Alexandros Kiparissides    ORCID iD: 0000-0003-2923-2481

# Improving the accuracy of flux balance analysis through the implementation of carbon availability constraints for intracellular reactions

Maximilian Lularevic[1,2]

Andrew J. Racher[2]

Colin Jaques[2]

Alex Kiparissides[1]

[1]Department of Biochemical Engineering, University College London, WC1E 6BT, London, UK

[2]Lonza Biologics PLC, 228 Bath Road, Slough SL1 4DX, UK

**Correspondence:** Dr Alexandros Kiparissides, Department, Department of Biochemical Engineering, University College London, WC1E 6BT, London, UK.

**E-mail**: mailto:alex.kiparissides@ucl.ac.uk

## Abstract

Constraint based modelling methods, such as Flux Balance Analysis (FBA), have been extensively used to decipher complex, information rich -omics datasets in order to elicit system-wide behavioral patterns of cellular metabolism. FBA has been successfully used to gain insight in a wide range of applications, such as range of substrate utilization, product yields and to design metabolic engineering strategies to improve bioprocess performance. A well-known challenge associated with large genome-scale metabolic networks (GEMs) is that they result in underdetermined problem formulations. Consequently, rather than unique solutions, FBA and related methods examine ranges of reaction flux values that are consistent with the studied physiological conditions. The wider the reported flux ranges, the higher the uncertainty in the determination of basic reaction properties, limiting interpretability of and confidence in the results. Herein we propose a new, computationally efficient approach that refines flux range predictions by constraining reaction fluxes based on

the elemental balance of carbon. We compared carbon constraint FBA (ccFBA) against experimentally measured intracellular fluxes using the latest CHO GEM (*i*CHO1766) and were able to substantially improve the accuracy of predicted flux values compared to FBA. ccFBA can be used as a stand-alone method but is also compatible with and complimentary to other constraint-based approaches.

## Graphical Abstract

Constraint based modelling methods, such as Flux Balance Analysis (FBA), have been extensively used to decipher complex, information rich -omics datasets in order to elicit system-wide behavioral patterns of cellular metabolism. FBA has been successfully used to gain insight in a wide range of applications, such as range of substrate utilization, product yields and to design metabolic engineering strategies to improve bioprocess performance.
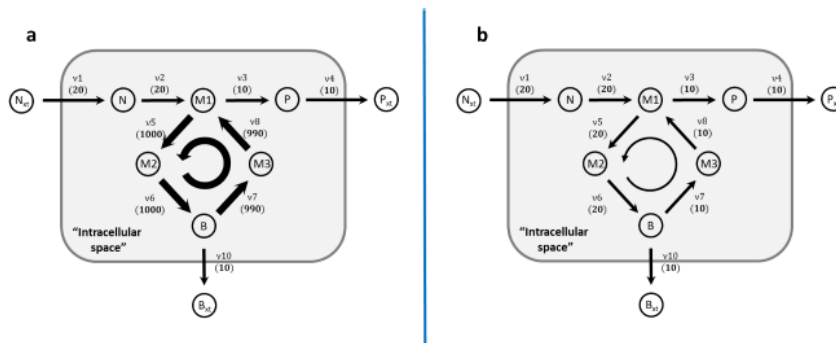


**Figure 1:** Example network illustrating a conventionally constraint network (a) and a carbon constraint network (b). The arrow thickness is not to scale and is only for visual accentuation of the differences in flux magnitude.

**Keywords:** Carbon constraining, CHO, Constraint-based modeling, Genome-scale metabolic network, FBA

## Introduction

Genome-scale metabolic network models (GEMs) have been successfully used to assign functionality to annotated genome sequences by providing the context of how the biochemical components of the cell interact on a molecular basis to effect the emergent cell phenotype. The gamut of biochemical reactions catalyzed by the enzymes identified in an annotated genome sequence form the basis of a metabolic network, which can be further refined by incorporating other –omics data such as transcriptomics, proteomics and/or metabolomics (Opdam et al., 2017; Ramon, Gollub, & Stelling, 2018). By assuming that the cell operates in a pseudo-steady

state, that is, there is no net metabolite production or consumption of intracellular metabolites, a series of mass balance constraints can be introduced and a linear programming (LP) problem can be formulated (Bonarius, Schmid, & Tramper, 1997; Varma & Palsson, 1994). This can be solved to quantify the distribution of carbon, energy and metabolic resources across the entire network for a given set of physiological conditions, commonly defined through metabolite uptake and secretion rates (Orth, Thiele, & Palsson, 2010). A variety of methods to augment and refine the resulting flux distributions through the incorporation of several types of biological activity data have been proposed (Beg et al., 2007; Covert & Palsson, 2002; Henry, Jankowski, Broadbelt, & Hatzimanikatis, 2006; Sánchez et al., 2017).

Since the stoichiometry of biochemical reactions is well-established and widely available in curated public databases, GEMs can be used as an accurate representation of the metabolic capabilities of a particular organism (Kelk, Olivier, Stougie, & Bruggeman, 2012). Constraint-based modeling has been successfully used to guide metabolic engineering strategies (Mishra et al., 2018), identify novel genes as antimicrobial drug targets (Mienda, Salihu, Adamu, & Idris, 2018), predict cellular phenotypes (Ramirez et al., 2017), analyze biological networks (Selvarasu et al., 2012) and study evolutionary processes (McCloskey, Palsson, & Feist, 2013; Pál et al., 2006) across more than 30 different organisms (Cook & Nielsen, 2017; N. C. Duarte et al., 2007; Natalie C Duarte & Herrg, 2004; Feist et al., 2007b; Förster, Famili, Fu, Palsson, & Nielsen, 2003; Hefzi et al., 2016; Reed, Vo, Schilling, & Palsson, 2003; Selvarasu et al., 2012).

Flux Balance Analysis (FBA), arguably the most prevalent constraint-based method, assumes the existence of an overarching metabolic objective, such as the maximization of biomass or of some metabolic product of interest (Schuetz, Kuepfer, & Sauer, 2007), resulting in a LP optimization problem. Information about the physiological state of the cell, under the conditions examined, is required in the form of uptake or secretion rates (also referred to as fluxes) ($v_{ex}$) for extracellular metabolites and the biomass growth rate. These are introduced through the upper ($v^U$) and lower ($v^L$) bound constraints for the respective exchange fluxes, where ($v^L \neq$

$v^U$) to account for experimental uncertainty. Additional information for the flux of intracellular reactions, if available, either through experimental data or relevant scientific literature, can be similarly integrated through the bound constraints of the respective reactions. The remaining, unknown reaction fluxes ($v_i$) are typically constrained to arbitrarily large numerical values (-1000 ≤ $v_i$ ≤ 1000 mmol gDCW$^{-1}$ h$^{-1}$) (Kelk et al., 2012). Originally, these arbitrarily large bounds were introduced during GEM development as a purely computational heuristic "trick" to test network connectivity and identify reactions carrying no flux. Since then it has extensively propagated in the scientific literature as a means to ensure that the "true" flux state of the cell is contained in the solution space. While factually the above statement is correct, the large numerical values commonly used are several orders of magnitude away from any physiologically observable flux and unnecessarily dilute the solution space with physiologically meaningless flux distributions.

Each reaction in the network for which ($v^L$ ≠ $v^U$) introduces additional degrees of freedom. Coupled with the high interconnectivity of metabolic networks and the limited number of reactions for which an exact flux value is known, this leads to a severely underdetermined LP problem. The result is numerous alternative internal flux distributions that satisfy the model's constraints and can achieve the same optimum (Orth et al., 2010; Lee et al., 2006; Soh et al., 2012). Consequently, a unique solution of the optimization problem carries little value. Instead, the range of permissible flux values through a particular reaction ($F_{v,i}$), under a given set of physiological conditions, can be studied either through Flux Variability Analysis (FVA) (Mahadevan & Schilling, 2003) and/or through Monte-Carlo sampling of the solution space (Price, Schellenberger, & Palsson, 2004).

As metabolic networks increase in size, the underlying linear programming problem becomes progressively more underdetermined which increases uncertainty in predicting an exact intracellular state relevant to the studied physiology. Reaction fluxes are likely to be overestimated and, particularly for networks with arbitrarily large bound constraints, internal loops or so-called futile cycles emerge in the majority of the reported flux distributions. Internal loops, also referred to as Type III

extreme pathways (Price, Famili, Beard, & Palsson, 2002), consist of a series of reactions that operate in a cycle, with unrealistically large flux values that allow them to generate metabolic resources at no cost. These internal loops can cause problems under two circumstances; (i) if they are linked to electron transfer between different electron carrier pools and (ii) if they are able to generate ATP or other currency metabolites (Maranas & Zomorrodi, 2016). This leads to a large amount of physiologically meaningless flux distributions, which confounds the determination of the "true" metabolic state. Several approaches have been proposed to tackle this problem, including: (i) identification and removal of internal loops (Chan, Wang, Dash, & Maranas, 2018; Schellenberger, Lewis, & Palsson, 2011), (ii) identification and removal of energy generating loops (Fritzemeier, Hartleb, Szappanos, Papp, & Lercher, 2017) (iii) reduction of the permissible flux ranges through the introduction of additional constraints based on thermodynamic properties (Henry et al., 2006), molecular crowding (Beg et al., 2007) or enzyme availability (Sánchez et al., 2017).

