Context is critical if researchers need to know the specificity or responsiveness of a target within a pathway. This is something that has yet to be captured by any informatics system, but it will invariably be tied to greater utility and even IP definitions in the near future. To a large extent, pathways data and knowledge management are seen to go hand in hand, both being used to make important decisions, and both relying on context. It is interesting to consider that perhaps we are witnessing the emergence of a new informatics paradigm, where knowledge becomes a formally defined and well-shared scientific and economic asset.

The third topic that was discussed by a few was the development of new IT strategies to improve informatics infrastructure within large companies. Several speakers presented new and effective methodologies to design and build complex informatics systems that will eventually become the backbone of drug discovery platforms. Richard Ashe described accelerated software development programs within GSK (http://www.gsk.com/) that can adapt more effectively to changing discovery requirements. Using lightweight methodologies and web-based services rather than monolithic designing principles, projects could be defined and completed faster within an ever-changing world of needs. Roy Dunbar, CIO of Eli Lilly (http://www.lilly.com/) proposes more effective collaborations between IT and scientists to create applications that have more utility. Juergen Seega from Abbott Laboratories (http://abbott.com/) described 'keeping the scientific and IT worlds in sync', rather than at odds. Taken together, IT is progressing from being hidden in the backwaters of software tinkering, to becoming a major strategic partner within drug discovery environments.

Other issues were addressed, including good laboratory practices as related to informatics, electronic notebooks, intelligent storage and use of high-throughput screening data, and IP portfolio management. Based on the presentations at the conference, it was clear that the applications and benefits of informatics for the pharmaceutical industry are growing, and that informatics promises to be a key factor for the future success of drug discovery.

CellDesigner: a process diagram editor for gene-regulatory and biochemical networks

Akira Funahashi*, Mineo Morohashi and **Hiroaki Kitano,** Kitano Symbiotic Systems Project, ERATO, Japan Science and Technology Corporation, Suite 6A, M31, 6-31-15 Jingumae, Shibuya-ku, Tokyo 150-0001 Japan. **Hiroaki Kitano,** The Systems Biology Institute, Suite 6A, M31, 6-31-15 Jingumae, Shibuya-ku, Tokyo 150-0001, Japan. **Naoki Tanimura,** Fuji Research Institute Corporation, 2-3 Kanda Nishikicho, Chiyoda-ku Tokyo 101-8443, Japan. **Hiroaki Kitano,** Sony Computer Science Laboratories, Inc., 3-14-15 Higashi-Gotanda, Shinagawa, Tokyo 141-0022, Japan. **Hiroaki Kitano and Mineo Morohashi,** School of Fundamental Science and Technology, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan. *e-mail: funa@symbio.jst.go.jp

Systems biology is characterized by synergistic integration of theory, computational modeling, and experiment [1]. Although software infrastructure is one of the most crucial components of systems biology research, there has been no common infrastructure or standard to enable integration of computational resources. To solve this problem, the Systems Biology Markup Language (SBML) [2] and Systems Biology Workbench (SBW) have been developed [3]. SBML is an open, XML-based format for representing biochemical reaction networks, and SBW is a modular, broker-based, message-passing framework for simplified intercommunication between applications. Several simulation and analysis software packages already support SBML (Level-1) and SBW, or are being developed to support them.

Identification of the logic and dynamics of gene-regulatory and biochemical networks is a major challenge of systems biology. We believe that such network building tools and simulation environments using standardized technologies play an important role in the software platform of systems biology. As one such approach, we have developed CellDesigner, a process diagram editor for gene-regulatory and biochemical networks.

In the following, we will introduce the main features of CellDesigner. The most crucial elements are that it is a system of

1478-5282/03/$ – see front matter PII: S1478-5282(03)02370-0

graphical representation, which is SBML-compliant and SBW-enabled. We expect that these features will become part of the standardized technology for systems biology.

Features of CellDesigner

Broadly classified, the current version of CellDesigner has the following features:

- Representation of biochemical semantics
- Detailed description of state transition of proteins
- SBML compliant (SBML Level-1, Version-1)
- Integration with SBW-enabled simulation modules
- Extreme portability as a Java application

The aim in developing CellDesigner is to supply a process diagram editor with standardized technology for every computing platform, so that it benefits as many users as possible. By using standardized technology, the model can be easily used with other applications, thereby reducing the need for users to create a specific model for each editing, simulation or analysis task. The main standardized features that CellDesigner supports are in 'graphical notation', 'model description', and 'application integration environment'. The standard for graphical notation plays an important role for efficient and accurate dissemination of knowledge [4], and the standard for model description will enhance the portability of models between software tools. Similarly, the standard for application integration environment will help software developers to provide the ability for their applications to communicate with other tools.

