Mini-Review

Genome-scale models as a vehicle for knowledge transfer from microbial to mammalian cell systems

Benjamin Strain, James Morrissey, Athanasios Antonakoudis, Cleo Kontoravdi *

Department of Chemical Engineering, Imperial College London, London SW7 2AZ, United Kingdom

A B S T R A C T

With the plethora of omics data becoming available for mammalian cell and, increasingly, human cell systems, Genome-scale metabolic models (GEMs) have emerged as a useful tool for their organisation and analysis. The systems biology community has developed an array of tools for the solution, interrogation and customisation of GEMs as well as algorithms that enable the design of cells with desired phenotypes based on the multi-omics information contained in these models. However, these tools have largely found application in microbial cells systems, which benefit from smaller model size and ease of experimentation. Herein, we discuss the major outstanding challenges in the use of GEMs as a vehicle for accurately analysing data for mammalian cell systems and transferring methodologies that would enable their use to design strains and processes. We provide insights on the opportunities and limitations of applying GEMs to human cell systems for advancing our understanding of health and disease. We further propose their integration with data-driven tools and their enrichment with cellular functions beyond metabolism, which would, in theory, more accurately describe how resources are allocated intracellularly.

1. Introduction

Genome-scale metabolic models (GEMs) are a comprehensive representation of the link between genotype and phenotype, summarising information on genome, proteome and metabolome of a cell [1]. This information is organised in the form of matrices relating genes to metabolic reactions and reactions to metabolites, as well as a set of gene-protein-reaction (GPR) associations [2,3], as shown in Fig. 1. The construction of the matrices and GPR associations relies on genomic, transcriptomic, proteomic and metabolomic data [4]. Given this information and a set of metabolite uptake/secretion rates that act as constraints, GEMs can calculate the rates of intracellular reactions, thus providing fluxomic information.

GEMs have been constructed for over 6000 organisms [5,6], including the well-studied *Escherichia coli* [7], *Mus musculus* [8,9], *Pichia pastoris* [10], *Saccharomyces cerevisiae* [11] and *Homo sapiens* [12–14], with reconstructions regularly updated to include more complete GPR associations and remove blocked reactions and dead-end metabolites. GEMs can be used to study cell metabolism, optimise bioprocesses, and design strains with enhanced or custom functionality. Historically, GEMs have found greater application in microbial organisms, owing to the smaller model size and relative ease of experimental validation/manipulation compared to mammalian cell systems. The smaller metabolic network size of microbial model systems such as, for example, *E.coli* cells, has also led to the development of a variety of solution methodologies and optimisation algorithms for the design of cells with desired phenotype (summarised in [15]). Although the transfer of the entire repertoire of techniques to mammalian cell systems is often hampered by increased model size and complexity leading to highly underdetermined models, there are already developments in the use of mammalian cell GEMs for strain and process engineering (*e.g.*, [16–18]), as well as recent algorithm development work applied for understanding the nutritional needs of bioprocessing-relevant organisms but also Atlantic salmon (*Salmo salar*) [19].

In this work, we outline the main remaining challenges in the application of GEMs and related toolkits to mammalian cell systems and zoom in on their application to human cells as a vehicle for understanding health and disease.

2. On the use of GEMs for understanding human cell systems

Advances in clinical sample analysis and *in vitro* disease models are now enabling the generation of similar datasets for human

* Corresponding author.
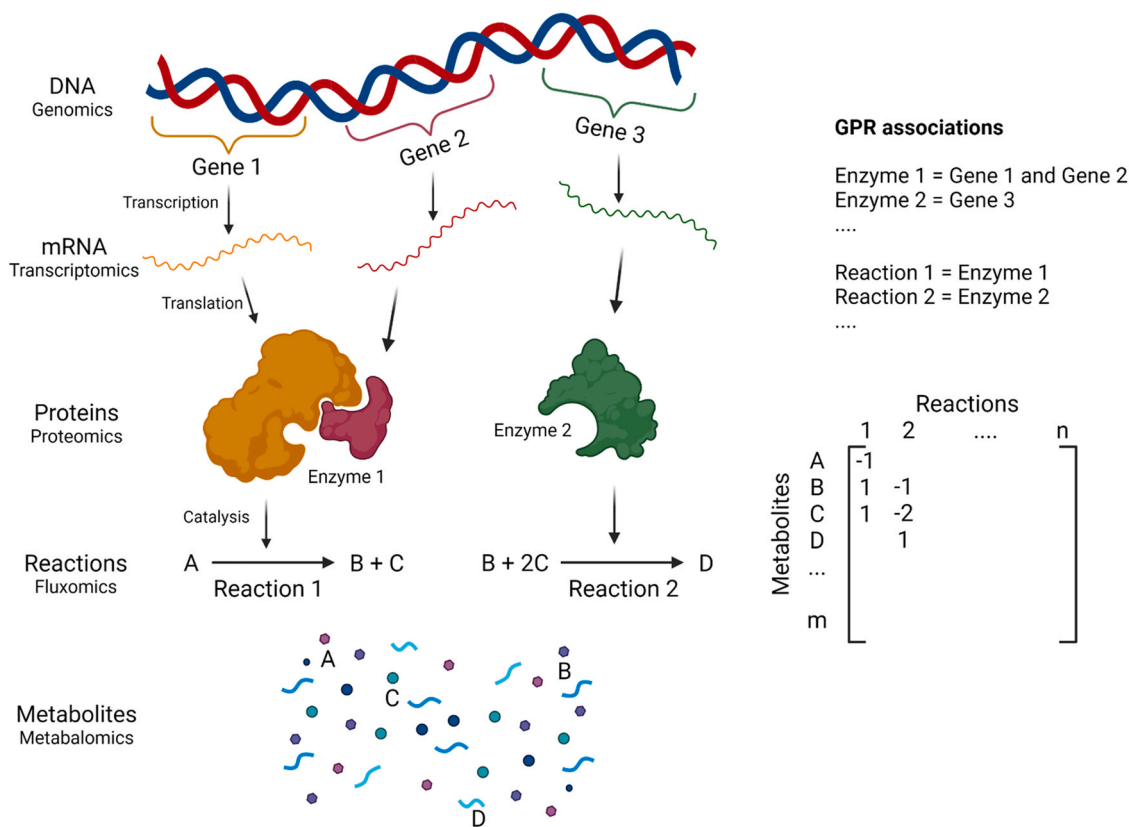*E-mail address:* cleo.kontoravdi@imperial.ac.uk (C. Kontoravdi).

