



Validation Through Collaboration: Encouraging Team Efforts to Ensure Internal and External Validity of Computational Models of Biochemical Pathways

Richard Fitzpatrick^{1,3} · Melanie I. Stefan^{1,2}

Accepted: 17 March 2022 / Published online: 11 May 2022
© The Author(s) 2022

Abstract

Computational modelling of biochemical reaction pathways is an increasingly important part of neuroscience research. In order to be useful, computational models need to be valid in two senses: First, they need to be consistent with experimental data and able to make testable predictions (external validity). Second, they need to be internally consistent and independently reproducible (internal validity). Here, we discuss both types of validity and provide a brief overview of tools and technologies used to ensure they are met. We also suggest the introduction of new collaborative technologies to ensure model validity: an incentivised experimental database for external validity and reproducibility audits for internal validity. Both rely on FAIR principles and on collaborative science practices.

Keywords Systems biology · Reproducibility · Computational modelling · Model validity

Mathematics Subject Classification 92-08 · 92C20 · 92C42 · 92C45

Introduction

The large number of molecules and interactions underpinning most biological phenomena call for *in silico* approaches to understand biochemical networks (Pollard 2013). This is especially true for neuroscience, where the interpretation of a molecular signalling network can have major implications on translational approaches to diseases and disorders. A good computational model makes testable predictions which can be used to narrow down the number of experimental investigations required to reach an understanding of a given phenomenon (Berro, 2018). With the newer multistate computational models and tools (Bazzazi et al., 2018; Boutillier et al., 2018; Harris et al., 2016; Stefan et al., 2014; Stites et al., 2015; Stefan et al., 2012; Pharris et al., 2019) we

can see the impact of modifying one aspect of a molecule's function on all others, and how that affects the biochemical network as a whole, without needing to construct many different computational models, or run multiple *in vivo* or *in vitro* experiments.

These powerful aspects of modelling have not reached their full potential within neuroscience. This may stem in part from a lack of clarity on how our modelling approaches represent the biological mechanisms they claim to simulate (Berro, 2018; Mogilner et al., 2006) and the soundness of the models themselves. Two major questions can be asked about the validity of a computational model:

1. How can we be sure that the model is representative of *in vivo* states?
2. How do we know the model is reliable?

The first question relates to the external validity of a model (how well the model fits with experimentally knowable data), the second relates to its internal validity (whether the model is soundly and consistently constructed).

This commentary provides two possible pathways to answer these questions about model validity. Both require a greater level of collaboration, both between biochemical

✉ Melanie I. Stefan
melanie.stefan@ed.ac.uk

¹ Centre for Discovery Brain Sciences, University of Edinburgh, Edinburgh, UK

² ZJU-UoE Institute, Zhejiang University, Haining, China

³ School of Biological Sciences, University of Edinburgh, Edinburgh, UK

modellers and between modellers and their experimental counterparts. We believe by fostering such connections models will be better utilised, better parameterised, and embedded more into the driving of neuroscientific inquiry.

External Validity: Comparing a Computational Model to Experimental Data

Computational models of biological systems are important tools: They can synthesise the current state of knowledge about a biological process into a coherent system. Using models, we can explore overall behaviours of a biological system that would be impossible to predict from just examining its component parts (Le Novère, 2015). Using computational models, we can quickly test a large number of possible scenarios. They are therefore especially useful for generating hypotheses about a system, and making testable predictions about its behaviour.

The predictive power of a computational model relies on the model being an accurate (enough) representation of biological reality. Modellers rely on experimental data to construct and constrain the model. Once a model is completed, experiments are needed to validate models, test model predictions, or select from competing models of the same process.

A biochemical modeller wants their model parameters to closely resemble the situation *in vivo*. This requires binding constants and concentrations specific to a specific cell type or functional component such as a dendritic spine. There are databases of biochemical parameters (Glont et al., 2020; Jeske et al., 2019; Sivakumaran et al., 2003; Wittig et al., 2012), but at this point they suffer from incomplete coverage, especially when it comes to data on signalling pathways. For instance, in one of our models of CaMKII activation (Stefan et al., 2012), only 27 % of the model parameters were taken directly from experimental papers, another 13 % came from previous modelling papers, 27 % were derived from measurements found in the literature, and the rest (33 %) has to be estimated in the course of model construction and validation.

