Cellular and Molecular Life Sciences (2020) 77:481–488
https://doi.org/10.1007/s00018-019-03383-z

REVIEW

Network reduction methods for genome-scale metabolic models

Dipali Singh[1,2] · Martin J. Lercher[1]

Received: 4 November 2019 / Revised: 4 November 2019 / Accepted: 12 November 2019 / Published online: 20 November 2019

Abstract

Genome-scale metabolic models (GSMs) provide a comprehensive representation of cellular metabolism. GSMs provide a mechanistic link between cellular genotypes and metabolic phenotypes, and are thus widely used to analyze metabolism at the systems level. GSMs consist of hundreds or thousands of reactions. They have thus largely been analyzed with computationally efficient constraint-based methods such as flux-balance analysis, limiting their scope and phenotype prediction accuracy. Computationally more demanding but potentially more informative methods, such as kinetic and dynamic modeling, are currently limited to small or medium-sized models. Thus, it is desirable to achieve unbiased stoichiometric reductions of large-scale metabolic models to small, coarse-grained model representations that capture significant metabolic modules. Here, we review published automated and semiautomated methods used for large-scale metabolic model reduction. The top-down methods discussed provide minimal networks that retain a set of user-protected phenotypes, but may reduce the model's metabolic and phenotypic versatility. In contrast, the two bottom-up approaches reviewed retain a more unbiased set of phenotypes; at the same time, these methods require the partitioning of the GSM into metabolic subsystems by the user, and make strong assumptions on the subsystems' connections and their states, respectively.

Keywords Genome-scale metabolic models · Metabolic networks · Network reduction methods · Flux-balance analysis · Elementary flux modes

Introduction

The increasing availability of genome sequences, combined with the wealth of biochemical knowledge [1], has paved the way for systematic metabolic network reconstructions at the cellular level, leading to in silico genome-scale metabolic models (GSMs). A GSM describes the interconversion of metabolites into biomass through genomically encoded enzymes in a given organism. The stoichiometry of these interconversions is mathematically represented in the form of a stoichiometric matrix *N,* whose rows correspond to internal metabolites and whose columns represent individual reactions [2]. GSMs also include information on the proteins required for the activity of each reaction (and on the encoding genes), and can thus link metabolic phenotypes to the organism's genotype.

GSMs have become a popular tool to study the system-level effects of genetic and environmental perturbations on metabolism and cell growth; they have been employed to design microbial strains with desired phenotypes, and to predict improved fermentation conditions, media, and other factors of interest (reviewed in [3–5]). GSMs are usually confined to structural analysis, based purely on network stoichiometry [2]; reasons are the computational efficiency of many such methods and the low number of parameters that have to be estimated. The prime example of such methods is flux-balance analysis (FBA) and its constraint-based extensions [6–12]. For further details on these methods, readers are referred to the reviews by Lewis et al. [6] and Singh et al. [13].

Currently available GSMs cover a wide variety of cells, including those of microorganisms, plants, and mammals [14]. These models typically increase in size over time, as new biological knowledge is integrated through the introduction of additional metabolites and reactions, or even complete metabolic pathways. An illustration is the expanding

✉ Dipali Singh
dipali.singh@quadram.ac.uk

[1] Department of Biology and Institute for Computer Science, Heinrich Heine University, 40225 Düesseldorf, Germany

[2] Present Address: Microbes in Food Chain, Quadram Institute Bioscience, Norwich Research Park, Norwich, UK

size of *E. coli* GSMs over the years, summarized in Table 1. Although the increasing size and complexity of GSMs better captures metabolic interactions and, therefore, metabolic capabilities, a variety of conceptually powerful analysis methods cannot be applied to large networks. Examples of computationally expensive constraint-based methods are the calculation of elementary flux modes [15] and of minimal cut sets [16].

All constraint-based methods typically applied to GSMs rely on a steady-state assumption, i.e., they assume that the concentration of internal metabolites and all reaction fluxes remain constant over time [6]. Moreover, they either ignore reaction kinetics or replace them by constant effective enzyme turnover numbers [22] or phenomenological functions of the growth rate [23, 24]. These methods cannot capture dynamic behaviors of biological systems, help to understand the role of metabolite concentrations in metabolic processes, or simulate non-balanced growth conditions that results in the accumulation of intracellular metabolites.