Fig. 1. Reconstruction of a generic GEM from the different layers of 'omics datasets'.

physiology and pathophysiology. It is therefore opportune to examine how learnings and techniques developed for analysing data from biotechnologically relevant organisms using GEMs can be applied in a clinical context to further our understanding of health and disease (Fig. 2). For example, within industry, it is commonplace to generate GEMs specific to a cell line *via* the integration of 'omics data to prune inactive reactions using techniques such as GIMME [20] (cell line specific model generation is reviewed in depth in [21]). The same holds true in health and disease research, with thousands of patient-derived GEMs having been published for cancer alone [22]. As these models are specific to each disease type, they can be effectively used to explore essential genes in diseased tissue and to identify drug targets. In a recent study, for instance, single-gene knockouts were performed on GEMs of NCI-60 cancer cell line panel to identify and rank genes responsible for the growth of cancerous cells in an effort to identify potential drug targets that would reduce the growth rate of cancer cells but not that of normal cells [23]. This type of analysis is not only limited to chronic diseases; it has also been used in infectious disease studies. In one body of work, flux balance analysis (FBA) was applied to human lung cells infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and host-specific essential genes and gene pairs were determined through *in silico* knockouts that were theorised to reduce viral biomass production without affecting the host biomass [24].

These examples highlight the potential of GEMs for analysing large clinical omics datasets in a systematic way that links multiple levels of information. In our opinion, it is possible to envisage the use of GEMs and related methodologies, such as strain design algorithms, in a health setting to generate optimal strategies that reduce disease-associated phenotypes, while improving desirable healthy phenotypes. Herein, we outline the main outstanding challenges towards this end.

2.1. Challenges when building mammalian GEMs

Eukaryotes are known to be more biologically complicated that prokaryotes, meaning that applying GEM techniques to eukaryotic organisms is more challenging with respect to obtaining accurate predictions. One of the key sources of difference in complexity between prokaryotes and eukaryotes is the presence of subcellular organelle structures, such as mitochondria, peroxisomes, and nucleus, that do not exist in prokaryotes. Any well annotated eukaryotic GEM must contain these structures and the reactions associated with them for truly accurate predictions [25]. Significantly however, the presence of sub compartments means there is a requirement to gap fill the model using intracellular transport reactions. These transport reactions are often poorly studied and can lead to models with many reversible reactions, which, in turn, may lead to futile cycles, freely exchanging metabolites and protons across compartments, and erroneous energy generation calculations [26]. These cycles have been shown to inflate maximal biomass production rates by 25 % and are known to be present in the majority of published genome scale models [27], with eukaryotic models at greater risk thanks to the increased presence of intracellular exchanges. Ultimately, these results highlight the importance of using an appropriate combination of gap filling algorithms (reviewed in depth here [28,29]) and manual curation when moving from prokaryotic to more complex, eukaryotic models of metabolism to ensure accurate predictive performance.

