Special Issue: Computation and Modeling

Opinion

Respectful Modeling: Addressing Uncertainty in Dynamic System Models for Molecular Biology

Areti Tsigkinopoulou,[1] Syed Murtuza Baker,[1] and Rainer Breitling[1,*]

Although there is still some skepticism in the biological community regarding the value and significance of quantitative computational modeling, important steps are continually being taken to enhance its accessibility and predictive power. We view these developments as essential components of an emerging 'respectful modeling' framework which has two key aims: (i) respecting the models themselves and facilitating the reproduction and update of modeling results by other scientists, and (ii) respecting the predictions of the models and rigorously quantifying the confidence associated with the modeling results. This respectful attitude will guide the design of higher-quality models and facilitate the use of models in modern applications such as engineering and manipulating microbial metabolism by synthetic biology.

Computational Models in Current Research: Success and Skepticism

Quantitative computational models of cellular pathways and circuits are essential tools for generating clear, testable predictions about the behavior of complex cellular machineries [1,2]. Their use is currently moving beyond the proof-of-concept stage towards real-world applications, such as engineering and optimizing biological microorganisms to produce specific chemicals and biofuels [3–5], as has been shown most successfully for various terpenoids [6] and for succinic acid [7]. Their applications in identifying potential drug targets in metabolic or signaling pathways are also rapidly advancing [8–10]. An increasing number of success stories demonstrate that computational models have much to offer to biologists, from surveying cellular development [11] and exploring signaling pathways [12] and genetic circuits [13] to investigating potential treatments for cancer [14,15]. However, modeling is not yet part of the mainstream of biological practice, even in fields such as synthetic biology that aim to embrace an engineering approach to manipulating biological complexity [16–18].

Major progress is also being made in the case of genome-scale models [19–21] which match enzyme-coding genes with predicted reactions in metabolic pathways. Until recently their nature limited the application of genome-scale models to stoichiometric constraint-based approaches, studying the fluxes through the system without being able to predict metabolite levels or the dynamics of responses to internal or external perturbations. However, efforts are being made to incorporate detailed information on enzyme regulation and kinetic mechanisms into such constraint-based models, thus transforming them into genome-scale kinetic models of metabolism [22–24].

Trends

The increased use of computational models in biology requires radical rethinking of modeling strategies, and this is leading to a new comprehensive framework that we term 'respectful modeling'.

Maximum likelihood-based routine parameter fitting should be avoided because it may lead to entrapment in local optima, missing qualitatively different model behaviors.

Uncertainty should be acknowledged and incorporated during the earliest stages of model building because it allows predictions to be made with specified confidence intervals.

The respectful modeling approach constructs models that are easier to maintain and update when new experimental evidence becomes available.

The respectful modeling approach will facilitate collaborations and promote the wider use of predictive models in areas such as synthetic biology and personalized medicine.

[1]Manchester Centre for Synthetic Biology of Fine and Speciality Chemicals (SYNBIOCHEM), Manchester Institute of Biotechnology, Faculty of Science and Engineering, University of Manchester, 131 Princess Street, Manchester M1 7DN, UK

 http://dx.doi.org/10.1016/j.tibtech.2016.12.008
© 2016 Elsevier Ltd. All rights reserved.

*Correspondence:
rainer.breitling@manchester.ac.uk
(R. Breitling).

We aim to outline 'respectful modeling' (Figure 1, Key Figure): an emerging set of closely related concepts and techniques that together enhance and advance earlier modeling approaches to make computational models both more approachable and more relevant for the work of experimental biologists.

Many molecular biologists still doubt the relevance of computational models of cellular systems. One sometimes even hears the truism that 'all models are wrong, but some are useful' [25], cited as if it meant 'all models are false and most are useless'. In fact, models are meant to be explicit descriptions of our implicit knowledge about the function of a biological system. Therefore, they are ideally no more wrong than less-formal reasoning about the same system, but they offer numerous advantages: most importantly, they put any assumptions about how a cellular mechanism works out into the open for everyone else to check and criticize, and they allow predictions about non-obvious ('emergent') behaviors of a system that follow compellingly from these assumptions, even when the complexity and non-linearity of the systems make simple back-of-the-envelope arguments impossible.

Nonetheless, beyond the anecdotal observation that models are sometimes flippantly dismissed as being 'always wrong', there are more serious indications that computational models of cellular pathways and circuits are not yet fully respected as scientific tools in molecular biology. For example, complex systems models are very rarely updated and developed further; with a few rare exceptions, such as some constraint-based models [26,27], there are not many

Key Figure

Respectful Modeling Manifests Itself in Various Steps of the Pipeline

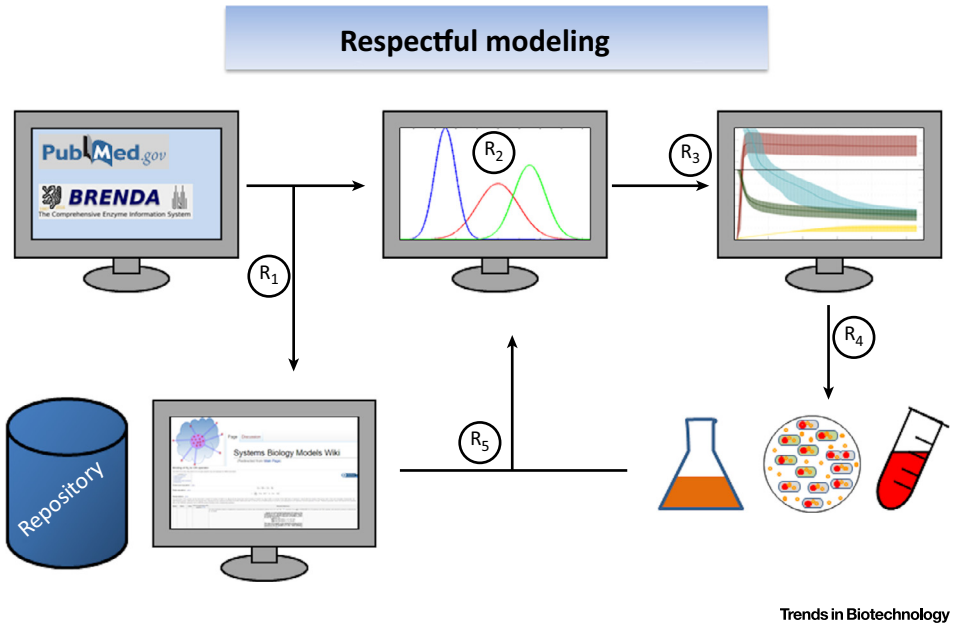

Figure 1. (R1) Respect for model construction: full documentation of all modeling decisions and sources in reusable form for future updating. (R2) Respect for limited knowledge: principled consideration of uncertainty of model parameter data. (R3) Respect for model predictions: quantitative assessment of confidence intervals for all model predictions. (R4) Respect for model hypotheses: using models to inform real-world experimental hypothesis testing. (R5) Respect for existing models: rigorous strategies for updating based on new experimental evidence.