In contrast to structural methods, parameter-dependent dynamic modeling (or kinetic modeling) takes into account kinetic and thermodynamic information in addition to stoichiometry [25]. Such methods can thus help to understand non-balanced growth and dynamic behaviors of the cellular system, provided that kinetic parameters can be measured or can be fitted to experimental data. Because of their computational complexity and the number of kinetic parameters required, most applications and studies using metabolic models focus on the core metabolism or on reduced models, or analyze specific pathways of interest [26, 27]. Although such studies are useful for our understanding of the dynamic mechanisms of metabolic pathways, they are limited in their reflection of system-level effects (which can, to some extent, be captured by schematic whole-cell models such as those examined by Scott et al. [28]).

To facilitate the application of dynamic modeling approaches and other computationally demanding methods to metabolic networks representing complete cellular systems, one needs to reduce GSMs stoichiometrically to small, coarse-grained models that still capture phenotypic and/or genotypic properties of interest and display sufficient functionality and flexibility [29, 30]. Most reduced models employed in previous studies were generated ad hoc [27], and it is not clear to what extent they faithfully reflect system-level properties of the respective complete GSMs.

Table 1 Genome-scale metabolic models of *Escherichia coli*

Year	Model	Reactions	Metabolites	References
2000	iJE660	627	438	[7]
2003	iJR904	931	625	[17]
2007	iAF1260	1721	1039	[18]
2011	iJO1366	1863	1136	[19]
2014	EcoCyc-18.0-GSM	2286	1453	[20]
2017	iML1515	2719	1192	[21]

One class of model reduction algorithms utilizes context-specific gene expression data to reduce a GSM to those reactions that are likely active in a specific experimental setting; examples are GIMME [31] and iMAT [32, 33]. Similar methods have been developed to reduce GSMs of multicellular organisms to tissue-specific models utilizing gene expression and other experimental data; examples are MBA [34] and the method by Shlomi et al. [32]. Such methods based on the integration of GSMs with large, context-specific experimental data sets have been reviewed by Semidán & Zoran [35] and by Opdam et al. [36]. In this review, we instead focus on unbiased computational methods that reduce metabolic networks exclusively based on the network structure itself.

Metabolic network reduction methods

Enzyme subsets: loss-less network compression

The importance of reducing network size while retaining system properties and functionalities of interest was realized as early as in 2000. Pfeiffer et al. [37] introduced the concept of enzyme subsets, defined as sets of reactions that carry flux in a fixed proportion at steady state. The corresponding enzymes are thus likely under coordinated regulatory control. Enzyme subsets are sets of fully coupled reactions [38] and can be identified efficiently [39].

The lumping of reactions in the same enzyme subsets can reduce the network size without compromising the network capabilities and flexibility. The collapsing of enzyme subsets thus represents a consistent and "loss-less" method to reduce the network size. However, the degree of network reduction achievable with this method is limited, and thus, other network reduction algorithms have been proposed.

Minimal reaction sets: environment-specific subnetworks

An early method for network reduction was introduced by Burgard et al. [40]. This computational procedure identifies the minimal set of metabolic reactions capable of supporting growth in a given in silico environment based on mixed integer linear programming (MILP). The presence or absence of reaction j in the reduced model is incorporated into the flux-balance analysis framework as a binary variable (y_j), which is then introduced into the problem formulation as a flux constraint. The inclusion of binary variables converts what otherwise would have been a linear programming (LP) problem into an MILP problem. The latter is then solved for the minimal number of active reactions by minimizing over

the sum of the binary variables y_i while meeting a target biomass production rate.

This method was applied to an early GSM of *E. coli* with 720 reactions, reducing the network size to 234 and 231 reactions required for maximal biomass yield on glucose and on acetate, respectively. Relaxing the yield requirement to $\geq 70\%$ of maximal yield reduced the minimal reaction set further by ten reactions on glucose and by two reactions on acetate. Differences between the minimal reaction sets demonstrated the versatility and redundancy of the *E. coli* network.

This early study demonstrated that only a relatively small fraction of reactions need to be active in a given condition, and that sets of reactions are likely turned on or off based on growth requirements and substrate availability. For example, in the above case of glucose, the metabolic network is capable to support growth by utilizing only 31% of the available metabolic reactions.

