BIOPHARMACEUTICS & DRUG DISPOSITION
Biopharm. Drug Dispos. **34**: 508–526 (2013)
Published online 4 December 2013 in Wiley Online Library
(wileyonlinelibrary.com) **DOI**: 10.1002/bdd.1875

Invited Review

Toward an integrated software platform for systems pharmacology

Samik Ghosh[a,b,*], Yukiko Matsuoka[a,c], Yoshiyuki Asai[d], Kun-Yi Hsin[d], and Hiroaki Kitano[a,b,d,*]
[a]*The Systems Biology Institute, 5F Falcon Building, 5-6-9 Shirokanedai, Minato, Tokyo 108-0071, Japan*
[b]*Disease Systems Modeling Laboratory, RIKEN Center for Integrative Medical Sciences, 1-7-22 Suehiro-Cho, Tsurumi, Yokohama 230-0045, Japan*
[c]*JST ERATO Kawaoka Infection-induced Host Response Project, 4-6-1 Shirokanedai, Minato, Tokyo 108-8639, Japan*
[d]*Okinawa Institute of Science and Technology, 1919-1, Tancha, Onna-son, Kunigami, Okinawa 904-0412, Japan*

ABSTRACT: Understanding complex biological systems requires the extensive support of computational tools. This is particularly true for systems pharmacology, which aims to understand the action of drugs and their interactions in a systems context. Computational models play an important role as they can be viewed as an explicit representation of biological hypotheses to be tested. A series of software and data resources are used for model development, verification and exploration of the possible behaviors of biological systems using the model that may not be possible or not cost effective by experiments. Software platforms play a dominant role in creativity and productivity support and have transformed many industries, techniques that can be applied to biology as well. Establishing an integrated software platform will be the next important step in the field. © 2013 The Authors. *Biopharmaceutics & Drug Disposition* published by John Wiley & Sons, Ltd.

Key words: systems pharmacology; systems biology; integrated software platform

Introduction

Systems biology emerged in the mid-1990s aimed at a system-level understanding of living organisms and their applications in various areas including medicine and biotechnology [1–4]. System-level studies are often built upon molecular and genetic-level findings, as well as involving 'omics' studies such as genomics, proteomics and metabolomics. Aspects of systems biology that many researchers are facing are essentially fights against the complexity of systems, the vastness of the data and scattered pieces of knowledge that all has to be integrated to make sense of and be useful. It is not possible for humans to simply look at them and expect to extract useful knowledge or to integrate them coherently without systematic aids from computational tools. Thus, computational tools are critically important in systems biology because anything we do requires computational aids to varying degree and aspects.

Lessons from industrial experiences in electronics, aviation and other areas strongly suggest that powerful tools that are aligned as a platform can drive transformation of an industry [5]. One of the best examples is computer graphics. Computer graphics used to be on hand for only a handful of computer geeks and artists who were eager to uptake new technology for their expression. However, with the development of a range of platforms from OpenGL (an open source graphics software code library) to commercial 3D rendering and animation packages that confer

*Correspondence to: The Systems Biology Institute, 5F Falcon Building, 5-6-9 Shirokanedai, Minato, Tokyo 108–0071 Japan.
E-mail: ghosh@sbi.jp;kitano@sbi.jp

Received 29 September 2013
Accepted 6 October 2013

them with industrial standards of representation, it has turned into a powerful set of tools in everyone's hands. As a result, large numbers of computer graphics designers and artists have emerged, the game industry has flourished, and industrial computer aided design and desktop publishing (DTP) has replaced the old approach, and today Hollywood studios are fortified with high performance computers. All these transformations and the explosion of industry-scale systems would not have been possible without the emergence of powerful software platforms.

Biological sciences are no exception. This is particularly true for systems biology and its application in areas such as systems drug design because their success depends on sophisticated data handling, modeling, integrated computational analysis and knowledge integration. One of the central approaches is to create computational models that enable us to predict the behaviors of biological systems at multiple scales thereby helping us to understand the mechanism behind them as well as to predict the impacts of perturbations, including drugs. This article reviews the current status and issues in software tools, the need for integrated software platforms and challenges for the future.

Software Needs in Systems Biology and Systems Pharmacology

What is expected in software platforms is the capacity to support novel biological discoveries, drug design and other life science research by their analysis, prediction, explanation, sharing and integration capabilities, thereby enabling research and development that would not be possible without such platforms. A set of software tools and resources has to be defined and provided to best achieve these objectives. A typical workflow for computational analysis in engineering involves data acquisition, modeling and analysis. Prediction and explanation capabilities are associated with this cycle, and integration and sharing of knowledge are involved to best achieve and sustain these capabilities. This cycle of workflow is relatively universal and it can be applied for biological research (Figure 1).

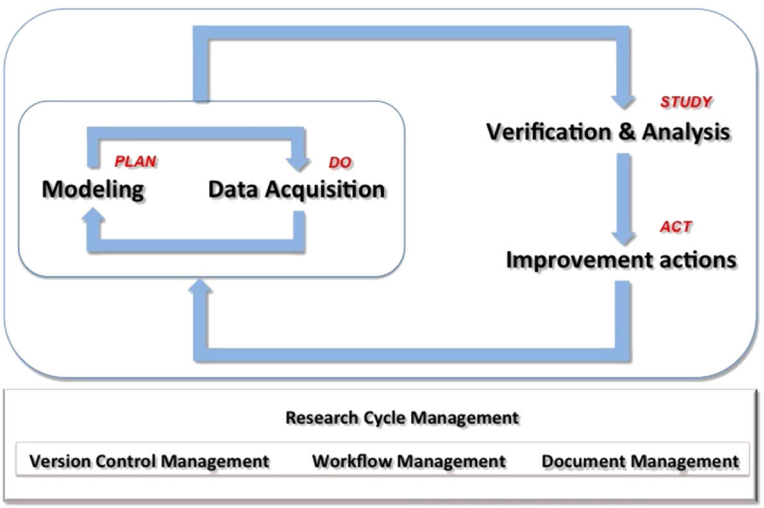

Figure 1. Workflow of computational tasks in systems biology. (a) Research cycle with computational modeling and analysis involved. In this cycle, modeling is not just an interaction map and computational model development, but also involves experimental planning because data suitable for model verification and analysis have to be acquired through experiments. At the same time, what can be actually measured and at what accuracy affect the proper modeling strategy. Acquired data and models will be used for various analyses. In most cases, analysis results are used for further improvement of models and experimental procedures. This cycle is consistent with what is known as the 'Deming cycle' (also known as the 'Shewhart cycle') of Plan-Do-Study-Act (PDSA) cycle that is the standard practice for business process quality improvement. Given the nature of biological model development and analysis, it is obvious that this process involves a highly interactive and flexible process. In some aspects, it requires a kind of extreme programming (XP) that assumes multiple iterations of different time spans so that potential problems can be detected earlier in the development phase and can be quickly fixed

DOI: 10.1002/bdd

The central pillar of the whole process is the development and verification of computational models that are represented at the proper level of abstraction and scope required to best answer the biological questions being asked. Computational models, or *in silico* models, play an important role in systems biology because these are a tangible implementation of hypotheses and verified knowledge of the biological systems and processes being investigated. In the engineering design process, computer models are assumed to be an *in silico* replica of a specific aspect of the artificial objects being manufactured. In the biological investigation process, computer models are supposed to be an *in silico* replica of a specific aspect of a biological system as recognized and hypothesized.

When a research project adopts this view, the entire research cycle has to be aligned to achieve the goal of creating and verifying the computational model. This means 'wet' experiments will be designed to provide the data required for model construction and verification. At the same time, the kind of models that can be developed and verified properly depends on the experimental capabilities available for the project. This process of interactive design of experiments and definition of model scope is one of the most important steps in systems biology research, and the parties involved need to share what are the central scientific questions being asked in the project, and hence the hypotheses to be tested.

Within this context, it is all about creating hypotheses and verifying them. There are two typical, contrasting but not contradicting, approaches for hypotheses generation. One is a top-down and data-driven approach, and the other is a bottom-up and often literature-based approach. The top-down data-driven approach makes the most use of large-scale data to construct possible hypotheses that are represented in the form of a network derived from the data set using certain algorithms. This is often called a 'weak method' meaning that the hypotheses are generated without prior knowledge of the domain, but merely from statistical inference from the data. Thus, networks mostly represent association and correlation among protein and genes rather than causality and mechanisms. The network can be used for further modeling, designing the next round of experiments, interpretations of experimental results, etc. On the contrary, the bottom-up approach tries to create detailed and highly precise models by integrating knowledge taken from the literature and well-curated databases. This approach, that can be called a 'strong method', creates in-depth mechanistic models, but is extremely labor-intensive and limited in coverage of the model. Analysis of the models created also requires a set of computational tools either involving numerical computation or a more quantitative and logic-oriented approach, and generates vast computational results as well.

