Novel column generation-based optimization approach for poly-pathway kinetic model applied to CHO cell culture

Erika Hagrot[a,b], Hildur Æsa Oddsdóttir[d], Meeri Mäkinen[a,b], Anders Forsgren[d], Véronique Chotteau[a,b,c,*]

[a] *Cell Technology Group, Department of Industrial Biotechnology/Bioprocess Design, School of Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, Stockholm, Sweden*
[b] *AdBIOPRO, VINNOVA Competence Centre for Advanced Bioproduction by Continuous Processing, Sweden*
[c] *WCPR, Wallenberg Centre for Protein Research, Sweden*
[d] *Department of Mathematics, Division of Optimization and Systems Theory, KTH Royal Institute of Technology, Stockholm, Sweden*

ARTICLE INFO

ABSTRACT

Mathematical modelling can provide precious tools for bioprocess simulation, prediction, control and optimization of mammalian cell-based cultures. In this paper we present a novel method to generate kinetic models of such cultures, rendering complex metabolic networks in a poly-pathway kinetic model. The model is based on subsets of elementary flux modes (EFMs) to generate macro-reactions. Thanks to our column generation-based optimization algorithm, the experimental data are used to identify the EFMs, which are relevant to the data. Here the systematic enumeration of all the EFMs is eliminated and a network including a large number of reactions can be considered. In particular, the poly-pathway model can simulate multiple metabolic behaviors in response to changes in the culture conditions.

We apply the method to a network of 126 metabolic reactions describing cultures of antibody-producing Chinese hamster ovary cells, and generate a poly-pathway model that simulates multiple experimental conditions obtained in response to variations in amino acid availability. A good fit between simulated and experimental data is obtained, rendering the variations in the growth, product, and metabolite uptake/secretion rates. The intracellular reaction fluxes simulated by the model are explored, linking variations in metabolic behavior to adaptations of the intracellular metabolism.

1. Introduction

Mathematical modelling can aid in the understanding of underlying biological mechanisms as well as provide tools for bioprocess simulation, prediction, control and optimization. Mammalian cell lines have a complex and flexible metabolism and can display varied metabolic behaviors depending on the culture conditions. A model that captures these variations would be a precious tool for the developments of media, feeds and/or processes in the biopharmaceutical industry. To ultimately function as a predictive tool, such a model must (as a first step) simulate and (in a next step) also predict variations in the metabolic behavior that occurs under a range of culture conditions. Kinetic models relate the culture conditions to the metabolic state of the cells in a quantitative manner, and are thereby well-suited for such purposes. Network-based models benefit from the incorporation of *a priori* knowledge about biochemical reaction pathways for which

detailed information is available in databases for many organisms. Defining the model is however a challenge, as it requires the determination of relevant reactions, metabolic pathways and mostly unknown and potentially complex kinetic equations (Almquist et al., 2014).

While the analysis of the intracellular metabolism of living cells demands expertise and techniques which are complicated and costly (Zamorano et al., 2010; Ben Yahia et al., 2015), the measurements of several extracellular metabolites can be achieved in many laboratories. Macroscopic models have been recognized as useful in this context; they exclude several details of the intracellular metabolism, yet can achieve simulation of rates and concentration profiles relevant to cell cultures (Provost and Bastin, 2004; Provost et al., 2005; Dorka et al., 2009; Gao et al., 2007; Naderi et al., 2011; Zamorano et al., 2013; Hagrot et al., 2017). The macroscopic kinetic model structure can be separated into two parts: (i) the macro-reactions that connect extracellular substrates to products; and (ii) the kinetic equations that

relate the macro-reaction fluxes to the culture conditions (Ben Yahia et al., 2015).

Macro-reactions can be derived from empirical knowledge alone or from a metabolic network, potentially in combination with experimental data and/or statistical analysis. In the latter case, methods from pathway analysis can be used to obtain elementary flux modes (EFMs) (Schuster and Hilgetag, 1994; Klamt and Stelling, 2003; Papin et al., 2004; Llaneras and Picó, 2010). An EFM is a stoichiometrically balanced linear combination of individual network reactions, and provides a route through the network that connects extracellular substrates to products. The experimental data can be taken into account by combining the EFMs with metabolic flux analysis (MFA), forming the EFMs-based MFA problem (Provost, 2006); the problem is solved via estimation of the macro-reaction fluxes such that the squared residuals between the EFM model and data are minimized. The problem is developed into a macroscopic kinetic model as the flux over each macro-reaction is described by a kinetic equation whose parameters become targets for the estimation.

Generalized Monod- or Michaelis-Menten-type equations have been frequently used as the starting point to formulate the kinetic equations in macroscopic models (Provost and Bastin, 2004; Naderi et al., 2011; Hagrot et al., 2017). Examples of variables that can be incorporated into these equations include the concentrations of medium components and metabolic by-products, as well as other process parameters. The parameters of the equations can be estimated from literature and/or by fitting the model to experimental data, typically using least squares or maximum likelihood functions (Ben Yahia et al., 2015). However, non-linear problems (as given by the Michaelis-Menten-type equations) are generally difficult to solve, especially when there is a large number of parameters; challenges may include multiple local minima and over-fitting issues (Ben Yahia et al., 2015). Fixing the saturation parameters yields a linear problem for which only the maximum flux rates of the equations need to be estimated (Provost and Bastin, 2004; Dorka et al., 2009; Hagrot et al., 2017). In particular, the strategy of setting the saturation parameters sufficiently small (or large) such that the inputs have little or no impact on the outputs have been applied in many cases, and justified under conditions of balanced growth (Provost and Bastin, 2004; Provost et al., 2005; Dorka et al., 2009; Zamorano et al., 2013; Ben Yahia et al., 2015).

The EFMs of a metabolic network can be systematically enumerated using, e.g., the Metatool algorithm (von Kamp and Schuster, 2006) or other software (Klamt et al., 2007; Schwarz et al., 2007), and then provide a comprehensive representation of all possible pathways through the network. With increasing size and complexity of the metabolic network, there is an explosion of possible routes and the EFM enumeration becomes computationally prohibitive (Klamt and Stelling, 2002). Models developed based on EFM enumeration are thereby limited to simplified networks. In this context, it has been suggested to strive for a reduced set of EFMs and to use experimental data to guide the simplification of the network prior to the enumeration (Gao et al., 2007; Niu et al., 2013; Naderi et al., 2011): based on a preliminary metabolic flux analysis, the reactions with insignificant flux are identified and removed, yielding small numbers of EFMs (Gao et al., 2007; Niu et al., 2013) that are even manually enumerable (Gao et al., 2007). However, the loss of network detail at an early stage of the model development limits the modelling capacity in many cases.

Pathway analysis of complex networks can be made possible by focusing on a subset of the EFMs, while accepting that the remaining EFMs stay unknown. An advantage here is that EFMs in the subset are based on a network with preserved size and complexity, rather than simplified reactions. The identification can be made feasible, e.g. by focusing on finding the EFMs using the fewest number of reactions (de Figueiredo et al., 2009), subsystem analysis (Kaleta et al., 2009) or random sampling (Machado et al., 2012; Tabe-Bordbar and Marashi, 2013). Another option is to let the identification be guided by experimental data (Jungers et al., 2011; Oddsdóttir et al., 2014).

Jungers et al. (2011) suggested a method that first generates a flux vector feasible for a metabolic state and then randomly decomposes this vector into a small set of EFMs. The resulting algorithm is later used to build a kinetic EFM model, simulating the dynamics of a CHO a cell culture based on a detailed metabolic network (Zamorano et al., 2013).

Column generation (CG) (Ford and Fulkerson, 1958; Dantzig and Wolfe, 1960; Gilmore and Gomory, 1961, 1963; Desrosiers et al., 1984), is a general optimization technique that can be used to efficiently solve large linear problems, and that has been successfully applied to areas such as vehicle and aircraft routing (Lübbecke and Desrosiers, 2005). The idea of the technique is that, as many variables may be zero in the optimal solution, only a subset of the variables needs to be taken into account when solving the problem. In previous work (Oddsdóttir et al., 2014, 2016) we applied CG, for the first time, to identify pathways (EFM subsets) in metabolic networks. Using our novel CG-based algorithm, the EFMs-based MFA problem can be efficiently solved as the EFM identification and data-fitting procedure are carried out dynamically, without prior enumeration of the EFMs or flux vector estimation (Oddsdóttir et al., 2014). The algorithm delivers a subset of EFMs and the corresponding macro-reaction fluxes that are able to optimally fit the data in a least-squares data-fitting sense, i.e., the solution is a global optimum to the EFMs-based MFA problem.

In our previous work, we introduced the poly-pathway model approach (Hagrot et al., 2017) aiming to capture the metabolic behavior of multiple experimental conditions in a single kinetic EFM model. The approach builds on the concept that the cells adapt their metabolism in response to external stimuli by using metabolic pathways in different combinations, e.g., to compensate for nutrient depletion or product accumulation. In the model, the metabolic pathways are represented by macro-reactions obtained from the EFMs of a metabolic reaction network. In this first proof-of-concept, the approach is demonstrated using a simplified network for which the complete set of EFMs is enumerated using Metatool algorithm. For the kinetic equations, we developed a strategy of flexible kinetics in which different equations are formulated and then identified from experimental data generated in parallel CHO cell cultures subjected to variations in amino acid availability. The simplification of the metabolic network imposed by the EFM enumeration step is identified as a key limitation in this study (Hagrot et al., 2017).

In the present work, a novel method based on CG is developed for the poly-pathway model approach. The CG algorithm (Oddsdóttir et al., 2014) is integrated into the modelling framework, such that complex metabolic networks can be used for the model development. The experimental data are used to identify the EFMs, which are relevant to the data, eliminating the systematic EFM enumeration.

The outline of the paper is as follows. In Section 2, we briefly present the theory behind the EFMs-based MFA-problem, the poly-pathway model approach and the CG algorithm. A metabolic network, for which the total enumeration of EFMs is prohibitive, is introduced in Section 3. The experimental data (Hagrot et al., 2017) are briefly presented in Section 4. In Section 5, we introduce our general modelling approach to build the model and identify its parameters, based on the theoretical part of Section 2, 3. Then we incorporate the CG algorithm into our modelling framework, perform EFM identification, formulate the kinetic equations, generate a poly-pathway model using cross-validation, and explore the detailed intracellular flux distributions simulated by the resulting model. Finally, Section 6 provides our conclusions.

2. Theory

2.1. EFMs-based MFA

A metabolic network is defined by the two stoichiometric matrices $\mathbf{A}_{ext}$ (rows of extracellular metabolites) and $\mathbf{A}_{int}$ (rows of intracellular

metabolites), and J_{irrev} (the set of irreversible reactions). The reaction flux vector $\mathbf{v}$ is a solution to the fundamental equation of MFA (1) and is consistent with the stoichiometry and irreversibility constraints of the metabolic network, the experimentally determined uptake and secretion rates in $\mathbf{q}_{ext}$, and the pseudo steady-state assumption (Quek et al., 2010),

$$\begin{bmatrix} \mathbf{A}_{ext} \\ \mathbf{A}_{int} \end{bmatrix} \mathbf{v} = \begin{bmatrix} \mathbf{q}_{ext} \\ \mathbf{0} \end{bmatrix}, \quad v_j \geq 0, \quad j \in J_{irrev}. \tag{1}$$

The complete set of EFMs in the metabolic network is defined by the column vectors $\mathbf{e}_l$ in the matrix $\mathbf{E} = [\mathbf{e}_1 \quad \mathbf{e}_2 \quad \ldots \quad \mathbf{e}_L]$. Non-zero entries $e_{j,l}$ in $\mathbf{e}_l$ indicate the stoichiometric contribution of reaction j in EFM l. By EFM, we refer to a linear combination of network reactions. Each EFM is associated with a macroscopic flux w_l in the column vector $\mathbf{w} = [w_1 \quad w_2 \quad \ldots \quad w_L]^T$. Each $\mathbf{v}$ can be expressed as a non-negative linear combination of the EFMs in which $\mathbf{w}$ are the weights: $\mathbf{v} = \mathbf{Ew}$. The experimental data relate to the fluxes in the macroscopic model by: $\mathbf{q}_{ext} = \mathbf{A}_{ext}\mathbf{v} = \mathbf{A}_{ext}\mathbf{Ew}$. Here, $\mathbf{A}_{ext}\mathbf{E}$ derives the macro-reactions from the EFMs; each column provides stoichiometric coefficients for the extracellular metabolites involved. Finally, the EFMs-based MFA problem is to

$$\underset{\mathbf{w}}{\text{minimize}} \quad \frac{1}{2} \sum_{k=1}^{D} \| \mathbf{q}_k - \mathcal{I}_k \mathbf{A}_{ext} \mathbf{Ew} \|_2^2$$
$$\text{subject to} \quad \mathbf{w} \geq 0. \tag{2}$$

Here, $\mathbf{q}_k$ corresponds to the $\mathbf{q}_{ext}$-vector calculated from the concentration measurements over one day in one experimental condition, and D is the total number of culture days for this particular condition. $\mathcal{I}_k$ is an identity matrix that can be defined such that rows corresponding to missing metabolite data for certain culture days are removed from the problem. Since (2) is a convex problem, global optimality is guaranteed. However, the solution is not necessarily unique.