examples of 'versioned' and iteratively improved models. Moreover, predictions of computational models in systems biology are usually reported without error bars. By contrast, the incorporation of uncertainty in models has long been common practice in other fields [28,29], from natural hazard insurance to economic forecasts, where the predictions and their associated confidence make a real difference to people's lives. For example, in the prediction of human-induced climate change, error bars and the exploration of multiple scenarios are the dominant feature in all predictions, to the extent that they are even presented and discussed in the popular press [30].

A Respectful Approach to Biological Modeling

What then is needed to establish the same 'respectful' attitude to modeling in biology such that it can meaningfully contribute to areas such as synthetic biology or personalized medicine? Two aspects seem to be central.

(i) Respect for the model itself as a resource that can grow and develop. If we respect our models, we should not treat them as one-off exercises. Instead, we need to make them understandable and reproducible by other scientists, and update them regularly and continuously instead of discarding them once new discoveries are made (or a new PhD student takes on the project). Therefore, reaction mechanisms and parameter values (such as assumed rate constants, enzyme concentrations, and substrate affinities) need to be documented to allow others to build on the work, and models that capture the full breadth of alternative hypotheses need to be maintained instead of fixating on one 'maximum likelihood' or preferred model. In this way, the model can improve iteratively as new data become available instead of having to be rebuilt each time.

(ii) Respect for the output of the model and for the predictions it generates. We should strive for accurate predictions that can be used for rational decision-making. When respecting our models, in the sense that we really care about their predictions, we need to care about the associated level of uncertainty, and we also need to identify the different levels of confidence associated with each alternative prediction. Otherwise, we could not place reasonable bets on different possible outcomes. Therefore, we need model predictions with confidence intervals that correctly reflect our current (limited) knowledge about the topology and dynamics of the biological system. We can only ignore confidence intervals if we are not interested in the real-world implications of our quantitative model predictions, in other words when we only treat them as little more than random numbers that illustrate the elegant model-building and analysis algorithms. Once we have a true stake in generating accurate quantitative predictions, because we want to apply them in the design of actual biological experiments or the engineering of real living systems, we will listen much more carefully to what the models have to tell us. Looking for the uncertainty in the model predictions is not a sign of disrespectful mistrust but an immediate consequence of taking the model serious as a predictive, rather than illustrative, tool.

In fact, these two aspects are closely related. They both imply that our models should not be rigidly fixed but need to capture alternative scenarios, alternative parameter values [31], alternative circuit topologies [32], and generally alternative hypotheses about the biological system that is being studied. A recent study on energy metabolism in the protozoan parasite *Trypanosoma brucei* [33,34], the causative agent of sleeping sickness, illustrates how an explicit treatment of uncertainty can yield new insights even for well-studied model organisms (Box 1). The 'respectful modeling' data identify several previously unexplored key experiments. In the trypanosome example, for instance, the exact level of permeability of the glycosomes for specific glycolytic intermediates could be determined by a more targeted measurement, or the sensitivity of trypanosomes to the inhibition of glycosomal enzymes with unexpectedly high **control coefficients** (see Glossary; i.e., triose phosphate isomerase) could be directly

Glossary

Bayesian statistical framework: a model analysis framework aiming to derive a posterior probability that expresses the modeler's degree of belief in various alternative hypotheses (e.g., different model topologies or parameter values). These beliefs can be revised in a principled way in view of new experimental information.

Bivariate distribution: a probability distribution from which two random variables are sampled simultaneously; a multivariate distribution is one from which *n* variables are sampled simultaneously. If the variables are independent, the results are identical to those obtained by sampling from the corresponding univariate (marginal) distributions; however, if there are correlations or other dependencies between the variables, they are also taken into account, and shape the scatter of the sampled values.

Control coefficient: the system property of an enzyme that expresses how much influence (i.e., control) an infinitesimal change in the activity of an enzyme has on a systemic variable, usually a flux or a metabolite concentration.

Ensemblemodeling: a modeling strategy that employs a collection of models to represent the range of alternative structures and parameter values that could plausibly describe a given biological system; an analysis of the collection of models reveals the range of possible behaviors of the system and allows the confidence intervals for model predictions to be calculated.

Gibrat's law of proportionate effect: a rule defined by the French engineer and econometrician Robert Gibrat stating that the proportional rate of growth of a firm is independent of its size. The law of proportionate growth eventually leads to a log-normal distribution of firm sizes. Analogous processes explain the widespread occurrence of log-normal distributions in biology.

Location and scale parameters (μ and σ): these parameters determine the shift and the spread of a distribution, respectively. In a normal distribution, these are equivalent to the mean and SD; however, for other types of distributions the exact mathematical relationship between

measured. Therefore, these two aspects represent different levels of alternatives: on one hand, the degree of confidence in our predictions can be determined by examining ensembles of equally plausible models. On the other hand, models can be updated and reused as our biological understanding evolves by changing our assumptions about which models (and parameter values) are plausible and which are not. The advantages of respectful modeling are obvious:

(i) It quantitatively indicates the robustness of each prediction, and this is helpful for identifying the most suitable experiments to carry out next: do we already have enough confidence in a prediction to build an expensive study around it, or are there specific uncertainties that we first need to reduce by targeted measurements?

(ii) It enables collaborative and iterative work on model building and model improvement by explicitly identifying and documenting alternative model structures and parameter values.

