Approaches to Computational Strain Design in the Multiomics Era

Peter C. St. John and Yannick J. Bomble*

National Renewable Energy Laboratory, Golden, CO, United States

Modern omics analyses are able to effectively characterize the genetic, regulatory, and metabolic phenotypes of engineered microbes, yet designing genetic interventions to achieve a desired phenotype remains challenging. With recent developments in genetic engineering techniques, timelines associated with building and testing strain designs have been greatly reduced, allowing for the first time an efficient closed loop iteration between experiment and analysis. However, the scale and complexity associated with multi-omics datasets complicates manual biological reasoning about the mechanisms driving phenotypic changes. Computational techniques therefore form a critical part of the Design-Build-Test-Learn (DBTL) cycle in metabolic engineering. Traditional statistical approaches can reduce the dimensionality of these datasets and identify common motifs among high-performing strains. While successful in many studies, these methods do not take full advantage of known connections between genes, proteins, and metabolic networks. There is therefore a growing interest in model-aided design, in which modeling frameworks from systems biology are used to integrate experimental data and generate effective and non-intuitive design predictions. In this mini-review, we discuss recent progress and challenges in this field. In particular, we compare methods augmenting flux balance analysis with additional constraints from fluxomic, genomic, and metabolomic datasets and methods employing kinetic representations of individual metabolic reactions, and machine learning. We conclude with a discussion of potential future directions for improving strain design predictions in the omics era and remaining experimental and computational hurdles.

Keywords: constraint-based methods, kinetic metabolic models, machine learning, multiomics, strain engineering

INTRODUCTION

The biorefinery concept involves the development of sustainable and low-impact production routes for major commodity chemicals and fuels from biomass (Bozell and Petersen, 2010). Biomass conversion using engineered microbes is a critical component of many production pathways, and offers the opportunity for high selectivity and yield (Nielsen and Keasling, 2016). However, optimizing microbial metabolism for a given process is time intensive and costly, limiting microbial conversions at present to only a few commercially successful compounds (Van Dien, 2013; Chubukov et al., 2016). This difficulty is primarily due to the complex relationship between genotype and phenotype, involving regulation at the metabolic, translational, and transcriptional levels. In recent years, the procedure of strain engineering has been formalized through the
Design-Build-Test-Learn (DBTL) cycle, which takes advantage of recent improvements in genetic engineering and high-throughput characterization in the Build and Test stages, respectively, to efficiently screen larger libraries of strain modifications (Liu et al., 2015). The Learn and Design stages use computational techniques to interpret experimental results and suggest further modification targets. The Learn step is perhaps the most weakly developed step of the DBTL cycle, and can take the form of a wide range of computational techniques from statistical analysis to detailed simulations (Nielsen and Keasling, 2016). In this minireview, we discuss recent research in methodology for integrating biological data – particularly in the form of multomics analyses – into developing new and efficient strain designs. We first review relevant experimental considerations from the Test stage and summarize the types of data available for informing strain designs. We next cover constraint based methods, kinetic simulations, and machine learning approaches, as well as recent studies that have used these methods in strain design. Lastly, we finish by discussing available software implementations and future directions for tackling the Learn step.

