
CHALMERS
UNIVERSITY OF TECHNOLOGY

# Mathematical modeling of proteome constraints within metabolism



Citation for the original published paper (version of record):
Chen, Y., Nielsen, J. (2021). Mathematical modeling of proteome constraints within metabolism. Current Opinion in Systems Biology, 25: 50-56. http://dx.doi.org/10.1016/j.coisb.2021.03.003

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Systems Biology

# Mathematical modeling of proteome constraints within metabolism

Yu Chen[1] and Jens Nielsen[1,2,3]

**Abstract**

Genome-scale metabolic models (GEMs) are widely used to predict phenotypes with the aid of constraint-based modeling. In order to improve the predictive power of these models, there have been many efforts on imposing biological constraints, among which proteome constraints are of particular interest. Here we describe the concept of proteome constraints and review proteome-constrained GEMs, as well as their advantages and applications. In addition, we discuss a key issue in the field, i.e., low coverage of enzyme-specific turnover rates, and subsequently provide a few solutions to solve it. We end with a discussion on the trade-off between model complexity and capability.

**Addresses**

[1] Department of Biology and Biological Engineering, Chalmers University of Technology, SE412 96, Gothenburg, Sweden

[2] Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK2800 Kgs. Lyngby, Denmark

[3] BioInnovation Institute, DK2200, Copenhagen N, Denmark

Corresponding author: Nielsen, Jens (nielsenj@chalmers.se)

**Current Opinion in Systems Biology** 2021, **25**:50–56

This review comes from a themed issue on **Mathematical Modelling**

Edited by **Stacey D. Finley** and **Vassily Hatzimanikatis**

For a complete overview see the Issue and the Editorial

Available online 17 March 2021

**https://doi.org/10.1016/j.coisb.2021.03.003**



**Keywords**

Genome-scale metabolic model, Constraint-based model, Proteome constraint, Turnover rate, $k_{cat}$.

## Introduction

Genome-scale metabolic models (GEMs) are organism-specific knowledge bases that account for genes, proteins, and biochemical reactions, which enable systematic analysis of metabolism [1]. GEMs have been used to predict phenotypes with the aid of optimization algorithms when mathematically converted to constraint-based models [2], where two types of constraints are used, i.e., mass conservation and flux bounds. In order to improve the predictive power, it is common to impose additional constraints, e.g., exchange fluxes. Additionally, an increasing number of efforts have focused on adding biologically meaningful constraints, e.g., proteome constraints.

The proteome constraint relies on the fact that cells have to optimally allocate finite proteome resources to various biological processes due to limited space, e.g., membrane area and cell volume. This is related to the concept of cellular resource allocation, which means that an increase in the requirement of proteome resources of an enzyme or a pathway would be a trade-off for other functions. Experimental evidence indicates that resource re-allocation could be an effective strategy for various organisms in response to perturbations [3], e.g., nutrient shift [4] and growth shift [5,6], which demonstrates the biological significance of proteome constraints. Accordingly, this suggests that proteome constraints could be a valuable addition to GEMs to improve model predictions.

Here, we describe how proteome constraints can be integrated into GEMs and review proteome-constrained GEMs (pcGEMs) that were constructed in recent years. Subsequently, we discuss one of the current challenges on implementing proteome constraints, i.e., unavailability of turnover rates, and provide potential strategies to deal with the issue.

## Frameworks to integrate proteome constraints with GEMs

Over the past years, different frameworks have been proposed to integrate proteome constraints into GEMs with two directions, i.e., coarse-grained and fine-grained frameworks [7]. The coarse-grained frameworks, e.g., MOMENT [8] and GECKO [9], impose phenomenological constraints, e.g., enzyme concentrations and activities, on metabolic enzymes (Figure 1a), which do not change the gene number of the original GEM. These models are sometimes referred to as enzyme-constrained GEMs (ecGEMs). In contrast, the fine-grained frameworks, e.g., ETFL [10], ME-model [11,12] and RBA [13], introduce additional biological processes, especially protein expression processes (Figure 1a), with very detailed descriptions, which can expand the scope of the original GEM and results in a self-replicating system [14]. Fine-grained pcGEMs are

therefore able to predict extra biological layers beyond metabolism when more biological data, e.g., ribosome efficiencies, are available [11,12].