(iii) It allows managing alternative hypotheses about the functions of complex biological systems within a unified modeling framework, instead of having slightly different, incompatible, and often incomparable models associated with each hypothesis.

(iv) Finally, and perhaps most importantly, it makes modeling approaches available for systems in which quantitative information is incomplete and uncertain, thus unveiling otherwise inaccessible biological phenomena. This connection may also help to bridge the conceptual gap between quantitative and qualitative modeling [35].

Box 1. 'Respectful Modeling' of Trypanosome Metabolism

In trypanosome parasites in the human bloodstream, energy metabolism is mostly restricted to glycolysis, and this takes place in unique, specialized organelles, the glycosomes; its enzymes are considered promising targets for newly developed drugs against sleeping sickness.

The steps followed during the model design and analysis are depicted in Figure I. Starting from a highly curated model of trypanosomal energy metabolism, the experimental uncertainty of every enzyme kinetic parameter was determined by an extensive exploration of the original literature. Parameter sources, evidence for alternative model topologies (extra reactions), and any calculations (e.g., for the parameter means and SDs) were documented in a dedicated Wikipedia-based database (A). Several versions of the model were created in which alternative groups of metabolites with different molecular weights could freely diffuse across the membrane, representing the uncertainty about possible glycosome permeability resulting from recent evidence that the organelle membrane contains non-selective pores (B).

For each of the alternative models, ranging from very tight to very leaky glycosomes, plausible combinations of parameter values were sampled according to the documented uncertainty by using a random number generator in accordance with the assumed probability distribution of each parameter. This resulted in a large collection of model variants, each using a different set of parameter values (C). The ensembles of models were then subjected to the same types of analysis as traditional dynamic models, for example determining steady-state concentrations of metabolites and calculating the control coefficients of enzymatic reactions that could identify the most promising drug targets (D).

The analysis of the resulting ensembles of model predictions provided several interesting insights that had remained 'hidden' in classical maximum-likelihood analyses of individual models. For example, it revealed unexpected fragilities in the existing models. Two metabolites, 3-phosphoglycerate and pyruvate, seemed to accumulate to impossible concentrations in many of the models, indicating that crucial regulatory loops are probably still unaccounted for in our current understanding of trypanosome metabolism. Moreover, control of glycolytic flux seemed to be more widely distributed between several key reaction steps instead of being largely restricted to the rate-limiting glucose uptake transporter. The results also showed that models that predicted steady-state metabolite concentrations and fluxes most closely matching experimental observations, for the largest number of plausible parameter sets, were those in which glycosomes were permeable to small metabolites up to the size of fructose 6-phosphate and fructose 1,6-bisphosphate. This result challenges the current consensus view of trypanosome glycolysis that crucially depends on highly controlled transmembrane fluxes, but is in good agreement with the presence of recently discovered nonspecific glycosomal pores.

these quantities can be more complex.

Markovchain Monte Carlo (MCMC) methods: methods for sampling random variables from a probability distribution; these are often employed to sample parameter values for models in systems biology. MCMC methods are particularly useful and efficient when sampling from distributions with complicated non-standard shapes.

Sensitivity analysis: a method for model analysis that quantifies the influence of individual parameters on the behavior of a system. In molecular biology, sensitivity analysis is often known as 'metabolic control analysis' (MCA), and the control coefficients of MCA are equivalent to the sensitivities of classical sensitivity analysis. Sensitivity analysis can be 'local', when exploring the effect of small variations in a parameter value around its given value in an otherwise fixed model, or 'global', when exploring the influence of a parameter in the context of all possible combinations of parameter values, quantifying how much the uncertainty about a given parameter contributes to the overall uncertainty about the behavior of a model. Global sensitivity analysis is closely related to some of the concepts of respectful modeling.

XML-based language: a programming language that uses the extensible mark-up language (XML) format. This format follows a set of rules to describe data in a way that makes them readable by both machines and humans.

