

REVIEW

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

Yukiko Matsuoka^{1,2,*}, Samik Ghosh¹ and Hiroaki Kitano^{1,3,4}

¹The Systems Biology Institute, 4-01 Shillman Hall, 3-14-9 Ohkubo, Shinjuku, Tokyo 169-0072, Japan

²ERATO 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

³Sony Computer Science Laboratories, Inc., 3-14-13 Higashi-Gotanda, Shinagawa, Tokyo 141-0022, Japan

⁴Okinawa 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'.

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

match the different components of the design phase. Further, an assembler would allow the assembly of the different components and allow further simulation iterations before final fabrication.

As seen from the workflow, modelling and simulation, together with standardized visual representation and the ability to store and share the biological circuit in a machine-readable format, are key elements of the process. Several challenges exist in integrating standardized model visualization and simulation techniques in the workflow.

First, visualization standards should have the capability to capture biological entities at different granularities, from promoter sequences, ribosome-binding site sequences, genes, proteins to higher-order molecular pathways and complexes.

Second, simulation tools should be able to define models at these multiple scales, capturing unknown interactions and parameters while integrating standardized, characterized biological components into networks of interacting modules.

Third, there exists a knowledge gap between visual representations of biological systems vis-à-vis corresponding mathematical models. While graphical representations (molecular maps) are intended to capture the known biology of processes, mathematical models encode only parts of the detailed molecular maps owing to the underlying complexity of the interactions, insufficient data to characterize processes, unknown parameters or shortcomings in mathematical representation of complex interactions. *In silico* modelling and simulation tools should provide mechanisms to reconcile such knowledge gaps in their framework. Moreover, software tools should provide mechanisms for incorporating restrictions in parts assembly and design phases of biological circuits.

In this respect, development of a common framework, encompassing different tools and schemata employed in the different phases, is required to accelerate the progress of synthetic biology. To share the results of modelling efficiently, we need a common language in representing: (i) mathematical contents; (ii) semantics, annotations; and (iii) visual representation of models.

Figure 1 shows a schematic of the biological engineering process with different schemata available for standardized exchange of information. As seen from the figure, standards encompassing mathematical content, like the Systems Biology Markup Language (SBML; Hucka *et al.* 2003), semantics and annotation, like Minimum Information Requested in the Annotation of biochemical Models (MIRIAM; Le Novère *et al.* 2005), and visualization, like the Systems Biology Graphical Notation (SBGN; Le Novère *et al.* 2008), can be employed for modelling and simulation of biological parts, devices as well as systems. In order to facilitate *in silico* simulation, a common language for the simulation systems would facilitate the sharing of results and usage of different simulation engines. Efforts are now being made to standardize simulation result description such as SBRML (Dada *et al.* 2009), which

can be integrated into the workflow as depicted in figure 1. The figure illustrates how modelling and simulation techniques can be employed at each step of the design and assembly process hierarchy of biological parts, devices and systems. However, such standards need to be enhanced to address the unique challenges of synthetic biology elucidated earlier.

In the remaining sections, we outline the different standardization efforts in the systems and synthetic biology communities and provide potential avenues for developing a common framework in a mutually beneficial manner.

3. STANDARDS AND TOOLS IN BIOLOGICAL ENGINEERING

3.1. Efforts in synthetic biology

The goal in synthetic biology is to develop engineered biological circuits with well-defined input–output behaviour. Thus, design, simulation and assembly tools that aid in the development of high-precision biological constructs form an integral part of this effort. Moreover, the ability to *re-use* well-characterized parts is a hallmark of any engineering discipline. In this respect, significant efforts have been undertaken in the standardization of biological parts and their systematic storage and retrieval, which we briefly review next.

As synthetic biology is an engineering discipline from the onset, efforts in standardization were in place at the very early stages of research activities. The Registry of Standard Biological Parts (<http://parts.mit.edu/>) is a database to create and maintain such building blocks, providing free access to an ‘open-commons’ of basic biological functions. The goal is to streamline the fabrication of complex constructs to programme synthetic biological systems. The registry stores the biological constructs, providing detailed *worksheets* of the input–output characteristics of biological components together with the *interfaces* for communication between them. It contains *biological parts*, which are combined to form *biological devices* that can be further connected to build engineered *biological systems*.

While the registry provides storage of the biological components, their properties and interfaces need to be defined in terms of standardized schemata that allows unambiguous definition of the parts and their behaviour. In this effort, the BioBrick parts, an open source genetic parts as defined via an open technical standards setting process that is led by the BioBricks Foundation (<http://bbf.openware.org/>), represent an effort to introduce the engineering principles of abstraction and standardization into synthetic biology. BioBricks standard biological parts are DNA sequences of defined structure and function; they share a common interface and are designed to be composed and incorporated into living cells such as *Escherichia coli* to construct new biological systems.

3.1.1. Challenges in parts characterization and assembly. One of the major challenges faced by the synthetic biology community with regard to standard formation is how to characterize parts features. While it is

