

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Metabolic Modelling of the Archaeon *Nitrososphaera viennensis*

verfasst von | submitted by

Irfan Suljagic BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2025

Studienkennzahl lt. Studienblatt |
Degree programme code as it appears on
the student record sheet:

UA 066 875

Studienrichtung lt. Studienblatt | Degree
programme as it appears on the student re-
cord sheet:

Masterstudium Bioinformatik

Betreut von | Supervisor:

Ass.-Prof. DI Dr. -Ing. Steffen Waldherr

1. Abstract

The main goal of this Master's thesis was to apply various approaches to model the metabolism of *Nitrososphaera viennensis*, an archaeon capable of oxidizing ammonia to nitrite. The starting point was a draft metabolic model developed by Rustem Musaev. The first part of the project focused on improving this model, which primarily involved balancing biochemical reactions within metabolic pathways and assigning appropriate gene annotations to enhance the model's consistency and reduce ambiguity.

Subsequent steps included performing Flux Balance Analysis (FBA) and, ultimately, Resource Balance Analysis (RBA) to enable a more accurate comparison of metabolite fluxes. This comparison aimed to provide clearer insights into the metabolic characteristics of *N. viennensis*. To facilitate the construction of the RBA model, several modifications were made to the RBAPy software-originally developed by Ana Bulović and colleagues-as well as to the Systems Biology Markup Language (SBML/XML) model files. These changes included adjustments to naming conventions and the addition of reactions necessary for modeling translation processes. With both FBA and RBA models in place, a comparative analysis was conducted, highlighting the similarities and differences between the two approaches. Additionally, a sensitivity analysis was performed as part of the RBA modeling process. This analysis underscored the importance of enzyme-related parameters, particularly catalytic constants, in influencing metabolic predictions and pointed to future directions for improving the accuracy of flux simulations.

While the project successfully produced meaningful and interpretable results, some aspects-such as metabolite flows in key metabolic pathways and the availability of enzyme kinetic data-remain to be clarified. These areas represent opportunities for future work and are essential for achieving a more accurate and comprehensive representation of the metabolic functioning and environmental interactions of *Nitrososphaera viennensis*.

1. Zusammenfassung

Das Hauptziel dieser Masterarbeit war die Anwendung verschiedener Ansätze zur Modellierung des Metabolismus von *Nitrososphaera viennensis*, einem Archaeon, das in der Lage ist, Ammoniak zu Nitrit zu oxidieren. Ausgangspunkt war das metabolische Modell, das von Rustem Musaev entwickelt wurde. Der erste Teil des Projekts konzentrierte sich auf die Verbesserung dieses Modells, was in erster Linie das Ausbalancieren biochemischer Reaktionen innerhalb von Stoffwechselwegen sowie die Zuweisung geeigneter Genannotationen umfasste, um die Konsistenz des Modells zu erhöhen und Unklarheiten zu reduzieren.

In den folgenden Schritten wurden eine Flux Balance Analysis (FBA) sowie abschließend eine Resource Balance Analysis (RBA) durchgeführt, um einen genaueren Vergleich der Metabolitflüsse zu ermöglichen. Ziel dieses Vergleichs war es, ein klareres Verständnis der metabolischen Eigenschaften von *N. viennensis* zu gewinnen. Zur Erstellung des RBA-Modells wurden mehrere Anpassungen an der ursprünglich von Ana Bulović und Kolleg:innen entwickelten Software RBAPy sowie an den Systems Biology Markup Language (SBML/XML)-Modelldateien vorgenommen. Diese Änderungen umfassten unter anderem die Anpassung von Benennungen und die Ergänzung von Reaktionen, die für die Modellierung von Translationsprozessen erforderlich sind. Mit den vorliegenden FBA- und RBA-Modellen wurde eine vergleichende Analyse durchgeführt, die die Gemeinsamkeiten und Unterschiede der beiden Ansätze herausstellte. Zusätzlich wurde im Rahmen der RBA-Modellierung eine Sensitivitätsanalyse durchgeführt. Diese hob die Bedeutung enzymbezogener Parameter, insbesondere katalytischer Konstanten, für die Vorhersage des Metabolismus hervor und zeigte mögliche zukünftige Verbesserungsrichtungen für die Genauigkeit der Flusssimulationen auf.

Obwohl das Projekt aussagekräftige und interpretierbare Ergebnisse lieferte, bleiben einige Aspekte wie der Metabolitfluss in zentralen Stoffwechselwegen und die Verfügbarkeit kinetischer Enzymdaten noch unklar. Diese Bereiche bieten Ansätze für zukünftige Arbeiten und sind entscheidend für eine genauere und umfassendere Darstellung der metabolischen Funktionen und Umweltinteraktionen von *Nitrososphaera viennensis*.

Table of Contents

1. Abstract	4
1. Zusammenfassung.....	6
2. Introduction.....	9
2.1. Background, Main Goals and Problem Description.....	9
2.2. <i>Nitrososphaera viennensis</i> and its Role in the Nitrogen Cycle.....	12
2.3. Principles and Paradigms of Constraint-based Modelling	14
3. Flux Balance Analysis (FBA) metabolic Model of <i>Nitrososphaera viennensis</i>	16
3.1. Methods.....	16
3.2. Results	18
4. Resource Balance Analysis (RBA) metabolic Model of <i>Nitrososphaera viennensis</i>	30
4.1. Methods.....	30
4.2. Results	32
4.3. Sensitivity Analysis of Enzyme catalytic Efficiency	36
5. Comparative Analysis between FBA and RBA metabolic model of <i>Nitrososphaera viennensis</i>	40
6. Discussion	45
7. List of References	47

2. Introduction

2.1. Background, Main Goals and Problem Description

Nitrososphaera viennensis belongs to a group of ammonia-oxidizing archaea whose metabolic network enables the conversion of ammonia into nitrite, making them key contributors to the global nitrogen cycle (see Lehtovirta-Morley (2018) and references therein). This organism is primarily found in soil environments and is one of the most studied representatives of this archaeal group. Given its ecological relevance, constructing an accurate model of its underlying metabolic processes-particularly those involved in ammonia oxidation-would be highly beneficial.

In contrast to its phylogenetic relative *Nitrosopumilus maritimus*, which inhabits marine environments and for which a validated flux balance analysis (FBA) model already exists (see Li *et al.* (2018) and references therein), *N. viennensis* still lacks a well-established genome-scale metabolic model. Although some modeling efforts have been made, such as the Master's thesis of Rustem Musaev, further work is needed to refine and expand these early attempts. This thesis therefore builds upon Musaev's work, aiming to improve the accuracy and consistency of the metabolic network and offer deeper insight into the organism's metabolism.

The primary goals of this project were to validate and improve the previous model, particularly by focusing on the correctness of biochemical reactions and gene annotations. Musaev's draft model included 433 genes and 782 reactions, assigned to either extracellular or intracellular compartments. Using Flux Balance Analysis (FBA)-a constraint-based mathematical approach used to estimate the most feasible distribution of metabolic fluxes under a given set of conditions while optimizing an objective function (typically biomass production; see Orth *et al.* (2010) and references therein)-several inconsistencies were identified.

In particular, a number of active reactions, i.e. reactions carrying flux, were found to be chemically unbalanced, which undermines the reliability of simulation results. Reducing such inconsistencies is essential to produce a more accurate and interpretable metabolic model.

Additionally, because the model includes genes encoding enzymes that catalyze metabolic reactions, it was considered essential to add proper gene annotations using publicly available databases such as UniProt. This ensures that the model components are universally recognizable and facilitates further comparative and functional analyses of the *N. viennensis* metabolic network. A more detailed description of these initial steps is provided in the subchapters that follow.

FBA-based models are inherently limited by the relatively small number of parameters they incorporate. While useful for estimating optimal metabolic states under defined conditions, they often fail to capture the complexity of biological systems operating under dynamic environmental and resource constraints. Typically, FBA models aim to maximize flux through a biomass reaction—a pseudo-reaction that reflects organismal growth but does not account for enzyme usage or resource allocation (see De Becker *et al.* (2022) and references therein). If one would want a more complex model and therefore a more realistic picture of organism's inner mechanisms, a more detailed model that relies on resource allocation would be beneficial.

Resource Balance Analysis (RBA) is one such approach. RBA models increase both the informational input and computational complexity by simulating how precursors, enzymes, and other cellular components are allocated and balanced within the cell. These models incorporate processes such as transcription, translation, protein synthesis and degradation, and enzymatic turnover rates to simulate more biologically realistic scenarios (see Giordano *et al.* (2016) and references therein). In recent years, there have been developed automation tools and packages that facilitate the creation of such models by prompting the user to provide relatively small amount of data/information (see Bulović *et al.* (2019) and references therein).

For this thesis, the RBAPy software package—developed by Ana Bulović and colleagues—was used to construct an RBA model of *N. viennensis*.

This required adapting the original Systems Biology Markup Language (SBML/XML) model by adding 20 translation-related reactions, the corresponding genes, and modifying gene-protein-reaction (GPR) rule formats to meet the input requirements of RBAPy. Further details on these modifications are provided in subsequent sections.

This part of the project also allowed the identification of sensitive parameters that strongly influence simulation outcomes, especially catalytic constants (k_{cat} values) of enzymes. These were shown to significantly affect predicted flux distributions and growth rates in the model.

Once both the improved FBA and the newly created RBA models were established, a comparative analysis was conducted to identify similarities and differences in metabolic behavior under equivalent conditions. This comparison forms a central part of the thesis and provides critical insights into the strengths and limitations of each modeling framework. Comparing growth predictions, flux distributions, sets of active reactions, and enzymes involved-along with available experimental expression data-provides a deeper understanding of the metabolism of *N. viennensis* and highlights areas for future refinement.

2.2. *Nitrososphaera viennensis* and its Role in the Nitrogen Cycle

Nitrogen cycle is a biogeochemical cycle whose initial steps include the fixation of nitrogen gas (N_2) from the atmosphere primarily by prokaryotes (nitrogen fixing bacteria), and its conversion to biologically usable ammonia (NH_3 ; see Bernhard (2010) and references therein). In the step that follows, ammonia-oxidizing bacteria (AOB) and ammonia-oxidizing archaea (AMA) convert this biologically available form of nitrogen to nitrite (NO_2) in the process called nitrification, which makes them a crucial link in the nitrogen cycle chain. The produced nitrite can ultimately be converted back to atmospheric nitrogen through a series of steps, some of which include conversion to nitrate (NO_3) with the same type of organisms as agents and denitrification of nitrate with the help of denitrifying bacteria.

Nitrososphaera viennensis belongs to AMA and therefore exhibits some of the features in common to this whole group of organisms. When it comes to AOA in general, although the first step of their ammonia oxidation process is relatively well-explained with a non-ambiguous prediction of the ammonia monooxygenase (AMO) with its 3 main copper-containing domains (AmoA, AmoB and AmoC) as the enzyme responsible for the conversion of ammonia to intermediate product hydroxylamine (NH_2OH), the next step which includes the oxidation of hydroxylamine to nitrite still leaves place for interpretation. There are several mechanisms, i.e. enzymes proposed that can play an important role in this part of the process such as F420 (cofactor proposed to have an active part in detoxification of produced reactive nitrogen species), multicopper oxidase (MCO) and copper-containing nitrite reductase (NirK) even though to this day no conclusive discoveries have been obtained (see Kerou *et al.* (2016) and references therein). Furthermore, some research has shown that even though ammonia oxidation is hindered when Cu^{2+} concentrations in the soil are low, *N. viennensis* is still more tolerable to low copper content than its close phylogenetic relative *N. maritimus* because it possesses more Cu transporters and genes encoding for Cu proteins (see Lehtovirta-Morley (2018) and references therein).