While GEMs have been developed for many different species across all domains of life (reviewed in [6]), given the complexity of building eukaryotic GEMs, there has been a lack of regularly updated and publicly curated GEMs for mammalian model organisms such as *Mus musculus* (mouse) and *Rattus norvegicus* (rat) [30]. Instead, recent research developing new modelling techniques using non-human
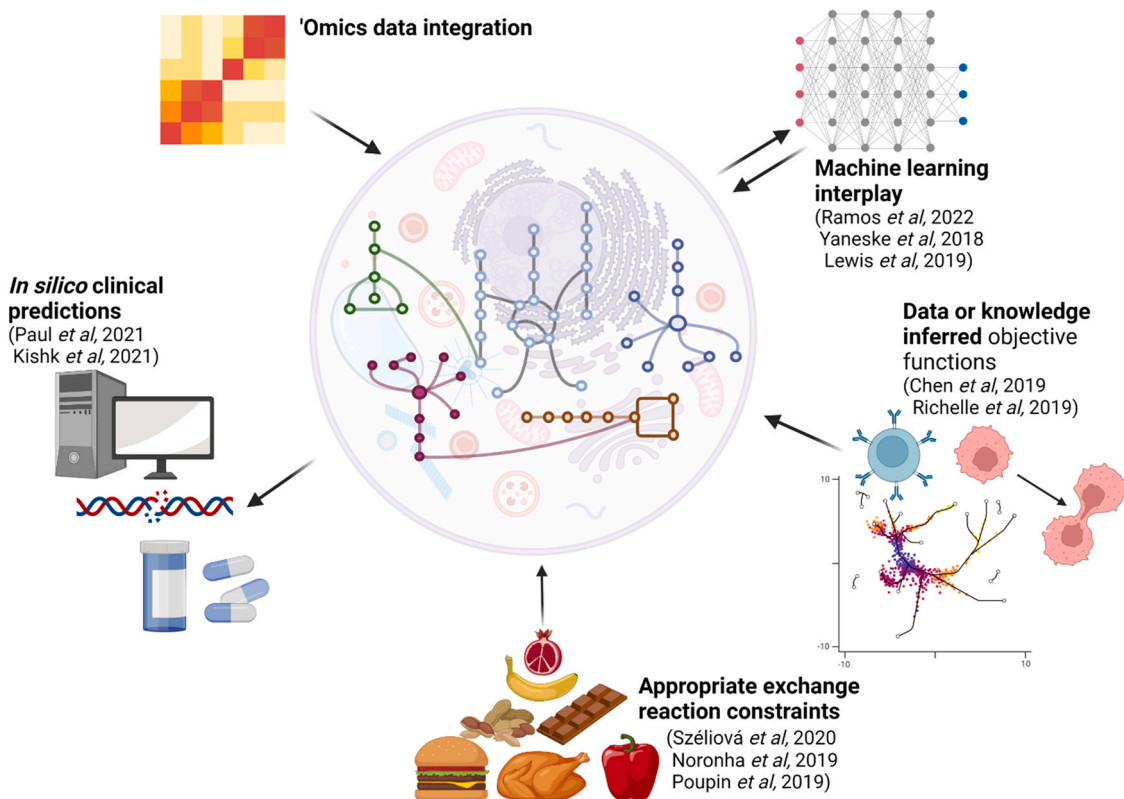
Fig. 2. Summary of areas where knowledge can be transferred from mammalian cell GEMs to wider health and disease research.

mammalian GEMs has predominantly focused on industrially relevant organisms such as Chinese Hamster Ovary (CHO) cells. To help address this, a framework has recently been published that combines multiple data sources, including the Kyoto Encyclopaedia of Genes and Genomes (KEGG) [31], and generates a coherent collection of GEMs for major model animals using the Human1 GEM as a template [30,32]. This approach allows for the straightforward development and maintenance of GEMs for multiple species. Since small rodents account for 90% animals used annually in medical research [33], the development of these models using a high-quality model as a backbone opens the possibility to better utilise GEMs in medical research settings, reduce the reliance on model animals and understand differences between model animal and human metabolism.

2.2. Determining effective exchange reaction constraints

One of the first challenges that occurs when running GEMs is gathering sufficient extracellular metabolomic data to effectively calculate metabolite uptake rates to constrain the GEM of interest. Within industrial biotechnology, the calculation of these uptake rates is straight-forward, thanks to the relative ease at which extracellular metabolomics can be measured in bioreactors. This means that industrially relevant cultured mammalian cells, such as CHO cell lines, often have detailed constraints for many exchange reactions. This has allowed researchers to understand how the accuracy of this data can affect GEM predictions. For instance, using the CHO cell GEM, researchers have demonstrated that the measurement of low exchange rates of essential amino acids has the biggest impact on the growth rate prediction [34] and that the highly accurate quantification of all uptake and secretion rates was essential for reliable predictions generated by FBA [35].

The generation of such extracellular time-course metabolomics is far more challenging in multicellular organisms. While researchers can culture the cells of interest *in vitro*, this may not be fully representative of how a tissue behaves *in vivo*. This means that generating *in vivo* constraints is of vital importance to accurately understand diseased states, toxicology, and nutrition. A potential method to do this is the use of nutrition databases to calculate the approximate composition of metabolites in a diet that are available for uptake by a cell. One such database is the Virtual Metabolic Human [36], which contains the composition of 11 pre-defined diets that can be downloaded as a flux rate (in mM per person per day). This data can be directly used to constrain the human metabolic model. Significantly, while this resource acts as an excellent baseline for constraining the human GEM to understand differences in diet, given that small changes in the exchange rates of essential amino acids can significantly impact the accuracy of predictions in the CHO GEM [35], it seems unlikely that such a database would provide enough accuracy to consistently give meaningful outputs from a GEM in all use cases.

To overcome this obstacle, techniques that rely on true *in vivo* measurements, such as arterio-venous blood metabolomics (AVBM) profiles, may be considered. In this approach, blood samples are taken from an artery directly before and a vein directly after the tissue type of interest. The difference in metabolite concentrations between these two samples is then presumed to be the amount of metabolite exchanged by the tissue of interest, which can be used to constrain the GEM. This approach has recently been applied to the genome scale modelling of multi-cellular organisms. In one body of work, researchers used AVBM measurements to constrain a GEM to study the global metabolism of liver and intestine of a minipig model of obesity, leading to the identification of upregulated pathways in obese subjects, such as tryptophan metabolism [37]. Nonetheless, while this approach may be appropriate for the genome scale modelling of animal models in the lab, it is highly invasive and unlikely to be acceptable for humans.

