Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Systems Biology

Dynamic models for metabolomics data integration

Polina Lakrisenko[1,2] and Daniel Weindl[1]

Abstract

As metabolomics datasets are becoming larger and more complex, there is an increasing need for model-based data integration and analysis to optimally leverage these data. Dynamic models of metabolism allow for the integration of heterogeneous data and the analysis of dynamic phenotypes. Here, we review recent efforts in using dynamic metabolic models for data integration, focusing on approaches based on ordinary differential equations that are applicable to both time-resolved and steady-state measurements and that do not require flux distributions as inputs. Furthermore, we discuss recent advances and current challenges. We conclude that much progress has been made in various areas, such as the development of scalable simulation tools, and although challenges remain, dynamic modeling is a powerful tool for metabolomics data analysis that is not yet living up to its full potential.

Addresses

[1]Institute of Computational Biology, Helmholtz Zentrum München GmbH German Research Center for Environmental Health, Neuherberg, 85764, Germany

[2] Center for Mathematics, Technische Universität München, Garching, 85748, Germany

Corresponding author: Weindl, Daniel (daniel.weindl@helmholtz-muenchen.de)

Current Opinion in Systems Biology 2021, **28:**100358

This review comes from a themed issue on **Big Data Acquisition & Analysis**

Edited by **Julio Banga** and **Jan Hasenauer**

For a complete overview see the Issue and the Editorial

Available online 20 July 2021

https://doi.org/10.1016/j.coisb.2021.100358

Keywords

Metabolic modeling, Kinetic modeling, Mechanistic modeling, Data integration, Dynamic model, Metabolomics.

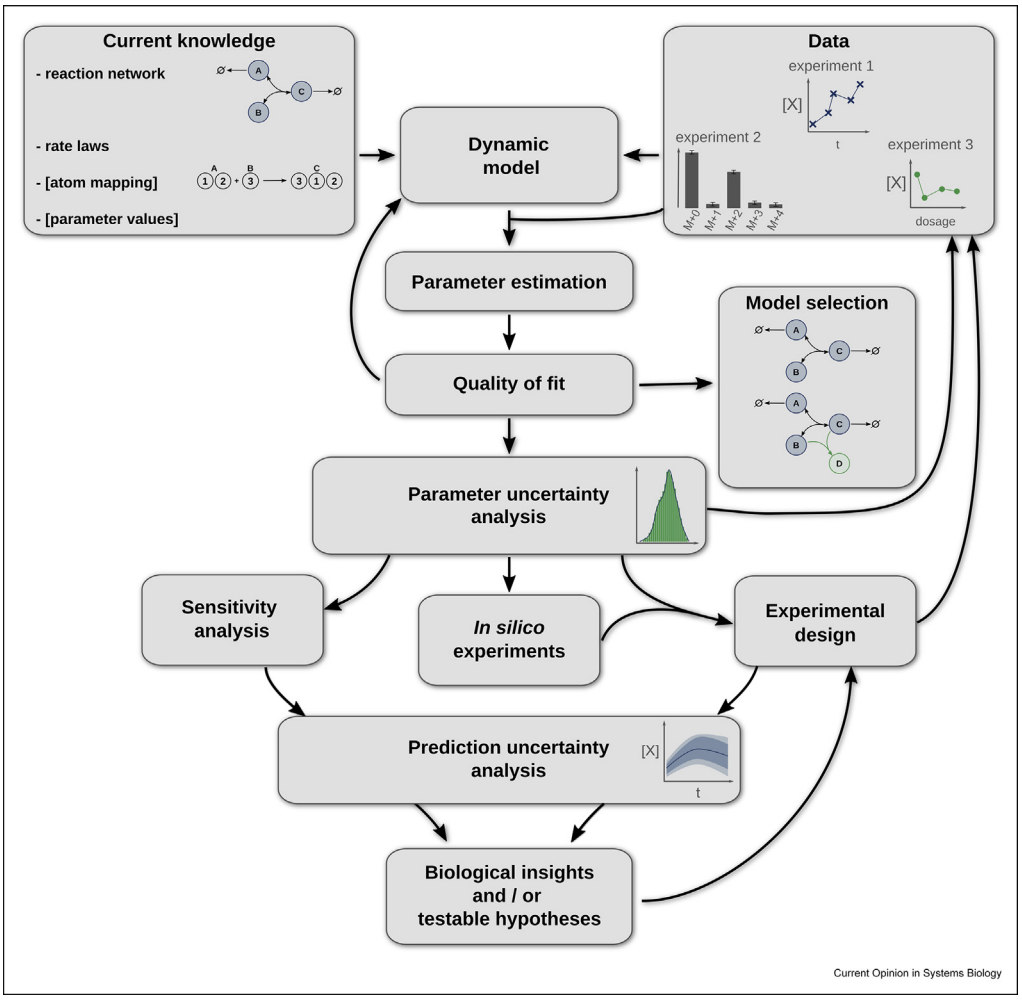
Introduction

Metabolism is a key determinant of cellular behavior, and metabolomics approaches are being applied in a wide range of domains [1]. Nowadays, metabolomics experiments yield rich datasets, which pose new challenges for their analysis. Current metabolomics assays provide information on hundreds to thousands of metabolites, and large numbers of samples can be measured within reasonable times [2]. Since the holistic interpretation of the resulting datasets using 'mental models' becomes impossible, computational models are required and increasingly used for data integration and interpretation. However, observations from individual experiments are still commonly analyzed independently and in a qualitative manner instead of being integrated into an overarching formal quantitative model.