Symbols and expressions

CellDesigner supports graphical notation and listing of symbols based on a proposal by Kitano [4]. Although several graphical notation systems have been already proposed [5–9], each has obstacles to becoming a standard. Kitano proposed a graphical notation for biological networks [4] designed to express sufficient information in a clearly visible and unambiguous way. The key components of Kitano's proposal are as follows:

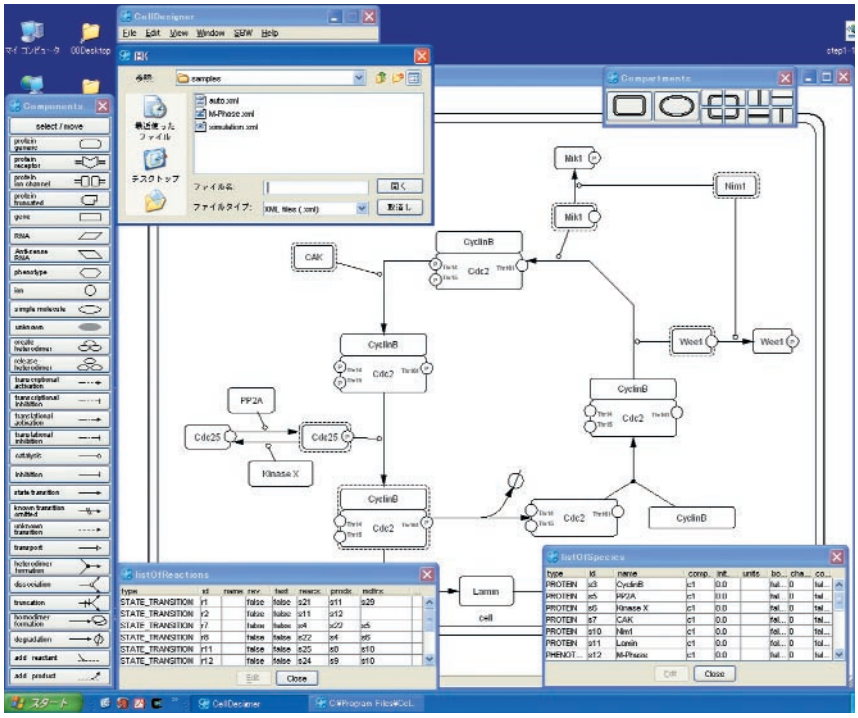

Figure 1. A screenshot of CellDesigner, showing a part of a yeast cell cycle process.

(1) Expressiveness
(2) Semantically unambiguous
(3) Visually unambiguous
(4) Extension capability

The goal of Kitano's proposal is to define a comprehensive system of notation for visually describing biological networks and processes, thereby contributing to the eventual formation of a standard notation. Software support is an important issue for standardization of a notation system. Even if the proposed notation system satisfies the requirements of biologists, lack of software support will drastically decrease its advantages. CellDesigner supports graphical notation for state-transition diagrams and residue-state representations (Figure 1).

SBML compliant

CellDesigner is an SBML-compliant application, which means that it supports SBML reading and writing capabilities. SBML is the standard model definition language in systems biology field; it is now used by the BASIS Project (http://www.basis.ncl.ac.uk/technology.html), the DARPA BioSPICE project (http://www.biospice.org) and the International *E. coli* Alliance (http://ieca2003.jtbcom.co.jp). CellDesigner can read all SBML Level-1 documents, so users can use existing SBML models such as KEGG database. We have already converted more than 5,000 metabolic pathways of KEGG to SBML (are available from http://www.systems-biology.org/). Other SBML models are available from the SBML model repository (http://www.sbw-sbml.org/ModelsWebPages/ModelRepository.htm). Users can also use their own SBML models created using CellDesigner on other SBML compliant applications.

