Computers in Biology and Medicine 105 (2019) 64–71

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

A simplified metabolic network reconstruction to promote understanding and development of flux balance analysis tools

Kristopher D. Rawls[a,1], Bonnie V. Dougherty[a,1], Edik M. Blais[a,1], Ethan Stancliffe[a,2], Glynis L. Kolling[a,b], Kalyan Vinnakota[c], Venkat R. Pannala[c], Anders Wallqvist[c], Jason A. Papin[a,*]

[a] Department of Biomedical Engineering, University of Virginia, Charlottesville, VA, 22908, USA
[b] Department of Medicine, Division of Infectious Diseases and International Health, University of Virginia, Charlottesville, VA, 22908, USA
[c] Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Materiel Command, Fort Detrick, Maryland, USA

ARTICLE INFO

Keywords:
Systems biology
Metabolic networks
Flux balance analysis
Computational modeling
Metabolic engineering

ABSTRACT

GEnome-scale Network REconstructions (GENREs) mathematically describe metabolic reactions of an organism or a specific cell type. GENREs can be used with a number of constraint-based reconstruction and analysis (COBRA) methods to make computational predictions on how a system changes in different environments. We created a simplified GENRE (referred to as *iSIM*) that captures central energy metabolism with nine metabolic reactions to illustrate the use of and promote the understanding of GENREs and constraint-based methods. We demonstrate the simulation of single and double gene deletions, flux variability analysis (FVA), and test a number of metabolic tasks with the GENRE. Code to perform these analyses is provided in Python, R, and MATLAB. Finally, with *iSIM* as a guide, we demonstrate how inaccuracies in GENREs can limit their use in the interrogation of energy metabolism.

1. Introduction

Genome-scale metabolic network reconstructions (GENREs) have emerged as powerful tools for the contextualization of high-throughput data, to guide discovery in biological systems, and to simulate the effect of genetic and environmental perturbations [1,2]. While sophisticated computational tools have been developed to facilitate the understanding and use of these network models [3,4], there remains a significant challenge for the development of intuition for how associated modeling analyses are applied and can be used. While several “simple” networks have been developed [5,6], these networks often fail to capture key features of energy metabolism and are too large (> 100 reactions) to readily develop intuition for associated modeling analyses.

GENREs represent metabolic reactions and corresponding genes and capture the stoichiometric relationships between metabolites and associated chemical transformations. With GENREs as a foundation, constraint-based reconstruction and analysis (COBRA) methods allow for computational predictions of metabolic phenotypes [3]. Constraints placed on the GENRE are based on genetic, environmental, or thermodynamic factors and reduce the space of possible phenotypes of the system. To constrain the network, experimental data such as transcriptomics or metabolomics can be integrated into GENREs, adding specificity to the network representing a particular environment or cell state. Recently developed methods [7–11] allow for the integration of context-specific ‘omics datasets, constraining general models of metabolism for more context-specific computational predictions.

Applying constraints to the model allows the generation of novel scientific hypotheses about the underlying biology of the system studied; therefore, it is important for the reconstructions to maintain mass balance and consider thermodynamic constraints appropriately. For example, a well-documented problem for several published reconstructions is the infinite synthesis of ATP due to thermodynamically infeasible loops [12,13]. A key challenge for the development and understanding of methods for the analysis of metabolic networks is that they are inherently complex and difficult to interrogate. Such complexity makes it a challenge to understand the impact of particular characteristics of a new method. Simple networks can be a crucial tool for new method development. To date, simple metabolic networks have captured representative catabolic and anabolic reactions, but other key characteristics of real metabolic networks are often neglected.

* Corresponding author.
E-mail address: papin@virginia.edu (J.A. Papin).
[1] These authors contributed equally to this work.
[2] Author is currently at the Division of Biology and Biomedical Sciences at Washington University in Saint. Louis. 660 South Euclid Avenue, St. Louis, MO 63110.

https://doi.org/10.1016/j.compbiomed.2018.12.010
Received 26 September 2018; Received in revised form 11 December 2018; Accepted 13 December 2018

Here we present *iSIM*, a simplified metabolic network that captures key features of energy metabolism that are frequently neglected in other similar simple network models. We demonstrate the application of several constraint-based techniques to this simplified network reconstruction, including how ATP production and metabolic flux ranges for each reaction in the network are altered through simulations of single and double gene deletions. Finally, we illustrate how *iSIM* can be used to help understand how thermodynamic errors can arise in metabolic networks and affect computational predictions. We provide source code for analyses in Python, R, and MATLAB. In summary, *iSIM* serves as a tool both for understanding and benchmarking GENREs and their associated methods. *iSIM* also demonstrates the need for quality control measures that are important to consider when reconstructing and analyzing genome-scale metabolic networks and in the development of constraint-based methods.

2. Methods

2.1. Stoichiometric matrix

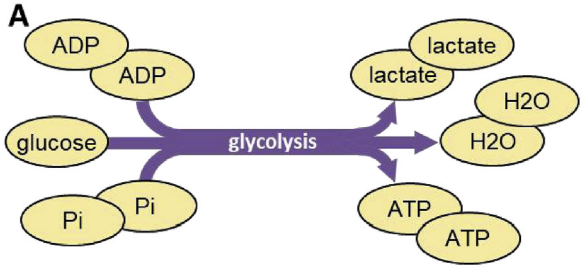
The stoichiometric matrix (*S* matrix) captures the stoichiometric coefficients for metabolic reactions in the GENRE with each metabolite represented in a row and each reaction in a column. For example, the simplified representation of glycolysis in Fig. 1A is represented as one column in the S matrix, and the metabolites in Fig. 1B are represented as rows in the matrix. The elements of the matrix are the stoichiometric coefficients of the metabolite consumed (negative) or produced (positive) in the reaction. Fig. 1C is a visual representation of an S matrix for the human GENRE [14] which contains over 8000 reactions, demonstrating how individual metabolites participate in many reactions in a network. Of note, there are several reactions in the general human metabolic network reconstruction that contain many metabolites (indicated by a large number of colors in one column) which represent lumped reactions or reactions that represent a large number of metabolites such as lipid synthesis and degradation.

