CHALMERS
UNIVERSITY OF TECHNOLOGY

Advances in constraint-based models: methods for improved predictive power based on resource allocation constraints

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Microbiology

Advances in constraint-based models: methods for improved predictive power based on resource allocation constraints

Eduard J Kerkhoven [1,2]

The concept of metabolic models with resource allocation constraints has been around for over a decade and has clear advantages even when implementation is relatively rudimentary. Nonetheless, the number of organisms for which such a model is reconstructed is low. Various approaches exist, from coarse-grained consideration of enzyme usage to fine-grained description of protein translation. These approaches are reviewed here, with a particular focus on user-friendly solutions that can introduce resource allocation constraints to metabolic models of any organism. The availability of k_{cat} data is a major hurdle, where recent advances might help to fill in the numerous gaps that exist for this data, especially for nonmodel organisms.

Addresses
[1] Department of Biology and Biological Engineering, Chalmers University of Technology, Kemivägen 10, SE412 96 Gothenburg, Sweden
[2] Novo Nordisk Foundation Center for Biosustainability, Chalmers University of Technology, Kemivägen 10, SE-412 96 Gothenburg, Sweden

Corresponding author: Eduard J Kerkhoven (eduardk@chalmers.se)

Current Opinion in Microbiology 2022, **68**:102168

This review comes from a themed issue on **Systems and Synthetic Biology**

Edited by **Rainer Breitling** and **Eriko Takano**

For complete overview of the section, please refer to the article collection, "Systems and Synthetic Biology"

Available online 9th June 2022

https://doi.org/10.1016/j.mib.2022.102168

Introduction

Investigations into metabolism can be aided by the use of genome-scale metabolic models (GEMs), which are organism-specific descriptions of complete metabolic networks, linking all metabolic reactions to their respective enzymes. These GEMs can simulate metabolic flux distributions by optimization of an objective function that describes the perceived cellular objective that propels metabolism, by flux balance analysis (FBA) [1].

Elementary to GEM simulations is that metabolism (as well as life in general) operates under myriad of constraints. These include so-called 'hard' constraints that remain unchanged as they abide by the laws of physics (e.g. conservation of mass and energy, thermodynamics), in addition to 'soft' constraints that can vary not only by organism, environmental condition, and the state of the cell (e.g. nutrient uptake rate, biomass composition), but can also change through processes like evolution or changes in gene expression.

A major limitation in the predictive power of these conventional GEMs is that they ignore that the capacity of a cell to support a metabolic flux is constrained by its resource allocation, chiefly as most metabolic reactions are catalyzed by enzymes. The synthesis of enzymes (i.e. proteins) is resource- and energy-expensive, their catalytic capacities are limited by their often modest kinetics, plus the quantity of enzymes is space-constraint, such that stringency in resource allocation is vital for optimal cell growth. The modeling problem of resource allocation can be narrowly defined by only considering protein allocation: "given a certain *budget*, what is the best way to distribute it", where budget refers to the total cellular protein level that is distributed over all its constituent proteins. This problem can be extended by not only considering protein allocation but also the resources it takes to produce proteins, where resources refer to the metabolic and energetic costs of protein synthesis. Regardless of whether one considers only the *budget* or also the *resources*, applying such constraints in a model of metabolism not only reduces simulated flux distributions to those that are most economic. It also limits the phenotypes that the model can simulate, where both contribute to more realistic results. Such models have already found numerous applications in e.g. unraveling the underlying mechanisms for observed metabolic phenotypes and the prediction of strain optimization strategies, as reviewed in more detail elsewhere [2].

The simplest approach to consider the economics of resource allocation assumes that cells aim to minimize the number of active fluxes to yield the most efficient flux distribution. Parsimonious FBA [3] is an example of

a computational approach that can find such minimal flux distributions by removing loops and thereby partially alleviates the protein allocation problem, but more advanced approaches have since been developed. Moreover, various approaches reach further and more explicitly consider the resource allocation constraints that affect metabolism. While most of these approaches are introduced with a proof-of-principle application in a model organism (often *Escherichia coli* or *Saccharomyces cerevisiae*), their valuable characteristics could be of benefit for models of many other organisms. Hence, here the field of resource allocation models since 2020 is reviewed with a particular focus on (i) approaches with software that allow for relatively straightforward extension of *any* GEM for *any* organism; (ii) recent advances in k_{cat} value determination that prominently affect model reliability, and disproportionally affects nonmodel organisms.

Acknowledging the workhorses

Principles of enzyme-constrained models

The protein allocation that is arguably closest to metabolism is the selection and quantity of enzymes to be synthesized. Efficient pathways are those involving enzymes with an optimal combination of low molecular weight and high catalytic capacity. As such, metabolic fluxes are constrained by:

$$v_{max,i} \leq k_{cat,i} \cdot [E]_i \tag{1}$$

$$\sum [E]_i \cdot MW_i \leq C \left[\frac{g_{\text{enzyme}}}{g_{\text{DCW}}} \right] \tag{2}$$

where the maximum rate of reaction i is a product of the catalytic capacity and concentration of its corresponding enzyme (assuming absence of isoenzymes), while the total enzyme content C (in gram enzyme per gram dry cell weight) is an overarching constraint and considers the molecular weight MW of each enzyme.

