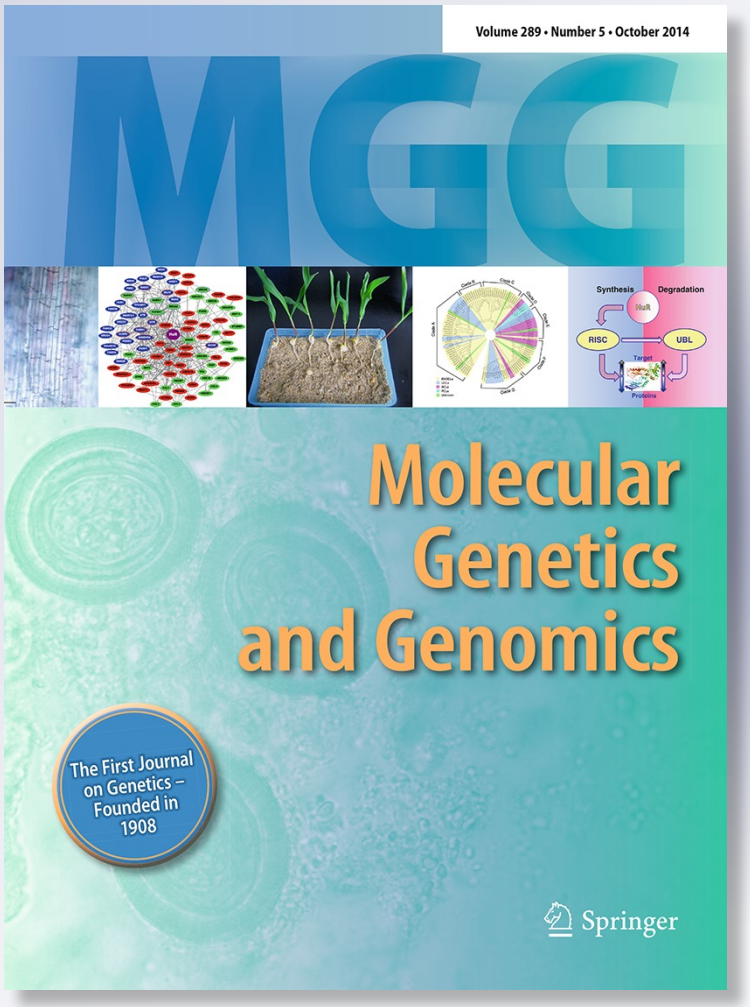
Bridging the gaps in systems biology

Marija Cvijovic, Joachim Almquist, Jonas Hagmar, Stefan Hohmann, Hans-Michael Kaltenbach, Edda Klipp, Marcus Krantz, Pedro Mendes, et al.

Molecular Genetics and Genomics

ISSN 1617-4615
Volume 289
Number 5

Mol Genet Genomics (2014)
289:727-734
DOI 10.1007/s00438-014-0843-3



Springer

Mol Genet Genomics (2014) 289:727–734
DOI 10.1007/s00438-014-0843-3

REVIEW

Bridging the gaps in systems biology

Marija Cvijovic · Joachim Almquist · Jonas Hagmar · Stefan Hohmann · Hans-Michael Kaltenbach · Edda Klipp · Marcus Krantz · Pedro Mendes · Sven Nelander · Jens Nielsen · Andrea Pagnani · Natasa Przulj · Andreas Raue · Jörg Stelling · Szymon Stoma · Frank Tobin · Judith A. H. Wodke · Riccardo Zecchina · Mats Jirstrand

Received: 31 January 2014 / Accepted: 21 March 2014 / Published online: 13 April 2014

Abstract Systems biology aims at creating mathematical models, i.e., computational reconstructions of biological systems and processes that will result in a new level of understanding—the elucidation of the basic and presumably conserved "design" and "engineering" principles of biomolecular systems. Thus, systems biology will move biology from a phenomenological to a predictive science. Mathematical modeling of biological networks and processes has already greatly improved our understanding of many cellular processes. However, given the massive amount of qualitative and quantitative data currently produced and number of burning questions in health care and biotechnology needed to be solved is still in its early phases. The field requires novel approaches for abstraction, for modeling bioprocesses that follow different biochemical and biophysical rules, and for combining different modules into larger models that still allow realistic simulation with the computational power available today. We have identified and discussed currently most prominent problems in systems biology: (1) how to bridge different scales of modeling abstraction, (2) how to bridge the gap between topological and mechanistic modeling, and (3) how to bridge the wet and dry laboratory gap. The future success of systems biology largely depends on bridging the recognized gaps.

Communicated by J. Graw.

