Biochemical Engineering Journal 221 (2025) 109770

Contents lists available at ScienceDirect

Biochemical Engineering Journal

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

Regular article

A constructive bottom-up approach for the elaboration of metabolic networks: Case study of photosynthetic cyanobacteria *Arthrospira* spirulina platensis PCC 8005

Maxime Maton [a], Baptiste Leroy [b], Alain Vande Wouwer [a,*]

[a] Systems, Estimation, Control and Optimization Department (SECO), University of Mons, Mons 7000, Belgium
[b] Department of Proteomic and Microbiology, Research Institute of Biosciences, University of Mons, Mons 7000, Belgium

ARTICLE INFO

Keywords:
Metabolic modeling
Metabolic flux analysis
Metabolic network
Cyanobacteria
Energy metabolism

ABSTRACT

Mathematical modeling has proven to be a highly effective tool for understanding microbial metabolism for which in-silico and experimental studies help to quantify intracellular mechanisms and pave the way for optimizing the production of molecules of interest. In that context, the development of metabolic networks turns out to be particularly interesting despite the challenges underlying their reconstruction. While the elaboration of genome-scale networks is computationally costly, small networks are often oversimplified and important biological mechanisms might be omitted, which limits their use in industrial applications. For this purpose, this study proposes a constructive bottom-up approach for the identification of metabolic networks of intermediate size, typically comprised of a couple of hundred reactions. It combines basic biological knowledge and a series of constraint-based methods in an iterative strategy, enabling the refinement of the network definition. The network is first validated using in-silico data, and subsequently refined using experimental data to enhance its biological relevance. Several case studies have been addressed to assess the efficiency of the methodology, and this paper focuses on the modeling of photosynthetic cyanobacteria *Arthrospira* sp. PCC 8005. The procedure is effective and provides promising results and metabolic analyses show consistent predictive capabilities of the network, in concordance with existing studies.

1. Background

Metabolic networks are used in various applications in biotechnology, medicine, and environmental science and enable scientists to gain insights into the operation of organisms and processes of energy production, nutrient utilization, and waste elimination, among others. However, constructing a metabolic network requires integrating high-quality, reliable, and available data from multiple sources, including genomics, transcriptomics, proteomics, and metabolomics. In addition, metabolic networks are highly interconnected, making their reconstruction time-consuming and their analyses computationally intensive. Besides, to ensure their accuracy and predictive capability, an iterative process of model refinement is needed, comparing model predictions with experimental observations and leading to updating the network structure when necessary.

With the advances in sequencing techniques and genome annotation methods, complete genome sequences are available for many microorganisms, enabling the generation of genome-scale metabolic reconstructions and leading to genome-scale metabolic models [1–3]. However, to ensure the consistency of such reconstructions and comprehend the organizational principles of metabolic interactions within cellular networks, examination of the structural and topological properties of metabolic networks is important. By the past, many studies [4–7] have covered this issue, e.g., the identification of groups of reactions unable to carry a flux due to the stoichiometry of the network under steady-state conditions or the establishment of enzyme subsets, i.e., groups of reactions operating together in fixed flux proportions at steady-state. Such analyses enable identifying potentially missing reactions as well as reactions under coordinated regulation and allow a continuous refinement of metabolic reconstructions through an iterative model-building process [8–11]. Similarly, frameworks based on convex analysis have been developed to examine structural and topological network properties. Those methods are mainly based on a series of constraints that govern the operation of the metabolic network at

* Corresponding author.
E-mail address: Alain.VANDEWOUWER@umons.ac.be (A. Vande Wouwer).

https://doi.org/10.1016/j.bej.2025.109770
Received 2 February 2025; Received in revised form 14 April 2025; Accepted 1 May 2025
Available online 8 May 2025

steady-state, including stoichiometric and thermodynamic constraints, which limit the range of possible network behaviors corresponding to different metabolic phenotypes. For instance, flux balance analysis is a mathematical approach for analyzing the flow of metabolites through the network and identifies an optimal flux distribution characterizing a particular phenotype [12]. Other techniques rely on the concept of extreme pathways [13] and elementary modes [14,15] and are often used as a quantitative measure of network robustness. However, even if constraint-based modeling and analysis have attracted huge interest in the last decades, most algorithms for the identification of elementary modes and extreme pathways do not scale well for genome-scale models of complex microorganisms due to the combinatorial explosion of the identified modes. In that context, several mathematical techniques and algorithms have emerged recently to reduce the number of modes and identify the most informative ones [16,17]. Also, besides these popular approaches, flux coupling analysis is concerned with describing dependencies between metabolic reactions and finding coupled reaction sets and blocked reactions in genome-scale metabolic systems [18,19]. The latter method is essentially used for exploring biological questions such as network evolution, gene essentiality, and gene regulation and requires solving a series of linear programs.

The above methods are efficient and take part in the model-building process. However, they require the computation of null-space matrices or involve the solution of a sequence of linear programs that can be computationally costly. Another challenge with genome-scale networks is the complexity involved in constructing the stoichiometric matrix, which serves as the foundation for many computational analyses. Therefore, building an accurate and balanced stoichiometric matrix can be time-consuming and prone to errors. Additionally, genome-scale networks depend on the availability of high-quality genomic data and comprehensive gene annotations, which are not always accessible or reliable, particularly for lesser-studied or newly isolated strains. Furthermore, although the models deduced from such networks provide a complete skeleton of the metabolic reactions taking place in an organism, these models are difficult to use for control and optimization strategies because of their complexity and their dimension [20]. Henceforth, to facilitate the analyses and the application of control strategies, small-size metabolic networks have been deduced from larger networks by focusing only on the main metabolic functions of the microorganisms [21–24,25,26]. Such networks often encompass essential pathways such as glycolysis, the citric acid cycle and other specific biosynthetic pathways and lead to simplified models that are used to understand the fundamental aspects of cellular metabolism and the principles of enzymatic activity and regulatory mechanisms. However, even though they are easier to manage than large-scale networks and require reduced computational resources, small-size metabolic networks are often oversimplified. Hence, important interactions that occur in the full network are omitted, and insights gained from the simplified networks might not be directly generalized to more complex systems. In addition, the behavior and effectiveness of the network can vary depending on specific environmental conditions and the physiological state of the organism, which are key aspects that are not represented in small-size networks.

A middle ground is then the development of metabolic networks of intermediate size, which are detailed enough to capture the complexity of the cellular metabolism while remaining straightforward to analyze at the process level. These networks can be constructed by means of different approaches in metabolic modeling. On the one hand, top-down approaches start with a comprehensive genome-scale or a large model and aim to simplify or reduce the network by applying reduction algorithms and techniques like transcriptomics or proteomics-based filtering for further refinement [27–30]. These methods typically rely on the structure of the genome-scale model. For this purpose, powerful algorithms have emerged in recent years, such as *NetworkReducer* [20], *MinNM* or *MetNet* [31] among others. Nevertheless, these methods inherit many of the same limitations as genome-scale networks themselves (i.e., availability of genomic data, errors or gaps in the original genomic reconstruction, the complexity and the size of the original genome-scale model as well as computational challenges associated with handling large matrices). On the other hand, bottom-up approaches involve constructing metabolic networks by carefully curating reactions based on biological insight, biochemical and physiological data, and experimental findings. These methods often simplify the networks in a context-specific manner. For instance, the elaboration of the network can start from condition-specific metabolome data [27] or mass spectroscopy data [32]. Then, metabolic networks are constructed by using Gaussian Graphical Models, statistical techniques for dimensionality reduction [32], or optimization-based methods to infer the active networks [27]. Also, more recently, studies have been conducted to leverage both top-down and bottom-up strategies, leading to a flexible methodology for constructing functional and testable models. This hybrid approach enables enhancing model accuracy and biological relevance and is based on graph-based transformation and stoichiometric and retro-synthetic analyses [33,34]. Nevertheless, it is worth noting that, in many cases, methods used for metabolic network construction (both top-down or bottom-up approaches) utilize genome-scale metabolic models (GEMs) as foundational components, though their usage can differ depending on the approach [29]. As a matter of fact, top-down approaches use GEMs as a starting point for simplification, bottom-up approaches begin with experimental data to construct smaller networks but still sometimes make use of GEMs to ensure a broader metabolic context (like *redGEM* and *lumpGEM* in [29]) and hybrid methods leverage both approaches using GEMs and experimental data to build and refine the model.

The methodology introduced in this paper offers an alternative to existing methods and can be considered as a constructive approach to derive a metabolic network. In contrast with current methods, it is a bottom-up approach with minimal reliance on genomic reconstructions. Therefore, it proves particularly valuable in cases where the strain under study is more specific and genomic reconstructions are not readily accessible in the literature. As illustrated in Fig. 1, the method is by essence iterative. When applicable, genomic studies are exploited only to identify key metabolites and detect specific reactions proper to the strain under study (in contrast to top-down approaches where genome-scale models are simplified using reduction algorithms). In that case, the modeling procedure is also based on lumping techniques [35–37] to overlook some intermediate metabolites and finally, it uses constraint-based methods [38–41] to ensure network consistency, as illustrated in Fig. 2 (more details will be given in the sequel of this article). Furthermore, it uses in-silico data and experimental data to quantify the intracellular mechanisms and refine the metabolic network definition.

In this study, this methodology is applied to the case of *Arthrospira* sp. PCC 8005, which is a photosynthetic prokaryote using light as an energy source, water as an electron donor, and CO_2 as a carbon source, making its metabolism interesting to model. This strain has notably been selected as the best nutritional resource for long-haul space exploration missions for its ability to convert CO_2 into edible biomass and oxygen and its ability to remove certain compounds from contaminated waters [42–44]. Therefore, because of its use in bioregenerative life-support systems, a deeper understanding of the metabolic capabilities of this organism is essential to fully harness its biotechnological potential. Also, studies on similar strains have been conducted in recent years, enabling a deep examination and assessment of the results of metabolic analyses, using existing studies as a benchmark. So far, the analysis of this specific strain has been scarce. A basic metabolic flux model containing 22 reactions has been developed in [45] to predict rate-limiting enzymes for the production of γ-linolenic acid. Subsequently, a more extensive metabolic network for *S. platensis* has been identified by [46] in a context-specific manner. However, in addition to inconsistencies regarding reaction reversibility, these models did not provide the microorganisms' whole-cell characteristics and metabolic capabilities.

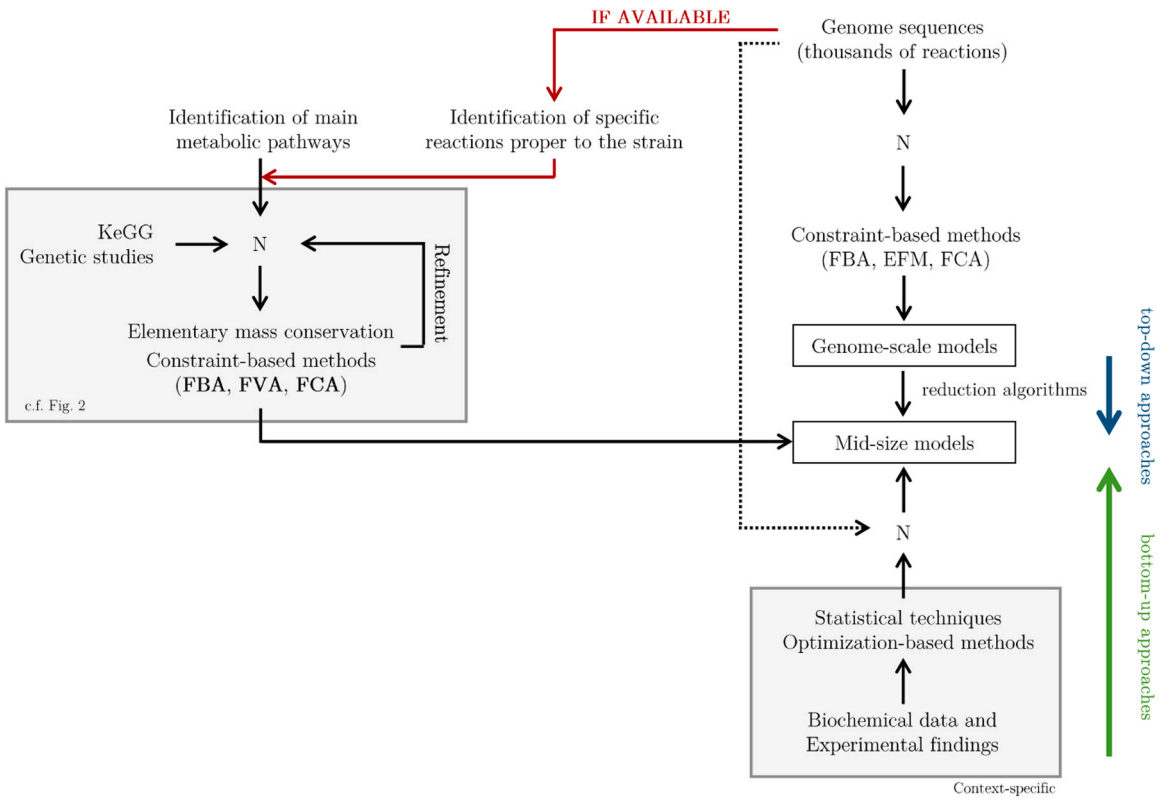

Fig. 1. Modeling procedure.

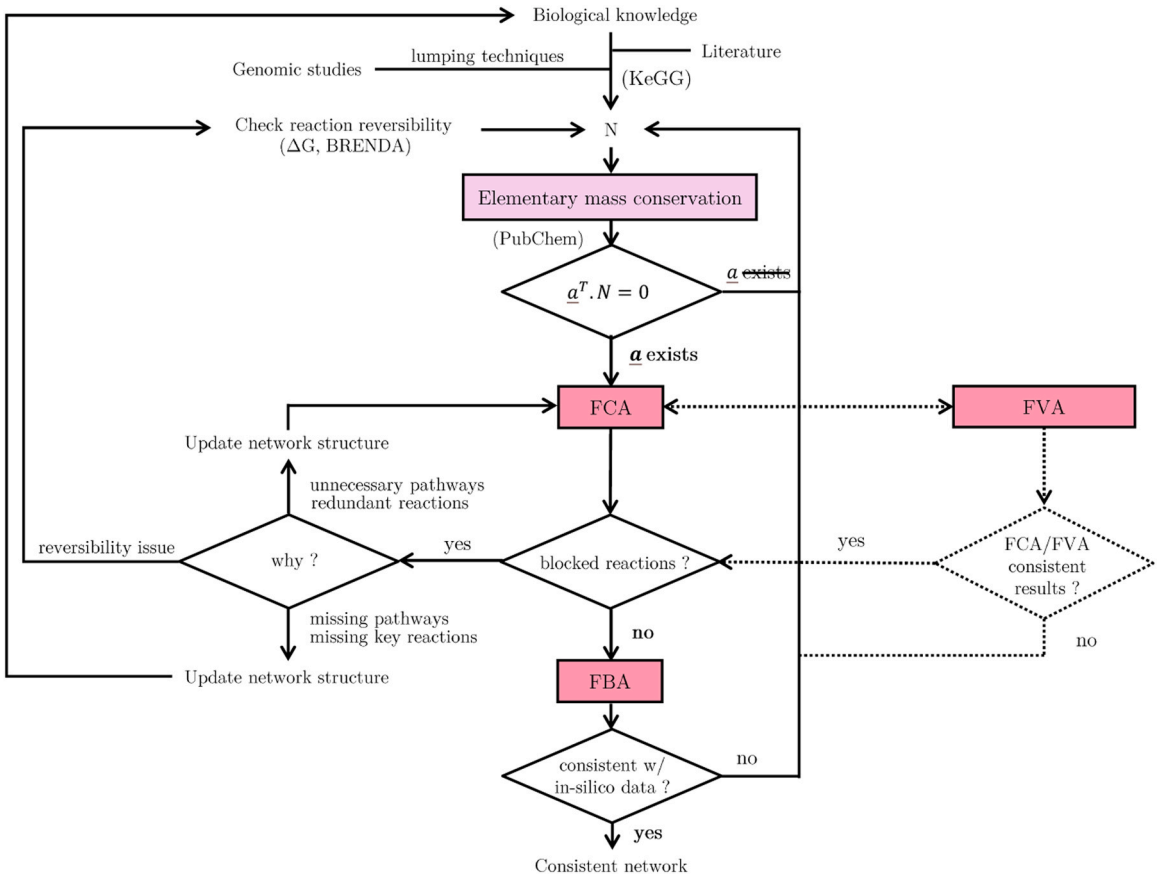

Fig. 2. Algorithmic scheme of the mathematical methods. Note that FCA is preferred to FVA for larger networks. The latter methods should give equivalent results (methods' cross-validation).