Model-based data analysis aims at analyzing experimental observations in the light of the current understanding of the observed system, as encoded in the model. Experimental observations are either used as model inputs or to infer model parameters. The behavior of the parameterized model is subsequently analyzed to obtain biological insights. Here, we focus mainly on the case of inferring model parameters from experimental data. Typical steps comprise (1) data acquisition, (2) model construction, (3) parameter inference, (4) uncertainty analysis, and (5) analyzing model fit and model-derived predictions (Figure 1).

A variety of metabolic modeling frameworks have emerged, which allow for the integration of varying types of data, each with its specific advantages and disadvantages. A broad overview of these existing metabolic modeling approaches is provided in recent reviews [3–5]. In this article, we focus on metabolomics data integration using dynamic models. These models, usually specified in the form of ordinary differential equations (ODEs), are particularly appealing because: (1) they allow for both the integration of various types of metabolomics data and other types of data; (2) they allow for the analysis of inherently dynamic phenotypes; and (3) they can provide quantitative dynamic information on latent quantities, such as metabolic fluxes or compartment-resolved metabolite levels. On the downside, dynamic models come with a larger number of *a priori* unknown parameters and are computationally more challenging, and therefore, are less scalable than other modeling approaches. When only considering steady-state measurements, these issues can be circumvented to a large extent, as exemplified by the recently developed K-FIT algorithm [6]. However, one is often interested in transient behavior or a case where a steady-state cannot be attained. Therefore, although methods tailored to steady-state measurements have provided

Figure 1



Typical steps for data integration using dynamic models. A model is developed based on current knowledge (brackets indicate optional components). Experimental observations of various types are used as model inputs or to estimate model parameters. If parameters are estimated, the model fit and parameter intervals are assessed. Competing hypotheses can be encoded in different candidate models to select the most plausible one given the data. The parameterized model is used in various downstream analyses to derive biological insights. Examples include sensitivity analysis or metabolic control analysis to determine which parameters have the highest impact on reaction fluxes or metabolite concentrations. *In silico* experiments can be performed to derive testable hypotheses and to design validation experiments. Considering uncertainties is crucial in all analyses to derive meaningful conclusions.

great insights in many applications, they are not discussed further here. Instead, we will focus on ODE-based approaches that are applicable to both time-resolved and steady-state measurements and that do not require a known flux distribution or a biochemical objective function as input. For a broader overview of existing approaches, see, for example, references [5,7,8].

In the following, we give a brief overview of different types of metabolomics data, review recent examples of metabolomics data integration and analysis using dynamic models, and discuss major challenges, as well as recent advances related to model construction, determining model parameters, and dealing with uncertainty.

Metabolomics data

Metabolomics measurements are usually performed using mass spectrometry, nuclear magnetic resonance, or enzymatic assays, with mass spectrometry being the most widespread. Metabolomics datasets can be very rich and diverse, depending on the analytical platforms and experimental protocols. There are the classical label-free metabolomics approaches, which assess metabolite levels, and there are stable-isotope-assisted approaches, which determine isotopic enrichment after application of stable-isotope-labeled tracers [1]. Recent mass spectrometry methods can quantify isotopic enrichment of more than 100 metabolites in a single run [9]. Various types of samples are analyzed, such as complete cell populations and extracellular media, and

in some cases, compartment-specific metabolite pools are accessible [10]. Single-cell metabolomics assays have also emerged [11,12]. Measurements can be time-resolved or for single timepoints, quantitative or semi-quantitative. Due to this heterogeneity, various metabolic modeling approaches have been developed that are, more or less, tailored or restricted to specific types of data [3]. However, in many studies, a combination of different types of data is acquired, and integration of all these data is possible with dynamic models [13,14].

Applications

Dynamic metabolic models have been applied in various contexts. Recurrently pursued objectives are (1) understanding dynamic processes [15,16], for example, through comparison of competing models [17]; (2) inferring control mechanisms and rate-limiting steps [13,18]; and (3) leveraging those to push some system of interest in specific directions, for example, for strain optimization or drug target identification) [14,19,20].