2.2. Gene-protein-reaction relationship rules

GENREs contain gene-protein-reaction (GPR) rules that describe a gene's relationship to a protein and the reaction catalyzed by the protein. These GPR rules allow for the generation of tissue-specific or organism specific models for further analyses. GPR rules are organized with Boolean logic relationships between genes, proteins, and reactions. Fig. 2 gives examples of different Boolean relationships between genes and reactions. Using these GPR rules, simulations of gene deletions can systematically account for the removal of specific genes and their associated reactions from the network to determine the effect on the system. For example, in Fig. 2C, if gene C2 is removed, reaction C is also removed because both genes C1 and C2 are essential for reaction C to function. This is not the case in the reaction illustrated in Fig. 2B, as either gene B1 or B2 catalyze reaction B. Therefore, if gene B1 is removed, reaction B can still occur. GPR rules also allow for the integration of gene or protein expression data with GENREs, where genes or proteins can be turned “on” or “off” based on expression data. Fig. 2D shows a complex GPR rule, where the gene D1 is needed, and either D2a or D2b. Each of the different types of GPR rules and their corresponding reaction is summarized in Fig. 2E.

2.3. Flux Balance Analysis

The *S*-matrix is often an underdetermined system with more reactions than metabolites, meaning there are many potential solutions to the system; consequently, constraints are used to further reduce the solution space. Optimization methods are used to find a solution that satisfies the constraints and is of particular interest, for example, the flux distribution that corresponds to maximum biomass yield. . Flux Balance Analysis (FBA) is a constraint-based method that calculates the maximum possible flux through a specified reaction, known as the objective function, subject to constraints on reaction fluxes. Given a stoichiometric matrix (S), lower and upper bounds (v_{lb} and v_{ub}) on reaction fluxes (v), and an objective (v_{obj}), FBA uses linear programming to solve the following optimization problem:



B

Metabolite	Abbreviation	Coefficient
Adenosine diphosphate	ADP	-2
Phosphate	Pi	-2
Glucose	Glucose	-1
Lactate	Lactate	2
Water	H20	2
Adenosine triphosphate	ATP	2

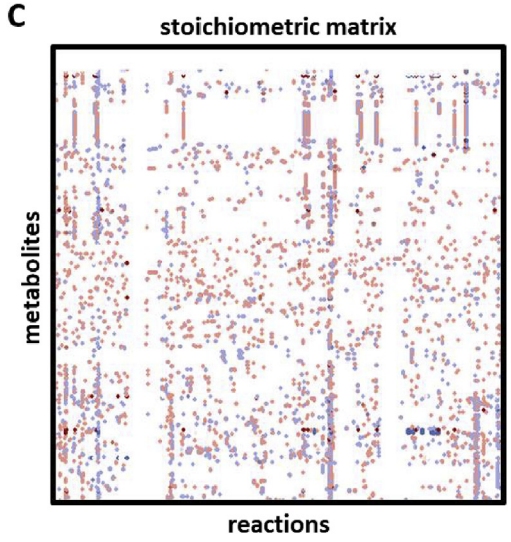


Fig. 1. The stoichiometric matrix captures stoichiometric relationships between metabolites in reactions. (**A**) Representation of a simplified glycolysis reaction, catalyzing the conversion of one unit of glucose and two units of adenosine diphosphate and two units of phosphate into two water molecules, two units of ATP and two units of lactate. (**B**) Stoichiometric coefficients describing the amounts of metabolites consumed (blue) and produced (red) by this simplified representation of glycolysis. (**C**) The stoichiometric matrix (S matrix) is a mathematical representation of reactions like the example in **A** formatted as a sparse matrix where each column represents a reaction and each row represents a metabolite. Each point in the sparse matrix accounts for the stoichiometric coefficient of each reaction-metabolite pair, where the color represents whether the metabolite was consumed (blue) or produced (red). The S matrix in C represents thousands of reactions within the human metabolic network [14].