At the same time, modeling and simulation has traditionally been a strong pillar in pharmacology in understanding the mechanisms of drug effects, side effects and adverse events. Recently, the community has increasingly acknowledged the need for a global, network-based understanding of drug–target interactions [82]. Systems pharmacology or network pharmacology aims to obtain insights into drug actions and adverse events in the context of the biological network around the therapeutic targets rather than in isolation [83–86]. Such an holistic approach requires an enhancement of existing PK/PD systems analysis techniques with deeply integrated computational systems biology approaches across and between multiple scales of organization – from chemistry to structural biology, pathology/cell physiology and to organ-level specialties (cardiology, nephrology etc.) and clinical care [83].

In particular, in a recent NIH study group white paper reviewing the state of the art in systems biology and pharmacology [87], special emphasis was laid on the emerging discipline of quantitative systems pharmacology (QSP), defined as '*.. as an approach to translational medicine that combines computational and experimental methods to elucidate, validate and apply new pharmacological concepts to the development and use of small molecule and biologic drugs. QSP will provide an integrated 'systems- level' approach to determining mechanisms of action of new and existing drugs in preclinical and animal models and in patients*.' Amongst the various expected outcomes envisaged in the paper, the working group highlighted the specific need for '*new approaches and tools to link preclinical and clinical studies of drugs and disease*'.

DOI: 10.1002/bdd

As mentioned earlier, various tools for modeling, simulation and pharmacodynamics systems analysis have been applied extensively in pharmacology. A comprehensive summary of such network modeling tools and their application in drug discovery has been reviewed recently by Csermely *et al*. [88]. While models occupy an unique position in systems biology and systems pharmacology, many insights can be gained during the process of creating the model, rather than from the model itself. A software platform should be designed flexibly to support various deviations of research processes from the assumed process. At the same time, certain workflows are assumed as a typical procedure of model development and verification. This review presents a view of how each task in the workflow can be interlinked to generate and verify the hypotheses-driven approach based on computational models.

In the next few sections, software tools, issues and challenges of typical data handling, model development and analysis procedures will be described.

Data Management

The proper acquisition and handling of data is critically important for both the generation of hypotheses and their verification. The quality of hypotheses depends on the nature of the data. The rapid development of high throughput experimental techniques (next generation sequencers, genome wide association study (GWAS) tools, proteomics, transcriptomics and metabolomics studies) is transforming life science research into *big data* science [6,7]. Vast amounts of data are generated on a regular basis in individual research laboratories as well as in large-scale systems biology projects spanning multiple groups spread across different countries. With such an exponential growth of experimental data production, data management becomes fundamental. This is particularly true for systems biology, where heterogeneous sources of data need to be aggregated, analysed and interpreted in a systems-level context.

The heterogeneity and complexity of the data pose the greatest challenges in data management for biological and preclinical research [8]. Figure 2

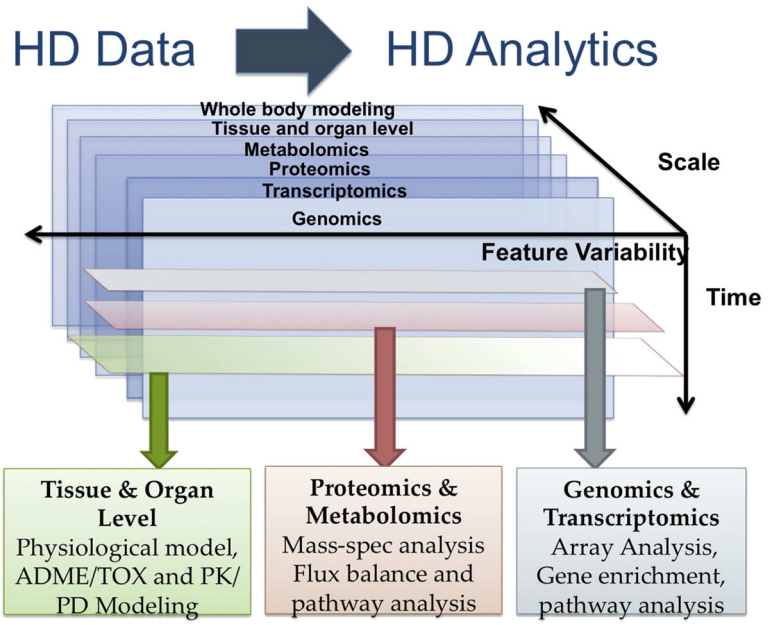

Figure 2. From HD data to HD analytics. Biological data are characterized by heterogeneity and diversity in multiple scales spanning across genomics, transcriptomics, proteomics and metabolomics to tissue, organ and whole body modeling. Each layer captures the dynamics across various features (parameters) across different temporal scales. Depending on the analysis focus, various techniques are applied for vertical or horizontal integration, including but not limited to PK/PD modeling, pathway and flux balance analysis to data-driven and statistical analysis of transcriptomics or proteomics data

DOI: 10.1002/bdd

captures the complexity and diversity of biological data across multiple scales. Such HD data (high dimension, high diversity, high definition) calls for HD analytics techniques necessitating vertical integration across pathways, cells, tissues and organs as well as horizontal integration linking transcriptomics, proteomics and metabolomics within a cell. This is further aggravated by huge volume, context dependency and provenance [9], lack of well-established exchange standards and globally unique identifiers, which hinder mapping and integration of the results [10]. While there are already plenty of tools for bioinformatics and data management for the life sciences [11–14], systems biology requires yet another set of tools and standards.

The role of standards and protocols in information exchange and management has been underscored by their widespread adoption in engineering, product manufacturing and recently in information technology. Data representation and communication standards for systems biology and bioinformatics have developed as a distinct field [15], leading to the development of a plethora of standards covering different stages of the research pipeline. Standards for data management have been focused on three core aspects: minimum information for data description, file format and ontology [8]. These standards combined enable the consistent and rich annotation of data generated, and can thus be used effectively for further analysis.

Informatics tools and services have played a central role in the analysis and interpretation of large-scale biomedical data both in basic research as well as in preclinical and clinical trials. Traditionally, the options for informatics tools have revolved around in-house custom tool development, the use of open source software (such as R programming language) or commercial data management software. Current data management systems include spreadsheet-based system, web-based document sharing systems and laboratory information management systems (LIMS), and Workflow Management Systems [8]. An integrated LIMS and Workflow Management System would play an even greater role as the number and complexity of experiments increases, and there are several examples of these systems such as Taverna [16], caGrid [17], Bio-Steer [18] and KNIME (Link 9). However, the lack of interoperability and interfaces with other software tools pose a serious bottleneck for wider adoption.

Closely associated with the management of a genome-wide data-set is their visualization and analysis. Several tools provide different levels of functionality to visualize molecular networks, expression profiles and genomic data [19,20], Cytoscape [21] being one of the most widely used software in the community.

It is the plethora, rather than the paucity, of databases, standards and tools that provides the major challenge towards developing a platform that is integrated and consistent. A key driver in a systems approach is the ability to integrate data from diverse sources and to apply computational analysis techniques to generate global insights of the biological systems and hypotheses to be tested. Thus, an integrated platform should reconcile the existence of multiple standards and tools, and develop technologies that allow them to co-exist and inter-operate.

Data-driven Network Inference

Data-driven, network-based approaches have been actively developed over the past decade [22] to infer a causal relationship between molecular entities from experimental data. A data-driven approach algorithmically generates a set of hypotheses on the possible relationships among genes and molecules depending on the data and inference methods used. The first generation studies were focused on finding patterns in gene expression profiles to distinguish disease and healthy states at the molecular level, as elucidated in the classical study in breast cancer prognosis [23]. The limitation of using a one-dimensional expression profile in reconstructing the complex interactions in living systems has led to the development of new methods correlating genome scale DNA variations with disease traits in identifying disease susceptible genes [24–26].

To further improve the approach, researchers have combined genotypic data with phenotypic data (gene expression, protein and metabolite states) to develop integrated models of disease networks [27–29]. With the growing availability of high-throughput data in multiple dimensions, recent efforts have focused on harnessing knowledge from

DOI: 10.1002/bdd

genotypes, gene expression, protein–protein interaction, DNA–protein binding and complex binding data to construct probabilistic, causal gene networks [30–32]. These models have been predominantly based on Bayesian inference techniques – computing the probability of a hypothesis (in this case the relationship between two molecular entities) based on some kind of evidence or observations (known as priors). However, several alternative techniques have also been applied with varying degrees of success such as mutual information (MI) approaches [33,34] and others [35–40].

Although numerous inference algorithms have been proposed and even incorporated in commercial data analysis packages, the reconstruction accuracy of such techniques requires a careful and systematic evaluation, benchmarking and verification. An interesting effort in standardized benchmarking for network reconstruction has been initiated through the DREAM project (Dialogue for Reverse Engineering Assessments and Methods) [Link 10] – which attempts to evaluate the different paradigms influencing network inference. Analysis of DREAM results (DREAM2 and DREAM3) reveals that algorithms complement each other in a highly context specific manner. A community-driven, *consensus based reverse-engineering*, which aggregates the results from the best performing algorithm has been advocated as a way forward for high quality network inference. One explanation for why such a community-based approach performs better than the best algorithm in the pool is the existence of strengths and weaknesses of each algorithm, thus multiple algorithms need to be used to compensate for such characteristics. This is an interesting observation and consistent with the reason why IBM's DeepQA system was successful in the Jeopardy! Challenge [41].