Despite differences of the frameworks, a commonly imposed constraint is on metabolic flux (Figure 1a), formulated as enzyme kinetics $v \leq k_{cat} \cdot E$, in which $v$ is metabolic rate, $E$ is the concentration of the enzyme that catalyzes the reaction, and $k_{cat}$ is the turnover rate of the enzyme. This means that each flux is constrained by the catalytic capability of the enzyme and the availability of the enzyme. In addition to this, fine-grained pcGEMs can constrain protein synthesis rates based on catalytic capabilities of machinery, e.g., ribosomes, and their abundances (Figure 1a). Figure 1b introduces the common equations across various frameworks when proteome constraints are used. It should be noted that proteome constraints will be active only after imposing upper limits on individual or total enzyme abundances (Equation (4) in Figure 1b). The total limit can also be imposed on enzymes that compete within shared compartments, e.g., membrane enzymes and mitochondrial enzymes. The upper bound for the total enzymes can be estimated by summing up the abundances of all involved enzymes based on proteomics data or databases such as the PaxDb [15]. Due to the inclusion of protein expression processes, fine-grained pcGEMs have additional equations related to the maximal catalytic rate of machinery in the processes (Equation (5) in Figure 1b), e g., ribosomes and chaperons, and balance of the components in the processes (Equation (6) in Figure 1b), as well as a balanced equation for each protein.

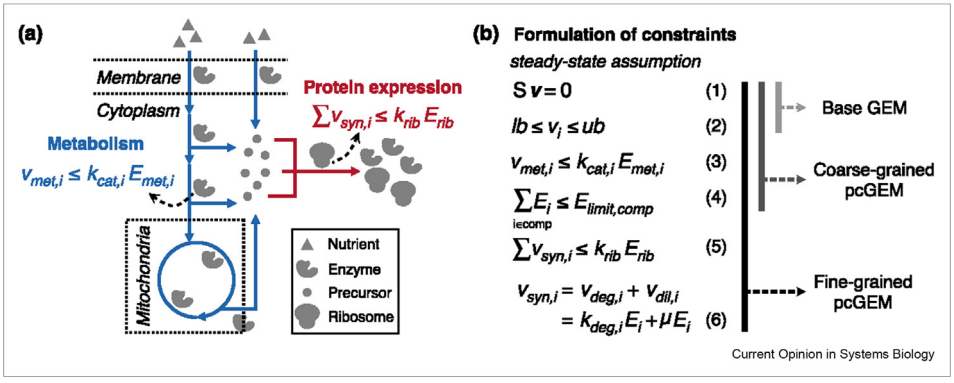
## Genome-scale models integrated with proteome constraints

Over the past decade, a number of pcGEMs have been developed using various frameworks (Table 1). All the models were derived from the corresponding base GEM of the organism, including bacteria, fungi, and human cell lines, suggesting that proteome constraints are relatively easy to integrate when a base GEM is available.

With the integration of proteome constraints in either a coarse- or fine-grained formalism, model predictions have been improved. In particular, variabilities of simulated metabolic fluxes of pcGEMs are much lower than those of the corresponding base GEMs [9,11,18,19,22,23], which allows more accurate prediction of fluxes [12,18,23]. Additionally, pcGEMs are capable of predicting phenotypes that base GEMs cannot unless some specific rates are constrained *a priori*, e.g., maximal growth at unlimited conditions [9,12], overflow metabolism in *Escherichia coli* [12] and *Clostridium ljungdahlii* [33], the Crabtree effect in *Saccharomyces cerevisiae* [9], and metabolic shift of arginine catabolism in *Lactococcus lactis* [27].

pcGEMs can also serve as frameworks for integrated analysis of transcriptomics and proteomics data [22,38].

**Figure 1**




Overview of proteome constraints. **(a)** Illustration of the integration of proteome constraints into GEMs based on coarse-grained and fine-grained approaches. The former only imposes phenomenological constraints on enzymatic reactions while the latter adds protein expression reactions. The rate ($v_{met,i}$) of a metabolic reaction (blue arrow) is constrained by the turnover rate ($k_{cat,i}$) and abundance ($E_{met,i}$) of the enzyme. The rate of a protein expression reaction (red arrow) is also constrained by such a type of equation. For example, the total rates of all protein synthesis reactions ($\Sigma v_{syn,i}$) are constrained by the catalytic rate ($k_{rib}$) and the abundance ($E_{rib}$) of ribosome, as well as the production rate of all precursors, e.g., amino acids and energy. **(b)** Formulation of constraints. Equation (1) is the flux balance constraint, in which $S$ is stoichiometric matrix and $\boldsymbol{v}$ is a vector of fluxes. Equation (2) imposes lower and upper bounds on each flux. Equation (3) imposes phenomenological constraints on enzymatic reactions. Equation (4) imposes upper limits on the total abundances of enzymes in certain compartments. Equation (5) imposes constraints on the reaction rate of protein expression processes by catalytic rate and abundance of the machinery. Equation (6) represents the mass balance of components in protein expression processes, e.g., enzymes and ribosomes. At a steady-state, the synthesis rate of a component ($v_{syn,i}$) is equal to degradation rate ($v_{deg,i}$) plus dilution rate ($v_{dil,i}$), in which degradation rate is equal to degradation constant ($k_{deg,i}$) multiplies abundance of the component ($E_i$) while dilution rate depends on growth rate ($\mu$) and also the abundance.