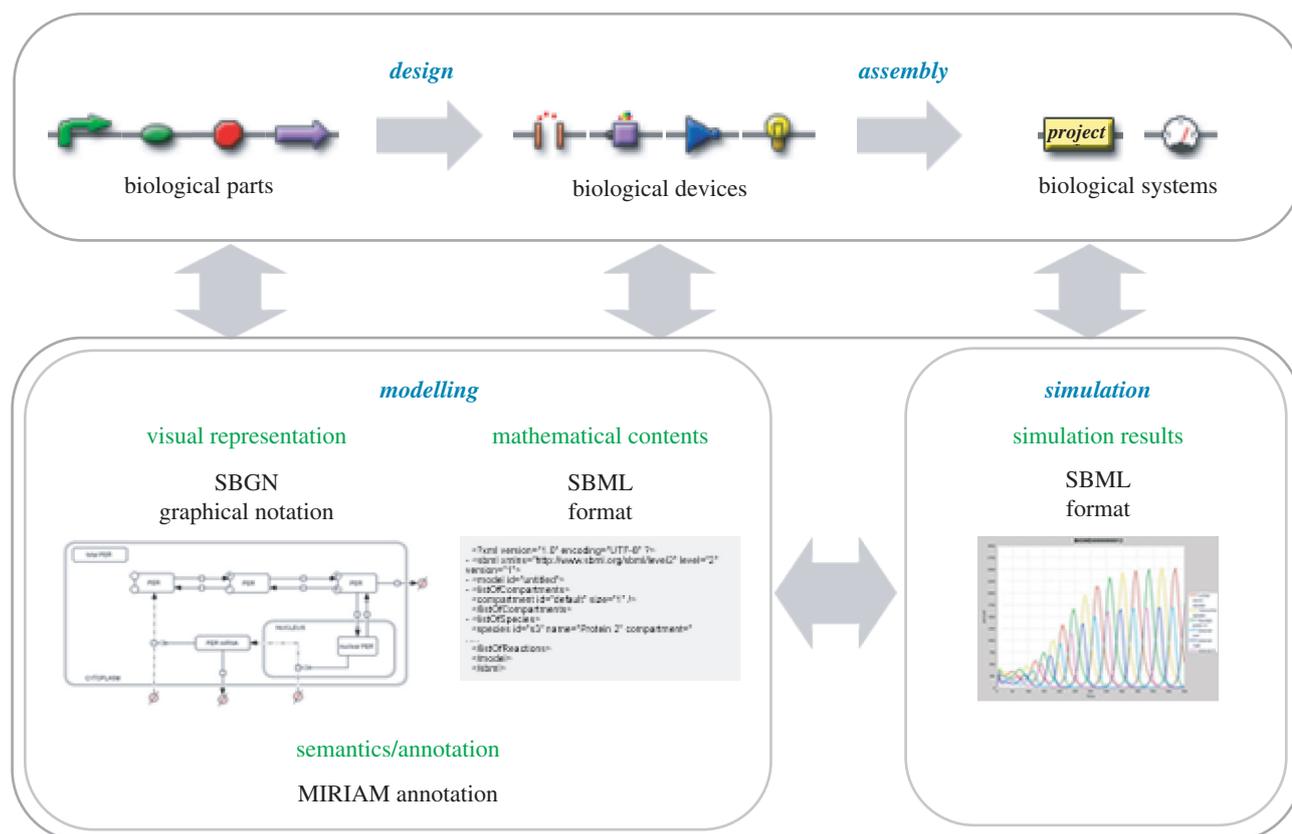


Figure 1. Workflow for the design and development of engineered biological circuits.

currently defined in terms of promoter structure and sequences, it is not a characterization in terms of function in the context of interacting networks.

A proper level of description that may smoothly interface with a network-level context is at the device behaviour level. This is at the same level of description as in electric design. In electric design, apart from circuit diagrams, there are datasheets for each component that specify basic parameters and modes of action. For example, in transistor specification, datasheets define various basic parameters and behaviour characteristics of a transistor. Such specifications can be used to generate properly parametrized equivalent circuits that provide a versatile definition of functional behaviour in a specific network context. Figure 2*a* is an illustrative example of an equivalent circuit at the device level, and figure 2*b* is an equivalent circuit of common emitter configuration of the same device in a network context. In the circuit-level equivalent circuit, parameters such as h_{fe} , i.e. current amplification ratio, and h_{ie} , i.e. input impedance, are defined. Electronic engineers can design and analyse circuits using these parameters without looking into details of implementations. This representation significantly enhanced our capability to design and analyse circuits and is particularly important when it has to be scaled up. At the same time, it should be noted that unlike electronic circuits in which an identical device can be used in multiple places in the system (such as using FET-type 2SK30A in different amplification modules), a synthetic biology device can be used only once in the system to avoid unexpected interferences. Thus, ‘device’ in the synthetic biology context shall be considered as ‘family of

parts’. Therefore, it is essential for each device description to have parametric features reflecting different parts of the family.

The ability to define biological parts unambiguously at the device as well as network levels is one of the key challenges in synthetic biology. Current efforts to define interfaces to biological devices in a molecule-independent manner use parameters such as polymerase per second, which is the flow rate of RNA polymerase molecules along the DNA, or by ribosomes per second which measures the flow rate of ribosomes along the mRNA molecule. Figure 3*a* shows an illustrative device description of a banana odour generator from the BioBricks registry. In this figure, ATF1 transcription activity is described as a function of input signal that activates ATF1 transcription. The output, isoamyl acetate production, is described as a function of isoamyl alcohol and ATF1 transcription activity. In order to characterize the device, it is required to define a parameter capturing the function of the device, instead of the kinetic constant of ATF1 catalysing isoamyl alcohol conversion into isoamyl acetate. It is not always useful to describe the kinetic constant for ATF1 enzyme alone because BioBrick is assumed to be the building blocks at the device level. Thus, biological equivalents of parameters such as h_{fe} need to be defined that provide a rate at which changes in signalling to transcription of ATF1 affect the rate of isoamyl acetate production (e.g. k_{sp} : signal–product amplification rate). Also, it should have a parameter (k_{ip} : input–product rate) that characterizes change in product (isoamyl acetate) per change in input (isoamyl alcohol). The above example is only provided to describe the level

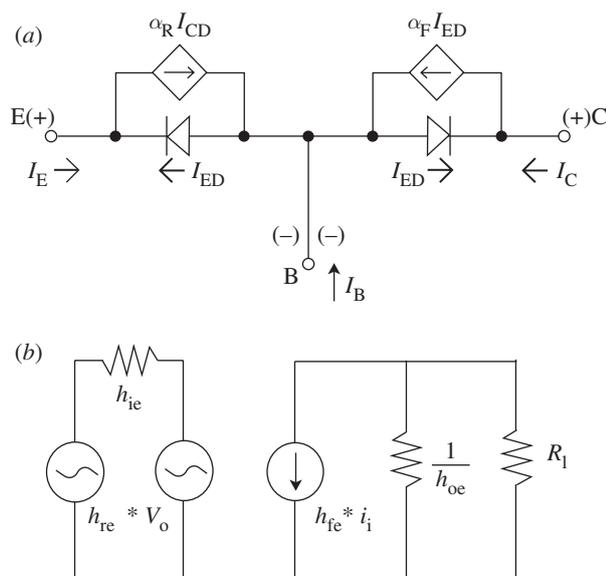


Figure 2. (a) Equivalent circuit for device level—an equivalent circuit for NPN transistor (from Wikipedia). (b) Equivalent circuit (h parameter representation) for common emitter topology configuration.

of abstraction at which an equivalent biological circuit shall be defined (figure 3b). With these parameters, Bio-Brick designers will be able to design scalable circuits without examining details of biological elementary reactions. This example highlights the need for developing a standard format for describing elementary building blocks at the device level, which can be integrated in synthetic biology parts databases, like BioBricks.

3.1.2. Software support in synthetic biology. With the development of parts databases on the one hand, efforts to build software support for synthetic biology are also underway. Software platforms such as BioJADE (Goler 2004), ATHENA, now called TINKERCELL (Chandran *et al.* 2009), and GENOCAD (Cai *et al.* 2007) have been designed specifically for synthetic biology needs. Tools like GENEDESIGN (Richardson *et al.* 2006) provide a genome version control system, while BRICKIT and CLOTHO provide mechanisms to manipulate DNA and protein coding sequences. Machine-readable language efforts have also been developed in Antimony and LBS. Table 1 provides an overview of the common software tools and platforms in synthetic biology.