Berndt et al. [16] developed a comparably large dynamic model of liver metabolism to, among others, study the response to various perturbations. The model describes the regulation of enzyme activities by allosteric effectors, hormone-dependent reversible phosphorylation, and variable protein abundances. A subset of parameters was estimated from measurements of intracellular and extracellular metabolite levels under 25 different experimental conditions. Carter et al. [17] performed model selection among four candidate models, calibrated on time-resolved measurements of metabolite levels, to infer the most likely inhibitory mechanism in a drug–drug interaction. Feldman-Salit et al. [18] trained a dynamic model of sulfur assimilation in *Arabidopsis thaliana* on steady-state metabolite measurements, which allowed them to infer dynamic control patterns. Millard et al. [13] analyzed time-resolved measurements of biomass, metabolite levels, and isotopic enrichment using a coarse-grained dynamic model that links glucose uptake to acetate metabolism and growth. Their analysis improved the understanding of control and regulation of acetate overflow and suggested the existence of a yet unknown regulatory program. Marín-Hernández et al. [19] applied a dynamic model of glycolysis, the pentose phosphate pathway, and glycogen metabolism of cancer and noncancer cells, with parameters inferred from various enzyme assays, to derive suggestions for new therapeutic targets. Ou et al. [20] applied a dynamic model, which integrated experimental proteomics, metabolomics, and fermentation kinetics data, to identify key regulators and guide metabolic engineering of *n*-butanol biosynthesis. Their comparison of the results from the dynamic model to those derived from a static model indicated higher accuracy of the former. Ramos et al. [14] integrated both extracellular and intracellular time-resolved metabolite measurements from different experimental settings using a dynamic model comprising 33 state variables and describing central carbon metabolism and cell growth. They identified different physiological states and performed *in silico* experiments to demonstrate how the use of their model can improve cell line engineering and medium design. Moon et al. [21] used a dynamic model to predict the dynamics of mitochondrial NADPH concentration and NADPH/$NADP^+$ ratio in response to oxidative stress. Parameters of the model were estimated from time-resolved NADPH measurements. Lövfors et al. [15] developed a model describing the hormonal regulation of lipolysis. The model, comprising 15 state variables, was calibrated to time-resolved *in vivo* and *in vitro* measurements of metabolite and phosphoprotein abundance after different stimulations. Horvath et al. [22] calibrated an exceptionally comprehensive model of *Escherichia coli* cell-free protein synthesis, comprising 148 metabolites, based on time-resolved metabolite and protein product measurements. An ensemble model was subsequently used to identify the pathways with a strong influence on product yield. Yilmaz et al. [23] developed a coarse-grained model of protein synthesis in Chinese hamster ovary cells based on elementary flux modes. The model was calibrated to time-resolved measurements of cell density and levels of extracellular metabolites and protein products.

These application examples demonstrate the potential of dynamic models to answer diverse research questions by integrating and jointly analyzing various types of data. However, many current applications of dynamic models that integrate time-resolved measurements only cover small parts of the metabolic network or use very coarse-grained representations. This may be due to various challenges discussed in the following sections.

Constructing kinetic models

Depending on the available data and research question, the model scope has to be defined, and a model needs to be constructed. Model construction includes choosing and specifying (1) the network topology and (2) kinetic rate laws.

Genome-scale metabolic reconstructions are a valuable resource for deriving the model topology. Such knowledge bases that describe the metabolic capabilities of an organism have been created for many species through automated reconstruction or extensive manual curation through community efforts [24]. For example, the recently published Human1 [25] unifies two lineages of human genome-scale metabolic reconstructions and provides a good example of open and transparent curation. Although such reconstructions are continuously improved, structural uncertainty remains, for example, due to enzymes of unknown function, enzyme

promiscuity, or unclear subcellular localization of enzymes. For most data-integration efforts, genome-scale dynamic models are still intractable. Therefore, integrating measurements from very different parts of the metabolic network, as commonly obtained from nontargeted metabolomics experiments, is still difficult. Automated algorithms have been devised to reduce the complexity of genome-scale metabolic reconstructions [26] or even generating targeted kinetic models [27] that are more tractable. Such algorithms can significantly speed up model development. However, care has to be taken that the (over-)reduced model complexity does not skew analysis results [28].

If stable-isotope-labeling data are to be analyzed, the model not only needs to account for the pools and reactions of unlabeled metabolites but also for those of the additionally occurring isotopic isomers (isotopologues). Deriving such isotopologue reaction networks requires knowledge of the mapping of substrate atoms to product atoms for all reactions involving potentially labeled compounds. A number of algorithms for automatically deriving such atom mappings have been developed (for example [29]), and recent genome-scale metabolic reconstructions also include atom mappings [30]. However, correct atom mappings are crucial for modeling, and manual curation is still required, which is tedious for bigger models. The resulting isotopologue reaction networks can be, depending on the network and tracer, vastly larger than the usual metabolic networks due to the combinatorial complexity of stable isotope incorporation.