SBW enabled

CellDesigner is also an SBW-enabled application. With SBW installed, CellDesigner can integrate with all SBW-enabled modules (Figure 2). For example, users can browse or modify a model converted from an existing database with CellDesigner, and launch a simulator from CellDesigner (by selecting 'Simulation Service' from the SBW menu) to run simulations in real time. There are many other SBW-enabled modules, such as

ODE-based simulator, stochastic simulator, Matlab translator, bifurcation analysis tool, and optimization module. These SBW-enabled modules are freely available from http://www.sbw-sbml.org/.

Supported environment

CellDesigner is implemented in Java, and it can run on many platforms that support JRE (Java Runtime Environment). Currently, CellDesigner can run on the following platforms:

- Windows (98SE or later)
- MacOS X (10.2.6)
- Linux (RedHat8.0, Vine 2.6)
- FreeBSD (4.8-RELEASE, 5.1-RELEASE)

The current version of CellDesigner requires JRE1.3.x or higher, and X-Window System for UNIX platforms.

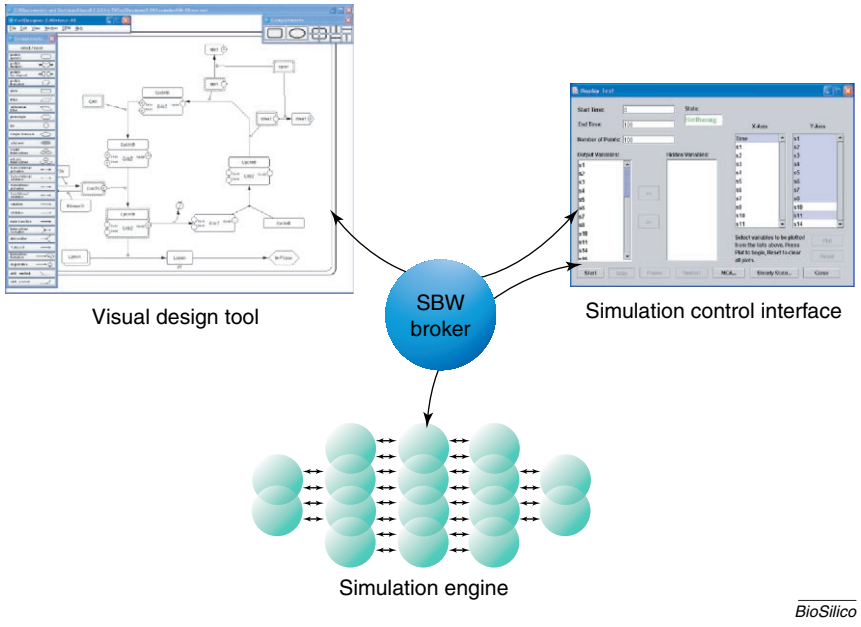

Figure 2. The relationship between SBW broker and some SBW-enabled modules.

How does it work?

Building models with CellDesigner is quite straightforward. To create a model, the user selects 'New' from the 'File' menu, inputs the name of an SBML document, and a new canvas appears. The user can then place a species, such as a protein, gene, RNA, ion, simple molecule and so on, by selecting the species in the 'Components' window and clicking on the canvas. A new window will appear asking the name of the species. The size of each species can be changed by clicking and dragging the corner of species. The user can also define the default size of each species from 'Show Palette option' from the 'Window' menu. Species can be moved by dragging and dropping.

To connect species with arrows to represent reactions, a type of reaction should first be selected from the 'Components' window, and a reactant species then clicked, followed by a product species. To add more reactants, the user can select 'Add reactant' from the 'Components' window, and then choose species and reaction. These are straightforward steps, which should not cause users any confusion.

CellDesigner can also represent common types of reactions, such as catalysis, inhibition activation and so forth. The procedure for representing such reactions is just the same as adding reactants or products to an existing reaction; that is, a species (modifier) is selected, followed by a reaction. The user can also easily edit the symbols for proteins with modification residues, and hence, can describe detailed state transitions between species of an identical protein by adding different modifications.

The models are stored in an SBML document, which contains all the necessary information referring to species, reactions, modifiers, layout information (geometry), state transitions of proteins, modification residues and so on. These SBML models can be used on other SBML-compliant applications.

What distinguishes CellDesigner's technology from others currently available?