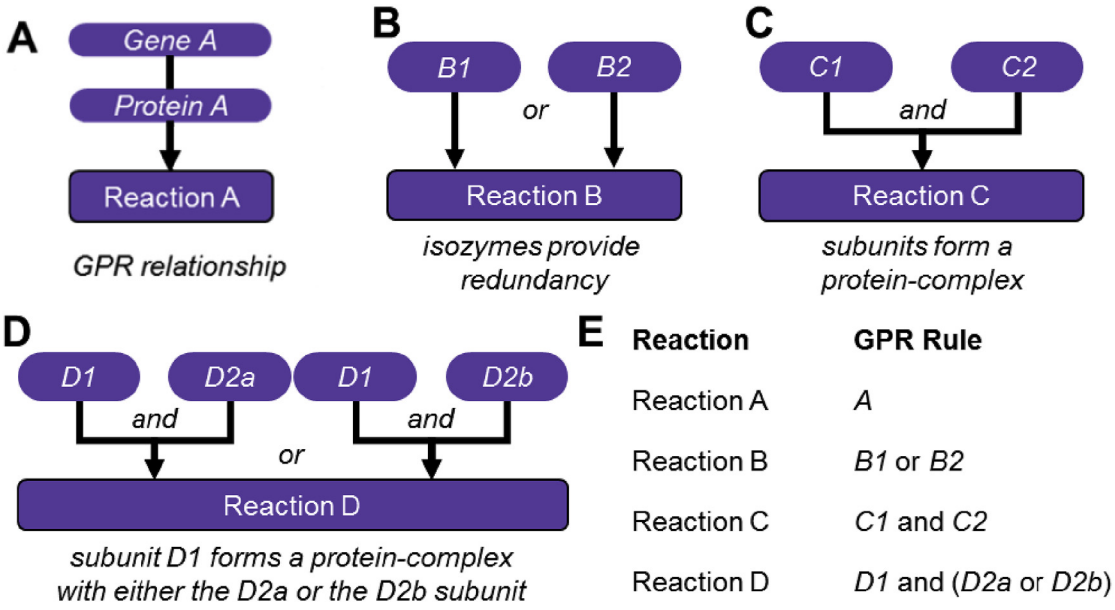


Fig. 2. Gene-protein-reaction (GPR) rules describe the relationship between genotype and phenotype. (**A**) Example of a GPR rule representing an enzymatic reaction catalyzed by the protein product of a single gene. Genes within GPR rules are often represented as Entrez gene identifiers, Ensembl transcripts, UniProt proteins, or Enzyme Commission numbers. (**B**) Example of a redundant GPR rule where either protein *B1* or protein *B2* can independently catalyze the same function. In this case, these isozymes are separated by an "or" statement in the GPR rule. (**C**) Example of a complex GPR rule where both C1 and C2 are required for the catalytic reaction to occur. In this case, two non-redundant subunits that form a protein complex are separated by an "and" statement. (**D**) Example of a complex GPR with redundancies where D1 can form a protein complex with either D2a or D2b. In this case, the GPR rule can be separated by unique protein complexes or first by subunits then by redundancies as represented in E. (**E**) Table summarizing genotype-phenotype relationships from A-D as Boolean GPR rules.

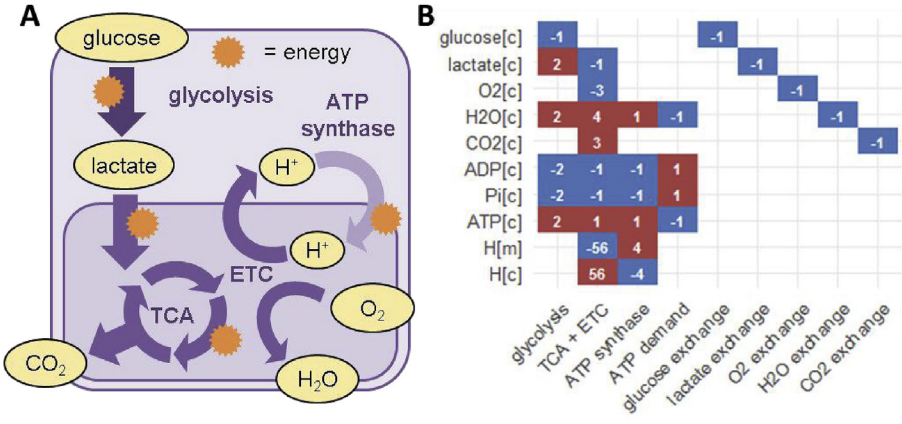


Fig. 3. *iSIM*, a prototypic metabolic network that represents simplified energy metabolism. (A) *iSIM* represents two major catabolic pathways for glucose that generate cellular energy via the addition of a phosphate group to ADP to generate ATP, indicated here in orange. Glucose is broken down into lactate while producing ATP and lactate is broken down to carbon dioxide through the tricarboxylic acid (TCA) cycle and the electron transport chain (ETC). The proton concentration gradient created by the ETC is then utilized to generate more ATP. **(B)** The stoichiometric matrix of *iSIM*. Each column represents a reaction and each row represents a metabolite, where the brackets indicate the cellular localization of metabolite: [c], cytosol; [m], mitochondria. Negative values (blue) indicate a metabolite is being consumed in the reaction and positive values (red) indicate a metabolite is produced. The first 4 columns represent biochemical reactions while the last 5 columns represent exchange reactions that allow metabolites to enter or leave the network.

$$max\ v$$

$$\text{s.t.} \quad S \cdot v = \frac{dC}{dt} = 0. \tag{1}$$

$$v_{lb} \leq v \leq v_{ub}\,. \tag{2}$$

Equation (1) describes the mass balance constraints, where S is the stoichiometric matrix multiplied by a flux vector (v). This equation represents the steady-state assumption inherent in many FBA simulations. Equation (2) describes reaction constraints applied to the FBA problem. These constraints are often applied to simulate conditions relevant to a biological question. For example, these constraints may be defined to represent a particular media condition (presence of glucose in the media, absence of fructose in the media) where the bounds are set to zero for metabolites absent in the media. Furthermore, these constraints are also often used to capture irreversibility of reactions as known (i.e., one of the bounds is set to zero so that flux through the corresponding reaction can only go in one direction).

2.4. Flux Variability Analysis

Flux variability analysis (FVA) is a constraint-based modeling technique that calculates the range of flux through each reaction that are feasible given a set of constraints. Unlike FBA, which is often used to calculate a single flux distribution to solve the associated optimization problem, FVA results in the calculation of a range of feasible fluxes through each reaction which still results in a defined value of the objective function. FVA is implemented in two steps; first, FVA constrains a metabolic network to require a minimum amount of flux through an objective function, which is usually a percentage of the maximum value determined by FBA. In the second step, FVA optimizes for the flux through each reaction in the network. This analysis enables the identification of the minimum and maximum flux values for each reaction in

Table 1
iSIM metabolic tasks simulate cellular functions by specifying lower bound values (v_{lb}) and upper bound values (v_{ub}) for individual reactions. (*) Minimum required ATP yields for feasible metabolic tasks are sub-optimal (less than maximum). (**) Infeasible metabolic tasks require unrealistic flux through ATP demand that should fail when simulated.