Having established a reaction network, mathematical expressions for describing reaction rates need to be chosen. Various approaches for specifying rate laws exist (reviewed in Ref. [8]). The choice of rate laws is a trade-off between 'mechanisticness' on the one hand and model complexity and the number of parameters on the other hand. Using a thermodynamics-based formalism [31,32] helps to create physically feasible models. Furthermore, such a parameterization can reduce the number of unknown parameters and simplify parameter estimation. The required thermodynamic parameters can be derived from experiments or approximated by group contribution methods or quantum-mechanical simulations [33].

Determining model parameters

Dynamic models usually come with a comparably high number of parameters whose values are not known *a priori*. Some parameter values can be retrieved from databases, while others need to be estimated. Often, a combination of both approaches is used.

Commonly used databases of experimentally determined kinetic parameters include BRENDA [34] and SABIO-RK [35]. Additionally, the BioModels Parameters database [36] provides structured access to parameter values that are used in models contained in the BioModels repository. If parameter values for a specific organism or experimental conditions are not available, values from closely related species or settings may still be useful [16]. Datanator provides a simple interface to find such related parameter values based on various similarity measures [37]. However, using parameters determined by *in vitro* assays or other makeshift parameter values comes with the caveat that they may not optimally reflect the *in vivo* situation [5]. As an alternative to taking the parameter values from databases as true values, they can instead be used as prior information during parameter estimation [8,18].

Given sufficiently informative data, model parameters can be inferred, for example, via optimization- [13–21,23] or sampling-based [22,28] approaches. However, this is computationally costly as, for most cases, it involves thousands to millions of numerical ODE simulations. Nevertheless, it has been demonstrated that parameter estimation is computationally tractable for large dynamic models. For example, through leveraging scalable algorithms, optimization-based maximum likelihood estimation was shown to be possible for a dynamic model of signaling pathways comprising over 1000 states and 4000 unknown parameters [38]. In some cases, the structure of the optimization problem can be exploited to further improve convergence and reduce computational costs. This has been demonstrated for optimization-based parameter inference from relative measurements [39], which are quite common in metabolomics datasets.

Scalable and highly optimized algorithms for model simulation and sensitivity analysis have been made available through easy-to-use toolboxes [40,41], and there exists a wide variety of algorithms and tools for parameter inference [42,43]. Many of these tools are able to exploit increasingly available high-performance computing resources, which is key for moving towards larger dynamic models [38,39]. The use of community standards for specifying models and parameter estimation problems [44–46] gives easy access to a majority of these tools.

Sometimes model parameters are estimated independently for different subsystems and only later combined in the full model [19,47]. Although this can be computationally cheaper, it might not yield the best fit to the data.

Dealing with uncertainty

Independently of the algorithm employed for parameter estimation, parameter estimates will be subject to uncertainty. This parameter uncertainty propagates to prediction uncertainty and can result in false conclusions [48]. Uncertainty analysis is, therefore, crucial.

Different methods to determine parameter confidence intervals exist. Commonly applied methods include those based on the Fisher information matrix (FIM), the profile likelihood approach, or sampling-based procedures [49]. The choice of method is a trade-off between accuracy and computational complexity with FIM-based methods being the cheapest computationally but least accurate, and sampling being the most expensive but also most informative. Profile likelihood is often the method of choice since it provides accurate results and is more scalable than sampling-based methods. Furthermore, as opposed to other methods, profile likelihood can be applied in spite of and to detect nonidentifiable parameters [49]. Implementations of these algorithms are available through several easy-to-use software tools [42].

Methods that are similar to those used for parameter uncertainty analysis are applied for prediction uncertainty analysis [48,49]. Sampling-based approaches account for the prediction uncertainty by design, as a sample from the prediction posterior is acquired while generating a sample from the parameter posterior. Where sampling or prediction profile likelihood is computationally too expensive, a cheaper ensemble-based approach can be applied instead [48]. During parameter estimation, multiple parameter vectors will be obtained that result in similarly good model fits. An ensemble of models can, for example, be built from these parameter vectors, and the spread of the resulting simulations can be used as an estimate for prediction uncertainty [48].

Recent applications of dynamic models of metabolism include examples of identifiability analysis [17,50], sampling-based assessment of parameter and prediction uncertainties [13], as well as other approaches to uncertainty analyses [15].

It is important to note that large parameter uncertainties do not have to manifest in large prediction uncertainties but may still allow for deriving robust predictions [18,38,51]. For example, Feldman-Salit et al. [18] generated an ensemble of models that reproduced the available data equally well and used it to derive predictions that were consistent across the ensemble. However, if no robust predictions can be derived, additional measurements need to be considered for parameter estimation. Optimal experimental design approaches can help to decide the most informative experiment [52].

