Recent advances in model-assisted metabolic engineering

Steinn Gudmundsson[1] and Juan Nogales[2,3*]

[1]University of Iceland, Department of Computer Science, School of Engineering and Natural Sciences, Bjargargata 1, 102 Reykjavik, Iceland

[2]Department of Systems Biology, Centro Nacional de Biotecnología, CSIC, Madrid, Spain

[3]Interdisciplinary Platform for Sustainable Plastics towards a Circular Economy-Spanish National Research Council (SusPlast-CSIC), Madrid, Spain

* Correspondence to: J. Nogales, j.nogales@csic.es

Abstract

Mathematical modelling of cellular processes enables predictions of biological phenotypes under perturbations and they are therefore widely used in metabolic engineering. Metabolic models can be roughly divided into two groups, genome-scale metabolic models that are based on steady-state assumptions and dynamic (kinetic) models that are frequently small in scale. Hybrid models attempt to bridge the gap between the two paradigms by integrating large experimental data sets with mechanistic models of metabolism, often using state-of-the-art machine learning algorithms. The new models hold great promise in significantly shortening the design-build-test-learn cycle of metabolic engineering. Here we review some recent developments in the field.

Highlights

- Metabolic models have become an essential tool in the metabolic engineering toolbox.
- Constraint-based metabolic models are useful for identifying global changes to metabolism while dynamic models help fine tune individual pathways.
- The many simplifying assumptions of constraint-based models are gradually being relaxed by the introduction of new modeling methodology.
- Hybrid models combine network structure with experimental data and machine learning algorithms for increased scope and improved accuracy.

Keywords

Metabolic models, Metabolic engineering, Model-based strain design, Omics data, Machine learning

Introduction

Circular economy has emerged as a key concept to address the global issues caused by dependence on non-renewable energy sources and increasing human population. In this scenario, microbial biotechnology is thought to play an important role by providing alternatives to current production chains [1]. The microbial metabolic space allows production of a large universe of metabolites. However, the titers, yield, and productivity of most compounds are rather low using naturally occurring microbial factories. Significant metabolic redesign is therefore often required to achieve cost-effective production of target compounds, a practice referred to as metabolic engineering [2,3]. Such efforts have furthermore expanded the known metabolic space via rational design of unnatural pathways with new stoichiometric balances, new-to-nature reactions, and new compounds [4-6].

The exploration of the microbial metabolic space, its optimization, and rational expansion requires holistic approaches that take system level properties into account. Mathematical modelling of cellular processes is increasingly used for optimizing microbial cell factories [7,8]. Metabolic models allow for a rational design process, integration of vast amounts of experimental data and can therefore reduce the amount of trial-and-error work involved in metabolic engineering [9,10]. Metabolic modelling formalisms can be roughly divided into two categories, constraint-based models (CBMs) and dynamic models (DMs), also known as kinetic models (Figure 1). CBMs are mathematical representations of cellular metabolism, which account for reaction stoichiometry and reversibility under the assumption of steady-state [11]. These models

are relatively easy to construct and work with. Despite their intrinsic simplicity, they turn out to be quite powerful tools to analyze biological networks at the genome-scale. However, they cannot directly address transient behaviors and do not provide information about metabolite concentrations, both of which are of importance in metabolic engineering. Dynamic models incorporate enzyme mechanisms and experimental data with reaction stoichiometry. They are usually described by a system of nonlinear differential equations and can provide detailed information about the time evolution of the system [12]. Recent advances in data acquisition and machine learning are driving the development of new methods for both steady-state and dynamic models. It is not unreasonable to expect that by merging big data sets with mechanistic models and state-of-the art machine learning algorithms we will soon witness a new era in genotype-to-phenotype predictions. Here we review recent developments in model-based metabolic engineering. In particular, methods that combine mechanistic models, large data sets and machine learning.

Constraint-based models

Constraint-based models are constructed through the systematic integration of genome annotation, omics data sets, and legacy knowledge such as reaction stoichiometry and gene-protein-reaction (GPR) rules. A CBM represents the metabolic capabilities of a particular organism and can be used to describe and predict the phenotype in response to environmental and/or genetic perturbations [11,13]. In the following, a CBM refers to a basic stoichiometric model of metabolism (Figure 1). Many phenotype prediction and strain optimization methods have been developed to date. Some of the methods assume that cell metabolism is shaped by specific biological goals. For instance, the widely used Flux Balance Analysis (FBA) is frequently used with the assumption that cell growth is the main biological objective [14]. Strain optimization methods find genetic perturbations resulting in overproduction of a target compound, compared