Metabolic Task	Reaction	vlb	vub
(1) glucose catabolism	glucose exchange	−1	∞
	lactate exchange	2	∞
	H_2O exchange	0	∞
	ATP demand*	1	∞
(2) glucose catabolism without oxygen	glucose exchange	−1	∞
	lactate exchange	2	∞
	H_2O exchange	0	∞
	ATP demand**	10	∞
(3) glucose catabolism with oxygen	glucose exchange	−1	∞
	O_2 exchange	-∞	∞
	CO_2 exchange	0	∞
	H_2O exchange	0	∞
	ATP demand*	10	∞

the network, providing feasible flux ranges that are consistent with the constraints imposed in the first step of FVA.

3. Results

3.1. iSIM: a simple metabolic network

We have created a prototypic GENRE, *iSIM,* to represent central energy metabolism (Fig. 3). *iSIM* includes 9 unique metabolites in the cytosolic and mitochondrial compartments and five exchange reactions for the consumption or secretion of glucose, lactate, O_2, H_2O, and CO_2 (Fig. 3B). To simplify the network, select linear metabolic reactions are aggregated as lumped reactions that do not explicitly include all the intermediate steps within a pathway. For example, the first metabolic reaction, 'glycolysis' (Fig. 3), represents the reactions involved in conversion of glucose to lactate and the generation of ATP, which includes multiple intermediates not considered here. The second metabolic reaction, 'TCA + ETC,' represents the tricarboxylic acid (TCA) cycle which produces CO_2 and H_2O coupled with the generation of the mitochondrial proton gradient via the electron transport chain (ETC). As a result of the simplification of reactions, there is only one metabolite in the mitochondrial compartment, H+, represented as H[m] (Fig. 3B). The third reaction, catalyzed by the protein complex ATP synthase, utilizes the proton concentration gradient generated by the ETC to regenerate ATP from ADP. The fourth reaction, ATP demand, represents the consumption of energy through the hydrolysis of ATP to ADP through an ATPase. To explore basic concepts of bioenergetics, maximization of flux through the ATP demand reaction was used as an objective function. The other three metabolic reactions described above facilitate the regeneration of ATP from ADP to fuel the ATP demand reaction. The remaining five reactions, glucose exchange, lactate exchange, O_2 exchange, H_2O exchange, and CO_2 exchange constrain the uptake and secretion of these metabolites in the network. Here, we use *iSIM* to demonstrate the utility of computational methods and GENREs, including the use of metabolic tasks to demonstrate network functionality, *in silico* genetic perturbations to identify genes of interest, and flux variability analysis to analyze possible flux distributions through a metabolic network.

3.2. Recapitulating biological functions with metabolic tasks

To test model functionality, organism-specific biological metabolic processes are summarized as metabolic tasks or metabolic objectives that the model should be able to perform. Metabolic tasks are formulated by specifying a set of input and output metabolites, such as an input of glucose and output of ATP, with constraints for upper and lower bounds of flux. If the model can produce a feasible flux distribution, then the model is capable of completing the task [15]. Flux in the model is defined as 1 unit = 1 fmol/cell/hour. For example, given that catabolism of glucose to lactate through the glycolytic pathway is a function often present in metabolic networks, we can formulate a metabolic task that allows uptake of 1 unit of glucose and requires secretion of 2 units of lactate and 2 units of water. This metabolic task could be tested in a given model by setting a lower bound of −1 on the glucose exchange reaction and a lower bound of +2 on the lactate and water exchange reactions. If there is a feasible solution for the model given the constraints, the task "passes". If not, we have identified an area for model curation or improvement. For each metabolic task, exchange metabolites in the network for which upper and lower bounds are not explicitly stated are set to 0, unless a parameter is included as an output which allows for the excretion of all metabolites. This parameter should be included in a metabolic task when not all potential metabolic byproducts are known.

To demonstrate the functionality of *iSIM*, we have curated three metabolic tasks representing glucose catabolism with ATP demand, the requirement of O_2 for large ATP production, and oxidative phosphorylation (Table 1). Building on the glucose catabolism metabolic task from above, we can require minimum activity through the ATP demand reaction, representing production and then consumption of ATP (Task 1). This metabolic task uses a lower bound of −1 on the glucose exchange reaction, a lower bound of +2 on the lactate and water exchange reactions, and a lower bound of +2 on the ATP demand reaction. To capture the necessity of O_2 for increased ATP production per unit of glucose (Task 2), we can formulate metabolic task that should not pass. If the ATP demand reaction from above has a lower bound of +10, the task should fail since glycolysis can only produce 2 ATP per 1 unit of glucose. However, if we remove the requirement for the secretion of lactate (lower bound to 0) and allow O_2 uptake (lower bound less than 0) (Task 3), the metabolic network will use the TCA + ETC reaction to meet the +10 lower bound on ATP flux. Table 1 provides a summary of lower and upper bounds for these metabolic tasks; the tasks marked as infeasible could not be completed or failed. These three metabolic tasks demonstrate the functionality of *iSIM* in capturing a simplified representation of central energy metabolism.

Metabolic tasks provide a way to demonstrate the functionality of the model and identify areas for improvement. Numerous reconstructions are published with a list of metabolic tasks or metabolic objectives that the model can perform, demonstrating the validity and applicability of the reconstruction [14,16,17]. If a task is unable to pass when the biological function is known to be present, the task identifies an area where the model may be misrepresenting or missing a metabolic reaction, allowing for further expansion and/or curation of the model.