While some of the proposed methods significantly increase the accuracy of FBA predictions (Henry, Broadbelt, & Hatzimanikatis, 2007), they either require the problem to be reformulated as a mixed integer linear programming problem (MILP) (Henry et al., 2007; Schellenberger, Lewis, et al., 2011) or introduce a series of constraints that require information from aspects of cellular physiology that are difficult to quantify experimentally (Beg et al., 2007; Sánchez et al., 2017). In the absence of the required delicate intracellular measurements, researchers have to rely either on heuristic assumptions or on literature derived data. On the other hand, MILP problems scale poorly with dimension, making their application in the context of large scale GEMs computationally expensive. Herein we propose a new method for refining the flux ranges reported by FVA that is computationally efficient, relies on existing experimental data or assumptions and is able to systematically reduce flux variability ($F_V$) to physiologically meaningful values. The method is based on the elemental balance of carbon throughout the network and ensures reported flux values are consistent with the amount of carbon being taken up or secreted (Zarecki et al., 2014). While the proposed method does not explicitly resolve internal loops or

futile cycles, it does limit the permissible flux through reactions that participate in such loops and therefore substantially reduces their overall impact. Carbon Constrained FBA (ccFBA) can be used as a stand-alone method or to compliment any of the above mentioned methods.

**Materials and Methods**

All simulations were performed using Matlab (R2016a) (Mathworks, Natick Massachusetts, USA) and the algorithms included in the COBRA Toolbox v2.0 (Schellenberger, Que, et al., 2011). LP optimization problems were solved using the Gurobi solver (Gurobi Optimization, 2016).

Genome-scale model and experimental data

The latest metabolic network reconstruction for CHO cells (Hefzi et al., 2016) was used in the present work in order to ensure (a) a well-curated large scale GEM and (b) the existence of multiple, diverse experimental datasets to use for validation. The *i*CHO1766 was pre-processed to eliminate dead-end metabolites from the network ensuring complete functionality remained in terms of maximum biomass production, maximum product formation and oxygen dependence. Reactions containing dead-end metabolites are incapable of carrying flux under any imposed constraints. The removal of these reactions does not have any impact on the predicted flux patterns or general behavior of the model (Thiele & Palsson, 2010). In addition, an exchange reaction and the corresponding intracellular transport reaction for ethanolamine were introduced ('etha_e    <=> '; 'etha_e <=> etha_c';) (Voelker & Frazier, 1986) in the network. The final metabolic network, after curation, contained 4642 reactions (excluding 595 exchange reactions) involving 2816 metabolites (1571 unique metabolites). The model is available in supplementary materials (SuppInfo3_iCHO1766_currated) as a SBML file. Datasets containing uptake, secretion and intracellular flux rates for central carbon metabolism (Glycolysis, Tricarboxylic Acid Cycle (TCA) and Pentose Phosphate Pathway (PPP) ) were retrieved from literature (Ahn & Antoniewicz, 2011; McAtee Pereira, Walther, Hollenbach, & Young, 2018; Templeton, Xu, Roush, & Chen, 2017) and used to validate the predictions of the proposed algorithm (Table 2). In cases where

experimentally measured fluxes corresponded to more than one reaction in *i*CHO1766, average FVA or sample means were used for comparison (Table 1 and SuppInfo4 for an example calculation).

**Table 1** : The validation process involved integrating experimental data of secretion/uptake rates through the appropriate constraints and comparing the resulting intracellular fluxes with the corresponding experimental values from each dataset. A comprehensive list of the upper and lower bound values used to constrain exchange reactions for each experimental dataset can be found in supplementary material (SuppInfo1). In the absence of explicit experimental information, generalized upper and lower bounds allowing only secretion ($0 \leq v_{ex,i} \leq 100$) were used for all remaining exchange reactions. Intracellular reactions were purposefully left unconstrained, i.e. the limits were set to arbitrarily large values ($-1000 \leq v_i \leq 1000$ or $0 \leq v_i \leq 1000$) according to their reversibility properties.

## **Table 2 :** Flux Balance Analysis

Constraint-based methods, such as FBA, represent the gamut of biochemical and transport reactions known to occur in a particular type of cells in the form of a stoichiometric matrix **S** of size ($m$ x $n$). Every row ($m$) of **S** represents a unique compound, while every column ($n$) represents a single reaction. Consequently, each element ($s_{ij}$) of **S** contains the stoichiometric coefficient of the $i^{th}$ metabolite in the $j^{th}$ reaction. The flux through all ($n$) reactions is represented by the (n x 1) vector **v**. Assuming the existence of a steady-state, that is the total amount of any compound being produced is equal to the total amount being consumed, a mass balance across the entire metabolic network yields (Orth, Thiele, et al., 2010):

$$S \cdot v = 0 \tag{E1}$$

The system of algebraic equations defined by equation 1 is underdetermined and therefore a unique solution cannot be found. FBA assumes that cells configure their metabolism in a manner that seeks to optimize a particular objective, such as the maximization of biomass or ATP (Blazeck & Alper, 2010). Therefore, the problem can be formulated as a linear programming (LP) optimization problem:

$$min/\max(z) \tag{E2}$$

$$s.t.\, S \cdot v = 0 \tag{E3}$$

$$v^{ub} \geq v \geq v^{lb} \tag{E4}$$

where $z$ corresponds to the flux or sum of fluxes being optimized subject to mass balance constraints (Equation 3) and set of inequality constraints (Equation 4). In the present study, maximization of biomass was used as the objective function ($z = v_{biomass}$). The set of inequality constraints described by Equation 4 is used to constrain flux values between an upper ($v^{ub}$) and lower ($v^{lb}$) bound.

## Calculating reaction flexibility

Despite the reformulation into an LP problem, the system is still underdetermined and the calculated flux distribution $v$ is never a unique solution to the problem. Therefore, unique solutions rarely hold much value in FBA problems. Instead, Flux Variability Analysis (FVA) (Mahadevan & Schilling, 2003) or Monte-Carlo sampling of the null space (Schellenberger & Palsson, 2009) are used to evaluate the network's performance under a given set of physiological conditions. The concept of reaction flexibility introduced in previous work (Kiparissides & Hatzimanikatis, 2016) can be adapted as a means to quantify uncertainty in determining an exact metabolic state through the permissible flux range (Fv) for each reaction in the network as shown in equation 5.

$$F_{v,i} = \left| v_i^{max} - v_i^{min} \right| \tag{E5}$$

For a metabolic network containing ($n$) reactions, the total flux variability across the entire network can be defined as:

$$F_v^{TOT} = \sum_{i=1}^{n} \left| v_i^{max} - v_i^{min} \right| \tag{E6}$$

Carbon Constrained FBA (ccFBA)

The problem of FBA solutions containing very high flux values, well outside the physiological range, can be largely be attributed to: (a) the existence of redundancy and plasticity in cellular metabolism (Güell, Sagués, & Serrano, 2014; Mahadevan & Lovley, 2008; Min et al., 2011), (b) the use of arbitrarily large bounds for intracellular fluxes allowing orders of magnitude more carbon to cycle through the network than what is available to the cell through its environment and (c) a lack of physiological constraints such as additional intracellular flux measurements. The latter is illustrated in the example network shown in Figure 1a below. In reality, fluxes v5-v8 would involve multiple substrates and potentially co-factor pairs such as ATP/ADP or $NAD^+$/NADH which could lead to excessive generation of energy and redox potential, as fluxes (v5-v8) would form a type III extreme pathway. This problematic behavior is further exacerbated by the fact that large GEMs contain a number of extracellular reactions that are allowed to import, export or extracellularly convert metabolites into other chemical species. ccFBA attempts to resolve the issues outlined above, by constraining the permissible flux through intracellular reactions based on the amount of carbon taken up by the cell under the studied physiological conditions (Figure 1b). In the carbon-constrained example network, fluxes v5-v8 are constrained based on the maximum carbon taken up supplied through v1 (Figure 1b). While at its core, ccFBA imposes constraints based on the elemental balance of carbon on a constraint-based model, several aspects of cellular metabolism need to be taken into account and are elaborated in the following section.