Deep Curation

Contrasting with the data-driven network inference is the deep curation approach. The deep curation approach creates a detailed molecular interaction map by integration of available knowledge using publications, databases and other information sources [42,43]. This is essentially a large-scale knowledge integration approach. Unlike the data-driven approach in which hypotheses about interactions are generated automatically, the deep curation approach manually or semi-manually constructs the model that means it is easier for researchers to add their own hypotheses onto the model. The map has to be precise and sufficiently in-depth in order to be useful. While a data-driven approach may generate networks representing inferred causality or an association of behaviors for molecules and genes, it does not provide mechanistic details nor confirm causality. A deep curation approach can provide mechanistic details for each interaction.

In-depth mechanistic-level models are essential not only for precise computer simulation and understanding of biological mechanisms, but also for the proper evaluation of potential drug targets (Figure 3). Models that only represent interactions without details of their mechanisms have limited utility. This is critical because only in the in-depth model, is it possible to identify precisely which interaction leads to target, thus a proper choice of lead compound and their optimization strategy can be defined. Deep curation requires an open-ended knowledge assembly from publications, pathway databases, as well as from high-throughput data [42–45]. The current 'Gold Standard' is manually curated maps carefully built based on literature and various data resources by a small group of people who spend months on the same pathway to the extent they would be familiar with almost every work reported [46]. Recent progress in standard formation such as SBML [47], SBGN [48] and MIRIAM [49] and software such as CellDesigner [50] have made it possible to develop such detailed models.

Challenges of the deep curation approach are up-dating and validation. Manually creating large-scale network maps from the literature is extremely labor-intensive and stressful work. Also, it is very difficult to maintain motivation for continuously up-dating the map to keep up with new discoveries for many years. Automated literature mining has been investigated extensively, but is nowhere near the stage needed to replace human curators. Validation is another challenge for deep curation. Although standards such as SBGN and commonly used tools such as CellDesigner have been made available, the interpretation of papers and data presented depends on someone reading them and transfering them

DOI: 10.1002/bdd

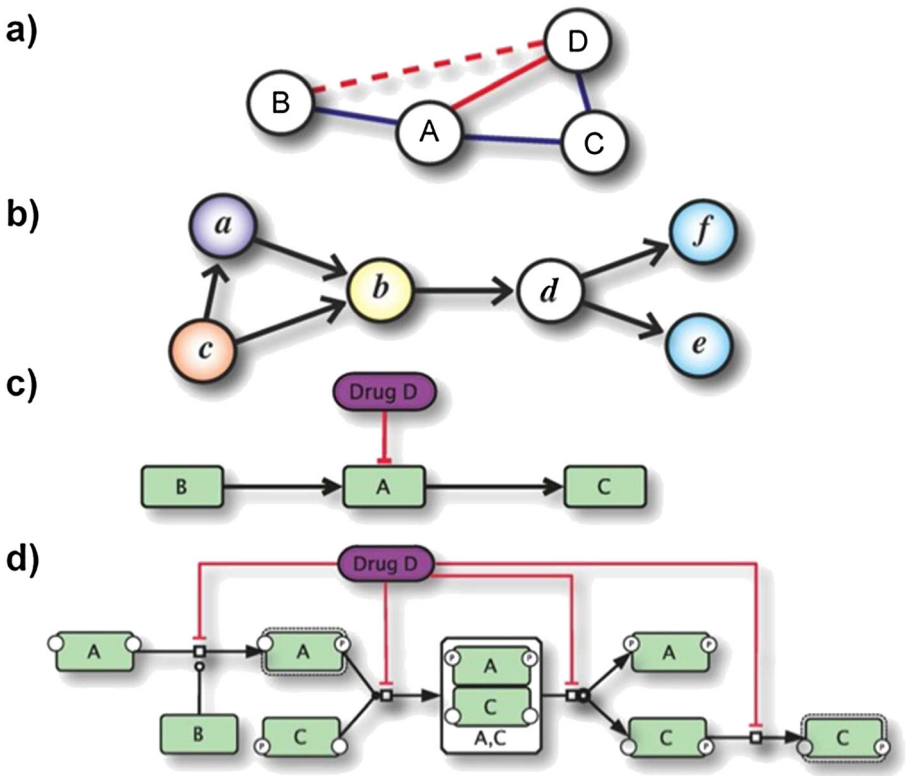

Figure 3. Molecular interactions at different levels of abstraction. (a) A simple visualization of a part of the protein–protein interaction (PPI) data. It is visualized as nodes and undirected arcs. (b) A possible network inferred by Bayesian inference. Nodes represent genes and are connected via a directed arc that shows statistically inferred association characteristics, such as correlational or anti-correlational. It should be noted that this does not necessary represent causal relationships. (c) Conventional diagram that indicates the flow of activation and inhibition. This type of diagram is often called an activity flow diagram. (d) SBGN process diagram representation that depicts the detailed mechanisms of interactions. For example, one may try to identify a drug target. With an abstract model of interactions, it may only represent an effect inhibition for a specific molecule (molecule A) as seen in (c). With an in-depth mechanistic model, it is now possible to capture the fact that there are four processes of interaction and that inhibition of any one of them would appear to inhibit the same molecule, but may have a different overall efficacy and collateral effects (d)

into interaction maps. Validation requires expert knowledge of the biology and the ability to decipher the literature evidence into the pathway diagram. Recruiting experts and assigning them for pathway curation, and their working together to build the integrated pathway itself is a big challenge. While community-based development and refinement of pathways has been proposed as seen in WikiPathways [51], following the success of Wikipedia, it has not been taken up seriously by the community so far, perhaps due to the issue of incentives in biological community.

In addition, there is an issue of how two parallel approaches can be integrated. The data-driven approach is comprehensive, unbiased and has the potential to uncover novel interactions that have not been identified before. Deep curation integrates knowledge from publications and databases so that novel discovery cannot be expected unless explicitly added, but each interaction incorporated into the map is backed by specific experiments and independently confirmed. It would be ideal if both approaches could be combined to enhance the strength of each approach. Certain percentages of interactions inferred by the data-driven approach are likely to be confirmed by the deep curation approach and some can be clearly rejected, then the remaining inferred interactions deserve further study, hence research resources can be focused on these hypotheses. Extensive software support and research are needed to enable the integration of the two approaches.

DOI: 10.1002/bdd

In silico Simulation Models

Simulation is an indispensable tool in all engineering designs and has been applied successfully in the automobile, aerospace and telecommunication industries for many decades. Computational fluid dynamics (CFD), for example, is an essential design process in aircraft design, ship design and automobile design. Any high rising building has to carry out 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 aircrafts' flown under simulated conditions before hitting the manufacturing floor [89]. While the application of advanced modeling and simulation techniques have resulted in immense cost-savings and standardized procedures for such R&D intensive industries, the pharmaceutical industry has historically lacked these approaches, leading to astronomical costs in drug development (~25% of its revenue, almost twice that of other knowledge-driven industries [89]).

While an appreciation and awareness for the potential benefits of the computational approach in the future of biological sciences and drug design has been on an increasing trajectory in industry and academic circles [55,60], it is important to keep in perspective the unique hurdles and significant challenges in applying *in silico* techniques in the life sciences. Identification of the specific needs for computational tools in the pharmaceutical industry, together with an open, collaborative mindset between all players would form the key stepping stones in developing safer, efficacious and cost-efficient drugs for complex diseases such as cancer.

Issues and Challenges

The adoption of simulation techniques in the life sciences requires careful and detailed consideration of the unique challenges faced in trying to fathom the complex interaction between the different entities at multiple scales – from cells, to tissues and organs to whole human body and host–pathogen interactions. There exists a series of issues that has to be addressed before simulation can be accepted as normal practice in the industry.

First, there is a set of fundamental technical issues to be solved to further improve the accuracy of simulation. Different flavors of simulation technologies exist, from deterministic, differential equation-based systems to non-deterministic, stochastic techniques, agent-based and discrete-event based simulations. Each presents a unique set of assumptions and system conditions which need to be considered before their successful application to specific biological problems, as elucidated schematically in Figure 4 [90]. Cellular modeling or physiological modeling with molecular details will require the integration of heterogeneous computational models that are on different spatial and temporal scales, and the basic equations still need to be defined.

The purpose and goal of a simulation system applied in drug design should be clearly defined: as in other fields such as Formula 1 aerodynamics design, where the goal is to design an aerodynamically optimal body with maximum down force and minimum drag. This forms a key step in determining the eventual success of a biological simulation system and in defining the boundaries of the system. For example, Merrimack Pharmaceuticals [91] used computer simulation to identify a novel drug target for a specific cancer that resulted in the development of a monoclonal antibody for ErbB3, now in clinical trial. Simulation models need to be designed sufficiently to capture the essential features to accomplish the task defined, but features that are unlikely to affect prediction accuracy of the given task may be ignored.

Sophisticated models with molecular details that can predict cellular behaviors under various conditions are crucial for elucidating system-level properties of cellular systems, such as their robustness and the underlying principles of cellular functions. Such models should be able to provide predictions on how cells and organs respond when certain perturbations, such as drug administration, are given. While there are some successful cases of computational modeling of limited scale biological networks, there is no established method for developing high precision models.