**Table 1**

**Published genome-scale models integrated with proteome constraints.**

| Model | Organism | Type | Turnover rate of metabolic enzyme | Year | Ref. |
|---|---|---|---|---|---|
| ec_iYO844 | *Bacillus subtilis* | Coarse-grained (GECKO) | Only 17 reactions were assigned with turnover rates, which were manually collected from the BRENDA [16]and SABIO-RK [17] databases, and literature. | 2019 | [18] |
| ecYeast8[a] | *Saccharomyces cerevisiae* | Coarse-grained (GECKO) | Turnover rates were automatically retrieved from the BRENDA database using the GECKO toolbox or manually collected from the literature. | 2019 | [19] |
| ec_iML1515 | *Escherichia coli* | Coarse-grained (GECKO) | Turnover rates were automatically retrieved from the BRENDA database using the GECKO toolbox or manually collected from the SABIO-RK database and the literature. | 2020 | [20] |
| ec-iBag597 | *Bacillus coagulans* | Coarse-grained (GECKO) | Turnover rates were automatically retrieved from the BRENDA database using the GECKO toolbox. | 2020 | [21] |
| EcSco-GEM | *Streptomyces coelicolor* | Coarse-grained (GECKO) | Turnover rates were automatically retrieved from the BRENDA database using the GECKO toolbox or manually collected from the literature. | 2020 | [22] |
| Cell line-specific ecGEMs | *Homo sapiens* | Coarse-grained (GECKO) | Turnover rates were automatically retrieved from the BRENDA database using the GECKO toolbox. | 2020 | [23] |
| *E. coli* MOMENT model | *Escherichia coli* | Coarse-grained (MOMENT) | Turnover rates were retrieved from the BRENDA and SABIO-RK databases. | 2012 | [8] |
| *E. coli* MOMENT model | *Escherichia coli* | Coarse-grained (MOMENT) | Turnover rates were retrieved from the BRENDA and SABIO-RK databases. Apparent turnover rates [24] were used as additional input. | 2020 | [25] |
| *S. elongatus* model | *Synechococcus elongatus* | Fine-grained | Turnover rates were retrieved from the BRENDA database or manually collected from the literature. | 2017 | [26] |
| pcLactis | *Lactococcus lactis* | Fine-grained | Turnover rates were automatically retrieved from the BRENDA database using the GECKO toolbox or manually collected from the literature. | 2020 | [27] |
| *E. coli* model | *Escherichia coli* | Fine-grained | Turnover rates were retrieved from the BRENDA and SABIO-RK databases. | 2021 | [28] |
| *E. coli* ETFL model | *Escherichia coli* | Fine-grained (ETFL) | Turnover rates were obtained from the study [29] and SABIO-RK database. | 2020 | [10] |
| yETFL | *Saccharomyces cerevisiae* | Fine-grained (ETFL) | Turnover rates were automatically retrieved from the BRENDA database using the GECKO toolbox. | 2021 | [30] |
| *T. maritima* ME | *Thermotoga maritima* | Fine-grained (ME-model) | Turnover rates were globally assumed to 15 reactions per second per protein complex. | 2012 | [11] |
| iJL1678b-ME[b] | *Escherichia coli* | Fine-grained (ME-model) | Effective catalytic rates obtained from the previous study [31] were adopted, in which an iterative workflow that integrates proteomics data with the ME-model was used to infer the effective catalytic rates. | 2018 | [32] |
| iJL965-ME | *Clostridium ljungdahlii* | Fine-grained (ME-model) | Turnover rates were globally assumed to 25 $s^{-1}$. | 2019 | [33] |
| *B. subtilis* RBA model | *Bacillus subtilis* | Fine-grained (RBA) | Apparent catalytic rates inferred using measured data [13] were adopted. | 2015 | [13] |
| *E. coli* RBA model | *Escherichia coli* | Fine-grained (RBA) | Apparent catalytic rates inferred using measured data [34,35] were adopted. | 2019 | [36] |

[a] This is the latest ecYeast model, with a previous version ecYeast7 published in connection with the development of the GECKO toolbox [9].

[b] This is the latest *E. coli* ME-model. Previous versions have been reviewed in Ref. [37].