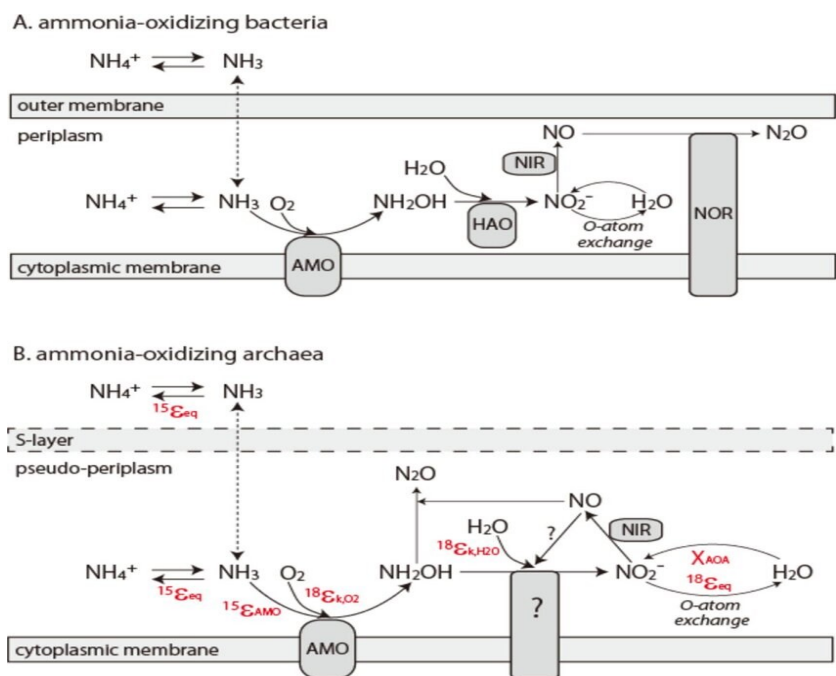


Figure 1. Differences in ammonia oxidation between AOA and AOB highlighting the fact that exact enzyme responsible for oxidation of NH_2OH to NO_2^- in AOA is still to be conclusively identified

(Source: Nishizawa *et al.* (2016))

Besides the above-described metabolic pathway, one other important machinery in the metabolism of *N. viennensis* represents the 3-hydroxypropionate/4-hydroxybutyrate cycle (HP/HB cycle), a carbon fixation cycle thought to exist only in archaea, although it has recently been discovered in the bacterial genus *Luminiphilus* as well (see Garritano *et al.* (2022) and references therein). In this process, carbon dioxide (CO_2) is fixed from the environment in the process catalyzed by the enzyme γ -class carbonic anhydrase (CA) and through a series of steps converted to acetyl-CoA which is used both to regenerate the cycle and as a compound that enters the Cytric Acid Cycle (TCA cycle), along with other intermediate products of HP/HB cycle like succinyl-CoA.

In the central carbon metabolism, sugars are broken down in order to, *inter alia*, make metabolic precursors and/or glucose in the Gluconeogenesis.

Although not really relevant for this thesis, there have also been some other discoveries in the metabolic pathways worth mentioning. Namely, *N. viennensis* showed to have glyoxylase/bleomycin resistance dyoxygenase family enzymes, whose activity is involved in the degradation of aromatic compounds and therefore the organism's resistance to drugs.

Moreover, there is evidence that the genome of the archaeon in question encodes for enzyme families like glycosyl transferase (GT), carbohydrate esterase (CE4) and E-type oligosaccharyl transferase AglB, all of which contribute to the cell surface modifications which are necessary for biofilm formations and adhesions (see Kerou *et al.* (2016) and references therein).

2.3. Principles and Paradigms of Constraint-based Modelling

Both Flux balance analysis (FBA) and Resource balance analysis (RBA) models belong to the group of constraint-based modelling approaches, meaning that they follow principles of steady-state condition and flux bounding constraints, instead of relying on more complex kinetic parameters. A system is said to be in a steady state when the velocities of the active reactions, i.e. reactions that produce or consume particular metabolites balance each other out. This condition can be represented by the equation:

$$Sv = 0$$

where S is the stoichiometric matrix representing the participation of specific metabolites in specific reactions, and v is the rate of reactions taking place.

Moreover, since both types of models rely on the two conditions mentioned above, they both use linear programming approach as a problem-solving method, which helps to find the optimal set of flux distributions while maximizing the chosen objective function.

This principle, besides needing an objective function to be optimized, needs linear constraints, that will restrict otherwise infinite solution space, and provide a biologically-meaningful metabolic phenotype/distribution of fluxes, or in other words, a solution to the linear problem. FBA in particular solves the linear problem/set of linear differential equations obtained by satisfying the steady-state condition by, in most cases, maximizing biomass production (see Sahu *et al.* (2021) and references therein). This biomass reaction can come in two different types, based on the building blocks for its composition.

Usually, if it consists of precursors or metabolites produced by elementary metabolic reactions, its size can get relatively big (several dozens of components), compared to the one composed mainly of macromolecules, whose size tends to be smaller and consists of 10-20 components (see von Kamp and Klamt (2023) and references therein).

Unlike FBA, RBA models follow the rules of resource allocation modelling and thus take matters like allocation of cellular resources and costs of protein synthesis into account. Although mainly using linear programming for solving the problem of finding the optimal metabolic phenotype, because of their complexity, they may also rely on other paradigms like quadratic optimization. Moreover, these types of models don't assume a predetermined biomass composition that lacks non-catalytic components, but rather treat this as a variable whose composition includes both catalytic and non-catalytic components (see De Becker *et al.* (2022) and references therein).

Besides this different consideration of biomass, RBA models also take other cellular factors like catalytic enzymatic activity, translation and transport of proteins, and other means of cellular production into account, which makes the inclusion of concentrations of proteins and RNA, along with some other components whose concentration reflects biological processes responsible for their production or degradation (mRNA), a crucial step in model parameterization (see Bodeit *et al.* (2023) and references therein).

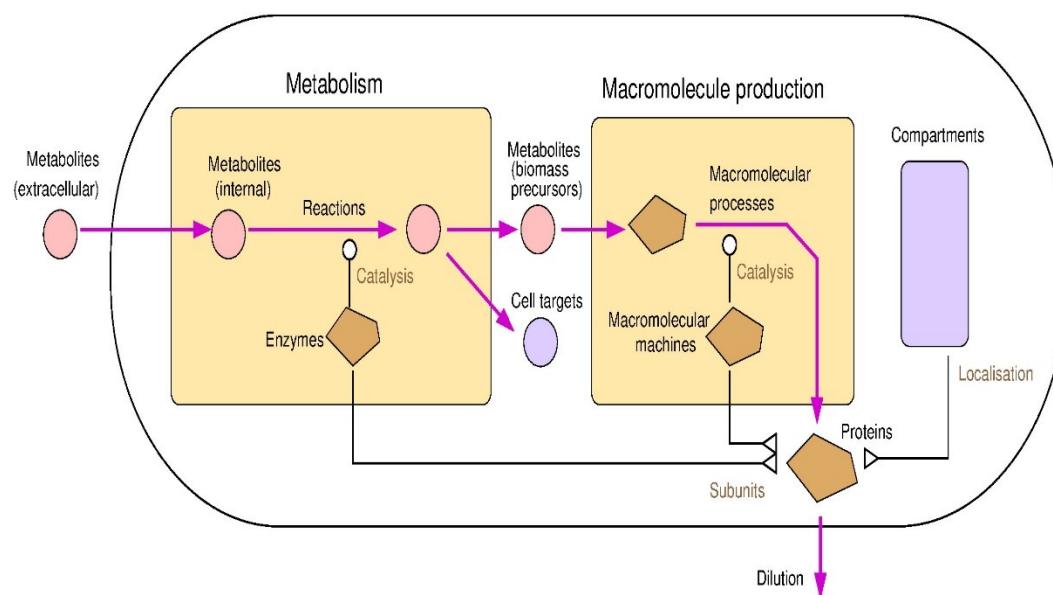


Figure 2. A schematic representation of the composition of RBA models

(Source: Bodeit *et al.* (2023))

3. Flux Balance Analysis (FBA) metabolic Model of *Nitrososphaera viennensis*

3.1. Methods

As already mentioned in the introduction, one of the first goals of this Master's thesis was to validate and improve the reactions in the model created by Rustem Musaev. The metabolism-describing model was stored in XML (Extendible Markup Language) file, and COBRApy (version 0.29.1) installed in Python (version 3.10.18) was used to read it. Additionally, a medium file was also necessary, since it contains information about the concentrations of compounds in the environment, which provide the source of nutrients/energy, and based on uptake of which the metabolism of *Nitrososphaera viennensis* was to be modelled. The mentioned file, stored in JSON (JavaScript Object Notation) format and loaded with json module (version 2.0.9), contained, among others, compounds such as: ammonia, sulfate, nitrite, phosphate, water; and ions like: calcium, zinc, magnesium, iron,...

For the optimization process, which produced the value of objective function along the flux distributions along metabolic reactions, GLPK (GNU Linear Programming Kit) (version 5.0) was used.

When it comes to the balancing of reactions that contained some amount of flux in the mentioned starting model, online databases such as ModelSEED, BioCyc, KEGG and MetaCyc were used as a reference and the reactions in questions were altered/corrected either by adding new metabolites to the equations, changing existing metabolite stoichiometries, or changing the formulas and charges of certain metabolites.

Once this step was completed successfully, the next goal was to validate unbalanced reactions that didn't contain any flux in the original model. The idea behind this was to remove unfavorable reactions from the model completely, and thereby increase its accuracy and consistency. In order to accomplish this goal, the gene-protein-reaction (GPR) rules, which represent the association between genes and enzymes/proteins they encode, of the mentioned reactions were taken into consideration (see Di Filippo *et al.* (2021) and references therein).

A work provided by Melina Kerou containing the annotation of genes and metabolic pathways these participate in was used as a reference material for this part. Had it been the case that the given unbalanced non-active reaction, i.e. reaction that didn't show any flux, didn't contain associated genes and enzymes, or if its GPR deviated significantly from the information provided in Kerou's work, the reaction was removed from the model.

Furthermore, the rest of the non-active unbalanced reactions not belonging to the above category, were balanced using the same procedure as with the active reactions.

Once the balancing corrections have been added to the model, the next part which included the addition of relevant gene annotation could be carried out. In this step, protein IDs from UniProt database were assigned to corresponding genes present in model's reactions, creating a model that is more comprehensible and consistent on a larger scale.

To check if the described alterations have proven to be successful, a metabolic model test suite MEMOTE (version 0.17.0) was used (see Lieven *et al.* (2020) and references therein), and afterwards, an FBA optimization could be performed, in order to obtain flux distributions which would then be visualized for some of the most important metabolic pathways of *Nitrososphaera viennensis* using the web tool Escher (see King *et al.* (2015) and references therein).

Lastly, a comparative analysis between the new improved model and the one from Musaev was performed, mainly to get insight into similarities and differences when it comes to active and inactive reactions, and furthermore, to compare two models' gene-protein associations with the data produced by proteomics experiment performed by Logan H. Hodgskiss and his colleagues (see Hodgskiss *et al.* (2025) and references therein). In these proteomics experiments, enzyme activity, i.e. amount was measured using Label-Free Quantification (see Zhu *et al.* (2010) and references therein), and so obtained LFQ values served as a reference in this Master's thesis for gene expression analysis and comparisons, for the purpose of which libraries Pandas (version 2.2.2), NumPy (version 1.26.4) and Matplotlib (version 3.9.2) were utilized.

