## DiVA字

http://www.diva-portal.org
[This is not an article, chapter, of conference paper!]
since their product results in the aforementioned volumetric productivity. Previous experimental work has shown that a very high viable cell density and a cell-specific productivity comparable to fed-batch reactors can be achieved in perfusion bioreactors coupled to alternating tangential flow systems [2]. The improvement of the cell-specific productivity is the object of ongoing work since it highly depends on the cell line, culture medium, and product of interest.

Hence, reliable and robust operation of perfusion bioreactors at a very high volumetric productivity is one of the main goals of the current research on continuous manufacturing of biologics. For this purpose, mathematical modeling and model-based control and optimization are proposed here as alternatives to the exhaustive screening of culture media that is still the industrial practice [3]. However, the use of mathematical models in the context of perfusion bioreactors is accompanied by significant challenges, such as the complexity and nonlinearity of these mathematical models and the lack of a widely accepted model structure that is guaranteed to capture the transient and steady-state behavior of the perfusion bioreactors under study. In addition, from the methodological point of view, there is no consensus regarding the modeling framework and the most appropriate procedures for model identification, experimental design, and use of the resulting mathematical model for control and optimization. This paper attempts to deal with some of these challenges via an integrated approach that encompasses procedures for modeling and identification of perfusion bioreactors, which complements another recently published paper about an integrated approach for experimental design, control, and optimization of perfusion bioreactors [4].

A framework that is typically used for modeling of biological reaction systems such as perfusion bioreactors is metabolic flux analysis (MFA). The goal of MFA is the computation of rates (also known as fluxes) for each reaction in the cell network so that a certain cost, such as the difference to the rates of variation of measured metabolites, is minimized. In addition, these reaction rates should be such that (i) the rates of variation of intracellular metabolites are equal to zero and (ii) the reaction rates of irreversible reactions are nonnegative [5]. The computation of reaction rates is the first step for the modeling of biological systems such as perfusion bioreactors since one typically constructs a model for each reaction.

Previous work on MFA has used the concept of elementary flux modes (EFMs), which correspond to the extreme rays of the polyhedral cone defined by the conditions (i) and (ii) satisfied by the reaction rates. This concept is useful since any vector of reaction rates satisfies (i) and (ii) if and only if it is a conical combination of EFMs [6]. However, the number of EFMs grows exponentially with the size of the reaction network [7]. Despite the existence of several approaches based on generation of EFMs followed by a selection procedure, these approaches still require the generation of all the EFMs as a first step [8, 9, 10]. Hence, these procedures for selection of EFMs would be applied to a very large number of EFMs and the computational cost would be very high in the case of large reaction networks. Fortunately, it has been shown that it is possible to obtain the reaction rates for one metabolic state, that is, the intracellular state induced by corresponding experimental conditions, as conical combinations of a minimal set of EFMs using the column generation method [11]. However, MFA for different metabolic states requires using the union of the EFMs that were obtained for each state [12]. Even if a procedure for selection of EFMs based on data from different metabolic states is used, the number of selected EFMs may need to increase linearly with the number of metabolic states to allow describing each state in terms of EFMs. This is inconvenient because it can lead to the selection of many EFMs in the case of many metabolic states. For this reason, the stoichiometries of these EFMs may be linearly dependent, and one may not be able to uniquely compute the rates of the EFMs. Also, since the model identification task amounts to finding the model of each reaction, this may have to be done for a large number of EFMs.

In this paper, we present an alternative type of flux modes that can model the system for any metabolic state using a unique and smaller set of flux modes that facilitates model identification. These flux modes allow modeling the biological reaction system only in terms of macroscopic reactions that involve the extracellular species despite the existence of complex intracellular reaction networks, which is a concept that has already been used in previous work [13]. For each flux mode, the corresponding reaction rate can then be modeled using Monod-type kinetics to account for activating or inhibitory effects of certain chemical species [14]. This type of kinetics has been frequently used in other modeling studies for animal cell cultures $[15,16]$. On the other hand, it is known that Monod-type kinetics results in challenging nonlinear estimation problems due to large correlation between parameters [17]. The procedure for model identification and parameter estimation that we propose attempts to deal with two challenges that are typically present in system identification: (i) the difficulty in guaranteeing global optimality of the parameter estimates due to their computation via numerical optimization algorithms that may either attain globally suboptimal solutions or result in intractable problems; and (ii) the identification of the correct model structure among a set of candidate model structures. Hence, the proposed procedure is in line with the purpose of other recently published methods but extends these methods in the sense that it circumvents the need to perform linear reparameterizations or to use modeling frameworks without a physical meaning $[18,19]$. To this end, the rational structure of the kinetic model is used to express the parameter estimation problem as a polynomial optimization problem where the cost and constraints are explicitly written as quadratic functions that involve only a few decision variables, which can be reformulated as a convex semidefinite program via the concept of sum-of-squares polynomials and sparse semidefinite relaxations [20]. This leads to a tractable method that computes the maximum-likelihood parameter estimates with posterior certification of global optimality.

The paper is organized as follows. Section 2 presents the generic model and the concept of flux modes that is used in this paper for modeling of perfusion bioreactors and shows how these flux modes and the corresponding reaction rates are computed. Then, Section 3 takes advantage of these computed reaction rates and shows how one can estimate globally optimal parameters and identify the structure of the Monod-type kinetic model that describes each reaction rate in a tractable way. Finally, Section 4 summarizes the conclusions of this paper.

## 2. Computation of flux modes and reaction rates

In this section, we present the model of perfusion bioreactors used throughout the paper and we propose to work with a reduced set of stoichiometries that are valid for any metabolic state, called basis flux modes (BFMs). By using BFMs, one can show that it is possible to model the biological reaction system for any metabolic state using a unique and much smaller set of stoichiometries than with EFMs, which results in unique computation of flux modes and reaction rates. Hence, the concept of BFMs is more convenient for model identification.

### 2.1. Reaction network

The quantitative relationships among $S$ species that participate in $R_{d}$ reactions are known as stoichiometries. Hence, the net production of the $s$ th species by the $i$ th reaction is given by the stoichiometric coefficient $\nu_{i, s}$. The stoichiometric matrix $\mathbf{N}_{d}$ of dimension $R_{d} \times S$ contains the
stoichiometric coefficients and is defined as ${ }^{1}$

$$
\mathbf{N}_{d}:=\left[\begin{array}{ccc}
\nu_{1,1} & \cdots & \nu_{1, S}  \tag{1}\\
\vdots & \vdots & \vdots \\
\nu_{R_{d}, 1} & \cdots & \nu_{R_{d}, S}
\end{array}\right]
$$

The species in the reaction system are composed of a number of quantities that are conserved by the reactions, namely atoms of different elements and electrical charges. Let $E_{d}$ denote the number of such conserved quantities, and let $\alpha_{s, e}$ denote the number of the $e$ th conserved quantity in the $s$ th species. The atomic matrix $\mathbf{A}$ of dimension $S \times E_{d}$ and rank $E>0$ is then defined as

$$
\mathbf{A}:=\left[\begin{array}{ccc}
\alpha_{1,1} & \cdots & \alpha_{1, E_{d}}  \tag{2}\\
\vdots & \vdots & \vdots \\
\alpha_{S, 1} & \cdots & \alpha_{S, E_{d}}
\end{array}\right] .
$$

For any stoichiometric matrix $\mathbf{N}_{d}$, the $R_{d}$ stoichiometries must satisfy a conservation equation for each one of the $E_{d}$ conserved quantities, that is,

$$
\begin{equation*}
\mathbf{A}^{\mathrm{T}} \mathbf{N}_{d}^{\mathrm{T}}=\mathbf{0}_{E_{d} \times R_{d}} \tag{3}
\end{equation*}
$$

which implies that the columns of $\mathbf{N}_{d}^{\mathrm{T}}$ lie in the null space of $\mathbf{A}^{\mathrm{T}}$ and the rank of $\mathbf{N}_{d}$, denoted as $R$, satisfies $R:=\operatorname{rank}\left(\mathbf{N}_{d}\right) \leq S-E<S$, as Theorem 3 in Appendix A shows.

Now let $\mathbf{r}_{d}(t)$ denote the $R_{d}$-dimensional vector of reaction rates that correspond to the stoichiometric matrix $\mathbf{N}_{d}$. These reaction rates are typically functions of the vector $\mathbf{c}(t)$ of concentrations of certain species and of the temperature $T(t)$, which correspond to the so-called reaction kinetics, but they are represented here as time-varying signals $\mathbf{r}_{d}(t)$, without indicating the explicit dependence on $\mathbf{c}(t)$ and $T(t)$. One can observe that the net production of the $S$ species by the reactions at time $t$ is given by $\mathbf{N}_{d}^{\mathrm{T}} \mathbf{r}_{d}(t)$.

The stoichiometries that correspond to the rows of the matrix $\mathbf{N}_{d}$ may be linearly dependent. Ideally, one would like to describe the net production by the reactions with linearly independent stoichiometries.

### 2.2. Transformation to independent stoichiometries

Recall that $\mathbf{N}_{d}$ is an $R_{d} \times S$ stoichiometric matrix of rank $R$, where $R_{d}$ is the number of possibly linearly dependent stoichiometries. The stoichiometries in the matrix $\mathbf{N}_{d}$ result from linear combinations, specified by an $R_{d} \times R$ matrix $\mathbf{L}_{N}$ of rank $R$, of $R$ linearly independent stoichiometries, specified by an $R \times S$ matrix $\mathbf{N}$ of rank $R$, which can be written as

$$
\begin{equation*}
\mathbf{N}_{d}=\mathbf{L}_{N} \mathbf{N} \tag{4}
\end{equation*}
$$

Since it would be useful to find the underlying independent stoichiometries, two methods M1 and M2 are presented below to decompose any matrix $\mathbf{N}_{d}$ with $R_{d}$ rows and rank $R$ as $\mathbf{N}_{d}=\mathbf{L}_{N} \mathbf{N}$,

[^0]where $\mathbf{N}$ is a matrix with $R$ linearly independent rows. Hence, both methods can be used to obtain independent stoichiometries. A singular value decomposition could also be used for this purpose, which would be done in MATLAB by executing the commands $R=r a n k(N d) ; ~[U, S, V]=s v d(N d)$; $L N=U * S(:, 1: R) ; N=V(:, 1: R)^{\prime} ;$. However, the goal here is to look for a more physically meaningful decomposition in the context of independent reactions by using rational bases of null spaces, that is, bases where the elements are ratios of small integers. These bases are obtained from the reduced row echelon form of the matrices for which the null spaces are computed, which can be done in MATLAB for some matrix $\mathbf{M}$ by executing the command $Z=n u l l(M, ' r ') ;$ Note that this property of rational bases of null spaces also implies that executing the MATLAB commands $Z=n u l l(M, ' r ')$; $W=n u l l\left(Z^{\prime}, ' r '\right)$ '; for any matrix $\mathbf{M}$ results in the same matrix $\mathbf{W}$ as executing $R=r a n k(M) ; Y=r r e f(f l i p l r(M)) ; W=f l i p l r(f l i p u d(Y(1: R,:))) ;$, as one can verify numerically, which means that $\mathbf{W}$ contains the nonzero rows of the reduced row echelon form of a flipped version of $\mathbf{M}$. This fact is used below to interpret the results achieved by methods M1 and M2.

Consequently, to obtain a model of the reaction system expressed only in terms of independent stoichiometries, the following steps are taken:

1. One computes an $R_{d} \times R$ matrix $\mathbf{L}_{N}$ of rank $R$ and an $R \times S$ matrix $\mathbf{N}$ of rank $R$ such that $\mathbf{N}_{d}=\mathbf{L}_{N} \mathbf{N}$. In addition, one computes an $S \times(S-R)$ matrix $\mathbf{A}_{N}$ of rank $S-R$ with columns that span the null spaces of $\mathbf{N}_{d}$ and $\mathbf{N}$ and an $R_{d} \times\left(R_{d}-R\right)$ matrix $\mathbf{K}_{N}$ of rank $R_{d}-R$ with columns that span the null spaces of $\mathbf{N}_{d}^{\mathrm{T}}$ and $\mathbf{L}_{N}^{\mathrm{T}}$ since they are used in subsequent sections. To this end, one can use one of the two following methods, among other possible methods:

M1. One finds a matrix $\mathbf{L}_{N}$ with columns that are a rational basis of the null space of $\mathbf{K}_{N}^{\mathrm{T}}$, where $\mathbf{K}_{N}$ is a matrix with columns that are a rational basis of the null space of $\mathbf{N}_{d}^{\mathrm{T}}$. Then, since $\mathbf{L}_{N}$ is of full column rank, one computes $\mathbf{N}=\left(\mathbf{L}_{N}^{\mathrm{T}} \mathbf{L}_{N}\right)^{-1} \mathbf{L}_{N}^{\mathrm{T}} \mathbf{N}_{d}$ and finds a matrix $\mathbf{A}_{N}$ with columns that are a rational basis of the null space of $\mathbf{N}$, which is also the null space of $\mathbf{N}_{d}$ since $\mathbf{N}_{d}=\mathbf{L}_{N} \mathbf{N}$. This can be done in MATLAB by executing the commands KN=null(Nd', 'r'); LN=null (KN', 'r') ; N=pinv(LN) *Nd; AN $=$ null ( $\mathrm{N}, \mathrm{r} \mathrm{r}^{\prime}$ ) ; . This results in a matrix $\mathbf{L}_{N}^{\mathrm{T}}$ where the linearly dependent stoichiometries that correspond to columns with pivots in the reduced row echelon form of a flipped version of $\mathbf{N}_{d}^{\mathrm{T}}$ are involved in only one independent stoichiometry and a matrix $\mathbf{N}^{\mathrm{T}}$ with the columns of $\mathbf{N}_{d}^{\mathrm{T}}$ that correspond to the columns with pivots in the said reduced row echelon form. Hence, this method is most useful when one aims to obtain a matrix $\mathbf{L}_{N}$ such that most of the linearly dependent stoichiometries are involved in only one linearly independent stoichiometry and a matrix $\mathbf{N}$ with rows similar to the rows of $\mathbf{N}_{d}$.
M2. One finds a matrix $\mathbf{N}^{\mathrm{T}}$ with columns that are a rational basis of the null space of $\mathbf{A}_{N}^{\mathrm{T}}$, where $\mathbf{A}_{N}$ is a matrix with columns that are a rational basis of the null space of $\mathbf{N}_{d}$. Then, since $\mathbf{N}$ is of full row rank, one computes $\mathbf{L}_{N}=\mathbf{N}_{d} \mathbf{N}^{\mathrm{T}}\left(\mathbf{N} \mathbf{N}^{\mathrm{T}}\right)^{-1}$ and finds a matrix $\mathbf{K}_{N}$ with columns that are a rational basis of the null space of $\mathbf{L}_{N}^{\mathrm{T}}$, which is also the null space of $\mathbf{N}_{d}^{\mathrm{T}}$ since $\mathbf{N}_{d}^{\mathrm{T}}=\mathbf{N}^{\mathrm{T}} \mathbf{L}_{N}^{\mathrm{T}}$. This can be done in MATLAB by executing the commands AN=null(Nd,'r'); N=null(AN','r')'; LN=Nd*pinv(N); KN=null (LN', 'r'); This results in a matrix $\mathbf{N}$ where the species that correspond to columns with pivots in the reduced row echelon form of a flipped version of $\mathbf{N}_{d}$ are involved in only one independent stoichiometry and a matrix $\mathbf{L}_{N}$ with the columns of $\mathbf{N}_{d}$ that correspond to the columns with pivots in the said reduced row echelon form.

Hence, this method is most useful when one aims to obtain a matrix $\mathbf{N}$ such that most of the species are involved in only one linearly independent stoichiometry and a matrix $\mathbf{L}_{N}$ with columns similar to the columns of $\mathbf{N}_{d}$.
2. At this point, one can write that

$$
\begin{equation*}
\mathbf{N}_{d}^{\mathrm{T}} \mathbf{r}_{d}(t)=\mathbf{N}^{\mathrm{T}} \mathbf{L}_{N}^{\mathrm{T}} \mathbf{r}_{d}(t) \tag{5}
\end{equation*}
$$

Let us denote the number of independent reactions as $R$. Then,

$$
\begin{equation*}
\mathbf{N}_{d}^{\mathrm{T}} \mathbf{r}_{d}(t)=\mathbf{N}^{\mathrm{T}} \mathbf{r}(t) \tag{6}
\end{equation*}
$$

with

$$
\begin{equation*}
\mathbf{r}(t)=\mathbf{L}_{N}^{\mathrm{T}} \mathbf{r}_{d}(t) \tag{7}
\end{equation*}
$$

Since it has been shown that the number of linearly independent stoichiometries is less than the number of species, that is, $R<S$, the number of independent reactions is also less than the number of species.

In summary, one can assume without loss of generality that:

- There are $R$ independent reactions, and the number of independent reactions is less than the number of species, that is, $R<S$.
- The stoichiometric matrix $\mathbf{N}$ is of dimension $R \times S$ and of rank $R$.
- The vector of independent reaction rates $\mathbf{r}(t)$ is of dimension $R$.


### 2.2.1. Illustrative example

The previous results are now illustrated by a simple example of a biological reaction system. In this system, the $S=18$ species in Table 1 participate in the $R_{d}=14$ reactions in Table 2 with the linearly dependent stoichiometries in the $R_{d} \times S$ matrix

$$
\mathbf{N}_{d}=\left[\begin{array}{rrrrrrrrrrrrrrrrrr}
-1 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0  \tag{8}\\
-1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 4 \\
1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 2 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & -2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 1 & -1 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & -1 & 1 & 0 & -1 & 4 \\
0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 6 \\
0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 0 & 6 \\
0 & -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & 0 & 1 \\
0 & -1 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 \\
-4 & 13 & 0 & 0 & -3 & 0 & 0 & 0 & -1 & 2 & 0 & 0 & -2 & 0 & 2 & -2 & -2 & -8 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 1 & -3 & 0 & 0 & 0 & 0 & 0 \\
0 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -4
\end{array}\right]
$$

of rank $R=13$. This reaction network corresponds to an amendment of the networks proposed by $[21,22]$ that was made to ensure that all the reactions satisfy the conservation of the $E_{d}=5$ quantities that correspond to the atoms of $\mathrm{C}, \mathrm{H}, \mathrm{N}, \mathrm{O}, \mathrm{P}$. The conserved quantities are given by the $S \times E_{d}$ matrix $\mathbf{A}$ of rank $E=5$, with

Assuming that one is interested in obtaining linearly independent stoichiometries in the rows of the matrix $\mathbf{N}$ that are similar to the linearly dependent stoichiometries in the rows of the matrix $\mathbf{N}_{d}$, the method described in M1 is used. One can then compute the $S \times(S-R)$ matrix $\mathbf{A}_{N}$ of rank $S-R$, the $R_{d} \times\left(R_{d}-R\right)$ matrix $\mathbf{K}_{N}$ of rank $R_{d}-R$, with

$$
\begin{align*}
& \mathbf{K}_{N}^{\mathrm{T}}=\left[\begin{array}{llllllllll}
0-1 & 3 & -1 & 0 & 0 & -1-1 & 0 & 0 & 0 & 0
\end{array} 10\right] \text {, } \tag{10}
\end{align*}
$$

the $R_{d} \times R$ matrix $\mathbf{L}_{N}$ of rank $R$, and the $R \times S$ matrix $\mathbf{N}$ of rank $R$, the rows of which are the $R$ linearly independent stoichiometries, which are exactly the linearly dependent stoichiometries in the rows of the matrix $\mathbf{N}_{d}$, with the exception of the stoichiometry in the second row of $\mathbf{N}_{d}$.

On the other hand, assuming that one is interested in obtaining linearly independent stoichiometries in the rows of the matrix $\mathbf{N}$ such that most of the species are involved in only one linearly independent stoichiometry, the method described in M2 is used. One can then compute the same matrices $\mathbf{A}_{N}$ and $\mathbf{K}_{N}$, the $R_{d} \times R$ matrix $\mathbf{L}_{N}$ of rank $R$, and the $R \times S$ matrix $\mathbf{N}$ of rank $R$ with the $R$ linearly independent stoichiometries of the reactions in Table 3. Indeed, one can observe that most of the species are involved in only one of the linearly independent stoichiometries in the rows of the matrix $\mathbf{N}$, with the exception of $\mathrm{H} 3 \mathrm{PO} 4_{\text {ext }}, \mathrm{H} 2 \mathrm{O}_{\text {ext }}, \mathrm{O} 2_{\mathrm{ext}}$, $\mathrm{Glc}_{\mathrm{ext}}$, and $\mathrm{Gln}_{\text {ext }}$.

In both cases, $\mathbf{K}_{N}$ corresponds to the single linear dependence among the linearly dependent stoichiometries in the rows of the matrix $\mathbf{N}_{d}$, while $\mathbf{A}_{N}$ corresponds to the conservation of quantities in the species by these stoichiometries. For this reason, a single linearly dependent stoichiometry can be removed from the rows of the matrix $\mathbf{N}_{d}$, while the matrix $\mathbf{A}_{N}$ can be related to the conservation of atoms of $\mathrm{C}, \mathrm{H}, \mathrm{N}, \mathrm{O}, \mathrm{P}$ in the species given by the matrix $\mathbf{A}$ as follows:

$$
\mathbf{A}_{N}^{\mathrm{T}}=\left[\begin{array}{rrrrr}
0 & 0 & 0 & 0 & \frac{1}{1}  \tag{12}\\
1 & 0 & -\frac{9}{5} & -\frac{4}{5} & \frac{7}{5} \\
0 & 0 & 1 & 0 & 0 \\
-1 & 0 & 1 & 1 & -3 \\
-2 & 1 & -\frac{11}{5} & \frac{4}{5} & -\frac{37}{5}
\end{array}\right] \mathbf{A}^{\mathrm{T}},
$$

which means that the fourth row of $\mathbf{A}_{N}^{\mathrm{T}}$ corresponds to the number of atoms of N and O minus the number of atoms of C minus 3 times the number of atoms of P in the species, for example.

### 2.3. Generic model of perfusion bioreactors

In perfusion bioreactors, suspended cells are cultivated in a liquid medium that is continuously renewed via the reactor inlets and outlets. The goal is to keep the cells in an environment that allows them to generate a certain product of interest from the nutrients and other species in the medium and to recover that product continuously. To avoid loss of productivity due to dilution of biomass as a result of the medium renewal and to facilitate the downstream processing of the product of interest, perfusion bioreactors are coupled to a cell retention device. This ensures that the biomass remains in the reactor and the other extracellular species, including the product of interest, are continuously harvested. In addition, a so-called bleed stream is typically used to avoid accumulation of dead cells. Figure 1 shows a schematic of a perfusion bioreactor coupled to a cell retention device.

Hence, a perfusion bioreactor is typically a constant-volume, continuous, agitated bioreactor that is intended to operate predominantly at steady state, but here its transient behavior is shown as a

Table 1: List of species in the illustrative example.

| Species | Label | Chemical formula |
| :---: | :---: | :---: |
| 1 | $\mathrm{H} 3 \mathrm{PO} 4_{\text {ext }}$ | $\mathrm{H}_{3} \mathrm{PO}_{4}$ |
| 2 | H 2 O ext | $\mathrm{H}_{2} \mathrm{O}$ |
| 3 | $\mathrm{O} 2{ }_{\text {ext }}$ | $\mathrm{O}_{2}$ |
| 4 | $\mathrm{Glc}_{\text {ext }}$ | $\mathrm{C}_{6} \mathrm{H}_{12} \mathrm{O}_{6}$ |
| 5 | Gln ${ }_{\text {ext }}$ | $\mathrm{C}_{5} \mathrm{H}_{10} \mathrm{~N}_{2} \mathrm{O}_{3}$ |
| 6 | Lacext | $\mathrm{C}_{3} \mathrm{H}_{6} \mathrm{O}_{3}$ |
| 7 | NH 4 ext | $\mathrm{NH}_{4}$ |
| 8 | Ala ${ }_{\text {ext }}$ | $\mathrm{C}_{3} \mathrm{H}_{7} \mathrm{NO}_{2}$ |
| 9 | CO 2 ext | $\mathrm{CO}_{2}$ |
| 10 | Nucext | $\mathrm{C}_{17} \mathrm{H}_{30} \mathrm{~N}_{4} \mathrm{O}_{19} \mathrm{P}_{3}$ |
| 11 | G6P | $\mathrm{C}_{6} \mathrm{H}_{13} \mathrm{O}_{9} \mathrm{P}$ |
| 12 | GA3P | $\mathrm{C}_{3} \mathrm{H}_{7} \mathrm{O}_{6} \mathrm{P}$ |
| 13 | R5P | $\mathrm{C}_{5} \mathrm{H}_{11} \mathrm{O}_{8} \mathrm{P}$ |
| 14 | Pyr | $\mathrm{C}_{3} \mathrm{H}_{4} \mathrm{O}_{3}$ |
| 15 | aKG | $\mathrm{C}_{5} \mathrm{H}_{6} \mathrm{O}_{5}$ |
| 16 | Glu | $\mathrm{C}_{5} \mathrm{H}_{9} \mathrm{NO}_{4}$ |
| 17 | Oxa | $\mathrm{C}_{4} \mathrm{H}_{4} \mathrm{O}_{5}$ |
| 18 | H | H |

Table 2: List of reactions with linearly dependent stoichiometries in the illustrative example.

| Reaction | Chemical equation |
| :---: | :---: |
| 1 | $\mathrm{H} 3 \mathrm{PO}^{\text {ext }}+\mathrm{Glc}_{\text {ext }} \leftrightarrow \mathrm{H}^{\text {2O }}$ ext +G 6 P |
| 2 | H3PO4ext $+\mathrm{G} 6 \mathrm{P} \leftrightarrow \mathrm{H} 2 \mathrm{O}_{\text {ext }}+2 \mathrm{GA} 3 \mathrm{P}$ |
| 3 | H 2 O ext $+\mathrm{G} 6 \mathrm{P} \leftrightarrow \mathrm{CO} 2$ ext $+\mathrm{R} 5 \mathrm{P}+4 \mathrm{H}$ |
| 4 | H 2 O ext $+\mathrm{GA} 3 \mathrm{P} \leftrightarrow \mathrm{H} 3 \mathrm{PO} 44_{\text {ext }}+\mathrm{Pyr}+2 \mathrm{H}$ |
| 5 | $\mathrm{Pyr}+2 \mathrm{H} \leftrightarrow \mathrm{Lac}_{\text {ext }}$ |
| 6 | Pyr + Glu $\leftrightarrow \mathrm{Ala}_{\text {ext }}+\mathrm{aKG}$ |
| 7 | H 2 O ext $+\mathrm{Pyr}+\mathrm{Oxa} \leftrightarrow 2 \mathrm{CO} 2$ ext $+\mathrm{aKG}+4 \mathrm{H}$ |
| 8 | 2 H 2 O ext $+\mathrm{aKG} \leftrightarrow \mathrm{CO} 2$ ext $+\mathrm{Oxa}+6 \mathrm{H}$ |
| 9 | 2 H 2 O ext $+\mathrm{aKG} \leftrightarrow 2 \mathrm{CO} 2$ ext $+\mathrm{Pyr}+6 \mathrm{H}$ |
| 10 | H 2 O ext $+\mathrm{Glu} \leftrightarrow \mathrm{NH} 4 \mathrm{e}_{\text {ext }}+\mathrm{aKG}+\mathrm{H}$ |
| 11 |  |
| 12 |  |
| 13 | $3 \mathrm{R} 5 \mathrm{P} \leftrightarrow 2 \mathrm{G} 6 \mathrm{P}+\mathrm{GA} 3 \mathrm{P}$ |
| 14 | $\mathrm{O} 2 e_{\text {ext }}+4 \mathrm{H} \leftrightarrow 2 \mathrm{H} 2 \mathrm{O}_{\text {ext }}$ |

Table 3: List of reactions with linearly independent stoichiometries in the illustrative example using the method described in M2.

| Reaction | Chemical equation |
| :---: | :---: |
| 1 | $0.5 \mathrm{Glc}_{\text {ext }} \leftrightarrow \mathrm{Lac}_{\text {ext }}$ |
| 2 |  |
| 3 | 0.5 H 2 O ext $+0.083333 \mathrm{Glc}_{\text {ext }}+0.5 \mathrm{Gln}_{\text {ext }} \leftrightarrow 0.25 \mathrm{O} 2{ }_{\text {ext }}+\mathrm{Ala}_{\text {ext }}$ |
| 4 | $\mathrm{O} 22_{\text {ext }}+0.16667 \mathrm{Glc}_{\text {ext }} \leftrightarrow \mathrm{H} 2 \mathrm{O}_{\text {ext }}+\mathrm{CO} 2 \mathrm{ext}^{\text {ext }}$ |
| 5 | 3 H 3 PO 4 ext +0.25 O 2 ext $+1.1667 \mathrm{Glc}_{\text {ext }}+2 \mathrm{Gln}_{\text {ext }} \leftrightarrow 6.5 \mathrm{H} 2 \mathrm{O}_{\text {ext }}+\mathrm{Nuc}_{\text {ext }}$ |
| 6 | H 3 PO 4 ext $^{+} \mathrm{Glc}_{\text {ext }} \leftrightarrow \mathrm{H} 2 \mathrm{O}_{\text {ext }}+\mathrm{G} 6 \mathrm{P}$ |
| 7 | $\mathrm{H} 3 \mathrm{PO}_{4}$ ext $+0.5 \mathrm{Glcext}_{\text {ext }} \leftrightarrow \mathrm{H} 2 \mathrm{O}$ ext +GA 3 P |
| 8 | $\mathrm{H} 3 \mathrm{PO} 4_{\text {ext }}+0.83333 \mathrm{Glcext}^{\leftrightarrow} \mathrm{H} 2 \mathrm{O}$ ext +R 5 P |
| 9 | 0.5 O 2 ext $+0.5 \mathrm{Glc}_{\text {ext }} \leftrightarrow \mathrm{H} 2 \mathrm{O}^{\text {ext }}+\mathrm{Pyr}$ |
| 10 | O 2 ext $+0.83333 \mathrm{Glc}_{\text {ext }} \leftrightarrow 2 \mathrm{H} 2 \mathrm{O}$ ext + aKG |
| 11 | 0.25 O 2 ext $+0.41667 \mathrm{Glc} \mathrm{ext}^{\text {ext }}+0.5 \mathrm{Gln}$ ext $\leftrightarrow 0.5 \mathrm{H} 2 \mathrm{O}_{\text {ext }}+\mathrm{Glu}$ |
| 12 | $1.5 \mathrm{O} 22_{\text {ext }}+0.66667 \mathrm{Glc}_{\text {ext }} \leftrightarrow 2 \mathrm{H} 2 \mathrm{O}$ ext +Oxa |
| 13 | $0.5 \mathrm{H} 2 \mathrm{O}_{\text {ext }} \leftrightarrow 0.25 \mathrm{O} 2$ ext +H |



Figure 1: Schematic of a perfusion bioreactor coupled to a cell retention device.
dynamic model in the form of differential-algebraic equations rather than a steady-state model in the form of algebraic equations. In this biological reaction system, there are $S_{e c}$ extracellular species and $S_{i c}$ intracellular species. Assuming that the species are ordered such that the extracellular species precede the intracellular species, the extracellular and intracellular species are selected from the complete set of species using the selection matrices $\mathbf{S}_{e c}=\left[\begin{array}{ll}\mathbf{I}_{S_{e c}} & \mathbf{0}_{S_{e c} \times S_{i c}}\end{array}\right]$ of dimension $S_{e c} \times S$ and $\mathbf{S}_{i c}=\left[\begin{array}{l}\mathbf{0}_{S_{i c} \times S_{e c}} \\ \mathbf{I}_{S_{i c}}\end{array}\right]$ of dimension $S_{i c} \times S$, respectively. The stoichiometries of the independent reactions that affect the intracellular and extracellular species are given by the matrices $\mathbf{N}_{i c}=\mathbf{N S}_{i c}^{\mathrm{T}}$ and $\mathbf{N}_{e c}=\mathbf{N S} \mathbf{S}_{e c}^{\mathrm{T}}$, respectively. For these systems, a pseudo steady-state assumption for the intracellular species is typically used. This means that the generic dynamic model of a perfusion bioreactor is described by the following system of differential-algebraic equations:

$$
\begin{align*}
\dot{\mathbf{c}}(t) & =\mathbf{N}_{e c}^{\mathrm{T}} \mathbf{r}_{c}(\mathbf{c}(t))+\mathbf{C}_{i n}(t) \boldsymbol{\omega}_{i n}(t)+\omega_{h}(t) \mathbf{R}(t) \mathbf{c}(t)-\omega_{p}(t) \mathbf{c}(t),  \tag{13a}\\
\mathbf{0}_{S_{i c}} & =\mathbf{N}_{i c}^{\mathrm{T}} \mathbf{r}_{c}(\mathbf{c}(t)), \tag{13b}
\end{align*}
$$

where $\mathbf{c}(t)$ is the $S_{e c}$-dimensional vector of bioreactor concentrations of extracellular species including biomass in moles per unit of volume, $\boldsymbol{\omega}_{i n}(t)$ is the $p$-dimensional vector of inlet rates in reactor volumes per unit of time, $\mathbf{C}_{i n}(t)$ is the $S_{e c} \times p$ matrix of inlet concentrations of extracellular species including biomass, with $\mathbf{C}_{i n, j}(t)$, the $j$ th column of $\mathbf{C}_{i n}(t)$, being the vector of inlet concentrations of those species in the $j$ th inlet for $j=1, \ldots, p, \mathbf{R}(t)$ is the $S_{e c}$-dimensional diagonal matrix of retention factors for the extracellular species (fractions of these species retained by the cell retention device), and $\omega_{p}(t)$ is the perfusion rate in reactor volumes per unit of time, which corresponds not only to the sum of the harvest rate $\omega_{h}(t)$ and bleed rate $\omega_{b}(t)$ but also to the sum of the inlet rates $\boldsymbol{\omega}_{i n}(t)$, that is, $\mathbf{1}_{p}^{\mathrm{T}} \boldsymbol{\omega}_{i n}(t)=\omega_{p}(t)=\omega_{h}(t)+\omega_{b}(t)$. Furthermore, $\mathbf{r}_{c}(\mathbf{c}(t))$ are independent reaction rates in moles per unit of volume per unit of time, and their dependence on the concentrations of the species in the bioreactor is typically the only unknown part of the model, which needs to be estimated from experimental results via appropriate model identification techniques. The relation between the vectors of independent reaction rates $\mathbf{r}_{c}(\mathbf{c}(t))$ and $\mathbf{r}(t)$ is given by $\mathbf{r}(t)=\frac{\mathbf{r}_{c}(\mathbf{c}(t))}{V C(t)}$. The
reaction rates $\mathbf{r}(t)$ are thus specific reaction rates in moles per cell per unit of time, expressed without the explicit dependence on the concentrations of the species in the bioreactor, where $V C(t)$ is the viable cell concentration in viable cells per unit of volume.

In the case of biomass, $c_{\text {biom }}(t)$ is the biomass concentration in moles per unit of volume, obtained from $V C(t)$ upon multiplying by a scaling factor $f_{\text {biom }}$ that expresses the moles of biomass per viable cell, that is, $c_{\text {biom }}(t)=f_{\text {biom }} V C(t)$. Biomass is selected from the extracellular species using the $S_{\text {ec }}$-dimensional selection vector $\mathbf{s}_{\text {biom }}$ such that $c_{\text {biom }}(t)=\mathbf{s}_{b i o m}^{\mathrm{T}} \mathbf{c}(t)$, the stoichiometries of the independent reactions that affect it are given by the vector $\mathbf{n}_{b i o m}=\mathbf{N}_{e c} \mathbf{s}_{b i o m}$, and it is expected that its retention factor is $R_{b i o m}(t)=\mathbf{s}_{\text {biom }}^{\mathrm{T}} \mathbf{R}(t) \mathbf{s}_{\text {biom }}=1$ and its $p$-dimensional row vector of inlet concentrations is $\mathbf{c}_{i n, b i o m}(t)=\mathbf{s}_{b i o m}^{\mathrm{T}} \mathbf{C}_{\text {in }}(t)=\mathbf{0}_{p}^{\mathrm{T}}$, which implies that

$$
\begin{align*}
\dot{c}_{\text {biom }}(t) & =\mathbf{n}_{\text {biom }}^{\mathrm{T}} \mathbf{r}_{c}(\mathbf{c}(t))+\mathbf{c}_{\text {in }, \text { biom }}(t) \boldsymbol{\omega}_{\text {in }}(t)+\omega_{h}(t) R_{\text {biom }}(t) c_{\text {biom }}(t)-\omega_{p}(t) c_{\text {biom }}(t) \\
& =\mathbf{n}_{\text {biom }}^{\mathrm{T}} \mathbf{r}_{c}(\mathbf{c}(t))-\omega_{b}(t) c_{\text {biom }}(t) \tag{14}
\end{align*}
$$

It is possible to rewrite the generic model (13) in different ways. For example, one can denote the specific rates of variation of the $S$ species due to reactions as the $S$-dimensional vector $\mathbf{q}(t)=\mathbf{N}^{\mathrm{T}} \mathbf{r}(t)$ and the overall inlet concentrations of extracellular species including biomass as the $S_{e c}$-dimensional vector $\mathbf{c}_{i n}(t)=\frac{1}{\omega_{p}(t)} \mathbf{C}_{i n}(t) \boldsymbol{\omega}_{i n}(t)$, which results in

$$
\left[\begin{array}{c}
\mathbf{S}_{e c}  \tag{15}\\
\mathbf{S}_{i c}
\end{array}\right] \mathbf{q}(t)=\frac{1}{V C(t)}\left[\begin{array}{c}
\dot{\mathbf{c}}(t)+\omega_{p}(t)\left(\mathbf{c}(t)-\mathbf{c}_{i n}(t)\right)-\omega_{h}(t) \mathbf{R}(t) \mathbf{c}(t) \\
\mathbf{0}_{S_{i c}}
\end{array}\right]
$$

In the case of biomass, its specific rate of variation due to reactions is $q_{b i o m}(t)=\mathbf{s}_{b i o m}^{\mathrm{T}} \mathbf{N}_{e c}^{\mathrm{T}} \mathbf{r}(t)$, and it is expected that its overall inlet concentration is $c_{\text {in,biom }}(t)=0$, which implies that

$$
\begin{equation*}
q_{b i o m}(t)=\frac{\dot{c}_{b i o m}(t)+\omega_{p}(t)\left(c_{b i o m}(t)-c_{i n, b i o m}(t)\right)-\omega_{h}(t) R_{b i o m}(t) c_{b i o m}(t)}{V C(t)}=\frac{\dot{c}_{b i o m}(t)}{V C(t)}+\omega_{b}(t) f_{b i o m} \tag{16}
\end{equation*}
$$

### 2.4. Basis flux modes

We would like to construct a minimal set of flux modes that is uniquely defined for any metabolic state of the cells and enables the computation of unique results for the rates of all the independent reactions. Furthermore, it should be possible to obtain any vector of rates that satisfies the pseudo steady-state assumption for the intracellular species if and only if it is a linear combination of the stoichiometries of this minimal set of flux modes. This explains why these flux modes are called basis flux modes (BFMs) in the remainder. Let $R_{m}$ denote the dimension of this set of BFMs, and let $\mathbf{N}_{m}$ of dimension $R_{m} \times S$ and rank $R_{m}$ denote the stoichiometric matrix that represents the stoichiometry of these BFMs.