Additionally, pcGEMs have provided mechanistic insights into cellular resource allocation under environmental changes [39–41]. Furthermore, fine-grained pcGEMs can predict more biological parameters, e.g., growth-condition dependent biomass composition [11,42] and translation machinery content [11,12], which cannot be predicted by coarse-grained pcGEMs. The coarse-grained pcGEMs are, however, more suitable for the prediction of metabolic engineering targets [18,20] due to their simple formalism, which makes simulation cheaper than for fine-grained pcGEMs [43].

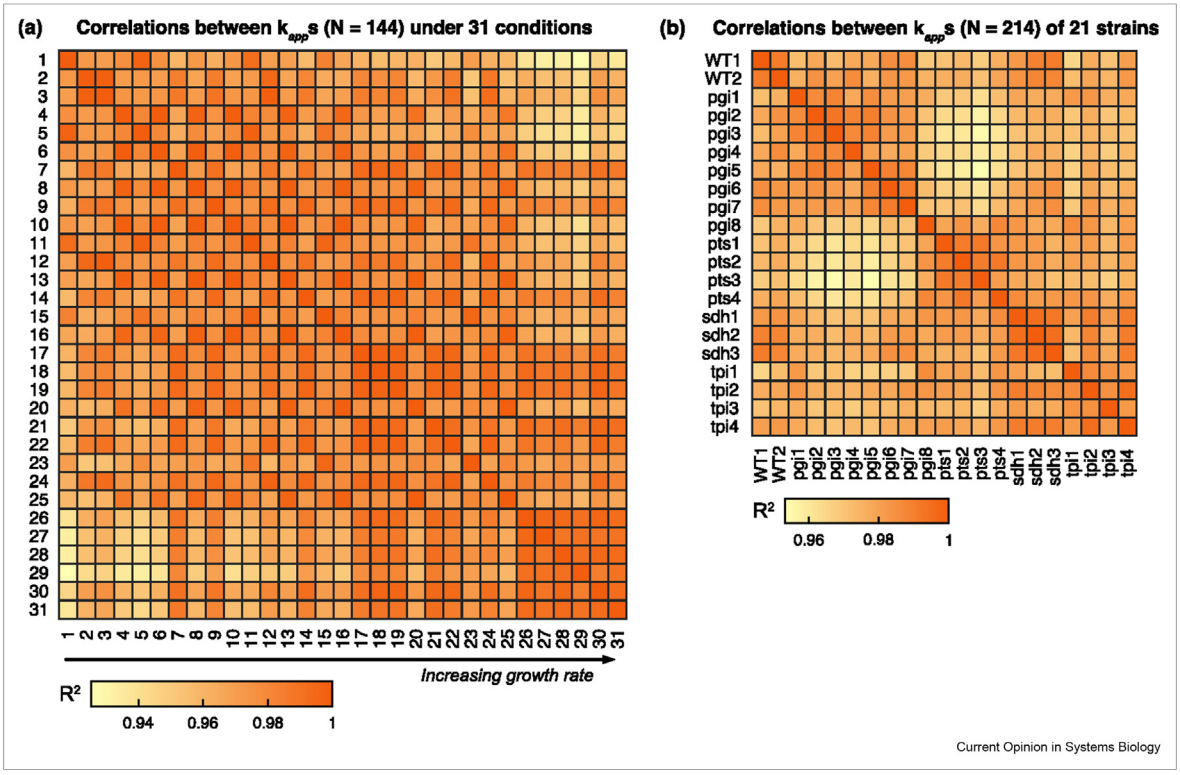
## Current challenge: turnover rates are essential parameters

Despite improved predictions and applications, issues remain with constructing and utilizing pcGEMs [37,44], of which defining the turnover rates is key as they are essential parameters in any type of pcGEMs but are assigned in distinct manners (Table 1). The simplest one is to assume a global turnover rate for all enzymes, which is the case in the first version of the *E. coli* ME-model [12], the *Thermotoga maritima* ME-model [11], and even the very recent *C. ljungdahlii* ME-model [33]. Although these models succeed in capturing phenotypes such as overflow metabolism, the predictions deviated from experimental measurements [12]. This could be circumvented, as done in other pcGEMs (Table 1), by adopting organism-specific and enzyme-specific turnover rates, which can be retrieved in databases such as BRENDA and SABIO-RK, and literature. Given that the retrieved turnover rates were mostly measured *in vitro*, it may be questionable whether *in vitro* turnover rates can be used in pcGEMs to simulate *in vivo* fluxes, but for *E. coli* there is a good correlation between *in vitro* and *in vivo* enzyme kinetics [29]. However, measured turnover rates normally only cover a small fraction of metabolic reactions even in well-studied organisms [45]. Here we discuss a few solutions that could be potentially helpful to address this challenge.

