molecular
systems
biology

REVIEW

Depicting combinatorial complexity with the molecular interaction map notation

Kurt W Kohn*, Mirit I Aladjem, Sohyoung Kim, John N Weinstein and Yves Pommier

Laboratory of Molecular Pharmacology, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, MD, USA

* Corresponding author. Laboratory of Molecular Pharmacology, Center for Cancer Research, National Cancer Institute, Building 37, Room 5068, NIH, Bethesda, MD 20892, USA. Tel.: + 1 301 496 2769; Fax: + 1 301 402 0752; E-mail: kohnk@dc37a.nci.nih.gov

To help us understand how bioregulatory networks operate, we need a standard notation for diagrams analogous to electronic circuit diagrams. Such diagrams must surmount the difficulties posed by complex patterns of protein modifications and multiprotein complexes. To meet that challenge, we have designed the molecular interaction map (MIM) notation (http://discover.nci.nih.gov/mim/). Here we show the advantages of the MIM notation for three important types of diagrams: (1) explicit diagrams that define specific pathway models for computer simulation; (2) heuristic maps that organize the available information about molecular interactions and encompass the possible processes or pathways; and (3) diagrams of combinatorially complex models. We focus on signaling from the epidermal growth factor receptor family (EGFR, ErbB), a network that reflects the major challenges of representing in a compact manner the combinatorial complexity of multimolecular complexes. By comparing MIMs with other diagrams of this network that have recently been published, we show the utility of the MIM notation. These comparisons may help cell and systems biologists adopt a graphical language that is unambiguous and generally understood.
Molecular Systems Biology 3 October 2006;
doi:10.1038/msb4100088
Subject Categories: bioinformatics; metabolic and regulatory networks
Keywords: EGFR; epidermal growth factor receptor; molecular interaction maps; network diagrams; signaling networks

Introduction

A standard notation for biomolecular interaction networks is urgently needed for three main purposes: (1) to define explicit models for computer simulation; (2) to organize available information about a network's molecular interactions; and (3) to diagram combinatorially complex processes. Although several diagram notations have been proposed, it is important to reach a consensus, so that diagrams can be widely understood, as is the case for electronic circuit diagrams.

Two of the best developed notations are the molecular interaction maps (MIMs) that we have described (Kohn, 1999, 2001; Kohn *et al*, 2006) and the 'process diagrams' described by Kitano *et al* (Kitano, 2003; Kitano *et al*, 2005). We recently discussed the strengths and weaknesses of the various notations that have been proposed (Kohn *et al*, 2006). These include, in addition to the MIM and process diagram notations, the computer-aided design (CAD)-like diagrams produced by CellDesigner (Funahashi *et al*, 2003), a software suite called CADLIVE (Kurata *et al*, 2003), the automated diagrams of Cook *et al* (2001), and BIOCARTA's connection diagrams (http://www.biocarta.com).

Here, we compare the MIM and process diagram notations in more detail and consider where each may be advantageous. We have previously demonstrated the utility of the MIM notation for computer simulation (Kohn, 1998, 2001; Kohn *et al*, 2004) and for organizing information (Kohn, 1999, 2001; Kohn and Bohr, 2002; Kohn *et al*, 2003, 2006; Pommier and Kohn, 2003; Pommier *et al*, 2004, 2006; Aladjem *et al*, 2004; Kohn and Pommier, 2005). We show here that the MIM notation is suitable as a standard for both of these purposes, and also for representation of complex combinatorial schemes.

Process diagrams show reactions in a manner that is direct and intuitive, requiring little or no description in accompanying text. MIM diagrams are also self-explanatory when one is familiar with the notation. A detailed description of the MIM notation with many examples was recently published and could serve as a reference and tutorial (Kohn *et al*, 2006).

To compare the graphic notations, we present MIM versions of recently published process diagrams of signaling from ErbB receptors (Kitano *et al*, 2005; Oda *et al*, 2005) and discuss their respective characteristics and advantages. This comparison shows advantages and flexibility of the MIM notation that may justify learning its nuances. It also illustrates how MIM diagrams can represent signaling from multiple receptor homo- and heterodimers, as well as the combinatorial complexity of a network.

We also clarify what we previously described as a distinction between 'explicit' and 'heuristic' MIMs (Kohn, 2001; Kohn *et al*, 2006). Rather than representing different *types* of diagrams, our current view, which we explain herein, is that they are alternative *interpretations* of the notation. The way in which an MIM is to be interpreted depends on the intended application, and must be specified.

For readers' convenience, we show the list of MIM symbols in Figure 1 and the rules of the MIM notation in Box 1.

Three interpretations of MIMs: explicit, heuristic, and combinatorial

The MIM notation allows three interpretations, each suited to a different purpose. The examples in Figure 2 explain the

distinctions between the 'explicit', 'heuristic', and 'combinatorial' interpretations.

The 'explicit' interpretation is that an interaction line applies only to the molecular species directly connected to it (Figure 2). This type of MIM defines the reaction paths for a particular model, explicitly depicting every reaction. In this way, it is like the process diagrams of Kitano and co-workers (Kitano, 2003; Kitano *et al*, 2005). The reactions shown in either of those two types of diagram can be translated into input for computer

Reaction symbols

⟷	Non-covalent binding (reversible)
═══	Covalent bond
━━▶	Stoichiometric conversion
━━○	Enzymatic catalysis
┌─▷	Transcription
─╱─	Cleavage of a covalent bond.
━━▶⊘	Degradation

Contingency symbols

━━▷	Stimulation
━━┤▷	Requirement
▌	Inhibition

Figure 1 MIM symbols used in this paper. For a complete list of symbol, see Kohn *et al* (2006) or http://discover.nci.nih.gov/mim/.

simulation (Figure 3 and Table I; further explanation is given in the next section).

In 'heuristic' and 'combinatorial' interpretations, on the other hand, an interaction line represents a functional connection between domains or sites that (unless otherwise indicated) is independent of the modification or binding states of the directly interacting species (Figure 2). Therefore, an interaction line in heuristic or combinatorial interpretation may define a large class of interactions, such as defined by Blinov *et al* (Blinov *et al*, 2005; Faeder *et al*, 2005).

The 'heuristic' interpretation serves as a compact information organizer, showing the possible reaction paths. It depicts what is known and reveals what still remains to be determined, thereby 'helping to discover or learn' (a meaning of 'heuristic' given in Webster's Unabridged Dictionary, 2nd edition, 1979). The influences of indirect interactions, so far as they are known, can be shown by means of contingency symbols (Figure 1).