**EXPERIMENTAL INPUTS**

A number of recent reviews have covered the growing usefulness of omics approaches in characterizing cell physiology (Petzold et al., 2015; Nielsen, 2017; Becker and Wittmann, 2018; Yurkovich and Palsson, 2018), and therefore we only briefly cover the relevant data generated in typical strain characterization experiments. Frequently used omics data include transcriptomics, proteomics, metabolomics, and fluxomics, which measure gene expression, protein expression, metabolite concentrations, and intracellular fluxes, respectively. Transcriptomics is typically performed using next-generation sequencing methods that quantify relative differences in RNA expression within a given biological sample (Petzold et al., 2015). Relative comparisons between samples are also possible using statistical techniques (Wagner et al., 2012). Due to the similar physical nature of RNA transcripts, transcriptomics approaches are among the easiest to perform at the genome-scale, but their distance from metabolic networks by several layers of regulation makes direct understanding of metabolic function using these data difficult. Proteomics is one step closer to the determination of metabolic fluxes and uses mass spectrometry to quantify protein expression through the amino acid sequences of digested peptides (Kolker et al., 2006). Similar to transcriptomics, proteomics experiments typically measure relative protein expression within a sample, although statistical and experimental methods for comparing relative protein expression between samples are possible (Petzold et al., 2015). Absolute quantification of protein expression is feasible but more difficult, with a range of accuracies depending on the method used (Arike et al., 2012). While more involved than transcriptomics due to protein’s 3D structure and lack of amplification techniques, proteomic analyses are still able to survey a similar fraction of the protein-coding genome (Haider and Pal, 2013). Metabolomics poses an even greater challenge, as the high turnover of metabolites requires fast quenching and processing of samples (Petzold et al., 2015). As a result, the scope of metabolomic analyses are typically restricted to a smaller fraction of the organism’s metabolism. Similar to transcriptomics and proteomics, metabolite concentrations are typically measured as relative quantities in high-throughput exploratory experiments (Lei et al., 2011). Absolute metabolite quantifications are possible in targeted metabolomic studies using external or isotope-labeled standards. Lastly, fluxomics is concerned with accurately measuring internal fluxes of key metabolic reactions directly using isotopic labeling. While an excellent indicator of metabolic state, fluxomics is performed with less frequency than the previously discussed methods due to its experimental difficulty (Blank, 2016). In addition to careful cell culture and sample processing, fluxomics requires an accurate mathematical model that tracks atom transitions during metabolic reactions (Wiechert, 2001). This mathematical model is used in conjunction with 13C isotope labeling patterns to infer fluxes through each reaction, and as a result, inferred fluxes have typically been restricted to the main reactions in central carbon metabolism. However, extensions of MFA to include genome-scale flux analysis have been proposed (Gopalakrishnan and Maranas, 2015). Some genome-scale MFA methods leverage metabolism’s bow-tie structure to constrain fluxes through peripheral pathways with a high degree of confidence (Garcia Martin et al., 2015; Ando and Garcia Martin, 2018).

Even with access to direct measurements of activity for a wide range cellular machinery components, using these data to enhance metabolic flux for a desired pathway remains challenging. We next discuss Learn techniques that synthesize these vast data sources together with generalized knowledge of biological function.

**LEARN METHODOLOGY**

The goal of the Learn and Design steps is to use the characterization of previously engineered strains to develop improved strain designs. In its most basic form, this step can be accomplished by examining biological features (i.e., differentially expressed genes) correlated with improved strain performance, and overexpressing those likely involved in the pathway of interest (Yoshikawa et al., 2012). Designs based on rational consideration of omics data have proven successful (Guan et al., 2017), validating the human-in-the-loop approach. However, model driven designs will likely be critical to speeding up the DBTL cycle and revealing non-intuitive targets (Vickers, 2016). In the next sections, we review several lines of research into model-driven interpretation of omics data. A schematic of these approaches is shown in Figure 1.

**Constraint-Based Methods**

Constraint-Based Reconstruction and Analysis (COBRA) methods use biological knowledge and data to place constraints on intracellular fluxes, and in recent years have expanded to consider a wide range of recent omics techniques. Here we
focus on extensions of COBRA methods that pertain to guiding strain designs from omics data, while a number of recent reviews have covered COBRA methods in greater depth (O’Brien et al., 2015; Campbell et al., 2017; Stalidzans et al., 2018). A central technique to COBRA methods is flux balance analysis (FBA), which assumes that metabolite concentrations in the cell reach a pseudo steady-state when compared to the time scales associated with substrate uptake and cell division (Orth et al., 2010). This assumption allows fluxes to be constrained by mass balance equations developed from databases of biochemical reaction stoichiometry. Mass balance constraints alone (in the absence of 13C isotope labeling or other product data) are often not sufficient to determine a unique vector of metabolic fluxes. By assuming a cellular objective such as maximizing biomass or ATP production, unique flux vectors can be predicted. The accuracy of these predicted flux values are dependent on the objective chosen, and some objective functions have shown good correlation with experimental omics data (Lewis et al., 2010). Since such models can be simulated quickly and rely primarily on well-curated databases of metabolic reactions, many genome-scale models (GEMs) of microbial metabolism have been created (Henry et al., 2007). These principles have been successfully applied to select the most promising pathways for the synthesis of a variety of products (Averesch et al., 2017). Further extensions to the COBRA framework will likely include even more cellular functionality. Toward this goal, whole-cell models that integrate gene expression, protein production, and cell cycle have been constructed (Karr et al., 2012).