Minimal reaction sets can also be identified using graph theoretical approaches [41]; Ref. [42] describes a recursive graph theoretical optimization approach to identify all minimal reaction sets of a given GSM. A very fast, approximate calculation of a minimal reaction set is provided using LP to find a set of active reactions that leads to a minimal sum of absolute fluxes while maintaining maximal biomass production [43], a method often termed parsimonious FBA.

While minimal reaction sets are useful for the detailed analysis of specific conditions, they do not conserve the versatility and flexibility of the underlying GSM. Moreover, the pathways employed in vivo are in many conditions not those that have the lowest total number of genes, but are determined by kinetic and thermodynamic properties of alternative pathways [44, 45].

NetworkReducer: a greedy approach to calculate minimal functional networks

NetworkReducer [29] is an algorithm for the automated reduction of a metabolic reconstruction to obtain a smaller network while retaining pre-specified metabolic modules of interest. NetworkReducer uses LP to iteratively eliminate reactions from a network while retaining specified properties and phenotypes. The algorithm consists of two steps: network pruning followed by network compression. The inputs to the algorithm are a metabolic network that is to be reduced and a set of properties and phenotypes, as listed below, that must be retained in the reduced network: protected metabolites and reactions; protected functions and phenotypes, such as maximal growth rate for a given nutrient uptake; the minimum degrees of freedom (*dof*, defined as the dimension of the null space of the network); and the minimum number of reactions to be retained in the reduced network.

The network pruning step checks the input network for the number of reactions and its *dof*. If it does not yet meet the minimality criteria, i.e., if the number of reactions and the *dof* are greater than the specified minima, then flux variability analysis (FVA) [46] is performed to determine the flux range of each reaction. FVA, like FBA, is an LP approach to assign fluxes to reactions. Contrary to FBA, which assigns a single flux value to each reaction, FVA determines the minimum and maximum admissible fluxes for each reaction at steady state.

NetworkReducer identifies reactions whose flux range does not include zero as essential; these cannot be removed from the network. From the remaining non-essential reactions that are not on the list of protected reactions, NetworkReducer chooses the one with the smallest flux range and tentatively removes it from the network. This reduced network is then re-checked for protected functions and phenotypes along with the minimality criteria. In case of any violations, the tentatively removed reaction is added back to the network and appended to the list of protected reactions. The pruning step is iterated until no more reactions can be removed or until the number of reactions or the degrees of freedom reaches its minimum limit. NetworkReducer belongs to the class of greedy algorithms, which solve an optimization problem by iteratively making local decisions (the removal of reactions with a low flux variability range in this case). Generally, such methods cannot guarantee that the final solution is indeed optimal and hence only solve the optimization problem approximately.

The subsequent network compression step involves lumping reactions that form fully coupled enzyme subsets, representing these with a pseudo-reaction described by the collapsed net stoichiometry, while retaining protected metabolites and reactions. Network compression does not affect protected functions and phenotypes.

NetworkReducer was applied to a GSM of *E. coli* (iAF1260) that contains 2382 reactions and 1668 internal metabolites. The protected phenotype included demand for at least 99% of the maximal growth rate limited by glucose uptake; the minimal degree of freedom was set to 1 to reduce network as much as possible. The reduced model after the network pruning step but without compression contained 438 metabolites and 455 reactions, while retaining protected reactions from central carbon metabolism. Network compression reduced the network size further to 85 metabolites and 105 reactions while retaining all functional and phenotypical properties of the pruned model.

In a second case study, NetworkReducer was applied to a GSM of the phototrophic cyanobacterium *Synechocystis sp. PCC 6803* consisting of 599 reactions and 519 internal metabolites. The aim of this application was to reduce the GSM while still retaining the maximal phototrophic growth and ethanol production. Thus, the protected reactions

included the Calvin cycle, light reactions, and pathways for biomass and ethanol production. The protected phenotypes included demand for at least 99% of the maximal growth rate and ethanol production limited by photon uptake. The reduced model contained 37 reactions and 38 internal metabolites and had 4 degrees of freedom. It was possible to efficiently compute the elementary flux modes of the reduced model, which would have been computationally extremely expensive for the full GSM. The computed elementary flux modes not only supported biomass or ethanol production separately, but also included a growth-coupled ethanol production mode that enforced ethanol secretion as a by-product of biomass formation.

