networks. Thus, the characteristics of middle-degree sub-networks strongly influence the statistical characteristics of the whole PIN. The whole network architecture seems to have tightly connected middle-degree nodes that are connected to high-degree nodes, and a large number of low-degree nodes are mostly connected to high-degree nodes (see Figure 2). Moreover, we used more stringent thresholds for middle- and high-degree nodes and found that changing the thresholds did not essentially affect the results (i.e., the average cluster coefficient, average shortest path length, or G_C) (see Tables S3 and S4).

The series of analyses thus far indicates that the functional role for proteins included in low-degree, middle-degree, and high-degree sub-networks are totally different. This means that the yeast and human PINs are not scale-free in terms of the composition of the functional role of proteins. Proteins with each functional group have a characteristic degree distribution. To investigate the degree distribution of proteins in each functional category, we annotated proteins in the yeast and human PINs by using the GO slim biological process ontology. As shown in Figures 4 and S5, there are different degree-distribution patterns for proteins from different functional categories. This suggests that a scale-free distribution emerges from the composition of different functional protein groups each of which has scale-dependent degree distributions. Thus, from the functional distribution, the yeast and human PINs are scale-rich.

***S*(*g*) value.** Before giving a definition for the $S(g)$ value, let us first define some notations. Let n be the number of nodes in a network and k_i be the degree of node i. $D=\{k_1,k_2,\ldots,k_n\}$ represents a given degree distribution and $G(D)$ denotes the set of all connected networks having the same degree distribution, D. For a network, g, having degree distribution D, graph-theoretic quantity $s(g)$ is defined as $s(g)=\Sigma_{(i,j)\in E(g)}k_ik_j$, where $E(g)$ is the set of links in the network. $s_{\max}$ is defined as $s_{\max}=\max\{s(g)\colon g\in G(D)\}$ and we calculated the value of $s_{\max}$ by using the algorithm devised by Alderson et al. [7]. $S(g)$, the value normalized against $s_{\max}$, is defined as $S(g)=s(g)/s_{\max}$ [8]. In this paper, we calculated the value of $S(g)$ in the yeast and human PINs.

Supporting Information

Figure S1 Statistics of sub-networks generated by MSD (yeast PIN). Red triangles and black squares show the values for the yeast PIN and random network, respectively. The results for random network were obtained by taking the average among 100 random networks. (A) Distribution of <C(kC)>. (B) Distribution of <L(kC)>. (C) Distribution of GC(kC). (D) Distribution of PLC(kC). The dashed line represents the probability that a randomly selected protein is a lethal protein.
Found at: doi:10.1371/journal.pcbi.1000550.s001 (1.37 MB TIF)

Figure S2 Statistics of sub-networks generated by MSD (human PIN). Red triangles and black squares show the values for the human PIN and random network, respectively. The results for random network were obtained by taking the average among 100 random networks. (A) Distribution of <C(kC)>. (B) Distribution of <L(kC)>. (C) Distribution of GC(kC). (D) Distribution of PDT(kC). The dashed line in black represents the probability that a randomly selected protein is a drug target.
Found at: doi:10.1371/journal.pcbi.1000550.s002 (1.35 MB TIF)

Figure S3 Degree Dependent Connectivity Chart with stringent thresholds. Pn(k) gives the probability that a link of a k-degree node is a link to a node in each sub-network of the yeast (left) and human (right) PINs. The value of Pn(k) is calculated for a sub-network consisting of high-degree nodes, that consisting of middle-degree nodes, and that consisting of low-degree nodes. (A) Distribution of Pn(k) for the high-degree sub-network. (B) Distribution of Pn(k) for the middle-degree sub-network. (C) Distribution of Pn(k) for the low-degree sub-network.
Found at: doi:10.1371/journal.pcbi.1000550.s003 (0.26 MB TIF)

Figure S4 Cloud topologies in yeast and human PINs with stringent thresholds. Grey, red, and blue nodes correspond to low-, middle-, and high-degree nodes. Grey, red, green, and blue links correspond to links between low- and high-degree nodes, those between middle-degree nodes, those between middle- and high-degree nodes, and those between high-degree nodes. For clarity, low- and middle-degree nodes that have no links to high-degree nodes have been omitted. (A) Altocumulus and stratus structures in the yeast PIN. (B) Stratus structure in the yeast PIN. (C) Altocumulus structure in the yeast PIN. (D) Altocumulus and stratus structure in the human PIN. (E) Stratus structure in the human PIN. (F) Altocumulus structure in the human PIN.
Found at: doi:10.1371/journal.pcbi.1000550.s004 (2.83 MB TIF)

Figure S5 Scale-richness in human PIN. Each diagram shows cumulative degree distributions of proteins in each functional group. The name above each diagram denotes the name of the functional category with which the cumulative degree distribution was examined.
Found at: doi:10.1371/journal.pcbi.1000550.s005 (0.45 MB TIF)

Table S1 Statistics of sub-networks in the yeast PIN. a. number of nodes b. average shortest path legth c. fraction of nodes contained in a largest component to all nodes contained in a sub-network d. average cluster coefficient e. betweeness centrality f. fraction of essential nodes to all nodes contained in a sub-network g. a sub-network consist of low-degree nodes h. a sub-network consist of middle-degree nodes i. a sub-network consist of high-degree nodes j. a sub-network consist of low- and middle-degree nodes k. a sub-network consist of low- and high-degree nodes.
Found at: doi:10.1371/journal.pcbi.1000550.s006 (0.04 MB DOC)

Table S2 Statistics of sub-networks in the human PIN. a. See Table S1. b. fraction of drug-target nodes contained in a sub-network to all nodes contained in the sub-network.
Found at: doi:10.1371/journal.pcbi.1000550.s007 (0.04 MB DOC)

Table S3 Statistics of sub-networks in yeast PIN with stringent thresholds for middle- and high-degree nodes. a. See Table S1.
Found at: doi:10.1371/journal.pcbi.1000550.s008 (0.04 MB DOC)

Table S4 Statistics of sub-networks in human PIN with stringent thresholds for middle- and high-degree nodes. a. See Table S2.
Found at: doi:10.1371/journal.pcbi.1000550.s009 (0.04 MB DOC)

Table S5 Degrees of the genes in yeast PIN belonging to each functional category. a. Mean degree among the proteins contained in each functional category. b. Number of proteins in each functional category. c. ***, **, and * represents that a given value is significantly higher (or lower) than average degree among proteins belonging other functional categories with $P<0.001$, $P<0.01$, and $P<0.05$, respectively, by the Wilcoxon rank-sum two-sample test with the Bonferronni correction.
Found at: doi:10.1371/journal.pcbi.1000550.s010 (0.06 MB DOC)

Table S6 Degrees of the genes in human PIN belonging to each functional category. a. See Table S5.

Found at: doi:10.1371/journal.pcbi.1000550.s011 (0.04 MB DOC)

Table S7 Middle degree proteins in human PIN and their functions.
Found at: doi:10.1371/journal.pcbi.1000550.s012 (2.79 MB DOC)

Acknowledgments

The authors wish to thank Dr. T. Shiraishi at Sony Computer Science Laboratories for his useful comments and discussion. We would also like to thank the anonymous reviewers for their valuable suggestions.

Author Contributions

Conceived and designed the experiments: TH HT YS SN HK. Performed the experiments: TH YS SN. Analyzed the data: TH HT YS SN. Wrote the paper: TH HK.

References

1. Henney A, Superti-Furga G (2008) A network solution. Nature 455: 730–731.
2. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, et al. (2007) The human disease network. Proc Natl Acad Sci USA 104: 8685–8690.
3. Yildirim MA, Goh KI, Cusick ME, Barabasi AL, Vidal M (2007) Drug-target network. Nat Biotechnol 25: 1119–1126.
4. Albert R (2005) Scale-free networks in cell biology. J Cell Sci 118: 4947–4957.
5. Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5: 101–113.
6. Willinger W, Alderson DL, Doyle JC, Li L (2004) More "normal" than normal: scaling distributions and complex systems. In: Ingaalls RG, et al. (2004) Proceedings of the 2004 Winter Simulation Conference. Washington D. C.: IEEE Press. pp 130–141.
7. Alderson DL, Li L, Willinger W, Doyle JC (2005) Understanding internet topology: principles, models, and validation. IEEE/ACM Transactions on Networking 13: 1205–1218.
8. Doyle JC, Alderson DL, Li L, Low S, Roughan M, et al. (2005) The "robust yet fragile" nature of the Internet. Proc Natl Acad Sci USA 102: 14497–14502.
9. Kitano H (2007) Towards a theory of biological robustness. Mol Syst Biol 3: 137.
10. Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, et al. (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature 430: 88–93.
11. Patil A, Nakamura H (2006) Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks. FEBS Lett 580: 2041–2045.
12. Vazquez A (2003) Growing network with local rules: preferential attachment, clustering hierarchy, and degree correlations. Phys Rev E Stat Nonlin Soft Matter Phys 67: 056104.
13. Maslov S, Sneppen K (2002) Specificity and stability in topology of protein networks. Science 296: 910–913.
14. Spirin V, Mirny LA (2003) Protein complexes and functional modules in molecular networks. Proc Natl Acad Sci USA 100: 12123–12128.
15. Batada NN, Reguly T, Breitkreutz A, Boucher L, Breitkreutz BJ, et al. (2006) Stratus not altocumulus: a new view of the yeast protein interaction network. PLoS Biol 4: e317.
16. Tanaka R (2005) Scale-rich metabolic networks. Phys Rev Lett 94: 168101.
17. Feldman I, Rzhetsky A, Vitkup D (2008) Network properties of genes harboring inherited disease mutations. Proc Natl Acad Sci USA 105: 4323–4328.
18. Yao L, Rzhetsky A (2008) Quantitative systems-level determinants of human genes targeted by successful drugs. Genome Res 18: 206–213.
19. Kitano H (2007) A robustness-based approach to systems-oriented drug design. Nat Rev Drug Discov 6: 202–210.
20. Rzhetsky A, Gomez SM (2001) Birth of scale-free molecular networks and the number of distinct DNA and protein domains per genome. Bioinformatics 17: 988–996.
21. Li L, Alderson DL, Willinger W, Doyle JC (2004) A first-principles approach to understanding the internet's router-level topology. In: Proceedings of the 2004 conference on Application, technologies, architectures, and protocols for computer communications. New York: ACM Press. pp 3–14.
22. Guldener U, Munsterkotter M, Oesterheld M, Pagel P, Ruepp A, et al. (2006) MPact: the MIPS protein interaction resource on yeast. Nucleic Acids Res 34: D436–441.
23. Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, et al. (2005) Towards a proteome-scale map of the human protein-protein interaction network. Nature 437: 1173–1178.
24. Tong AHY, Lesage G, Bader GD, Ding H, Xu H, et al. (2004) Global mapping of the yeast genetic interaction network. Science 303: 808–813.
25. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, et al. (2006) DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res 34: D668–672.
26. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393: 440–442.

The Systems Biology Graphical Notation

Nicolas Le Novère[1], Michael Hucka[2], Huaiyu Mi[3], Stuart Moodie[4], Falk Schreiber[5,6], Anatoly Sorokin[7], Emek Demir[8], Katja Wegner[9], Mirit I Aladjem[10], Sarala M Wimalaratne[11], Frank T Bergman[12], Ralph Gauges[13], Peter Ghazal[4,14], Hideya Kawaji[15], Lu Li[1], Yukiko Matsuoka[16], Alice Villéger[17,18], Sarah E Boyd[19], Laurence Calzone[20], Melanie Courtot[21], Ugur Dogrusoz[22], Tom C Freeman[14,23], Akira Funahashi[24], Samik Ghosh[16], Akiya Jouraku[24], Sohyoung Kim[10], Fedor Kolpakov[25,26], Augustin Luna[10], Sven Sahle[13], Esther Schmidt[1], Steven Watterson[4,22], Guanming Wu[27], Igor Goryanin[4], Douglas B Kell[18,28], Chris Sander[8], Herbert Sauro[12], Jacky L Snoep[29], Kurt Kohn[10] & Hiroaki Kitano[16,30,31]

Circuit diagrams and Unified Modeling Language diagrams are just two examples of standard visual languages that help accelerate work by promoting regularity, removing ambiguity and enabling software tool support for communication of complex information. Ironically, despite having one of the highest ratios of graphical to textual information, biology still lacks standard graphical notations. The recent deluge of biological knowledge makes addressing this deficit a pressing concern. Toward this goal, we present the Systems Biology Graphical Notation (SBGN), a visual language developed by a community of biochemists, modelers and computer scientists. SBGN consists of three complementary languages: process diagram, entity relationship diagram and activity flow diagram. Together they enable scientists to represent networks of biochemical interactions in a standard, unambiguous way. We believe that SBGN will foster efficient and accurate representation, visualization, storage, exchange and reuse of information on all kinds of biological knowledge, from gene regulation, to metabolism, to cellular signaling.