The 'combinatorial' interpretation is that all of the possible interactions do in fact occur, subject only to any restrictions indicated by contingency symbols (Figure 2D). The combinatorial interpretation shows implicitly the large number of reaction paths that can take place concurrently in the actual expression of a network. This corresponds closely to the 'reaction class' or 'rule-based' convention described by B Goldstein, WS Hlavacek and co-workers (Blinov *et al*, 2005; Faeder *et al*, 2005; ML Blinov, personal communication; further explanation in a later section below). The combinatorial interpretation of MIMs and the 'rule-based' description of combinatorial networks both define large numbers of concurrent reaction paths for computer simulation. Interactions in combinatorially interpreted MIMs, like 'rules', can in principle serve as generators of reaction events and molecular species (Blinov, personal communication).
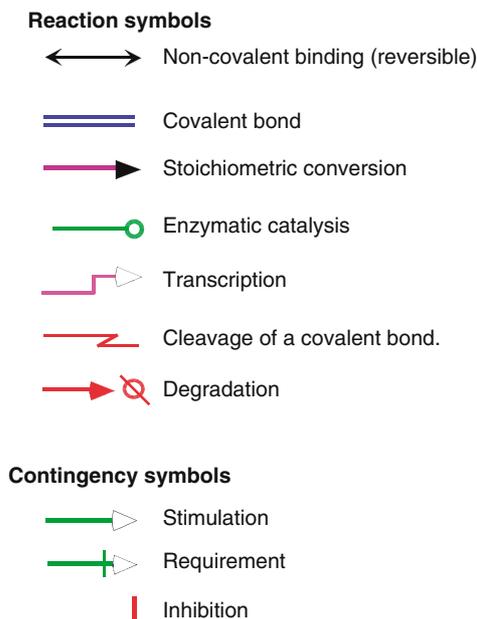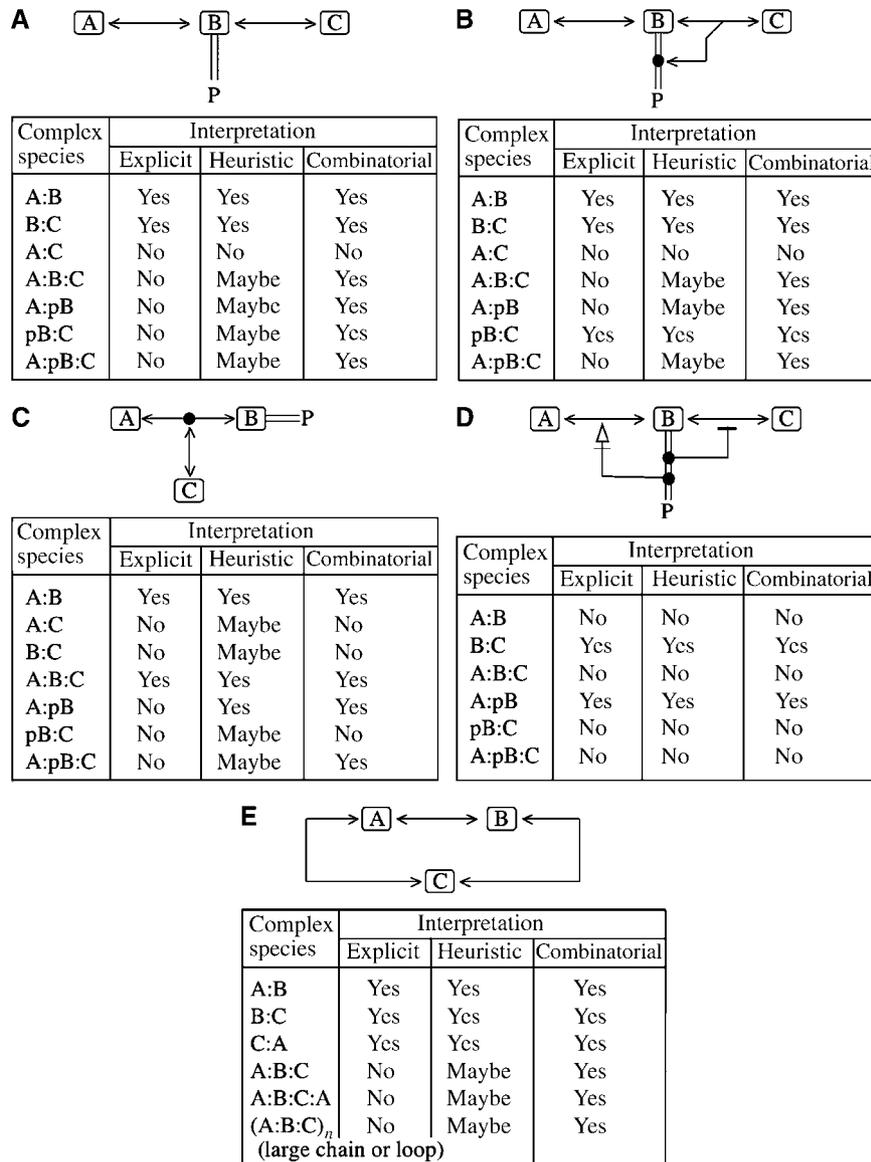
Box 1 Rules and definitions of the MIM notation

1. A named molecular species generally appears in only one place on a map. (Exempt from this rule are molecules, such as GTP or ubiquitin, that act in a similar manner in a large number of different reactions. For clarity, the named species and its interactions must sometimes be duplicated upon translocation from one cell compartment to another.)
2. Interactions between molecular species are shown by different types of connecting lines, distinguished by different arrowheads or other terminal symbols (Figure 1).
3. Interaction lines can change direction (but not by more than 90° at a corner—this restriction prevents ambiguities at branch points).
4. When lines cross, it is as if they do not touch.
5. Symbol definitions are not affected by color. Color is optional: it can be used as an independent visual parameter to guide the eye and/or emphasize particular features of the network. We use red for inhibitions and other negative actions; the net effect of a sequence of interactions (whether positive or negative) can then be determined by whether the number of red-colored steps is even or odd. We use green for stimulatory or catalytic actions, blue for covalent modifications, and purple for transcription/translation.
6. A small filled circle ('node') on an interaction line indicates the consequence or product of the interaction. Thus, the consequence of binding between two molecules is production of a dimer, which is represented by a node on the binding interaction line. The consequence of a modification (e.g., phosphorylation) is production of the modified (e.g., phosphorylated) molecule; the phosphorylated product is represented by a node placed on the modification line.
7. Multiple nodes on an interaction line represent exactly the same molecular species. To avoid ambiguity, a node should not be placed at a line crossing.
8. An isolated node (a node that is not on a line) is an abbreviation that represents another copy of the same molecular species that is defined at the other end of the line pointing to the node (to avoid ambiguity, only one arrow should point to an isolated node).
9. Molecular interactions are of two types, reactions and contingencies, as listed in Figure 1. Reactions operate on molecular species; contingencies operate on reactions or on other contingencies.
10. A line without arrowheads is a 'state-combination' symbol. A node on this line represents the combination of states defined by the symbols at the two ends of the line. For example, in the upper left of Figure 3, there is an arrowless line connecting a node representing the EGF:EGFR complex with a node representing phosphorylated-EGFR; the node within this line represents the EGF:phosphorylated-EGFR complex. The dimer of this complex is designated species 5. (Note that for convenience, there are two nodes on the dimerization line, both of which refer to species 5. Also note that in the text, we use a colon to indicate binding.)
11. MIMs may be interpreted as 'explicit', 'heuristic', or 'combinatorial.'

For a more detailed description with examples, see Kohn *et al* (2006) or http://discover.nci.nih.gov/mim/.

A

Complex species	Interpretation		
	Explicit	Heuristic	Combinatorial
A:B	Yes	Yes	Yes
B:C	Yes	Yes	Yes
A:C	No	No	No
A:B:C	No	Maybe	Yes
A:pB	No	Maybe	Yes
pB:C	No	Maybe	Yes
A:pB:C	No	Maybe	Yes

B

Complex species	Interpretation		
	Explicit	Heuristic	Combinatorial
A:B	Yes	Yes	Yes
B:C	Yes	Yes	Yes
A:C	No	No	No
A:B:C	No	Maybe	Yes
A:pB	No	Maybe	Yes
pB:C	Yes	Yes	Yes
A:pB:C	No	Maybe	Yes

C

Complex species	Interpretation		
	Explicit	Heuristic	Combinatorial
A:B	Yes	Yes	Yes
A:C	No	Maybe	No
B:C	No	Maybe	No
A:B:C	Yes	Yes	Yes
A:pB	No	Yes	Yes
pB:C	No	Maybe	No
A:pB:C	No	Maybe	Yes

D

Complex species	Interpretation		
	Explicit	Heuristic	Combinatorial
A:B	No	No	No
B:C	Yes	Yes	Yes
A:B:C	No	No	No
A:pB	Yes	Yes	Yes
pB:C	No	No	No
A:pB:C	No	No	No

E

| Complex species	Interpretation		
	Explicit	Heuristic	Combinatorial
A:B	Yes	Yes	Yes
B:C	Yes	Yes	Yes
C:A	Yes	Yes	Yes
A:B:C	No	Maybe	Yes
A:B:C:A	No	Maybe	Yes
$(A:B:C)_n$ (large chain or loop)	No	Maybe	Yes

Figure 2 Examples to illustrate the 'explicit', 'heuristic', and 'combinatorial' interpretations of MIMs. For any MIM, the interpretation that applies to it must be stated. In the 'heuristic' column, some molecular species are marked 'maybe', which means that either it is not known whether those species form or that further information is provided in the text annotations. The meaning of each panel is defined in the table below each diagram. The tables list the molecular species that are included, excluded, or left indeterminate by each interpretation.

For any MIM, one must state whether it should be interpreted as explicit, heuristic, or combinatorial. In order to make the distinctions clear, we next refer to the examples shown in Figure 2.

In explicit interpretation, an interaction involves only the species that are connected directly to a particular interaction line, subject to any contingencies impacting on that line. For example, in Figure 2D, a binding interaction line connects species **A** and **B**. But there is a contingency on this line, specifying that binding requires phosphorylation of **B**. Therefore, **A:pB** is permitted, but not **A:B** (**pB**=phosphorylated **B**). A binding line connects **B** and **C**, with contingency that phosphorylation of **B** inhibits. Therefore, **B:C** is permitted, but not **pB:C**.