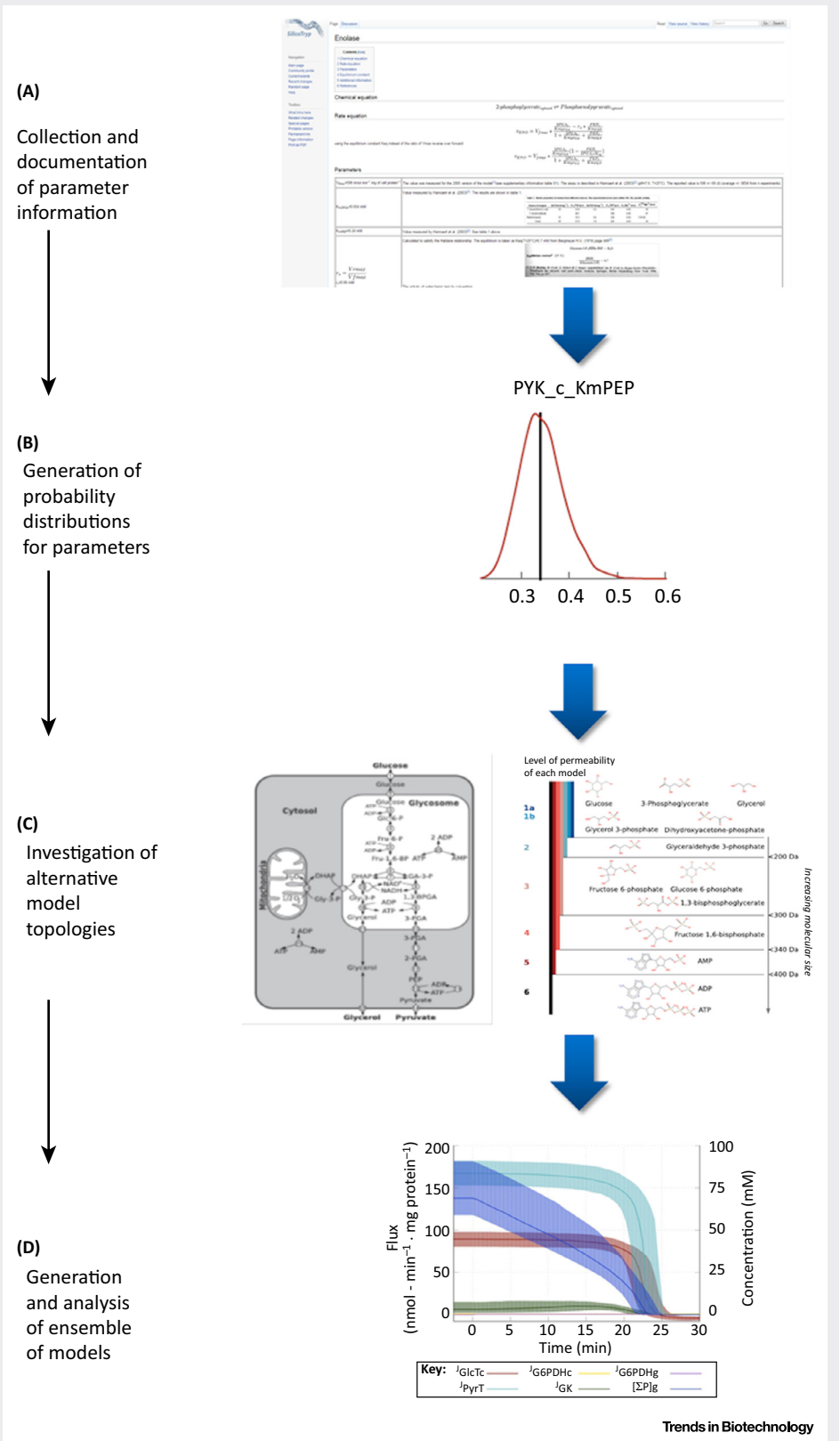

Figure I. Flowchart of Steps Followed During Dynamic Modeling of *Trypanosoma brucei* Energy Metabolism with Explicit Consideration of Parameter Uncertainty [33,34,50].

Respectful Modeling Showcases and Applications

Several important steps have recently been taken to provide the ingredients for a respectful modeling approach.

(i) Model descriptions are being standardized. This practice is becoming applied to every aspect of the model building process, from file formats to variable names and equation structures and to mandatory archiving of the resulting models. The systems biology markup language (SBML) [36], a machine-readable **XML-based language**, is supported by almost all software tools in the field and is prevailing as a model representation format that facilitates model sharing and replication. In addition, models are starting to comply with some basic community guidelines that describe the minimum information that needs to accompany a published model (MIRIAM) [37] together with conventions that facilitate data integration in model building, such as SBtab [38]. Moreover, centralized model repositories such as BioModels [39], JWS Online [40], the CellML model repository [41], and the BiGG database [42] have been created to enable the distribution and curation of biological system models.

(ii) Parameter information is being rigorously documented. The inclusion of supplementary material on the sources and values of the parameters (e.g., enzyme and substrate affinities, transcription rates, etc.) in published models is increasing [43–46]. For example, recently published updated models of central carbon metabolism in trypanosome parasites included a Wikipedia page dedicated to each biochemical reaction of the system [33,34,47]. In this way, detailed information on the sources for each parameter value was provided, together with descriptions of the underlying calculations and assumptions, as well as alternative model versions with different topologies. This approach not only increases the accountability of modelers for the crucial decisions made during model building; it also greatly facilitates the reconstruction, validation, and updating of models by successive generations of researchers.

Box 2. Generating Probability Distributions To Describe Uncertainty

Describing the uncertainty associated with our knowledge of parameter values in molecular systems models is challenging. The natural choice for the shape describing the range of plausible parameter values is often a log-normal distribution: there will be a most likely value (the mode of the distribution), negative values are not allowed, and the distribution is symmetrical, in the sense that values that are x-fold larger than the most-likely estimate are equally plausible as values that are x-fold smaller (Figure I). More specifically, the mode is the value x_o for which the condition $f(x_o{*}\delta) = f(x_o/\delta)$ is fulfilled for all real numbers δ (where f is the probability density function).

In contrast to common assumptions, the log-normal distribution is ubiquitous in nature across different fields and is often a far better description of the data than a normal distribution. This is particularly seen in the medical and biological sciences, when the mean values are low and the variances are large, but the values cannot be negative, irrespective of whether they are latency periods of infectious diseases in epidemiology, species abundances in ecological studies, or enzyme kinetic parameters (as in the present discussion) [74]. The reason for this predominance of log-normal distributions is simple: when the change in a variable (observable) at each step of the process is proportional to its current value (i.e., when the process follows **Gibrat's law of proportionate effect**), the resulting observations will be log-normally distributed [75]. In the life sciences, biochemistry, biophysics, and population ecology are disciplines where processes characterized by Gibrat's law are most obvious: reaction velocities, surface and volume measurements, and population growth (also at the molecular level) are regulated by factors that act in a multiplicative way rather than in the additive way that would be required for data to become normally distributed.

The challenge for the biological systems modeler is to decide on the appropriate values describing the distribution – what is the most likely value (described by the mode of the distribution), and how rapidly does the plausibility of the values decrease when moving away from this value (described by the SD)? These in turn determine the **location and scale parameters** μ and σ of the log-normal distribution. In many cases a good estimate of the most likely parameter value exists, for example from actual experimental measurements. In other cases, related parameters have been measured, for example the kinetics of similar enzymes or even of the same enzyme in different species or conditions. In the extreme case, no measurements are available at all; but even in that case, an informed guess is usually possible. For instance, the most plausible K_m value of an uncharacterized novel enzyme might be the average of all K_m values ever

recorded in a comprehensive enzyme database, such as BRENDA, which comprehensively records available experimental data on enzyme kinetic parameters from all domains of life. The spread of the distribution, in other words its SD, would in such a case reasonably be determined by the range of all reported values (Figure II).

It is also necessary to decide if there are any hard (biophysical) thresholds beyond which a parameter value is not only implausible but impossible; in other words, if the distribution should be truncated at the extremes. In every case it is important to remember that determining plausible parameter distributions is not a mechanical exercise but must be based on actual biological and biophysical reasoning. If there are arguments to support the idea that a given log-normal distribution does not match the expected range of plausible values, then it needs to be adjusted accordingly.