While time is a key component of biological systems, spatial dynamics also play an important

DOI: 10.1002/bdd

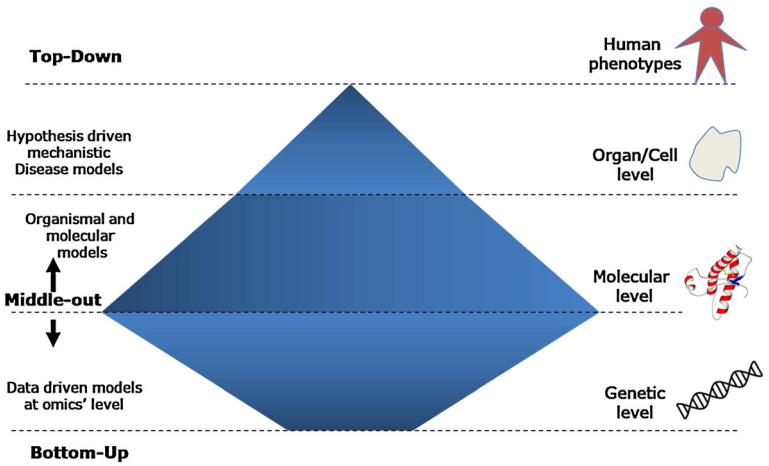

Figure 4. Multi-scale modeling and simulation. Depending on the level of focus, modeling and simulation techniques can be applied from the genetic level upwards to the molecular level (bottom-up) or starting from abstract physiological models (at organ or tissue level) downwards to molecular and pathway levels (top-down)

role in elucidating their behavior. This is particularly true for multi-scale models involving cells, tissues and organs where the spatial heterogeneity of different components governs their behavior in time. The PDE solvers available in engineering tools such as Matlab® can be applied to model biological systems in time and space. In addition, finite element method (FEM) based PDE solvers may be used to compute the elastic displacement of the system. There some popular software, e.g. OpenFEM (Link 4) and FreeFEM (Link 5), for this task.

It is often the case that the stochastic behavior of molecules significantly affects the outcome of cellular and physiological behaviors of the system. Thus, stochastic computation plays an important role in biological simulation. In typical cases, Gillespie's algorithm is used for approximation of the chemical master equation (CME), captures the stochastic (random) behavior of molecular interaction and has been successful in elucidating the dynamics behind gene transcription and translation processes [56,57]. Some such examples can be seen in the *E. coli* fate decision under phage infection [61].

Several other techniques, such as agent-based modeling [62] and process algebra (PetriNets) [63] have also been applied to study the behavior of specific biological systems.

Apart from the challenges in model calibration and verification, large-scale simulation systems require computational resources and efficient numerical algorithms. Further, standards for simulation experimental definition (MIASE [64]), parametric conditions (SED-ML) and result dissemination (SBRML) need to be developed for high precision simulation models. In this direction, the systems biology markup language (SBML) [47] has been widely adopted as the standard format for the storage and sharing of dynamic models. This has led to the development of inter-operability among a suite of SBML compliant simulation and analysis tools.

Lessons from Computational Fluid Dynamics

While there are some successful cases of computational modeling of limited scale biological networks [54,55,58–60], there is no established method for developing high precision models. It should be remembered that even the most successful computational approach so far, computational fluid dynamics (CFD), took decades of research to become practically useful. So, it is perhaps too optimistic to expect such a technology to be easily obtainable, although there is no need to think that it will never be realized. At the same time, we need to move forward to solve a series of problems. We can learn some lessons from CFD as to how we

DOI: 10.1002/bdd

can improve biological simulation, as there are several reasons that make CFD indispensable today.

First, CFD uses the Navier-Storkes equation that is a very suitable motion equation to describe most of fluid dynamics. It assumes a homogeneous medium flowing through objects. In addition, it is fundamentally a monolayer system. Cellular modeling or physiological modeling with molecular details requires the integration of heterogeneous computational models that are on different spatial and temporal scales, and the basic equations still need to be defined. Most interaction-network simulations use the Michaelis–Menten equation or a similar equation that assumes a certain ideal condition. However, these assumptions might be unwarranted in a crowded molecular environment in which reactions and molecular movements are constrained in space. The challenging issue of the integration of different computational models for interaction networks and cellular structures, macroscopic dynamics at the cellular and molecular levels and processes with different timescales has to be resolved.

Second, integration of computational modeling and experimental data acquisition has to be promoted. Looking at the design process for a Formula 1 racing car, an iterative design cycle is established. First, numbers of designs are tested with CFD. Some designs will be tested using a wind tunnel. One or two designs are actually implemented and tested in a test course, and one design is selected for the final production. In this process, CFD models are calibrated against wind tunnel data for further improvement of accuracy, instead of data from test course or from actual racing telemetry data. This is because only a wind tunnel enables highly controlled experiments so that high precision data with controlled initial and boundary conditions can be obtained. This comparison delivers two messages. First, we need to develop highly controllable experimental systems comparable to the wind tunnel in aerodynamics. This means that we need to be able precisely to control exposure to chemical substances and other environmental conditions. Micro-fluidics technologies offer interesting opportunities to develop experimental systems that may meet the needs of modeling. Second, efforts need to be made to create a high precision model against well-controlled experimental systems, instead of relatively uncontrollable systems.

Third, the structural, spatial and temporal dynamics of both interaction networks and cellular structures need to be identified in order to define proper models. Whether cellular microstructures need to be modeled depends on the purpose and specific biological processes being modeled. The dynamics of cellular structure and interaction networks need to be measured by taking comprehensive, high-resolution quantitative measurements of the intracellular status, such as the concentrations, interactions, modifications and localizations of molecules, and of cellular structures at each coordinate in four dimensions under various conditions. In addition, the problem still remains as to how to identify unknown interactions from such data sets. These problems are very fundamental and require collaborative effort from the community.

Model Analysis

After models have been verified using experimental data, further analysis can provide in-depth insights on the intrinsic and dynamic nature of the system. Conventional time course simulation with a defined initial state provides us with an idea on how the system behaves dynamically under the specific condition. However, it only provides us with a snapshot of the dynamics the system may exhibit. In-depth insights on how systems may behave can be captured by systematic analysis of the system's reactions under different initial conditions and the propensity to change in reaction to the variations. It can also compute possible system behavior for conditions that cannot be set or are too expensive to be tested experimentally.

Different mathematical techniques have been developed to analyse the behavior of complex biological models [65,66], particularly focused on sensitivity analysis, phase space analysis and metabolic flux analysis.

Sensitivity analysis

The sensitivity of the system against various parameter changes is one of the properties that affects robustness and fragility of the system. It can reveal not only the stability of the system for

DOI: 10.1002/bdd

various perturbations, but also the controllability of the system.

Phase space analysis

As living systems operate under cellular homeostasis and homeodynamics, capturing the behavior of a complex biological model under equilibrium (steady-state or quasi steady state) conditions and the delineation of boundaries of different dynamic states within a set of parameter axes can provide insights into the dynamic properties of the system. Bifurcation analysis (analysis of a system of ODEs under parameter variation), phase-plane analysis (e.g. null-clines and local stability) helps to predict system behavior (equilibrium or oscillations) under a different parameter space. While model analysis is supported by many ODE solver systems such as Matlab®, specific tools are used widely in the community, e.g. Auto (Link 11), XPPAut (Link 12), BUNKI (Link 13) to name a few.

Metabolic control analysis (MCA)

This is a powerful quantitative framework for understanding the relationship between the properties of a metabolic network (steady state) characterized by its stoichiometric structure and the component reactions. It has been applied widely for analysis of cellular metabolism, particularly the control and regulation thereof. Developing on from the stoichiometric structure, a constraint-based modeling technique has been applied in metabolic engineering called Flux Balance Analysis (FBA) [67,68]. This does not require details of enzyme kinetics or metabolite concentrations, but aims to compute metabolic fluxes across a network that maximizes certain system properties (e.g. growth rates) under constraint conditions. Notably, FBA has been shown accurately to predict the growth rate of the bacteria *Escherichia coli* under different culture conditions [68].

While most of the model analysis techniques elucidated here focus on dynamic systems represented as a set of ODEs, it is pertinent to mention that alternative analyses have also been developed, based on statistical network analysis [63]. In particular, Boolean network modeling of genetic regulatory networks have gained wide acceptance in the modeling community based on the pioneering work by Kauffman [69]. Several Boolean network simulators for biological systems have been developed including NetBuilder [Link 14], BooleanNet [Link 15], SimBoolNet [70], to name a few.

One of the central difficulties for such a series of analysis is the nature of biological systems which are very high dimensional and non-linear. The choice of axis of phase space affects what dynamic properties can be captured, but such choices are made primarily by the intuition of researchers. Thus, using such analysis software does not ensure that proper insights can be obtained. In addition, the implication of analysis results requires proper translation from mathematical terms to biological terms. For example, when a system exhibits a saddle node bifurcation at a certain parameter region, it implies that the system can be unstable and its dynamics can qualitatively change depending on which paths are followed near the bifurcation point. While such a discovery is mathematically exciting, how can this be translated into experimental set-ups to verify that such findings are possible, but not trivial.