2.3. Determining appropriate objective functions

The common selection of a biomass maximisation as an objective function for performing FBA of mammalian GEMs is a methodology that largely remains from microbial GEMs, despite that fact it is well known to not be representative of the true 'objective' of a mammalian cell, especially outside the exponential growth phase [38]. This lack of suitability of a biomass objective is even more apparent for *in vivo* systems where, unless the tissue of interest is cancerous, cells rarely maximise their proliferation. As a result, researchers trying to model *in vivo* systems must consider the use of alternate objective functions and draw inspiration from mammalian biotechnology solutions. For instance, an unconventional objective function based on the minimisation of non-essential nutrient uptake has been designed for the CHO cell GEM [39]. This method directly estimates essential amino acid uptake fluxes by solving for the "essential minimum" consumption requirements based on cellular growth measurements. This unconventional objective function was shown to distinguish metabolic differences between three distinct CHO cell lines not directly observed using the conventional biomass maximisation. This highlights how the use of more appropriate objective functions may render GEM outputs more information rich, improving their practical application in health and disease research.

The identification of more appropriate objectives may either be achieved applying well established knowledge around the tissues of interest (*e.g.*, a GEM of a B cell may be set to maximise antibody production) or by inferring cell functions through data, such as the analysis carried out by Richelle et al. [40]. In this work, the functions of a cell were inferred from transcriptomics data by considering the gene expression level associated with a metabolic pathway and the number of reactions involved. During this work, a list of tasks was curated resulting in a collection of 210 tasks covering seven major metabolic activities of a cell (energy generation, nucleotide, carbohydrates, amino acid, lipid, vitamin & cofactor and glycan metabolism). These tasks were used to protect selected metabolic features using context-specific model generation algorithms for human, CHO, and mouse cell GEMs. The results highlight that these context-specific models better capture the actual biological variability across cell lines. Similar methodologies can therefore be considered when trying to determine the 'goal' of a tissue when selecting an objective function.

In addition to the lack of suitability of maximising biomass, it is important to consider that the biomass formation of mammalian cells is highly variable, depending on factors such as environmental conditions, cell type or culture phase, meaning the biomass equation must be customised for optimal model performance. For example, research in CHO cells has demonstrated cell lines display highly variable total protein content, cell dry mass and lipid composition across cell lines [35]. Moreover, work using the human GEM showed that metabolite composition and associated coefficients of the biomass function had a large impact on the growth rate prediction accuracy of cancer cell lines. In addition, metabolite composition of the biomass equations significantly impacted gene essentiality accuracy [41], meaning a new biomass equation should arguably be determined in each case. To this end, tools originally designed for microbial systems may be used, such as BOFdat [42], to generate custom biomass reactions for mammalian cell systems based on experimental 'omics data.

2.4. On the integration of data-driven modelling with GEMs

In recent years, advances in artificial intelligence and machine learning have revolutionised many areas of biological research [43]. Such approaches have started to be coupled with GEMs to help improve predictions and aid model output analysis. The coupling of GEMs with data- driven methods has been proposed as a method to effectively reduce the solution space by predicting biologically relevant constraints from experimental data (reviewed in depth in [15,44,45]). As with the previously discussed methodological areas of genome-scale modelling, this coupling of machine learning with GEMs to improve predictions is at a more advanced stage in biotechnologically relevant mammalian cell systems than it is in human health and disease research. For instance, a recently published method, termed HybridFBA, coupled unsupervised machine learning with a CHO cell GEM. In this approach additional flux constraints were deduced by Principal Component Analysis (PCA) of experimental flux data [46]. Specifically, the authors used each principal component to impose a constraint on the direction of variation of groups of fluxes. This method was shown to significantly improve growth rate predictions compared to standard FBA and was used to design a culture feed *in silico* that led to desired phenotype from target cell lines. This highlights how the coupling of mammalian cell GEMs with machine learning algorithms can improve their performance.

In addition, machine learning methods may be used to better analyse outputs and extract meaning from complex model predictions. For example, flux distribution predictions may be analysed using supervised and unsupervised machine learning methodologies to pick apart key aspects of metabolism that may influence a diseased phenotype of interest. This methodology has already been well applied within health and disease research using GEMs [15,44,45]. For example, researchers have used unsupervised learning with GEMs to identify the fluxes that explain most of the data variation in breast cancer patients, reduce dimensionality and create patient groupings [47]. Furthermore, researchers have applied personalised FBA models of patient tumours to predict metabolite production rates. These were input into machine learning classifiers for the identification of metabolite biomarkers associated with radiation resistance. The results demonstrated improved classification accuracy and identification of clinical patient subgroups, marking a significant step toward personalised classifiers for radiation treatment response [48]. These approaches demonstrate the power of using these two techniques synergistically.