1. Reduced models circumvent uncertain turnover rates

When it is impossible to assign turnover rates to all enzymes in a genome-scale, model reduction can be an effective way to improve the coverage of high-quality turnover rates. A recent study proposed such a concept to reduce GEMs to only include reactions of

**Figure 2**




Correlation analysis on log10 transformed *in vivo* turnover rates. **(a)** Dataset is from. 31 various growth conditions of *E. coli* [29]. **(b)** Dataset is from 21 various *E. coli* strains [47]. The codes are available here (https://github.com/Yu-sysbio/correlate_kapp), where the information of numbered conditions in panel **(a)** is also available.

energy metabolism so that most of the reactions have manually curated turnover rates, which can capture overflow metabolism of *E. coli* and the Crabtree effect of *S. cerevisiae* [46]. This indicates that the energy metabolism is sufficient to interpret key phenotypes. The concept of reduced models has also been applied to investigate metabolic shifts in *L. lactis*, and it was found that the shift in arginine catabolism could abide by the resource allocation principles [27]. Therefore, we argue that genome-scale models might have more uncertainties and be unnecessary for specific studies where reduced models, that focus on particular pathways or organelles, e.g., mitochondrion, may be better suited.

2. Proteomics data enable estimation of *in vivo* turnover rates on a large scale

It is possible to infer *in vivo* turnover rates, also called apparent catalytic rates ($k_{app}$s), from genome-wide absolute proteomics and fluxomics data. These can be calculated simply by dividing the flux through each enzyme by the abundance of the enzyme [45]. The *in vivo* turnover rates have been estimated for *E. coli* under diverse growth conditions [29] and for various *E. coli* strains [47]. The inclusion of genome-wide data can improve the coverage of turnover rates of pcGEMs and even result in condition- or strain-specific pcGEMs. By analyzing the two datasets, we can see strong correlations of the *in vivo* turnover rates across conditions and strains (Figure 2). This means that the *in vivo* turnover rates estimated from one condition or strain could be applicable to others in case data is unavailable. However, it should be noted that this has not been observed in other organisms especially eukaryotes that have more complex regulation. Another approach that also uses proteomics data to infer *in vivo* turnover rates implements an iterative workflow within a ME-model of *E. coli* by minimizing the difference between simulated and measured proteomics data [31]. Altogether, these approaches show the possibility to estimate genome-wide *in vivo* turnover rates, which can be directly adopted in pcGEMs.

As *in vivo* turnover rates become available another question arises: whether *in vitro* or *in vivo* turnover rates outperform in pcGEM simulations? It was shown that using maximal *in vivo* turnover rates predicted proteomics data much better [47], and we, therefore, encourage adopting *in vivo* turnover rates in pcGEMs. However, *in vivo* data are currently only available for *E. coli* and *Bacillus subtilis* [13] under specific conditions due to the limited availability of absolute proteomics data. In most cases with the lack of *in vivo* enzyme kinetics, *in vitro* data have to be used in pcGEMs, which are in principle not applicable to cases with changed *in vivo* turnover rates. This is therefore a general weakness of current pcGEMs.

3. Machine learning for predicting turnover rates

Although data-driven estimation of *in vivo* turnover rates improves the coverage, there are still a lot of enzymes that cannot be estimated due to coverage issues of proteomics experiments. For example, only hundreds of turnover rates have been estimated compared with thousands of reactions in the model [47]. To improve further the coverage, machine learning approaches [48,49] have been utilized [47], which enable predictions of genome-scale turnover rates based on enzyme biochemistry, structure, and network context [50]. With the integration of predicted turnover rates, pcGEMs predict unseen proteomics data with higher precision [47], which demonstrates that machine learning can effectively solve the coverage issue of turnover rates in pcGEMs.

## Conclusion

As more kinetic parameters become available, we believe that an increasing number of pcGEMs will be constructed in the future. A key direction will thus focus on the trade-off between model complexity and predictive capability. From our overview of published pcGEMs (Table 1), we find that coarse-grained pcGEMs, especially GECKO models have been adopted by various research groups, probably because they are easier to build. In contrast, fine-grained pcGEMs do not spread widely in the community, even though toolboxes for automated construction have been provided, e.g., COBRAme [32] and RBApy [36]. Also, recent efforts focus on the simplification of pcGEMs, including the sMOMENT framework [25] and simplified proteome-constrained model of *E. coli* [28]. Altogether, simple models seem to be gaining the most attention. However, fine-grained models are valuable due to the established connections between various biological processes, which can serve as mathematical frameworks for cross-evaluating various types of data. Recently, a whole-cell model of *E. coli* was developed [51], which includes even more biological processes than fine-grained pcGEM. The model is able to link heterogeneous data and therefore confirms that most of the data are cross-consistent.

## Conflict of interest statement

Nothing declared.