Currently, many other applications exist that include pathway design features, as follows:

- BioUML (Cuttingedge, http://www.biouml.org)
- Freehand templates (BioCarta, http://www.biocarta.com)
- JDesigner (ERATO Kitano Symbiotic Systems Project and Keck Graduate Institute, http://www.cds.caltech.edu/~hsauro/JDesigner.htm)
- PathwayAssist (Startagene, http://www.startagene.com/softwaresolutions)
- VisualCell (Gene Network Sciences, http://www.gnsbiotech.com)
- Virtual Cell (National Resource for Cell Analysis and Modeling, http://www.nrcam.uchc.org)

The advantages of CellDesigner over other pathway design tools are that it is based on standard technology (i.e. is SBML-compliant and SBW-enabled), supports clearly expressive and unambiguous graphical notation systems, which is aimed at contributing to eventual standard formation, and it runs on many platforms (e.g. Windows, MacOS X, Linux, etc.).

As described above, the aim of the development of CellDesigner is to supply a process diagram editor with standardized technology for every computing platform, so that it will benefit as many biological researchers as possible. Some of the existing applications are SBML-compliant, and some are written in Java so that they can be run on multiple platforms. However, only

JDesigner is SBW-enabled, and very few applications support graphical notation.

Of course, these tools are powerful in some respects and they are not intended to support the same features as CellDesigner. Some of them have the facility to create pathways, and some also include a simulation engine or database integration module. CellDesigner does not include a simulation engine in itself, but it can connect with other SBW-enabled applications that do, so the simulation facility was not necessary to include in CellDesigner. Furthermore, we have been converting existing databases to SBML, such as KEGG, and all SBML-compliant applications can easily browse, edit and simulate CellDesigner models.

The overriding advantage of CellDesigner is that it uses open and standard technology. The models created by CellDesigner can be used on many other standard applications, and its graphical notation system will make the representation of models more efficient and accurate.

How does the technology still need to be improved?

CellDesigner is aimed to be used by those who are not familiar with computer models or mathematics, so we designed it to be as user-friendly as possible. We want it to be 'easy to create a model, execute simulation and use analysis tools'. This will be achieved by extending the development of corresponding SBW-enabled modules. Improvement of the graphical–user interface is also required, including the mathematical equation editor, so that the user can easily write equations by selecting and dragging a species.

Conclusion

We have introduced CellDesigner, as a process diagram editor for gene-regulatory and biochemical networks based on standardized technologies and with wide transportability to other SBML-compliant applications and SBW-enabled modules. CellDesigner also aims to support standard graphical notation. These technologies are still changing and growing so standardization is not yet strictly defined. As we are in partnership with the SBML and SBW development groups, we will go through with these standardization projects and hence improve the quality of CellDesigner. Finally, CellDesigner runs on multiple platforms such as Windows, Linux and MacOS X, and it is freely available from http://www.systems-biology.org/.

Acknowledgements

This research is in part supported by the Rice Genome and Simulation Project (Ministry of Agriculture, Japanese government), the International Standard Development area of the International Joint Research Grant (NEDO, Japanese Ministry of Economy, Trade, and Industry), the Exploratory Research for Advanced Technology (ERATO) program (Japan Science and Technology Corporation), and through the special coordination funds for promoting science and technology from the Japanese government's Ministry of Education, Culture, Sports, Science and Technology.

References

1 Kitano, H. (2002) Systems biology: A brief overview. *Science* 295, 1662–1664
2 Hucka, M. *et al*. (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 19, 524–531
3 Hucka, M. *et al*. (2001) The ERATO systems biology workbench: an integrated environment for multiscale and multitheoretic simulations in systems biology. In *Foundations of Systems Biology*, pp. 125–143, MIT Press
4 Kitano, H. (2003) A graphical notation for biological networks. *BioSilico* 1, 169–176
5 Kohn, K. (2001) Molecular interaction maps as information organizers and simulation guides. *Chaos* 11, 84–97
6 Kohn, K.W. (1999) Molecular interaction map of the mammalian cell cycle control and DNA repair systems. *Mol. Biol. Cell* 10, 2703–2734
7 Pirson, I. *et al*. (2000) The visual display of regulatory information and networks. *Trends Cell Biol.* 10, 404–408
8 Cook, D.L. *et al*. (2001) A basis for a visual language for describing, archiving and analyzing functional models of complex biological systems. *Genome Biol*. 2(4), RESEARCH0012
9 Maimon, R. and Browning, S. (2001) Diagrammatic notation and computational structure of gene networks. In *Proceedings of the Second International Conference on Systems Biology*, pp. 311–317, Omnipress