3.2. Results

As already mentioned in the Methods subchapter, the first step was to balance the reactions which showed non-zero flux values upon performing the optimization process of the biomass objective function. The following table summarizes these reactions along with the necessary alterations:

Reaction Name	Flux (mmol/gDW/hour)	Original Equation	Performed Modifications	Flux after Balancing
adenyl - sulfate reductase	-0.007481631961898134	$AMP + H^+ + Sulfite + Ox - Thioredoxin \leftarrow APS + Red - Thioredoxin$	Change of formula and charge for Red-Thioredoxin to $C_{10}H_{15}N_4O_4R_4S_2$ and 1 respectively	-0.007481066958098967
NADPH-thioredoxin reductase	-0.007481631961898133	$NADP + Red - Thioredoxin \leftarrow NADPH + H^+ + Ox - Thioredoxin$	Change of formula and charge for Red-Thioredoxin to $C_{10}H_{15}N_4O_4R_4S_2$ and 1 respectively	-0.007481066958098966
methanofuran b synthesis	0.00016477826708709764	$2 L - Glutamate + 4 - [N - gamma - L - glutamyl -] - p - (beta - aminoethyl)phenoxy - methyl] - 2 - (aminomethyl)furan \leftrightarrow 2 H_2O + methanofuran b)$	Change of formula and charge for Methanofuran b to $C_{29}H_{38}N_5O_{11}$ And -1 respectively	0.00016476582323161283
CoA:apo-[acyl-carrier-protein] pantetheinephosphotransferase	0.00016477826708709764	$CoA + apo - ACP \leftrightarrow Adenosine 3 - 5 - bisphosphate + ACP$	Change of charge for apo-ACP to -2	0.0001647658232316128
L-glutamate:coenzyme F420-1 ligase (GDP-forming)	0.00016477826708709764	$L - Glutamate + GTP + Coenzyme F_{420-1} \rightarrow Phosphate + GDP + H^+ + Coenzyme F_{420}$	Removal of H^+ from the right-hand side of the equation	0.00016476582323161283
(2S)-lactyl-2-diphospho-5'-guanosine:7,8-didemethyl-8-hydroxy-5-deazariboflavin lactylphosphotransferase	0.0001647782670870976	$7,8 - didemethyl - 8 - hydroxy - 5 - deazariboflavin + lactyl - (2) - diphospho - (5') - guanosine \rightarrow H^+ + GMP + Coenzyme F_{420-0}$	Addition of 1 H^+ to the right-hand side of the equation	0.00016476582323161283
L-glutamate:coenzyme gamma-F420-2 (ADP-forming)	0.00016477826708709764	$ATP + L - Glutamate + Coenzyme F_{420} \rightarrow ADP + Phosphate + H^+ + Coenzyme F_{420-3}$	Addition of 1 H^+ to the right-hand side of the equation	0.0001647658232316128

Reaction Name	Flux (mmol/gDW/hour)	Original Equation	Performed Modifications	Flux after Balancing
2,3-bis-(O-phytanyl)-sn-glycerol 1-phosphate:NADP+ oxidoreductase	-0.0012747635506682 629	$8 \text{ NADP} + 2,3 - \text{Bis} - (\text{O} - \text{phytanyl}) - \text{sn} - \text{glycerol} - 1 - \text{phosphate} \leftrightarrow 8 \text{ NADPH} + 8 \text{ H}^+ + 2,3 - \text{Bis} - \text{O} - (\text{geranylgeranyl})\text{glycerol} 1 - \text{phosphate}$	Removal of 1 H ⁺ from the right-hand side of the equation	0.00459789 1367832805
CTP:2,3-bis-O-(geranylgeranyl)-sn-glycerol-1-phosphate cytidyltransferase	0.00063738177533413 14	$\text{CTP} + 2,3 - \text{Bis} - \text{O} - (\text{geranylgeranyl})\text{glycerol} 1 - \text{phosphate} \leftrightarrow \text{PPi} + \text{CDP} - 2,3 - \text{bis} - \text{O} - (\text{geranylgeranyl}) - \text{sn} - \text{glycerol}$	Addition of 1 H ⁺ to the right-hand side of the equation	0.00063733 3641033163 6
2,3-bis-(O-phytanyl)-sn-glycerol 1-phosphate:NAD+ oxidoreductase	-30.416715164649943	$8 \text{ NAD} + 2,3 - \text{Bis} - (\text{O} - \text{phytanyl}) - \text{sn} - \text{glycerol} 1 - \text{phosphate} \leftrightarrow 8 \text{ NADH} + 8 \text{ H}^+ + 2,3 - \text{Bis} - \text{O} - (\text{geranylgeranyl})\text{glycerol} 1 - \text{phosphate}$	Removal of 1 H ⁺ from the right-hand side of the equation	-0.0058725 5864989913 2

Table 1. Summary of balancing of active reactions in FBA

After this initial step of model optimization, reactions which showed no flux were analysed and compared with Kerou's annotation data to check if they play a significant role in the archaeon's metabolism, i.e. if they can be safely removed from the model without disrupting the model's accuracy and predictive power. It was determined that some of these inactive reactions didn't contain GPR rules, and some only contained hypothetical annotation which hasn't been experimentally validated. These reactions were removed from the model altogether and are listed in the table below:

Reaction Name/ Reaction ID	Equation
rxn04656	$L - 2 - \text{Amino adipate} \leftrightarrow N2 - \text{Acetyl} - L - \text{amino adipate}$
RXN-10622.c	$\text{Cyclic de} - \text{hypoxanthine futalosine} \leftarrow 1,4 - \text{Dihydroxy} - 6 - \text{naphthoate} + \text{UNKNOWN}$
RXN-9771.c	$\text{H}_2\text{O} + \text{Iminoaspartate} \leftarrow \text{NH}_3 + \text{Oxaloacetate}$
RXN-13329.c	$\text{H}_2\text{O} + \text{Iminoglycine} \leftarrow \text{NH}_3 + \text{Glyoxalate}$
enterobactin Fe(III) binding (spontaneous)	$\text{Enterobactin} + \text{Fe}_3 \rightarrow \text{Fe} - \text{enterochlin}$

Table 2. Reactions removed from the model

The rest of the reactions with zero flux values were balanced the same way shown in Table 1 (through addition/removal of metabolites and changing of compounds' charges and formulas).

After the adequate UniProt IDs have then been assigned to genes present in the model, a MEMOTE report was generated in order to check model's consistency. The report showed that the improved version of the model has a stoichiometric consistency of 100%, mass balance consistency of 99.6%, charge balance consistency of 99.3%, metabolite connectivity of 99.5% and unbounded flux in default medium consistency of 92.3%. Along with the added UniProt gene annotations, the MEMOTE score for this new improved model, which is calculated as the weighted sum of scores for all individual categories (Consistency, Metabolites-, Reactions-, Gene annotation, etc.) divided by maximum achievable score, was 89%, which highlights the improvement compared to Musaev's model with the score of 72%. It should be noted that neither mass balance consistency nor charge balance consistency reach 100%. This matter, however, doesn't represent an issue for the overall model predictive power and further analysis, since usually chemical formulas and charges of some metabolites in the generic reactions, like exchange reactions, or lumped biomass, protein synthesis and nucleic acid synthesis reactions aren't completely exact or known. For example, metabolite charges can show variability, especially when it comes to parameters like pH, and molecules in reactions can be present in different charge states, depending on what state from the ion 'pool' they take on. Additionally, central metabolic and more significant reactions which predominantly determine model's flux distributions are properly balanced. Therefore, minor inconsistencies shown in the report can be safely disregarded.

Upon solving the model, the biomass objective function showed the optimized flux value of $0.03004359160483208 \text{ h}^{-1}$, which means the organism takes to produce 1 unit of biomass, i.e. has a growth rate of approximately 0.0300 h^{-1} . This value is reasonably similar to the empirical data obtained from experimental studies, which underlines its biological realism (see Melcher *et al.* (2023) and references therein). Furthermore, resulting flux distributions were analysed in three important metabolic pathways which are shown in the figures that follow.

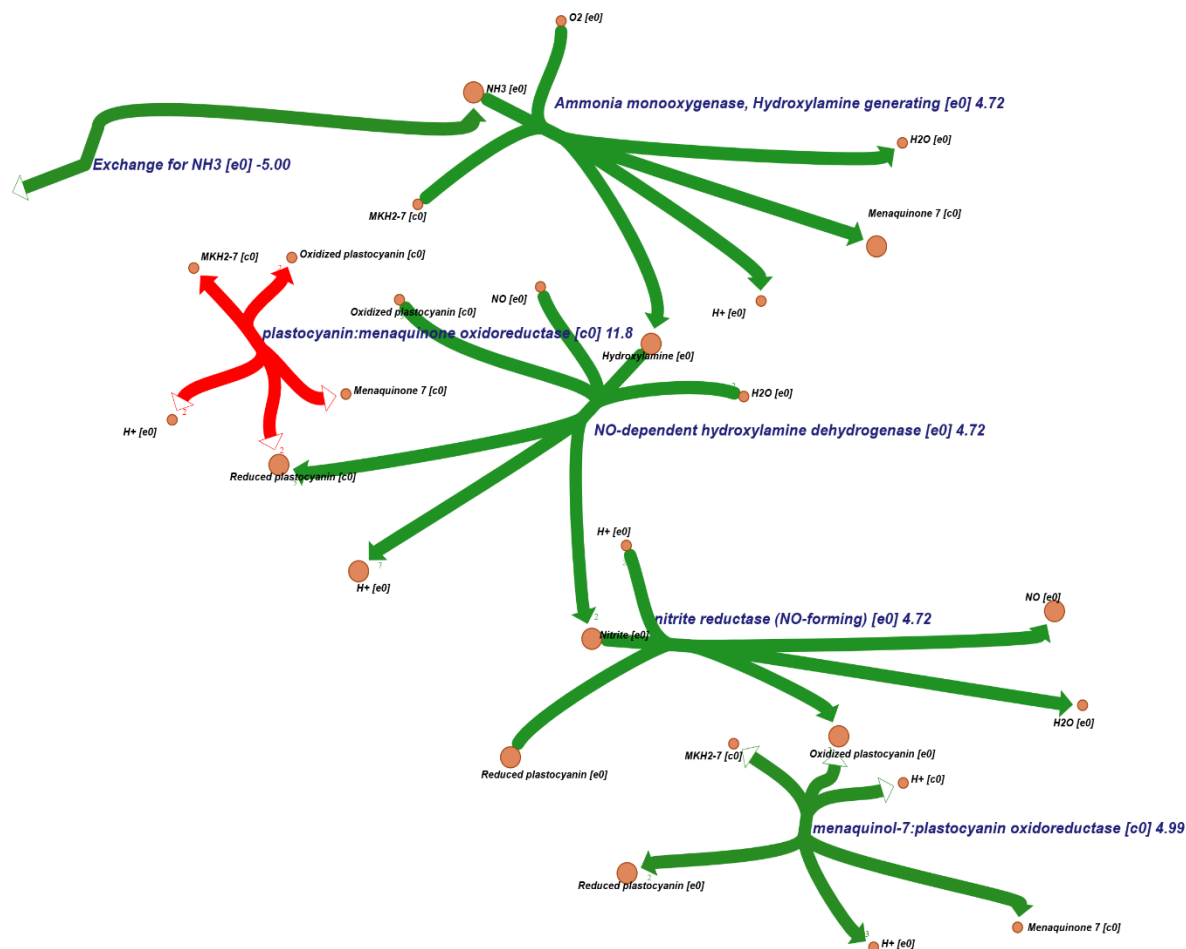


Figure 3. 1. FBA flux distribution through NH₃ oxidation cycle

In the above picture one can see that all reactions participating in the NH₃ oxidation cycle, one of the most important metabolic pathways in *N. viennensis*, show significant amount of flux, with one of the electron transport chain reactions showing the highest value of 11.8 mmol/gDW/hour (colored in red). This indicates that the model recognizes the oxidation of ammonia, one of the key features of *N. viennensis*, as an important metabolic pathway and favors matter flow through its reactions. Enzymes catalysing these reactions are presented in the following table:

Reaction	Enzyme
Exchange for NH ₃	No annotation
Ammonia monooxygenase, Hydroxylamine generating	ammonia monooxygenase, subunit C
NO-dependent hydroxylamine dehydrogenase	putative Copper binding protein, plastocyanin/azurin family
plastocyanin:menaquinone oxidoreductase	putative Copper binding protein, plastocyanin/azurin family
nitrite reductase (NO-forming)	multicopper oxidase type 3
menaquinol-7:plastocynin oxidoreductase	putative cytochrome b/b6 & putative Rieske 2Fe-2S domain cytochrome b6-f complex iron-sulfur subunit & putative Copper binding protein, plastocyanin/azurin family