Combinatorial interpretation includes interactions regardless of the states of binding or modification of the directly interacting species. For example, in Figure 2A, where explicit interpretation allows only **A:B** and **B:C**, combinatorial interpretation allows **A** bound to **B**, regardless of whether **B** is phosphorylated and/or bound to **C** (we call this property 'transitive', because the interaction symbol applies indirectly to species 'down the line'). The indirect interactions however may affect the reaction rate constants quantitatively; such indirect quantitative effects can be explained in text annotations or in a reaction class table, such as Tables 1 and 2 of Blinov *et al* (2006a).

Heuristic interpretation is definitive for those interactions that would be allowed by explicit interpretation, but is non-
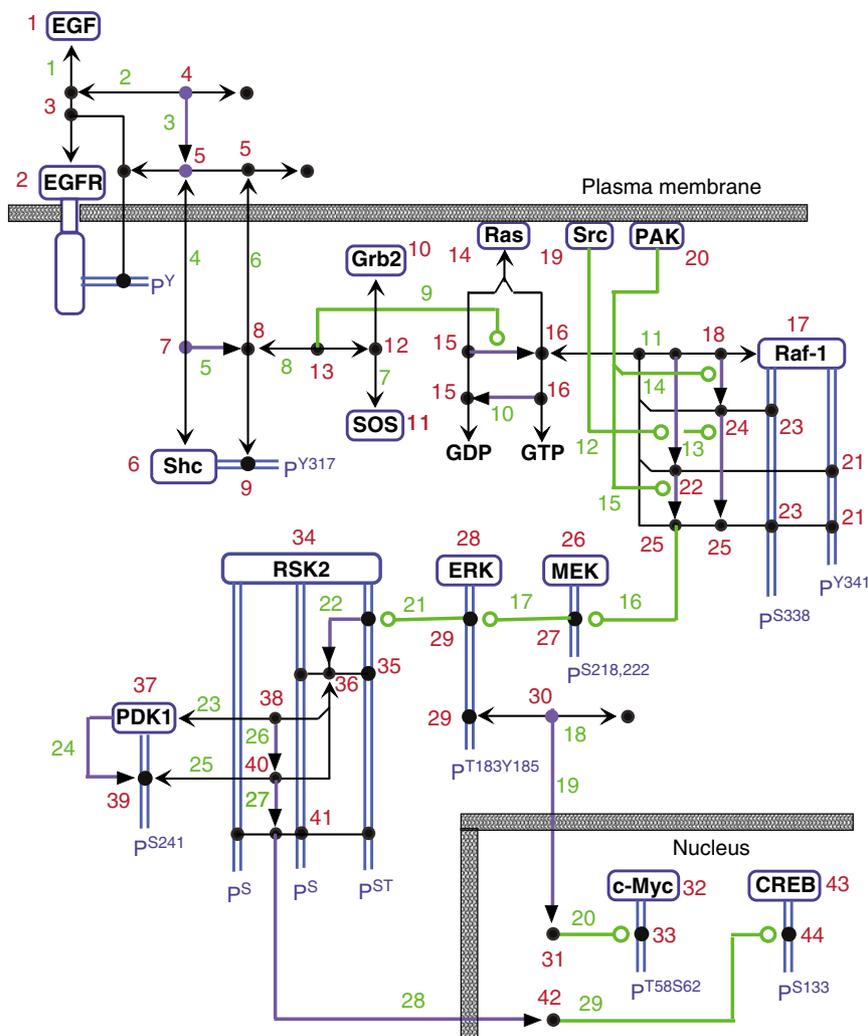
Figure 3 Explicit MIM of signaling from the epidermal growth factor receptor (EGFR, ErbB1). With few minor exceptions, this diagram shows exactly the same reactions as in the process diagram in Figure 1b of Kitano *et al* (2005). Molecular species are numbered in red and reactions are numbered in green italics. A detailed description of the MIM can be found in Supplementary information or at http://discover.nci.nih.gov. The meaning of each reaction can also be seen from the connection table (Table I). (Species numbers 14, 21, and 23 in the MIM do not appear in Table I, because they serve only to help with the explanation in Supplementary information. Species 14 is part of species 15 and 16; 21 is part of 22 and 25; and 23 is part of 24 and 25.) Modified with permission from Kohn and Aladjem (2006).

committal for the indirect interactions allowed by the combinatorial interpretation. In Figure 2, these indirect interactions are assigned the value 'maybe' in the heuristic column: each of the combinatorial possibilities may or may not occur, either because of lack of knowledge, or because contingency symbols have been omitted to avoid excessive crowding of the diagram. These uncertainties may then be clarified in text annotations.

It may be useful to note some additional points from the examples in Figure 2. Figure 2B shows the case in which **C** can bind **B** or **pB** explicitly. Figure 2C asserts that **A** and **B** can bind to form **A:B** and that **A:B** (indicated by the node on the interaction line) can bind **C**. Direct binding of **C** to **A** or **B** is excluded in the explicit or combinatorial interpretations. However, these bindings are not excluded in the heuristic interpretation, because the binding site for **C** might be on **A** or **B** (or both). Figure 2D illustrates how known contingencies can be indicated by means of symbols for stimulation, requirement, or inhibition.

Figure 2E shows the interesting case of a cycle of binding interactions. The explicit interpretation is clear. The heuristic and combinatorial interpretations however include the possibility that a cycle can begin and end at different copies of the same molecular species. Thus, they include linear or cyclic multimers of the form …A:B:C:A:B:C. Molecular rings or chains of this kind have also been considered in 'rule-based' iterations (ML Blinov, personal communication). Such multimer structures may be what gives rise to the discrete bodies or foci commonly seen in cell nuclei.

An explicit MIM, like a process diagram, defines a model for computer simulation: signaling from the EGF receptor, ErbB1

Process diagrams represent network models that show every reaction explicitly and that can in principle be simulated (Kitano, 2003; Kitano *et al*, 2005); this can be performed also

Table I Connection table of the reactions in the explicit MIM shown in Figure 3

Rxn	Reactants		Products		Rxn	Reactants		Products	
1a	1	2	3		16a	25	26	26a	
1b	3		1	2	16b	26a		25	26
2a	3	3	4		16c	26a		25	27
2b	4		3	3	17a	27	28	28a	
3	4		5		17b	28a		27	28
4a	5	6	7		17c	28a		27	29
4b	7		5	6	18a	29	29	30	
5	7		8		18b	30		29	29
6a	5	9	8		19	30		31	
6b	8		5	9	20a	31	32	32a	
7a	10	11	12		20b	32a		31	32
7b	12		10	11	20c	32a		31	33
8a	8	12	13		21a	29	34	34a	
8b	13		8	12	21b	34a		29	34
9a	13	15	15a		21c	34a		29	35
9b	15a		13	15	22	35		36	
9c	15a		13	16	23a	36	37	38	
10	16		15		23b	38		36	37
11a	16	17	18		24	37		39	
11b	18		16	17	25a	36	39	40	
12a	18	19	18a		25b	40		36	39
12b	18a		18	19	26	38		40	
12c	18a		19	22	27	40		41	
13a	19	24	24a		28	41		42	
13b	24a		19	24	29a	42	43	43a	
13c	24a		19	25	29b	43a		42	43
14a	18	20	18b		29c	43a		42	44
14b	18b		18	20					
14c	18b		20	24					
15a	20	22	22a						
15b	22a		20	22					
15c	22a		20	25					

The numbers refer to the reaction and species identification numbers in Figure 3. For reversible binding, the letter 'a' or 'b' is appended to refer to association and dissociation, respectively. For enzyme reactions, suffixes 'a' and 'b' refers to production and dissociation of the enzyme–substrate complex, respectively; suffix 'c' refers to conversion of the enzyme–substrate complex to products. A letter suffix added to a reactant or product species refers to the enzyme–substrate complex (whose existence is implied by the enzyme reaction symbol in the MIM).