These tools focus on specific aspects of sequence design with minimal information on interactions at a network level. These will be powerful and useful tools for designing systems on a small scale, where possible interactions can be intuitively followed. However, scaling up of synthetic biology modules to form network elements will quickly make dynamical behaviours intractable. Moreover, higher level descriptions need to be developed at the level of device function and mode of action as device-level part characterizations become more commonplace. It is envisaged that as building blocks are assembled and scaled up to relatively complex networks, computer-assisted network design and

modelling tools such as CELLDISIGNER (Funahashi *et al.* 2003) would be useful to capture network dynamics generated from assembled parts.

The efforts outlined earlier focus on the definition of standard parts, their common storage and retrieval and tools to simulate them. However, a significant aspect of biological engineering is the visual representation of the components (from parts to devices and finally systems), ability to simulate different components *in silico* and tools to support such graphical representation and simulation. In §3.2, we overview the different efforts in the systems biology community for visualization and simulation standards before providing insights into the development of a common computational framework across the synthetic and systems biology domains.

3.2. Efforts in systems biology

The focus in systems biology has been on understanding the mechanistic behaviour of biological components in a holistic manner and aggregating data from literature and experimental systems. Major efforts in systems biology have focused on the representation of biological pathways and their molecular interactions, together with the development of simulation schemata. Pathway standards have been developed, aiming to facilitate collaboration and data exchange among various research communities. The development of standards for computational platforms in biological engineering can be broadly defined in terms of four key feature elements, which we elucidate next:

- (i) standardization of representation of mathematical contents,
- (ii) semantics and annotations,
- (iii) visual representation of models, and
- (iv) simulation of biochemical networks.

3.2.1. Standardization of representation of mathematical contents. With the rapid increase in the volume of high throughput data available to systems biologists, efforts for defining standards for storage, analysis and exchange of large datasets have been undertaken on a community-wide scale. Standards include Gene Ontology (Ashburner *et al.* 2000) for describing gene functions, SBML and CellML (Lloyd *et al.* 2008) for describing biochemical reaction networks and Minimum Information About a Simulation Experiment (MIASE, <http://www.ebi.ac.uk/compneur-srv/miase/>), to name a few. While the goal of all these standards (extensively reviewed in Brazma *et al.* 2006; Strömbäck *et al.* 2007) is to define a consistent schema for data exchange, individual standards are targeted towards addressing specific issues. The ability to represent biological knowledge in a mathematically consistent format is key to performing *in silico* simulation and analysis of their dynamic behaviours. In this respect, SBML and CellML are two standards adopted across the systems biology community.

SBML is a machine-readable format for representing pathway models (<http://sbml.org>; Hucka *et al.* 2003).

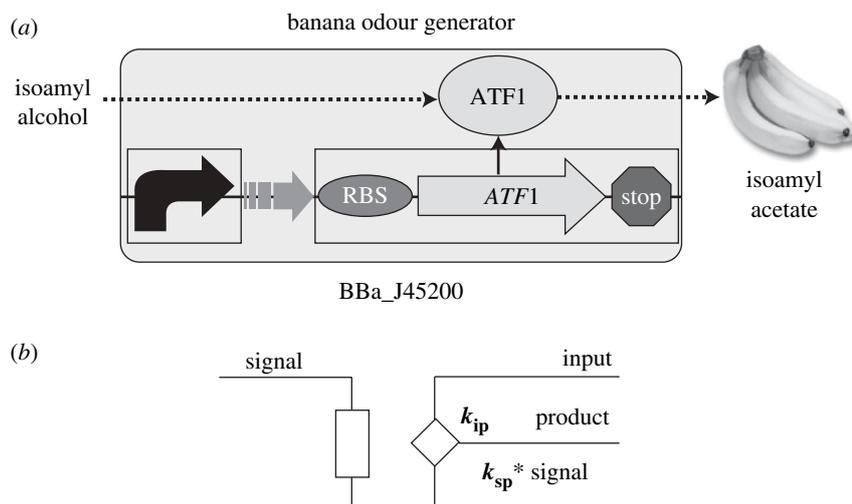


Figure 3. (a) An example of device description from BioBrick (taken from BioBrick registry). (b) An example of equivalent circuit for BBa_J45200. Shown only to provide the sense of abstraction, instead of exact formalism.

Table 1. Software tools and standards in synthetic biology.

software tool	overview	references	availability
ASMPARTS	a computational tool to produce models of biological systems by assembling models from biological parts	Rodrigo <i>et al.</i> (2007 <i>a,b</i>)	http://soft.synth-bio.org/asmparts.html
Antimony	a human-readable and human-writable language for describing biological modules		http://antimony.sourceforge.net
BIOJADE	a design and simulation tool for synthetic biological systems	Goler (2004)	http://web.mit.edu/jagoler/www/biojade
BIO MORTAR	a laboratory management system designed specifically to deal with BioBricks		http://igem.uwaterloo.ca/BioMortar
BRICKIT	portable Web-based registry that helps synthetic biologists to plan, organize and track their local BioBrick samples		http://brickit.wiki.sourceforge.net
CLOTHO	a design environment to manipulate DNA sequence information and store the manipulated data as packaged 'parts' back to part repositories		http://biocad-server.eecs.berkeley.edu/wiki/index.php/Clotho_Development
GENEDESIGN/ BIOSTUDIO	a suite of algorithms that allow users to edit several features of protein coding sequences in an integrated development environment with a genome version control system	Richardson <i>et al.</i> (2006)	http://www.genedesign.org
GENETDES	a tool to design transcriptional networks with targeted behaviour	Rodrigo <i>et al.</i> (2007 <i>a,b</i>)	http://soft.synth-bio.org/genetdes.html
GENOCAD	a Web-based application guiding users through the design of part-based genetic systems	Cai <i>et al.</i> (2007)	http://www.genocad.org
LBS	a language for biological systems. It aims to provide a language for programming systems at the logical level of interactions between genes and proteins		http://www.inf.ed.ac.uk/publications/report/1270.html
OPENCELL (PCE NV)	an environment for creating and simulating arbitrary mathematical models	Beard <i>et al.</i> (2009)	http://www.cellml.org/tools/opencell
SYNBIOSS	a software suite for the quantitative simulation of biochemical networks using hybrid stochastic algorithms	Hill <i>et al.</i> (2008)	http://synbioess.sourceforge.net
TINKERCELL (ATHENA)	a tool for building, simulating and analysing genetic circuits	Chandran <i>et al.</i> (2009)	http://www.tinkercell.com

It was developed by an international community of systems biologists and software developers aiming to provide a common intermediate format for data sharing among various computer modelling software

applications. SBML is neutral with respect to programming languages and software encoding; however, it is encoded using XML (Bray *et al.* 2008). By supporting SBML as a format for reading and writing models,

different software tools (including programs for building and editing models, simulation programs, databases and other systems) can directly communicate and store the *same* computable representation of those models. Currently, there are over 160 software packages supporting SBML (http://sbml.org/SBML_Software_Guide/SBML_Software_Summary).