2.2. The poly-pathway modelling framework

The poly-pathway model approach takes its starting point in (2). However, rather than finding $\mathbf{w}$ for a single experimental condition, the aim is to combine data from multiple conditions via the introduction of kinetic equations. It is not obvious how to formulate the kinetic equations for macroscopic reactions. In cell culture kinetics, the Monod-type equations capable of simulating saturation, inhibition and metabolite limitation have been frequently applied (Ben Yahia et al., 2015). In our previous implementation (Hagrot et al., 2017), the kinetic equations are formulated based on the general equation,

$$w_{k,l} = w_{max,l} \cdot f_l(\mathbf{c}_{0,l,k}, \mathbf{K}_l), \tag{3}$$

where

$$f_l(\mathbf{c}_{0,l,k}, \mathbf{K}_l) = \prod_{i \in M_{ext,s,l}} \underbrace{\frac{\frac{c_{0,k,i}}{K_{s,i,l}}}{\frac{c_{0,k,i}}{K_{s,i,l}} + 1}}_{\text{substrate saturation}} \cdot \prod_{i \in M_{ext,p,l}} \underbrace{\frac{1}{\frac{c_{0,k,i}}{K_{p,i,l}} + 1}}_{\text{product inhibition}} \cdot \prod_{i \in M_{ext,r,l}} \underbrace{\frac{1}{\frac{c_{0,k,i}}{K_{r,i,l}} + 1}}_{\text{metabolite inhibition}}.$$

$M_{ext,s,l}$ and $M_{ext,p,l}$ are the sets of extracellular substrates and products in macro-reaction l, respectively. $M_{ext,r,l}$ is a set of extracellular metabolites freely selected and not necessarily involved in the macro-reaction. $\mathbf{c}_{0,l,k}$ are the concentrations of extracellular metabolites in measurement k; for substrates in $M_{ext,s,l}$, products in $M_{ext,p,l}$ and metabolites in $M_{ext,r,l}$. The Eq. (3) is linear in the parameter $w_{max,l}$, which represents the maximum flux over macro-reaction l, and non-linear in the parameters $\mathbf{K}_l$. In the function f_l $K_{s,i,l}$ is a saturation parameter for metabolite i in $M_{ext,s,l}$, and $K_{p,i,l}$ and $K_{r,i,l}$ are inhibition parameters for a product or metabolite i in $M_{ext,p,l}$ and $M_{ext,r,l}$, respectively. The effects of saturation or inhibition on the macroscopic flux rate imposed by a single metabolite are explained in the supplementary material (Fig. S4). The problem in (2) now incorporating (3) for a total of D culture days
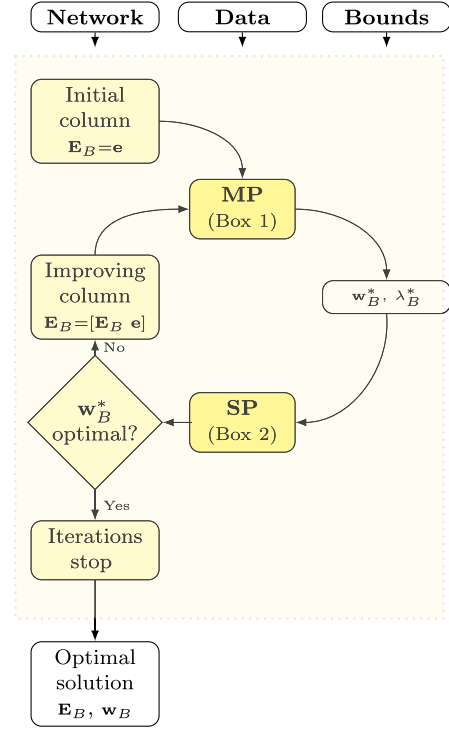

Fig. 1. CG algorithm. Flow chart showing the CG-based identification of an EFM subset using the CG algorithm.

in multiple experimental conditions is to

$$\underset{\mathbf{w}_{max}, \mathbf{K}}{\text{minimize}} \quad \frac{1}{2} \sum_{k=1}^{D} \| \mathbf{q}_k - \mathcal{I}_k \mathbf{A}_{ext} \mathbf{EF}(\mathbf{c}_{0,k}, \mathbf{K}) \mathbf{w}_{max} \|_2^2$$
$$\text{subject to} \quad \mathbf{w}_{max} \geq \mathbf{0}, \tag{4}$$

where $\mathbf{F}$ is a diagonal matrix with f_l on its diagonal, non-linear in the parameters $\mathbf{K}$. (4) is a complicated non-linear and non-convex optimization problem for which optimality can only be determined locally.

2.3. CG-based identification of EFM subsets in metabolic networks

Here, we use the workflow of the CG algorithm (Fig. 1), originally developed by Oddsdóttir et al. (2014) for the purpose of identifying EFMs in metabolic networks. The algorithm solves the data-fitting problem (2) for a single metabolic state by finding a subset $\mathbf{E}_B$ of the EFMs in $\mathbf{E}$ and an associated flux vector $\mathbf{w}_B$ that fits the experimental data optimally. Two inputs are necessary: (i) the metabolic network; and (ii) the experimental data. Optionally, bounds on the uptake/secretion rates for unmeasured metabolites can be provided (Oddsdóttir et al., 2016). Inside the algorithm, two optimization problems (the Master Problem (MP) and the subproblem (SP)) are solved iteratively. Initially, $\mathbf{E}_B$ is empty. As a first step, an initial column $\mathbf{e}$ is generated and added to $\mathbf{E}_B$. At each iteration, the MP (Box 1) finds a solution $\mathbf{w}_B^*$ for the current $\mathbf{E}_B$ and forwards this solution to the SP (together with the dual variable λ_B^*). Solving the SP (Box 1) determines if the current $\mathbf{E}_B$ and $\mathbf{w}_B^*$ constitute an optimal solution to the MP. If the answer is no, the SP identifies a new column $\mathbf{e}$ whose addition to $\mathbf{E}_B$ decreases the objective function value in the MP and adds $\mathbf{e}$ to $\mathbf{E}_B$. If the solution is optimal, the algorithm terminates. The solution is globally optimal but not necessarily unique.

3. Metabolic network

A metabolic network of CHO cell metabolism with 126 reactions is used. For this, the network presented by Zamorano et al. (2010) is further

Box 1. Master Problem (MP)

$$
\underset{\mathbf{w}_B, \mathbf{z}^{UB}, \mathbf{z}^{LB}}{\text{minimize}} \quad \frac{1}{2} \sum_{k=1}^{D} \| \mathbf{q}_k - \mathcal{I}_k \mathbf{A}_{ext} \mathbf{E}_B \mathbf{w}_B \|_2^2 +
$$

$$
+ \mathbf{M}_{UB}^T \mathbf{z}^{UB} + \mathbf{M}_{LB}^T \mathbf{z}^{LB}, \tag{5a}
$$

$$
\text{subject to} \quad \mathbf{w}_B \geq \mathbf{0} \tag{5b}
$$

$$
\mathbf{z}^{UB} - \mathbf{A}_{nm} \mathbf{E}_B \mathbf{w}_B \geq - \mathbf{q}_{nm}^{UB} \tag{5c}
$$

$$
\mathbf{z}^{LB} + \mathbf{A}_{nm} \mathbf{E}_B \mathbf{w}_B \geq \mathbf{q}_{nm}^{LB} \tag{5d}
$$

$$
\mathbf{z}^{UB} \geq \mathbf{0} \tag{5e}
$$

$$
\mathbf{z}^{LB} \geq \mathbf{0} \tag{5f}
$$

The bounds for unmeasured metabolites are represented via the penalty function in (5a) with $\mathbf{M}_{UB}$ and $\mathbf{M}_{LB}$ set sufficiently large to satisfy these requirements. $\mathbf{A}_{nm}$ in (5c) and (5d) is a stoichiometric matrix for the unmeasured metabolites and $\mathbf{q}_{nm}^{UB}$ and $\mathbf{q}_{nm}^{LB}$ are the assumed upper and lower bounds for their uptake/secretion rates, respectively. The problem is handled in extended space; thereby, all the macro-reactions/EFMs are considered irreversible and all the macroscopic fluxes are positive (5b).

developed and adapted as follows: (i) it is compartmentalized, i.e., a mitochondrial subspace is introduced; (ii) several reactions are made reversible; (iii) the reactions involved in biomass formation are modified; and (iv) a mAb synthesis reaction is added. In the resulting network 74 of the 126 reactions are reversible, 29 of the metabolites are extracellular and 89 intracellular, and fifteen subsystems are represented (Fig. 2). Additional details and abbreviations are available in the supplementary material (Section S2 and Tables S1–S6). Energetic co-factors (e.g. ATP, NADH and NADPH) are involved in the metabolism (Tables S4 and S5, Fig. S19). Similarly to the work of Zamorano et al. (2010), these co-factors are excluded from the network in the generation of the model because they are not internally balanced. The co-factor balances are calculated based on the final model and are presented in Supplementary Table S17.

4. Experimental data

The experimental data are generated and presented in previous work (Hagrot et al., 2017), based on the original amino acid cocktail of the Irvine Scientific medium. Ten amino acids (alanine, asparagine,

Box 2. Subproblem (SP)

$$
\underset{\mathbf{e}}{\text{minimize}} \quad \sum_{k=1}^{D} ((\mathbf{A}_{ext}^T \mathcal{I}_k^T (\mathcal{I}_k \mathbf{A}_{ext} \mathbf{E}_B \mathbf{w}_B^* - \mathbf{q}_k))^T +
$$

$$
+ (\lambda_{UB}^k - \lambda_{LB}^k)^T \mathbf{A}_{nm}) \mathbf{e}, \tag{6a}
$$

$$
\text{subject to} \quad \mathbf{A}_{int} \mathbf{e} = \mathbf{0}, \tag{6b}
$$

$$
\mathbf{1}^T \mathbf{e} \leq 1, \tag{6c}
$$

$$
\mathbf{e} \geq 0. \tag{6d}
$$

The objective function (6a) is derived from the optimality conditions for a globally optimal solution to the MP. If the objective function value of (6a) is negative, then the EFM ($\mathbf{e}$) found will be included in the MP. If the value is zero, $\mathbf{E}_B$ is an EFM subset that solves the MP to optimality and the CG algorithm terminates. The constraints are formed such that the column $\mathbf{e}$ fulfills the pseudo-steady state assumption (6b), the problem is bounded (6c), and the reactions are irreversible (6d). $\mathbf{1}$ is a vector consisting only of ones. λ_{UB}^k and λ_{LB}^k are the dual variables corresponding to the constraints (5c) and (5d), respectively.

aspartate, glutamine, glutamate, glycine, serine, threonine, cysteine, and proline) are selected from preliminary experiments in our lab, showing effects on the metabolism. Individual amino acids are removed (A0, N0, D0, Q0, E0, G0 and S0), reduced at 50% (C50) or doubled in concentration (A200, N200, D200, Q200, T200, S200 and P200), creating sixteen different culture media (Table 1). Except for amino acids deliberately omitted in the media, none of the measured metabolites are depleted during the cultures - depletion would imply incorrect flux measurements since those are calculated from daily sampling analyses.

The cells are cultured in parallel TubeSpin bioreactors in pseudo-perfusion mode with daily medium renewals. The cell-specific rates of growth (μ) and uptake/secretion of extracellular metabolites and mAb (q_{ext}) are calculated from concentration measurements in samples collected before and after each medium renewal according to,

$$
\mu = \frac{ln\left(\dfrac{X_v}{X_{v,0}}\right)}{t - t_0}, \tag{7}
$$

$$
q_{ext} = \mu \cdot \frac{c - c_0}{X_v - X_{v,0}}, \tag{8}
$$

$X_{v,0}$ is the viable cell concentration after medium renewal and X_v is the viable cell concentration before the following medium renewal, c_0 and c are the corresponding metabolite concentrations, and t_0 and t are the corresponding consecutive time points of sample collection. The measurements of six metabolites are either not available (carbon dioxide ($CO_{2,ext}$), ethanolamine ($Ethn_{ext}$), choline (Cho_{ext}) and urea ($Urea_{ext}$)) or of poor accuracy (arginine (Arg_{ext}) and histidine (His_{ext})). Bounds for the corresponding uptake/secretion rates are defined based on partial measurement and/or literature as described in the supplementary material (Table S7).

5. Results and discussion

5.1. Overview

The present section gives an overview of the different parts applied for the modelling. These parts are then detailed in dedicated sections.

i) Concept of poly-pathway modelling using the CG-algorithm

The concept to generate a poly-pathway model using the CG-algorithm is as follows.