M. Cvijovic (✉)
Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, Chalmers Tvärgata 3, 412 96 Göteborg, Sweden
e-mail: marija.cvijovic@chalmers.se

J. Almquist · J. Hagmar · M. Jirstrand
Fraunhofer-Chalmers Centre, Göteborg, Sweden
e-mail: joachim.almquist@fcc.chalmers.se

J. Hagmar
e-mail: jonas.hagmar@gmail.com

M. Jirstrand
e-mail: mats.jirstrand@fcc.chalmers.se

S. Hohmann
University of Gothenburg, Göteborg, Sweden
e-mail: stefan.hohmann@gu.se

H.-M. Kaltenbach · J. Stelling
ETH, Zürich, Switzerland
e-mail: hans-michael.kaltenbach@bsse.ethz.ch

J. Stelling
e-mail: joerg.stelling@bsse.ethz.ch

E. Klipp · M. Krantz · J. A. H. Wodke
Humboldt University, Berlin, Germany
e-mail: edda.klipp@biologie.hu-berlin.de

M. Krantz
e-mail: marcus.krantz@biologie.hu-berlin.de

J. A. H. Wodke
e-mail: jwodke@googlemail.com

P. Mendes
Manchester University, Manchester, UK
e-mail: pedro.mendes@manchester.ac.uk

S. Nelander
Uppsala Univeristy, Uppsala, Sweden
e-mail: sven.nelander@igp.uu.se

J. Nielsen
Chalmers University of Technology, Göteborg, Sweden
e-mail: nielsenj@chalmers.se

A. Pagnani · R. Zecchina
Politecnico di Torino, Turin, Italy
e-mail: andrea.pagnani@gmail.com

Keywords Systems biology · Modeling · Sensitivity analysis · Model merging · Model standards

Introduction

By use of mathematical models, systems biology aims at a new level of understanding—the understanding of how higher-level functions emerge from the combined action of many biomolecules. Thus, systems biology moves biology from a phenomenological to a predictive science. Mathematical modeling of biological networks and processes has already greatly improved our understanding of many cellular processes (Klipp et al. 2005; Barberis et al. 2007; Nelander et al. 2008; Jörnsten et al. 2011; Novak et al. 2007; Erjavec et al. 2008; Agren et al. 2012; Krantz et al. 2009; Kotte et al. 2010; Jol et al. 2012; Smallbone et al. 2010; Arabidopsis Interactome Mapping Consortium 2011). However, there are still immense challenges in bridging experimental work and mathematical modeling. Advancing the field requires novel approaches for abstraction, for modeling bioprocesses that follow different biochemical and biophysical rules, and for combining a set of smaller models into larger models that still allow simulation with the computational power available today. To address current challenges, we present and discuss three prominent problems in systems biology: (1) how to bridge the gap between topological and mechanistic modeling, (2) how to bridge different scales of modeling abstraction, and (3) how to bridge the gap between the wet and dry laboratory.

Bridging the gap between topological and mechanistic modeling

Understanding how complex functions emerge from the interaction of biomolecules would allow us to predict the effects of drugs or to rationally design microbes for biotechnological applications and industrial exploitation. Thus, models that have predictive power are ultimately needed. Those models should capture and describe the important cellular processes at a sufficient mechanistic level providing new molecular insight.

Unfortunately, the current state of mechanistic modeling allows models only to capture small biological subsystems, i.e., most often they only grasp a limited number of molecular interactions or they can only describe known biological behavior, but they have limited predictive power. In fact, we are still far away from having predictive models for real applications, such as for medicine or biotechnology. Developing such models represents one of the biggest challenges for the field of systems biology.

Our current modeling capabilities are limited due to the following reasons: (1) model development requires a certain critical level of prior knowledge, e.g., about the biological components and their interactions—knowledge, which we typically do not have for larger and more complex systems; (2) a general lack of methods that can guide us in finding the correct level of abstraction given the particular purpose of the modeling task; and (3) some problems are computationally intractable (NP-hard) and hence the solutions for large-scale data can only be obtained approximately (i.e., heuristically).

Along with the limitations on the modeling side, particularly in more complex systems, there is a problem of an inverse relation between the system size and the available biological insight: while there is mostly a good understanding about the biological components, very often there are large gaps in network topology.

Thus, to deliver what is ultimately expected from systems biology—predictive models of biological systems—we need novel modeling and computational approaches that can develop such models despite provable computational intractability and existing uncertainties about the molecular details of a given biological system. These models do not necessarily need to be correct in all mechanistic detail, but they should have predictive power, at least for certain aspects of the modeled system.