MinNW: an exact approach to find minimal functional networks

Röhl and Bockmayr [30] implemented an algorithm that uses MILP to obtain an exact solution to the reduction problem solved approximately by NetworkReducer. The input to the algorithm is almost identical to that of NetworkReducer: a metabolic network that is to be reduced, a list of protected metabolites and reactions, protected functions and phenotypes, and the minimal accepted degrees of freedom. Contrary to NetworkReducer, this approach does not specify a minimal number of reactions to be retained in the reduced network. The MILP formulation allows to find the globally minimal number of active reactions, and does not require the identification of "candidate" reactions for removal through FVA. Each reaction i is associated with a binary variable a_i that forces the flux through this reaction to zero if $a_i = 0$; reversible reactions are associated with two binary variables, one for each flux direction. The optimization problem solved by MinNW is then to minimize the sum of all a_i, where a_i is fixed at 1 for all protected reactions. MinLP thereby finds a solution with the minimal number of active reactions while retaining protected metabolites, reactions, and functions. The minimal solution may be not unique; to compute all possible minimal networks, the MILP algorithm is applied iteratively, excluding already found minimal networks in each iteration.

This exact approach was applied to the GSMs of *E. coli* (iAF1260) and *Synechocystis sp. PCC 6803,* the same networks analyzed with NetworkReducer. The reduced network of *E. coli* consisted of 416 reactions, 39 reactions less than the pruned network from NetworkReducer. The reduced network of *Synechocystis sp.* consisted of 453 reactions, 9 reactions less than the pruned network from NetworkReducer. While MILP problems are computationally more complex than LP problems, MinNW was found to be 5–10 times faster than NetworkReducer in these applications. A noteworthy feature of minNW in comparison to NetworkReducer is its ability to compute all minimal subnetworks. When applied to the *Helicobacter pylori* GSM consisting of 501 reactions, minNW identified 16 minimal reaction sets with the desired biomass yield, each comprising 321 reactions. In this way, network reduction by minNW allows the exploration of alternative metabolic routes and provides an insight into the metabolic versatility of the studied organism.

redGEM: expanding reduced subnetworks into a consistent global network

The redGEM [27] algorithm combines graph-based search and optimization methods to reduce the size of metabolic networks, focusing on selected metabolic subsystems. In contrast to the top-down approaches discussed above, which start from the full GSM, it represents a bottom-up approach that begins with a subsystem of interest and extends it systematically. redGEM aims to not only retain the relevant stoichiometry of the full GSM in a reduced core model, but aims to produce a set of reduced networks that are consistent with the original GSM in terms of flux profiles, essential genes and reactions, thermodynamically feasible ranges of metabolites, and ranges of Gibbs free energy of reactions.

The inputs to the redGEM algorithm are: the GSM; pre-specified metabolic subsystems (sets of reactions and metabolites) of interest; the maximal degree of connection (D) between the subsystems to be retained, set by the user, where D_{ij} is defined as the number of consecutive reactions that link subsystems S_i and S_j; and available physiological information such as organism specific cofactor pairs, possible by-products, or relevant extracellular metabolites and media. In its first phase, redGEM uses graph algorithms to connect the metabolic subsystems of interest and maintain the network linkage. In the second phase, MILP is used to restore key cellular functionality and properties, such as biomass production and by-product yields.

The core stoichiometric matrix is derived for each subsystem excluding all cofactor pairs and inorganic compounds. For each subsystem pair S_i and S_j, a graph search is then performed to find all links up to length D between the subsystems. This will not find any reactions and metabolites between two subsystems if the length of the shortest path between the subsystems S_i and S_j is greater than D. The core network for redGEM is then composed of all metabolites and reactions from all subsystems along with connecting reactions and metabolites; at this stage, cofactors and inorganic compounds are added back to the model. Depending on the pre-specified subsystems, the core network might have metabolic routes to generate some biomass components; however, it will likely not be capable of producing all biomass components as defined in the full GSM. Moreover, the graph theoretical approaches used to generate the core network do not ensure mass

conservation, and hence, the resulting core model may consist of reactions that are not elementally balanced or are not able to carry flux in steady state.