EFM identification

The macro-reactions of the metabolic reaction network are created by an EFM approach. Given the metabolic network, the experimental data and the bounds, the EFM subsets are generated using the CG algorithm (see Fig. 3). Each of the sixteen data sets listed in Table 1 have been measured in a culture where a given variation in the amino acid availability has been applied. One EFM subset is generated for each data set. Among all the possible EFMs of the metabolic reaction network, each subset includes only the EFMs relevant to the information present in the data (Oddsdóttir et al., 2014, 2016). The subsets are then combined into a set of unique EFMs by forming the union $\mathbf{E}_{union}$ of the sixteen EFM subsets, generating the macro-reaction network.

Kinetic model development

The kinetic model is based on the macro-reactions network combined with kinetics equation. To determine the kinetics part, we apply our strategy of flexible kinetics (Hagrot et al., 2017): one or several potential kinetic equations with saturation and/or inhibition effects are attributed for each EFM in $\mathbf{E}_{union}$, generating a much larger potential model, LargeM, composed of the macro-reactions and several kinetics alternative per macro-reaction. The potential kinetic equations vary in the effect of the metabolites. The saturation parameters $\mathbf{K}_s$ of
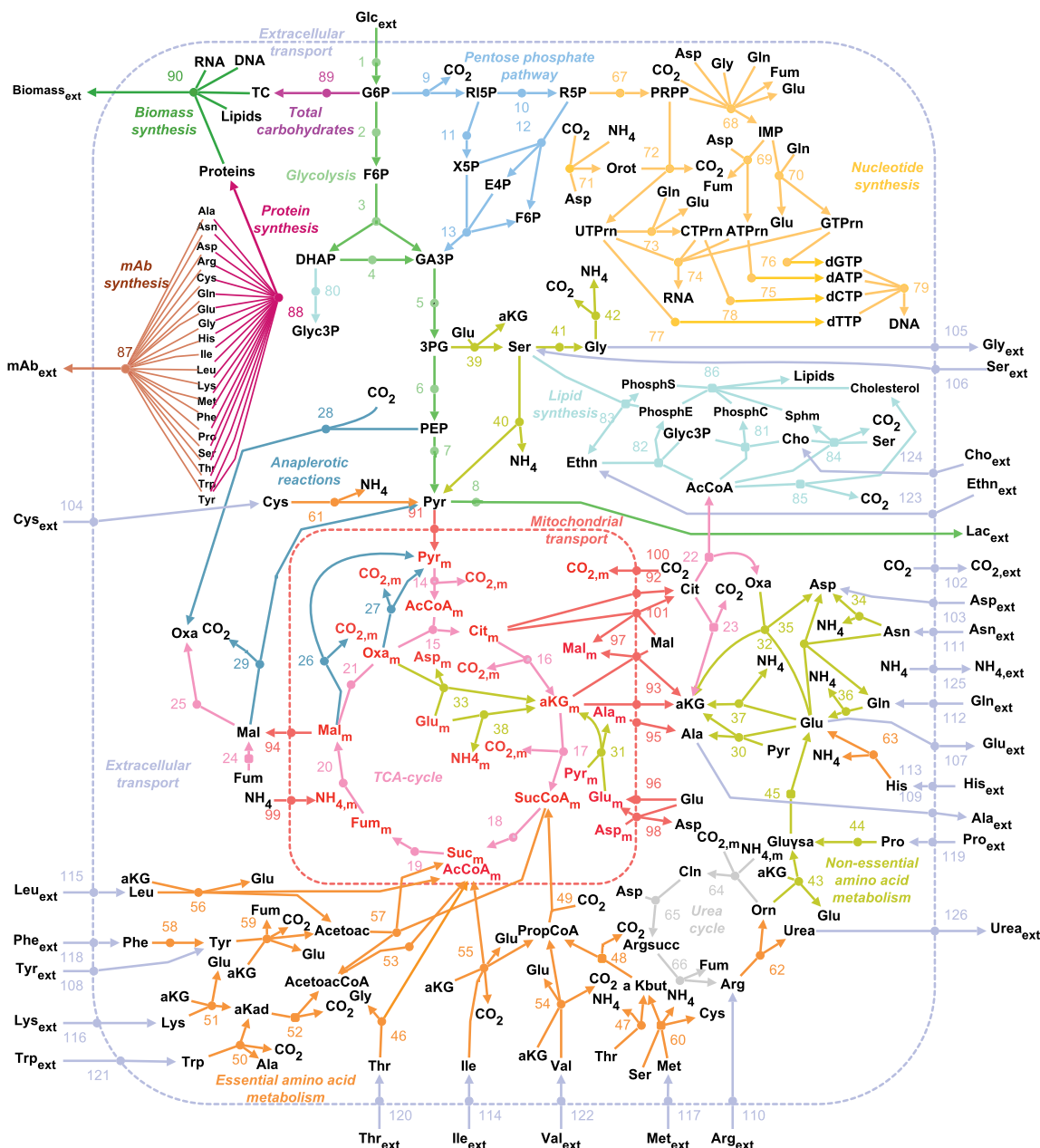
Fig. 2. Metabolic network map. The subscript m denotes mitochondrial metabolites and these metabolites are shown in red. The subscript ext denotes extracellular metabolites. The boundaries between compartments are indicated by dashed lines. Subsystems are indicated by color. The visualization is created with VANTED (Rohn et al., 2012b) (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

Table 1
Media used in pseudo-perfusion culture. The media #1–15 are referred to by code Xy, where X is the one-letter abbreviation of the varied amino acid and y is its concentration level in percent relative to the concentration in the control medium #16 (which is referred to as Ctrl).

#	Amino acid	Level (%)	Code	#	Amino acid	Level (%)	Code
1	Alanine	0	A0	9	Glutamate	0	E0
2	Alanine	200	A200	10	Glycine	0	G0
3	Asparagine	0	N0	11	Threonine	200	T200
4	Asparagine	200	N200	12	Serine	0	S0
5	Aspartate	0	D0	13	Serine	200	S200
6	Aspartate	200	D200	14	Proline	200	P200
7	Glutamine	0	Q0	15	Cysteine	50	C50
8	Glutamine	200	Q200	16	—	—	Ctrl

the kinetics are taken from the literature while the inhibition parameters K_p and K_r are obtained by fitting the fluxes estimated by the model to the experimental data in LargeM. The maximal flux rates w_{max} are estimated by linear regression to fit the model to the experimental data. Adjustment of K_p and K_r and weighting are introduced to improve the fit between the model and the data.

Model reduction

LargeM model is reduced into model RedM by removing the equations associated with very low maximum flux rates, since these have no influence on the model.

ii) Workflow for the model identification

The workflow includes the determination of the macro-reactions and all the potential kinetics in Step 1 followed by the identification of the final reduced model using a cross-validation approach in Step 2.

Determination of the macro-reactions and potential kinetics - Step 1
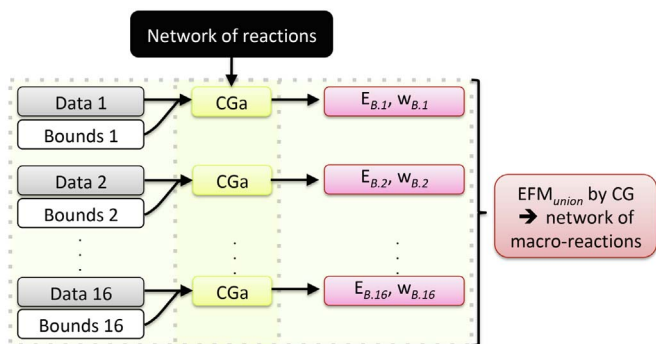
Fig. 3. Determination of the network of macro-reactions by CG. The CG-based algorithm (CGa, yellow), replaces the conventional enumeration of EFMs. The identification of each EFM subset (pink) is carried out independently for each experimental condition. The subsets are then combined into the EFM union.

In Step 1, the procedures described above of *EFM identification* and *Kinetic model development* are sequentially applied, generating an initial LargeM model, LargeM$_{initial}$, see Fig. 4. The macro-reaction network defined by $\mathbf{E}_{union}$ and the kinetics of LargeM are then used as inputs for Step 2.

Identification of the final model - Step 2

A cross-validation approach is applied to identify the final model, see Fig. 5. The data are randomly distributed in a training data set and a testing data set. For the training data set d, the macro-reaction network defined by $\mathbf{E}_{union}$ and the kinetics of Step 1 are used to determine a large model LargeM^d. Using the *Model reduction* method described above, LargeM^d is reduced into RedM^d. RedM^d is then used on the testing data set to simulate the flux rates and the error between the simulated and measured data is computed. This is repeated in a 4-fold approach. From all the repetition exercises, the RedM^d providing the smallest simulation error for the testing data is selected as final model RedM$_{final}$.

5.2. EFM identification

5.2.1. EFM subsets generated by column generation

The CG algorithm is run in MATLAB (version R2018a) based on previous implementations (Oddsdóttir et al., 2014, 2016). For the present work, the glpk solver (GLPK and GNU,) is implemented to efficiently solve the subproblems. In addition, an upper limit of 1000 pmol/cell, day is applied on all the macroscopic fluxes (i.e., $0 \leq w_l \leq 1000$), without loss of generality since this value is very large. Using the CG algorithm, we solve a series of sixteen independent EFMs-based MFA problems (one for each experimental condition) yielding sixteen EFM subsets, with corresponding sets of macro-reactions and macroscopic flux estimates. The total number of EFMs in each subset varies between 25 and 29 (in extended space), including 1–3 EFMs involved in biomass synthesis and 1 EFM in mAb synthesis (Table 2). The CG algorithm starts from an empty set of EFMs, and at each iteration the addition of a new candidate EFM is evaluated (Fig. 1). When the optimality criterion is achieved, the iterations stop (Oddsdóttir et al., 2014). We do not restrict the number of EFMs to be found. Nonetheless, a rather small set of EFMs and a small number of EFMs involved in biomass or mAb synthesis are found. These small numbers in comparison with systematic enumeration of millions of EFMs are potentially linked to the amount of information present in the experimental data. As described in Section 2, an EFM subset delivered by the CG algorithm fits the data of the associated experimental condition optimally in the least-squares sense. The fit is good for all the metabolites except lysine for which the specific uptake rates are generally overestimated compared to the data.

5.2.2. The EFM union

The sixteen EFM subsets are combined into $\mathbf{E}_{union}$ (Table S8), which included 125 EFMs in total, representing a network of 125 macro-reactions. It can be noted that $\mathbf{E}_{union}$ is in itself a subset of $\mathbf{E}$, and as it contains all the sixteen EFM subsets it can optimally fit the data of each experimental condition. A comparison between the sixteen EFM subsets and $\mathbf{E}_{union}$ reveals that the majority of the EFMs in $\mathbf{E}_{union}$ occurs in one EFM subset only, while a small number of EFMs occurs in all the sixteen subsets (see Fig. 6). A complete overview of this comparison is available in the supplementary material (Tables S9 and S10). EFMs representing the reversible conversion of glutamine into glutamate and ammonium, the reversible conversion between serine and glycine, and the conversion of proline into glutamate are common to all the sixteen subsets. Detailed characteristics of the EFMs are illustrated in the supplementary material (i.e. the stoichiometric coefficients of the macro-reactions (Fig. S2), and the comparison of EFMs between the sixteen EFM subsets (Fig. S3)).