**Figure 1**

ccFBA requires all sources of carbon uptake and their detailed chemical formulae to be known and an appropriately constrained FBA model. Therefore, availability of experimental measurements for major carbon sources (e.g. glucose, glycerol, etc.) in the form of uptake and secretion rates need to be provided. In the absence of experimental measurements for some carbon uptake rates, relevant values from literature can be used. In addition, detailed knowledge of culture media composition is necessary in order to determine compounds available for uptake and compounds

completely absent from the simulated experiment or condition. Below we present an overview of the basic concept of the proposed methodology using a simple example before presenting the entire algorithm in detail. Let us consider the case of a reversible reaction involving two substrates, two products and a co-factor pair (Reaction 1):

$$1\ G3P + 1\ P + 1\ NAD^+ \Leftrightarrow 1\ BPG + 1\ H^+ + 1\ NADH \tag{R1}$$
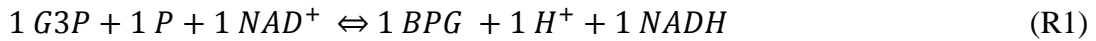
Using the chemical formula for each of the involved species this can be written as:

$$C_3H_5O_6P + P_i + C_{21}H_{26}N_7O_{14}P_2 \Leftrightarrow C_3H_4O_{10}P_2 + H + C_{21}H_{27}N_7O_{14}P_2 \tag{R2}$$
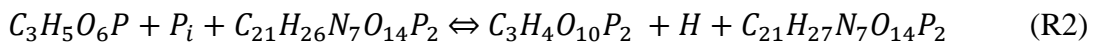
Let us further assume that the total amount of carbon taken up by the cell ($C_{TOT}$) is equal to 9 mmol$_c$ gDCW$^{-1}$ h$^{-1}$. In this particular reaction, NAD$^+$ and NADH act as co-factor pair and consequently their carbon atoms do not immediately participate (i.e. are exchanged or cleaved) in the reaction. Hence, co-factor pairs such as NAD$^+$/NADH, should not be considered when estimating the permissible carbon flux through biochemical reactions. Without loss of generality, we can apply the same rationale to the most commonly met co-factor pairs in metabolic networks. Consequently, excluding the co-factor pair, only 3 carbon atoms (N$^C$ = 3) are being exchanged in reaction 1. Therefore, taking into consideration the total amount of carbon being taken up by the cell ($C_{TOT}$ = 9 mmol$_c$ gDCW$^{-1}$ h$^{-1}$) and the inherent assumption of FBA for the existence of a metabolic pseudo-steady state, the maximum carbon flux through the reaction (R1) can be estimated as:

$$v_{max} = \frac{C_{TOT}}{N^C} = 3\ mmol\ gDCW^{-1}\ h^{-1} \tag{E7}$$

Equation 7, provides a hard upper bound on the permissible flux through reaction 1, based on the amount of carbon cycling through the metabolic network. As reaction 1 is bidirectional, the resulting bound constraints can be written as ($-3$ mmol $gDCW^{-1}h^{-1} \leq v_x \leq 3$ mmol $gDCW^{-1}h^{-1}$). This approach can be extended to the entire metabolic network; the steps of the ccFBA algorithm are detailed below:

1. Identification of all exchange (or "drain") reactions contained in the FBA model. These are reactions in the general form:

$$m_i^{ex} \xleftrightarrow{v_j^{EX}} \; ' \; '$$
(E8)

Where $m_i^{ex}$ is the extracellular concentration of metabolite (*i*) and $v_j^{EX}$ is the flux through exchange reaction (*j*).

2. Exchange reactions containing carbon need to be split based on directionality (uptake or secretion). The upper ($v^{UB}$) and lower ($v^{LB}$) bounds of exchange reactions involving carbon containing metabolites that are present in the culture media ($v_j^{EX,prs}$) should be adjusted to allow uptake and secretion:

$$v_j^{LB,prs} \leq v_j^{EX,prs} \leq v_j^{UB,prs}, \text{ where } v_j^{UB,prs} \geq 0 \text{ and } v_j^{LB,prs} \leq 0$$
( E9)

Similarly, the bounds for exchange reactions involving carbon containing metabolites absent from the culture media ($v_j^{EX,abs}$) should be adjusted to allow only secretion:

$$v_j^{LB,abs} \leq v_j^{EX,abs} \leq v_j^{UB,abs}, \text{ where } v_j^{LB,abs} \geq 0$$
(E10)

3. Available experimental data for uptake and secretion rates are integrated in the model by adjusting the bounds of the appropriate exchange reaction constraints. Variability in the reported experimental values should be reflected in the set upper and lower bounds so that all experimental values fall within the bounds. If the available experimental data does not contain values for all exchange reactions set as uptakes, relevant values should be sourced from literature. If such values are not available for conditions similar to the physiology being studied, the lower bound should be set to allow, at most, a carbon uptake rate equal to the highest experimentally measured carbon uptake rate (usually a primary substrate such as glucose). This can be calculated as:

$$v_{uk}^{LB} \leq v_{C_{max}}^{LB} \frac{N_{C_{max}}^{C}}{N_{uk}^{C}} \qquad\qquad (E11)$$

where $v_{uk}^{LB}$ is the unknown lower bound of the exchange reaction we wish to calculate, $v_{C_{max}}^{LB}$ is the lower bound of the exchange reaction carrying the highest amount of carbon, $N_{C_{max}}^{C}$ is the number of carbon atoms participating in the highest carbon uptake and $N_{uk}^{C}$ is the number of carbon atoms participating in the unknown uptake we wish to calculate.

4. The total carbon consumed ($C_{TOT}$) is calculated through the sum of products of each uptake rate ($v_u^{LB,prs}$) multiplied with the number of carbon atoms present in the taken up metabolite ($N_u^C$).

$$C_{TOT} = \sum_{u=1}^{\# \, uptakes} |v_u^{LB,prs} \cdot N_u^C| \qquad\qquad (E12)$$

Only the lower bound is considered in the calculation of $C_{TOT}$ to ensure that the maximum feasible amount of taken up carbon is estimated.

5. The maximum allowable flux ($v_j^{cc}$) through each intracellular reaction ($j$) can be calculated based on the total amount of carbon available ($C_{TOT}$) and the number of carbon atoms participating in each reaction ($N_j^C$).

$$v_j^{cc} = \frac{C_{TOT}}{N_j^C}$$
(E13)

Consequently, unidirectional intracellular reactions can be constrained between [0, $v_j^{cc}$], while bidirectional intracellular reactions can be constrained between [$-v_j^{cc}$, $v_j^{cc}$]. It is important to note, that when calculating ($v_j^{cc}$) substrates that do not participate in the reaction through their carbon atoms, such as co-factors, should be excluded from the calculation. This includes currency metabolites such as NAD$^+$/NADH, NADP$^+$/NADPH, FAD/FADH2, ATP/ADP-AMP but also coenzyme A, cytochromes and quinones. The algorithm (available on github: https://github.com/usMiggs/ccFBA) accounts for the most common non-

carbon contributing metabolites but a user defined list can be additionally provided (SuppInfo4). Furthermore, some metabolic reconstructions include generalized R-groups, particularly as part of lipid metabolism, which can contain a variable amount of carbon atoms. In such cases, the lowest possible number of carbon atoms is used for a more conservative constraining approach (see SuppInfo4 for more details). When the number of carbon atoms ($N^C$) can not be defined either due to missing chemical formulae or in the case of inorganic reactions, the bounds should be left at their default value.

6. Finally the model is tested for solvability with the given constraints and compared to experimental rates (e.g. growth rate) if available. If a feasible solution cannot be found with the bounds imposed by ccFBA, the bounds can be uniformly relaxed by introducing an incremental, percentage increase of the total amount of carbon available to the model ($C_{TOT}$) until a feasible solution is achieved (see SuppInfo4 for more details). Optionally, a ccFVA can be performed which may lead to a further reduction of $F_v^{TOT}$ (see SuppInfo4 for more details). It is recommended to compare the amount of carbon being taken up through the set exchange rates against the amount of carbon being secreted through the respective exchange rates (carbon closure). This enables an accurate approximation of unmeasured secretion rates (e.g. $CO_2$ in mammalian cell culture) but also may identify unknown carbon sources or inconsistencies in the experimental data.

**Results & Discussion**

A series of studies that report experimentally measured values for both extracellular (uptake/secretion) and intracellular fluxes were used to evaluate the performance of ccFBA (Templeton et al. 2017, McAtee Pereira et al. 2018, and Ahn & Antoniewicz 2011). The curated *i*CHO1766 model was constrained to reflect each experimental dataset using the reported values for extracellular reactions only and the ability of ccFBA/ccFVA to predict intracellular fluxes was compared to that of traditional

FBA/FVA. Rates for all major carbon sources and sinks such as glucose, lactate, amino acids, biomass and product formation were constrained based on experimental data while all other metabolite exchange rates were constrained to allow only secretion. After imposing the constraints on exchange fluxes, all experimental datasets were tested for solvability without carbon constraining as well as their ability to predict biomass and product yields (solving for maximum biomass or product yield without imposing upper bound constraints on them). Predictions of biomass and product formation were generally very close (e.g. experimentally determined growth rate = [0.0168 – 0.0238 h$^{-1}$], FBA value = 0.0284) to or even within the experimentally measured values (e.g. experimentally determined growth rate = [0.0150 – 0.0213 h$^{-1}$], FBA value = 0.0198). However, the wide flux ranges ($F_V$) reported for intracellular reactions, limit our ability to draw any conclusions regarding the underlying metabolic phenotype.