Compounding this data scarcity is that much of these experimentally derived sources for reaction constants and concentrations come from decades-old research. This work is often of excellent quality, but does not cover many more recently discovered molecules and interactions. The urgent need for new experimental data for models is being approached in interesting ways, with frameworks such as FindSim (Viswan et al., 2018) encouraging the integration of multiscale models with experimental datasets, and the FAIR initiative improving the extraction of data from published studies to improve discovery, standardisation, and enable

the re-use of this data (Wilkinson et al., 2016). These cannot generate data which is not there though. As our *in vivo* techniques have improved, there has been a major shift in analysing the function of molecules *in situ*. The actual mechanisms that underpin function are often included quite late in how we currently construct and perceive biological theory (Lazebnik, 2002; Kennedy, 2017). This contrasts with biochemical modellers, who are almost always concerned with mechanisms (Chen et al., 2010; van Riel, 2006).

Modellers do have an array of tools to work around this problem. We can run parameter sensitivity analyses to pinpoint the parameters that matter to a reaction network (Zi et al., 2008), and then estimate values that fit with experimental outcomes. We can assess the robustness a reaction network, determining sloppy parameters, whose 'true' value does not matter much to the behaviour of the model (Gutenkunst et al., 2007). Indeed it may be enough to focus experimental efforts on a few parameters that the model is most sensitive to, instead of measuring every single model parameter, which would be both experimentally costly and, for some parameters, unnecessary (Gutenkunst et al., 2007; Transtrum et al., 2015).

And yet, as we move to larger and more complex models, the question of how well these sensitivity and parameter identification analyses scale to larger contexts is still open (Babtie and Stumpf, 2017). Parameterisation techniques for large models have evolved rapidly to reduce the intractable computational load and to accommodate the sparsity of absolute datasets (Schmiester et al., 2020), but they do not eliminate the need to experimentally determine some key parameters.

Once built, a model can be tested against experimental data to establish its validity. A good model can make predictions that go beyond the currently available data, and may even go beyond currently possible experimental techniques - this is indeed quite common in fields such as physics (see for instance Englert & Brout, 1964; Higgs, 1964; Guralnik et al., 1964; ATLAS Collaboration, 2012; CMS Collaboration, 2012). This means that a modeller may predict the outcome of an experiment, but may never be able to conduct the experiment itself, or even see it conducted.

Taken together, experiments can provide useful information for computational modelling, from early on in model development to years after a model has been completed and distributed. For a fast-paced field with rich data such as neuroscience, where molecular understanding is constantly evolving, the ability to test hypotheses quickly and robustly against prior evidence is a valuable asset that modelling affords. There is a risk that a lack of biologically acquired parameters decouples modellers and experimenters. This causes work to be unnecessarily duplicated, or unfeasible avenues of research undertaken that could have been shown to be unwise with a single simulation or a simple *in vivo* test

We suspect all modellers have a “wish list” of experiments that would improve and accelerate their model development, or test model predictions. But not all modellers have access to the necessary infrastructure and skills. Finding experimental collaborators is not always easy: There is not necessarily complete overlap between the experiments that would be informative to a computational modeller and the experiments that interest the experimentalist.

Incentivised Experimental Database

As we have seen, there is a gap between computational models and the experimental data needed to both constrain those models and test their outcomes. This is partly because existing data is not always published and shared, but partly also because some experiments have just never been conducted.

What is needed is a way of incentivising these experiments, to persuade our experimentalist colleagues that there is some benefit to them for carrying them out. We propose one such way is to take some lessons from the past. Could we not present our experimental wishlist, specifying the data we need to complete or check our models, and offer a cash incentive for providing this data?

Offering cash rewards for solving scientific problems is not without precedent. Historically, Challenge Prizes drove major advancements in problems of navigation (e.g. The UK Longitude Act of 1714 [Lon, 1714](#)) and aviation (e.g. the Orteig Prize [Brady, 2002](#)). The Millennium Prize offers \$1 million for the solution of any of seven stated mathematical problems ([Jaffe, 2006](#)). Even more recently, foundations such as Nesta offer considerable sums of money for the solving of defined problems within a range of different fields ([Puttick et al., 2014](#)).

What we are suggesting is not quite financially on the same scale as this but captures some of that spirit; of incentivising innovation to accelerate improvements in biochemical modelling. Our “problem” is that we have limited access to specific biochemical data necessary to accelerate the construction and testing of complex dynamical models of biochemical systems. Our solution is to reward experimenters who “solve” parts of this problem through the provision of this data.