Conclusions

Increasingly detailed metabolic reconstructions for more and more organisms provide the basis for constructing genome-scale models but also for deriving targeted submodels. Efficient implementations of scalable algorithms enable simulation and sensitivity analysis for dynamic models with as many as a few thousand state variables, even on personal computers. Adopting community standards for specifying models or parameter estimation problems not only facilitates reproducibility and reusability of models and data [46], but also grants easy access to a wide ecosystem of tools for model simulation, parameter inference, and further analyses.

Computationally challenging parameter estimation is often considered the major bottleneck for data-based dynamic (metabolic) modeling [5]. It is true that for very large models, this still is intractable, and more scalable algorithms are required. Results from other fields of application suggest that a better leveraging of both existing tools and computational resources can facilitate work with models and datasets that are an order of magnitude larger than those currently used in the field of dynamic modeling of metabolism.

Aside from any algorithms, successful model-based data analysis depends on matching models and data. For obtaining informative data, experiments should ideally be designed with a specific modeling goal in mind, and the available data should be taken into account when building models. Nevertheless, parameter uncertainty will always remain. Robust predictions may still be possible despite large parameter uncertainties. Therefore, dynamic modeling should not be ruled out prematurely due to supposedly too sparse data. However, awareness and assessment of parameter and prediction uncertainties are crucial, although often neglected.

In summary, we feel that dynamic modeling is a valuable yet underused tool for integrative metabolomics data analysis that could help to derive a more comprehensive, quantitative, and dynamic understanding of metabolism in many applications. We expect that with the increasing availability of scalable and interoperable tools, dynamic modeling will become more accessible and will play a more prominent role in the future. Given the efforts on integrating small dynamic models with genome-scale models [53] and the increasing size of fully dynamic models, we think that data integration using genome-scale dynamic models will become feasible in the future, although there is still a long way to go.

Conflict of interest statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the German Federal Ministry of Education and Research (BMBF) within the e:Med funding scheme (junior research alliance PeriNAA, grant no. 01ZX1916A).

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Jang C, Chen L, Rabinowitz JD: **Metabolomics and isotope tracing**. *Cell* 2018, **173**:822–837, https://doi.org/10.1016/j.cell.2018.03.055.

2. Liu X, Zhou L, Shi X, Xu G: **New advances in analytical methods for mass spectrometry-based large-scale metabolomics study**. *Trac Trends Anal Chem* 2019, **121**:115665, https://doi.org/10.1016/j.trac.2019.115665.

3. ** Volkova S, Matos MRA, Mattanovich M, Marín de Mas I: **Metabolic modelling as a framework for metabolomics data integration and analysis**. *Metabolites* 2020, **10**, https://doi.org/10.3390/metabo10080303.
Broad review of metabolic modeling approaches for data integration.

4. Antoniewicz MR: **A guide to metabolic flux analysis in metabolic engineering: methods, tools and applications**. *Metab Eng* 2021, **63**:2–12, https://doi.org/10.1016/j.ymben.2020.11.002.

5. Foster CJ, Wang L, Dinh HV, Suthers PF, Maranas CD: **Building kinetic models for metabolic engineering**. *Curr Opin Biotechnol* 2021, **67**:35–41, https://doi.org/10.1016/j.copbio.2020.11.010.

6. Gopalakrishnan S, Dash S, Maranas C: **K-fit: an accelerated kinetic parameterization algorithm using steady-state fluxomic data**. *Metab Eng* 2020, **61**:197–205, https://doi.org/10.1016/j.ymben.2020.03.001.

7. Strutz J, Martin J, Greene J, Broadbelt L, Tyo K: **Metabolic kinetic modeling provides insight into complex biological questions, but hurdles remain**. *Curr Opin Biotechnol* 2019, **59**:24–30, https://doi.org/10.1016/j.copbio.2019.02.005.

8. Saa PA, Nielsen LK: **Formulation, construction and analysis of kinetic models of metabolism: a review of modelling frameworks**. *Biotechnol Adv* 2017, **35**:981–1003, https://doi.org/10.1016/j.biotechadv.2017.09.005.

9. Shi X, Xi B, Jasbi P, Turner C, Jin Y, Gu H: **Comprehensive isotopic targeted mass spectrometry: reliable metabolic flux analysis with broad coverage**. *Anal Chem* 2020, **92**:11728–11738, https://doi.org/10.1021/acs.analchem.0c01767.

10. Nonnenmacher Y, Palorini R, d'Herouël AF, Krämer L, Neumann-Schaal M, Chiaradonna F, Skupin A, Wegner A, Hiller K: **Analysis of mitochondrial metabolism in situ: combining stable isotope labeling with selective permeabilization**. *Metab Eng* 2017, **43**:147–155, https://doi.org/10.1016/j.ymben.2016.12.005.