to a base strain, e.g. the wild type. Most methods published to date aim to identify designs which couple product secretion to growth, so-called growth-coupled designs. This approach has multiple advantages such as robustness against detrimental mutations and simple selection [15]. Growth-coupling has furthermore been demonstrated to be possible for almost all metabolites in five major production organisms [16]. Since the publication of the first growth-coupling algorithm, OptKnock [17], numerous algorithms for identifying growth-coupled strategies with CBMs have been published. Some of the algorithms identify only knockouts, but other algorithms consider knock-ins, overexpression and down-regulation as well. Alter and Ebert recently identified some of the underlying metabolic principles involved in growth-coupled designs, including carbon drain and cofactor and proton balancing [18]. Most of the strain design methods generate a large list of potential solutions, thus a systematic characterization and ranking of potential strategies is needed. A recent framework addressing this issue has been published, establishing simple criteria for the scoring and ranking of strategies [19]. Algorithms for growth-coupled production continue to be developed, e.g., GeneReg which identifies growth-coupled designs based on changes in gene expression by taking gene-protein-reaction rules directly into account [20]. The OptCouple algorithm extends the original OptKnock algorithm by identifying strategies that combine knock-outs, knock-ins and medium supplementation [21]. The concept of multi-objective optimization has been also explored. The MOMO algorithm identifies reaction deletions that optimize several functions simultaneously, including the concurrent maximization of a product and of biomass, or the maximization of a target product while minimizing the formation of a given by-product [22]. ModCell2 is a framework for modular cell design [23] that also employs multi-objective optimization. It identifies the genetic modifications needed to design modular cells (chassis) that can couple with a variety of production modules. A novel approach to strain design is based on evolutionary game theory [30]. The method identifies gene-associated reaction knockouts without the assumption of growth maximization. The algorithm considers a game between two players. One player corresponding to the host strain, attempts to avoid

overproduction of the target compound while the other player corresponds to the metabolic engineer that attempts to manipulate the network in order to disrupt the activity of the first player [31].

Pathway analysis methods based on the identification of elementary flux modes (EFMs), and the concept of Minimal Cut Sets (MCS) for strain design [24] have also turned out to be useful tools in metabolic engineering. The computation of EFMs and MCSs is computationally demanding for large networks but recent algorithm improvements have reduced the time complexity considerably, paving the way for widespread application of EFM and MCS in metabolic engineering [25-28]. While growth-coupled designs are frequently of interest, coupling production to growth is not always possible *in vivo*, e.g. due to GPR relationships, and in many bioprocesses it is not desirable, e.g., when producing toxic metabolites. The Metabolic Valve enumerator (MoVe) algorithm is an MCS-based method developed to address this situation [29]. MoVe uses a metabolic model to identify genetic intervention strategies which decouple two desired phenotypes such as growth and product formation.

Although CBMs have been used successfully in many metabolic engineering projects, the basic assumptions frequently made, steady-state conditions, lack of allosteric regulation and fixed capacity of enzymes, among others, limit the usefulness of these models. For example, they do not provide information about temporal dynamics and by ignoring enzyme kinetics, pathway bottlenecks are ignored. Numerous methods have been proposed in recent years that attempt to mitigate these limitations. We highlight some of these new methods in the following, emphasizing methods that involve high-throughput data.

Dynamic models

The steady-state assumption of CBMs means that they cannot be used to study temporal behavior, limiting the use of CBMs for modeling many conditions of interest, e.g. those found in bioreactors. Dynamic Flux Balance Analysis (dFBA) is an extension of FBA that simulates changes in the extracellular environment by assuming that intracellular concentrations reach a steady state rapidly in response to extracellular changes [30]. Multiple dFBA approaches have been proposed, see [31] for a recent overview and description of a state-of-the-art interior-point method applicable to genome-scale models. The mcPECASO framework uses two-stage dynamic-FBA to identify growth and phenotypic targets that optimize titer, rate and yield values [32] and represents an interesting alternative to static strain design approaches.

More sophisticated dynamic models account for detailed kinetic information of a given network including metabolic fluxes, enzyme and metabolite levels and allosteric interactions (Figure 1). DMs therefore have broader applicability than CBMs. They enable predictions of cell behavior over time, in response to genetic and environmental perturbations, and model nonlinear behavior of the underlying system. However, these models contain many parameters that have to be obtained experimentally, often with considerable effort [33] which again limits their widespread use [34]. For large DMs the experimental effort becomes prohibitive and the parameter values are therefore estimated indirectly. A number of reviews focusing on the construction and analysis of dynamic models have recently been published [12,35-37].

In the context of metabolic engineering, the goal is to predict the behavior of a particular biological system under genetic or environmental perturbations. In practice, this goal is achieved by identifying parameter values for defined kinetic expressions, resulting in some desired biotechnological output, e.g., overproduction of a target compound. The classical framework for elucidating parameters responsible for the control of metabolic fluxes over time is metabolic

control analysis (MCA) [38]. The functional states are quantified using control coefficients, which provide information about changes in the metabolic flux or metabolite concentration in response to changes in enzymatic activity. Rational metabolic design approaches take advantage of this information to identify rate-limiting steps of the network which correspond to potential targets for engineering. The ORACLE framework is based on MCA and uses uncertainty analysis in the study of metabolic pathways [39]. It is an ensemble method thatprovides an alternative to full-scale parameter estimation methods and has facilitated the construction of large-scale kinetic models. The ORACLE framework generates many feasible versions of the same model by sampling the parameter space and performs statistical analysis of the results. It has e.g., been used to predict genetic targets for the production of 1,4-butanediol in *E. coli* [40] and more recently to increase stress endurance in *P. putida* [41]. Alternatively, DMs can be obtained from CBMs by creating a reduced stoichiometric model that captures the properties of the CBM that are most relevant to the engineering task at hand. The reduced model can then be used with methods that automatically construct DMs from stoichiometric models, standardized rate laws, and regulatory interactions [42].