It is known that flux modes satisfy some properties: (i) they are obtained from linear combinations of reactions in the reaction system, which also implies that they satisfy the same conservation of quantities as the original reactions, and (ii) their stoichiometry involves only the extracellular species, and not the intracellular species. The first property implies that the rows of $\mathbf{N}_{m}$ are in a space that is orthogonal to the column space of $\mathbf{A}_{N}$, since this space of dimension $q:=S-R$ spans the null space of $\mathbf{N}_{d}$. The second property implies that the rows of $\mathbf{N}_{m}$ are in a space that is orthogonal to the row space of $\mathbf{S}_{i c}$. In other words, any candidate for $\mathbf{N}_{m}$ is such that the $R_{m}$ columns of $\mathbf{N}_{m}^{\mathrm{T}}$ span the null space of $\left[\begin{array}{c}\mathbf{A}_{N}^{\mathrm{T}} \\ \mathbf{S}_{i c}\end{array}\right]$, which can be expressed mathematically as

$$
\left[\begin{array}{c}
\mathbf{A}_{N}^{\mathrm{T}}  \tag{17}\\
\mathbf{S}_{i c}
\end{array}\right] \mathbf{N}_{m}^{\mathrm{T}}=\left[\begin{array}{c}
\mathbf{0}_{q \times R_{m}} \\
\mathbf{0}_{S i c} \times R_{m}
\end{array}\right]
$$

If we denote the rank of $\mathbf{A}_{N}^{\mathrm{T}} \mathbf{S}_{e c}^{\mathrm{T}}$ as $q_{e c}$ and the rank of $\mathbf{A}^{\mathrm{T}} \mathbf{S}_{e c}^{\mathrm{T}}$ as $E_{e c}$, Theorem 4 in Appendix A shows the following implications:

- From the definitions of $\mathbf{S}_{e c}$ and $\mathbf{S}_{i c}$, the matrix $\mathbf{N}_{m}^{\mathrm{T}}$ can also be constructed as $\left[\begin{array}{l}\mathbf{S}_{e c} \mathbf{N}_{m}^{\mathrm{T}} \\ \mathbf{S}_{i c} \mathbf{N}_{m}^{\mathrm{T}}\end{array}\right]$, where $\mathbf{S}_{i c} \mathbf{N}_{m}^{\mathrm{T}}=\mathbf{0}_{S_{i c} \times R_{m}}$ and the $R_{m}$ columns of $\mathbf{S}_{e c} \mathbf{N}_{m}^{\mathrm{T}}$ span the null space of $\mathbf{A}_{N}^{\mathrm{T}} \mathbf{S}_{e c}^{\mathrm{T}}$.
- The rank of $\left[\begin{array}{c}\mathbf{A}_{N}^{\mathrm{T}} \\ \mathbf{S}_{i c}\end{array}\right]$ is $q_{e c}+S_{i c} \geq E_{e c}+S_{i c}>S_{i c}$, that is, greater than the number of intracellular species.
- The number of BFMs, the number of columns of $\mathbf{N}_{m}^{\mathrm{T}}$, and the dimension of the null spaces of $\left[\begin{array}{c}\mathbf{A}_{N}^{\mathrm{T}} \\ \mathbf{S}_{i c}\end{array}\right]$ and $\mathbf{A}_{N}^{\mathrm{T}} \mathbf{S}_{e c}^{\mathrm{T}}$ are $R_{m}=S_{e c}-q_{e c} \leq S_{e c}-E_{e c}<S_{e c}$, that is, less than the number of extracellular species.

Once a candidate for $\mathbf{N}_{m}$ is chosen, one would like to know which linear combinations of the independent stoichiometries in the rows of $\mathbf{N}$ correspond to the stoichiometries of the BFMs in the rows of $\mathbf{N}_{m}$. In other words, one would like to compute the $R \times R_{m}$ matrix $\mathbf{E}_{m}$ of rank $R_{m}$ such that

$$
\begin{equation*}
\mathbf{N}^{\mathrm{T}} \mathbf{E}_{m}=\mathbf{N}_{m}^{\mathrm{T}} \tag{18}
\end{equation*}
$$

Since the stoichiometric matrix $\mathbf{N}$ is of full row rank and the rows of $\mathbf{N}_{m}$ are in the row space of $\mathbf{N}$, it is guaranteed that there is a unique solution for $\mathbf{E}_{m}$ for a given matrix $\mathbf{N}_{m}$. In Appendix A, Theorem 5 shows that the rank of $\mathbf{N}_{i c}^{\mathrm{T}}$ is equal to $R-R_{m}$ and the $R_{m}$ columns of $\mathbf{E}_{m}$ span the null space of $\mathbf{N}_{i c}^{\mathrm{T}}$, which indicates the desired relation between the concept of BFMs and the pseudo steady-state assumption for the intracellular species.

The solution for $\mathbf{E}_{m}$ can still be uniquely obtained if we consider only the $S_{i c}$ intracellular species and a subset of $S_{a}$ extracellular species, represented by the $S_{a} \times S$ selection matrix $\mathbf{S}_{a}=$ $\left[\begin{array}{ll}\mathbf{I}_{S_{a}} & \mathbf{0}_{S_{a} \times\left(S-S_{a}\right)}\end{array}\right]$ and the stoichiometric matrix $\mathbf{N}_{a}=\mathbf{N} \mathbf{S}_{a}^{\mathrm{T}}$. In that case, it is required that the stoichiometric matrix $\left[\begin{array}{ll}\mathbf{N}_{a} & \mathbf{N}_{i c}\end{array}\right]$ be of full row rank $R$, which is the case if and only if $\mathbf{S}_{a} \mathbf{N}_{m}^{\mathrm{T}}$ is of full column rank $R_{m}$. Note that this can occur only if $S_{a} \geq R_{m}=S_{e c}-q_{e c}$, that is, only if at least $S_{e c}-q_{e c}$ extracellular species are part of the subset of $S_{a}$ extracellular species. Theorem 6 in Appendix A shows the more general fact that $\operatorname{rank}\left(\left[\begin{array}{ll}\mathbf{N}_{a} & \mathbf{N}_{i c}\end{array}\right]\right)-R=\operatorname{rank}\left(\mathbf{S}_{a} \mathbf{N}_{m}^{\mathrm{T}}\right)-R_{m}$.

A question that can be asked at this point is whether the true reaction system can always be described correctly by using the stoichiometry of the BFMs represented by a particular $\mathbf{N}_{m}$. It is possible to show that this is always the case as in Theorem 7 since there exists some $R_{m}$-dimensional vector of rates $\boldsymbol{\psi}(t)$ such that

$$
\left[\begin{array}{c}
\mathbf{q}(t)  \tag{19}\\
\mathbf{r}(t)
\end{array}\right]=\left[\begin{array}{c}
\mathbf{N}_{m}^{\mathrm{T}} \\
\mathbf{E}_{m}
\end{array}\right] \boldsymbol{\psi}(t)
$$

which means that the true reaction system can always be recovered by using

$$
\begin{equation*}
\boldsymbol{\psi}(t)=\left(\mathbf{E}_{m}^{\mathrm{T}} \mathbf{E}_{m}\right)^{-1} \mathbf{E}_{m}^{\mathrm{T}} \mathbf{r}(t) \tag{20}
\end{equation*}
$$

The previous remarks and the fact that $\boldsymbol{\psi}_{c}(\mathbf{c}(t))$ are reaction rates related to the BFMs in moles per unit of volume per unit of time imply that the model (13) can be equivalently written as

$$
\begin{equation*}
\dot{\mathbf{c}}(t)=\mathbf{N}_{e c}^{\mathrm{T}} \mathbf{E}_{m} \boldsymbol{\psi}_{c}(\mathbf{c}(t))+\mathbf{C}_{i n}(t) \boldsymbol{\omega}_{i n}(t)+\omega_{h}(t) \mathbf{R}(t) \mathbf{c}(t)-\omega_{p}(t) \mathbf{c}(t), \tag{21}
\end{equation*}
$$

Table 4: List of BFMs in the illustrative example.

| BFM | Chemical equation |
| :---: | :---: |
| 1 | 0.5 Glcext $^{\text {e }}$ ¢ Lacext |
| 2 | $2 \mathrm{H} 2 \mathrm{O}_{\text {ext }}+0.5 \mathrm{Gln} \mathrm{ext}^{\text {e }}$ ( $0.5 \mathrm{O} 2 \mathrm{ext}^{\text {ext }}+0.41667 \mathrm{Glc}_{\text {ext }}+\mathrm{NH} 4_{\text {ext }}$ |
| 3 | 0.5 H 2 O ext $+0.083333 \mathrm{Glc}_{\text {ext }}+0.5 \mathrm{Gln}_{\text {ext }} \leftrightarrow 0.25 \mathrm{O} 2_{\text {ext }}+\mathrm{Ala}$ ext |
| 4 | $\mathrm{O} 22_{\text {ext }}+0.16667 \mathrm{Glc}_{\text {ext }} \leftrightarrow \mathrm{H} 2 \mathrm{O}_{\text {ext }}+\mathrm{CO} 2_{\text {ext }}$ |
| 5 | $3 \mathrm{H} 3 \mathrm{PO} 4{ }_{\text {ext }}+0.25 \mathrm{O} 2_{\text {ext }}+1.1667 \mathrm{Glc}_{\text {ext }}+2 \mathrm{Gln}_{\text {ext }} \leftrightarrow 6.5 \mathrm{H} 2 \mathrm{O}_{\text {ext }}+\mathrm{Nuc}_{\text {ext }}$ |

while $\boldsymbol{\psi}(t)=\frac{\boldsymbol{\psi}_{c}(\mathbf{c}(t))}{V C(t)}$ are specific reaction rates related to the BFMs in moles per cell per unit of time, expressed without the explicit dependence on the concentrations of the species in the bioreactor.

In the case of biomass, it is expected that $R_{\text {biom }}(t)=1$ and $\mathbf{c}_{\text {in,biom }}(t)=\mathbf{0}_{p}^{\mathrm{T}}$, which implies that

$$
\begin{align*}
\dot{c}_{\text {biom }}(t) & =\mathbf{n}_{\text {biom }}^{\mathrm{T}} \mathbf{E}_{m} \boldsymbol{\psi}_{c}(\mathbf{c}(t))+\mathbf{c}_{\text {in }, \text { biom }}(t) \boldsymbol{\omega}_{i n}(t)+\omega_{h}(t) R_{\text {biom }}(t) c_{b i o m}(t)-\omega_{p}(t) c_{\text {biom }}(t) \\
& =\mathbf{n}_{\text {biom }}^{\mathrm{T}} \mathbf{E}_{m} \boldsymbol{\psi}_{c}(\mathbf{c}(t))-\omega_{b}(t) c_{b i o m}(t) . \tag{22}
\end{align*}
$$

Note that the connection between stoichiometric and atomic matrices and balances for species and elements has been investigated in the literature for many years [23]. However, these concepts have typically been used without any distinction between groups of species, such as intracellular or extracellular. Nevertheless, the use of knowledge about the stoichiometry to detect structural invariants in biological reaction networks has been documented [24]. Furthermore, the use of knowledge about the space that is spanned by the stoichiometries of a reaction system to identify these stoichiometries has been proposed previously in slightly different contexts. For example, target factor analysis has been used to identify target stoichiometries based on the fact that they are valid if and only if they belong to the row space of a data matrix and the null space of an atomic matrix $[25,26]$. The main difference in this paper is that the prior knowledge of a reaction network is used instead of data to identify the stoichiometries of BFMs.

### 2.4.1. Illustrative example

The illustrative example in Section 2.2 .1 is considered again to show how the corresponding BFMs can be computed. Out of the $S=18$ species, $S_{e c}=10$ are extracellular and $S_{i c}=8$ are intracellular. The species are ordered as follows: extracellular, intracellular. The matrix $\mathbf{A}$ results in the $S_{e c} \times E_{d}$ matrix $\mathbf{S}_{e c} \mathbf{A}$ of rank $E_{e c}=5$, while the matrix $\mathbf{A}_{N}$ results in the $S_{e c} \times q$ matrix $\mathbf{S}_{e c} \mathbf{A}_{N}$ of rank $q_{e c}=5$. One can then execute the MATLAB command Nm=null ([AN';Sic] ,'r')'; to compute the $R_{m} \times S$ matrix

$$
\mathbf{N}_{m}=\left[\begin{array}{rrrrrrrrrrrrrrrrrrrrrrrrrrrrrr}
0 & 0 & 0 & -\frac{1}{2} & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0  \tag{23}\\
0 & -2 & \frac{1}{2} & \frac{5}{12} & -\frac{1}{2} & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -\frac{1}{2} & \frac{1}{4} & -\frac{1}{12} & -\frac{1}{2} & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & -\frac{1}{6} & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-3 & \frac{13}{2} & -\frac{1}{4} & -\frac{7}{6} & -2 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right]
$$

of rank $R_{m}=5$ with the $R_{m}$ linearly independent stoichiometries of the BFMs in Table 4 and the $R \times R_{m}$ matrix $\mathbf{E}_{m}$ of rank $R_{m}$.

### 2.5. Estimating the reaction rates from measurements

Now assume that measurements $\tilde{\mathbf{q}}_{a}(t)$ of the rates of variation $\mathbf{S}_{a} \mathbf{q}(t)$ of $S_{a}$ extracellular species due to reactions are available and corrupted by zero-mean noise. Also, assume that the covariance
matrix of the noise that corrupts $\tilde{\mathbf{q}}_{a}(t)$ is known and given by the $S_{a} \times S_{a}$ matrix $\boldsymbol{\Sigma}_{\tilde{\mathbf{q}}_{a}}{ }^{2}$
In the context of MFA, we would like to compute simultaneously (i) estimates $\hat{\mathbf{q}}$ of the rates of variation of all the $S$ species due to reactions such that they fit the measurements in the sense of minimization of the sum of weighted squared errors, (ii) estimates $\hat{\mathbf{r}}$ of the $R$ rates of the independent reactions, and (iii) estimates $\hat{\mathbf{r}}_{d}$ of the original $R_{d}$ rates of reaction. These estimates should be related as follows:

$$
\begin{align*}
\hat{\mathbf{q}} & =\mathbf{N}^{\mathrm{T}} \hat{\mathbf{r}}  \tag{24}\\
\hat{\mathbf{r}} & =\mathbf{L}_{N}^{\mathrm{T}} \hat{\mathbf{r}}_{d} \tag{25}
\end{align*}
$$

It is also known that a subset of the original $R_{d}$ rates of reaction given by the $R_{i r} \times R_{d}$ selection matrix $\mathbf{S}_{i r}$ is nonnegative, since this subset of $R_{i r}$ reactions is assumed to be irreversible. Furthermore, recall that there is a pseudo steady-state assumption for the intracellular species, which correspond to the stoichiometries given by the matrix $\mathbf{N}_{i c}$.

These considerations lead to the formulation of the following optimization problem with the estimated rates of variation of the species $\hat{\mathbf{q}}$ and the estimated rates of reaction $\hat{\mathbf{r}}$ and $\hat{\mathbf{r}}_{d}$ as decision variables:

$$
\begin{align*}
\min _{\hat{\mathbf{q}}, \hat{\mathbf{r}}, \hat{\mathbf{r}}_{d}} & \left(\mathbf{S}_{a} \hat{\mathbf{q}}-\tilde{\mathbf{q}}_{a}\right)^{\mathrm{T}} \boldsymbol{\Sigma}_{\tilde{\mathbf{q}}_{a}}^{-1}\left(\mathbf{S}_{a} \hat{\mathbf{q}}-\tilde{\mathbf{q}}_{a}\right),  \tag{26a}\\
\text { s.t. } & {\left[\begin{array}{cc}
\mathbf{I}_{S} & -\mathbf{N}^{\mathrm{T}} \\
\mathbf{0}_{S_{i c} \times S} & \mathbf{N}_{i c}^{\mathrm{T}}
\end{array}\right]\left[\begin{array}{c}
\hat{\mathbf{q}} \\
\hat{\mathbf{r}}
\end{array}\right]=\left[\begin{array}{c}
\mathbf{0}_{S} \\
\mathbf{0}_{S_{i c}}
\end{array}\right], }  \tag{26b}\\
& {\left[\mathbf{I}_{R}-\mathbf{L}_{N}^{\mathrm{T}}\right]\left[\begin{array}{c}
\hat{\mathbf{r}} \\
\hat{\mathbf{r}}_{d}
\end{array}\right]=\mathbf{0}_{R} }  \tag{26c}\\
& \mathbf{S}_{i r} \hat{\mathbf{r}}_{d} \geq \mathbf{0}_{R_{i r}} . \tag{26d}
\end{align*}
$$

The optimization problem above corresponds to a standard convex quadratic programming problem. Next, we analyze different formulations of problem (26) that are equivalent to it but result in alternative parameterizations with different interpretations. For this, we will often use the following known result: any optimization problem with decision variables $\mathbf{x}$ and constrained by the linear equality constraints $\mathbf{A x}=\mathbf{0}$ can be converted to an equivalent optimization problem, in which these constraints are discarded and $\mathbf{x}$ is replaced by $\mathbf{F} \boldsymbol{\psi}$, where $\boldsymbol{\psi}$ are the new decision variables and the columns of $\mathbf{F}$ are a basis of the null space of $\mathbf{A}[27]$.

### 2.5.1. Using EFMs

The first method presented here uses EFMs to solve problem (26). For this, one needs to change the problem such that it is fully expressed in terms of irreversible reactions. This is achieved by denoting the estimates of the nonnegative rates of these $R_{n}:=2 R_{d}-R_{i r}$ irreversible reactions as $\hat{\mathbf{r}}_{n}$, related to $\hat{\mathbf{r}}_{d}$ as follows:

$$
\hat{\mathbf{r}}_{d}=\left[\begin{array}{lll}
\mathbf{S}_{i r}^{\mathrm{T}} & \mathbf{S}_{r}^{\mathrm{T}} & -\mathbf{S}_{r}^{\mathrm{T}} \tag{27}
\end{array}\right] \hat{\mathbf{r}}_{n},
$$

[^1]where $\hat{\mathbf{r}}_{n} \geq \mathbf{0}_{R_{n}}$ and the $R_{r}:=R_{d}-R_{i r}$ columns of $\mathbf{S}_{r}^{\mathrm{T}}$ span the null space of $\mathbf{S}_{i r}$. By replacing $\hat{\mathbf{q}}$ by $\mathbf{N}^{\mathrm{T}} \hat{\mathbf{r}}, \hat{\mathbf{r}}$ by $\mathbf{L}_{N}^{\mathrm{T}} \hat{\mathbf{r}}_{d}$, and $\hat{\mathbf{r}}_{d}$ by $\left[\begin{array}{lll}\mathbf{S}_{i r}^{\mathrm{T}} & \mathbf{S}_{r}^{\mathrm{T}} & -\mathbf{S}_{r}^{\mathrm{T}}\end{array}\right] \hat{\mathbf{r}}_{n}$, one obtains the equivalent problem

$$
\begin{align*}
\min _{\hat{\mathbf{r}}_{n}} & \left(\mathbf { S } _ { a } \mathbf { N } ^ { \mathrm { T } } \mathbf { L } _ { N } ^ { \mathrm { T } } \left[\begin{array}{lll}
\mathbf{S}_{i r}^{\mathrm{T}} & \mathbf{S}_{r}^{\mathrm{T}} & \left.\left.-\mathbf{S}_{r}^{\mathrm{T}}\right] \hat{\mathbf{r}}_{n}-\tilde{\mathbf{q}}_{a}\right)^{\mathrm{T}} \boldsymbol{\Sigma}_{\tilde{\mathbf{q}}_{a}}^{-1}\left(\mathbf { S } _ { a } \mathbf { N } ^ { \mathrm { T } } \mathbf { L } _ { N } ^ { \mathrm { T } } \left[\begin{array}{lll}
\mathbf{S}_{i r}^{\mathrm{T}} & \mathbf{S}_{r}^{\mathrm{T}} & \left.\left.-\mathbf{S}_{r}^{\mathrm{T}}\right] \hat{\mathbf{r}}_{n}-\tilde{\mathbf{q}}_{a}\right) \\
\text { s.t. } & \mathbf{N}_{i c}^{\mathrm{T}} \mathbf{L}_{N}^{\mathrm{T}}\left[\begin{array}{llll}
\mathbf{S}_{i r}^{\mathrm{T}} & \mathbf{S}_{r}^{\mathrm{T}} & \left.-\mathbf{S}_{r}^{\mathrm{T}}\right] \hat{\mathbf{r}}_{n}=\mathbf{0}_{S_{i c}} \\
& \hat{\mathbf{r}}_{n} \geq \mathbf{0}_{R_{n}}
\end{array}\right.
\end{array} .\right.\right.
\end{array} .\right.\right. \tag{28a}
\end{align*}
$$

Then, note that any $\hat{\mathbf{r}}_{n} \geq \mathbf{0}_{R_{n}}$ is such that $\mathbf{N}_{i c}^{\mathrm{T}} \mathbf{L}_{N}^{\mathrm{T}}\left[\begin{array}{lll}\mathbf{S}_{i r}^{\mathrm{T}} & \mathbf{S}_{r}^{\mathrm{T}} & -\mathbf{S}_{r}^{\mathrm{T}}\end{array}\right] \hat{\mathbf{r}}_{n}=\mathbf{0}_{S_{i c}}$ if and only if there exists an $R_{n} \times R_{m}^{n}$ matrix $\mathbf{E}_{m}^{n} \geq \mathbf{0}_{R_{n} \times R_{m}^{n}}$ such that each column of $\mathbf{E}_{m}^{n}$ is an extreme ray of the polyhedral cone defined by $\left\{\mathbf{e} \geq \mathbf{0}_{R_{n}}: \mathbf{N}_{i c}^{\mathrm{T}} \mathbf{L}_{N}^{\mathrm{T}}\left[\begin{array}{lll}\mathbf{S}_{i r}^{\mathrm{T}} & \mathbf{S}_{r}^{\mathrm{T}} & -\mathbf{S}_{r}^{\mathrm{T}}\end{array}\right] \mathbf{e}=\mathbf{0}_{S_{i c}}\right\}$ and

$$
\begin{equation*}
\hat{\mathbf{r}}_{n}=\mathbf{E}_{m}^{n} \hat{\boldsymbol{\psi}}_{n} \tag{29}
\end{equation*}
$$

for some $\hat{\boldsymbol{\psi}}_{n} \geq \mathbf{0}_{R_{m}^{n}}$, or in other words, $\hat{\mathbf{r}}_{n}$ can be expressed as a conical combination of the columns of $\mathbf{E}_{m}^{n}$ and is a ray of the polyhedral cone [6, 7].

This means that, upon computing a matrix $\mathbf{E}_{m}^{n}$ that satisfies the aforementioned requirements, the optimization problem can be further reformulated as

$$
\begin{align*}
\min _{\hat{\boldsymbol{\psi}}_{n}} & \left(\mathbf { S } _ { a } \mathbf { N } ^ { \mathrm { T } } \mathbf { L } _ { N } ^ { \mathrm { T } } \left[\begin{array}{lll}
\mathbf{S}_{i r}^{\mathrm{T}} & \mathbf{S}_{r}^{\mathrm{T}} & \left.\left.-\mathbf{S}_{r}^{\mathrm{T}}\right] \mathbf{E}_{m}^{n} \hat{\boldsymbol{\psi}}_{n}-\tilde{\mathbf{q}}_{a}\right)^{\mathrm{T}} \boldsymbol{\Sigma}_{\tilde{\mathbf{q}}_{a}}^{-1}\left(\mathbf { S } _ { a } \mathbf { N } ^ { \mathrm { T } } \mathbf { L } _ { N } ^ { \mathrm { T } } \left[\begin{array}{lll}
\mathbf{S}_{i r}^{\mathrm{T}} & \mathbf{S}_{r}^{\mathrm{T}} & \left.\left.-\mathbf{S}_{r}^{\mathrm{T}}\right] \mathbf{E}_{m}^{n} \hat{\boldsymbol{\psi}}_{n}-\tilde{\mathbf{q}}_{a}\right), \\
\text { s.t. } & \hat{\boldsymbol{\psi}}_{n} \geq \mathbf{0}_{R_{m}^{n}} .
\end{array}\right.\right.
\end{array},=\right.\right.\text { (30a) } \tag{30a}
\end{align*}
$$

In this particular case, one can observe that the columns of $\mathbf{E}_{m}^{n}$ represent the EFMs and the variables $\hat{\boldsymbol{\psi}}_{n}$ can be interpreted as rates of reaction related to the EFMs. Now, there are two possibilities for the computation of the EFMs in the columns of $\mathbf{E}_{m}^{n}$ :

1. The matrix $\mathbf{E}_{m}^{n}(t)$ is computed individually for each set of measurements $\tilde{\mathbf{q}}_{a}(t)$. This means that, if there exist different sets of measurements $\tilde{\mathbf{q}}_{a}(1), \ldots, \tilde{\mathbf{q}}_{a}(N)$ for each one of $N$ experimental conditions, there exist corresponding matrices $\mathbf{E}_{m}^{n}(1), \ldots, \mathbf{E}_{m}^{n}(N)$, where each matrix $\mathbf{E}_{m}^{n}(t)$ has a relatively small number of columns. This is what is done in the case of the column generation method, for example [12]. However, if $S_{a}$ is no larger than the difference $R_{n}-S_{i c}$ between the number of columns and rows of $\mathbf{N}_{i c}^{\mathrm{T}} \mathbf{L}_{N}^{\mathrm{T}}\left[\begin{array}{lll}\mathbf{S}_{i r}^{\mathrm{T}} & \mathbf{S}_{r}^{\mathrm{T}}-\mathbf{S}_{r}^{\mathrm{T}}\end{array}\right]$, each matrix $\mathbf{E}_{m}^{n}(t)$ may have up to $S_{a}$ columns [7]. To make sure that each one of the vectors $\hat{\mathbf{r}}_{n}(1), \ldots, \hat{\mathbf{r}}_{n}(N)$ is a conical combination of the columns of the same matrix $\mathbf{E}_{m}^{n}$, this matrix must correspond to the union of the columns of $\mathbf{E}_{m}^{n}(1), \ldots, \mathbf{E}_{m}^{n}(N)$. In the worst case, the number of columns of $\mathbf{E}_{m}^{n}$ may be equal to $N S_{a}$, which equals the number of individual measurements in $\tilde{\mathbf{q}}_{a}(1), \ldots, \tilde{\mathbf{q}}_{a}(N)$.
2. The matrix $\mathbf{E}_{m}^{n}$ is pre-computed such that it is valid for any set of measurements $\tilde{\mathbf{q}}_{a}(t)$. This means that the columns of $\mathbf{E}_{m}^{n}$ must correspond to all the EFMs for the considered reaction network to make sure that any vector $\hat{\mathbf{r}}_{n}(t)$ that satisfies the constraints in (28) is a conical combination of the columns of the same matrix $\mathbf{E}_{m}^{n}$. However, the number of EFMs grows exponentially with the size of the reaction network [7].
In both cases, if the same matrix $\mathbf{E}_{m}^{n}$ must be valid for many sets of measurements $\tilde{\mathbf{q}}_{a}(t)$, the number of columns of $\mathbf{E}_{m}^{n}$ and EFMs may become very large. This motivates the use of BFMs as described next.

### 2.5.2. Using BFMs with no active inequality constraints

In general, the solution to problem (26) may be determined by the inequality constraints, that is, the solution may be such that some inequality constraints are active. In that case, the solution cannot be expressed analytically and needs to be computed via optimization. However, let us analyze first what occurs when the solution is not determined by the inequality constraints, and let us also change the notation from $\hat{\mathbf{q}}, \hat{\mathbf{r}}$ to $\tilde{\mathbf{q}}, \tilde{\mathbf{r}}$ for this case. In this case, the optimization problem is simpler and can be expressed as

$$
\begin{array}{ll}
\min _{\tilde{\mathbf{q}}, \tilde{\mathbf{r}}} & \left(\mathbf{S}_{a} \tilde{\mathbf{q}}-\tilde{\mathbf{q}}_{a}\right)^{\mathrm{T}} \boldsymbol{\Sigma}_{\tilde{\mathbf{q}}_{a}}^{-1}\left(\mathbf{S}_{a} \tilde{\mathbf{q}}-\tilde{\mathbf{q}}_{a}\right) \\
\text { s.t. } & {\left[\begin{array}{cc}
\mathbf{I}_{S} & -\mathbf{N}^{\mathrm{T}} \\
\mathbf{0}_{S_{i c} \times S} & \mathbf{N}_{i c}^{\mathrm{T}}
\end{array}\right]\left[\begin{array}{c}
\tilde{\mathbf{q}} \\
\tilde{\mathbf{r}}
\end{array}\right]=\left[\begin{array}{c}
\mathbf{0}_{S} \\
\mathbf{0}_{S_{i c}}
\end{array}\right] .} \tag{31b}
\end{array}
$$

For the linear equality constraints (31b), the columns of $\left[\begin{array}{l}\mathbf{N}_{m}^{\mathrm{T}} \\ \mathbf{E}_{m}\end{array}\right]$ are a basis of the null space of $\left[\begin{array}{cc}\mathbf{I}_{S} & -\mathbf{N}^{\mathrm{T}} \\ \mathbf{0}_{S_{i c} \times S} & \mathbf{N}_{i c}^{\mathrm{T}}\end{array}\right]$. Then, if these constraints are discarded, $\tilde{\mathbf{q}}$ is replaced by $\mathbf{N}_{m}^{\mathrm{T}} \tilde{\boldsymbol{\psi}}$, and $\tilde{\mathbf{r}}$ is replaced by $\mathbf{E}_{m} \tilde{\boldsymbol{\psi}}$, where $\tilde{\boldsymbol{\psi}}$ are the new decision variables, the constrained optimization problem (31) can be reformulated as the equivalent unconstrained problem

$$
\begin{equation*}
\min _{\tilde{\boldsymbol{\psi}}}\left(\mathbf{S}_{a} \mathbf{N}_{m}^{\mathrm{T}} \tilde{\boldsymbol{\psi}}-\tilde{\mathbf{q}}_{a}\right)^{\mathrm{T}} \boldsymbol{\Sigma}_{\tilde{\mathbf{q}}_{a}}^{-1}\left(\mathbf{S}_{a} \mathbf{N}_{m}^{\mathrm{T}} \tilde{\boldsymbol{\psi}}-\tilde{\mathbf{q}}_{a}\right) \tag{32}
\end{equation*}
$$

In this particular case, one can observe that the variables $\tilde{\psi}$ can be interpreted as rates of reaction related to the $R_{m}$ BFMs that correspond to the rows of $\mathbf{N}_{m}$. If the matrix $\mathbf{S}_{a} \mathbf{N}_{m}^{\mathrm{T}}$ is of full column rank (which is possible only if $S_{a} \geq R_{m}=S_{e c}-q_{e c}$ ), then the Hessian matrix $2 \mathbf{N}_{m} \mathbf{S}_{a}^{\mathrm{T}} \boldsymbol{\Sigma}_{\tilde{\mathbf{q}}_{a}}^{-1} \mathbf{S}_{a} \mathbf{N}_{m}^{\mathrm{T}}$ is positive definite and invertible for all $\tilde{\boldsymbol{\psi}}$ and the solution $\tilde{\boldsymbol{\psi}}^{*}$ is given by the stationarity condition $\mathbf{0}_{R_{m}}=2 \mathbf{N}_{m} \mathbf{S}_{a}^{\mathrm{T}} \boldsymbol{\Sigma}_{\tilde{\mathbf{q}}_{a}^{-1}} \mathbf{S}_{a} \mathbf{N}_{m}^{\mathrm{T}} \tilde{\boldsymbol{\psi}}^{*}-2 \mathbf{N}_{m} \mathbf{S}_{a}^{\mathrm{T}} \boldsymbol{\Sigma}_{\tilde{\mathbf{q}}_{a}}^{-1} \tilde{\mathbf{q}}_{a}$, thus

$$
\begin{equation*}
\tilde{\boldsymbol{\psi}}^{*}=\left(\mathbf{N}_{m} \mathbf{S}_{a}^{\mathrm{T}} \boldsymbol{\Sigma}_{\tilde{\mathbf{q}}_{a}}^{-1} \mathbf{S}_{a} \mathbf{N}_{m}^{\mathrm{T}}\right)^{-1} \mathbf{N}_{m} \mathbf{S}_{a}^{\mathrm{T}} \boldsymbol{\Sigma}_{\tilde{\mathbf{q}}_{a}}^{-1} \tilde{\mathbf{q}}_{a} \tag{33}
\end{equation*}
$$

and the solution to the original problem corresponds to the optimal estimated rates of variation of the species

$$
\begin{equation*}
\tilde{\mathbf{q}}^{*}=\mathbf{N}_{m}^{\mathrm{T}} \tilde{\boldsymbol{\psi}}^{*}=\mathbf{N}_{m}^{\mathrm{T}}\left(\mathbf{N}_{m} \mathbf{S}_{a}^{\mathrm{T}} \boldsymbol{\Sigma}_{\tilde{\mathbf{q}}_{a}}^{-1} \mathbf{S}_{a} \mathbf{N}_{m}^{\mathrm{T}}\right)^{-1} \mathbf{N}_{m} \mathbf{S}_{a}^{\mathrm{T}} \boldsymbol{\Sigma}_{\tilde{\mathbf{q}}_{a}}^{-1} \tilde{\mathbf{q}}_{a} \tag{34}
\end{equation*}
$$

and the optimal estimated rates of the independent reactions

$$
\begin{equation*}
\tilde{\mathbf{r}}^{*}=\mathbf{E}_{m} \tilde{\boldsymbol{\psi}}^{*}=\mathbf{E}_{m}\left(\mathbf{N}_{m} \mathbf{S}_{a}^{\mathrm{T}} \boldsymbol{\Sigma}_{\tilde{\mathbf{q}}_{a}}^{-1} \mathbf{S}_{a} \mathbf{N}_{m}^{\mathrm{T}}\right)^{-1} \mathbf{N}_{m} \mathbf{S}_{a}^{\mathrm{T}} \boldsymbol{\Sigma}_{\tilde{\mathbf{q}}_{a}}^{-1} \tilde{\mathbf{q}}_{a} \tag{35}
\end{equation*}
$$

which satisfy the original linear equality constraints (31b).
Note that $\tilde{\boldsymbol{\psi}}^{*}, \tilde{\mathbf{q}}^{*}$, and $\tilde{\mathbf{r}}^{*}$ result from a linear transformation of $\tilde{\mathbf{q}}_{a}$ that takes into account the weighting matrix $\boldsymbol{\Sigma}_{\tilde{\mathbf{q}}_{a}}^{-1}$, and $\tilde{\boldsymbol{\psi}}^{*}$ determine unique estimates for $\tilde{\mathbf{q}}^{*}$ and $\tilde{\mathbf{r}}^{*}$. Moreover, if the measured data $\tilde{\mathbf{q}}_{a}$ are such that $\tilde{\mathbf{q}}_{a}=\mathbf{S}_{a} \mathbf{N}_{m}^{\mathrm{T}} \boldsymbol{\psi}$ for some $\boldsymbol{\psi}$, then $\tilde{\boldsymbol{\psi}}^{*}=\boldsymbol{\psi}, \mathbf{S}_{a} \tilde{\mathbf{q}}^{*}=\mathbf{S}_{a} \mathbf{N}_{m}^{\mathrm{T}} \boldsymbol{\psi}=\tilde{\mathbf{q}}_{a}$, and $\tilde{\mathbf{r}}^{*}=\mathbf{E}_{m} \boldsymbol{\psi}$.