Table 3. 1. NH₃ oxidation reactions and corresponding catalysing enzymes

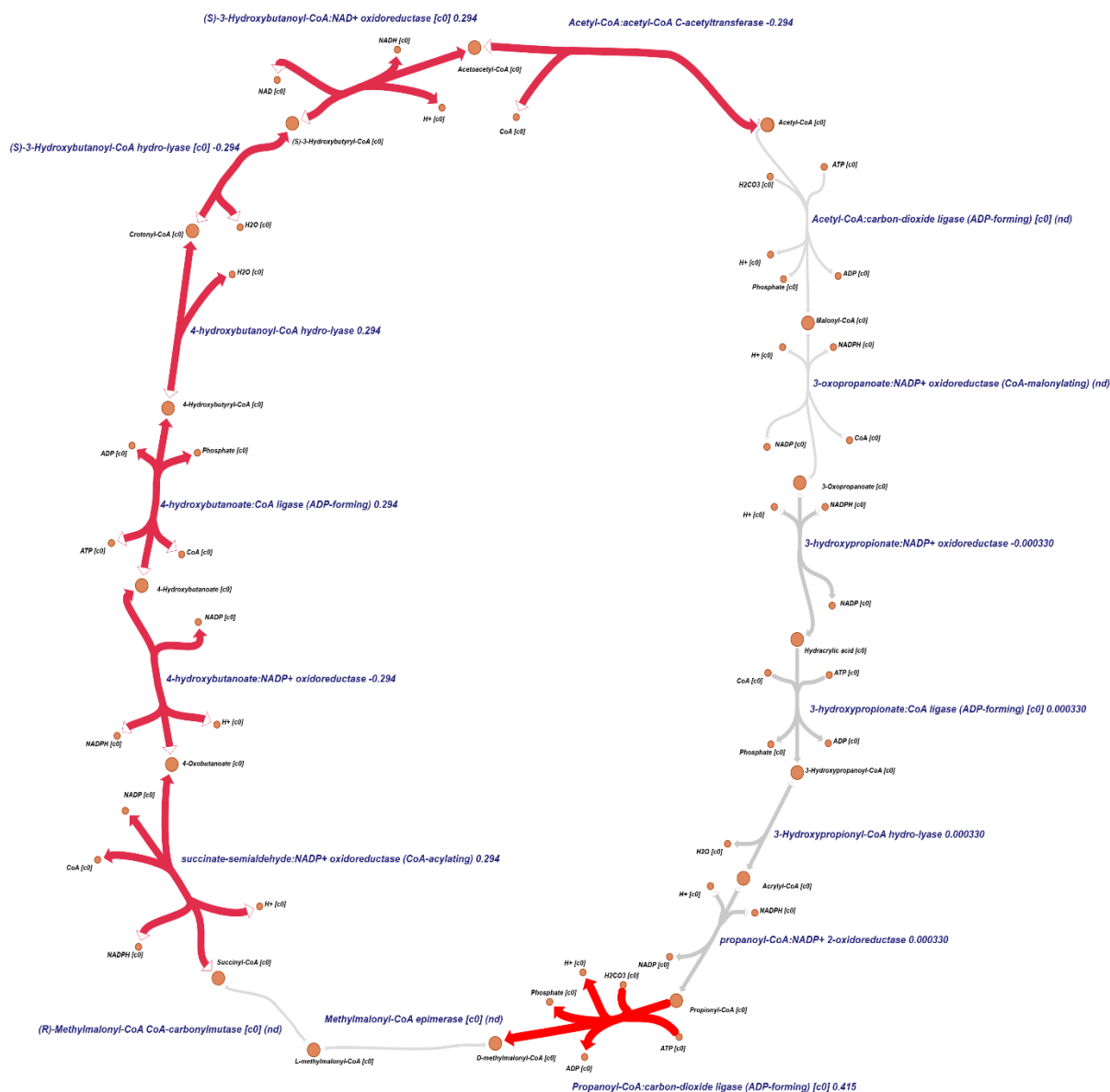


Figure 3. 2. FBA flux distribution through HP/HB cycle

From the figure above it is evident that reactions of metabolic pathway responsible for carbon acquisition also show non-zero flux values, indicating that model predicts them to be of significant importance.

However, several reactions in this cycle don't carry any flux, with evident disruption occurring in the reactions responsible for converting D-methylmalonyl-CoA to L-methylmalonyl-CoA and ultimately to Succinyl-CoA which is then utilized for further steps of the cycle. Instead, model favors the reaction that uses D-methylmalonyl-CoA in order to produce Propionyl-CoA and Oxaloacetate as part of the optimal solution, while the necessary Succinyl-CoA is produced by alternative reaction, namely, the one where CoA, Succinate and ATP serve as reactants. Similarly, model predicts that in the optimal solution 3-Oxopropanoate is not produced by the initial reactions of HP/HB cycle which convert Acetyl-CoA to 3-Oxopropanoate through an intermediate compound, Malonyl-CoA, but through a reaction where CO₂, NADH and Acetyl-CoA serve as reactants (3-Oxopropanoate:NAD⁺ oxidoreductase (decarboxylating, CoA-acetylating)).

Enzymes catalysing the HP/HB cycle reactions are listed in the following table:

Reaction	Enzyme
Acetyl-CoA:acetyl-CoA C-acetyltransferase	acetoacetyl-CoA thiolase subunit A & putative beta-ketothiolase beta subunit
Acetyl-CoA:carbon-dioxide ligase	acetyl-CoA/ propionyl-CoA carboxylase, carboxyltransferase subunit & acetyl-CoA/propionyl-CoA carboxylase, biotin carboxylation subunit & biotin carboxyl carrier protein of putative acetyl-CoA/ propionyl-CoA carboxylase
3-oxopropanoate: NADP ⁺ oxidoreductase (CoA-malonylating)	No annotation
3-hydroxypropionate:NADP ⁺ oxidoreductase	malonic semialdehyde reductase
3-hydroxypropionate:CoA ligase (ADP-forming)	3-hydroxypropionyl-CoA synthetase
3-Hydroxypropionyl-CoA hydro-lyase	3-hydroxybutyryl-CoA dehydratase
propanoyl-CoA:NADP ⁺ 2-oxidoreductase	putative NAD-dependent alcohol dehydrogenase
Propanoyl-CoA:carbon-dioxide ligase (ADP-forming)	acetyl-CoA/ propionyl-CoA carboxylase, carboxyltransferase subunit & acetyl-CoA/propionyl-CoA carboxylase, biotin carboxylation subunit & biotin carboxyl carrier protein of putative acetyl-CoA/ propionyl-CoA carboxylase
Methylmalonyl-CoA epimerase	methylmalonyl-CoA epimerase

Reaction	Enzyme
(R)-Methylmalonyl-CoA CoA-carboxylmutase	methylmalonyl-CoA mutase, large subunit
succinate-semialdehyde:NADP ⁺ oxidoreductase (CoA-acylating)	succinate-semialdehyde dehydrogenase
4-hydroxybutanoate:NADP ⁺ oxidoreductase	No annotation
4-hydroxybutanoate:CoA ligase (ADP-forming)	acetyl-CoA synthetase (ADP-forming)
4-hydroxybutanoyl-CoA hydro-lyase	4-hydroxybutyryl-CoA dehydratase
(S)-3-Hydroxybutanoyl-CoA hydro-lyase	3-hydroxybutyryl-CoA dehydratase
(S)-3-Hydroxybutanoyl-CoA:NAD ⁺ oxidoreductase	hydroxyacyl-CoA dehydrogenase

Table 3. 2. HP/HB cycle reactions and corresponding catalysing enzymes

When it comes to the Cytric Acid Cycle (TCA cycle), performing the FBA optimization produced following fluxes through reactions of this fundamental metabolic pathway:

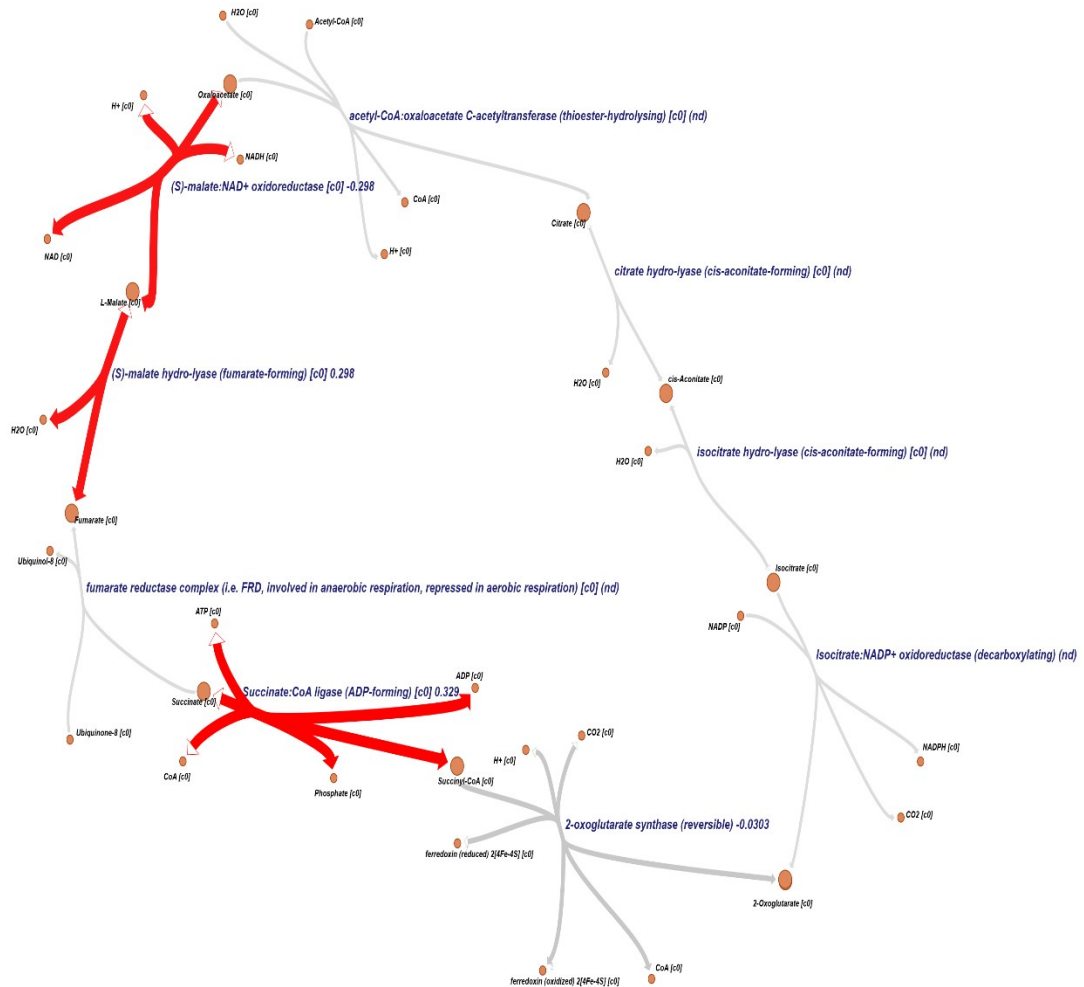


Figure 3. 3. FBA flux distribution through TCA cycle

Similarly to the HP/HB cycle, central carbon metabolism pathway also shows that some key reactions don't contain flux in the optimal solution. The most evident is the disruption in flux occurring in the reaction that converts Oxaloacetate to Citrate. Instead, model favors the reaction that produces Pyruvate and D-methylmalonyl-CoA from Oxaloacetate and predicts the flux value of -0.298 mmol/gDW/hour in it.

Furthermore, even though 2-Oxoglutarate is produced/consumed in the initial reaction of the cycle as shown above, optimal solution actually predicts that this compound is more abundant in reactions used in aminoacid biosynthesis pathways, some of which showed greater flux values; e.g. L-Aspartate:2-oxoglutarate aminotransferase with -0.117 mmol/gDW/hour. Enzymes catalysing the reactions shown in the above figure are listed in the table below, while the matter of incomplete flux distribution through this cycle is discussed in more detail in the Discussion chapter.