Integrated Software Platform

The process of data acquisition and management, model development by data-driven and deep curation approaches, and model verification and analysis requires a series of software and data resources to be interlinked and used in an integrated manner. Looking at systems biology and biomedical research in general, there are numbers of software tools and resources that can be used independently. However, their interoperability is missing. Back in the 1990s, a group of researchers recognized the need for community-wide efforts for data and model representation standards so that data and models can be used among different software tools. Efforts that resulted in the successful development and acceptance include the formation of data representation standards (MIAME [71], etc.), model representation standards (SBML [72], BioPAX [73], etc.), graphical representation standards (SBGN [74]) and model curation standards (MIRIAM [49]), etc. Workshops such as COMBINE (Link 2), a joint forum for standardization efforts, are now held regularly.

DOI: 10.1002/bdd

The establishment of a series of standards has dramatically improved the situation by enabling data and models to be reused by multiple software, promoted healthy competition among software tools, and helped to build a pipeline of tools for efficient analysis. However, the problem still remains of a lack of interoperability and inconsistent user interfaces. Software tools are being developed by independent research groups and companies without explicit agreement of how they can be operated smoothly when users have to use multiple tools in a single workflow. Thus, users often have to convert data format to adjust differences between tools, learn operating procedures for each tool, and sometimes even have to adjust operating environments. This impedes users from using these tools easily and does not help the wide utilization of tools either.

Recently, an alliance called 'The Garuda Alliance' (Link 1) was formed to rectify this situation. The aim of The Garuda Alliance is to create a platform and a set of guidelines that enable a high level of interoperability, consistent user experiences and a broader reachability of tools and resources. To achieve these objectives, the alliance will provide the Garuda Core that provides defined and comprehensive application program interfaces (APIs), a wide range of program and widget parts and a series of design guidelines (Figure 5).

The developers of tools can use well-defined APIs to make their tools operational on the Garuda Core and can easily attain a high level of interoperability. Garuda compliant software also needs to adopt user interface guidelines so that users can use a range of tools without extra learning efforts. Initially, software such as CellDesigner, Panther pathway database [75], bioCompendium (Link 3), PathText [76], Edinburgh SBSI tools (Link 6) and other software tools will be provided as Garuda compliant software, and the list will increase rapidly. Given the already widely used software and data resources, the numbers of Garuda users will be substantial from the beginning. This helps developers to reach out their tools to broader users through the Garuda download site, and users can obtain a wide selection of tools that are consistently designed. In this aspect, The Garuda Alliance (Link 1) can provide users with a one-stop-service type resource in systems biology and bioinformatics field.

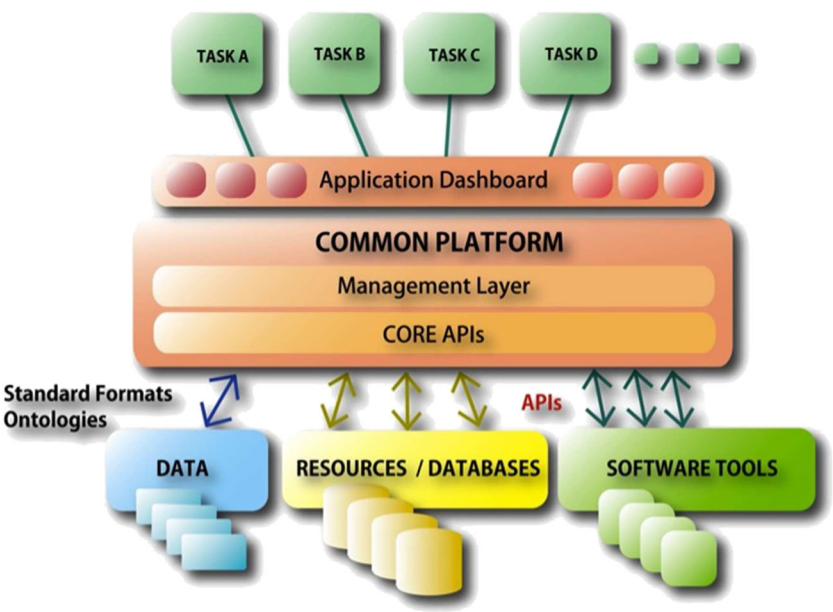

Figure 5. A conceptual diagram of the Garuda Platform. The integrated platform requires a well defined and widely accepted common software and application program interfaces (APIs). In the case of the Garuda platform, a set of APIs defined for the Garuda core that are derived from CellDesigner APIs will play a pivotal role. Numbers of Garuda application can talk to each other through this core software using Garuda APIs

DOI: 10.1002/bdd

There are other alliances that share a similar objective, but with a different emphasis such as Sage Bionetworks (Link 7) that focuses on data annotation and sharing. The Sage Bionetworks is currently focusing on establishing a platform for data acquisition, data curation, data adjustment, reformatting, and eventually modeling using open collaborative approach. While proper data handling can be achieved with software tools and well-defined protocols, achieving effective data sharing requires community-wide efforts. Most data are not fully utilized often because data published are not properly annotated, not stored in a proper repository, etc. A certain mechanism must be imposed for effective solution and the software platform design has to be an integral part of the solution to be practical and enforceable. Along a similar line of thought is the European effort for building biological data management infrastructure ELIXIR (Link 8) that is funded by EMBL and multiple European countries.

From Molecules to Physiology

One of the applications that can be effectively enabled by such a platform is an integrated modeling from molecular to physiology level because it has to smoothly integrate software for multiple levels of abstractions and various model development and analysis approaches. Application of computational models to clinically relevant studies often requires that the model be extended to embrace tissue, organ and whole body. While a cellular level model often plays a pivotal role, integration from the molecular level to the whole body level is critical because changes in the molecular and genetic level may affect the physiological outcome that is relevant to specific disease manifestation.

Research on how such integration can be accomplished is a major interest in systems biology and the physiology community. While there are numerous studies attempting to make such an integration, especially heart models, more comprehensive and community-driven projects have been launched. The IUPS (International Union of Physiological Sciences) Physiome Project has been trying to promote basic science and technological foundation for integrated physiological models for years. A couple of new initiatives started in 2010 are the Virtual Physiological Human (VPH) project in Europe and the High Definition Physiology (HD-Physiology) project in Japan.

For example, the HD-Physiology Project funded by the Japanese government is trying to develop a comprehensive platform for the virtual integration of models from the molecular to the whole body level, and focuses on developing a combined model of whole heart electrophysiology interconnected with cellular, pathway and molecular level modes and a whole body metabolism model. The core modeling tools are CellDesigner and PhysioDesigner. CellDesigner is a modeling software specialized for molecular and gene regulatory networks from sub-cellular to multi-cellular levels, and PhysioDesigner is a modeling software for the physiological level that is from the multi-cellular to the whole body level. Both software packages comply with standards such as SBML, SBGN and provide extensions to interface with other software platforms. They are connected to other software via the Garuda platform.

One of the typical use cases may be to simulate an effect of a certain drug on cardiac events and possibly to predict its difference due to genetic variation. In this case, first, an ADME/PK model has to be developed as a whole body metabolism to estimate drug dose changes in cardiac cells under the defined regimen. Then, pathway and cellular level models will be used to compute the effects of drugs on ion transport, signaling and other cellular behaviors that impacts on electrophysiology at the tissue level. Cellular level models are aggregated to compute tissue or organ level behavior that will provide organ level behaviors such as arrhythmia (Figure 6a). There are specific modeling tasks at each layer of modeling that have to be linked to provide a coherent simulation outcome.

The challenges are that such models have to deal with different time scales (Figure 6b), physical size, modeling principles and heterogeneity of the system. For example, modeling heartbeat alone from the molecular level requires the integration of the whole heart electrophysiology, heart muscle tissue structure, cardiomyocyte electrophysiology and contraction, intra-cellular signaling and ion channel behaviors, and molecular-level behaviors of key molecules, and their