"*Un bon croquis vaut mieux qu'un long discours*" ("A good sketch is better than a long speech"), said Napoleon Bonaparte. This claim is nowhere as true as for technical illustrations. Diagrams naturally engage innate cognitive faculties[1] that humans have possessed since before the time of our cave-drawing ancestors. Little wonder that we find ourselves turning to them in every field of endeavor. Just as with written human languages, communication involving diagrams requires that authors and readers agree on symbols, the rules for arranging them and the interpretation of the results. The establishment and widespread use of standard notations have permitted many fields to thrive. One can hardly imagine today's electronics industry, with its powerful, visually oriented design and automation tools, without having first established standard notations for circuit diagrams. Such was not the case in biology[2]. Despite the visual nature of much of the information exchange, the field was permeated with *ad hoc* graphical notations having little in common between different researchers, publications, textbooks and software tools. No standard visual language existed for describing biochemical interaction networks, inter- and intracellular signaling gene regulation—concepts at the core of much of today's research in molecular, systems and synthetic biology. The closest to a standard is the notation long used in many metabolic and signaling pathway maps, but in reality, even that lacks uniformity between sources and suffers from undesirable ambiguities (Fig. 1). Moreover, the existing tentative representations, however well crafted, were ambiguous, and only suitable for specific needs, such as representing metabolic networks or signaling pathways or gene regulation.

The molecular biology era, and more recently the rise of genomics and other high-throughput technologies, have brought a staggering increase in data to be interpreted. It also favored the routine use of software to help formulate hypotheses, design experiments and interpret results. As a group of biochemists, modelers and computer scientists working in systems biology, we believe establishing standard graphical notations is an important step toward more efficient and accurate transmission of biological knowledge among our different communities. Toward this goal, we initiated the SBGN project in 2005, with the aim of developing and standardizing a systematic and unambiguous graphical notation for applications in molecular and systems biology.

Historical antecedents

Graphical representation of biochemical and cellular processes has been used in biochemical textbooks as far back as sixty years ago[3], reaching an apex in the wall charts hand drawn by Nicholson[4] and Michal[5]. Those graphs describe the processes that transform a set of inputs into a set of outputs, in effect being process, or state transition, diagrams. This style was emulated in the first database systems that depicted metabolic networks, including EMP[6], EcoCyc[7] and KEGG[8]. More notations have been 'defined' by virtue of their implementation in specialized software tools such as pathway and network designers (e.g., NetBuilder[9], Patika[10], JDesigner[11], CellDesigner[12]). Those

A list of affiliations appears at the end of the paper.

Published online 7 August 2009; Corrected after print 11 August 2009; doi:10.1038/nbt1558

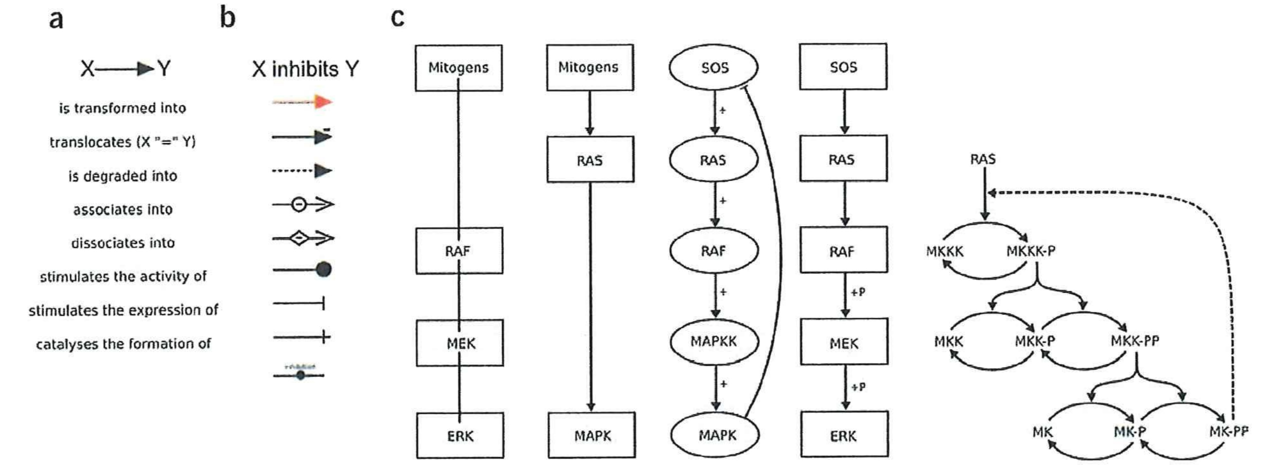

Figure 1 Inconsistency and ambiguity of current nonstandardized notations. (**a**) Eight different meanings associated with the same symbol in a chart describing the role of cyclin in cell regulations (http://www.abcam.com/ps/pdf/nuclearsignal/cell_cycle.pdf). (**b**) Nine different symbols found in the literature to represent the same meaning. (**c**) Five different representations of the MAP kinase cascade found in the scientific literature, depicting progressive levels of biological and biochemical knowledge. From left to right: relations[30], directionality of influence[31], directionality of effect[32], biochemical effect[33], chemical reactions[34]. In the last diagram, different instances of an identical arrowhead style represent catalysis, production and inhibition.

graphical notations were not standardized, and their understanding relied mainly on relating examples with one's preexisting knowledge of biochemical processes. Although the classical graphs adequately conveyed information about biochemistry, other types of diagrams were needed to represent signaling pathways, and incomplete or indirect information, as coming from molecular biology or genomics. Those conventions effectively mimicked the empirical notations used by biologists, describing either the relationships between elements[13,14] or the flow of activity or influence[15–17]. Lists of standard glyphs (**Box 1**) to represent identified concepts were then provided. The efforts to create rigidly defined schema were pioneered by Kurt Kohn with his Molecular Interaction Maps (MIM), which defined not only a set of symbols but also a syntax to describe interactions and relationships of molecules[18,19]. The MIM notation influenced other proposals[14]. Several proposals followed to describe process diagrams, not only with standard symbols but also defined grammars[20–23].

The SBGN project

Despite the popularity of some of the efforts mentioned above, none of the notations has acquired the status of a community standard. This can be attributed partly to the fact that the efforts only went as far as to propose notations, or implement them in software. Several of us have been involved in the development of the Systems Biology Markup Language (SBML)[24], from which we learned that establishing a standard is extremely difficult without an explicit, concerted, effort to engage a community and build a consensus among participants. We organized the SBGN project with this lesson in mind.

For SBGN to be successful, it must satisfy a majority of technical and practical needs and be embraced by a diverse community of biologists, biochemists, bioinformaticians, geneticists, theoreticians and software engineers. Early in the project's history, we established the following overarching principles to help steer SBGN toward those aims, ranked by rough hierarchical order of precedence.

The notation should

- be free of intellectual property restrictions to allow free use by the community;
- be syntactically and semantically consistent and unambiguous;
- support representation of diverse common biological objects, their properties and their interactions;
- keep the number of symbols and syntax to a minimum to help comprehension and learning by humans;
- be visually consistent and concise, using discriminable symbols;
- support modularity to help cope with diagram size and complexity;
- support the automated generation of diagrams by software starting from mathematical models.

Many of the design principles above resonate with research on visual languages[25,26] and studies aimed at understanding end-user needs in pathway visualization[27], although we derived them from our collective hands-on experiences with developing notations and software. In addition to these principles, we also sought to avoid many problems (**Table 1**) that affect some existing notations.

SBGN aims to specify the connectivity of the graphs and the types of the nodes and edges, but not the precise layout of the graphs. The semantics of an SBGN diagram does not depend on the relative position of the symbols. Furthermore, it does not depend on colors, patterns, shades, shapes and thickness of edges (**Fig. 2**). Similarly, the labels of symbols are not regulated and are only required to be unique within a map.

Finally, it was clear at the outset that it would be impossible to design a perfect and complete language from the beginning. Apart from the prescience this would require, it also would likely require a vast language that most newcomers would shun as being too complex. Thus, the SBGN community decided to stratify language development into levels. A level in SBGN represents a usable set of functionalities that the user community agrees is sufficient for a reasonable set of tasks and goals. Capabilities and features that cannot be agreed upon and are judged insufficiently critical to require inclusion in a given level are postponed to a higher level. In this way, SBGN development is envisioned to proceed in stages, with each higher SBGN level adding richness compared to the levels below it, while maintaining compatibility whenever possible. Furthermor only the actual usage of SBGN languages will tell us how well they work for the diverse communities involved, and this experience will certainly shape the evolution of the notation.

The three languages of SBGN

Molecular entities possess many properties that affect their interactions with other entities. Attempting to represent all the possible reactions and interactions in the same diagram is often futile, usually resulting in an incomprehensible jumble. The different styles of notations described above were attempts to control this complexity by presenting only what was needed in a specific context, or what was available through specific views of the system[14]. Each view focuses on only a portion of the semantics of the overall system, trading off diagram comprehensibility against completeness of biological knowledge.

SBGN follows this strategy and defines three orthogonal and complementary types of diagrams that can be seen as three alternative projections of the underlying more complex biological information. The process diagram draws its inspiration from process-style notations, borrowing ideas from the work of CellDesigner[28] and EPE[22]. By contrast, the entity relationship diagram is based to a large extent on Kohn's MIM notation[18,19]. The SBGN activity flow diagram depicts only the cascade of activity, thus making the notation similar to the reduced representations often used in the current literature to describe signaling pathways and gene regulatory networks. In **Figure 2**, we illustrate the three views applied to a very simple example. The characteristics of the SBGN languages are summarized in **Table 2**.

The idea of having three diagram types naturally begs the question of whether they could be merged into one, at least in paper form. The answer is no, for at least two reasons. First, a single diagram type would bring us back to the problem of dealing with unreasonable numbers of interactions as described above. Second, each SBGN language reflects fundamental differences in the underlying formal description of the phenomena. The meanings are so different that merging diagram types would compromise their representational robustness.

Having multiple visual languages is not uncommon in engineering (consider, for example, block diagrams and circuit diagrams in electronics, UML class, state sequence and deployment diagrams in software engineering), and this supports the idea that having three sublanguages in SBGN will be manageable in practice. In SBGN, the sharing of symbols representing identical concepts further reduces the differences between the three languages to differences in syntax and semantics. We believe that this, combined with careful design, will mitigate some of the difficulties of learning SBGN. However, it is to be noted that the clean orthogonality of the languages makes their overlap very limited, mostly to modulatory arcs, and node decorations.

Box 1 Glossary

SBGN diagrams are a specific set of graphs and thus make use of concepts from graph theory. The following list defines the terms used most often. We are aware of the unavoidable circularity of such definitions.

- **Arc.** A directed edge, that is, an edge that is not symmetrical in shape.
- **Edge.** A line joining two nodes.
- **Glyph.** A symbol that conveys information nonverbally.
- **Graph.** A set of nodes connected with edges.
- **Node.** A point that terminates a line or curve or comprises the intersection of two or more lines or curves.

SBGN process diagram

A process diagram represents all the molecular processes and interactions taking place between biochemical entities, and their results. This type of diagram depicts how entities transition from one form to another as a result of different influences; thus, it portrays the temporal qualities of molecular events occurring in biochemical reactions. In this way, the approach underlying process diagrams is the same as in the familiar textbook drawings of metabolic pathways. The main drawback of process diagrams is that a given entity must appear multiple times in the same diagram if it exists under several states; therefore, the notation is sensitive to the combinatorial explosion of possible entities and reactions, as is often the case in signaling pathways.