### 2.5.3. Using BFMs with active inequality constraints

Now we recover the original optimization problem (26) with inequality constraints and assume that the solution to this problem is determined by the inequality constraints. However, we express it in terms of $\hat{\boldsymbol{\psi}}$ instead of $\hat{\mathbf{q}}, \hat{\mathbf{r}}$, since $\hat{\mathbf{q}}=\mathbf{N}_{m}^{\mathrm{T}} \hat{\boldsymbol{\psi}}=\mathbf{N}^{\mathrm{T}} \mathbf{E}_{m} \hat{\boldsymbol{\psi}}$ and $\hat{\mathbf{r}}=\mathbf{E}_{m} \hat{\boldsymbol{\psi}}$. This results in the following optimization problem:

$$
\begin{array}{ll}
\min _{\hat{\boldsymbol{\psi}}, \hat{\mathbf{r}}_{d}} & \left(\mathbf{S}_{a} \mathbf{N}_{m}^{\mathrm{T}} \hat{\boldsymbol{\psi}}-\tilde{\mathbf{q}}_{a}\right)^{\mathrm{T}} \boldsymbol{\Sigma}_{\tilde{\mathbf{q}}_{a}}^{-1}\left(\mathbf{S}_{a} \mathbf{N}_{m}^{\mathrm{T}} \hat{\boldsymbol{\psi}}-\tilde{\mathbf{q}}_{a}\right), \\
\text { s.t. } & {\left[\mathbf{E}_{m}\right.} \\
& \left.-\mathbf{L}_{N}^{\mathrm{T}}\right]\left[\begin{array}{c}
\hat{\boldsymbol{\psi}} \\
\hat{\mathbf{r}}_{d}
\end{array}\right]=\mathbf{0}_{R}  \tag{36c}\\
& \mathbf{S}_{i r} \hat{\mathbf{r}}_{d} \geq \mathbf{0}_{R_{i r}}
\end{array}
$$

Note that $\left[\begin{array}{ll}\mathbf{E}_{m} & -\mathbf{L}_{N}^{\mathrm{T}}\end{array}\right]$ is an $R \times\left(R_{m}+R_{d}\right)$ matrix of rank $R$, which implies that its null space is of dimension $R_{m}+R_{k}$, where $R_{k}:=R_{d}-R$. For the linear equality constraints (36b), the columns of $\left[\begin{array}{cc}\mathbf{I}_{R_{m}} & \left.\begin{array}{c}\mathbf{0}_{R_{m} \times R_{k}} \\ \mathbf{L}_{N}\left(\mathbf{L}_{N}^{\mathrm{T}} \mathbf{L}_{N}\right)^{-1} \mathbf{E}_{m} \\ \mathbf{K}_{N}\end{array}\right] \text { are a basis of the null space of }\left[\begin{array}{ll}\mathbf{E}_{m} & -\mathbf{L}_{N}^{\mathrm{T}}\end{array}\right] \text {, where } \mathbf{K}_{N} \text { is a matrix of }\end{array}\right.$ rank $R_{k}$ with columns that span the null space of $\mathbf{N}_{d}^{\mathrm{T}}$. Then, if these constraints are discarded and $\hat{\mathbf{r}}_{d}$ is replaced by $\mathbf{L}_{N}\left(\mathbf{L}_{N}^{\mathrm{T}} \mathbf{L}_{N}\right)^{-1} \mathbf{E}_{m} \hat{\boldsymbol{\psi}}+\mathbf{K}_{N} \hat{\mathbf{r}}_{r}$, where $\hat{\boldsymbol{\psi}}$ and $\hat{\mathbf{r}}_{r}$ are the new decision variables, the optimization problem with equality constraints (36) can be reformulated as the equivalent problem without equality constraints

$$
\begin{align*}
\min _{\hat{\boldsymbol{\psi}}, \hat{\mathbf{r}}_{r}} & \left(\mathbf{S}_{a} \mathbf{N}_{m}^{\mathrm{T}} \hat{\boldsymbol{\psi}}-\tilde{\mathbf{q}}_{a}\right)^{\mathrm{T}} \boldsymbol{\Sigma}_{\tilde{\mathbf{q}}_{a}}^{-1}\left(\mathbf{S}_{a} \mathbf{N}_{m}^{\mathrm{T}} \hat{\boldsymbol{\psi}}-\tilde{\mathbf{q}}_{a}\right),  \tag{37a}\\
\text { s.t. } & \mathbf{S}_{i r}\left[\mathbf{L}_{N}\left(\mathbf{L}_{N}^{\mathrm{T}} \mathbf{L}_{N}\right)^{-1} \mathbf{E}_{m} \quad \mathbf{K}_{N}\right]\left[\begin{array}{c}
\hat{\boldsymbol{\psi}} \\
\hat{\mathbf{r}}_{r}
\end{array}\right] \geq \mathbf{0}_{R_{i r}} . \tag{37b}
\end{align*}
$$

Again, the variables $\hat{\boldsymbol{\psi}}$ can be interpreted as rates of reaction related to the $R_{m}$ BFMs that correspond to the rows of $\mathbf{N}_{m}$. From the solution $\hat{\boldsymbol{\psi}}^{*}, \hat{\mathbf{r}}_{r}^{*}$ to this optimization problem, one can then compute $\hat{\mathbf{q}}^{*}=\mathbf{N}_{m}^{\mathrm{T}} \hat{\boldsymbol{\psi}}^{*}, \hat{\mathbf{r}}^{*}=\mathbf{E}_{m} \hat{\boldsymbol{\psi}}^{*}$, and $\hat{\mathbf{r}}_{d}^{*}=\mathbf{L}_{N}\left(\mathbf{L}_{N}^{\mathrm{T}} \mathbf{L}_{N}\right)^{-1} \hat{\mathbf{r}}^{*}+\mathbf{K}_{N} \hat{\mathbf{r}}_{r}^{*}$.

Obviously, any matrix with $R_{m}+R_{k}$ columns that span the null space of $\left[\begin{array}{ll}\mathbf{E}_{m} & \left.-\mathbf{L}_{N}^{\mathrm{T}}\right]\end{array}\right]$ can be used in lieu of $\left[\begin{array}{cc}\mathbf{I}_{R_{m}} & \mathbf{0}_{R_{m} \times R_{k}} \\ \mathbf{L}_{N}\left(\mathbf{L}_{N}^{\mathrm{T}} \mathbf{L}_{N}\right)^{-1} \mathbf{E}_{m} & \mathbf{K}_{N}\end{array}\right]$, for example a nonnegative matrix $\left[\begin{array}{l}\mathbf{M}_{r} \\ \mathbf{E}_{m}\end{array}\right]$. Hence, an alternative reformulation of the optimization problem with equality constraints (36) as an equivalent problem without equality constraints is

$$
\begin{align*}
\min _{\hat{\boldsymbol{\psi}}_{r}} & \left(\mathbf{S}_{a} \mathbf{N}_{m}^{\mathrm{T}} \mathbf{M}_{r} \hat{\boldsymbol{\psi}}_{r}-\tilde{\mathbf{q}}_{a}\right)^{\mathrm{T}} \boldsymbol{\Sigma}_{\tilde{\mathbf{q}}_{a}}^{-1}\left(\mathbf{S}_{a} \mathbf{N}_{m}^{\mathrm{T}} \mathbf{M}_{r} \hat{\boldsymbol{\psi}}_{r}-\tilde{\mathbf{q}}_{a}\right),  \tag{38a}\\
\text { s.t. } & \mathbf{S}_{i r} \mathbf{E}_{m}^{r} \hat{\boldsymbol{\psi}}_{r} \geq \mathbf{0}_{R_{i r}} \tag{38b}
\end{align*}
$$

From the solution $\hat{\boldsymbol{\psi}}_{r}^{*}$ to this optimization problem, one can then compute $\hat{\mathbf{q}}^{*}=\mathbf{N}_{m}^{\mathrm{T}} \mathbf{M}_{r} \hat{\boldsymbol{\psi}}_{r}^{*}$, $\hat{\mathbf{r}}^{*}=\mathbf{E}_{m} \mathbf{M}_{r} \hat{\boldsymbol{\psi}}_{r}^{*}$, and $\hat{\mathbf{r}}_{d}^{*}=\mathbf{E}_{m}^{r} \hat{\boldsymbol{\psi}}_{r}^{*}$.

### 2.6. Example of a perfusion bioreactor

Consider the example of a biological reaction system that consists in a perfusion bioreactor with Chinese hamster ovary ( CHO ) cells that produce monoclonal antibodies. In this system, the
$S=128$ species in Table 5 participate in the $R_{d}=131$ reactions in Table 6 with the linearly dependent stoichiometries in the $R_{d} \times S$ matrix $\mathbf{N}_{d}$ of rank $R=117$. This reaction network corresponds to a slight amendment of the network proposed by [12] that was made to ensure that all the reactions satisfy the conservation of the $E_{d}=6$ quantities that correspond to the atoms of C , H, N, O, P, S. One can observe that this is a more complex version of the network in Sections 2.2.1 and 2.4.1. Due to the large size of the reaction network, it becomes rather challenging to compute all the EFMs, thus it is beneficial to use the concept of BFMs. Out of the $S$ species, $S_{e c}=33$ are extracellular and $S_{i c}=95$ are intracellular. Out of the $S_{e c}$ extracellular species, $S_{a}=25$ species are measured. Out of the $R_{d}$ reactions, $R_{i r}=74$ are irreversible. The species are ordered as follows: measured extracellular, unmeasured extracellular, intracellular. The measured extracellular species are ordered as follows: biomass, glucose, 20 amino acids (ordered alphabetically), ammonium, monoclonal antibodies, lactate. The conserved quantities are given by the $S \times E_{d}$ matrix $\mathbf{A}$ of rank $E=6$ that results in the $S_{e c} \times E_{d}$ matrix $\mathbf{S}_{e c} \mathbf{A}$ of rank $E_{e c}=6$.

One can then use the method described in M1 of Section 2.2 to compute the $S \times q$ matrix $\mathbf{A}_{N}$ of $\operatorname{rank} q=11$ that results in the $S_{e c} \times q$ matrix $\mathbf{S}_{e c} \mathbf{A}_{N}$ of rank $q_{e c}=8$, the $R_{d} \times R_{k}$ matrix $\mathbf{K}_{N}$ of rank $R_{k}=14$, the $R_{d} \times R$ matrix $\mathbf{L}_{N}$ of rank $R$, and the $R \times S$ matrix $\mathbf{N}$ of rank $R$ with the $R$ linearly independent stoichiometries of the reactions in Table 7.

In addition, one can then compute the $R_{m} \times S$ matrix $\mathbf{N}_{m}$ of rank $R_{m}=25$ with the $R_{m}$ linearly independent stoichiometries of the BFMs and the $R \times R_{m}$ matrix $\mathbf{E}_{m}$ of rank $R_{m}$. Interestingly, $S_{a}=R_{m}=25$ in this case, that is, the number of measured species is equal to the number of BFMs. Recall that, when the solution to the optimization problem (26) is not determined by the inequality constraints, the condition for the unique computation of $R_{m}$ estimates $\hat{\boldsymbol{\psi}}$ of the reaction rates related to the BFM from the $S_{a}$ measurements of the rates of variation $\tilde{\mathbf{q}}_{a}$ of the measured extracellular species is $S_{a} \geq R_{m}$, which means that the condition is marginally satisfied.

Regarding the matrix $\mathbf{N}_{m}$ of independent stoichiometries of the BFMs, we present below two different options that can be chosen among others. Then, once a matrix $\mathbf{N}_{m}$ is chosen, the resulting computation of reaction rates is unique. These two options for $\mathbf{N}_{m}$ are as follows:

1. Initially a matrix $\mathbf{N}_{m}$ is obtained such that $\mathbf{S}_{a} \mathbf{N}_{m}^{\mathrm{T}}=-\mathbf{I}_{R_{m}}$. All the resulting columns of $\mathbf{E}_{m}$ correspond to admissible BFMs, in the sense that each one satisfies the pseudo steady-state assumption for the intracellular species and the irreversibility constraints. The only exceptions are: (i) the columns that correspond to biomass and monoclonal antibodies, which need to be multiplied by -1 (to indicate production instead of consumption) and added to a conical combination of the columns that correspond to the 9 essential amino acids (and the column that corresponds to glucose in the case of biomass production) to yield admissible BFMs; and (ii) the column that corresponds to lactose, which needs to be multiplied by -1 and added to the column that corresponds to glucose times the appropriate positive coefficient (0.5). The resulting BFMs are listed in Table 8.
2. The EFMs that were identified in at least 4 metabolic states in the work by [12] (except the ones that correspond to production of biomass and monoclonal antibodies) are used as BFMs. This amounts to 20 BFMs . Since none of these BFMs involves the essential amino acids lysine, threonine, and histidine, two more BFMs that consist in consumption of lysine and threonine are added, where these BFMs correspond to the most frequent EFMs that involved these two amino acids in the work by [12]. Since none of the EFMs in this previous work involved the amino acid histidine, the three remaining BFMs that are added to the set of 25 BFMs consist in two BFMs for production of biomass and one BFM for production of monoclonal

Table 5: List of species in the perfusion bioreactor under study.

| Species | Label | Chemical formula | Species | Label | Chemical formula |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | Biomass $_{\text {ext }}$ | $\mathrm{C}_{0.9611} \mathrm{H}_{1.8716} \mathrm{~N}_{0.2295}$ | 65 | Leu | $\mathrm{C}_{6} \mathrm{H}_{13} \mathrm{NO}_{2}$ |
|  |  | $\mathrm{O}_{0.5488} \mathrm{P}_{0.0189} \mathrm{~S}_{0.0052}$ | 66 | Lys | $\mathrm{C}_{6} \mathrm{H}_{14} \mathrm{~N}_{2} \mathrm{O}_{2}$ |
| 2 | $\mathrm{Glc}_{\text {ext }}$ | $\mathrm{C}_{6} \mathrm{H}_{12} \mathrm{O}_{6}$ | 67 | Met | $\mathrm{C}_{5} \mathrm{H}_{11} \mathrm{NO}_{2} \mathrm{~S}$ |
| 3 | Ala ${ }_{\text {ext }}$ | $\mathrm{C}_{3} \mathrm{H}_{7} \mathrm{NO}_{2}$ | 68 | Phe | $\mathrm{C}_{9} \mathrm{H}_{11} \mathrm{NO}_{2}$ |
| 4 | Argext | $\mathrm{C}_{6} \mathrm{H}_{14} \mathrm{~N}_{4} \mathrm{O}_{2}$ | 69 | Pro | $\mathrm{C}_{5} \mathrm{H}_{9} \mathrm{NO}_{2}$ |
| 5 | $\mathrm{Asn}_{\text {ext }}$ | $\mathrm{C}_{4} \mathrm{H}_{8} \mathrm{~N}_{2} \mathrm{O}_{3}$ | 70 | Thr | $\mathrm{C}_{4} \mathrm{H}_{9} \mathrm{NO}_{3}$ |
| 6 | Aspext | $\mathrm{C}_{4} \mathrm{H}_{7} \mathrm{NO}_{4}$ | 71 | Trp | $\mathrm{C}_{11} \mathrm{H}_{12} \mathrm{~N}_{2} \mathrm{O}_{2}$ |
| 7 | Cysext | $\mathrm{C}_{3} \mathrm{H}_{7} \mathrm{NO}_{2} \mathrm{~S}$ | 72 | Val | $\mathrm{C}_{5} \mathrm{H}_{11} \mathrm{NO}_{2}$ |
| 8 | Glnext | $\mathrm{C}_{5} \mathrm{H}_{10} \mathrm{~N}_{2} \mathrm{O}_{3}$ | 73 | aKbut | $\mathrm{C}_{4} \mathrm{H}_{6} \mathrm{O}_{3}$ |
| 9 | $\mathrm{Glu}_{\text {ext }}$ | $\mathrm{C}_{5} \mathrm{H}_{9} \mathrm{NO}_{4}$ | 74 | PropCoA | $\mathrm{C}_{24} \mathrm{H}_{40} \mathrm{~N}_{7} \mathrm{O}_{17} \mathrm{P}_{3} \mathrm{~S}$ |
| 10 | Gly ext | $\mathrm{C}_{2} \mathrm{H}_{5} \mathrm{NO}_{2}$ | 75 | aKad | $\mathrm{C}_{6} \mathrm{H}_{8} \mathrm{O}_{5}$ |
| 11 | $\mathrm{His}_{\text {ext }}$ | $\mathrm{C}_{6} \mathrm{H}_{9} \mathrm{~N}_{3} \mathrm{O}_{2}$ | 76 | AcetoacCoA | $\mathrm{C}_{25} \mathrm{H}_{40} \mathrm{~N}_{7} \mathrm{O}_{18} \mathrm{P}_{3} \mathrm{~S}$ |
| 12 | Il ${ }_{\text {ext }}$ | $\mathrm{C}_{6} \mathrm{H}_{13} \mathrm{NO}_{2}$ | 77 | Acetoac | $\mathrm{C}_{4} \mathrm{H}_{6} \mathrm{O}_{3}$ |
| 13 | Leuext | $\mathrm{C}_{6} \mathrm{H}_{13} \mathrm{NO}_{2}$ | 78 | GluySA | $\mathrm{C}_{5} \mathrm{H}_{9} \mathrm{NO}_{3}$ |
| 14 | Lysext | $\mathrm{C}_{6} \mathrm{H}_{14} \mathrm{~N}_{2} \mathrm{O}_{2}$ | 79 | Orn | $\mathrm{C}_{5} \mathrm{H}_{12} \mathrm{~N}_{2} \mathrm{O}_{2}$ |
| 15 | $\mathrm{Met}_{\text {ext }}$ | $\mathrm{C}_{5} \mathrm{H}_{11} \mathrm{NO}_{2} \mathrm{~S}$ | 80 | Cln | $\mathrm{C}_{6} \mathrm{H}_{13} \mathrm{~N}_{3} \mathrm{O}_{3}$ |
| 16 | Pheext | $\mathrm{C}_{9} \mathrm{H}_{11} \mathrm{NO}_{2}$ | 81 | Argsucc | $\mathrm{C}_{10} \mathrm{H}_{18} \mathrm{~N}_{4} \mathrm{O}_{6}$ |
| 17 | $\mathrm{Prog}_{\text {ext }}$ | $\mathrm{C}_{5} \mathrm{H}_{9} \mathrm{NO}_{2}$ | 82 | Urea | $\mathrm{CH}_{4} \mathrm{~N}_{2} \mathrm{O}$ |
| 18 | $\mathrm{Ser}_{\text {ext }}$ | $\mathrm{C}_{3} \mathrm{H}_{7} \mathrm{NO}_{3}$ | 83 | PRPP | $\mathrm{C}_{5} \mathrm{H}_{13} \mathrm{O}_{14} \mathrm{P}_{3}$ |
| 19 | Thrext | $\mathrm{C}_{4} \mathrm{H}_{9} \mathrm{NO}_{3}$ | 84 | IMP | $\mathrm{C}_{10} \mathrm{H}_{13} \mathrm{~N}_{4} \mathrm{O}_{8} \mathrm{P}$ |
| 20 | Trpext | $\mathrm{C}_{11} \mathrm{H}_{12} \mathrm{~N}_{2} \mathrm{O}_{2}$ | 85 | Orot | $\mathrm{C}_{5} \mathrm{H}_{4} \mathrm{~N}_{2} \mathrm{O}_{4}$ |
| 21 | Tyrext | $\mathrm{C}_{9} \mathrm{H}_{11} \mathrm{NO}_{3}$ | 86 | UTPrn | $\mathrm{C}_{9} \mathrm{H}_{15} \mathrm{~N}_{2} \mathrm{O}_{15} \mathrm{P}_{3}$ |
| 22 | Valext | $\mathrm{C}_{5} \mathrm{H}_{11} \mathrm{NO}_{2}$ | 87 | ATPrn | $\mathrm{C}_{10} \mathrm{H}_{16} \mathrm{~N}_{5} \mathrm{O}_{13} \mathrm{P}_{3}$ |
| 23 | NH4 ext | $\mathrm{NH}_{4}$ | 88 | GTPrn | $\mathrm{C}_{10} \mathrm{H}_{16} \mathrm{~N}_{5} \mathrm{O}_{14} \mathrm{P}_{3}$ |
| 24 | $m A b_{\text {ext }}$ | $\mathrm{CH}_{1.9584} \mathrm{~N}_{0.2654} \mathrm{O}_{0.5135} \mathrm{~S}_{0.0073}$ | 89 | CTPrn | $\mathrm{C}_{9} \mathrm{H}_{16} \mathrm{~N}_{3} \mathrm{O}_{14} \mathrm{P}_{3}$ |
| 25 | $\mathrm{Lac}_{\text {ext }}$ | $\mathrm{C}_{3} \mathrm{H}_{6} \mathrm{O}_{3}$ | 90 | dATP | $\mathrm{C}_{10} \mathrm{H}_{16} \mathrm{~N}_{5} \mathrm{O}_{12} \mathrm{P}_{3}$ |
| 26 | CO 2 ext | $\mathrm{CO}_{2}$ | 91 | dGTP | $\mathrm{C}_{10} \mathrm{H}_{16} \mathrm{~N}_{5} \mathrm{O}_{13} \mathrm{P}_{3}$ |
| 27 | Ethnext | $\mathrm{C}_{2} \mathrm{H}_{7} \mathrm{NO}$ | 92 | dTTP | $\mathrm{C}_{10} \mathrm{H}_{17} \mathrm{~N}_{2} \mathrm{O}_{14} \mathrm{P}_{3}$ |
| 28 | Choext | $\mathrm{C}_{5} \mathrm{H}_{14} \mathrm{NO}$ | 93 | dCTP | $\mathrm{C}_{9} \mathrm{H}_{16} \mathrm{~N}_{3} \mathrm{O}_{13} \mathrm{P}_{3}$ |
| 29 | Urea ext | $\mathrm{CH}_{4} \mathrm{~N}_{2} \mathrm{O}$ | 94 | Ethn | $\mathrm{C}_{2} \mathrm{H}_{7} \mathrm{NO}$ |
| 30 | $\mathrm{H} 3 \mathrm{PO} 4_{\text {ext }}$ | $\mathrm{H}_{3} \mathrm{PO}_{4}$ | 95 | Cho | $\mathrm{C}_{5} \mathrm{H}_{14} \mathrm{NO}$ |
| 31 | $\mathrm{H} 2 \mathrm{SO} 4_{\text {ext }}$ | $\mathrm{H}_{2} \mathrm{SO}_{4}$ | 96 | Glyc3P | $\mathrm{C}_{3} \mathrm{H}_{9} \mathrm{O}_{6} \mathrm{P}$ |
| 32 | O 2 ext $^{\text {ex }}$ | $\mathrm{O}_{2}$ | 97 | PhosphE | $\mathrm{C}_{41} \mathrm{H}_{80} \mathrm{NO}_{8} \mathrm{P}$ |
| 33 | $\mathrm{H}_{2} \mathrm{O}_{\text {ext }}$ | $\mathrm{H}_{2} \mathrm{O}$ | 98 | PhosphC | $\mathrm{C}_{44} \mathrm{H}_{84} \mathrm{NO}_{8} \mathrm{P}$ |
| 34 | G6P | $\mathrm{C}_{6} \mathrm{H}_{13} \mathrm{O}_{9} \mathrm{P}$ | 99 | PhosphS | $\mathrm{C}_{42} \mathrm{H}_{80} \mathrm{NO}_{10} \mathrm{P}$ |
| 35 | F6P | $\mathrm{C}_{6} \mathrm{H}_{13} \mathrm{O}_{9} \mathrm{P}$ | 100 | Sphm | $\mathrm{C}_{41} \mathrm{H}_{81} \mathrm{~N}_{2} \mathrm{O}_{6} \mathrm{P}$ |
| 36 | DHAP | $\mathrm{C}_{3} \mathrm{H}_{7} \mathrm{O}_{6} \mathrm{P}$ | 101 | Cholesterol | $\mathrm{C}_{27} \mathrm{H}_{46} \mathrm{O}$ |
| 37 | GA3P | $\mathrm{C}_{3} \mathrm{H}_{7} \mathrm{O}_{6} \mathrm{P}$ | 102 | Proteins | $\mathrm{C}_{4.8194} \mathrm{H}_{9.633} \mathrm{~N}_{1.3493} \mathrm{O}_{2.6254} \mathrm{~S}_{0.0337}$ |
| 38 | 3PG | $\mathrm{C}_{3} \mathrm{H}_{7} \mathrm{O}_{7} \mathrm{P}$ | 103 | DNA | $\mathrm{C}_{9.785} \mathrm{H}_{16.285} \mathrm{~N}_{3.715} \mathrm{O}_{13} \mathrm{P}_{3}$ |
| 39 | PEP | $\mathrm{C}_{3} \mathrm{H}_{5} \mathrm{O}_{6} \mathrm{P}$ | 104 | RNA | $\mathrm{C}_{9.5} \mathrm{H}_{15.715} \mathrm{~N}_{3.715} \mathrm{O}_{14} \mathrm{P}_{3}$ |
| 40 | Pyr | $\mathrm{C}_{3} \mathrm{H}_{4} \mathrm{O}_{3}$ | 105 | Lipids | $\mathrm{C}_{40.475} \mathrm{H}_{76.975} \mathrm{~N}_{0.925}$ |
| 41 | AccoA | $\mathrm{C}_{23} \mathrm{H}_{38} \mathrm{~N}_{7} \mathrm{O}_{17} \mathrm{P}_{3} \mathrm{~S}$ |  |  | $\mathrm{O}_{6.95} \mathrm{P}_{0.85}$ |
| 42 | Cit | $\mathrm{C}_{6} \mathrm{H}_{8} \mathrm{O}_{7}$ | 106 | TC | $\mathrm{C}_{6} \mathrm{H}_{10} \mathrm{O}_{5}$ |
| 43 | aKG | $\mathrm{C}_{5} \mathrm{H}_{6} \mathrm{O}_{5}$ | 107 | NH4 | $\mathrm{NH}_{4}$ |
| 44 | SucCoA | $\mathrm{C}_{25} \mathrm{H}_{40} \mathrm{~N}_{7} \mathrm{O}_{19} \mathrm{P}_{3} \mathrm{~S}$ | 108 | CO2 | $\mathrm{CO}_{2}$ |
| 45 | Suc | $\mathrm{C}_{4} \mathrm{H}_{6} \mathrm{O}_{4}$ | 109 | $\mathrm{H} 3 \mathrm{PO}^{4}$ | $\mathrm{H}_{3} \mathrm{PO}_{4}$ |
| 46 | Fum | $\mathrm{C}_{4} \mathrm{H}_{4} \mathrm{O}_{4}$ | 110 | H2O | $\mathrm{H}_{2} \mathrm{O}$ |
| 47 | Mal | $\mathrm{C}_{4} \mathrm{H}_{6} \mathrm{O}_{5}$ | 111 | CoA | $\mathrm{C}_{21} \mathrm{H}_{34} \mathrm{~N}_{7} \mathrm{O}_{16} \mathrm{P}_{3} \mathrm{~S}$ |
| 48 | Oxa | $\mathrm{C}_{4} \mathrm{H}_{4} \mathrm{O}_{5}$ | 112 | H2SO4 | $\mathrm{H}_{2} \mathrm{SO}_{4}$ |
| 49 | R15P | $\mathrm{C}_{5} \mathrm{H}_{11} \mathrm{O}_{8} \mathrm{P}$ | 113 | O2 | $\mathrm{O}_{2}$ |
| 50 | R5P | $\mathrm{C}_{5} \mathrm{H}_{11} \mathrm{O}_{8} \mathrm{P}$ | 114 | H | H |
| 51 | X5P | $\mathrm{C}_{5} \mathrm{H}_{11} \mathrm{O}_{8} \mathrm{P}$ | 115 | $\mathrm{Pyr}_{\mathrm{m}}$ | $\mathrm{C}_{3} \mathrm{H}_{4} \mathrm{O}_{3}$ |
| 52 | E4P | $\mathrm{C}_{4} \mathrm{H}_{9} \mathrm{O}_{7} \mathrm{P}$ | 116 | $\mathrm{AcCoA}_{\mathrm{m}}$ | $\mathrm{C}_{23} \mathrm{H}_{38} \mathrm{~N}_{7} \mathrm{O}_{17} \mathrm{P}_{3} \mathrm{~S}$ |
| 53 | Glu | $\mathrm{C}_{5} \mathrm{H}_{9} \mathrm{NO}_{4}$ | 117 | Oxam | $\mathrm{C}_{4} \mathrm{H}_{4} \mathrm{O}_{5}$ |
| 54 | Asp | $\mathrm{C}_{4} \mathrm{H}_{7} \mathrm{NO}_{4}$ | 118 | $\mathrm{Cit}_{\mathrm{m}}$ | $\mathrm{C}_{6} \mathrm{H}_{8} \mathrm{O}_{7}$ |
| 55 | Gly | $\mathrm{C}_{2} \mathrm{H}_{5} \mathrm{NO}_{2}$ | 119 | $\mathrm{aKG}_{\mathrm{m}}$ | $\mathrm{C}_{5} \mathrm{H}_{6} \mathrm{O}_{5}$ |
| 56 | Ser | $\mathrm{C}_{3} \mathrm{H}_{7} \mathrm{NO}_{3}$ | 120 | $\mathrm{SucCoA}_{\mathrm{m}}$ | $\mathrm{C}_{25} \mathrm{H}_{40} \mathrm{~N}_{7} \mathrm{O}_{19} \mathrm{P}_{3} \mathrm{~S}$ |
| 57 | Tyr | $\mathrm{C}_{9} \mathrm{H}_{11} \mathrm{NO}_{3}$ | 121 | $\mathrm{Suc}_{\mathrm{m}}$ | $\mathrm{C}_{4} \mathrm{H}_{6} \mathrm{O}_{4}$ |
| 58 | Cys | $\mathrm{C}_{3} \mathrm{H}_{7} \mathrm{NO}_{2} \mathrm{~S}$ | 122 | $\mathrm{Fum}_{\mathrm{m}}$ | $\mathrm{C}_{4} \mathrm{H}_{4} \mathrm{O}_{4}$ |
| 59 | Ala | $\mathrm{C}_{3} \mathrm{H}_{7} \mathrm{NO}_{2}$ | 123 | $\mathrm{Mal}_{\mathrm{m}}$ | $\mathrm{C}_{4} \mathrm{H}_{6} \mathrm{O}_{5}$ |
| 60 | Arg | $\mathrm{C}_{6} \mathrm{H}_{14} \mathrm{~N}_{4} \mathrm{O}_{2}$ | 124 | $\mathrm{Glu}_{\mathrm{m}}$ | $\mathrm{C}_{5} \mathrm{H}_{9} \mathrm{NO}_{4}$ |
| 61 | Asn | $\mathrm{C}_{4} \mathrm{H}_{8} \mathrm{~N}_{2} \mathrm{O}_{3}$ | 125 | $\mathrm{Alam}_{\mathrm{m}}$ | $\mathrm{C}_{3} \mathrm{H}_{7} \mathrm{NO}_{2}$ |
| 62 | Gln | $\mathrm{C}_{5} \mathrm{H}_{10} \mathrm{~N}_{2} \mathrm{O}_{3}$ | 126 | $\mathrm{Asp}_{\mathrm{m}}$ | $\mathrm{C}_{4} \mathrm{H}_{7} \mathrm{NO}_{4}$ |
| 63 | His | $\mathrm{C}_{6} \mathrm{H}_{9} \mathrm{~N}_{3} \mathrm{O}_{2}$ | 127 | $\mathrm{NH} 4{ }_{\mathrm{m}}$ | $\mathrm{NH}_{4}$ |
| 64 | Ile | $\mathrm{C}_{6} \mathrm{H}_{13} \mathrm{NO}_{2}$ | 128 | CO 2 m | $\mathrm{CO}_{2}$ |

Table 6: List of reactions with linearly dependent stoichiometries in the perfusion bioreactor under study.