Table II Reaction classes defined by Blinov *et al* (Blinov, 2006 #2766) and shown or implied in the combinatorial MIM (Figure 5)

Step	Reaction class	Implied
1	Ligand–receptor binding	
2	Receptor dimerization	
3	Receptor tyrosine phosphorylation	
4	Receptor tyrosine dephosphorylation	
5	Binding of PLCγ to Y992	
6	Transphosphorylation of PLCγ	
7	Binding of PLCγP to Y992	5(6)
8	Dephosphorylation PLCγP	
9	Binding of Grb2 to pY1068	
10	Binding of Sos to Grb2-pY1068	12(9)
11	Binding of Sos-Grb2 to pY1068	9(12)
12	Binding of Sos to Grb2	
13	Binding of Shc to pY1148/1173	
14	Phosphorylation of Shc	(13)
15	Binding of ShcP to pY1148/1173	13(14)
16	Dephosphorylation of ShcP	
17	Binding of Grb2 to ShcP-pY1148/1173	21(15)
18	Binding of Grb2-ShcP to pY1148/1173	13(21)
19	Binding of Sos to Grb2-ShcP-pY1148/1173	12(17 or 18)
20	Binding of Sos-Grb2-ShcP to pY1148/1173	15(22 or 23)
21	Binding of Grb2 to ShcP	
22	Binding of Sos to Grb2-ShcP	12(21)
23	Binding of Sos-Grb2 to ShcP	21(12)
24	Binding of Sos-Grb2 to ShcP-pY1148/1173	21(12 and 15)
25	Inactivation of PLCγP	

The steps that do not have an entry in the 'Implied' column in the table appear as direct interactions in the MIM in Figure 5. The meaning of the entries in the 'Implied' column is as follows: the first number (which is not enclosed in parentheses) indicates the step that defines the direct binary interaction; the numbers in parentheses indicate the steps that must precede the direct interaction step; these interactions are implied in the combinatorial MIM (Figure 5) without having to be shown explicitly. For example, step 22 involves the binding of SOS to Grb2, where Grb2 has already bound ShcP. We denote this reaction symbolically as 12(21), because step 12 is SOS binding to Grb2 and step 21 is Grb2 binding ShcP. Similarly, step 23 (binding of ShcP to Grb2, where Grb2 has already bound SOS) is denoted by 21(12). The advantage of this symbolic notation is that it indicates the order of events and also tells us which molecules interact directly. In a further example, step 20 refers to a binding of ShcP to ErbB1 phosphotyrosine-1148/1173, where the ShcP already exists in the ternary complex, ShcP–Grb2–SOS. This ternary complex however can form in two ways: by way of step 22 or 23. Therefore, we write 15(22 OR 23), meaning that the direct binary step is 15, and the pre-existing reactions are step 22 or 23.

with the explicit form of MIMs (Kohn, 1998, 2001; Kohn *et al*, 2004). In order to compare the two diagram notations directly, we discuss here an explicit MIM version (Figure 3) of a process diagram of EGF receptor signaling from Figure 1b in Kitano *et al* (2005). We show how the MIM defines the topology (Table I) of the network and defines a set of differential equations (Supplementary Table 1) that could be used for simulation.

To show that the explicit MIM in Figure 3 is an unambiguous description of the network model's topology, we express its component reactions in a connection table (Table I) that contains all the information that a computer program needs (other than rate constant values and initial conditions) to simulate the network. Table I is made up of the reaction and species numbers assigned to the symbols in Figure 3. This representation is suitable for 'micro-world models', which consist solely of mass action terms (Kholodenko and Westerhoff, 1995; Kohn, 2001). Micro-world models have no stimulation or inhibition terms and no Michaelis–Menten terms. Thus, everything is modeled as direct molecular events: binding, dissociation, or stoichiometric conversion/translocation. We have used this modeling procedure in two published computational studies (Kohn, 1998; Kohn *et al*, 2004).

The MIM notation compresses the association and dissociation reactions of reversible binding into a single symbol (a double-arrowed line). The two reactions are represented in the connection table (Table I) by the interaction number shown in Figure 3, followed by an 'a' or 'b' suffix, respectively.

The notation compresses enzyme action into a single symbol that represents three component reactions: (a) binding between enzyme and substrate; (b) dissociation of the enzyme:substrate complex; and (c) conversion of the enzyme:substrate complex to products. This manner of representing enzyme actions in three component reactions has two advantages: First, it makes the connection table homogeneous in that all reactions are simple mass action terms. Second, it avoids the assumption of a quasi-steady-state inherent in Michaelis–Menten expressions.

The three component reactions of an enzyme action in Table I are labeled with suffixes 'a', 'b', and 'c' placed after the number assigned to the enzyme action symbol in Figure 3. For example, the three component reactions of the enzyme action '*9*' in Figure 3 are labeled 9a for enzyme:substrate binding, 9b for enzyme:substrate dissociation, and 9c for conversion to

products. Enzyme–substrate complexes are not shown explicitly in Figure 3, but are included in the connection table, where they are assigned the species number of a reactant followed by a letter (thus, the enzyme–substrate complex for enzyme action '9' in Figure 3 is marked '15a' in Table I).

The connection table (Table I) defines a set of differential equations, listed in Supplementary Table 1. This example illustrates how an explicit MIM defines a set of ordinary differential equations suitable as a basis for simulation. Actually carrying out such a simulation study however still requires choice of rate constant parameters and/or exploration of parameter space for parameter sets that confer plausible behavior (Kohn *et al*, 2004). The rate constant selections must be thermodynamically consistent to avoid violations of the second law when the network contains closed loops or when two paths lead co-energetically from one point to another. It will be useful to develop facilities to translate the graphical interactions into a textual form, or systems biology markup language (SBML), that can generate a stoichiometry matrix to assure consistency. In the network depicted in Figure 3, however, there are no thermodynamic problems, because there are no closed loops or parallel paths (other than the two paths for production of doubly phosphorylated Raf-1, which involve ATP hydrolysis and therefore are energetically independent).

One may question whether microworld models that assume mass action behavior in a homogeneous system can adequately represent processes occurring within the grossly inhomogeneous structure of the cell. Moreover, rate constants determined in chemical systems may differ greatly from those existing in the cell, where molecular crowding can markedly affect activity coefficients (Ellis, 2001; Minton, 2001; Hancock, 2004). Molecular crowding also enhances protein–protein binding interactions and may contribute to the formation of various types of nuclear bodies (Hancock, 2004), functionally integrated chromatin-associated foci (Pilch *et al*, 2003; Au and Henderson, 2005), and clusters of membrane-associated proteins on membrane rafts (Cary and Cooper, 2000; Parton and Hancock, 2004; Rajendran and Simons, 2005). Molecular crowding however may be uneven in the cell, and micro-world models may yield useful approximations if most of the reactions take place in relatively uncrowded regions. These models have particular clarity, and it may be premature to give up on them. Even so, the integrated behavior of multi-molecular systems such as those that control transcription or translation may require statistical mechanical expressions, such as developed by Shea and Ackers (Shea and Ackers, 1985; Wolf and Eeckman, 1998). Simulation of such structures may require additional facilities, such as those provided by SBML (Finney and Hucka, 2003; Hucka *et al*, 2003; Machne *et al*, 2006). The issues involved in the simulation of cell-signaling dynamics were thoroughly reviewed recently by Kholodenko (2006).

Heuristic MIMs and process diagrams as information organizers

Comprehensive process diagrams—such as signaling from EGF receptors—are too complicated for meaningful simulation at this time; they can however serve to organize large amounts of information about molecular interactions (Oda *et al*, 2005). Heuristic MIMs are also effective information organizers, but in a different way (Kohn, 1999, 2001; Kohn *et al*, 2006). Process diagrams are equivalent to explicit MIMs, as discussed above. As information organizers, however, heuristic MIMs have the advantage of 'transitivity' (as already explained in Figure 2 and associated text). Process diagrams specify particular reaction sequences or pathways, show all of the direct reactions, and include symbols for each and every reactant and product. Heuristic MIMs, on the other hand, focus on the interactions between *sites*, independent of other binding or modification states of the directly interacting molecules. Therefore, heuristic MIMs include by implication many possible reaction sequences occurring simultaneously, whereas process diagrams depict a narrow subset of the possible reactions. We will compare these two types of information-organizing diagrams directly in Figures 4 and 5. First, however, we will point out the main differences.

In contrast to process diagrams or explicit MIMs, an interaction between molecular species or domains in heuristic MIMs may apply regardless of the binding or modification states of the directly interacting molecules. Because a binding or modification site often cannot 'see' what is happening in other sites or domains of the same molecule, the heuristic MIM interpretation assumes that a direct interaction between sites or domains may occur regardless of bindings or modifications that may exist at other sites of the directly interacting molecules. As already explained in Figure 2 and associated text, heuristic and combinatorial MIM interpretations differ only in that the heuristic interpretation is non-committal with respect to the possible indirect interactions, whereas the combinatorial interpretation asserts that all of them do occur. When binding or modification states are known to affect each other—by stimulation or inhibition—this is indicated by means of contingency symbols (Figures 1 and 2; for further examples, see Kohn *et al*, 2006).