Reaction	Enzyme
Succinate:CoA ligase (ADP-forming)	succinyl-CoA ligase (ADP-forming) beta subunit & succinyl-CoA ligase (ADP-forming) alpha subunit
fumarate reductase complex	FeS-center protein of succinate dehydrogenase/fumarate reductase & succinate dehydrogenase flavoprotein subunit/fumarate reductase
(S)-malate hydro-lyase (fumarate-forming)	putative fumarate hydratase
(S)-malate:NAD ⁺ oxidoreductase	malate dehydrogenase
acetyl-CoA:oxaloacetate C-acetyltransferase (thioester-hydrolysing)	citrate synthase
citrate hydro-lyase (cis-aconitate-forming)	aconitate hydratase
isocitrate hydro-lyase (cis-aconitate-forming)	aconitate hydratase
Isocitrate:NADP ⁺ oxidoreductase (decarboxylating)	2:oxoacid-ferredoxin oxidoreductase, fused alpha and gamma subunit & 2:oxoacid-ferredoxin oxidoreductase, beta subunit
2-oxoglutarate synthase (reversible)	2:oxoacid-ferredoxin oxidoreductase, fused alpha and gamma subunit OR 2:oxoacid-ferredoxin oxidoreductase, beta subunit

Table 3. 3. TCA cycle reactions and corresponding catalysing enzymes

As the final step of FBA modelling, a comparison between reaction fluxes and associated genes was made with the experimentally obtained data described in the Methods subchapter. The main idea behind this was to test model's predictive power by comparing real-life protein activity with the estimations carried out by the model and get a clearer picture where these two domains differ. In this regard, the most interesting part was to determine how many and what reactions contained flux values deviating from 0, but whose catalysing enzymes showed LFQ values equal to 0, i.e. whose corresponding enzymes showed no expression in the proteomics data. This will also be important for the RBA model, because it will highlight the fact that modelling approaches sometimes favor the usage of pathways and metabolic processes that weren't verified in real-life experimental conditions. For this purpose, fluxes of reactions in the optimal solution were categorized and their GPR rules were mapped to experimental data obtained by Logan H. Hodgskiss. The following table summarizes the number of reactions in each category:

	Flux not equal to 0	Flux equal to 0	
Expression in Proteomics Data	151	249	
No Expression in Proteomics Data	41	130	
	Sum: 192	Sum: 379	Total: 571

Table 4. Classification of reactions based on FBA flux values and corresponding enzyme expression levels

Lastly, the comparison of genes in active reactions between the original model from Musaev and the new updated one was performed and the following results have been obtained:

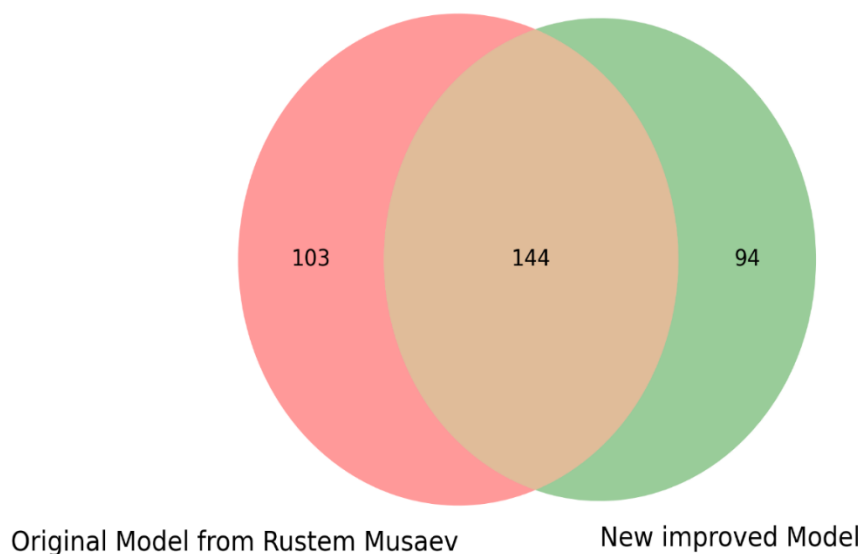


Figure 4. Venn diagram depicting comparison of number of distinct genes present in active reactions

From the above figure it is evident that there are 144 genes associated with active reactions in common for both models, indicating that there is a relatively high number of reactions containing flux in the optimal solution that are the same in both models. Out of these 144 genes, when compared to the proteomics data, 108 genes showed expression, i.e. their encoded proteins had non-zero LFQ values, while for 36 of them zero expression values had been experimentally determined. Some of these shared genes/enzymes include: NADH-quinone oxidoreductase subunit M, histidinol dehydrogenase, 4-hydroxybutyryl-CoA dehydratase, 3-isopropylmalate dehydrogenase, threonine ammonia-lyase, etc.

4. Resource Balance Analysis (RBA) metabolic Model of *Nitrososphaera viennensis*

4.1. Methods

In order to perform RBA metabolic modelling, RBAPy software package (version 2.0.1) and CPLEX optimization solver (version 22.1.1) were used. The RBA model was constructed by closely following the instructions provided in Ana Bulović's doctoral dissertation (see Bulović (2023) and references therein), as well as the tutorials available in the RBAPy GitHub repository. Creating an RBA model with the software mentioned above is a relatively automated process which only requires a handful of input parameters that should be provided by the user, including the metabolic model in the SBML file format, chaperone sequences, and rRNA and tRNA information of the desired organism. Sequences and molecule IDs of the mentioned parameters were obtained from online databases like KEGG, UniProt and NCBI and stored as FASTA files. Additionally, SBML/XML model file's writing conventions had to be readjusted in order to be compatible with the software's expectations. Namely, for the biomass and mRNA reaction, the chemical formula value was originally set to 'NULL' which isn't a valid format for the RBAPy, so these values had to be removed completely. Secondly, the Boolean rule for the GPR rule of the reaction 'R_Complex_4' had to be altered, so that the input parameters could be correctly loaded.

The final preparation step in the model creation included a brief developmental phase where software's internal module files were updated to ensure compatibility with the latest Python release.

Now that the input parameters were ready, the first run of RBA model generation was performed, and produced an incomplete model that had to be refined by altering the newly automatically created helper files in the TSV text format, along with the addition of 20 new reactions to the model.

These reactions represent the charging process of uncharged aminoacid tRNAs, and can be depicted by the following pseudo-reaction equation:



When it comes to the alterations of the above mentioned helper files, two crucial steps were to change the TSV files defining the metabolites, i.e. defining which metabolic precursors are utilized or generated by cellular processes in the formation of biomass components, and process locations in terms of cellular compartments. ChEBI IDs were used for successful modification of the ‘metabolites.tsv’ file, while UniProt compartment IDs were used in the ‘location_map.tsv’ file.

All the steps mentioned above allowed the creation of the complete RBA model of *N. viennensis* which consisted of several XML files, including among others ‘enzymes.xml’, ‘parameters.xml’, and ‘metabolism.xml’, and a ‘medium.tsv’ file which could later also be modified to match the desired compound concentrations in the organism’s environment.

As mentioned in the Introduction chapter, the last part of RBA modelling was to perform sensitivity analysis, in order to see if there are enzymes whose catalytic activity plays a significant role in the optimal growth rate prediction of the model (see Wendering *et al.* (2023) and references therein). This process required to carry out simulations for each enzyme in the model by changing its default efficiency value of 200000 h⁻¹ to either 400000 h⁻¹ or 100000 h⁻¹ and reporting the outcome of the optimal growth rate upon model optimization. The model files modified during this step included ‘enzymes.xml’ and ‘parameters.xml’. These simulations facilitated the identification of enzymes for which incorporating precise k_{cat} values would be most beneficial, thereby enabling a clearer and more accurate representation of metabolic activity.

4.2. Results

After modifying the ‘medium.tsv’ file to contain metabolites described in the Methods part of the previous chapter, adding corresponding changes to the ‘metabolites.xml’ file, and performing the optimization procedure, the model predicted optimal growth rate of $0.0389862060546875 \text{ h}^{-1}$ and flux distributions shown in the figures that follow:

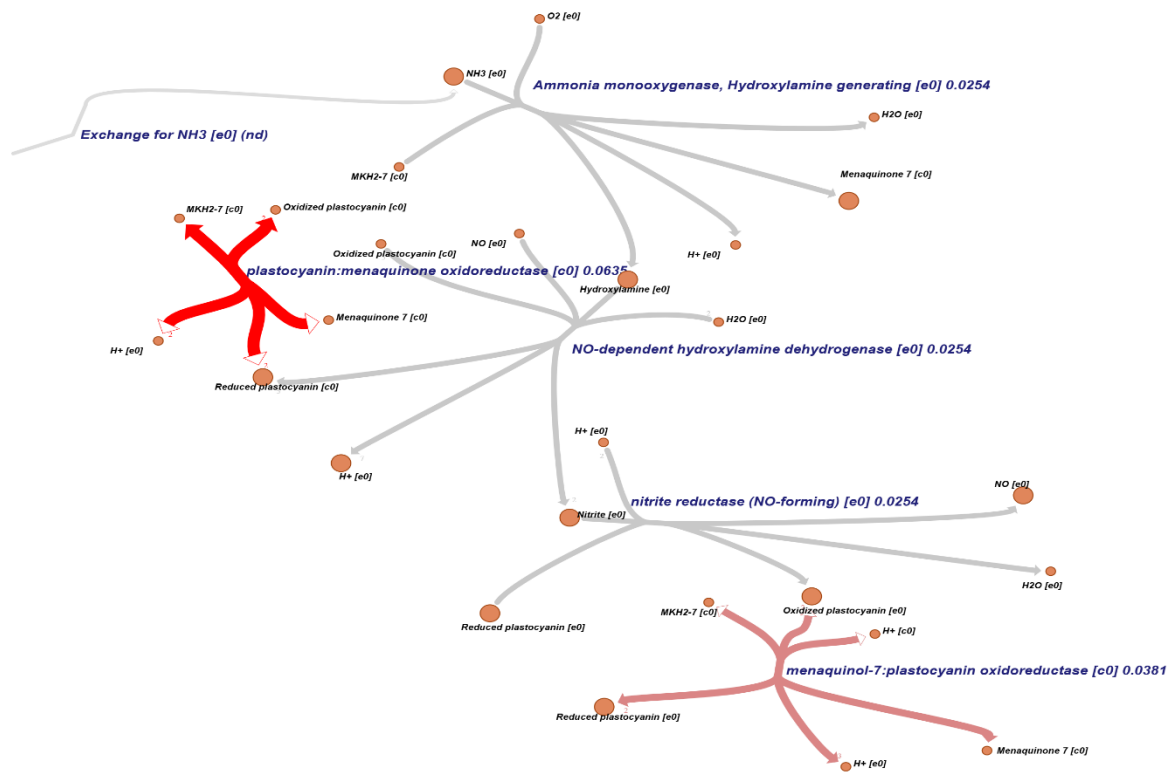


Figure 5. 1. RBA flux distribution through NH_3 oxidation cycle

As it was the case with the FBA optimal solution, all reactions taking place in the metabolic pathway responsible for NH_3 oxidation show non-zero fluxes, with plastocyanin:menaquinone oxidoreductase reaction having the highest flux value of 0.0635 mmol/gDW/hour.