| Reaction | Chemical equation $R$ | Reaction | Chemical equation |
| :---: | :---: | :---: | :---: |
| 1 | $\mathrm{Glc}_{\text {ext }}+\mathrm{H} 3 \mathrm{PO} 4 \rightarrow \mathrm{G6P}+\mathrm{H} 2 \mathrm{O} \longrightarrow 7$ | 73 | Gln + UTPrn $\rightarrow$ Glu + CTPrn |
| 2 | $\mathrm{G6P} \leftrightarrow \mathrm{F6P}$ | 74 | 0.285 UTPrn +0.285 ATPrn +0.215 GTPrn +0.215 CT- |
| 3 | $\mathrm{F} 6 \mathrm{P}+\mathrm{H} 3 \mathrm{PO} 4 \rightarrow \mathrm{DHAP}+\mathrm{GA} 3 \mathrm{P}+\mathrm{H} 2 \mathrm{O}$ |  | Prn $\rightarrow$ RNA |
| 4 | DHAP $\leftrightarrow$ GA3P 7 | 75 | ATPrn $+2 \mathrm{H} \rightarrow \mathrm{dATP}+\mathrm{H} 2 \mathrm{O}$ |
| 5 | $\mathrm{GA} 3 \mathrm{P}+\mathrm{H} 2 \mathrm{O} \leftrightarrow 3 \mathrm{PG}+2 \mathrm{H}$ | 76 | GTPrn $+2 \mathrm{H} \rightarrow$ dGTP +H 2 O |
| 6 | $3 \mathrm{PG} \leftrightarrow \mathrm{PEP}+\mathrm{H} 2 \mathrm{O}$ - 7 | 77 | $\mathrm{UTPrn}+\mathrm{CO} 2+8 \mathrm{H} \rightarrow \mathrm{dTTP}+3 \mathrm{H} 2 \mathrm{O}$ |
| 7 | $\mathrm{PEP}+\mathrm{H} 2 \mathrm{O} \rightarrow \mathrm{Pyr}+\mathrm{H} 3 \mathrm{PO} 4 \rightarrow 78$ | 78 | CTPrn $+2 \mathrm{H} \rightarrow \mathrm{dCTP}+\mathrm{H} 2 \mathrm{O}$ |
| 8 | Pyr $+2 \mathrm{H} \rightarrow$ Lacext ${ }^{\text {P }}$ ( ${ }^{\text {P }}$ | 79 | $0.285 \mathrm{dATP}+0.215 \mathrm{dGTP}+0.285 \mathrm{dTTP}+0.215 \mathrm{dCTP}$ |
| 9 | $\mathrm{G} 6 \mathrm{P}+\mathrm{H} 2 \mathrm{O} \rightarrow \mathrm{R} 15 \mathrm{P}+\mathrm{CO} 2+4 \mathrm{H}$ |  | $\rightarrow$ DNA |
| 10 | R15P $\leftrightarrow$ R5P 8 | 80 | DHAP $+2 \mathrm{H} \rightarrow$ Glyc3P |
| 11 | R15P $\leftrightarrow$ X5P 8 | 81 | $18 \mathrm{AcCoA}+\mathrm{Cho}+\mathrm{Glyc} 3 \mathrm{P}+23 \mathrm{H} \rightarrow$ PhosphC +17 H 2 O |
| 12 | $\mathrm{R} 5 \mathrm{P}+\mathrm{X} 5 \mathrm{P} \leftrightarrow \mathrm{F} 6 \mathrm{P}+\mathrm{E} 4 \mathrm{P}$ |  | $+18 \mathrm{CoA}$ |
| 13 | $\mathrm{X} 5 \mathrm{P}+\mathrm{E} 4 \mathrm{P} \leftrightarrow \mathrm{F} 6 \mathrm{P}+\mathrm{GA} 3 \mathrm{P}$ | 82 | 18 AcCoA + Ethn + Glyc3P + $26 \mathrm{H} \rightarrow$ PhosphE +17 H 2 O |
| 14 | $\mathrm{CoA}+\mathrm{Pyr}_{\mathrm{m}} \rightarrow \mathrm{AcCoA}_{\mathrm{m}}+\mathrm{CO} 2_{\mathrm{m}}$ |  | $+18 \mathrm{CoA}$ |
| 15 | $\mathrm{H} 2 \mathrm{O}+\mathrm{AcCoA}_{\mathrm{m}}+\mathrm{Oxam}_{\mathrm{m}} \rightarrow \mathrm{CoA}+2 \mathrm{H}+\mathrm{Cit}_{\mathrm{m}} \quad 83$ | 83 | Ser + PhosphE $\rightarrow$ Ethn + PhosphS |
| 16 | $\mathrm{Cit}_{\mathrm{m}} \rightarrow 2 \mathrm{H}+\mathrm{aKG} \mathrm{m}^{+\mathrm{CO} 2 \mathrm{~m}} 8$ | 84 | $16 \mathrm{AcCoA}+\mathrm{Ser}+\mathrm{Cho}+\mathrm{Glyc} 3 \mathrm{P}+19 \mathrm{H} \rightarrow \mathrm{Sphm}+2$ |
| 17 | $\mathrm{CoA}+\mathrm{aKG}_{\mathrm{m}} \rightarrow \mathrm{SucCoA} \mathrm{m}+\mathrm{CO} 2 \mathrm{~m}$ |  | $\mathrm{CO} 2+16 \mathrm{H} 2 \mathrm{O}+16 \mathrm{CoA}$ |
| 18 | $\mathrm{H} 2 \mathrm{O}+\mathrm{SucCoA}_{\mathrm{m}} \leftrightarrow \mathrm{CoA}+2 \mathrm{H}+\mathrm{Suc}_{\mathrm{m}} 8^{\text {a }}$ | 85 | $16 \mathrm{AcCoA} \rightarrow$ Cholesterol $+5 \mathrm{CO} 2+5 \mathrm{H} 2 \mathrm{O}+16 \mathrm{CoA}+8$ |
| 19 | $\mathrm{Suc}_{\mathrm{m}} \leftrightarrow 2 \mathrm{H}+\mathrm{Fum}_{\mathrm{m}}$ |  | H |
| 20 | $\mathrm{H} 2 \mathrm{O}+\mathrm{Fum}_{\mathrm{m}} \leftrightarrow \mathrm{Mal}_{\mathrm{m}} 8^{\text {a }}$ | 86 | 0.2 PhosphE + 0.5 PhosphC + 0.075 PhosphS + 0.075 |
| 21 | $\mathrm{Malm}_{\mathrm{m}} \leftrightarrow 2 \mathrm{H}+\mathrm{Oxam}_{\mathrm{m}}$ |  | Sphm +0.15 Cholesterol $\rightarrow$ Lipids |
| 22 | $\mathrm{Cit}+\mathrm{CoA}+2 \mathrm{H} \leftrightarrow \mathrm{AcCoA}+\mathrm{Oxa}+\mathrm{H} 2 \mathrm{O}$ | 87 | $0.0087873 \mathrm{Glu}+0.0084944 \mathrm{Asp}+0.01406 \mathrm{Gly}+0.024897$ |
| 23 | $\mathrm{Cit} \rightarrow \mathrm{aKG}+\mathrm{CO} 2+2 \mathrm{H}$ |  | Ser $+0.0090803 \mathrm{Tyr}+0.0052724 \mathrm{Cys}+0.012009$ Ala |
| 24 | Fum $+\mathrm{H} 2 \mathrm{O} \leftrightarrow \mathrm{Mal}$ |  | $+0.0058582 \mathrm{Arg}+0.0073228$ Asn $+0.0090803 \mathrm{Gln}+$ |
| 25 | $\mathrm{Mal} \leftrightarrow \mathrm{Oxa}+2 \mathrm{H}$ |  | $0.0043937 \mathrm{His}+0.0055653 \mathrm{Ile}+0.015231 \mathrm{Leu}+0.013181$ |
| 26 | $\mathrm{Mal}_{\mathrm{m}} \leftrightarrow 2 \mathrm{H}+\mathrm{Pyrm}_{\mathrm{m}}+\mathrm{CO} 2 \mathrm{~m}$ |  | Lys $+0.0020504 \mathrm{Met}+0.0070299$ Phe +0.013474 Pro + |
| 27 | $\mathrm{Oxam}_{\mathrm{m}} \leftrightarrow \mathrm{Pyrm}_{\mathrm{m}}+\mathrm{CO} 2 \mathrm{~m}$ |  | 0.017282 Thr $+0.0038079 \mathrm{Trp}+0.018746 \mathrm{Val} \rightarrow \mathrm{mAb}$ ext |
| 28 | $\mathrm{PEP}+\mathrm{CO} 2+\mathrm{H} 2 \mathrm{O} \leftrightarrow \mathrm{Oxa}+\mathrm{H} 3 \mathrm{PO} 48^{2}$ | 88 | $0.11952 \mathrm{Glu}+0.084994 \mathrm{Asp}+0.075289 \mathrm{Gly}+0.055603 \mathrm{Ser}$ |
| 29 | Mal $\leftrightarrow \mathrm{Pyr}+\mathrm{CO} 2+2 \mathrm{H}$ |  | $+0.019101 \mathrm{Tyr}+0.014495 \mathrm{Cys}+0.073479 \mathrm{Ala}+0.045166$ |
| 30 | $\mathrm{Pyr}+\mathrm{Glu} \leftrightarrow \mathrm{aKG}+\mathrm{Ala}$ |  | $\mathrm{Arg}+0.045623 \mathrm{Asn}+0.050412 \mathrm{Gln}+0.012685 \mathrm{His}+$ |
| 31 | $\mathrm{Pyr}_{\mathrm{m}}+\mathrm{Glu}_{\mathrm{m}} \leftrightarrow \mathrm{aKG}_{\mathrm{m}}+$ Ala $_{\mathrm{m}}$ |  | $0.037452 \mathrm{Ile}+0.076276 \mathrm{Leu}+0.086585 \mathrm{Lys}+0.019229$ |
| 32 | $\mathrm{Oxa}+\mathrm{Glu} \leftrightarrow \mathrm{aKG}+\mathrm{Asp}$ |  | Met +0.034729 Phe +0.043283 Pro +0.045641 Thr + |
| 33 | Oxam $+\mathrm{Glu}_{\mathrm{m}} \leftrightarrow \mathrm{aKG}_{\mathrm{m}}+\mathrm{Aspm}^{\text {m }}$ |  | $0.0058582 \mathrm{Trp}+0.054579 \mathrm{Val} \rightarrow$ Proteins |
| 34 | $\mathrm{Asn}+\mathrm{H} 2 \mathrm{O}+\mathrm{H} \rightarrow \mathrm{Asp}+\mathrm{NH} 4$ | 89 | $\mathrm{G} 6 \mathrm{P} \rightarrow \mathrm{TC}+\mathrm{H} 3 \mathrm{PO} 4$ |
| 35 | $\mathrm{Glu}+\mathrm{Asn} \leftrightarrow \mathrm{Asp}+\mathrm{Gln}$ 90 | 90 | 0.15299 Proteins +0.001405 DNA +0.0041309 RNA + |
| 36 | $\mathrm{Gln}+\mathrm{H} 2 \mathrm{O}+\mathrm{H} \leftrightarrow \mathrm{Glu}+\mathrm{NH} 4$ |  | 0.0026611 Lipids $+0.010513 \mathrm{TC} \rightarrow$ Biomass ext |
| 37 | $\mathrm{Glu}+\mathrm{H} 2 \mathrm{O} \leftrightarrow \mathrm{aKG}+\mathrm{NH} 4+\mathrm{H} \longrightarrow 9$ | 91 | $\mathrm{Pyr} \rightarrow \mathrm{Pyr}_{\mathrm{m}}$ |
| 38 | $\mathrm{H} 2 \mathrm{O}+\mathrm{Glum}_{\mathrm{m}} \leftrightarrow \mathrm{H}+\mathrm{aKG}_{\mathrm{m}}+\mathrm{NH} 4_{\mathrm{m}}$ | 92 | $\mathrm{Citm}_{\mathrm{m}} \leftrightarrow \mathrm{Cit}$ |
| 39 | $3 \mathrm{PG}+\mathrm{Glu}+\mathrm{H} 2 \mathrm{O} \rightarrow \mathrm{aKG}+\mathrm{Ser}+\mathrm{H} 3 \mathrm{PO} 4+2 \mathrm{H}$ | 93 | aKGm $\mathrm{aKG}^{\text {aKg }}$ |
| 40 | $\mathrm{Ser}+\mathrm{H} \rightarrow \mathrm{Pyr}+\mathrm{NH} 4$ | 94 | $\mathrm{Malm}_{\mathrm{m}} \leftrightarrow \mathrm{Mal}$ |
| 41 | $\mathrm{Ser}+\mathrm{H} 2 \mathrm{O} \leftrightarrow \mathrm{Gly}+\mathrm{CO} 2+4 \mathrm{H}$ | 95 | Alam $\leftrightarrow$ Ala |
| 42 | $\mathrm{Gly}+2 \mathrm{H} 2 \mathrm{O} \leftrightarrow \mathrm{NH} 4+2 \mathrm{CO} 2+5 \mathrm{H}$ | 96 | Glu $\leftrightarrow \mathrm{Glu}_{\mathrm{m}}$ |
| 43 | $\mathrm{aKG}+$ Orn $\leftrightarrow$ Glu + GluySA 9 | 97 | $\mathrm{Mal}+\mathrm{aKG}_{\mathrm{m}} \rightarrow \mathrm{aKG}+\mathrm{Mal}_{\mathrm{m}}$ |
| 44 | Pro $+\mathrm{H} 2 \mathrm{O} \leftrightarrow$ GluySA +2 H | 98 | $\mathrm{Glu}+\mathrm{Aspm} \leftrightarrow \mathrm{Asp}+\mathrm{Glu}_{\mathrm{m}}$ |
| 45 | GluySA $+\mathrm{H} 2 \mathrm{O} \leftrightarrow \mathrm{Glu}+2 \mathrm{H}$ | 99 | NH4 $\leftrightarrow \mathrm{NH} 4 \mathrm{~m}$ |
| 46 | $\mathrm{Thr}+\mathrm{CoA} \rightarrow \mathrm{Gly}+\mathrm{AcCoA}_{\mathrm{m}}$ | 100 | $\mathrm{CO} 2 \leftrightarrow \mathrm{CO} 2 \mathrm{~m}$ |
| 47 | Thr $+\mathrm{H} \rightarrow$ aKbut + NH4 101 | 101 | $\mathrm{Mal}+\mathrm{Cit}_{\mathrm{m}} \leftrightarrow \mathrm{Cit}+\mathrm{Mal}_{\mathrm{m}}$ |
| 48 | aKbut $+\mathrm{CoA} \rightarrow$ PropCoA $+\mathrm{CO} 2 \rightarrow 102$ | 102 | $\mathrm{O} 2+4 \mathrm{H} \rightarrow 2 \mathrm{H} 2 \mathrm{O}$ |
| 49 | PropCoA $+\mathrm{CO} 2 \rightarrow \mathrm{SucCoA}_{\mathrm{m}}$ ( 103 | 103 | $\mathrm{CO} 2 \leftrightarrow \mathrm{CO} 2$ ext |
| 50 | Trp $+9 \mathrm{H} 2 \mathrm{O} \rightarrow \mathrm{Ala}+\mathrm{aKad}+\mathrm{NH} 4+2 \mathrm{CO} 2+11 \mathrm{H} \quad 10$ | 104 | Aspext $\leftrightarrow$ Asp |
| 51 | $2 \mathrm{aKG}+\mathrm{Lys}+\mathrm{H} 2 \mathrm{O} \rightarrow 2 \mathrm{Glu}+\mathrm{aKad}+2 \mathrm{H}$ | 105 | Cysext $\leftrightarrow$ Cys |
| 52 | $\mathrm{aKad}+\mathrm{H} 2 \mathrm{O}+\mathrm{CoA} \rightarrow$ AcetoacCoA $+2 \mathrm{CO} 2+4 \mathrm{H}$ | 106 | Gly $\leftrightarrow$ Gly ${ }_{\text {ext }}$ |
| 53 | AcetoacCoA $+\mathrm{CoA}+2 \mathrm{H} \rightarrow 2 \mathrm{AcCoA}_{\mathrm{m}} 107$ | 107 | Serext $\leftrightarrow$ Ser |
| 54 | $\begin{aligned} & \mathrm{aKG}+\mathrm{Val}+2 \mathrm{H} 2 \mathrm{O}+\mathrm{CoA} \rightarrow \mathrm{Glu}+\mathrm{PropCoA}+2 \mathrm{CO} 210 \\ & +6 \mathrm{H} \end{aligned}$ | $\begin{aligned} & 108 \\ & 109 \end{aligned}$ | Glu $\leftrightarrow$ Gluext $^{\text {ext }}$ <br> $\mathrm{Tyr}_{\text {ext }} \leftrightarrow \mathrm{Tyr}$ |
| 55 | $\begin{aligned} & \mathrm{aKG}+\mathrm{Hle}+\mathrm{H} 2 \mathrm{O}+2 \mathrm{CoA} \rightarrow \mathrm{Glu}+\mathrm{PropCoA}+\mathrm{CO} 2+1 \\ & 2 \mathrm{H}+\mathrm{AcCoA} \end{aligned}$ | $\begin{aligned} & 110 \\ & 111 \end{aligned}$ | $\begin{aligned} & \text { Ala } \leftrightarrow \text { Ala ext } \\ & \text { Argext } \leftrightarrow \text { Arg } \end{aligned}$ |
| 56 | $\mathrm{aKG}+\mathrm{Leu}+\mathrm{H} 2 \mathrm{O}+\mathrm{CoA} \rightarrow \mathrm{Glu}+\mathrm{Acetoac}+2 \mathrm{H}+1$ | 112 | $\mathrm{Asn}_{\text {ext }} \leftrightarrow$ Asn |
|  | $\mathrm{AcCoA}_{\mathrm{m}}$ | 113 | Glnext $\leftrightarrow$ Gln |
| 57 | Acetoac $+\mathrm{SucCoA}_{\mathrm{m}} \rightarrow$ AcetoacCoA $+\mathrm{Suc}_{\mathrm{m}}$ | 114 | $\mathrm{His}_{\text {ext }} \rightarrow$ His |
| 58 | $\mathrm{Phe}+\mathrm{H} 2 \mathrm{O} \rightarrow$ Tyr +2 H | 115 | $\mathrm{Ile}_{\text {ext }} \rightarrow$ Ile |
| 59 | $\underset{\mathrm{H}}{\mathrm{aKG}}+\mathrm{Tyr}+5 \mathrm{H} 2 \mathrm{O} \rightarrow \mathrm{Fum}+\mathrm{Glu}+\text { Acetoac }+\mathrm{CO} 2+81$ | $\begin{aligned} & 116 \\ & 117 \end{aligned}$ | Leuext $\rightarrow$ Leu <br> Lys ${ }^{\text {ext }} \rightarrow$ Lys |
| 60 | $\mathrm{Ser}+\mathrm{Met}+2 \mathrm{H} 2 \mathrm{O} \rightarrow \mathrm{Cys}+\mathrm{aKbut}+\mathrm{NH} 4+\mathrm{CO} 2+5 \mathrm{H} 1$ | 118 | Met ext $^{\text {e }}$ Met |
| 61 | $\mathrm{Cys}+5 \mathrm{H} 2 \mathrm{O} \rightarrow \mathrm{Pyr}+\mathrm{NH} 4+\mathrm{H} 2 \mathrm{SO} 4+7 \mathrm{H}$ | 119 | $\mathrm{Ph}_{\text {ext }} \rightarrow$ Phe |
| 62 | $\mathrm{Arg}+\mathrm{H} 2 \mathrm{O} \rightarrow \mathrm{Orn}+$ Urea 1 | 120 | $\mathrm{Pro}_{\text {ext }} \rightarrow$ Pro |
| 63 | $\mathrm{His}+4 \mathrm{H} 2 \mathrm{O} \rightarrow \mathrm{Glu}+2 \mathrm{NH} 4+\mathrm{CO} 2 \rightarrow 1$ | 121 | Thrext $\rightarrow$ Thr |
| 64 | $\mathrm{Orn}+\mathrm{NH} 4_{\mathrm{m}}+\mathrm{CO} 2 \mathrm{~m} \rightarrow \mathrm{Cln}+\mathrm{H} 2 \mathrm{O}+\mathrm{H} \quad 12$ | 122 | $\operatorname{Trp}_{\text {ext }} \rightarrow$ Trp |
| 65 | $\mathrm{Asp}+\mathrm{Cln} \rightarrow$ Argsucc +H 2 O | 123 | Valext $\rightarrow$ Val |
| 66 | Argsucc $\rightarrow$ Fum + Arg 12 | 124 | Ethnext $\rightarrow$ Ethn |
| 67 | $\mathrm{R} 5 \mathrm{P}+2 \mathrm{H} 3 \mathrm{PO} 4 \rightarrow \mathrm{PRPP}+2 \mathrm{H} 2 \mathrm{O}$ | 125 | Choext $\rightarrow$ Cho |
| 68 | $\begin{aligned} & \mathrm{Asp}+\mathrm{Gly}+2 \mathrm{Gln}+\mathrm{PRPP}+3 \mathrm{CO} 2+4 \mathrm{H} \rightarrow \mathrm{Fum}+21 \\ & \mathrm{Glu}+\mathrm{IMP}+2 \mathrm{H} 3 \mathrm{PO} 4+4 \mathrm{H} 2 \mathrm{O} \end{aligned}$ | 126 | $\mathrm{NH} 4 \leftrightarrow$ Urea $\rightarrow$ NH4 ext |
| 69 | Asp $+\mathrm{IMP}+2 \mathrm{H} 3 \mathrm{PO} 4 \rightarrow \mathrm{Fum}+\mathrm{ATPrn}+3 \mathrm{H} 2 \mathrm{O} \xrightarrow{2}$ | 128 |  |
| 70 | $\mathrm{Gln}+\mathrm{IMP}+2 \mathrm{H} 3 \mathrm{PO} 4 \rightarrow \mathrm{Glu}+\mathrm{GTPrn}+\mathrm{H} 2 \mathrm{O}+2 \mathrm{H}$ | 129 | H 2 SO 4 ext $\leftrightarrow$ H 2 SO 4 |
| 71 | Asp $+\mathrm{NH} 4+\mathrm{CO} 2 \rightarrow$ Orot $+2 \mathrm{H} 2 \mathrm{O}+3 \mathrm{H}$ | 130 | $\mathrm{H} 2 \mathrm{O} \leftrightarrow \mathrm{H} 2 \mathrm{O}_{\text {ext }}$ |
| 72 | PRPP + Orot $\rightarrow$ UTPrn $+\mathrm{CO} 2+\mathrm{H} 2 \mathrm{O}$ | 131 | O 2 ext $\leftrightarrow \mathrm{O} 2$ |

Table 7: List of reactions with linearly independent stoichiometries in the perfusion bioreactor under study.

| Reaction | Chemical equation | Reaction | Chemical equation |
| :---: | :---: | :---: | :---: |
| 1 | $\mathrm{Glc}_{\text {ext }}+\mathrm{H} 3 \mathrm{PO} 4 \leftrightarrow \mathrm{G6P}+\mathrm{H} 2 \mathrm{O}$ | 65 | UTPrn $+\mathrm{CO} 2+8 \mathrm{H} \leftrightarrow \mathrm{dTTP}+3 \mathrm{H} 2 \mathrm{O}$ |
| 2 | $\mathrm{F} 6 \mathrm{P}+\mathrm{H} 3 \mathrm{PO} 4 \leftrightarrow \mathrm{DHAP}+\mathrm{GA} 3 \mathrm{P}+\mathrm{H} 2 \mathrm{O}$ | 66 | CTPrn $+2 \mathrm{H} \leftrightarrow \mathrm{dCTP}+\mathrm{H} 2 \mathrm{O}$ |
| 3 | DHAP $\leftrightarrow$ GA3P 6 | 67 | $0.285 \mathrm{dATP}+0.215 \mathrm{dGTP}+0.285 \mathrm{dTTP}+0.215 \mathrm{dCTP}$ |
| 4 | $\mathrm{GA} 3 \mathrm{P}+\mathrm{H} 2 \mathrm{O} \leftrightarrow 3 \mathrm{PG}+2 \mathrm{H}$ |  | $\leftrightarrow$ DNA |
| 5 | Pyr $+2 \mathrm{H} \leftrightarrow \mathrm{Lac}_{\text {ext }}$ | 68 | DHAP $+2 \mathrm{H} \leftrightarrow$ Glyc3P |
| 6 | $\mathrm{G} 6 \mathrm{P}+\mathrm{H} 2 \mathrm{O} \leftrightarrow \mathrm{R} 15 \mathrm{P}+\mathrm{CO} 2+4 \mathrm{H}$ | 69 | 18 AcCoA + Cho + Glyc3P + $23 \mathrm{H} \leftrightarrow$ PhosphC +17 H 2 O |
| 7 | R15P $\leftrightarrow$ R5P |  | $+18 \mathrm{CoA}$ |
| 8 | R15P $\leftrightarrow$ X5P 7 | 70 | $18 \mathrm{AcCoA}+\mathrm{Ethn}+\mathrm{Glyc} 3 \mathrm{P}+26 \mathrm{H} \leftrightarrow \mathrm{PhosphE}+17 \mathrm{H} 2 \mathrm{O}$ |
| 9 | $\mathrm{R} 5 \mathrm{P}+\mathrm{X} 5 \mathrm{P} \leftrightarrow \mathrm{F} 6 \mathrm{P}+\mathrm{E} 4 \mathrm{P}$ |  | $+18 \mathrm{CoA}$ |
| 10 | $\mathrm{X} 5 \mathrm{P}+\mathrm{E} 4 \mathrm{P} \leftrightarrow \mathrm{F} 6 \mathrm{P}+\mathrm{GA} 3 \mathrm{P}$ | 71 | Ser + PhosphE $\leftrightarrow$ Ethn + PhosphS |
| 11 |  | 72 | $16 \mathrm{AcCoA}+\mathrm{Ser}+$ Cho + Glyc3P $+19 \mathrm{H} \leftrightarrow$ Sphm +2 |
| 12 | $\mathrm{H} 2 \mathrm{O}+\mathrm{SucCoA}_{\mathrm{m}} \leftrightarrow \mathrm{CoA}+2 \mathrm{H}+\mathrm{Suc}_{\mathrm{m}}$ |  | $\mathrm{CO} 2+16 \mathrm{H} 2 \mathrm{O}+16 \mathrm{CoA}$ |
| 13 | $\mathrm{Sucm}_{\mathrm{m}} \leftrightarrow 2 \mathrm{H}+\mathrm{Fum}_{\mathrm{m}}{ }^{\text {a }}$ | 73 | $16 \mathrm{AcCoA} \leftrightarrow$ Cholesterol $+5 \mathrm{CO} 2+5 \mathrm{H} 2 \mathrm{O}+16 \mathrm{CoA}+8$ |
| 14 | $\mathrm{H} 2 \mathrm{O}+\mathrm{Fum}_{\mathrm{m}} \leftrightarrow \mathrm{Mal}_{\mathrm{m}}$ |  |  |
| 15 | $\mathrm{Cit}+\mathrm{CoA}+2 \mathrm{H} \leftrightarrow \mathrm{AcCoA}+\mathrm{Oxa}+\mathrm{H} 2 \mathrm{O} \quad 7$ | 74 | 0.2 PhosphE + 0.5 PhosphC + 0.075 PhosphS + 0.075 |
| 16 | Cit $\leftrightarrow \mathrm{aKG}+\mathrm{CO} 2+2 \mathrm{H}$ |  | Sphm +0.15 Cholesterol $\leftrightarrow$ Lipids |
| 17 | Fum $+\mathrm{H} 2 \mathrm{O} \leftrightarrow$ Mal ${ }^{\text {a }}$ | 75 | $0.0087873 \mathrm{Glu}+0.0084944 \mathrm{Asp}+0.01406 \mathrm{Gly}+0.024897$ |
| 18 | $\mathrm{Oxam}_{\mathrm{m}} \leftrightarrow \mathrm{Pyr}_{\mathrm{m}}+\mathrm{CO} 2 \mathrm{~m}$ |  | Ser $+0.0090803 \mathrm{Tyr}+0.0052724$ Cys +0.012009 Ala |
| 19 | $\mathrm{PEP}+\mathrm{CO} 2+\mathrm{H} 2 \mathrm{O} \leftrightarrow \mathrm{Oxa}+\mathrm{H} 3 \mathrm{PO} 4$ |  | $+0.0058582 \mathrm{Arg}+0.0073228 \mathrm{Asn}+0.0090803 \mathrm{Gln}+$ |
| 20 | $\mathrm{Mal} \leftrightarrow \mathrm{Pyr}+\mathrm{CO} 2+2 \mathrm{H}$ |  | $0.0043937 \mathrm{His}+0.0055653 \mathrm{Ile}+0.015231 \mathrm{Leu}+0.013181$ |
| 21 | $\mathrm{Pyr}_{\mathrm{m}}+\mathrm{Glu}_{\mathrm{m}} \leftrightarrow \mathrm{aKG}_{\mathrm{m}}+\mathrm{Ala}_{\mathrm{m}}$ |  | Lys $+0.0020504 \mathrm{Met}+0.0070299$ Phe +0.013474 Pro + |
| 22 | $\mathrm{Oxa}+\mathrm{Glu} \leftrightarrow \mathrm{aKG}+\mathrm{Asp}$ |  | $0.017282 \mathrm{Thr}+0.0038079 \mathrm{Trp}+0.018746 \mathrm{Val} \leftrightarrow \mathrm{mAb}$ ext |
| 23 | $\mathrm{Oxam}_{\mathrm{m}}+\mathrm{Glu}_{\mathrm{m}} \leftrightarrow \mathrm{aKG}_{\mathrm{m}}+\mathrm{Aspm}$ | 76 | $0.11952 \mathrm{Glu}+0.084994 \mathrm{Asp}+0.075289 \mathrm{Gly}+0.055603 \mathrm{Ser}$ |
| 24 | $\mathrm{Glu}+\mathrm{Asn} \leftrightarrow \mathrm{Asp}+\mathrm{Gln}$ |  | $+0.019101 \mathrm{Tyr}+0.014495 \mathrm{Cys}+0.073479 \mathrm{Ala}+0.045166$ |
| 25 | $\mathrm{Gln}+\mathrm{H} 2 \mathrm{O}+\mathrm{H} \leftrightarrow \mathrm{Glu}+\mathrm{NH} 4$ |  | Arg +0.045623 Asn $+0.050412 \mathrm{Gln}+0.012685 \mathrm{His}+$ |
| 26 | $\mathrm{H} 2 \mathrm{O}+\mathrm{Glu}_{\mathrm{m}} \leftrightarrow \mathrm{H}+\mathrm{aKG}_{\mathrm{m}}+\mathrm{NH} 4 \mathrm{~m}$ |  | $0.037452 \mathrm{Ile}+0.076276 \mathrm{Leu}+0.086585 \mathrm{Lys}+0.019229$ |
| 27 | $3 \mathrm{PG}+\mathrm{Glu}+\mathrm{H} 2 \mathrm{O} \leftrightarrow \mathrm{aKG}+\mathrm{Ser}+\mathrm{H} 3 \mathrm{PO} 4+2 \mathrm{H}$ |  | Met +0.034729 Phe +0.043283 Pro +0.045641 Thr + |
| 28 | Ser $+\mathrm{H} \leftrightarrow \mathrm{Pyr}+\mathrm{NH} 4$ |  | $0.0058582 \mathrm{Trp}+0.054579 \mathrm{Val} \leftrightarrow$ Proteins |
| 29 | $\mathrm{Ser}+\mathrm{H} 2 \mathrm{O} \leftrightarrow \mathrm{Gly}+\mathrm{CO} 2+4 \mathrm{H} \quad 7$ | 77 | $\mathrm{G} 6 \mathrm{P} \leftrightarrow \mathrm{TC}+\mathrm{H} 3 \mathrm{PO} 4$ |
| 30 | $\mathrm{Gly}+2 \mathrm{H} 2 \mathrm{O} \leftrightarrow \mathrm{NH} 4+2 \mathrm{CO} 2+5 \mathrm{H}$ | 78 | 0.15299 Proteins +0.001405 DNA +0.0041309 RNA + |
| 31 | $\mathrm{aKG}+$ Orn $\leftrightarrow \mathrm{Glu}+$ GluySA |  | 0.0026611 Lipids $+0.010513 \mathrm{TC} \leftrightarrow$ Biomass ext |
| 32 | Pro $+\mathrm{H} 2 \mathrm{O} \leftrightarrow$ GluySA $+2 \mathrm{H} \quad 7$ | 79 | Pyr $\leftrightarrow$ Pyrm |
| 33 | GluySA $+\mathrm{H} 2 \mathrm{O} \leftrightarrow \mathrm{Glu}+2 \mathrm{H}$ | 80 | Malm ${ }_{\text {mal }}$ |
| 34 | $\mathrm{Thr}+\mathrm{CoA} \leftrightarrow \mathrm{Gly}+\mathrm{AcCoA}_{\mathrm{m}}$ ( 8 | 81 | $\mathrm{Ala}_{\mathrm{m}} \leftrightarrow$ Ala |
| 35 | $\mathrm{Thr}+\mathrm{H} \leftrightarrow \mathrm{aKbut}+\mathrm{NH} 4$ | 82 | Glu $\leftrightarrow \mathrm{Glu}_{\mathrm{m}}$ |
| 36 | aKbut $+\mathrm{CoA} \leftrightarrow \mathrm{PropCoA}+\mathrm{CO} 2$ | 83 | $\mathrm{Mal}+\mathrm{aKG}_{\mathrm{m}} \leftrightarrow \mathrm{aKG}+\mathrm{Mal}_{\mathrm{m}}$ |
| 37 | PropCoA + CO2 $\leftrightarrow \mathrm{SucCoA}_{\mathrm{m}}$ ( $\mathrm{Cl}^{\text {d }}$ | 84 | $\mathrm{Glu}+\mathrm{Aspm} \leftrightarrow \mathrm{Asp}+\mathrm{Glu}_{\mathrm{m}}$ |
| 38 | $\mathrm{Trp}+9 \mathrm{H} 2 \mathrm{O} \leftrightarrow \mathrm{Ala}+\mathrm{aKad}+\mathrm{NH} 4+2 \mathrm{CO} 2+11 \mathrm{H}$ | 85 | NH4 $\leftrightarrow \mathrm{NH} 4 \mathrm{~m}$ |
| 39 | $2 \mathrm{aKG}+\mathrm{Lys}+\mathrm{H} 2 \mathrm{O} \leftrightarrow 2 \mathrm{Glu}+\mathrm{aKad}+2 \mathrm{H}$ | 86 | $\mathrm{CO} 2 \leftrightarrow \mathrm{CO} 2 \mathrm{~m}$ |
| 40 | $\mathrm{aKad}+\mathrm{H} 2 \mathrm{O}+\mathrm{CoA} \leftrightarrow$ AcetoacCoA $+2 \mathrm{CO} 2+4 \mathrm{H}$ | 87 | $\mathrm{Mal}+\mathrm{Cit}_{\mathrm{m}} \leftrightarrow \mathrm{Cit}+\mathrm{Mal}_{\mathrm{m}}$ |
| 41 | AcetoacCoA $+\mathrm{CoA}+2 \mathrm{H} \leftrightarrow 2 \mathrm{AcCoA} \mathrm{m}^{2} 8$ | 88 | $\mathrm{O} 2+4 \mathrm{H} \leftrightarrow 2 \mathrm{H} 2 \mathrm{O}$ |
| 42 | $\begin{aligned} & \mathrm{aKG}+\mathrm{Val}+2 \mathrm{H} 2 \mathrm{O}+\mathrm{CoA} \leftrightarrow \mathrm{Glu}+\mathrm{PropCoA}+2 \mathrm{CO} 2 \\ & +6 \mathrm{H} \end{aligned}$ | $\begin{aligned} & 89 \\ & 90 \end{aligned}$ | $\begin{aligned} & \mathrm{CO} 2 \leftrightarrow \mathrm{CO} 22_{\text {ext }} \\ & \text { Aspext } \leftrightarrow \text { Asp } \end{aligned}$ |
| 43 | $\begin{aligned} & \mathrm{aKG}+\mathrm{Ile}+\mathrm{H} 2 \mathrm{O}+2 \mathrm{CoA} \leftrightarrow \mathrm{Glu}+\text { PropCoA }+\mathrm{CO} 2+9 \\ & 2 \mathrm{H}+\mathrm{AcCoA}_{\mathrm{m}} \end{aligned}$ | $\begin{aligned} & 91 \\ & 92 \end{aligned}$ | $\text { Cysext } \leftrightarrow \text { Cys }$ <br> Gly $\leftrightarrow$ Gly |
| 44 | $\mathrm{aKG}+\mathrm{Leu}+\mathrm{H} 2 \mathrm{O}+\mathrm{CoA} \leftrightarrow \mathrm{Glu}+\mathrm{Acetoac}+2 \mathrm{H}+9$ |  | Ser $_{\text {ext }} \leftrightarrow \leftrightarrow$ Ser |
|  | $\mathrm{AcCoAm}_{m}$ | 94 | Glu $\leftrightarrow \mathrm{Glu}_{\text {ext }}$ |
| 45 | Acetoac $+\mathrm{SucCoA}_{\mathrm{m}} \leftrightarrow$ AcetoacCoA $+\mathrm{Sucm}_{\mathrm{m}}{ }^{\text {c }}$ | 95 | Tyrext $\leftrightarrow$ Tyr |
| 46 | $\mathrm{Phe}+\mathrm{H} 2 \mathrm{O} \leftrightarrow \mathrm{Tyr}+2 \mathrm{H}$ | 96 | Ala $\leftrightarrow \mathrm{Ala}_{\text {ext }}$ |
| 47 | $\underset{\mathrm{H}}{\mathrm{aKG}}+\mathrm{Tyr}+5 \mathrm{H} 2 \mathrm{O} \leftrightarrow \mathrm{Fum}+\mathrm{Glu}+$ Acetoac $+\mathrm{CO} 2+89$ | 97 98 | $\text { Argext }^{4} \leftrightarrow \mathrm{Arg}$ |
| 48 | $\mathrm{Ser}+\mathrm{Met}+2 \mathrm{H} 2 \mathrm{O} \leftrightarrow \mathrm{Cys}+\mathrm{aKbut}+\mathrm{NH} 4+\mathrm{CO} 2+5 \mathrm{H} 9$ |  | Gln ext $\leftrightarrow$ Gln |
| 49 | $\mathrm{Cys}+5 \mathrm{H} 2 \mathrm{O} \leftrightarrow \mathrm{Pyr}+\mathrm{NH} 4+\mathrm{H} 2 \mathrm{SO} 4+7 \mathrm{H}$ | 100 | $\mathrm{His}_{\text {ext }} \leftrightarrow \mathrm{His}$ |
| 50 | $\mathrm{Arg}+\mathrm{H} 2 \mathrm{O} \leftrightarrow \mathrm{Orn}+$ Urea 101 | 101 | $\mathrm{Ile}_{\text {ext }} \leftrightarrow$ Ile |
| 51 | $\mathrm{His}+4 \mathrm{H} 2 \mathrm{O} \leftrightarrow \mathrm{Glu}+2 \mathrm{NH} 4+\mathrm{CO} 2$ | 102 | $L^{\text {Leu }}$ ext $\leftrightarrow$ Leu |
| 52 | $\mathrm{Orn}+\mathrm{NH} 4 \mathrm{~m}+\mathrm{CO} 2 \mathrm{~m} \leftrightarrow \mathrm{Cln}+\mathrm{H} 2 \mathrm{O}+\mathrm{H} \quad 1$ | 103 | Lys ext $\leftrightarrow$ Lys |
| 53 | Asp $+\mathrm{Cln} \leftrightarrow$ Argsucc +H 2 O | 104 | Metext $\leftrightarrow$ Met |
| 54 | Argsucc $\leftrightarrow$ Fum + Arg 1 | 105 | $\mathrm{Ph}_{\text {ext }} \leftrightarrow$ Phe |
| 55 | $\mathrm{R} 5 \mathrm{P}+2 \mathrm{H} 3 \mathrm{PO} 4 \leftrightarrow \mathrm{PRPP}+2 \mathrm{H} 2 \mathrm{O}$ | 106 | Proext $\leftrightarrow$ Pro |
| 56 | $\mathrm{Asp}+\mathrm{Gly}+2 \mathrm{Gln}+\mathrm{PRPP}+3 \mathrm{CO} 2+4 \mathrm{H} \leftrightarrow \mathrm{Fum}+21$ | $107$ | Thrext $\leftrightarrow$ Thr |
| 57 | $\mathrm{Asp}+\mathrm{IMP}+2 \mathrm{H} 3 \mathrm{PO} 4 \leftrightarrow \mathrm{Fum}+\mathrm{ATPrn}+3 \mathrm{H} 2 \mathrm{O}$ | 109 | $\underset{\text { Trpext }}{\text { Val }}$ ext $\leftrightarrow$ Trp |
| 58 | $\mathrm{Gln}+\mathrm{IMP}+2 \mathrm{H} 3 \mathrm{PO} 4 \leftrightarrow \mathrm{Glu}+\mathrm{GTPrn}+\mathrm{H} 2 \mathrm{O}+2 \mathrm{H} \quad 1$ | 110 | Ethnext $\leftrightarrow$ Ethn |
| 59 | $\mathrm{Asp}+\mathrm{NH} 4+\mathrm{CO} 2 \leftrightarrow$ Orot $+2 \mathrm{H} 2 \mathrm{O}+3 \mathrm{H}$ | 111 | Choext $\leftrightarrow$ Cho |
| 60 | $\mathrm{PRPP}+$ Orot $\leftrightarrow \mathrm{UTPrn}+\mathrm{CO} 2+\mathrm{H} 2 \mathrm{O}$ | 112 | $\mathrm{NH} 4 \leftrightarrow \mathrm{NH}_{4}$ ext |
| 61 | Gln + UTPrn $\leftrightarrow$ Glu + CTPrn 11 | 113 | Urea $\leftrightarrow$ Urea ext |
| 62 | 0.285 UTPrn +0.285 ATPrn +0.215 GTPrn +0.215 CT- 11 | 114 | H 3 PO 4 ext $\leftrightarrow \mathrm{H} 3 \mathrm{PO} 4$ |
|  | $\operatorname{Prn} \leftrightarrow$ RNA 11 | 115 | H 2 SO 4 ext $\leftrightarrow$ H2SO4 |
| 63 | ATPrn $+2 \mathrm{H} \leftrightarrow$ dATP +H 2 O | 116 | $\mathrm{H} 2 \mathrm{O} \leftrightarrow \mathrm{H}_{2} \mathrm{O}_{\text {ext }}$ |
| 64 | GTPrn $+2 \mathrm{H} \leftrightarrow \mathrm{dGTP}+\mathrm{H} 2 \mathrm{O}$ | 117 | $\mathrm{O} 22_{\text {ext }} \leftrightarrow \mathrm{O} 2$ |

