PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

DEVELOPMENT AND VALIDATION OF A GENOME-SCALE METABOLIC MODEL FOR THE MALOLACTIC BACTERIUM *OENOCOCCUS OENI*

SEBASTIÁN MENDOZA FARÍAS

Thesis submitted to the Office of Research and Graduate Studies in partial fulfillment of the requirements for the Degree of Master of Science in Engineering

Advisors:
EDUARDO E. AGOSIN

Santiago de Chile, October, 2014

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

DEVELOPMENT AND VALIDATION OF A GENOME-SCALE METABOLIC MODEL FOR THE MALOLACTIC BACTERIUM *OENOCOCCUS OENI*

SEBASTIÁN MENDOZA F.

Members of the Committee:

EDUARDO AGOSIN

RICARDO PEREZ

ALEJANDRO MAASS

JORGE VALDÉS

JUAN DE DIOS ORTÚZAR

Thesis submitted to the Office of Research and Graduate Studies in partial fulfillment of the requirements for the Degree of Master of Science in Engineering

Santiago de Chile, October, 2014

“Imagination is more important than knowledge. Knowledge is limited. Imagination encircles the world”

Albert Einstein

ACKNOWLEDGMENTS

I am grateful to Dr. Eduardo Agosin for inspiring me in my first years of university with his love for science and his teachings about how biotechnology could transform this world in a better place to live. Thanks for inviting me to participate in this exciting investigation and for his impeccable guidance in this process. I also want to acknowledge my evaluation committee, Drs Jorge Valdes, Alejandro Mass, Ricardo Perez and Juan de Dios Ortúzar who were always willing to help and contribute with valuable feedback to further enrich this investigation. I am especially grateful to Jorge Valdes for his valuable advice along the entire investigation. Thanks also to Professor Francisco Melo for given me his support to enter to the Master program.

I also wish to thank Pablo Cañón, Benjamín Sánchez, Paulina Torres, Francisco Saitua, Guillermo Gutierrez, Aline Ovalle and Paz Tapia for their willingness to help and valuable advice relating the contents and structure of this thesis. I am also grateful to my lab partners who always made me feel like if I was at home with them.

I deeply thank my parents for their unconditional and affectionate support, for their drive to make me become a happier person every day and for trusting me in my not always conventional decisions. Also, I deeply thank my brothers for being that place where I can always come back when life gets hard. I am very grateful to my girlfriend Catalina Becerra for encouraging me when I needed the most and for her dedicated demonstrations of love which always make me a happier person.

Finally I would like to thank the Chilean Fund for Scientific and Technological Research, FONDECYT, for partly financing this project (Grant 1130822).

GENERAL INDEX

Page

TABLE INDEX vii

FIGURE INDEX xii

ABSTRACT xxi

RESUMEN xxii

1 INTRODUCTION 1
- 1.1 Oenology-related lactic acid bacteria 1
- 1.2 Impact of *Oenococcus oeni* on wine quality 1
- 1.3 *O. oeni* PSU-1 and its genome 3
- 1.4 Genome-scale metabolic models as an opportunity for studying *O. oeni* metabolism 4
- 1.5 Hypothesis 5
- 1.6 Objectives 5

2 MODELING AND SCOPE 6
- 2.1 Flux Balance Analysis 6
  - 2.1.1 Mathematical representation of metabolism 6
  - 2.1.2 Steady state assumption 9
  - 2.1.3 Capacity constraints 9
  - 2.1.4 Definition and optimization of an objective function 10
- 2.2 Genome-scale metabolic models 12
  - 2.2.1 Applications of genome-scale metabolic models 13
  - 2.2.2 Limitations of genome-scale metabolic models 16

3 MATERIALS AND METHODS 18
- 3.1 Procedure 18
- 3.2 Sensitivity Analysis 20
  - 3.2.1 By varying all coefficients at the same time 20
  - 3.2.2 By varying one coefficient at time 20

3.3 Softwares and Databases 20
3.3.1 The Pathway Tools software 20
3.3.2 COBRA Toolbox 21
3.3.3 Marvin Suite 21
3.3.4 Scripts for refinement and analysis 21
3.3.5 TransportDB 22
3.3.6 The MetaCyc Database 22