In the earliest approach to implement these constraints—FBAwMC (FBA with molecular crowding)—Eq. (1) was not explicitly considered [4]. Rather, an overall *crowding coefficient* was assigned, representing an amalgamation of k_{cat} values, enzyme molecular weights, substrate and product concentrations, and any other factors that modify enzyme activities. This approach has later been extended with thermodynamic penalties for reversible reactions [5], and dynamic assignment of the proteome to four sections: ribosome; biosynthetic enzymes; carbon intake; housekeeping [6]. However, the explicit incorporation of both Eqs. (1) and (2) in so-called enzyme-constrained models (ec-models) has recently been the more dominant approach, after a first implementation in 2011 [7].

Enzyme-constrained model reconstruction workflows

As an expansion of conventional GEMs, ec-models consider the protein allocation (or *budget*) problem by requiring each enzyme-catalyzed reaction to use an enzyme-specific pseudometabolite [8]. The stoichiometric coefficient of enzyme usage in these reactions is defined as $1/k_{cat}$, such that fast enzymes with high k_{cat} values require less usage. The molecular weight of the protein is considered in when drawing the usages of all enzymes from either an overall pool of total protein, or alternatively from enzyme levels measured by absolute proteomics. New ec-model reconstruction approaches are often presented as a software solution (Table 1). This not only allows for replicating the model accompanying each publication, but also for relatively easy extension of other conventional GEMs to ec-models. Main differences between the various approaches are (i) the exact model implementation of Eqs. (1)–(2); (ii) the source of k_{cat} values. Notably, single-enzyme reactions (i.e. no subunits and no isoenzymes) are simplest to expand with enzymatic constraint, as they involve only single molecular weights and k_{cat} values. For isoenzymes, AutoPACMEN [9] always assumes that the enzyme with the lowest cost,

$$c_i = \frac{MW_i}{E_i} \tag{3}$$

will be used, while in GECKO [10,11], DLKcat [12] and ECMpy [13] isoenzymatic reactions are split and thereby explicitly allowing the model to *choose* which isoenzyme should be used. For GECKO and DLKcat this results in significantly increased model sizes, while ECMpy has only one single added constraint. While all model formulations allow for proteome prediction, only DLKcat and GECKO can directly constrain the model with measured quantitative proteomics. Since 2020, ec-models have been constructed for a variety of different organisms, including prokaryotes [13–17] and eukaryotes [11,18–21], indicating wide applicability of this model formalism (Table 2).

Improving coverage of available k_{cat} values

Critical for explicitly including Eq. (1) is access to reliable and well-covered k_{cat} data [22], that can be obtained by different approaches, but is often gathered by querying the BRENDA [23] and SABIO-RK [24] databases. However, for many reactions no k_{cat} value is known: in *E. coli* less than 10% of its ± 2000 enzyme-reaction pairs have experimentally measured k_{cat} values [25], while for nonmodel organisms the situation is even more dire. When data are missing, GECKO and AutoPACMEN adopt k_{cat} values by fuzzy matching to similar reactions (by Enzyme Commission number, or substrate) or organisms, but this can cause model predictions to deviate significantly from experimental observations [8]. Using random k_{cat} values drastically

Table 1

Recent software solutions to reconstruct enzyme-constrained models.

Software	AutoPACMEN	ECMpy	GECKO	DLKcat
Method	sMOMENT	ECMpy	GECKO	GECKO
k_{cat} source	BRENDA, SABIO-RK	User provided	BRENDA	Deep learning
Method of adding enzyme constraints	To each enzyme catalyzed reaction, add enzyme usage as substrate, and one total protein exchange reaction	Add one enzyme constraint	Split isoenzyme reactions. To each enzyme catalyzed reaction, add isoenzyme-specific enzyme usage as substrates	Split isoenzyme reactions. To each enzyme catalyzed reaction, add isoenzyme-specific enzyme usage as substrates
Model complexity	Medium: add one enzyme usage pseudoreaction per metabolic reaction, and one exchange reaction for total enzyme content.	Low: no metabolites or reactions added, only a total protein constraint outside the S-matrix.	High: split isoenzymic reactions, add enzyme usage pseudoreactions for each (iso)enzyme, and exchange reaction for each (iso)enzyme.	High: split isoenzymic reactions, add enzyme usage pseudoreactions for each (iso)enzyme, and exchange reaction for each (iso)enzyme
Predict proteome	No, due to absence of isoenzyme specificity	Yes	Yes	Yes
Constrain proteome	No, due to absence of isoenzyme specificity	No, only total protein content constraint	Yes	Yes
Year	2020	2022	2018, 2021	2021
Reference	[9]	[13]	[10,11]	[12]
Platform	Python	Python	MATLAB / Python	MATLAB / Python

increased the prediction error in an ec-model of *Bacillus subtilis* [26]. ECMpy does not have its own k_{cat} sourcing capability, and if no k_{cat} is provided for a particular reaction its enzymatic cost will be set to zero.