Recent updates to MCA include the development of a method to compute simultaneous confidence intervals for flux control coefficients, enabling the quantification of the sensitivity of enzyme levels on metabolite concentrations [43]. This method is noteworthy since it is the first method to assign statistical significance to the output of ensemble modeling in metabolic engineering. The NRA method is a constraint-based metabolic control analysis framework for rational strain engineering [44] that makes use of the confidence intervals. NRA enables physiologically relevant bounds and design constraints to be imposed on the system and identifies thermodynamically and kinetically consistent metabolic engineering targets. The method can be used for a wide range of optimization criteria and with various physiological constraints using large-scale kinetic models.

Hybrid models and machine learning tools in metabolic engineering

In this section we describe methods that extend the basic constraint-based and dynamic models from previous sections (Figure 1). For CBMs, this can involve thermodynamic constraints, modeling cell behavior at multiple scales, e.g. metabolism and macromolecular synthesis, or the incorporation of experimental data, typically large-scale omics data sets, in order to improve model accuracy. In case of DMs, the data is typically used to infer values of kinetic parameters that cannot be obtained by direct experiments.

Flux balance analysis of CBMs can result in fluxes that correspond to thermodynamically infeasible cycles. Several approaches have been proposed to overcome this problem such as TFA which adds constraints on flux directionality so that it is consistent with the corresponding change in Gibb's energy [45]. This ensures that flux values are guaranteed to be thermodynamically feasible and furthermore, provides a link between fluxes and metabolite concentrations. Python and Matlab implementations of TFA have recently become available [46] but it should be noted that TFA makes use of experimental data that may not be directly available for the organism under study.

A major limitation of CBMs is the lack of regulatory information. The OptRAM algorithm extends traditional growth-coupling strain design algorithms by identifying engineering strategies for transcription factors as well as for metabolic genes [47]. The algorithm does not require an existing regulatory network to identify transcription factor manipulations but is able to infer the regulatory network directly from transcriptomic data. The authors validated their method experimentally for ethanol production in yeast. Another frequently made assumption in CBMs is that the production of metabolites is only limited by carbon uptake, ignoring the role of enzymatic levels, and enzymatic activities in determining fluxes. A number of methods for integrating

transcriptomic or proteomic data with CBMs have been published to date that attempt to address this issue [48]. The general methodology is to take transcript (or protein) levels as proxy for enzyme load and modulate fluxes accordingly. The expectation is that cellular processes such as gene regulation that are not directly included in the original stoichiometric model will then be taken implicitly into account. While it is reasonable to assume that such strategies can lead to improved prediction accuracy, much work remains to be done in order to understand how best to achieve this goal [49,50].

The GECKO method [51] takes enzyme capacity into account by adding new constraints to the model. The enzyme constraints are derived from experimentally determined enzyme turnover numbers and abundance values obtained from proteomics data, when available. An advantage of this method is that the resulting model can be used directly with most existing software for CBMs. Follow-up work used an enzyme-constrained model and Bayesian statistical learning to identify enzymes which limit the growth of yeast at superoptimal temperatures [52]. The enzyme that was predicted to be the most rate-limiting was replaced by a thermotolerant homolog, resulting in increased growth rate compared to the wild type. GECKO has recently been used to generate a catalogue of enzyme constrained models from existing CBMs and now supports continuous and version-controlled updates of such models [53].

Experimentally determined enzyme turnover values are mostly based on *in-vitro* measurements and do not necessarily reflect *in-vivo* conditions. An alternative to sourcing enzyme turnover values experimentally has recently been proposed [54]. In this method, regression models were used to predict effective turnover rates in *E. coli* using features derived from enzyme biochemistry, structural properties and metabolic network properties. The method was tested on two modeling frameworks by predicting quantitative proteomic data and was found to outperform methods based on *in-vitro* turnover numbers.

Metabolic-expression models are CBMs that have been combined with mechanistic models of gene expression (ME-models) [55]. An interesting use case of ME-models in metabolic engineering is ranking strain designs obtained from CBMs, by taking protein cost and kinetic variability into account [56]. The DynamicME algorithm combines ME-models with dynamic FBA and enables time-course simulation of cell metabolism and protein expression. The algorithm correctly predicted the substrate utilization hierarchy on mixed carbon substrates [57]. The ETFL framework incorporates thermodynamic constraints are in ME-models [58]. This formalism was used in a dynamic setting to explain the intracellular mechanism underlying diauxic growth in *E. coli* [59].