The SBGN process diagram level 1 specification defines six major classes of glyphs: entity pool nodes, process nodes, container nodes, reference nodes, connecting arcs and logical operators (**Supplementary Note 1**). In **Figure 3a**, we show a complete example of an SBGN process diagram. The number of symbols in level 1 of the SBGN process diagram notation has been purposefully limited so that they could be easily memorized. The notation may be enriched (perhaps using subclasses of symbols) in higher levels of SBGN.

Table 3 lists software projects that are already developing support for SBGN process diagram level 1 (see also **Supplementary Note 2**). Some of these rely on manual design of the pathways, whereas others, such as Arcadia, automatically generate SBGN PD from SBML models that have been annotated with terms from the Systems Biology Ontology[29]. The encoding of SBGN diagrams using computer-readable formats, a crucial step toward exchange and reuse of SBGN

Table 1 Features of *ad hoc* graphical notations, and the problems they create

Feature	Problem(s)
Different line thicknesses distinguish different types of processes or elements Dotted or dashed line styles distinguish different types of processes or elements	1. Rescaling a diagram can make line thicknesses and styles impossible to discern 2. Photocopying or faxing a diagram can cause differences in line thicknesses and styles to disappear 3. Differences in line thickness and style are difficult to make consistent in diagrams drawn by hand
Different colors distinguish different types of processes or elements	1. Photocopying or faxing a diagram will cause color differences to be indistinguishable 2. Color characteristics are difficult to achieve and keep consistent when drawing diagrams by hand
Identical line terminators (e.g., a single arrow) indicate different effects or processes depending on context	1. Greater ambiguity is introduced into a diagram 2. Interpreting a diagram requires more thought on the part of the reader 3. Automated verification of diagrams is more difficult due to lack of distinction between different processes or elements
Ad hoc symbols introduced at will by author	Interpreting a diagram requires the reader to search for additional information explaining the meaning of the symbols

diagrams, is currently supported in different formats such as SBML, GML and GraphML by different tools, and a general XML-based exchange format for SBGN is currently under discussion.

SBGN entity relationship diagram

The SBGN notation for entity relationships puts the emphasis on the influences that entities have upon each other's transformations rather than the transformations themselves. One can imagine that each of the relationships represents a specific conclusion of a scientific experiment or article. Their addition on a map represents the knowledge we have of the effects the entities have upon each other. Contrary to the process diagrams, where the different processes affect each other, the relationships are independent, and this independence is the key to avoiding the combinatorial explosion inherent to process diagrams. Unlike in process diagrams, a given entity may appear only once. Readers can better grasp at first sight all the possible influences and interactions affecting an entity, without having to explore the whole diagram to discover the different states an entity may be in, or to trace all the edges to find the relevant process nodes.

The relationship symbols in entity relationship diagrams support the representation of interactions and state variable assignments, thus allowing the notation to describe certain processes that cannot be expressed in process diagrams, such as allosteric modulation. In process diagrams, one can represent the formation of a ligand-receptor complex, but it is not possible to state that the complex is more active than the receptor alone without additional markup; entity relationships support this by allowing the interaction with the ligand to modulate the assignment of the variable representing the activity. The trade-off is that the temporal course is difficult to follow in entity relationships, because the sequence of events is not explicitly described (**Fig. 2a,b**).

The specification of SBGN entity relationship diagram level 1 defines three major classes of glyphs: entity nodes, statements and influences (logical operators are entity nodes). We summarize the symbols and the rules for their assembly (**Supplementary Note 3**). In **Fig. 3b**, we show a complete example of an SBGN entity relationship diagram.

SBGN activity flow diagram

A strategy often used for coping with biochemical network complexity or with incomplete or indirect knowledge is to selectively ignore the biochemical details of processes, instead representing the influences between entities directly. SBGN's activity flow diagrams permit modulatory arcs to directly link different activities, rather than entities and processes or relationships as described previously. Instead of displaying the details of biochemical reactions with process nodes and connecting arcs, the activity flow diagrams show only influences such as 'stimulation' and 'inhibition' between the activities displayed by the molecular entities (**Fig. 2c**). For example, a signal 'stimulates' the activity of a receptor, and this activity in turn 'stimulates' the activity of an intracellular transducing protein (note that activity flow retains the sequential chains of influences). Because most signaling pathway diagrams in the current literature are essentially activity flow diagrams, we expect many biologists will find this type of diagram familiar.

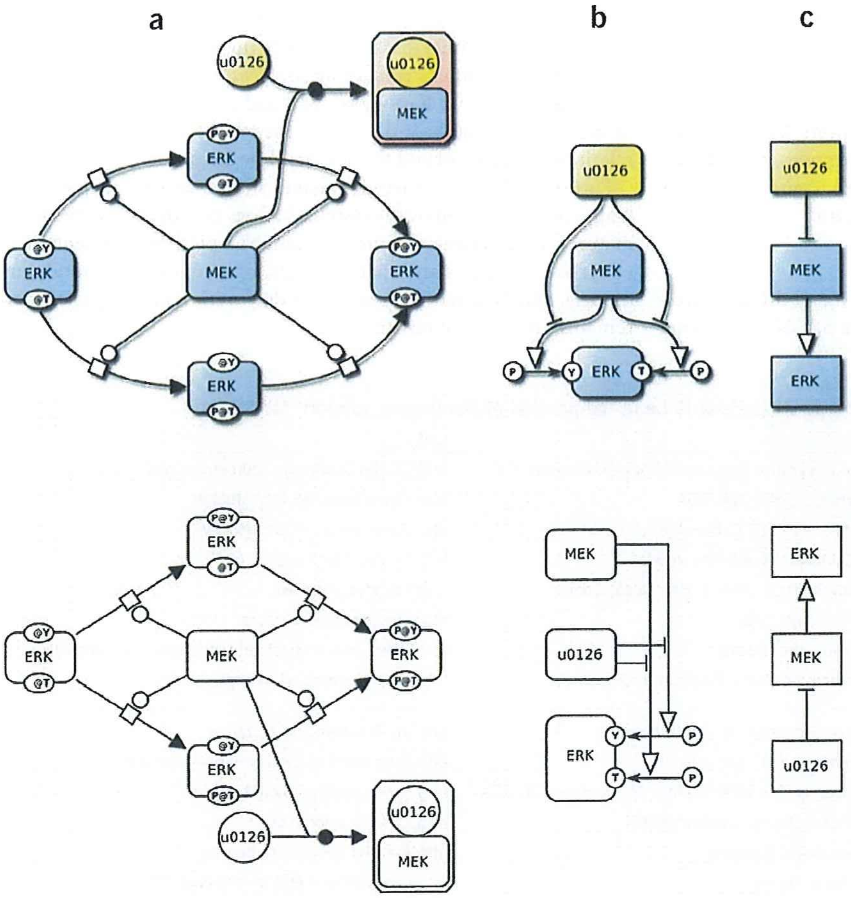

Figure 2 Simple example of protein phosphorylation catalyzed by an enzyme and modulated by an inhibitor. The semantics of an SBGN diagram does not depend on the relative position of the symbols, or on colors, patterns, shades, shapes and thickness of edges. Therefore, the upper and lower diagrams are identical as far as SBGN is concerned, and have to be interpreted exactly the same way. (**a**) Process diagrams, explicitly displaying the four forms of ERK, phosphorylated and nonphosphorylated on the tyrosine and the threonine, as well as the processes of phosphorylation by MEK and the inhibition of MEK by complexation with u0126. Note that the inhibition in this diagram emerges from the sequestration of MEK and is not explicitly represented. The phosphorylation sites are represented by variables, which in this example are labeled simply as 'Y' and 'T' (but in general could be anything desired by the diagram author), shown adorning the main symbols for ERK. (**b**) Entity relationship diagrams, showing ERK and the assignment of its phosphorylations (at the tyrosine and threonine residues), as well as the relationships between those and MEK and u0126. Note that ERK appears only once in this diagram; the different possible states are not explicitly depicted. (**c**) Activity flow diagrams depicting the activation of ERK by MEK and the inhibition of MEK by u0126. In this notation, only the relevant activities of u0126, MEK and ERK are represented, as well as abstract representations of the influences of activities upon each other, whereas the biochemical details are omitted.

Table 2 Comparison between the three languages of SBGN

	Process diagram	Entity relationship diagram	Activity flow diagram
Purpose	Represent processes that convert physical entities into other entities, change their states or change their location	Represent the interactions between entities and the rules that control them	Represent the influence of biological activities on each other
Building block	Different states of physical entities are represented separately	Physical entities are represented only once	Different activities of physical entities are represented separately
Ambiguity	Unambiguous transcription into biochemical events	Unambiguous transcription into biochemical events	Ambiguous interpretation in biochemical terms
Level of description	Mechanistic descriptions of processes	Mechanistic description of relationships	Conceptual description of influences
Temporality	Representation of sequential events	Absence of sequentiality between events	Representation of sequential influences
Pitfalls	Sensitive to combinatorial explosion of states and processes	Creation, destruction and translocation are not easily represented	Not suitable to represent association, dissociation, multistate entities
Advantages	The best for representing temporal/mechanistic aspects of processes such as metabolism	The best for representing signaling involving multistate entities	The best for functional genomics and signaling with simple activities

By ignoring processes and entity states, the number of nodes in an activity flow diagram is greatly reduced compared to an equivalent process diagram (**Fig. 2a,c**). Activity flow diagrams are also especially convenient for representing the effects of perturbations, whether genetic or environmental, because the complete mechanisms of the perturbations may not be known, or are irrelevant to the goals of a given study. The drawback is that activity flow diagrams may contain a high level of ambiguity. For instance, the biochemical basis of a positive or negative influence in a given system is left undefined. For this reason, this type of SBGN diagram should not exist alone; it should be associated, when possible, with detailed entity relationship and process diagrams, and used only for viewing purposes. We expect it will often be possible to generate activity flow diagrams mechanically from process diagrams and entity relationships, and have already performed preliminary work in that direction.

The SBGN activity flow diagram level 1 specification defines four major classes of glyphs: activity nodes, container nodes, modulating arcs and logical operators (**Supplementary Note 4**). **Figure 3c** shows a complete example of an SBGN activity flow diagram.

Participation and future prospects

The SBGN website (http://sbgn.org/) is a portal for all things related to SBGN. Interested persons can get involved in SBGN discussions by joining the SBGN discussion list (sbgn-discuss@sbgn.org). Face-to-face meetings of the SBGN community, generally held as satellite workshops of larger conferences, are announced on the website as well as the mailing list.

Standardizing a notation for depicting networks of biochemical interactions has so far remained an elusive goal, despite numerous but isolated efforts in that direction. Only with such a standardized notation will biologists, modelers and computer scientists be able to exchange accurate descriptions of complex systems—a task that continues to grow more demanding as our collective knowledge expands. SBGN blends many influences from past efforts, and also introduces many new ideas designed to overcome limitations of other notations.

Using a community-based approach involving many interested groups and individuals (including some who have been involved in previous efforts), we have developed and released the first version of the three languages of the SBGN, the process diagram, the entity relationship diagram and the activity flow diagram.

Future levels of the three languages should address major challenges currently faced by the systems biology community, as the field matures and diversifies. To cite but a few examples, the representation of spatial structures and spatial events, of composed and modular models, and of dynamic creation or destruction of compartments remains unchartered territory.