5.3. Kinetic model development

The poly-pathway model approach is based on the idea that the fluxes over the pathways are regulated or switched on and off depending on the culture conditions. Since the macro-reactions are issued from several metabolic reactions, the determination of the kinetics is a challenging task in particular for the inhibition effects. Rather than *a priori* assuming one kinetic equation for each macro-reaction, we provide several different kinetic equations as 'potential' kinetics and then let the importance of each kinetic equation be automatically determined based on the data via the parameter estimation step, leading to their final selection. The kinetics of substrate saturation are Michaelis-Menten equations with parameters taken from the reported information about the transport of the metabolites into the cells (see supplementary material S12). For the saturation effects, the parameters are not changed but an effect is potentially removed as alternative kinetics. The potential kinetics of inhibition effect are designed based on the effects of the metabolites observed in our experimental data and/or indication from general biochemistry knowledge of CHO or mammalian cells. During the modelling development Step 1, some of the parameters of inhibition kinetics are tuned to improve the model fitting to the experimental data. During these iterations, some parameters are set to the very small value of 1e-05, which is equivalent to have the maximum flux in (3). Using the general kinetic Eq. (3), one or several kinetic equations are defined for each macro-reaction as given in Tables 3 and S11. Each additional equation is added to the problem in (4) by extending $\mathbf{E}$, $\mathbf{w}_{max}$, and $\mathbf{F}$ with the corresponding EFM column, w_{max}-parameter and kinetic function, respectively:

$$\mathbf{E} = [\mathbf{E}_{union} \quad \mathbf{E}_{added}], \quad \mathbf{F} = \begin{bmatrix} \mathbf{F}_{union} \\ \mathbf{F}_{added} \end{bmatrix}, \quad \mathbf{w}_{max} = \begin{bmatrix} \mathbf{w}_{max,union} \\ \mathbf{w}_{max,added} \end{bmatrix}. \quad (9)$$

5.3.1. Estimation of $\mathbf{w}_{max}$

The parameters in $\mathbf{w}_{max}$ are estimated by solving the following constrained linear-least-squares problem using the function `lsqlin` with the interior-point algorithm (MATLAB, R2018a):

$$\underset{\mathbf{w}_{max}}{\text{minimize}} \quad \frac{1}{2} \sum_{k=1}^{D} \| \mathcal{P}_k \mathcal{N}_k (\mathbf{q}_k - \mathbf{A}_{ext} \mathbf{E} \mathbf{F} \mathbf{w}_{max}) \|_2^2$$

$$\text{subject to} \quad 0 \leq \mathbf{w}_{max} \leq 1000$$

$$\mathbf{q}_{nm}^{LB} \leq \mathbf{q}_{nm} \leq \mathbf{q}_{nm}^{UB}. \quad (10)$$

A normalization is applied to attenuate the effect of the variation in order of magnitude between the measured rates, i.e., the normalization matrix $\mathcal{N}_k$ is as a diagonal matrix in which the entries are $|q_{k,i}|^{-1}$ when $|q_{k,i}| > 0.01$ and otherwise 1 ((Hagrot et al., 2017)). We also introduce weighting factors (Table S13) to adjust the influence of the data sets on
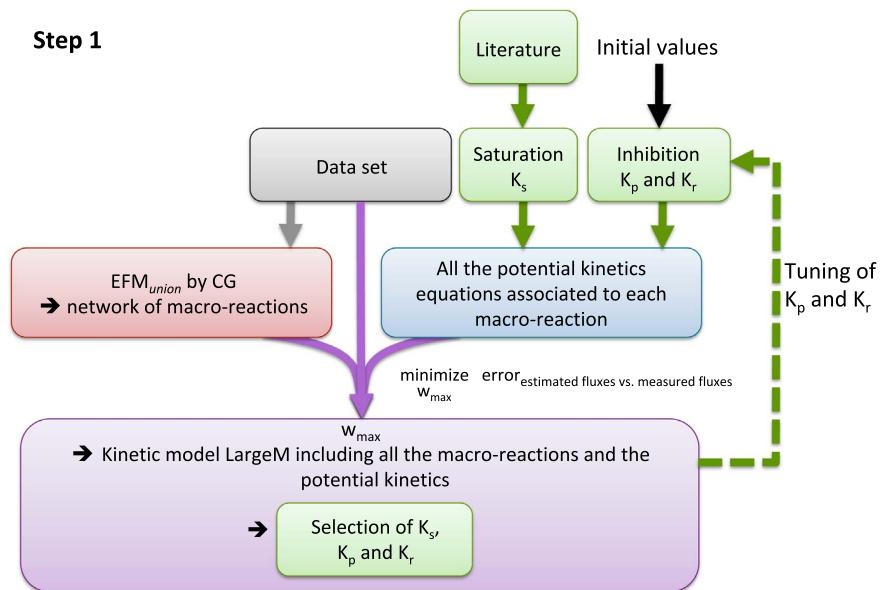
Fig. 4. Determination of the macro-reactions and kinetics - Step 1. The macro-reactions of the metabolic reaction network are created by an EFM approach: given the metabolic network, the experimental data and the bounds, the $\mathbf{E}_{union}$ is generated using the CG algorithm (see Fig. 3). One or several potential kinetic equations with saturation and/or inhibition effects are attributed for each EFM in $\mathbf{E}_{union}$, noted as 'All the potential kinetics equations associated to each macro-reaction'. This generates a large potential model, LargeM, composed of the macro-reactions and several kinetics alternative per macro-reaction. The potential kinetic equations vary in the effect of the metabolites. The saturation parameters $\mathbf{K_s}$ of the kinetics are taken from the literature while the inhibition parameters $\mathbf{K_p}$ and $\mathbf{K_r}$ are obtained by fitting the fluxes estimated by the model to the experimental data in LargeM.

the parameter estimation (i.e., the weighting matrix $\mathcal{P}_k$ is a diagonal matrix in which the entries are the weights given to each row in (10)). The fluxes with higher information content are attributed a higher weight as commonly used to select the weights during an iterative modelling exercise (Montgomery et al., 2015) so that the model fit is improved. This strategy is used to improve the fit for particular variations in some experimental conditions that are not well captured

in the preliminary estimations. Furthermore, the data with a larger uncertainty are attributed a lower weight, e.g. lysine uptake. The resulting model is an initial and large model (Step 1), LargeM$_{initial}$, including the macro-reactions (or EFMs) and the kinetics. The simulated fluxes generated with this model are compared to the measured fluxes in Supplementary material Figs. S5–S7 showing a good agreement of the model to the data.
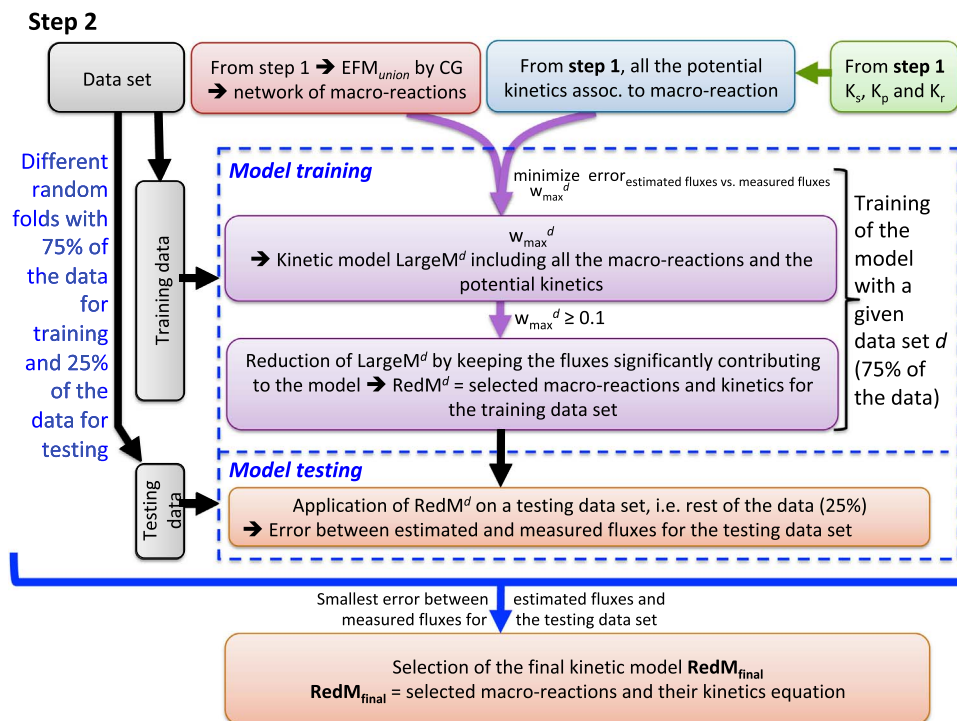
Fig. 5. Identification of the reduced model - Step 2. A cross-validation approach is applied to identify the final model. The data are randomly distributed in a training data set and a testing data set. For each training data set d, the macro-reaction network defined by $\mathbf{E}_{union}$ and the kinetics of Step 1 are used to determine a large model LargeM^d. The LargeM^d is reduced into RedM^d. RedM^d is then used on the testing data set to simulate the flux rates and the error between the simulated and measured data is computed. From all the repetition exercises, the RedM^d providing the smallest simulation error for the testing data is selected as final model RedM$_{final}$.

Table 2
Number of EFMs in subsets and union. The number of EFMs in the sixteen EFM subsets delivered by the CG algorithm and in the E_{union}. 'Removed' refers to the number of EFMs removed during the iterations, i.e., EFMs that are firstly included in an EFM subset and then removed by the CG algorithm.

	#EFMs					
Subset	Total	Reversible	Extended space	Biomass synthesis	mAb synthesis	Removed
1	24	4	28	2	1	4
2	22	3	25	3	1	3
3	22	4	26	2	1	5
4	22	3	25	2	1	1
5	24	2	26	3	1	1
6	23	3	26	3	1	3
7	24	3	27	1	1	3
8	25	3	28	2	1	4
9	24	4	28	3	1	4
10	24	3	27	3	1	1
11	22	3	25	2	1	2
12	23	3	26	3	1	2
13	25	3	28	3	1	0
14	25	4	29	2	1	2
15	22	3	25	2	1	3
16	25	4	29	3	1	2
Union	119	6	125	16	4	–

5.4. Model reduction

In LargeM, many values of $\mathbf{w}_{max}$ are very small indicating that only a fraction of the equations significantly contributed to the model. A model reduction is carried out by discarding the equations with $w_{max,j} < w_{cut}$, where w_{cut} is the cut-off level, such that the model is not significantly affected. Fig. 7 illustrates the selection of $w_{cut} = 0.1$, which corresponds to a reduced model including approximately 10% of the equations of LargeM. In the resulting reduced model RedM, the values of $\mathbf{w}_{max}$ are recalculated.

## 5.5. Identification of the final model

In cross-validation, part of the data (test set) is excluded from the model training and the remaining data (training set) are used to train the model. The ability of the model to predict the test set is then assessed (Hastie et al., 2001). In k-fold cross-validation, the data set is split into k folds and the cross-validation is repeated k times so that each fold is the test set once.

Here, the model LargeM$_{initial}$ is used as initial model to establish the macro-reaction network and the potential kinetics. We perform a k-fold cross-validation with $k = 4$. First, the whole data set is divided into four folds of nearly equal size such that, in each fold, each experimental condition is represented at similar proportion as in the whole data set.

Three of the folds (75% data) are used as training set d and the fourth fold as test set (25% data). Thus, both the training and test data are in the same experimental range. The model generation includes the estimation of $\mathbf{w}_{max}$ using the training set, resulting into LargeM^d, and a reduction of LargeM^d into RedM^d. RedM^d is then applied to the test data set and the sum of the relative errors between the simulated and the measured fluxes, error^d, is computed. The cross-validation is repeated systematically until each fold has been the test set once. This procedure is repeated 10 times; each time with random assignment of data into the folds and generation of a model RedM^d. The RedM^d corresponding to the smallest error^d is selected as the final model RedM$_{final}$. RedM$_{final}$ model includes 132 kinetics equations while the corresponding LargeM^d model includes 1227 kinetics equations before the reduction (see Table S14). Among the 125 macro-reactions of LargeM^d, 65 are finally retained in RedM$_{final}$ model. Each of the macro-reaction is associated with one or several kinetics, which have been automatically selected among all the potential kinetics based on the experimental data. The simulation obtained with RedM$_{final}$ model is visualized in Figs. 8 to 10.

RedM$_{final}$ model predicts the test data fairly well (Figs. 8 to 10). For most metabolites, the difference between the simulations of the training data and the testing data are barely distinguishable. They are also comparable for many of the metabolites and experimental conditions to the simulation obtained in the initial LargeM$_{initial}$ model (Figs. S5 to S7), where all the data are used for the model training and the model is not reduced.

The average relative and absolute errors for each medium and metabolite of simulation of the test data set obtained with RedM$_{final}$ are given in Tables S15 and S16 in the supplementary material. 72% of the relative errors are smaller than 10% and 85% smaller than 20% while 8% of the relative errors are higher than 50% (Table 15). Notice that for very small fluxes the relative error can be misleading due to division by a very small number however the absolute errors for these components are in the range of the other components as confirmed by the visual observation of the simulation in Figs. 8–10. From the relative errors and visual observation, the rates of the control (Ctrl, #16) are captured very well, as are many of the shifts triggered by varying the initial concentration of selected amino acids in the other experimental conditions (#1–15). The models in the present work and the model in our previous work (Hagrot et al., 2017) are based on the same set of experimental data, except that data of mAb production are included in the present models. As can be seen in Fig. 8 the present model simulated the mAb production rate well, capturing also the cell-specific rate increase observed for glutamine omission (Q0). For the simulation of EAAs uptake, major improvements in the simulation accuracy are achieved. The metabolic network used to build the poly-pathway in our previous proof-of-concept (Hagrot et al., 2017) lacked EAA catabolism due to the restriction imposed in the network size. In the present work, in which EAA catabolism is included in the model, the simulation
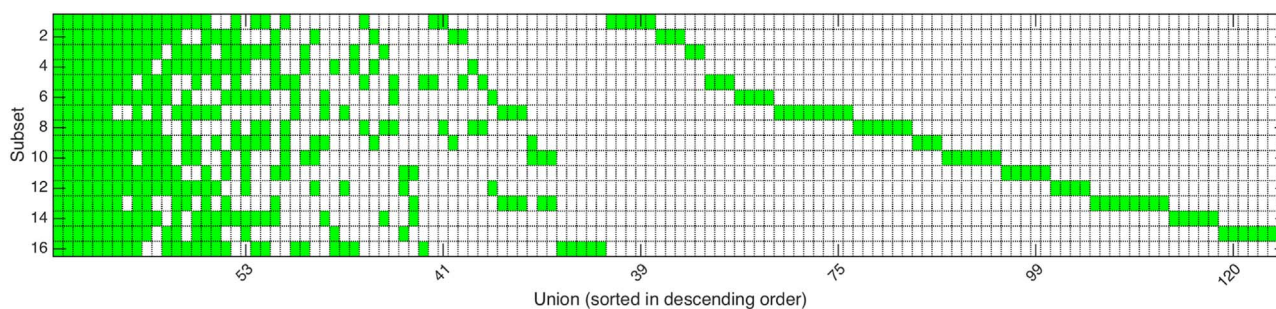
Fig. 6. Comparison between EFMs in subsets and union. A comparison between EFMs in the sixteen EFM subsets (y-axis) and in the union E_{union} (x-axis) with green color indicating that the EFM is found in both the EFM subset and the union. In the figure, the 125 unique EFMs in E_{union} have been sorted in descending order according to the number of matched EFM subsets, and occurs along the x-axis in the following order: 4, 8, 16, 18, 26, 27, 1, 3, 22, 30, 31, 2, 9, 17, 20, 28, 5, 15, 32, 53, 10, 29, 47, 36, 13, 19, 44, 65, 6, 11, 24, 33, 42, 50, 55, 97, 100, 25, 34, 41, 43, 46, 54, 56, 62, 66, 67, 69, 85, 89, 91, 7, 12, 14, 21, 23, 35, 37, 38, 39, 40, 45, 48, 49, 51, 52, 57, 58, 59, 60, 61, 63, 64, 68, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 86, 87, 88, 90, 92, 93, 94, 95, 96, 98, 99, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125.