Table 8: List of BFMs in the perfusion bioreactor under study obtained in option 1 for the generation of BFMs.

| BFM | Chemical equation |
| :---: | :---: |
| 1 | 0.014204 Glc $_{\text {ext }}+0.03234$ His $_{\text {ext }}+0.00573$ Ile $_{\text {ext }}+0.01167$ Leu $_{\text {ext }}+0.013247$ Lys $_{\text {ext }}+0.0051595$ Met $_{\text {ext }}+$ $0.0082357 \mathrm{Phe}_{\text {ext }}+0.0069828 \mathrm{Thr}_{\text {ext }}+0.028883 \operatorname{Trp}_{\text {ext }}+0.0083503 \mathrm{Val}_{\text {ext }}+0.0019424 \mathrm{CO} 2_{\text {ext }}+0.00053223$ $\mathrm{Ethn}_{\text {ext }}+0.0015302$ Cho ext $+0.01887 \mathrm{H} 3 \mathrm{PO} 4_{\mathrm{ext}}+0.13401 \mathrm{H} 2 \mathrm{O}_{\text {ext }} \leftrightarrow$ Biomass $_{\text {ext }}$ |
| 2 | $\mathrm{Glc}_{\text {ext }}+6 \mathrm{O} 22_{\text {ext }} \leftrightarrow 6 \mathrm{CO} 2_{\text {ext }}+6 \mathrm{H} 2 \mathrm{O}_{\text {ext }}$ |
| 3 | $\mathrm{Ala}_{\text {ext }}+3 \mathrm{O} 2_{\text {ext }} \leftrightarrow 2.5 \mathrm{CO} 2_{\text {ext }}+0.5 \mathrm{Urea}_{\text {ext }}+2.5 \mathrm{H} 2 \mathrm{O}_{\text {ext }}$ |
| 4 | $\mathrm{Arg}_{\text {ext }}+5.5 \mathrm{O} 2_{\text {ext }} \leftrightarrow 4 \mathrm{CO} 2 \mathrm{ext}^{\text {ext }}+2 \mathrm{Urea}_{\text {ext }}+3 \mathrm{H} 2 \mathrm{O}$ ext |
| 5 | $\mathrm{Asn}_{\text {ext }}+3 \mathrm{O} 2_{\text {ext }} \leftrightarrow 3 \mathrm{CO} 2_{\text {ext }}+\mathrm{Urea}_{\text {ext }}+2 \mathrm{H} 2 \mathrm{O}_{\text {ext }}$ |
| 6 | Asp $_{\text {ext }}+3 \mathrm{O} 2$ ext $^{\text {ext }} 3.5 \mathrm{CO} 2_{\text {ext }}+0.5 \mathrm{Urea}_{\text {ext }}+2.5 \mathrm{H} 2 \mathrm{O}_{\text {ext }}$ |
| 7 | $\mathrm{Cys}_{\mathrm{ext}}+4.5 \mathrm{O} 22_{\text {ext }} \leftrightarrow 2.5 \mathrm{CO} 2_{\text {ext }}+0.5 \mathrm{Urea}_{\text {ext }}+\mathrm{H} 2 \mathrm{SO} 4_{\text {ext }}+1.5 \mathrm{H} 2 \mathrm{O}_{\text {ext }}$ |
| 8 | $\mathrm{Gln} \mathrm{ext}^{\text {ext }}+4.5 \mathrm{O} 2 \mathrm{ext}^{\text {ext }} 44 \mathrm{CO} 22_{\text {ext }}+\mathrm{Urea}_{\text {ext }}+3 \mathrm{H} 2 \mathrm{O}$ ext |
| 9 | $\mathrm{Glu}_{\text {ext }}+4.5 \mathrm{O} 2_{\text {ext }} \leftrightarrow 4.5 \mathrm{CO} 22_{\text {ext }}+0.5 \mathrm{Urea}_{\text {ext }}+3.5 \mathrm{H} 2 \mathrm{O}_{\text {ext }}$ |
| 10 | $\mathrm{Gly} \mathrm{ext}_{\text {ext }}+1.5 \mathrm{O} 2$ ext $\leftrightarrow 1.5 \mathrm{CO} 22_{\text {ext }}+0.5 \mathrm{Urea}_{\text {ext }}+1.5 \mathrm{H} 2 \mathrm{O}$ ext |
| 11 | $\mathrm{His}_{\text {ext }}+5 \mathrm{O} 2_{\text {ext }} \leftrightarrow 4.5 \mathrm{CO} 2_{\text {ext }}+1.5 \mathrm{Urea}_{\text {ext }}+1.5 \mathrm{H} 2 \mathrm{O}_{\text {ext }}$ |
| 12 | $\mathrm{Ile}_{\text {ext }}+7.5 \mathrm{O} 2_{\text {ext }} \leftrightarrow 5.5 \mathrm{CO} 2_{\text {ext }}+0.5 \mathrm{Urea}_{\text {ext }}+5.5 \mathrm{H} 2 \mathrm{O}_{\text {ext }}$ |
| 13 | $\mathrm{Leu}_{\text {ext }}+7.5 \mathrm{O} 2$ ext $\leftrightarrow 5.5 \mathrm{CO} 22_{\text {ext }}+0.5 \mathrm{Urea}_{\text {ext }}+5.5 \mathrm{H} 2 \mathrm{O}_{\text {ext }}$ |
| 14 | $\mathrm{Lys}_{\text {ext }}+7 \mathrm{O} 2_{\text {ext }} \leftrightarrow 5 \mathrm{CO} 2_{\text {ext }}+\mathrm{Urea}_{\text {ext }}+5 \mathrm{H} 2 \mathrm{O}_{\text {ext }}$ |
| 15 | $\mathrm{Met}_{\text {ext }}+7.5 \mathrm{O} 2_{\text {ext }} \leftrightarrow 4.5 \mathrm{CO} 2_{\text {ext }}+0.5 \mathrm{Urea}_{\text {ext }}+\mathrm{H} 2 \mathrm{SO} 4_{\text {ext }}+3.5 \mathrm{H} 2 \mathrm{O}$ ext |
| 16 | $\mathrm{Phe}_{\text {ext }}+10 \mathrm{O} 2_{\text {ext }} \leftrightarrow 8.5 \mathrm{CO} 22_{\text {ext }}+0.5 \mathrm{Urea}_{\text {ext }}+4.5 \mathrm{H} 2 \mathrm{O}$ ext |
| 17 | $\mathrm{Pro}_{\text {ext }}+5.5 \mathrm{O} 2_{\text {ext }} \leftrightarrow 4.5 \mathrm{CO} 22_{\text {ext }}+0.5 \mathrm{Urea}_{\text {ext }}+3.5 \mathrm{H} 2 \mathrm{O}_{\text {ext }}$ |
| 18 | $\mathrm{Ser}_{\text {ext }}+2.5 \mathrm{O} 22_{\text {ext }} \leftrightarrow 2.5 \mathrm{CO} 22_{\text {ext }}+0.5 \mathrm{Urea}_{\text {ext }}+2.5 \mathrm{H} 2 \mathrm{O}$ ext |
| 19 | $\mathrm{Thr}_{\text {ext }}+4 \mathrm{O} 2_{\text {ext }} \leftrightarrow 3.5 \mathrm{CO} 2_{\text {ext }}+0.5 \mathrm{Urea}_{\text {ext }}+3.5 \mathrm{H} 2 \mathrm{O}_{\text {ext }}$ |
| 20 | Trpext $+11.5 \mathrm{O} 2_{\text {ext }} \leftrightarrow 10 \mathrm{CO} 22_{\text {ext }}+\mathrm{Urea}_{\text {ext }}+4 \mathrm{H} 2 \mathrm{O}_{\text {ext }}$ |
| 21 | $\mathrm{Tyr}_{\text {ext }}+9.5 \mathrm{O} 2_{\text {ext }} \leftrightarrow 8.5 \mathrm{CO} 22_{\text {ext }}+0.5 \mathrm{Ur}_{\text {eaext }}+4.5 \mathrm{H} 2 \mathrm{O}$ ext |
| 22 | $\mathrm{Val}_{\text {ext }}+6 \mathrm{O} 22_{\text {ext }} \leftrightarrow 4.5 \mathrm{CO} 2_{\text {ext }}+0.5 \mathrm{Urea}_{\text {ext }}+4.5 \mathrm{H} 2 \mathrm{O}_{\text {ext }}$ |
| 23 | $\mathrm{NH} 44_{\text {ext }}+0.5 \mathrm{CO} 2_{\text {ext }}+0.25 \mathrm{O} 2_{\text {ext }} \leftrightarrow 0.5 \mathrm{Ur}^{\text {ex }}$ ext $+\mathrm{H} 2 \mathrm{O}_{\text {ext }}$ |
| 24 | $\begin{aligned} & 0.041952 \text { His }_{\text {ext }}+0.0055653 \text { Ile ext }+0.015231 \mathrm{Leu}_{\text {ext }}+0.013181 \mathrm{Lys}_{\text {ext }}+0.0073228 \text { Met }_{\text {ext }}+0.01611 \text { Phe }_{\text {ext }} \\ & +0.017282 \text { Thr }_{\text {ext }}+0.016451 \text { Trpext }_{\text {ext }}+0.018746 \text { Val }_{\text {ext }}+0.018997 \mathrm{CO} 2_{\text {ext }}+0.15451 \mathrm{H} 2 \mathrm{O}_{\text {ext }} \leftrightarrow \mathrm{mAb} \text { ext } \end{aligned}$ |
| 25 | $0.5 \mathrm{Glc}_{\text {ext }} \leftrightarrow \mathrm{Lac}_{\text {ext }}$ |

antibodies. One of the BFMs for production of biomass uses glucose and a combination of all the amino acids as reactants, while the other BFM for production of biomass uses only glucose and a combination of essential amino acids as reactants. Regarding the BFM for production of monoclonal antibodies, it uses a combination of all the amino acids as reactants. Once again, all the resulting columns of $\mathbf{E}_{m}$ correspond to admissible BFMs. The resulting BFMs are listed in Table 9.

In either option, the reaction system under study can always be represented as a reaction network with 25 BFMs , and each measured species participates in only a few BFMs. In option 1, each BFM corresponds to a stoichiometric coefficient 1 for only one measured species. In option 2, each BFM involves only a few species in a way that is biologically meaningful according to previous work by [12].

In summary, the rates of the 25 BFMs can be uniquely computed from the rates of variation of 25 measured extracellular species. However, it must be stressed that the concept of BFMs remains a virtual concept that is not guaranteed to correspond to true reactions since the stoichiometries of BFMs are constructed as linear combinations of the stoichiometries of independent reactions in the reaction network (as shown in (18)), the reaction rates of BFMs are also linear combinations of the rates of independent reactions (as shown in (20)), and these linear combinations may include negative coefficients. In particular, this means that the rates of certain BFMs may be negative, even if the stoichiometries of these BFMs seem to indicate that their rates should be nonnegative. Nevertheless, the 5 -fold reduction (from 125 EFMs in [12] to 25 BFMs ) in the number of flux modes that need to be modeled alleviates the modeling effort that is needed in the subsequent sections of this paper.

Table 9: List of BFMs in the perfusion bioreactor under study obtained in option 2 for the generation of BFMs.

| BFM | Chemical equation |
| :---: | :---: |
| 1 |  |
| 2 | $\mathrm{Ser}_{\text {ext }}+\mathrm{O} 2_{\text {ext }} \leftrightarrow \mathrm{Gly}_{\text {ext }}+\mathrm{CO} 2_{\text {ext }}+\mathrm{H} 2 \mathrm{O}_{\text {ext }}$ |
| 3 | $2 \mathrm{Phe}_{\text {ext }}+\mathrm{O} 2_{\text {ext }} \leftrightarrow 2 \mathrm{Tyr}_{\text {ext }}$ |
| 4 | $\mathrm{Ile}_{\text {ext }}+3 \mathrm{O} 2_{\text {ext }} \leftrightarrow \mathrm{Glu}_{\text {ext }}+\mathrm{CO} 22_{\text {ext }}+2 \mathrm{H} 2 \mathrm{O}$ ext |
| 5 | Glcext $^{\text {e }}$, $2 \mathrm{Lac}_{\text {ext }}$ |
| 6 | $\mathrm{PrO}_{\text {ext }}+\mathrm{O} 2_{\text {ext }} \leftrightarrow \mathrm{Glu}_{\text {ext }}$ |
| 7 |  |
| 8 | $2 \mathrm{Glu}_{\text {ext }}+3 \mathrm{O} 22_{\text {ext }} \leftrightarrow 2 \mathrm{Asp}$ ext $+2 \mathrm{CO} 2{ }_{\text {ext }}+2 \mathrm{H} 2 \mathrm{O}$ ext |
| 9 | 12 NH 4 ext $+9 \mathrm{CO} 2_{\text {ext }} \leftrightarrow 2 \mathrm{Gly}_{\text {ext }}+5 \mathrm{Urea}_{\text {ext }}+9 \mathrm{H} 2 \mathrm{O}_{\text {ext }}$ |
| 10 | $\mathrm{Val}_{\text {ext }}+3 \mathrm{O} 2_{\text {ext }} \leftrightarrow \mathrm{Asp}_{\text {ext }}+\mathrm{CO} 2_{\text {ext }}+2 \mathrm{H} 2 \mathrm{O}$ ext |
| 11 | $\mathrm{Ala}_{\text {ext }}+3 \mathrm{Asp}_{\text {ext }}+2 \mathrm{Leu}_{\text {ext }}+3 \mathrm{CO} 22_{\text {ext }} \leftrightarrow 6 \mathrm{Glu}_{\text {ext }}$ |
| 12 | $4 \mathrm{Aspext}^{\text {ex }}$, $4 \mathrm{NH} 4_{\text {ext }}+\mathrm{O} 2_{\text {ext }} \leftrightarrow 4 \mathrm{Asn}_{\text {ext }}+6 \mathrm{H} 2 \mathrm{O}_{\text {ext }}$ |
| 13 | $4 \mathrm{Cys}_{\text {ext }}+5 \mathrm{O} 2_{\text {ext }}+10 \mathrm{H} 2 \mathrm{O}$ ext $\leftrightarrow 4 \mathrm{NH} 4_{\text {ext }}+4 \mathrm{Lac}_{\text {ext }}+4 \mathrm{H} 2 \mathrm{SO} 4{ }_{\text {ext }}$ |
| 14 | Aspext $^{\text {a }}$ Ala ${ }_{\text {ext }}+\mathrm{CO} 2{ }_{\text {ext }}$ |
| 15 | Ala ${ }_{\text {ext }}+\mathrm{Asp} \mathrm{exp}_{\text {ext }}+2 \mathrm{Tyr}_{\text {ext }}+7 \mathrm{O} 2_{\text {ext }} \leftrightarrow 4 \mathrm{Glu}_{\text {ext }}+5 \mathrm{CO} 2{ }_{\text {ext }}$ |
| 16 | $2 \mathrm{Cys}_{\text {ext }}+3 \mathrm{O} 2 e_{\text {ext }}+2 \mathrm{H} 2 \mathrm{O}$ ext $\leftrightarrow 2 \mathrm{Ala}$ ext $+2 \mathrm{H} 2 \mathrm{SO} 4_{\text {ext }}$ |
| 17 | Argext $^{+} \mathrm{O} 2_{\text {ext }} \leftrightarrow \mathrm{Gln}_{\text {ext }}+\mathrm{Urea}_{\text {ext }}$ |
| 18 | $\mathrm{Met}_{\text {ext }}+\mathrm{Ser}_{\text {ext }}+\mathrm{Tyr}_{\text {ext }}+6 \mathrm{O} 2$ ext $^{\text {e }}$ ( $\mathrm{Cys}_{\text {ext }}+2 \mathrm{Glu}_{\text {ext }}+4 \mathrm{CO} 2{ }_{\text {ext }}+2 \mathrm{H} 2 \mathrm{O}$ ext |
| 19 | 2 Ala ext $+\mathrm{O} 2_{\text {ext }} \leftrightarrow 2 \mathrm{Ser}_{\text {ext }}$ |
| 20 | $\mathrm{Met}_{\text {ext }}+\mathrm{CO} 2_{\text {ext }}+\mathrm{H} 2 \mathrm{O}$ ext $\leftrightarrow \mathrm{Cys}_{\text {ext }}+\mathrm{Lac}_{\text {ext }}$ |
| 21 |  |
| 22 |  |
| 23 | 0.014204 Glc $_{\text {ext }}+0.03234$ His $_{\text {ext }}+0.00573$ Ile $_{\text {ext }}+0.01167 \mathrm{Leu}_{\text {ext }}+0.013247 \mathrm{Lys}_{\text {ext }}+0.0051595 \mathrm{Met}_{\mathrm{ext}}+$ 0.0082357 Phe $_{\text {ext }}+0.0069828$ Thr $_{\text {ext }}+0.028883 \operatorname{Trp}_{\text {ext }}+0.0083503 \mathrm{Val}_{\text {ext }}+0.0019424 \mathrm{CO} 2_{\text {ext }}+0.00053223$ $\mathrm{Ethn}_{\text {ext }}+0.0015302 \mathrm{Cho}_{\text {ext }}+0.01887 \mathrm{H} 3 \mathrm{PO} 4_{\text {ext }}+0.13401 \mathrm{H} 2 \mathrm{O}_{\text {ext }} \leftrightarrow$ Biomass $_{\text {ext }}$ |
| 24 | $0.014204 \mathrm{Glc}_{\text {ext }}+0.011242 \mathrm{Ala}_{\text {ext }}+0.0013577 \mathrm{Arg}_{\text {ext }}+0.0022176 \mathrm{Cys}_{\text {ext }}+0.0019408$ His $_{\text {ext }}+0.00573 \mathrm{Ile}_{\text {ext }}$ +0.01167 Leu $_{\text {ext }}+0.013247$ Lys $_{\text {ext }}+0.0029419$ Met $_{\text {ext }}+0.0053133$ Phe $_{\text {ext }}+0.038173$ Pro $_{\text {ext }}+0.092324$ Ser $_{\text {ext }}+0.0069828$ Thr $_{\text {ext }}+0.00089628$ Trpext +0.0029223 Tyr $_{\text {ext }}+0.0083503$ Val $_{\text {ext }}+0.00053223$ Ethn $_{\text {ext }}$ +0.0015302 Cho ext +0.01887 H3PO $4_{\text {ext }} \leftrightarrow$ Biomass $_{\text {ext }}+0.013085 \mathrm{CO} 2_{\text {ext }}+0.10061 \mathrm{H} 2 \mathrm{O}_{\text {ext }}$ |
| 25 |  |

## 3. Parameter estimation for Monod-type kinetics

The specific reaction rates related to the BFMs that were presented in the previous section are typically modeled using Monod-type kinetics to account for activating or inhibitory effects of certain chemical species [12]. This means that each specific reaction rate is expressed as a sum of Monod terms, where each Monod term may be subject to activation or inhibition by several species, and the inhibition mechanism may be any combination of competitive, non-competitive, or uncompetitive inhibition. In this section, the goal is to develop methods to identify the activation and inhibition mechanisms and compute globally optimal parameter estimates in an efficient way.

### 3.1. Monod terms

We start by observing that each Monod term $v$ can be written in different ways, depending on the activation and inhibition mechanisms. Consider a particular case, where a certain Monod term is either activated by the species $S$ and inhibited by the species I or activated by the species S and A . In what follows, $v_{\max }$ is the maximum rate of the Monod term, $c_{\mathrm{S}}, c_{\mathrm{A}}$, and $c_{\mathrm{I}}$ are the concentrations of the species $\mathrm{S}, \mathrm{A}$, and I , and $k_{\mathrm{S}}, k_{\mathrm{A}}$, and $k_{\mathrm{I}}$ are the corresponding kinetic constants. Next, we observe how the Monod term would be written in different cases [28].

In the case of no activation by A and no inhibition by I:

$$
\begin{equation*}
v=v_{\max } \frac{c_{\mathrm{S}}}{k_{\mathrm{S}}+c_{\mathrm{S}}} \tag{39}
\end{equation*}
$$

In the case of non-competitive inhibition by I:

$$
\begin{equation*}
v=v_{\max } \frac{c_{\mathrm{S}}}{k_{\mathrm{S}}\left(1+\frac{c_{\mathrm{I}}}{k_{\mathrm{I}}}\right)+c_{\mathrm{S}}\left(1+\frac{c_{\mathrm{I}}}{k_{\mathrm{I}}}\right)} \tag{40}
\end{equation*}
$$

In the case of competitive inhibition by I:

$$
\begin{equation*}
v=v_{\max } \frac{c_{\mathrm{S}}}{k_{\mathrm{S}}\left(1+\frac{c_{\mathrm{I}}}{k_{\mathrm{I}}}\right)+c_{\mathrm{S}}} \tag{41}
\end{equation*}
$$

In the case of uncompetitive inhibition by I:

$$
\begin{equation*}
v=v_{\max } \frac{c_{\mathrm{S}}}{k_{\mathrm{S}}+c_{\mathrm{S}}\left(1+\frac{c_{\mathrm{I}}}{k_{\mathrm{I}}}\right)} \tag{42}
\end{equation*}
$$

In the case of non-cooperative activation by A :

$$
\begin{equation*}
v=v_{\max } \frac{c_{\mathrm{S}} c_{\mathrm{A}}}{k_{\mathrm{S}}\left(c_{\mathrm{A}}+k_{\mathrm{A}}\right)+c_{\mathrm{S}}\left(c_{\mathrm{A}}+k_{\mathrm{A}}\right)} . \tag{43}
\end{equation*}
$$

It is known that all the inhibition mechanisms are particular cases of the so-called mixed inhibition. More generally, the Monod term for mixed inhibition by I is written as:

$$
\begin{equation*}
v=v_{\max } \frac{c_{\mathrm{S}}}{k_{\mathrm{S}}\left(1+\frac{c_{\mathrm{I}}}{k_{\mathrm{I}}}\right)+c_{\mathrm{S}}\left(1+\frac{c_{\mathrm{I}}}{k_{\mathrm{I}}^{I}}\right)} \tag{44}
\end{equation*}
$$

where for the case of no activation and no inhibition one uses $\frac{1}{k_{\mathrm{I}}}=\frac{1}{k_{\mathrm{I}}^{\prime}}=0$, for the case of noncompetitive inhibition one uses $\frac{1}{k_{\mathrm{I}}}=\frac{1}{k_{\mathrm{I}}^{\prime}}$, for the case of competitive inhibition one uses $\frac{1}{k_{\mathrm{T}}^{\prime}}=0$, and for the case of uncompetitive inhibition one uses $\frac{1}{k_{1}}=0$.

Similarly, one can show that, in the case of activation by A, the Monod term is written as:

$$
\begin{equation*}
v=v_{\max } \frac{c_{\mathrm{S}} c_{\mathrm{A}}}{k_{\mathrm{S}}\left(c_{\mathrm{A}}+k_{\mathrm{A}}\right)+c_{\mathrm{S}}\left(c_{\mathrm{A}}+k_{\mathrm{A}}^{\prime}\right)}, \tag{45}
\end{equation*}
$$

where the case of no activation and no inhibition is obtained with $k_{\mathrm{A}}=k_{\mathrm{A}}^{\prime}=0$ and the case of non-cooperative activation is obtained with $k_{\mathrm{A}}=k_{\mathrm{A}}^{\prime}$.

Hence, any Monod term with activation by S and inhibition by I is given by:

$$
\begin{equation*}
v=\frac{b c_{\mathrm{S}}}{c_{\mathrm{S}}+a_{1} c_{\mathrm{S}} c_{\mathrm{I}}+a_{2} c_{\mathrm{I}}+a_{3}}, \tag{46}
\end{equation*}
$$

with unknown parameters $a_{1}=\frac{1}{k_{\mathrm{I}}^{\prime}}, a_{2}=\frac{k_{\mathrm{S}}}{k_{\mathrm{I}}}, a_{3}=k_{\mathrm{S}}, b=v_{\max }$. This parameterization of Monod terms includes all the combinations of inhibition mechanisms without assuming which one is truly present.

In addition, any Monod term with activation by S and A is given by:

$$
\begin{equation*}
v=\frac{b c_{\mathrm{S}} c_{\mathrm{A}}}{c_{\mathrm{S}} c_{\mathrm{A}}+a_{1} c_{\mathrm{S}}+a_{2} c_{\mathrm{A}}+a_{3}} \tag{47}
\end{equation*}
$$

with unknown parameters $a_{1}=k_{\mathrm{A}}^{\prime}, a_{2}=k_{\mathrm{S}}, a_{3}=k_{\mathrm{S}} k_{\mathrm{A}}, b=v_{\max }$.

### 3.2. Monod-type kinetics

If we add several Monod terms to express the rate of a certain reaction, we obtain the so-called Monod-type kinetics. The general expression for the model of a reaction with Monod-type kinetics using $L$ Monod terms is

$$
\begin{equation*}
\tilde{\psi}(t)=e(t)+\sum_{l=1}^{L} \frac{n_{l}(t, \boldsymbol{\theta})}{d_{l}(t, \boldsymbol{\theta})}, \quad t=1, \ldots, N \tag{48}
\end{equation*}
$$

with

$$
\begin{align*}
& n_{l}(t, \boldsymbol{\theta}):=b_{l} \prod_{i \in \mathcal{A}_{l}} c_{i}(t), \quad l=1, \ldots, L, \quad t=1, \ldots, N  \tag{49a}\\
& d_{l}(t, \boldsymbol{\theta}):=\sum_{c=1}^{\left|\mathcal{C}_{l}\right|} a_{l, c-1} \prod_{i \in\left(\mathcal{C}_{l}\right)_{c}} c_{i}(t), \quad l=1, \ldots, L, \quad t=1, \ldots, N \tag{49b}
\end{align*}
$$

where:

- $N$ is the sample size,
- $\tilde{\psi}(t)$ is the specific reaction rate for the sample $t$ corrupted by additive noise,
- $e(t)$ is the additive noise for the sample $t$,

For example, for a reaction rate expressed as a sum of $L=2$ Monod terms, where each term is activated by one species and activated or inhibited by another species, which implies that $\left|\mathcal{S}_{1}\right|=$ $\left|\mathcal{S}_{2}\right|=2$ and $\left|\mathcal{C}_{1}\right|=\left|\mathcal{C}_{2}\right|=4$, the Monod-type kinetics is given by:

$$
\begin{align*}
n_{l}(t, \boldsymbol{\theta}) & :=b_{l} \prod_{i \in \mathcal{A}_{l}} c_{i}(t), \quad l=1,2, \quad t=1, \ldots, N  \tag{50a}\\
d_{l}(t, \boldsymbol{\theta}) & :=\prod_{i \in \mathcal{A}_{l}} c_{i}(t)+a_{l, 1} \prod_{i \in\left(\mathcal{C}_{l}\right)_{2}} c_{i}(t)+a_{l, 2} \prod_{i \in\left(\mathcal{C}_{l}\right)_{3}} c_{i}(t)+a_{l, 3} \prod_{i \in\left(\mathcal{C}_{l}\right)_{4}} c_{i}(t), \quad l=1,2, \quad t=1, \ldots, N . \tag{50b}
\end{align*}
$$

In particular, if the first term is activated by species 1 and inhibited by species 3 , and the second term is activated by species 2 and 4 , then $\mathcal{A}_{1}=\{1\}, \mathcal{A}_{2}=\{2,4\}, \mathcal{S}_{1}=\{1,3\}, \mathcal{S}_{2}=\{2,4\}$, $\mathcal{C}_{1}=\{\{1\},\{1,3\},\{3\},\{ \}\}, \mathcal{C}_{2}=\{\{2,4\},\{2\},\{4\},\{ \}\}$, which means that

$$
\begin{align*}
n_{1}(t, \boldsymbol{\theta}) & :=b_{1} c_{1}(t), \quad n_{2}(t, \boldsymbol{\theta}):=b_{2} c_{2}(t) c_{4}(t), \quad t=1, \ldots, N,  \tag{51a}\\
d_{1}(t, \boldsymbol{\theta}) & :=c_{1}(t)+a_{1,1} c_{1}(t) c_{3}(t)+a_{1,2} c_{3}(t)+a_{1,3}, \\
d_{2}(t, \boldsymbol{\theta}) & :=c_{2}(t) c_{4}(t)+a_{2,1} c_{2}(t)+a_{2,2} c_{4}(t)+a_{2,3}, \quad t=1, \ldots, N . \tag{51b}
\end{align*}
$$

### 3.3. Toward tractable global parameter estimation

Now that the model structure expressed by the Monod-type kinetics has been presented, we would like to estimate the unknown parameters $\boldsymbol{\theta}$ that provide the globally optimal fit of measured rates $\tilde{\psi}(t)$ and predicted rates $\hat{\psi}(t \mid \boldsymbol{\theta}):=\sum_{l=1}^{L} \frac{n_{l}(t, \boldsymbol{\theta})}{d_{l}(t, \boldsymbol{\theta})}$ in a computationally tractable way. In this context, global optimality means that the estimated parameters globally minimize the mean squared error (MSE) $\hat{J}(\boldsymbol{\theta})=\sum_{t=1}^{N} \hat{e}(t \mid \boldsymbol{\theta})^{2} / N$, where $\hat{e}(t \mid \boldsymbol{\theta}):=\tilde{\psi}(t)-\hat{\psi}(t \mid \boldsymbol{\theta})$ is the prediction error. For the sake of simplicity, the prediction error $\hat{e}(t \mid \boldsymbol{\theta})$ and the predicted rate $\hat{\psi}(t \mid \boldsymbol{\theta})$ are denoted as $\hat{e}(t)$ and $\hat{\psi}(t)$, respectively, and the definitions $\hat{\mathbf{e}}:=\left[\begin{array}{lll}\hat{e}(1) & \cdots & \hat{e}(N)\end{array}\right]^{\mathrm{T}}$ and $\hat{\boldsymbol{\psi}}:=\left[\begin{array}{lll}\hat{\psi}(1) & \cdots & \hat{\psi}(N)\end{array}\right]^{\mathrm{T}}$ are used in the remainder of this paper. Hence, the parameters $\boldsymbol{\theta}:=\left(a_{1,1}, \ldots, a_{1,\left|\mathcal{C}_{1}\right|-1}, \ldots, a_{L, 1}, \ldots, a_{L,\left|\mathcal{C}_{L}\right|-1}\right.$, $\left.b_{1}, \ldots, b_{L}\right)$ are the global solution to the problem

$$
\begin{align*}
& \min _{\hat{\mathbf{e}}, \boldsymbol{\theta}} \sum_{t=1}^{N} \frac{\hat{e}(t)^{2}}{N}  \tag{52a}\\
& \text { s.t. } \hat{e}(t)=\tilde{\psi}(t)-\sum_{l=1}^{L} \frac{n_{l}(t, \boldsymbol{\theta})}{d_{l}(t, \boldsymbol{\theta})}, \quad t=1, \ldots, N  \tag{52b}\\
& \quad a_{1,1} \geq 0, \quad \ldots \quad a_{1,\left|\mathcal{C}_{1}\right|-1} \geq 0, \quad \ldots \quad a_{L, 1} \geq 0, \quad \ldots \quad a_{L,\left|\mathcal{C}_{L}\right|-1} \geq 0 \tag{52c}
\end{align*}
$$

with $n_{l}(t, \boldsymbol{\theta})$ and $d_{l}(t, \boldsymbol{\theta})$ defined as in (49). An equivalent reformulation of this problem is

$$
\begin{align*}
& \min _{\hat{\mathbf{e}}, \hat{\mathbf{d}}, \boldsymbol{\theta}} \sum_{t=1}^{N} \frac{\hat{e}(t)^{2}}{N},  \tag{53a}\\
& \text { s.t. } \prod_{l_{l=1}^{L}}^{L} d_{l}(t, \boldsymbol{\theta})-\hat{d}(t)=0, \quad t=1, \ldots, N,  \tag{53b}\\
& \sum_{l=1}^{L} n_{l}(t, \boldsymbol{\theta}) \prod_{\substack{l^{\prime}=1 \\
l^{\prime} \neq l}}^{L} d_{l^{\prime}}(t, \boldsymbol{\theta})+(\hat{e}(t)-\tilde{\psi}(t)) \hat{d}(t)=0, \quad t=1, \ldots, N,  \tag{53c}\\
& a_{1,1} \geq 0, \quad \cdots \quad a_{1,\left|\mathcal{C}_{1}\right|-1} \geq 0, \quad \ldots \quad a_{L, 1} \geq 0, \quad \ldots \quad a_{L,\left|\mathcal{C}_{L}\right|-1} \geq 0, \tag{53d}
\end{align*}
$$

with the common denominators $\hat{\mathbf{d}}:=\left[\begin{array}{lll}\hat{d}(1) & \cdots & \hat{d}(N)\end{array}\right]^{\mathrm{T}}$. With this reformulation, not only the cost but also the constraints become polynomial functions of the decision variables since $n_{l}(t, \boldsymbol{\theta})$ and $d_{l}(t, \boldsymbol{\theta})$ are linear in $\boldsymbol{\theta}$.

In the remainder, we assume that the Monod-type kinetics for a reaction rate is expressed as a sum of $L=2$ Monod terms, where each term is activated by one species and activated or inhibited by another species, which implies that the parameters are $\boldsymbol{\theta}:=\left(a_{1,1}, a_{1,2}, a_{1,3}, a_{2,1}, a_{2,2}, a_{2,3}, b_{1}, b_{2}\right)$, with $a_{1,1}, a_{1,2}, a_{1,3}, a_{2,1}, a_{2,2}, a_{2,3}$ nonnegative. The use of 2 Monod terms is appropriate not only to make the problem more computationally tractable, but also because the addition of more parameters to the model would risk to overparameterize the model, especially if we consider that the available sample sizes are typically small. In addition, a model with 2 Monod terms is still sufficiently complex to capture the potential case of one Monod term for a forward reaction and another Monod term for the corresponding backward reaction.

For this, we compute the global solution to the following optimization problem:

$$
\begin{align*}
& \min _{\hat{\mathbf{e}}, \hat{\mathbf{d}}, \boldsymbol{\theta}} \sum_{t=1}^{N} \frac{\hat{e}(t)^{2}}{N},  \tag{54a}\\
& \text { s.t. } d_{1}(t, \boldsymbol{\theta}) d_{2}(t, \boldsymbol{\theta})-\hat{d}(t)=0, \quad t=1, \ldots, N,  \tag{54b}\\
& \quad n_{1}(t, \boldsymbol{\theta}) d_{2}(t, \boldsymbol{\theta})+n_{2}(t, \boldsymbol{\theta}) d_{1}(t, \boldsymbol{\theta})+(\hat{e}(t)-\tilde{\psi}(t)) \hat{d}(t)=0, \quad t=1, \ldots, N,  \tag{54c}\\
& \quad a_{1,1} \geq 0, \quad a_{1,2} \geq 0, \quad a_{1,3} \geq 0, \quad a_{2,1} \geq 0, \quad a_{2,2} \geq 0, \quad a_{2,3} \geq 0, \tag{54~d}
\end{align*}
$$

with $n_{l}(t, \boldsymbol{\theta})$ and $d_{l}(t, \boldsymbol{\theta})$ defined as in (50). In this problem, the cost and constraints are quadratic functions of the decision variables since $n_{l}(t, \boldsymbol{\theta})$ and $d_{l}(t, \boldsymbol{\theta})$ are linear in $\boldsymbol{\theta}$.

For example, if the first term is activated by species 1 and inhibited by species 3 , and the second term is activated by species 2 and 4 , the optimization problem is (54), with $n_{l}(t, \boldsymbol{\theta})$ and $d_{l}(t, \boldsymbol{\theta})$ defined as in (51).

To reduce the model overparameterization in the case of small sample sizes, one may consider using the sample variance $\hat{\sigma}^{2}(\boldsymbol{\theta})=\sum_{t=1}^{N} \hat{e}(t)^{2} /\left(N-n_{p}\right)$ as the cost instead of the MSE $\hat{J}(\boldsymbol{\theta})$ in a final step of the estimation procedure, where $n_{p}$ is the number of nonzero parameters.

If the additive noise $e(t)$ that corrupts the measured rates $\tilde{\psi}(t)$ is independent, identically distributed (i.i.d.) zero-mean Gaussian noise with variance $\sigma^{2}$, the global solution to the optimization
problem (54) provides maximum-likelihood estimates. In this case, the Cramér-Rao lower bound

$$
\begin{equation*}
\operatorname{Var}[\hat{\boldsymbol{\theta}}] \succeq \sigma^{2}\left(\frac{\partial \hat{\boldsymbol{\psi}}}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta})^{\mathrm{T}} \frac{\partial \hat{\boldsymbol{\psi}}}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta})\right)^{-1} \tag{55}
\end{equation*}
$$

with $\hat{\boldsymbol{\psi}}$ defined as in the first paragraph of this section, allows one to achieve the lower bound for the variance of the estimator $\boldsymbol{\theta}$ that corresponds to the estimate $\boldsymbol{\theta}$ when the problem (54) is solved to global optimality. In (55), the true variance $\sigma^{2}$ can be replaced by $\hat{\sigma}^{2}(\boldsymbol{\theta})$ if $\sigma^{2}$ is unknown.

To solve (54) efficiently to global optimality, one can reformulate it via the concept of sum-ofsquares (SOS) polynomials as a hierarchy of sparse linear matrix inequality (LMI) feasibility problems of increasing relaxation order $d$. This reformulation as a hierarchy of sparse SOS relaxations using semidefinite programs (SDPs) is detailed in Appendix B. Note that a similar reformulation has been proposed in the context of maximum-likelihood and Bayesian point estimation problems for linear models that share a similar rational structure $[29,30]$.