Recently, available data and genome sequences of *S. platensis* (different strains) have enabled the development of genome-scale metabolic models [47–49]. While the first models lack details, the others are difficult to comprehend. In this context, the suggested procedure bridges these two extremes while facilitating the learning and understanding of the studied strain cell's metabolism, particularly for non-expert users (through a step-by-step constructive approach). After carefully refining the network structure, the identified metabolic network comprises 198 reactions and 171 metabolites and includes energy aspects such as the proton motive force, barely outlined in existing networks [50].

The paper is organized as follows. Sec. 2 explains the complete methodology for the construction of mid-size metabolic networks and reviews the mathematical methods applied during the procedure. Sec. 3 presents the case study of *Arthrospira* sp. PCC 8005, emphasizes the main metabolic routes of the photosynthetic microorganisms and presents several iterations of the bottom-up method to derive the final metabolic network. Sec. 4 discusses the results of metabolic analyses using in-silico data under different light regimes and Sec. 5 validates the network structure against experimental data. Finally, conclusions are drawn in Sec. 6.

2. Constructive modeling approach

The modeling procedure is illustrated in Figs. 1 and 2. Fig. 1 positions the approach within the landscape of existing methods and Fig. 2 depicts the algorithmic scheme of the mathematical methods. The first step of the methodology involves collecting metabolic reactions defining the metabolism of the microorganism from the literature. When genome sequences of the strain are available, specific reactions proper to the strain under study might be added. In that case, lumping techniques are applied to overlook nonessential intermediate metabolites and identify only the main reactions and key metabolites. Furthermore, because the reversibility of chemical reactions is essential to ensure the consistency of the network, genetic studies and biochemical databases (such as KeGG) are used to refine the development of the network and to develop a fully generalized network. At this stage, it is possible to build the stoichiometric matrix N that characterizes the metabolic network and finally apply constraint-based methods to validate the network or update the network structure when necessary. As mentioned previously, the proposed modeling strategy stands out from existing methods (classical top-down and bottom-up approaches) due to its minimal reliance on genome-scale reconstruction.

2.1. Core pathway selection

Although most metabolic routes are similar for most organisms, other pathways are very specific and can affect the whole metabolism. Essentially, organisms are broadly classified into different categories based on various characteristics, such as cell structure, mode of nutrition, reproduction, and evolutionary history. For instance, regarding the case study presented in Sec. 3, *Arthrospira* sp. PCC 8005 are photosynthetic prokaryotic cyanobacteria. Consulting literature and surveys is the most relevant way to gather information and identify the main metabolic routes forming the network. For this purpose, biochemistry textbooks, online biochemical databases, and genome sequences (when available) can be used.

When genome sequences are used, lumping techniques must be applied to reduce the final network's dimensionality and complexity. In this study, lumping involves combining similar chemical species or reactions into a single entry to simplify the mathematical modeling. The intermediate metabolites that are eliminated from the model are typically those that do not serve as precursors for the synthesis of other essential compounds within the network. These metabolites serve as transient intermediates in specific pathways without contributing to broader connectivity or downstream processes. By removing these non-essential intermediates, the complexity of the network is reduced while retaining its core functional and mechanistic properties.

Furthermore, the reversibility of reactions in metabolic networks is a critical aspect that influences the efficiency, flexibility, and regulation of metabolic pathways simultaneously. Reversible reactions allow cells to adapt to environmental changes, maintain metabolic balance, and optimize resource utilization. For *Arthrospira* sp. PCC 8005, this aspect is particularly important to account for because cyanobacteria behave differently depending on light conditions. For instance, experimental observation demonstrates a large accumulation of lipids and carbohydrates during the day and their consumption at night. In mathematical modeling, considering reaction reversibility is a standard approach in network reconstruction, and can be addressed in multiple ways. As mentioned in many studies [51–54,55], the most consistent approach is to define concentration ranges for metabolites and to evaluate the thermodynamic feasibility of flux solutions. In this study, the reversibility of the reactions is stated specifically based on the standard Gibbs free energy change $\Delta G'^{\circ}$ and the enzyme regulation, i.e., the enzyme's capability to catalyze a metabolic reaction in both directions depending on cellular conditions. For this purpose, databases BRENDA [56] and KEGG [57] are relevant and helpful resources. In addition, reaction reversibility can be tuned and cross-validated by using a constraint-based method to identify blocked reactions in the network. See Sec. 2.2.2 for more information.

2.2. Mathematical methods

After collecting a representative set of biochemical reactions, it is possible to infer the corresponding stoichiometric matrix of the network. In metabolic engineering, a metabolic network is defined by a $m \times n$ stoichiometric matrix N where m represents the number of internal metabolites, and n corresponds to the number of biochemical reactions. Based on the mass balance principle, the general equation of internal metabolite dynamics is expressed as follows:

$$\frac{d\underline{C}}{dt} = N\underline{\nu} - \mu\underline{C} \tag{1}$$

where $\underline{C} \in \mathbf{R}^m$ denotes the vector of metabolite concentrations, $\underline{\nu} \in \mathbf{R}^n$ is the vector of network fluxes and $\mu\underline{C}$ represents the dilution occurring due to the cell growth. As stated before, $N \in \mathbf{R}^{m \times n}$ is the stoichiometric matrix where the element N_{ij} is related to the metabolite i taking part in reaction j.

2.2.1. Elementary mass conservation

To ensure mass conservation, a variable z can be defined. It represents the total number of relevant atoms involved in the metabolic reactions and taking part in the cellular metabolism. Consider the general equation of internal metabolite dynamics in Eq. (1) for which the dilution term is neglected because it does not affect the fundamental mass balance:

$$\frac{d\underline{C}}{dt} = N\underline{\nu}(\underline{C}) \tag{2}$$

A non-negative vector $\underline{a}$ is introduced, which contains information about the number of atoms (of carbon, nitrogen, or phosphorus) in each metabolite:

$$\underline{z} = \underline{a}^T\underline{C} \tag{3}$$

By differentiating $\underline{z}$ with respect to time, it follows:

$$\frac{d\underline{z}}{dt} = \underline{a}^T\frac{d\underline{C}}{dt} = \underline{a}^T N\underline{\nu}(\underline{C}) \tag{4}$$

Henceforth, if mass conservation applies, then $\underline{z}$ must be constant over time, so its time derivative is null:

$$\underline{a}^T N\underline{\nu}(\underline{C}) = 0,\ \forall\, \underline{\nu}(\underline{C}) \rightarrow \underline{a}^T N = 0 \tag{5}$$

Therefore, if the vector $\underline{a}$ exists, then there is mass conservation when constructing the network. It is important to note that for the mass balance of biomass and macromolecules (proteins, nucleic acids, lipids), an average atomic composition must be considered in agreement with the monomerical composition used to state their biosynthetic reactions. Also, the elemental composition of biomass is highly variable over time due to factors such as species-specific differences, growth phase, and environmental conditions. This variability is critical and must be considered in metabolic analyses, as it directly influences the stoichiometry of biochemical reactions. However, an average elemental composition is often assumed as a practical simplification to ensure consistency and facilitate the verification of elementary mass conservation in such analyses.

2.2.2. Constraint-based methods

On the one hand, it is reasonable to assume that internal metabolites reach steady-state instantaneously in comparison with extracellular compounds so that their transient behavior can be neglected. On the other hand, the influence of dilution due to cell growth is negligible compared to the internal and exchange reaction rates so that $\mu\underline{C}$ vanishes. Henceforth, by applying the pseudo-steady state assumption (QSSA), Eq. (1) becomes the following system of linear equations:

$$N\underline{v} = 0 \tag{6}$$

In addition, network fluxes are often subject to positivity constraints assuming that direct reactions prevail over their reverse counterparts:

$$\underline{v} \geq 0 \tag{7}$$

In any realistic large-scale metabolic model, there are more biological reactions than compounds, leading to under-determined systems for which no unique solution exists. Constraint-based methods may alleviate this issue by applying additional constraints to limit the possible solutions and explore the solution space.

Flux balance analysis is undoubtedly the most popular constraint-based method that involves computing an optimal flux distribution $\underline{v}$ which maximizes or minimizes an objective function. This method assumes that cells regulate their fluxes in order to optimize some relevant biological objectives, e.g., maximization of biomass growth or the ATP production rate [12]. Despite the efficiency of FBA in finding optimal solutions, the latter may not correspond to the actual flux distribution, which, moreover, might not be unique.

Flux variability analysis is another constraint-based method that computes the maximal and minimal values of each reaction flux and so the intervals bounding internal fluxes [58]. In this way, it ensures that the actual flux distribution belongs to the interval. The extreme values of the set of admissible fluxes can be obtained using the FVA approach defined as a series of 2 n linear programs.

Flux coupling analysis is another framework for studying the topological and flux connectivity features of genome-scale or large metabolic networks. This method requires the solution of a sequence of linear programs and enables the reduction of the dimensionality of the network by the identification of pairs of metabolic fluxes. To reduce the number of linear programs to solve, it is suggested to develop an algorithm that identifies only blocked reactions, defined as reactions incapable of carrying a flux under steady-state conditions for a particular uptake scenario. In the context of this study, blocked reactions are used to identify incomplete pathways, to pinpoint errors or omissions in the metabolic reconstruction, and also to detect potential issues with reaction reversibility. This last feature appears particularly interesting because the actual Gibbs free energy change depends on the actual concentrations of reactants and products, which can affect the reaction's directionality. Furthermore, depending on the uptake scenario, it is feasible to identify inconsistencies in the network and redundant reactions. Mathematically, the set of blocked reactions is identified by solving the following linear programming problem once for every flux:

$$maximize \quad v_j \tag{8}$$

$$subject\ to \quad \sum_{j=1}^{N} N_{ij} v_j = 0,\ \forall i \in M \tag{9}$$

$$v_j^{uptake} \leq v_j^{uptake^{max}},\ \forall j \in N_{transport} \tag{10}$$

$$v_j \geq 0,\ \forall j \in N \tag{11}$$

It consists in the maximization of a particular flux v_j for a metabolic network comprised of a set of M = {1,…,M} metabolites and a set of N = {1,…,N} metabolic reactions. This formulation requires that reversible reactions are expressed as two irreversible reactions in opposite directions, constraining all fluxes to positive values. Therefore, if the maximum value of the flux is zero, then the reaction is said to be unusable or blocked. In addition, it is possible to limit the uptake of resources to the network using the constraint (10), and the maximum uptake rate of any metabolite absent from the external medium is set to zero. Also, it is important to note that the solution of the linear program depends on the steady-state assumption, imposed uptake/secretion scenarios, growth requirements, and energy production requirements. Therefore, a blocked reaction does not necessarily mean that the reaction is incorrect, inconsistent, or irrelevant in describing the metabolism of the cell; it could only mean the reaction is not activated regarding the actual uptake/secretion scenario. To evaluate the merits and the effective operation of flux coupling analysis, the algorithm has been performed on existing networks for the modeling of *Tisochrysis lutea* [59], *Chlorella sorokiniana* [60] and the mixotrophic growth of *Chlorella sp.* [61]. Results of the algorithm are consistent, demonstrating its reliability. Therefore, it can confidently be used for the current network under study.

The latter constraint-based techniques are complementary approaches, even though the methods and contexts differ. For instance, FCA focuses on understanding reaction dependencies and relationships, while FVA aims to determine the range of feasible fluxes for each reaction. In this way, it is also possible to identify blocked reactions using FVA. However, while unusable reactions in FCA are considered in the context of network structure and inter-dependencies, they are considered in the context of flux variability and constraint satisfaction in FVA. Also, depending on the size of the metabolic network, some of these methods are more appropriate than others. In particular, FCA is faster than FVA for larger networks [19]. For the present study, this set of methods allows the cross-validation of the results.

2.3. Cellular energy metabolism

Another focus of this study, in addition to the elaboration of mid-size metabolic network, is the modeling of the proton motive force (PMF), which is an electrochemical gradient that drives the synthesis of adenosine triphosphate (ATP), a molecule that serves as the primary energy currency in cells.

Essentially, organisms can synthesize ATP through various mechanisms depending on the availability of oxygen and the presence of light. First of all, as it does in all living organisms, the microorganisms under study generate ATP through a process called substrate-level phosphorylation, where ATP is generated directly during the metabolic pathways such as glycolysis and the citric acid cycle. This metabolic process generates ATP directly through enzymatic reactions without the involvement of an electron transport chain or a proton gradient or membrane-bound ATP synthase complexes. Therefore, it provides a means to generate ATP quickly in the cytoplasm, particularly under anaerobic conditions or when light is not available for photosynthesis.

In aerobic organisms, ATP is also produced through a process called oxidative phosphorylation, where cells use oxygen as the final electron acceptor in the electron transport chain (ETC) to produce ATP.

Furthermore, in photosynthetic cells, ATP can be generated under light conditions through a process called photophosphorylation, involving the oxidation of H_2O to O_2, with $NADP^+$ as the ultimate electron acceptor. In these cases, the synthesis of ATP is driven by the proton motive force, denoted PMF, which is a critical component of energy generation and is involved in various cellular processes. This electrochemical gradient is generated across a particular membrane and is based on the chemiosmotic theory that states that the movement of electrons through the electron transport chain creates a proton gradient across a membrane, which drives the synthesis of ATP. As electrons move through the electron transport chain, protons (i.e., H^+ ions) are pumped across the membrane. The pumping of protons creates a higher concentration of H^+ on one side of the membrane, producing a proton gradient that will serve as the driving force for ATP synthesis from ADP and P_i, catalyzed by a membrane-bound ATP synthase complex. For the process of oxidative phosphorylation, the membrane is the mitochondrial membrane in eukaryotic cells or the plasma membrane in prokaryotic cells. For the process of photophosphorylation, the membrane is the thylakoid membrane of chloroplasts in eukaryotic cells or the thylakoid membrane of thylakoid-like membrane structures inside the prokaryotic cells that serve as sites for light-dependent reactions.