DOI: 10.1002/bdd

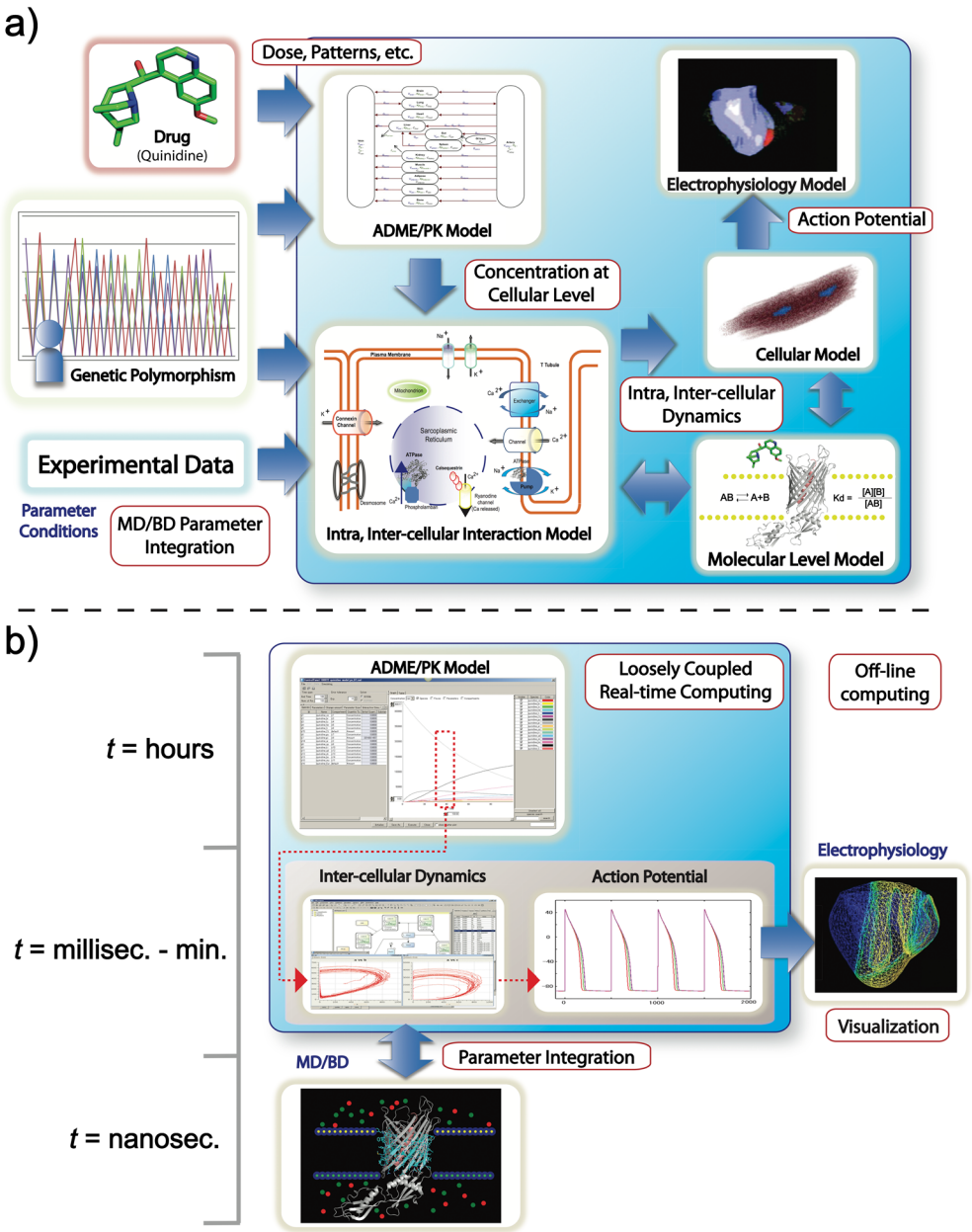

Figure 6. A possible use case and multiple time scale integration for the HD-Physiology Project. (a) A possible use case of an integrated model is to evaluate the effects of a drug for cardiac events. One can set a simulation condition of a specific drug dosage with a specific interval as a hypothetical regimen. The ADME/PK model can compute drug distribution and metabolism, so that change in a drug dose at cardiomyocytes can be simulated. Pathway and cellular level models assume the computed drug dosage level as an environmental factor for the simulation of ion channel, signaling and energy metabolism to compute membrane potential and cellular contraction. In some cases, genetic polymorphisms may change the behaviors of the cell. For novel protein structures for ion channel and other critical molecules, *in silico* simulation of molecular interactions may be used to better estimate the interaction parameters that are not experimentally known. Computed membrane potential will be used to reproduce organ level electrophysiology of arrhythmia. (b) Three different time scales have to be coupled for such simulation. The ADME/PK will be simulated on the scale from minutes to days. Cellular and pathway level simulations are mostly of the order from milli-seconds to hours. Molecular dynamics is computed in the order from nano- to micro-seconds. With such a difference in time-scale, loosely coupled simulation and pre-computed values will be used for the final integrated computation

DOI: 10.1002/bdd

differences due to genetic polymorphism. In addition, precise heart modeling at the organ and tissue level will include fibre directions, neuronal systems, coronary arteries, blood fluid dynamics and other biological sub-systems that are functioning in an integrated manner. The models of ADME/PK that are calculated based on a number of molecular properties as well as the atomic interaction between the given biologically relevant receptors and drug also need to be considered in order computationally to evaluate the pharmacophore features relating to the bioactivity and potential toxicity of a candidate drug. Many of the ADMET properties that are driven by the physiochemical properties of a candidate drug are calculated using the *in silico* method [77], such as QSAR (quantitative structure-activity relationship) modeling. Such ADMET properties correlate with the drug's bioavailability, distribution and clearance in the body, and can be applied as a parametric component to a specific cell model.

Inevitably, different numerical solution methods need to be used, but function coherently. For example, the fluid dynamics of blood in a heart can be described by PDEs such as the Navier–Stokes equation and Poisson equation, ECG derived from the cardiac electrical activity with the 3D torso model will be computed using PDEs as well, but most of intracellular signaling and whole body ADME model will be calculated by ODEs [52,53]. Close linkage of ODE and PDE is critical in such a model, but stochastic computation may have to be involved in case the stochastic behaviors of molecules play a critical role.

This is in sharp contrast to computational fluid dynamics (CFD) in which models are monolithic, monolayer and mono-physics. Biological models are inevitably multi-scale, hierarchical, heterogeneous and governed by different computational principles for each layer. Thus, proper modeling and computational architecture has to be designed to produce coherent and accurate results.

Beyond Software: Network of Intelligence

Creating and making the best use of software and data resources will certainly promote research for novel discoveries and efficient drug design. However, the impact of creating a widely accepted software platform may go far beyond productivity improvement in each research group, because it can potentially connect research groups globally. One of the desires of biomedical researchers is to find a way collaboratively to accomplish a major mission that cannot be accomplished by a single research group or a project alone. For example, creating and maintaining a comprehensive and in-depth model of biological systems at multiple spatio-temporal scales is often beyond the scope of any single research group. Even though one managed to develop one, it is not feasible to continue to maintain such models and to up-date new discoveries in the field.

Some alternative approaches have been proposed that includes Web2.0 services such as Wikipedia. There are several such attempts including WikiPathways [78], WikiGenes [79], etc. for large scale pathway maps. However, many of such efforts are at best struggling. One of the possible reasons is the sociological and incentive issue for contributing one's knowledge and data. The basic question is why should you spend time to share knowledge when such a contribution is not properly acknowledged.

One of a few successful community-based resource development efforts is a project on tuberculosis metabolic network reconstruction of the Open Source Drug Discovery project in India (http://www.osdd.net/). In this project, the Payao system was used to promote collective pathway development and annotation [80]. The key for success is the high motivation of participants because every year over 1 million outbreaks are reported in India and the expectation is that participation in such a project may provide better job opportunities. Furthermore, how such a success can be repeated in other domains where such burning motivations are not prevalent is a major issue and other motivations have to be considered.

Building such a 'network of intelligence', which supports a knowledge continuum, is key for the success of quantitative systems pharmacology (QSP) as a discipline and its application towards the development of personalized medicine. The marriage of systems biology approaches with network-based pharmacology analysis systems

DOI: 10.1002/bdd

across academia, industry and governmental regulatory bodies such as the US FDA, is critical towards translating the promises into reality. As outlined by Rogers *et al*. [92], '*..analogy comes to mind that we have two railroad companies attempting to span a continent, beginning at different ends and meeting in the middle. It is obviously necessary to arrive at the same place with the same gauge tracks*'.

How to frame and motivate open-ended collaboration is not well understood. Consistency of the software platform and the interoperability of various tools and devices is the first step, but is not enough to transform the field into the next stage. The successful formation of virtual big science may be the key for solving many of the significant biomedical problems such as an in-depth understanding of particular aspects of cellular function, drug discovery of neglected diseases, etc. [81]. The NIH study group on quantitative systems pharmacology (QSP) [87] recommended the establishment of inter-disciplinary research and training programs on numerous scales from individual teams to multi-investigator, multi-center programs. While establishing such a network of intelligence is going to be a key component of future research, a comprehensive, consistent and community-wide software platform is the prerequisite for such evolution in research paradigm, and we are now about to establish such a platform.

Conflict of Interest

The authors have declared that there is no conflict of interest.