Resource Balance Analysis (RBA) [60] shares similarities with both GECKO and the ME-model formalism in the sense that they all extend FBA by imposing additional constraints, e.g., on enzyme capacity. RBA enables quantitative predictions of resource allocation in constraint-based models, including abundance of enzymes, transporters and ribosomes. This can be used to identify cell functionality that is superflous under given industrial process conditions. Deletion of the unused functionality would then free up resources for additional growth and/or synthesis of the target product. A package for the automatic generation of RBA models from CBMs is available [61].

Recent developments in machine learning and the large amount of publicly available omics data sets help advance dynamic modeling approaches. Dynamic models ranging from the small-scale to almost genome-scale can now be parametrized automatically and used in strain design [62]. The PathParser tool [63] performs thermodynamic and kinetic analysis of metabolic pathways and provides estimates of protein cost. Metabolomics, fluxomic and proteomic data are used as inputs together with enzymatic constants obtained from online databases. The method was used to analyze the Calvin cycle and photorespiration in a cyanobacterium but can potentially be extended to genome-scale models. In another example, a Bayesian inference method has

been developed for predicting steady-state fluxes and metabolite concentrations in metabolic networks, using metabolomic, proteomic and fluxomic data [34]. The K-FIT algorithm [64] performs parametrization of genome-scale kinetic models using 13C fluxomic data.

Until recently, the analysis of high-dimensional biological data was hampered by the lack of suitable tools. During the last decade, developments within the field of machine learning in data visualization, deep neural networks, data fusion, model interpretation and more have resulted in new tools that hold great promise for dealing with disparate omics data sets. The advent of hybrid metabolic models and efficient parametrization methods is likely to shorten the design-build-test-learn cycle significantly, in particular the design and learn stages [9,65,66].

Conclusions

Chemical production with cell factories is an important step towards replacing non-renewable carbon and energy sources. To achieve sustainable, cost efficient microbial or plant production systems, significant redesign of the underlying cellular processes is almost always required. When redesigning genetic and regulatory circuits, metabolic engineers are increasingly relying on mathematical models of the underlying processes. Hybrid models bridge the gap between genome-scale metabolic models and dynamic models. They capture a mechanistic description of metabolism, retaining some of the scope and simplicity of CBMs and the kinetic details of DMs. The new generation of metabolic models leverage off recent advances modelling, machine learning, increasing data availability and expanding computational capabilities. Whether the new modeling methodologies will lead to a new dawn in metabolic engineering remains to be seen but judging by the many exiting studies that have come out recently, we believe that there is good reason for optimism.

Highlighted references:

*5. Ding S, Tian Y, Cai P, Zhang D, Cheng X, Sun D, Yuan L, Chen J, Tu W, Wei D-Q, et al.: novoPathFinder: a web server for designing novel-pathways with integrating GEM-model. Nucleic Acids Research 2020, 48:W477-W487.

This paper describes novoPathFinder, a retrosynthesis tool, implemented as a web server, that enables predictions of novel pathways, given source and target compounds in the context of a given GEM.

**9. Czajka JJ, Oyetunde T, Tang YJ: Integrated knowledge mining, genome-scale modeling, and machine learning for predicting Yarrowia lipolytica bioproduction. Metabolic Engineering 2021, 67:227-236.

A holistic approach involving GEM analysis and ML was applied to identify engineering targets involved in the overproduction of a variety of products. Production data from literature was compiled and used to train the GEM-ML framework.

**18. Alter TB, Ebert BE: Determination of growth-coupling strategies and their underlying principles. BMC Bioinformatics 2019, 20:447.

This paper identifies some of the basic principles that drive growth-coupling designs, including carbon drain and cofactor balancing. The findings help guide synthetic engineering of such strategies.

* 23. Garcia S, Trinh CT: Multiobjective strain design: A framework for modular cell engineering. Metabolic Engineering 2019, 51:110-120.

The algorithm described in this paper identifies modular strain designs consisting of core metabolic pathways (chassis) and production pathways which are broken down into modules that can be individually fine-tuned. The algorithm holds promise for designing distributed catalysts in the context of microbiomes.

**29. Venayak N, von Kamp A, Klamt S, Mahadevan R: MoVE identifies metabolic valves to switch between phenotypic states. Nature Communications 2018, 9:5332.

MoVE identifies dynamically controlled metabolic valves in GEMs which decouple growth and production phenotypes, enabling high flux for each of the phenotypes. MoVe showed that decoupling of growth and production phenotypes was possible for a majority of natural chemicals in *E. coli* and *S. cerevisiae.*

**34. Raj K, Venayak N, Mahadevan R: Novel two-stage processes for optimal chemical production in microbes. Metabolic Engineering 2020, 62:186-197.

This paper describes a method to engineer a two-step bioprocess where growth and production are decoupled in time using dynamic pathway regulation. The method identifies optimal growth and production targets during the first and second stages, respectively, which optimize a user-defined combination of titer, rate and yield.