Even though the mechanism of ATP synthesis is well documented, the corresponding chemical reactions differ according to studies or biochemistry textbooks, which leads to confusion, particularly in mathematical modeling. In that respect, this study proposes to differentiate the cytoplasmic protons $H^+_{\mathbf{n}}$ (namely H^+) and the periplasmic protons $H^+_{\mathbf{p}}$ to attempt to model the proton gradient (see Fig. 3). The synthesis of ATP from ADP and P_i catalyzed by a membrane-bound ATP synthase complex is thus given as follows:

$$ADP + P_i + \eta\, H_p^+ \quad \rightarrow ATP + H_2O + \eta\, H_n^+ \tag{12}$$

Where η represents the number of protons required to produce one molecule of ATP and varies depending on the specific process and pathway. In photosynthetic microorganisms, 4 protons are required to form one molecule of ATP ($\rightarrow\eta = 4$).

Regarding photophosphorylation, the first step of light-driven phosphorylation is the so-called oxygenic photosynthesis, where light energy is captured by chlorophyll and other pigments to enable the conversion of CO_2 and H_2O into carbohydrates, releasing oxygen as a byproduct. During this process, electrons flow through a series of membrane-bound carriers, including cytochromes, quinones, and iron-sulfur proteins embedded in the thylakoid membrane (see Fig. 4). At least 8 photons must be absorbed to drive four electrons from H_2O to $NADP^+$, leading to the movement of 12 protons across the thylakoid membrane. In this way, the reactions relative to photophosphorylation are obtained:

$$2\, H_2O + 2\, NADP^+ + 8\, photons \rightarrow O_2 + 2\, NADPH + 12\, H_p^+ \tag{13}$$

$$ADP + P_i + 4\, H_p^+ \rightarrow ATP + H_2O + 4\, H_n^+ \tag{14}$$

Therefore, this proton gradient can produce approximately 3 molecules of ATP, which is in concordance with the approximate stoichiometry of photophosphorylation established in [62] stating experimental measurements yield values of about 3 ATP per O_2 produced.

In cellular respiration, the synthesis of ATP also involves other cofactors (NADH and $FADH_2$) that contribute to the generation of the proton gradient. The summary of the flow of electrons and protons through the four complexes of the respiratory chain is given in Fig. 5, and the detailed chemical reactions can be found in [62]. The oxidation of NADH induces the movement of 10 protons (through complexes I, III, and IV) across the membrane, while the oxidation of $FADH_2$ results in the translocation of 6 protons (through complexes III and IV) across the membrane. Thus, the simplified equations are written as follows:

$$NADH + 11\, H_n^+ + 0.5\, O_2 \rightarrow NAD^+ + 10\, H_p^+ + H_2O \tag{15}$$

$$FADH_2 + 6\, H_n^+ + 0.5\, O_2 \rightarrow FAD^+ + 6\, H_p^+ + H_2O \tag{16}$$

$$ADP + P_i + 4\, H_p^+ \rightarrow ATP + H_2O + 4\, H_n^+ \tag{17}$$

Therefore, 2.5 molecules of ATP are produced per NADH, and 1.5 molecules of ATP are produced per $FADH_2$. The previous chemical reactions can be compared with the global reactions related to photophosphorylation and oxidative phosphorylation, commonly found in the literature.

3. Case study of Arthrospira sp. PCC 8005

3.1. Core metabolic pathways

This section presents a simplified overview of the main metabolic routes of *Arthrospira* sp. under photoautotrophic conditions. These metabolic pathways and the corresponding chemical reactions are mainly derived from the reference textbook in biochemistry [62]. Unlike eukaryotic plants and algae that possess chloroplasts as distinct organelles within their cells, *Arthrospira* sp. PCC 8005 is a prokaryotic photosynthetic organism having specialized structures that are the sites for light-dependent reactions. These thylakoid-like membranes are spread throughout the cell's cytoplasm and contain chlorophyll and other pigments necessary for photosynthesis.

The main metabolic pathways for photosynthetic organisms are the Calvin-Benson cycle (fluxes v_2 to v_5 in Table 2), the Embden-Meyerhof-Parnas pathway (fluxes v_6 to v_{13}), the Pentose Phosphate pathway (fluxes v_{18} to v_{26}), the tricarboxylic acid cycle (fluxes v_{27} to v_{36}) and the GS-GOGAT pathway (fluxes v_{38} to v_{43}). Furthermore, as discussed previously, the cellular energy metabolism is given by fluxes v_{120} to v_{123}.

Another way to regulate ammonium level is via the urea cycle-like pathway. Although photosynthetic organisms primarily manage nitrogen through the GS-GOGAT pathway, certain cyanobacteria, including *Arthrospira* sp. PCC 8005, possess the enzymatic machinery to synthesize urea under specific conditions (confirmed by the proteomic study in [43] and the analysis of the genome sequence of the similar strain NIES-39 [48]). Instead of serving as an excretory pathway, as in ureotelic organisms, the urea cycle-like pathway in cyanobacteria likely

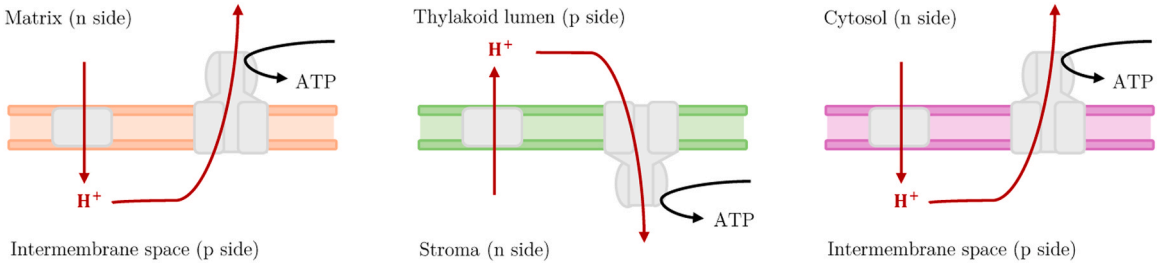

Fig. 3. Schematic representation of the proton motive force across the mitochondrial membrane (left-hand side), the thylakoid membrane (center) and the plasma membrane (right-hand side).

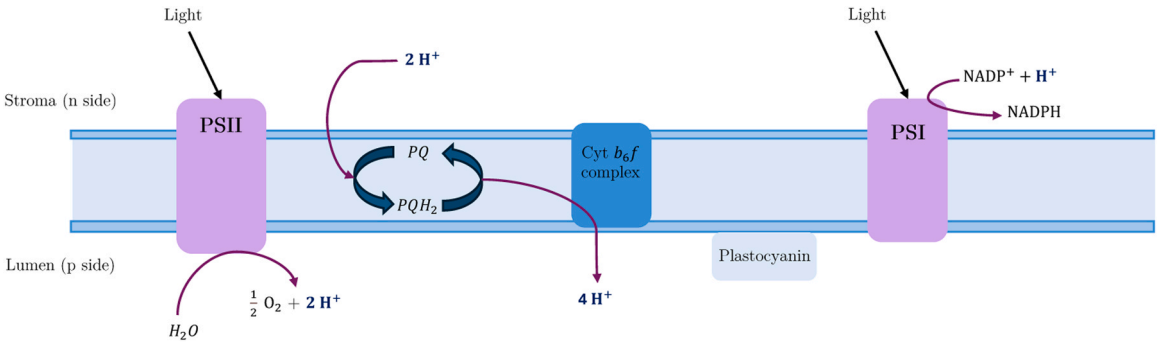

Fig. 4. Illustration of the process of photophosphorylation. Photosystem II (PSII) absorbs photons, splits water into O_2, H^+ protons and electrons while generating a proton gradient for ATP synthesis. Photosystem I (PSI) then uses the excited electrons to reduce $NADP^+$ into NADPH.

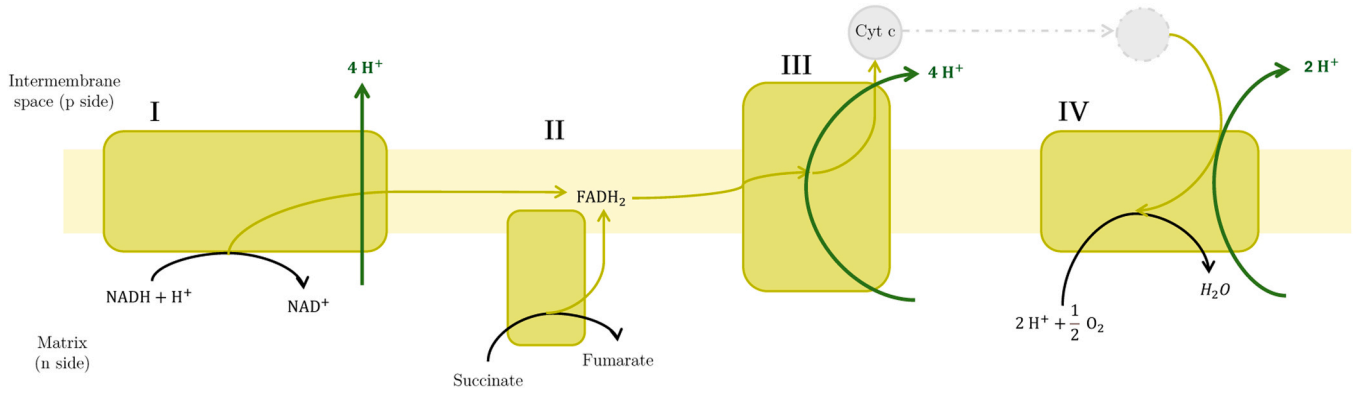

Fig. 5. Flow of electrons and protons through the 4 complexes of the respiratory chain.

plays a role of nitrogen recycling and metabolic flexibility. The corresponding metabolic reactions are given by fluxes v_{44} to v_{49} in Table 2.

Most of the previous pathways (other than the Calvin-Benson cycle) are catabolic, meaning that the major metabolic fuels are degraded to enter the citric acid cycle and yield their electrons to the respiratory chain, allowing the synthesis of ATP. Nevertheless, anabolic pathways are also fundamental to describe the whole metabolism of an organism. Those pathways use chemical energy in the form of ATP, NADH or NADPH to synthesize cellular components from simple precursor molecules. The most important anabolic pathways are the biosynthesis of carbohydrates, the biosynthesis of lipids, and the biosynthesis of amino acids and nucleotides. Finally, information pathways (proteins, DNA, and RNA metabolism) must be included in the metabolic network in order to form biomass. Most of the anabolic pathways can be found in [62], especially metabolic reactions characterizing the metabolism of amino-acids (fluxes v_{70} to v_{101} in Table 2) and the nucleotide salvage pathway (fluxes v_{102} to v_{118}).

As highlighted in Sec. 1, a benefit of the methodology presented in this paper is its minimal reliance on genomic reconstructions (not always accessible or reliable for lesser-studied or newly-isolated strains). As a consequence, in the first instance, and for the sole purpose of demonstrating feasibility (i.e., handling situations where genome sequences are not available), simplified chemical reactions are considered for some anabolic pathways. Then, in the second instance, because genome sequences of similar strains are available, the latter will be exploited to better estimate the composition of the main components of cell material and to identify specific metabolic routes.

When genome sequences are not available, it can be assumed that the synthesis of carbohydrates begins with glucose-6-phosphate (G6P), an intermediate in glycolysis, which is converted into glucose-1-phosphate (G1P) by the enzyme phosphoglucomutase. Glucose-1-phosphate is then activated by nucleotides such as UTP to form UDP-glucose, a building block for synthesizing polysaccharides like glycogen, which serve as carbohydrate reserves in cells. Regarding the synthesis of lipids, it can be assumed that it starts with glycerol-3-phosphate (G3P), which is derived from glycolysis through the reduction of dihydroxyacetone phosphate. Then, glycerol-3-phosphate is acylated with two molecules of acyl-CoA, forming phosphatidic acid via acyltransferases. The acyl-CoA molecules are derived from fatty acids, which are themselves synthesized from acetyl-CoA, a product of central carbon metabolism. Phosphatidic acid is a precursor for various lipids. The latter can form triglycerides or alternatively be modified to produce phospholipids, key components of cell membranes. Finally, the elemental composition of proteins and biomass can be estimated using data from similar organisms. Experimental methods, like protein hydrolysis for amino acid profiles or solvent extraction for lipids, can also be useful to determine key components without genomic data. Therefore, if desired, all major metabolic pathways could be expressed without relying on genomic data.

However, in this particular case, because genome sequences are available for *Arthrospira* sp. (different strains), the composition of the main components of cell material (proteins, carbohydrates, lipids, and nucleic acids) and other macromolecules can be provided by genomic studies [47–49], as well as the average formula of biomass. Doing so, by using the genome sequence of *S. platensis* C1 PCC 9438 [47] and applying lumping techniques, the reactions describing the synthesis of carbohydrates are given by fluxes v_{133} to v_{146}, the reactions for the synthesis of lipids are represented by fluxes v_{147} to v_{166} and the elemental composition of the macromolecules and the biomass is depicted by fluxes v_{173} to v_{178} in Table 2 (note that the stoichiometry will be adjusted later using experimental data). More specific pathways relative to the strain under study are also derived from this genome sequence and will be detailed in the sequel. For instance, the synthesis of chlorophyll, especially important for photosynthetic organisms, which is

depicted by fluxes v_{167} to v_{171}.

3.2. Iterative procedure

As mentioned previously, the method proposed is iterative, and model refinement is possible by comparing model predictions with experimental observations, which may include available experimental data or well-established findings from literature or surveys. This section outlines a couple of iterations before obtaining a final draft for the metabolic network. Table 1 summarizes the iterations.

3.2.1. First iteration

The first iteration considers the core metabolic pathways described in Sec. 3.1. The reversibility of the reactions is analyzed by considering the standard Gibbs free energy change found in literature, and elementary-mass conservation is verified. This first iteration leads to a metabolic network with 157 metabolites and 167 reactions.

In order to check network consistency, as well as structural and topological network properties, constraint-based methods can be applied. For this purpose, a particular uptake scenario is considered, for which light, a carbon source (HCO_3^-), and a nitrogen source (HNO_3^-) are available. To avoid infeasible solutions, a set of exchange reactions is added to the network (fluxes v_{179} to v_{198}), as well as a metabolic reaction that includes ATP maintenance requirements (flux v_{124}) to simulate non-growth-associated energy demands. In addition, the objective function of the flux balance analysis should represent a biologically relevant process. Accordingly, a flux balance analysis with biomass maximization is suggested, and a flux coupling analysis is performed to prevent large computation times. After careful analysis, the network provides inconsistent results. As a matter of fact, the biomass production rate obtained from the FBA is null, which is not expected considering the uptake scenario, and flux coupling analysis provides several blocked reactions, meaning there are gaps in the proposed network structure.