## Acknowledgements

This work was supported by the European Union's Horizon 2020 research and innovation program under Grant Agreement No. 686070 and the Novo Nordisk Foundation (grant no. NNF10CC1016517). The authors also acknowledge funding from the Knut and Alice Wallenberg Foundation.

## References

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

• of special interest
•• of outstanding interest


1. Kim WJ, Kim HU, Lee SY: **Current state and applications of microbial genome-scale metabolic models**. *Curr Opin Struct Biol* 2017, **2**:10–18.

2. Bordbar A, Monk JM, King ZA, Palsson BO: **Constraint-based models predict metabolic and associated cellular functions**. *Nat Rev Genet* 2014, **15**:107–120.

3. Basan M: **Resource allocation and metabolism: the search for governing principles**. *Curr Opin Microbiol* 2018, **45**:77–83.

4. Erickson DW, Schink SJ, Patsalo V, Williamson JR, Gerland U, Hwa T: **A global resource allocation strategy governs growth transition kinetics of *Escherichia coli***. *Nature* 2017, **551**: 119–123.

5. Metzl-Raz E, Kafri M, Yaakov G, Soifer I, Gurvich Y, Barkai N: **Principles of cellular resource allocation revealed by condition-dependent proteome profiling**. *Elife* 2017, **6**.

6. Zavřel T, Faizi M, Loureiro C, Poschmann G, Stühler K, Sinetova M, Zorina A, Steuer R, Červený J: **Quantitative insights into the cyanobacterial cell economy**. *Elife* 2019, **8**.

7. Yang L, Yurkovich JT, King ZA, Palsson BO: **Modeling the multi-scale mechanisms of macromolecular resource allocation**. *Curr Opin Microbiol* 2018, **45**:8–15.

8. Adadi R, Volkmer B, Milo R, Heinemann M, Shlomi T: **Prediction of microbial growth rate versus biomass yield by a metabolic network with kinetic parameters**. *PLoS Comput Biol* 2012, **8**, e1002575.

9. 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**. *Mol Syst Biol* 2017, **13**:935.

10. • Salvy P, Hatzimanikatis V: **The ETFL formulation allows multi-omics integration in thermodynamics-compliant metabolism and expression models**. *Nat Commun* 2020, **11**:30.
This study proposes a formulation called ETFL to integrate expression, thermodynamics, and growth-dependent variables into metabolism.

11. Lerman JA, Hyduke DR, Latif H, Portnoy VA, Lewis NE, Orth JD, Schrimpe-Rutledge AC, Smith RD, Adkins JN, Zengler K, *et al.*: **In silico method for modelling metabolism and gene product expression at genome scale**. *Nat Commun* 2012, **3**:929.

12. O'Brien EJ, Lerman JA, Chang RL, Hyduke DR, Palsson B: **Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction**. *Mol Syst Biol* 2013, **9**:693.

13. Goelzer A, Muntel J, Chubukov V, Jules M, Prestel E, Nölker R, Mariadassou M, Aymerich S, Hecker M, Noirot P, *et al.*: **Quantitative prediction of genome-wide resource allocation in bacteria**. *Metab Eng* 2015, **32**:232–243.

14. Molenaar D, van Berlo R, de Ridder D, Teusink B: **Shifts in growth strategies reflect tradeoffs in cellular economics**. *Mol Syst Biol* 2009, **5**:323.

15. Wang M, Herrmann CJ, Simonovic M, Szklarczyk D, von Mering C: **Version 4.0 of PaxDb: protein abundance data, integrated across model organisms, tissues, and cell-lines**. *Proteomics* 2015, **15**:3163–3168.

16. Jeske L, Placzek S, Schomburg I, Chang A, Schomburg D: **BRENDA in 2019: a European ELIXIR core data resource**. *Nucleic Acids Res* 2019, **47**:D542–D549.

17. Wittig U, Rey M, Weidemann A, Kania R, Müller W: **SABIO-RK: an updated resource for manually curated biochemical reaction kinetics**. *Nucleic Acids Res* 2018, **46**:D656–D660.

18. • Massaiu I, Pasotti L, Sonnenschein N, Rama E, Cavaletti M, Magni P, Calvio C, Herrgård MJ: **Integration of enzymatic data in *Bacillus subtilis* genome-scale metabolic model improves phenotype predictions and enables in silico design of poly-γ-glutamic acid production strains**. *Microb Cell Fact* 2019, **18**:3.
This study presents an ecGEM of *B. subtilis* and shows improved predictions compared with the base GEM. Additionally, the ecGEM enables the identification of gene deletion targets to optimize the biosynthesis of the poly-γ-glutamic acid (γ-PGA) polymer.