## Flux variability with (ccFVA) and without (FVA) carbon based constraints

A total of seven experimental datasets, corresponding to seven distinct sets of culture conditions (Table 2) were considered. Compared to traditional FVA, ccFVA reported a lower flux variability ($F_V$) for 58% ± 7% of the reactions in $i$CHO1766 across all 7 experimental datasets. This result compares well with the value reported by Sanchez et al. using the GECKO method (60% of the reactions showed reduced flux variability) (Sánchez et al., 2017). Furthermore, ccFVA led to a significant reduction in total flux variability ( $F_V^{TOT}$, Equation 6) with the extent of reduction being dependent on two factors: (i) the arbitrary value used to constrain intracellular fluxes and (ii) the total amount of carbon consumed by the cell ($C_{TOT}$). According to common FBA practice, arbitrarily large bounds are applied to unknown intracellular fluxes constraining bidirectional reactions between [-1000, 1000] mmol gDCW$^{-1}$ h$^{-1}$ and unidirectional reactions between [0 , 1000] mmol gDCW$^{-1}$ h$^{-1}$ (Bordel, Agren, & Nielsen, 2010; Chaudhary, Tøndel, Puchałka, Santos, & Bhatnagar, 2016; Kelk et al., 2012; Maranas & Zomorrodi, 2016; Opdam et al., 2017; Yang et al., 2016). In the present study, three different arbitrary values for the bounds of unknown

intracellular fluxes (10, 100, and 1000 mmol gDCW$^{-1}$ h$^{-1}$) were considered (SuppInfo2 Figures S1 and S2).

Our results (SuppInfo2 Figure S3) indicate a negative linear dependence between the reduction in $F_V^{TOT}$ and $C_{TOT}$. The lower the arbitrary value used to constrain unknown intracellular fluxes the higher the percentage reduction in $F_V^{TOT}$ with increasing carbon uptake ($C_{TOT}$), as indicated by the increasing negative gradient of the linear fit from $R_{1000}^{Fv,TOT}$ (slope = -0.018) to $R_{10}^{Fv,TOT}$ (slope = -1.511). ccFVA was able to reduce total flux variability ($F_V^{TOT}$) by at least 85% compared to normal FVA across the set of 21 simulations considered (7 experimental datasets using 3 different arbitrary values to constrain unknown intracellular fluxes). Crucially when constraining unknown intracellular fluxes using the values most commonly met in literature, between 100 – 1000 mmol gDCW$^{-1}$ h$^{-1}$, ccFVA resulted in a reduction of total flux variability of at least 95%. The substantial difference in the size of the resulting solution space on a reaction basis can be seen for the stationary growth condition data as reported by Ahn & Antoniewicz (2011) in Figure 2 (results for all other conditions can be found in supplementary materials SuppInfo2 Figures S4 - S9). Over 50% of the reactions of the FVA treated model use their maximum flux range (in this example 1000 mmol gDCW$^{-1}$ h$^{-1}$). A reduction of this magnitude has significant implications when considered in the context of sampling the space of feasible solutions of a GEM using Monte-Carlo methods. The benefits of using ccFVA to adjust the bounds of intracellular reactions prior to sampling are twofold: (i) a significantly higher proportion of the sampled flux distributions will contain physiologically relevant solutions, devoid of internal loops and unrealistically high flux values and (ii) the resulting solution space will be considerably smaller increasing computational efficiency and leading to faster convergence of sample statistics.

**Figure 2**

One of the main benefits of ccFBA/ccFVA is the ability to mitigate the impact of internal loops which are prevalent in large-scale, highly interconnected networks (Price, Reed, & Palsson, 2004). A closer observation of the FVA results presented in

Figure 3 (A-G: grey, blue, and green bars) highlights six central carbon reactions that are allowed to carry physiologically unrealistic amounts of flux and are likely to be involved in internal loops. These include isocitrate dehydrogenase (IDH), malate dehydrogenase (MDH), fumarase (FUS), ribose-5-phosphate isomerase (PPI), lactate dehydrogenase (LDH), and triose phosphate isomerase (TPI) indicated by their minimum and maximum permissible flux being equal to the set upper and lower bounds. While ccFVA does not explicitly attempt to remove internal loops, it reduces the flux carried through such reactions to a physiologically meaningful range (Figure 3 A-G: red bars) and consequently limits their impact in terms of potential electron or ATP drains (Maranas & Zomorrodi, 2016).

In addition, the implementation of ccFVA does not modify the underlying LP problem formulation or increase its complexity as it acts directly on the bounds of intracellular reactions. The benefits of this approach are twofold: (i) it makes ccFBA fully compatible with other methods, such as ll-FVA (Schellenberger, Lewis, et al., 2011), tFBA (Henry et al., 2007), GECKO (Sánchez et al., 2017), FCF (Burgard, Nikolaev, Schilling, & Maranas, 2004) or lll-FVA (Chan et al., 2018) and can be used in tandem to further refine the predicted flux distributions and (ii) it makes ccFBA/ccFVA computationally efficient compared to approaches that recast or modify the original LP formulation (for example through the implementation of additional constraints). This results in execution times significantly ($p>0.001$) faster than other approaches (Table 3).

**Figure 3**

**Table 3 :** In order to investigate how ccFVA scales with model size, a series of simulations involving models of increasing complexity were performed. The predictions (SuppInfo2 Figure S10) and computational efficiency (Table 3) of ccFVA were compared against traditional FVA and loopless-FVA (Schellenberger, Lewis, et al., 2011). As expected, the computational time increases with model size for all three methods. Interestingly, ccFVA outperforms normal FVA for the largest model examined (*i*CHO1766) which implies that the reduction in flux variability achieved by

the carbon based constraints increases the efficiency of the subsequently run FVA algorithm.

## Comparing ccFVA predictions with intracellular flux measurements

The upper and lower bounds calculated by ccFBA are based on the total amount of carbon ($C_{TOT}$) being taken up by the cell. An inherent assumption being made is that any biochemical reaction within the metabolic network has access to all of the carbon entering the cell. While an improvement over arbitrarily selected bounds, this is a generous assumption that will result in the overestimation of the permissible flux range for many nodes in the network. Therefore, despite the significant reduction in total flux variability compared to normal FVA, the permissible flux ranges calculated using ccFVA are still expected to contain the "true" phenotypic flux distribution. In order to validate this statement, the results of ccFVA were compared against intracellular flux measurements conducted using ^{13}C labelled glucose for each of the seven experimental datasets considered (Figure 4). Not only were nearly all experimental flux ranges contained within the permissible flux range calculated by ccFVA, they were also mostly within the same order of magnitude as the calculated ccFVA flux ranges highlighting the reliability of the proposed method.

Only two reactions had measured flux values that were not within the flux range calculated by ccFVA (isocitrate dehydrogenase and citrate synthase) and this was the case in only one of the seven datasets considered (Figure 4G). However, in both cases the experimentally measured flux values were very close to the range reported by ccFVA indicating that the observed discrepancy could be caused by an experimental or conversion calculation error. [13]C measurements are prone to a number of measurement errors such as suboptimal metabolism quenching, unaccounted carbon sources in the media (Mairinger et al., 2018), or errors in conversion calculations. An example of the latter is the conversion of experimental flux values from the conventionally reported g cells$^{-1}$ h$^{-1}$ into mmol gDCW$^{-1}$ h$^{-1}$ usually used in GEMs. This conversion requires the dry cell weight of the culture to be determined based on the viable cell concentration. The cell size and dry cell weight of individual cells varies through cultures leading to difficulties in estimating

dry cell weight accurately (Pan, Dalm, Wijffels, & Martens, 2017) and is a known potential source of error (Ahn & Antoniewicz, 2011). A comprehensive list of potential errors in ^{13}C measurements has been reported by Mairinger and co-workers (Mairinger et al., 2018). Despite this minor discrepancy, the predictions of biomass and product yields were not affected by ccFBA/ccFVA in any of the seven experimental datasets considered compared to the reported experimental values further increasing confidence in the algorithm.