Building predictive models—ways ahead

Incorporating uncertainty into mathematical models

When building mathematical models of biochemical networks, at least three interlinked levels of uncertainty can be encountered (Schaber et al. 2009): (1) uncertainty in the network structure, e.g., which components and reactions to include and at which level of detail (2) uncertainty with respect to the choice of kinetics of reactions, e.g., mass

R. Zecchina
e-mail: riccardo.zecchina@polito.it

N. Przulj
Imperial College London, London, UK
e-mail: natasha@imperial.ac.uk

A. Raue
University of Freiburg, Freiburg im Breisgau, Germany
e-mail: andreas.raue@fdm.uni-freiburg.de

S. Stoma
INRIA, Paris, France
e-mail: szymon.stoma@gmail.com

F. Tobin
Tobin Consulting LLC, New Jersey, USA
e-mail: frank@tobins.org

action, Michaelis–Menten, hill, and (3) uncertainty of the parameter values. Obviously, kinetic laws depend to some extent on the chosen granularity (e.g., two lumped mass-action reactions may conveniently be described by one Michaelis–Menten-type reaction) and the nature of the parameters is determined by the choice of the rate law.

The vast majority of today's mechanistic models does not incorporate nor try to quantify the uncertainty inherent in the studied systems. Simulation approaches based on Monte Carlo methods and sensitivity analysis of ordinary differential equations (ODEs) can be used to understand implications on system behavior of uncertainty in parameters, initial conditions, and to some extent also network structure (Wang et al. 2004; Mišković and Hatzimanikatis 2011). However, without combining proper descriptions of uncertainty with experimental data, proper differentiation between what we have learned about the system and what remains uncertain in an integrated modeling and experimental study will not be possible.

A formal way of specifying uncertainty is through the use of probability distributions. The most common use of concepts from statistics and probability theory in systems biology is to describe variability resulting from the stochastic nature of single reactions, cell-to-cell variation, noise, and variability introduced in the process of taking single or repeated measurements. However, probability distributions can also be used to characterize uncertainty in a broader sense, in particular in combination with measurement data. Assigning probability distributions to entities such as parameters, state variables, and system and measurement noise variables in dynamic models of biological networks provides understanding of both the quantity itself and its precision or uncertainty. This approach becomes even more useful if one not only uses prior probability distributions and studies their time evolution but also updates those using time series experimental data to obtain so-called *posterior* distributions. This more elaborate way of representing the system under investigation naturally comes with the cost of more complex computations and requires more advanced mathematics. Nevertheless, the gain of obtaining not only single numerical values for quantities but also some measure of quality or quantified uncertainty is a tremendous advantage. To fully exploit the described methodology, there is a need to augment systems of ODEs, in favor of systems of stochastic differential equations (SDEs) (Kloeden and Platen 1992). These are more complex mathematical objects than ODEs since they represent not only a single solution trajectory, but also a family of solutions, so-called realizations, whose statistical properties can be captured in terms of time-varying distributions for the state variables. From a modeling perspective, the mechanistic part of the SDEs is similar to the ODE description with the exception of additional stochastic variables representing uncertainty in parameters, reaction rates, or network structure. The computational effort to fully solve the SDEs for a realistic modeling problem, i.e., to compute the time-varying prior/posterior probability distributions for the state variables, is not to be underestimated and is in most cases not even tractable (requires the solution of a high-dimensional partial differential equation known as the Kolmogorov forward equation). However, often quantities such as the peaks or mean of a distribution, together with a measure of its spread is what is required to draw conclusions about the accuracy with which the model characterizes a system under study. Hence, much of today's research in this area targets methods for efficient and accurate approximate computations of the distributions or their derived features. This approach is commonly used in the field of nonlinear filtering (Jazwinski 1970) with longstanding successful applications in various engineering fields and tools such as extended and unscented Kalman filters and particle filters (Bar-Shalom et al. 2001; Ristic et al. 2004).

Bringing constraints into models

Kinetic parameters in biological network models are thermodynamically dependent. K_m-values and maximal rates of reversible reactions are linked by the equilibrium constants of those reactions. Along a series of reactions, the drop or gain of free energy provides a restriction for possible choices of kinetic parameters. If we would estimate the parameters independently from each other, we run into the danger of violating thermodynamic constraints. This flaw can be prevented by, e.g., parameter balancing (Lubitz et al. 2010) or tackled by network embedded thermodynamic analysis (Kümmel et al. 2006).

The accumulating information on parameter values is collected in databases such as Brenda (Chang et al. 2008) and SABIO-RK (Wittig et al. 2006) and can be used to obtain typical distributions of parameter values specific for reactions, organisms, or experimental conditions. Such distributions can be considered as prior probability densities (short: priors) for estimating parameter values in a Bayesian approach. New experimental data can be quantified with a likelihood function and a combination of both types of information can be used to calculate a posterior density distribution (Liebermeister and Klipp 2006). This approach takes into account the uncertainty of parameter values as discussed above and at the same time makes use of information obtained in previous and unrelated experimental work.