3. A case for resource allocation models

3.1. Benefits of resource allocation

There has been a drive in the systems biology community away from classical stoichiometric network study and towards the study of metabolism through an optimised cellular economy. Resource allocation models (RAMs), as recently reviewed in [49–52] and with key methods summarised in Table 1, can describe many aspects of metabolism and cellular behaviour [53,54], where simple stoichiometric balances fall short. So far this drive towards RAMs has been almost exclusively carried out in microbial systems, due to their relative simplicity. Enzyme constrained FBA (ecFBA) models are also considered in this review due to their similarities with RAMs and are included in Table 1.

One key advantage of a RAM approach to metabolic modelling is that the additional constraints greatly reduce the feasible solution space by placing more restrictive bounds on fluxes. This lowers the variability of metabolic fluxes and guides flux towards more biologically feasible solutions. This would be particularly useful for mammalian GEMs, which contain many thousands of reactions [9,55,56][refs], and hence have extremely large potential solution spaces.

As well as reducing the solution space of metabolic models, the additional constraints also predict and explain key phenotypes that are not possible with traditional stoichiometric models [57–60]. Classical models ignore costs related to synthesis and usage of proteins and are limited only by the stoichiometry of metabolites

Table 1
Resource allocation models and the current challenges in their application to mammalian systems.

Method	Method Class	Description	Current challenges for application to mammalian systems
FBAwMC [57]	ecFBA	Global constraint on enzyme solvency capacity and kinetics	Achieved already [76]
MOMENT [81]	ecFBA	Inclusion of enzyme concentration in solvency capacity and kinetics	More accurate k_{cat} values, further genome annotation
GECKO [60]	ecFBA	Kinetic and solvency capacity of enzymes with integration of proteomic data	More accurate k_{cat} values, further genome annotation, quantitative proteomic data
RBA [82]	RAM	Inclusion and constraining of translation, replication and transcription machinery	Accurate parameterisation
CAFBA [83]	RAM	Global constraint modelling tradeoff between growth and biosynthetic cost	Accurate parameterisation
ME models [53,84]	RAM	Addition and coupling of transcription and translation with metabolism	Further genome annotation, quantitative proteomic data, knowledge of expression machinery, computational burden
ETFL [85]	RAM	Integration of expression machinery with thermodynamics	Further genome annotation, quantitative proteomic data, knowledge of expression machinery, computational burden

exchanged by the cell with its surroundings, meaning, even if a reaction is unlikely to occur due to the production of an expensive catalysing enzyme, the model is unable to account for this. The communal usage of resources drastically effects the distribution of fluxes through the model. Phenomena such as overflow metabolism do not make sense from a purely stoichiometric point of view and can only be explained in the context of the trade-off between 'inefficient' metabolism, protein cost and cell growth [58,61,62]. The ability to predict overflow metabolism is an important feature of mammalian cell modelling, such as the Warburg Effect in cancer cells. Being able to better predict peripheral overflow metabolism would be beneficial in the metabolic modelling for clinical research of diseases such as cancer, where non central pathways are known to play a key role [63–66].

In addition to cancer cell biology, overflow metabolism is important in biopharmaceutical production using mammalian cells *e.g.*, CHO cells. CHO cells typically undergo a lactate-producing phase, in which overflow metabolism is high, followed by a lactate consuming phase as growth rate subsides [67]. The accumulation of lactate is toxic to cell cultures, causing the addition of base to maintain pH set point and subsequently raising osmolality and lowering growth rates [68,69]. The ability to accurately capture lactate producing and consuming phases through metabolic modelling would aid in process and cell line optimisation.

More recently, other phenomena have been effectively modelling through proteome allocation, for example arginine catabolism in L. lactis [70]. The application of resource allocation to mammalian metabolism would be able elucidate features that have yet to be observed in traditional metabolic modelling.

A further benefit of expanding classical models with resource allocation machinery is the ability to incorporate omics data more effectively. With the increased availability of omics data, GEMs provide an excellent framework for the integration of this data into a combined workflow. As RAMs can consider transcription and translation machinery, transcriptomics and proteomics can be used to constrain metabolism in a more targeted manner, as opposed to current methods which rely on assumptions on the link between reaction rate and gene expression/protein translation [71–73].

The broadened scope of RAMs allows a more complete understanding of cell behaviour and the relationship between cellular processes. This allows predictions that could not be captured with classical models, such as identifying bottlenecks and gene engineering targets as well as biological parameters *e.g.*, condition-dependent biomass composition [74,75] and transcription/translation machinery [74].

3.2. Challenges in implementation to mammalian systems

While the benefits of RAMs and ecFBA in mammalian systems are numerous, there are obstacles on the path to achieving this goal. One of the main challenges is the scarcity of enzyme data. EcFBA, in particular, rely on the choice of turnover number (k_{cat}) values, which are difficult to source for mammalian cells. For example, Yeo et al. were able to find k_{cat} values for 16 % of enzymes in their CHO GEM [76], and several of these were taken from other organisms (*e.g.*, rodent and human) when there was no Chinese hamster data available. Additionally, *in vitro* k_{cat} measurements may differ from those *in vivo*, although the two have been shown to be correlated [77]. These factors render the application of ecFBA to mammalian cell systems difficult and prevent their full utilisation. A potential solution is to use machine learning approaches for k_{cat} prediction [78], which the enzyme amino acid sequences and the structures of their substrates are used to estimate k_{cat} values. Another solution is to infer the apparent k_{cat} value (k_{app}) *in vivo*, using measured proteomics and transcriptomics data [77,79].