19. Lu H, Li F, Sánchez BJ, Zhu Z, Li G, Domenzain I, Marcišauskas S, Anton PM, Lappa D, Lieven C, *et al.*: **A consensus *S. cerevisiae* metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism**. *Nat Commun* 2019, **10**:1–13.

20. • Ye C, Luo Q, Guo L, Gao C, Xu N, Zhang L, Liu L, Chen X: **Improving lysine production through construction of an *Escherichia coli* enzyme-constrained model**. *Biotechnol Bioeng* 2020, **117**:3533–3544.
This study presents an ecGEM of *E. coli* and uses the model to improve lysine production.

21. Chen Y, Sun Y, Liu Z, Dong F, Li Y, Wang Y: **Genome-scale modeling for *Bacillus coagulans* to understand the metabolic characteristics**. *Biotechnol Bioeng* 2020, **117**:3545–3558.

22. Sulheim S, Kumelj T, van Dissel D, Salehzadeh-Yazdi A, Du C, van Wezel GP, Nieselt K, Almaas E, Wentzel A, Kerkhoven EJ: **Enzyme-constrained models and omics analysis of *Streptomyces coelicolor* reveal metabolic changes that enhance heterologous production**. *iScience* 2020, **23**:101525.

23. •• Robinson JL, Kocabaş P, Wang H, Cholley PE, Cook D, Nilsson A, Anton M, Ferreira R, Domenzain I, Billa V, *et al.*: **An atlas of human metabolism**. *Sci Signal* 2020, **13**:1482.
This study presents the first ecGEMs for human cells and uses the models for improved prediction of gene essentiality in different cancer cell lines.

24. Valgepea K, Adamberg K, Seiman A, Vilu R: ***Escherichia coli* achieves faster growth by increasing catalytic and translation rates of proteins**. *Mol Biosyst* 2013, **9**:2344.

25. • Bekiaris PS, Klamt S: **Automatic construction of metabolic models with enzyme constraints**. *BMC Bioinf* 2020, **21**:19.
This study develops the MOMENT approach and the AutoPACMEN toolbox, which pave the way for the simplified construction and analysis of ecGEMs.

26. Reimers A-M, Knoop H, Bockmayr A, Steuer R: **Cellular trade-offs and optimal resource allocation during cyanobacterial diurnal growth**. *Proc Natl Acad Sci U S A* 2017, https://doi.org/10.1073/pnas.1617508114.

27. • Chen Y, van Pelt-KleinJan E, van Olst B, Douwenga S, Boeren S, Bachmann H, Molenaar D, Nielsen J, Teusink B: **Proteome constraints reveal targets for improving microbial fitness in nutrient-rich environments**. *Mol Syst Biol* 2021 Apr, **17**, e10093, https://doi.org/10.15252/msb.202010093. PMID: 33821549.
This study develops a fine-grained pcGEM of *L. lactis* together with a reduced model and uses the models to predict active constraints on growth, including glucose transporter and arginine catabolism pathway. These two constraints are validated based on comparative proteomics analysis between wildtype and evolved strains.

28. Grigaitis P, Olivier BG, Fiedler T, Teusink B, Kummer U, Veith N: **Protein cost allocation explains metabolic strategies in *Escherichia coli***. *J Biotechnol* 2021, **327**: 54–63.

29. Davidi D, Noor E, Liebermeister W, Bar-Even A, Flamholz A, Tummler K, Barenholz U, Goldenfeld M, Shlomi T, Milo R: **Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro kcat measurements**. *Proc Natl Acad Sci U S A* 2016, **113**:3401–3406.

30. Oftadeh O, Salvy P, Masid M, Curvat M, Miskovic L, Hatzimanikatis V: **A genome-scale metabolic model of *Saccharomyces cerevisiae* that integrates expression constraints and reaction thermodynamics**. *bioRxiv* 2021, https://doi.org/10.1101/2021.02.17.431671.

31. Ebrahim A, Brunk E, Tan J, O'Brien EJ, Kim D, Szubin R, Lerman JA, Lechner A, Sastry A, Bordbar A, *et al.*: **Multi-omic data integration enables discovery of hidden biological regularities**. *Nat Commun* 2016, **7**:13091.

32. Lloyd CJ, Ebrahim A, Yang L, King ZA, Catoiu E, O'Brien EJ, Liu JK, Palsson BO, COBRAme: **A computational framework for genome-scale models of metabolism and gene expression**. *PLoS Comput Biol* 2018, **14**, e1006302.

33. Liu JK, Lloyd C, Al-Bassam MM, Ebrahim A, Kim J-N, Olson C, Aksenov A, Dorrestein P, Zengler K: **Predicting proteome allocation, overflow metabolism, and metal requirements in a model acetogen**. *PLoS Comput Biol* 2019, **15**, e1006848.