Model reduction/model expansion

An important aspect of modeling is the adaptation of the model complexity to the available knowledge and data

used for estimation and validation. Drafting a model from mechanistic hypotheses often leads to models with a quite large number of parameters to be determined. The situation becomes even more problematic when certain gaps in knowledge of how the system is wired are to be filled with partly competing or overlapping mechanistic hypotheses. To reduce the complexity of a model to better describe the available data, various methods of model reduction can be applied. Here, knowledge about orders of magnitude of different parameters as well as qualitative behavior of the system under study is important. A proper sensitivity analysis is also an important tool to rank the importance of certain parts of the model on the overall system behavior to gain information of how model can be simplified without killing the model's descriptive power. Building dynamic models by aggregating subsystems developed separately is a natural method for obtaining comprehensive models (Kuepfer et al. 2007). However, if such models are to be used for parameter estimation and model validation using time series data, then it is important to understand if the system is persistently excited by the perturbations or experimental situation at hand, i.e., if the generated data contain enough information to infer values of the parameters to be determined. A simple example is to consider a reaction rate given by a Michaelis–Menten expression. If the experimental conditions are such that this particular rate only operates in its linear range (and never becomes saturated), there will be no information in the data allowing us to distinguish between V_{max} and K_m and we will only be able to determine their ratio. To address comprehensive dynamic models, the frequent use of toolbox for model reduction techniques including linearization, separation of time-scales leading to applications of singular perturbation theory, quasi-steady-state assumptions, is required.

Bridging different scales of modeling abstraction

Mathematical models are often developed for a specific purpose and may be targeted at a small part of the overall cellular biochemical network. In the future, these models will become more useful if they can be integrated into larger, more unified models. For example, the Unicellsys project (Alberghina and Cirulli 2010) aims at an integrated model of stress, nutrient signaling on cell cycle, and growth of a single yeast cell and its impact on population. The Virtual Liver project (Abbott 2010) endeavors the integration of processes from signaling to whole organ level. There is also a major effort underway to develop accurate myocyte level ion channel models and link them with accurate 3D models of the heart to be able to model cardiac electrophysiological effects (Bassingthwaighte et al. 2009).

The complexity of biological systems is so extensive that separate groups of scientists are required to curate separate parts and pathways and to build the corresponding models. It is useful to model the details in separate models of appropriate granularities and sizes and later merge them; this process is likely to be carried out repeatedly as new knowledge is incorporated in the separate parts. By necessity, the merging of models requires combining the data on which they are based, as well as recalibration and validation efforts for the merged, resulting models.

Model merging

The development of models, even if entirely new, benefits from the parallel exploration of alternative hypotheses. For example, a signaling pathway may not be well known and it may be useful to explore models of two alternative subnetworks (Flöttmann et al. 2008). At some point, these alternative versions of a model will need to be resolved and combined—which requires model integration. Since, model development is, in general, a time-consuming process, one approach to building new models is to start from existing models. The reuse of existing models is facilitated by model databases such as JWS online (Olivier and Snoep 2004), BioModels (Le Novère et al. 2006), the CellML model repository (Lloyd et al. 2008), BioMet Toolbox (Cvijovic et al. 2010), and other such libraries. The ability to develop models in this way will improve collaborative processes and increase model robustness.

Generally, one could combine models with the same level of detail— the horizontal merging problem or combine models with different levels of details—the vertical merging problem. Often both are needed. In horizontal merging, one of the aims is to create models that have larger scope than the ingoing models, while keeping the same level of detail. For example, combining models of two signaling pathways provides a new model with more coverage of the network. In vertical merging, another aim could be to represent different levels of organization that is related to the same phenomenon. Horizontal combination of two models at the same level of detail allows covering a number of processes that would otherwise be ignored when the two parts are modeled independently. Such separation means that the interaction between the components of the two subsets is lost, but it is needed explicitly when they are combined. Without that interaction, new biological phenomena may not be able to be modeled.

Ensuring the consistency of the combined model with respect to the original models is a major challenge (Liebermeister 2008). One issue is the matching of biologically identical compartments; for species and reactions shared between the two models, this problem can be solved by the use of proper annotation (Krause et al. 2010). Even though overlapping species and reactions in the models to be merged can be matched, there may be still conflicts and

ambiguity. For instance, species may have different initial concentrations or environmental conditions. There can also be contradictions in parameter values or the structural form of rate expressions for some of the overlapping reactions. All such conflicts must be resolved either manually or with automated assistance, if available. For example, semanticSBML (Krause et al. 2010) and its predecessor SBMLmerge (Schulz et al. 2006) are two tools that can assist in the merging of models encoded in Systems Biology Markup Language—SBML (Hucka et al. 2003).