3.3. Genetic perturbations and gene essentiality

To identify gene targets of interest in a metabolic network, GENREs are analyzed with constraint-based methods to simulate the effects of gene knockouts using the gene-protein-reaction (GPR) relationships. In an *in silico* gene knockout screen, each gene in the model is removed through the identification of reactions that are catalyzed by that gene through the GPR relationships and these associated reactions are disabled by setting the lower and upper bounds to 0. In *iSIM*, for simplicity, each reaction is associated with only one gene. Therefore, a gene knockout screen will remove each reaction systematically and evaluate the effect on the objective function of ATP demand. Table 2 shows the maximum values through the objective function (ATP demand), after disabling each gene and the reaction associated with that gene. Removing the gene associated with glycolysis (phosphofructokinase (PFK)), disables the glycolysis reaction which in turn completely blocks ATP production since glucose is the only carbon source entering the system. The same holds for removing the gene for glucose exchange,

Table 2
Maximum possible flux through the ATP demand reaction predicted by FBA after deleting individual reactions from *iSIM*.

Gene Deletion	Reaction Deletion	ATP demand (fmol/cell/hr)
GLUT	glucose exchange	0
LDH	lactate exchange	32
ETC	O_2 exchange	2
AQP	H_2O exchange	2
CO2	CO_2 exchange	2
PFK	glycolysis	0
CS	TCA + ETC	2
ATPV	ATP synthase	2
MYH2	ATP demand	0

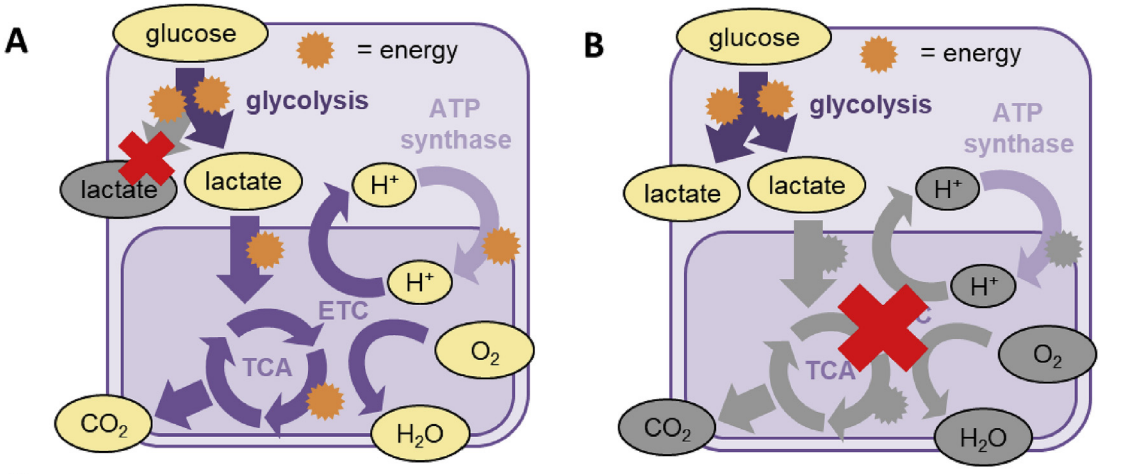
glucose transporter (GLUT), which blocks the import of the sole carbon substrate into the model. Removing the gene associated with the TCA + ETC reaction, citrate synthase (CS), disables the TCA + ETC reaction and decreases the flux through ATP demand from 32 to 2. Finally, removing the gene associated with ATP demand (with a key ATPase in muscle contraction bioenergetics (heavy chain 2 (MYH2)) as representative of several other ATPases) completely removes flux through the network since there is no demand reaction.

Double or triple gene knockout *in silico* simulations can reveal potential combinatorial effects that are not apparent with single gene knockout simulations. With the exception of disabling the glucose exchange and glycolysis reactions, removing individual reactions in *iSIM* did not reduce ATP demand flux to 0 (Table 2). To identify potential knockout combinations that inhibit flux through the ATP demand reaction, we used *iSIM* to simulate a system-wide double-knockout screen for 36 unique pairs of genes (Supplemental Table 1). By comparing double versus single knockout predictions, we identified five double gene knockout combinations that reduced ATP production more than any of the individual simulated gene knockouts (Fig. 4C). For example, removing only the lactate exchange reaction maintained a maximal ATP demand since lactate was utilized by the TCA + ETC reaction to generate ATP (Fig. 4A). However, if the lactate exchange reaction is removed in combination with any of the reactions associated with oxidative phosphorylation (O_2 exchange, CO_2 exchange, H_20 exchange, ATP synthase, or ETC + TCA), the flux through the ATP demand reaction drops to zero where removing any of these reactions on their own (Fig. 4B) did not completely reduce ATP production. Double knockouts provide an indication of the redundant nature of the network, which corresponds to multiple pathways for ATP synthesis.

In silico gene knockout studies provide insight into the flexibility and potential genetic targets of the network. For *iSIM*, we identified multiple single genes (Table 2) and double genes (Fig. 4C) that could be targeted to reduce the amount of ATP flux through the system. *iSIM* demonstrates that double knockouts can identify pairs of genes that do not additively inhibit the objective function of the model, identifying non-obvious targets. In real biological networks, single and double gene knockouts are used to identify gene targets to prevent biomass synthesis which could, for example, serve as potential drug targets for microbial pathogens or cancer cells.

