Depicting signaling cascades

To the editor:

In a paper in the August issue (*Nat. Biotechnol.* **23**, 961–966, 2005), Kitano *et al.* discuss the use of process diagrams to map signal-transduction cascades. They have used the formalism of process diagrams to specify pathway maps that are both readable and precise, and they have developed a map depicting hundreds of species and reactions involved in signaling by the epidermal growth factor receptor (EGFR)[1]. However, this map, as expansive as it is, omits the vast majority of species and reactions that could potentially be generated during signaling. We submit that comprehensive process diagrams for this, or any other signaling system, are very likely to be of unmanageable size. The reason is combinatorial complexity, a hallmark of signal-transduction cascades[2–5]. Although Kitano *et al.* discuss this problem in their paper and suggest some solutions (e.g., modules for concise representation of subnetworks of a signaling system), we feel their solutions are inadequate in that explicit representation of all species at some level is still required.

Here, we wish to call attention to an alternative method of representation that we believe better addresses the problem of combinatorial complexity. This method involves the use of graphical reaction rules to represent the protein-protein interactions in a system and their consequences[6,7]. A rule illustrates features of species relevant for a particular type of reaction that can result from a protein-protein interaction, whereas a process diagram illustrates individual species and reactions.

Before discussing rules further, we should clarify the limitations of process diagrams. Let us consider the map of Figure 3e in the original Kitano *et al.* paper, which depicts 18 species and 32 reactions involved in EGFR signaling. These species and reactions correspond, more or less, to those included in the mathematical model of Kholodenko *et al.*[8], and they arise from interactions among five proteins: EGFR, its ligand epidermal growth factor (EGF), the adapters Grb2 and Shc, and the guanine nucleotide exchange factor Sos. The map, as we will elaborate shortly, presents an arguably oversimplified picture of signaling events. However, it is already challenging to decipher because a fairly large number of pictograms and intersecting arrows are needed to illustrate the various species and reactions. How complicated would the map be if it presented a more comprehensive picture of signaling?

Interactions of the proteins considered in Figure 3e of Kitano *et al.* can potentially generate not tens of species but hundreds to thousands of species, and even more reactions[4,9–11]. A focus on the 18 species of the map is appropriate only if several limiting assumptions hold true. These assumptions, upon which the model of Kholodenko *et al.*[8] (and derivative models such as that of Schoeberl *et al.*[12]) are based, include the following: first, simultaneous phosphorylation of tyrosines of both receptors in a ligand-induced receptor dimer; second, association of at most one adapter with a given receptor dimer at a time; and third, no dissociation of receptor dimers if receptors are phosphorylated.

In recent work[11], we discuss the validity of these assumptions and consider the impact of relaxing them. The result is an extended model for Sos activation that predicts the dynamics of a network of 356 species and 3,749 unidirectional reactions, all of which arise from protein-protein interactions underlying the map of Kitano *et al.*