Constraint-Based Reconstruction and Analysis methods therefore represent an extensible and computationally efficient framework for connecting omics data of different types and have been used to successfully interpret omics data and improve strain designs in a number of studies (Wisselink et al., 2010; Brunk et al., 2016). An advantage of COBRA methods is their limited number of parameters that must be fit from experimental data, and therefore they are often able to suggest strain designs without substantial experimental support. In particular, these methods are especially efficient in determining metabolic changes that couple product production to cell growth (Long et al., 2015). The accuracy of constraint-based models in predicting de novo experimental results has not been rigorously evaluated and would serve a useful study in measuring the progress in our understanding of cellular behavior. However, even modest success rates from predictive tools are useful in guiding experimental efforts where the search space is vast. A limitation of constrained-based methods is that they are often less suitable for suggesting improvements to fine-tune the enzyme expression of an existing pathway. Such a task typically requires a kinetic description of the reactions in question, which we discuss in the next section.
Kinetic Metabolic Models

The goal of kinetic metabolic models is to capture the dynamic behavior of individual enzymes and integrate these expressions into the behavior of the full metabolic network. These models allow the direct prediction steady-state flux distributions as a function of enzyme expression, which typically serve as the most reliable experimental data for validation. However, models that explicitly incorporate enzyme kinetics (if parameterized correctly) are capable of predicting finer details of pathway dynamics, including the effect of slight changes in enzyme activity on metabolic flux. In constraint-based models, metabolite pools are assumed to be in a pseudo steady-state, and thus the rate rules governing flux through each reaction can be ignored. While the steady-state assumption may be justified, the specific steady-state reached inside the cell is determined, among a multitude of factors, both by external metabolite conditions as well as the kinetics and expression levels of metabolic enzymes. Kinetic modeling frameworks therefore seek to estimate these reaction rate rules from observed metabolic phenotypes to predict how enzyme perturbation will affect steady-state concentrations and fluxes.

Small-scale kinetic models of core carbon metabolism can leverage enzyme kinetics in vitro and time-course metabolite concentration measurements in fitting parameter values (Chassagnole et al., 2002). However, transient cellular responses are difficult to measure at the genome-scale, and direct enzyme kinetic measurements are sparser for peripheral pathways. Large-scale dynamic metabolic simulations are therefore largely based on steady-state flux and concentration data (Vasilakou et al., 2016). Because of these limited data, quantifying parameter uncertainty is therefore a critical challenge in large-scale kinetic models (Tummler and Klipp, 2018). Metabolic ensemble modeling addresses this challenge directly by finding distributions in parameter values that all reproduce the observed experimental data (Tran et al., 2008). This approach has been used to suggest subsequent enzymes in a linear pathway for overexpression (Contador et al., 2009), and an ensemble-based kinetic model of Escherichia coli has demonstrated superior predictive ability of steady-state flux distributions (Khodayari et al., 2014).

Smaller-scale, hand-curated kinetic models can use rate rules for individual enzymes with experimentally validated functional forms. However, traditional rate rule expressions (such as Michaelis–Menten kinetics) become difficult to construct for reactions with many participating species. Accordingly, larger-scale kinetic models typically choose a generalizable framework for constructing rate rule expressions. These frameworks range in computational complexity and faithfulness to the underlying enzyme-substrate system, and we leave a detailed comparison of these approaches to a number of recently published reviews (Heijnen, 2005; Hadlich et al., 2009; Du et al., 2016; Saa and Nielsen, 2017). Software available for kinetic modeling has continued to improve, and typically allows the user to specify reaction stoichiometry and rate rules independently from the chosen simulation algorithm. Such software includes COPASI (Hoops et al., 2006), CellDesigner (Funahashi et al., 2008), and MATCONT (Dhooge et al., 2003).