With the knowledge that enzyme activities can vary in multiple orders of magnitude [27], fuzzy k_{cat} matching is not ideal. Beyond sourcing measured catalytic capacities, one can estimate k_{cat} values from quantitative proteomics and flux data sets [25]. Following Eq. (1), condition-specific apparent enzyme activities (k_{app}) can be computed when v_i and $[E]_i$ are known. When testing numerous conditions, maximum apparent activity ($k_{app,max}$) will in theory approach in vivo k_{cat} if the enzyme is used at full capacity in (at least) one of the conditions.

By comparing 31 growth conditions in *E. coli* with FBA as source of flux data, computed $k_{app,max}$ values were representative of reported k_{cat} [28], while the use of four knockout strains and flux data from ^{13}C-labeled substrates yielded similar results [29•]. In *S. cerevisiae*, only a weak correlation was found that could be somewhat improved when discarding k_{cat} values that were measured via heterologous expression [30••]. This approach yields reasonable estimates of k_{cat} values, with models performing better than those populated with k_{cat} from fuzzy matching, but remains a tedious process, especially for nonmodel organisms. Moreover, this approach should be treated with caution as its primary reliance on extensive fitting raises the risk that information on aspects such as potential undersaturation of enzymes and preparatory expression will be lost.

Predict k_{cat} values through machine learning

Even though species-specific k_{cat} coverage is poor, the BRENDA and SABIO-RK databases still contain a wealth of measured k_{cat} data. Notwithstanding various challenges including assay conditions affecting kinetic activities [31], this gathered information can be leveraged to predict k_{cat} values. Indeed, in *E. coli* machine learning predicted k_{cat} from a variety of features, including GEM network properties, enzyme structural properties, biochemical mechanism information, assay conditions, and absolute proteomics data [29•]. Particularly the latter requirement limits the use of this approach for other organisms. Contrastingly, deep learning does not rely on feature selection, and the DLKcat neural network was able to predict k_{cat} values from protein sequences and reaction substrates alone [12]. The DLKcat neural network is trained on 16 000+ in vitro measured k_{cat} entries from the BRENDA and SABIO-RK databases. Incorporation of DLKcat predicted in vitro k_{cat} values and Bayesian inference to reconcile the predicted in vitro k_{cat} values with in vivo measured phenotypes showed superior performance over fuzzy k_{cat} matching in various yeast ec-models. With

Table 2

Published enzyme- and proteome-constrained models since 2020.

Organism	Type	Software	Brief description	Year	Ref
Aspergillus niger	ec	GECKO	More accurate growth rate predictions due to considering enzyme constraints	2021	[53]
Bacillus coagulans	ec	GECKO	Not used for simulating intracellular fluxes as the total protein concentration was not known, indicating this as a crucial parameter in ec-models	2020	[14]
CHO cells	ec	[a]	Manual implemented MOMENT, to decipher clone- and media-specific lactate metabolism	2020	[18]
Cupriavidus necator	pc	RBApy	Extensive analysis of protein allocation and usage by incorporating multi-omics data	2021	[43]
Escherichia coli	ec	GECKO	Used to predict genetic targets for improved lysine production. Overexpression of top-demanding enzymes doubled lysine titers	2020	[15]
Escherichia coli	ec	[a]	Extension of GECKO with thermodynamic constraints, called ETGEM	2021	[16]
Escherichia coli	ec	ECMpy	Introduced ECMpy, predicting growth rates with less error than GECKO and sMOMENT	2022	[13]
Escherichia coli	pc	[a]	Introduced PAM (protein allocation model) framework, an extension on GECKO formulation to find optimal partitioning of a limited proteome to different protein entities	2020	[46]
Escherichia coli	pc	ETFL	Introduced ETFL, simulating both enzyme and mRNA levels, by formulating a mixed-integer linear problem	2020	[40]
Escherichia coli	pc	ETFL	Dynamic ETFL, showing the time-dependent diauxie phenotype	2021	[44]
Escherichia coli	pc	[a]	Modified GECKO to include protein synthesis and turnover, with results suggesting that cellular growth is mainly regulated by protein translation	2021	[47]
Homo sapiens	ec	GECKO	Cell-specific ec-models were reconstructed, giving accurate predictions of exchange fluxes for 78% of metabolite components from the growth medium	2020	[19]
Lactococcus lactis	pc	[a]	Chemostat proteomics and flux data from strains evolved under nutrient-rich conditions were integrated, confirming that glucose and arginine uptake was growth limiting	2020	[48]
Rhizophagus irregularis	ec	[a]	Introduced the eMOMENT approach that is a slight modification of MOMENT, to also consider promiscuous enzymes	2020	[20]
Saccharomyces cerevisiae	ec	GECKO	Integrated with a Boolean module describing glucose and nitrogen signaling, accurately predicting growth-rate dependent isoenzyme preference	2021	[54]
Saccharomyces cerevisiae	ec/pc	[a]	The ecMitoYeast model is a hybrid of ec- and pc-model formulation. While based on ec-model formulation, it was expanded with protein transport over the mitochondrial membrane, a detailed representation of the proton motive force, and notably the dynamic allocation of mitochondrial enzyme cofactors such as iron-sulfur clusters	2021	[45]
Saccharomyces cerevisiae	pc	[a]	Major extension from pc to whole cell model, including 26 cellular processes	2020	[55]
Saccharomyces cerevisiae	pc	[a]	Describes cofactor requirements and reproduced the effect of iron deficiency	2021	[49]
Saccharomyces cerevisiae	pc	yETFL	Expanded ETFL with e.g. eukaryotic compartmentalized expression system	2021	[41]
Saccharomyces cerevisiae	pc	[a]	Whole-cell model that besides pc formulation also includes compartment-specific protein allocation, by considering membrane space and compartment volume constraints	2022	[50]
Streptomyces coelicolor	ec	GECKO	Proteomics data was applied to further constrain fluxes, to e.g. unravel mutant-specific growth rate differences	2020	[17]
Saccharomyces cerevisiae, *Kluyveromyces marxianus*, *Yarrowia lipolytica*	ec	GECKO	Release of GECKO 2.0 with updated fuzzy k_{cat} matching and improved model generation for other organisms. Ec-models of three yeast species were compared under environmental stress, demonstrating different long-term stress responses when comparing enzyme usages	2021	[11]
14 yeast species	ec	GECKO	Phylogenetically diverse yeast species see largely overlapping flux control coefficients for many enzymes in primary metabolism, with Erg3 as notable exception	2021	[21]