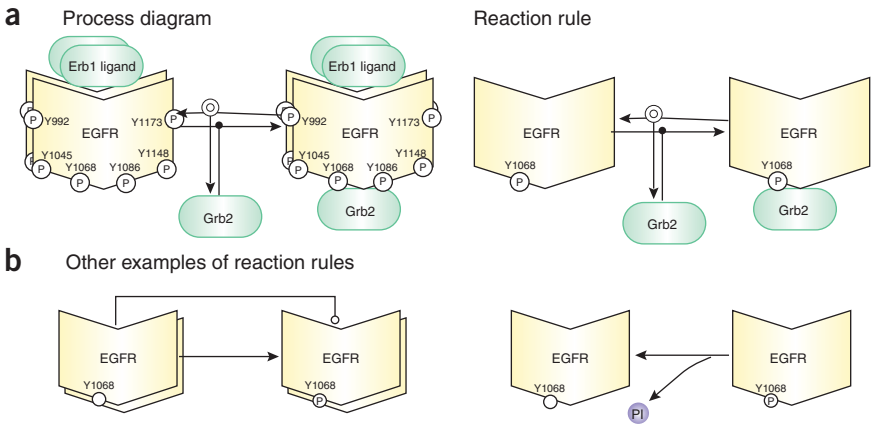

Figure 1 A process diagram and three graphical reaction rules drawn using CellDesigner[16]. (**a**) The process diagram illustrates Grb2 binding to a particular EGFR-containing species: three species and two unidirectional reactions are depicted. The adjacent reaction rule, drawn in a style consistent with the diagrammatic conventions of Kitano *et al.*, also pertains to Grb2 interaction with EGFR. It is one of the rules used to generate our model for EGFR signaling[11], and it indicates that Grb2-EGFR association via Y1068 in EGFR depends only on phosphorylation of this residue. By convention, it is assumed that the interaction represented in a rule is independent of all features not explicitly indicated. Thus, multiple species may qualify as reactants in a type of reaction defined by a rule. The exact number of reactions generated by the rule depends on the graph grammar of which the rule is a part (that is, the rule set and seed species that generate a model)[7]. Within the scope of our model[11], the rule shown here generates 312 distinct unidirectional reactions. (**b**) These reaction rules, which are also included in the rule set used to generate our model for EGFR signaling[11], represent transphosphorylation of one EGFR in a receptor dimer by the neighboring receptor and receptor dephosphorylation, which is catalyzed by phosphatases assumed to be present in excess. The left rule indicates that EGFR-catalyzed phosphorylation of Y1068 depends on dimerization of EGFR. In contrast, the right rule indicates that receptor dephosphorylation is spontaneous and independent of the state of EGFR aggregation. These rules generate 144 and 156 reactions, respectively, in our model for EGFR signaling[11].

We have found that consideration of this additional complexity is necessary if the model is to make accurate predictions about network dynamics and the role of specific components, such as individual sites of tyrosine phosphorylation[11,13].

Drawing a process diagram with 356 species to represent the interactions of only five proteins[11] would be inefficient and difficult to accomplish or read. Moreover, there are no obvious modules that could be introduced to simplify the process diagram, because the reaction network is highly branched[11]. In any case, a module has the drawback that protein-protein interactions are either altogether hidden (when the module is closed) or obscured by the possibly large number of species and reactions that can arise from the interactions (when the module is open).

Given that protein-protein interactions can generate myriad species and reactions for combinatorial reasons, what can be done to capture the essence of these interactions without ignoring their combinatorial complexity? To address this problem, we have proposed that protein-protein interactions and their effects be represented in the form of reaction rules that are generators of species and reactions[14,15]. More recently, we have introduced graphical reaction rules[6,7], in which graphs similar to the pictograms of process diagrams are used to represent features of proteins and protein complexes. Graphical rules were introduced to allow the connectivity of proteins in a complex to be explicitly represented, and they also provide a means to comprehensibly visualize protein–protein interactions, as illustrated in **Figure 1**.

In summary, process diagrams are useful for representing the individual species and reactions that can arise in a signaling system. However, representation at this microscopic level of detail may not be practical. In the face of combinatorial complexity, diagrams can be overly complicated or hide information about protein-protein interactions. An alternative approach is to represent not the species and reactions resulting from the interactions of proteins in a system but rather the interactions themselves. This task can be accomplished relatively easily using graphical reaction rules. A set of rules can be interpreted to obtain a mathematical model that accounts comprehensively for the species and reactions logically consistent with the rules, even when large numbers of species and reactions are possible[7,14,15]. We are currently extending the BioNetGen software package[14,15] to provide tools for drawing and interpreting graphical reaction rules (http://cellsignaling.lanl.gov/). In the future, we believe such model-generation tools will play an important role in obtaining a mechanistic understanding of cellular information processing and in manipulating signaling systems for therapeutic and biotechnological purposes.