Table 3

Overview of kinetic equations defined with starting point in the general kinetic Eq. (3) in Section 2.2. Kinetic equations based on substrate saturation effects are defined for each macro-reaction from the EFM union as defined in ①. In addition, alternative kinetic equations are formulated for macro-reactions that fulfilled certain criteria as defined in ②-⑦.

		Macro-reaction	*Note*
①	1a	$S \xrightarrow{\ S\ } P$	**P** does not include biomass or mAb.
	1b	$S \xrightarrow{\ EAAs\ Gln\ Cys\ } Biomass + P_2 + \ldots + P_p$	
	1c	$S \xrightarrow{\ EAAs\ Gln\ Cys\ } mAb + P_2 + \ldots + P_p$	
②	2a	$S \xrightarrow{\ EAAs\ Cys\ } Biomass + P_2 + \ldots + P_p$	
	2b	$S \xrightarrow{\ EAAs\ Cys\ } mAb + P_2 + \ldots + P_p$	
	2c	$S \xrightarrow{\ EAAs\ Cys\ Gln\ X\ } Biomass + P_2 + \ldots + P_p$	X is Cys, Ser, Gln, Asp, or Asn.
	2d	$S \xrightarrow{\ EAAs\ Cys\ X\ } mAb + P_2 + \ldots + P_p$	X is Glu, Gly, Asp, or Gln.
	2e	$S \xrightarrow{\ EAAs\ Cys\ Gln\ X\ } Biomass + P_2 + \ldots + P_p$	X is Ser, Gln, Asp, or Asn.
	2f	$S \xrightarrow{\ EAAs\ Cys\ X\ } mAb + P_2 + \ldots + P_p$	X is Glu, Gly, Asp, or Gln.
③	3a	$S \xrightarrow{\ S\ X\ } X + P_2 + \ldots + P_p$	X is Ala, Aln, Asp, Gln, Glu, Gly, or Ser.
	3b	$S \xrightarrow{\ S\ Y\ } X + P_2 + \ldots + P_p$	X/Y are Asp/Asn, Glu/Gln, Ala/Asn, Ala/Cys, Ser/Cys.
④	4a	$X + S_2 + \ldots + S_s \xrightarrow{\ S_2,\ldots,S_s\ } P$	X is Ala, Asp, Cys, Glu, or Gly.
	4b	$X + S_2 + \ldots + S_s \xrightarrow{\ S\ Y\ } P$	X/Y are Asp/Asn, Gln/Glu, or Gly/Ser.
	4c	$X + S_2 + \ldots + S_s \xrightarrow{\ S\ Y\ } P$	X/Y are Pro/Gln, Asn/Gln, Asp/Gln, Gly/Gln, Ile/Gln, Leu/Gln, Met/Gln, Phe/Gln, Lys/Gln, Val/Gln, Gly/Ser, Ser/Gln or Gln/Glu.
⑤	5	$Asn + S_2 + \ldots + S_s \xrightarrow{\ S_2,\ldots,S_s\ Asn\ } P$	
⑥	6a	$S \rightarrow X \xrightarrow{\ S\ X\ } P$	X is Asp, Ala, Asn, Gln, Glu, Gly, or Ser.
	6b	$S \rightarrow Pyr \xrightarrow{\ S\ X\ } P$	X is Cys, Gln, Gly or Ser.
⑦	7	$Glc \xrightarrow{\ Gln\ } Lac$	

S (or $S_1 + S_2 + \ldots + S_s$) are all the substrates of the macro-reaction.
P (or $P_1 + P_2 + \ldots + P_p$) are all the products of the macro-reaction.
X and *Y* represent metabolites as specified under *Note*.
EAAs: Ile, Leu, Met, Phe, Thr, Trp, Tyr, Val and Lys.
↓: saturation effect.
⊥: inhibition effect.

accuracy for the uptake of the EAAs isoleucine, leucine, phenylalanine, threonine, tryptophan, tyrosine, valine and methionine is greatly improved (Fig. 10). The RedM$_{final}$ poly-pathway model is also able to capture the increased specific uptake rate for several of those EAAs triggered by glutamine omission (Q0). The model fit is degraded for a couple of cases such as the simulation of alanine in A0 medium and glutamate in Q200, which will be worth further investigation in future work.
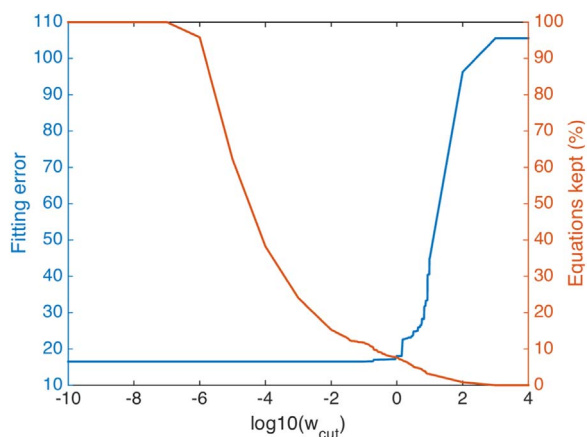
Fig. 7. The x-axis shows the logarithm (base 10) of the cut-off level w_{cut}. The *left* y-axis shows the error (2-norm) indicating the fit between model and data. The *right* y-axis shows the percentage of equations in LargeM$_{initial}$ that are kept. $w_{cut} = 0.1$ is chosen for the model reduction ($log10(0.1) = -1$).

5.6. Simulation of the flux distributions

The poly-pathway model can be translated into a flux distribution over the network of the 126 metabolic reactions presented in Fig. 2. These distributions provide a simulation of the intracellular reaction fluxes underlying the uptake and secretion rates. To illustrate this, the simulated flux distribution of the control medium and the variations triggered by depleting asparagine, glutamine, and serine are detailed

below in comparison with the control.

5.6.1. Control (Ctrl)

The flux distribution in the control medium is shown in Fig. 11. High fluxes over the glycolysis pathway reactions v_1-v_8 leads to secretion of lactate, which is typical for the CHO cell metabolism (Gódia and Cairo, 2006). Part of the G6P enters the PPP (v_9). While a small flux continues over v_{67}, contributing to nucleotide synthesis, the remainder is returned back to the glycolysis via v_{11}-v_{13}. Based on isotope tracer measurements in a CHO cell perfusion culture experiment (Goudar et al., 2010), the flux from G6P towards the PPP is 41% of the consumed glucose; this value is 22% for the control in our model. In our case, the flux from G6P continuing in the glycolysis versus the PPP can not be determined from our extracellular metabolite data. Even though carbon dioxide is produced in the first step of the PPP (v_9), we do not measure extracellular carbon dioxide and its secretion rate is free to vary within the intervals specified by the bounds in Table S7.

In line with typical CHO cell metabolism (Gódia and Cairo, 2006), only a small fraction of the pyruvate generated in the glycolysis enters the mitochondrion via v_{91}, supplying AcCoA for the TCA-cycle (here the flux over v_{14} corresponds to 19% of the glucose uptake rate which can be compared to 83% in the study by Goudar et al., 2010). Additional mitochondrial AcCoA is supplied via the EAA catabolism. AcCoA is required for the lipid synthesis, however, as it cannot cross the membrane it must be transported as citrate (Martens, 2007), leaving the cycle such that AcCoA can be regenerated in the cytosol (v_{22}). Based on their isotope tracer measurements, Goudar et al. (2010) estimated that 24% of the TCA-cycle flux (v_{15}) is channeled to lipid synthesis, a number which is very close to 22% simulated in the present model.
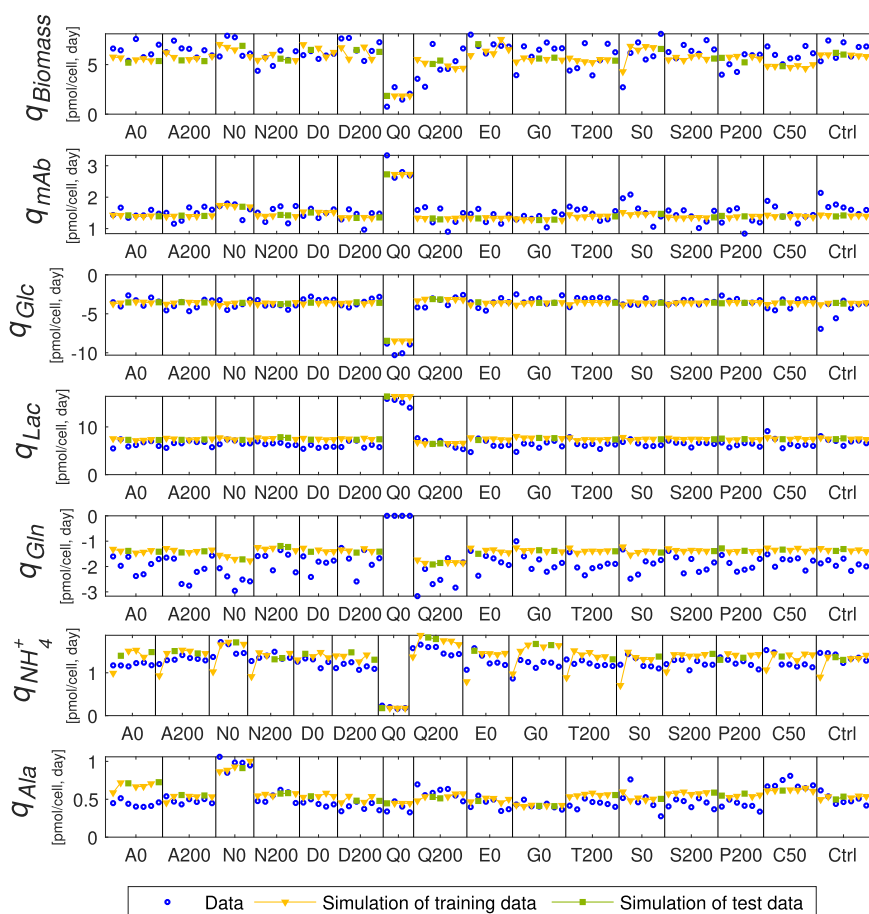
Fig. 8. Comparison of the simulated fluxes of the training set (yellow triangle) and the test set (green squares) with the experimental data (blue circle) of metabolic rates in pmol/cell, day (y-axis), for biomass, mAb, glucose, lactate, glutamine, ammonia and alanine. For each experimental condition (A0, A200, N0, ..., Ctrl), the data and simulated values are shown along the x-axis in ascending order according to time (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).
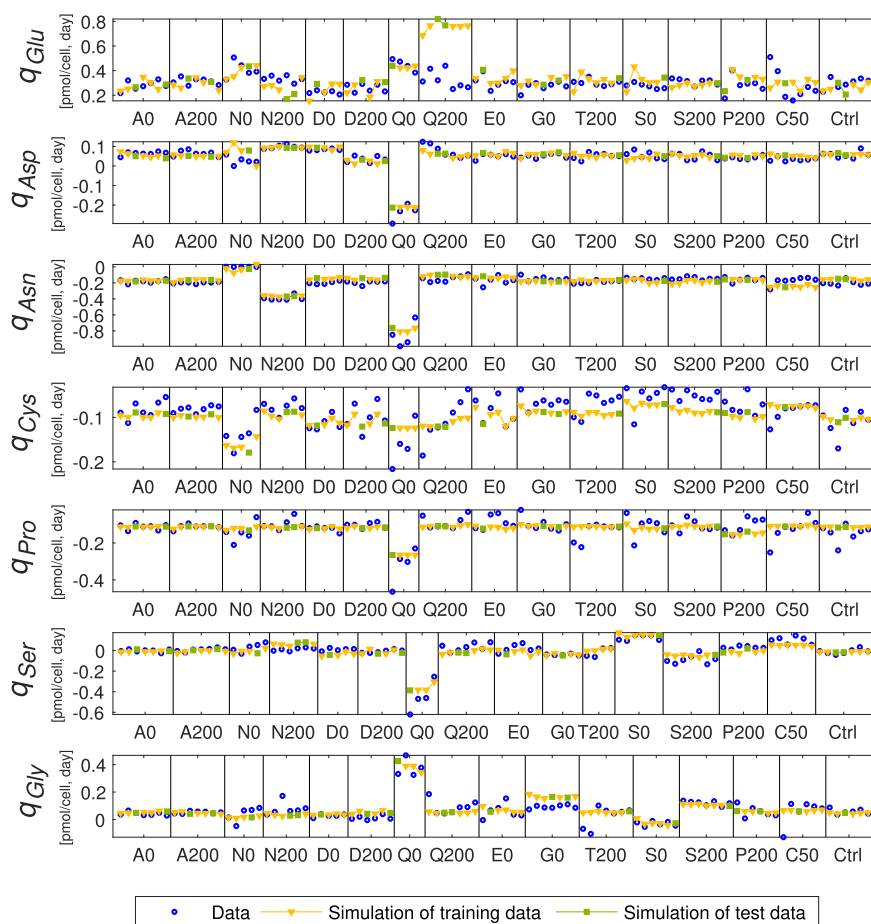
Fig. 9. Comparison of the simulated fluxes of the training set (yellow triangle) and the test set (green squares) with the experimental data (blue circle) of metabolic rates in pmol/cell, day (y-axis), for glutamate, aspartate, asparagine, cysteine, proline, serine and glycine. For each experimental condition (A0, A200, N0, …, Ctrl), the data and simulated values are shown along the x-axis in ascending order according to time (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