To correct this, redGEM employs lumpGEM [47] to expand the core network to the biomass building blocks through elementally balanced lumped reactions. lumpGEM is an MILP algorithm developed to identify elementally balanced minimal subnetworks capable of producing target compounds from a set of defined core metabolites. In this analysis, all metabolites of the core network generated through redGEM are regarded as precursor metabolites. Each subnetwork computed by lumpGEM is represented as a lumped reaction, which captures the overall stoichiometry of the subnetwork while preserving the elemental balance. lumpGEM ensures the thermodynamic consistency of the lumped pathways; moreover, it is able to identify alternative subnetworks for the synthesis of the same biomass component and can hence help to retain some level of metabolic and physiological flexibility in the reduced model.

Like the two previously discussed network reduction approaches, redGEM was also applied to a GSM of *E. coli* (iJO1366, consisting of 1863 reactions and 1136 metabolites). The input specified six central metabolism subsystems, consisting of glycolysis, pentose phosphate pathway, citric acid cycle, glyoxylate cycle, pyruvate metabolism, and oxidative phosphorylation. These subsystems in total consisted of 185 reaction and 126 metabolites. The subsystems were then connected to generate core networks with different degrees of connection (D). At $D = 1$, the core network generated by redGEM consisted of 243 reactions and 156 metabolites. These 156 metabolites from the core network were provided as precursors to lumpGEM, which generated 254 unique lumped reactions for the synthesis of 79 biomass components, while the remaining 23 biomass components were already present in the core network. In total, the reduced model consisted of 497 reactions and 235 metabolites, and was capable of producing all biomass components, with the same growth rate as the full GSM.

In contrast to the other methods discussed so far, redGEM is not aimed at identifying a minimal network and retains a much higher metabolic flexibility than the alternative methods (with the exception of fully coupled enzyme subsets). This metabolic flexibility is reflected in a larger size of the compressed network, e.g., in comparison to those generated by NetworkReducer and minNW. redGEM may thus be better suited to provide reduced base models for studies that aim to capture a wide range of metabolic versatility. Users should keep in mind that the limitation of subsystem connections to a maximal length will likely lose some network flexibility and may impact the networks faithful representation of the metabolic and phenotypic flexibility of the full GSM.

DRUM: representing subnetworks through collapsed EFMs

Dynamic reduction of unbalanced metabolism (DRUM) [48] differs from the previously discussed algorithms by taking a more radical approach to network reduction. DRUM splits the network into subnetworks, where each subnetwork is assumed to be at (quasi-)steady state and is eventually represented by a single lumped reaction; the metabolites that connect the subnetworks are assumed not to be in steady state and are thus allowed to accumulate and to behave dynamically. DRUM was developed specifically for the dynamic modeling of temporally varying metabolic processes that lead to the accumulation of intracellular metabolites, as exemplified by the circadian cycle of microalgae. Despite the specific motivation for its development, DRUM can also be used for the same general purpose that motivated the network reduction methodologies discussed above (with the exception of minimal reaction sets), the generation of metabolic models small enough to be amenable to a range of computationally expensive modeling approaches.

The first stage of the DRUM approach consists of splitting the metabolic network into user-defined subnetworks, each consisting of highly interconnected reactions. Starting with subsystems allows DRUM to calculate EFMs, which would be computationally challenging or even impossible when applied to full GSMs. EFMs are calculated separately for each subnetwork, treating inputs and outputs of the subnetwork as external metabolites. Each subnetwork is then reduced to a set of macroscopic reactions, each of which reflects the net stoichiometry of one EFM. Based on user-defined criteria, the set of EFMs for each subnetwork can be reduced to those EFMs deemed biologically most relevant; e.g., the biomass producing subnetwork may be represented only by the EFM with the highest biomass yield. To further transform the reduced network into a dynamic metabolic model, kinetics for the macroscopic reactions are obtained, and the network is transformed into a system of ordinary differential equations (ODEs).

This approach was applied to the metabolic network of the unicellular photoautotrophic microalga *Tisochrysis lutea*, containing 157 metabolites and 162 reactions. Metabolic reactions were grouped by metabolic functions to obtain six subnetworks corresponding to photosynthesis, upper part of glycolysis, carbohydrate synthesis, lower part of glycolysis, lipid synthesis, and biomass synthesis. The latter subnetwork included reactions for the synthesis of protein, DNA, RNA, and chlorophyll. All subnetworks were assumed to be at quasi-steady state. Metabolites that connected the subnetworks, such as GAP, PEP, reductants, energy components, end products (such as lipids and biomass), and inputs (such as light, CO_2, and Pi) were assumed not to be at steady state and, hence, were treated as external metabolites for each

subnetwork. The biomass production subnetwork had 24 EFMs, out of which the EFM with the highest carbon yield was selected. In total, the reduced network consisted of 8 macroscopic reactions and 16 metabolites. It was converted to an ODE system, assuming simple proportional kinetics. The reduced model was fitted to experimental data to estimate the kinetic parameter values, and was then used to simulate the accumulation of lipids and carbohydrates during a day/night cycle.