Another property of heuristic or combinatorial MIMs is that they can be 'canonical' ('generic' may be a better term), meaning independent of cell type or cell state, and inclusive of multiple event sequences occurring in parallel. These MIMs show the interactions that *can* occur if the potentially interacting molecules are in the same place at the same time: they show what each domain or site can 'see'. (For an example of the generic property and how alternative pathways can be shown on the same MIM by highlighting, see Figure 14 of Kohn *et al*, 2006.) An MIM can be made specific to a particular cell type or cell state by deleting the molecules that are not expressed and the interactions that do not occur owing to lack of colocalization in time or place.

The process diagram notation defines a variety of symbols for different types of elementary state nodes. In MIM diagrams, it is not necessary to specify so many different symbols, because the nature of a molecular species is adequately defined by the interactions in which it engages. This is possible in MIM diagrams because all of the interactions of a given molecular species connect to the same symbol.

Process diagrams use a special symbol to indicate the activated state of a molecular species. The MIM notation does not explicitly indicate activation, because a given molecular
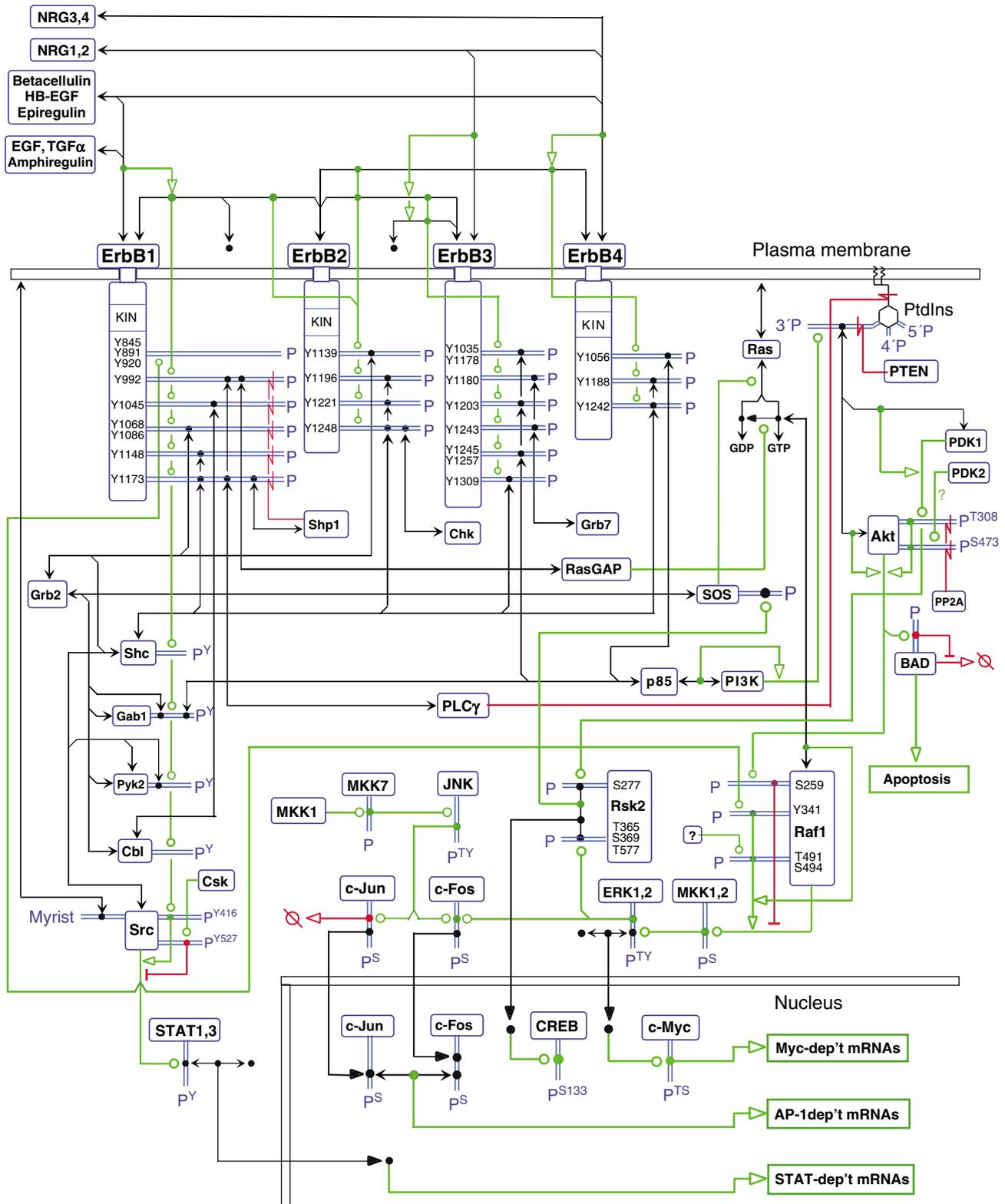
Figure 4 Heuristic MIM of part of the network of signaling from EGF receptors, which is presented in process diagram notation in Figure 1 of Oda *et al* (2005) (ErbB1=EGFR). This MIM displays an integral core of the network. It shows in particular how interactions of multiple receptor family members can be denoted.

species may be active with respect to one action, while being inactive with respect to another. MIMs thus rely on the interaction patterns themselves to define activity state.

Kitano *et al* (2005) presented a graph-theoretic description of their process diagrams. Explicit MIMs can be described in a similar way. Our 'molecular species' correspond to their 'state

A
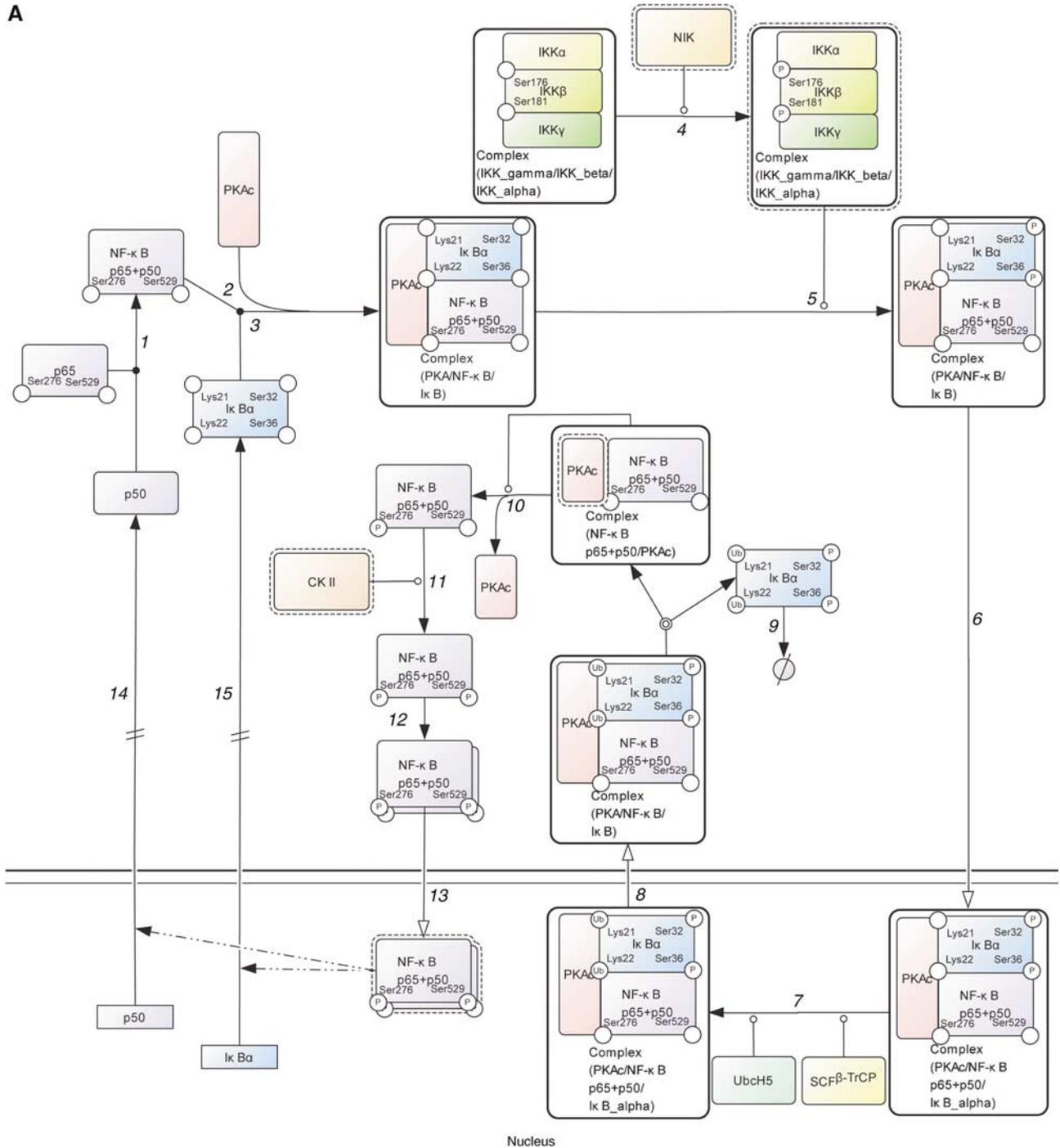
Figure 5 A heuristic MIM version of a process diagram of NFκB-related interactions. The process diagram was presented as Supplement Figure 3 in Kitano *et al* (2005). (**A**) Process diagram reproduced with numbers added corresponding to identification numbers of the interactions in panel B. (**B**) Heuristic MIM version.