A second issue is the aforementioned complexity of mammalian biology compared to simpler systems for which RAMs are more developed. There still exists a knowledge gap for protein sequences and gene-protein-reactions associations in mammalian cells, preventing the construction of effective transcription/translation machinery and integration into metabolism. This could be overcome by considering a reduced system, for example central carbon metabolism, for which biological understanding is more complete. This can then be expanded to consider peripheral pathways when the required data becomes available.

A third issue is the computational burden of fine-grained RAMs. As an example, one of the original *E.coli* RAMs [80], contains around 80,000 reactions from an original GEM of around 2000 reactions. Applying this 40-fold change to Recon 2.2 [55], one of the latest human GEMs, would result in a model of around 300,000 reactions. This makes simulation more computationally expensive, which is particularly problematic for sampling-based approaches. Again, focusing on a reduced system would alleviate this computational burden. Overcoming these challenges is imperative to progress mammalian cell metabolic modelling and to access the benefits that RAMs can offer to the community.

4. Concluding remarks

Herein, we summarised the main challenges for applying GEMs and related methodologies to mammalian cell systems, including human cell systems representative of health and disease states. These centred around (a) model size, which makes it cumbersome to apply advanced methodologies and algorithms developed for microbial cell systems in the absence of significant computational power, (b) time course data availability, which may be limited to *in vitro* studies to avoid intrusive sampling, and (c) the choice of appropriate objective functions that are representative of highly specialised human cells. Potential solutions involve (a) the integration

of data-driven elements with GEMs, either to derive appropriate constraints that restrict the solution space or to analyse and visualise GEM results, and (b) the development of RAMs for mammalian and, eventually, human cell systems. The flexibility that RAMs offer means that models are widely applicable, beyond exponential cell growth, where traditional metabolic modelling approaches are less effective. The main factors restricting mammalian RAM development include lack of data and, again, computational burden for large models. However, it is possible to make small steps towards the goal of creating full-scale mammalian cell RAMs using microorganism models as inspiration.

CRediT authorship contribution statement

BS: Conceptualization, Investigation, Visualization, Writing – original draft. **JM:** Conceptualization, Investigation, Visualization, Writing – original draft. **AA:** Conceptualization, Investigation, Visualization, Writing – original draft. **CK:** Conceptualization, Investigation, Supervision, Writing – review & editing.

Conflict of interest

The authors have no conflict of interest to declare.

Acknowledgements

References

[1] Nielsen J. Systems biology of metabolism. Annu Rev Biochem 2017;86(1):245–75.

[2] Maranas C, Zomorrodi A. Flux balance analysis and LP problems. Optimization methods in metabolic networks. 2016. p. 53–80.

[3] Di Filippo M, Damiani C, Pescini D, GPRuler. Metabolic gene-protein-reaction rules automatic reconstruction. PLoS Comput Biol 2021;17(11):e1009550.

[4] Haggart CR, et al. Whole-genome metabolic network reconstruction and constraint-based modeling. Methods Enzymol 2011;500:411–33.

[5] Martínez VS, et al. The topology of genome-scale metabolic reconstructions unravels independent modules and high network flexibility. PLoS Comput Biol 2022;18(6):e1010203.

[6] Gu C, et al. Current status and applications of genome-scale metabolic models. Genome Biol 2019;20(1):121.

[7] Edwards JS, Palsson BO. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. Proc Natl Acad Sci USA 2000;97(10):5528–33.

[8] Sheikh K, Förster J, Nielsen LK. Modeling hybridoma cell metabolism using a generic genome-scale metabolic model of *Mus musculus*. Biotechnol Prog 2005;21(1):112–21.

[9] Khodaee S, et al. iMM1865: a new reconstruction of mouse genome-scale metabolic model. Sci Rep 2020;10(1):6177.

[10] Tomàs-Gamisans M, Ferrer P, Albiol J. Fine-tuning the *P. pastoris* iMT1026 genome-scale metabolic model for improved prediction of growth on methanol or glycerol as sole carbon sources. Microb Biotechnol 2018;11(1):224–37.

[11] Förster J, et al. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. Genome Res 2003;13(2):244–53.

[12] Duarte NC, et al. Global reconstruction of the human metabolic network based on genomic and bibliomic data. Proc Natl Acad Sci USA 2007;104(6):1777–82.

[13] Quek L-E, et al. Reducing Recon 2 for steady-state flux analysis of HEK cell culture. J Biotechnol 2014;184:172–8.

[14] Zhang C, et al. Elucidating the reprograming of colorectal cancer metabolism using genome-scale metabolic modeling. Front Oncol 2019;9.

[15] Antonakoudis A, et al. The era of big data: genome-scale modelling meets machine learning. Comput Struct Biotechnol J 2020;18:3287–300.

[16] Kol S, et al. Multiplex secretome engineering enhances recombinant protein production and purity. Nat Commun 2020;11(1):1908.

[17] Schinn S-M, et al. A genome-scale metabolic network model and machine learning predict amino acid concentrations in Chinese Hamster Ovary cell cultures. Biotechnol Bioeng 2021;118(5):2118–23.