Of all the methods reviewed, DRUM results in the smallest reduced network. This is because each user-specified subsystem is typically represented by a small number of EFMs. This reduction is achieved through the separation of the subsystems, which allows the metabolites that connect the subsystems to accumulate. Thus, while DRUM conserves the subsystems, the fluxes connecting them may violate the steady-state assumption; this may lead to systematic differences between the behavior of the reduced and the full system.

Discussion

In this work, we have reviewed six metabolic network reduction methods. Table 2 briefly summarizes the features of each method. The most simple and exact network reduction method is the "loss-less compression" of GSMs through the lumping of fully coupled enzyme subsets [37]. This method conserves all potential phenotypes in steady state, although some metabolites or reactions of interest may not be represented explicitly in the reduced network.

The minimal reaction sets introduced by Burgard et al. [40] were not specifically aimed at reducing network size, but rather aimed at extracting the minimal reaction set needed to meet the growth demand. This work is also important from a biological point of view, because although GSMs comprise up to several thousand reactions, only a relatively small proportion of reactions are utilized to support growth and biomass production in a given environment. While the minimal reaction sets [40] retain certain cellular functionalities (in particular the growth rate in a specified environment), this method does not retain reactions and metabolites of interest in the reduced network, does not conserve network redundancy or *dof*, and does not retain any metabolic flexibility. NetworkReducer to some degree adds these features to the algorithm. It not only allows to retain desired metabolites and reactions, but also allows to retain a desired amount of network flexibility by adding the criterion of a minimal *dof*. In addition, it allows to protect multiple phenotypes through added functional constraints. Neither the minimal reaction set approach nor NetworkReducer can capture all minimal subnetworks, as there may exist different subnetworks that fulfill the requirements and have the same number of active reactions. The minNW method by Röhl and Bockmayr [30] overcomes these limitations to some extent. It achieves the same goals as NetworkReducer (only faster and more accurately), but further evaluates all minimal subnetworks by excluding already found subnetworks in each iteration. All three methods will often exclude phenotypes available to the full GSM from the reduced model unless these are explicitly protected in the algorithms' inputs. Thus, although NetworkReducer and especially minNW with its alternative optima retain some degree of network flexibility, much of the systems phenotypic flexibility may get lost in the process.

Minimal reaction sets [40], NetworkReducer [29], and MinNW [30] are top-down approaches, scaling a larger network down to a smaller size. In contrast, redGEM and DRUM are bottom-up approaches, starting from smaller subsystems and expanding them to re-establish the properties of a full GSM. redGEM and DRUM also add some variety to the reduction methods, as they are not completely based on LP/MILP approach, but make use of graph algorithms and EFMs, respectively.

Both redGEM and DRUM need user-defined subsystems of the network as input data. This presents a challenge for their application, as the manual definition of subsystems requires in-depth metabolic knowledge on the organism of

Table 2 Comparison of approaches and features of model reduction methods discussed in this review

Algorithm	Method	Reaction protection	Metabolite protection	*dof* protection	Growth/biomass support	Computes all minimal subnetworks	Multiple phenotype protection	Flux variability protection
Enzyme subsets [37]		–	–	+	+	+	+	+
Minimal reaction sets [40]	MILP	–	–	–	+	+	–	–
NetworkReducer [29]	LP/FVA	+	+	+	+	–	+	+
MinNW [30]	MILP	+	+	+	+	+	+	–
redGEM [27]	Graph alg. and MILP	+	+	+	+	+	+	+
DRUM [48]	EFM	–	+	–	+	+	+	–

interest, and as a high degree of interconnectedness especially in central metabolism makes the definition of subsystems difficult. Moreover, both algorithms achieve the network reduction through an approximate treatment of the reactions that interconnect subsystems: redGEM limits the length of subsystem connections, while DRUM drops the steady-state assumption for interconnecting metabolites and allows them to accumulate. This approximate treatment of subsystem connections may alter the metabolic and phenotypic flexibility of the reduced model in comparison to the full GSM. Thus, modelers that aim to apply metabolic network reduction methods have to choose between relatively rigid representations that "project" the full GSM on subnetworks that represent only protected functionalities (minimal reaction sets, NetworkReducer, minNW) and more flexible representations that only approximately reflect the functionalities of the full GSM (redGEM, DRUM).