We envisage this working thusly. Modellers submit a wishlist of experiments to a database, with explicit instructions on the biological background, the model, the data needed, and (if available) a suggested experimental design. These experiments are sorted into categories related to difficulty and the experimental methodology required to implement them. The relative difficulty or complexity of the experiment is linked to a cash reward, which not only compensates for the time and resources used but provides extra income for the lab to continue their own research. These “microgrants” would be split

into two components – money up front for the experiment, with the bonus provided upon submission of raw data and documentation following FAIR principles (regardless of the nature of the outcome). The dataset publication would also include authorship and contribution information of all experimental collaborators, as well as a link to the original data request and the model it stemmed from, thus giving credit and facilitating provenance tracking for model parameters.

These ‘arranged’ collaborations may even prove more fruitful in connecting researchers exploring the same phenomena through different approaches. This leads to our second parallel intention for this database: prediction testing.

As initiatives like FindSim ([Viswan et al., 2018](#)) emphasise, models ideally exist in a dynamic cycle with experimental research. Model outcomes produce predictions that are tested experimentally, which provides data to update the model to drive further predictions. This ideal scenario is rarely achieved however, and what we have is a mostly decoupled system where model predictions are not seen by experimental researchers, or only discovered after convergently reaching the same outcome. Our database would encourage modellers to post the major predictions of their models which can then be validated by experimental work. This approach allows for a gradual and visible increase in the utility of modelling alongside experimental work. As models receive higher fidelity parameter sets, the predictions made will have more weight and power to guide real conceptual breakthroughs.

This provides another use of the modeller “wish list”, supplying experimenters with readily available predictions and a clearly defined direction for potentially fruitful future research. The results of these investigations are importantly just as valuable if they contradict the model as when they agree, leading in each case to publishable findings and model enhancement.

Thus, the incentivised experiment database provides a mechanism for long-term mutually beneficial cycles of models and experiments to arrive at a deeper understanding of biological questions. Importantly, the cycle does not involve fixed teams of researchers. At any point, another experimentalist can claim an open experiment and contribute their data. And a modeller can pick up and refine an existing model. This brings us to the second issue around model validity: The “internal validity” of a computational model, i.e. its ability to be reproduced.

Internal Validity: Ensuring Reproducibility of Computational Model

The importance of reproducible research has received much attention in recent years across all areas of science ([Baker, 2016](#)), including computational modelling of biological

systems (Mendes, 2018). Reproducibility is an important condition for model sharing and reuse (Cucurull-Sanchez et al., 2019; Scharm et al., 2018).

Much work has been done on how to ensure the reproducibility of computational work. Standards for reproducible computational research have been formulated both as stand-alone guidelines (Sandve et al., 2013; Elofsson et al., 2019), and within the FAIR framework (Wilkinson et al., 2016).

Community efforts to ensure reproducible modelling of biochemical reaction systems include efforts to standardise model specification (e.g. Hucka et al., 2003; Zhang et al., 2020; Hedley et al., 2001; Le Novère et al., 2009; Touré et al., 2020; Schreiber et al., 2020; Waltemath et al., 2020), model databases (Glont et al., 2018; Malik-Sheriff et al., 2019), and standards for model annotation and documentation (Waltemath et al., 2011; Bergmann et al., 2014; Waltemath et al., 2020).

There is now also a journal specifically designed for replication studies of previous computational work (Hinsen & Rougier, 2019).

There are, however, persistent problems with model reproducibility. Not all modellers use available standards and share their code. Even for those who attempt to, there are often additional assumptions, e.g. about simulation parameters, model interfacing, or model data analysis that are not made explicit and that hinder future reproducibility (Waltemath et al., 2020).

To carefully annotate and document a model and ensure its reproducibility involves time and work. There is as yet little incentive or reward to doing this - the benefits of a model being reproducible often become clear years after it is first published, and scientific career paths are not currently structured to invite this level of foresight.

In theory, pre-publication peer review could pick up problems around reproducibility, but peer reviewers do not always have access to the code, computational resources, and time it would take to reproduce a model that they are reviewing. Post-publication review or replication studies (Hinsen & Rougier, 2019) are valuable, but may come too late to salvage the original model.

There is thus room for a new robust process to ascertain model reproducibility pre-publication, and even pre-submission, so that any reproducibility gaps can be caught and addressed early.

Reproducibility Audits

In order to increase model reproducibility, we suggest the introduction of pre-publication reproducibility audits for modelling projects. This means that each project should involve a reproducibility auditor, whose role is to ensure the model is reproducible before it is published.

A reproducibility auditor would be a person familiar with the biological framework and modelling methodologies used, but who was not involved in the original model development. They would ideally not be part of the same research group as the original model developers, so that they do not share the workflows and implicit assumptions prevalent in that group.