4 RESULTS AND DISCUSSION 23
4.1 Generation of draft reconstruction 23
4.1.1 Obtaining the genome sequence and annotation of *O. oeni* 23
4.1.2 Identifying candidate metabolic functions and obtaining of candidate metabolic reactions 24
4.1.3 Assembly of draft reconstruction and collection of experimental data 25
4.1.4 General analysis of draft reconstruction 26
4.2 Refinement of *O. oeni* GEM 28
4.2.1 Determination and verification of substrate and cofactor usage ... 28
4.2.2 Obtaining charged formula for each metabolite 30
4.2.3 Determination of reaction stoichiometry and directionality 30
4.2.4 Reconciliation of draft reconstruction with the KEGG database .. 30
4.2.5 Adding information for gene and reaction localization 31
4.2.6 Drawing of the metabolic map 32
4.2.7 Determination of biomass composition 32
4.2.8 Adding of ATP-maintenance 34
4.2.9 Membrane transporters 35
4.2.10 Malolactic fermentation 44
4.2.11 TCA cycle 44
4.2.12 Heterolactic fermentation 46
4.2.13 General features of the GEM of *O. oeni* 50
4.3 Conversion of *O. oeni* GENRE to a computable format 54
4.3.1 Initialization of COBRA Toolbox. 54
4.3.2 Loading of reconstruction into MATLB 54
4.3.3 Setting of objective function and simulations constraints 59

4.4 Network Evaluation 60
4.4.1 Troubleshooting of non-produced biomass precursors 60
4.4.2 Elimination of reactions for which no evidence exists 67
4.4.3 Genes, reactions and compound deletion analysis 67
4.4.4 Analysis of minimal requirements 76
4.4.5 Analysis of specific consumption/production rates 95
4.4.6 Sensitivity analysis 106
4.5 Data assembly and dissemination 111

5 GENERAL CONCLUSION AND PERPECTIVES 112

ABBREVIATIONS 115

REFERENCES 116

APPENDIX 125

APPENDIXA: SUPPLEMENTARY TABLES 126

APPENDIX B: SUPPLEMENTARY FIGURES 134

TABLE INDEX

Table 4-1: **General features of the draft reconstruction of Oenococcus oeni**. The draft reconstruction contains 1040 reactions and 1219 metabolites resulting in an overdetermined system with no solution. This results from the presence of 202 internal reactions that contain more than one dead-end, causing an important increase in the number of metabolites. Many of these reactions describe metabolic processes that are not of interest, or contain unspecific metabolites. Despite that 84 exchange reactions were included in this reconstruction, several others exchange reactions were missing for important nutrients such L-malate. .27

Table 4-2: **Exchange metabolites between cytoplasm and extracellular space in O. oeni reconstruction after manual refinement**. The mechanism of transport is described in the second column for each metabolite transported. In case of multiple mechanisms of transport, all transporters are mentioned with the corresponding reference in the last column. Also, the direction of transport is described in the third column 38

Table 4-3: **Comparison between the initial draft reconstruction and the final reconstruction after manual refinement**. 53

Table 4-4: **In vivo amino acid requirements for different strains of Oenococcus oeni and comparison with in silico nutritional requirements**. Columns two to five show the number of strains that require a particular amino acid for each of the four studies analyzed. The sixth column shows the total number of strains that require a particular amino acid. If more than half of the strains require a particular amino acid, that amino acid is considered to be essential in the seventh column. Last column shows in silico nutritional requirements for each amino acid. 78

Table 4-5: ***In vivo*** **nutritional requirements for different strains of** ***Oenococcus oeni*** **and comparison with** ***in silico*** **nutritional requirements**. From column two until column 4 it is shown how many strains require a particular nutrient for each of

the 2 studies analyzed. If discrepancies exists between the two studies, results from Terrade et al., 2009 . Last column shows *in silico* nutritional requirements for each amino acid... 79

Table 4-6: **In silico minimal culture medium determined using the script FindMinimalMedium.m**. The script FindMinimalMedium.m simulates an experiment of successive transfers, returning an in silico minimal medium, i.e. a medium for which if one of the nutrients is removed, Oenococcus oeni is not able to grow. Different groups of nutrients can be found in this minimal medium such as vitamins, amino acids, carbon sources, nucleotides and mineral. 84