Once the definitional aspect of the component models is handled, it is necessary to evaluate the properties of the output model. The main purpose of the merged model should be to describe certain data or behavior that the original, separated models could not handle. But there may be further criteria resulting from the merging and this may require a recalibration of model parameters, preferably using the original data. This is critical, because the original data sets when brought together into the newly merged model may need to be interpreted differently due to the interaction between the component models. Moreover, each individual model is designed for a particular purpose using particular data, a specific level of abstraction, and different parameter optimization strategies. Therefore, both models and data must be available and accessible. Data stored in relevant databases, e.g., *ArrayExpress* (Parkinson et al. 2010), *caBig* (http://cabig.nci.nih.gov/), *ImmunoblotDB* (http://www.immunoblot.de/) can be standardized, which is of special importance when the data have been obtained using different biological protocols, e.g., micro-array, 2D-gels, mass spectrometry. A model should be annotated with the unique access number of the data used for its calibration. Clearly, with the merged model serving a new purpose, the original data might be reinterpreted, thereby giving new biological insight. It is not only just standardization of data that is critical, but also its availability. If the merged model is to be recalibrated and revalidated, the legacy data, therefore, needs to be available (which currently is often not the case).

Yet another aspect of model merging is the combination of models that are based on different mathematical frameworks. One might combine (1) ODEs with stochastic kinetics (often represented as master equations, but sometimes as SDEs) (Resat et al. 2009), (2) discrete (e.g., graph theoretic methods, Boolean networks) and continuous systems (e.g., ODEs), or (3) ODEs with partial differential equations (PDEs), which is formally fairly straightforward, but may well have serious multi-scale numerical problems as a result (Flöttmann et al. 2008).

One common aspect of merged models with different mathematical frameworks is that they require very different numerical approaches toward integration of equations. Such hybrids may need to not only resolve issues with scale (temporal or spatial), but the numerical robustness of the resulting combined integration may not always be as good as it was in the component problems. A potential multi-scale integration problem (Southerna et al. 2008) occurs when different time or length scales are combined in the same problem, and the numerical solution is often very inefficient unless appropriate numerical methods are used to improve the accuracy, speed, memory requirements, etc. Multi-scale integration is a ubiquitous problem in many areas of mathematical modeling and scientific computation, often computationally intensive, yet still not fully understood and an area of active research in the numerical analysis community (Barth et al. 2001). As better techniques become available, they will accelerate further model integration efforts.

The role of standards in systems biology

Currently, both models and data are often available without proper annotation, which makes automation of their integration infeasible. Annotation of model elements is required for matching parts of the two models (Stein 2001). In this particular area, systems biology strongly benefits from the application of computer science methods (http://www.w3.org/TR/rdf-concepts/, Smith and Hucka 2010). Still, various community standards are already available and partly integrated in common tools (Parkinson et al. 2010). There is a need for unifying syntax as well as semantics, and the latter is crucial for the model integration problem. For example, SBML is the de facto standard for exchanging dynamic models at the biochemical level, although there are many models represented in CellML, and in other emerging or forgotten standards, and some are stored in none standardized format. Using Resource Description Framework (RDF) relations (Southerna et al. 2008), modified versions of the same biochemical species (such as a protein and its phosphorylated version) are also possible to annotate. SBML, in particular, provides advanced annotation capabilities that allow the assignment of one or more unique database identifiers to each species, parameter, and reaction as well as expressing relationships among them. Also, standardized graphical notation is crucial for efficient and accurate communication of biological knowledge between researchers. Here, the Systems Biology Graphical Notation (SBGN) is an emerging standard for graphical notation (Jansson and Jirstrand 2010; Le Novère et al. 2009).

Finally, even when appropriate experimental data exist, it can be difficult to use them due to limited or inaccurate documentation and/or annotation. The implicit (or "hidden") data obscured by faulty documentation are, therefore, not communicated from experimentalist to modeler.

Standards for annotation of experimental data including detailed specification of experimental conditions are needed. A large effort has already been made to define standards for specific subjects, and these are collected in the Minimum Information for Biological and Biomedical Investigations (MIBBI, http://www.mibbi.org) (Taylor et al. 2009). The definition of standards and their comprehensive usage is primarily a communication issue that can be bridged by cross-disciplinary training, as accomplished in bioinformatics or biophysics.

Bridging the gap between wet and dry

Different scientific cultures lead to different approaches to—and expectations on—experimental design and model predictions. Importantly, models are built on specific hypotheses and cannot be used to verify the hypotheses they are based on, which instead only allows for a check of internal consistency. Failure to understand how—and why—models are built may lead to not only over interpretation of simulation results, but also to the collection of data that are not optimal for model building and validation. Data suitable for modeling might on its own not be the most efficient way to reach biological conclusions, but the iterative process between in silico hypothesis generation and experimental evaluation leads to an even better understanding of the analyzed system than possible by experimental research only. Besides, qualitative data are of limited use for systems biology and quantitative biological measurements as traditionally taken are often based on measurements of relative changes. Even highly precise, fold induction measurements are better than qualitative data for the purpose of model fitting and relatively little effort would be required to relate such measurements to entities per cell. In addition, it is imperative to stress the importance of time-resolved data; not only to be certain to hit the peak change, but also more importantly to be able to decipher causality which should be reflected in the temporal order. Overall, a stronger understanding of the modeling process would allow experimentalists to increase the usefulness and impact of the data they produce—with little or no additional cost.