When a computational model is ready for publication, the model developers send the model and write-up to their reproducibility auditor. The auditor attempts to run the model and reproduce the figures in the paper based on the information given to them, and identifies gaps in reproducibility. Both parties then work together to improve the documentation and ensure model reproducibility.

Once this is achieved, they submit the manuscript describing the model for publication together, with a short report of the reproducibility audit (steps taken, results obtained) in the appendix, and an author contribution statement specifying the role of the auditor.

The benefit for modellers is an external confirmation of reproducibility prior to publishing the model: possible gaps in documentation can thus be caught and fixed early.

For the field as such, the benefit is that there is an extra quality control step pre-publication. Journals could highlight this by introducing a “reproducibility audited” badge, similar to existing open science badges (Kidwell et al., 2016).

For the reproducibility auditor, the benefit is an opportunity to establish a collaboration with another research group in their field and learn first-hand about their modelling methods and process. This could be especially useful for early-career scientists just starting in a particular field. A simple reproducibility audit could even be an exercise assigned in a computational biology, biomedical informatics, or neuroscience class.

Standardisation efforts would benefit from reproducibility audits in two ways: First, reproducibility audits provide a good incentive for adhering to standards and good practices, thereby popularising the standards. Second, reproducibility audits will generate feedback on both the usefulness and usability of the standards used, and can thus feed back into standards development.

How will modellers find their reproducibility auditors? This could be done fairly informally within existing collaborative network. Some institutions or consortia may also create the role of a reproducibility auditor or make reproducibility audits part of the role of their research integrity advisors (Winchester, 2018).

For researchers not having access to these channels, there could also be a centralised online forum where people can post a short description of the project, techniques used, skills expected of auditor, expected workload associated with auditing and an indicative time frame. This may provide especially valuable opportunities for scientists who

may not have access to mainstream scientific networks to act as auditors and thereby gain experience and establish collaborative ties.

Turning Recommendations into Best Practice

How can both initiatives be incentivised, monitored and validated?

If a scientist invests time and resources in contributing to the experiment database, or serve as a reproducibility auditor, what is in it for them? We see several possibilities. The “microgrant” model of (small) cash incentives for solving particular problems is not entirely new. In data science, the Kaggle platform (<https://www.kaggle.com/>) challenges users to analyse data sets, sometimes (but not always) in competition for cash prizes. Over the last years, a wealth of interesting research papers have come out of Kaggle challenges. The potential for a similar crowd-sourcing approach to biomedical challenges has been recognised previously (Saez-Rodriguez et al., 2016). Additional incentives could be provided by opportunities for shared authorship and the establishment of new collaborations.

In the same way that Kaggle has been successfully used as a learning resource (Serrano et al., 2018), this is also a possibility here: Contributions to the incentivised experimental database or reproducibility audits could be a learning experience, for instance within the framework of an undergraduate assignment or Honours project. It could also be an opportunity for on-the-job learning for PhD students or postdoctoral researchers new to a particular field.

Our ideas fit within the bigger context of biocuration, in which datasets are structured, curated and annotated such that they adhere to FAIR-TLC principles (Howe et al., 2008; International Society for Biocuration, 2018). This movement has over the past decade or so striven to reframe data as an asset, that requires quality control and trust in those carrying this out (International Society for Biocuration, 2018; Gabrielsen, 2020).

Modellers arguably conduct biocuration in the course of the construction of their models. Parameters are chosen based on an expert assessment of the experiment that produced them, the use of these data are clearly defined, and there are standardised naming conventions for annotating model parts (Le Novère et al., 2005).

Our reproducibility audits therefore fall under the quality control aspect of curation, and can take much from recent guidance in this area (Tang et al., 2019). Here again, the idea of using curation as a teaching tool has already been brought up: Undergraduates have been shown to be just as capable at biocuration as experts after training (Mitchell et al., 2015), and from our own experience quickly acquire the proficiency

required to critically evaluate model data inputs and outputs. This, if combined with the experimental database idea, could harness student expertise in driving model validation, expose potential researchers early in their career to FAIR principles, and introduce modelling methods as a way of structuring existing data into valuable and usable formats.

Ultimately, the ideas laid out here have to be tested and evaluated. If an incentivised experimental database with an attached microgrant scheme were to be implemented, it would make sense to monitor not only application and success rates, but also completion of projects, data submission, and subsequent use, for instance in models and follow-up publications.

The use of reproducibility audits could easily be tracked if journals were to introduce a “reproducibility audited” badge. This would also allow an analysis of the impact (e.g. model downloads from repositories or paper citations) of audited vs non-audited models.