Table 3 List of software systems known to provide support, or to be in the process of developing support, for SBGN

Name	Organization	Link
Arcadia	Manchester Centre for Integrative Systems Biology, Manchester, UK	http://arcadiapathways.sourceforge.net/
Athena	University of Washington, Seattle, WA, USA	http://www.codeplex.com/athena/
BioModels Database	European Bioinformatics Institute, Cambridge, UK	http://www.ebi.ac.uk/biomodels/
BioUML	Institute of Systems Biology, Novosibirsk, Russia	http://www.biouml.org/
ByoDyn	Institut Municipal d'Investigació Mèdica, Barcelona, Spain	http://byodyn.imim.es/
CellDesigner	The Systems Biology Institute, Tokyo	http://www.celldesigner.org/
Dunnart	Monash University, Melbourne, Australia	http://www.csse.monash.edu.au/~mwybrow/dunnart/
Edinburgh Pathway Editor	Edinburgh Centre for Bioinformatics, Edinburgh, UK	http://www.pathwayeditor.org/
JWS Online	Stellenbosch University, Stellenbosch, South Africa	http://jjj.biochem.sun.ac.za/
NetBuilder	STRI, University of Hertfordshire, Hatfield, UK	http://strc.herts.ac.uk/bio/maria/Apostrophe/
PANTHER	Artificial Intelligence Center, SRI international, Menlo Park, CA, USA	http://www.pantherdb.org/pathway/
Reactome	European Bioinformatics Institute, Cambridge, UK	http://www.reactome.org/
Vanted	IPK Gatersleben, Gatersleben, Germany	http://vanted.ipk-gatersleben.de/
VISIOweb	Bilkent University, Ankara, Turkey	http://www.bilkent.edu.tr/~bcbi/pvs.html

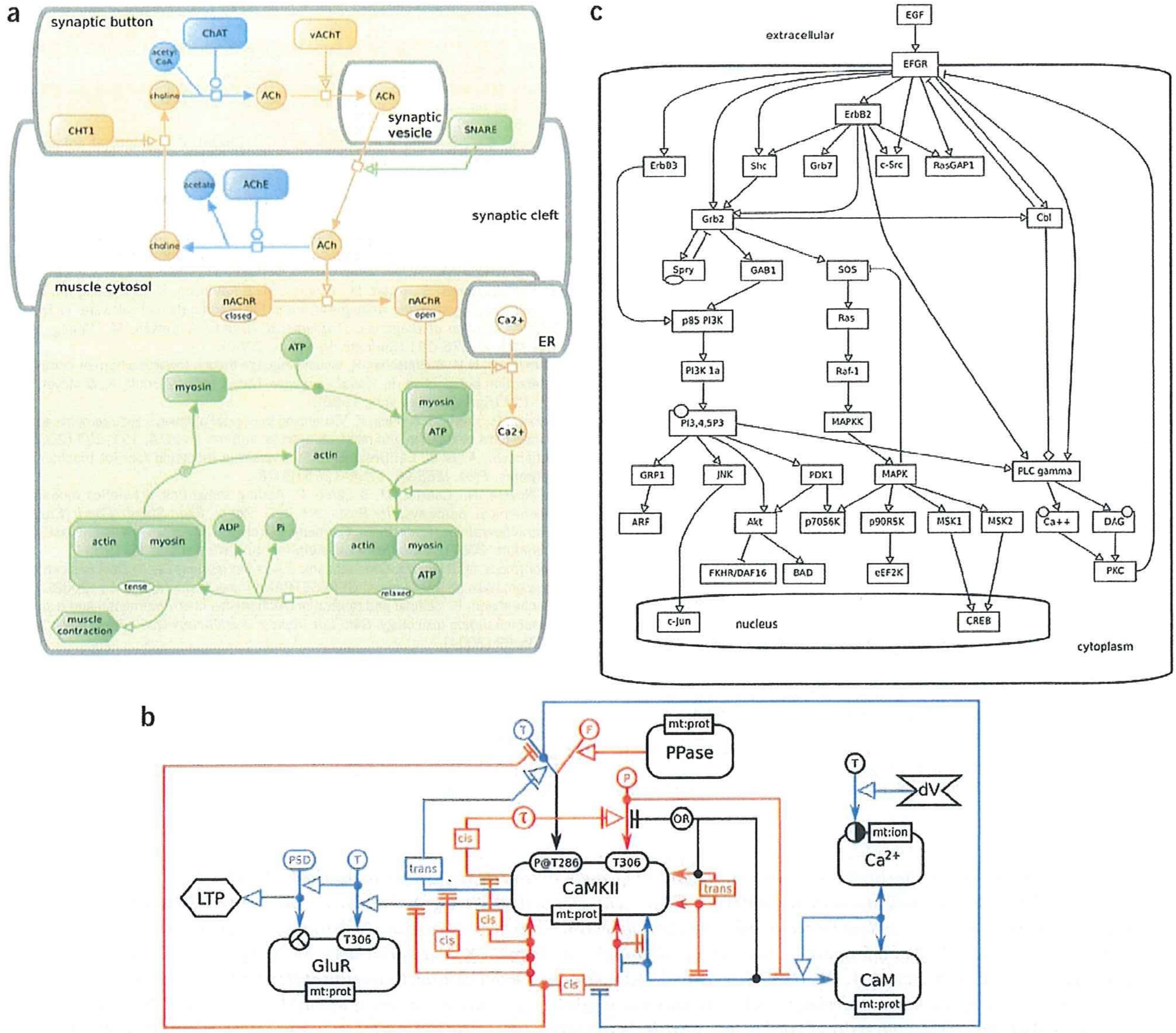

Figure 3 Example of complete SBGN diagrams. (**a**) Process diagram representing the synthesis of the neurotransmitter acetylcholine in the synaptic button of a nerve terminal, its release in the synaptic cleft, degradation in the synaptic cleft, the post-synaptic stimulation of its receptors and the subsequent effect on muscle contraction. Colors are used to enhance the biological semantics, blue representing catalytic reactions, orange for transport between compartments (including unrepresented ions, through channels) and green for the function of contractile proteins. However, it is important to note that those colors are not part of SBGN process diagram notation, and must not change the interpretation of the graph. (**b**) SBGN entity relationship diagram representing the transduction, by calcium/calmodulin kinase II, of the effect of voltage-induced increase of intracellular calcium onto the long-term potentiation (LTP) of the neuronal synapses, triggered by a translocation of glutamate receptors. The diagram describes the various relationships between the phosphorylations of the kinase monomers and their conformation. Colors highlight the direction of the relationships relative to the phenotype; blue relationships enhance LTP whereas red ones preclude this enhancement. (**c**) SBGN activity flow diagram representing the cascade of signals triggered by the epidermal growth factor, and going from the plasma membrane to the nucleus. The diagram is derived from reference 30.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

The development of SBGN was mainly supported by a grant from the New Energy and Industrial Technology Development Organization of the Japanese government. SBGN workshops also benefited from funding by the following organizations: the UK Biotechnology and Biological Sciences Research Council, the National Institute of Advanced Industrial Science and Technology of Japan, the Okinawa Institute of Science and Technology), the European Media Laboratory, Heidelberg, Germany and the Beckman Institute at the California Institute of Technology, Pasadena, California, USA. Attendance at the meetings by Japanese authors was supported by the Japan Science and Technology Agency and by the genome network project of the Japanese Ministry of Education, Culture, Sports, Science, and Technology. I.G., S.M. and A.S. acknowledge support by the British Engineering and Physical Sciences Research Council. F.T.B. acknowledges support by the National Institutes of Health (NIH; grant 1R01GM081070-01). The contributions of M.I.A., S.K., A.L. and K.K. were supported by the Intramural Research Program of the NIH, Center for Cancer Research, National Cancer Institute (NIH).

AUTHOR CONTRIBUTIONS

N.L.N., M.H., H.M., S.M., F.S. and A.S. contributed equally to the redaction of SBGN specifications.

Published online at http://www.nature.com/naturebiotechnology/.

Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Larkin, J.H. & Simon, H.A. Why a diagram is (sometimes) worth ten thousands words. *Cogn. Sci.* **11**, 65–100 (1987).
2. Lazebnik, Y. Can a biologist fix a radio?—Or,what I learned while studying apoptosis. *Cancer Cell* **2**, 179–182 (2002).
3. Gortner, R.A. *Outlines of Biochemistry.* (Wiley, New York, 1949).
4. Dagley, S. & Nicholson, D.E. *An Introduction to Metabolic Pathways* (Wiley, New York, 1970).
5. Michal, G. Biochemical Pathways (wall chart). (Boehringer Mannheim, Mannheim, Germany, 1984)
6. Sel'kov, E.E., Goryanin, I.I., Kaimatchnikov, N.P., Shevelev, E.L. & Yunus, I.A. Factographic data bank on enzymes and metabolic pathways. *Studia Biophysica* **129**, 155–164 (1989).
7. Karp, P.D. & Paley, S.M. Representations of metabolic knowledge: pathways. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 203–211 (1994).
8. Goto, S. *et al.* Organizing and computing metabolic pathway data in terms of binary relations. *Pac. Symp. Biocomput.* **PSB97**, 175–186 (1997).
9. Brown, C.T. *et al.* New computational approaches for analysis of cis-regulatory networks. *Dev. Biol.* **246**, 86–102 (2002).
10. Demir, E. *et al.* PATIKA: an integrated visual environment for collaborative construction and analysis of cellular pathways. *Bioinformatics* **18**, 996–1003 (2002).
11. Sauro, H.M. *et al.* Next generation simulation tools: the Systems Biology Workbench and BioSPICE integration. *OMICS* **7**, 355–372 (2003).
12. Funahashi, A., Morohashi, M., Kitano, H. & Tanimura, N. CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. *BIOSILICO* **1**, 159–162 (2003).
13. Kohn, K.W. Functional capabilities of molecular network components controlling the mammalian G1/S cell cycle phase transition. *Oncogene* **16**, 1065–1075 (1998).
14. Kitano, H. A graphical notation for biochemical networks. *BIOSILICO* **1**, 169–176 (2003).
15. Pirson, I. *et al.* The visual display of regulatory information and networks. *Trends Cell Biol.* **10**, 404–408 (2000).
16. Cook, D.L., Farley, J.F. & Tapscott, S.J. A basis for a visual language for describing, archiving and analyzing functional models of complex biological systems. *Genome Biol.* **2**, research0012.1–0012.10 (2001).
17. Longabaugh, W.J.R., Davidson, E.H. & Bolouri, H. Visualization, documentation, analysis, and communication of large-scale gene regulatory networks. *Biochim. Biophys. Acta* **1789**, 363–374 (2009).
18. Kohn, K.W. Molecular interaction map of the mammalian cell cycle control and DNA repair systems. *Mol. Biol. Cell* **10**, 2703–2734 (1999).
19. Kohn, K.W., Aladjem, M.I., Weinstein, J.N. & Pommier, Y. Molecular interaction maps of bioregulatory networks: a general rubric for systems biology. *Mol. Biol. Cell* **17**, 1–13 (2006).
20. Demir, E. *et al.* An ontology for collaborative construction and analysis of cellular pathways. *Bioinformatics* **20**, 349–356 (2004).
21. Kitano, H., Funahashi, A., Matsuoka, Y. & Oda, K. Using process diagrams for the graphical representation of biological networks. *Nat. Biotechnol.* **23**, 961–966 (2005).
22. Moodie, S.L., Sorokin, A.A., Goryanin, I.I. & Ghazal, P. A graphical notation to describe the logical interactions of biological pathways. *J. Integr. Bioinform.* [Au: Correct publication is only one page?] **3**, 36–46 (2006).
23. Raza, S. A logic-based diagram of signalling pathways central to macrophage activation. *BMC Syst. Biol.* **2**, 36 (2008).
24. Hucka, M. *et al.* The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19**, 524–531 (2003).
25. Britton, C., Jones, S., Kutar, M., Loomes, M. & Robinson, B. Evaluating the intelligibility of diagrammatic languages used in the specification of software. in *Theory and Application of Diagrams: Diagrams 2000* (eds. Anderson, M., Cheng, P. & Haarslev, V.) 376–391 (Springer, New York, 2000).
26. Narayanan, N.H. & Hübscher, R. Visual language theory: towards a human-computer interaction perspective. in *Visual Language Theory* (eds. Marriott, K. & Meyer, B.) 85–127 (Springer, New York, 1998).
27. Saraiya, P., North, C. & Duca, K. Visualizing biological pathways: requirements analysis, systems evaluation and research agenda. *Inform. Visual* **4**, 191–205 (2005).
28. Funahashi, A. *et al.* CellDesigner 3.5: a versatile modeling tool for biochemical networks. *Proc. IEEE* **96**, 1254–1265 (2008).
29. Le Novère, N., Courtot, M. & Laibe, C. Adding semantics in kinetics models of biochemical pathways. in *Proc. 2nd Intl. Symp. Exp. Stand. Cond. Enzyme Characterizations*, Rüdesheim, Germany, March 19–23, 2006 (Beilstein Institute, Frankfurt, 2007). <http://www.beilstein-institut.de/index.php?id=196>
30. Anonymous. MAP kinases. Gene set bank. *Riken BioResource Center DNA Bank* <http://www.brc.riken.go.jp/lab/dna/en/GENESETBANK/index.html> (August 19, 2008).
31. Riechelmann, H. Cellular and molecular mechanisms in environmental and occupational inhalation toxicology. *GMS Cur. Topics. Otorhinolaryngol.—Head Neck Surg.* **3**, Doc02 (2004).
32. Schlessinger, J. Epidermal growth factor receptor pathway. *Sci. Signal.* (connections map in the database of cell signaling, as seen May 29, 2009). <http://stke.sciencemag.org/cgi/cm/stkecm;CMP_14987>
33. Anonymous. MAPK signaling pathway, *Homo sapiens. KEGG Pathway hsa04010* <http://www.genome.jp/kegg/pathway/hsa/hsa04010.html> (July 15, 2009).
34. Kholodenko, B. Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. *Eur. J. Biochem.* **267**, 1583–1588 (2000).