Regardless of the framework chosen, a major hurdle in using kinetic models for interpretation of omics data is the computational effort required in parameter estimation. In metabolic ensemble modeling, parameters are sampled at random and retained in the final ensemble only if they match all the considered experimental data (Tran et al., 2008). As a result, as more data is added or the model expanded, the computational costs increase substantially. Methods for improving the computational speed of the approach have been developed (Greene et al., 2017), but calculating steady states of the dynamic model remains a computational bottleneck. Ensemble-based inference approaches are therefore typically applied to smaller, core-carbon metabolic networks (Khodayari et al., 2014). A recent genome-scale kinetic modeling study optimized only a single parameter set due to the added cost of ensemble-based parameter estimation (Khodayari and Maranas, 2016). However, this single parameter set demonstrated a superior ability to reproduce a wide range of experimental observations compared with constraint-based methods (Khodayari and Maranas, 2016). The ensemble modeling sampling approach has been recently formalized as a form of Bayesian inference (Saa and Nielsen, 2016), demonstrating that detailed posterior distributions in parameter estimates and model predictions could be found. Kinetic models therefore offer a promising future direction for incorporating vast quantities of omics data into metabolic reconstructions if computational bottlenecks can be circumvented (St. John et al., 2018). While difficult to fit, the added parameters from kinetic representations give these models more expressive power in fitting experimental data.

A factor complicating the analysis of experimental data with kinetic models is the stochasticity introduced by low cell volumes and small copy numbers of several key enzymes (Levine and Hwa, 2007; Kiviet et al., 2014). Cell to cell heterogeneity therefore imposes unique challenges in understanding microbial kinetics that might be resolved through the use of explicit stochastic simulation algorithms (Gillespie, 1977) as implemented in a variety of software packages (Hoops et al., 2006; Sanft et al., 2011; Abel et al., 2016) In the subsequent section we discuss machine learning approaches that add even more parameters to be fit, but may prove useful as high-throughput strain construction and characterization techniques improve.

Machine Learning

Machine learning methods for interpreting omics data have taken a wide range of forms, largely due to the many varied biological questions that can be asked. In this section, we focus on methods that predict future targets for strain engineering. Integrative omics analyses attempt to draw connections between disparate omics data sources, either with or without prior biological knowledge (Berger et al., 2013; Bersanelli et al., 2016). These methods have been used to predict key regulatory genes correlated with metabolic productivity (Larsen et al., 2018), and inferred regulatory networks have also been incorporated into FBA models (Chandrasekaran and Price, 2010). Other studies have used machine learning to understand and predict metabolic performance from hyperparameters associated with cell growth. Wu et al. (2016) explored methods for machine learning in
meta-analysis to predict likely pathway success as a function of the complexity of the engineered pathway and other factors. In Kim et al. (2016), machine learning methods are used both for data reconciliation between omics sources, as well as to directly map the genotype-phenotype relationship. Another interesting study used machine learning methods as a replacement for the traditional rate equation frameworks discussed in the previous section (Costello and Martin, 2018). In that study, rate equations were learned directly from time-series metabolomics and were successful in predicting medium-producing strains given high and low-producing varieties. Costello and Martin (2018) also quantified the amount of data required for accurate rate determination at approximately 10 strains. Given the rapid advancement of machine learning methods and biological data collection, these approaches may offer flexible and efficient ways of directly incorporate biological data in new strain designs.

**DISCUSSION**

Since Learn lags behind the rest of the DBTL methodology in the development of validated and standardized techniques, feasible computational techniques are still being explored and improved upon. As a result, software libraries for performing the analyses described in this minireview are relatively scarce. As the most mature method of the three, COBRA methods have relatively strong software support in both the MATLAB (Heirendt et al., 2017) and Python (Ebrahim et al., 2013) ecosystems. Dependent packages have also been created for a number of the COBRA extensions for integrating or predicting omics-level data. Kinetic models, alternatively, have relatively poor support in the software landscape. This is likely due to the multitude of kinetic frameworks available as well as their slow (but parallelizable) convergence, requiring hardware-dependent simulation strategies. For machine learning, several actively developed packages are available that implement common approaches. Scikit-Learn for Python implements a variety of machine learning strategies under a consistent API (Pedregosa et al., 2011).

Deep learning frameworks such as Tensorflow or PyTorch simplify the process of constructing deep neural networks and training them on specialized hardware. Compared to the availability of general-purpose machine learning, omics-specific machine learning analyses have substantially fewer libraries under active development. However, creating and distributing standardized Learn work flows will be critical to enabling the reproducible analyses required of the iterative DBTL cycle. Such standardized approaches will necessarily require the development and maintenance of software and best practices in the metabolic modeling community.

**AUTHOR CONTRIBUTIONS**

PSJ and YB contributed to the conception and writing of the manuscript. PSJ created the figure. YB supervised the research.

**FUNDING**

We thank the U.S. Department of Energy Bioenergy Technologies Office for funding under Contract DE-AC36-08GO28308 with the National Renewable Energy Laboratory. This work was authored by Alliance for Sustainable Energy, LLC, the Manager and Operator of the National Renewable Energy Laboratory for the United States Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government.