11. Ali A, Abouleila Y, Shimizu Y, Hiyama E, Emara S, Mashaghi A, Hankemeier T: **Single-cell metabolomics by mass spectrometry: advances, challenges, and future applications**. *Trac Trends Anal Chem* 2019, **120**:115436, https://doi.org/10.1016/j.trac.2019.02.033.

12. Thiele C, Wunderling K, Leyendecker P: **Multiplexed and single cell tracing of lipid metabolism**. *Nat Methods* 2019, **16**:1123–1130, https://doi.org/10.1038/s41592-019-0593-6.

13. Millard P, Enjalbert B, Uttenweiler-Joseph S, Portais J-C, Létisse F: **Control and regulation of acetate overflow in *Escherichia coli***. *eLife* 2021, **10**, e63661, https://doi.org/10.7554/eLife.63661.

14. Ramos JRC, Rath AG, Genzel Y, Sandig V, Reichl U: **A dynamic model linking cell growth to intracellular metabolism and extracellular by-product accumulation**. *Biotechnol Bioeng* 2020, **117**:1533–1553, https://doi.org/10.1002/bit.27288.

15. * Lövfors W, Ekström J, Jönsson C, Strålfors P, Cedersund G, Nyman E: **A multi-level model analysis of lipolysis and fatty acid release from adipocytes in vitro and from adipose tissue in vivo**. *bioRxiv* 2020, https://doi.org/10.1101/2020.12.18.423229.
Model of lipolysis integrating time-resolved metabolite measurements.

16. Berndt N, Bulik S, Wallach I, Wünsch T, König M, Stockmann M, Meierhofer D, Holzhütter H-G: **HEPATOKIN1 is a biochemistry-based model of liver metabolism for applications in medicine and pharmacology**. *Nat Commun* 2018, **9**:2386, https://doi.org/10.1038/s41467-018-04720-9.

17. Carter SJ, Ferecskó AS, King L, Ménochet K, Parton T, Chappell MJ: **A mechanistic modelling approach for the determination of the mechanisms of inhibition by cyclosporine on the uptake and metabolism of atorvastatin in rat hepatocytes using a high throughput uptake method**. *Xenobiotica* 2020, **50**:415–426, https://doi.org/10.1080/00498254.2019.1652781.

18. * Feldman-Salit A, Veith N, Wirtz M, Hell R, Kummer U: **Distribution of control in the sulfur assimilation in *Arabidopsis thaliana* depends on environmental conditions**. *New Phytol* 2019, **222**:1392–1404, https://doi.org/10.1111/nph.15704.
Ensemble-based analysis of control patterns.

19. Marín-Hernández Á, Gallardo-Pérez JC, Reyes-García MA, Sosa-Garrocho M, Macías-Silva M, Rodríguez-Enríquez S, Moreno-Sánchez R, Saavedra E: **Kinetic modeling of glucose central metabolism in hepatocytes and hepatoma cells**. *Biochim Biophys Acta Gen Sub* 2020, **1864**:129687, https://doi.org/10.1016/j.bbagen.2020.129687.

20. Ou J, Bao T, Ernst P, Si Y, Prabhu SD, Wu H, Zhang JJ, Zhou L, Yang S-T, Liu XM: **Intracellular metabolism analysis of clostridium cellulovorans via modeling integrating proteomics, metabolomics and fermentation**. *Process Biochem* 2020, **89**:9–19, https://doi.org/10.1016/j.procbio.2019.10.032.

21. Moon SJ, Dong W, Stephanopoulos GN, Sikes HD: **Oxidative pentose phosphate pathway and glucose anaplerosis support maintenance of mitochondrial nadph pool under mitochondrial oxidative stress**. *Bioeng Transl Med* 2020, **5**, e10184, https://doi.org/10.1002/btm2.10184.

22. ** Horvath N, Vilkhovoy M, Wayman JA, Calhoun K, Swartz J, Varner JD: **Toward a genome scale sequence specific dynamic model of cell-free protein synthesis in *Escherichia coli***. *Metabol Eng Commun* 2020, **10**, e00113, https://doi.org/10.1016/j.mec.2019.e00113.
Data integration using a large model of *E. coli* cell-free protein synthesis.

23. Yilmaz D, Parulekar SJ, Cinar A: **A dynamic EFM-based model for antibody producing cell lines and model based evaluation of fed-batch processes**. *Biochem Eng J* 2020, **156**:107494, https://doi.org/10.1016/j.bej.2020.107494.

24. Fang X, Lloyd CJ, Palsson BO: **Reconstructing organisms in silico: genome-scale models and their emerging applications**. *Nat Rev Microbiol* 2020, **18**:731–743, https://doi.org/10.1038/s41579-020-00440-4.

25. Robinson JL, Kocabaş P, Wang H, Cholley P-E, Cook D, Nilsson A, Anton M, Ferreira R, Domenzain I, Billa V, Limeta A, Hedin A, Gustafsson J, Kerkhoven EJ, Svensson LT, Palsson BO, Mardinoglu A, Hansson L, Uhlén M, Nielsen J: **An atlas of human metabolism**. *Sci Signal* 2020, **13**, https://doi.org/10.1126/scisignal.aaz1482.