This reformulation takes advantage of the fact that, in the problem (54), each equality constraint corresponds to a quadratic polynomial in the decision variables $\hat{\mathbf{e}}, \mathbf{d}, \boldsymbol{\theta}$ that involves at most $\left|\mathcal{C}_{1}\right|+$ $\left|\mathcal{C}_{2}\right|+2=10$ of these variables, and the cost function can be written as a sum of quadratic polynomials that involve only a few variables. This allows the use of sparse semidefinite relaxations if each equality constraint is transformed into a pair of inequality constraints to obtain a basic semi-algebraic set. For this, we need to define $p$ index subsets $I_{k}$ with the corresponding $n_{k}:=$ $\left|I_{k}\right|$ variables $\mathbf{x}\left(I_{k}\right)=\left\{x_{i}: i \in I_{k}\right\}$, for $k=1, \ldots, p$, such that $\cup_{k=1}^{p} I_{k}=\{1, \ldots, n\}$, where $n$ is the number of decision variables, such that the index subsets $I_{1}, \ldots, I_{p}$ satisfy the conditions in Theorem 8 in Appendix B.

Hence, for problem (54), we introduce the following definitions:

$$
\begin{align*}
f(\mathbf{x}) & :=J(\mathbf{x})-\tau,  \tag{56a}\\
g_{j}(\mathbf{x}) & := \begin{cases}-h_{j}^{d}(\mathbf{x}), & j=1, \ldots, N, \\
-h_{j-N}^{e}(\mathbf{x}), & j=N+1, \ldots, 2 N, \\
h_{j-2 N}^{d}(\mathbf{x}), & j=2 N+1, \ldots, 3 N, \\
h_{j-3 N}^{e}(\mathbf{x}), & j=3 N+1, \ldots, 4 N, \\
\theta_{j-4 N}, & j=4 N+1, \ldots, 4 N+6,\end{cases} \tag{56b}
\end{align*}
$$

with $\mathbf{x}:=(\hat{\mathbf{e}}, \hat{\mathbf{d}}, \boldsymbol{\theta})=\left(\hat{e}(1), \ldots, \hat{e}(N), \hat{d}(1), \ldots, \hat{d}(N), a_{1,1}, a_{1,2}, a_{1,3}, a_{2,1}, a_{2,2}, a_{2,3}, b_{1}, b_{2}\right)$ and

$$
\begin{align*}
J(\mathbf{x}): & =\sum_{t=1}^{N} \frac{\hat{e}(t)^{2}}{N},  \tag{56c}\\
h_{t}^{d}(\mathbf{x}): & =\left(\prod_{i \in \mathcal{A}_{1}} c_{i}(t)+a_{1,1} \prod_{i \in\left(\mathcal{C}_{1}\right)_{2}} c_{i}(t)+a_{1,2} \prod_{i \in\left(\mathcal{C}_{1}\right)_{3}} c_{i}(t)+a_{1,3} \prod_{i \in\left(\mathcal{C}_{1}\right)_{4}} c_{i}(t)\right) \\
& \left(\prod_{i \in \mathcal{A}_{2}} c_{i}(t)+a_{2,1} \prod_{i \in\left(\mathcal{C}_{2}\right)_{2}} c_{i}(t)+a_{2,2} \prod_{i \in\left(\mathcal{C}_{2}\right)_{3}} c_{i}(t)+a_{2,3} \prod_{i \in\left(\mathcal{C}_{2}\right)_{4}} c_{i}(t)\right) \\
& -\hat{d}(t), \quad t=1, \ldots, N,  \tag{56d}\\
h_{t}^{e}(\mathbf{x}): & =b_{1}\left(\prod_{i \in \mathcal{A}_{2}} c_{i}(t)+a_{2,1} \prod_{i \in\left(\mathcal{C}_{2}\right)_{2}} c_{i}(t)+a_{2,2} \prod_{i \in\left(\mathcal{C}_{2}\right)_{3}} c_{i}(t)+a_{2,3} \prod_{i \in\left(\mathcal{C}_{2}\right)_{4}} c_{i}(t)\right) \prod_{i \in \mathcal{A}_{1}} c_{i}(t) \\
& +b_{2}\left(\prod_{i \in \mathcal{A}_{1}} c_{i}(t)+a_{1,1} \prod_{i \in\left(\mathcal{C}_{1}\right)_{2}} c_{i}(t)+a_{1,2} \prod_{i \in\left(\mathcal{C}_{1}\right)_{3}} c_{i}(t)+a_{1,3} \prod_{i \in\left(\mathcal{C}_{1}\right)_{4}} c_{i}(t)\right) \prod_{i \in \mathcal{A}_{2}} c_{i}(t) \\
& +(\hat{e}(t)-\tilde{\psi}(t)) \hat{d}(t), \quad t=1, \ldots, N . \tag{56e}
\end{align*}
$$

Then, problem (54) is equivalent to that of computing the maximum $\tau$ such that $f(\mathbf{x})$ is strictly positive $\forall \mathbf{x} \in \mathbb{K}=\left\{\mathbf{x}: g_{j}(\mathbf{x}) \geq 0, \forall j=1, \ldots, 4 N+6\right\}$. The previous definitions seem to suggest the choice of $n_{k}=\left|\mathcal{C}_{1}\right|+\left|\mathcal{C}_{2}\right|+2=10$ variables $\mathbf{x}\left(I_{k}\right)=(\hat{e}(k), \hat{d}(k), \boldsymbol{\theta})$ and corresponding index subsets $I_{k}=\{k, N+k, 2 N+1, \ldots, 2 N+8\}$, for $k=1, \ldots, p$, with $p:=N$. To satisfy Condition 4 of Theorem 8 in Appendix B for these index subsets, additional constraints must be added. For this, we redefine $\mathbb{K}=\left\{\mathbf{x}: g_{j}(\mathbf{x}) \geq 0, \forall j=1, \ldots, m\right\}$, with $m:=5 N+6$, by adding the quadratic polynomials

$$
\begin{equation*}
g_{j}(\mathbf{x}):=-\bar{h}_{j-4 N-6}(\mathbf{x}), \quad j=4 N+7, \ldots, m \tag{56f}
\end{equation*}
$$

with

$$
\begin{equation*}
\bar{h}_{t}(\mathbf{x})=-r^{2}+\hat{e}(t)^{2}+\hat{d}(t)^{2}+a_{1,1}^{2}+a_{1,2}^{2}+a_{1,3}^{2}+a_{2,1}^{2}+a_{2,2}^{2}+a_{2,3}^{2}+b_{1}^{2}+b_{2}^{2}, \quad t=1, \ldots, N, \tag{56~g}
\end{equation*}
$$

where $r$ is some finite constant. If $r$ is chosen large enough to ensure that the minimizers $\mathbf{x}^{*}$ of problem (54) are such that $\left\|\mathbf{x}\left(I_{k}\right)^{*}\right\| \leq r$, for $k=1, \ldots, p$, then the new constraints are redundant because adding them does not change the minimizers. Moreover, the polynomials ( 56 g ) are chosen to be quadratic since the polynomials with compact superlevel sets (as in Condition 4 of Theorem 8) are at least of degree 2 and the polynomials that specify the other constraints $g_{j}(\mathbf{x}) \geq 0$ are also of degree $2 v_{j}=2$.

Some comments about the boundedness of $\left\|\mathbf{x}\left(I_{k}\right)^{*}\right\|$, for $k=1, \ldots, p$, are necessary at this point. Since $\mathbf{x}\left(I_{k}\right)=(\hat{e}(k), \hat{d}(k), \boldsymbol{\theta})$, this boundedness implies that the prediction error $\hat{e}(t)$, the prediction common denominator $\hat{d}(t)$, and the parameters $\boldsymbol{\theta}$ are bounded. It seems reasonable to assume that the parameters $\boldsymbol{\theta}$ are bounded if the data are scaled to units such that the magnitude of the non-zero parameters is expected to be approximately the same as the magnitude of 1 , that is, below 2 but not too close to 0 . Regarding the prediction errors, they are expected to have the same magnitude as the noise realizations, which are assumed to be realizations of a normally distributed random variable with zero mean and variance $\sigma^{2}$ in this paper. More precisely, in our implementation of the estimation procedure, we assume that the sum of squares of the variables $\mathbf{x}\left(I_{k}\right)$ is bounded by $r^{2}=2^{2}\left(\left|\mathcal{C}_{1}\right|+\left|\mathcal{C}_{2}\right|\right)=2^{2}(4+4)=32$.

Since all the conditions in Theorem 8 in Appendix B are satisfied for the formulation (56) and the index subsets $I_{k}=\{k, N+k, 2 N+1, \ldots, 2 N+8\}$, for $k=1, \ldots, p$, it is possible to state the following theorem:

Theorem 1. Solving problem (54) to global optimality is equivalent to solving the $S D P$ (B.3) for some integer relaxation order $d \geq 1$, with $f(\mathbf{x})$ and $g_{j}(\mathbf{x})$, for $j=1, \ldots, m$, given in (56).

Proof. This follows from the fact that solving problem (54) to global optimality is equivalent to the problem that consists in computing the global minimum of $J(\mathbf{x})$ subject to $g_{j}(\mathbf{x}) \geq 0$, for $j=1, \ldots, m$, and thus can be formulated as the $\operatorname{SDP}$ (B.3) for some integer relaxation order $d \geq v_{j}=1$ as described in Appendix B.

A certificate of the representation in terms of SOS polynomials for some order $d$ can be obtained upon convergence of the SDP as shown in Theorem 9 in Appendix B, which is a certificate of global optimality of the solution $\mathbf{x}^{*}:=\left(\hat{\mathbf{e}}^{*}, \hat{\mathbf{d}}^{*}, \boldsymbol{\theta}^{*}\right)$ and the MSE $\tau^{*}=J^{*}$. This certification uses the rank deficiency of the coefficient matrices $\mathbf{Q}_{0, k}^{*}$ and the rank of the moment matrices $\mathbf{L}_{0, k}^{*}$, for $k=1, \ldots, p$, which are in the primal and dual solutions to the SDP, respectively. If a unique global
solution can be certified, the rank deficiency of the coefficient matrices and the rank of the moment matrices are equal to 1 , as shown in Theorem 9.

If the certificate of global optimality is obtained, the globally optimal solution $\mathbf{x}^{*}$ is extracted from the null space of $\mathbf{Q}_{0, k}^{*}$ and the row space of $\mathbf{L}_{0, k}^{*}$, for $k=1, \ldots, p$. If the certificate of global optimality is not obtained, there are two options: either (i) the relaxation order $d$ is incremented and a larger SDP is formulated and solved, or (ii) the relaxation order $d$ is considered to be sufficiently large (for example, equal to some bound $d_{\max }$ ) such that increasing the relaxation order would lead to an SDP with an excessively large size. Then, in the latter option (ii), one ignores the fact that the rank of the moment matrices and the rank deficiency of the coefficient matrices are not equal to 1 and extracts a unique solution $\mathbf{x}^{*}$ as if the null space of $\mathbf{Q}_{0, k}^{*}$ and the row space of $\mathbf{L}_{0, k}^{*}$ were of dimension 1. Although this solution is not guaranteed to be the global solution, its optimality gap can be computed, as discussed in Remark 2 in Appendix B.

Hence, an important issue is the size of the SDP that needs to be solved to compute and certify a global optimum for a given relaxation order $d$ since the size of the SDP strongly affects the tractability of this problem. The following theorem summarizes the problem size for a given relaxation order $d$.

Theorem 2. Suppose that a global optimum for problem (54) is computed and certified for some relaxation order $d \geq 1$. Then, an $S D P$ with $N\binom{10+2 d}{10}-(N-1)\binom{8+2 d}{8}$ equality constraints, $N$ LMIs of size $\binom{10+d}{10}$, and $5 N+6$ LMIs of size $\binom{9+d}{10}$ has been solved.
Proof. This results from the fact that the SDP (B.3) has been solved for $d \geq v_{j}=1$, which is an SDP with $p\binom{n_{k}+2 d}{n_{k}}-\sum_{k=1}^{p-1}\binom{\left|I_{k} \cap I_{k+1}\right|+2 d}{\left|I_{k} \cap I_{k+1}\right|}=N\binom{10+2 d}{10}-(N-1)\binom{8+2 d}{8}$ equality constraints, $p=N$ LMIs of size $\binom{n_{k}+d}{n_{k}}=\binom{10+d}{10}$, and $m=5 N+6$ LMIs of size $\binom{n_{k}+d-v_{j}}{n_{k}}=\binom{9+d}{10}$ since $n_{k}=10$, $\left|I_{k} \cap I_{k+1}\right|=8$, and $v_{j}=1$.

Note that, thanks to the sparse representation, the input size of this SDP is linear in the sample size $N$ for any order $d$. In particular, now suppose that a global optimum for problem (54) is computed and certified for the relaxation order $d=2$ (in fact, this is always the case in the example of Section 3.5). This implies that an SDP with $506 N+495$ equality constraints, $N$ LMIs of size 66 , and $5 N+6$ LMIs of size 11 has been solved. Since the complexity of SDPs is polynomial in their input size, that is, the number of constraints and the size of the LMIs, it means that it has been possible to compute and certify a global solution $\mathbf{x}^{*}$ in polynomial time.

The procedure for global parameter estimation of Monod-type kinetics is summarized in Algorithm 1.

### 3.4. Use of the model for simulation, control, and optimization

As mentioned throughout this section, the goal of the described procedure is to identify the correct model and compute maximum-likelihood parameter estimates for the reaction rate of each BFM , that is, for each macroscopic reaction in the bioreactor. Suppose that the parametric model $\hat{\psi}(t \mid \boldsymbol{\theta})$ is identified for one BFM at a time. The variance of the function $\hat{\psi}(t \mid \hat{\boldsymbol{\theta}})$ of the estimator $\hat{\boldsymbol{\theta}}$ is given by Gauss' approximation formula $\operatorname{Var}[\hat{\psi}(t \mid \hat{\boldsymbol{\theta}})] \simeq \frac{\partial \hat{\psi}}{\partial \boldsymbol{\theta}}(t \mid \boldsymbol{\theta}) \operatorname{Var}[\hat{\boldsymbol{\theta}}] \frac{\partial \hat{\psi}}{\partial \boldsymbol{\theta}}(t \mid \boldsymbol{\theta})^{\mathrm{T}}$, with a lower bound for $\operatorname{Var}[\hat{\boldsymbol{\theta}}]$ given by (55). Then, it is possible to obtain the column vector $\hat{\boldsymbol{\psi}}(t \mid \boldsymbol{\Theta})$ of reaction rates of all the BFMs with elements $\hat{\psi}(t \mid \boldsymbol{\theta})$ and the covariance matrix $\operatorname{Var}[\hat{\boldsymbol{\psi}}(t \mid \hat{\boldsymbol{\Theta}})]$ that consists in a diagonal matrix with diagonal elements $\operatorname{Var}[\hat{\psi}(t \mid \hat{\boldsymbol{\theta}})]$, where $\boldsymbol{\Theta}$ denotes the matrix of parameters for all BFMs

```
Algorithm 1: Procedure for global parameter estimation of Monod-type kinetics.
    Input: Specific reaction rate \(\tilde{\psi}(t)\) and concentrations \(\mathbf{c}(t)\) for the samples \(t=1, \ldots, N\), and
                    sets \(\mathcal{A}_{l}\) of activating species and \(\mathcal{C}_{l}\) of combinations of elements of the set of
                    activating and inhibiting species for the Monod terms \(l=1,2\).
    Output: MSE \(J^{*}\) and estimated parameters \(\boldsymbol{\theta}^{*}:=\left(a_{1,1}^{*}, a_{1,2}^{*}, a_{1,3}^{*}, a_{2,1}^{*}, a_{2,2}^{*}, a_{2,3}^{*}, b_{1}^{*}, b_{2}^{*}\right)\).
    Compute coefficients of the polynomials \(f(\mathbf{x})\) and \(g_{j}(\mathbf{x})\) in the variables \(\mathbf{x}:=(\hat{\mathbf{e}}, \hat{\mathbf{d}}, \boldsymbol{\theta})\), for
    \(j=1, \ldots, m\), as given in (56), and fix an integer relaxation order \(d \geq 1\) (for example \(d=2\) )
    and a maximum relaxation order \(d_{\max }\).
    2 Formulate SDP (B.3) for the order \(d\) using the computed coefficients of \(f(\mathbf{x})\) and \(g_{j}(\mathbf{x})\), for
    \(j=1, \ldots, m\).
    Solve SDP (B.3).
    Use the solution to the SDP as described in Theorem 9 to check whether a solution
    \(\mathbf{x}^{*}:=\left(\hat{\mathbf{e}}^{*}, \hat{\mathbf{d}}^{*}, \boldsymbol{\theta}^{*}\right)\) and an MSE \(\tau^{*}=J^{*}\) with certified global optimality can be obtained from
    the solution to the SDP.
    if Global optimality is certified then
        Compute the global minimum \(\tau^{*}=J^{*}\) and global minimizers \(\mathbf{x}^{*}\) (including \(\boldsymbol{\theta}^{*}\) ) from the
        null space of \(\mathbf{Q}_{0, k}^{*}\) and the row space of \(\mathbf{L}_{0, k}^{*}\), for \(k=1, \ldots, p\), as described in Theorem 9 .
    else if \(d=d_{\text {max }}\) then
        Compute \(\tau^{*}=J^{*}\) and \(\mathbf{x}^{*}\) (including \(\boldsymbol{\theta}^{*}\) ) without certification of global optimality from
        the null space of \(\mathbf{Q}_{0, k}^{*}\) and the row space of \(\mathbf{L}_{0, k}^{*}\), for \(k=1, \ldots, p\), as if these spaces were
        of dimension 1.
    else
        Set \(d \leftarrow d+1\) and return to Step 2 .
```

and $\hat{\boldsymbol{\Theta}}$ denotes its estimator. Finally, one can compute $\hat{\mathbf{q}}(t \mid \boldsymbol{\Theta})=\mathbf{N}_{m}^{\mathrm{T}} \hat{\boldsymbol{\psi}}(t \mid \boldsymbol{\Theta})$ for the rates of variation of all extracellular species and the covariance matrix $\operatorname{Var}[\hat{\mathbf{q}}(t \mid \hat{\boldsymbol{\Theta}})]=\mathbf{N}_{m}^{\mathrm{T}} \operatorname{Var}[\hat{\boldsymbol{\psi}}(t \mid \hat{\boldsymbol{\Theta}})] \mathbf{N}_{m}$.

However, note that this parametric model does not guarantee that the irreversibility constraints for the irreversible reactions are satisfied. To make sure that the model used for simulation, control, and optimization satisfies these constraints, the following procedure can be used. In problem (37), the measurements $\tilde{\mathbf{q}}_{a}(t)$ are replaced by the parametric model $\mathbf{S}_{a} \hat{\mathbf{q}}(t \mid \boldsymbol{\Theta})$ and the noise covariance matrix $\boldsymbol{\Sigma}_{\tilde{\mathbf{q}}_{a}}$ is replaced by $\mathbf{S}_{a} \operatorname{Var}[\hat{\mathbf{q}}(t \mid \hat{\boldsymbol{\Theta}})] \mathbf{S}_{a}^{\mathrm{T}}$, which results in the problem

$$
\begin{align*}
\min _{\hat{\boldsymbol{\psi}}(t), \hat{\mathbf{r}}_{r}(t)} & \left(\mathbf{S}_{a} \mathbf{N}_{m}^{\mathrm{T}} \hat{\boldsymbol{\psi}}(t)-\mathbf{S}_{a} \hat{\mathbf{q}}(t \mid \boldsymbol{\Theta})\right)^{\mathrm{T}}\left(\mathbf{S}_{a} \operatorname{Var}[\hat{\mathbf{q}}(t \mid \hat{\boldsymbol{\Theta}})] \mathbf{S}_{a}^{\mathrm{T}}\right)^{-1}\left(\mathbf{S}_{a} \mathbf{N}_{m}^{\mathrm{T}} \hat{\boldsymbol{\psi}}(t)-\mathbf{S}_{a} \hat{\mathbf{q}}(t \mid \boldsymbol{\Theta})\right),  \tag{57a}\\
\text { s.t. } & \mathbf{S}_{i r}\left[\mathbf{L}_{N}\left(\mathbf{L}_{N}^{\mathrm{T}} \mathbf{L}_{N}\right)^{-1} \mathbf{E}_{m} \quad \mathbf{K}_{N}\right]\left[\begin{array}{l}
\hat{\boldsymbol{\psi}}(t) \\
\hat{\mathbf{r}}_{r}(t)
\end{array}\right] \geq \mathbf{0}_{R_{i r}}, \quad t=1, \ldots, N . \tag{57b}
\end{align*}
$$

From the solution $\hat{\boldsymbol{\psi}}^{*}(t), \hat{\mathbf{r}}_{r}^{*}(t)$ to this optimization problem, one can then compute $\hat{\mathbf{q}}^{*}(t)=$ $\mathbf{N}_{m}^{\mathrm{T}} \hat{\boldsymbol{\psi}}^{*}(t), \hat{\mathbf{r}}^{*}(t)=\mathbf{E}_{m} \hat{\boldsymbol{\psi}}^{*}(t)$, and $\hat{\mathbf{r}}_{d}^{*}(t)=\mathbf{L}_{N}\left(\mathbf{L}_{N}^{\mathrm{T}} \mathbf{L}_{N}\right)^{-1} \mathbf{E}_{m} \hat{\boldsymbol{\psi}}^{*}(t)+\mathbf{K}_{N} \hat{\mathbf{r}}_{r}^{*}(t)$. Note that this solution is guaranteed to satisfy the irreversibility constraints and can be used for simulation, control, and optimization.

Another relevant issue is related to the fact that the described procedure is able to compute maximum-likelihood parameter estimates for the reaction rate of each BFM, but in general it does
not compute maximum-likelihood parameter estimates for the rates of variation of the extracellular species, when in fact we are interested in predicting the evolution of these extracellular species. For this reason, one may wonder at this point why it is not ensured during the execution of the model identification procedure that (i) the irreversibility constraints are guaranteed for all the possible concentration values and (ii) the computation of maximum-likelihood parameter estimates is performed for the extracellular species rather than for each BFM.

In both cases, the answer is related to intractability of the problem that would be obtained if one tried to achieve (i) and (ii) at the same time. In that case, instead of solving the smaller problem (52) individually and separately for each BFM, one would have to solve the problem that can be obtained by replacing $\hat{\boldsymbol{\psi}}_{r}(t)$ by the parametric model $\hat{\boldsymbol{\psi}}_{r}\left(t \mid \boldsymbol{\Theta}_{r}\right)$ in problem (38) and minimizing the MSE as in problem (52), as follows:

$$
\begin{align*}
\min _{\boldsymbol{\Theta}_{r}} & \sum_{t=1}^{N} \frac{1}{N}\left(\mathbf{S}_{a} \mathbf{N}_{m}^{\mathrm{T}} \mathbf{M}_{r} \hat{\boldsymbol{\psi}}_{r}\left(t \mid \boldsymbol{\Theta}_{r}\right)-\tilde{\mathbf{q}}_{a}(t)\right)^{\mathrm{T}} \boldsymbol{\Sigma}_{\tilde{\mathbf{q}}_{a}}^{-1}\left(\mathbf{S}_{a} \mathbf{N}_{m}^{\mathrm{T}} \mathbf{M}_{r} \hat{\boldsymbol{\psi}}_{r}\left(t \mid \boldsymbol{\Theta}_{r}\right)-\tilde{\mathbf{q}}_{a}(t)\right),  \tag{58a}\\
\text { s.t. } & \mathbf{S}_{i r} \mathbf{E}_{m}^{r} \hat{\boldsymbol{\psi}}_{r}\left(t \mid \boldsymbol{\Theta}_{r}\right) \geq \mathbf{0}_{R_{i r}}, \forall t . \tag{58b}
\end{align*}
$$

This is clearly a much larger problem since it requires identifying the model and estimating the parameters $\boldsymbol{\Theta}_{r}$ simultaneously for the entire reaction network. If we consider the nonlinearity of the parametric model $\hat{\boldsymbol{\psi}}_{r}\left(t \mid \boldsymbol{\Theta}_{r}\right)$ with respect to most parameters $\boldsymbol{\Theta}_{r}$ and the large number of possible models for each BFM, which results in a nonconvex problem with an extremely large number of possible models for the entire reaction network, we conclude that problem (58) is indeed intractable. Consequently, although the procedure in this section does not guarantee the irreversibility constraints for all the possible concentration values during the execution of the model identification procedure and does not yield statistically optimal parameter estimates for the rates of variation of the extracellular species, it seems to be the best available option to identify the model of perfusion bioreactors. Nevertheless, when the model is used subsequently for simulation, control, and optimization, it is possible to ensure that the irreversibility constraints are satisfied.

Note that the same paradigm has already been proposed for identification of reaction kinetics using the so-called incremental approach, in which the reaction rates are first estimated from concentration data and then used to identify the model structure and estimate parameters for one reaction at a time [31, 32]. This incremental approach was indeed suggested as a way to deal with the computational complexity associated with the simultaneous approach, in which the identification of the model structure and parameter estimation is performed by fitting the prediction of the full kinetic model for all the reactions to the experimental data.

The following subsections show that, although the model identification and parameter estimation for the Monod-type kinetics that is used to describe each BFM remains a challenging problem, it is a tractable problem for which good results can be achieved.

### 3.5. Simulation example

In this simulation example, we suppose that:

- The true specific reaction rate is given by two Monod terms.
- The first term is subject to activation by species 1 and uncompetitive inhibition by species 3 $\left(v_{\max , 1}=2, k_{\mathrm{S}, 1}=0.24, k_{\mathrm{I}, 1}=1.12\right)$.
- The second term is subject to non-cooperative activation by species 2 and 4 ( $v_{\text {max,2 }}=1$, $\left.k_{\mathrm{S}, 2}=0.48, k_{\mathrm{A}, 2}=0.56\right)$.
- The measured rate $\tilde{\psi}(t)$ is corrupted by additive noise $e(t)$.
- The sample size is $N=16$.

These assumptions correspond to the model (51), with $a_{1,1}=\frac{1}{k_{\mathrm{I}, 1}}, a_{1,2}=0, a_{1,3}=k_{\mathrm{S}, 1}$, $a_{2,1}=k_{\mathrm{A}, 2}, a_{2,2}=k_{\mathrm{S}, 2}, a_{2,3}=k_{\mathrm{S}, 2} k_{\mathrm{A}, 2}, b_{1}=v_{\max , 1}, b_{2}=v_{\max , 2}$. Furthermore, the concentrations $c_{i}(t)$ of the species $i=1,2,3,4$ are generated randomly, with $\log _{10}\left(c_{i}(t)\right)$ being realizations of a random variable with mean $\log _{10}(0.4)$ and variance $0.5^{2}$, and the additive noise $e(t)$ is i.i.d. Gaussian noise with zero mean and variance $\sigma^{2}=0.01^{2}$.

Then, the measured rates $\tilde{\psi}(t)$ and the corresponding concentrations $c_{i}(t)$ are used by the proposed approach to formulate the $\operatorname{SDP}(\mathrm{B} .3)$, where $f(\mathbf{x})$ and $g_{j}(\mathbf{x})$, for $j=1, \ldots, m$, are given in (56). The implementation was performed on MATLAB R2018a running on an Intel Core i7 1.9 GHz processor, and MOSEK 8.1 was used as SDP solver. This procedure yields the following result:

```
Testing uncompetitive inhibition for Monod term 1, non-cooperative activation for Monod term 2:
With measurement noise:
Minimizing with polynomial basis up to degree 2:
Status returned by the solver: OPTIMAL
Optimal value of the cost function: 7.8169e-04
Dual solution: moment matrices with rank 1 and 1
Primal solution: coefficient matrix with rank deficiency 1
MSE =
4.8856e-05
sol =
0.9031 0 0.2424 0.5472 0.5018 0.3240 2.0254 1.0083
true_sol =
0.8929 0 0.2400 0.5600 0.4800 0.2688 2.0000 1.0000
```

As one can observe, the estimated parameters in sol are similar to the true parameters true_sol, and the MSE is of the same order of magnitude as the variance $\sigma^{2}$. Similar results are obtained for all combinations of activation and inhibition mechanisms (no activation and no inhibition, non-competitive, competitive, and uncompetitive inhibition, and non-cooperative activation). Furthermore, each problem is solved to global optimality and a unique global solution is computed in approximately 20 s from the solution to the SDP for the relaxation order $d=2$. This is certified by the fact that the coefficient matrices $\mathbf{Q}_{0, k}^{*}$ are of rank deficiency 1 and the moment matrices $\mathbf{L}_{0, k}^{*}$ are of rank 1 , for $k=1, \ldots, p$ (see Theorem 9 in Appendix B).

### 3.6. Experimental example

The proposed procedure that was used above for global parameter estimation with simulation data is also applied to global parameter estimation with experimental data. These data result from experiments with pseudo-perfusion bioreactors detailed in [12]. In summary, these bioreactors do not have inlets and outlets, but the medium is renewed once a day using a feed solution with the same composition, and concentration measurements are obtained before and after each medium
renewal for each one of the $S_{a}=25$ measured extracellular species. This experimental procedure results in a metabolic state similar to steady state after some days ( 4 to 7 days). Then one can obtain replicated measurements for that experimental condition by continuing the experiments for some more days ( 3 to 6 days). According to (15), this allows computing the measurements $\tilde{\mathbf{q}}_{a}$ of the rates of variation of $S_{a}$ extracellular species between consecutive medium renewals (at times $\bar{t}$ ) from the measurements $\tilde{V C}$ of viable cell concentrations and $\tilde{\mathbf{c}}_{a}$ of $S_{a}$ bioreactor concentrations of extracellular species including biomass (with $\tilde{c}_{\text {biom }}=f_{\text {biom }} \tilde{V C}$ ) after the preceding medium renewal and before the following medium renewal (at times $t_{0}$ and $t_{f}$ ) as

$$
\begin{equation*}
\tilde{\mathbf{q}}_{a}(\bar{t})=\mathbf{S}_{a} \mathbf{S}_{e c}^{\mathrm{T}} \frac{\dot{\mathbf{c}}(\bar{t})+\omega_{p}(\bar{t})\left(\mathbf{c}(\bar{t})-\mathbf{c}_{i n}(\bar{t})-\omega_{h}(\bar{t}) \mathbf{R}(\bar{t} \mathbf{c}(\bar{t})\right.}{V C(t)}=\frac{\dot{\mathbf{c}}_{a}(\bar{t})}{V C(t)}=\tilde{\mu}(\bar{t}) \frac{\tilde{\mathbf{c}}_{a}\left(t_{f}\right)-\tilde{\mathbf{c}}_{a}\left(t_{0}\right)}{V C\left(t_{f}\right)-V C\left(t_{0}\right)}, \tag{59}
\end{equation*}
$$

by using the fact that $\omega_{p}(\bar{t})=\omega_{h}(\bar{t})=0$, the measured growth rate $\tilde{\mu}(\bar{t})=\frac{\log \left(\tilde{V C}\left(t_{f}\right)\right)-\log \left(\tilde{V C}\left(t_{0}\right)\right)}{t_{f}-t_{0}}$, and the approximations $V C(\bar{t})=\frac{\left.\tilde{V C( } t_{f}\right)-\tilde{V C}\left(t_{0}\right)}{\left.\left.\log \left(\tilde{V C( } t_{f}\right)\right)-\log \left(\tilde{V C( } t_{0}\right)\right)}$ and $\dot{\mathbf{c}}_{a}(\bar{t})=\frac{\tilde{\mathbf{c}}_{a}\left(t_{f}\right)-\tilde{\mathbf{c}}_{a}\left(t_{0}\right)}{t_{f}-t_{0}}$. In the case of biomass, $\dot{c}_{\text {biom }}(\bar{t})=f_{\text {biom }} \frac{\left.\tilde{V C( } t_{f}\right)-\tilde{V C}\left(t_{0}\right)}{t_{f}-t_{0}}$ implies that $\tilde{q}_{\text {biom }}(\bar{t})=\frac{\dot{c}_{b i o m}(\bar{t})}{V C(t)}=f_{\text {biom }} \tilde{\mu}(\bar{t})$.

From the $S_{a}$ measured rates $\tilde{\mathbf{q}}_{a}(\bar{t})$ of variation of extracellular species and the knowledge of a set of BFMs, one can compute $R_{m}$ measured reaction rates $\tilde{\boldsymbol{\psi}}(\bar{t})$ related to the BFMs using (33).

In the work of [12], this procedure has been repeated for 16 feed solutions with different amino acid compositions that induce 16 different metabolic states. Hence, one can use the experimental data from these 16 different conditions to construct a single kinetic model. In the remainder of this example, we assume that the variable $t$ denotes the sample $t$, where each sample corresponds to a single feed solution and $\tilde{\mathbf{q}}_{a}(t)$ and $\tilde{\boldsymbol{\psi}}(t)$ consist in the mean of the replicated measurements $\tilde{\mathbf{q}}_{a}(\bar{t})$ and $\tilde{\boldsymbol{\psi}}(\bar{t})$ for that experimental condition. Figure 2 shows a total of 100 measurements $\tilde{\mathbf{q}}_{a}(\bar{t})$ divided in 16 groups according to the corresponding feed solution, as well as the resulting mean $\tilde{\mathbf{q}}_{a}(t)$ for each experimental condition. Figure 3 shows the same information for the measurements $\tilde{\boldsymbol{\psi}}(\bar{t})$ and the resulting mean $\tilde{\boldsymbol{\psi}}(t)$ for each experimental condition. Due to the imprecision in the measurements of Arg and His, certain constant rate values were assumed in each condition.

Note that one can also compute the estimated rates $\hat{\mathbf{q}}(t)$ of variation of extracellular species and estimated reaction rates $\hat{\boldsymbol{\psi}}(t)$ related to the BFMs from $\tilde{\mathbf{q}}_{a}(t)$ using (37). Figures 2 and 3 compare $\tilde{\mathbf{q}}_{a}(t)$ and $\tilde{\boldsymbol{\psi}}(t)$ with $\hat{\mathbf{q}}_{a}(t)$ and $\hat{\boldsymbol{\psi}}(t)$. It was decided to use $\tilde{\boldsymbol{\psi}}(t)$ rather than $\hat{\boldsymbol{\psi}}(t)$ as data for parameter estimation since the inequality constraints of problem (37) seem to introduce bias in $\hat{\boldsymbol{\psi}}(t)$ for certain BFMs and conditions (see Figure 3).

Then, as discussed before, for each reaction rate $\psi(t)$, one can determine which sum of 2 Monod terms (where each term is activated by one species and activated or inhibited by another species) and corresponding parameters $\boldsymbol{\theta}$ result in predicted rates $\hat{\psi}(t \mid \boldsymbol{\theta})$ that provide the best fit to the measured rates $\tilde{\psi}(t)$. Recall that these Monod terms depend on not only the unknown parameters $\boldsymbol{\theta}$ but also the concentrations $c_{i}(t)$ of the species involved in the Monod terms as activating or inhibiting species. In this example, we postulate that the concentrations $\mathbf{c}(t)$ for the sample $t$ correspond to the mean of the concentrations $\tilde{\mathbf{c}}\left(t_{0}\right)$ after the medium renewals for that experimental condition. Figure 4 shows the 100 measurements $\tilde{\mathbf{c}}\left(t_{0}\right)$ divided in 16 groups according to the corresponding feed solution, as well as the resulting mean $\mathbf{c}(t)$ for each experimental condition. Since these concentrations are expected to be similar to the ones in the feed solution for the sample $t$, this allows relating the reaction rates to the composition of the feed solution. Furthermore, since the composition of the feed solution is known, which increases the precision of the concentrations $\mathbf{c}(t)$, it is reasonable to assume that only the measured rates $\tilde{\psi}(t)$ are corrupted by noise.


Figure 2: Measured rates $\tilde{\mathbf{q}}_{a}(\bar{t})$ of variation of extracellular species (indicated in the $y$-axis) between consecutive medium renewals (blue circles), grouped according to the corresponding feed solution (indicated in the $x$-axis and separated by vertical lines), mean $\tilde{\mathbf{q}}_{a}(t)$ of the measurements for each experimental condition (blue solid horizontal lines) $\pm$ standard deviation (blue dashed horizontal lines), and rates $\mathbf{S}_{a} \hat{\mathbf{q}}(t)$ estimated via problem (37) (red horizontal lines). Constant values are assumed for Arg and His in each condition.


Figure 3: Measured reaction rates $\tilde{\psi}(\bar{t})$ related to the BFMs (indicated in the $y$-axis) between consecutive medium renewals (blue circles), grouped according to the corresponding feed solution (indicated in the $x$-axis and separated by vertical lines), mean $\boldsymbol{\psi}(t)$ of the measurements for each experimental condition (blue solid horizontal lines) $\pm$ standard deviation (blue dashed horizontal lines), and rates $\hat{\boldsymbol{\psi}}(t)$ estimated via problem (37) (red horizontal lines). Constant values are assumed for Arg and His in each condition.


Figure 4: Concentrations $\tilde{\mathbf{c}}\left(t_{0}\right)$ of extracellular species (indicated in the $y$-axis) after the medium renewals (blue circles), grouped according to the corresponding feed solution (indicated in the $x$-axis and separated by vertical lines), and mean $\mathbf{c}(t)$ of the measurements for each experimental condition (blue solid horizontal lines) $\pm$ standard deviation (blue dashed horizontal lines). Zero values are assumed for Arg and His in all conditions.