Conclusions

We show here that the fostering of collaborative practices has potential for improving the validity of biochemical models, both external and internal. These collaborations are designed to be mutually beneficial, bringing researchers in similar fields closer together whilst also addressing the challenge of model validity. Furthermore, they have the potential to both accelerate model development and increase the number of biologically-derived parameters available to all researchers. The specifics of how to exactly implement these ideas should also be collaboratively decided. By setting out a possible framework we now invite readers to discuss and debate how we can best turn these ideas into reality.

Information Sharing Statement

This article is a think piece, for which no data was produced or analyzed. Readers who nonetheless want to access the data can do so here: <http://tinyurl.com/nein2022data>.

Acknowledgements We thank the anonymous reviewers for their constructive comments.

Declarations

Conflicts of Interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source,

provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- (1714). The discovery of longitude at sea act 1714 (13 Anne, c. 14). Royal Greenwich Observatory Archives RGO 14/1.
- ATLAS Collaboration. (2012). Observation of a new particle in the search for the standard model higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1), 1–29. <https://doi.org/10.1016/j.physletb.2012.08.020>
- Babtie, A. C., & Stumpf, M. P. (2017). How to deal with parameters for whole-cell modelling. *Journal of The Royal Society Interface*, 14(133), 20170237.
- Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, 533, 452–454. <https://doi.org/10.1038/533452a>
- Bazzazi, H., Zhang, Y., Jafarnejad, M., & Popel, A. S. (2018). Computational modeling of synergistic interaction between $\alpha\beta3$ integrin and vegfr2 in endothelial cells: Implications for the mechanism of action of angiogenesis-modulating integrin-binding peptides. *Journal of theoretical biology*, 455, 212–221.
- Bergmann, F. T., Adams, R., Moodie, S., Cooper, J., Glont, M., Golebiewski, M., Hucka, M., Laibe, C., Miller, A. K., Nickerson, D. P. et al. (2014). Combine archive and omex format: one file to share all information to reproduce a modeling project. *BMC bioinformatics*, 15(1), 369.
- Berro, J. (2018). “essentially, all models are wrong, but some are useful” - a cross-disciplinary agenda for building useful models in cell biology and biophysics. *Biophysical reviews*, 10(6), 1637–1647.
- Boutillier, P., Maasha, M., Li, X., Medina-Abarca, H. F., Krivine, J., Feret, J., Cristescu, I., Forbes, A. G., & Fontana, W. (2018). The kappa platform for rule-based modeling. *Bioinformatics*, 34(13), i583–i592.
- Brady, T. (2002). The orteg prize. *Journal of Aviation/Aerospace Education & Research*, 12(1), 9.
- Chen, W. W., Niepel, M., & Sorger, P. K. (2010). Classic and contemporary approaches to modeling biochemical reactions. *Genes & development*, 24(17), 1861–1875.
- CMS Collaboration. (2012). Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1), 30–61. <https://doi.org/10.1016/j.physletb.2012.08.021>
- Cucurull-Sanchez, L., Chappell, M. J., Chelliah, V., Cheung, S. Y. A., Derks, G., Penney, M., Phipps, A., Malik-Sheriff, R. S., Timmis, J., Tindall, M. J., Graaf, P. H., Vicini, P., & Yates, J. W. T. (2019). Best practices to maximize the use and reuse of quantitative and systems pharmacology models: Recommendations from the united kingdom quantitative and systems pharmacology network. *CPT: Pharmacometrics & Systems Pharmacology*, 8(5), 259–272. <https://doi.org/10.1002/psp4.12381>
- Elofsson, A., Hess, B., Lindahl, E., Onufriev, A., van der Spoel, D., & Wallqvist, A. (2019). Ten simple rules on how to create open access and reproducible molecular simulations of biological systems. *PLoS computational biology*, 15, e1006649. <https://doi.org/10.1371/journal.pcbi.1006649>
- Englert, F., & Brout, R. (1964). Broken symmetry and the mass of gauge vector mesons. *Physical Review Letters*, 13(9), 321–323. <https://doi.org/10.1103/physrevlett.13.321>
- Gabrielsen, A. M. (2020). Openness and trust in data-intensive science: the case of biocuration. *Medicine, Health Care and Philosophy*, 23(3), 497–504.
- Glont, M., Arankalle, C., Tiwari, K., Nguyen, T. V., Hermjakob, H., & Malik-Sheriff, R. S. (2020). Biomodels parameters: a treasure trove of parameter values from published systems biology models. *Bioinformatics*.
- Glont, M., Nguyen, T. V. N., Graesslin, M., Hälke, R., Ali, R., Schramm, J., Wimalaratne, S. M., Kothamachu, V. B., Rodriguez, N., Swat, M. J., Eils, J., Eils, R., Laibe, C., Malik-Sheriff, R. S., Chelliah, V., Le Novère, N., & Hermjakob, H. (2018). Biomodels: expanding horizons to include more modelling approaches and formats. *Nucleic Acids Research*, 46(D1), D1248–D1253. <https://doi.org/10.1093/nar/gkx1023>
- Guralnik, G. S., Hagen, C. R., & Kibble, T. W. B. (1964). Global conservation laws and massless particles. *Physical Review Letters*, 13(20), 585–587.
- Gutenkunst, R. N., Waterfall, J. J., Casey, F. P., Brown, K. S., Myers, C. R., & Sethna, J. P. (2007). Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput Biol*, 3(10), e189.
- Harris, L. A., Hogg, J. S., Tapia, J.-J., Sekar, J. A., Gupta, S., Korsunsky, I., Arora, A., Barua, D., Sheehan, R. P., & Faeder, J. R. (2016). Bionetgen 2.2: advances in rule-based modeling. *Bioinformatics*, 32(21), 3366–3368.
- Hedley, W. J., Nelson, M. R., Bellivant, D. P., & Nielsen, P. F. (2001). A short introduction to CellML. Philosophical Transactions of the Royal Society of London. *Series A: Mathematical, Physical and Engineering Sciences*, 359(1783), 1073–1089. <https://doi.org/10.1098/rsta.2001.0817>
- Higgs, P. W. (1964). Broken symmetries and the masses of gauge bosons. *Physical Review Letters*, 13(16), 508–509. <https://doi.org/10.1103/physrevlett.13.508>
- Hinsen, K., & Rougier, N. P. (2019). Rescience (r)evolution. <https://doi.org/10.5281/ZENODO.3069619>
- Howe, D., Costanzo, M., Fey, P., Gojobori, T., Hannick, L., Hide, W., Hill, D. P., Kania, R., Schaeffer, M., St Pierre, S. et al. (2008). The future of biocuration. *Nature*, 455(7209), 47–50.
- Hucka, M., Finney, A., Sauro, H. M., Bolouri, H., Doyle, J. C., Kitano, H., Doyle, J., Arkin, A. P., Bornstein, B. J., Bray, D., Cornish-Bowden, A., Cuellar, A., Dronov, S., Gilles, E. D., Ginkel, M., Gor, V., Goryanin, I. I., Hedley, W. J., Hodgman, T. C., Hofmeyr, J.-H., Hunter, P. J., Juty, N. S., Kasberger, J., Kremling, A., Kummer, U., Le Novère, N., Loew, L. M., Lucio, D., Mendes, P., Minch, E., Mjolsness, E. D., Nakayama, Y., Nelson, M. R., F.Nielsen, P., Sakurada, T., Schaff, J. C., Shapiro, B. E., Shimizu, T. S., D.Spence, H., Stelling, J., Takahashi, K., Tomita, M., Wagner, J., & Wang, J. (2003). The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics*, 19(4), 524–531.
- International Society for Biocuration (2018). Biocuration: Distilling data into knowledge. *PLoS Biology*, 16(4), e2002846. <https://doi.org/10.1371/journal.pbio.2002846>
- Jaffe, A. M. (2006). The millennium grand challenge in mathematics. *Notices of the AMS*, 53(6).
- Jeske, L., Placzek, S., Schomburg, I., Chang, A., & Schomburg, D. (2019). Brenda in 2019: a european elixir core data resource. *Nucleic acids research*, 47(D1), D542–D549.
- Kennedy, M. B. (2017). Biochemistry and neuroscience: the twain need to meet. <https://doi.org/10.1016/j.conb.2017.01.004>, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5447485/>
- Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T. M., Fiedler, S., & Nosek, B. A. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency.