*42. van Rosmalen RP, Smith RW, Martins dos Santos VAP, Fleck C, Suarez-Diez M: Model reduction of genome-scale metabolic models as a basis for targeted kinetic models. Metabolic Engineering 2021, 64:74-84.

The authors address the difficulty of parameter estimation and simulation in large-scale dynamic models by creating a subset of the GEMs that capture the main properties of the original model and automatically constructing a DM from the reduced model.

Figure legend

Figure 1. An overview of metabolic models for metabolic engineering. A: Constraint-based models and **B:** Dynamic models are important computational tools that are frequently used to guide metabolic engineering efforts. However, the basic assumption of steady-state and the lack of kinetic information in constraint-based models and the limited scope of dynamic models, can lead to erroneous predictions. **C:** Hybrid models that combine network structure with experimental data and machine learning algorithms increase the scope of the metabolic models and subsequently provide more accurate predictions of engineering targets.

Acknowledgements

This work was supported by the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreements no. 814650 (SynBio4Flav) 870294 (MixUp) and 101000733 (Promicon), funding from the Spanish Ministry of Science and Innovation in the RobExplode

project: PID2019- 108458RB- I00 (AEI /10.13039/501100011033) and the Icelandic Research fund, project grant number 207088-051 (ThermoExplore*)*.

Conflict of interest statement

Nothing declared

References

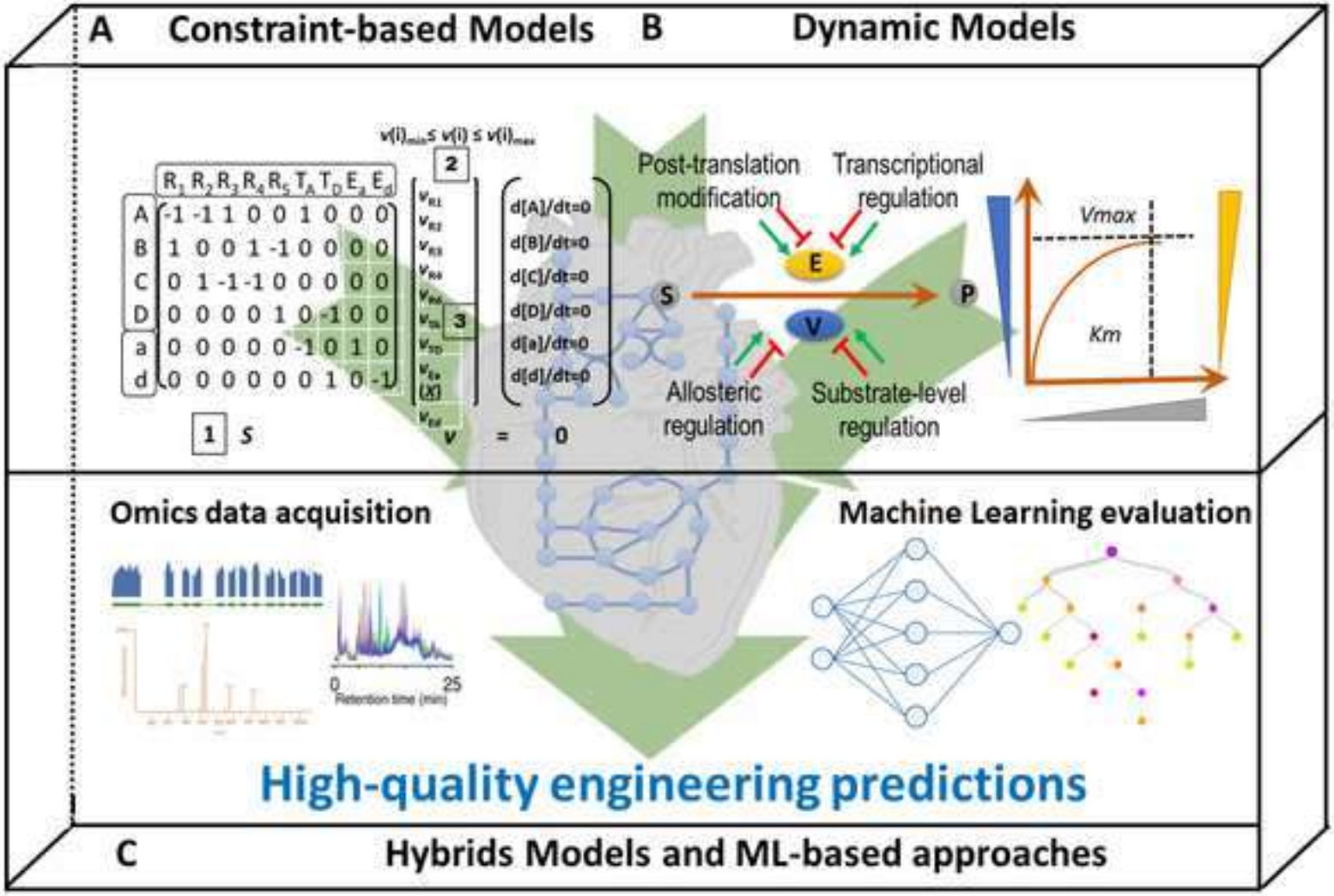
1. O'Connor KE: **Microbiology challenges and opportunities in the circular economy**. *Microbiology* 2021, **167**.
2. Alam K, Hao J, Zhang Y, Li A: **Synthetic biology-inspired strategies and tools for engineering of microbial natural product biosynthetic pathways**. *Biotechnology Advances* 2021, **49**:107759.
3. Yang D, Park SY, Park YS, Eun H, Lee SY: **Metabolic Engineering of Escherichia coli for Natural Product Biosynthesis**. *Trends in Biotechnology* 2020, **38**:745-765.
4. Prather KLJ: **Chemistry as biology by design**. *Microbial Biotechnology* 2019, **12**:30-31.
5. Ding S, Tian Y, Cai P, Zhang D, Cheng X, Sun D, Yuan L, Chen J, Tu W, Wei D-Q, et al.: **novoPathFinder: a webserver of designing novel-pathway with integrating GEM-model**. *Nucleic Acids Research* 2020, **48**:W477-W487.
6. Peretó J: **Transmetabolism: the non-conformist approach to biotechnology**. *Microbial Biotechnology* 2021, **14**:41-44.
7. Hafner J, Payne J, MohammadiPeyhani H, Hatzimanikatis V, Smolke C: **A computational workflow for the expansion of heterologous biosynthetic pathways to natural product derivatives**. *Nature Communications* 2021, **12**:1760.
8. Machado D, Herrgård MJ: **Co-evolution of strain design methods based on flux balance and elementary mode analysis**. *Metabolic Engineering Communications* 2015, **2**:85-92.
9. Czajka JJ, Oyetunde T, Tang YJ: **Integrated knowledge mining, genome-scale modeling, and machine learning for predicting Yarrowia lipolytica bioproduction**. *Metabolic Engineering* 2021, **67**:227-236.
10. Suthers PF, Foster CJ, Sarkar D, Wang L, Maranas CD: **Recent advances in constraint and machine learning-based metabolic modeling by leveraging stoichiometric balances, thermodynamic feasibility and kinetic law formalisms**. *Metabolic Engineering* 2021, **63**:13-33.
11. Fang X, Lloyd CJ, Palsson BO: **Reconstructing organisms in silico: genome-scale models and their emerging applications**. *Nature Reviews Microbiology* 2020, **18**:731-743.