Generally in mammalian cell lines, the flux in the TCA-cycle is replenished at aKG, which is supported via glutamine uptake (Gódia and Cairo, 2006). As can be seen in the present simulation (Fig. 11), the conversion of glutamine into glutamate into aKG results from combined fluxes over the NEAA metabolism and mitochondrial transport reactions. Overall, NEAA metabolism generates aKG at 2 pmol/cell, day, while the TCA-cycle consumes aKG at 1.7 pmol/cell, day. EAA catabolism feeds additional succinate into the cycle. Typically, in industrial cell lines, the resulting TCA overflow is handled via the malate shunt: the export of malate out of the mitochondrion with subsequent conversion into pyruvate (Gódia and Cairo, 2006). In our network, malate and oxaloacetate are interconvertible (v_{21} and v_{25}), as are PEP and pyruvate (v_7). In our simulation, there is a net generation of malate (at 0.27 pmol/cell, day) and oxaloacetate (at 2.7 pmol/cell, day) in the TCA-cycle. Overall, the anaplerotic (v_{26}-v_{29}) convert oxaloacetate and malate (at 1.6 and 0.27 pmol/cell, day) into PEP and pyruvate (at 1.7 and 0.2 pmol/cell, day).

5.6.2. Glutamine omission (Q0)

In case of glutamine omission (Q0), the glutamine transport v_{112} shifts towards secretion at a small rate compared to uptake in the control, and the fluxes are altered over the majority of the network reactions. The biomass synthesis (v_{90}) and the reactions of the nucleotide and lipid syntheses are greatly reduced, while the mAb synthesis (v_{87}) is increased. This is in accordance with the decreased specific growth rate and increased specific mAb productivity experimentally observed here. The uptake of glucose v_1 and fluxes over the glycolysis and PPP are greatly increased, leading to increased generation of pyruvate in v_7 and v_{40} and lactate generation/secretion. An increased

amount of pyruvate enters the mitochondrion via v_{91}, leading to increased aKG$_m$ and aKG generation by v_{16} and v_{23}. All the fluxes of the TCA-cycle are increased (with exception for v_{22} that supplies AcCoA for lipid synthesis) and there is a net outflow of malate from the mitochondrion. Consistent with experimental observation, the model simulates increased uptake of asparagine (v_{111}), as well as increases of asparagine degradation (v_{34}), asparagine synthesis (v_{35}), and glutamine synthesis (v_{36}). Thereby, part of the glutamine synthesized in v_{36} is converted to asparagine via v_{35}. Reaction v_{103} shifts direction from aspartate secretion towards uptake. Overall, the net glutamate generation of the NEAA metabolism in the control shifts to consumption in Q0. Despite the increased fluxes over some of the ammonium-generating reactions (e.g. v_{40} and v_{34}), in total, less ammonium is generated as significant amount is used in the generation of glutamate and glutamine by v_{36} and v_{37}. A substantial amount of serine is generated and consumed via v_{30} and v_{40}, such that glutamate is converted into aKG. Additional serine is supplied via an increased uptake (v_{106}). Serine is used in methionine catabolism (v_{60}) and converted into glycine (v_{41}), which is also generated in v_{42} and secreted (v_{105}). Overall, the combined activity of the EAA catabolism generates significantly higher amounts of glutamate, ammonium and pyruvate, and consumes an increased amount of aKG compared to the control. Increased amounts of SucCoA$_m$, Suc$_m$ and AcCoA$_m$ are generated for entry into the TCA-cycle. The fluxes in the urea cycle (v_{64} − v_{66}) as well as uptake and catabolism of arginine (v_{110} and v_{62}) are all increased, leading to an increased fixation of ammonium, urea secretion (v_{126}), and the generation of glutamate from ornithine via v_{43} and v_{45}.

Wahrheit et al. (2014) have studied the effects of glutamine availability during CHO cell batch cultivations using dynamic MFA.
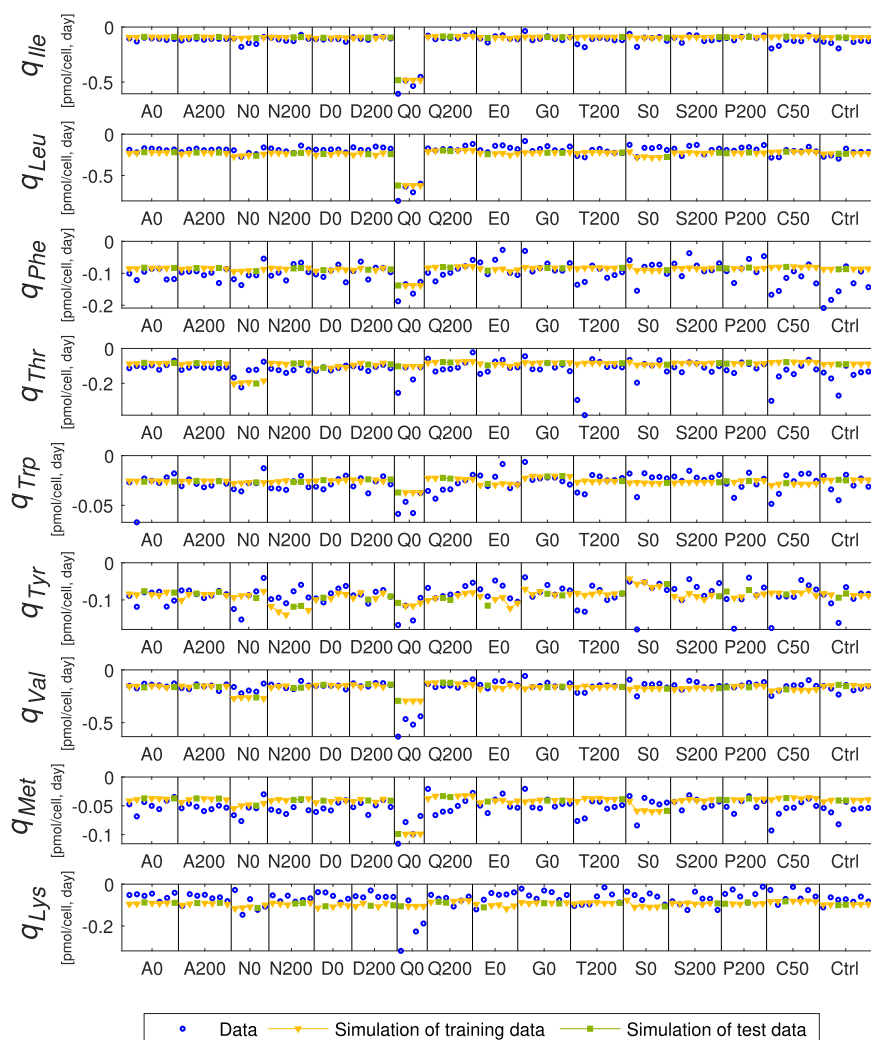
Fig. 10. Comparison of the simulated fluxes of the training set (yellow triangle) and the test set (green squares) with the experimental data (blue circle) of metabolic rates in pmol/cell, day (y-axis), for isoleucine, leucine, phenylalanine, threonine, tryptophan, tyrosine, valine, methionine and lysine. For each experimental condition (A0, A200, N0, ..., Ctrl), the data and simulated values are shown along the x-axis in ascending order according to time (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article).

The conditions during the early stages of the glutamine free cultures (Wahrheit et al., 2014) can be compared to Q0 in the present work. Consistent with the present experimental results, the glutamine free cultivations reported in Wahrheit et al. (2014) are characterized by an initial slight production of glutamine, low production of ammonium, high production of glutamate and glycine and high consumption of serine. Increased asparagine uptake and shift towards aspartate uptake is not initially observed by Wahrheit et al. (2014), yet occurs at a later stage of the culture. In particular, the reversal of the glutamate dehydrogenase reaction to provide glutamate for the glutamine synthesis (Wahrheit et al., 2014) corresponds with the increased flux over glutamine-generating v_{36} simulated in the present work. Similar to the present simulation, Linz et al. (1997) observed an enhanced consumption of the other amino acids in a BHK cell line cultured at low glutamine concentration.

5.6.3. Asparagine omission (N0)

In case of asparagine omission (N0), the simulated asparagine uptake (v_{111}) is almost null. The model simulates both reduced asparagine degradation (v_{34}) and asparagine synthesis (v_{35}), resulting in a shift in the NEAA metabolism from net asparagine consumption to net asparagine generation, compared to the control. Furthermore, the NEAA metabolism consumes more glutamine, and generates increased amounts of glutamate and aKG. The increase in alanine secretion

experimentally observed can be traced to the increased flux over v_{30}, for which additional pyruvate is primarily obtained from increased anaplerotic flux (v_{29}) and cysteine catabolism (v_{61}). In the EAA metabolism, an increased uptake and catabolism of arginine generates additional glutamate from ornithine.

5.6.4. Serine omission (S0)

The serine omission triggers direction shifts for the transports of serine (v_{106}) and glycine, from small serine uptake to its secretion and from glycine secretion to its uptake (v_{105}). The conversion of serine into glycine (v_{41}) is decreased, leaving additional serine available for secretion. In addition to glycine taken up from the medium, additional glycine is now available compared to the control due to a decreased consumption (v_{42}).