However, these inconclusive results can be explained. First, in the presence of light, cyanobacteria reduce atmospheric CO_2 to triose sugars before converting them into glucose via the Calvin-Benson cycle and gluconeogenesis. The latter pathway is missing from the network and needs to be considered (addition of fluxes v_{14} to v_{17} - found in [62] - and adjustment of the reversibility of several reactions of the Embden-Meyerhof-Parnas pathway). Also, flux coupling analysis has revealed 39 blocked reactions, which are reactions unable to carry flux under the defined uptake scenario. These reactions often indicate gaps in the identified metabolic network. By inspecting the blocked reactions carefully, it is possible to identify missing pathways or transport steps necessary for the synthesis or utilization of intermediate metabolites. Indeed, addressing these gaps improves the accuracy and completeness of the network reconstruction, ensuring that all essential metabolic connections are captured. In this case, it appears that several reactions enabling the metabolism of amino acids and reactions belonging to the nucleotide salvage pathways are inactive, which explains why biomass production is impossible when applying FBA. The main reasons are the following:

- The Lewis acid-base reaction v_{40}

  Initially, the reaction v_{40} was set irreversible, assuming that ammonia NH_3 primarily acts as a weak base and reacts with a proton H^+ to form the ammonium NH_4^+. However, acid-base equilibria are inherently reversible, with the direction of the equilibrium shifting based on factors such as the concentration of reactants and products, the pH of the medium, the temperature, or the presence of other ions or compounds. While the equilibrium generally favors ammonium formation under physiological conditions, FCA revealed blocked reactions under the specific uptake scenario. Therefore, to ensure a more generalized and functional metabolic network, the reaction was modeled as reversible.
- The missing pathways of tetrahydrofolate metabolism and S fixation

  The formation of several amino acids requires H_2S and tetrahydrofolate (THF) or its byproducts (CHO-THF; CH_2-THF; CH_3-THF; DHF) as reactants. In addition, some of them also intervene in the nucleotide salvage pathway. Because the synthesis of such intermediate metabolites is not considered in the network, this leads to several inactive reactions that are essential in the formation of biomass, which explains the result of the FBA. Therefore, they need to be synthesized to avoid incomplete pathways or gaps in the network reconstruction, leading to infeasible solutions. In that regard, the pathways of tetrahydrofolate metabolism and S fixation are added to the network using genome sequences and lumping techniques. The corresponding metabolic reactions are represented by fluxes v_{50} to v_{69} in Table 2.

3.2.2. Second iteration

After making the Lewis acid-base reaction reversible and adding the pathways of neoglucogenesis, tetrahydrofolate metabolism and S fixation, the metabolic network is comprised of 170 metabolites and 191 reactions. Elementary-mass conservation is still verified. For the same uptake scenario as the one considered in Sec. 3.2.1, FBA provides plausible results (e.g., the specific growth rate is non-zero), which highlights the benefits of the iterative procedure. More details on the outcomes of the metabolic analyses will be given in Sec. 4 and 5. However, FCA still identifies unusable/blocked reactions.

In this case, the unusable reactions are due to an excessively detailed representation of the metabolism of tetrahydrofolate. In this way, the reactions detailing the formation of dihydrofolate from glutamate, para-aminobenzoate, and GTP (fluxes v_{53} to v_{61} - obtained from the genome sequence in [47] and using lumping techniques) are not strictly necessary and can be removed from the network because these reactions involve intermediate metabolites that are not currently used in the synthesis of other essential components. However, it is worth noting that it depends on the scope and focus of the model. If more specific pathways are added later on, these metabolites might become relevant and necessary. Henceforth, it is worth noticing that flux coupling analysis can be used as a method to identify blocked reactions and, thus, inconsistencies in the network structure, but also as a method to simplify over-detailed networks.

3.2.3. More iterations

The third iteration involves removing unnecessary reactions from the pathway of tetrahydrofolate metabolism. Besides, more iterations are needed to improve the structure of the metabolic network. In this regard, the fourth iteration consists of removing redundant reactions due to lumping techniques or the conflicting nomenclature of chemical compounds from the literature. The fifth iteration allows for discarding reactions involving specific metabolites that do not take place in the biosynthesis of carbohydrates, lipids, proteins, or other macromolecules. The sixth iteration allows the addition of cofactor metabolism reactions (fluxes v_{125} to v_{132} - found in [62]) and the reaction of formation of cyanophycin, a key biopolymer that stores excess nitrogen during periods of nitrogen abundance (flux v_{172}). More iterations can be

Table 1
Information related to metabolic networks from the iterative procedure. The number of metabolites is comprised of internal metabolites and intracellular energetic cofactors, and the number of reactions includes the intracellular reactions and the transport reactions.

Iteration	# metabolites	# reactions	# blocked reactions
1	157	167	39
2	170	191	16
3	161	182	7
4	161	179	7
5	154	172	0
6	155	182	0

made, but they are not detailed in this paper for the sake of clarity.

3.2.4. Final metabolic network

After several iterations, and a compromise between the level of detail and the size of the network, a final draft for the metabolic network can be obtained. It comprises 198 reactions (including intracellular reactions and transport reactions) and 171 metabolites (including internal metabolites and intracellular energetic cofactors). For the sake of completeness and to enable the development of enriched networks in further research, the final network also considers the synthesis of specific precursors taking place in the synthesis of important biomolecules and vitamins. The metabolic reactions of the network are given in Table 2. This network models the metabolism of the cyanobacterium under a photoautotrophic regime. This strain is particularly interesting due to its use in bioregenerative life-support systems whose main objectives are to supply a crew with food, water, and oxygen while eliminating waste.

4. In-silico validation

Validating the metabolic network using data is a crucial step to ensure its reliability and predictive power. In this section, in-silico data are used to assess the network's consistency and functionality. Thereafter, the validation process is performed using real experimental data (see Sec. 5), allowing for a more accurate and biologically relevant refinement of the model.

As a first validation or when experimental data are not available, a common strategy for validating a metabolic network is to leverage in-silico data, i.e., data generated through computational models and simulations rather than obtained from experiments, and explore the behavior of the network under defined conditions. The idea is then to analyze whether the model's predictions align with known biological behavior under imposed scenarios, such as nutrient uptake or environmental conditions, in order to assess the network's validity. This approach enables identifying inconsistencies or unrealistic outputs, ensuring the network reflects biological reality and not just mathematical consistency.

This study focuses only on the quasi-steady state under constant light regimes (i.e., continuous illumination), where the pseudo-steady-state assumption (QSSA) is valid, enabling the use of classical metabolic analysis tools. The metabolism adaptation under dynamic conditions, such as permanent fluctuating light, will be considered in future research. Nevertheless, even though this in-silico validation is static and assumes a steady-state framework, it offers insights into the photoautotrophic metabolism of cyanobacteria under the day-night cycle. Indeed, during the day, when light levels are sufficient to meet carbon and energy demands, the autotrophic regime occurs. At night, the metabolism shifts to a heterotrophic regime, breaking down carbon storage molecules into precursor metabolites and energy for growth and maintenance. Otherwise, when light is not intense enough to meet the carbon and energy growth demands, the regime is said to be mixotrophic and is characterized by the consumption of carbon storage molecules. In this section, diverse uptake scenarios are explored to assess the robustness of the constructed network and quantify the intracellular mechanisms. Note that a flux map of the network had been created using *CellDesigner* [63] and can be provided upon request.

4.1. Day phase

The uptake scenarios considered in this section correspond to an autotrophic regime where cyanobacteria use sunlight, CO_2, and water to produce organic compounds, among others. To ensure the metabolic consistency of the network, we first perform simulations under a strict carbon-only uptake scenario, where only light and HCO_3^- are provided, and no nitrogen source is available. In this case, as expected, biomass production is not feasible, since nitrogen is essential for the synthesis of amino acids and nucleotides. However, this scenario allows us to validate our metabolic model by ensuring that, under nitrogen limitation, only core metabolic pathways, such as glycolysis, the tricarboxylic acid cycle, and the pentose phosphate pathway, are activated. Then, to

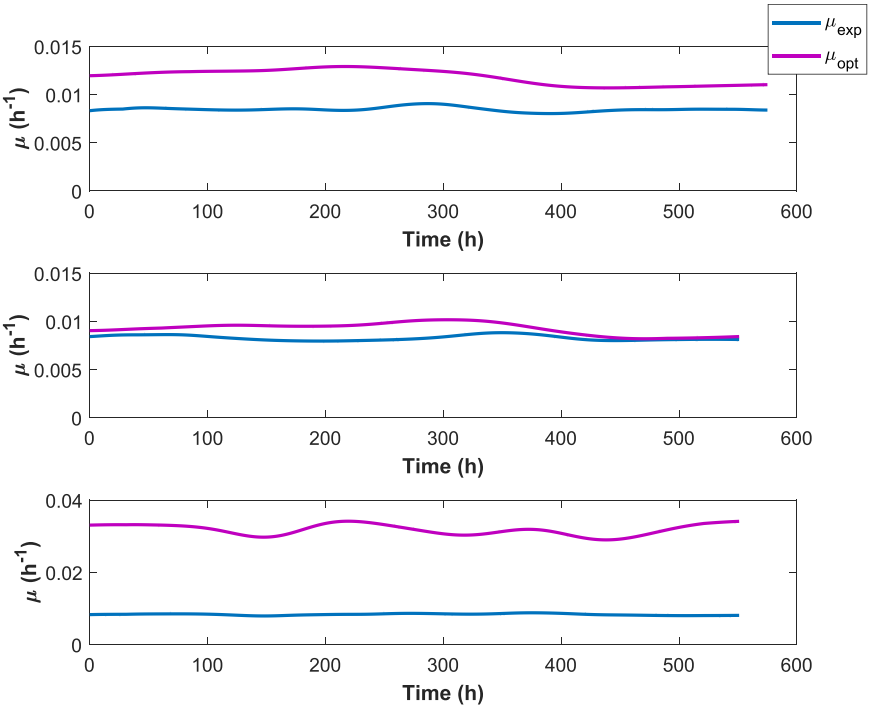

Fig. 6. Experimental specific growth rate (μ_{exp}) and specific growth rate from dFBA (μ_{opt}) for different feeding strategies: 30mM-N nitrate (top); 15mM-N nitrate + 15mM-N nitrite (center) and 15mM-N nitrate + 15mM-N urea (bottom).

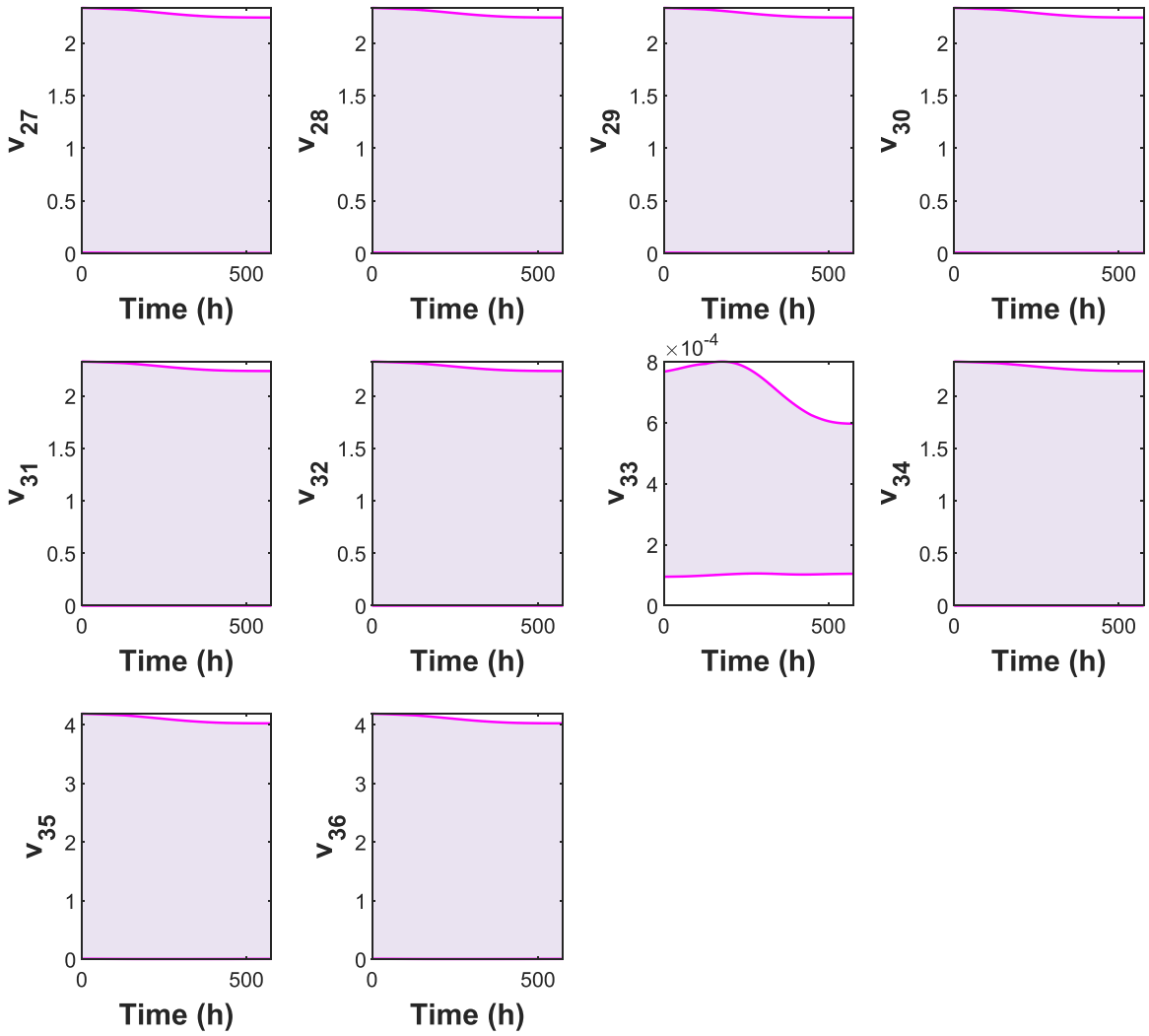

Fig. 7. Dynamic flux variability analysis under the 30mM-N nitrate feeding condition. Fluxes are expressed in mmol.g_{DW}^{-1}.h^{-1}.

further explore nitrogen metabolism, simulations are performed considering a nitrogen source in addition to light and CO_2, enabling the activation of amino acid biosynthesis and macromolecule formation, which are required for biomass growth. Simulations are performed considering a light with an average wavelength of 600 *n*m and a power density of 100 W.m^{-2}, giving a photon flux density of approximately 5.10^{-4} E.m^{-2}.s^{-1}.

4.1.1. Carbon-only uptake scenario

In this section, energy and the carbon input are imposed to mimic an autotrophic regime. Flux variability analysis and flux balance analysis are performed to examine the intracellular mechanisms and predict the metabolic distribution of the fluxes under constant environmental conditions. In order to perform the flux balance analysis, the biomass growth is maximized. Even though biomass formation is infeasible under a carbon-only uptake scenario due to the absence of a nitrogen source, performing FBA with biomass optimization remains informative. This approach ensures that the model does not allocate flux toward biomass formation, confirming that elemental balance is properly enforced. In addition, it allows us to analyze how core metabolic pathways operate under nitrogen starvation. By maintaining a consistent optimization objective across the different uptake scenarios, we can also assess how metabolic fluxes shift when nitrogen becomes available, providing insights into the network's response to nutrient limitations.