**ACKNOWLEDGMENTS**

The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.

**REFERENCES**

Abel, J. H., Drawert, B., Hellander, A., and Petzold, L. R. (2016). GillesPy: a python package for stochastic model building and simulation. IEEE Life Sci. Lett. 2, 35–38. doi: 10.1109/lisls.2017.2652448 doi: 10.1109/lisls.2017.2652448

Åkesson, M., Forster, J., and Nielsen, J. (2004). Integration of gene expression data into genome-scale metabolic models. Met. Eng. 6, 285–293. doi: 10.1016/j.ymben.2003.12.002

Ando, D., and Garcia Martin, H. (2018). Two-scale 13C metabolic flux analysis for metabolic engineering. Synth. Metab. Pathways 1671, 333–352. doi: 10.1007/978-1-4939-7295-1_21

Aripe, L., Valgepea, K., Peil, L., Nahku, R., Adamberg, K., and Vilu, R. (2012). Comparison and applications of label-free absolute proteome quantification methods on Escherichia coli. J. Proteomics 75, 5437–5448. doi: 10.1016/j.jprot.2012.06.020

Averesch, N. I. H., Martinez, V. S., Nielsen, L. K., and Krömer, J. O. (2017). Toward synthetic biology strategies for adipic acid production: an in silico tool for combined thermodynamics and stoichiometric analysis of metabolic networks. ACS Synth. Biol. 7, 490–509. doi: 10.1021/acssynbio.7b00304

Basan, M., Hui, S., Okano, H., Zhang, Z., Shen, Y., Williamson, J. R., et al. (2015). Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. Nature 528, 99–104. doi: 10.1038/nature15765

Becker, J., and Wittmann, C. (2018). From systems biology to metabolically engineered cells — an omics perspective on the development of industrial microbes. Curr. Opin. Microbiol. 45, 180–188. doi: 10.1016/j.mib.2018.06.001

Becker, S. A., and Palsson, B. O. (2008). Context-specific metabolic networks are consistent with experiments. PLoS Comput. Biol. 4:e1000082. doi: 10.1371/journal.pcbi.1000082

Berger, B., Peng, J., and Singh, M. (2013). Computational solutions for omics data. Nat. Rev. Genet. 14, 333–346. doi: 10.1038/nrg3433

Bersanelli, M., Mosca, E., Remondini, D., Giampieri, E., Sala, C., Castellani, G., et al. (2016). Methods for the integration of multi-omics data: mathematical aspects. BMC Bioinformatics 17:S15. doi: 10.1186/s12859-015-0857-9

Blank, L. M. (2016). Let’s talk about flux or the importance of (intracellular) reaction rates. Microbial Biotechnol. 10, 28–30. doi: 10.1111/1751-7915.12455

Bozell, J. J., and Petersen, G. R. (2010). Technology development for the production of bio-based products from biorefinery carbohydrates—the us department of energy’s “top 10” revisited. Green Chem. 12:539. doi: 10.1039/b922014c
Brunk, E., George, K. W., Alonso-Gutierrez, J., Thompson, M., Baidoo, E., Wang, G., et al. (2016). Characterizing strain variation in engineered E. coli using a multi-omics-based workflow. Cell Syst. 2, 335–346. doi: 10.1016/j.cels.2016.04.004

Campbell, K., Xia, J., and Nielsen, J. (2017). The impact of systems biology on bioprocessing. Trends Biotechnol. 35, 1116–1168. doi: 10.1016/j.tibtech.2017.08.011

Chandrasekaran, S., and Price, N. D. (2010). Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in Escherichia coli and Mycobacterium tuberculosis. Proc. Natl. Acad. Sci. U.S.A. 107, 17845–17850. doi: 10.1073/pnas.1005139107

Chang, R. L., Andrews, K., Kim, D., Wang, C., Godzik, A., and Palsson, B. O. (2013). Structural systems biology evaluation of metabolic thermotolerance in Escherichia coli. Science 340, 1220–1223. doi: 10.1126/science.1236012

Chassagnole, C., Noisommit-Rizzi, N., Schmid, J. W., Mauch, K., and Reuss, M. (2002). Dynamic modeling of the central carbon metabolism of Escherichia coli. Biotechnol. Bioeng. 79, 53–73. doi: 10.1002/bit.10288