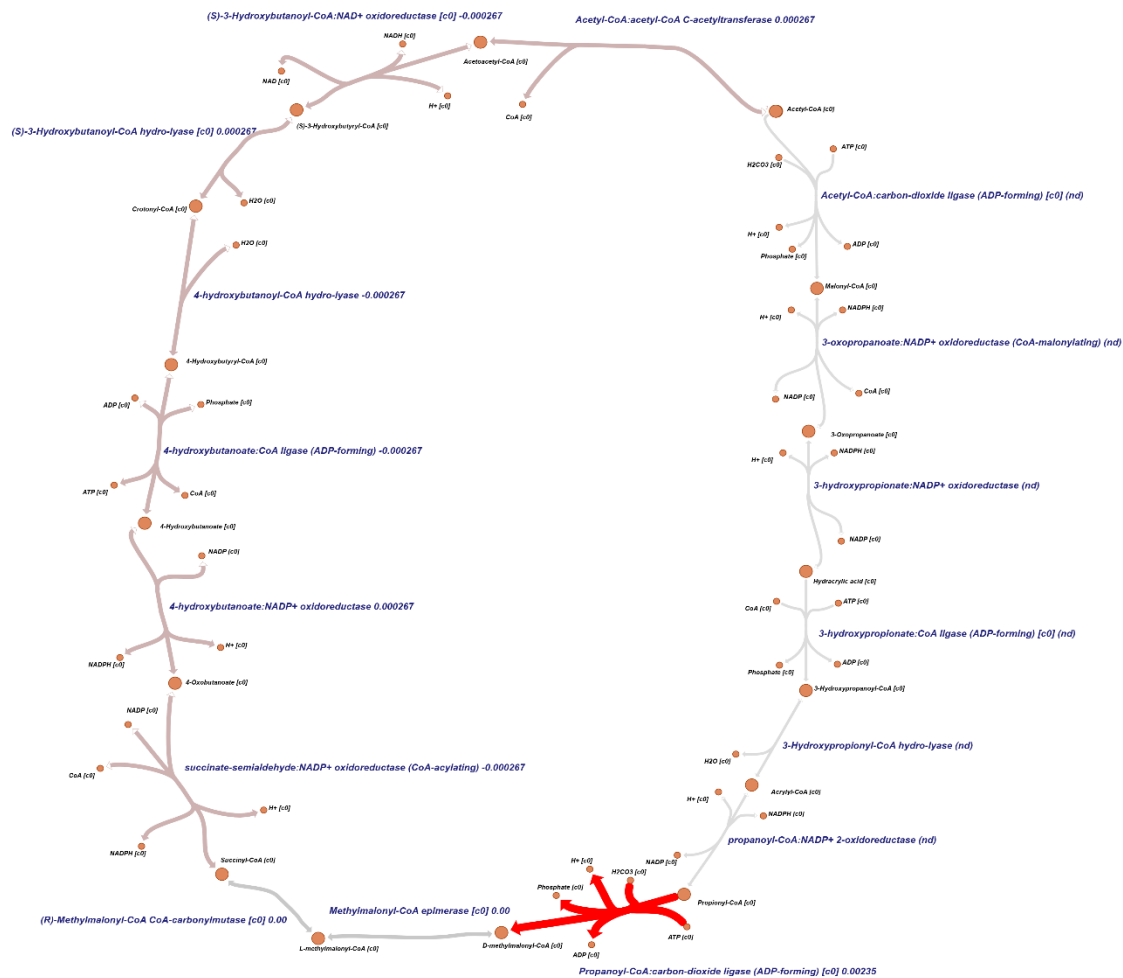


Figure 5.2. RBA flux distribution through HP/HB cycle

When it comes to HP/HB cycle, one can see many similarities with the FBA model, mostly in the way 3-Oxopropionate enters the cycle, which is through 3-Oxopropionate:NAD⁺ oxidoreductase (decarboxylating, CoA-acetylating) reaction. However, unlike the FBA model which contained a disruption of flux through two reactions that are responsible for converting D-methylmalonyl-CoA to Succinyl-CoA, the optimal solution here predicts that, even relatively small, there is flux through these key steps in the HP/HB cycle, namely, approximately -2.5166×10^{-7} mmol/gDW/hour in the reaction (R)-Methylmalonyl-CoA CoA-carboxylmutase and approximately 2.5166×10^{-7} mmol/gDW/hour in the reaction Methylmalonyl-CoA epimerase.

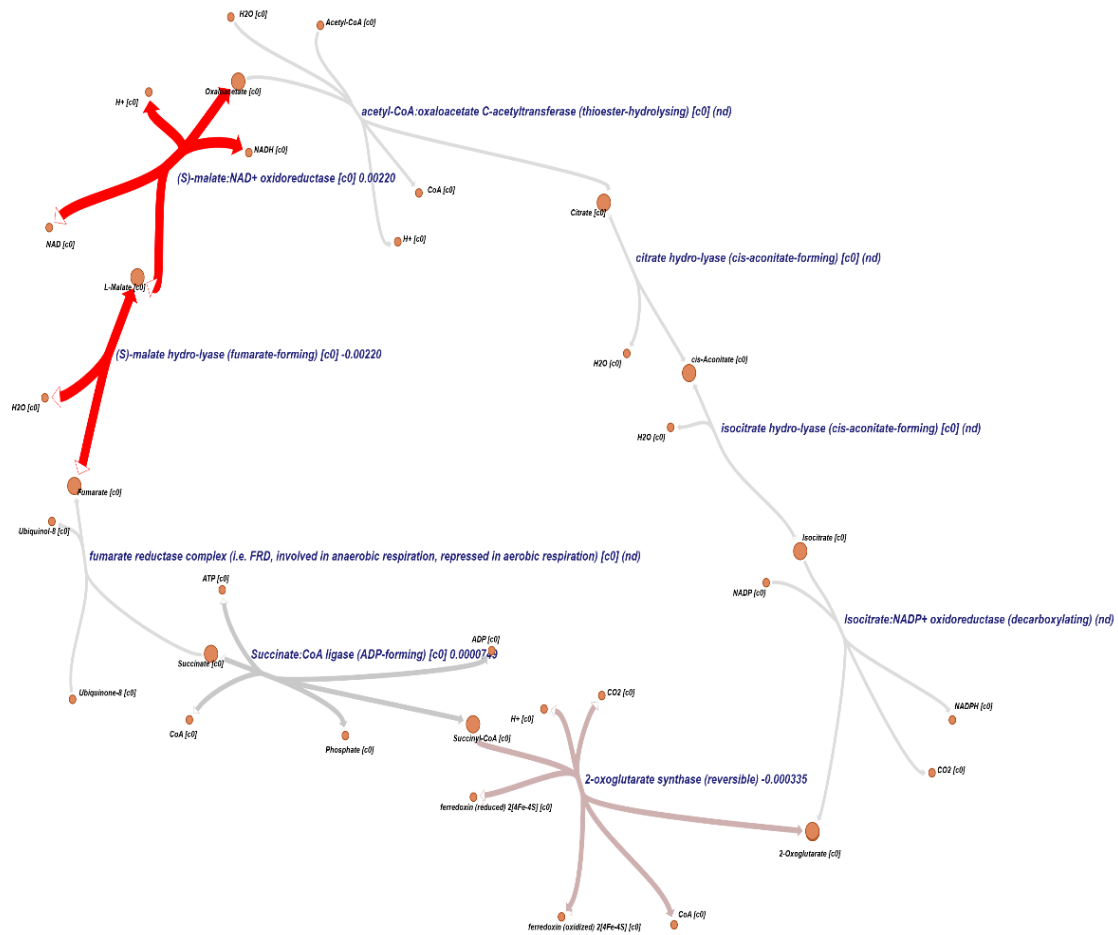


Figure 5.3. RBA flux distribution through TCA cycle

In the figure above it can be seen that RBA model predicts the same flux distribution in TCA cycle reactions as the FBA model, with the most interesting matter being that, in the optimal solution, reaction converting Succinate to Fumarate using Ubiquinol as a component of the electron transport chain doesn't contain any flux. According to both FBA and RBA model, reaction responsible for inclusion of Fumarate to the cycle, containing the flux value of 0.0000810 mmol/gDW/hour, is the one that uses Menaquinol as the electron carrier.

When genes and corresponding enzymes in the optimal solution space reactions were compared to LFQ values from the proteomics data, the following summary has been obtained for the RBA model:

	Flux not equal to 0	Flux equal to 0	
Expression in Proteomics Data	136	264	
No Expression in Proteomics Data	47	124	
	Sum: 183	Sum: 388	Total: 571

Table 5. Classification of reactions based on RBA flux values and corresponding enzyme expression levels

For the sake of comparison between two models, 20 additionally added aminoacid tRNA charging reactions were excluded from this analysis, but it is worth noting that upon optimization, no flux was detected in them.

4.3. Sensitivity Analysis of Enzyme catalytic Efficiency

After simulating the effects of halved and doubled efficiency for each enzyme on optimal growth rate, results have been obtained that for majority of these simulations, optimal growth rate remained the same. This indicates that changed efficiency of most of the enzymes in the model, whose default efficiency value was set to 200000 h^{-1} , doesn't really influence the outcome of model's prediction. However, several enzymes have been shown to reduce or increase the optimal growth rate once their efficiency value has been halved, doubled or both.

The figure on the following page visualizes the effects of changed efficiency of these several enzymes:

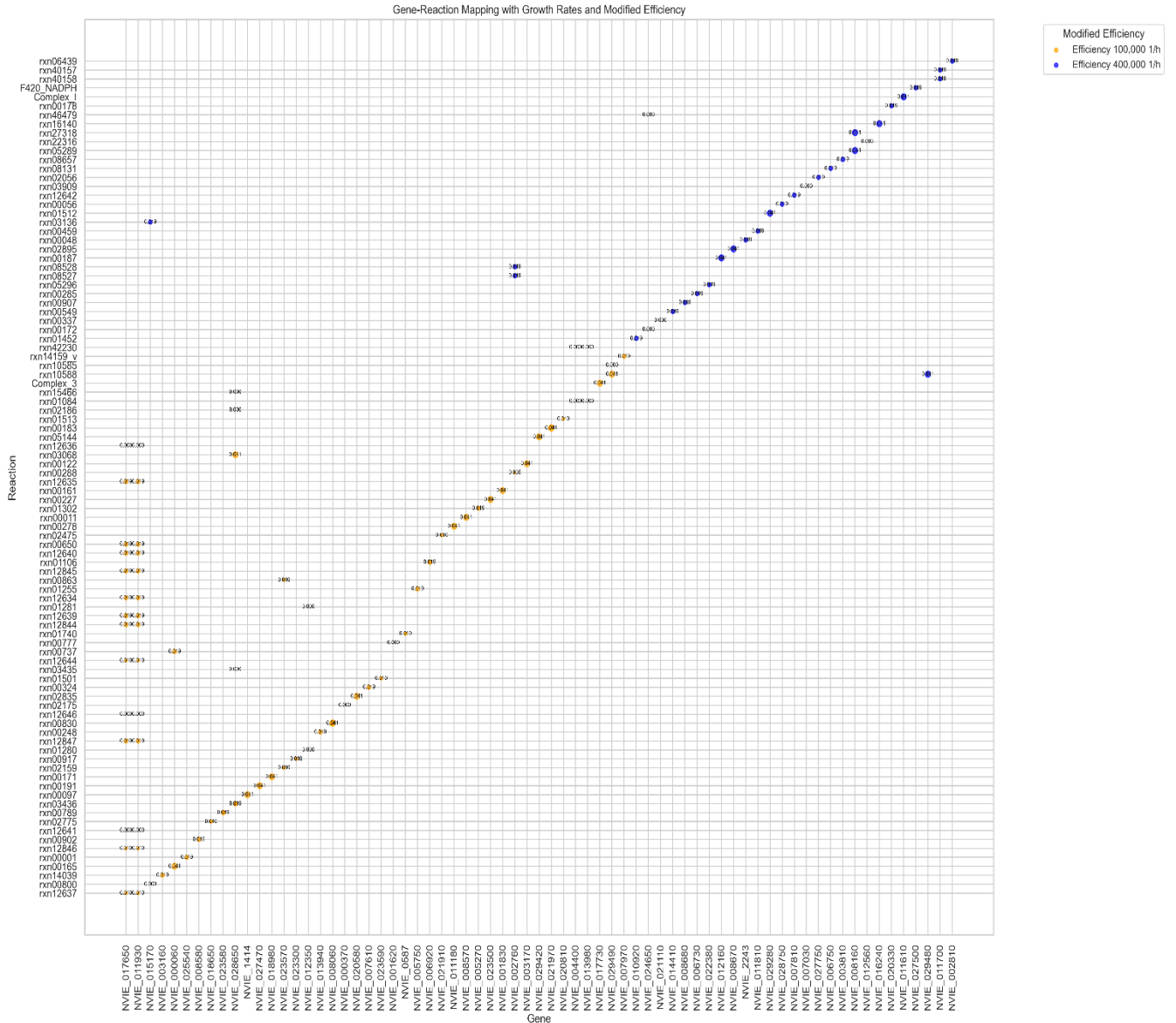


Figure 6. Effects of enzymes' modified efficiency on optimal growth rate

The above figure shows that the genes/enzymes listed on the x-axis can be considered sensitive parameters since once their default efficiency has been doubled, represented by blue color, or halved, represented by yellow color, model's prediction of optimal growth rate remarkably changes from the original value of $0.0389862060546875 \text{ h}^{-1}$. To make the figure more comprehensible, one can take a look at the example of the gene NVIE_017650 which encodes for enzyme leucyl aminopeptidase.