Table 4-7: **Alternatives for nutrients present in the in silico minimal medium.** Only four nutrients present in the in silico minimal medium could be replaced by others in order to maintain a minimal medium. Other 9 carbon sources could replace α-D-glucose.....86

Table 4-8: **In silico essential nutrients for Oenococcus oeni**. A nutrient is essential if its omission from the chemically defined medium resulted in no growth of the microorganism.....88

Table 4-9: **Summary of classification results**. True positives are essential metabolites which were classified as essential metabolites. True negatives are non essential metabolites which were classified as non essential metabolites. False positives are non-essential metabolites which were classified as essential metabolites. False negatives are essential metabolites which were classified as non essential metabolites. Most metabolites were classified correctly with the exception of alanine, lysine, leucine and threonine which were classified as essential metabolites but were not essential in vivo, and nicotinate and manganse sulfate, that were classified as not essential but were essential in vivo 93

Table 4-10: Confusion matrix of minimal medium requirement predictions. 14 essential metabolites were predicted as such (true positives) while 24 non-essential metabolites were predicted as such (true negatives). On the other hand, four non-essential metabolites were predicted as essential (false positives) and two essential metabolites were predicted as non-essential (false negatives).....94

Table 4-11: Statistical measures of performance. Sensitivity (proportion of essential metabolites predicted as such), specificity (proportion of non-essential metabolites predicted as such), precision (proportion of metabolites predicted as essential being essential in vivo) and accuracy (proportion of correct results) are indicated. The model F-1 score indicates a high overall performance 94

Table 4-12: Specific consumption rates of L-malic acid, fructose, glucose, citric acid and specific production rates of L-lactate, D-lactate and acetate determined from a batch culture of O. oeni growing at different pH levels and ethanol concentrations. Initial and final concentrations of each compound were extracted from four batch cultures using different pH levels (3.5 and 4.0) and ethanol concentrations (0% and 10% v/v). The specific rate was determined dividing the difference between final and initial concentrations by the corresponding fermentation time (final concentration was measured 24 hrs after measurement of initial concentration) and the mean biomass concentration. ..97

Table 4-13: Specific growth rates determined from a batch culture of O. oeni growing at different pH levels and ethanol concentrations. Initial and final biomass concentration was extracted from 4 batch cultures using different pH levels (3.5 and 4.0) and ethanol concentrations (0% and 10% v/v). Final concentration was measured 24 hrs after measurement of initial concentration. Growth rates were determined calculating the slope of the logarithm of biomass concentration versus culture time. 99

Table 4-14: **In silico CO2 specific production rates used to adjust fructose and glucose consumption**. These rates were determined by optimizing the genome-scale metabolic model of O. oeni constrained with four sets of restrictions corresponding to the four batch culture conditions (two different pH leves: 3.5 and 4.0 and two different ethanol concentration: 0% and 10% v/v) 100

Table 4-15: **Main in silico CO2 sources at pH 4.0, in absence (0%) and presence (10% v/v) of ethanol.** In parenthesis the percentages of these contributions to total CO2 are shown. At pH 4.0, the model predicts that the main CO2 source is malolactic reaction (28% and 36% of the total CO2, without and with ethanol respectively). Interestingly the reaction producing diacetyl is activated only when ethanol is present in the mediu. 101

Table 4-16: **Specific rates predicted by the Oenococcus oeni genome-scale metabolic model using constraints corresponding to a medium at pH 4.0 without ethanol**.....102

Table 4-17: **Specific rates predicted by Oenococcus oeni genome-scale metabolic model using constraints corresponding to a medium at pH 3.5 without ethanol**. 103

Table 4-18: **Specific rates predicted by *Oenococcus oeni* genome-scale metabolic model using constraints corresponding to a medium at pH 4.0 with ethanol**.....104

Table 4-19: **Specific rates predicted by Oenococcus oeni genome-scale metabolic model using constraints corresponding to a medium at pH 4.0 with ethanol.**.....105

Table 4-20: **Sensitivity analysis of the biomass formation equation imported from the Lactococcus lactis genome-scale metabolic model**.....108

Table A-1: **Reactions found in O. oeni KEGG metabolic maps but not in the O. oeni draft reconstruction**. Reactions that were added to the model were classified as accepted reactions while reactions that were not were classified as rejected reactions. 126