Acknowledgements This work was funded by the Deutsche Forschungsgemeinschaft (German Research Foundation) through grants CRC 680 and CRC 1310.

Compliance with ethical standards

Conflict of interest The authors declare that there are no competing interests associated with the manuscript.

References

1. Jeske L, Placzek S, Schomburg I, Chang A, Schomburg D (2019) BRENDA in 2019: a European ELIXIR core data resource. Nucleic Acids Res 47:D542–D549
2. Heinrich R, Schuster S (1996) The regulation of cellular systems. Chapman and Hall, London
3. Raman K, Chandra N (2009) Flux balance analysis of biological systems: applications and challenges. Brief Bioinform 10:435
4. Gianchandani EP, Chavali AK, Papin JA (2010) The application of flux balance analysis in systems biology. Wiley interdisciplinary reviews: systems biology and medicine. Wiley, Hoboken, pp 372–382
5. Zhang C, Hua Q (2016) Applications of genome-scale metabolic models in biotechnology and systems medicine. Front Physiol 6:413
6. Lewis NE, Nagarajan H, Palsson BO (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. Nat Rev Microbiol 10:291–305
7. Edwards JS, Palsson BO (2000) The *Escherichia coli* MG1655 *in silico* Metabolic Genotype: its Definition, Characteristics and Capabilities. Proc Natl Acad Sci USA 97:5528–5533
8. Poolman MG, Miguet L, Sweetlove LJ, Fell DA (2009) A genome-scale metabolic model of *Arabidopsis* and some of its properties. Plant Physiol 151:1570–1581
9. Boyle NR, Morgan JA (2009) Flux balance analysis of primary metabolism in *Chlamydomonas reinhardtii*. BMC Syst Biol 3:4
10. Zomorrodi AR, Suthers PF, Ranganathan S, Maranas CD (2012) Mathematical optimization applications in metabolic networks. Metab Eng 14:672–686
11. Bockmayr A, Müller AC (2013) Fast thermodynamically constrained flux variability analysis. Bioinformatics 29:903–909
12. Villanova V et al (2017) Investigating mixotrophic metabolism in the model diatom *Phaeodactylum tricornutum*. Philos Trans R Soc 372:20160404. https://doi.org/10.1098/rstb.2016.0404
13. Singh D, Carlson R, Fell D, Poolman M (2015) Modelling metabolism of the diatom *Phaeodactylum tricornutum*. Biochem Soc Trans 43:1182–1186
14. Ye C, Xu N, Dong C, Ye Y, Zou X, Chen X, Guo F, Liu J (2017) IMGMD: a platform for the integration and standardisation of *in silico* microbial genome-scale metabolic models. Sci Rep 7:727
15. Schuster S, Hilgetag C (1994) On Elementary Flux Modes in Biochemical Systems at Steady State. Journal of Biological Systems 2:165–182
16. Klamt S, Gilles ED (2004) Minimal cut sets in biochemical reaction networks. Bioinformatics 20:226–234
17. Reed JL, Vo TD, Schilling CH, Palsson BO (2003) An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). Genome Biol 4:54
18. Feist FM et al (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol 3:121
19. Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Nam H, Palsson BO (2011) A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism-2011. Mol Syst Biol 7:535. https://doi.org/10.1038/msb.2011.65
20. Weaver DS, Keseler IM, Mackie A, Paulsen IT, Karp PD (2014) A genome-scale metabolic flux model of *Escherichia coli* K-12 derived from the EcoCyc database. BMC Syst Biol 8:79
21. Monk JM et al (2017) iML1515, a knowledgebase that computes Escherichia coli traits. Nat Biotechnol 35:904–908
22. Beg QK, Vazquez A, Ernst J, de Menezes MA, Bar-Joseph Z, Barabsi A-L, Oltvai ZN (2007) Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity. Proc Natl Acad Sci 104:12663–12668
23. Goelzer A et al (2015) Quantitative prediction of genome-wide resource allocation in bacteria. Metab Eng 32:232–243
24. O'Brien EJ, Lerman JA, Chang RL, Hyduke DR, Palsson B (2013) Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. Mol Syst Biol 9:693. https://doi.org/10.1038/msb.2013.52
25. Saa PA, Nielsen LK (2017) Formulation, construction and analysis of kinetic models of metabolism: a review of modelling frameworks. Biotechnol Adv 35:981–1003
26. Soh KC, Miskovic L, Hatzimanikatis V (2012) From network models to network responses: integration of thermodynamic and kinetic properties of yeast genome-scale metabolic networks. FEMS Yeast Res 12:129–143
27. Ataman M, Georgios F, Vassily H (2017) redGEM: systematic reduction and analysis of genome-scale metabolic reconstructions for development of consistent core metabolic models. PLoS Comput Biol 13:1–22
28. Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T (2010) Interdependence of cell growth and gene expression: origins and consequences. Science 330:1099–1102
29. Erdrich P, Steuer R, Klamt S (2015) An algorithm for the reduction of genome-scale metabolic network models to meaningful core models. BMC Syst Biol 9:48
30. Röhl A, Bockmayr A (2017) A mixed-integer linear programming approach to the reduction of genome-scale metabolic networks. BMC Bioinformatics 18:2
31. Becker SA, Palsson BO (2008) Context-specific metabolic networks are consistent with experiments. PLoS Comput Biol 4:1–10
32. Shlomi T, Cabili M, Herrgard M, Palsson B, Ruppin E (2008) Network-based prediction of human tissue-specific metabolism. Nat Biotechnol 26:1003–1010