It is evident that in every reaction listed on the y-axis this enzyme participates in, once its efficiency has been reduced to 100000 h^{-1} , model's original optimal growth rate prediction value changes. On the other hand, if the efficiency of the enzyme bifunctional fructose-1,6-bisphosphatase/aldolase (NVIE_014410) is doubled in the D-Fructose-1,6-bisphosphate 1-phosphohydrolase reaction (rxn00549), the optimal growth rate value changes to approximately 0.019 h^{-1} .

Since all of the enzymes shown in the above figure showed to be of importance when it comes to the outcomes of model predictions, it can be stated that it should be reasonable to find their catalytic information and include it in the model if one wants a more accurate and complete representation of metabolism of *N. viennensis*.

Some of the shown enzymes were searched for in BRAunschweig ENzyme DAtabase (BRENDA), with the purpose of checking if there is any information available for them in terms of their k_{cat} or turnover numbers, which would at least approximately match the results obtained from the sensitivity analysis. Turnover numbers for a handful of enzymes have been found on BRENDA and the results are summarized in the table that follows:

Enzyme	Turnover number	Substrate on which turnover number was measured	Modified enzyme efficiency	Reactions in which change of efficiency affected growth rate
Hydroxyacyl-CoA dehydrogenase	374400 h ⁻¹	3-acetoacyl-CoA	400000 h ⁻¹	(S)-3-Hydroxybutanoyl-CoA:NADP+oxidoreductase
Methionine aminopeptidase	302400 h ⁻¹	L-Met-L-Ser-L-His-L-Arg-L-Trp-L-Asp-L-Trp	400000 h ⁻¹	RXNO-6978.c
Threonine ammonia lyase	117324 h ⁻¹	L-serine	100000 h ⁻¹	L-serine ammonia-lyase and L-threonine ammonia-lyase (2-oxobutanoate-forming)
Ketol-acid reductoisomerase	65880 h ⁻¹	2-acetolactate	100000 h ⁻¹	(S)-2-Aceto-2-hydroxybutanoate:NADP+ oxidoreductase (isomerizing) and (R)-2,3-Dihydroxy-3-methylpentanoate:NADP+ oxidoreductase (isomerizing)
Leucyl aminopeptidase	93600 h ⁻¹	Asp-Ile-Gly-Gly	100000 h ⁻¹	RXNO-6987.c and Gly-Phe aminopeptidase and RXNO-6977.c among others

Table 6. BRENDA turnover numbers for some of the sensitive enzymes

It is important to note, however, that most of the examined enzymes contained a wide range of plausible turnover numbers in BRENDA, as well as several different substrates these were measured on. This variability makes it difficult to directly compare BRENDA values with the model-derived estimates, as the experimental context and substrate specificity often differ. Still, the collected data offers a helpful point of reference for evaluating the reliability of the model outputs. A more in-depth analysis and interpretation of these findings will be provided in the Discussion chapter.

5. Comparative Analysis between FBA and RBA metabolic model of *Nitrososphaera viennensis*

With detailed results provided by both modelling approaches, it is now possible to compare them and analyse key similarities and differences. Firstly, as an overview, there are 277 reactions containing flux in both models, 48 containing flux only in the FBA model, and 36 containing flux only in the RBA model. The second step included performing a linear regression analysis with the help of which a scatter plot was created which summarizes flux values of active reactions in both FBA and RBA model and compares them with one another:

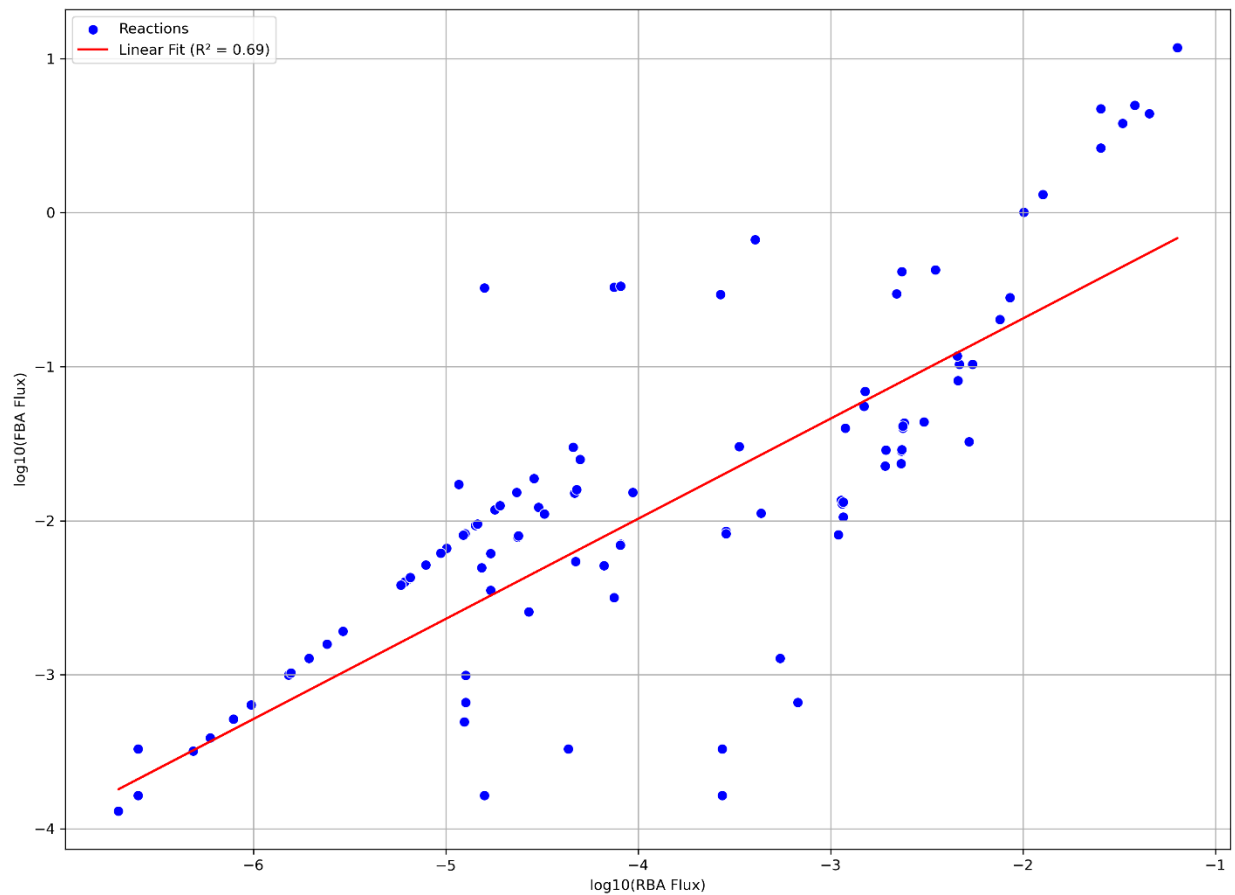


Figure 7. Scatter plot of linear regression analysis comparing flux values from two modelling approaches

In the above figure flux values of active reactions in the FBA model, transformed to log10 scale, are presented on the y-axis, and flux values of active reactions in the RBA model, also transformed to log10 scale, are presented on the x-axis. Linear regression analysis predicts a moderate to strong relationship between the two sets of fluxes, with coefficient of determination (R^2) having a value of 0.69, meaning that approximately 69% of variation in FBA flux values can be explained by the RBA flux values. This further suggests that reactions exhibiting low flux values in the FBA model tend to show similarly low fluxes in the RBA model, while reactions with high flux values in the FBA model also display correspondingly high fluxes in the RBA model. However, one can observe that there are several outliers present, and eight of these reactions with highest flux differences are shown in the table that follows:

Reaction	Flux difference
Plastocyanin:menaquinone oxidoreductase	11.735498
Menaquinol-7:plastocyanin oxidoreductase	4.947590
Ammonia monooxygenase, Hydroxylamine generating	4.694199
NO-dependent hydroxylamine dehydrogenase	4.694199
Nitrite reductase (NO-forming)	4.694199
O ₂ transport (diffusion)	1.300245
Passive diffusion of water	-4.346255
CO ₂ transporter via diffusion	0.995844

Table 7. Eight reactions with the highest flux value differences between FBA and RBA model

From the above table one can see that the reactions which showed the highest flux value differences between two modelling approaches are either the ones included in the NH₃ oxidation or the ones responsible for transport; uptake of essential compounds and secretion of metabolic byproducts. This suggests that resource allocation constraints that are considered in RBA play a crucial role in shaping the flux distribution for these reactions.

This fact highlights how metabolic potential (FBA) can diverge from physiological realism (RBA) under resource constraints, and indicates that NH₃ oxidation and compound transport are bottlenecks in terms of proteome investment, and organisms may regulate these processes to balance growth with resource expenditure.

As described in the previous chapters, reactions most worthy of further analysis are the ones that show flux values different from zero, but the genes of which showed no expression in the experimental data. Since they have already been obtained in both models, it can now be shown how many of them are shared between the two.

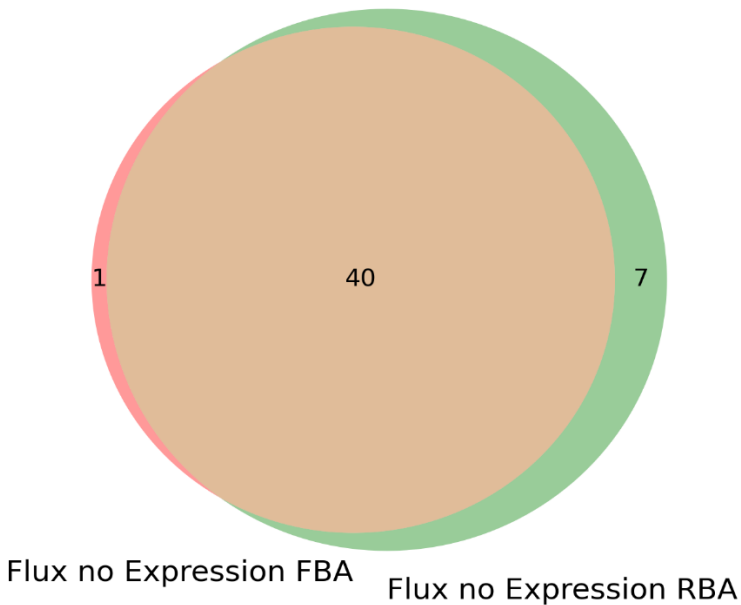


Figure 8. Venn diagram depicting comparison of number of active reactions with no gene expression in the proteomics data

From the above figure it can be understood that out of 48 active reactions obtained from both models, whose genes' LFQ values were 0, 40 are shared/the same in both FBA and RBA.

The following table lists some of these reactions along with brief description of their functions:

Reaction	Function
UDP-glucose:NAD+ 6-oxidoreductase	Biosynthesis of glycosaminoglycans, compounds present in extracellular matrix, and biosynthesis cell wall components (source: NCBI)
Prephenate:NAD+ oxidoreductase (decarboxylating)	Tyrosine biosynthesis (source: InterPro)
2-(Nomega-L-arginino) succinate arginine-lyase (fumarate-forming)	Biosynthesis of arginine and secondary metabolites (source: KEGG)
(S)-dihydroorotate amidohydrolase	Pyrimidine metabolism and cofactor biosynthesis (source: KEGG)
L-glutamine:D-fructose-6-phosphate isomerase (deaminating)	Glutamate and aminosugars metabolism (source: https://pmc.ncbi.nlm.nih.gov/articles/PMC2144271/)
Chorismate pyruvate mutase	Tyrosine and phenylalanine biosynthesis (source: KEGG)
5-phosphoribosylamine:diphosphate phospho-alpha-D-ribosyltransferase (glutamate-amidating)	De novo purine biosynthesis and purine metabolism (source: KEGG)
UDP-glucose 4-epimerase	Galactose metabolism (source: https://www.jbc.org/article/S0021-9258(20)82392-8/fulltext)
Nicotinate-nucleotide:pyrophosphate phosphoribosyltransferase (carboxylating)	Cofactor biosynthesis and metabolism of nicotinate and nicotinamide (source: KEGG)

Table 8. Common reactions between FBA and RBA whose genes aren't expressed in the proteomics experiment

The presence of flux through biosynthetic pathways in both FBA and RBA models, despite the absence of detectable protein expression under defined conditions without extracellular supply, suggests that *N. viennensis* may maintain minimal, tightly regulated flux through these pathways-possibly using very low enzyme levels that escape proteomic detection. Another possible explanation could be that due to relatively scarce data on archaeal metabolism in general, it is still hard to state with certainty if modelled pathways, especially some of the ones listed above, represent an accurate image of life in reality.