FIGURE INDEX

Figure 2-1: **Example of mathematical representation of biochemical network. Here, the matrix represents the glycolysis pathway**. Each reaction of the pathway (top) is represented in a determined column of the matrix (bottom). For example, the highlighted reaction glyceraldehyde 3-phosphate dehydrogenase (GAPD) is represented in column six. The reactants have a negative number in the matrix while products are positive. To the right, reactions of this pathway are showed as a graph where nodes represent metabolites and links represent reactions. Figure modified from Schellenberger et al (2011).....8

Figure 2-2: **Conceptual basis of constraint-based modeling**. With no constraints, the flux distribution of a biological network may lie at any point in a solution space. When mass balance constraints imposed by the stoichiometric matrix S and capacity constraints imposed by lower and upper bounds are applied to a network it defines an allowable solution space. The network may acquire any flux distribution within this space, but points outside this space are denied by the constraints. Through optimization of the objective function, FBA identifies a single optimal flux distribution that lies on the edge of the allowable solution space. Extracted from Orth et al. (2010) 12

Figure 3-1: **Overview of the procedure to iteratively reconstruct a high-quality metabolic network**. The protocol comprises 5 stages involving a total of 96 steps. In the first stage a draft reconstruction is built. This draft reconstruction is often generated automatically therefore could contain incorrect information. In the second stage each reaction in this draft reconstruction is revised manually to repair possible autogenerated errors. Additionally missing reactions, not included initially in the draft reconstruction, are added from literature and databases. In the third stage, the reconstruction is converted into a computable format in order to be evaluated in the fourth stage. In the fourth stage, *in silico* simulations of phenotypes are tested. If there are inconsistencies between *in*

silico results and literature, stage 2 must be carried out again in order to further refine the network. This iterative process ends when *in silico* results are considered to agree with experimental data. Figure modified from Thiele & Palsson (2010).....18

Figure 4-1: **Example of information contained in genome annotation**. The gene OEOE_0304 is located between the pair bases 292.470 and 293.648 in the complementary strand and encodes for a galactokinase which catalyzes the formation of alpha-D-galactose 1-phosphate from D-galactose (Enzyme Commission number 2.7.1.6).....24

Figure 4-2: **Overview of information visualized with Pathway Tools for gene OEOE_0304 of *Oenococcus oeni* PSU-1**. Gene OEOE_0304 encodes for a galactokinase (candidate metabolic function), which transforms α-D-galactose to α-D-galactose 1-phosphate (candidate metabolic reaction). Additionally, this figure shows the gene-reaction scheme representing the relationship between gene OEOE_0304, the gene product and the reaction catalyzed. Moreover the E.C number of the reaction is specified as well as the pathway it belongs to. Finally gene context is also showed. This figure was extracted from the Pathway Tools platform. 25

Figure 4-3: **Overview of genome-scale metabolic reconstruction of *Oenococcus oeni* PSU-1**. Each color represents a specific metabolic pathway. Reactions to the right are reactions that do not belong to any metabolic pathway. Nevertheless, these reactions are not necessarily disconnected from the network.....32

Figure 4-4: **Representation of transporters incorporated into *Oenococcus oeni* reconstruction for acetate (A), (S)-lactate (B), ethanol (C), (R)-lactate (D), 2,3-butanediol (E) and diacetyl (F)**. Created with Pathways Toolss 37

Figure 4-5: **Malolactic stoichiometric equation included in the genome-scale metabolic model of *Oenococcus oeni***. (S)-malate (or L-malate) is converted to (S)-lactate (or L-lactate) and CO_2. Made with Pathways Tools.....44

Figure 4-6: **Defective TCA cycle of *Oenococcus oeni.*** Genes associated to malate dehydrogenase (E.C. 1.1.1.37) and fumarase (E.C. 4.2.1.2) were the only ones found. All the other TCA reactions lack associations with genes. 45

Figure 4-7: **Reactions added to the *Oenococcus oeni* reconstruction to complete the heterolactic fermentation pathway**. The transformation of acetoin to 2-3 butanediol (A) and the transformation of diacetyl to acetoin (B) were added to the reconstruction to complete this pathway. Two isoenzymes are responsible of carrying out these reactions. Genes encoding these reactions were found in KEGG database.....47