After analysis of the flux map, it is observed that autotrophy is characterized by high fluxes in the photosynthetic pathways and by the activation of the Calvin-Benson cycle. Upper glycolysis operates in the glyconeogenic direction to produce carbohydrates and sugar precursor metabolites such as phosphoenolpyruvate, glucose-6-phosphate and ribose-5-phosphate, essential for growth. Also, the pentose phosphate pathway is in the reductive mode, and the citric acid cycle enables the formation of metabolite precursors for biomass growth. These results are in concordance with [64]. As detailed in Sec. 2.3, ATP is generated through glycolysis and the citric acid cycle on the one hand. On the other hand, ATP is driven by the proton motive force (depicted by the electrochemical gradient H_p^+/H_n^+) and is synthesized via photophosphorylation and oxidative phosphorylation. Besides, it can be shown that the influence of light on qualitative flux distribution is minimal as long as light remains the limiting factor. Otherwise, the excessive light energy is dissipated by means of photophosphorylation and the rest of the metabolism remains relatively unchanged in terms of flux distribution. In this way, it enhances the robustness of the photosynthetic pathways against light variation, maintaining a consistent energy supply to the metabolism. As expected, biomass is not produced (i.e., specific growth rate is zero) because no nitrogen source is provided to the organism.

4.1.2. Carbon-nitrogen uptake scenario

In this section, the day phase is still addressed, but a nitrogen source is provided to the organism, in addition to light and a carbon source. In this case, a net assimilation of 10 mol of nitrate NO_3^- is also considered,

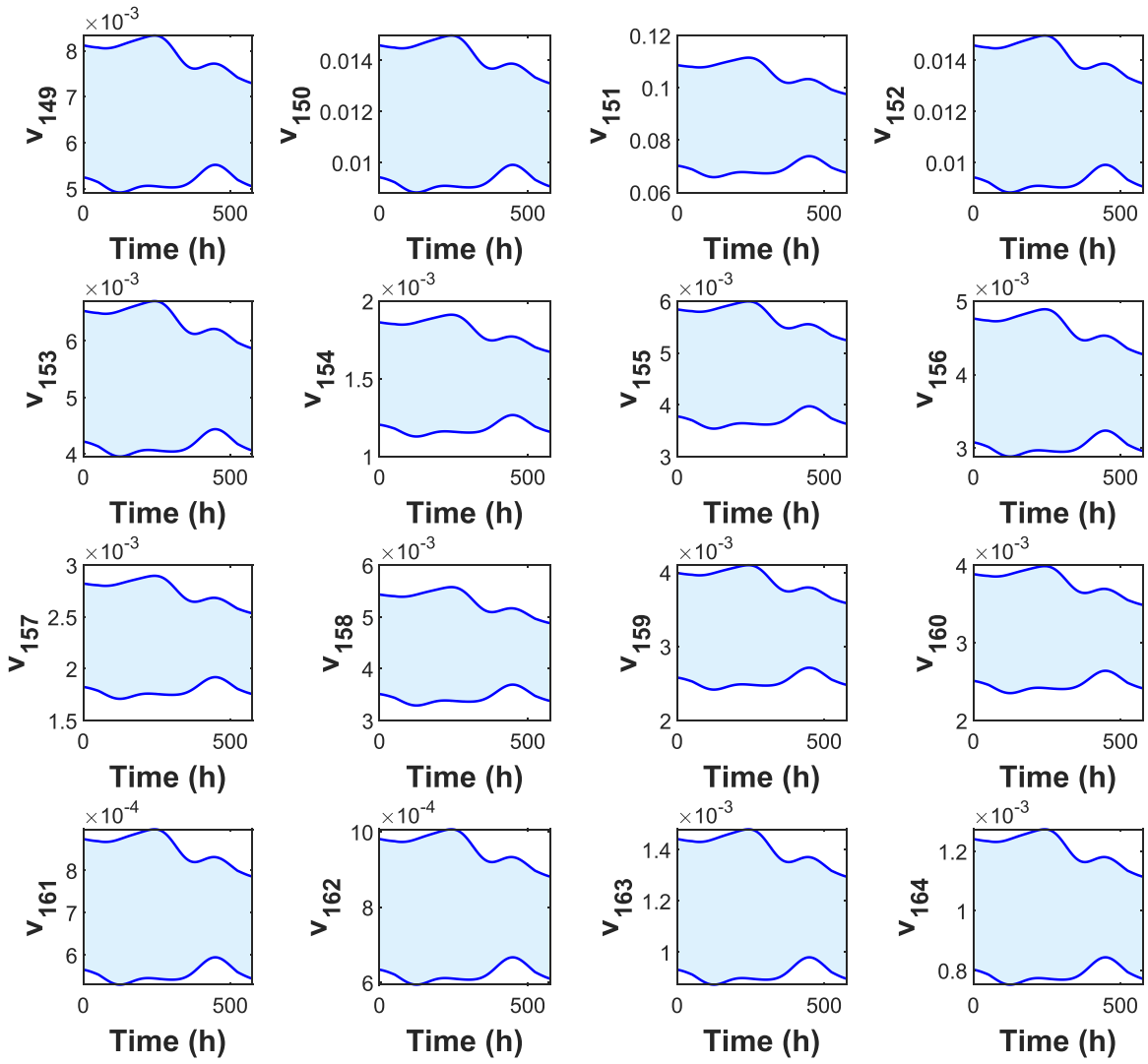

Fig. 8. Dynamic flux variability analysis under the 30mM-N nitrate feeding condition. Fluxes are expressed in mmol.g_{DW}^{-1}.h^{-1}.

known as the preferential nitrogen source for cyanobacteria. After analysis of the corresponding flux map, the same conclusions as before can be drawn regarding the Calvin-Benson cycle, the Embden-Meyerhof-Parnas pathway, PPP and the routes of TCA. Also, it is interesting to note that glyceraldehyde-3-phosphate (G3P) is mainly produced via the Calvin-Benson cycle. The reductive PPP contributes a little, but glycolysis (i.e., the breakdown of fructose-1,6-bisphosphate) is not a major pathway for G3P production, particularly when the organism is in a photosynthetic state. Furthermore, for this uptake scenario, the nitrogen source allows the biosynthesis of amino acids and the activation of the nucleotide salvage pathway. Also, the GS-GOGAT pathway and the urea cycle-like pathway are involved in nitrogen assimilation. Consequently, biomass is produced.

4.2. Dark phase

The uptake scenarios analyzed below correspond to a heterotrophic regime, meaning that carbon storage molecules are degraded into precursor metabolites and energy, essentially for growth and maintenance. To ensure the validity of the metabolic network and facilitate comparisons with existing studies, the heterotrophic regime (equivalent to a chemotrophic regime in this case) is simulated by imposing a net assimilation of 100 mol of glucose, one of the organic carbon sources represented in the network. As in previous analyses, simulations are first conducted with glucose as the sole energy to examine the behavior of the network under carbon-only conditions and verify that core metabolic pathways function as expected under heterotrophic conditions. Subsequently, a source of nitrogen is introduced to enable biosynthetic processes and biomass formation.

In heterotrophic conditions during the dark phase, optimizing biomass formation remains a relevant objective, as many cyanobacteria are capable of growth when both carbon and nitrogen sources are available. However, depending on the physiological state of the organism, alternative metabolic objectives may be more biologically meaningful. For instance, if only a carbon source is available, cells are unlikely to prioritize growth and may focus on maximizing ATP production to sustain vital cellular functions. Another plausible objective could be the accumulation of storage compounds, such as glycogen, which would later serve as an energy reserve for the next light phase.

4.2.1. Carbon-only uptake scenario

Flux variability analysis and flux balance analysis are performed by setting the transport reaction of glucose and no light energy is considered due to the dark phase. The corresponding flux map is not represented for the sake of clarity, but the operation of the main pathways is resumed thereafter. The absence of light inhibits the activation of the Calvin-Benson cycle, which normally produces glucose from CO_2 during the day phase ($v_2 = 0$). When glucose is used as the carbon source, upper

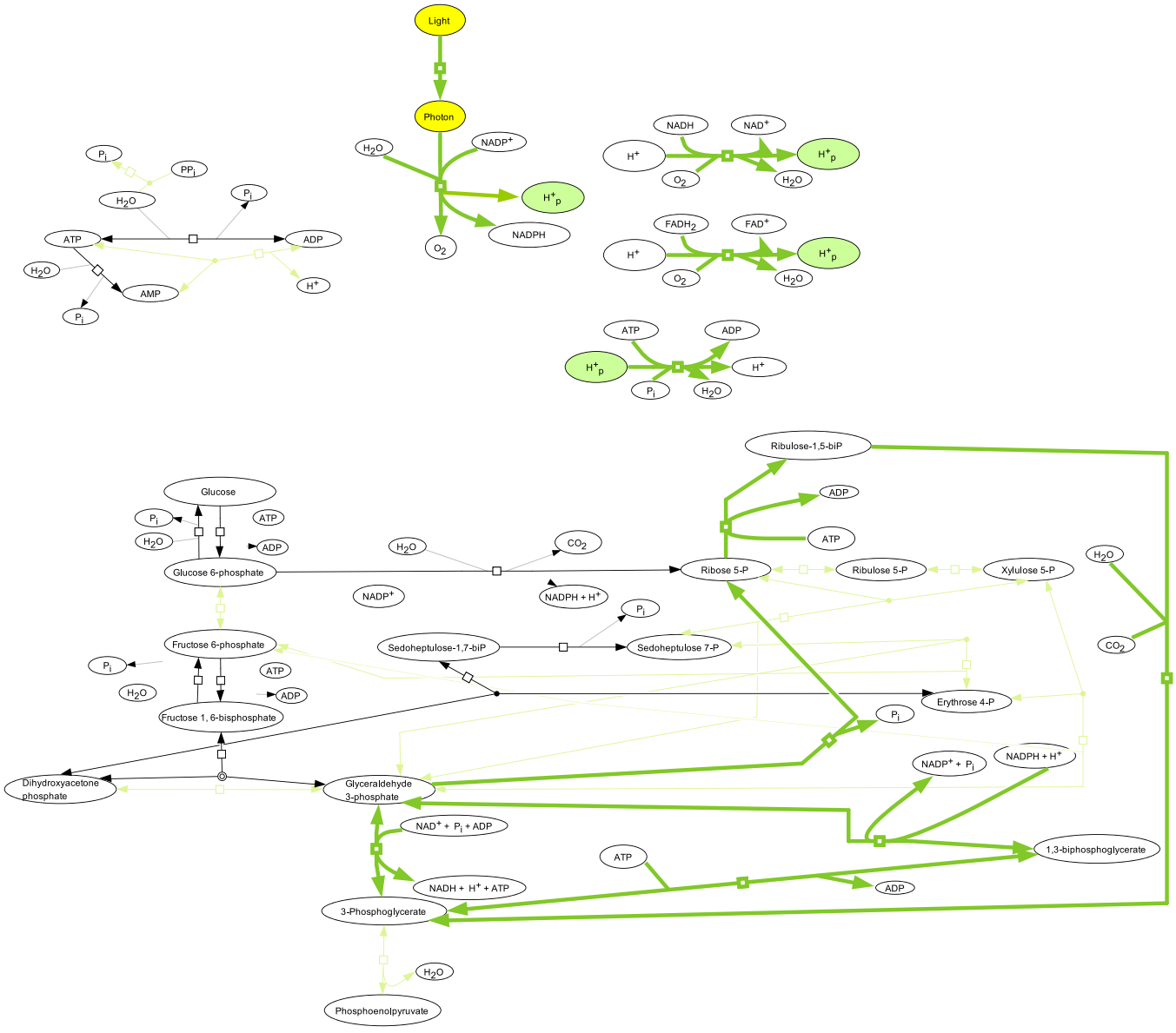

Fig. 9. Simplified flux map at a specific time instant. The flux map depicts the energy cellular metabolism, glycolysis and the pentose phosphate pathway.

glycolysis is in the downward direction, and the pentose phosphate pathway is in the oxidative mode. Also, the primary carbon flux is through the tricarboxylic acid cycle, which generates precursor metabolites for growth and energy via oxidative phosphorylation. Indeed, during the night, ATP is not synthesized by photophosphorylation. Furthermore, when proceeding in the downward direction, glycolysis also contributes to energy demands, and it can be established that approximately 40 % of the carbon is lost through respiration [9]. These latter conclusions are supported by studies covering the modeling of photosynthetic bacteria.

4.2.2. *Carbon-nitrogen uptake scenario*

This uptake scenario deals with the dark phase, considering glucose as a carbon source and nitrate as a nitrogen source. Same conclusions can be drawn as in Sec. 4.1.2, except the Calvin-Benson cycle is not activated due to the absence of light. However, an interesting observation can be made, which is in agreement with [64] modeling unicellular microalgae. As mentioned in the previous section, PPP is activated in the oxidative mode in heterotrophic regimes, mainly to fulfill the NADPH requirements, enabling the synthesis of macromolecules such as amino acids, nucleotides, and lipids. Nevertheless, after careful analysis of the flux map corresponding to carbon-nitrogen uptake scenario during dark phase, it is observed that the presence of NADPH/NAD^+ transhydrogenase alpha subunit (EC number: 1.6.1.2) converting NADH to NADPH via v_{126} is usually preferred as source of NADPH, implying a nearly null flux into the pentose phosphate pathway via v_{18}. This result contradicts experimental findings proving that PPP remains the primary route for NADPH synthesis. This allows highlighting the limitation of in-silico data and it reminds the importance of experimental data. This discrepancy can arise from missing regulatory mechanisms or oversimplified assumptions in the model. Also, the more data, the better the flux predictions provided by metabolic analyses. As a reminder, the identified metabolic network is underdetermined and a calculability/observability analysis might be conducted to determine the number of fluxes that can be uniquely calculated [65]. In the present study, the inaccuracy pointed out can be corrected by limiting the flux of NADPH/NAD^+ transhydrogenase complex in the FBA approach, leading to results consistent with experimental findings.

This in-silico validation confirms network consistency with metabolic predictions that reflect known biological behavior. Essentially,

autotrophic and heterotrophic fluxes differ in the arrangement of the core carbon network including the Calvin-Benson cycle, the Embden-Meyerhof-Parnas pathway, the pentose phosphate pathway and the citric acid cycle. The remaining pathways involved in the biosynthesis of amino acids and macromolecules, and the information pathways show relatively consistent flux patterns regardless of growth conditions. In that respect, it suggests that anabolic processes operate independently of growth conditions, which is explained by the classical bow tie structure of microorganisms.

5. Experimental validation

Although in-silico validation helps identify potential gaps or errors in the network structure and confirm network consistency, in-silico data have limitations. Therefore, validating the network against experimental data is crucial to confirm that it accurately represents actual metabolic behavior.

5.1. Cell line and media

Data were provided by the Department of Proteomics and Microbiology of the University of Mons. More information relative to the cell line, the media, the bioreactor operation mode and the analysis methods can be found in [44].