12. Foster CJ, Wang L, Dinh HV, Suthers PF, Maranas CD: **Building kinetic models for metabolic engineering**. *Current Opinion in Biotechnology* 2021, **67**:35-41.
13. Gu C, Kim GB, Kim WJ, Kim HU, Lee SY: **Current status and applications of genome-scale metabolic models**. *Genome Biology* 2019, **20**:121.
14. Orth JD, Thiele I, Palsson BØ: **What is flux balance analysis?** *Nature Biotechnology* 2010, **28**:245-248.
15. Shepelin D, Hansen ASL, Lennen R, Luo H, Herrgård MJ: **Selecting the Best: Evolutionary Engineering of Chemical Production in Microbes**. *Genes* 2018, **9**:249.
16. von Kamp A, Klamt S: **Growth-coupled overproduction is feasible for almost all metabolites in five major production organisms**. *Nature Communications* 2017, **8**:15956.
17. Burgard AP, Pharkya P, Maranas CD: **Optknock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization**. *Biotechnology and Bioengineering* 2003, **84**:647-657.
18. Alter TB, Ebert BE: **Determination of growth-coupling strategies and their underlying principles**. *BMC Bioinformatics* 2019, **20**:447.
19. Schneider P, Klamt S: **Characterizing and ranking computed metabolic engineering strategies**. *Bioinformatics* 2019, **35**:3063-3072.
20. Razaghi-Moghadam Z, Nikoloski Z: **GeneReg: A constraint-based approach for design of feasible metabolic engineering strategies at the gene level**. *Bioinformatics* 2020.
21. Jensen K, Broeken V, Hansen ASL, Sonnenschein N, Herrgård MJ: **OptCouple: Joint simulation of gene knockouts, insertions and medium modifications for prediction of growth-coupled strain designs**. *Metabolic Engineering Communications* 2019, **8**:e00087.
22. Andrade R, Doostmohammadi M, Santos JL, Sagot M-F, Mira NP, Vinga S: **MOMO - multi-objective metabolic mixed integer optimization: application to yeast strain engineering**. *BMC Bioinformatics* 2020, **21**:69.
23. Garcia S, Trinh CT: **Multiobjective strain design: A framework for modular cell engineering**. *Metabolic Engineering* 2019, **51**:110-120.
24. Klamt S, Gilles ED: **Minimal cut sets in biochemical reaction networks**. *Bioinformatics* 2004, **20**:226-234.
25. Apaolaza I, Valcarcel LV, Planes FJ: **gMCS: fast computation of genetic minimal cut sets in large networks**. *Bioinformatics* 2018, **35**:535-537.
26. Klamt S, Mahadevan R, von Kamp A: **Speeding up the core algorithm for the dual calculation of minimal cut sets in large metabolic networks**. *BMC Bioinformatics* 2020, **21**:510.
27. Miraskarshahi R, Zabeti H, Stephen T, Chindelevitch L: **MCS2: minimal coordinated supports for fast enumeration of minimal cut sets in metabolic networks**. *Bioinformatics* 2019, **35**:i615-i623.
28. Schneider P, von Kamp A, Klamt S: **An extended and generalized framework for the calculation of metabolic intervention strategies based on minimal cut sets**. *PLOS Computational Biology* 2020, **16**:e1008110.
29. Venayak N, von Kamp A, Klamt S, Mahadevan R: **MoVE identifies metabolic valves to switch between phenotypic states**. *Nature Communications* 2018, **9**:5332.
30. Mahadevan R, Edwards JS, Doyle FJ: **Dynamic Flux Balance Analysis of Diauxic Growth in Escherichia coli**. *Biophysical Journal* 2002, **83**:1331-1340.
31. Scott F, Wilson P, Conejeros R, Vassiliadis VS: **Simulation and optimization of dynamic flux balance analysis models using an interior point method reformulation**. *Computers & Chemical Engineering* 2018, **119**:152-170.
32. Raj K, Venayak N, Mahadevan R: **Novel two-stage processes for optimal chemical production in microbes**. *Metabolic Engineering* 2020, **62**:186-197.