**Figure 4**

## Random sampling of feasible flux distributions with (ccFVA) and without (FVA) carbon based constraints

Once the region of feasible flux distributions has been defined either via FVA or ccFVA, random sampling of flux distributions from the resulting solution space can be used to enhance behavioral understanding of the network state (Mo, Palsson, & Herrgard, 2009; Schellenberger & Palsson, 2009; Shlomi, Benyamini, Gottlieb, Sharan, & Ruppin, 2011). Each of the 7 experimental conditions considered were uniformly sampled using the ACHR sampler included in the COBRA toolbox (Schellenberger, Que, et al., 2011). Prior to ACHR sampling, the GEM was constrained using the experimentally reported extracellular fluxes in combination with either normal FVA ($RS_{FVA}$) or ccFVA ($RS_{FVA}^{cc}$) results. Each experimental condition was sampled until 200,000 unique flux distributions had been retrieved using a set of warm-up points equal to four times the number of reactions in the GEM. The resulting sampled flux distributions were compared to determine if carbon based constraints improve the reliability and efficiency of flux sampling.

Figure 5 (A-G) presents the estimated sample means with their respective standard deviations for reactions in central carbon metabolism and compares them against ^{13}C experimental data. Overall, random sampling of feasible flux distributions led to narrower predicted flux ranges compared to the FVA ranges irrespective of the method used to constrain the model. The sampled flux distributions calculated from models using ccFVA constraints ($RS_{FVA}^{cc}$) were substantially narrower and in better agreement with experimentally measured values when compared with samples

constrained with normal FVA ($RS_{FVA}$). Moreover, the sampled distributions for all reactions where the permissible flux range predicted by FVA (Figure 3) coincided with the initially set bounds (-1000 ≤ $v_i$ ≤ 1000 mmol gDCW$^{-1}$ h$^{-1}$) remained very wide (Table 4) for all $RS_{FVA}$ samples. Crucially, for the CM dataset (Figure 5A) the sampled means derived from $RS_{FVA}$ could not predict the correct reaction directionality for four reactions (indicated with a § in Table 4). This inability to predict reaction directionalities with $RS_{FVA}$ could be observed in all datasets considered (Figure 5) leading to a mismatch for 33 out of 149 (22%) experimentally determined directionalities. In comparison sample means derived from $RS_{FVA}^{cc}$ failed to predict only 14 out of 149 (9%) experimentally measured reaction directionalities, all for reactions that $RS_{FVA}$ had also failed to predict (except LDH reaction Figure 5F). This is an improvement of > 50% and highlights the benefits of using ccFVA not only to reduce flux variability but also to obtain improved predications in terms of reaction directionality.

**Figure 5**

**Table 4**

## Conclusion

Herein we presented ccFBA/ccFVA, a fast and simple method to accurately constrain stoichiometric metabolic networks using physiologically relevant assumptions. We demonstrated ccFVA's ability to reduce total flux variability by >85% (initial bounds 1000 mmol gDCW$^{-1}$ h$^{-1}$) and >95% (initial bounds ≤100 mmol gDCW$^{-1}$ h$^{-1}$) compared to standard FVA techniques. A set of seven distinct experimental datasets containing $^{13}$C measurements for central carbon fluxes were retrieved from literature and used to validate the predictive capabilities of ccFVA. We showed that permissible flux ranges estimated by ccFVA contained the experimentally measured intracellular fluxes in the majority of cases and lead to quantitative predictions in the same order of magnitude as $^{13}$C measurements. This can be attributed to ccFVA's ability to mitigate the impact of internal loops or futile cycles by constraining the amount of flux able to pass through any single reaction based on the amount of carbon entering the cell. Moreover, when used in combination with random sampling, ccFVA

substantially improved our ability to predict reaction directionalities compared to normal FVA. ccFBA is an easy to use and computationally efficient method for reducing flux variability in and increasing the reliability of constrained-based metabolic networks. It can be used as a stand-alone method or as a complimentary tool to most other methods currently available for stoichiometric metabolic network analysis. Finally, the method can be expanded to consider additional elemental balances such as nitrogen and phosphorous which could be beneficial when studying other organisms such as algae or specific types of bacterial strains. Future updates of ccFBA will include additional elemental balances and will be made available online through the github repository.

## Acknowledgements

## Nomenclature

| Abbreviation | Explanation | Reaction ID $i$CHO1766 |
|---|---|---|
| IDH | aconitate hydratase / isocitrate dehydrogenase (NAD$^+$ and NADP$^+$) | 2001 / 1591 / 1593 |
| ADH | 2-oxoglutarate dehydrogenase | 1589 / 2561 |
| CS | citrate synthase | 1590 |
| MDH | malate dehydrogenase | 1594 |
| FUS | fumarase | 2191 |
| SDH | succinate dehydrogenase | 1595 |

| | | |
|---|---|---:|
| PDH | pyruvate dehydrogenase | 2457 |
| TKT2 | transketolase | 1803 |
| TAL | transaldolase | 2563 |
| TKT1 | transketolase | 1802 |
| PPI | ribose-5-phosphate isomerase | 2506 |
| PPE | ribulose 5-phosphate 3-epimerase | 2505 |
| G6PDH | L-gulonate 3-dehydrogenase / 6-phosphogluconolactonase / phosphogluconate dehydrogenase | 1797 / 2464 / 2675 |
| LDH | lactate dehydrogenase | 1720 |
| HK | hexokinase | 2313 |
| PK | pyruvate kinase | 1723 |
| ENO | phosphoglycerate mutase / enolase | 1722 / 2157 |
| GAPDH | glyceraldehyde-3-phosphate dehydrogenase / phosphoglycerate kinase | 1719 / 1721 |
| TPI | triose-phosphate isomerase | 1724 |
| PFK | phosphofructokinase / fructose-bisphosphate aldolase | 2462 / 1716 |
| PGI | glucose-6-phosphate isomerase | 2674 |
| GALK | galactokinase / UDP-glucose: alpha-D-galactose-1-phosphate uridylyltransferase / phosphoglucomutase | 2258 / 2595 / 2465 |

## References

Ahn, W. S., & Antoniewicz, M. R. (2011). Metabolic flux analysis of CHO cells at growth and non-growth phases using isotopic tracers and mass spectrometry. *Metabolic Engineering*, *13*(5), 598–609. http://doi.org/10.1016/j.ymben.2011.07.002

Beg, Q. K., Vazquez, A., Ernst, J., de Menezes, M. A., Bar-Joseph, Z., Barabasi, A.-L., & Oltvai, Z. N. (2007). Intracellular crowding defines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity. *Proceedings of the National Academy of Sciences*, *104*(31), 12663–12668. http://doi.org/10.1073/pnas.0609845104

Blazeck, J., & Alper, H. (2010). Systems metabolic engineering: Genome-scale models and beyond. *Biotechnology Journal*, *5*(7), 647–659. http://doi.org/10.1002/biot.200900247

Bonarius, H. P. J., Schmid, G., & Tramper, J. (1997). Flux analysis of underdetermined metabolic networks: The quest for the missing constraints. *Trends in Biotechnology*, *15*(8), 308–314. http://doi.org/10.1016/S0167-7799(97)01067-6

Bordel, S., Agren, R., & Nielsen, J. (2010). Sampling the solution space in genome-scale metabolic networks reveals transcriptional regulation in key enzymes. *PLoS Computational Biology*, *6*(7), 16. http://doi.org/10.1371/journal.pcbi.1000859

Burgard, a P., Nikolaev, E. V, Schilling, C. H., & Maranas, C. D. (2004). Flux Coupling Analysis of Genome-Scale Metabolic Network. *Genome Research*, *14*, 301–312. http://doi.org/10.1101/gr.1926504.

Chan, S. H. J., Wang, L., Dash, S., & Maranas, C. D. (2018). Accelerating flux balance calculations in genome-scale metabolic models by localizing the application of loopless constraints. *Bioinformatics*, *34*(24), 4248–4255. http://doi.org/10.1093/bioinformatics/bty446

Chaudhary, N., Tøndel, K., Puchałka, J., Santos, V. A. P. M. dos, & Bhatnagar, R. (2016). Characterizing the optimal flux space of Genome Scale Metabolic Reconstructions through Modified Latin-Hypercube sampling. *Molecular BioSystems*, *12*(September 2015), 994–1005. http://doi.org/10.1039/C5MB00457H

Cook, D. J., & Nielsen, J. (2017). Genome-scale metabolic models applied to human health and disease. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*, *9*(6), e1393. http://doi.org/10.1002/wsbm.1393

Covert, M. W., & Palsson, B. (2002). Transcriptional regulation in constraints-based metabolic models of Escherichia coli. *Journal of Biological Chemistry*, *277*(31), 28058–28064. http://doi.org/10.1074/jbc.M201691200

Duarte, N. C., Becker, S. A., Jamshidi, N., Thiele, I., Mo, M. L., Vo, T. D., ... Palsson, B. O. (2007). Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences*, *104*(6), 1777–1782. http://doi.org/10.1073/pnas.0610772104

Duarte, N. C., & Herrg, M. J. (2004). Reconstruction and Validation of. *Genome Research*, 1–12. http://doi.org/10.1101/gr.2250904.

Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., ... Palsson, B. (2007a). A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*, *3*(121), 1–18. http://doi.org/10.1038/msb4100155

Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., ... Palsson, B. Ø. (2007b). A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*, *3*(121), 1–18. http://doi.org/10.1038/msb4100155

Förster, J., Famili, I., Fu, P., Palsson, B. Ø., & Nielsen, J. (2003). Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network. *Genome Research*, *13*(2), 244–53. http://doi.org/10.1101/gr.234503

Fritzemeier, C. J., Hartleb, D., Szappanos, B., Papp, B., & Lercher, M. J. (2017). Erroneous energy-generating cycles in published genome scale metabolic networks: Identification and removal. *PLOS Computational Biology*, *13*(4), e1005494. http://doi.org/10.1371/journal.pcbi.1005494

Güell, O., Sagués, F., & Serrano, M. Á. (2014). Essential Plasticity and Redundancy of Metabolism Unveiled by Synthetic Lethality Analysis. *PLoS Computational Biology*, *10*(5). http://doi.org/10.1371/journal.pcbi.1003637

Gurobi Optimization, I. (2016). Gurobi Optimizer Reference Manual.

Hefzi, H., Ang, K. S., Hanscho, M., Bordbar, A., Ruckerbauer, D., Lakshmanan, M., ... Lewis, N. E. (2016). A Consensus Genome-scale Reconstruction of Chinese Hamster Ovary Cell Metabolism. *Cell Systems*, *3*(5), 434–443.e8. http://doi.org/10.1016/j.cels.2016.10.020

Henry, C. S., Broadbelt, L. J., & Hatzimanikatis, V. (2007). Thermodynamics-Based Metabolic Flux Analysis. *Biophysical Journal*, *92*(5), 1792–1805. http://doi.org/10.1529/biophysj.106.093138

Henry, C. S., Jankowski, M. D., Broadbelt, L. J., & Hatzimanikatis, V. (2006). Genome-scale thermodynamic analysis of Escherichia coli metabolism. *Biophysical Journal*, *90*(4), 1453–61. http://doi.org/10.1529/biophysj.105.071720

Kelk, S. M., Olivier, B. G., Stougie, L., & Bruggeman, F. J. (2012). Optimal flux

spaces of genome-scale stoichiometric models are determined by a few subnetworks. *Scientific Reports*, *2*, 44–46. http://doi.org/10.1038/srep00580

Kiparissides, A., & Hatzimanikatis, V. (2016). Thermodynamics-based Metabolite Sensitivity Analysis in metabolic networks. *Metabolic Engineering*, (November), 0–1. http://doi.org/10.1016/j.ymben.2016.11.006

Mahadevan, R., & Lovley, D. R. (2008). The degree of redundancy in metabolic genes is linked to mode of metabolism. *Biophysical Journal*, *94*(4), 1216–1220. http://doi.org/10.1529/biophysj.107.118414

Mahadevan, R., & Schilling, C. H. (2003). The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic Engineering*, *5*(4), 264–276. http://doi.org/10.1016/j.ymben.2003.09.002

Mairinger, T., Wegscheider, W., Peña, D. A., Steiger, M. G., Koellensperger, G., Zanghellini, J., & Hann, S. (2018). Comprehensive assessment of measurement uncertainty in 13C-based metabolic flux experiments. *Analytical and Bioanalytical Chemistry*, *410*(14), 3337–3348. http://doi.org/10.1007/s00216-018-1017-7

Maranas, C. D., & Zomorrodi, A. R. (2016). *Optimization Methods in Metabolic Networks*. Hoboken, NJ: John Wiley & Sons, Inc. http://doi.org/10.1002/9781119188902

McAtee Pereira, A. G., Walther, J. L., Hollenbach, M., & Young, J. D. (2018). 13C Flux Analysis Reveals that Rebalancing Medium Amino Acid Composition can Reduce Ammonia Production while Preserving Central Carbon Metabolism of CHO Cell Cultures. *Biotechnology Journal*, *1700518*, 1–7. http://doi.org/10.1002/biot.201700518

McCloskey, D., Palsson, B. Ø., & Feist, A. M. (2013). Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli. *Molecular Systems Biology*, *9*(1), 661. http://doi.org/10.1038/msb.2013.18

Mienda, B. S., Salihu, R., Adamu, A., & Idris, S. (2018). Genome-scale metabolic models as platforms for identification of novel genes as antimicrobial drug targets. *Future Microbiology*, *13*(4), 455–467. http://doi.org/10.2217/fmb-2017-0195

Min, Y., Jin, X., Chen, M., Pan, Z., Ge, Y., & Chang, J. (2011). Pathway knockout and redundancy in metabolic networks. *Journal of Theoretical Biology*, *270*(1), 63–69. http://doi.org/10.1016/j.jtbi.2010.11.012

Mishra, P., Lee, N. R., Lakshmanan, M., Kim, M., Kim, B. G., & Lee, D. Y. (2018). Genome-scale model-driven strain design for dicarboxylic acid production in Yarrowia lipolytica. *BMC Systems Biology*, *12*(Suppl 2). http://doi.org/10.1186/s12918-018-0542-5

Mo, M. L., Palsson, B. Ø., & Herrgard, M. J. (2009). Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Systems Biology*, *3*, 37. http://doi.org/10.1186/1752-0509-3-37

Opdam, S., Richelle, A., Kellman, B., Li, S., Zielinski, D. C., & Lewis, N. E. (2017). A Systematic Evaluation of Methods for Tailoring Genome-Scale Metabolic Models. *Cell Systems*, *4*(3), 318–329.e6. http://doi.org/10.1016/j.cels.2017.01.010

Orth, J. D., Palsson, B. Ø., & Fleming, R. M. T. (2010). Reconstruction and Use of Microbial Metabolic Networks: the Core Escherichia coli Metabolic Model as an Educational Guide. *EcoSal Plus*, *4*(1). http://doi.org/10.1128/ecosalplus.10.2.1

Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010). What is flux balance analysis? *Nature Biotechnology*, *28*(3), 245–248. http://doi.org/10.1038/nbt.1614

Pál, C., Papp, B., Lercher, M. J., Csermely, P., Oliver, S. G., & Hurst, L. D. (2006). Chance and necessity in the evolution of minimal metabolic networks. *Nature*, *440*(7084), 667–670. http://doi.org/10.1038/nature04568

Pan, X., Dalm, C., Wijffels, R. H., & Martens, D. E. (2017). Metabolic characterization of a CHO cell size increase phase in fed-batch cultures. *Applied Microbiology and Biotechnology*. http://doi.org/10.1007/s00253-017-8531-y

Price, N. D., Famili, I., Beard, D. A., & Palsson, B. Ø. (2002). Extreme pathways and Kirchhoff's second law. *Biophysical Journal*, *83*(5), 2879–2882. http://doi.org/10.1016/S0006-3495(02)75297-1

Price, N. D., Reed, J. L., & Palsson, B. Ø. (2004). Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Reviews. Microbiology*, *2*(11), 886–97. http://doi.org/10.1038/nrmicro1023

Price, N. D., Schellenberger, J., & Palsson, B. O. (2004). Uniform sampling of steady-state flux spaces: Means to design experiments and to interpret enzymopathies. *Biophysical Journal*, *87*(4), 2172–2186. http://doi.org/10.1529/biophysj.104.043000

Ramirez, A. K., Lynes, M. D., Shamsi, F., Xue, R., Tseng, Y. H., Kahn, C. R., … Dreyfuss, J. M. (2017). Integrating Extracellular Flux Measurements and Genome-Scale Modeling Reveals Differences between Brown and White Adipocytes. *Cell Reports*, *21*(11), 3040–3048. http://doi.org/10.1016/j.celrep.2017.11.065

Ramon, C., Gollub, M. G., & Stelling, J. (2018). Integrating – omics data into genome-scale metabolic network models : principles and. *Portland Press*, *62*(August), 563–574. http://doi.org/10.1042/EBC20180011

Reed, J. L., Vo, T. D., Schilling, C. H., & Palsson, B. O. (2003). An expanded genome-

scale model of Escherichia coli K-12 (iJR904 GSM/GPR). *Genome Biology*, *4*(9), R54. http://doi.org/10.1186/gb-2003-4-9-r54

Sánchez, B. J., Zhang, C., Nilsson, A., Lahtvee, P., Kerkhoven, E. J., & Nielsen, J. (2017). Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Molecular Systems Biology*, *13*(8), 935. http://doi.org/10.15252/msb.20167411

Schellenberger, J., Lewis, N. E., & Palsson, B. (2011). Elimination of thermodynamically infeasible loops in steady-state metabolic models. *Biophysical Journal*, *100*(3), 544–553. http://doi.org/10.1016/j.bpj.2010.12.3707