The photo-bioreactor (PBR) started under batch mode, and continuous feeding started after 7 days (only data in continuous mode were provided). For the continuous PBR experiments, cyanobacteria were incubated in a radially illuminated 2 L cylindrical double jacket reactor under a constant light power density of 125 $W.m^{-2}$ with the agitation settled at 150 rpm. Note that this constant light regime is achieved using artificial lighting systems that provide uniform light conditions, meaning that the quasi-steady state assumption holds. The pH was automatically maintained at 8.5 with HCl (0.5 M), whereas the dilution rate was kept around 0.2 per day with fresh Zarrouk medium (30mM-N $NaNO_3$ or a mixture of nitrogen sources). Experimental data in continuous mode are provided for different feeding strategies (30mM-N nitrate; 15mM-N nitrate + 15mM-N nitrite and 15mM-N nitrate + 15mM-N urea).

Experimental data include biomass and oxygen productivity; residual nitrogen concentration (i.e., NO_2^-, NO_3^-, NH_4^+ and urea); pigment concentration (i.e., $chlorophyll_a$); TC, TOC, TIC and TN in biomass and supernatant, and protein, lipid, and carbohydrate content in biomass. O_2 is monitored in exhaust gas (purging with air to prevent an increase in the oxygen saturation of the medium).

5.2. Numerical results

As mentioned previously, this study focuses on the quasi-steady state under constant light regimes, using the full spectrum of classical metabolic analysis tools. In this case, dynamic metabolic frameworks, such as dynamic flux balance analysis (dFBA) and dynamic flux variability analysis (dFVA), are used because experimental uptake and secretion rates are computed from concentration profiles over time. These methods extend traditional FBA and FVA by integrating time-dependent metabolic constraints. Several tests have been performed to validate the metabolic network. For direct validation, it is proposed to perform dFBA, dFVA and flux consistency checks via the analysis of a flux map and to compare the results of such analyses with published data from literature and databases.

Initially, a series of fundamental tests is conducted to verify the coherence of the network. For instance, flux variability analysis is achieved by imposing only experimental uptake rates and ensuring experimental secretion rates belong to the predicted intervals. Also, dFBA with biomass optimization is performed by imposing both experimental uptake and secretion rates and ensuring that the predicted optimal value for the growth rate is higher than the measured growth rate. Doing these analyses is a prudent approach to enhance confidence in analyses and verify the model accurately reflects the organism's growth potential and efficiency. Fig. 6 shows the experimental and optimized specific growth rates for different feeding strategies, highlighting the capability of the network to predict the growth rates.

In this study, the growth behavior of *Arthrospira* sp. PCC 8005 was investigated under continuous photobioreactor conditions using different nitrogen sources. Cultures were operated in steady-state continuous mode, where the specific growth rate μ is expected to be equal to the dilution rate D ($= 0.083\ h^{-1}$) for all nitrogen regimes. For each feeding strategy, two growth profiles are compared: the experimental growth rate ($\mu_{exp} \approx$ D), and the growth rate predicted by the metabolic model assuming biomass maximization.

Across conditions, the model-predicted growth rates are higher than the experimental values, which is expected given that in vivo systems do not necessarily operate at maximal efficiency. Under nitrate-only and nitrate-nitrite conditions, the predicted growth closely matches the experimental value, suggesting that the cells are operating near their metabolic limits. This is consistent with the high energetic cost of nitrate and nitrite assimilation, which requires one or two reduction steps, respectively, prior to incorporation into biomass. The energetic burden of assimilating these oxidized nitrogen forms likely limits the cell's ability to grow faster, which is consistent with the simulated optimal growth rates close to the dilution rate.

In contrast, the model predicts higher growth under the nitrate-urea condition, as examined in [43]. Urea assimilation bypasses the need for reduction and is therefore energetically more favorable, which explains the higher growth rate obtained by performing dFBA. However, the experimental growth rate is constrained by the fixed dilution rate of the PBR. This highlights the difference between metabolic potential, as predicted by the model, and the actual performance under controlled conditions. Furthermore, it is worth noting that such differences are commonly observed in constraint-based modeling, particularly for underdetermined networks, where optimized and experimental growth rates can differ by a factor of 1.5–3 depending on the system and conditions [12]. These results, particularly the strong agreement between predicted and measured growth rates under nitrate-only and nitrate-nitrite conditions underscore the model's reliability in capturing the metabolic behavior of *Arthrospira* across diverse nitrogen sources.

Subsequently, all experimental constraints are applied, and dFVA is conducted to explore the range of possible fluxes through each reaction in the network. Fig. 7 and Fig. 8 present the dFVA results for some metabolic reactions under the 30mM-N nitrate feeding condition. Specifically, reactions v_{27} to v_{36} are related to the tricarboxylic acid cycle and reactions v_{149} to v_{164} are related to the biosynthesis of lipids. Narrow flux intervals indicate higher prediction confidence, whereas broader intervals suggest uncertainty and imply that the model could benefit from additional constraints. In particular, when a flux interval narrows to a single value, it indicates that the corresponding flux is calculable, as defined through calculability analysis.

As expected, reactions involved in biomass precursor biosynthesis (e.g., lipids, amino acids, nucleotides, and carbohydrates) exhibit minimal variability, due to their direct coupling with the fixed specific growth rate. In contrast, reactions from central carbon metabolism (glycolysis, tricarboxylic acid cycle, and pentose phosphate pathway) and cellular energy metabolism display broader flux ranges, reflecting the intrinsic metabolic flexibility in carbon routing and cofactors regeneration (e.g., ATP, NADH, NADPH) within the network constraints. These trends align with biological expectations and support the structural and functional validity of the refined network. Moreover, these results are consistent with the findings of [47], who performed a similar FVA analysis on a genome-scale metabolic network (875 reactions), and also reported narrow intervals for anabolic pathways, and broader ranges for central carbon metabolism.

FBA at a specific time instant is also performed to analyze the flux distribution of the model through the flux map. It is performed by

imposing the experimental uptake and secretion rates and the growth rate while maximizing ATP production. After careful analysis of the flux map, it is observed that the regime is autotrophic and is characterized by high fluxes in the photosynthetic pathways and the activation of the Calvin-Benson cycle, as expected. More specifically, upper glycolysis operates in the glyconeogenic direction to produce carbohydrates and sugar precursor metabolites (PEP, G6P and R5P), essential for growth. The pentose phosphate pathway (PPP) is in the reductive mode and ATP synthesis is driven by the PMF and is synthesized via photophosphorylation and oxidative phosphorylation (in addition to substrate-level phosphorylation), as discussed in Sec. 4.1.2. The simplified flux map is given in Fig. 9. Overall, the observed flux distribution not only reflects biologically plausible behavior but also aligns closely with established literature, underscoring the robustness of the model and its relevance in capturing key aspects of autotrophic metabolism.

6. Conclusion

This work proposes a constructive methodology for the identification of metabolic networks of intermediate size. The approach combines basic biological knowledge with a series of constraint-based methods in an iterative procedure, allowing progressive refinement of the network structure. A representative case study has been examined and the network structure and metabolic fluxes have been characterized in photosynthetically grown cells of *Arthrospira* spirulina platensis PCC 8005. Validation using experimental data has enabled further refinement of the network, validating its ability to accurately reflect actual cellular behavior. The predictive capabilities of the model show strong agreement with experimental observations, reinforcing the biological relevance and robustness of the approach. This investigation highlights the value of the methodology in constructing consistent and biologically meaningful networks. Moreover, metabolic analysis reveals key physiological constraints, particularly in the balancing of reducing equivalents, while also exposing the limitations inherent to in-silico simulations. Compared to existing metabolic networks for photosynthetic organisms, the proposed network offers several advantages - notably, a more rigorous mathematical modeling of the proton motive force, as well as an improved representation of lipid and carbohydrate biosynthesis pathways in networks of comparable size.

Metabolic Network of Arthrospira Sp. PCC 8005

Table 2
Metabolic reactions for the metabolism of Arthrospira spirulina platensis PCC 8005 (photoautotrophic regime). Protons H^+ involved in most metabolic reactions correspond to H_n^+.

Flux	Metabolic reaction
Carbon entry	
v_1	$HCO_3^- + H^+ \rightarrow CO_2 + H_2O$
Calvin cycle	
v_2	3 ribulose−1,5-biP + 3 CO_2 + H_2O → 6 3-P-glycerate
v_3	6 3-P-glycerate + 6 ATP ↔ 6 1,3-biphosphoglycerate + 6 ADP
v_4	6 1,3-biphosphoglycerate + 6 NADPH,H^+ ↔ 6 glyceraldehyde−3-P + 6 $NADP^+$ + 6 P_i
v_5	5 glyceraldehyde−3-P → 3 ribose−5-P + 2 P_i
Glycolysis	
v_6	glucose + ATP → glucose−6-P + ADP
v_7	glucose−6-P ↔ fructose−6-P
v_8	fructose−6-P + ATP → fructose−1,6-biP + ADP
v_9	fructose−1,6-biP ↔ glyceraldehyde−3-P + dihydroxyacetone-P
v_{10}	glyceraldehyde−3-P ↔ dihydroxyacetone-P
v_{11}	glyceraldehyde−3-P + NAD^+ + P_i + ADP ↔ 3-P-glycerate + NADH,H^+ + ATP
v_{12}	3-P-glycerate ↔ P-enolpyruvate + H_2O

(*continued on next column*)

Table 2 (*continued*)

Flux	Metabolic reaction
v_{13}	P-enolpyruvate + ADP → pyruvate + ATP
Neoglucogenesis	
v_{14}	glucose−6-P + H_2O → glucose + P_i
v_{15}	fructose−1,6-biP + H_2O → fructose−6-P + P_i
v_{16}	oxaloacetate + GTP → P-enolpyruvate + GDP + CO_2
v_{17}	pyruvate + HCO_3^- + ATP → oxaloacetate + ADP + P_i
PPP	
v_{18}	glucose−6-P + 2 $NADP^+$ + H_2O → ribose−5-P + 2 NADPH,H^+ + CO_2
v_{19}	ribose−5-P ↔ ribulose−5-P
v_{20}	ribose−5-P + ATP → ribulose−1,5-biP + ADP
v_{21}	xylulose−5-P ↔ ribulose−5-P
v_{22}	ribose−5-P + xylulose−5-P ↔ glyceraldehyde−3-P + sedoheptulose−7-P
v_{23}	glyceraldehyde−3-P + sedoheptulose−7-P ↔ fructose−6-P + erythrose−4-P
v_{24}	erythrose−4-P + xylulose−5-P ↔ fructose−6-P + glyceraldehyde−3-P
v_{25}	dihydroxyacetone-P + erythrose−4-P ↔ sedoheptulose−1,7-biP
v_{26}	sedoheptulose−1,7-biP → sedoheptulose−7-P + P_i
TCA cycle	
v_{27}	acetyl-CoA + oxaloacetate + H_2O ↔ citrate + CoASH
v_{28}	citrate ↔ H_2O + cis-aconitate
v_{29}	cis-aconitate + H_2O ↔ isocitrate
v_{30}	isocitrate + NAD^+↔α-ketoglutarate + NADH, H^+ + CO_2
v_{31}	α-ketoglutarate ↔ succinate-semialdehyde + CO_2
v_{32}	succinate-semialdehyde + NAD^+ + H_2O ↔ succinate + NADH + 2 H^+
v_{33}	succinate + GTP + CoASH ↔ succinyl-CoA + GDP + P_i
v_{34}	succinate + FAD^+↔ fumarate + $FADH_2$
v_{35}	fumarate + H_2O ↔ malate
v_{36}	malate + NAD^+↔ oxaloacetate + NADH,H^+
Anaplerotic reaction	
v_{37}	P-enolpyruvate + CO_2 + H_2O ↔ oxaloacetate + P_i
GS-GOGAT pathway	
v_{38}	α-ketoglutarate + glutamine + NADPH,H^+↔ 2 glutamate + $NADP^+$ + H_2O
v_{39}	glutamate + ATP + NH_4^+→ glutamine + ADP + P_i + H^+
v_{40}	NH_3 + H^+↔ NH_4^+
v_{41}	NH_4^+ + α-ketoglutarate + NADPH,H^+→ glutamate + $NADP^+$
v_{42}	NO_3^- + 2 Fe_{red} + 2 H^+↔ NO_2^- + 2 Fe_{ox} + H_2O
v_{43}	NO_2^- + 6 Fe_{red} + 7 H^+↔ NH_4^+ + 6 Fe_{ox} + 2 H_2O
Urea cycle-like pathway	
v_{44}	H_2O + urea → CO_2 + 2 NH_4^+
v_{45}	2 ATP + NH_4^+ + HCO_3^-→ 2 ADP + P_i + carbamyl-P
v_{46}	carbamyl-P + ornithine → citrulline + P_i
v_{47}	ATP + citrulline + aspartate → AMP + PP_i + arginosuccinate
v_{48}	arginosuccinate → fumarate + arginine
v_{49}	arginine + H_2O → ornithine + urea
S fixation	
v_{50}	ATP + SO_4^{2-} → APS + PP_i
v_{51}	APS + NADH,H^+→ AMP + NAD^+ + SO_3
v_{52}	2 H^+ + 3 NADPH,H^+ + SO_3↔ H_2S + 3 H_2O + 3 $NADP^+$
THF metabolism	
v_{53}	GTP + H_2O → formate + 7,8-dihydroneopterin−3'-triP
v_{54}	7,8-dihydroneopterin−3'-triP + H_2O → dihydroneopterin-P + PP_i + 2 H^+
v_{55}	dihydroneopterin-P + H_2O → dihydroneopterin + P_i
v_{56}	dihydroneopterin → glycoaldehyde + 6-hydroxymethyl-dihydropterin
v_{57}	6-hydroxymethyl-dihydropterin + ATP → 2-amino−4-hydroxy−6-hydroxymethyl−7,8-dihydropteridine-biP + AMP + H^+
v_{58}	glutamine + chorismate ↔ glutamate + 4-amino−4-deoxychorismate
v_{59}	4-amino−4-deoxychorismate ↔ para-aminobenzoate + pyruvate + H^+

(*continued on next page*)

Table 2 (*continued*)