It has been found out that aromatic aminoacids like tyrosine and phenylalanine are synthesised in alternative ways in certain archaeal species, using pathways and precursors that differ from the ones that are common and generally well-known (see Hochuli *et al.* (1999) references therein). This highlights a limitation in relying solely on steady-state models and suggests that integrating proteomic detection limits and regulatory constraints could further improve model accuracy.

6. Discussion

This Master's thesis showed and summarized that both modelling approaches have advantages and disadvantages. The main area for which it can be stated that needs further work and consideration would be to refine the models in a way that would include all biochemical reactions of key metabolic pathways to be active. As shown in the previous chapters, both modelling approaches occasionally predicted that some reactions in central metabolic pathways, such as the TCA and HP/HB cycle, did not carry flux. For example, the conversion of oxaloacetate to citrate in the TCA cycle was bypassed in favor of oxaloacetate conversion to pyruvate, while 2-oxoglutarate was preferentially diverted into amino acid biosynthesis. This suggests that the models tend to reroute fluxes toward biosynthetically favorable pathways, which may reflect cellular priorities under certain conditions but could also indicate that key core metabolic reactions are underrepresented in the current model formulations. Since these core metabolism pathways represent major and almost universal way an organism operates on a metabolic level, it is unlikely that models' predictions about fluxes through them match the realistic and real-life picture. Moreover, some of the reactions along which the flux is rerouted may not actually be present in the organism. This matter can be most likely explained by the fact that the original model, created by Rustem Musaev and used as a starting point for this Master's thesis, was mostly constructed using orthology search between *N. viennensis* and phylogenetically related organisms, and therefore might include reactions and pathways that don't represent a biological reality for *N. viennensis* specifically. It goes without saying, that due to relatively scarce information/knowledge present about archaeal metabolism in general, it can be rather difficult to confirm the real-life presence of some of these reactions with certainty.

Similarly, given the presence of flux through biosynthetic pathways in both FBA and RBA, despite the absence of detectable protein expression, suggests that *N. viennensis* may sustain minimal fluxes below proteomic detection limits or utilize alternative routes not yet well characterized in archaea. This also underlines certain areas left for further improvement and optimization by possibly integrating regulatory constraints, detection thresholds, and archaeal-specific metabolic variants.

When it comes to a comparison between both modelling approaches used in this work, RBA modelling usually takes more factors into account, and could therefore be considered a better starting point for further refinements. Sensitivity analysis showed that enzyme turnover numbers can play a significant role when it comes to model's predictions, which suggests that incorporating their features could create a more realistic representation of functioning of *N. viennensis* on a deeper level. However, since catalytic rate constants are often unavailable for a large number of enzymes, some studies have proposed techniques like machine learning to estimate their values in the area of physiological and metabolic modelling (see Kroll *et al.* (2023) and references therein).

Since this approach proved successful, the idea of using it for *N. viennensis* should certainly be considered for future work and potential improvement strategies. Additionally, it is worth noting that some other relatively automated techniques have been proposed in the area of metabolic modelling that would also be useful to apply and explore when it comes to the work summarized in this thesis. One such approach/software is coralME which considers proteome allocation and gene expression in more detail, although more complex and detailed parameterization is also required (see Tibocha-Bonilla *et al.* (2024) and references therein).

It goes without saying that this concept warrants further investigation, which could potentially yield new results that help address the gaps in the current work. Taken together, these considerations highlight that even though the present work provided valuable insights, it still leaves place for methodological refinement and conceptual development. Addressing these aspects will be essential for achieving models that more closely reflect the complex metabolic reality of *N. viennensis* and living systems in general.

7. List of References

- 1) Bernhard, A. 2010. The Nitrogen Cycle: Processes, Players, and Human Impact. *Nature Education Knowledge*, 3(10): 25
- 2) Bodeit, O., Samir, I.B., Karr, J.R., Goelzer, A., Liebermeister, W. 2023. RBAtools: a programming interface for Resource Balance Analysis models. *Bioinformatics Advances*, 3: vbad056
- 3) Bulović, A., Fischer, S., Dinh, M., Golib, F., Liebermeister, W., Poirier, C., Tournier, L., Klipp, E., Fromion, V., Goelzer, A. 2019. Automated generation of bacterial resource allocation models. *Metab Eng.*, 55: 12-22
- 4) Bulović, A. 2023. Modeling growth and adaptation in bacteria. [Doctoral dissertation, Humboldt-Universität zu Berlin].
<https://edoc.hu-berlin.de/items/86807f3a-3537-4610-a835-430086298c76>
- 5) De Becker, K., Totis, N., Bernaerts, K., Waldherr, S. 2022. Using resource constraints derived from genomic and proteomic data in metabolic network models. *Current Opinion in Systems Biology*, 29: 100400
- 6) Di Filippo, M., Damiani, C., Pescini, D. 2021. GPRuler: Metabolic gene-protein-reaction rules automatic reconstruction. *PLoS Comput Biol.*, 17(11): e1009550
- 7) Garritano, A.N., Song, W., Thomas, T. 2022. Carbon fixation pathways across the bacterial and archaeal tree of life. *PNAS Nexus*, 1: pgac226
- 8) Giordano, N., Mairet, F., Gouzé, J.L., Geiselmann, J., de Jong, H. 2016. Dynamical Allocation of Cellular Resources as an Optimal Control Problem: Novel Insights into Microbial Growth Strategies. *PLoS Comput Biol.*, 12(3): e1004802
- 9) Hochuli, M., Patzelt, H., Oesterhelt, D., Wüthrich, K., Szyperski, T. 1999. Amino Acid Biosynthesis in the Halophilic Archaeon *Haloarcula hispanica*. *J Bacteriol.*, 181(10): 3226-3237

- 10) Hodgskiss, L.H., Kerou, M., Luo, Z.H., Bayer, B., Maier, A., Weckwerth, W., Naegele, T., Schleper, C. 2025. Metabolic Response of a Chemolithoautotrophic Archaeon to Carbon Limitation. *bioRxiv*, doi.org/10.1101/2025.02.13.638070
- 11) Kerou, M., Offre, P., Valledor, L., Abby, S.S., Melcher, M., Nagler, M., Weckwerth, W., Schleper, C. 2016. Proteomics and comparative genomics of *Nitrososphaera viennensis* reveal the core genome and adaptations of archaeal ammonia oxidizers. *Proc. Natl. Acad. Sci. U.S.A.*, 113(49): E7937-E7946
- 12) King, Z.A., Dräger, A., Ebrahim, A., Sonnenschein, N., Lewis, N.E., Palsson, B.O. 2015. Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. *PLoS Comput Biol.*, 11(8): e1004321
- 13) Kroll, A., Rousset, Y., Hu, X.P., Liebrand, N.A., Lercher, M.J. 2023. Turnover number predictions for kinetically uncharacterized enzymes using machine and deep learning. *Nat Commun.*, 14: 4139
- 14) Lehtovirta-Morley, L.E. 2018. Ammonia oxidation: Ecology, physiology, biochemistry and why they must all come together. *FEMS Microbiology Letters*, 365: fny058
- 15) Li, F., Xie, W., Yuan, Q., Luo, H., Li, P., Chen, T., Zhao, X., Wang, Z., Ma, H. 2018. Genome-scale metabolic model analysis indicates low energy efficiency in marine ammonia-oxidizing archaea. *AMB Express*, 8: 106
- 16) Lieven, C., Beber, M.E., Olivier, B.G., Bergmann, F.T., Ataman, M., Babaei, P., Bartell, J.A., Blank, L.M., Chauhan, S., Correia, K., Diener, C., Dräger, A., Eber, B.E., Edirisinghe, J.N., Faria, J.P., Feist, A.M., Fengos, G., Fleming, R.M.T., García-Jiménez, B., Hatzimanikatis, V., van Helvoirt, W., Henry, C.S., Hermjakob, H., Herrgård, M.J., Kaafarani, A., Kim, H.U., King, Z., Klamt, S., Klipp, E., Koehorst, J.J., König, M., Lakshmanan, M., Lee, D.Y., Lee, S.Y., Lee, S., Lewis, N.E., Liu, F., Ma, H., Machado, D., Mahadevan, R., Maia, P., Mardinoglu, A., Medlock, G.L., Monk, J.M., Nielsen, J., Nielsen, L.K., Nogales, J., Nookaew, I., Palsson, B.O., Papin, J.A., Patil, K.R., Poolman, M., Price, N.D., Resendis-Antonio, O., Richelle, A., Rocha, I., Sánchez, B.J., Schaap, P.J., Malik Sheriff, R.S., Shoaie, S., Sonnenschein, N., Teusink, B., Vilaça, P., Vik, J.O., Wodke, J.A.H., Xavier, J.C., Yuan, Q., Zakhartsev, M., Zhang, C. 2020. MEMOTE for standardized genome-scale metabolic model testing. *Nature Biotechnology*, 38: 272-276

- 17) Melcher, M., Hodgskiss, L.H., Mardini, M.A., Schleper, C., Rittmann, S.K.R. 2023. Analysis of biomass productivity and physiology of *Nitrososphaera viennensis* grown in continuous culture. *Frontiers in Microbiology*, 14: doi.org/10.3389/fmicb.2023.1076342
- 18) Nishizawa, M., Sakai, S., Konno, U., Nakahara, N., Takaki, Y., Saito, Y., Imachi, H., Tasumi, E., Makabe, A., Koba, K., Takai, K. 2016. Nitrogen and Oxygen Isotope Effects of Ammonia Oxidation by Thermophilic Thaumarchaeota from a Geothermal Water Stream. *Applied and Environmental Microbiology*, 82(15): AEM.00250
- 19) Orth, J.D., Thiele, I., Palsson, B. Ø. 2010. What is flux balance analysis? *Nat Biotechnol*, 28(3): 245-248
- 20) Sahu, A., Blätke, M.A., Szymański, J.J., Töpfer, N. 2021. Advances in flux balance analysis by integrating machine learning and mechanism-based models. *Computational and Structural Biotechnology Journal*, 19: 4626-4640
- 21) Tibocha-Bonilla, J.D., Santibáñez-Palominos, R., Weng, Y., Kumar, M., Zengler, K. 2024. Proteome allocation of the microbiome reveals how diet and metabolic dysbiosis impact disease. *BioRxiv*, doi.org/10.1101/2024.11.21.624724
- 22) von Kamp, A., Klamt, S. 2023. Balancing biomass reaction stoichiometry and measured fluxes in flux balance analysis. *Bioinformatics*, 39: btad600
- 23) Wendering, P., Arend, M., Razaghi-Moghadam, Z., Nikoloski, Z. 2023. Data integration across conditions improves turnover number estimates and metabolic predictions. *Nat Commun.*, 14(1): 1485
- 24) Zhu, W., Smith, J.W., Huang, C.M. 2010. Mass spectrometry-based label-free quantitative proteomics. *J Biomed Biotechnol.*, 2010: 840518

Supplementary Files

Relevant and most important files that were created and used for the preparation of this Master's thesis will be attached along with it.

These files include Jupyter Notebooks used for initial model optimization consisting of addition, alteration and deletion of model reactions, along with the annotation process.

Furthermore, files and folders of the RBA model, and also a Jupyter Notebook used for its creation will be attached.