- PLoS biology, 14, e1002456. <https://doi.org/10.1371/journal.pbio.1002456>
- Lazebnik, Y. (2002). Can a biologist fix a radio?—or, what I learned while studying apoptosis. *Cancer cell*, 2(3), 179–182.
- Le Novère, N., Finney, A., Hucka, M., Bhalla, U. S., Campagne, F., Collado-Vides, J., Crampin, E. J., Halstead, M., Klipp, E., Mendes, P. et al. (2005). Minimum information requested in the annotation of biochemical models (miriam). *Nature biotechnology*, 23(12), 1509–1515.
- Le Novère, N., Hucka, M., Mi, H., Moodie, S., Schreiber, F., Sorokin, A., Demir, E., Wegner, K., Aladjem, M. I., Wimalaratne, S. M., Bergman, F. T., Gauges, R., Ghazal, P., Kawaji, H., Li, L., Matsuoka, Y., Villéger, A., Boyd, S. E., Calzone, L., Courtot, M., Dogrusoz, U., Freeman, T. C., Funahashi, A., Ghosh, S., Jouraku, A., Kim, S., Kolpakov, F., Luna, A., Sahle, S., Schmidt, E., Watterson, S., Wu, G., Goryanin, I., Kell, D. B., Sander, C., Sauro, H., Snoep, J. L., Kohn, K., & Kitano, H. (2009). The systems biology graphical notation. *Nat Biotechnol*, 27(8), 735–741. <https://doi.org/10.1038/nbt.1558>
- Le Novère, N. (2015). Quantitative and logic modelling of molecular and gene networks. *Nature reviews. Genetics*, 16, 146–158. <https://doi.org/10.1038/nrg3885>
- Malik-Sheriff, R. S., Glont, M., Nguyen, T. V. N., Tiwari, K., Roberts, M. G., Xavier, A., Vu, M. T., Men, J., Maire, M., Kananathan, S., Fairbanks, E. L., Meyer, J. P., Arankalle, C., Varusai, T. M., Knight-Schrijver, V., Li, L., Dueñas-Roca, C., Dass, G., Keating, S. M., Park, Y. M., Buso, N., Rodriguez, N., Hucka, M., & Hermjakob, H. (2019). BioModels—15 years of sharing computational models in life science. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkz1055>
- Mendes, P. (2018). Reproducible research using biomodels. *Bulletin of mathematical biology*, 80, 3081–3087. <https://doi.org/10.1007/s11538-018-0498-z>
- Mitchell, C. S., Cates, A., Kim, R. B., & Hollinger, S. K. (2015). Undergraduate biocuration: developing tomorrow's researchers while mining today's data. *Journal of Undergraduate Neuroscience Education*, 14(1), A56.
- Mogilner, A., Wollman, R., & Marshall, W. F. (2006). Quantitative modeling in cell biology: what is it good for? *Developmental cell*, 11(3), 279–287.
- Pharris, M. C., Patel, N. M., VanDyk, T. G., Bartol, T. M., Sejnowski, T. J., Kennedy, M. B., Stefan, M. I., & Kinzer-Ursem, T. L. (2019). A multi-state model of the camkii dodecamer suggests a role for calmodulin in maintenance of autophosphorylation. *PLoS computational biology*, 15(12), e1006941.
- Pollard, T. D. (2013). No question about exciting questions in cell biology. *PLoS Biol*, 11(23), e1001734.
- Puttick, R., Baeck, P., & Colligan, P. (2014). The teams and funds making innovation happen in governments around the world. *Londres: Nesta & Bloomberg Philanthropies*.
- van Riel, N. A. (2006). Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments. *Briefings in bioinformatics*, 7(4), 364–374.
- Saez-Rodriguez, J., Costello, J. C., Friend, S. H., Kellen, M. R., Mangravite, L., Meyer, P., Norman, T., & Stolovitzky, G. (2016). Crowdsourcing biomedical research: leveraging communities as innovation engines. *Nature Reviews Genetics*, 17(8), 470–486.
- Sandve, G. K., Nekrutenko, A., Taylor, J., & Hovig, E. (2013). Ten simple rules for reproducible computational research. *PLoS computational biology*, 9, e1003285. <https://doi.org/10.1371/journal.pcbi.1003285>