Figure 4-8: **First half of the heterolactic fermentation**. Hexoses (β-D-fructofuranose and β-D-glucose) are converted to two compounds with 3 carbons: acetylphosphate and D-glyceraldehyde. Made with Pathways Tools.....48

Figure 4-9: **Second half of the heterolactic fermentation**. Citrate is converted to oxaloacetate through citrate lyase. The main end products of this pathway are generated from pyruvate: D-lactate, diacetyl and 2,3-butanodiol. Made with Pathways Tools 49

Figure 4-10: **Comparison of metabolite connectivity of *Oenococcus oeni* draft reconstruction (A) and *Oenococcus oeni* reconstruction after manual refinement (B)**. The number of metabolites participating in just one reaction diminishes from 631 to 234 after manual refinement. This resulted from elimination of non specific and disconnected reactions that do not contribute to describe the *Oenococcus oeni* metabolism. 51

Figure 4-11: **Metabolite connectivity in GEM of *Escherichia coli*.** It can be appreciated that many metabolites participate in a small number of reactions while few metabolites participate in a large number of reactions which is the general behavior of high quality GEMs. There are 87 metabolites participating in just one reaction and 991 metabolites participating in two reactions. The reason for the large number of metabolites participating in two reactions is because 480 metabolites are involved in exchange reactions, including transport between extracellular space, periplasm and cytoplasm. On the other hand, there are few metabolites participating in a large number of reactions such as the proton that participates in 1031 reactions. 52

Figure 4-12: **Fields contained in the *Oenococcus oeni* model once it is loaded using COBRA Toolbox**. The model contains several fields describing reactions and metabolites present in the model but lacks important fields such as genes, rules, grRules and rxnGeneMat. 55

Figure 4-13: **Extract of rxnNames (left), grRules (center) and rules (right) MATLAB arrays corresponding to the fields of *O. oeni* GE**M. 57

Figure 4-14: **Comparison of fields contained in the *Oenococcus oeni* model before and after running the scripts to add the fields metFormulas, genes, rules, grRules and rxnGeneMat**. The model loaded into Matlab without any later change lacked important fields such as metFormulas, genes, rules, grRules and rxnGeneMat (left). These fields were added to the model running the scripts *GenerateGeneField.m, GenerateRulesFields.m and GenerateRxnGenMat.m* and *SeparateNamesAndFormulas.m* (right). 59

Figure 4-15: **Algorithm used for identifying non-produced metabolites in case of having a model with biomass equation but not functional**. First a demand reaction is created for the metabolite of interest. Then, the flux for that reaction is

maximized. If the flux for the demand reaction is zero, for each substrate in the equation producing the metabolite of interest, a demand reaction is added and maximized. If the flux for the later demand reaction is not zero for the substrate *i*, the substrate *i* is labeled as produced metabolite. Otherwise, the substrate *i* is labeled as non produced metabolite and is analyzed if there are reactions producing the substrate *i*. If there are not reactions producing the substrate *i*, the substrate is labeled as root non produced metabolite. Otherwise, it is analyzed if there is flux through any of the reactions producing the substrate *i*. If there is no reaction having flux, the algorithm analyzes the substrate *i* in a recursive way. The algorithm returns the produced metabolites, the non produced metabolites and the root non produced metabolites that must be reviewed in order to make the model functional. 63

Figure 4-16: **Example of an unfeasible metabolic network caused by the lack of a consuming reaction**. In this example, the reaction labeled as *Rxn* consumes metabolites A and B and produces metabolites C and D. Despite the reactions supplying substrates A and B being present, if the production of D is maximized, there will be no flux through the reaction *Rxn* because there is no reaction that consumes C. In this case, the network is unfeasible and there will be no solution when FBA is applied because mass balances cannot be accomplished. 64

Figure 4-17: **Reactions found in MetaCyc leading to the production of 2-succinylbenzoate**. The algorithm *SearchMetaCycReactions.m* search in MetaCyc for reactions producing the metabolism of interest and then the results are showed with a node graph. Blue indicates the target metabolite, in this case 2-succinylbenzoate. Red indicates metabolites not present in the model (the names of these nodes correspond to the MetaCyc IDs). Purple indicates reactions utilizing metabolites not present in the model. Yellow indicates metabolites present in the model. Green indicates reactions utilizing metabolites present in the model. 66