Chubukov, V., Mukhopadhyay, A., Petzold, C. J., Keasling, J. D., and Martin, H. G. (2015). A kinetic model of the central carbon metabolism in Escherichia coli. Metab. Eng. 32, 12–22. doi: 10.1016/j.ymben.2015.08.006

Greene, J. L., Wäechter, A., Tyo, K. E., and Broadbelt, L. J. (2017). Acceleration of genome-scale metabolic pathway analysis using multi-omics data. Metab. Eng. 25, 50–62. doi: 10.1016/j.men.2015.05.014

Kim, M., Bai, N., Zorzaquinio, V., and Tagkopoulos, I. (2016). Multi-omics integration accurately predicts cellular state in unexplored conditions for Escherichia coli. Nat. Commun. 7:13090. doi: 10.1038/ncomms13090

King, Z. A., Lu, J., Dräger, A., Miller, P., Fedorowicz, S., Lerman, J. A., et al. (2015). BiGG models: a platform for integrating, standardizing and sharing genome-scale models. Nucleic Acids Res. 44, D515–D522. doi: 10.1093/nar/gkv1049

Kiviet, D. J., Nghe, P., Walker, N., Boulineau, S., Sunderlikova, V., and Tans, S. J. (2014). Stochasticity of metabolism and growth at the single-cell level. Nature 514, 376–379. doi: 10.1038/nature13582

Kolker, E., Higdon, R., and Hogan, J. M. (2006). Protein identification and expression analysis using mass spectrometry. Trends Microbiol. 14, 229–235. doi: 10.1016/j.tim.2006.03.005

Larsen, P. E., Zerbs, S., Laible, P. D., Collart, F. R., Korajczyk, P., Dai, Y., et al. (2018). Modeling the Pseudomonas sulfur regulon by quantifying the storage and communication of information. Mysteins 3:e00189-17. doi: 10.11128/mysteins.00189-17

Lei, Z., Hulman, D. V., and Sumner, L. W. (2011). Mass spectrum strategies in metabolomics. J. Biol. Chem. 286, 25435–25442. doi: 10.1074/jbc.R111.236961

Lerman, J. A., Hyduke, D. R., Latif, H., Portnoy, V. A., Lewis, N. E., Orth, J. D., et al. (2012). In silico model for modelling metabolism and gene product expression at genome scale. Nat. Commun. 3:929. doi: 10.1038/ncomms1928

Levine, E., and Hwa, T. (2007). Stochastic fluctuations in metabolic pathways. Proc. Natl. Acad. Sci. U.S.A. 104, 9224–9229. doi: 10.1073/pnas.0610971104

Lewis, N. E., Hixson, K. K., Conrad, T. M., Lerman, J. A., Charusanti, P., Polpitiya, A. D., et al. (2010). Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. Mol. Syst. Biol. 6:390. doi: 10.1038/msb.2010.47

Liu, R., Bassolo, M. C., Zeitoun, R. I., and Gill, R. T. (2015). Genome scale engineering techniques for metabolic engineering. Metab. Eng. 32, 143–154. doi: 10.1016/j.men.2015.09.013

Long, M. R., Ong, W. K., and Reed, J. I. (2015). Computational methods in metabolic engineering for strain design. Curr. Opin. Biotechnol. 34, 135–141. doi: 10.1016/j.copbio.2014.12.019

Nielsen, J. (2017). Systems biology of metabolism. Annu. Rev. Biochem. 86, 245–275. doi: 10.1146/annurev-biochem-061516-044757

Nilsson, A., Nielsen, J., and Palsson, B. O. (2017). Metabolic models of protein allocation call for the kinetome. Cell Syst. 5, 538–541. doi: 10.1016/j.cels.2017.11.013

O’Brien, E. J., Lerman, J. A., Chang, R. L., Hyduke, D. R., and Palsson, B. O. (2014). Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. Mol. Syst. Biol. 9, 693–693. doi: 10.1038/msb.2013.52

O’Brien, E. J., Monk, J. M., and Palsson, B. O. (2015). Using genome-scale models to predict biological capabilities. Cell 161, 971–987. doi: 10.1016/j.cell.2015.05.019

Orth, J. D., Thiele, I., and Palsson, B. O. (2010). What is flux balance analysis? Nat. Biotechnol. 28, 245–248. doi: 10.1038/nbt.1614