nodes'; our 'reactions' correspond to their 'transition nodes'; our 'complex species' (represented as a filled circle on an interaction line) correspond to their 'complex state nodes'. Edges would be defined in the same way for both notations (Aguda and Sauro, 2004). The full description of a reaction in both methods is 'one or more state nodes connected by edges connected through a reaction node'.

In summary, heuristic MIMs show the interactions that *can* occur if the potentially interacting molecules are in the same place at the same time, or more precisely, if the relevant domains or sites can access each other. Such MIMs are independent of cell type or cell state ('generic' property), and have a generality that can encompass abnormal or uncertain conditions ('heuristic' property) as well as combinatorial
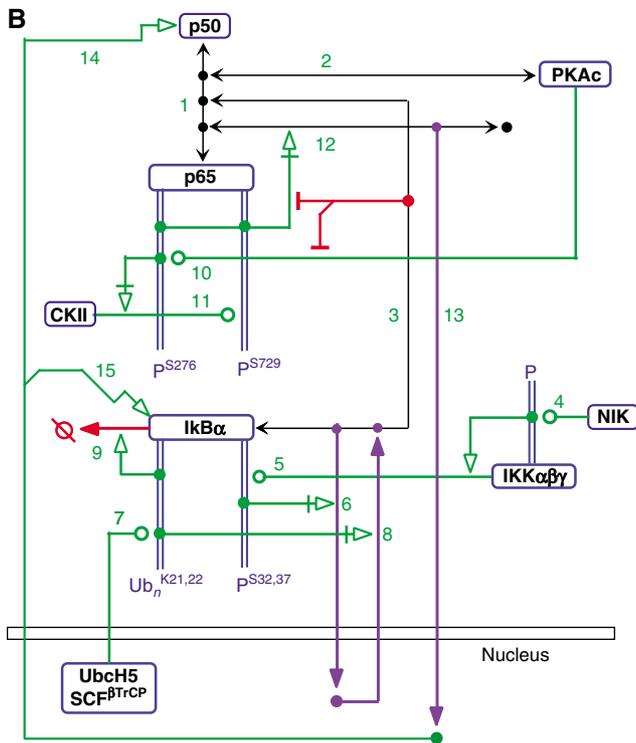
Figure 5 Continued.

complexity ('transitive' property). An MIM specific to a particular cell type or cell state can be derived from a heuristic or combinatorial MIM by deleting the molecules that are not expressed and the interactions that do not occur because the potential reactants do not occur at the same time and place.

A heuristic MIM of signaling from the EGF receptor family

The limitations of process diagrams, particularly in regard to the difficulty posed by the inherent combinatorial complexity of the networks, were recently discussed by Blinov *et al* (2006a, b), who use the 'rule-based' method to meet this difficulty. We will show how this difficulty is also overcome by the heuristic and combinatorial interpretations of MIMs.

To compare the process diagram and heuristic MIM notations with respect to their ability to organize large amounts of molecular interaction information, we prepared a heuristic MIM corresponding to a portion of the reactions shown in the large EGFR network diagram recently presented by Oda *et al* (Oda *et al*, 2005) (their Figure 1; our Figure 4). The MIM contains a subset of the reactions so as to fit legibly on one page and yet include most of the best established pathways. A similar comparison between MIM and process diagram is provided for the NF-κB signaling pathway in Figure 5.

The process diagram of the EGFR network depicts separately and in full the molecular species in each and every reaction (Oda *et al*, 2005). A given molecular species or complex therefore often appears several times in different places in this diagram. To gain a comprehensive view of the interactions of a

particular molecular species, one must therefore survey all of its occurrences wherever it may be located on the diagram. In the MIM notation, on the other hand, each molecular species generally is depicted in only one place on the map, so that all of the interactions involving this species can be traced from a single location (Figure 4).

As a named molecular species is in only one place on an MIM, it can easily be found, even in a complicated map, by way of an index of map coordinates (Kohn, 1998) or a search function that identifies the single location. Moreover, its icon (cartouche) on an on-line map (eMIM) can link to information about that species in other databases (http://discover.nci. nih.gov/mim/).

Figure 4 reveals another capability of the MIM notation: the ability to represent the complexity of EGFR family homo- and heterodimer actions in a compact manner. This is very difficult to show clearly in a compact manner using other notations, such as process diagrams (Oda *et al*, 2005).

An important feature of heuristic and combinatorial MIMs, as already mentioned, is that a given binding or modification symbol on a map may apply to many multimolecular complexes, differing with respect to the binding and modification states of the directly interacting species, sites, or domains ('transitive' property). Such MIMs therefore encompass the combinatorial complexity of a network, as we will discuss in the next section.

Whereas the process diagram of the EGFR network (Oda *et al*, 2005) specifies a particular set of interaction paths, the MIM in Figure 4 encompasses a large number of possible pathway combinations. That is, the process diagram specifies a particular model. In contrast, the heuristic MIM shown in Figure 4 encompasses several possible models (specific models could be distinguished by highlighting, as in Figure 14 of Kohn *et al*, 2006).

It may be useful to reiterate in a more specific way these subtle, but important, differences between the process diagram of Oda *et al* and the corresponding heuristic MIM in Figure 4. A major difference is that the interactions in a heuristic MIM are interpreted in a transitive manner (defined above), whereas this is not the case for process diagrams (nor for explicit MIMs). An example from Figure 4 will further clarify what we mean. The double-arrowed line that signifies reversible binding between SOS and Grb2 implies at least 16 binding interactions. (The actual number is substantially larger, but to simplify the example, we count association/dissociation as a single interaction, we count binding to different phosphotyrosines in the same molecule as a single interaction, and we ignore the multiplicity of receptor monomer and dimer states.) With this simplification, the 16 interactions implied by the binding interaction line connecting Grb2 and SOS in Figure 4 are

(1) SOS:Grb2
(2) SOS:(Grb2:ErbB1-P)
(3) SOS:(Grb2:ErbB2-P)
(4) SOS:(Grb2:Shc)
(5) SOS:(Grb2:(Shc:ErbB1-P))
(6) SOS:(Grb2:(Shc:ErbB2-P))
(7) SOS:(Grb2:(Shc:ErbB3-P))
(8) SOS:(Grb2:(Shc:ErbB4-P))

and a similar set of eight interactions with phosphorylated SOS (SOS-P) instead of unphosphorylated SOS. (We use a colon to represent binding.) This example also illustrates how the MIM notation can deal with interactions involving alternative receptor family members.

In summary, the direct interactions for the heuristic MIM (Figure 4) were taken from the process diagram of Oda *et al*, but the interpretation of the heuristic MIM differs from that of the process diagram in that the heuristic MIM includes the combinatorial complexity of the network.