Another major standardization effort for machine-readable representation of biological pathways is CellML (<http://www.cellml.org>; Lloyd *et al.* 2008). It is an XML-based markup language originally developed by the Auckland Bioengineering Institute at the University of Auckland and affiliated research groups. It is similar to SBML, but more suited for multi-scale biological modelling capturing the structure and underlying mathematics of cellular models in a generic manner. CellML is growing in popularity as a portable description format for computational models, and groups throughout the world are using CellML for modelling or developing software tools based on it. Currently, there are a set of open-access tools and model storage databases based on CellML available at <http://www.cellml.org/tools>, including VIRTUAL CELL (Loew & Schaff 2001), a Java-based modelling and simulation environment that imports and exports CellML.

While the current standards are largely geared towards mathematical modelling of biological pathways and molecular interactions, it is envisaged that future developments would be able to incorporate synthetic biology constructs, such as protein coding sequences, biological parts and device-level information in their formalism. Some software tools, such as OPENCELL/PCEnv (table 1), support mathematical modelling for systems and synthetic biology constructs while using CellML as the native format for model storage. Similar efforts to enhance existing standards and synergize their usage would accelerate the adoption of consistent standards in biological engineering.

3.2.2. Semantics and annotation. For models to be informative, they need to be properly annotated with sufficient information attached to them enabling third parties to effectively use such models. BioPAX (<http://www.biopax.org>) is a collaborative effort to create a data exchange format for biological pathway data with ontological annotations. The main purpose is to facilitate data access, sharing and integration from multiple pathway databases by biologists. BioPAX supports representation of metabolic and signalling pathways, molecular and genetic interactions and gene regulation. Relationships between genes, small molecules, complexes and their states (e.g. post-translational protein modifications, mRNA splice variants, cellular location) are described, including biological events. BioPAX is complementary to other standard pathway information exchange languages, including SBML and CellML, as it focuses on large qualitative pathways and their integration rather than on mathematical modelling.

Since BioPAX targets annotation on pathways for existing organisms, it does not directly translate into

the synthetic biology domain. However, BioPAX-like ontological annotation may need to be initiated for BioBricks. This is important as the semantics of each device (BioBricks) may become unmanageable once engineered circuits exceed certain levels of complexity and millions of variations arise for similar functional devices. The current registry of biological parts can be a potential starting point for such ontological annotations. Other effort in annotation, particularly for model annotations, exists in the MIRIAM project (<http://www.ebi.ac.uk/miriam/>; Le Novère *et al.* 2005). The MIRIAM standard defines the minimum information that has to be attached to the model so that the model can be informative by itself.

The Systems Biology Ontology (SBO; <http://www.ebi.ac.uk/sbo/main/>) project endeavours to enhance the semantics of models, regardless of modelling approaches (refer to Brazma *et al.* (2006) and Strömback *et al.* (2007), for details on semantics approaches in systems biology). Again, standards like MIRIAM and SBO, although designed for molecular network models, can be extended to synthetic biology devices and circuit descriptions, particularly to define standard semantics for biological parts and devices.

3.2.3. Visual representation of models. Clear and unambiguous visualization is a fundamental step in applying computational techniques to biological models for scientific discovery. It is important to have standard visual representation languages for describing events and concepts in biology, such as biochemical interaction network, inter- and intracellular signalling and gene regulation. Most of the field is permeated with ad hoc graphical notations that have little in common between different researchers, publications, textbooks and software tools. While simplified notations can be used for purposes of elucidation, standardized representation of biological entities is of paramount importance for exchange between computational tools.

Thus, it is imperative to define a comprehensive set of graphical symbols that have precise semantics and detailed syntactic rules defining their use and are insensitive to restrictions of any medium or software, so that they can be used across a large array of applications to enhance data sharing, exchange and integration.

Definition of a common lingua franca is an important step in the standardization of biological representations and technologies. Biology has traditionally been a descriptive science, where the role of pictures and diagrams cannot be overstated. A community-wide effort is currently underway to define the SBGN (<http://sbgn.org>). The goal of SBGN is to define a set of visual glyphs and syntax, so that anyone can understand what the diagram exactly means much in the same vein as electrical circuit diagrams used by chip designers. The SBGN project was initiated by a group of biochemists, molecular biologists, modellers and computer scientists, with the aim of developing and standardizing a systematic and unambiguous graphical notation for applications in systems biology. SBGN is expected to be used not only by systems biologists,

but also by biologists of all disciplines, educators, publishers and students. Level 1 specification of the SBGN process diagram was released in August 2008 (Le Novère *et al.* 2008). The SBGN entity relationship diagram and activity flow diagram specifications now exist as draft proposals and are expected to be released in 2009. Currently, SBGN specification only defines visual icons and their syntax. Specification of the file format on how such graphics shall be stored and exchanged is the subject of future development.

Another effort in the direction of visualization is the SBML layout extension (Gauges *et al.* 2006). While the SBML file format does not provide for the storage of visual information for reaction graphs, the extension aims to provide a schema for describing the position and size of objects associated with biochemical reactions, thus providing the potential to render complex graphical standards with the SBML schema.

Early efforts to develop consistent schematics for synthetic biology constructs are represented by BOGL (http://openwetware.org/wiki/Endy:Notebook/BioBrick_Open_Graphical_Language), a graphical language for the formal description of standard biological parts. It aims to define symbolic notations for different biological parts, such as selection markers and restriction markers (Shetty *et al.* 2008). The standardization of visual representations in biological engineering presents new areas of research, particularly in enhancing existing standards to incorporate multi-level views—from detailed sequence level, binding site domain level views to molecular interactions, pathways and large-scale networks, in a consistent format. Visualization tools in the biological domain need to support such a *semantically zoomable* (Hu *et al.* 2007) multi-dimensional view of biological parts, devices and systems in the future.