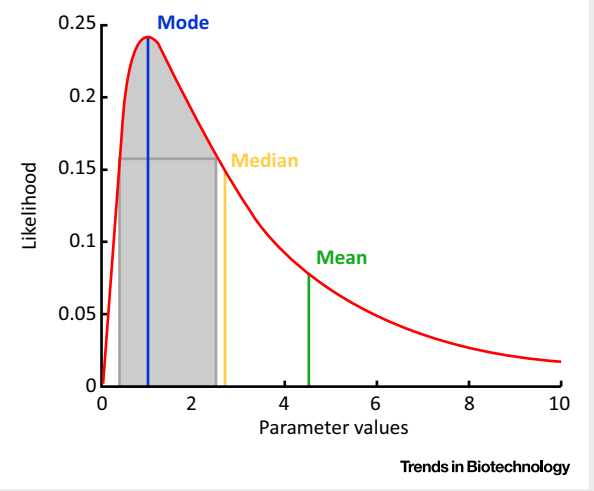

Figure I. The Properties of a Log-Normal Distribution with $\mu = \sigma = 1$. The mode represented by the blue line is equal to 1, the mean is equal to 4.48 (green line), and the median is equal to 2.7 (yellow line). For any value of δ the product, and the quotient of the mode and δ, have equal probability of being sampled [f(1*δ) = f(1/δ); illustrated for δ = 2.5 by the edges of the grey area].

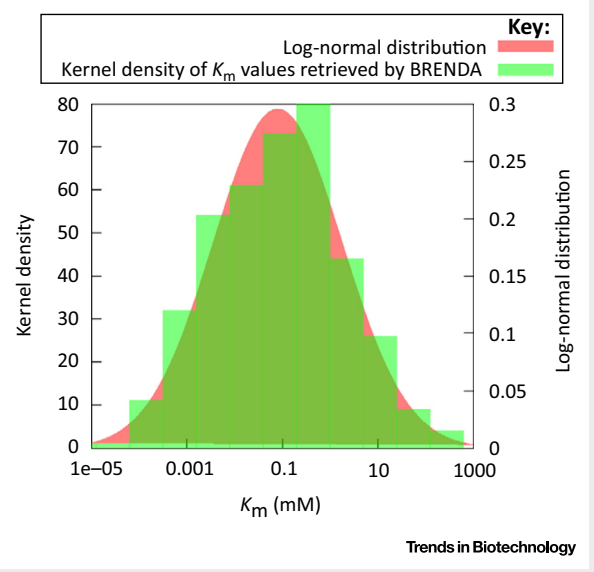

Figure II. Distribution of K_m Values Retrieved from the BRENDA Database (www.brenda-enzymes.org/). K_m values retrieved from the database are shown in green, and the corresponding fitted log-normal distribution is shown in red [34].

(iii) Moreover, the detailed documentation of parameter sources helps to quantify the uncertainty associated with each value such that model predictions can take these uncertainties into consideration intuitively. A formal acknowledgement of uncertainty in computational models of biological systems not only describes the beliefs and confidence of modelers in model structure and model parameters but also explicitly identifies alternative structures and parameter values and their associated plausibility [33,34]. This approach mirrors the process of scientific progress by continuously exploring alternative hypotheses. Thus, instead of paralyzing any analysis using the model, the explicit acknowledgement of uncertainty actually enables the flexible evolution of biological models which at each stage honestly represent our current knowledge about a biological system. This is in contrast to traditional modeling approaches where alternative hypotheses and assessments of uncertainty are managed only in an *ad hoc* process taking place implicitly in the brains of the modelers and their collaborating expert biologists. Furthermore, acknowledging uncertainty and integrating it during the model-building phase allows making predictions that are associated with specified confidence intervals, which can guide further experimentation [48,49].

(iv) To transition from the collected information about parameter uncertainty to the resulting quantitative assessment of our confidence in specific model predictions, data-driven parameter sampling strategies have been formulated [33,34,47,50,51]. They employ informative

Box 3. Thermodynamic Consistency

Once the plausible values for each parameter have been described, there is another important factor that needs to be considered to decide if combinations of parameters are plausible: thermodynamic consistency. For example, imagine a reaction that is known to have an equilibrium constant very close to 1; in other words, the standard Gibbs free energy $\Delta G^{\circ} = 0$. We are trying to determine the kinetic parameters for the forward and backward component of this reaction, but we know little about the rate of the reaction, and therefore we sample each of the two parameters from a very broad distribution. If we do not take the additional thermodynamic information into account, we will often end up sampling the forward reaction rate from the 'fast' end of the spectrum, and the backward rate from the 'slow' end, or vice versa. Thermodynamic consistency requires that we discard such samples and only keep samples where the two reaction rates are very similar (how similar will in turn depend on our uncertainty about the equilibrium constant) (Figure I).