3.4. Flux variability analysis

As described above, flux variability analysis (FVA) is a method to calculate the range of possible fluxes through each reaction in a network given a specific value for the objective function. This analysis enables a quantitative assessment of the flexibility in the network as well as identifying essential, non-essential, flexible, and non-flexible reactions for a defined objective function. Essential reactions are reactions with a non-zero lower limit on flux, indicating that the reaction must carry flux in order to meet the objective function, whereas non-essential reactions have a flux range that includes zero, meaning the reaction can but does not have to carry flux in order to meet the objective function. Flexible reactions are reactions where the upper and lower range on flux are not equal, indicating that the reaction can carry



C

Gene A	Reaction A	Flux A (fmol/cell/hr)	Gene B	Reaction B	Flux B (fmol/cell/hr)	Flux A and B (fmol/cell/hr)
LDH	lactate exchange	32	ATPV	ATP synthase	2	0
LDH	lactate exchange	32	CO2	CO_2 exchange	2	0
LDH	lactate exchange	32	AQP	H_2O exchange	2	0
LDH	lactate exchange	32	ETC	O_2 exchange	2	0
LDH	lactate exchange	32	CS	TCA + ETC	2	0

Fig. 4. Maximum possible flux through the ATP demand reaction predicted by FBA after performing a pairwise gene deletion screen with *iSIM*. (A) Network schematic representing flux through the network with the removal of lactate exchange from the network. (B) Network schematic representing the flux through the network with removal of the TCA + ETC reaction from the network. (C) Of the 36 possible reaction pairs, 5 double gene knockouts that reduced ATP production more than either single gene knockouts are shown.

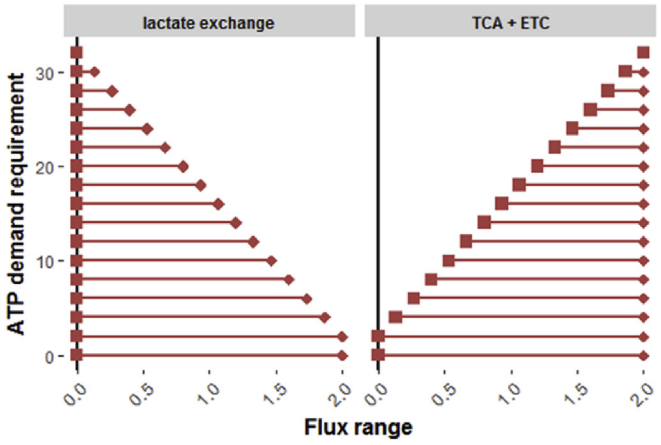


Fig. 5. Flux Variability Analysis in *iSIM*. Flux variability analysis (FVA) was performed requiring increasing amounts of ATP demand flux, as shown along the y-axis. Maximum (♦), minimum (■), and the range of flux (−) are shown for two reactions in the network, lactate exchange and TCA + ETC. As the ATP demand requirement increases, maximum possible flux values for lactate exchange decrease, indicating a decrease in lactate fermentation, while the minimum required flux values for glucose oxidation increase. The box indicates the flux values for an objective function with 50% of maximal ATP production.

a range of fluxes in order to meet the objective function, whereas non-flexible reactions have upper and lower range that are the same, indicating that the reaction must carry a specific flux through the network in order to meet the objective function. As above, FBA can be used to calculate a maximum flux of 32 units through the ATP demand reaction per 1 unit of glucose entering the *iSIM* network. To determine which reactions in the network are essential, non-essential, flexible and non-flexible, FVA can be used to determine feasible flux ranges for the remaining eight reactions. For this example, the lower bound on the flux through the objective function (ATP demand) was set at 50% of the maximum (16 flux units). Given that the lactate exchange reaction was involved in all double gene knockouts which completely inhibited flux through the ATP demand reaction, we have chosen to illustrate the flux range through this reaction. When requiring 50% of maximal flux through the ATP demand reaction, the lactate exchange reaction has a range of 0.000–1.066 units of flux while the glycolysis reaction has a range of 0.933–2.000 units of flux (Fig. 5). Therefore, lactate exchange is a non-essential (lower bound of zero), flexible (upper and lower bounds are not equal) reaction in order to maintain a minimum of 16 units of flux through the ATP demand reaction while the glycolysis reaction is an essential (positive lower bound), flexible (upper and lower bounds are not equal) reaction.

To better understand the relationship between ATP production and glucose breakdown, FVA was performed using incremental requirements of flux through the ATP demand reaction, from 0 to 100% of maximal flux or 0–32 units of flux (Fig. 5). Flux variability, or the range of possible flux values, decreased for both the lactate exchange reaction and the TCA + ETC reaction when the minimum flux through the ATP demand reaction was greater than 2. As the minimum flux through the ATP demand reaction increased to 32, the flux through the lactate exchange reaction decreased to 0, completely disabling flux through the reaction.

FVA is a useful approach to determine ranges of fluxes through individual reactions in the network, demonstrating the flexibility of metabolic networks to meet specific objective functions. With *iSIM*, we can see a tradeoff in the values of flux through the reactions involved in either anaerobic glycolysis or oxidative phosphorylation, based on the ATP needs of the cell. Combined with other approaches described above, FVA can also be used to identify how networks adapt to changes in the environment or genetic perturbations.

3.5. The importance of balancing thermodynamics in reactions

Errors can be unintentionally incorporated into metabolic network reconstructions through thermodynamically infeasible reactions and stoichiometrically unbalanced reactions that can affect computational predictions. Since all living organisms rely on external nutrients for energy to fuel biological processes, it is important that 1 unit of glucose in mammalian network reconstructions generates 32–36 ATP under ideal conditions with an unlimited supply of oxygen and at least 1 unit of ATP in the absence of oxygen [18]. In HMR2 [19] and other human GENREs [15,20], 1 unit of glucose yields an infinite amount of ATP with an unlimited supply of oxygen due to thermodynamically infeasible loops, an error which has been corrected in Recon 2.2 [13]. *iSIM*, a prototypic metabolic network consisting of nine reactions, captures theoretical ATP yields but the addition of two reactions can generate unrealistic ATP yields.