Discrepancy between data produced and data needed in models

Direct inference of network functionality from network topology is a nontrivial problem (Cotterell and Sharp 2010). Even a completely defined topological network does not provide more than a static view of the analyzed system. The topology defines the possibilities within the network, but does not include the information on cause-and-effect that is absolutely required to understand the network's function. Nevertheless, tools are being developed that can extract biological information even from such static network representations. In addition, tools for inference of missing parts of network topologies are also being developed, but the predictions need to be experimentally validated. Large efforts have been made to define reaction topology using large-scale methods. To determine the information flow through the network, the topology must be complemented with the cause-and-effect information, resulting in a causal topology. This requires experimental data not only on which reactions may occur but also on how they are regulated. Typically such data cannot be generated with high-throughput methods, but must be addressed by dedicated experimentation; something which would be time consuming but still feasible.

A conceptually greater challenge is the discrepancy between the states used in mathematical models and the states explored experimentally. One of the best examples comes from the global protein–protein interaction (PPI) studies, in which experiments give information on whether a single pair of proteins interacts. The experiments may reveal a large number of such interaction partners, but give no information about which combinations may/must occur and in which cellular context. In contrast, mathematical models typically contain highly defined specific state variables, in which reactions are defined for each possible combination of interaction partners. Even dedicated experiments explore only very few such combinations. The discrepancy between the combinatorial state space in the models and the relatively small state space explored experimentally leads to uncertainty and implicit assumptions when data are mapped onto the models. For example, let us have experimental evidence that a protein has three interaction partners. A model would typically either include or exclude, explicitly, the single proteins, the tertiary complex, and/or the intermediate states. However, the experimental evidence proves neither the absence nor the presence of any dimers, trimers, or the tetramer. Hence, any definition of model structure would be based on guessing. While arbitrary model reductions may be required, it becomes an issue when these are implicit and cannot be distinguished from the real knowledge-based model. Since it is inconceivable that the empirical data will ever cover the entire possible state space of the network, it is necessary to adapt the network description according to the available data (Tiger et al. 2012). It will be critical to consequently adjust the modeling strategy and also to include new types of data or scientific approaches such as molecular modeling. Therefore, this is one of the largest challenges for the modeling community in systems biology.

Finding common languages for different scientific communities

To advance and stimulate research in systems biology, we propose two strategies that could contribute to overcoming the current problems often encountered in multi-disciplinary projects. One is to start training undergraduate students in systems biology and providing them with detailed knowledge in biology, chemistry, physics, and computer science (Wingreen and Botstein 2006). The alternative is to qualify highly skilled specialists with excellent understanding of their complementary disciplines and strong experience in interdisciplinary work, which is crucial to reach a systems level of understanding. To do so, closer collaboration between different research areas would be necessary on all levels, i.e., for the general laboratory organization as well as for individual scientists. Different laboratories interested in answering the same biological questions need to establish common project planning, including the definition of all aspects to be addressed, and allowing for building multi-disciplinary networks. Within these networks, elementary cross-disciplinary training at the beginning of a project will enable scientists to gain a better understanding of the biological problem and help define a common language. Paired working on the same topic/project(s) with complementary tools will be very helpful to ensure smooth communication and trust between the involved researchers. In particular, graduate students will benefit from learning by doing interdisciplinary research projects with dual supervision (mentors from complementary disciplines).

Conclusions

The future success of systems biology largely depends on bridging the current gaps between different scales of modeling abstraction, topological and mechanistic modeling, and wet and dry laboratories. Novel modeling and computational approaches need to deliver predictive models and also provide the means to incorporate uncertainty. Furthermore, the incorporation/integration of constraints in the modeling process is necessary to limit the model space given the sparsity of the data situation and often encountered mismatch between the complexity of a proposed class of models and information content in available data. To balance the available information content in measurement data with proposed model classes is also the objective of model reduction or expansion, and both these topics are needed for further development. Another area of great importance is model merging, which deals with the problem of obtaining more comprehensive descriptions of the operations of biological systems. The development of standards for both modeling and experimentation is crucial for facilitating model reuse and for increasing the pace of mapping out, describing, and better understanding biological systems by computational approaches. Related issues are the importance of finding common languages to enhance efficient communication between wet and dry laboratory scientists, as well as the importance of forming multi-disciplinary networks of different laboratories to attack the challenging problems lying ahead.