[1]EMBL European Bioinformatics Institute, Hinxton, UK. [2]Engineering and Applied Science, California Institute of Technology, Pasadena, California, USA. [3]SRI International, Menlo Park, California, USA. [4]Centre for Systems Biology at Edinburgh, University of Edinburgh, Edinburgh, UK. [5]Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany. [6]Institute of Computer Science, University of Halle, Halle, Germany. [7]School of Informatics, University of Edinburgh, Edinburgh, UK. [8]Memorial Sloan Kettering Cancer Center - Computational Biology Center, New York, NY, USA. [9]Science and Technology Research Institute, University of Hertfordshire, Hatfield, UK. [10]National Cancer Institute, Bethesda, Maryland, USA. [11]Auckland Bioengineering Institute, University of Auckland, Auckland, New Zealand. [12]Department of Bioengineering, University of Washington, Seattle, Washington, USA. [13]BIOQUANT, University of Heidelberg, Heidelberg, Germany. [14]Division of Pathway Medicine, University of Edinburgh Medical School, Edinburgh, UK. [15]Riken OMICS Science Center, Yokohama City, Kanagawa, Japan. [16]The Systems Biology Institute, Tokyo, Japan. [17]School of Computer Science, University of Manchester, Manchester, UK. [18]Manchester Interdisciplinary Biocentre, Manchester, UK. [19]Clayton School of Information Technology, Faculty of Information Technology, Monash University, Melbourne, Victoria, Australia. [20]U900 INSERM, Paris Mines Tech, Institut Curie, Paris, France. [21]Terry Fox Laboratory, British Columbia Cancer Research Center, Vancouver, British Columbia, Canada. [22]Bilkent Center for Bioinformatics, Bilkent University, Ankara, Turkey. [23]The Roslin Institute, University of Edinburgh, Midlothian, UK. [24]Department of Biosciences and Informatics, Keio University, Hiyoshi, Kouhoku-ku, Yokohama, Japan. [25]Institute of Systems Biology, Novosibirsk, Russia. [26]Design Technological Institute of Digital Techniques SB RAS, Novosibirsk, Russia. [27]Ontario Institute for Cancer Research, Toronto, Ontario, Canada. [28]School of Chemistry, University of Manchester, Manchester, UK. [29]Department of Biochemistry, Stellenbosch University, Matieland, South Africa. [30]Sony Computer Science Laboratories, Tokyo, Japan. [31]Okinawa Institute of Science and Technology, Okinawa, Japan. Correspondence should be addressed to N.L.N. (lenov@ebi.ac.uk).

Corrigendum: Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins

Bernd Wollscheid, Damaris Bausch-Fluck, Christine Henderson, Robert O'Brien, Miriam Bibel, Ralph Schiess, Ruedi Aebersold & Julian D Watts
Nat. Biotechnol. 27, 378–386 (2009); published online 5 April 2009; corrected after print 9 September 2009

In the version of this article initially published, in Methods, p. 385, line 5, the concentration of $MgCl_2$, given as 0.5 M, is incorrect. The correct concentration is 0.5 mM $MgCl_2$. The error has been corrected in the HTML and PDF versions of the article.

Corrigendum: Multi-site assessment of the precision and reproducibility of multiple reaction monitoring–based measurements of proteins in plasma

Terri A Addona, Susan E Abbatiello, Birgit Schilling, Steven J Skates, D R Mani, David M Bunk, Clifford H Spiegelman, Lisa J Zimmerman, Amy-Joan L Ham, Hasmik Keshishian, Steven C Hall, Simon Allen, Ronald K Blackman, Christoph H Borchers, Charles Buck, Helene L Cardasis, Michael P Cusack, Nathan G Dodder, Bradford W Gibson, Jason M Held, Tara Hiltke, Angela Jackson, Eric B Johansen, Christopher R Kinsinger, Jing Li, Mehdi Mesri, Thomas A Neubert, Richard K Niles, Trenton C Pulsipher, David Ransohoff, Henry Rodriguez, Paul A Rudnick, Derek Smith, David L Tabb, Tony J Tegeler, Asokan M Variyath, Lorenzo J Vega-Montoto, Åsa Wahlander, Sofia Waldemarson, Mu Wang, Jeffrey R Whiteaker, Lei Zhao, N Leigh Anderson, Susan J Fisher, Daniel C Liebler, Amanda G Paulovich, Fred E Regnier, Paul Tempst & Steven A Carr
Nat. Biotechnol. 27, 633–641 (2009); published online 28 June 2009; corrected after print 9 September 2009

In the version of this article initially published, the following acknowledgment was inadvertently left out: "The UCSF CPTAC team gratefully acknowledges the support of the Canary Foundation for providing funds to purchase a 4000 QTRAP mass spectrometer." The acknowlegment has been added to the HTML and PDF versions of the article.

Erratum: Synergistic drug combinations tend to improve therapeutically relevant selectivity

Joseph Lehár, Andrew S Krueger, William Avery, Adrian M Heilbut, Lisa M Johansen, E Roydon Price, Richard J Rickles, Glenn F Short III, Jane E Staunton, Xiaowei Jin, Margaret S Lee, Grant R Zimmermann & Alexis A Borisy
Nat. Biotechnol. 7, 659–666 (2009); published online 5 July 2009; corrected after print 8 July 2009

In the version of this article initially published, in the legend of Figure 5b, line 2, "stress" is followed by a period. The period should be a comma, so that the sentence reads, "In response to stress, lymphoctyes…." The error has been corrected in the HTML and PDF versions of the article.

Erratum: The Systems Biology Graphical Notation

Nicolas Le Novère, Michael Hucka, Huaiyu Mi, Stuart Moodie, Falk Schreiber, Anatoly Sorokin, Emek Demir, Katja Wegner, Mirit I Aladjem, Sarala M Wimalaratne, Frank T Bergman, Ralph Gauges, Peter Ghazal, Hideya Kawaji, Lu Li, Yukiko Matsuoka, Alice Villéger, Sarah E Boyd, Laurence Calzone, Melanie Courtot, Ugur Dogrusoz, Tom C Freeman, Akira Funahashi, Samik Ghosh, Akiya Jouraku, Sohyoung Kim, Fedor Kolpakov, Augustin Luna, Sven Sahle, Esther Schmidt, Steven Watterson, Guanming Wu, Igor Goryanin, Douglas B Kell, Chris Sander, Herbert Sauro, Jacky L Snoep, Kurt Kohn & Hiroaki Kitano
Nat. Biotechnol. 27, 735–741(2009); published online 7 August 2009; corrected after print 11 August 2009

In the version of this article initially published, the wrong versions of Figures 1, 2 and 3 were used. The error has been corrected in the HTML and PDF versions of the article.

Erratum: Table of Contents

Nat. Biotechnol. 27, i (2009); published online 7 August 2009; corrected after print 7 August 2009

In the PDF version of the table of contents initially published, a news article titled "Genzyme's Lumizyme clears bioequivalence hurdles" was omitted. The error has been corrected in the PDF version of the table of contents.

Molecular Systems Biology 5; Article number 281; doi:10.1038/msb.2009.36
Citation: ***Molecular Systems Biology*** **5:281**
© 2009 EMBO and Macmillan Publishers Limited All rights reserved 1744-4292/09
www.molecularsystemsbiology.com

molecular systems biology

REPORT

Robustness and fragility in the yeast high osmolarity glycerol (HOG) signal-transduction pathway

Marcus Krantz[1,2,*], Doryaneh Ahmadpour[1,6], Lars-Göran Ottosson[1,6], Jonas Warringer[1], Christian Waltermann[3], Bodil Nordlander[1], Edda Klipp[3], Anders Blomberg[1], Stefan Hohmann[1] and Hiroaki Kitano[2,4,5]

[1] Department of Cell and Molecular Biology, University of Gothenburg, Göteborg, Sweden, [2] The Systems Biology Institute, Tokyo, Japan, [3] Theoretical Biophysics, Institute of Biology, Humboldt University, Berlin, Germany, [4] Sony Computer Science Laboratories, Tokyo, Japan and [5] Okinawa Institute of Science and Technology, Okinawa, Japan
[6] These authors contributed equally to this work
[*] Corresponding author. Cell and Molecular Biology, University of Gothenburg, Box 462, SE-40530 Gothenburg, Sweden. Tel.: +467 364 547 67; Fax: +463 178 625 99; E-mail: marcus.krantz@cmb.gu.se

Received 28.5.08; accepted 13.5.09

Cellular signalling networks integrate environmental stimuli with the information on cellular status. These networks must be robust against stochastic fluctuations in stimuli as well as in the amounts of signalling components. Here, we challenge the yeast HOG signal-transduction pathway with systematic perturbations in components' expression levels under various external conditions in search for nodes of fragility. We observe a substantially higher frequency of fragile nodes in this signal-transduction pathway than that has been observed for other cellular processes. These fragilities disperse without any clear pattern over biochemical functions or location in pathway topology and they are largely independent of pathway activation by external stimuli. However, the strongest toxicities are caused by pathway hyperactivation. *In silico* analysis highlights the impact of model structure on *in silico* robustness, and suggests complex formation and scaffolding as important contributors to the observed fragility patterns. Thus, *in vivo* robustness data can be used to discriminate and improve mathematical models.

Molecular Systems Biology 5: 281; published online 16 June 2009; doi:10.1038/msb.2009.36
Subject Categories: metabolic & regulatory networks; signal transduction
Keywords: gTow; HOG; robustness; signal transduction; systems biology

Introduction

Robustness is an intrinsic feature of life as all cellular systems have to maintain functionality in the face of naturally occurring external and internal fluctuations. The resilience of cellular genetic networks lets the cell tolerate a certain level of environmental or mutational perturbations. This robustness can be achieved either by maintaining the cellular status stable against various fluctuations, or by adapting to external changes by triggering a series of cellular responses (Kitano, 2004; Stelling *et al*, 2004). The decision to respond and adapt is relayed via signal transduction systems, which, upon activation by specific stimuli, produce distinct regulatory signals in the form of changes in levels of activated signal-pathway components. A critical aspect of such processes is distinguishing a genuine signal from stochastic fluctuations in protein levels and activity, as misinterpretation of these has potentially disastrous fitness consequences. Thus, the robustness of cells to maintain such a function despite variations in dosage of the components is of primary importance for survival. Despite their importance for viability and fitness, little is known about how signalling systems distinguish between signals and natural fluctuations, or to what extent such fluctuations are tolerated.