[a] model was reconstructed via a (semi-)manual approach without the use or presentation of software that could reconstruct models for other organisms.

DLKcat only requiring protein sequences and reaction substrates renders this approach widely applicable for any organism.

An interesting extension to establishing k_{cat} values is the consideration of temperature on enzyme activity and denaturation [32•]. For this, enzyme optimal temperatures (T_{opt}) were predicted by machine learning [33], while Bayesian inference dealt with uncertainty in T_{opt} and melting temperatures (T_m). Heterologous expression of a thermotolerant squalene epoxidase confirmed that the native yeast gene limits growth at superoptimal temperatures, as predicted by this modeling approach [32•].

Expand with nonmetabolic reactions

Consider protein translation affecting enzyme levels

While constraints of enzyme levels and catalytic activities addresses the protein allocation problem, and also known as *course-grained* approaches, the biosynthesis of proteins also raises the resource allocation problem that can be addressed by *fine-grained* approaches that explicitly integrate cellular processes into a GEM [34]. Provision of metabolic precursors and energy by metabolism are required to synthesize the very enzymes that constrain metabolism. Moreover, protein translation requires ribosomes that are themselves synthesized by protein translation. The concentration of E_i in Eq. (1) is therefore not constant but determined in the model:

$$v_{syn,i} \leq k_{ribo} \cdot [E]_{ribo} \quad (4)$$

$$\sum v_{syn,i} = v_{deg,i} + v_{dil,i} = k_{deg,i}\,[E]_i + \mu\,[E]_i \quad (5)$$

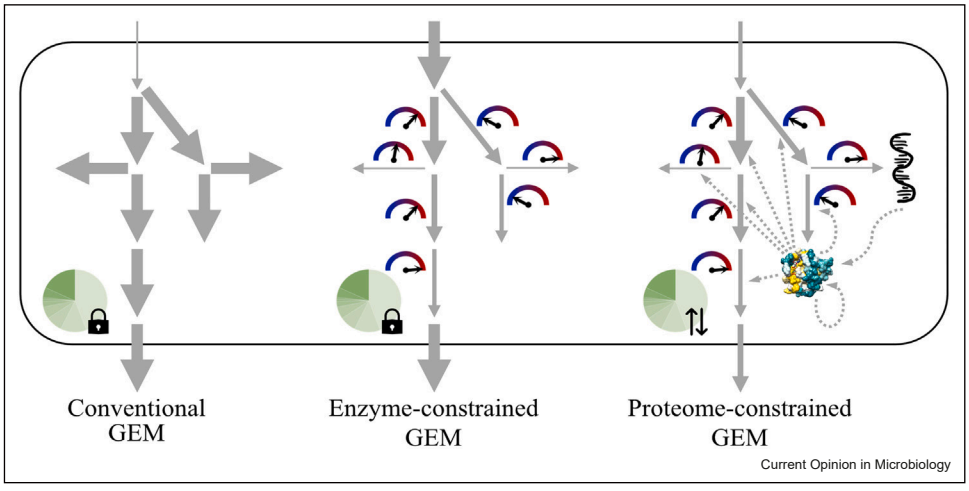
where v_{syn}, v_{deg} and v_{dil} are the rates of protein-specific translation, degradation and (growth-related) dilution, respectively. Protein synthesis (Eq. (4)) is dependent on ribosome concentration and maximum translation rate (k_{ribo}). The constraint-based framework meanwhile dictates (Eq. (5)) that the protein-specific synthesis rates are equal to the sum of the protein-specific degradation constants (k_{deg}) and growth rate (μ), both multiplied by the enzyme concentration. Variations of these equations can also be used to describe other cellular processes.

Accordingly, fine-grained models consider both the biosynthesis of macro-molecular constituents and explicitly simulate their catalytic role, and are also known as proteome-constrained models, multiscale extensions to GEMs or resource allocation models (Figure 1).