Flux	Metabolic reaction
v_{60}	para-aminobenzoate + 2-amino−4-hydroxy−6-hydroxymethyl−7,8-dihydropteridine-biP → 7,8-dihydropteroate + PP_i
v_{61}	glutamate + 7,8-dihydropteroate + ATP → DHF + P_i + ADP + H^+
v_{62}	DHF + NADPH,H^+↔ THF + $NADP^+$
v_{63}	CH_2-THF + $NADP^+$↔ 5,10-CH=THF + NADPH,H^+
v_{64}	H_2O + 5,10-CH=THF ↔ CHO-THF + H^+
v_{65}	CH_2-THF + NADPH,H^+↔ CH_3-THF + $NADP^+$
v_{66}	CH_3-THF + H_2O ↔ 5-formyl-THF + H^+
v_{67}	5-formyl-THF + ATP + H_2O → ADP + H^+ + CHO-THF + P_i
v_{68}	CHO-THF + H_2O ↔ formate + H^+ + THF
v_{69}	ATP + formate + THF → ADP + CHO-THF + P_i
AA metabolism	
v_{70}	glutamate + ATP + NADPH,H^+ + NADH,H^+↔ proline + ADP + P_i + $NADP^+$ + NAD^+ + H_2O
v_{71}	glutamate + acetyl-CoA → N-acetylglutamate + CoASH
v_{72}	N-acetylglutamate + ATP + NADPH,H^+ + glutamate + H_2O → ornithine + acetate + $NADP^+$ + ADP + P_i + α-ketoglutarate
v_{73}	glutamine + 2 ATP + CO_2 + 2 H_2O → carbamyl-P + glutamate + 2 ADP + P_i
v_{74}	oxaloacetate + glutamate ↔α-ketoglutarate + aspartate
v_{75}	aspartate + glutamine + ATP → asparagine + glutamate + AMP + PP_i
v_{76}	aspartate + ATP + NADPH,H^+→ aspartate semialdehyde + ADP + $NADP^+$ + P_i
v_{77}	aspartate semialdehyde + pyruvate + succinyl-CoA + glutamate + NADPH,H^+ + H_2O ↔α-ketoglutarate + diaminopimelate + succinate + CoASH + $NADP^+$
v_{78}	diaminopimelate + H^+↔ lysine + CO_2
v_{79}	aspartate semialdehyde + NADPH,H^+↔ homoserine + $NADP^+$
v_{80}	homoserine + ATP + H_2O → threonine + ADP + P_i
v_{81}	succinyl-CoA + homoserine + cysteine ↔ homocysteine + pyruvate + NH_3 + succinate + CoASH
v_{82}	homocysteine + CH_3-THF ↔ methionine + THF
v_{83}	serine + acetyl-CoA + H_2S + H^+↔ cysteine + CoASH + acetate
v_{84}	serine → pyruvate + NH_4^+
v_{85}	threonine ↔α-ketobutyrate + NH_4^+ + H_2O
v_{86}	threonine + NAD^+ + CoASH → glycine + NADH,H^+ + acetyl-CoA
v_{87}	pyruvate + glutamate + α-ketobutyrate + NADPH,H^+↔ isoleucine + α-ketoglutarate + $NADP^+$ + CO_2 + H_2O
v_{88}	pyruvate + glutamate → alanine + α-ketoglutarate
v_{89}	2 pyruvate + NADPH,H^+↔ oxoisovalerate + $NADP^+$ + CO_2 + H_2O
v_{90}	glutamate + oxoisovalerate ↔ valine + α-ketoglutarate
v_{91}	acetyl-CoA + glutamate + oxoisovalerate + NAD^+ + H_2O ↔ leucine + α-ketoglutarate + CoASH + NADH,H^+ + CO_2
v_{92}	3-P-glycerate + glutamate + NAD^+ + H_2O ↔ serine + α-ketoglutarate + NADH,H^+ + P_i
v_{93}	serine + THF ↔ glycine + CH_2-THF + H_2O
v_{94}	2 P-enolpyruvate + erythrose−4-P + ATP + NADPH,H^+→ chorismate + ADP + $NADP^+$ + 4 P_i
v_{95}	ribose−5-P + ATP → 5'-P-ribosyl−1PP + AMP
v_{96}	5'-P-ribosyl−1PP + glutamine + serine + chorismate ↔ tryptophan + glyceraldehyde−3-P + pyruvate + glutamate + PP_i + CO_2 + 2 H_2O
v_{97}	glutamate + chorismate → phenylalanine + α-ketoglutarate + CO_2 + H_2O
v_{98}	phenylalanine + O_2 + NADH,H^+→ tyrosine + NAD^+ + H_2O
v_{99}	glutamate + chorismate + NAD^+↔ tyrosine + α-ketoglutarate + NADH,H^+ + CO_2
v_{100}	5'-P-ribosyl−1PP + glutamine + ATP + 2 NAD^+ + 3 H_2O → histidine + α-ketoglutarate + ACR + 2 NADH,H^+ + 2 PP_i + P_i

(*continued on next column*)

Table 2 (*continued*)

Flux	Metabolic reaction
v_{101}	glycine + NAD^+ + THF ↔ NH_4^+ + CH_2-THF + NADH, H^+ + CO_2
Nucleotide salvage pathway	
v_{102}	5'-P-ribosyl−1PP + carbamyl-P + aspartate + NAD^+→ UMP + NADH,H^+ + PP_i + P_i + CO_2 + H_2O
v_{103}	UMP + ATP → UDP + ADP + H^+
v_{104}	UDP + ATP ↔ UTP + ADP + H^+
v_{105}	UTP + ATP + glutamine → CTP + ADP + P_i + glutamate
v_{106}	CMP + ATP ↔ CDP + ADP + H^+
v_{107}	CDP + ATP ↔ CTP + ADP + H^+
v_{108}	5'-P-ribosyl−1PP + 2 glutamine + aspartate + glycine + 4 ATP + CHO-THF + CO_2 + 2 H_2O → ACR + fumarate + 2 glutamate + 4 ADP + THF + PP_i + 4 P_i
v_{109}	ACR + CHO-THF ↔ IMP + THF + H_2O
v_{110}	IMP + ATP + NAD^+ + glutamine + 2 H_2O → GMP + glutamate + AMP + NADH,H^+ + PP_i
v_{111}	GMP + ATP → GDP + ADP + H^+
v_{112}	GDP + ATP ↔ GTP + ADP + H^+
v_{113}	aspartate + IMP + GTP ↔ AMP + fumarate + GDP + P_i
v_{114}	UDP + 2 ATP + CH_2-THF + NADPH,H^+→ 2 ADP + dTTP + DHF + $NADP^+$ + P_i
v_{115}	dTDP + ATP → dTTP + ADP + H^+
v_{116}	CDP + ATP + NADPH,H^+→ dCTP + ADP + $NADP^+$ + H_2O
v_{117}	GDP + ATP + NADPH,H^+→ dGTP + ADP + $NADP^+$ + H_2O
v_{118}	ATP + NADPH,H^+↔ dATP + $NADP^+$ + H_2O
ATP synthesis and cofactor metabolism	
v_{119}	2 Fe_{red} + 2 H^+ + $NADP^+$→ 2 Fe_{ox} + NADPH,H^+
v_{120}	2 H_2O + 2 $NADP^+$ + 8 photons → O_2 + 2 NADPH + 12 H_p^+
v_{121}	NADH + 11 H_n^+ + 0.5 O_2→ NAD^+ + 10 H_p^+ + H_2O
v_{122}	$FADH_2$ + 6 H_n^+ + 0.5 O_2→ FAD^+ + 6 H_p^+ + H_2O
v_{123}	ADP + P_i + 4 H_p^+→ ATP + H_2O + 4 H_n^+
v_{124}	ATP + H_2O ↔ ADP + P_i + MAINT
v_{125}	ATP + AMP ↔ 2 ADP + H^+
v_{126}	NADPH + NAD^+↔ $NADP^+$ + NADH
v_{127}	PP_i + H_2O → 2 P_i
v_{128}	ATP + 2 H_2O → AMP + 2 P_i
v_{129}	NAD^+ + H^+ + 2 e^-↔ NADH
v_{130}	$NADP^+$ + H^+ + 2 e^-↔ NADPH
v_{131}	FAD^+ + 2 H^+ + 2 e^-↔ $FADH_2$
v_{132}	2 H_2O ↔ O_2 + 4 H^+ + 4 e^-
Carbohydrates synthesis	
v_{133}	glucose−6-P + UTP → UDP-glucose + PP_i
v_{134}	UDP-glucose → UDP-galactose
v_{135}	fructose−6-P + GTP → GDP-mannose + PP_i
v_{136}	GDP-mannose + NADPH,H^+→ GDP-fucose + $NADP^+$ + H_2O
v_{137}	glucose−6-P + dTTP + NADPH,H^+→ dTDP-rhamnose + PP_i + $NADP^+$ + H_2O
v_{138}	UDP-glucose + 2 NAD^+ + H_2O → UDP-glucuronate + 2 NADH,H^+
v_{139}	UDP-glucuronate ↔ UDP-galacturonate
v_{140}	UDP-glucuronate + H^+→ UDP-xylose + CO_2
v_{141}	fructose−6-P + glutamine + acetyl-CoA + UTP → UDP-N-acetylglucosamine + glutamate + CoASH + PP_i
v_{142}	UDP-N-acetylglucosamine + glutamate + 3 alanine + P-enolpyruvate + NADPH,H^+ + diaminopimelate + 4 ATP → UDP-N-acetylpentapeptide + $NADP^+$ + 4 ADP + 5 P_i
v_{143}	UDP-N-acetylglucosamine + P-enolpyruvate + ATP + CTP + 3 H_2O → CMP-N-acetylneuraminate + UDP + ADP + PP_i + 2 P_i
v_{144}	glucose−6-P + H_2O → $cyclitol_{av}$ + P_i
v_{145}	UDP-N-acetylglucosamine + UDP-N-acetylpentapeptide + 5 glycine → $peptidoglycan_{av}$ + alanine + UMP + UDP + P_i
v_{146}	glucose−6-P + ATP → ADP + PP_i + glycogen

(*continued on next page*)

Table 2 (*continued*)

Flux	Metabolic reaction
Lipids synthesis	
v_{147}	pyruvate + NAD^+ + CoASH → acetyl-CoA + CO_2 + $NADH,H^+$
v_{148}	acetyl-CoA + H_2O → acetate + CoASH + H^+
v_{149}	dihydroxyacetone-P + $NADH,H^+$↔ glycerol−3-P + NAD^+
v_{150}	acetyl-CoA + ACP → acetyl-ACP + CoASH
v_{151}	acetyl-CoA + ATP + CO_2 + H_2O → malonyl-CoA + ADP + P_i
v_{152}	acetyl-ACP + 7 malonyl-CoA + 14 $NADPH,H^+$→ palmitic-ACP + 7 CoASH + 14 $NADP^+$ + 7 CO_2 + 7 H_2O
v_{153}	malonyl-CoA + palmitic-ACP + 2 $NADPH,H^+$→ stearic-ACP + CoASH + 2 $NADP^+$ + CO_2 + H_2O
v_{154}	palmitic-ACP + $NADP^+$→ palmitoleic-ACP + $NADPH,H^+$
v_{155}	stearic-ACP + $NADH,H^+$→ oleic-ACP + NAD^+
v_{156}	oleic-ACP + $NADH,H^+$→ linoleic-ACP + NAD^+
v_{157}	linoleic-ACP + $NADH,H^+$→γ-linolenic-ACP + NAD^+
v_{158}	glycerol−3-P + 0.922 palmitic-ACP + 0.212 palmitoleic-ACP + 0.059 stearic-ACP + 0.101 oleic-ACP + 0.271 linoleic-ACP + 0.435 γ-linolenic-ACP → phosphatidic $acid_{av}$ + 2 ACP
v_{159}	phosphatidic $acid_{av}$ + H_2O → $diacylglycerol_{av}$ + P_i
v_{160}	$diacylglycerol_{av}$ + UDP-glucose → monogalactosyldiacylglycerol + UDP
v_{161}	2 monogalactosyldiacylglycerol → digalactosyldiacylglycerol + $diacylglycerol_{av}$
v_{162}	$diacylglycerol_{av}$ + UDP-sulfoquinovose → sulfoquinovosyldiacylglycerol + UDP
v_{163}	phosphatidic $acid_{av}$ + glycerol−3-P + CTP → glycerol + CMP + PP_i + P_i
v_{164}	glycerol−3-P + 0.956 palmitic-ACP + 0.573 palmitoleic-ACP + 0.299 stearic-ACP + 0.425 oleic-ACP + 0.377 linoleic-ACP + 0.370 γ-linolenic-ACP + H_2O → $triglyceride_{av}$ + 3 ACP + P_i
v_{165}	UDP-glucose + SO_3 + H^+→ H_2O + UDP-sulfoquinovose
v_{166}	acetate + CoASH + ATP → acetyl-CoA + AMP + PP_i
Chlorophyll synthesis	
v_{167}	8 glutamate + 8 ATP + 8 $NADPH,H^+$ + 2.5 O_2→ protoporphyrin + 8 ADP + 8 P_i + 21 H_2O + 8 $NADP^+$ + 6 CO_2 + 4 NH_3
v_{168}	12 acetyl-CoA + 11 $NADPH,H^+$ + 12 ATP + 4 H_2O → phytyl-PP + 12 CoASH + 11 $NADP^+$ + 12 ADP + 4 P_i + 4 CO_2 + 3 PP_i
v_{169}	methionine + ATP + H_2O → S-ad.methionine + PP_i + P_i
v_{170}	protoporphyrin + Mg^{2+} + S-ad.methionine + 5 $NADPH,H^+$ + 3 O_2 + phytyl-PP + ATP → chlorophyll + 3 H^+ + S-ad.homocysteine + 5 $NADP^+$ + 4 H_2O + PP_i + ADP + P_i
v_{171}	S-ad.homocysteine + ATP + H_2O → homocysteine + AMP + ADP
Nitrogen stockage	
v_{172}	aspartate + arginine → cyanophycin + H_2O
Synthesis of biomass macromolecules	
v_{173}	0.732 dTDP-rhamnose + 0.124 UDP-glucose + 0.051 $cyclitol_{av}$ + 0.030 $peptidoglycan_{av}$ + 0.043 glycogen + 0.022 CMP-N-acetylneuraminate + 5 ATP → $carbohydrate_{av}$ + 5 ADP + 5 P_i + 0.732 dTDP + 0.124 UDP + 0.022 CMP
v_{174}	0.058 isoleucine + 0.094 leucine + 0.036 lysine + 0.018 methionine + 0.039 phenylalanine + 0.054 threonine + 0.009 tryptophan + 0.076 valine + 0.053 arginine + 0.014 histidine + 0.101 alanine + 0.098 aspartate + 0.132 glutamate + 0.007 cysteine + 0.086 glycine + 0.038 proline + 0.054 serine + 0.033 tyrosine + 2 GTP + 55 ATP + 2 H_2O → $protein_{av}$ + 2 GDP + 55 ADP + PP_i + 55 P_i
v_{175}	0.320 monogalactosyldiacylglycerol + 0.131 digalactosyldiacylglycerol + 0.147 sulfoquinovosyldiacylglycerol + 0.216 glycerol + 0.186 $triglyceride_{av}$ + ATP → $lipid_{av}$ + ADP + P_i

(*continued on next column*)

Table 2 (*continued*)

Flux	Metabolic reaction
v_{176}	0.262 ATP + 0.322 GTP + 0.216 UTP + 0.200 CTP + 2 ATP → RNA_{av} + PP_i + 2 ADP + 2 P_i
v_{177}	0.279 dTTP + 0.222 dCTP + 0.222 dGTP + 0.279 dATP + 5 ATP → DNA_{av} + PP_i + 5 ADP + 5 P_i
v_{178}	0.8404 $protein_{av}$ + 0.1230 $carbohydrate_{av}$ + 0.0182 $lipid_{av}$ + 0.0130 RNA_{av} + 0.0039 DNA_{av} + 0.0016 chlorophyll + 97 ATP → $biomass_{av}$ + 97 ADP + 97 P_i
Transport reactions	
v_{179}	HCO_3^- *ext*↔ HCO_3^-
v_{180}	NO_2^- *ext*↔ NO_2^-
v_{181}	photon $_{ext}$→ photon
v_{182}	CO_2 $_{ext}$↔ CO_2
v_{183}	O_2 $_{ext}$↔ O_2
v_{184}	H_2O_{ext}↔ H_2O
v_{185}	P_{iext}↔ P_i
v_{186}	NO_3^- *ext*↔ NO_3^-
v_{187}	SO_4^{2-} *ext*↔ SO_4^{2-}
v_{188}	NH_4^+ *ext*↔ NH_4^+
v_{189}	Mg^{2+} *ext*↔ Mg^{2+}
v_{190}	H^+ *ext*↔ H^+
v_{191}	urea $_{ext}$↔ urea
v_{192}	glucose $_{ext}$→ glucose
v_{193}	chlorophyll → chlorophyll $_{ext}$
v_{194}	cyanophycin → cyanophycin $_{ext}$
v_{195}	carbohydrate ↔ carbohydrate $_{ext}$
v_{196}	lipid ↔ lipid $_{ext}$
v_{197}	biomass → biomass $_{ext}$
v_{198}	MAINT → MAINT $_{ext}$

CRediT authorship contribution statement

B. Leroy: Data curation. **A. Vande Wouwer:** Writing – review & editing, Validation, Supervision. **M. Maton:** Writing – original draft, Software, Methodology, Formal analysis.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

References

[1] M. Covert, C. Schilling, I. Famili, G.I. Edwards, J.S.E. Selkov, B. Palsson, Metabolic modeling of microbial strains in silico, Trends Biochem. Sci. 26 (2001) 179–186.