Contingencies implied by colocalization

We said that heuristic MIMs by default assume that interactions involving a particular molecule occur independently of each other, unless contingency symbols are applied to indicate otherwise. Sometimes however, contingencies may be obvious enough to allow contingency symbols to be omitted, thereby simplifying the diagram. This happens when potentially interacting species are brought together to the same place; then the default assumption is that the actions that cause these species to colocalize stimulate their interaction. In Figure 4, for example, Shc is shown binding to ErbB1 (Y1148 or Y1173) and to be phosphorylated by ErbB1 (homodimer or ErbB1:ErbB2 heterodimer). The binding brings together (colocalizes) the phosphorylatable site(s) of Shc and the kinase domain of ErbB1 (or ErbB2). In the absence of a contingency symbol to indicate the contrary, the default assumption is that the colocalization brought about by the binding stimulates the phosphorylation. In another example from the same figure, SOS is shown catalyzing guanine nucleotide exchange in Ras; Ras is shown binding to plasma membrane, and SOS can be recruited to the plasma membrane via its binding through Grb2 to the ErbBs (or more indirectly via Shc). The default assumption then is that the consequent colocalization at the plasma membrane favors the SOS action. Although symbols can be added to make these contingencies explicit, diagrams are often simpler and easier to read without them. This is especially true in Figure 4 for the action of SOS on Ras, because SOS can be recruited to the plasma membrane by way of many different adapter–receptor combinations (this will be discussed further in the following section in the context of combinatorial complexity).

Further examples in Figure 4 of stimulation implied by colocalization are the actions of PI3K:p38 and PCLγ on phosphatidylinositols at the plasma membrane. These contingencies could be shown by adding appropriate symbols, but it would complicate the diagram unnecessarily.

A few known contingencies however are shown explicitly. For example, a contingency symbol indicates that p85 stimulates the activity of the kinase domain of PI3 K. On the other hand, no such symbol appears for the binding of Grb2 to SOS, because the domain of SOS that binds Grb2 does not materially alter the intrinsic activity of the catalytic domain: Grb2 enhances the action of SOS solely by bringing SOS to the plasma membrane, where its substrate is located. This convention is consistent with the principle that heuristic MIMs show what each interacting domain 'sees': the kinase domain of PI3 K senses the binding of p85, whereas the catalytic domain of SOS does not sense the binding of Grb2.
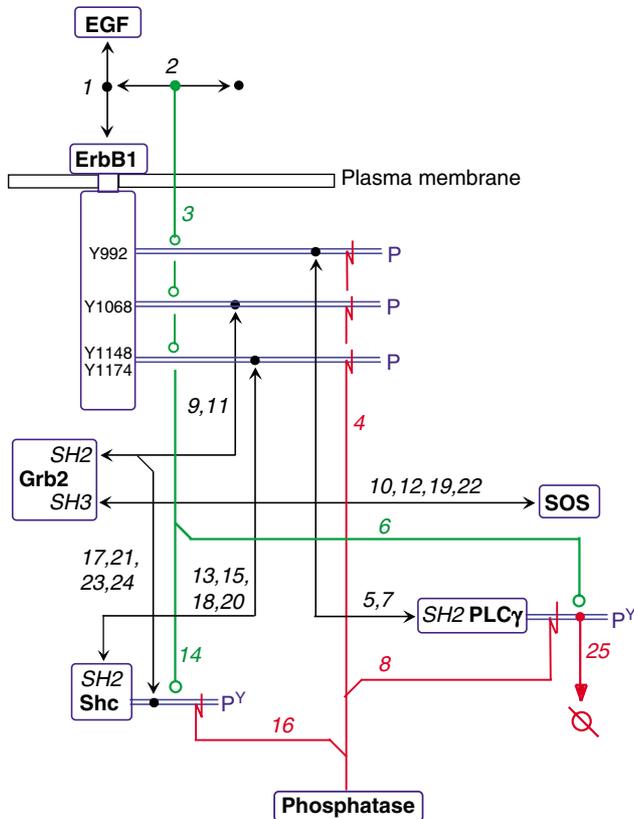
Combinatorial complexity

A protein molecule may exist in many different complexes that are composed of a variety of molecules in a variety of modification states. A given protein site or domain may therefore function in the context of many different complexes. Blinov *et al* (2006a) recently studied the effect of this diversity in a computational model of a small part of the EGFR network, particularly the early events in signaling from the receptor. The model included many of the possible molecular complexes and their interactions and was a generalization of the model analyzed by Kholodenko *et al* (1999), from which the rate constants were taken. To make the computation feasible, Blinov *et al* grouped the reactions into classes with rate constants dependent upon the directly interacting sites, but independent of many of the modifications and bindings at other sites.

The model of Blinov *et al* is a highly branched network so complicated that its full graphical representation, even in this relatively simple case, was impractical. Receptor dimerization, for example, including ligands associated with two phosphorylation sites in various combinations, comprised ∼600 different reactions (even though interactions at Y992 were not included in this enumeration).

The full repertoire of reactions in this networks can however be represented in a combinatorial MIM (Figure 6). This is made possible by the assumption of transitivity: that is, that unless otherwise indicated, a binary interaction symbol includes all of the possible modification and binding patterns of the directly interacting pair (or of the interacting species defined in a 'reaction class'). For a 'rule-based' model, such as used by Blinov *et al* (Blinov *et al*, 2005, 2006a; Faeder *et al*, 2005), the rate constants can be associated with the binding or enzyme action symbols on an MIM, as we have carried out in Figure 5 and Table II, and as we will explain further in the next section. An essential feature of the combinatorial model of Blinov *et al* is that domain bindings or site modifications are assumed to affect each other only in well-defined cases, which are grouped as reaction classes (Table II). In the combinatorial MIM (Figure 6), the identification number assigned to each reaction class listed in Table II is marked next to the corresponding interaction line.

In Figure 6, we show an MIM corresponding to the combinatorial model of the early events in EGFR signaling studied by Blinov *et al* (2006a). Blinov *et al* divided the reactions into 25 classes, listed in Table II. Each numbered step is a reaction class, consisting of many different reactions, all of which are assigned the same rate constant in their rule-based model. In their tables, Blinov *et al* (2006a) show the number of reactions in each class and the assigned rate constants. When all the combinatorial possibilities of the 25 reaction classes are included, the total number of reactions added up to 3749!

Interaction-*1* in Figure 6 (Blinov's step 1) represents ligand– receptor binding (and dissociation, as the interaction is taken to be reversible). Interaction-*1* includes reversible binding of ligand to receptor in any of its modification states, and in

Figure 6 Combinatorial MIM of the combinatorial model analyzed by Blinov *et al* (2006a). The numbers correspond to the reaction class or step numbers assigned by Blinov *et al* and listed in Table II.

complex with any of the combination of molecules that its cytoplasmic domain may bind. As Blinov *et al* assume that receptor dimers can dissociate even when phosphorylated and/or bound to cytosolic proteins, interaction-*1* includes bindings and modifications of receptor monomer, as well as dimer. In all, Blinov *et al* enumerate 48 binding reactions in this class.

Interaction-*2* represents reversible dimerization between ligand and receptor. Dimerization is assumed to require that both molecules of receptor have bound ligand, but that all other combinations of possible bindings or modifications are included. According to Blinov's count, this step comprises a total of 600 reactions. (The double-arrowed line connecting the ligand:receptor node to the isolated node means that ligand:receptor in any cytoplasmic state can bind to another copy of ligand:receptor in any of these states.)

Interaction-*3* refers to receptor tyrosine phosphorylations, including phosphorylation of any site without regard to the status of the other sites on both members of the homodimer. As the reaction occurs *in trans* (one member of the dimer phosphorylating the other), we show the reaction to be catalyzed by the homodimer.

Interaction-*4* refers to dephosphorylation of any site on the receptor, regardless of other phosphorylations and/or bindings.

The SH2 sites of Grb2 can bind phosphotyrosine-1068 of ErbB1 (monomer or dimer) (interaction-*9*), and the SH3 site of Grb2 can bind cytosolic SOS (interaction-*12*). These two

bindings can coexist, as there are no contingency symbols in Figure 5 to indicate otherwise. Moreover, the two bindings can form in either order. Blinov *et al* assigned each order of formation to a separate reaction class with separate rate constants (their steps 10 and 11; the same numbers identify the reactions in Figure 6 and Table II). Reaction class 10 is Sos binding to the SH3 site of Grb2 (step 9) after the SH2 sites of Grb2 have bound to pY1068 of ErbB1 (step 12). The reaction class includes all possible modification or binding states other than those specified in other reaction classes or otherwise excluded in the specification of a particular model.