- Scharm, M., Gebhardt, T., Touré, V., Bagnacani, A., Salehzadeh-Yazdi, A., Wolkenhauer, O., & Waltemath, D. (2018). Evolution of computational models in BioModels database and the physiome model repository. *BMC Systems Biology*, 12(1). <https://doi.org/10.1186/s12918-018-0553-2>
- Schmiester, L., Schälte, Y., Fröhlich, F., Hasenauer, J., & Weindl, D. (2020). Efficient parameterization of large-scale dynamic models based on relative measurements. *Bioinformatics*, 36(2), 594–602.
- Schreiber, F., Sommer, B., Czauderna, T., Golebiewski, M., Gorochowski, T. E., Hucka, M., Keating, S. M., König, M., Myers, C., Nickerson, D., & Waltemath, D. (2020). Specifications of standards in systems and synthetic biology: status and developments in 2020. *Journal of Integrative Bioinformatics*, 17(2-3). <https://doi.org/10.1515/jib-2020-0022>
- Serrano, E., Molina, M., Manrique, D., & Bajo, J. (2018). Challenge-based learning in computational biology and data science. In *ICTERI Workshops* (pp. 725–733).
- Sivakumaran, S., Hariharaputran, S., Mishra, J., & Bhalla, U. S. (2003). The database of quantitative cellular signaling: management and analysis of chemical kinetic models of signaling networks. *Bioinformatics*, 19(3), 408–415.
- Stefan, M. I., Bartol, T. M., Sejnowski, T. J., & Kennedy, M. B. (2014). Multi-state modeling of biomolecules. *PLoS Comput Biol*, 10(9), e1003844.
- Stefan, M. I., Marshall, D. P., & Le Novère, N. (2012). Structural analysis and stochastic modelling suggest a mechanism for calmodulin trapping by camkii. *PLoS One*, 7(12), e29406.
- Stites, E. C., Aziz, M., Creamer, M. S., Von Hoff, D. D., Posner, R. G., & Hlavacek, W. S. (2015). Use of mechanistic models to integrate and analyze multiple proteomic datasets. *Biophysical journal*, 108(7), 1819–1829.
- Tang, Y. A., Pichler, K., Füllgrabe, A., Lomax, J., Malone, J., Munoz-Torres, M. C., Vasant, D. V., Williams, E., & Haendel, M. (2019). Ten quick tips for biocuration. *PLOS Computational Biology*, 15(5), e1006906. <https://doi.org/10.1371/journal.pcbi.1006906>
- Touré, V., Dräger, A., Luna, A., Dogrusoz, U., & Rougny, A. (2020). The systems biology graphical notation: Current status and applications in systems medicine. In *Reference Module in Biomedical Sciences*. Elsevier. <https://doi.org/10.1016/b978-0-12-801238-3.11515-6>
- Transtrum, M. K., Machta, B. B., Brown, K. S., Daniels, B. C., Myers, C. R., & Sethna, J. P. (2015). Perspective: Slowness and emergent theories in physics, biology, and beyond. *The Journal of chemical physics*, 143(1), 07B201_1.
- Viswan, N. A., HarshaRani, G. V., Stefan, M. I., & Bhalla, U. S. (2018). Findsim: A framework for integrating neuronal data and signaling models. *Frontiers in neuroinformatics*, 12, 38.
- Waltemath, D., Adams, R., Bergmann, F. T., Hucka, M., Kolpakov, F., Miller, A. K., Moraru, I. I., Nickerson, D., Sahle, S., Snoep, J. L., & Le Novère, N. (2011). Reproducible computational biology experiments with sed-ml—the simulation experiment description markup language. *BMC systems biology*, 5, 198. <https://doi.org/10.1186/1752-0509-5-198>
- Waltemath, D., Golebiewski, M., Blinov, M. L., Gleeson, P., Hermjakob, H., Hucka, M., Inau, E. T., Keating, S. M., König, M., Krebs, O., Malik-Sheriff, R. S., Nickerson, D., Oberortner, E., Sauro, H. M., Schreiber, F., Smith, L., Stefan, M. I., Wittig, U., & Myers, C. J. (2020). The first 10 years of the international coordination network for standards in systems and synthetic biology (COMBINE). *Journal of Integrative Bioinformatics*, 17(2-3). <https://doi.org/10.1515/jib-2020-0005>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., &