Here, we approach this issue through a system-wide robustness study of the HOG pathway of *Saccharomyces cerevisiae*, which is one of the most extensively studied eukaryotic signal-transduction cascades. It is activated by high osmolarity and is essential under this condition. The signalling pathway, which is depicted in Figure 1, consists of a MAP kinase (MAPK) core module, upon which two independent upstream branches converge. The first of these consists of

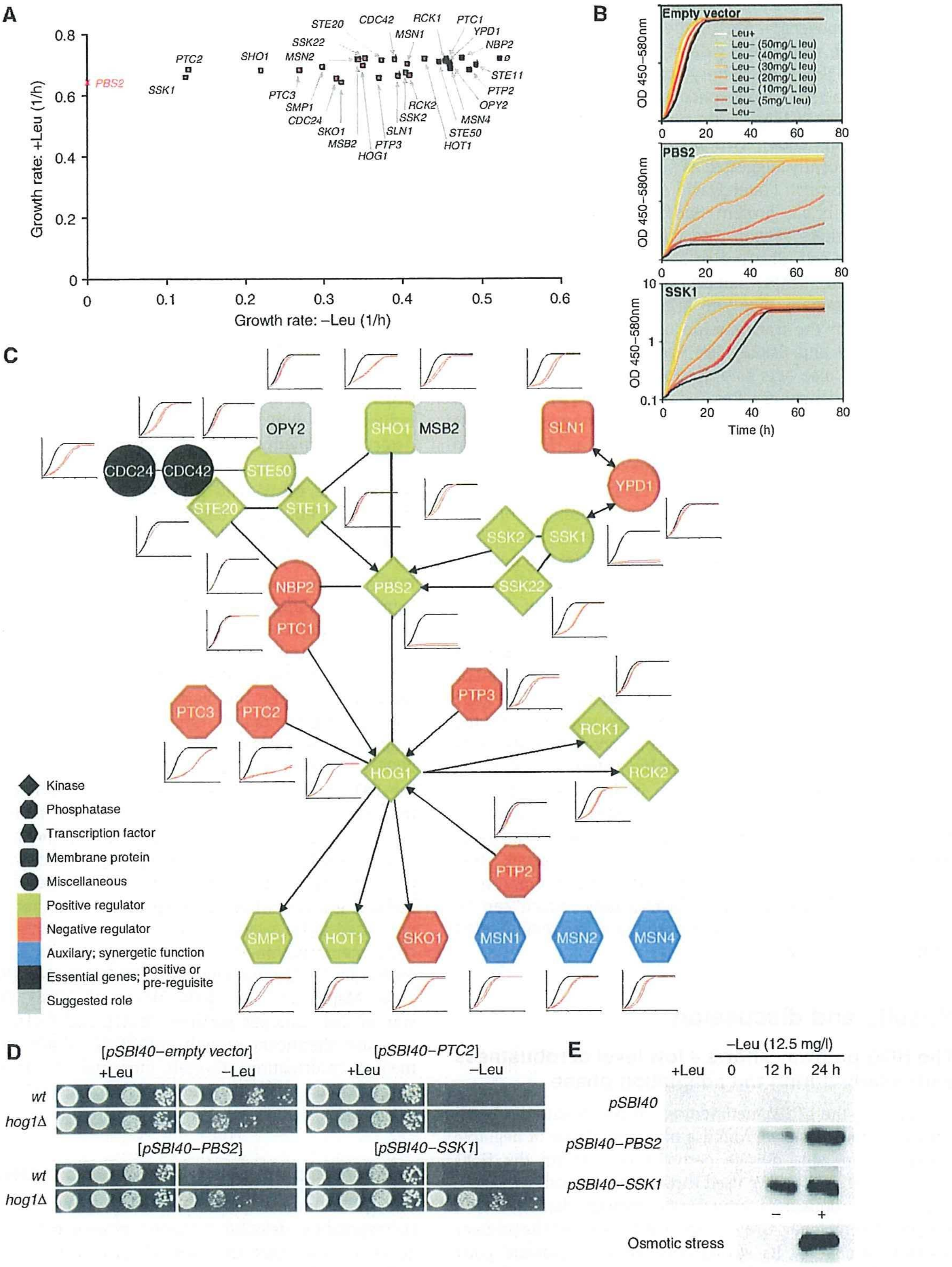

Figure 1 **(A)** gToW growth phenotypes occur in the absence of leucine, and red squares indicate a significant growth-rate defect as compared with the empty plasmid control (Ø). *PBS2*s growth rate could only be determined in the presence of leucine. **(B)** The severity of the growth defect increases with the level of leucine starvation and **(C)** they spread over different pathway functions. Graphs in (C) indicate growth with (black) or without (red) leucine. **(D)** Phenotypes caused by *PBS2* and *SSK1*, unlike *PTC2*, are partially suppressed by the deletion of *HOG1*. **(E)** Overexpression of Pbs2p and Ssk1p causes dual phosphorylation of Hog1p, which after leucine limitation (12.5 mg/l), reaches levels comparable to those caused by osmotic stress (+) within 24 h. The empty plasmid control (pSBI40) remains similar to unstressed cells (−).

a phosphotransfer module, including the histidine kinase and presumed osmosensor Sln1p, the phosphotransfer protein Ypd1p and the response regulator Ssk1p. When active, this module keeps Ssk1p phosphorylated and inactive. When the module is inactive, dephosphorylated Ssk1p binds to and activates the MAPK kinase kinases Ssk2p and Ssk22p (Saito and Tatebayashi, 2004). Two mucin-like proteins Msb2p and Hkr1p were recently suggested as putative osmosensors of the second input branch (Tatebayashi *et al*, 2007). On activation, the transmembrane protein Sho1p is believed to receive signals from these sensors and convey these signals to the interior of the cell. Sho1p also assembles the MAPK kinase Pbs2p and the MAPK kinase kinase Ste11p through its cytoplasmic domain. The Sho branch also requires Cdc42p and Ste20p for the transmission of the signal. Once active, Ste11p, Ssk2p and Ssk22p are each able to phosphorylate Pbs2p, which also acts as a scaffold for the MAPK module. Pbs2p in turn phosphorylates and activates the MAPK Hog1p, which has numerous targets, including the cytoplasmic kinase Rck2p and several transcription factors such as Hot1p, Sko1p and Smp1p. Active Hog1p accumulates in the nucleus and partakes in transcription (Hohmann, 2002).

The methodology used here links the expression of each target gene of interest to that of a defective allele of a metabolic gene. This allele, *leu2-d*, has a defective promoter and needs to be present in a high copy number to support high Leu2p levels and thus unperturbed growth in the absence of leucine (Schneider and Guarente, 1991). By placing a target gene of interested on the same episomal plasmid as *leu2-d*, the copy number and thus the expression level of this target gene can be controlled via the leucine concentration in the media. The gene in question is still under the control of its normal promoter, allowing expression that is regulated but increases proportionally with the increase in copy number (Moriya *et al*, 2006; Torres *et al*, 2007). If the increase in target gene product interferes with cellular function, a negative pressure on plasmid copy number will balance the positive pressure conferred by the metabolic gene, resulting in a genetic tug-of-war (gToW; Supplementary Figure S1). Such a compromise will result in a decrease in cellular fitness, which can be measured precisely using high-resolution microcultivation (Warringer *et al*, 2003).

Results and discussion

The HOG pathway shows a low level of robustness particularly during the adaptation phase

On applying the gToW methodology to investigate the HOG-pathway robustness, we found a high prevalence of negative impacts from gene dosage perturbations within the HOG pathway (Figure 1A). We used three physiological 'windows' to assess the robustness; growth rate, growth adaptation time and growth efficiency. Growth rate was considered the primary readout because of its strong correlation to plasmid copy number (Moriya *et al*, 2006). Adaptation defects (prolonged growth lag) turned out to be similar to growth-rate defects, although even more pronounced and frequent (Supplementary Figure S2a). There was almost no effect on growth efficiency (Supplementary Figure S2b). Altogether, overexpression of 22 out of 29 HOG-pathway components caused a significant defect in at least one of these three growth variables ($P<0.001$, see Supplementary information). These phenotypes were strongly linked to leucine starvation and thus to selective pressure for high plasmid copy numbers (Figure 1B). They were also observed for all protein classes, and for both positive and negative regulators of the pathway (Figure 1C). The negative effects from increases in gene dosage were emphasized by a complete lack of positive fitness effects mediated by any of the gToW constructs. The highest frequency and strength of defects were observed during adaptation, which may reflect the delicate balance of signal transmission in initiating proliferation.

The high frequency (76%) of HOG-pathway gToW-imposed defects stands in stark contrast to the cell-cycle system, for which only 25–30% of the gToW constructs caused a clear growth retardation (Moriya *et al*, 2006), and to a global *GAL1* promoter-driven overexpression study using galactose induction in which a mere 15% of the targets conferred detectable growth defects (Sopko *et al*, 2006). Interestingly, little correlation in terms of cellular toxicity was found between the gToW- and galactose-driven overexpressions of the same, but GST tagged proteins, even on galactose (Supplementary Figure S3). The sole exception was Ssk1p that scored as highly toxic with both methods. The lack of correlation may be explained by the varying absolute levels of overexpressions (Supplementary Figure S3e) in the different screens and by the influence of the GST tag on protein function.

The high prevalence of nodes of fragility within the HOG pathway may be partly explained by the very nature of signal transduction, as overexpression defects are enriched among components that transduce adaptation signals, that is kinases, phosphatases and transcription factors (Sopko *et al*, 2006). However, the high fraction of such components in the HOG pathway cannot be the sole reason for the high frequency of fitness defects, as the system is also sensitive to overexpression of more than half of the components, which do not partake in (de)phosphorylation or transcription. In fact, each such component that does not impair normal growth when overexpressed is either known or presumed to be a targeting or activating partner of catalytic components within the pathway, that is Nbp2p for Ptc1p, Ste50p for Ste11p and Opy2p for Ste50p/Ste11p (Posas *et al*, 1998; Mapes and Ota, 2004; Wu *et al*, 2006). Overexpression of their catalytic partners, Ste11p and Ptc1p, also failed to cause significant growth defects ($P>0.001$; see Supplementary information). Overall, these paired components are more robust against overexpression ($P=0.022$, Fisher's exact test).

Robustness analysis at the genetic extremes

The gToW targets' phenotypes vary much more than the corresponding deletion mutants' phenotypes. Four of the genes included here are essential (*CDC24*, *CDC42*, *SLN1* and *YPD1*), but the twenty-five viable deletion strains show at the most mild growth defects during normal conditions. Only in the presence of NaCl stress does deletion of those genes cause strong but viable phenotypes. However, these deletion phenotypes have little in common with those caused by overexpression (Supplementary Figure S4). The strongest gToW phenotype is caused by Pbs2p, which is well known to be highly important for growth on salt as well as severely toxic when overexpressed. Hog1p is likewise important for osmotic tolerance but, unlike Pbs2p, not toxic when overexpressed. In addition, Ssk1p is severely toxic when overexpressed, but dispensable for osmotic tolerance. Although their roles in the osmotic stress response are well known, the toxicity mechanisms of Ssk1p and Pbs2p overexpression remain to be mechanistically resolved. However, both activate Hog1p constitutively and are suppressed by the deletion of *HOG1*, indicating that most, if not all, of the toxicity stems from the downstream pathway hyperactivation (Figure 1D and E).