6. Conclusions and perspectives

In this work, a kinetic model based on a network of 126 metabolic reactions is created. Using our novel column generation-based method, subsets of EFMs relevant to the experimental data are identified instead of systematically including all the EFMs. By including flexible kinetics, the resulting poly-pathway model can describe multiple metabolic states. It is usual that certain fluxes cannot be resolved from extracellular data alone (Martens, 2007), e.g. the fluxes from G6P continuing in the glycolysis or the pentose phosphate pathway (PPP),
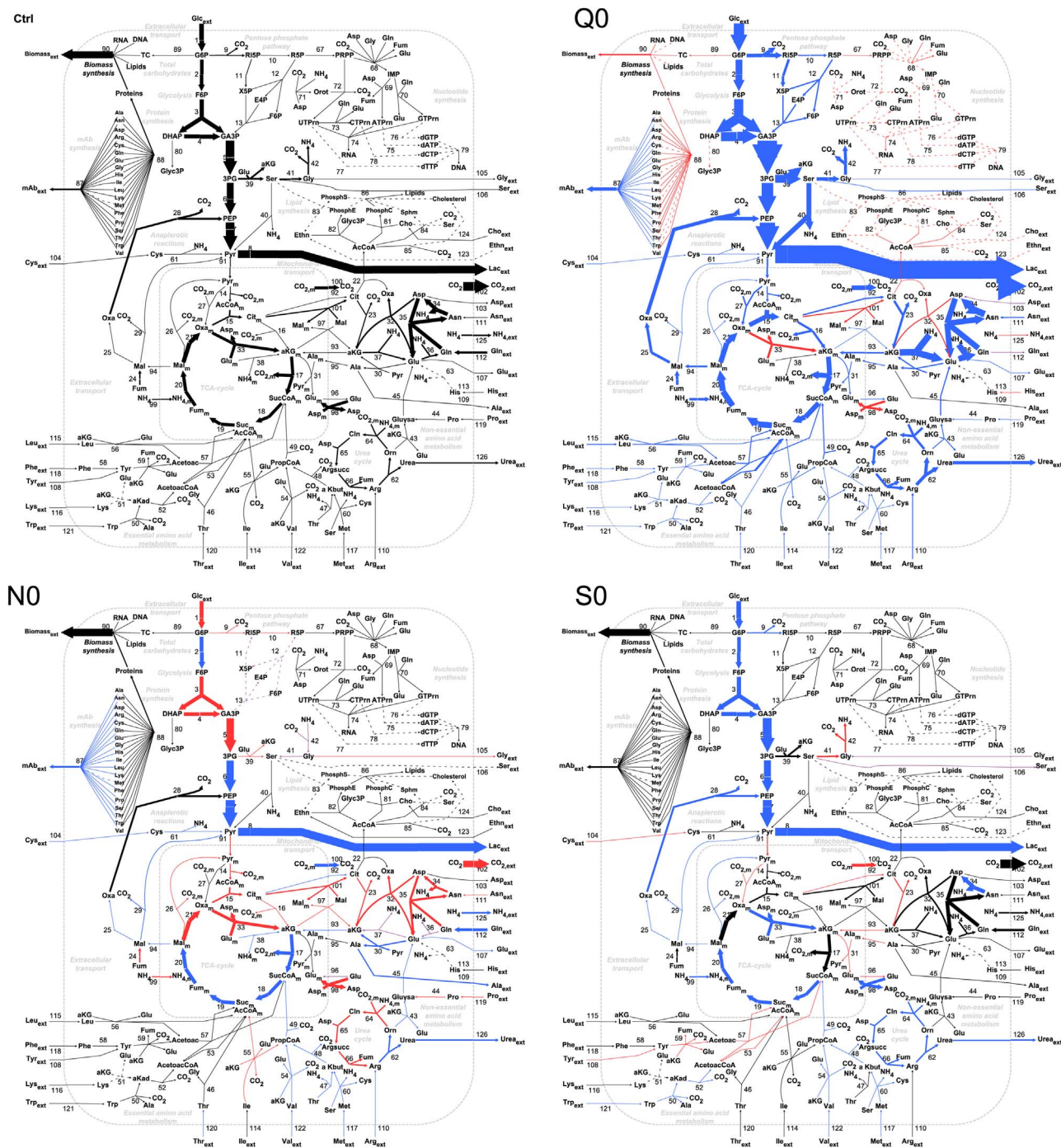
Fig. 11. Flux distributions over the metabolic network in the control medium, subjected to glutamine omission (Q0), asparagine omission (Q0) and serine omission (S0). The thickness of the arrows is proportional to the reaction flux. Dashed lines represent fluxes for which $|v_j| < 0.0025$. Fluxes that are significantly different compared to the control are colored in the following way: blue indicates an increase, red indicates a decrease, and purple indicates a shift in the direction of the flux. The flux visualizations are created with VANTED (Rohn et al., 2012b) and FluxMap (Rohn et al., 2012a).

fluxes over parallel reactions of the TCA-cycle, the anaplerotic reactions or some reversible reactions of amino acid metabolism cannot be resolved.

The integration of the CG algorithm adds a new dimension to the poly-pathway modelling framework previously presented in Hagrot et al. (2017). It enables here the use of a much more complex and comprehensive metabolic network than what has been typically used for kinetic models of mammalian cell cultures (Provost and Bastin, 2004; Provost et al., 2005; Gao et al., 2007; Dorka et al., 2009; Naderi et al., 2011; Ghorbaniaghdam et al., 2013, 2014; Hagrot et al., 2017; Nolan and Lee, 2011; Robitaille et al., 2015). In particular, the models

based on macroscopic reactions (e.g. EFMs) have been limited to simple networks of approximately 40 reactions or less (Provost and Bastin, 2004; Provost et al., 2005; Gao et al., 2007; Dorka et al., 2009; Naderi et al., 2011; Hagrot et al., 2017), as the enumeration of EFMs in more complex networks leads to a very high number of EFMs and becomes computationally prohibitive. The CG algorithm avoids a systematic EFM enumeration and provides a set of EFMs relevant to the information of the experimental data while the properties of the original network are preserved.

Compared to the network of 34 reactions in our previous work (Hagrot et al., 2017), the present network of 126 reactions includes

additional reactions, metabolites, and intracellular pools, as well as compartmentalization and product formation. A major improvement compared to Hagrot et al. (2017) is observed in the ability of the model to simulate essential amino acid uptake, resulting from the inclusion of EAA catabolism in the metabolic network. Despite the increased network complexity, our CG algorithm can efficiently generate reduced sets of EFMs that optimally fit the data of each experimental condition. Interestingly, even though the network is much more complex, the combined set of EFMs used to build the poly-pathway model is smaller (125 EFMs) than the EFM set obtained by systematic enumeration (379 EFMs) in our previous work (Hagrot et al., 2017) using the well-established Metatool method.

We formulate flexible and diverse macro-reaction kinetics that are applied in a systematic manner, including *a priori* information and/or data-driven kinetics selection, that are identified via parameter estimation. The generation of EFMs by CG is very rapid. For each medium, EFMs are obtained by CG in a few seconds using non-optimized Matlab2018a code (on a Macbook Pro computer). The generation of RedM$_{final}$ model in the present cross-validation of ten 4-fold repetitions, which includes the generation of a kinetic model forty times, takes 13 min (using non-optimized code).

The resulting poly-pathway model accurately simulates the different metabolic states observed in the CHO cell cultures under varied amino acid availability, demonstrating the ability of this approach to capture diverse metabolic behavior in one single model. The method supports as well the inclusion of bound values of parameters/components for which information is scarce, uncertain or taken from an external source, e.g. from the literature. This allows a better identification of the model parameters and offers a possibility to use larger models since the measurements can potentially be replaced by intervals. Furthermore, using the poly-pathway model, the influence of amino acid availability on the experimentally observed metabolic behavior can be linked to varied use of certain reactions in the internal structure of the model network, information which is important to optimize the cell metabolism, culture media or processes.

For future modelling applications requiring predictive power, the kinetic parameters will become increasingly important, e.g. for process optimization or feedback control. Parameter estimation in large non-linear systems is a challenging task, requiring specialized mathematical tools and strategies or alternative kinetics models such as for instance ensemble modelling (Tran et al., 2008). In the present approach, CG-based selection of the EFMs of a poly-pathway kinetic model constituted a first step. In future work, the kinetics modelling could be further developed, e.g. the determination of the number of kinetics and the selection of their parameters could be automated and/or guided by sensitivity analysis and more advanced parameter estimation strategies (Ben Yahia et al., 2015). Future development could also benefit from a simultaneous EFM identification and parameter estimation. Due to the non-linearity of the kinetic equations, this is not a straightforward implementation. A heuristic approach is presented by Oddsdóttir (2015) and will be implemented in the future.

The information of the intracellular components could be beneficial for the development and validation of the model with the present approach. The segregation of extracellular vs. intracellular is motivated by the biology. From a mathematical point of view, in the present work, this relates to the fact that extracellular components are measured while the intracellular components are not (and they are mathematically eliminated during the EFM exercise). Therefore our model frame allows the inclusion of measured intracellular component fluxes and concentrations, and this could be used in future work.

The main goal of the present work is to show a new way to use EFMs in a kinetic model thanks to the novel CG-based method. In the present work, the modelling approach takes into account the internal structure of the metabolism given by its stoichiometric reaction network to identify the relevant metabolic pathways and fluxes taking place in the experiments. The present model is not able to describe pathways which are not included in the reaction network and do not occur in the data set. To include a given behavior in the model, it is necessary that experimental data corresponding to this behavior (potentially available as interval range) are provided for the model identification. The model presented herein has been developed for scenarios of varied amino acid availability. To capture metabolic behavior in response to other environmental perturbations than the ones studied here will require relevant experimental data and a reconsideration of the network and kinetic design will be necessary. Furthermore the model does not include more detailed regulatory mechanisms from information e.g. of the transcriptome or proteome, which could be of interest. The algorithm accepts metabolic networks of varying size and complexity, from simplified descriptions (Oddsdóttir et al., 2014) up to genome-scale (in preparation), opening up for applications in the omics field. For this type of application, it will be relevant to impose objective such as those used for microorganism culture, e.g. the maximization of growth rate or the optimization of the complexity in their metabolism to achieve maximum growth rate (Groot et al., 2017). To give a perspective from the microorganism field, the omics information has recently been included in different approaches to model these organisms such as incorporating the transcriptions factors, post-translational regulation and enzyme level regulation mechanisms (Millard et al., 2017); generating kinetics based on qualitative knowledge in the proteome (Erickson et al., 2017); or integrating the regulatory mechanisms relevant to the modeled cell type into the network and/or kinetic design underlying regulatory network (Chubukov et al., 2014). These could potentially be accounted for in future work for kinetic models of animal cells however the high complexity of these latter will represent a challenge.

Finally, the CG algorithm is not limited to the construction of kinetic EFM models or models of mammalian cell cultures; it could be applied for other applications and to other organisms for which the identification of an EFM subset is desired.

Research data availability

Datasets generated during the current study are available in the Mendeley Data repository.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.mec.2018.e00083

A file with supplementary material is available containing supporting information about the metabolic network and bounds, the EFM union, the kinetic equations, the poly-pathway model, simulations by LargeM$_{initial}$, the co-factors.

References

Almquist, J., Cvijovic, M., Hatzimanikatis, V., Nielsen, J., Jirstrand, M., 2014. Kinetic models in industrial biotechnology – improving cell factory performance. Metab. Eng. 24, 38–60. http://dx.doi.org/10.1016/j.ymben.2014.03.007.

Ben Yahia, B., Malphettes, L., Heinzle, E., 2015. Macroscopic modeling of mammalian cell growth and metabolism. Appl. Microbiol. Biot. 99, 7009–7024. http://dx.doi.org/10.1007/s00253-015-6743-6.

Chubukov, V., Gerosa, L., Kochanowski, K., Sauer, U., 2014. Coordination of microbial metabolism. Nat. Rev. Microbiol. 12, 327–340. http://dx.doi.org/10.1038/nrmicro3238, (URL ⟨http://dx.doi.org/10.1038/nrmicro3238⟩, http://arXiv:1011.1669v3arXiv:1011.1669v3).

Dantzig, G.B., Wolfe, P., 1960. Decomposition principle for linear programs. Oper. Res. 8, 101–111, (URL ⟨https://www.jstor.org/stable/167547?seq=1#page_scan_tab_contents⟩).

Desrosiers, J., Soumis, F., Desrochers, M., 1984. Routing with time windows by column generation. Networks 14, 545–565. http://dx.doi.org/10.1002/net.3230140406, (URL ⟨http://doi.wiley.com/10.1002/net.3230140406⟩).

Dorka, P., Fischer, C., Budman, H., Scharer, J.M., 2009. Metabolic flux-based modeling of mAb production during batch and fed-batch operations. Bioprocess Biosyst. Eng. 32, 183–196. http://dx.doi.org/10.1007/s00449-008-0236-2, (URL ⟨http://www.ncbi.nlm.nih.gov/pubmed/18560901⟩).

Erickson, D.W., Schink, S.J., Patsalo, V., Williamson, J.R., Gerland, U., Hwa, T., 2017. A global resource allocation strategy governs growth transition kinetics of Escherichia coli. Nature 551, 119–123. http://dx.doi.org/10.1038/nature24299, (URL ⟨http://dx.doi.org/10.1038/nature24299⟩, http://arXiv:NIHMS150003arXiv:NIHMS150003.).

de Figueiredo, L.F., Podhorski, A., Rubio, A., Kaleta, C., Beasley, J.E., Schuster, S., Planes, F.J., 2009. Computing the shortest elementary flux modes in genome-scale metabolic networks. Bioinformatics 25, 3158–3165. http://dx.doi.org/10.1093/bioinformatics/btp564.

Ford, L.R., Fulkerson, D.R., 1958. A suggested computation for maximal multi-commodity network flows. Manag. Sci. 5, 97–101, (URL ⟨https://www.jstor.org/stable/2626975?seq=1#page_scan_tab_contents⟩).

Gao, J., Gorenflo, V.M., Scharer, J.M., Budman, H.M., 2007. Dynamic metabolic modeling for a MAB bioprocess. Biotechnol. Prog. 23, 168–181. http://dx.doi.org/10.1021/bp060089y, (URL ⟨http://www.ncbi.nlm.nih.gov/pubmed/17269685⟩).