The regeneration of ATP from ADP, catalyzed by ATP synthase and driven by the proton gradient between the mitochondria and the cytosol, is an important reaction in central energy metabolism. In *iSIM*, ATP synthase catalyzes the reaction ADP[c] + P_i [c] + 4 H[c] → ATP [c] + H_2O[c] + 4 H[m] to generate ATP by the movement of protons between the mitochondrial and cytosolic compartments. This reaction represents a simplification of the 2.7 protons necessary for the formation of one ATP in mammalian mitochondria [21] with the one proton necessary to transport lactate into the mitochondria. To demonstrate the creation of a thermodynamically infeasible loop, two reactions were added to *iSIM*: a hypothetical reversible transport reaction representing the passive diffusion of lactate across the mitochondrial membrane (lactate[c] ↔ lactate[m]) and a transport reaction representing the symport of lactate and a proton across the mitochondrial membrane (lactate[m] + H[m] ↔ lactate[c] + H[c]). These two reactions create a loop where lactate is transported into the mitochondria through the

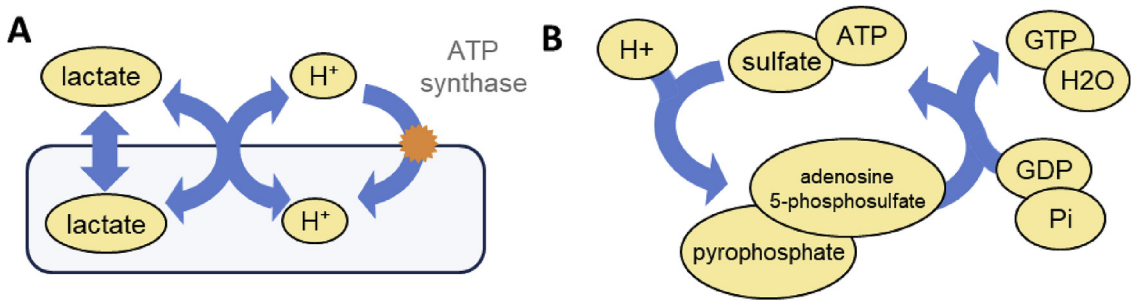


Fig. 6. Introducing errors into metabolic network reconstructions. (A) Including two new reactions in *iSIM* creates a thermodynamically infeasible loop that creates infinite amounts of ATP. Lactate is transported into the mitochondrial with no energy cost but is transported out of the mitochondria with a proton which is then used to generate ATP using the ATP synthase reaction. (B) As identified by Fritzemeier et al. [7], erroneous energy generating cycles often include groups of reactions which produce energy metabolites without an input. In this case, GTP is produced for "free" through each cycle of these two reactions.

first reaction and then back out of the mitochondria with a proton through the second reaction (Fig. 6A). ATP synthase then utilizes this H [c] to generate ATP. The addition of these two reactions represents two errors: the addition of an infeasible but stoichiometrically balanced reaction and the addition of a thermodynamically infeasible reaction. First, lactate cannot freely diffuse through the cell membrane due to its positive charge and therefore must be transported with an H. Secondly, although the transporter responsible for transporting lactate with an H can operate reversibly, the transporter only transports down a concentration gradient, which in this case is from lactate produced through glycolysis in the cytosol to the mitochondria. Together, these two reactions can produce ATP with no energy source. To resolve this issue, lactate should only be allowed to enter the mitochondria through irreversible active transport which requires a proton for entry.

To avoid introducing thermodynamically infeasible loops during the network reconciliation and manual curation process, it is important to check tasks regularly to ensure realistic ATP production. To avoid proton movement problems as described above for the lactate transporter, we included a metabolic task that should fail: regeneration of ATP from ADP without a carbon-based fuel source. Second, excess oxygen and inorganic ions should not increase ATP yields above 36 units of flux [18] for 1 unit of glucose, a result which was observed with multiple human GENREs. Various algorithms exist to include thermodynamic constraints with metabolic network reconstructions and are discussed in detail elsewhere [22].

3.6. The importance of maintaining stoichiometric balance in reactions

As demonstrated, infinite loops can be introduced into networks through thermodynamically infeasible reactions. Stoichiometrically unbalanced reactions can produce metabolites without the necessary inputs, which can then be used to maximize the objective function, or as a source for ATP generation [12,23]. As noted by Fritzemerier et al., an example of a pair of non-stoichiometrically balanced reactions are the reactions catalyzed by sulfate adenyltransferase which were included in multiple models generated by ModelSEED (Fig. 6B). Together, this pair of reactions can generate GTP without consuming energy.

Here, we demonstrate the importance of the fundamental assumptions of thermodynamically and stoichiometrically balanced reactions which form metabolic network reconstructions. Failures in these assumptions, as seen with the reversible lactate transporter or the sulfate adenyltransferase reactions, can lead to infeasible ATP or energy production and inaccurate results.

4. Discussion

GENREs allow for the system-wide integration of genetic and metabolic information in a mathematical formalism, enabling the prediction of phenotypes with constraint-based analytical methods. Prototypic networks are often created to illustrate new modeling method concepts. While additional small networks [25,26] allow for predictions that aren't immediately apparent, the larger number of reactions (>100 reactions) makes it difficult to trace pathways in the network and understand basic COBRA concepts. In addition, most such systems capture catabolic and anabolic functions while neglecting energy-generating metabolism. We created *iSIM*, a simplified metabolic network, to illustrate key considerations for metabolic network reconstructions and associated modeling methods.

In particular, *iSIM* highlights the importance of thermodynamically and stoichiometrically balanced reactions in a metabolic network reconstruction. Various algorithms have been proposed to automate the identification of thermodynamically infeasible reactions [24] and these methods can be used with metabolic tasks in network reconstructions to ensure feasible ATP yields. Recent work highlights energy generating cycles that were removed from the most recent human metabolic network reconstructions to produce feasible ATP yields for a variety of carbon sources [12].