33. Foster CJ, Gopalakrishnan S, Antoniewicz MR, Maranas CD: **From Escherichia coli mutant 13C labeling data to a core kinetic model: A kinetic model parameterization pipeline**. *PLOS Computational Biology* 2019, **15**:e1007319.
34. St. John PC, Strutz J, Broadbelt LJ, Tyo KEJ, Bomble YJ: **Bayesian inference of metabolic kinetics from genome-scale multiomics data**. *PLOS Computational Biology* 2019, **15**:e1007424.
35. Islam MM, Schroeder WL, Saha R: **Kinetic modeling of metabolism: Present and future**. *Current Opinion in Systems Biology* 2021, **26**:72-78.
36. Kim OD, Rocha M, Maia P: **A Review of Dynamic Modeling Approaches and Their Application in Computational Strain Optimization for Metabolic Engineering**. *Frontiers in Microbiology* 2018, **9**.
37. Lopatkin AJ, Collins JJ: **Predictive biology: modelling, understanding and harnessing microbial complexity**. *Nature Reviews Microbiology* 2020, **18**:507-520.
38. Hatzimanikatis V, Bailey JE: **MCA Has More to Say**. *Journal of Theoretical Biology* 1996, **182**:233-242.
39. Miskovic L, Hatzimanikatis V: **Production of biofuels and biochemicals: in need of an ORACLE**. *Trends in Biotechnology* 2010, **28**:391-397.
40. Andreozzi S, Miskovic L, Hatzimanikatis V: **iSCHRUNK – In Silico Approach to Characterization and Reduction of Uncertainty in the Kinetic Models of Genome-scale Metabolic Networks**. *Metabolic Engineering* 2016, **33**:158-168.
41. Tokic M, Hatzimanikatis V, Miskovic L: **Large-scale kinetic metabolic models of Pseudomonas putida KT2440 for consistent design of metabolic engineering strategies**. *Biotechnology for Biofuels* 2020, **13**:33.
42. van Rosmalen RP, Smith RW, Martins dos Santos VAP, Fleck C, Suarez-Diez M: **Model reduction of genome-scale metabolic models as a basis for targeted kinetic models**. *Metabolic Engineering* 2021, **64**:74-84.
43. Hameri T, Boldi M-O, Hatzimanikatis V: **Statistical inference in ensemble modeling of cellular metabolism**. *PLOS Computational Biology* 2019, **15**:e1007536.
44. Tsouka S, Ataman M, Hameri T, Miskovic L, Hatzimanikatis V: **Constraint-based metabolic control analysis for rational strain engineering**. *Metabolic Engineering* 2021, **66**:191-203.
45. Henry CS, Broadbelt LJ, Hatzimanikatis V: **Thermodynamics-Based Metabolic Flux Analysis**. *Biophysical Journal* 2007, **92**:1792-1805.
46. Salvy P, Fengos G, Ataman M, Pathier T, Soh KC, Hatzimanikatis V: **pyTFA and matTFA: a Python package and a Matlab toolbox for Thermodynamics-based Flux Analysis**. *Bioinformatics* 2018, **35**:167-169.
47. Shen F, Sun R, Yao J, Li J, Liu Q, Price ND, Liu C, Wang Z: **OptRAM: In-silico strain design via integrative regulatory-metabolic network modeling**. *PLOS Computational Biology* 2019, **15**:e1006835.
48. Dahal S, Yurkovich JT, Xu H, Palsson BO, Yang L: **Synthesizing Systems Biology Knowledge from Omics Using Genome-Scale Models**. *PROTEOMICS* 2020, **20**:1900282.
49. Magazzù G, Zampieri G, Angione C: **Multimodal regularized linear models with flux balance analysis for mechanistic integration of omics data**. *Bioinformatics* 2021.
50. Machado D, Herrgård M: **Systematic Evaluation of Methods for Integration of Transcriptomic Data into Constraint-Based Models of Metabolism**. *PLOS Computational Biology* 2014, **10**:e1003580.
51. Sánchez BJ, Zhang C, Nilsson A, Lahtvee P-J, Kerkhoven EJ, Nielsen J: **Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints**. *Molecular Systems Biology* 2017, **13**:935.