3.2.4. Simulation of biochemical networks. While standardization is an integral part of the process of computational systems biology, the development of software tools for model building, distribution and running simulations is another important dimension. In this direction, plenty of model building and simulation tools are available. CELLERATOR (<http://www.cellerator.org>; Shapiro *et al.* 2003), COPASI (<http://www.copasi.org>; Hoops *et al.* 2006) and DIZZY (http://www.systemsbiology.org/Technology/Data_Visualization_and_Analysis/Dizzy; Ramsey *et al.* 2005) in the academic community and SIMBIOLOGY from Mathworks Inc. (<http://www.mathworks.com/products/simbiology/>) and PHYSIOLAB (Entelos Inc.; <http://www.entelos.com/physiolabModeler.php>) exist, each catering to different modelling techniques (refer to Hucka *et al.* (2004) for a review on SBML compliant simulators).

One of the most popular and widely used tools in this category is CELLDISIGNER (Funahashi *et al.* 2003)—a modelling and simulation tool to visualize, model and simulate gene-regulatory and biochemical networks. Two major characteristics embedded in CELLDISIGNER boost its usability to create/import/export models: (i) solidly defined and comprehensive graphical representation, specifically process diagram-based notations (Kitano *et al.* 2003; Kitano *et al.* 2005) of network

models and (ii) SBML as a model-describing basis, which functions as inter-tool media to import/export SBML-based models. Moreover, CELLDISIGNER provides the ability to embed or smoothly connect via Systems Biology Workbench (Sauro *et al.* 2003) different simulation/analysis packages that allow the simulation of the pathways using various simulation techniques such as COPASI, SBML ODE SOLVER (Machné *et al.* 2006), etc.

The simulation tools in the systems biology space currently focus on simulation of biological pathways and networks represented as biochemical reactions. On the other hand, simulation tools in synthetic biology allow the study of the dynamic behaviour of specific building blocks (like transcription constructs). As mentioned earlier, large-scale assembly of biological constructs would require multi-level modelling and simulation capabilities. For example, CELLDISIGNER currently does not have the capability to assist parts design and simulation. However, the software supports a plug-in architecture through which it is possible to establish close links with tools such as BIOJADE or to develop plug-ins for assembly and design of biological parts. Such synergistic integration across tools is essential to create consistent design platforms for synthetic biology.

4. BUILDING A COMMON FRAMEWORK

As reviewed in the previous section, various standards and technologies are already in place for practical use to cope with distinct levels of biological modelling and simulation. The goal of standardization is to fit together different pieces in a consistent manner to build a useful whole (Brazma *et al.* 2006). Integration of the various approaches and efforts to build a common framework to share accumulating knowledge in models is critical for advance in biological science across various disciplines. However, as elucidated in previous sections, several challenges need to be addressed to enhance existing standards in mathematical representations, visualization and modelling tools to accommodate the unique features of synthetic biology.

For representation and exchange of biochemical network models, SBML can be used as a good medium. The essential strength of the SBML format lies in the simulation of biological networks. SBML is defined as a set of standards to facilitate effective and efficient sharing of models with biochemical reactions. As mentioned earlier, semantic annotation of biological models is an important step in developing consistent, unambiguous representation. SBML, in conjunction with annotation standards such as MIRIAM, provides the ability to store and share information in a seamless, unambiguous fashion, which can be used as a medium for study and analysis in both the synthetic and systems biology communities.

There have already been some attempts to provide the tools for synthetic biological modelling, which allow converting the models into SBML format, such as ATHENA (Chandran *et al.* 2009) and ASMPARTS (Rodrigo *et al.* 2007a,b).

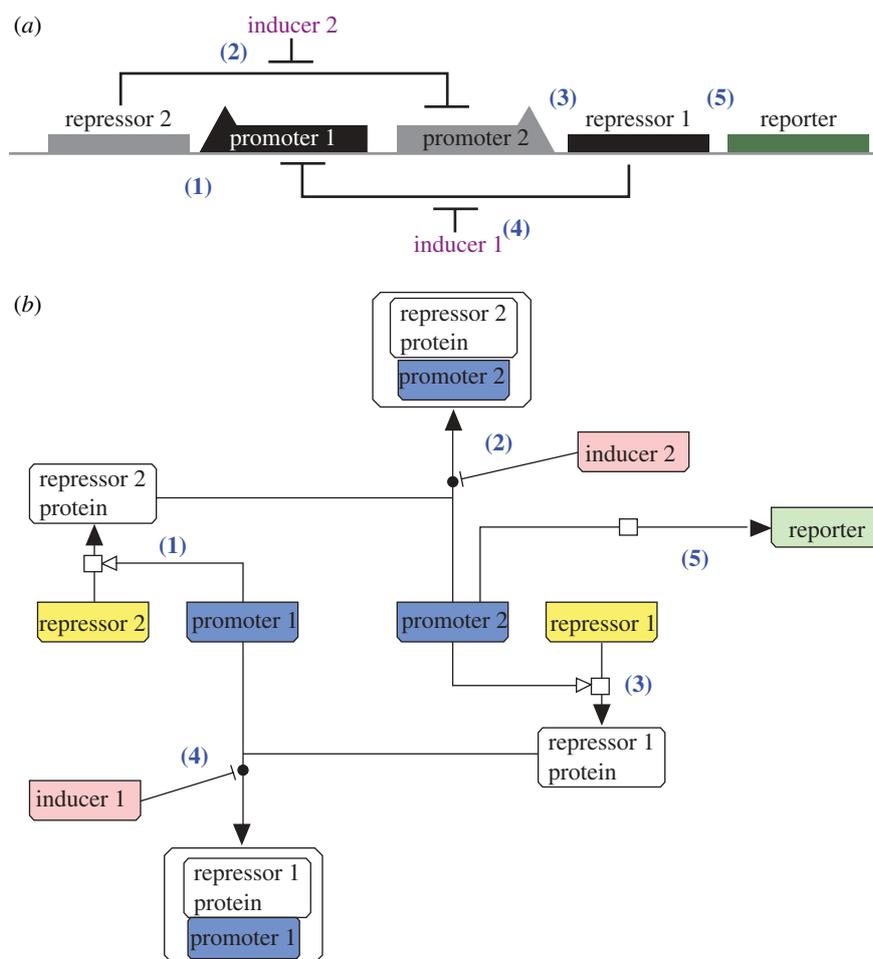


Figure 4. (a) Genetic toggle switch design taken from Gardner *et al.* (2000). (b) A possible representation of the genetic toggle switch in the process diagram format. Numbers in parentheses show the corresponding ‘processes’ in both (a) and (b) diagrams. (1) Promoter 1 promotes repressor 2 translation, (2) inducer 2 induces the promoter 2 activities, (3) promoter 2 promotes repressor 1 translation, (4) inducer 1 induces the promoter 1 activities, (5) promoter 2 promotes reporter translation. Arrows with filled heads describe the state transition, while arrows with open heads describe the stimulation to the process. T-shaped arcs represent inhibition.