Schellenberger, J., & Palsson, B. (2009). Use of randomized sampling for analysis of metabolic networks. *Journal of Biological Chemistry*, *284*(9), 5457–5461. http://doi.org/10.1074/jbc.R800048200

Schellenberger, J., Que, R., Fleming, R. M. T., Thiele, I., Orth, J. D., Feist, A. M., … Palsson, B. Ø. (2011). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature Protocols*, *6*(9), 1290–1307. http://doi.org/10.1038/nprot.2011.308

Schuetz, R., Kuepfer, L., & Sauer, U. (2007). Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. *Molecular Systems Biology*, *3*(119). http://doi.org/10.1038/msb4100162

Selvarasu, S., Ho, Y. S., Chong, W. P. K., Wong, N. S. C., Yusufi, F. N. K., Lee, Y. Y., … Lee, D.-Y. (2012). Combined in silico modeling and metabolomics analysis to characterize fed-batch CHO cell culture. *Biotechnology and Bioengineering*, *109*(6), 1415–1429. http://doi.org/10.1002/bit.24445

Shlomi, T., Benyamini, T., Gottlieb, E., Sharan, R., & Ruppin, E. (2011). Genome-scale metabolic modeling elucidates the role of proliferative adaptation in causing the warburg effect. *PLoS Computational Biology*, *7*(3), 1–8. http://doi.org/10.1371/journal.pcbi.1002018

Templeton, N., Xu, S., Roush, D. J., & Chen, H. (2017). 13C metabolic flux analysis identifies limitations to increasing specific productivity in fed-batch and perfusion. *Metabolic Engineering*, *44*(August), 126–133. http://doi.org/10.1016/j.ymben.2017.09.010

Thiele, I., & Palsson, B. Ø. (2010). A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, *5*(1), 93–121. http://doi.org/10.1038/nprot.2009.203

Varma, A., & Palsson, B. O. (1994). Metabolic flux balancing: Basic concepts, scientific and practical use. *Bio/Technology*, *12*(10), 994–998. http://doi.org/10.1038/nbt1094-994

Voelker, D. R., & Frazier, J. L. (1986). Isolation and characterization of a Chinese

hamster ovary cell line requiring ethanolamine or phosphatidylserine for growth and exhibiting defective phosphatidylserine synthase activity. *Journal of Biological Chemistry*, *261*(3), 1002–1008.

Yang, L., Ma, D., Ebrahim, A., Lloyd, C. J., Saunders, M. A., & Palsson, B. O. (2016). solveME: Fast and reliable solution of nonlinear ME models. *BMC Bioinformatics*, *17*(1), 1–10. http://doi.org/10.1186/s12859-016-1240-1

Zarecki, R., Oberhardt, M. A., Yizhak, K., Wagner, A., Segal, E. S., Freilich, S., … Ruppin, E. (2014). Maximal sum of metabolic exchange fluxes outperforms biomass yield as a predictor of growth rate of microorganisms. *PLoS ONE*, *9*(5), 1–10. http://doi.org/10.1371/journal.pone.0098372

## Tables

**Table 1: Example for reported experimental flux corresponding to multiple reactions within the *i*CHO1766 model.**

| Enzyme | Reaction | *i*CHO1766 ID | *i*CHO1766 Reactions |
|--------|----------|---------------|----------------------|
| PFK | F6P → DHAP + GAP | 2462;1716 | 1) 'atp_c + f6p_c -> h_c + adp_c + fdp_c '<br>2) 'fdp_c <=> dhap_c + g3p_c ' |

**Table 2: Process information on experimental data used in this study. Further information can be found in the respective original publications.**

| Code | Media | Cell type | Culture type | Culture phase | Source |
|------|-------|-----------|--------------|---------------|--------|
| CM | Sanofi in-hose chemically defined CHO cell growth media | | | | |
| LA | Adjusted CM media with changes in 5 AA concentrations | mAb producing CHO cells | Batch culture in<br><br>250 mL Shakeflasks | Exponential growth phase | McAtee Pereira et al., 2018 |
| LAp | Same composition as LA except suplimented with 4mM Ammonia as control | | | | |

| | | | | | |
|---|---|---|---|---|---|
| **EXP** | | | | Exponential growth phase | |
| | DMEM + 10% FBS + 1% penicillin–streptomycin | CHO-K1 cells | Fed-batch in T-25 flasks | | Ahn & Antoniewicz, 2011 |
| **STAT** | | | | Stationary phase | |
| **FB** | | | Fed-batch in 3L glass bioreactor | | |
| | Proprietary chemically defined CHO cell media and feeds | IgG1 producing GS-CHO | | Stationary phase | Templeton et al., 2017 |
| **PF** | | | Perfusion in 3L glass bioreactor with ATF | | |

**Table 3: Comparison of computational efficiency for FVA, ll-FVA and ccFVA using Gurobi solver. Efficiency is reported as execution time in seconds. In addition to the *i*CHO1766 model the performance comparison was expanded to a wider range of publically available constraint-based models including the *E.coli* core model (Orth, Palsson, & Fleming, 2010) and the *i*MM904 yeast model (Mo et al., 2009). Run times were measured on a standard laptop with an Intel Core i7-5600U CPU processor @ 2.60 GHz and 16GB of RAM on Windows 7 64-bit operating system.**

| | | Execution time (s) | | |
|---|---|---|---|---|
| | **# of Reactions** | **FVA** | **ll-FVA** | **ccFVA** |
| ***E.coli* core** | 95 | 0.97[a] | 4.20[a] | 2.01[a] |
| **iMM904** | 1577 | 45.25[a] | 10263[b] | 59.32[a] |
| ***i*CHO1766** | 5237 | 871.02[a] | >86400[c] | 475.34[a] |

[a] average across six consecutive simulations. [b] average across three consecutive simulations. [c] simulation was aborted at the reported time.

**Table 4: Sample mean data for selected reactions of the CM data set shown in Figure 5A (McAtee Pereira et al., 2018). All values shown are in mmol gDCW$^{-1}$ h$^{-1}$. Reactions indicated with a section sign (§) show differences in predicted directionalities between RS$_{FVA}$ and $RS_{FVA}^{cc}$.**

| | TPI | GAPDH§ | LDH | PPI | FUS | MDH§ | IDH§ |
|---|---|---|---|---|---|---|---|
| 13C mean | 0.29 | -0.60 | -0.29 | -0.02 | 0.35 | 0.21 | 0.29 |

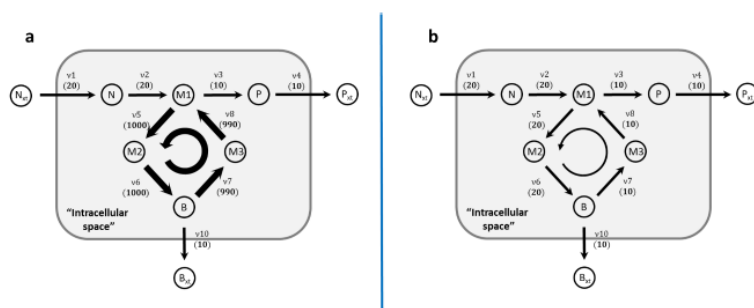| | | | | | | | |
|---|---|---|---|---|---|---|---|
| ccFVA SM | 0.39 | -0.82 | -0.62 | -0.02 | 0.13 | 0.45 | 0.16 |
| FVA SM | 47.58 | 80.87 | -18.41 | -11.92 | 3.44 | -70.03 | -29.07 |

# Figures

**Figure 1: Example network illustrating a conventionally constraint network (a) and a carbon constraint network (b). The arrow thickness is not to scale and is only for visual accentuation of the differences in flux magnitude.**



Figure 1: Example network illustrating a conventionally constraint network (a) and a carbon constraint network (b). The arrow thickness is not to scale and is only for visual accentuation of the differences in flux magnitude.
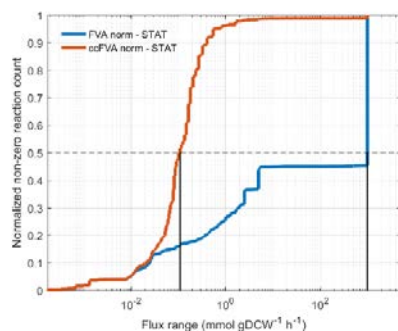
**Figure 2: Comparison of reaction flux ranges between ccFVA and FVA treated *i*CHO1766 model under experimental constraints. Shown in this graph are the flux ranges for the stationary growth condition as reported by Ahn & Antoniewicz, 2011. Results for all other conditions can be found in Supplementary materials SuppInfo2 Figures S4 - S9. For the purpose of this illustration, all bidirectional reactions were split into two individual reactions resulting in a maximum flux range of 1000 mmol gDCW$^{-1}$ h$^{-1}$. The y-axis shows the normalized number of non-zero flux reactions as the number differs between the two different methods. The x- axis shows the flux range on a log-scale with a maximum flux range of 1000 mmol gDCW$^{-1}$ h$^{-1}$.**