Further Information

Online links
Link 1: http://www.garuda-alliance.org
Link 2: http://sbml.org/Events/Forums/COMBINE_2010
Link 3: http://biocompendium.embl.de/
Link 4: http://support.sdtools.com/gf/project/openfem
Link 5: http://www.freefem.org/
Link 6: http://csbe.bio.ed.ac.uk/sbsi.php
Link 7: http://www.sagebase.org/
Link 8: http://www.elixir-europe.org/page.php
Link 9: http://www.knime.org/
Link 10: http://www.the-dream-project.org/
Link 11: http://indy.cs.concordia.ca/auto/
Link 12: www.math.pitt.edu/~bard/xpp/xpp.html
Link 13: http://bunki.ait.tokushima-u.ac.jp:50080/
Link 14: http://homepages.stca.herts.ac.uk/~erdqmjs/NetBuilder%20home/NetBuilder/
Link 15: http://atlas.bx.psu.edu/booleannet/booleannet.html

References

1. Kitano H. Systems biology: a brief overview. *Science* 2002; **295**: 1662–1664.
2. Kitano H. Computational systems biology. *Nature* 2002; **420**: 206–210.
3. Kitano H. Perspectives on systems biology. *Future Generat Comput Syst* 2000; **18**: 199–216.
4. Ideker T, Galitski T, Hood L. A new approach to decoding life: systems biology. *Annu Rev Genomics Hum Genet* 2001; **2**: 343–372. doi:10.1146/annurev.genom.2.1.343.
5. Evans D, Hagiu A, Schmalensee R. Invisible Engines: How Software Platforms Drive Innovation and Transform Industries. MIT Press: USA, 2006. http://mitpress.mit.edu/books/invisible-engines
6. Lee TL. Big data: open-source format needed to aid wiki collaboration. *Nature* 2008; **455**: 461. doi:10.1038/455461c.
7. Galperin MY, Cochrane GR. The 2011 nucleic acids research database issue and the online molecular biology database collection. *Nucleic Acids Res* 2011; **39**: D1–D6. doi:10.1093/nar/gkq1243.
8. Mayer G. Data management in systems biology I – Overview and bibliography *CoRR* 2009; abs/0908.0411.
9. Zhao J, Miles A, Klyne G, Shotton D. Linked data and provenance in biological data webs. *Brief Bioinform* 2009; **10**: 139–152. doi:10.1093/bib/bbn044.
10. Van Deun K, Smilde AK, van der Werf MJ, Kiers HA, Van Mechelen I. A structured overview of simultaneous component based data integration. *BMC Bioinformatics* 2009; **10**: 246. doi:10.1186/1471-2105-10-246.
11. Brown F. Saving big pharma from drowning in the data pool. *Drug Discov Today* 2006; **11**: 1043–1045. doi:10.1016/j.drudis.2006.10.002.
12. Bry F, Kröger P. A computational biology database digest: data, data analysis, and data management. *Distrib Parallel Databases* 2003; **13**: 7–42.
13. Field D, Tiwari B, Snape J. Bioinformatics and data management support for environmental genomics. *PLoS Biol* 2005; **3**: e297. doi:10.1371/journal.pbio.0030297.

14. Keator DB. Management of information in distributed biomedical collaboratories. *Methods Mol Biol* 2009; **569**: 123. doi:10.1007/978-1-59745-524-4_1.
15. Brazma A, Krestyaninova M, Sarkans U. Standards for systems biology. *Nat Rev Genet* 2006; **7**: 593–605. doi:10.1038/nrg1922.
16. Oinn T, Addis M, Ferris J, *et al*. Taverna: a tool for the composition and enactment of bioinformatics workflows. *Bioinformatics* 2004; **20**: 3045–3054. doi:10.1093/bioinformatics/bth361.
17. Saltz J, Oster S, Hastings S, *et al*. caGrid: design and implementation of the core architecture of the cancer biomedical informatics grid. *Bioinformatics* 2006; **22**: 1910–1916. doi:10.1093/bioinformatics/btl272.
18. Lee S, Wang TD, Hashmi N, Cummings MP. BioSTEER: a semantic web workflow tool for grid computing in the life sciences. *Future Generations Computer Systems* 2007; **23**(3): 497–509.
19. Gehlenborg N, O'Donoghue SI, Baliga NS, *et al*. Visualization of omics data for systems biology. *Nat Methods* 2010; **7**: S56–S68. doi:10.1038/nmeth.1436.
20. Krzywinski M, Schein J, Birol I, *et al*. Circos: an information aesthetic for comparative genomics. *Genome Res* 2009; **19**: 1639–1645. doi:10.1101/gr.092759.109.
21. Kohl M, Wiese S, Warscheid B. Cytoscape: software for visualization and analysis of biological networks. *Methods Mol Biol* 2011; **696**: 291–303. doi:10.1007/978-1-60761-987-1_18.
22. Schadt EE, Friend SH, Shaywitz DA. A network view of disease and compound screening. *Nat Rev Drug Discov* 2009; **8**: 286–295. doi:10.1038/nrd2826.
23. van't Veer LJ, Dal H, van der Vijver MJ, *et al*. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002; **415**: 530–536. doi:10.1038/415530a.
24. Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. *Science* 2008; **322**: 881–888. doi:10.1126/science.1156409.
25. Dewan A, Liu M, Hartman S, *et al*. HTRA1 promoter polymorphism in wet age-related macular degeneration. *Science* 2006; **314**: 989–992. doi:10.1126/science.1133807.
26. Yang Z, Camp NJ, Sun H, *et al*. A variant of the HTRA1 gene increases susceptibility to agerelated macular degeneration. *Science* 2006; **314**: 992–993. doi:10.1126/science.1133811.
27. Chesler EJ, Lu L, Shou S, *et al*. Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nat Genet* 2005; **37**: 233–242. doi:10.1038/ng1518.
28. Monks SA, Leonardson A, Zhu H, *et al*. Genetic inheritance of gene expression in human cell lines. *Am J Hum Genet* 2004; **75**: 1094–1105. doi:10.1086/426461.
29. Morley M, Molony CM, Weber TM, *et al*. Genetic analysis of genome-wide variation in human gene expression. *Nature* 2004; **430**: 743–747. doi:10.1038/nature02797.
30. Zhu J, Lum PY, Lamb J, *et al*. An integrative genomics approach to the reconstruction of gene networks in segregating populations. *Cytogenetic Genome Res* 2004; **105**: 363–374. doi:10.1159/000078209.
31. Zhu J, Wiener MC, Zhang C, *et al*. Increasing the power to detect causal associations by combining genotypic and expression data in segregating populations. *PLoS Comput Biol* 2007; **3**: e69. doi:10.1371/journal.pcbi.0030069.
32. Zhu J, Zhang B, Smith EN, *et al*. Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nat Genet* 2008; **40**: 854–861. doi:10.1038/ng.167.
33. Margolin AA, Nemenman I, Basso K, *et al*. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 2006; **7**(Suppl 1): S7. doi:10.1186/1471-2105-7-S1-S7.
34. Faith JJ, Hayete B, Thaden JT, *et al*. Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol* 2007: **5**: e8. doi:10.1371/journal.pbio.0050008.
35. Shen-Orr SS, Milo R, Mangan S, Alon U. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat Genet* 2002; **31**: 64–68. doi:10.1038/ng881.
36. Alon U. Network motifs: theory and experimental approaches. *Nature Rev Genet* 2007; **8**: 450–461. doi:10.1038/nrg2102.
37. Fadda A, Fierro AC, Lemmens K, Monsieurs P, Engelen K, Marchal K. Inferring the transcriptional network of Bacillus subtilis. *Mol BioSystems* 2009; **5**: 1840–1852. doi:10.1039/b907310h.
38. Cho BK, Zengler K, Qiu Y, *et al*. The transcription unit architecture of the *Escherichia coli* genome. *Nat Biotechnol* 2009; **27**: 1043–1049. doi:10.1038/nbt.1582.
39. Mendoza-Vargas A, Olvera L, Olvera M, *et al*. Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in *E. coli*. *PLoS One* 2009; **4**: e7526. doi:10.1371/journal.pone.0007526.
40. Lemmens K, De Bie T, Dhollander T, *et al*. DISTILLER: a data integration framework to reveal condition dependency of complex regulons in *Escherichia coli*. *Genome Biol* 2009; **10**: R27. doi:10.1186/gb-2009-10-3-r27.
41. Ferrucci D, Brown E, Chu-Carroll J, *et al*. Building Watson: an overview of the DeepQA project. *AI Magazine* 2010; **31**: 59–79.
42. Oda K, Kitano H. A comprehensive map of the toll-like receptor signaling network. *Mol Syst Biol* 2006; **2**: 2006.0015. doi:msb4100057 [pii] 10.1038/msb4100057 [doi].
43. Oda K, Matsuoka Y, Funahashi A, Kitano H. A comprehensive pathway map of epidermal growth factor receptor signaling. *Mol Syst Biol* 2005; **1**: 2005 0010. doi:msb4100014 [pii] 10.1038/msb4100014 [doi].