There are several models that have been converted into SBML format and curated and stored in public databases. Elowitz & Leibler’s (2000) classic ‘repressilator’ model, for example, is already registered and available at BioModels database (<http://biomodels.net>; Le Novère *et al.* 2006). While efforts are underway to convert models of biological circuits into SBML format (Rodrigo *et al.* 2007*a,b*; Chandran *et al.* 2009), which can then be simulated using various SBML compliant simulation tools, a concerted effort is imperative to provide consistent visual representation of the various BioBricks—biological parts, devices and systems.

On the other hand, on-going efforts in the SBGN community provide a platform for defining a common visual language for biological systems—from engineered circuits to cellular pathways. The SBGN schema envisages supporting the representation of systems at different scales through process diagrams, activity flow or entity relationship diagrams. These different diagrammatic schemata allow the representation of knowledge at different levels of abstraction depending on the scope, accuracy of knowledge and other design requirements.

We provide some illustrative examples of possible representation of classical biological constructs in current process diagram notation. In the process diagram, nodes represent the states of biological entities and arcs describe biological process between the states. As a first example of developing an SBGN process diagram compliant representation of synthetic biological circuits, we consider the classical toggle switch model by Gardner *et al.* (2000). The genetic toggle switch model, constructed by Gartner *et al.*, toggles between stable transcription from either of two promoters in response to external signals (figure 4*a,b*). While figure 4*a* captures the toggle switch behaviour of the system, figure 4*b* endeavours to give a more mechanistic view, showing, for example, the mechanism of repression of the promoters by complex formation with the repressors. It is possible to capture different mechanisms of repression in the process diagram notation, where the biology of the process is known.

Another example illustrated here is a translational switch (Isaacs *et al.* 2004). While the classical diagram (figure 5*a*) captures the structural changes of the entities in the biological process, the process diagram

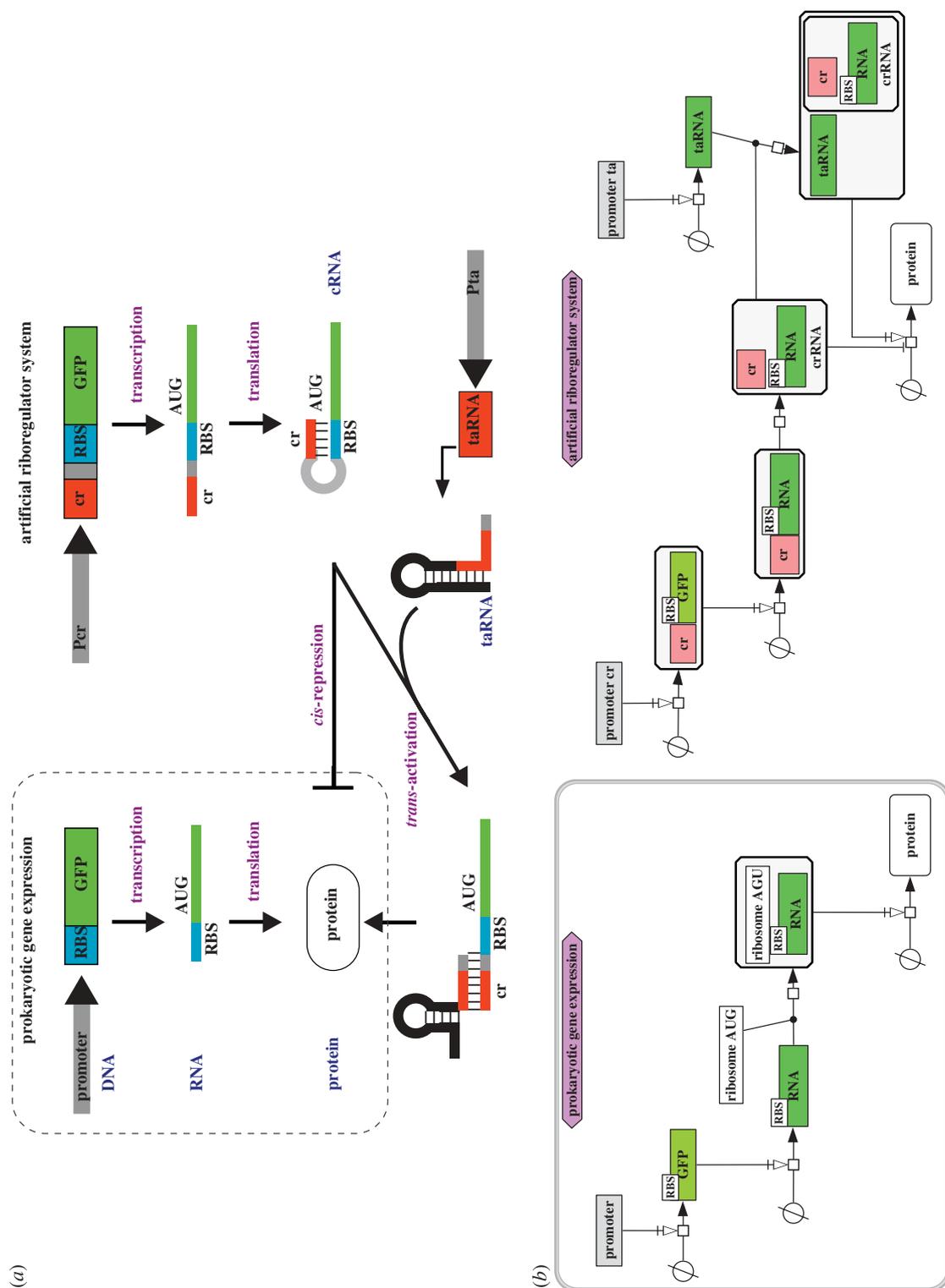


Figure 5. (a) An example of translational switch (Isaacs *et al.* 2004). (b) An example of equivalent riboregulator system represented in the process diagram format. (a) The artificial riboregulator system, shown in the graphical representation typically adopted by synthetic biology, and (b) interpretation in the process diagram notation. The circle with slash glyph represents the source of the creation of the entity or a state. The arrow with bar indicates necessary stimulation to the process. Lines connected with black dots represent the binding (e.g. RNA and ribosome, taRNA and crRNA).