Figure 2: Comparison of reaction flux ranges between ccFVA and FVA treated *i*CHO1766 model under experimental constraints. Shown in this graph are the flux ranges for the stationary growth condition as reported by Ahn & Antoniewicz, 2011. For the purpose of this illustration, all bidirectional reactions were split into two individual reactions resulting in a maximum flux range of 1000 mmol gDCW$^{-1}$ h$^{-1}$. The y-axis shows the normalized number of non-zero flux reactions as the number differs between the two different methods. The x- axis shows the flux range on a log-scale with a maximum flux range of 1000 mmol gDCW$^{-1}$ h$^{-1}$.
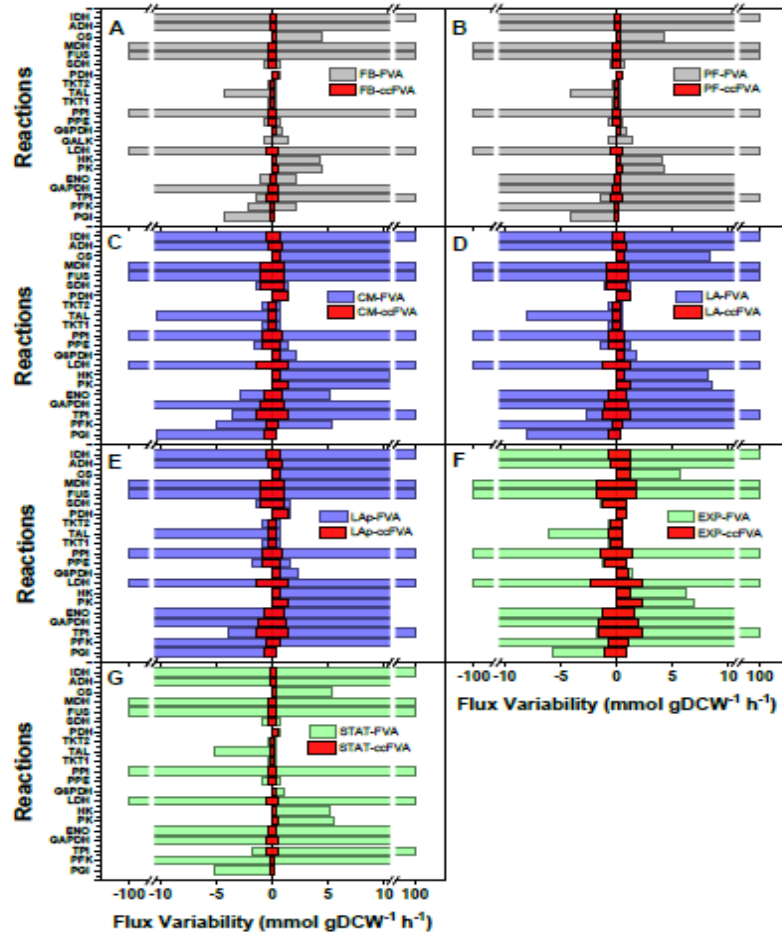
Figure 3: Flux Variability Analysis. (A-G): Comparison between normal FVA (blue, green, grey bars correspond to data from Templeton et al., 2017; Ahn & Antoniewicz, 2011; McAtee Pereira et al., 2018 respectively) and ccFVA (red bars) for central carbon metabolism reactions (glycolysis, TCA, PPP) in *i*CHO1766 across seven distinct experimental datasets. The initial arbitrary bounds $v^{LB}$ and $v^{UB}$ were set to 0 and 100 mmol gDCW$^{-1}$ h$^{-1}$ for unidirectional and -100 – 100 mmol gDCW$^{-1}$ h$^{-1}$ for bidirectional reactions. The discontinued x-axis was chosen in order to increase visibility on the very small flux ranges compared to the ones reaching the large initially set bounds.
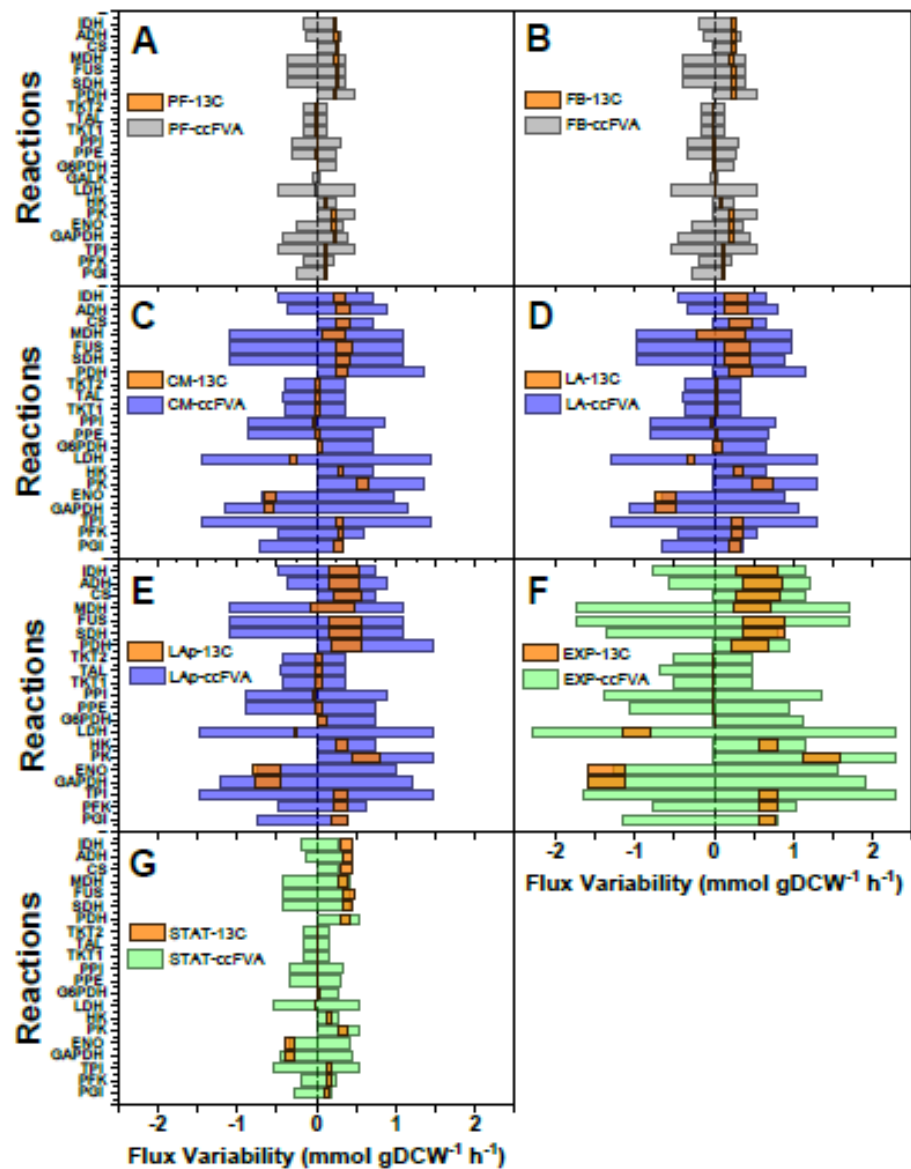
**Figure 4: Comparison of ccFVA to intracellular flux measurements for central carbon metabolism (Glycolysis, TCA, PPP). (A-G): Comparison between ccFVA (blue, green, grey bars correspond to data from Templeton et al., 2017; Ahn & Antoniewicz, 2011; McAtee Pereira et al., 2018 respectively) and $^{13}$C measurements (orange bars).**

**Figure 5: Random Sampling of Feasible Solution Space:** Comparison of sample means derived from FVA (●) and ccFVA (◆) constrained *i*CHO1766 in compariosn to intracelullarly measured fluxes (▉). A total of 200,000 samples were generated for each scenario per experimental condition. Plotted are the means with their respective standard deviation. Reactions marked with a section sign (§) indicate erroneous identification of reaction directionality by samples derived from models constrained with normal FVA ($RS_{FVA}$) but correct prediction by models constrained with ccFVA ($RS_{FVA}^{cc}$). Reactions marked with a double dagger sign (‡) indicate erroneous identification of reaction directionality by samples derived from models constrained with both, $RS_{FVA}$ and $RS_{FVA}^{cc}$. Reactions marked with a dagger symbol (†)indicate erroneous identification of reaction directionality by samples derived from $RS_{FVA}^{cc}$ but correct prediction by models constrained with $RS_{FVA}$. A-B: Results for Templeton's data (FB and PF). C-E: Results for McAtee's data (CM, LA, and LAp respectivly). F-G: Results for Ahn's data (EXP and STAT).