The specification of the other reaction steps or classes can be gleaned from Table II and Figure 6, which use the same step numbers.

In their simulations, Blinov *et al* treat the interactions involving PLCγ at ErbB1 phosphotyrosine-992 differently from the interactions at the other ErbB1 sites. They carry out simulations in which the interactions of PLCγ are included or not. Figure 6 includes all of the interactions, and therefore is generic with respect to how the reaction subsets are segregated into classes in a particular model.

Conclusions

The MIM notation has the characteristics and flexibility required for a standard diagram representation of complex biological networks. Here, we have demonstrated how MIMs can represent networks in three demanding types of applications, in each case drawn from recently published network studies. These demonstrations argue that the MIM notation is suitable to become a standard graphic notation (1) for definition of complex models to be used in computer simulation of bioregulatory networks, (2) for compact, detailed, and illuminating representation of available information about molecular interactions in a complex network, and (3) for representation of the combinatorial complexity of network models. The advantages of the MIM notation, we think, justify the effort to learn the rules of the notation.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

References

Aguda BD, Sauro HM (2004) Computer simulation of MAPK signal transduction. *Methods Mol Biol* **250:** 167–176

Aladjem MI, Pasa S, Parodi S, Weinstein JN, Pommier Y, Kohn KW (2004) Molecular interaction maps—a diagrammatic graphical language for bioregulatory networks. *Sci STKE*, 2004: pe8

Au WW, Henderson BR (2005) The BRCA1 RING and BRCT domains cooperate in targeting BRCA1 to ionizing radiation-induced nuclear foci. *J Biol Chem* **280:** 6993–7001

Blinov ML, Faeder JR, Goldstein B, Hlavacek WS (2006a) A network model of early events in epidermal growth factor receptor signaling that accounts for combinatorial complexity. *Biosystems* **83:** 136–151

Blinov ML, Yang J, Faeder JR, Hlavacek WS (2005) Graph theory for rule-based modeling of biochemical networks. In *Proc BioCONCUR 2005*, Mishra B, Priami C (eds) pp 1–19

Blinov ML, Yang J, Faeder JR, Hlavacek WS (2006b) Depicting signaling cascades. *Nat Biotechnol* **24:** 137–138; author reply 138

Cary LA, Cooper JA (2000) Molecular switches in lipid rafts. *Nature* **404:** 945–947

Cook DL, Farley JF, Tapscott SJ (2001) A basis for a visual language for describing, archiving and analyzing functional models of complex biological systems. *Genome Biol* **2:** RESEARCH0012

Ellis RJ (2001) Macromolecular crowding: obvious but underappreciated. *Trends Biochem Sci* **26:** 597–604

Faeder JR, Blinov ML, Hlavacek WS (2005) Graphical rule-based representation of signal-transduction networks. In *ACM Symposium on Applied Computing*, Liebrock LM (ed), pp 133–140. New York: ACM Press

Finney A, Hucka M (2003) Systems biology markup language: level 2 and beyond. *Biochem Soc Trans* **31:** 1472–1473

Funahashi A, Morohashi M, Kitano H (2003) CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. *Biosilico* **1:** 159–162

Hancock R (2004) Internal organisation of the nucleus: assembly of compartments by macromolecular crowding and the nuclear matrix model. *Biol Cell* **96:** 595–601

Hucka M, Finney A, Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, Cuellar AA, Dronov S, Gilles ED, Ginkel M, Gor V, Goryanin II, Hedley WJ, Hodgman TC, Hofmeyr JH, Hunter PJ, Juty NS, Kasberger JL, Kremling A, Kummer U, Le Novere N, Loew LM, Lucio D, Mendes P, Minch E, Mjolsness ED, Nakayama Y, Nelson MR, Nielsen PF, Sakurada T, Schaff JC, Shapiro BE, Shimizu TS, Spence HD, Stelling J, Takahashi K, Tomita M, Wagner J, Wang J (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19:** 524–531

Kholodenko BN (2006) Cell-signalling dynamics in time and space. *Nat Rev Mol Cell Biol* **7:** 165–176

Kholodenko BN, Demin OV, Moehren G, Hoek JB (1999) Quantification of short term signaling by the epidermal growth factor receptor. *J Biol Chem* **274:** 30169–30181

Kholodenko BN, Westerhoff HV (1995) The macroworld versus the microworld of biochemical regulation and control. *Trends Biochem Sci* **20:** 52–54

Kitano H (2003) A graphical notation for biochemical networks. *Biosilico* **1:** 169–176

Kitano H, Funahashi A, Matsuoka Y, Oda K (2005) Using process diagrams for the graphical representation of biological networks. *Nat Biotechnol* **23:** 961–966

Kohn KW (1998) Functional capabilities of molecular network components controlling the mammalian G1/S cell cycle phase transition. *Oncogene* **16:** 1065–1075

Kohn KW (1999) Molecular interaction map of the mammalian cell cycle control and DNA repair systems. *Mol Biol Cell* **10:** 2703–2734

Kohn KW (2001) Molecular interaction maps as information organizers and simulation guides. *Chaos* **11:** 84–97

Kohn KW, Aladjem MI (2006) Circuit diagrams for biological networks. *Mol Systems Biol*, doi:10.1038/msb4100044

Kohn KW, Aladjem MI, Pasa S, Parodi S, Pommier Y (2003) Molecular interaction map of mammalian cell cycle control. *Encyclopedia Hum Genome* **1:** 457–474

Kohn KW, Aladjem MI, Weinstein JN, Pommier Y (2006) Molecular interaction maps of bioregulatory networks: a general rubric for systems biology. *Mol Biol Cell* **17:** 1–13

Kohn KW, Bohr VA (2002) Genomic instability and DNA repair. In *Cancer Handbook*, Chapter 7, Alison M (ed) London: Nature Publishing Group (Macmillan)

Kohn KW, Pommier Y (2005) Molecular interaction map of the p53 and Mdm2 logic elements, which control the Off–On switch of p53 in response to DNA damage. *Biochem Biophys Res Commun* **331:** 816–827

Kohn KW, Riss J, Aprelikova O, Weinstein JN, Pommier Y, Barrett JC (2004) Properties of switch-like bioregulatory networks studied by simulation of the hypoxia response control system. *Mol Biol Cell* **15:** 3042–3052

Kurata H, Matoba N, Shimizu N (2003) CADLIVE for constructing a large-scale biochemical network based on a simulation-directed notation and its application to yeast cell cycle. *Nucleic Acids Res* **31:** 4071–4084

Machne R, Finney A, Muller S, Lu J, Widder S, Flamm C (2006) The SBML ODE Solver Library: a native API for symbolic and fast numerical analysis of reaction networks. *Bioinformatics* **22:** 1406–1407

Minton AP (2001) The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. *J Biol Chem* **276:** 10577–10580

Oda K, Matsuoka Y, Funahashi A, Kitano H (2005) A comprehensive pathway map of epidermal growth factor receptor signaling. *Mol Syst Biol* **1:** 2005.0010

Parton RG, Hancock JF (2004) Lipid rafts and plasma membrane microorganization: insights from Ras. *Trends Cell Biol* **14:** 141–147

Pilch DR, Sedelnikova OA, Redon C, Celeste A, Nussenzweig A, Bonner WM (2003) Characteristics of gamma-H2AX foci at DNA double-strand breaks sites. *Biochem Cell Biol* **81:** 123–129

Pommier Y, Kohn KW (2003) [Cell cycle and checkpoints in oncology: new therapeutic targets]. *Med Sci (Paris)* **19:** 173–186

Pommier Y, Sordet O, Antony S, Hayward RL, Kohn KW (2004) Apoptosis defects and chemotherapy resistance: molecular interaction maps and networks. *Oncogene* **23:** 2934–2949

Pommier Y, Weinstein JN, Aladjem MI, Kohn KW (2006) Chk2 molecular interaction map and rationale for Chk2 inhibitors. *Clin Cancer Res* **12:** 2657–2661

Rajendran L, Simons K (2005) Lipid rafts and membrane dynamics. *J Cell Sci* **118:** 1099–1102

Shea MA, Ackers GK (1985) The OR control system of bacteriophage lambda. A physical–chemical model for gene regulation. *J Mol Biol* **181:** 211–230

Wolf DM, Eeckman FH (1998) On the relationship between genomic regulatory element organization and gene regulatory dynamics. *J Theor Biol* **195:** 167–186