Acknowledgments This work was supported by the European Commission (Contract No. 223137, FutureSysBio).

References

Abbott A (2010) Germans cook up liver project. Nature 468:879

Agren R, Bordel S, Mardinoglu A, Pornputtapong N, Nookaew I, Nielsen J (2012) Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. PLoS Comput Biol 8(5):e1002518

Alberghina L, Cirulli C (2010) Proteomics and systems biology to tackle biological complexity: yeast as a case study. Proteomics 24:4337–4341

Barberis M, Klipp E, Vanoni M, Alberghina L (2007) Cell size at S phase initiation: an emergent property of the G1/S network. PLoS Comput Biol 3(4):e64

Bar-Shalom Y, Rong Li X, Kirubarajan T (2001) Estimation with applications to tracking and navigation. Wiley-Interscience, Hoboken

Barth T, Chan T, Haimes R (2001) Multiscale and multiresolution methods: theory and applications. Springer, New York

Bassingthwaighte J, Hunter P, Noble D (2009) The cardiac physiome: perspectives for the future. Exp Physiol 94:597–605

Chang A, Scheer M, Grote A, Schomburg I, Schomburg D (2008) BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009. Nucl Acids Res 37:D588–D592

Cotterell J, Sharp J (2010) An atlas of gene regulatory networks reveals multiple three-gene mechanisms for interpreting morphogen gradients. Mol Syst Biol 6:425

Cvijovic M, Olivares-Hernández R, Agren R, Dahr N, Vongsangnak W, Nookaew I, Patil KR, Nielsen J (2010) BioMet toolbox: genome-wide analysis metabolism. Nucl Acid Res 38:W144–W149

Erjavec N, Cvijovic M, Klipp E, Nyström T (2008) Selective benefits of damage partitioning in unicellular systems and its effects on aging. Proc Natl Acad Sci USA 105(48):18764–18769

Flöttmann M, Schaber J, Hoops S, Klipp E, Mendes P (2008) ModelMage: a tool for automatic model generation, selection and management. Genome Inf 20:52–63

Hucka M, Finney A, Sauro H, Bolouri H, Doyle JC, Kitano H, Arkin AP, Bornstein BJ, Bray D, Cornish-Bowden A, Cuellar AA, Dronov S, Gilles ED, Ginkel M, Gor V, Goryanin II, Hedley WJ, Hodgman TC, Hofmey JH, Hunter PJ et al (2003) The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. Bioinformatics 19:524–531

Jansson A, Jirstrand M (2010) Biochemical modeling with systems biology graphical notation. Drug Discov Today 15(9–10):365–370

Jazwinski AH (1970) Stochastic processes and filtering theory. Academic Press, New York

Jol SJ, Kümmel A, Terzer M, Stelling J, Heinemann M (2012) System-level insights into yeast metabolism by thermodynamic analysis of elementary flux modes. PLoS Comput Biol 8(3):e1002415

Jörnsten R, Abenius T, Kling T, Schmidt L, Johansson E, Nordling TE, Nordlander B, Sander C, Gennemark P, Funa K, Nilsson B, Lindahl L, Nelander S (2011) Network modeling of the transcriptional effects of copy number aberrations in glioblastoma. Mol Syst Biol 7:486

Klipp E, Nordlander B, Krüger R, Gennemark P, Hohmann S (2005) Integrative model of the response of yeast to osmotic shock. Nat Biotech 23(8):975–982

Kloeden PE, Platen E (1992) Numerical solution of stochastic differential equations. Springer-Verlag, New York

Kotte O, Zaugg JB, Heinemann M (2010) Bacterial adaptation through distributed sensing of metabolic fluxes. Mol Syst Biol 6:355

Krantz M, Ahmadpour D, Ottosson LG, Warringer J, Waltermann C, Nordlander B, Klipp E, Blomberg A, Hohmann S, Kitano H (2009) Robustness and fragility in the yeast high osmolarity glycerol (HOG) signal-transduction pathway. Mol Syst Biol 5:281

Krause F, Uhlendorf J, Lubitz T, Schulz M, Klipp E, Liebermeister W (2010) Annotation and merging of SBML models with semanticSBML. Bioinformatics 26:421–422

Kuepfer L, Peter M, Sauer U, Stelling J (2007) Ensemble modeling for analysis of cell signaling dynamics. Nat Biotech 25(9):1001–1006

Kümmel A, Panke S, Heinemann M (2006) Putative regulatory sites unraveled network-embedded thermodynamic analysis of metabolome data. Mol Syst Biol 2(2006):0034

Le Novère N, Bornstein B, Broicher A, Courtot M, Donizelli M, Dharuri H, Li L, Sauro H, Schilstra M, Shapiro B, Snoep J, Hucka M (2006) BioModels database: a free, centralized database of curated, published, quantitative kinetic models of biochemical and cellular systems. Nucl Acid Res 34:D689–D691