52. Li G, Hu Y, Jan Z, Luo H, Wang H, Zelezniak A, Ji B, Nielsen J: **Bayesian genome scale modelling identifies thermal determinants of yeast metabolism**. *Nature Communications* 2021, **12**:190.
53. Domenzain I, Sánchez B, Anton M, Kerkhoven EJ, Millán-Oropeza A, Henry C, Siewers V, Morrissey JP, Sonnenschein N, Nielsen J: **Reconstruction of a catalogue of genome-scale metabolic models with enzymatic constraints using GECKO 2.0**. *bioRxiv* 2021:2021.2003.2005.433259.
54. Heckmann D, Lloyd CJ, Mih N, Ha Y, Zielinski DC, Haiman ZB, Desouki AA, Lercher MJ, Palsson BO: **Machine learning applied to enzyme turnover numbers reveals protein structural correlates and improves metabolic models**. *Nature Communications* 2018, **9**:5252.
55. Yang L, Yurkovich JT, King ZA, Palsson BO: **Modeling the multi-scale mechanisms of macromolecular resource allocation**. *Current Opinion in Microbiology* 2018, **45**:8-15.
56. Dinh HV, King ZA, Palsson BO, Feist AM: **Identification of growth-coupled production strains considering protein costs and kinetic variability**. *Metabolic Engineering Communications* 2018, **7**:e00080.
57. Yang L, Ebrahim A, Lloyd CJ, Saunders MA, Palsson BO: **DynamicME: dynamic simulation and refinement of integrated models of metabolism and protein expression**. *BMC Systems Biology* 2019, **13**:2.
58. Salvy P, Hatzimanikatis V: **The ETFL formulation allows multi-omics integration in thermodynamics-compliant metabolism and expression models**. *Nature Communications* 2020, **11**:30.
59. Salvy P, Hatzimanikatis V: **Emergence of diauxie as an optimal growth strategy under resource allocation constraints in cellular metabolism**. *Proceedings of the National Academy of Sciences* 2021, **118**:e2013836118.
60. Goelzer A, Fromion V, Scorletti G: **Cell design in bacteria as a convex optimization problem**. *Automatica* 2011, **47**:1210-1218.
61. Bulović A, Fischer S, Dinh M, Golib F, Liebermeister W, Poirier C, Tournier L, Klipp E, Fromion V, Goelzer A: **Automated generation of bacterial resource allocation models**. *Metabolic Engineering* 2019, **55**:12-22.
62. Kavvas ES, Yang L, Monk JM, Heckmann D, Palsson BO: **A biochemically-interpretable machine learning classifier for microbial GWAS**. *Nature communications* 2020, **11**:2580-2580.
63. Wu C, Jiang H, Kalra I, Wang X, Cano M, Maness P, Yu J, Xiong W: **A generalized computational framework to streamline thermodynamics and kinetics analysis of metabolic pathways**. *Metabolic Engineering* 2020, **57**:140-150.
64. Gopalakrishnan S, Dash S, Maranas C: **K-FIT: An accelerated kinetic parameterization algorithm using steady-state fluxomic data**. *Metabolic Engineering* 2020, **61**:197-205.
65. Zhang J, Petersen SD, Radivojevic T, Ramirez A, Pérez-Manríquez A, Abeliuk E, Sánchez BJ, Costello Z, Chen Y, Fero MJ, et al.: **Combining mechanistic and machine learning models for predictive engineering and optimization of tryptophan metabolism**. *Nature Communications* 2020, **11**:4880.
66. Roy S, Radivojevic T, Forrer M, Marti JM, Jonnalagadda V, Backman T, Morrell W, Plahar H, Kim J, Hillson N, et al.: **Multiomics Data Collection, Visualization, and Utilization for Guiding Metabolic Engineering**. *Frontiers in Bioengineering and Biotechnology* 2021, **9**.

Figure


A Constraint-based Models
B Dynamic Models
C Hybrids Models and ML-based approaches
Post-translation modification
Transcriptional regulation
Allosteric regulation
Substrate-level regulation
Vmax
Km
Omics data acquisition
Machine Learning evaluation
Retention time (min)
High-quality engineering predictions