[2] N. Price, J. Papin, C. Schilling, B. Palsson, Genome-scale microbial in silico models: the constraints-based approach, Trends Biotechnol. 21 (2003) 162–169.

[3] J. Reed, B. Palsson, Thirteen years of building constraint-based in silico models of Escherichia coli, J. Bacteriol. 185 (2003) 2692–2699.

[4] B. Kholodenko, S. Schuster, J. Rohwer, M. Cascante, H. Westerhoff, Composite control of cell function: Metabolic pathways behaving as single control units, FEBS Lett. 368 (1995) 1–4.

[5] J. Rohwer, S. Schuster, H. Westerhoff, How to recognize monofunctional units in a metabolic system, J. Theor. Biol. 179 (1996) 213–228.

[6] T. Pfeiffer, I. Sanchez-Valdenebro, J. Nuno, F. Montero, S. Schuster, Metatool: For studying metabolic networks, Bioinformatics 15 (1999) 251–257.

[7] S. Klamt, J. Stelling, M. Ginkel, E. Gilles, Fluxanalyzer: exploring structure, pathways, and flux distributions in metabolic networks on interactive flux maps, Bioinformatics 19 (2003) 261–269.

[8] I. Thiele, B. Palsson, A protocol for generating a high-quality genome-scale metabolic reconstruction, Nat. Protoc. 5 (1) (2010) 93–121.

[9] A. Montagud, E. Navarro, P. Fernandez de Cordoba, J. Urchueguia, K. Patil, Reconstruction and analysis of genome-scale metabolic model of a photosynthetic bacterium, BMC Syst. Biol. 4 (156) (2010) 1–16.

[10] M. Wu, C. Chan, Human metabolic network: reconstruction, simulation, and applications in systems biology, Metabolites 2 (2012) 242–253.

[11] D. Machado, S. Andrejev, M. Tramontano, K. Patil, Fast automated reconstruction of for microbial species and communities, Nucleic Acids Res. 46 (15) (2018) 7542–7553.

[12] J. Orth, I. Thiele, B. Palsson, What is flux balance analysis, Comput. Biol. 3 (2010) 245–248.
[13] C. Schilling, D. Letscher, B. Palsson, Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective, J. Theor. Biol. 203 (2000) 229–248.
[14] S. Schuster, C. Hilgetag, On elementary flux modes in biochemical reaction systems at steady-state, J. Biol. Syst. 2 (1994) 165–182.
[15] S. Schuster, D. Fell, T. Dandekar, A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks, Nat. Biotechnol. 18 (2000) 326–332.
[16] H. Oddsdottir, E. Hagrot, V. Chotteau, A. Forsgren, Robustness analysis of elementary flux modes generated by column generation, Math. Biosci. 273 (2016) 45–56.
[17] M. Maton, P. Bogaerts, A. Vande Wouwer, A systematic elementary flux mode selection procedure for deriving macroscopic bioreaction models from metabolic networks, J. Process Control 118 (2022) 170–174.
[18] A. Burgard, E. Nikolaev, C. Schilling, C. Maranas, Flux coupling analysis of genome-scale metabolic network reconstructions, Genome Res. 14 (2) (2004) 301–312.
[19] A. Larhlimi, L. David, J. Selbig, A. Bockmayr, F2c2: a fast tool for the computation of flux coupling in genome-scale metabolic networks, BMC Bioinforma. 13 (2012) 57.
[20] P. Erdrich, R. Steur, S. Klamt, An algorithm for the reduction of genome-scale metabolic network models to meaningful core models, Cell. Mol. Life Sci. 9 (48) (2015) 1–12.
[21] A. Provost, G. Bastin, Dynamic metabolic modelling under the balanced growth condition, J. Process Control 14 (7) (2004) 717–728.
[22] G. Bastin, Quantitative analysis of metabolic networks and design of minimal bioreaction models, ARIMA, Rev. Afr. De. la Rech. En. Inform. Et. Math. ématiques Appliquées 9 (1) (2008) 41–55.
[23] F. Zamorano, A. Vande Wouwer, G. Bastin, A detailed metabolic flux analysis of an underdetermined network of cho cells, J. Biotechnol. 150 (2010) 497–508.
[24] F. Zamorano, A. Vande Wouwer, R. Jungers, G. Bastin, Dynamic metabolic models of cho cell cultures through minimal sets of elementary modes, J. Biotechnol. 164 (3) (2013) 409–422.
[25] S. Fernandes de Sousa, J. Robitaille, G. Bastin, M. Jolicoeur, A. Vande Wouwer, Application of dynamic metabolic flux convex analysis to cho-dxb11 cell fed-batch cultures, IFAC-PapersOnLine 49 (7) (2016) 466–471.
[26] S. Fernandes de Sousa, G. Bastin, M. Jolicoeur, A. Vande Wouwer, Dynamic metabolic flux analysis using a convex analysis approach: Application to hybridoma cell cultures in perfusion, Biotechnol. Bioeng. 113 (5) (2016) 1102–1112.
[27] T. Cakir, M. Khatibipour, Metabolic network discovery by top-down and bottom-up approaches and paths for reconciliation, Front. Bioeng. Biotechnol. 2 (2014) 00062.
[28] O. Hadicke, S. Klamt, Ecolicore2: a reference network model of the central metabolism of Escherichia coli and relationships to its genome-scale parent model, Nat. Sci. Rep. 7 (2017) 39647.
[29] D. Singh, M. Lercher, Network reduction methods for genome-scale metabolic models, Cell. Mol. Life Sci. 77 (2020) 481–488.
[30] G. Sambamoorthy, K. Raman, Minreact: a systematic approach for identifying minimal metabolic networks, Bioinformatics 36 (15) (2020) 4309–4315.
[31] N. Cocco, M. Llabrés, M. Reyes-Prieto, M. Simeoni, Metnet: a two-level approach to reconstructing and comparing metabolic networks, PLoS ONE 16 (2) (2021) e0246962.
[32] I. Koo, X. Wei, X. Shi, Z. Zhou, S. Kim, X. Zhang, Constructing metabolic association networks using high-dimensional mass spectroscopy data, Chemom. Intell. Lab. Syst. 138 (2014) 193–202.
[33] D. Petrovsky, K. Malsagova, V. Rudnev, L. Kulikova, V. Pustovoyt, E. Balakin, K. Yurku, A. Kaysheva, Bioinformatics methods for constructing metabolic networks, Processes 11 (2023) 3430.
[34] G. Jansen, T. Qi, V. Latora, G. Amoutzias, D. Delneri, S. Oliver, G. Nicosia, Minimisation of metabolic networks defines a new functional class of genes, Nat. Commun. 15 (2024) 9076.
[35] E. Martinez, Lumping of components and reactions in complex reaction networks, Chem. Eng. Commun. 93 (1) (1990) 1–24.
[36] P. Pepiot, L. Cai, H. Pitsch, Model reduction and lumping procedures, Computer Aided, Chem. Eng. 45 (2019) 799–827.
[37] R. Lui, J. Yang, Y. Zhao, L. Cai, An automated chemical lumping method of functional isomers based on the reaction class concept for the reduction of large chemical kinetic mechanisms, Fuel 371 (2024) 132033.
[38] K. Kauffman, P. Prakash, J. Edwards, Advances in flux balances analysis, Curr. Opin. Biotechnol. 14 (5) (2003) 491–496.
[39] K. Raman, N. Chandra, Flux balance analysis of biological systems: applications and challenges, Brief. Bioinforma. 10 (4) (2009) 435–449.
[40] L. David, S. Marashi, A. Larhlimi, B. Mieth, A. Bockmayr, Ffca: a feasibility-based method for flux coupling analysis of metabolic networks, BMC Bioinforma. 12 (2011) 236.
[41] M. Yasemi, M. Jolicoeur, Modelling cell metabolism: a review on constraint-based steady-state and kinetic approaches, Processes 9 (2) (2021) 322.
[42] L. Hendrickx, H. De Wever, V. Hermans, F. Mastroleo, N. Morin, A. Wilmotte, P. Janssen, M. Mergeay, Microbial ecology of the closed artificial ecosystem melissa: reinventing and compartmentalizing the earth's food and oxygen regeneration system for long-haul space exploration missions, Res. Microbiol. 157 (2006) 77–86.
[43] F. Deschoenmaeker, R. Facchini, B. Leroy, H. Badri, C. Zhang, R. Wattiez, Proteomic and cellular views of arthrospira sp. pcc 8005 adaptation to nitrogen depletion, Microbiology 160 (2014) 1224–1236.
[44] F. Deschoenmaeker, G. Bayon-Vicente, N. Sachdeva, O. Depraetere, J. Cabrera, B. Leroy, K. Muylaert, R. Wattiez, Impact of different nitrogen sources on the growth of arthrospira sp. pcc 8005 under batch and continuous cultivation - a biochemical, transcriptomic and proteomic profile, Bioresour. Technol. 237 (2017) 78–88.
[45] A. Meechai, S. Pongakarakun, P. Deshnium, S. Cheevadhanarak, S. Bhumiratana, Metabolic flux distribution for gamma-linolenic acid synthetic pathways in spirulina platensis, Biotechnol. Bioprocess Eng. 9 (2004) 506–513.
[46] G. Cogne, J. Gros, C. Dussap, Identification of a metabolic network strcture representative of arthrospira (spirulina) platensis metabolism, Biotechnol. Bioeng. 84 (2003) 667–676.
[47] A. Klanchui, C. Khannapho, A. Phodee, S. Cheevadhanarak, A. Meechai, iak692: a genome-scale metabolic model of spirulina platensis c1, BMC Syst. Biol. 6 (71) (2012), https://doi.org/10.1186/1752-0509-6-71.
[48] K. Yoshikawa, S. Aikawa, Y. Kojima, Y. Toya, C. Furusawa, A. Kondo, H. Shimizu, Construction of a genome-scale metabolic model of arthrospira platensis nies-39 and metabolic design for cyanobacterial bioproduction, PLoS ONE 10 (12) (2015) e0144430, https://doi.org/10.1371/journal.pone.0144430.
[49] A. Klanchui, S. Dulsawat, K. Chaloemngam, S. Cheevadhanarak, P. Prommeenate, A. Meechai, An improved genome-scale metabolic model of arthrospira platensis c1 (iak888) and its application in glycogen overproduction, Metabolites 8 (4) (2018), https://doi.org/10.3390/metabo8040084. PMID: 30486288; PMCID: PMC6315860.
[50] K. Rawls, B. Dougherty, E. Blais, E. Stancliffe, G. Kolling, K. Vinnakota, V. Pannala, A. Wallqvist, J. Papin, A simplified metabolic network reconstruction to promote understanding and development of flux balance analysis tools, Comput. Biol. Med. 105 (2019) 64–71.
[51] C. Henry, L. Broadbelt, V. Hatzimanikatis, Thermodynamics-based metabolic flux analysis, Biophys. J. 92 (2007) 1792–1805.
[52] J. Schellenberger, N. Lewis, B. Palsson, Elimination of thermodynamically infeasible loops in steady-state metabolic models, Biophys. J. 100 (2011) 544–553.
[53] S. Peres, V. Fromion, Thermodynamic Approaches in Flux Analysis, Springer, US, New York, NY, 2020, pp. 359–367.
[54] M. Gollub, H.-M. Kaltenbach, J. Stelling, Probabilistic thermodynamic analysis of metabolic networks, Bioinformatics 37 (18) (2021) 2938–2945.
[55] C. Tomi-Andrino, R. Norman, T. Millat, P. Soucaille, K. Winzer, D. Barrett, J. King, D.-H. Kim, Physicochemical and metabolic constraints for thermodynamics-based stoichiometric modelling under mesophilic growth conditions, PLoS Comput. Biol. 17 (1) (2021) e1007694.
[56] A. Chang, L. Jeske, S. Ulbrich, J. Hofmann, J. Koblitz, I. Schomburg, M. Neumann-Schaal, D. Schomburg, Brenda: the elixir core data resource in 2021: new developments and updates, Nucleic Acids Res. 49 (2021) D498–D508, https://doi.org/10.1093/nar/gkaa1025. PubMed: 33211880.
[57] M. Kanehisa, S. Goto, Kegg: kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 28 (1) (2000) 27–30, https://doi.org/10.1093/nar/28.1.27. PMID: 10592173; PMCID: PMC102409.
[58] R. Mahadevan, C. Schilling, The effects of alternate optimal solutions in constraint-based genome-scale metabolic models, Metab. Eng. 5 (4) (2003) 264–276.
[59] C. Baroukh, R. Muñoz-Tamayo, J. Steyer, O. Bernard, Drum: a new framework for metabolic modeling under non-balanced growth. application to the carbon metabolism of unicellular microalgae, PLoS One 9 (8) (2014) e104499.
[60] C. Baroukh, V. Turon, O. Bernard, Dynamic metabolic modeling of heterotrophic and mixotrophic microalgal growth on fermentative wastes, PLoS Comput. Biol. 13 (6) (2017) e1005590.
[61] B. AssisPessi, C. Baroukh, A. Bacquet, O. Bernard, A universal dynamical metabolic model representing mixotrophic growth of chlorella sp. on wastes, Water Res. 229 (2023) 119388.
[62] D. Nelson, M. Cox Lehninger. Principles of Biochemistry, 7th Edition, W.H. Freeman and Company, New York, 2008.
[63] A. Funahashi, N. Tanimura, M. Morohashi, H. Kitano, Celldesigner: a process diagram editor for gene-regulatory and biochemical networks (Biosilico) 1 (2003) 159–162.
[64] C. Baroukh, R. Muñoz-Tamayo, J. Steyer, O. Bernard, Mathematical modeling of unicellular microalgae and cyanobacteria metabolism for biofuel production, Curr. Opin. Biotechnol. 33 (2015) 198–205.
[65] S. Klamt, S. Schuster, E. Gilles, Calculability analysis in underdetermined metabolic networks illustrated by a model of the central metabolism in purple nonsulfur bacteria, Biotechnol. Bioeng. 77 (7) (2002) 734–751.