Le Novère N, Hucka M, Mi H, Moodie S, Schreiber F, Sorokin A, Demir E, Wegner K, Aladjem MI, Wimalaratne SM, Bergman FT, Gauges R, Ghazal P, Kawaji H, Li L, Matsuoka Y, Villéger A, Boyd SE, Calzone L, Courtot M et al (2009) The systems biology graphical notation. Nat Biotech 27(8):735–741

Liebermeister W (2008) Validity and combination of biochemical models. In Proceedings of 3rd International ESCEC workshop on experimental standard conditions on enzyme characterizations, Beilstein Institute, Frankfurt

Liebermeister W, Klipp E (2006) Bringing metabolic networks to life: integration of kinetic, metabolic, and proteomic data. Theor Biol Med Mode 3:42

Lloyd C, Lawson J, Hunter P, Nielsen P (2008) The CellML model repository. Bioinformatics 24:2122–2123

Lubitz T, Schulz M, Klipp E, Liebermeister W (2010) Parameter balancing in kinetic models of cell metabolism. J Phys Chem B 114(49):16298–16303

Mišković L, Hatzimanikatis V (2011) Modeling of uncertainties in biochemical reactions. Biotechnol Bioeng. 108(2):413–423

Nelander S, Wang W, Nilsson B, She QB, Pratilas C, Rosen N, Gennemark P, Sander C (2008) Models from experiments: combinatorial drug perturbations of cancer cells. Mol Syst Biol 4:216

Novak B, Tyson JJ, Gyorffy B, Csikasz-Nagy A (2007) Irreversible cell-cycle transitions are due to systems-level feedback. Nat Cell Biol 9(7):724–728

Olivier B, Snoep J (2004) Web-based kinetic modelling using JWS online. Bioinformatics 20:2144

Parkinson H, Sarkans U, Kolesnikov N, Abeygunawardena N, Burdett T, Dylag M, Emam I, Farne A, Hastings E, Holloway E, Kurbatova N, Lukk M, Malone J, Mani R, Pilicheva E, Rustici G, Sharma A, Williams E, Adamusiak T, Brandizi M et al (2010) ArrayExpress update—an archive of microarray and high-throughput sequencing-based functional genomics experiments. Nucl Acid Res 39:D1002–D1004

Resat H, Petzold L, Pettigrew M (2009) Kinetic modeling of biological systems. Methods Mol Biol 541:311–335

Ristic B, Arulampalam S, Gordon Neil (2004) Beyond the Kalman filter: particle filters for tracking applications. Artech Print on Demand

Schaber J, Liebermeister W, Klipp E (2009) Nested uncertainties in biochemical models. IET Syst Biol 3(1):1–9

Schulz M, Uhlendorf J, Klipp E, Liebermeister W (2006) SBMLmerge, a system for combining biochemical network models. Genome Inf 17:62–71

Smallbone K, Simeonidis E, Swainston N, Mendes P (2010) Towards a genome-scale kinetic model of cellular metabolism. BMC Syst Biol 28(4):6

Smith L, Hucka M (2010) SBML level 3 hierarchical model composition, COMBINE 2010, informatics forum, University of Edinburgh, 09 Oct 2010

Southerna J, Pitt-Francisb J, Whiteley J, Stokeley D, Kobashid H, Nobes R, Kadookad Y, Gavaghan D (2008) Multi-scale computational modelling in biology and physiology. Prog Biophys Mol Biol 96:60–89

Stein L (2001) Genome annotation: from sequence to biology. Nat Rev Genet 2:493–501

Taylor CF, Field D, Sansone SA, Aerts J, Apweiler R, Ashburner M, Ball CA, Binz PA, Bogue M, Booth T, Brazma A, Brinkman RR, Michael Clark A, Deutsch EW, Fiehn O, Fostel J, Ghazal P, Gibson F, Gray T, Grimes G et al (2009) Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. Nat Biotech 26(8):889–896

Tiger CF, Krause F, Cedersund G, Palmér R, Klipp E, Hohmann S, Kitano H, Krantz M (2012) A framework for mapping, visualisation and automatic model creation of signal-transduction networks. Mol Syst Biol 24(8):578

Wang L, Birol I, Hatzimanikatis V (2004) Metabolic control analysis under uncertainty: framework development and case studies. Biophys J 87(6):3750–3763

Wingreen N, Botstein D (2006) Back to the future: education for systems-level biologists. Nat Rev Mol Cell Biol 7(11):829–832

Wittig U, Golebiewski M, Kania R, Krebs O, Mir S, Weidemann A, Anstein S, Saric J, Rojas I (2006) SABIO-RK: integration and curation of reaction kinetics data. Bioinformatics 4075:94–103 (Lecture notes)