Proteome-constrained model reconstruction

Various approaches have been developed to consider Eqs. (4)–(5), including RBA (resource balance analysis [35]), ME (metabolism and expression [36,37]) and ETFL (expression and thermodynamics flux). Generally, they are part metabolic network, part description of gene expression and protein translation, and coupling coefficients are subsequently used to relate Eq. (4) to Eq. (1). As pc-models have significant higher complexity than ec-models they have often been introduced as manual ad-hoc reconstructions for a particular organism, instead of providing a software that can be applied for

Figure 1



Comparison of various model types. Where conventional GEMs (left) have a fixed biomass composition, their internal fluxes are not under capacity constraints (indicated by the thick arrows), while typically the nutrient uptake is constraint (thin arrow). Their biomass composition (green chart) is fixed. In enzyme-constrained GEMs (middle), the internal reactions are each under their own unique capacity constraint (various arrow thickness), while also here the biomass composition is fixed. Proteome-constrained GEMs (right) include reactions describing macromolecule synthesis, such as transcription and translation. Reaction enzymatic constraints are affected by the amount of enzyme synthesized, and the biomass composition is variable.

Table 3

Recent software solutions to reconstruct proteome-constrained models.

Software	RBApy	COBRAme	(y)ETFL
Thermodynamics	No	No	Yes
Workflow	Automated	Manual	Manual
Optimization	Iterative LP	Iterative LP	MILP
Organisms for which pc-model was reconstructed using the software	*Escherichia coli* [38••] *Bacillus subtilis* [38••] *Cupriavidus necator* [43]	*Escherichia coli* [39] *Clostridium ljungdahlii* [56]	*Escherichia coli* [40] *Saccharomyces cerevisiae* [41]
Year	2019	2018	2020, 2021
Reference	[38••]	[39]	[40,41]
Platform	Python	Python (cobrapy)	Python

reconstructing models for other organisms. More recently this has been addressed, so there are now four different software solutions for reconstructing pc-models (Table 3).

RBApy in particular is a very user-friendly software that can extend conventional GEMs with translation and transcription, and initially only requires the GEM in SBML format and the organism Uniprot identifier, although RBApy is currently only suitable for prokaryotes [38••]. Additional curation is conveniently facilitated using dedicated 'helper files' formatted to contain relevant information, while RBApy can be extended to consider further processes, such as protein secretion and chaperoning. COBRAme [39] and (y)ETFL [40•,41] are less automated, although example tutorials are provided. COBRAme is an extension of cobrapy [42], rendering it more accessible to cobrapy users, but otherwise it is not too dissimilar from RBApy. ETFL (and the yeast variant yETFL) differ mostly in two important aspects: consideration of thermodynamics constraints, and the approach by which growth-related variables (see e.g. Eq. (5)) are considered. The other pc-model formulations require sequential solving of a series of linear problems (LPs), where the growth rate predicted from each iteration is used as variable for the next simulation, requiring extra computational power. In contrast, the ETFL model formalism defines a mixed-integer linear (MILP) problem that can directly be solved with a MILP solver, although those mathematical problems are also still significantly more computationally expansive compared to the relatively simple LPs of ec-models. While the three pc-models solutions have been published in recent years, they have not yet been widely used to reconstruct models for different organisms [40•,41,43,44], while more models were reconstructed ad hoc [45–50] (Table 2).

Additional resource allocation constraints

Beyond enzymatic, gene expression and protein translation constraints, cellular resources are also allocated to other processes. The mammalian protein secretory pathway was introduced to GEMs and was able to accurately predict the influence on various interventions on protein secretion [51•]. Coupling the transport of proteins over the yeast mitochondrial membrane allowed to quantify this effect on the proton motive force [45].

Conclusions

The principle of resource allocation constraints has been around for a while but had not been as widely adopted as perhaps one imagines their benefit over conventional GEMs. Contributing to this has likely been the lack of user-friendly and adaptable software solutions to generate these models. Ec-models are now easier to reconstruct with AutoPACMEN and GECKO, reflected with ec-model reconstructed for 22 distinct species since 2020 (Table 2). This will likely accelerate for nonmodel organisms with the advance of predicting k_{cat} values from deep learning. To date, pc-models have been reconstructed for 7 species (Tables 2–3), but it is anticipated that this number will also rise in the years to come, due to the required software solutions having become available. Irrespective of the formalism that is chosen for model reconstruction, on the model application side progress has already been made to develop computational tools such as MEWpy [52] that can handle a wide variety of ec- and pc-model formalisms.

As reconstruction of the models has been made easier, and can be reconstructed with feasible parameters, a challenge remains to populate such models with experimentally measured data. Albeit deep-learning derived k_{cat} values are significantly better than fuzzy matching, fine-tuning their value in DLKcat-derived models still relies on the availability of proteomics data. Merely generating resource allocation constraint models for many more organisms will likely be of limited impact if this is not accompanied by increasing availability of (multi-)omics data for such nonmodel organisms.