- Mons, B. (2016). The fair guiding principles for scientific data management and stewardship. *Scientific data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>
- Winchester, C. (2018). Give every paper a read for reproducibility. *Nature*, 557, 281. <https://doi.org/10.1038/d41586-018-05140-x>
- Wittig, U., Kania, R., Golebiewski, M., Rey, M., Shi, L., Jong, L., Algae, E., Weidemann, A., Sauer-Danzwith, H., Mir, S. et al. (2012). Sabio-rk-database for biochemical reaction kinetics. *Nucleic acids research*, 40(D1), D790–D796.
- Zhang, F., Smith, L. P., Blinov, M. L., Faeder, J., Hlavacek, W. S., Tapia, J. J., Keating, S. M., Rodriguez, N., Dräger, A., Harris, L. A., Finney, A., Hu, B., Hucka, M., & Meier-Schellersheim, M. (2020). Systems biology markup language (SBML) level 3 package: multistate, multicomponent and multicompartment species, version 1, release 2. *Journal of Integrative Bioinformatics*, 1(ahead-of-print).
- Zi, Z., Zheng, Y., Rundell, A. E., & Klipp, E. (2008). Sbml-sat: a systems biology markup language (sbml) based sensitivity analysis tool. *BMC bioinformatics*, 9, 342. <https://doi.org/10.1186/1471-2105-9-342>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.