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2829.
St. John and Bomble Computational Strain Design Using Multiomics

Petzold, C. J., Chan, L. J. G., Nhan, M., and Adams, P. D. (2015). Analytics for metabolic engineering. *Front. Bioeng. Biotechnol.* 3:135. doi: 10.3389/fbioe.2015.00135

Saa, P. A., and Nielsen, L. K. (2016). Construction of feasible and accurate kinetic models of metabolism: a bayesian approach. *Sci. Rep.* 6:29635. doi: 10.1038/srep29635

Saa, P. A., and Nielsen, L. K. (2017). Formulation, construction and analysis of kinetic models of metabolism: a review of modelling frameworks. *Biotechnol. Adv.* 35, 981–1003. doi: 10.1016/j.biotechadv.2017.09.005

Sanft, K. R., Wu, S., Roh, M., Fu, J., Lim, R. K., and Petzold, L. R. (2011). StochKit2: software for discrete stochastic simulation of biochemical systems with events. *Bioinformatics* 27, 2457–2458. doi: 10.1093/bioinformatics/btr401

Shlomi, T., Cabili, M. N., Herrgård, M. J., Palsson, B. Ö., and Ruppin, E. (2008). Network-based prediction of human tissue-specific metabolism. *Nat. Biotechnol.* 26, 1003–1010. doi: 10.1038/nbt.1487

St. John, P., Strutz, J., Broadbelt, L. J., Tyo, K. E. J., and Bomble, Y. J. (2018). Bayesian inference of metabolic kinetics from genome-scale multiomics data. *bioRxiv* https://doi.org/10.1101/450163

Stalidzans, E., Seiman, A., Peebo, K., Komasilovs, V., and Pentjuss, A. (2018). Model-based metabolism design: constraints for kinetic and stoichiometric models. *Biochem. Soc. Trans.* 46, 261–267. doi: 10.1042/bst20170263

Tran, L. M., Rizk, M. L., and Liao, J. C. (2008). Ensemble modeling of metabolic networks. *Biophys. J.* 95, 5666–5617. doi: 10.1529/biophysj.108.135442

Tummler, K., and Klipp, E. (2018). The discrepancy between data for and expectations on metabolic models: How to match experiments and computational efforts to arrive at quantitative predictions? *Curr. Opin. Syst. Biol.* 8, 1–6. doi: 10.1016/j.coisb.2017.11.003

Van Dien, S. (2013). From the first drop to the first truckload: commercialization of microbial processes for renewable chemicals. *Curr. Opin. Biotechnol.* 24, 1061–1068. doi: 10.1016/j.copbio.2013.03.002

Vasiliakou, E., Machado, D., Theorell, A., Rocha, I., Nõh, K., Oldiges, M., et al. (2016). Current state and challenges for dynamic metabolic modeling. *Curr. Opin. Microbiol.* 33, 97–104. doi: 10.1016/j.mib.2016.07.008

Vickers, C. (2016). Bespoke design of whole-cell microbial machines. *Microb. Biotechnol.* 10, 35–36. doi: 10.1111/1751-7915.12460

Wagner, G. P., Kin, K., and Lynch, V. J. (2012). Measurement of mRNA abundance using rna-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* 131, 281–285. doi: 10.1007/s12064-012-0162-3

Wiechert, W. (2001). 13C metabolic flux analysis. *Metab. Eng.* 3, 195–206. doi: 10.1006/mben.2001.0187

Wisseling, H. W., Cipollina, C., Oud, B., Crimi, B., Heijnen, J. J., Pronk, J. T., et al. (2010). Metabolome, transcriptome and metabolic flux analysis of arabinose fermentation by engineered saccharomyces cerevisiae. *Metab. Eng.* 12, 537–551. doi: 10.1016/j.mib.2017.03.002

Yoshikawa, K., Furusawa, C., Hirasawa, T., and Shimizu, H. (2012). “Design of superior cell factories based on systems wide omics analysis,” in *Systems Metabolic Engineering*, eds C. Wittmann and S. Lee (Dordrecht: Springer), 57–81. doi: 10.1007/978-94-007-4534-6_3

Yurkovich, J. T., and Palsson, B. O. (2018). Quantitative -omic data empowers bottom-up systems biology. *Curr. Opin. Biotechnol.* 51, 130–136. doi: 10.1016/j.copbio.2018.01.009

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Copyright © 2019 St. John and Bomble. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.**