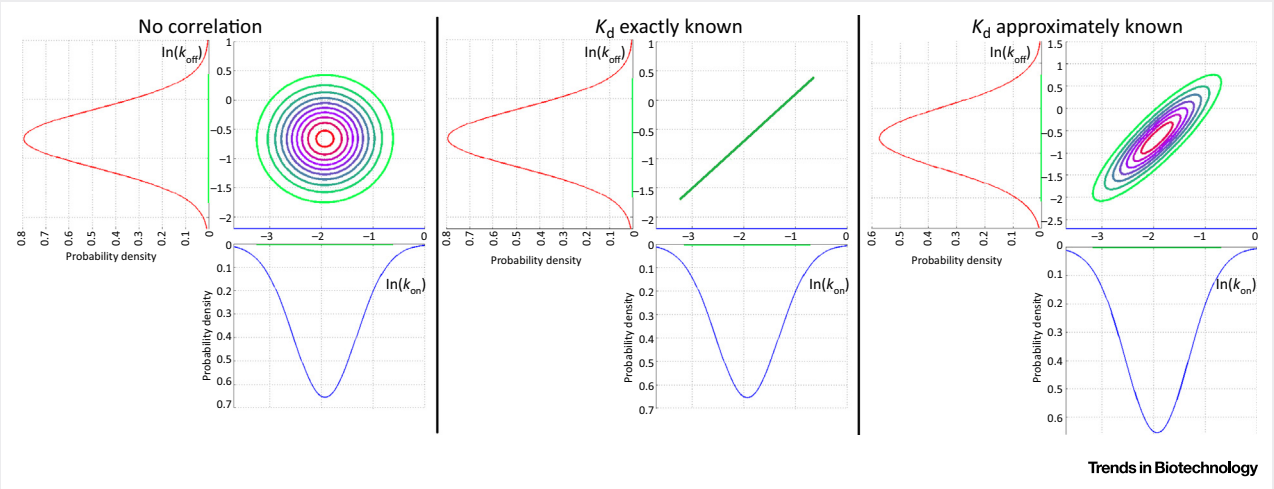

Figure I. Sampling Strategy for Maintaining Thermodynamic Consistency of Parameter Sets. To ensure the thermodynamic feasibility of a parameter combination in the case of interconnected parameters (i.e., the forward and backward reaction constants k_{on} and k_{off}), a **bivariate distribution** is created. When the two marginal distributions are non-correlated, the points generated to represent parameter pairs form a circle. When the equilibrium dissociation constant, K_d, is exactly known, the parameters k_{on} and k_{off} are tightly correlated through the K_d value, and the points form a straight line [51]. Finally, if K_d is approximately known (i.e., a distribution of values for K_d is available) the resulting points of the bivariate system form an ellipse. The thickness and orientation of the ellipse depend on the magnitude of the correlation between the two marginal distributions and on the degree of uncertainty in the values of k_{on}, k_{off}, and K_d (A. Tsigkinopoulou *et al.*, unpublished). This case represents a realistic modeling scenario because the parameter values are usually approximately known. The first case (no correlation) does not respect thermodynamic consistency and is therefore undesirable. The second case, although taking into account the dependency of the two parameters, is in most cases unrealistic because the value of a parameter such as the equilibrium constant is rarely exactly known.

distributions to describe what the modelers (and their biologist collaborators) consider to be plausible values for each parameter. These distributions correspond to the priors in a **Bayesian statistical framework** [52,53], and can indeed be used for a Bayesian statistical analysis [54,55] to update the parameter values when new experimental evidence becomes available, but, most importantly, they can be used in the next step to create an entire ensemble of plausible models by sampling values for each parameter from its corresponding distribution (Box 2). Experimental data, biological background knowledge, and biophysical plausibility (Box 3) can all contribute to defining the most appropriate distributions that capture our current state of knowledge accurately, not exaggerating the uncertainty, but also not being overconfident about specific values or connections in the network.

(iv) Approaches that explicitly acknowledge model uncertainty have been developed and employed [31], such as **Markovchain Monte Carlo** (MCMC) methods [56–58], **ensemblemodeling** [59,60], and global **sensitivity analysis** [52,61,62]. Such techniques are based on the concept of sampling parameters from their associated probability distributions, thereby creating a collection of models that can undergo further analysis in a similar way as typical dynamic models. In contrast to analyses of the local effects of variable parameter values around their preferred value, the entire resulting landscape of solutions can be surveyed without focusing only on one optimal solution – which may in the future be rejected once new data become available. This approach is also highly preferable over methods that try to fit a single 'maximum likelihood' set of parameters based on the best match to experimental data [54]. Such a 'fitting' strategy is very popular in applications including transcriptional dynamics [63,64], epigenetics [65], neuronal dynamics [66], and population-level epidemiology [67].

Trends in Biotechnology

Figure 2. By Focusing Only on the 'Best-Fit Solution', A Modeler Might End Up Trapped in a Local Optimum. The modeler might thus miss interesting alternatives, which, in the future, might be supported by new experimental data. Instead, by surveying the entire landscape of solutions without focusing on one particular peak, respectful (Bayesian) modeling remains adaptable to future developments, especially when studying complex systems with rugged likelihood surfaces, where the availability of new experimental data can result in rapid changes in the relative likelihood of alternative model scenarios (courtesy of Danai Triantafyllopoulou).

However, fitted parameters are also very likely to be trapped in transient and spurious global optima, given that the likelihood surface for complex biological systems is inherently extremely rugged (highly non-convex), and are always in danger of leading to overfitting and a failure to identify alternative hypothesis that could explain the results equally well [1,31]. Moreover, parameter values, once fitted, are rarely updated in the face of new experimental results. Avoiding fitting parameter values according to the maximum likelihood is particularly important in the case of biological models as their highly rugged likelihood surface means that models with considerably different sets of parameter values can have very similar likelihoods (Figure 2). If one set is preferred based on a momentarily higher likelihood, alternative options are easily overlooked later on. To avoid this pitfall, all options are kept in view in a respectful modeling approach, and the model can be easily adapted. In this regard, and in important details of the parameter distributions [68], the respectful modeling approach closely approximates the Bayesian inference processes supposedly implemented in the human neocortex [69].

Outstanding Questions

Are there limits to the use of log-normal distributions as the best description of the plausibility of different parameter values?

In which scenarios can respectful modeling and traditional parameter fitting approaches be combined for even more powerful modeling strategies?

How should respectful modeling be incorporated into automated model-building pipelines?

Concluding Remarks

It is probably clear by now that all of these developments offer reasonable and useful modifications to the way we build and analyze computational models for molecular biology. However, why will they make a difference? They bring us closer to the ideal expressed in another famous saying about computational modeling: 'Models are not meant to be descriptions, pathetic descriptions, of nature; they are designed to be accurate descriptions of our pathetic thinking about nature' (J. Black, cited in [70]). To achieve this ideal, each model has to allow us to capture the 'pathetic' aspects of our thinking, the uncertainties and incompleteness of the evidence, and to evolve as our thinking evolves on the basis of new experiments [71,72]. In the future of molecular systems biology and modeling, nothing is certain except uncertainty itself (see Outstanding Questions). The increasing use of automated model building strategies will only increase the challenge [73] as models grow in size and the specific refinement of individual parameters by targeted experiments becomes even less feasible than it is now. The adoption of a respectful modeling framework will promote and facilitate collaboration within the biological community, stimulating the use of models not only as a tool for fundamental research but also as a valuable guide for the predictive engineering of biological systems and their informed manipulation by increasingly personalized drugs.

Supplemental Information

Supplemental information associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.tibtech.2016.12.008.

