilitate the management of complex chemistry through the use of predefined modules. In particular, the generalization of all signaling interactions into a common interface in an object-oriented class means that any two pathways can be interconnected using the same high-level operation (such as click-and-drag) that is then specialized into specific reaction steps according to the class definition. From the user's point of view, complex signaling networks could be set up by connecting pathway blocks drawn from a database, without compromising the chemical details.

Three lines of current research are converging to facilitate more complete and more accurate descriptions of signaling networks. These are technical and experimental advances in obtaining mechanistic and kinetic details, development of databases of signaling pathways, and development of simulators for modeling signaling networks. This study draws upon the existing body of chemical-level descriptions of signaling pathways to inform design decisions for the development of signaling databases and simulators. The current analysis reveals a fortuitous symmetry and uniformity of organization of chemical reactions in signaling that is well suited to database and simulator construction and may simplify current designs. Although the level of this analysis is chemical, it provides an organizational tool for developing block-diagram descriptions that may contribute to higher-level analysis of pathways.

REFERENCES

Alberts,B., Bray,D., Lewis,J., Raff,M., Roberts,K. and Watson,J.D. (1994) *Molecular Biology of the Cell*, 3rd edition, Garland, New York.

Arkin,A., Ross,J. and McAdams,H.H. (1998) Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics*, **149**, 1633–1648.

Asthagiri,A.R. and Lauffenburger,D.A. (2000) Bioengineering models of cell signaling. *Annu. Rev. Biomed. Eng.*, **2**, 31–53.

Bhalla,U.S. (1998) The network within: signaling pathways. In Bower,J.M. and Beeman,D. (eds), *The Book of GENESIS: Exploring Realistic Neural Models with the GEneral NEural SImulation System*. Springer, New York, pp. 169–190.

Bhalla,U.S. and Iyengar,R. (1999) Emergent properties of networks of biological signaling pathways. *Science*, **283**, 381–387.

Blackwell,K.T. (2000) Evidence for a distinct light-induced calcium-dependent potassium current in *Hermissenda crassicornis*. *J. Comput. Neurosci.*, **9**, 149–170.

Bray,D. (1995) Protein molecules as computational elements in living cells. *Nature*, **376**, 307–312.

Brondello,J.-M., Pouyssegur,J. and McKenzie,F.R. (1999) Reduced MAP Kinase Phosphatase-1 degradation after p42/p44MAPK-dependent phosphorylation. *Science*, **286**, 2514–2517.

Gillespie,D.T. (1977) Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, **81**, 2340–2361.

Hedley,W.J., Nelson,M.R., Bullivant,D.P. and Nielsen,P.F. (2001) A short introduction to CellML. *Phil. Trans. R. Soc. London A*, **359**, 1073–1089.

Huang,C.Y.F. and Ferrell,J.E.J. (1996) Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc. Natl Acad. Sci. USA*, **93**, 10078–10083.

Hucka,M., Sauro,H., Finney,A., Bolouri,H., Doyle,J. and Kitano,H. The ERATO systems biology workbench: An integrated environment for multiscale and multitheoretic simulations in systems biology. In Kitano,H. (ed.), *First International Conference on Systems Biology*. MIT Press, Cambridge, MA, **2000**.

Ideker,T., Thorsson,V., Ranish,J.A., Christmas,R., Buhler,J., Eng,J.K., Bumgarner,R., Goodlett,D.R., Aebersold,R. and Hood,L. (2001) Integrated genomic and proteomic analysis of a systematically perturbed metabolic network. *Science*, **282**, 929–934.

International Union of Biochemistry and Molecular Biology, (1992) Nomenclature Committee. *Enzyme Nomenclature*, Academic Press, San Diego. Web version at http://www.chem.qmw.ac.uk/iubmb/enzyme/

Jeong,H., Mason,S.P., Barabási,A.-L. and Oltvai,Z.N. (2001) Lethality and centrality in protein networks. *Nature*, **411**, 41–42.

Kholodenko,B.N. (2000) Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. *Eur. J. Biochem.*, **267**, 1583–1588.

Kuroda,S., Schweighofer,N. and Kawato,M. (2001) Exploration of signal transduction pathways in cerebellar long-term depression by kinetic simulation. *J. Neurosci.*, **21**, 5693–5702.

Lamb,T.D. and Pugh,JrE.N. (1992) A quantitative account of the activation steps involved in phototransduction in amphibian photoreceptors. *J. Physiol. Lond.*, **449**, 719–758.

Lauffenburger,D.A. and Linderman,J.J. (1993) *Receptors: Models for Binding, Trafficking and Signaling*. Oxford University Press, Oxford.

Levchenko,A., Bruck,J. and Sternberg,P.W. (2000) Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties. *Proc. Natl Acad. Sci. USA*, **97**, 5818–5823.

Morton-Firth,C.J., Shimizu,T.S. and Bray,D. (1999) A free-energy-based stochastic simulation of the Tar receptor complex. *J. Mol. Biol.*, **286**, 1059–1074.

Pieroni,J.P., Jacobowitz,O., Chen,J. and Iyengar,R. (1993) Signal recognition and integration by G_s-stimulated adenylyl cyclases. *Curr. Op. Neurobiol.*, **3**, 345–351.

Ryu,S.H., Kim,U.-H., Wahl,M.I., Brown,A.B., Carpenter,G., Huang,K.-P. and Rhee,S.G. (1990) Feedback regulation of phosphlipase C-beta by protein kinase C.. *J. Biol. Chem.*, **265**, 17941–17945.

Schaff,J., Fink,C.C., Slepchenko,B., Carson,J.H. and Loew,L.M. (1997) A general computational framework for modeling cellular structure and function. *Biophys. J.*, **73**, 1135–1146.

Shimizu,T.S., Le Novere,N., Levin,M.D., Beavil,A.J., Sutton,B.J. and Bray,D. (2000) Molecular model of a lattice of signalling proteins involved in bacterial chemotaxis. *Nat. Cell. Biol.*, **2**, E199–201.

Singer,W.D., Brown,H.A. and Sternweis,P.C. (1997) Regulation of eukaryotic phosphatidylinositol-specific phospholipase C and phospholipase D. *Annu. Rev. Biochem.*, **66**, 475–509.

Stiles,J.R. and Bartol,T.M. (2000) Monte Carlo methods for simulating realistic synaptic microphysiology using MCell. In De Schutter,E. (ed.), *Computational Neuroscience: Realistic Modeling for Experimentalists*. CRC Press, Boca Raton, pp. 87–128.

Tomita,M., Hashimoto,K., Takahashi,K., Shimizu,T.S., Matsuzaki,Y., Miyoshi,F., Saito,K., Tanida,S., Yugi,K., Venter,J.C. and Hutchison,C.A. 3rd. (1999) E-CELL: software environment for whole-cell simulation. *Bioinformatics*, **15**, 72–84.

Tucker,C.L., Gera,J.F. and Uetz,P. (2001) Towards an understanding of complex protein networks. *Trends Cell Biol.*, **11**, 102–106.