Conflict of interest statement

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

Acknowledgements

E.J.K. is supported by the European Union's Horizon 2020 Research and Innovation Program [grant number 814650], the Research Council for Environment, Agricultural Sciences and Spatial Planning (Formas) [grant number 2018-00597], the Swedish Research Council (VR) [grant number 2019-04624] and the Novo Nordisk Foundation [grant number NNF20CC0035580].

References and recommended reading

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

- • of special interest
- •• of outstanding interest

1. Orth JD, Thiele I, Palsson BØ: **What is flux balance analysis?** *Nat Biotechnol* 2010, **28**:245-248, https://doi.org/10.1038/nbt.1614

2. Lu H, Kerkhoven EJ, Nielsen J: **Multiscale models quantifying yeast physiology: towards a whole-cell model**. *Trends Biotechnol* (3) 2021, **40**:291-305, https://doi.org/10.1016/j.tibtech.2021.06.010

3. Lewis NE, Hixson KK, Conrad TM, *et al*.: **Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models**. *Mol Syst Biol* 2010, **6**:390, https://doi.org/10.1038/msb.2010.47

4. Beg QK, Vazquez A, Ernst J, *et al*.: **Intracellular crowding defines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity**. *Proc Natl Acad Sci U S A* 2007, **104**:12663-12668, https://doi.org/10.1073/pnas.0609845104

5. Schultz A, Qutub AA: **Predicting internal cell fluxes at sub-optimal growth**. *BMC Syst Biol* 2015, **9**:1-12, https://doi.org/10.1186/s12918-015-0153-3

6. Mori M, Hwa T, Martin OC, *et al*.: **Constrained allocation flux balance analysis**. *PLoS Comput Biol* 2016, **12**:1-24, https://doi.org/10.1371/journal.pcbi.1004913

7. Shlomi T, Benyamini T, Gottlieb E, *et al*.: **Genome-scale metabolic modeling elucidates the role of proliferative adaptation in causing the warburg effect**. *PLoS Comput Biol* 2011, **7**:1-8, https://doi.org/10.1371/journal.pcbi.1002018

8. Chen Y, Nielsen J: **Mathematical modeling of proteome constraints within metabolism**. *Curr Opin Syst Biol* 2021, **25**:50-56, https://doi.org/10.1016/j.coisb.2021.03.003

9. Bekiaris PS, Klamt S: **Automatic construction of metabolic models with enzyme constraints**. *BMC Bioinformatics* 2020, **21**:1-13, https://doi.org/10.1186/s12859-019-3329-9

10. Sánchez BJ, Zhang C, Nilsson A, *et al*.: **Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints**. *Mol Syst Biol* 2017, **13**:935, , https://doi.org/10.15252/msb.20167411

11. Domenzain I, Sánchez B, Anton M, *et al*.: **Reconstruction of a catalogue of genome-scale metabolic models with enzymatic constraints using GECKO 2.0**. *bioRxiv* 2021,2021.03.05.433259, , https://doi.org/10.1101/2021.03.05.433259

12. Li F, Yuan L, Lu H, *et al*.: **Deep learning based kcat prediction enables improved enzyme constrained model reconstruction**. *bioRxiv* 2021,2021.08.06.455417, https://doi.org/10.1101/2021.08.06.455417

13. Mao Z, Zhao X, Yang X, *et al*.: **ECMpy, a simplified workflow for constructing enzymatic constrained metabolic network model**. *Biomolecules* 2022, **12**:65, https://doi.org/10.3390/biom12010065

14. Chen Y, Sun Y, Liu Z, *et al*.: **Genome-scale modeling for Bacillus coagulans to understand the metabolic characteristics**. *Biotechnol Bioeng* 2020, **117**:3545-3558, https://doi.org/10.1002/bit.27488

15. • Ye C, Luo Q, Guo L, *et al*.: **Improving lysine production through construction of an Escherichia coli enzyme-constrained model**. *Biotechnol Bioeng* 2020, **117**:3533-3544, https://doi.org/10.1002/bit.27485.

The authors used an ec-model to predict which enzymes where highly used, and overexpressing some of the top-used enzymes improved lysine production, demonstrating the use of ec-models in biotechnology.

16. Yang X, Mao Z, Zhao X, *et al*.: **Integrating thermodynamic and enzymatic constraints into genome-scale metabolic models**. *Metab Eng* 2021, **67**:133-144, https://doi.org/10.1016/j.ymben.2021.06.005

17. Sulheim S, Kumelj T, van Dissel D, *et al*.: **Enzyme-constrained models and omics analysis of streptomyces coelicolor reveal metabolic changes that enhance heterologous production**. *iScience* 2020, **23**:101525, https://doi.org/10.1016/j.isci.2020.101525

18. Yeo HC, Hong J, Lakshmanan M, Lee DY: **Enzyme capacity-based genome scale modelling of CHO cells**. *Metab Eng* 2020, **60**:138-147, https://doi.org/10.1016/j.ymben.2020.04.005

19. Robinson JL, Kocabaş P, Wang H, *et al*.: **An atlas of human metabolism**. *Sci Signal* 2020, **13**:1-12, https://doi.org/10.1126/scisignal.aaz1482