Understanding toxicity mechanisms

In order to gain more insight into the mechanisms of toxicity of the two main nodes of fragility, Ssk1p and Pbs2p, we compared the observed *in vivo* sensitivity profiles with the *in silico* sensitivities with respect to nuclear, dually phosphorylated Hog1p predicted by the Hog model by Klipp *et al* (2005) (Figure 2A). As the *in vivo* differences in fragility between Ssk1p and Ssk2p and the fragility of the Pbs2p node cannot be captured using the original model, we studied seven variants with alternative motifs of regulation involving Ssk1p and Pbs2p (Supplementary Figure S5) and scored the relative improvements of each in the light of our data on overexpression (Figure 2C). We found that the *in silico* sensitivity

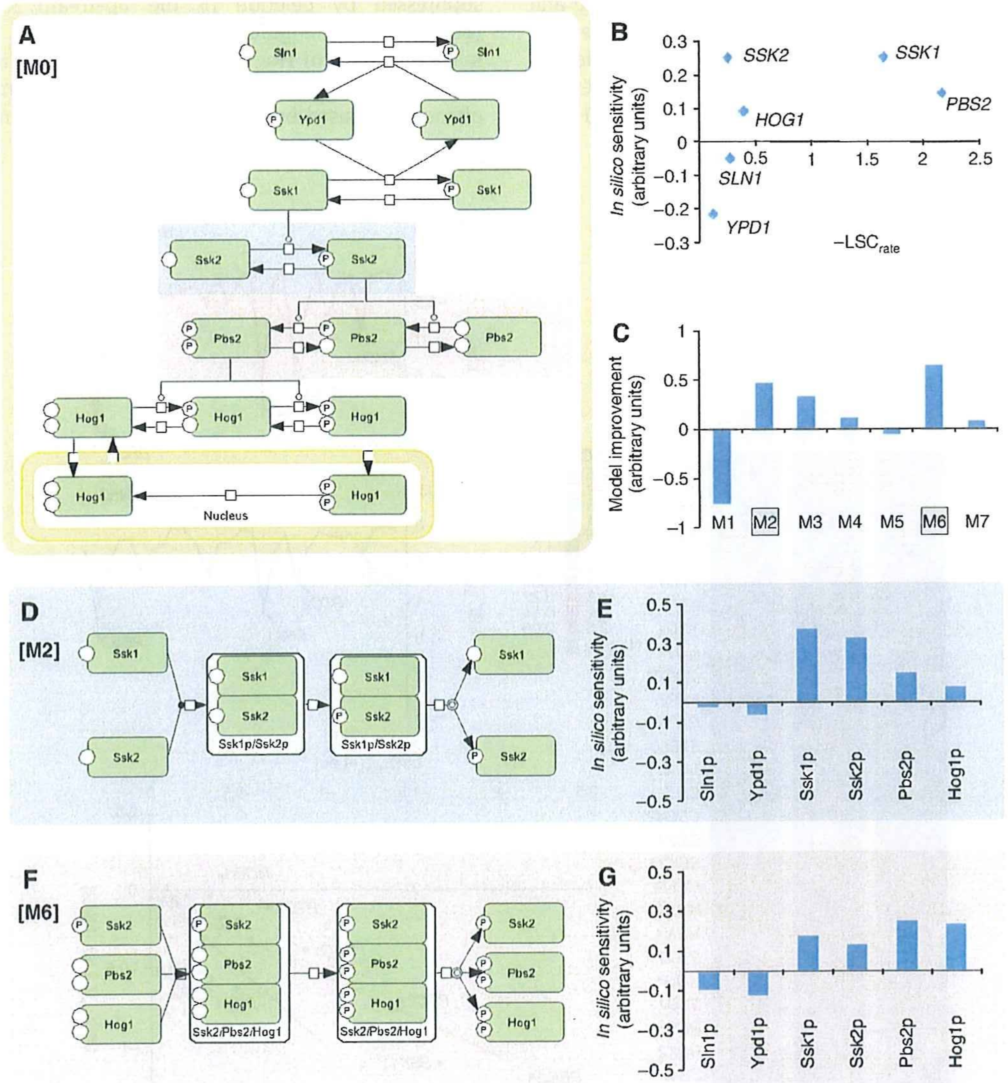

Figure 2 (**A**) The Hog part of the mathematical osmoregulatory model by Klipp *et al* (2005) (**B**) *In vivo* growth-rate defects in leucine-free medium are compared with *in silico* increases in basal levels of nuclear, dually phosphorylated Hog1p as a result of gene overexpression. The model does not capture the fragility of the Pbs2p node or distinguish the sensitivities of *SSK1* and *SSK2*. (**C**) The relative improvement in model performance by the inclusion of regulatory motifs around Ssk1p (M1–M3) and Pbs2p (M4–M7) (Supplementary Figure S5). (**D**) Requirement of dimerization before the phosphorylation of Ssk2p yielded the best improvements for Ssk1p (**E**; the corresponding sensitivity profile). (**F**) Explicit modelling of the scaffold function yields the best improvements for Pbs2p (**G**; the corresponding sensitivity profile). The LSC scores the growth difference compared with wild type, with a negative value, indicating a growth defect.

of Ssk1p is enhanced most when the dimerization of Ssk1p with Ssk2p is required for the phosphorylation and activation of Ssk2p (Figure 2D and E). In addition, explicit modelling of Pbs2p's function as a scaffold best improves its performance regarding the fragility of the Pbs2p node (Figure 2F and G). However, we observed no improvement of the *in silico* robustness through the implementation of the known dimerization of Ssk1p alone (Supplementary Figure S5; M1), suggesting that it is unlikely to contribute to the robustness pattern.

If *in vivo* toxicity stemmed from the indiscriminate interaction between protein pairs, we would expect the effect of overexpression to be roughly symmetrical for transient interactions or biased towards the component with lower expression levels in case of sustained interactions. Here, we see neither. Ssk1p has both a much stronger phenotype and higher basal expression level than Ssk2p (Supplementary Figure S6a). Ssk22p is even less abundant and the effect of Ssk1p overexpression is suppressed in *ssk2*Δ (Supplementary Figure S6d). As the phosphorylated, inactive state of Ssk1p has been reported to be stabilized by Ypd1p (Janiak-Spens *et al*, 2000), and the gToW overexpression brings Ssk1p into parity with Ypd1p levels (Supplementary Figure S6b), it may be the depletion of the stabilizing Ypd1p that leads to an accumulation of dephosphorylated and active Ssk1p (Supplementary Figure S7). Consistently, deletion of either *SLN1* or *YPD1* is lethal owing to the resulting constitutive activity of Ssk1p/Ssk2p and the HOG pathway. Pbs2p likewise stands out as being much more sensitive than its neighbours. Although this toxicity may stem from a disrupted balance with negative regulators, such as the Nbp2p–Ptc1p phosphatase complex, the high basal abundance of Pbs2p argues against the depletion of Nbp2–Ptc1 as the sole source of toxicity (Supplementary Figure S6a). However, the toxicity stems from the amplification of an existing residual signal, as it can be suppressed by deletion of the upstream kinase Ssk2p (Supplementary Figure S6d). As for Ssk1p, it leads to hyperactivation of Hog1p (Figure 1E) and this is the source of its toxicity (Figure 1D). In contrast to Ssk1p and Pbs2p, the phenotype caused by the phosphatase Ptc2p is not mediated

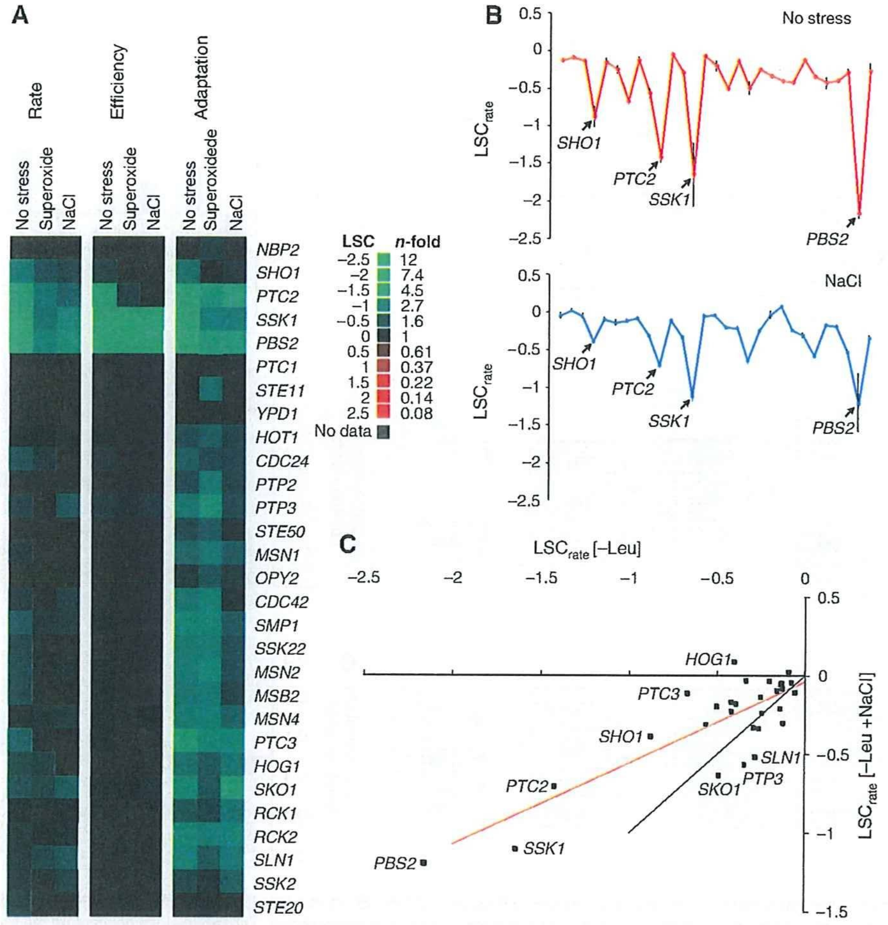

Figure 3 (**A**) Hierarchical clustering of the growth-phenotype profiles in the presence or absence of environmental perturbations. The phenotypic effect is indicated by colour. (**B**) Experimentally measured growth-rate-toxicity profiles in the presence or absence of an external pathway activator (LSC_{rate} ± s.d., n=2; gene order as in Supplementary Table SI). (**C**) There is a strong correlation between the phenotypic effect under the different conditions, although the relative gToW effect is milder under adverse growth conditions—NaCl (r^2=0.57, k=0.5 (red line; the black line indicates 1:1 correlation)).

through the activation of the HOG pathway and cannot be suppressed by the deletion of *HOG1* (Figure 1D), as would be expected because its overexpression phenotype is stronger than the deletion phenotype of Hog1p (Supplementary Figure S4b). Instead, the mechanism of its toxicity should be found outside the context of the HOG pathway.

The robustness of the HOG pathway is partly dependent on the environmental stress

To determine whether the HOG-pathway robustness is dependent on pathway activation status, we probed the HOG-pathway robustness during NaCl stress, which is known to activate the pathway, and superoxide stress (paraquat addition), which does not activate the pathway. Both stresses were applied in doses causing a similar (40–50%) reduction in reference strain growth rate. Interestingly, the relative gToW sensitivity patterns during these two stresses were very similar to that observed during unstressed conditions, indicating that the nodes of fragility remain regardless of pathway activation. The phenotypic correlations between these growth conditions ranged from 0.57 to 0.81 (Figure 3; r^2 for LSC rate). The impact of the HOG gToW perturbations were significantly stronger under normal growth conditions than under either NaCl or paraquat stress (P=0.005 and 7.4×10^{-5}, respectively, paired *t*-test of LSC rate). Apart from this general dampening effect, which is observed under different stress conditions, robustness is largely independent of pathway activation by environmental perturbations. However, we see an indication of interaction between the genetic and environmental perturbations. Although paraquat and NaCl stress give similar trends in the dampening of the phenotypes, the variance around this trend seems higher under NaCl (Supplementary Figure S8). An appealing interpretation would be conditional alleviation or aggravation, which would be expected if the effect of the genetic and environmental perturbation cancel out or act synergistically, respectively. We find it interesting that the targets furthest from the trend line under NaCl stress are *SLN1*, *SKO1* and *PTP3*, all known negative regulators of the osmotic stress response, on the negative side and *HOG1* and *PTC3* on the positive side. *PTC3* is equally and surprisingly alleviated by both paraquat and NaCl stress.

In summary, we used the gToW method to qualitatively capture nodes of fragility from overexpression within the HOG pathway. The quantitative correlation to the level of overexpression is more difficult to assess due to additional levels of gene, mRNA and protein regulation. However, previous results by Moriya *et al* show a correlation between growth phenotype, plasmid copy number and relative protein overexpression. Here, we report that expression changes have very strong impact on signalling. The system robustness against overexpression is heavily dependent on the target component and neighbouring nodes show very different fragility. In the HOG pathway, overexpression of Pbs2p and Ssk1p yield the strongest effects, whereas none of their neighbours; Ssk2p, Ste11p or Hog1p, are similarly sensitive. The *in silico* analysis of model variants clearly shows that model structure has a strong impact on the fragility of different nodes. Our results suggest that the stable formation of an Ssk1p–Ssk2p dimer and Pbs2p's scaffold function contribute to the fragility of their respective nodes. Although robustness information alone cannot be used to reject model structures, it provides information complementary to dynamic data that can be used to discriminate models, and should prove a valuable tool in any modelling endeavour.