Table 10: Rules that relate the reactants and products in the chemical equation of each BFM to the activating and inhibiting species of the corresponding Monod terms. The species $\mathrm{S}_{s 1}, \mathrm{~S}_{s 2}, \mathrm{~S}, \mathrm{~A}$, and I refer only to measured extracellular species, therefore the subscript ext is omitted in this table. The set of essential amino acids is denoted as EAA $=\{$ His, Ile, Leu, Lys, Met, Phe, Thr, Trp, Val $\}$.

| Type | Chemical equation of BFM | Monod term |
| :---: | :---: | :---: |
| 1a | $\left\|\nu_{i, s 1}\right\| \mathrm{S}_{s 1}+\left\|\nu_{i, s 2}\right\| \mathrm{S}_{s 2}+\sum_{\substack{s \notin\{s 1, s 2\} \\ \nu i, s<0}}\left\|\nu_{i, s}\right\| \mathrm{S}_{s} \leftrightarrow \sum_{s: \nu_{i, s}>0}\left\|\nu_{i, s}\right\| \mathrm{S}_{s}$ | Activation by $\mathrm{S}_{s 1}$ and $\mathrm{S}_{s 2}$, <br> $\mathrm{S}_{s 1} \notin\{$ Asn, Gln, Ser $\}, \mathrm{S}_{s 2} \notin\{$ Asn, Gln, Ser $\}$ |
| 1b | $\begin{aligned} & \forall s\left(\nu_{i, s}=0 \vee \mathrm{~S}_{s} \notin\{\text { Biomass, } \mathrm{mAb}\}\right) \\ & \sum_{s: \nu_{i, s}<0}\left\|\nu_{i, s}\right\| \mathrm{S}_{s} \leftrightarrow\left\|\nu_{i, s 1}\right\| \mathrm{S}_{s 1}+\left\|\nu_{i, s 2}\right\| \mathrm{S}_{s 2}+\sum_{\substack{s \notin\{s 1, s 2\}: \\ \nu_{i, s}>0}}\left\|\nu_{i, s}\right\| \mathrm{S}_{s} \end{aligned}$ | Activation by $\mathrm{S}_{s 1}$ and $\mathrm{S}_{s 2}$, <br> $\mathrm{S}_{s 1} \notin\{\mathrm{Asn}, \mathrm{Gln}, \mathrm{Ser}\}, \mathrm{S}_{s 2} \notin\{\mathrm{Asn}, \mathrm{Gln}, \mathrm{Ser}\}$ |
| 2a | $\forall s\left(\nu_{i, s}=0 \vee \mathrm{~S}_{s} \notin\{\right.$ Biomass, mAb\}$)$ <br> $\left\|\nu_{i, s 1}\right\| \mathrm{S}_{s 1}+\sum_{\substack{s \notin\{s 1\} \\ \nu_{i, s}<0}}\left\|\nu_{i, s}\right\| \mathrm{S}_{s} \leftrightarrow \sum_{s: \nu_{i, s}>0}\left\|\nu_{i, s}\right\| \mathrm{S}_{s}$ | Activation by $\mathrm{S}_{s 1}$ and A, $\mathrm{S}_{s 1} \notin\{\text { Asn, Gln, Ser }\}, \mathrm{A} \in\{\text { Asn, Gln, Ser }\}$ |
| 2b | $\begin{aligned} & \forall s\left(\nu_{i, s}=0 \vee \mathrm{~S}_{s} \notin\{\text { Biomass, } \mathrm{mAb}\}\right) \\ & \sum_{s: \nu_{i, s}<0}\left\|\nu_{i, s}\right\| \mathrm{S}_{s} \leftrightarrow\left\|\nu_{i, s 1}\right\| \mathrm{S}_{s 1}+\sum_{\substack{s \notin\{s 1\}: \\ \nu_{i, s}>0}}\left\|\nu_{i, s}\right\| \mathrm{S}_{s} \end{aligned}$ | Activation by $\mathrm{S}_{s 1}$ and A, <br> $\mathrm{S}_{s 1} \notin\{$ Asn, Gln, Ser $\}, \mathrm{A} \in\{$ Asn, Gln, Ser $\}$ |
| 3 a | $\begin{aligned} & \forall s\left(\nu_{i, s}=0 \vee \mathrm{~S}_{s} \notin\{\text { Biomass, mAb }\}\right) \\ & \left\|\nu_{i, s 1}\right\| \mathrm{S}_{s 1}+\sum_{\substack{s \notin\{s 1\}: \\ \nu_{i, s}<0}}^{\left\|\nu_{i, s}\right\| \mathrm{S}_{s} \leftrightarrow \sum_{s: \nu_{i, s}>0}\left\|\nu_{i, s}\right\| \mathrm{S}_{s}} \\ & \forall s\left(\nu_{i, s}=0 \vee \mathrm{~S}_{s} \notin\{\text { Biomass, mAb\}})\right. \end{aligned}$ | Activation by $\mathrm{S}_{s 1}$, Inhibition by I, <br> $\mathrm{S}_{s 1} \notin\{$ Asn, Gln, Ser $\}$, <br> $\mathrm{I} \in\{$ Ala, Asn, Asp, Cys, Gln, Glu, Gly, Ser $\}$ |
| 3 b | $\begin{aligned} & \sum_{s: \nu_{i, s}<0}\left\|\nu_{i, s}\right\| \mathrm{S}_{s} \leftrightarrow\left\|\nu_{i, s 1}\right\| \mathrm{S}_{s 1}+\sum_{\substack{s \notin\{s 1\} \\ \nu_{i, s}>0}}\left\|\nu_{i, s}\right\| \mathrm{S}_{s} \\ & \forall s\left(\nu_{i, s}=0 \vee \mathrm{~S}_{s} \notin\{\text { Biomass, mAb\})}\right. \end{aligned}$ | Activation by $\mathrm{S}_{s 1}$, Inhibition by I, <br> $\mathrm{S}_{s 1} \notin\{$ Asn, Gln, Ser $\}$, <br> $\mathrm{I} \in\{$ Ala, Asn, Asp, Cys, Gln, Glu, Gly, Ser $\}$ |
| 3 c | $\begin{aligned} & \sum_{s: \nu_{i, s}<0}\left\|\nu_{i, s}\right\| \mathrm{S}_{s} \leftrightarrow \sum_{s: \nu_{i, s}>0}\left\|\nu_{i, s}\right\| \mathrm{S}_{s} \\ & \forall s\left(\nu_{i, s}=0 \vee \mathrm{~S}_{s} \notin\{\text { Biomass, mAb\}})\right. \end{aligned}$ | Activation by S, Inhibition by I, <br> $\mathrm{S} \in\{$ Asn, Gln, Ser $\}, \mathrm{I} \in\{$ Ala, Asn, Asp, Cys, Gln $\}$ |
| 4a | $\sum_{s: \nu_{i, s}<0}\left\|\nu_{i, s}\right\| \mathrm{S}_{s} \leftrightarrow \text { Biomass }+\sum_{\substack{s: \nu_{i, s}>0, \mathrm{~S}_{s} \notin\{\text { Biomass }\}}}\left\|\nu_{i, s}\right\| \mathrm{S}_{s}$ | Activation by S and A , $\mathrm{S} \in \mathrm{EAA} \cup\{\mathrm{Cys}\}, \mathrm{A} \in\{\text { Asn, Asp, Gln, Ser }\}$ |
| 4b | $\sum_{s: \nu_{i, s}<0}\left\|\nu_{i, s}\right\| \mathrm{S}_{s} \leftrightarrow \mathrm{mAb}+\sum_{\substack{s: \nu_{i, s}>0, \mathrm{~S}_{s} \notin\{\mathrm{mAb}\}}}\left\|\nu_{i, s}\right\| \mathrm{S}_{s}$ | Activation by S and A , $S \in E A A \cup\{C y s\}, A \in\{$ Asp, Gln, Glu, Gly $\}$ |
| 5a | $\sum_{s: \nu_{i, s}<0}\left\|\nu_{i, s}\right\| \mathrm{S}_{s} \leftrightarrow \text { Biomass }+\sum_{\substack{s: \nu_{i, s}>0, \mathrm{~S}_{s} \notin\{\text { Biomass }\}}}\left\|\nu_{i, s}\right\| \mathrm{S}_{s}$ | Activation by S , Inhibition by I, $\mathrm{S} \in \mathrm{EAA} \cup\{\mathrm{Cys}\}, \mathrm{I} \in\{$ Asn, Asp, Cys, Gln, Ser $\}$ |
| 5b | $\sum_{s: \nu_{i, s}<0}\left\|\nu_{i, s}\right\| \mathrm{S}_{s} \leftrightarrow \mathrm{mAb}+\sum_{\substack{s: \nu_{i, s}>0, \mathrm{~S}_{s} \notin\{\mathrm{mAb}\}}}\left\|\nu_{i, s}\right\| \mathrm{S}_{s}$ | Activation by S , Inhibition by I, <br> $\mathrm{S} \in \mathrm{EAA} \cup\{\mathrm{Cys}\}, \mathrm{I} \in\{\mathrm{Asp}, \mathrm{Cys}, \mathrm{Gln}, \mathrm{Glu}, \mathrm{Gly}\}$ |

Clearly, another important issue that arises at this point concerns the choice of the Monod terms for each BFM. In this example, we relate the reactants and products in the chemical equation of the BFM to the activating and inhibiting species of the Monod terms under consideration according to the rules listed in Table 10. Note that all the activating and inhibiting species are measured extracellular species. The application of these rules to the 25 BFMs in Table 9 leads to the Monod terms listed in Table 11. Then, for each BFM, the estimation procedure in this section is applied to each combination of 2 Monod terms related to that BFM.

In this example with experimental data, one expects a structural mismatch between the structure of the model expressed by any sum of 2 Monod terms and the true reaction rate in the experimental system where the data were collected. For this reason, it is not surprising that the proposed procedure for global parameter estimation cannot certify a globally optimal solution for the relaxation order $d=2$, in contrast to the simulation example. However, one can extract a unique solution as if the null space of $\mathbf{Q}_{0, k}^{*}$ and the row space of $\mathbf{L}_{0, k}^{*}$ were of dimension 1 , although this solution is not guaranteed to be the global solution, as mentioned in Section 3.3 for the case $d=d_{\max }$. Then, the resulting solution is used as an initial guess for the solution to problem (54) via a nonlinear optimization algorithm, which is still expected to be much better than a random initial guess. Note that, when the initial guess for one of the nonnegative parameters $a_{1,1}, a_{1,2}, a_{1,3}, a_{2,1}, a_{2,2}, a_{2,3}$ is close to zero, that parameter is set to zero during the subsequent execution of the nonlinear optimization algorithm.

Table 11: List of Monod terms considered for each BFM in Table 9 according to the rules in Table 10.


Table 11: List of Monod terms considered for each BFM in Table 9 according to the rules in Table 10. (cont.)


By using the described methodology, one can obtain the kinetic model for each BFM shown in Table 12 that results in the good fit between measured and predicted reaction rates of each BFM and relatively precise predicted reaction rates that are shown in Figure 5, despite the large imprecision in the estimation of some kinetic parameters. Using this information, the parameters with a confidence interval that includes zero could be set to zero and discarded from the model in a subsequent step if a simpler model is necessary. From the predicted reaction rate $\hat{\psi}(t \mid \boldsymbol{\theta})$ of each BFM and the corresponding variance $\operatorname{Var}[\hat{\psi}(t \mid \hat{\boldsymbol{\theta}})]$, one can obtain the predicted reaction rates $\hat{\boldsymbol{\psi}}(t \mid \boldsymbol{\Theta})$ of all the BFMs, the predicted rates of variation of extracellular species $\hat{\mathbf{q}}(t \mid \boldsymbol{\Theta})=\mathbf{N}_{m}^{\mathrm{T}} \hat{\boldsymbol{\psi}}(t \mid \boldsymbol{\Theta})$, and the corresponding standard deviations, which equal the square root of the diagonal elements of $\operatorname{Var}[\hat{\boldsymbol{\psi}}(t \mid \hat{\boldsymbol{\Theta}})]$ and $\operatorname{Var}[\hat{\mathbf{q}}(t \mid \hat{\boldsymbol{\Theta}})]$. The resulting fit between measured and predicted rates of variation of extracellular species and precision of the predicted rates of variation, shown in Figure 6, are also good in general, except for some amino acids and some experimental conditions. Again, the implementation was performed on MATLAB R2018a running on an Intel Core i7 1.9 GHz processor, and MOSEK 8.1 was used as SDP solver. Since a total of 30899 combinations of Monod terms had to be evaluated and the computation of the initial guess took approximately 20 s for each combination of Monod terms as in Section 3.5, it took approximately one week to complete this model identification procedure. However, note that, since the different combinations of Monod terms are evaluated independently, this computation time could be reduced by using parallel computation. In addition, the relatively high computational load seems to be acceptable if one considers that parameter estimation is typically executed offline and the identification task proposed in this section is rather challenging due to the large number of modeled species, the complexity of models of biological reaction systems based on Monod-type kinetics, and the intended goal of global optimality.

## 4. Conclusions

This paper has presented an integrated approach that includes methods for computation of flux modes and reaction rates and system identification applied to modeling of perfusion bioreactors. The contributions can be summarized as follows:

- A framework for modeling of perfusion bioreactors using flux modes has been presented. This framework enables the computation of a unique set of flux modes for all the metabolic states and the unique computation of the corresponding reaction rates, which leads to a much more compact description of perfusion bioreactors and facilitates model identification and parameter estimation.
- A method for computationally tractable computation of maximum-likelihood parameter estimates of Monod-type kinetics has been presented. This method allows posterior certification of global optimality of the estimates and identification of the activation or inhibition mechanism. In an example with synthetic data, it was possible to guarantee global optimality of the estimates and identify the activation or inhibition mechanism. The reformulation of the problem as a semidefinite program takes advantage of the sparse structure of the optimization problem that results from the maximum-likelihood approach for the case of Monod-type kinetics. Furthermore, the procedure for global parameter estimation has been applied to real data from CHO cell bioreactors, resulting in a good fit between measurements and model.

This paper foresees that the ideal operation of perfusion bioreactors would consist in an approach for optimization of steady-state setpoints that takes advantage of a steady-state model obtained

Table 12: Kinetic model expressed by the sum of 2 Monod terms that is identified for each BFM, and corresponding kinetic parameters $\boldsymbol{\theta}$ := $\left(a_{1,1}, a_{1,2}, a_{1,3}, a_{2,1}, a_{2,2}, a_{2,3}, b_{1}, b_{2}\right) \pm$ their standard deviation.

| BFM | Sum of Monod terms | $a_{1,1}$ | $a_{1,2}$ | $a_{1,3}$ | $a_{2,1}$ | $a_{2,2}$ | $a_{2,3}$ | $b_{1}$ | $b_{2}$ |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | $\frac{b_{1} c_{\mathrm{NH} 4}}{c_{\mathrm{NH} 4}+a_{1,1}{ }^{c} \mathrm{NH} 4^{c} \mathrm{Asn}+a_{1,2}{ }^{c} \mathrm{Asn}+a_{1,3}}+\frac{b_{2} c_{\mathrm{Glu}}}{c_{\mathrm{Glu}}+a_{2,1}{ }^{c} \mathrm{Glu}^{c}{ }^{\mathrm{Gln}}+a_{2,2}{ }^{c}{ }_{\mathrm{Gln}}+a_{2,3}}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.125 \\ & \pm 0.0263 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.467 \\ & \pm 0.227 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & -0.582 \\ & \pm 0.0355 \end{aligned}$ | $\begin{aligned} & -0.295 \\ & \pm 0.0403 \end{aligned}$ |
| 2 | $\frac{b_{1} c_{\mathrm{Ser}}}{c_{\mathrm{Ser}}+a_{1,1}{ }^{c}{ }^{\mathrm{S} e r}{ }^{c} \mathrm{Asn}+a_{1,2}{ }^{c} \mathrm{Asn}+a_{1,3}}+\frac{b_{2} c_{\mathrm{Asn}}}{c_{\mathrm{Asn}}+a_{2,1}{ }^{c} \mathrm{Asn}^{c}{ }^{\mathrm{Gln}}+a_{2,2}{ }^{c} \mathrm{Gln}+a_{2,3}}$ | $\begin{aligned} & 0.384 \\ & \pm 0.833 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.479 \\ & \pm 0.606 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & -0.231 \\ & \pm 0.0489 \end{aligned}$ | $\begin{aligned} & -0.802 \\ & \pm 0.146 \end{aligned}$ |
| 3 | $\frac{b_{1} c_{\mathrm{Phe}}}{c_{\mathrm{Phe}^{+a}+a_{1,1}{ }^{c} \mathrm{Phe}^{c} \mathrm{Cys}}+a_{1,2}{ }^{c} \mathrm{Cys}^{+a_{1,3}}}+\frac{b_{2} c_{\mathrm{Phe}}}{c_{\mathrm{Phe}^{+}+a_{2,1}{ }^{c} \mathrm{Phe}^{c} \mathrm{Gln}^{+a_{2},{ }^{c} \mathrm{Gln}^{+}+a_{2,3}}}}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.0177 \\ & \pm 0.285 \end{aligned}$ | $\begin{aligned} & 0.419 \\ & \pm 6.87 \end{aligned}$ | $\begin{aligned} & -0.16 \\ & \pm 1.52 \end{aligned}$ | $\begin{aligned} & 0.35 \\ & \pm 0.203 \end{aligned}$ |
| 4 | $\frac{b_{1} c_{\mathrm{Ue}}}{c_{\mathrm{Ile}}+a_{1,1}{ }^{c}{ }_{\mathrm{Ile}}{ }^{c} \mathrm{Ass}^{+a_{1,2}{ }_{\mathrm{Asn}}+a_{1,3}}}+\frac{b_{2} c_{\mathrm{Gln}}}{c_{\mathrm{Gln}}+a_{2,1}{ }^{c} \mathrm{Gln}^{c} \mathrm{Cys}+a_{2,2}{ }^{c} \mathrm{Cys}+a_{2,3}}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.0637 \\ & \pm 0.0504 \end{aligned}$ | $\begin{aligned} & 0.699 \\ & \pm 0.58 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.0554 \\ & \pm 0.0357 \end{aligned}$ | $\begin{aligned} & 0.856 \\ & \pm 0.297 \end{aligned}$ | $\begin{aligned} & -0.0291 \\ & \pm 0.00194 \end{aligned}$ |
| 5 | $\frac{b_{1} c_{\mathrm{Glc}}}{c_{\mathrm{Glc}}+a_{1,1}{ }^{c} \mathrm{Glc}^{c}{ }^{c} \mathrm{Cys}+a_{1,2}{ }^{\mathrm{Cys}}+a_{1,3}}+\frac{b_{2} c_{\mathrm{Gln}}}{c_{\mathrm{Gln}}+a_{2,1}{ }^{c} \mathrm{Gln}^{c} \mathrm{Asp}+a_{2,2}{ }^{c} \mathrm{Asp}+a_{2,3}}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 24.1 \\ & \pm 350 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm \end{aligned}$ | $\begin{aligned} & 0.0442 \\ & \pm 0.0324 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm \\ & \hline \end{aligned}$ | $\begin{aligned} & 5.69 \\ & \pm 79.4 \end{aligned}$ | $\begin{aligned} & -0.413 \\ & \pm 0.0206 \end{aligned}$ |
| 6 | $\frac{b_{1} c_{\mathrm{Gll}} c_{\mathrm{Gln}}}{c_{\mathrm{Glu}}{ }^{c} \mathrm{Gln}^{c}+a_{1,1}{ }^{c} \mathrm{Glu}^{+}+a_{1,2} c_{\mathrm{Gln}}+a_{1,3}}+\frac{b_{2} c_{\mathrm{Glu}}}{c_{\mathrm{Glu}}+a_{2,1}{ }^{c} \mathrm{Glu}^{c}{ }_{\mathrm{Gly}}+a_{2,2}{ }^{c} \mathrm{Gly}+a_{2,3}}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.0122 \\ & \pm 0.00848 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 1 \\ & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & -0.0729 \\ & \pm 0.0148 \end{aligned}$ | $\begin{aligned} & 0.862 \\ & \pm 0.0264 \end{aligned}$ |
| 7 |  | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.000224 \\ & \pm 0.000132 \end{aligned}$ |  | $\begin{aligned} & 0.0091 \\ & \pm 0.00514 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.00989 \\ & \pm 0.00016 \end{aligned}$ | $\begin{aligned} & -0.182 \\ & 9 \pm 0.00157 \end{aligned}$ |
| 8 | $\frac{b_{1} c_{\mathrm{Glu}}}{{ }^{c_{\mathrm{Glu}}}+a_{1,1^{c}{ }^{c} \mathrm{Glu}^{c}{ }^{\mathrm{Gln}}+a_{1,2}{ }^{c} \mathrm{Gln}+a_{1,3}}+\frac{b_{2} c_{\mathrm{Ser}}}{c_{\mathrm{Ser}}+a_{2,1}{ }^{c} \mathrm{Ser}^{c} \mathrm{Asn}+a_{2,2}{ }^{c} \mathrm{Asn}+a_{2,3}}}$ | $\begin{aligned} & 0.993 \\ & \pm 0.52 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.148 \\ & \pm 0.0696 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 1.92 \\ & \pm 0.156 \end{aligned}$ | $\begin{aligned} & 0.626 \\ & \pm 0.0766 \end{aligned}$ |
| 9 | $\frac{b_{1} c_{\mathrm{NH} 4}}{c_{\mathrm{NH} 4}+a_{1,1} c_{\mathrm{NH}}{ }^{c}{ }^{\mathrm{C} 1 \mathrm{n}}+a_{1,2}{ }^{c} \mathrm{Gln}+a_{1,3}}+\frac{b_{2} c_{\mathrm{Asn}}}{c_{\mathrm{Asn}}+a_{2,1}{ }^{c} \mathrm{Asn}^{c} \mathrm{Asn}^{c}+a_{2,2}{ }_{\mathrm{Asn}}+a_{2,3}}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.184 \\ & \pm 0.0326 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 10 \\ & \pm 253 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.88 \\ & \pm 0.045 \end{aligned}$ | $\begin{aligned} & 0.666 \\ & \pm 15.5 \end{aligned}$ |
| 10 | $\frac{b_{1} c_{\mathrm{Val}}^{c_{\mathrm{Val}}+a_{1,1}{ }^{c} \mathrm{Val}^{c} \mathrm{Asn}+a_{1,2}{ }^{c} \mathrm{Asn}+a_{1,3}}}{+}+\frac{b_{2} c_{\mathrm{Gln}}}{c_{\mathrm{Gln}}+a_{2,1}{ }^{c} \mathrm{Gln}^{c} \mathrm{Cys}+a_{2,2}{ }^{c} \mathrm{Cys}+a_{2,3}}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.0756 \\ & \pm 0.0744 \end{aligned}$ | $\begin{aligned} & 0.786 \\ & \pm 0.867 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.0334 \\ & \pm 0.0476 \end{aligned}$ | $\begin{aligned} & 0.623 \\ & \pm 0.309 \end{aligned}$ | $\begin{aligned} & -0.0245 \\ & \pm 0.00226 \end{aligned}$ |
| 11 | $\frac{b_{1} c_{\text {Leu }} c_{\mathrm{Gln}}}{c_{\text {Leu }}{ }^{c} \mathrm{Gln}^{c}+a_{1,1}{ }^{c} \text { Leu }+a_{1,2}{ }^{\mathrm{Gln}}+a_{1,3}}+\frac{b_{2} c_{\text {Leu }}}{c_{\text {Leu }}+a_{2,1}{ }^{c}{ }_{\text {Leu }}{ }^{\text {Ass }}+a_{2,2}{ }^{c} \mathrm{Asn}+a_{2,3}}$ | $\begin{aligned} & 0.0367 \\ & \pm 0.0328 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm \end{aligned}$ | $\begin{aligned} & 0.0797 \\ & \pm 0.0941 \end{aligned}$ | $\begin{aligned} & 1.27 \\ & \pm 1.35 \end{aligned}$ | $\begin{aligned} & -0.0132 \\ & \pm 0.00111 \end{aligned}$ | $\begin{aligned} & 0.544 \\ & \pm 0.332 \end{aligned}$ |
| 12 | $\frac{b_{1} c_{\mathrm{Asn}}}{c_{\mathrm{Asn}}+a_{1,1}{ }^{c} \mathrm{Asn}^{c} \mathrm{Cys}^{+}+a_{1,2}{ }^{c} \mathrm{Cys}+a_{1,3}}+\frac{b_{2} c_{\mathrm{Asn}}}{c_{\mathrm{Asn}}+a_{2,1}{ }^{c} \mathrm{Asn}^{c} \mathrm{Gln}^{+}+a_{2,2}{ }^{c} \mathrm{Gln}+a_{2,3}}$ | $\begin{aligned} & 9.44 \\ & \pm 68.7 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm \end{aligned}$ | $\begin{aligned} & 0.471 \\ & \pm 0.633 \end{aligned}$ | $\begin{aligned} & 0.782 \\ & \pm \quad 1.4 \end{aligned}$ | $\begin{aligned} & 0.0841 \\ & \pm 0.469 \end{aligned}$ | $\begin{aligned} & -0.367 \\ & \pm 0.385 \end{aligned}$ |
| 13 | $\frac{b_{1} c_{\mathrm{NH} 4}}{c_{\mathrm{NH} 4}+a_{1,1} c_{\mathrm{NH}}{ }^{c} \mathrm{Assn}^{+a_{1,2}{ }^{c} \mathrm{Asn}+a_{1,3}}+\frac{b_{2} c_{\mathrm{Asn}}}{c_{\mathrm{Asn}}+a_{2,1}{ }^{c} \mathrm{Asn}^{c} \mathrm{Gln}^{+}+a_{2,2}{ }^{c} \mathrm{Gln}+a_{2,3}}}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.192 \\ & \pm 0.421 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.666 \\ & \pm 0.712 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm \\ & \hline \end{aligned}$ | $\begin{aligned} & 0.434 \\ & \pm 0.0739 \end{aligned}$ | $\begin{aligned} & 1.13 \\ & \pm 0.151 \end{aligned}$ |
| 14 |  | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 1.67 \\ & \pm 1.11 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 10.6 \\ & \pm 0.325 \end{aligned}$ | $\begin{aligned} & 0.201 \\ & \pm 0.02 \end{aligned}$ |
| 15 | $\frac{b_{1} c_{\mathrm{Ala}}}{c_{\mathrm{Ala}}+a_{1,1}{ }^{c} \mathrm{Ala}^{c} \mathrm{Gln}^{+a_{1,2}{ }^{c} \mathrm{Gln}+a_{1,3}}}+\frac{b_{2} c_{\mathrm{Gln}}}{{ }_{\mathrm{Gln}}+a_{2,1}{ }^{c} \mathrm{Gln}^{c} \mathrm{Cys}+a_{2,2}{ }^{c} \mathrm{Cys}}+a_{2,3}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 1.21 \\ & \pm 0.784 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm \\ & \hline 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm \\ & \hline \end{aligned}$ | $\begin{aligned} & -5.24 \\ & \pm 0.203 \end{aligned}$ | $\begin{aligned} & -0.053 \\ & \pm 0.0127 \end{aligned}$ |
| 16 | $\frac{b_{1} c_{\mathrm{Ala}}}{c_{\mathrm{Ala}}{ }^{+a_{1,1}} 1^{c} \mathrm{Ala}^{c} \mathrm{Gln}^{+a_{1,2}{ }^{c} \mathrm{Gln}}+a_{1,3}}+\frac{b_{2} c_{\mathrm{Gln}}}{c_{\mathrm{Gln}}+a_{2,1}{ }^{c} \mathrm{Gln}^{c}{ }^{c} \mathrm{Cys}}+a_{2,2}{ }^{c} \mathrm{Cys}^{+a_{2,3}}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & -3.53 \\ & \pm 0.142 \end{aligned}$ | $\begin{aligned} & 0.198 \\ & \pm 0.0106 \end{aligned}$ |
| 17 | $\frac{b_{1} c_{\mathrm{Asn}}}{{ }^{c_{\mathrm{Asn}}+a_{1,1^{c}} \mathrm{Asn}^{c} \mathrm{Cys}+a_{1,2^{c} \mathrm{Cys}}+a_{1,3}}+\frac{b_{2}{ }^{c} \mathrm{Asn}}{{ }^{c_{\mathrm{Asn}}}+a_{2,1}{ }^{c} \mathrm{Asn}{ }^{c} \mathrm{Gln}+a_{2,2}{ }^{c} \mathrm{Gln}+a_{2,3}}}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.597 \\ & \pm 0.655 \end{aligned}$ | $\begin{aligned} & \begin{array}{l} 0.182 \\ \pm 0.0762 \end{array} \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{gathered} -0.915 \\ \pm 0.4 \end{gathered}$ | $\begin{aligned} & 1.82 \\ & \pm 0.224 \end{aligned}$ |
| 18 | $\frac{b_{1} c_{\mathrm{Cys}}}{{ }^{c_{\mathrm{Cys}}+a_{1,1}{ }^{c} \mathrm{Cys}{ }^{c} \mathrm{Gln}^{+}+a_{1,2}{ }^{c} \mathrm{Gln}^{+a_{1,3}}}+\frac{b_{2} c_{\mathrm{Gln}}}{c_{\mathrm{Gln}}+a_{2,1}{ }^{c} \mathrm{Gln}^{c} \mathrm{Ala}}+a_{2,2}{ }^{c} \mathrm{Ala}+a_{2,3}}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.00606 \\ & \pm 0.00402 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.0528 \\ & \pm 0.0345 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 12.1 \\ & \pm 0.429 \end{aligned}$ | $\begin{aligned} & -0.589 \\ & \pm 0.0562 \end{aligned}$ |
| 19 | $\frac{b_{1} c_{\mathrm{Ala}}}{c_{\mathrm{Ala}}+a_{1,1^{c} \mathrm{Ala}^{c} \mathrm{Gln}^{c}+a_{1,2}{ }^{c} \mathrm{Gln}+a_{1,3}}+\frac{b_{2} c_{\mathrm{Gln}}}{c_{\mathrm{Gln}}+a_{2,1}{ }^{c} \mathrm{Gln}^{c} \mathrm{Gln}+a_{2,2}{ }^{c} \mathrm{Gln}+a_{2,3}}}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.848 \\ & \pm 0.544 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 3.99 \\ & \pm 0.201 \end{aligned}$ | $\begin{aligned} & 0.0513 \\ & \pm 0.013 \end{aligned}$ |
| 20 | $\frac{b_{1} c_{\mathrm{Cys}}}{c_{\mathrm{Cys}}+a_{1,1}{ }^{c} \mathrm{Cys}{ }^{\mathrm{G} l u}+a_{1,2}{ }^{c} \mathrm{Glu}+a_{1,3}}+\frac{b_{2}{ }^{c} \mathrm{Gln}}{c_{\mathrm{Gln}}+a_{2,1}{ }^{c} \mathrm{Gln}^{c} \mathrm{Ala}+a_{2,2}{ }^{c} \mathrm{Ala}+a_{2,3}}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.0448 \\ & \pm 0.0288 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.0533 \\ & \pm 0.0373 \end{aligned}$ | $\begin{aligned} & 0.0648 \\ & \pm 0.063 \end{aligned}$ | $\begin{aligned} & -13.2 \\ & \pm 0.81 \end{aligned}$ | $\begin{aligned} & \begin{array}{l} 0.738 \\ \pm 0.0796 \end{array} \end{aligned}$ |
| 21 | $\frac{b_{1} c_{\mathrm{Thr}}}{c_{\mathrm{Thr}}+a_{1,1}{ }^{c} \mathrm{Thr}^{c} \mathrm{Gln}+a_{1,2}{ }^{c} \mathrm{Gln}+a_{1,3}}+\frac{b_{2} c_{\mathrm{Ser}}}{c_{\mathrm{Ser}}+a_{2,1}{ }^{c} \mathrm{Ser}^{c}{ }^{\mathrm{Gln}}+a_{2,2}{ }^{c} \mathrm{Gln}+a_{2,3}}$ | $\begin{aligned} & 0.00425 \\ & \pm 0.0561 \end{aligned}$ | $\begin{aligned} & 0.0117 \\ & \pm 0.1 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.0167 \\ & \pm 0.0295 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.247 \\ & \pm 2.12 \end{aligned}$ | $\begin{aligned} & -0.346 \\ & \pm 3.19 \end{aligned}$ |
| 22 | $\frac{b_{1} c_{\mathrm{Asp}}}{c_{\text {Asp }}+a_{1,1^{c}{ }^{c} \mathrm{Asp}^{c} \mathrm{Gln}^{+}+a_{1,2}{ }^{c} \mathrm{Gln}^{+a_{1,3}}}+\frac{b_{2} c_{\mathrm{Lys}}}{c_{\mathrm{Lys}}+a_{2,1} c_{\mathrm{Lys}}{ }^{\mathrm{G} l y}+a_{2,2} c_{\mathrm{Gly}}+a_{2,3}}}$ | $\begin{aligned} & 4.92 \\ & \pm 17.6 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.533 \\ & \pm 0.396 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm \end{aligned}$ | $\begin{aligned} & 0.181 \\ & \pm 0.0225 \end{aligned}$ | $\begin{aligned} & -0.0393 \\ & \pm 0.0148 \end{aligned}$ |
| 23 | $\frac{b_{1} c_{\mathrm{Lys}}{ }^{c} \mathrm{Gln}}{c_{\mathrm{Lys}}{ }^{c} \mathrm{Gln}+a_{1,1} c_{\mathrm{Lys}}+a_{1,2}{ }^{c} \mathrm{Gln}^{+}+a_{1,3}}+\frac{b_{2} c_{\mathrm{Thr}}}{c_{\mathrm{Thr}}+a_{2,1}{ }^{c} \mathrm{Thr}^{c} \mathrm{Cys}^{+a_{2,2}{ }^{c} \mathrm{Cys}}+a_{2,3}}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.00237 \\ & \pm 0.000604 \end{aligned}$ |  | $\begin{aligned} & 0.00287 \\ & \pm 0.00177 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.00292 \\ & \pm 0.00181 \end{aligned}$ | $\begin{aligned} & -0.488 \\ & \pm 0.00554 \end{aligned}$ | $\begin{aligned} & 5.52 \\ & \pm 0.0257 \end{aligned}$ |
| 24 | $\frac{b_{1} c_{\mathrm{Lys}} c_{\mathrm{Gln}}}{c_{\mathrm{Lys}}{ }^{c} \mathrm{Gln}+a_{1,1} c_{\mathrm{Lys}}+a_{1,2}{ }_{\mathrm{Gln}}+a_{1,3}}+\frac{b_{2} c_{\mathrm{Thr}}}{c_{\mathrm{Thr}}+a_{2,1}{ }^{c} \mathrm{Thr}^{c} \mathrm{Asn}^{+}+a_{2,2}{ }^{c} \mathrm{Asn}^{+}+a_{2,3}}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.0209 \\ & \pm 0.011 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \end{aligned}$ | $\begin{aligned} & 0.0957 \\ & \pm 0.0392 \end{aligned}$ | $\begin{aligned} & 0.911 \\ & \pm 0.139 \end{aligned}$ | $\begin{aligned} & -5.84 \\ & \pm 0.431 \end{aligned}$ |
| 25 | $\frac{b_{1} c_{\operatorname{Trp}}{ }^{c}{ }_{\mathrm{Gln}}}{{ }^{{ }_{\operatorname{Trp}}{ }^{c} \mathrm{Gln}+a_{1,1}{ }^{c} \mathrm{Trp}^{+}+a_{1,2}{ }^{c} \mathrm{Gln}+a_{1,3}}+\frac{b_{2} c_{\mathrm{Ile}}}{{ }_{\mathrm{Ile}}+a_{2,1}{ }^{c}{ }_{\mathrm{Ile}}{ }^{c_{\mathrm{Cys}}+a_{2,2}{ }^{c} \mathrm{Cys}+a_{2,3}}}}$ | $\begin{aligned} & 0 \\ & \pm 0 \\ & \hline \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \\ & \hline \end{aligned}$ | $\begin{aligned} & 0.0201 \\ & \pm 0.00783 \\ & \hline \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm \end{aligned}$ | $\begin{aligned} & 0 \\ & \pm 0 \\ & \hline \end{aligned}$ | $\begin{aligned} & 1.82 \\ & \pm 1.77 \\ & \hline \end{aligned}$ | $\begin{aligned} & -0.106 \\ & \pm 0.0101 \\ & \hline \end{aligned}$ | $\begin{aligned} & 8 \\ & \pm 5.01 \\ & \hline \end{aligned}$ |



Figure 5: Measured reaction rates $\tilde{\psi}(\bar{t})$ related to the BFMs (indicated in the $y$-axis) between consecutive medium renewals (blue circles), grouped according to the corresponding feed solution (indicated in the $x$-axis and separated by vertical lines), mean $\tilde{\boldsymbol{\psi}}(t)$ of the measurements for each experimental condition (blue solid horizontal lines) $\pm$ standard deviation (blue dashed horizontal lines), and predicted rates $\hat{\boldsymbol{\psi}}(t \mid \boldsymbol{\Theta})$ (green solid horizontal lines) $\pm$ standard deviation (green dashed horizontal lines). Constant values are assumed for Arg and His in each condition.