20. Wendering P, Nikoloski Z: **Genome-scale modeling specifies the metabolic capabilities of Rhizophagus irregularis**. *mSystems* 2022, **7**:e0121621, https://doi.org/10.1128/msystems.01216-21

21. Lu H, Li F, Yuan L, *et al*.: **Yeast metabolic innovations emerged via expanded metabolic network and gene positive selection**. *Mol Syst Biol* 2021, **17**:1-23, https://doi.org/10.15252/msb.202110427

22. Li Z, Zhang C, Wang Z, *et al*.: **High-throughput and reliable acquisition of in vivo turnover number fuels precise metabolic engineering**. *Synth Syst Biotechnol* 2022, **7**:541-543, https://doi.org/10.1016/j.synbio.2021.12.006

23. Chang A, Jeske L, Ulbrich S, *et al*.: **BRENDA, the ELIXIR core data resource in 2021: new developments and updates**. *Nucleic Acids Res* 2021, **49**:D498-D508, https://doi.org/10.1093/nar/gkaa1025

24. Wittig U, Kania R, Golebiewski M, *et al*.: **SABIO-RK - database for biochemical reaction kinetics**. *Nucleic Acids Res* 2012, **40**:790-796, https://doi.org/10.1093/nar/gkr1046

25. Davidi D, Milo R: **Lessons on enzyme kinetics from quantitative proteomics**. *Curr Opin Biotechnol* 2017, **46**:81-89, https://doi.org/10.1016/j.copbio.2017.02.007

26. Massaiu I, Pasotti L, Sonnenschein N, *et al*.: **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**:1-20, https://doi.org/10.1186/s12934-018-1052-2

27. Bar-Even A, Noor E, Savir Y, *et al*.: **The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters**. *Biochemistry* 2011, **50**:4402-4410, https://doi.org/10.1021/bi2002289

28. Davidi D, Noor E, Liebermeister W, *et al*.: **Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro k cat measurements**. *Proc Natl Acad Sci* 2016, **113**:3401-3406, https://doi.org/10.1073/pnas.1514240113

29. • Heckmann D, Campeau A, Lloyd CJ, *et al*.: **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, https://doi.org/10.1073/pnas.2001562117.

The authors find that *E. coli in vivo* k_{cat} values are robust against gene knockout. Metabolic adaptation to gene loss is probably mostly achieved by other mechanisms.

30. •• Chen Y, Nielsen J: **In vitro turnover numbers do not reflect in vivo activities of yeast enzymes**. *Proc Natl Acad Sci U S A* 2021, **118**:2-4, https://doi.org/10.1073/pnas.2108391118.

The authors encounter during determination of $k_{app,max}$ values from proteomics and flux data that yeast k_{cat} values that are measured after heterologous expression are unreliable.

31. Van Eunen K, Bouwman J, Daran-Lapujade P, *et al*.: **Measuring enzyme activities under standardized in vivo-like conditions for**

systems biology. *FEBS J* 2010, **277**:749-760, https://doi.org/10.1111/j.1742-4658.2009.07524.x

32. • Li G, Hu Y, Zrimec Jan, *et al*.: **Bayesian genome scale modelling identifies thermal determinants of yeast metabolism**. *Nat Commun* 2021, **12**:1-12, https://doi.org/10.1038/s41467-020-20338-2.
Metabolism is also constraint by growth temperatures and the authors here demonstrate how a modified ec-model is able to accurately predict that squalene epoxidase is limiting growth in yeast at elevated temperatures.

33. Li G, Rabe KS, Nielsen J, Engqvist MKM: **Machine learning applied to predicting microorganism growth temperatures and enzyme catalytic optima**. *ACS Synth Biol* 2019, **8**:1411-1420, https://doi.org/10.1021/acssynbio.9b00099

34. Yang L, Yurkovich JT, King ZA, Palsson BO: **Modeling the multi-scale mechanisms of macromolecular resource allocation**. *Curr Opin Microbiol* 2018, **45**:8-15, https://doi.org/10.1016/j.mib.2018.01.002

35. Goelzer A, Muntel J, Chubukov V, *et al*.: **Quantitative prediction of genome-wide resource allocation in bacteria**. *Metab Eng* 2015, **32**:232-243, https://doi.org/10.1016/j.ymben.2015.10.003

36. Lewis NE, Nagarajan H, Palsson BO: **Constraining the metabolic genotype–phenotype relationship using a phylogeny of in silico methods**. *Nat Rev Microbiol* 2012, **10**:291-305, https://doi.org/10.1038/nrmicro2737

37. Lerman J a, Hyduke DR, Latif H, *et al*.: **In silico method for modelling metabolism and gene product expression at genome scale**. *Nat Commun* 2012, **3**:929, https://doi.org/10.1038/ncomms1928

38. •• Bulović A, Fischer S, Dinh M, *et al*.: **Automated generation of bacterial resource allocation models**. *Metab Eng* 2019, **55**:12-22, https://doi.org/10.1016/j.ymben.2019.06.001.
RBApy is a user-friendly and intuitive software for generating proteome-constrained models for any prokaryote, and can furthermore be extended with descriptions of other cellular processes.