33. Zur H, Ruppin E, Shlomi T (2010) iMAT: an integrative metabolic analysis tool. Bioinformatics 26:3140–3142
34. Jerby L, Shlomi T, Ruppin E (2010) Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. Molecular Systems Biology 6:401
35. Robaina Estevez S, Nikoloski Z (2014) Generalized framework for context-specific metabolic model extraction methods. Front Plant Sci 5:491
36. Opdam S, Richelle A, Kellman B, Li S, Zielinski DC, Lewis NE (2017) A systematic evaluation of methods for tailoring genome-scale metabolic models. Cell Syst 4:318–329
37. Pfeiffer T, Sanchez-Valdenebro I, Nuno JC, Montero F, Schuster S (1999) Metatool: for studying metabolic networks. Bioinformatics 15:251–257
38. Burgard AP, Nikolaev EV, Schilling CH, Maranas CD (2004) Flux coupling analysis of genome-scale metabolic network reconstructions. Genome Res 14:301–312
39. Larhlimi A, David L, Selbig J, Bockmayr A (2012) F2C2: a fast tool for the computation of flux coupling in genome-scale metabolic networks. BMC Bioinform 13:57
40. Burgard AP, Vaidyaraman S, Maranas CD (2001) Minimal Reaction Sets for *Escherichia coli* metabolism under different growth requirements and uptake environments. Biotechnol Prog 17:791–797
41. Jonnalagadda S, Balagurunathan B, Srinivasan R (2011) Graph theory augmented math programming approach to identify minimal reaction sets in metabolic networks. Comput Chem Eng 35:2366–2377
42. Jonnalagadda S, Srinivasan R (2014) An efficient graph theory based method to identify every minimal reaction set in a metabolic network. BMC Syst Biol 8:28
43. Holzhütter HG (2004) The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. Eur J Biochem 271:2905–2922
44. Adadi R, Volkmer B, Milo R, Heinemann M, Shlomi T (2012) Prediction of microbial growth rate versus biomass yield by a metabolic network with kinetic parameters. PLoS Comput Biol 8:e1002575
45. Basan M, Hui S, Okano H, Zhang Z, Shen Y, Williamson JR, Hwa T (2015) Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. Nature 528:99–104
46. Mahadevan R, Schilling CH (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab Eng 5:264–276
47. Ataman M, Vassily H (2017) lumpGEM: systematic generation of subnetworks and elementally balanced lumped reactions for the biosynthesis of target metabolites. PLoS Comput Biol 13:1–21
48. Baroukh C, Muñoz-Tamayo R, Steyer J-P, Bernard O (2014) DRUM: a new framework for metabolic modeling under non-balanced growth. application to the carbon metabolism of unicellular microalgae. PLoS One 9:1–15