[18] Antonakoudis A, et al. Synergising stoichiometric modelling with artificial neural networks to predict antibody glycosylation patterns in Chinese hamster ovary cells. Comput Chem Eng 2021;154:107471.

[19] Weston BR, Thiele I. A nutrition algorithm to optimize feed and medium composition using genome-scale metabolic models. Metab Eng 2023.

[20] Becker SA, Palsson BO. Context-specific metabolic networks are consistent with experiments. PLoS Comput Biol 2008;4(5):e1000082.

[21] Robaina Estévez S, Nikoloski Z. Generalized framework for context-specific metabolic model extraction methods. Front Plant Sci 2014;5:491.

[22] Uhlen M, et al. A pathology atlas of the human cancer transcriptome. Science 2017;357(6352):eaan2507.

[23] Paul A, et al. Exploring gene knockout strategies to identify potential drug targets using genome-scale metabolic models. Sci Rep 2021;11(1):213.

[24] Kishk A, Pacheco MP, Sauter T. DCcov: repositioning of drugs and drug combinations for SARS-CoV-2 infected lung through constraint-based modeling. iScience 2021;24(11):103331.

[25] Klitgord N, Segrè D. The importance of compartmentalization in metabolic flux models: yeast as an ecosystem of organelles. Genome Inform 2009:41–55.

[26] Thiele I, Palsson B. A protocol for generating a high-quality genome-scale metabolic reconstruction. Nat Protoc 2010;5(1):93–121.

[27] Fritzemeier CJ, et al. Erroneous energy-generating cycles in published genome scale metabolic networks: identification and removal. PLoS Comput Biol 2017;13(4):e1005494.

[28] Orth JD, Palsson BØ. Systematizing the generation of missing metabolic knowledge. Biotechnol Bioeng 2010;107(3):403–12.

[29] Pan S, Reed JL. Advances in gap-filling genome-scale metabolic models and model-driven experiments lead to novel metabolic discoveries. Curr Opin Biotechnol 2018;51:103–8.

[30] Wang H, et al. Genome-scale metabolic network reconstruction of model animals as a platform for translational research. Proc Natl Acad Sci USA 2021;118:30.

[31] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28(1):27–30.

[32] Robinson JL, et al. An atlas of human metabolism. Sci Signal 2020;13:624.

[33] Daneshian M, et al. Animal use for science in Europe. Altex 2015;32(4):261–74.

[34] Széliová D, et al. Error propagation in constraint-based modeling of Chinese hamster ovary cells. Biotechnol J 2021;16(4):2000320.

[35] Széliová D, et al. What CHO is made of: variations in the biomass composition of Chinese hamster ovary cell lines. Metab Eng 2020;61:288–300.

[36] Noronha A, et al. The Virtual Metabolic Human database: integrating human and gut microbiome metabolism with nutrition and disease. Nucleic Acids Res 2018;47(D1):D614–24.

[37] Poupin N, et al. Arterio-venous metabolomics exploration reveals major changes across liver and intestine in the obese Yucatan minipig. Sci Rep 2019;9(1):12527.

[38] Feist AM, Palsson BO. The biomass objective function. Curr Opin Microbiol 2010;13(3):344–9.

[39] Chen Y, et al. An unconventional uptake rate objective function approach enhances applicability of genome-scale models for mammalian cells. NPJ Syst Biol Appl 2019;5:25.

[40] Richelle A, et al. Increasing consensus of context-specific metabolic models by integrating data-inferred cell functions. PLoS Comput Biol 2019;15(4):e1006867.

[41] Moscardó García M, et al. Importance of the biomass formulation for cancer metabolic modeling and drug prediction. iScience 2021;24(10):103110.

[42] Lachance J-C, et al. BOFdat: generating biomass objective functions for genome-scale metabolic models from experimental data. PLoS Comput Biol 2019;15(4):e1006971.

[43] Jones DT. Setting the standards for machine learning in biology. Nat Rev Mol Cell Biol 2019;20(11):659–60.

[44] Zampieri G, et al. Machine and deep learning meet genome-scale metabolic modeling. PLoS Comput Biol 2019;15(7):e1007084.

[45] Kim Y, Kim GB, Lee SY. Machine learning applications in genome-scale metabolic modeling. Curr Opin Syst Biol 2021;25:42–9.

[46] Ramos JRC, et al. Genome-scale modeling of Chinese hamster ovary cells by hybrid semi-parametric flux balance analysis. Bioprocess Biosyst Eng 2022;45(11):1889–904.

[47] Yaneske E, Angione C. The poly-omics of ageing through individual-based metabolic modelling. BMC Bioinform 2018;19(14):415.

[48] Lewis JE, Kemp ML. Integration of machine learning and genome-scale metabolic modeling identifies multi-omics biomarkers for radiation resistance. Nat Commun 2021;12(1):2700.

[49] De Becker K, et al. Using resource constraints derived from genomic and proteomic data in metabolic network models. Curr Opin Syst Biol 2022;29:100400.

[50] Kerkhoven EJ. Advances in constraint-based models: methods for improved predictive power based on resource allocation constraints. Curr Opin Microbiol 2022;68:102160.

[51] Chen Y, Nielsen J. Mathematical modeling of proteome constraints within metabolism. Curr Opin Syst Biol 2021;25:50–6.