34. Schmidt A, Kochanowski K, Vedelaar S, Ahrné E, Volkmer B, Callipo L, Knoops K, Bauer M, Aebersold R, Heinemann M: **The quantitative and condition-dependent *Escherichia coli* proteome**. *Nat Biotechnol* 2016, **34**:104–110.

35. Haverkorn van Rijsewijk BRB, Nanchen A, Nallet S, Kleijn RJ, Sauer U: **Large-scale 13 C-flux analysis reveals distinct transcriptional control of respiratory and fermentative metabolism in *Escherichia coli***. *Mol Syst Biol* 2011, **7**:477.

36. 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**. *Metab Eng* 2019, **55**:12–22.

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

38. Du B, Yang L, Lloyd CJ, Fang X, Palsson BO: **Genome-scale model of metabolism and gene expression provides a multi-scale description of acid stress responses in *Escherichia coli***. *PLoS Comput Biol* 2019, **15**, e1007525.

39. Chen K, Gao Y, Mih N, O'Brien EJ, Yang L, Palsson BO: **Thermosensitivity of growth is determined by chaperone-mediated proteome reallocation**. *Proc Natl Acad Sci U S A* 2017, **114**:11548–11553.

40. Chen K, Anand A, Olson C, Sandberg TE, Gao Y, Mih N, Palsson BO: **Bacterial fitness landscapes stratify based on proteome allocation associated with discrete aero-types**. *PLoS Comput Biol* 2021, **17**, e1008596.

41. Yang L, Mih N, Anand A, Park JH, Tan J, Yurkovich JT, Monk JM, Lloyd CJ, Sandberg TE, Seo SW, *et al.*: **Cellular responses to reactive oxygen species are predicted from molecular mechanisms**. *Proc Natl Acad Sci U S A* 2019, **116**:14368–14373.

42. Lloyd C, Monk J, Yang L, Ebrahim A, Palsson B: **Computation of condition-dependent proteome allocation reveals variability in the macro and micro nutrient requirements for growth**. *bioRxiv* 2020, https://doi.org/10.1101/2020.03.23.003236.

43. Dinh HV, King ZA, Palsson BO, Feist AM: **Identification of growth-coupled production strains considering protein costs and kinetic variability**. *Metab Eng Commun* 2018, **7**, e00080.

44. 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**. *Metab Eng* 2020, https://doi.org/10.1016/j.ymben.2020.11.013.

45. Davidi D, Milo R: **Lessons on enzyme kinetics from quantitative proteomics**. *Curr Opin Biotechnol* 2017, **46**:81–89.

46. •• Chen Y, Nielsen J: **Energy metabolism controls phenotypes by protein efficiency and allocation**. *Proc Natl Acad Sci U S A* 2019, **116**:17592–17597.

This study proposes a concept to reduce GEMs to focus on energy metabolism pathways. Protein costs of independent pathways can be estimated with manually curated turnover rates. The reduced models of *E. coli* and *S. cerevisiae* are able to interpret metabolic shifts and increased growth, and predict target enzymes for improving growth.

47. •• Heckmann D, Campeau A, Lloyd CJ, Phaneuf PV, Hefner Y, Carrillo-Terrazas M, Feist AM, Gonzalez DJ, Palsson BO: **Kinetic profiling of metabolic specialists demonstrates stability and consistency of in vivo enzyme turnover numbers**. *Proc Natl Acad Sci U S A* 2020, **117**:23182–23190.

This study estimates *in vivo* turnover rates for specialist *E. coli* strains using proteomics and fluxomics data, which are robust against genetic perturbations. In addition, by combining the data-driven estimation and machine learning prediction of turnover rates, pcGEMs are able to predict better the proteomics data.

48. Kim GB, Kim WJ, Kim HU, Lee SY: **Machine learning applications in systems metabolic engineering**. *Curr Opin Biotechnol* 2020, **64**:1–9.

49. Rana P, Berry C, Ghosh P, Fong SS: **Recent advances on constraint-based models by integrating machine learning**. *Curr Opin Biotechnol* 2020, **64**:85–91.

50. 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**. *Nat Commun* 2018, **9**:5252.

51. •• Macklin DN, Ahn-Horst TA, Choi H, Ruggero NA, Carrera J, Mason JC, Sun G, Agmon E, Defelice MM, Maayan I, *et al.*: **Simultaneous cross-evaluation of heterogeneous *E. coli* datasets via mechanistic simulation**. *Science (80-)* 2020:369.

This study presents a whole-cell model of *E. coli* and demonstrates that the model allows for cross-evaluation of heterogeneous datasets.