References

1. Almquist, J. *et al.* (2014) Kinetic models in industrial biotechnology – improving cell factory performance. *Metab. Eng.* 24, 38–60
2. Wolkenhauer, O. (2014) Why model? *Frontiers in Physiology* 5, 21
3. Fisher, A.K. *et al.* (2014) A review of metabolic and enzymatic engineering strategies for designing and optimizing performance of microbial cell factories. *Comput. Struct. Biotechnol. J.* 11, 91–99
4. Lee, J.W. *et al.* (2012) Systems metabolic engineering of microorganisms for natural and non-natural chemicals. *Nat. Chem. Biol.* 8, 536–546
5. Dai, Z. and Nielsen, J. (2015) Advancing metabolic engineering through systems biology of industrial microorganisms. *Curr. Opin. Biotechnol.* 36, 8–15
6. Chung, B.K.-S. *et al.* (2013) Genome-scale in silico modeling and analysis for designing synthetic terpenoid-producing microbial cell factories. *Chem. Eng. Sci.* 103, 100–108
7. Ahn, J.H. *et al.* (2016) Production of succinic acid by metabolically engineered microorganisms. *Curr. Opin. Biotechnol.* 42, 54–66
8. Iadevaia, S. *et al.* (2010) Identification of optimal drug combinations targeting cellular networks: integrating phospho-proteomics and computational network analysis. *Cancer Res.* 70, 6704–6714
9. Breitling, R. and Takano, E. (2015) Synthetic biology advances for pharmaceutical production. *Curr. Opin. Biotechnol.* 35, 46–51
10. Takano, E. and Breitling, R. (2015) Synthetic biology of antibiotic production. In *Reviews in Cell Biology and Molecular Medicine* (Meyers, R.A., ed.), John Wiley & Sons
11. Fard, A.T. *et al.* (2016) Not just a colourful metaphor: modelling the landscape of cellular development using Hopfield networks. *NPJ Syst. Biol. Appl.* 2, 16001
12. Fey, D. *et al.* (2015) Signaling pathway models as biomarkers: patient-specific simulations of JNK activity predict the survival of neuroblastoma patients. *Sci. Signal.* 8, ra130
13. Hu, C.Y. *et al.* (2015) Generating effective models and parameters for RNA genetic circuits. *ACS Synth. Biol.* 4, 914–926
14. Barbolosi, D. *et al.* (2016) Computational oncology – mathematical modelling of drug regimens for precision medicine. *Nat Rev Clin Oncol* 13, 242–254
15. Folger, O. *et al.* (2011) Predicting selective drug targets in cancer through metabolic networks. *Mol. Syst. Biol.* 7, 501