Michael L. Blinov, Jin Yang, James R. Faeder & William S. Hlavacek

Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA. e-mail: wish@lanl.gov

1. Oda, K., Matsuoka, Y., Funahashi, A. & Kitano, H. *MMol. Syst. Biol.* published online 25 May 2005; doi:10.1038/msb4100014
2. Endy, D. & Brent, R. *Nature* **409**, 391–395 (2001).
3. Bray, D. *Science* **299**, 1189–1190 (2003).
4. Hlavacek, W.S., Faeder, J.R., Blinov, M.L., Perelson, A.S. & Goldstein, B. *Biotechnol. Bioeng.* **84**, 783–794 (2003).
5. Goldstein, B., Faeder, J.R. & Hlavacek, W.S. *Nat. Rev. Immunol.* **4**, 445–456 (2004).
6. Faeder, J.R., Blinov, M.L. & Hlavacek, W.S. In *Proc. 2005 ACM Symp. Appl. Computing* (Santa Fe, NM, March 13-17, 2005) Liebrock, L.M., Editor. ACM Press, New York, NY, pp. 133-140. doi:10.1145/1066677.1066712
7. Blinov, M.L., Yang, J., Faeder, J.R. & Hlavacek, W.S. *In Proc. BioCONCUR 2005* (Mishra, B. & Priami, C., eds) 1–19 (San Francisco, CA, August 27, 2005).
8. Kholodenko, B.N., Demin, O.V., Moehren, G. & Hoek, J.B. *J. Biol. Chem.* **274**, 30169–30181 (1999)
9. Conzelmann, H. *et al. Syst. Biol.* **1**, 159–169 (2004).
10. Borisov, N.M., Markevich, N.I., Hoek, J.B. & Kholodenko, B.N. *Biophys. J.* **89**, 951–966 (2005).
11. Blinov, M.L., Faeder, J.R., Goldstein, B. & Hlavacek, W.S. *BioSystems* doi:10.1016/j.biosystems.2005.06.014, **in press** (2005).
12. Schoeberl, B., Eichler-Jonsson, C., Gilles, E.D. & Müller, G. *Nat. Biotechnol.* **20**, 370–375 (2002).
13. Faeder, J.R., Blinov, M.L., Goldstein, B. & Hlavacek, W.S. *Syst. Biol.* **2**, 5–15 (2005).
14 Blinov, M.L., Faeder, J.R., Goldstein, B. & Hlavacek, W.S. *Bioinformatics* **20**, 3289–3291 (2004).
15. Faeder, J.R., Blinov, M.L., Goldstein, B. & Hlavacek, W.S. *Complexity* **10**, 22–41 (2005)
16. Funahashi, A., Tanimura, N., Morohashi, M. & Kitano, H. *BIOSILICO* **1**, 159–162 (2003).

Kitano *et al.* respond:

The first issue raised by Blinov *et al.* suggests that pathway maps are too simplistic to represent the protein combinatorial explosion in signal cascades. They detail Figure 3e in our article to illustrate their point; however, this figure was used solely to demonstrate the look-and-feel of how to represent pathways as process diagrams. Therefore, we used part of the diagram in a Hanahan and Weinberg paper[1], which is also a pathway extensively used in simulation studies[2,3]. It was not argued that this was a comprehensive representation of the EGFR pathway. Our recent interaction map published in *Molecular Systems Biology*[4] was intended to be a comprehensive EFGR map of experimentally validated interactions. We did not enumerate all possible interactions and molecular states and recognize that there are interactions not listed in the map due to lack of experimental validation, despite theoretical and intuitive possibilities. The process diagram is neutral on what should be described in the map. It defines the graphical representation of an interaction map; thus, the oversimplification critique does not apply to the process diagram itself as construction of these maps relies on experimental evidence.

The second issue raised was that describing all combinatorial states of molecules and resulting complexes would result in a combinatorial explosion making a rule-based approach more appropriate for modeling. We would argue that this depends on the intended use of the map. The process diagram was motivated by an experimentalist's need partly to represent detailed interactions, including residue modification state, to improve experimental design, and partly to visualize their data in the context of a pathway map where each combinatorial state has been explicitly described, regardless of the level of complexity. It is imperative that software tools make such complex and large-scale maps accessible to users.

Although the rule-based approach has attracted much attention as a viable approach for dynamical simulation[5,6], it may not allow users to project experimental data on to each combinatorial state without expansion. As illustrated by Blinov *et al.* wherever the rule-based approach is shown to be effective, the process diagram can then be used to expand graphical notation to represent rules and the network generated from the rule. We would like to incorporate such features into the process diagram and are receptive to constructive critiques to create standard graphical notations; to this end, we have formed an international alliance to standardize graphical notation called Systems Biology Graphical Notation (http://www.sbgn.org/).

1. Hanahan, D. & Weinberg, R.A. *Cell* **100**, 57–70 (2000).
2. Kholodenko, B.N., Demin, O.V., Moehren, G. & Hoek, J.B. *J. Biol. Chem.* **274**, 30169–30181 (1999).
3. Schoeberl, B., Eichler-Jonsson, C., Gilles, E.D. & Muller, G. *Nat. Biotechnol.* **20**, 370–375 (2002).
4. Oda, K., Matsuoka, Y., Funahashi, A. & Kitano, H. *Mol. Systems Biol.* doi: 10.1038/msb4100014 (25 May 2005).
5. Lok, L. & Brent, R. *Nat. Biotechnol.* **23**, 131–136 (2005).
6. Borisov, N.M., Markevich, N.I., Hoek, J.B. & Kholodenko, B.N. *Biosystems* doi:10.1016/j.biosystems.2005.03.006 (17 October 2005).