39. Lloyd CJ, Ebrahim A, Yang L, *et al*.: **COBRAme: a computational framework for genome-scale models of metabolism and gene expression**. *PLoS Comput Biol* 2018, **14**:1-14, https://doi.org/10.1371/journal.pcbi.1006302

40. • Salvy P, Hatzimanikatis V: **The ETFL formulation allows multi-omics integration in thermodynamics-compliant metabolism and expression models**. *Nat Commun* 2020, **11**:30, https://doi.org/10.1038/s41467-019-13818-7.
The ETFL framework includes a clever reformulation of the growth rate variable, allowing the model to be analyzed as a mixed-integer linear problem, greatly simplifying model simulations.

41. Oftadeh O, Salvy P, Masid M, *et al*.: **A genome-scale metabolic model of Saccharomyces cerevisiae that integrates expression constraints and reaction thermodynamics**. *Nat Commun* 2021, **12**:4790, https://doi.org/10.1038/s41467-021-25158-6

42. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR: **COBRApy: constraints-based reconstruction and analysis for python**. *BMC Syst Biol* 2013, **7**:74, https://doi.org/10.1186/1752-0509-7-74

43. Jahn M, Crang N, Janasch M, *et al*.: **Protein allocation and utilization in the versatile chemolithoautotroph cupriavidus necator**. *Elife* 2021, **10**:1-26, https://doi.org/10.7554/eLife.69019

44. Salvy P, Hatzimanikatis V: **Emergence of diauxie as an optimal growth strategy under resource allocation constraints in cellular metabolism**. *Proc Natl Acad Sci U S A* 2021, **118**:e2013836118, https://doi.org/10.1073/pnas.2013836118

45. Malina C, Di Bartolomeo F, Kerkhoven EJ, Nielsen J: **Constraint-based modeling of yeast mitochondria reveals the dynamics of protein import and iron-sulfur cluster biogenesis**. *iScience* 2021, **24**:103294, https://doi.org/10.1016/j.isci.2021.103294

46. Alter TB, Blank LM, Ebert BE: **Proteome regulation patterns determine escherichia coli wild-type and mutant phenotypes**. *mSystems* 2021, **6**:e00625-20, https://doi.org/10.1128/msystems.00625-20

47. Grigaitis P, Olivier BG, Fiedler T, *et al*.: **Protein cost allocation explains metabolic strategies in Escherichia coli**. *J Biotechnol* 2021, **327**:54-63, https://doi.org/10.1016/j.jbiotec.2020.11.003

48. Chen Y, Pelt-KleinJan E, Olst B, *et al*.: **Proteome constraints reveal targets for improving microbial fitness in nutrient-rich environments**. *Mol Syst Biol* 2021, **17**:1-13, https://doi.org/10.15252/msb.202010093

49. Chen Y, Li F, Mao J, *et al*.: **Yeast optimizes metal utilization based on metabolic network and enzyme kinetics**. *Proc Natl Acad Sci U S A* 2021, **118**:e2020154118, https://doi.org/10.1073/pnas.2020154118

50. Elsemman IE, Rodriguez Prado A, Grigaitis P, *et al*.: **Whole-cell modeling in yeast predicts compartment-specific proteome constraints that drive metabolic strategies**. *Nat Commun* 2022, **13**:1-12, https://doi.org/10.1038/s41467-022-28467-6

51. • Gutierrez JM, Feizi A, Li S, *et al*.: **Genome-scale reconstructions of the mammalian secretory pathway predict metabolic costs and limitations of protein secretion**. *Nat Commun* 2020, **11**:1-10, https://doi.org/10.1038/s41467-019-13867-y.
As most of the ec- and pc-model approaches are on microbes, this paper shows that these modelling frameworks are also well suited to study mammalian cells. Beyond translation, the model even included protein secretion.

52. Pereira V, Cruz F, Rocha M: **MEWpy: a computational strain optimization workbench in Python**. *Bioinformatics* 2021, **37**:2494-2496, https://doi.org/10.1093/bioinformatics/btab013

53. Zhou J, Zhuang Y, Xia J: **Integration of enzyme constraints in a genome-scale metabolic model of Aspergillus niger improves phenotype predictions**. *Microb Cell Fact* 2021, **20**:1-16, https://doi.org/10.1186/s12934-021-01614-2

54. Österberg L, Domenzain I, Münch J, *et al*.: **A novel yeast hybrid modeling framework integrating Boolean and enzyme-constrained networks enables exploration of the interplay between signaling and metabolism**. *PLoS Comput Biol* 2021, **17**:1-27, https://doi.org/10.1371/journal.pcbi.1008891

55. Ye C, Xu N, Gao C, *et al*.: **Comprehensive understanding of Saccharomyces cerevisiae phenotypes with whole-cell model WM_S288C**. *Biotechnol Bioeng* 2020, **117**:1562-1574, https://doi.org/10.1002/bit.27298

56. Liu JK, Lloyd C, Al-Bassam MM, *et al*.: **Predicting proteome allocation, overflow metabolism, and metal requirements in a model acetogen**. *PLoS Comput Biol* 2019, **15**:1-16, https://doi.org/10.1371/journal.pcbi.1006848