Supplementary information

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

Acknowledgements

The authors would like to thank Hisao Moriya and Yuki Yoshida for their advice in creating the initial gToW set for this study, and Raul Garcia Salcedo for advice with the immuno-qPCR. MK was supported by JSPS as a postdoctoral fellow, and is currently financed by a repatriation grant from SSF. CW is supported by the IRTG 'Genomics and Systems Biology of Molecular Networks' from DFG. SH is supported by the Swedish Research Council, the 'Quantitative Biology' platform at GU and by the EC-funded 'QUASI' project. The collaboration between SH and HK is supported jointly by Vinnova and JST. AB is supported by the Swedish Research Council.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Hohmann S (2002) Osmotic stress signaling and osmoadaptation in yeasts. *Microbiol Mol Biol Rev* **66:** 300–372

Janiak-Spens F, Sparling DP, West AH (2000) Novel role for an HPt domain in stabilizing the phosphorylated state of a response regulator domain. *J Bacteriol* **182:** 6673–6678

Kitano H (2004) Biological robustness. *Nature Reviews* **5:** 826–837

Klipp E, Nordlander B, Kruger R, Gennemark P, Hohmann S (2005) Integrative model of the response of yeast to osmotic shock. *Nat Biotechnol* **23:** 975–982

Mapes J, Ota IM (2004) Nbp2 targets the Ptc1-type 2C Ser/Thr phosphatase to the HOG MAPK pathway. *EMBO J* **23:** 302–311

Moriya H, Shimizu-Yoshida Y, Kitano H (2006) *In vivo* robustness analysis of cell division cycle genes in Saccharomyces cerevisiae. *PLoS Genet* **2:** e111

Posas F, Witten EA, Saito H (1998) Requirement of STE50 for osmostress-induced activation of the STE11 mitogen-activated protein kinase kinase kinase in the high-osmolarity glycerol response pathway. *Mol Cell Biol* **18:** 5788–5796

Saito H, Tatebayashi K (2004) Regulation of the osmoregulatory HOG MAPK cascade in yeast. *J Biochem* **136:** 267–272

Schneider JC, Guarente L (1991) Vectors for expression of cloned genes in yeast: regulation, overproduction, and underproduction. *Methods Enzymol* **194:** 373–388

Sopko R, Huang D, Preston N, Chua G, Papp B, Kafadar K, Snyder M, Oliver SG, Cyert M, Hughes TR, Boone C, Andrews B (2006) Mapping pathways and phenotypes by systematic gene overexpression. *Mol Cell* **21:** 319–330

Stelling J, Sauer U, Szallasi Z, Doyle III FJ, Doyle J (2004) Robustness of cellular functions. *Cell* **118:** 675–685

Tatebayashi K, Tanaka K, Yang HY, Yamamoto K, Matsushita Y, Tomida T, Imai M, Saito H (2007) Transmembrane mucins Hkr1 and Msb2 are putative osmosensors in the SHO1 branch of yeast HOG pathway. *EMBO J* **26:** 3521–3533

Torres EM, Sokolsky T, Tucker CM, Chan LY, Boselli M, Dunham MJ, Amon A (2007) Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science* **317**: 916–924

Warringer J, Ericson E, Fernandez L, Nerman O, Blomberg A (2003) High-resolution yeast phenomics resolves different physiological features in the saline response. *Proc Natl Acad Sci USA* **100**: 15724–15729

Wu C, Jansen G, Zhang J, Thomas DY, Whiteway M (2006) Adaptor protein Ste50p links the Ste11p MEKK to the HOG pathway through plasma membrane association. *Genes Dev* **20**: 734–746

JOURNAL
OF
THE ROYAL
SOCIETY
Interface

Consistent design schematics for biological systems: standardization of representation in biological engineering

Yukiko Matsuoka, Samik Ghosh and Hiroaki Kitano

J. R. Soc. Interface 2009 **6**, S393-S404 first published online 3 June 2009
doi: 10.1098/rsif.2009.0046.focus

Interface focus IET

J. R. Soc. Interface (2009) **6**, S393–S404
doi:10.1098/rsif.2009.0046.focus
Published online 3 June 2009

REVIEW

Consistent design schematics for biological systems: standardization of representation in biological engineering

Yukiko Matsuoka[1,2,*], Samik Ghosh[1] and Hiroaki Kitano[1,3,4]

[1] *The Systems Biology Institute, 4-01 Shillman Hall, 3-14-9 Ohkubo, Shinjuku, Tokyo 169-0072, Japan*
[2] *ERATO Kawaoka Infection-induced Host Responses Project, Japan Science and Technology Agency, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan*
[3] *Sony Computer Science Laboratories, Inc., 3-14-13 Higashi-Gotanda, Shinagawa, Tokyo 141-0022, Japan*
[4] *Okinawa Institute of Science and Technology, 7542 Onna, Onna-Son, Kunigami, Okinawa 904-0411, Japan*

The *discovery by design* paradigm driving research in synthetic biology entails the engineering of de novo biological constructs with well-characterized input–output behaviours and interfaces. The construction of biological circuits requires iterative phases of design, simulation and assembly, leading to the fabrication of a biological device. In order to represent engineered models in a consistent visual format and further simulating them *in silico*, standardization of representation and model formalism is imperative. In this article, we review different efforts for standardization, particularly standards for graphical visualization and simulation/annotation schemata adopted in systems biology. We identify the importance of integrating the different standardization efforts and provide insights into potential avenues for developing a common framework for model visualization, simulation and sharing across various tools. We envision that such a synergistic approach would lead to the development of global, standardized schemata in biology, empowering deeper understanding of molecular mechanisms as well as engineering of novel biological systems.

Keywords: systems biology; standardization; biological engineering; graphical notation

1. INTRODUCTION: SYNTHETIC BIOLOGY AND SYSTEMS BIOLOGY

Synthetic biology aims at designing artificial genetic circuits for specific functions. It may take the form of the bottom-up design and construction paradigm as elucidated by Isaacs *et al.* (2003), Stricker *et al.* (2008), Gardner *et al.* (2000), Hasty *et al.* (2002), Guido *et al.* (2006), Deans *et al.* (2007), etc., or engineering of existing genomes to fit specific purposes (Itaya *et al.* 2005). In the bottom-up approach, it is essential that genetic circuits be designed and proved to be functional before being actually implemented on biological materials, in the same way as electric circuits, robotics systems and aircraft are designed and built. Therefore, it is imperative to develop a series of industrial-strength software platforms that enable such design processes.

At the same time, a hallmark of matured engineering fields is the development and maintenance of standards and modularized components that can be re-used and cross applied for various circuits. Such components are openly publicized and exchanged in the market. It is often the case that software components are shared in the community as open source software or at low cost as shareware.

An interesting attempt to foster development of technologies and the science behind them is to create competition in organized or emergent form. RoboCup (Kitano *et al.* 1997), for example, is an organized effort to foster competition and collaboration to speed up the

*Author and address for correspondence: Systems Biology Institute, 4-01 Shillman Hall, 3-14-9 Ohkubo, Shinjuku, Tokyo 169-0072, Japan (myukiko@sbi.jp).

Electronic supplementary material is available at http://dx.doi.org/10.1098/rsif.2009.0046.focus or via http://rsif.royalsocietypublishing.org.

One contribution to a Theme Supplement 'Synthetic biology: history, challenges and prospects'.

Received 4 February 2009
Accepted 5 May 2009

development of artificial intelligence (AI) and robotics, which can be used in the real world. It sets the goal: 'By the year 2050, develop a team of fully autonomous humanoid robots that can win against the human world soccer champion team' and organizes annual competitions to benchmark technologies and exchange them for further progress (Kitano *et al.* 1997). This project has had a dramatic impact on the AI and robotics communities in accelerating research, and some of the research results have been quickly transferred into industry.

In the synthetic biology area where the goal is to design and construct novel biological circuits by combining basic building blocks, the process of developing standardized and well-characterized components has been a dominant paradigm. Unambiguous characterization of biological parts (Peccoud *et al.* 2008) as well as standardization of parts assembly and design processes are significant challenges in efforts to streamline the fabrication of biological circuits. Community efforts in this direction have been undertaken through the BioBricks Foundation (http://bbf.openwetware.org/) and OpenWetWare initiative (http://openwetware.org/). Community-wide adoption of the standards has been fostered through the International Genetically Engineered Machines competition, iGEM (Brown 2007), a competition and collaboration forum to foster development of synthetic biology. Such efforts shall pave the way for enhancing research and education in this discipline.

Clear characterization of biological parts and establishment of standards for deriving kinetic models of parts and devices are some of the key challenges in the synthetic biology community. At the same time, it would be required to endorse standards for the description of biological modules and circuit components, as initiated in systems biology (Hucka *et al.* 2003; Le Novère *et al.* 2008), for initiatives like iGEM to have wider impact on dissemination of design knowledge and adoption of consistent schematics. Efforts in this direction have been initiated through the Provisional BioBrick Language (POBOL, http://pobol.org/), which aims to define a data exchange standard for standard biological parts.

Development of consistent, standardized schemata for the representation of biological parts is an important direction of research, particularly as the field matures. Here we introduce the various standardization activities in the systems biology community for consistent schematic representations and discuss the possible scope for mutual deployment of standards and technologies in the synthetic biology area. It should be noted here that there are possibly other up-front issues that the synthetic biology community needs to address today, rather than worrying about standardization of representation. However, the point we wish to address here is a potential future need to develop standard descriptions when the field matures enough and discuss some of the current efforts and possible future directions.

2. SCOPE OF MODELLING AND SIMULATION IN DESIGNING BIOLOGICAL CIRCUITS

Modelling and simulation are indispensable tools in all engineering designs and have been successfully applied in the automobile, aerospace and telecommunication industries for many decades. Computational fluid dynamics (CFD), for example, is an essential design process in aircraft, ship and automobile design. Any high-rise building has to undergo a series of structural integrity simulations even to be approved for construction; chipmakers model, modify and simulate their designs on computers before sending them to the fabrication plants; 'virtual cars' are driven and 'virtual aircraft' flown under simulated conditions before hitting the manufacturing floor (The Economist 2005). In the field of sciences, modelling is a practice of quantitative hypothesis testing, which enables researchers to test and prove the scientific hypotheses. Models capture and communicate knowledge and theories in a concrete form that can be simulated before building prototypes.

Systematic modelling and simulation of prototypes have been notable features in all engineering design problems. However, their adoption in the biological sciences has been traditionally sparse. In recent years, the role of *in silico* modelling and simulation in understanding biological systems at a 'systems level' has gained traction in both academic and pharmaceutical research communities (British Telecommunications 2007; PricewaterhouseCoopers 2008). Various flavours of simulation techniques have been applied in understanding systems behaviour of biological processes at multiple scales (Ramsey *et al.* 2005; Hoops *et al.* 2006), from molecular maps of cellular pathways, to tissues and organs, to the simulation of drug regimens in virtual patient populations (Rullmann *et al.* 2005).

While systems biology emphasizes application of computational techniques for obtaining insights into the mechanism of various biological processes, synthetic biology endeavours to develop de novo biological circuits to engineer the behaviour of living systems. In this perspective, modular model development and simulation-driven validation of prototypes form the cornerstones of this discipline. In order to develop engineered biological circuits, like synthetic gene regulatory circuits, a computational platform comprising tools for designing, simulating and assembling biological circuits from existing parts would be required. A typical workflow for engineering a de novo biological circuitry can be envisaged as follows.

(i) *Design of the proposed biological circuitry.* The design would be at an abstract level of view, using graphical representation of different fundamental entities and standard format for storage, retrieval and exchange of the design.

(ii) *Simulation of biological circuit design.* The simulation tools would allow the designer to test the response of the biological circuit under different conditions of input (environmental signals etc.), explore the parameter space of different components to quantify biological robustness of the design and test various hypotheses.

(iii) *Assembly of biological circuit.* Once the design and simulation phases, working in multiple iterations, have confirmed the desired performance of the biological circuit, biological components would be accessed from a central repository to