In summary, *iSIM* serves as a tool for understanding constraint-based methods commonly used with GENREs. With access to the network reconstruction in several commonly used languages for constraint-based modeling methods (Matlab, R, Python, Supplemental Information), *iSIM* can be a useful didactic tool for illustrating new methods and promoting understanding of key concepts.

Author contributions

EB and JP conceived this study. EB, KR, and BD contributed to model design. EB wrote the initial draft of the paper. KR and BD significantly revised the manuscript and edited figures. EB, KR, BD, and ES wrote and edited the code in Python, Matlab and R. KR, BD, EB, ES, GK, VP, KV, AW, JP edited the final manuscript.

Conflicts of interest

None declared.

Acknowledgements

The opinions and assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the U.S. Army or of the U.S. Department of Defense. This paper has been approved for public release with unlimited distribution. Support for this project was provided by the United States Department of Defense (W81XWH-14-C-0054 to JP), the National Science Foundation Graduate Research Fellowship Program (awarded to BD), and the National Science Foundation Multi-scale Systems Bioengineering Research Experience for Undergraduates (awarded to ES).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2018.12.010.

References

[1] M.A. Oberhardt, B.Ø. Palsson, J.A. Papin, Applications of genome-scale metabolic reconstructions, Mol. Syst. Biol. 5 (2009) 320.
[2] A. Mardinoglu, F. Gatto, J. Nielsen, Genome-scale modeling of human metabolism – a systems biology approach, Biotechnol. J. 8 (2013) 985–996.
[3] S.A. Becker, et al., Quantitative prediction of cellular metabolism with constraint-based models: the COBRA toolbox, Nat. Protoc. 2 (2007) 727–738.
[4] R. Agren, et al., The RAVEN toolbox and its use for generating a genome-scale metabolic model for Penicillium chrysogenum, PLOS Comput. Biol. 9 (2013) e1002980.
[5] O. Hädicke, S. Klamt, EColiCore2: a reference network model of the central metabolism of Escherichia coli and relationships to its genome-scale parent model, Sci. Rep. 7 (2017) 39647–39647.
[6] J.D. Orth, R.M.T. Fleming, B.Ø. Palsson, Reconstruction and use of microbial metabolic networks: the core Escherichia coli metabolic model as an educational guide, EcoSal Plus 4 (2010).
[7] R. Agren, et al., Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling, Mol. Syst. Biol. 10 (2014).
[8] H. Zur, E. Ruppin, T. Shlomi, iMAT: an integrative metabolic analysis tool, Bioinformatics 26 (2010) 3140–3142.
[9] Y. Wang, J.A. Eddy, N.D. Price, Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE, BMC Syst. Biol. 6 (2012) 153.
[10] L. Jerby, T. Shlomi, E. Ruppin, Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism, Mol. Syst. Biol. 6 (2010) 401.
[11] A. Schultz, A.A. Qutub, Reconstruction of tissue-specific metabolic networks using CORDA, PLOS Comput. Biol. 12 (2016) e1004808.
[12] C.J. Fritzemeier, D. Hartleb, B. Szappanos, B. Papp, M.J. Lercher, Erroneous energy-generating cycles in published genome scale metabolic networks: identification and removal, PLOS Comput. Biol. 13 (2017) e1005494.
[13] N. Swainston, et al., Recon 2.2: from reconstruction to model of human metabolism, Metabolomics 12 (2016) 109.
[14] E.M. Blais, et al., Reconciled rat and human metabolic networks for comparative toxicogenomics and biomarker predictions, Nat. Commun. 8 (2017) 14250.

[15] I. Thiele, et al., A community-driven global reconstruction of human metabolism, Nat. Biotechnol. 31 (2013) 419–425.
[16] C. Gille, et al., HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology, Mol. Syst. Biol. 6 (2010) 411.
[17] A. Karlstädt, et al., CardioNet: a human metabolic network suited for the study of cardiomyocyte metabolism, BMC Syst. Biol. 6 (2012) 114.
[18] Metabolism at a Glance, third ed.. Wiley.com Available at: https://www.wiley.com/en-us/Metabolism+at+a+Glance%2C+3rd+Edition-p-9781118682074. (Accessed: 25th July 2018).
[19] A. Mardinoglu, et al., Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease, Nat. Commun. 5 (2014) 3083.
[20] N.C. Duarte, et al., Global reconstruction of the human metabolic network based on genomic and bibliomic data, Proc. Natl. Acad. Sci. Unit. States Am. 104 (2007) 1777–1782.
[21] I.N. Watt, M.G. Montgomery, M.J. Runswick, A.G.W. Leslie, J.E. Walker, Bioenergetic cost of making an adenosine triphosphate molecule in animal mitochondria, Proc. Natl. Acad. Sci. Unit. States Am. 107 (2010) 16823–16827.
[22] M. Ataman, V. Hatzimanikatis, Heading in the right direction: thermodynamics-based network analysis and pathway engineering, Curr. Opin. Biotechnol. 36 (2015) 176–182.
[23] R.G.A. van Heck, M. Ganter, V.A.P. Martins dos Santos, J. Stelling, Efficient reconstruction of predictive consensus metabolic network models, PLoS Comput. Biol. 12 (2016).
[24] J. Schellenberger, N.E. Lewis, B.Ø. Palsson, Elimination of thermodynamically infeasible loops in steady-state metabolic models, Biophys. J. 100 (2011) 544–553.
[25] L. Heirendt, et al., Creation and Analysis of Biochemical Constraint-based Models: the COBRA Toolbox v3.0, (2017) Q-Bio 171004038.
[26] A. Ebrahim, J.A. Lerman, B.O. Palsson, D.R. Hyduke, COBRApy: constraints-based reconstruction and analysis for Python, BMC Syst. Biol. 7 (2013) 74.