[52] Dahal S, Zhao J, Yang L. Recent advances in genome-scale modeling of proteome allocation. Curr Opin Syst Biol 2021;26:39–45.

[53] O'Brien EJ, et al. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. Mol Syst Biol 2013;9:693.

[54] Massaiu I, 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):3.

[55] Swainston N, et al. Recon 2.2: from reconstruction to model of human metabolism. Metabolomics 2016;12:109.

[56] Hefzi H, et al. A consensus genome-scale reconstruction of Chinese hamster ovary cell metabolism. Cell Syst 2016;3(5):434–43. e8.

[57] Beg QK, et al. Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity. Proc Natl Acad Sci USA 2007;104(31):12663–8.

[58] Molenaar D, et al. Shifts in growth strategies reflect tradeoffs in cellular economics. Mol Syst Biol 2009;5:323.

[59] Zhuang K, Vemuri GN, Mahadevan R. Economics of membrane occupancy and respiro-fermentation. Mol Syst Biol 2011;7:500.

[60] Sánchez BJ, et al. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. Mol Syst Biol 2017;13(8):935.

[61] Scott M, et al. Interdependence of cell growth and gene expression: origins and consequences. Science 2010;330(6007):1099–102.

[62] Basan M, et al. Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. Nature 2015;528(7580):99–104.

[63] Han X, et al. Cancer causes metabolic perturbations associated with reduced insulin-stimulated glucose uptake in peripheral tissues and impaired muscle microvascular perfusion. Metabolism 2020;105:154169.

[64] Läsche M, Emons G, Gründker C. Shedding new light on cancer metabolism: a metabolic tightrope between life and death. Front Oncol 2020:**10**.

[65] Vanhove K, et al. The metabolic landscape of lung cancer: new insights in a disturbed glucose metabolism. Front Oncol 2019:**9**.

[66] Stine ZE, et al. Targeting cancer metabolism in the era of precision oncology. Nat Rev Drug Discov 2022;21(2):141–62.

[67] Zagari F, et al. Lactate metabolism shift in CHO cell culture: the role of mitochondrial oxidative activity. New Biotechnol 2013;30(2):238–45.

[68] Brunner M, et al. Elevated pCO(2) affects the lactate metabolic shift in CHO cell culture processes. Eng Life Sci 2018;18(3):204–14.

[69] Ahleboot Z, et al. Designing a strategy for pH control to improve CHO cell productivity in bioreactor. Avicenna J Med Biotechnol 2021;13(3):123–30.

[70] Chen Y, et al. Proteome constraints reveal targets for improving microbial fitness in nutrient-rich environments. Mol Syst Biol 2021;17(4):e10093.

[71] Kim MK, et al. E-Flux2 and SPOT: validated methods for inferring intracellular metabolic flux distributions from transcriptomic data. PLoS One 2016;11(6):e0157101.

[72] Zur H, Ruppin E, Shlomi T. iMAT: an integrative metabolic analysis tool. Bioinformatics 2010;26(24):3140–2.

[73] Jensen PA, Papin JA. Functional integration of a metabolic network model and expression data without arbitrary thresholding. Bioinformatics 2011;27(4):541–7.

[74] Lerman JA, et al. *In silico* method for modelling metabolism and gene product expression at genome scale. Nat Commun 2012;3(1):929.

[75] Lloyd CJ, et al. Computation of condition-dependent proteome allocation reveals variability in the macro and micro nutrient requirements for growth. PLoS Comput Biol 2021;17(6):e1007817.

[76] Yeo HC, et al. Enzyme capacity-based genome scale modelling of CHO cells. Metab Eng 2020;60:138–47.

[77] Davidi D, et al. Global characterization of *in vivo* enzyme catalytic rates and their correspondence to *in vitro* kcat measurements. Proc Natl Acad Sci USA 2016;113(12):3401–6.

[78] Li F, et al. Deep learning-based kcat prediction enables improved enzyme-constrained model reconstruction. Nat Catal 2022;5(8):662–72.

[79] Heckmann D, et al. Kinetic profiling of metabolic specialists demonstrates stability and consistency of *in vivo* enzyme turnover numbers. Proc Natl Acad Sci USA 2020;117(37):23182–90.

[80] Thiele I, et al. Multiscale modeling of metabolism and macromolecular synthesis in *E. coli* and its application to the evolution of codon usage. PLoS One 2012;7(9):e45635.

[81] Adadi R, et al. Prediction of microbial growth rate *versus* biomass yield by a metabolic network with kinetic parameters. PLoS Comput Biol 2012;8(7):e1002575.

[82] Goelzer A, Fromion V, Scorletti G. Cell design in bacteria as a convex optimization problem. Automatica 2011;47(6):1210–8.

[83] Mori M, et al. Constrained allocation flux balance analysis. PLoS Comput Biol 2016;12(6):e1004913.

[84] Thiele I, et al. Genome-scale reconstruction of *Escherichia coli*'s transcriptional and translational machinery: a knowledge base, its mathematical formulation, and its functional characterization. PLoS Comput Biol 2009;5(3):e1000312.

[85] Salvy P, Hatzimanikatis V. The ETFL formulation allows multi-omics integration in thermodynamics-compliant metabolism and expression models. Nat Commun 2020;11(1):30.