26. * van Rosmalen R, Smith R, Martins dos Santos V, Fleck C, Suarez-Diez M: **Model reduction of genome-scale metabolic models as a basis for targeted kinetic models**. *Metab Eng* 2021, **64**:74–84, https://doi.org/10.1016/j.ymben.2021.01.008.
Automated reduction of genome-scale models.

27. * Masid M, Ataman M, Hatzimanikatis V: **Analysis of human metabolism by reducing the complexity of the genome-scale models using redHUMAN**. *Nat Commun* 2020, **11**:2821, https://doi.org/10.1038/s41467-020-16549-2.
Workflow for automated generation of reduced models of human metabolism.

28. Hameri T, Fengos G, Hatzimanikatis V: **The effects of model complexity and size on metabolic flux distribution and control: case study in escherichia coli**. *BMC Bioinf* 2021, **22**:134, https://doi.org/10.1186/s12859-021-04066-y.

29. Jaworski W, Szymkuć S, Mikulak-Klucznik B, Piecuch K, Klucznik T, Kaźmierowski M, Rydzewski J, Gambin A, Grzybowski BA: **Automatic mapping of atoms across both simple and complex chemical reactions**. *Nat Commun* 2019, **10**:1434, https://doi.org/10.1038/s41467-019-09440-2.

30. Brunk E, Sahoo S, Zielinski DC, Altunkaya A, Dräger A, Mih N, Gatto F, Nilsson A, Preciat Gonzalez GA, Aurich MK, Prlić A, Sastry A, Danielsdottir AD, Heinken A, Noronha A, Rose PW, Burley SK, Fleming RMT, Nielsen J, Thiele I, Palsson BO: **Recon3D enables a three-dimensional view of gene variation in human metabolism**. *Nat Biotechnol* 2018, **36**:272–281, https://doi.org/10.1038/nbt.4072.

31. Gawthrop PJ, Pan M, Crampin EJ: **Modular dynamic biomolecular modelling: the unification of stoichiometry, thermodynamics, kinetics and data**. *bioRxiv* 2021, https://doi.org/10.1101/2021.03.24.436792.

32. Mason JC, Covert MW: **An energetic reformulation of kinetic rate laws enables scalable parameter estimation for biochemical networks**. *J Theor Biol* 2019, **461**:145–156, https://doi.org/10.1016/j.jtbi.2018.10.041.

33. * Joshi RP, McNaughton A, Thomas DG, Henry CS, Canon SR, McCue LA, Kumar N: **Quantum mechanical methods predict accurate thermodynamics of biochemical reactions**. *ACS Omega* 2021, **6**:9948–9959, https://doi.org/10.1021/acsomega.1c00997.
Prediction of thermodynamics parameters from first-principles.

34. Chang A, Jeske L, Ulbrich S, Hofmann J, Koblitz J, Schomburg I, Neumann-Schaal M, Jahn D, Schomburg D: **BRENDA, the ELIXIR core data resource in 2021: new developments and updates**. *Nucleic Acids Res* 2020, **49**:D498–D508, https://doi.org/10.1093/nar/gkaa1025.

35. Wittig U, Rey M, Weidemann A, Kania R, Müller W: **SABIO-RK: an updated resource for manually curated biochemical reaction kinetics**. *Nucleic Acids Res* 2017, **46**:D656–D660, https://doi.org/10.1093/nar/gkx1065.

36. Glont M, Arankalle C, Tiwari K, Nguyen TVN, Hermjakob H, Malik-Sheriff RS: **BioModels Parameters: a treasure trove of parameter values from published systems biology models**. *Bioinformatics* 2020, **36**:4649–4654, https://doi.org/10.1093/bioinformatics/btaa560.

37. * Roth YD, Lian Z, Pochiraju S, Shaikh B, Karr JR: **Datanator: an integrated database of molecular data for quantitatively modeling cellular behavior**. *Nucleic Acids Res* 2020, **49**:D516–D522, https://doi.org/10.1093/nar/gkaa1008.
Web-based tool to retrieve various types of molecular data.

38. * Fröhlich F, Kessler T, Weindl D, Shadrin A, Schmiester L, Hache H, Muradyan A, Schütte M, Lim J-H, Heinig M, Theis FJ, Lehrach H, Wierling C, Lange B, Hasenauer J: **Efficient parameter estimation enables the prediction of drug response using a mechanistic pan-cancer pathway model**. *Cell Systems* 2018, **7**:567–579, https://doi.org/10.1016/j.cels.2018.10.013. e6.
Software for scalable simulation and sensitivity analysis of ODE models.