16. Yadav, V.G. *et al.* (2012) The future of metabolic engineering and synthetic biology: towards a systematic practice. *Metab. Eng.* 14, 233–241
17. Stanford, N.J. *et al.* (2015) RobOKoD: microbial strain design for (over)production of target compounds. *Front. Cell Dev. Biol.* 3, 17
18. Medema, M.H. *et al.* (2012) Computational tools for the synthetic design of biochemical pathways. *Nat Rev Micro* 10, 191–202
19. Hyotylainen, T. *et al.* (2016) Genome-scale study reveals reduced metabolic adaptability in patients with non-alcoholic fatty liver disease. *Nat Commun* 7, 8994
20. Fong, S.S. (2014) Computational approaches to metabolic engineering utilizing systems biology and synthetic biology. *Comput. Struct. Biotechnol. J.* 11, 28–34
21. Motamedian, E. *et al.* (2016) Reconstruction of a charge balanced genome-scale metabolic model to study the energy-uncoupled growth of *Zymomonas mobilis* ZM1. *Mol. BioSyst.* 12, 1241–1249
22. Link, H. *et al.* (2014) Advancing metabolic models with kinetic information. *Curr. Opin. Biotechnol.* 29, 8–14
23. Chakrabarti, A. *et al.* (2013) Towards kinetic modeling of genome-scale metabolic networks without sacrificing stoichiometric, thermodynamic and physiological constraints. *Biotechnol. J.* 8, 1043–1057
24. Smallbone, K. and Mendes, P. (2013) Large-scale metabolic models: from reconstruction to differential equations. *Ind. Biotechnol.* 9, 179–184
25. Box, G.E.P. (1979) *Robustness in the Strategy of Scientific Model Building,* University of Wisconsin-Madison Mathematics Research Center
26. Reed, J.L. *et al.* (2003) An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biology* 4, R54
27. Swainston, N. *et al.* (2016) Recon 2.2: from reconstruction to model of human metabolism. *Metabolomics* 12, 109
28. Peterson, J.T. and Freeman, M.C. (2016) Integrating modeling, monitoring, and management to reduce critical uncertainties in water resource decision making. *J. Environ. Manage.* 183, 361–370
29. Arnone, E. *et al.* (2014) Parameter incertainty in shallow rainfall-triggered landslide modeling at basin scale: a probabilistic approach. *Proc. Earth. Planet. Sci.* 9, 101–111
30. Antilla, L. (2005) Climate of scepticism: US newspaper coverage of the science of climate change. *Global Environ. Change* 15, 338–352
31. Vanlier, J. *et al.* (2013) Parameter uncertainty in biochemical models described by ordinary differential equations. *Math. Biosci.* 246, 305–314
32. Babtie, A.C. *et al.* (2014) Topological sensitivity analysis for systems biology. *Proc. Natl. Acad. Sci. U.S.A.* 111, 18507–18512
33. Achcar, F. *et al.* (2013) Explicit consideration of topological and parameter uncertainty gives new insights into a well-established model of glycolysis. *FEBS J.* 280, 4640–4651
34. Achcar, F. *et al.* (2012) Dynamic modelling under uncertainty: the case of *Trypanosoma brucei* energy metabolism. *PLoS Comput Biol* 8, e1002352
35. Samaga, R. and Klamt, S. (2013) Modeling approaches for qualitative and semi-quantitative analysis of cellular signaling networks. *Cell Communication and Signaling* 11, 43
36. Hucka, M. *et al.* (2015) The systems biology markup language (SBML): language specification for level 3 version 1 core. *Journal of Integrative Bioinformatics* 12, 266
37. Novere, N.L. *et al.* (2005) Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat Biotech* 23, 1509–1515
38. Lubitz, T. *et al.* (2016) SBtab: a flexible table format for data exchange in systems biology. *Bioinformatics* 32, 2559–2561
39. Li, C. *et al.* (2010) BioModels.net Web Services, a free and integrated toolkit for computational modelling software. *Briefings in Bioinformatics* 11, 270–277
40. van Gend, C. *et al.* (2007) Data and model integration using JWS online. *In Silico Biology* 7, 27–35
41. Cuellar, A. *et al.* (2015) The CellML 1.1 specification. *Journal of Integrative Bioinformatics* 12, 259
42. King, Z.A. *et al.* (2016) BiGG models: a platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.* 44, D515–522
43. Pai, A. and You, L. (2009) Optimal tuning of bacterial sensing potential. *Mol. Syst. Biol.* 5, 286-286
44. Mattioni, M. and Le Novère, N. (2013) Integration of biochemical and electrical signaling – multiscale model of the medium spiny neuron of the striatum. *PLoS ONE* 8, e66811
45. Singh, V.K. and Ghosh, I. (2006) Kinetic modeling of tricarboxylic acid cycle and glyoxylate bypass in *Mycobacterium tuberculosis*, and its application to assessment of drug targets. *Theor. Biol. Med. Model.* 3, 27
46. Smallbone, K. and Corfe, B.M. (2014) A mathematical model of the colon crypt capturing compositional dynamic interactions between cell types. *Int. J. Exp. Pathol.* 95, 1–7
47. Breitling, R. *et al.* (2013) Modeling challenges in the synthetic biology of secondary metabolism. *ACS Synth. Biol.* 2, 373-378
48. Engelhardt, B. *et al.* (2016) Learning (from) the errors of a systems biology model. *Sci. Rep.* 6, 20772
49. Ruess, J. *et al.* (2013) Designing experiments to understand the variability in biochemical reaction networks. *J. R. Soc. Interface* 10, 20130588
50. Kerkhoven, E.J. *et al.* (2013) Handling uncertainty in dynamic models: the pentose phosphate pathway in *Trypanosoma brucei*. *PLoS Comput Biol* 9, e1003371
51. Liebermeister, W. and Klipp, E. (2005) Biochemical networks with uncertain parameters. *IEE P. Syst. Biol.* 152, 97–107
52. Oakley, J.E. and O'Hagan, A. (2004) Probabilistic sensitivity analysis of complex models: a Bayesian approach. *J. Roy. Stat. Soc. Ser. B. (Stat. Method.)* 66, 751–769
53. Bernardo, J.M. and Smith, A.F.M. (2008) *Bayesian Theory,* John Wiley & Sons
54. Calderhead, B. *et al.* (2013) Bayesian approaches for mechanistic ion channel modeling. In *In Silico Systems Biology* (Schneider, V. M., ed.), pp. 247–272, Humana Press
55. Xu, T.-R. *et al.* (2010) Inferring signaling pathway topologies from multiple perturbation measurements of specific biochemical species. *Sci. Signal.* 3, ra20
56. Hastings, W.K. (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57, 97–109
57. Schellenberger, J. and Palsson, B.Ø. (2009) Use of randomized sampling for analysis of metabolic networks. *J. Biol. Chem.* 284, 5457–5461
58. Kramer, A. *et al.* (2014) MCMC_CILB – an advanced MCMC sampling package for ode models. *Bioinformatics* 30, 2991–2992
59. Tran, L.M. *et al.* (2008) Ensemble modeling of metabolic networks. *Biophys. J.* 95, 5606–5617
60. Jia, G. *et al.* (2012) Ensemble kinetic modeling of metabolic networks from dynamic metabolic profiles. *Metabolites* 2, 891
61. Marino, S. *et al.* (2008) A methodology for performing global uncertainty and sensitivity analysis in systems biology. *J. Theor. Biol.* 254, 178–196
62. Wainwright, H.M. *et al.* (2014) Making sense of global sensitivity analyses. *Comput. Geosci.* 65, 84–94
63. wa Maina, C. *et al.* (2014) Inference of RNA polymerase II transcription dynamics from chromatin immunoprecipitation time course data. *PLoS Comput Biol* 10, e1003598
64. Honkela, A. *et al.* (2015) Genome-wide modeling of transcription kinetics reveals patterns of RNA production delays. *Proc. Natl. Acad. Sci. U.S.A.* 112, 13115–13120
65. Bintu, L. *et al.* (2016) Dynamics of epigenetic regulation at the single-cell level. *Science* 351, 720–724
66. Amador, A. *et al.* (2013) Elemental gesture dynamics are encoded by song premotor cortical neurons. *Nature* 495, 59–64
67. Eggo, R.M. *et al.* (2016) Respiratory virus transmission dynamics determine timing of asthma exacerbation peaks: evidence from a population-level model. *Proc. Natl. Acad. Sci. U.S.A.* 113, 2194–2199

68. Buzsaki, G. and Mizuseki, K. (2014) The log-dynamic brain: how skewed distributions affect network operations. *Nat. Rev. Neurosci.* 15, 264–278
69. Probst, D. *et al.* (2015) Probabilistic inference in discrete spaces can be implemented into networks of LIF neurons. *Frontiers in Computational Neuroscience* 9, 13
70. Gunawardena, J. (2014) Models in biology: 'accurate descriptions of our pathetic thinking'. *BMC Biol.* 12, 29
71. Kirk, P.D.W. *et al.* (2015) Systems biology (un)certainties. *Science* 350, 386–388
72. Gustafsson, C. and Vallverdú, J. (2016) The best model of a cat is several cats. *Trends Biotechnol.* 34, 207–213
73. Sunnåker, M. *et al.* (2013) Automatic generation of predictive dynamic models reveals nuclear phosphorylation as the key Msn2 control mechanism. *Sci. Signal.* 6, ra41
74. Limpert, E. *et al.* (2001) Log-normal distributions across the sciences: keys and clues. *Bioscience* 51, 341–352
75. Grönholm, T. and Annila, A. (2007) Natural distribution. *Math. Biosci.* 210, 659–667