Figure 6: Measured rates $\tilde{\mathbf{q}}_{a}(\bar{t})$ of variation of extracellular species (indicated in the $y$-axis) between consecutive medium renewals (blue circles), grouped according to the corresponding feed solution (indicated in the $x$-axis and separated by vertical lines), mean $\tilde{\mathbf{q}}_{a}(t)$ of the measurements for each experimental condition (blue solid horizontal lines) $\pm$ standard deviation (blue dashed horizontal lines), and predicted rates $\mathbf{S}_{a} \hat{\mathbf{q}}(t \mid \boldsymbol{\Theta})$ (green solid horizontal lines) $\pm$ standard deviation (green dashed horizontal lines). Constant values are assumed for Arg and His in each condition.
via experimental design and model identification techniques that are designed for the purpose of steady-state optimization. Hence, the methods in this paper and [4] pave the way for rational design of models for perfusion bioreactors that are suited to their reliable and optimal operation.

## Acknowledgements

This work was supported by the VINNOVA Competence Centre AdBIOPRO, contract 201605181.

## References

[1] J. Pollock, S. V. Ho, S. S. Farid, Fed-batch and perfusion culture processes: Economic, environmental, and operational feasibility under uncertainty, Biotechnol. Bioeng. 110 (1) (2013) 206-219.
[2] M. F. Clincke, C. Mölleryd, P. K. Samani, E. Lindskog, E. Fäldt, K. Walsh, V. Chotteau, Very high density of Chinese hamster ovary cells in perfusion by alternating tangential flow or tangential flow filtration in WAVE bioreactor ${ }^{\text {TM }}$-part II: Applications for antibody production and cryopreservation, Biotechnol. Prog. 29 (3) (2013) 768-777.
[3] V. Chotteau, Perfusion processes, in: M. Al-Rubeai (Ed.), Animal Cell Culture, Springer, 2015, pp. 407-443.
[4] D. Rodrigues, An integrated approach for experimental design, control, and optimization of perfusion bioreactors, in: 21st IFAC World Congress, Berlin, Germany, 2020, pp. 17119-17124.
[5] G. N. Stephanopoulos, A. A. Aristidou, J. Nielsen, Metabolic engineering: Principles and methodologies, Academic Press, San Diego, CA, 1998.
[6] S. Klamt, G. Regensburger, M. P. Gerstl, C. Jungreuthmayer, S. Schuster, R. Mahadevan, J. Zanghellini, S. Müller, From elementary flux modes to elementary flux vectors: Metabolic pathway analysis with arbitrary linear flux constraints, PLOS Comput. Biol. 13 (4) (2017) e1005409.
[7] R. M. Jungers, F. Zamorano, V. D. Blondel, A. V. Wouver, G. Bastin, Fast computation of minimal elementary decompositions of metabolic flux vectors, Automatica 47 (6) (2011) 1255-1259.
[8] Z. I. T. A. Soons, E. C. Ferreira, I. Rocha, Identification of minimal metabolic pathway models consistent with phenotypic data, J. Process Control 21 (10) (2011) 1483-1492.
[9] L. Hebing, T. Neymann, T. Thüte, A. Jockwer, S. Engell, Efficient generation of models of fed-batch fermentations for process design and control, IFAC-PapersOnLine 49 (7) (2016) 621626.
[10] T. Abbate, S. F. Sousa, L. Dewasme, G. Bastin, A. V. Wouwer, Inference of dynamic macroscopic models of cell metabolism based on elementary flux modes analysis, Biochem. Eng. J. 151 (2019) 107325.
[11] H. A. Oddsdóttir, E. Hagrot, V. Chotteau, A. Forsgren, On dynamically generating relevant elementary flux modes in a metabolic network using optimization, J. Math. Biol. 71 (4) (2015) 903-920.
[12] E. Hagrot, H. A. Oddsdóttir, M. Mäkinen, A. Forsgren, V. Chotteau, Novel column generationbased optimization approach for poly-pathway kinetic model applied to CHO cell culture, Metab. Eng. Commun. 8 (2018) e00083.
[13] J. E. Haag, A. V. Wouwer, P. Bogaerts, Dynamic modeling of complex biological systems: a link between metabolic and macroscopic description, Math. Biosci. 193 (1) (2005) 25-49.
[14] G. Bastin, D. Dochain, On-line Estimation and Adaptive Control of Bioreactors, Elsevier, Amsterdam, 1990.
[15] C. Kontoravdi, S. P. Asprey, E. N. Pistikopoulos, A. Mantalaris, Application of global sensitivity analysis to determine goals for design of experiments: an example study on antibodyproducing cell cultures, Biotechnol. Prog. 21 (4) (2005) 1128-1135.
[16] Z. Amribt, H. Niu, P. Bogaerts, Macroscopic modelling of overflow metabolism and model based optimization of hybridoma cell fed-batch cultures, Biochem. Eng. J. 70 (2013) 196-209.
[17] C. Liu, J. M. Zachara, Uncertainties of Monod kinetic parameters nonlinearly estimated from batch experiments, Environ. Sci. Technol. 35 (1) (2001) 133-141.
[18] M. Wang, E. W. Jacobsen, V. Chotteau, H. Hjalmarsson, A multi-step least-squares method for nonlinear rational models, in: Proc. 2019 American Control Conference (ACC), Philadelphia, PA, 2019.
[19] M. Wang, R. S. Risuleo, E. W. Jacobsen, V. Chotteau, H. Hjalmarsson, Identification of nonlinear kinetics of macroscopic bio-reactions using multilinear Gaussian processes, Comput. Chem. Eng. 133 (2020) 106671.
[20] J. B. Lasserre, Moments, Positive Polynomials and Their Applications, Imperial College Press, London, UK, 2010.
[21] A. Provost, G. Bastin, Dynamic metabolic modelling under the balanced growth condition, J. Process Control 14 (7) (2004) 717-728.
[22] A. Provost, G. Bastin, S. N. Agathos, Y. J. Schneider, Metabolic design of macroscopic bioreaction models: application to Chinese hamster ovary cells, Bioprocess Biosyst. Eng. 29 (5-6) (2006) 349-366.
[23] D. R. Schneider, G. V. Reklaitis, On material balances for chemically reacting systems, Chem. Eng. Sci. 30 (2) (1975) 243-247.
[24] J. Behre, L. F. Figueiredo, S. Schuster, C. Kaleta, Detecting structural invariants in biological reaction networks, in: J. van Helden, A. Toussaint, D. Thieffry (Eds.), Bacterial Molecular Networks: Methods and Protocols, Springer, New York, 2012, pp. 377-407.
[25] D. Bonvin, D. W. T. Rippin, Target factor analysis for the identification of stoichiometric models, Chem. Eng. Sci. 45 (12) (1990) 3417-3426.
[26] M. Amrhein, B. Srinivasan, D. Bonvin, Target factor analysis of reaction data: Use of data pre-treatment and reaction-invariant relationships, Chem. Eng. Sci. 54 (5) (1999) 579-591.
[27] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, Cambridge, 2004.
[28] A. Cornish-Bowden, Fundamentals of Enzyme Kinetics, 4th Edition, Wiley-Blackwell, Weinheim, Germany, 2012.
[29] D. Rodrigues, M. R. Abdalmoaty, H. Hjalmarsson, Toward tractable global solutions to maximum-likelihood estimation problems via sparse sum-of-squares relaxations, in: Proc. 58th IEEE Conference on Decision and Control (CDC), Nice, France, 2019.
[30] D. Rodrigues, M. R. Abdalmoaty, H. Hjalmarsson, Toward tractable global solutions to Bayesian point estimation problems via sparse sum-of-squares relaxations, in: Proc. 2020 American Control Conference (ACC), Denver, CO, 2020.
[31] A. Bardow, W. Marquardt, Incremental and simultaneous identification of reaction kinetics: Methods and comparison, Chem. Eng. Sci. 59 (13) (2004) 2673-2684.
[32] M. Brendel, D. Bonvin, W. Marquardt, Incremental identification of kinetic models for homogeneous reaction systems, Chem. Eng. Sci. 61 (16) (2006) 5404-5420.
[33] T. S. Motzkin, The arithmetic-geometric inequality, in: Proc. Symposium on Inequalities, New York, NY, 1967, pp. 205-224.
[34] M. Putinar, Positive polynomials on compact semi-algebraic sets, Ind. Univ. Math. J. 42 (1993) 969-984.
[35] M. Kojima, S. Kim, M. Maramatsu, Sparsity in sums of squares of polynomials, Math. Program. 103 (2005) 45-62.
[36] J. B. Lasserre, Convergent SDP-relaxations in polynomial optimization with sparsity, SIAM J. Optim. 17 (2006) 822-843.

## Appendix A. Rank conditions for reactions and BFMs

Theorem 3. The rank of $\mathbf{N}_{d}$, denoted as $R$, satisfies $R:=\operatorname{rank}\left(\mathbf{N}_{d}\right) \leq S-E<S$.
Proof. The rank of $\mathbf{N}_{d}$ cannot be greater than the dimension of the null space of $\mathbf{A}^{\mathrm{T}}$. Since $\mathbf{A}^{\mathrm{T}}$ has $S$ columns, one can infer from the rank-nullity theorem that the dimension of the null space of $\mathbf{A}^{\mathrm{T}}$ is $S-E$, which implies that $R:=\operatorname{rank}\left(\mathbf{N}_{d}\right) \leq S-E<S$ since $E>0$.

Theorem 4. The following statements hold:

- From the definitions of $\mathbf{S}_{\text {ec }}$ and $\mathbf{S}_{i c}$, the matrix $\mathbf{N}_{m}^{\mathrm{T}}$ can also be constructed as $\left[\begin{array}{l}\mathbf{S}_{e c} \mathbf{N}_{m}^{\mathrm{T}} \\ \mathbf{S}_{i c} \mathbf{N}_{m}^{\mathrm{T}}\end{array}\right]$, where $\mathbf{S}_{i c} \mathbf{N}_{m}^{\mathrm{T}}=\mathbf{0}_{S_{i c} \times R_{m}}$ and the $R_{m}$ columns of $\mathbf{S}_{e c} \mathbf{N}_{m}^{\mathrm{T}}$ span the null space of $\mathbf{A}_{N}^{\mathrm{T}} \mathbf{S}_{e c}^{\mathrm{T}}$.
- The rank of $\left[\begin{array}{c}\mathbf{A}_{N}^{\mathrm{T}} \\ \mathbf{S}_{i c}\end{array}\right]$ is $q_{e c}+S_{i c} \geq E_{e c}+S_{i c}>S_{i c}$, that is, greater than the number of intracellular species.
- The number of BFMs, the number of columns of $\mathbf{N}_{m}^{\mathrm{T}}$, and the dimension of the null spaces of $\left[\begin{array}{c}\mathbf{A}_{N}^{\mathrm{T}} \\ \mathbf{S}_{i c}\end{array}\right]$ and $\mathbf{A}_{N}^{\mathrm{T}} \mathbf{S}_{e c}^{\mathrm{T}}$ are $R_{m}=S_{e c}-q_{e c} \leq S_{e c}-E_{e c}<S_{e c}$, that is, less than the number of extracellular species.

Proof. The definitions of $\mathbf{A}_{N}, \mathbf{S}_{i c}$, and $\mathbf{S}_{e c}$ imply that $\mathbf{N}_{m}^{\mathrm{T}}=\left[\begin{array}{l}\mathbf{S}_{e c} \\ \mathbf{S}_{i_{c}}\end{array}\right] \mathbf{N}_{m}^{\mathrm{T}}=\left[\begin{array}{c}\mathbf{S}_{e c} \mathbf{N}_{m}^{\mathrm{T}} \\ \mathbf{S}_{i c} \mathbf{N}_{m}^{\mathrm{m}}\end{array}\right]$ and the rank of $\left[\begin{array}{l}\mathbf{A}_{N}^{\mathrm{T}} \\ \mathbf{S}_{i c}\end{array}\right]$ is $q_{e c}+S_{i c}$, where $q_{e c} \leq S-R$, while the definition of $\mathbf{N}_{m}$ implies that $\mathbf{S}_{i c} \mathbf{N}_{m}^{\mathrm{T}}=\mathbf{0}_{S_{i c} \times R_{m}}$. From the rank-nullity theorem, one can infer that the dimension of the null spaces of $\left[\begin{array}{l}\mathbf{A}_{N}^{\mathrm{T}} \\ \mathbf{S}_{i c}\end{array}\right]$ and $\mathbf{A}_{N}^{\mathrm{T}} \mathbf{S}_{e c}^{\mathrm{T}}$ is $S-q_{e c}-S_{i c}=S_{e c}-q_{e c}$, which implies that $R_{m}=S_{e c}-q_{e c}$ and the $R_{m}$ columns of $\mathbf{S}_{e c} \mathbf{N}_{m}^{\mathrm{T}}$ span the null space of $\mathbf{A}_{N}^{\mathrm{T}} \mathbf{S}_{e c}^{\mathrm{T}}$.

Furthermore, note that the columns of the matrix $\mathbf{A}$ of rank $E$ lie in the null space of $\mathbf{N}_{d}$, which is spanned by the column space of $\mathbf{A}_{N}$ of dimension $q=S-R$. Hence, the rank of $\mathbf{A}$ cannot be greater than the rank of $\mathbf{A}_{N}$, which implies that $E \leq q=S-R$ as well. Along the same lines, the column space of $\mathbf{S}_{e c} \mathbf{A}$ of dimension $E_{e c}>0$ lies in the column space of $\mathbf{S}_{e c} \mathbf{A}_{N}$ of dimension $q_{e c}$, from which we can conclude that $E_{e c} \leq q_{e c}$. Hence, $\left[\begin{array}{c}\mathbf{A}_{N}^{\mathrm{T}} \\ \mathbf{S}_{i c}\end{array}\right]$ is of rank $q_{e c}+S_{i c} \geq E_{e c}+S_{i c}>S_{i c}$, and $R_{m}=S_{e c}-q_{e c} \leq S_{e c}-E_{e c}<S_{e c}$.
Theorem 5. The definitions in Section 2.4 imply that $\operatorname{rank}\left(\mathbf{N}_{i c}^{\mathrm{T}}\right)=R-R_{m}$ and the $R_{m}$ columns of $\mathbf{E}_{m}$ span the null space of $\mathbf{N}_{i c}^{\mathrm{T}}$.

Proof. One can use Sylvester's rank inequality and Frobenius' rank inequality to prove that

$$
\begin{gather*}
R-R_{m}=S_{i c}+q_{e c}+R-S=\operatorname{rank}\left(\left[\begin{array}{c}
\mathbf{A}_{N}^{\mathrm{T}} \\
\mathbf{S}_{i c}
\end{array}\right]\right)+\operatorname{rank}\left(\mathbf{N}^{\mathrm{T}}\right)-S \\
\leq \operatorname{rank}\left(\left[\begin{array}{c}
\mathbf{A}_{N}^{\mathrm{T}} \\
\mathbf{S}_{i c}
\end{array}\right] \mathbf{N}^{\mathrm{T}}\right)=\operatorname{rank}\left(\mathbf{N}_{i c}^{\mathrm{T}}\right),  \tag{A.1}\\
\quad \operatorname{rank}\left(\mathbf{N}_{i c}^{\mathrm{T}}\right)+R_{m}=\operatorname{rank}\left(\mathbf{S}_{i c} \mathbf{N}^{\mathrm{T}}\right)+\operatorname{rank}\left(\mathbf{N}^{\mathrm{T}} \mathbf{E}_{m}\right) \\
\quad \leq \operatorname{rank}\left(\mathbf{S}_{i c} \mathbf{N}^{\mathrm{T}} \mathbf{E}_{m}\right)+\operatorname{rank}\left(\mathbf{N}^{\mathrm{T}}\right)=R, \tag{A.2}
\end{gather*}
$$

which shows that $\operatorname{rank}\left(\mathbf{N}_{i c}^{\mathrm{T}}\right)=R-R_{m}$ as claimed.
For the second part of the theorem, note that $\mathbf{E}_{m}$ lies in the null space of $\mathbf{N}_{i c}^{\mathrm{T}}$, the columns of which contain the stoichiometries of the independent reactions that affect the intracellular species, since

$$
\begin{equation*}
\mathbf{N}_{i c}^{\mathrm{T}} \mathbf{E}_{m}=\mathbf{S}_{i c} \mathbf{N}^{\mathrm{T}} \mathbf{E}_{m}=\mathbf{S}_{i c} \mathbf{N}_{m}^{\mathrm{T}}=\mathbf{0}_{S_{i c} \times R_{m}} . \tag{A.3}
\end{equation*}
$$

To show that the $R_{m}$ columns of $\mathbf{E}_{m}$ span the null space of $\mathbf{N}_{i c}^{\mathrm{T}}$, we need to notice that the rank of $\mathbf{E}_{m}$ is equal to $R_{m}$, which is also the dimension of the null space of $\mathbf{N}_{i c}^{\mathrm{T}}$. This results from the fact that the rank of $\mathbf{N}_{i c}^{\mathrm{T}}$ is equal to $R-R_{m}$, as shown in the first part of the theorem.

Theorem 6. The definitions in Section 2.4 imply that $\operatorname{rank}\left(\left[\begin{array}{ll}\mathbf{N}_{a} & \mathbf{N}_{i c}\end{array}\right]\right)-R=\operatorname{rank}\left(\mathbf{S}_{a} \mathbf{N}_{m}^{\mathrm{T}}\right)-R_{m}$.
 Then, one can use Sylvester's rank inequality and Frobenius' rank inequality to prove that

$$
\begin{gather*}
\operatorname{rank}\left(\left[\begin{array}{ll}
\mathbf{N}_{a} & \mathbf{N}_{i c}
\end{array}\right]\right)+R_{m}-R=\operatorname{rank}\left(\left[\begin{array}{c}
\mathbf{N}_{a}^{\mathrm{T}} \\
\mathbf{N}_{i c}^{\mathrm{T}}
\end{array}\right]\right)+\operatorname{rank}\left(\mathbf{E}_{m}\right)-R \\
\leq \operatorname{rank}\left(\left[\begin{array}{c}
\mathbf{N}_{a}^{\mathrm{T}} \\
\mathbf{N}_{i c}^{\mathrm{T}}
\end{array}\right] \mathbf{E}_{m}\right)=\operatorname{rank}\left(\mathbf{S}_{a} \mathbf{N}_{m}^{\mathrm{T}}\right),  \tag{A.4}\\
R-R_{m}+\operatorname{rank}\left(\mathbf{S}_{a} \mathbf{N}_{m}^{\mathrm{T}}\right)=\operatorname{rank}\left(\mathbf{S}_{i c}\left[\begin{array}{ll}
\mathbf{S}_{a}^{\mathrm{T}} & \left.\mathbf{S}_{i c}^{\mathrm{T}}\right]
\end{array}\right]\left[\begin{array}{l}
\mathbf{N}_{a}^{\mathrm{T}} \\
\mathbf{N}_{i c}^{\mathrm{T}}
\end{array}\right)+\operatorname{rank}\left(\left[\begin{array}{l}
\mathbf{N}_{a}^{\mathrm{T}} \\
\mathbf{N}_{T c}^{\mathrm{T}}
\end{array}\right] \mathbf{E}_{m}\right)\right. \\
\leq \operatorname{rank}\left(\begin{array}{ll}
\left.\mathbf{S}_{i c}\left[\begin{array}{ll}
\mathbf{S}_{a}^{\mathrm{T}} & \mathbf{S}_{i c}^{\mathrm{T}}
\end{array}\right]\left[\begin{array}{l}
\mathbf{N}_{a}^{\mathrm{T}} \\
\mathbf{N}_{i c}^{\mathrm{T}}
\end{array}\right] \mathbf{E}_{m}\right)+\operatorname{rank}\left(\left[\begin{array}{l}
\mathbf{N}_{a}^{\mathrm{T}} \\
\mathbf{N}_{i c}^{\mathrm{T}}
\end{array}\right]\right)=\operatorname{rank}\left(\left[\left[\begin{array}{ll}
\mathbf{N}_{a} & \mathbf{N}_{i c}
\end{array}\right]\right),\right.
\end{array},\right. \tag{A.5}
\end{gather*}
$$

${ }^{814}$ which shows that $\operatorname{rank}\left(\left[\begin{array}{ll}\mathbf{N}_{a} & \mathbf{N}_{i c}\end{array}\right]\right)-R=\operatorname{rank}\left(\mathbf{S}_{a} \mathbf{N}_{m}^{\mathrm{T}}\right)-R_{m}$ as claimed.
Theorem 7. There exists some $R_{m}$-dimensional vector of rates $\boldsymbol{\psi}(t)$ such that

$$
\left[\begin{array}{c}
\mathbf{q}(t)  \tag{A.6}\\
\mathbf{r}(t)
\end{array}\right]=\left[\begin{array}{c}
\mathbf{N}_{m}^{\mathrm{T}} \\
\mathbf{E}_{m}
\end{array}\right] \boldsymbol{\psi}(t) .
$$

Proof. Recall that the true rates of variation of the $S$ species due to reactions are given by

$$
\begin{equation*}
\mathbf{q}(t)=\mathbf{N}^{\mathrm{T}} \mathbf{r}(t) . \tag{A.7}
\end{equation*}
$$

It is also known from the dynamic model (13) that $\mathbf{S}_{i c} \mathbf{q}(t)=\mathbf{0}_{S_{i c}}$, which imposes the constraint

$$
\begin{equation*}
\mathbf{N}_{i c}^{\mathrm{T}} \mathbf{r}(t)=\mathbf{S}_{i c} \mathbf{N}^{\mathrm{T}} \mathbf{r}(t)=\mathbf{S}_{i c} \mathbf{q}(t)=\mathbf{0}_{S_{i c}} . \tag{A.8}
\end{equation*}
$$

If we combine both constraints, we observe that

$$
\left[\begin{array}{cc}
\mathbf{I}_{S} & -\mathbf{N}^{\mathrm{T}}  \tag{A.9}\\
\mathbf{0}_{S_{i c} \times S} & \mathbf{N}_{i c}^{\mathrm{T}}
\end{array}\right]\left[\begin{array}{l}
\mathbf{q}(t) \\
\mathbf{r}(t)
\end{array}\right]=\left[\begin{array}{c}
\mathbf{0}_{S} \\
\mathbf{0}_{S_{i c}}
\end{array}\right] .
$$

Since the $R_{m}$ columns of $\mathbf{E}_{m}$ span the null space of $\mathbf{N}_{i c}^{\mathrm{T}}$, this implies that the $R_{m}$ columns of $\left[\begin{array}{c}\mathbf{N}_{m}^{\mathrm{T}} \\ \mathbf{E}_{m}\end{array}\right]$ span the null space of $\left[\begin{array}{cc}\mathbf{I}_{S} & -\mathbf{N}^{\mathrm{T}} \\ \mathbf{0}_{S_{i c} \times S} & \mathbf{N}_{i c}^{\mathrm{T}}\end{array}\right]$ and there is always some $R_{m}$-dimensional vector of rates $\boldsymbol{\psi}(t)$ such that

$$
\left[\begin{array}{c}
\mathbf{q}(t)  \tag{A.10}\\
\mathbf{r}(t)
\end{array}\right]=\left[\begin{array}{l}
\mathbf{N}_{m}^{\mathrm{T}} \\
\mathbf{E}_{m}
\end{array}\right] \boldsymbol{\psi}(t) .
$$

## Appendix B. Sum-of-squares polynomials for global optimization

This appendix summarizes the discussion about the concept of sum-of-squares polynomials and its application to global optimization. For a more comprehensive discussion, the reader is referred to previous papers that apply the same concept in another context related to parameter estimation [29, 30].

A polynomial $p(\mathbf{x})$ of degree $2 d$ in the $n$ variables $\mathbf{x}:=\left(x_{1}, \ldots, x_{n}\right)$ is a sum-of-squares (SOS) polynomial if it can be written as a sum of squares of polynomials of degree up to $d$ in $\mathbf{x}$. The concept of SOS polynomials is useful for optimization because $p(\mathbf{x})$ is an SOS polynomial if and only if there exists a positive semidefinite matrix $\mathbf{Q}$ such that $p(\mathbf{x})=\mathbf{v}_{d}(\mathbf{x})^{\mathrm{T}} \mathbf{Q} \mathbf{v}_{d}(\mathbf{x})=\operatorname{tr}\left(\mathbf{v}_{d}(\mathbf{x}) \mathbf{v}_{d}(\mathbf{x})^{\mathrm{T}} \mathbf{Q}\right)$, where $\mathbf{v}_{d}(\mathbf{x})$ is the $s(n, d)$-dimensional vector of monomials of degree up to $d$ in the $n$ variables $\mathbf{x}$, with $s(n, d):=\binom{n+d}{n}$ [20]. Hence, constraining $p(\mathbf{x})$ to the set of SOS polynomials amounts to satisfying the linear matrix inequality (LMI) $\mathbf{Q} \succeq \mathbf{0}_{s(n, d) \times s(n, d)}$, which can be done via a convex semidefinite program (SDP) [27].

However, it is not generally true that a nonnegative polynomial is an SOS polynomial [33]. On the other hand, if $f(\mathbf{x})$ of degree $2 v_{0}$ or $2 v_{0}-1$ is a strictly positive polynomial on a compact basic semi-algebraic set $\mathbb{K}$ specified by some polynomials $g_{j}(\mathbf{x})$ of degree $2 v_{j}$ or $2 v_{j}-1$, with $c_{d}:=\max _{j=1, \ldots, n_{c}} v_{j}$, that is, if $f(\mathbf{x})>0 \forall \mathbf{x} \in \mathbb{K}=\left\{\mathbf{x}: g_{j}(\mathbf{x}) \geq 0, \forall j=1, \ldots, n_{c}\right\}$ and $\mathbb{K}$ satisfies some technical assumptions, then $f(\mathbf{x})$ can be represented as a combination of SOS polynomials up to some degree $2 d$, where $d \geq v:=\max _{j=0,1, \ldots, n_{c}} v_{j}$ is the relaxation order [34].

A sparse representation can be obtained by taking advantage of the fact that each polynomial $g_{j}(\mathbf{x})$ may involve only a few variables, and $f(\mathbf{x})$ may be written as a sum of polynomials that also involve only a few variables [35]. For this, we define $p$ index subsets $I_{k}$ with the corresponding $n_{k}:=\left|I_{k}\right|$ variables $\mathbf{x}\left(I_{k}\right)=\left\{x_{i}: i \in I_{k}\right\}$, for $k=1, \ldots, p$, such that $\cup_{k=1}^{p} I_{k}=\{1, \ldots, n\}$. This important result about sparse representation is summarized in the following theorem [36].

Theorem 8. Consider the basic semi-algebraic set $\mathbb{K}:=\left\{\mathbf{x}: g_{j}(\mathbf{x}) \geq 0, \forall j=1, \ldots, n_{c}\right\}$ and assume that the index subsets $I_{1}, \ldots, I_{p}$ satisfy the following conditions:

1. The polynomial $f(\mathbf{x})$ can be written as a sum of $p$ polynomials that involve only the variables $\mathbf{x}\left(I_{1}\right), \ldots, \mathbf{x}\left(I_{p}\right)$, that is, $f(\mathbf{x})=\sum_{k=1}^{p} f_{k}\left(\mathbf{x}\left(I_{k}\right)\right)$.
2. The running intersection property holds, that is, for all $k=1, \ldots, p-1, I_{k+1} \cap\left(\cup_{j=1}^{k} I_{j}\right) \subseteq I_{s}$ for some $s \leq k$.
3. For all $j=1, \ldots, n_{c}$, there exists some $K_{j} \in\{1, \ldots, p\}$ that indicates that $g_{j}(\mathbf{x})$ involves only the variables $\mathbf{x}\left(I_{K_{j}}\right)$, that is, $g_{j}(\mathbf{x})=c_{j}\left(\mathbf{x}\left(I_{K_{j}}\right)\right)$.
4. For all $k=1, \ldots, p$, there exists some $q_{k} \in\left\{1, \ldots, n_{c}\right\}$ such that the set $\left\{\mathbf{x}\left(I_{k}\right): c_{q_{k}}\left(\mathbf{x}\left(I_{k}\right)\right) \geq 0\right\}$ is compact.

If $f(\mathbf{x})$ is strictly positive $\forall \mathbf{x} \in \mathbb{K}$, then

$$
\begin{equation*}
f(\mathbf{x})=\sum_{k=1}^{p} p_{0, k}\left(\mathbf{x}\left(I_{k}\right)\right)+\sum_{j=1}^{n_{c}} g_{j}(\mathbf{x}) p_{j}\left(\mathbf{x}\left(I_{K_{j}}\right)\right) \tag{B.1}
\end{equation*}
$$

for some SOS polynomials $p_{0,1}\left(\mathbf{x}\left(I_{1}\right)\right), \ldots, p_{0, p}\left(\mathbf{x}\left(I_{p}\right)\right)$ and $p_{1}\left(\mathbf{x}\left(I_{K_{1}}\right)\right), \ldots, p_{n_{c}}\left(\mathbf{x}\left(I_{K_{n_{c}}}\right)\right)$.
Proof. The proofs of Theorems 8-9 can be found in the references before each theorem and are not replicated.

Remark 1. This representation can be used to relax the verification of positivity of $f(\mathbf{x}) \forall \mathbf{x} \in \mathbb{K}$ as a hierarchy of sparse LMI feasibility problems of increasing order d [36]. To introduce the sparse relaxations, note that the monomials $\mathbf{x}^{\boldsymbol{\alpha}}:=x_{1}^{\alpha_{1}} \ldots x_{n}^{\alpha_{n}}$ of degree up to $2 d$ in the variables $\mathbf{x}\left(I_{k}\right)$ involve powers $\boldsymbol{\alpha}:=\left(\alpha_{1}, \ldots, \alpha_{n}\right)$ in the set $\overline{\mathcal{X}}_{d, k}:=\mathcal{X}_{d} \cap\left\{\left(\alpha_{1}, \ldots, \alpha_{n}\right) \in \mathbb{N}_{0}^{n}: \alpha_{i} \neq 0 \Rightarrow i \in I_{k}\right\}$, for $k=1, \ldots, p$, where $\mathcal{X}_{d}:=\left\{\left(\alpha_{1}, \ldots, \alpha_{n}\right) \in \mathbb{N}_{0}^{n}: 0 \leq \alpha_{1}+\ldots+\alpha_{n} \leq 2 d\right\}$. We define $\overline{\mathcal{X}}_{d}:=\cup_{k=1}^{p} \overline{\mathcal{X}}_{d, k}$ and use $f_{\boldsymbol{\alpha}}$ and $g_{j, \boldsymbol{\alpha}}$ to denote the coefficients of $f(\mathbf{x})$ and $g_{j}(\mathbf{x})$ such that $f(\mathbf{x})=\sum_{\boldsymbol{\alpha} \in \overline{\mathcal{X}}_{d}} f_{\boldsymbol{\alpha}} \mathbf{x}^{\boldsymbol{\alpha}}$ and $g_{j}(\mathbf{x})=\sum_{\boldsymbol{\alpha} \in \overline{\mathcal{X}}_{v_{j}}} g_{j, \boldsymbol{\alpha}} \mathbf{x}^{\boldsymbol{\alpha}}$, for $j=1, \ldots, n_{c}$. Moreover, the matrices $\mathbf{R}_{v, k, \boldsymbol{\alpha}}$ are defined such that $\sum_{\boldsymbol{\alpha} \in \overline{\mathcal{X}}_{d-v}} \mathbf{R}_{v, k, \boldsymbol{\alpha}} \mathbf{x}^{\boldsymbol{\alpha}}=\mathbf{v}_{d-v}\left(\mathbf{x}\left(I_{k}\right)\right) \mathbf{v}_{d-v}\left(\mathbf{x}\left(I_{k}\right)\right)^{\mathrm{T}}$, for $v=0, \ldots, d$ and $k=1, \ldots, p$. If Theorem 8 applies and $f(\mathbf{x})$ is strictly positive $\forall \mathbf{x} \in \mathbb{K}$, then there exists a positive integer $d$ such that

$$
\begin{equation*}
f_{\boldsymbol{\alpha}}=\sum_{k=1}^{p} \operatorname{tr}\left(\mathbf{R}_{0, k, \boldsymbol{\alpha}} \mathbf{Q}_{0, k}\right)+\sum_{j=1}^{n_{c}} \sum_{\substack{\boldsymbol{\beta} \in \overline{\mathcal{X}}_{d-v_{j}} \\ \boldsymbol{\alpha}-\boldsymbol{\beta} \in \overline{\mathcal{X}}_{v_{j}}}} g_{j, \boldsymbol{\alpha}-\boldsymbol{\beta}} \operatorname{tr}\left(\mathbf{R}_{v_{j}, K_{j}, \boldsymbol{\beta}} \mathbf{Q}_{j}\right), \quad \boldsymbol{\alpha} \in \overline{\mathcal{X}}_{d} \tag{B.2a}
\end{equation*}
$$

and

$$
\begin{align*}
& \mathbf{Q}_{0, k} \succeq \mathbf{0}_{s\left(n_{k}, d\right) \times s\left(n_{k}, d\right)}, \quad k=1, \ldots, p  \tag{B.2b}\\
& \mathbf{Q}_{j} \succeq \mathbf{0}_{s\left(n_{K_{j}}, d-v_{j}\right) \times s\left(n_{K_{j}}, d-v_{j}\right)}, \quad j=1, \ldots, n_{c} \tag{B.2c}
\end{align*}
$$

This result is very useful for the problem of computing $J^{*}$, an accurate approximation of the global minimum of $J(\mathbf{x})$ subject to the constraints $g_{j}(\mathbf{x}) \geq 0$, for $j=1, \ldots, n_{c}$, or equivalently, the maximum value $\tau$ such that $f(\mathbf{x})=J(\mathbf{x})-\tau>0 \forall \mathbf{x} \in \mathbb{K}=\left\{\mathbf{x}: g_{j}(\mathbf{x}) \geq 0, \forall j=1, \ldots, n_{c}\right\}$. Such a problem can be formulated as the SDP

$$
\begin{equation*}
\min _{\tau, \mathbf{Q}_{0,1}, \ldots, \mathbf{Q}_{0, p}, \mathbf{Q}_{1}, \ldots, \mathbf{Q}_{n_{c}}}-\tau, \quad \text { s.t. (B.2). } \tag{B.3}
\end{equation*}
$$

Hence, if $n_{k}$ and the maximum degree $v$ of the polynomials are relatively small, the SDP can be solved efficiently since the relaxation order $d$ that provides a sparse representation in terms of SOS polynomials is usually not much larger than $v$. If this representation exists for some order $d$, a certificate can be obtained upon convergence of the SDP. The result about the sparse representation for the order $d$ is stated as follows [36]:

Theorem 9. Denote the optimal values of the dual variables for the constraints (B.2a) as $\mu_{\boldsymbol{\alpha}}^{*}$ $\forall \boldsymbol{\alpha} \in \overline{\mathcal{X}}_{d}$ and of the dual variables for the LMIs (B.2b) as $\mathbf{L}_{0, k}^{*} \forall k=1, \ldots, p$. If $\exists G: G=$ $\operatorname{rank}\left(\mathbf{L}_{0, k}^{*}\right)=\operatorname{rank}\left(\sum_{\boldsymbol{\alpha} \in \overline{\mathcal{X}}_{d-c_{d}}} \mathbf{R}_{c_{d}, k, \boldsymbol{\alpha}} \mu_{\boldsymbol{\alpha}}^{*}\right) \forall k=1, \ldots, p$, then $f(\mathbf{x})=J(\mathbf{x})-J^{*}$ can be represented as in (B.1) with $p_{0, k}\left(\mathbf{x}\left(I_{k}\right)\right)$ of degree $2 d$, for $k=1, \ldots, p$, and $p_{j}\left(\mathbf{x}\left(I_{K_{j}}\right)\right)$ of degree $2\left(d-v_{j}\right)$, for $j=1, \ldots, n_{c}$. In addition, the global minimum $J^{*}=\tau^{*}$ and $G$ global minimizers $\mathbf{x}^{*}$ can be computed using the fact that $\mathbf{v}_{d}\left(\mathbf{x}\left(I_{k}\right)^{*}\right)$ lie both in the null space of $\mathbf{Q}_{0, k}^{*}$ and in the row space of $\mathbf{L}_{0, k}^{*}, \forall k=1, \ldots, p$.
Remark 2. Note that the rank condition in Theorem 9 is sufficient but not necessary. In addition, even if the rank condition is not satisfied, $\tau^{*} \leq J^{*}$ [36]. Consequently, if a solution $\mathbf{x}^{*}$ is extracted from the solution to the SDP and satisfies the constraints $g_{j}\left(\mathbf{x}^{*}\right) \geq 0$, for $j=1, \ldots, n_{c}$, then $\tau^{*} \leq J^{*} \leq J\left(\mathbf{x}^{*}\right)$. This implies that $J\left(\mathbf{x}^{*}\right)-\tau^{*}$ can be seen as the optimality gap, that is, an upper bound on the difference $J\left(\mathbf{x}^{*}\right)-J^{*} \geq 0$, which may be very small even if the rank condition is not satisfied. Hence, this remark supports the procedure used in steps 7-8 of Algorithm 1.


[^0]:    ${ }^{1}$ In this paper, we use stoichiometric matrices where the rows correspond to the reactions and the columns correspond to the species, which are easier to relate to the chemical equations of the reactions. Note that, in some references $[11,12]$, the transposed matrices where the rows correspond to the species and the columns correspond to the reactions are called "stoichiometric matrices".

[^1]:    ${ }^{2}$ For the sake of simplicity, the dependence on the time $t$ is omitted in most of this section.