44. Caron E, Ghosh S, Matsuoka Y, *et al*. A comprehensive map of the mTOR signaling network. *Mol Syst Biol* 2010; **6**: 453. doi:msb2010108 [pii] 10.1038/msb.2010.108.
45. Kaizu K, Ghosh S, Matsuoka Y, *et al*. A comprehensive molecular interaction map of the budding yeast cell cycle. *Mol Syst Biol* 2010; **6**: 415, doi: msb201073 [pii] 10.1038/msb.2010.73.
46. Bauer-Mehren A, Furlong LI, Sanz F. Pathway databases and tools for their exploitation: benefits, current limitations and challenges. *Mol Syst Biol* 2009; **5**: 290. doi:msb200947 [pii] 10.1038/msb.2009.47.
47. Hucka M, Finney A, Bornskin BJ, *et al*. Evolving a lingua franca and associated software infrastructure for computational systems biology: the Systems Biology Markup Language (SBML) project. *Systems Biol (Stevenage)* 2004; **1**: 41–53.
48. Le Novere N, Hucke M, Mi H, *et al*. The systems biology graphical notation. *Nat Biotechnol* 2009; **27**: 735–741. doi:10.1038/nbt.1558.
49. Le Novere N, Finney A, Hucka M, *et al*. Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat Biotechnol* 2005; **23**: 1509–1515. doi:10.1038/nbt1156.
50. Kitano H, Funahashi A, Matsuoka Y, Oda K. Using process diagrams for the graphical representation of biological networks. *Nat Biotechnol* 2005; **23**: 961–966. doi:10.1038/nbt1111.
51. Pico AR, Kelder T, van Iersel MP, Hanspers K, Conklin BR, Evelo C. WikiPathways: pathway editing for the people. *PLoS Biol* 2008; **6**: e184. doi:10.1371/journal.pbio.0060184.
52. Mendes P, Hoops S, Sahle S, Gauges R, Dada J, Kummer U. Computational modeling of biochemical networks using COPASI. *Methods Mol Biol* 2009; **500**: 17–59. doi:10.1007/978-1-59745-525-1_2.
53. Machne R, Finney A, Muller S, Lu J, Widder S, Flamm C. The SBML ODE Solver Library: a native API for symbolic and fast numerical analysis of reaction networks. *Bioinformatics* 2006; **22**: 1406–1407. doi:10.1093/bioinformatics/btl086.
54. Lopez-Aviles S, Kapuy O, Novak B, Uhlmann F. Irreversibility of mitotic exit is the consequence of systems-level feedback. *Nature* 2009; **459**: 592–595. doi:10.1038/nature07984.
55. Schoeberl B, Pace EA, Fitzgerald JB, *et al*. Therapeutically targeting ErbB3: a key node in ligand-induced activation of the ErbB receptor-PI3K axis. *Sci Signal* 2009; **2**: ra31. doi:10.1126/scisignal.2000352.
56. McAdams HH, Arkin A. Stochastic mechanisms in gene expression. *Proc Natl Acad Sci U S A* 1997; **94**: 814–819.
57. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A. Regulation of noise in the expression of a single gene. *Nat Genet* 2002; **31**: 69–73. doi:10.1038/ng869.
58. Novak B, Tyson JJ. Numerical analysis of a comprehensive model of M-phase control in Xenopus oocyte extracts and intact embryos. *J Cell Sci* 1993; **106**(Pt 4): 1153–1168.
59. Tyson JJ, Chen K, Novak B. Network dynamics and cell physiology. *Nat Rev Mol Cell Biol* 2001; **2**: 908–916.
60. Schoeberl B, Faber AC, Li D, *et al*. An ErbB3 antibody, MM-121, is active in cancers with ligand-dependent activation. *Cancer Res* 2010; **70**: 2485–2494. doi:10.1158/0008-5472.CAN-09-3145.
61. Arkin A, Ross J, McAdams HH. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics* 1998; **149**: 1633–1648.
62. Emonet T, Macal CM, North MJ, Wickersham CE, Cluzel P. AgentCell: a digital single-cell assay for bacterial chemotaxis. *Bioinformatics* 2005; **21**: 2714–2721. doi:10.1093/bioinformatics/bti391.
63. Hofestadt R, Thelen S. Quantitative modeling of biochemical networks. *Stud Health Technol Inform* 2011; **162**: 3–16.
64. Waltemath D, Adams R, Beard DA, *et al*. Minimum Information About a Simulation Experiment (MIASE). *PLoS Comput Biol* 2011; **7**: e1001122, doi:10.1371/journal.pcbi.1001122.
65. Klipp ERH, Kowald A, Wierling C, Lehrach H. Systems Biology in Practice: Concepts, Implementation and Application. John Wiley & Sons: Germany, 2005.
66. Haefner JW. Modeling Biological Systems: Principles and Applications (1st edn). Kluwer Academic: UK, 1996.
67. Edwards JS, Palsson BO. How will bioinformatics influence metabolic engineering? *Biotechnol Bioengineer* 1998; **58**: 162–169.
68. Edwards JS, Ibarra RU, Palsson BO. *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat Biotechnol* 2001; **19**: 125–130. doi:10.1038/84379.
69. Kauffman SA. Metabolic stability and epigenesis in randomly constructed genetic nets. *J Theor Biol* 1969; **22**: 437–467.
70. Zheng J, Zhang D, Przkycki PF, Zielinski R, Capala J, Prztycka TM. SimBoolNet – a Cytoscape plugin for dynamic simulation of signaling networks. *Bioinformatics* 2010; **26**: 141–142. doi:10.1093/bioinformatics/btp617.
71. Brazma A, Hingamp P, Quackenbush J, *et al*. Minimum information about a microarray experiment (MIAME)–toward standards for microarray data. *Nat Genet* 2001; **29**: 365–371. doi:10.1038/ng1201–365.
72. Hucka M, Finney A, Sauro HM, *et al*. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 2003; **19**: 524–531.
73. Demir E, Cary MP, Paley S, *et al*. The BioPAX community standard for pathway data sharing. *Nat Biotechnol* 2010; **28**: 935–942. doi:10.1038/nbt.1666.

74. Le Novere N, Hucka M, Mi H, *et al*. The systems biology graphical notation. *Nat Biotechnol* 2009; **27**: 735–741. doi:nbt.1558 [pii] 10.1038/nbt.1558 [doi].
75. Mi H, Thomas P. PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. *Methods Mol Biol* 2009; **563**: 123–140. doi:10.1007/978-1-60761-175-2_7.
76. Kemper B, Matsuzaki T, Matsuoka Y, *et al*. PathText: a text mining integrator for biological pathway visualizations. *Bioinformatics* 2010; **26**: i374–i381. doi:10.1093/bioinformatics/btq221.
77. Smith DA. Metabolism, Pharmacokinetics and Toxicity of Functional Groups Metabolism, Pharmacokinetics and Toxicity of Functional Groups. Royal Society of Chemistry Publishing: London, 2010; 61–94.
78. Pico AR, Kelder T, van Iersel MP, Hanspers K, Conklin BR, Evelo C. WikiPathways: pathway editing for the people. *PLoS Biol* 2008; **6**: e184. doi:08-PLBI-CP-0724 [pii] 10.1371/journal.pbio.0060184.
79. Maier H, Dohr S, Grote K, *et al*. LitMiner and WikiGene: identifying problem-related key players of gene regulation using publication abstracts. *Nucleic Acids Res* 2005; **33**: W779–782. doi:10.1093/nar/gki417.
80. Matsuoka Y, Ghosh S, Kikuchi N, Kitano H. Payao: a community platform for SBML pathway model curation. *Bioinformatics* 2010; **26**: 1381–1383. doi: btq143 [pii] 10.1093/bioinformatics/btq143.
81. Kitano H, Ghosh S, Matsuoka Y. Social engineering for virtual 'big science' in systems biology. *Nat Chem Biol* 2011; **7**: 323–326. doi:10.1038/nchembio.574.
82. Harrold MJ, Ramanathan M, Mager ED. Network-based approaches in drug discovery and early development. *Clin Pharmacol Ther* 2013. doi:10.1038/clpt.2013.176.
83. Wist DA, Berger IS, Iyengar R. Systems pharmacology and genome medicine: a future perspective. *Genome Med* 2009; **1**: 11. doi:10.1186/gm11.
84. Berger SI, Iyengar R. Role of systems pharmacology in understanding drug adverse events. *Wiley Interdiscip Rev Syst Biol Med* 2011; **3**(2): 129–135. doi:10.1002/wsbm.114.
85. Zhao S, Iyengar R. Systems pharmacology: network analysis to identify multiscale mechanisms of drug action. *Annu Rev Pharmacol Toxicol* 2012; **52**: 505–521. doi:10.1146/annurev-pharmtox-010611-134520.
86. Jones H, Rowland-Yeo K. Basic concepts in physiologically based pharmacokinetic modeling in drug discovery and development. *CPT Pharmacometrics Syst Pharmacol* 2013; **2**: e63. doi:10.1038/psp.2013.41.
87. Sorger PK, Allerheiligen RBS, *et al*. Quantitative and Systems Pharmacology in the Post-genomic Era: New Approaches to Discovering Drugs and Understanding Therapeutic Mechanisms. *An NIH White Paper by the QSP Workshop Group*. 2011.
88. Csermely P, Korcsmáros T, Kiss HJ, London G, Nussinov R. Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol Ther* 2013; **138**(3): 333–408. doi:10.1016/j.pharmthera.2013.01.016.
89. Models that take drugs. The Economist report, 11 June 2005.
90. Michelson S. Assessing the Impact of Predictive Biosimulation on Drug Discovery and Development. *Journal of Bioinformatics and Computational Chemistry* 2003; **1**(1): 169–177.
91. Merrimack Pharmaceuticals Initiates Enrollment In A Phase 1 Study Of MM-121, An ErbB3 Antagonist. *Medical News Today*, 12 August 2008.
92. Rogers M, Lyster P, Okita R. NIH Support for the emergence of quantitative and systems pharmacology. *CPT Pharmacometrics Syst Pharmacol* 2013; **2**: e37. doi:10.1038/psp.2013.13.