(figure 5*b*) clearly identifies each step of the event and illustrates the mechanism of causes and effects of the processes.

As can be observed from these diagrams (figures 4 and 5), the genetic toggle switch as well as translational switch can be represented in process diagram manner. As both diagrams (figures 4*b* and 5*b*) are constructed in CELLDESIGNER and stored in the SBML format, it is possible to simulate the characteristics of the models once the dynamics of the processes is described in mathematical formulae in SBML format.

The visual elements (glyphs) in the current SBGN standard proposal may need to be enhanced to accommodate the different components used in the synthetic biology community, as the current standard evolves in an inter-community-wide collaborative manner. Sharing of symbols representing identical biological elements would further help in developing a common graphical lingua franca for biological engineering, on the same lines as in electrical circuit diagrams and other advanced engineering disciplines. We strongly believe that careful collaboration on the visual as well as model representation aspects between the two communities would foster the development of a standard graphical notation schema and accelerate the application of computational techniques.

5. SUMMARY

While the paradigm of systems biology endeavours a *holistic* understanding of the working principles of complex biological networks, *discovery by design* forms a key essence in synthetic biology, motivated by Richard Feynman's phrase 'What I cannot create, I do not understand' (Simpson 2006). In this perspective, the two disciplines hold the potential of complementing each other—analysis, modelling and simulation of biological networks can provide insights into the design and synthesis of de novo biological circuits. In this article, we provided an overview of the role of standardization in developing systematic and consistent computational platforms for biological systems. Particularly, graphical notations for visualization and schemata for mathematical modelling of such systems will play a pivotal role in enforcing engineering rigours in the study of biology. As elucidated here, existing standards of model representation (SBML) and graphical visualization (SBGN) prevalent in the systems biology community can be extended to incorporate synthetic biological constructs and models. Such collaborative efforts would pave the path towards a common, standardized schematic framework for understanding as well as engineering biological systems.

At the same time, development of a standard specification for genetic building blocks will force the community to describe each BioBrick in a well-defined form as exemplified in the equivalent circuit concept in electronics. Such a practice will not only benefit the synthetic biology community, but also the systems biology community because it triggers accumulation of knowledge on canonically defined genetic circuits. When synthetic biology matures as an engineering

field, the issues discussed in this paper will be the common practice and that is when it can be regarded as precision engineering.

This research is, in part, supported by ERATO-SORST Program of the Japan Science and Technology Agency (JST), Genome Network Project of the Ministry of Education, Culture, Sports, Science and Technology, NEDO Fund for International Standard Formation from the New Energy Development Organization and the Okinawa Institute of Science and Technology.

REFERENCES

- Ashburner, M. *et al.* 2000 Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29. (doi:10.1038/75556)
- Beard, D. A. *et al.* 2009 CellML metadata standards, associated tools and repositories. *Phil. Trans. R. Soc. A* **367**, 1845–1867. (doi:10.1098/rsta.2008.0310)
- Bray, T., Paoli, J., Sperberg-McQueen, C. M., Maler, E. & Yergeau, F. (eds) 2008 Extensible Markup Language (XML) 1.0, 5th edn. See <http://www.w3.org/TR/REC-xml/>.
- Brazma, A., Krestyaninova, M. & Sarkans, U. 2006 Standards for systems biology. *Nat. Rev. Genet.* **7**, 593–605. (doi:10.1038/nrg1922)
- British Telecommunications 2007 Pharma futurology: joined-up healthcare, 2016 and beyond. See http://www2.bt.com/static/i/media/pdf/BT_Pharma_Lowres.pdf.
- Brown, J. 2007 The iGEM competition: building with biology. *Synth. Biol.* **1**, 3–6. (doi:10.1049/iet-stb:20079020)
- Cai, Y., Hartnett, B., Gustafsson, C. & Peccoud, J. 2007 A syntactic model to design and verify synthetic genetic constructs derived from standard biological parts. *Bioinformatics* **23**, 2760–2767. (doi:10.1093/bioinformatics/btm446)
- Chandran, D., Bergmann, F. T. & Sauro, H. M. 2009 Athena: modular CAD/CAM software for synthetic biology. (<http://arxiv.org/0902.2598>)
- Dada, J. O., Paton, N. W. & Mendes, P. 2009 Systems Biology Results Markup Language (SBRML) level 1: structure and facilities for results representation. See <http://www.comp-sys-bio.org/static/SBRML-specs-15-04-2009.pdf>
- Deans, T. L., Cantor, C. R. & Collins, J. J. 2007 A tunable genetic switch based on RNAi and repressor proteins for regulating gene expression in mammalian cells. *Cell* **130**, 363–372. (doi:10.1016/j.cell.2007.05.045)
- Elowitz, M. B. & Leibler, S. 2000 A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338. (doi:10.1038/35002125)
- Funahashi, A., Morohashi, M., Kitano, H. & Tanimura, N. 2003 CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. *Biosilico* **1**, 159–162. (doi:10.1016/S1478-5382(03)02370-9).
- Gardner, T. S., Cantor, C. R. & Collins, J. J. 2000 Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342. (doi:10.1038/35002131)
- Gauges, R., Rost, U., Sahle, S. & Wegner, K. 2006 A model diagram layout extension for SBML. *Bioinformatics* **22**, 1879–1885. (doi:10.1093/bioinformatics/btl195)
- Goler, J. A. 2004 BioJADE: a design and simulation tool for synthetic biological systems. Master's thesis, MIT Computer Science and Artificial Intelligence Laboratory, MIT-CSAIL-TR-2004-036.
- Guido, N. J., Wang, X., Adalsteinsson, D., McMillen, D., Hasty, J., Cantor, C. R., Elston, T. C. & Collins, J. J.