39. Schmiester L, Schälte Y, Fröhlich F, Hasenauer J, Weindl D: **Efficient parameterization of large-scale dynamic models based on relative measurements**. *Bioinformatics* 2019, **36**:594–602, https://doi.org/10.1093/bioinformatics/btz581.

40. Fröhlich F, Weindl D, Schälte Y, Pathirana D, Paszkowski L, Lines GT, Stapor P, Hasenauer J: **AMICI: high-performance sensitivity analysis for large ordinary differential equation models**. *Bioinformatics* 2021, https://doi.org/10.1093/bioinformatics/btab227,btab227.

41. Somogyi ET, Bouteiller J-M, Glazier JA, König M, Medley JK, Swat MH, Sauro HM: **libRoadRunner: a high performance SBML simulation and analysis library**. *Bioinformatics* 2015, **31**:3315–3321, https://doi.org/10.1093/bioinformatics/btv363.

42. Mitra ED, Hlavacek WS: **Parameter estimation and uncertainty quantification for systems biology models**. *Curr Opin Struct Biol* 2019, **18**:9–18, https://doi.org/10.1016/j.coisb.2019.10.006.

43. Villaverde AF, Fröhlich F, Weindl D, Hasenauer J, Banga JR: **Benchmarking optimization methods for parameter estimation in large kinetic models**. *Bioinformatics* 2018, **35**:830–838, https://doi.org/10.1093/bioinformatics/bty736.

44. Schreiber F, Sommer B, Czauderna T, Golebiewski M, Gorochowski TE, Hucka M, Keating SM, König M, Myers C, Nickerson D, Waltemath D: **Specifications of standards in systems and synthetic biology: status and developments in 2020**. *J Integr Bioinformat* 2020, **17**, https://doi.org/10.1515/jib-2020-0022.

45. Schmiester L, Schälte Y, Bergmann FT, Camba T, Dudkin E, Egert J, Fröhlich F, Fuhrmann L, Hauber AL, Kemmer S, Lakrisenko P, Loos C, Merkt S, Müller W, Pathirana D, Raimúndez E, Refisch L, Rosenblatt M, Stapor PL, Städter P, Wang D, Wieland F-G, Banga JR, Timmer J, Villaverde AF, Sahle S, Kreutz C, Hasenauer J, Weindl D: **PEtab—interoperable specification of parameter estimation problems in systems biology**. *PLoS Comput Biol* 2021, **17**:1–10, https://doi.org/10.1371/journal.pcbi.1008646.

46. Porubsky VL, Goldberg AP, Rampadarath AK, Nickerson DP, Karr JR, Sauro HM: **Best practices for making reproducible biochemical models**. *Cell Systems* 2020, **11**:109–120, https://doi.org/10.1016/j.cels.2020.06.012.

47. O'Donovan SD, Lenz M, Vink RG, Roumans NJT, de Kok TMCM, Mariman ECM, Peeters RLM, van Riel NAW, van Baak MA, Arts ICW: **A computational model of postprandial adipose tissue lipid metabolism derived using human arteriovenous stable isotope tracer data**. *PLoS Comput Biol* 2019, **15**:1–23, https://doi.org/10.1371/journal.pcbi.1007400.

48. Villaverde AF, Raimúndez-Álvarez E, Hasenauer J, Banga JR: **A comparison of methods for quantifying prediction uncertainty in systems biology**. *IFAC-PapersOnLine* 2019, https://doi.org/10.1016/j.ifacol.2019.12.234.

49. * Wieland F-G, Hauber AL, Rosenblatt M, Tönsing C, Timmer J: **On structural and practical identifiability**. *Curr Opin Struct Biol* 2021, **25**:60–69, https://doi.org/10.1016/j.coisb.2021.03.005.
Methods for identifiability and uncertainty analysis.

50. Roy M, Finley SD: **Metabolic reprogramming dynamics in tumor spheroids: insights from a multicellular, multiscale model**. *PLoS Comput Biol* 2019, **15**:1–36, https://doi.org/10.1371/journal.pcbi.1007053.

51. Christodoulou D, Link H, Fuhrer T, Kochanowski K, Gerosa L, Sauer U: **Reserve flux capacity in the pentose phosphate pathway enables Escherichia coli's rapid response to oxidative stress**. *Cell Syst* 2018, **6**:569–578, https://doi.org/10.1016/j.cels.2018.04.009.

52. Nimmegeers P, Bhonsale S, Telen D, Van Impe J: **Optimal experiment design under parametric uncertainty: a comparison of a sensitivities based approach versus a polynomial chaos based stochastic approach**. *Chem Eng Sci* 2020, **221**:115651, https://doi.org/10.1016/j.ces.2020.115651.

53. Ben Guebila M, Thiele I: **Dynamic flux balance analysis of whole-body metabolism for type 1 diabetes**. *Nat Comput Sci* 2021, **1**:348–361, https://doi.org/10.1038/s43588-021-00074-3.