Ghorbaniaghdam, A., Henry, O., Jolicoeur, M., 2013. A kinetic-metabolic model based on cell energetic state: study of CHO cell behavior under Na-butyrate stimulation. Bioprocess Biosyst. Eng. 36, 469–487. http://dx.doi.org/10.1007/s00449-012-0804-3, (URL ⟨http://www.ncbi.nlm.nih.gov/pubmed/22976819⟩).

Ghorbaniaghdam, A., Chen, J., Henry, O., Jolicoeur, M., 2014. Analyzing clonal variation of monoclonal antibody-producing CHO cell lines using an in silico metabolomic platform. PloS One 9, e90832. http://dx.doi.org/10.1371/journal.pone.0090832, (URL ⟨http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3954614&tool=pmcentrez&rendertype=abstract⟩).

Gilmore, P.C., Gomory, R.E., 1961. A linear programming approach to the cutting-stock problem. Oper. Res. 9, 849–859.

Gilmore, P.C., Gomory, R.E., 1963. A linear programming approach to the cutting stock problem-Part II. Oper. Res. 11, 863–888.

GLPK, GLPK (GNU Linear Programming Kit). URL ⟨http://www.gnu.org/software/glpk/glpk.html⟩.

Gódia, F., Cairo, J.J., 2006. Cell Metabolism. In: Ozturk, S., Hu, W.S. (Eds.), Cell Culture Technology for Pharmaceutical and Cell-Based Therapies. CRC Press Taylor & Francis Group, Boca Raton, FL, pp. 81–112.

Goudar, C., Biener, R., Boisart, C., Heidemann, R., Piret, J., de Graaf, A., Konstantinov, K., 2010. Metabolic flux analysis of CHO cells in perfusion culture by metabolite balancing and 2D [13C, 1H] COSY NMR spectroscopy. Metab. Eng. 12, 138–149. http://dx.doi.org/10.1016/j.ymben.2009.10.007, (URL ⟨http://www.ncbi.nlm.nih.gov/pubmed/19896555⟩).

Groot, D.H.D., Planqu, R., Boxtel, C.V., Bruggeman, F.J., Teusink, B., 2017. Maximal growth rate requires minimal metabolic complexity. bioRxiv. doi:10.1101/167171.

Hagrot, E., Oddsdóttir, H.Æ., Hosta, J.G., Jacobsen, E.W., Chotteau, V., 2017. Poly-pathway model, a novel approach to simulate multiple metabolic states by reaction network-based model - Application to amino acid depletion in CHO cell culture. J. Biotechnol. 259, 235–247. http://dx.doi.org/10.1016/j.jbiotec.2017.05.026, (URL ⟨http://dx.doi.org/10.1016/j.jbiotec.2017.05.026⟩).

Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning. Springer.

Jungers, R.M., Zamorano, F., Blondel, V.D., Vande Wouwer, A., Bastin, G., 2011. Fast computation of minimal elementary decompositions of metabolic flux vectors. Automatica 47, 1255–1259. http://dx.doi.org/10.1016/j.automatica.2011.01.011, (URL ⟨http://linkinghub.elsevier.com/retrieve/pii/S0005109811000264⟩).

Kaleta, C., de Figueiredo, L.F., Schuster, S., 2009. Can the whole be less than the sum of its parts? Pathway analysis in genome-scale metabolic networks using elementary flux patterns. Genome Res., 1872–1883. http://dx.doi.org/10.1101/gr.090639.108.

von Kamp, A., Schuster, S., 2006. Metatool 5.0: fast and flexible elementary modes

analysis. Bioinformatics 22, 1930–1931. http://dx.doi.org/10.1093/bioinformatics/btl267, (URL ⟨http://www.ncbi.nlm.nih.gov/pubmed/16731697⟩).

Klamt, S., Stelling, J., 2002. Combinatorial complexity of pathway analysis in metabolic networks. Mol. Biol. Rep. 29, 233–236. http://dx.doi.org/10.1023/A:1020390132244, (URL ⟨http://europepmc.org/abstract/MED/12241063⟩).

Klamt, S., Stelling, J., 2003. Two approaches for metabolic pathway analysis? Trends Biotechnol. 21, 64–69, (URL ⟨http://www.ncbi.nlm.nih.gov/pubmed/12573854⟩).

Klamt, S., Saez-Rodriguez, J., Gilles, E.D., 2007. Structural and functional analysis of cellular networks with CellNetAnalyzer. BMC Syst. Biol. 1, 2.

Linz, M., Zeng, A.P., Wagner, R., Deckwer, W.D., 1997. Stoichiometry, kinetics, and regulation of glucose and amino acid metabolism of a recombinant BHK cell line in batch and continuous cultures. Biotechnol. Prog. 13, 453–463.

Llaneras, F., Picó, J., 2010. Which metabolic pathways generate and characterize the flux space? A comparison among elementary modes, extreme pathways and minimal generators. J. Biomed. Biotechnol. 2010, 753904. http://dx.doi.org/10.1155/2010/753904, (URL ⟨http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2868190&tool=pmcentrez&rendertype=abstract⟩).

Lübbecke, M.E., Desrosiers, J., 2005. Selected topics in column generation. Oper. Res. 53, 1007–1023. http://dx.doi.org/10.1287/opre, (URL ⟨http://www.jstor.org/stable/25146936⟩⟨http://www.jstor.org/stable/25146936?seq=1&cid=pdf-reference#references_tab_contents⟩⟨http://about.jstor.org/terms⟩).

Machado, D., Soons, Z., Patil, K.R., Ferreira, E.C., Rocha, I., 2012. Random sampling of elementary flux modes in large-scale metabolic networks. Bioinformatics, 28. http://dx.doi.org/10.1093/bioinformatics/bts401.

Martens, D.E., 2007. Metabolic flux analysis of mammalian cells. In: Al-Rubeai, M., Fussenegger, M. (Eds.), Systems Biology. Springer, 275–299, (chapter 9).

Millard, P., Smallbone, K., Mendes, P., 2017. Metabolic regulation is sufficient for global and robust coordination of glucose uptake , catabolism , energy production and growth in Escherichia coli. PLoS Comput. Biol., 1–24. http://dx.doi.org/10.1371/journal.pcbi.1005396.

Montgomery, D.C., Peck, E.A., Vining, G.G., 2015. Introduction to Linear Regression Analysis. John Wiley & Sons.

Naderi, S., Meshram, M., Wei, C., McConkey, B., Ingalls, B., Budman, H., Scharer, J., 2011. Development of a mathematical model for evaluating the dynamics of normal and apoptotic Chinese hamster ovary cells. Biotechnol. Prog. 27, 1197–1205. http://dx.doi.org/10.1002/btpr.647, (URL ⟨http://www.ncbi.nlm.nih.gov/pubmed/21618458⟩).

Niu, H., Amribt, Z., Fickers, P., Tan, W., Bogaerts, P., 2013. Metabolic pathway analysis and reduction for mammalian cell cultures—towards macroscopic modeling. Chem. Eng. Sci. 102, 461–473. http://dx.doi.org/10.1016/j.ces.2013.07.034, (URL ⟨http://dx.doi.org/10.1016/j.ces.2013.07.034⟩).

Nolan, R.P., Lee, K., 2011. Dynamic model of CHO cell metabolism. Metab. Eng. 13, 108–124. http://dx.doi.org/10.1016/j.ymben.2010.09.003, (URL ⟨http://www.ncbi.nlm.nih.gov/pubmed/20933095⟩).

Oddsdóttir, H.Æ., 2015. Macroscopic Modeling of Metabolic Reaction Networks and Dynamic Identification of Elementary Flux Modes by Column Generatio (Ph.D. thesis). KTH Royal Institute of Technology. http://dx.doi.org/10.1017/CBO9781107415324.004, arXiv:arXiv:1011.1669v3.

Oddsdóttir, H.Æ., Hagrot, E., Chotteau, V., Forsgren, A., 2014. On dynamically generating relevant elementary flux modes in a metabolic network using optimization. J. Math. Biol. http://dx.doi.org/10.1007/s00285-014-0844-1, (URL ⟨http://link.springer.com/10.1007/s00285-014-0844-1⟩).

Oddsdóttir, H.Æ., Hagrot, E., Chotteau, V., Forsgren, A., 2016. Robustness analysis of elementary flux modes generated by column generation. Math. Biosci. 273, 45–56. http://dx.doi.org/10.1016/j.mbs.2015.12.009.

Papin, J.A., Stelling, J., Price, N.D., Klamt, S., Schuster, S., Palsson, B.O., 2004. Comparison of network-based pathway analysis methods. Trends Biotechnol. 22, 400–405. http://dx.doi.org/10.1016/j.tibtech.2004.06.010.

Provost, A., 2006. Metabolic Design of Dynamic Bioreaction Models (Ph.D. thesis). Université Catholique de Louvain.

Provost, A., Bastin, G., 2004. Dynamic metabolic modelling under the balanced growth condition. J. Process Contr. 14, 717–728. http://dx.doi.org/10.1016/j.jprocont.2003.12.004, (URL ⟨http://linkinghub.elsevier.com/retrieve/pii/S0959152403001392⟩).

Provost, A., Bastin, G., Agathos, S., Schneider, Y.J., 2005. Metabolic design of macroscopic models: application to CHO cells. In: Proceedings of the 44th IEEE Conference on Decision and Control, pp. 2982–2989 URL ⟨http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?Arnumber=1582618⟩, doi:10.1109/CDC.2005.1582618.

Quek, L.E., Dietmair, S., Krömer, J.O., Nielsen, L.K., 2010. Metabolic flux analysis in mammalian cell culture. Metab. Eng. 12, 161–171. http://dx.doi.org/10.1016/j.ymben.2009.09.002, (URL ⟨http://www.ncbi.nlm.nih.gov/pubmed/19833223⟩).

Robitaille, J., Chen, J., Jolicoeur, M., 2015. A Single Dynamic Metabolic Model Can Describe mAb Producing CHO Cell Batch and Fed-Batch Cultures on Different Culture Media. PLoS ONE 10, e0136815. http://dx.doi.org/10.1371/journal.pone.0136815, (URL ⟨http://dx.plos.org/10.1371/journal.pone.0136815⟩).

Rohn, H., Hartmann, A., Junker, A., Junker, B.H., Schreiber, F., 2012a. FluxMap: a VANTED add-on for the visual exploration of flux distributions in biological networks. BMC Syst. Biol. 6, 33. http://dx.doi.org/10.1186/1752-0509-6-33.

Rohn, H., Junker, A., Hartmann, A., Grafahrend-Belau, E., Treutler, H., Klapperstück, M., Czauderna, T., Klukas, C., Schreiber, F., 2012b. VANTED v2: a framework for systems biology applications. BMC Syst. Biol. 6, 139. http://dx.doi.org/10.1186/1752-0509-6-139.

Schuster, S., Hilgetag, C., 1994. On Elementary Flux Modes in Biochemical Reaction Systems at Steady State. J. Biol. Syst. 2, 165–182.

Schwarz, R., Liang, C., Kaleta, C., Kühnel, M., Hoffmann, E., Kuznetsov, S., Hecker, M.,

Griffiths, G., Schuster, S., Dandekar, T., 2007. Integrated network reconstruction, visualization and analysis using YANAsquare. BMC Bioinf. doi:10.1186/1471-2105-8-313.

Tabe-Bordbar, S., Marashi, S.A., 2013. Finding elementary flux modes in metabolic networks based on flux balance analysis and flux coupling analysis: application to the analysis of Escherichia coli metabolism. Biotechnol. Lett. 35, 2039–2044. http://dx.doi.org/10.1007/s10529-013-1328-x.

Tran, L.M., Rizk, M.L., Liao, J.C., 2008. Ensemble Modeling of Metabolic Networks. Biophys. J. 95, 5606–5617. http://dx.doi.org/10.1529/biophysj.108.135442 (URL ⟨http://dx.doi.org/10.1529/biophysj.108.135442⟩).

Wahrheit, J., Nicolae, A., Heinzle, E., 2014. Dynamics of growth and metabolism controlled by glutamine availability in Chinese hamster ovary cells. Appl. Microbiol. Biotechnol. 98, 1771–1783. http://dx.doi.org/10.1007/s00253-013-5452-2, (URL ⟨http://www.ncbi.nlm.nih.gov/pubmed/24362913⟩).

Zamorano, F., Vande Wouwer, A., Bastin, G., 2010. A detailed metabolic flux analysis of an underdetermined network of CHO cells. J. Biotechnol. 150, 497–508. http://dx.doi.org/10.1016/j.jbiotec.2010.09.944, (URL ⟨http://www.ncbi.nlm.nih.gov/pubmed/20869402⟩).

Zamorano, F., Vande Wouwer, A., Jungers, R.M., Bastin, G., 2013. Dynamic metabolic models of CHO cell cultures through minimal sets of elementary flux modes. J. Biotechnol. 164, 409–422. http://dx.doi.org/10.1016/j.jbiotec.2012.05.005, (URL ⟨http://www.ncbi.nlm.nih.gov/pubmed/22698821⟩).