- 2006 A bottom-up approach to gene regulation. *Nature* **439**, 856–860. (doi:10.1038/nature04473)
- Hasty, J., McMillen, D. & Collins, J. J. 2002 Engineered gene circuits. *Nature* **420**, 224–230. (doi:10.1038/nature01257)
- Hill, A. D., Tomshine, J. R., Weeding, E. M., Sotiropoulos, V. & Kaznessis, Y. N. 2008 SynBioSS: the synthetic biology modeling suite. *Bioinformatics* **24**, 2551–2553. (doi:10.1093/bioinformatics/btn468)
- Hoops, S. *et al.* 2006 COPASI—a COmplex Pathway Simulator. *Bioinformatics* **22**, 3067–3074. (doi:10.1093/bioinformatics/btl485)
- Hu, Z., Mellor, J., Wu, J., Kanehisa, M., Stuart, J. M. & DeLisi, C. 2007 Towards zoomable multidimensional maps of the cell. *Nat. Biotechnol.* **25**, 547–554. (doi:10.1038/nbt1304)
- Hucka, M., Finney, A., Sauro, H., Bolouri, H. & Doyle, J. 2003 The Systems Biology Markup Language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* **19**, 524–531. (doi:10.1093/bioinformatics/btg015)
- Hucka, M. *et al.* 2004 Evolving a lingua franca and associated software infrastructure for computational systems biology: the Systems Biology Markup Language (SBML) project. *IEE Proc. Syst. Biol.* **1**, 41–53. (doi:10.1049/sb:20045008)
- Isaacs, F. J., Hasty, J., Cantor, C. R. & Collins, J. J. 2003 Prediction and measurement of an autoregulatory genetic module. *Proc. Natl Acad. Sci. USA* **100**, 7714–7719. (doi:10.1073/pnas.1332628100)
- Isaacs, F. J., Dwyer, D. J., Ding, C., Pervouchine, D. D., Cantor, C. R. & Collins, J. J. 2004 Engineered riboregulators enable post-transcriptional control of gene expression. *Nat. Biotech.* **22**, 841–847. (doi:10.1038/nbt986)
- Itaya, M., Tsuge, K., Koizumi, M. & Fujita, K. 2005 Combining two genomes in one cell: stable cloning of the *Synechocystis* PCC6803 genome in the *Bacillus subtilis* 168 genome. *Proc. Natl Acad. Sci. USA* **102**, 15 971–15 976. (doi:10.1073/pnas.0503868102)
- Kitano, H. 2003 A graphical notation for biochemical networks. *Biosilico* **1**, 169–176. (doi:10.1016/S1478-5382(03)02380-1)
- Kitano, H., Funahashi, A., Matsuoka, Y. & Oda, K. 2005 Using process diagrams for the graphical representation of biological networks. *Nat. Biotechnol.* **23**, 961–966. (doi:10.1038/nbt1111)
- Kitano, H., Asada, M., Kuniyoshi, Y., Noda, I., Osawa, E. & Matsuura, H. 1997 RoboCup: a challenge problem for AI. *AI Mag.* **18**, 73.
- Le Novère, N. *et al.* 2005 Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat. Biotechnol.* **23**, 1509–1515. (doi:10.1038/nbt1156)
- Le Novère, N. *et al.* 2006 BioModels database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. *Nucleic Acids Res.* **34**, D689–D691. (doi:10.1093/nar/gkj092)
- Le Novère, N. *et al.* 2008 Systems biology graphical notation: process diagram level 1. *Nat. Precedings*. (hdl:10101/npre.2008.2320.1).
- Lloyd, C. M., Lawson, J. R., Hunter, P. J. & Nielsen, P. 2008 The CellML model repository. *Bioinformatics* **24**, 2122–2123. (doi:10.1093/bioinformatics/btn390)
- Loew, L. M. & Schaff, J. C. 2001 The virtual cell: a software environment for computational cell biology. *Trends Biotechnol.* **19**, 401–406. (doi:10.1016/S0167-7799(01)01740-1)
- Machné, R., Finney, A., Müller, S., Lu, J., Widder, S. & Flamm, C. 2006 The SBML ODE Solver Library: a native API for symbolic and fast numerical analysis of reaction networks. *Bioinformatics* **22**, 1406–1407. (doi:10.1093/bioinformatics/btl086)
- Peccoud, J. *et al.* 2008 Targeted development of registries of biological parts. *PLoS ONE* **3**, e2671. (doi:10.1371/journal.pone.0002671)
- PricewaterhouseCoopers 2008 Pharma 2020: virtual R7D— which path will you take? See <http://www.pwc.com/extweb/pwcpublishings.nsf/docid/91BF330647FFA402852572F2005ECC22>.
- Ramsey, S., Orrell, D. & Bolouri, H. 2005 Dizzy: stochastic simulation of large-scale genetic regulatory networks. *J. Bioinformatics Comput. Biol.* **3**, 415–436. (doi:10.1093/bioinformatics/btm231)
- Richardson, S. M., Wheelan, S. J., Yarrington, R. M. & Boeke, J. D. 2006 GeneDesign: rapid, automated design of multikilobase synthetic genes. *Genome Res.* **16**, 550–556. (doi:10.1101/gr.4431306)
- Rodrigo, G., Carrera, J. & Jaramillo, A. 2007a Asmparts: assembly of biological model parts. *Syst. Synth. Biol.* **1**, 167–170. (doi:10.1007/s11693-008-9013-4)
- Rodrigo, G., Carrera, J. & Jaramillo, A. 2007b Genetdes: automatic design of transcriptional networks. *Bioinformatics* **23**, 1857–1858. (doi:10.1093/bioinformatics/btm237)
- Rullmann, J. A., Struemper, H., Defranoux, N. A., Ramanujan, S., Meeuwisse, C. M. & van Elsas, A. 2005 Systems biology for battling rheumatoid arthritis: application of the Entelos PhysioLab platform. *IEE Proc. Syst. Biol.* **152**, 256–262. (doi:10.1049/ip-syb:20050053)
- Sauro, H. M., Hucka, M., Finney, A., Wellock, C. & Bolouri, H. 2003 Next generation simulation tools: the Systems Biology Workbench and BioSPICE integration. *Omics* **7**, 355–372. (doi:10.1089/153623103322637670)
- Shapiro, B. E., Levchenko, A., Meyerowitz, E. M., Wold, B. J. & Mjolsness, E. D. 2003 Cellerator: extending a computer algebra system to include biochemical arrows for signal transduction simulations. *Bioinformatics* **19**, 677–678. (doi:10.1093/bioinformatics/btg042)
- Shetty, R. P., Endy, D. & Knight, T. 2008 Engineering BioBrick vectors from BioBrick parts. *J. Biol. Eng.* **2**, 5. (doi:10.1186/1754-1611-2-5)
- Simpson, M. L. 2006 Cell-free synthetic biology: a bottom-up approach to discovery by design. *Mol. Syst. Biol.* **2**, 69. (doi:10.1038/msb4100104)
- Stricker, J., Cookson, S., Bennett, M. R., Mather, W. H., Tsimring, L. S. & Hasty, J. 2008 A fast, robust and tunable synthetic gene oscillator. *Nature* **456**, 516–519. (doi:10.1038/nature07389)
- Strömbäck, L., Hall, D. & Lambrich, P. 2007 A review of standards for data exchange within systems biology. *Proteomics* **7**, 857–867. (doi:10.1002/pmic.200600438)
- The Economist 2005 Models that take drugs. *The Economist Report*, 11 June 2005.