Figure 4-18: **Reaction deletion analysis for genome-scale metabolic model of *Oenococcus oeni***. Most reactions (81%) have no effect on growth as was expected based on the fact that metabolic network of microorganisms are robust enough to maintain growth even when a particular metabolic function is deleted. On the other hand, 15% of the reactions are considered to be essential as their individual deletion causes no growth. In the middle, only 3% of reactions slightly decrease growth (90-100%) and 1% of reactions strongly decrease growth (0-90%). These two last groups of reactions are interesting as it could provide insight in order adjust predicted growth rate 71

Figure 4-19: **Essential reactions organized according to metabolic processes**. The most fragile metabolic processes susceptible to disruption by removal of a particular reaction are those related to fatty acid biosynthesis 72

Figure 4-20: **Gene deletion analysis for genome-scale metabolic model of *Oenococcus oeni***. Most genes (96%) have no effect on growth as was expected based on the fact that the metabolic network of microorganisms is robust enough to maintain growth even when a particular metabolic function is deleted. On the other hand, 3% of the genes are considered to be essential as their individual deletion causes no growth. In the middle, only 1% of genes slightly decrease growth (90-100%) and less than 1% of genes strongly decrease growth (0-90%). These two last groups of reactions are interesting as they could provide insights on adjusting predicted growth rate.....73

Figure 4-21: **Essential reactions organized according to metabolic processes**. The most fragile metabolic processes susceptible to be disrupted through removal of a particular reaction are those related to fatty acid biosynthesis. Heterolactic fermentation appears as a sensitive metabolic process that could be affected by the deletion of genes. 74

Figure 4-22: Compound deletion analysis for genome-scale metabolic model of *Oenococcus oeni*. Most compounds (72%) have no effect on growth as was expected based on the fact that metabolic networks of microorganisms are robust enough to maintain growth even when a particular component is deleted. In this case the percentage of essential compounds is smaller than the percentage of essential reactions or genes because in most causes a particular compound is involved in more than one reaction and its deletion causes the elimination of all the reactions that participate. On the other hand 24% of the compounds are considered essential as their individual deletion causes no growth. In the middle, only 3% of compounds slightly decrease growth (90-100%) and 1% of compounds strongly decrease growth (0-90%). These two last groups of reactions are interesting as they could provide insight adjusting predicted growth rate 75

Figure 4-23: Algorithm implemented in MATLAB to determine minimal medium through the simulation of successive transfers. Before running the algorithm the model must be able to produce biomass in a chemically defined medium. The algorithm consists in removing each substrate one at a time and in turn. Once a particular substrate is removed the model is optimized and the growth is analyzed. If the model is able to produce biomass the substrate is considered to be a non-essential nutrient and the next substrate is removed. Otherwise, the substrate is considered to be an essential nutrient and the substrate is reincorporated into the medium before removing the next substrate. The algorithm finishes when all the substrates are analyzed and it returns the minimal medium for which the suppression of growth is achieved if one of the nutrients is removed 83

Figure 4-24: Algorithm implemented in MATLAB to determine alternatives nutrients of minimal nutrient requirements. Because the *in silico* minimal medium determined with the script *FindMinimalMedium* depends on the order in which the nutrients are omitted, alternative nutrients are determined replacing each

of the essential substrates for each of the nutrients that were not considered to be essential. If the model is able to produce biomass replacing substrate i for substrate j, the substrate j is an alternative for substrate i. The algorithm finishes when all substrates are analyzed and it returns the alternatives for each nutrient present in the minimal medium. 83

Figure 4-25: **Partial lysine biosynthesis pathway**. Here, the last five reactions of the lysine biosynthesis pathway are shown. None of these reactions are able to carry flux when the model is optimized because the model lacks feasible reactions producing succinyl-CoA and consuming conenzyme A and 2-oxoglutarate. Sink and demand reactions were tested for these metabolites showing that their incorporation allows the biosynthesis of L-lysine. Unfortunately their incorporation also diminishes the accuracy of specific consumption/production rates predictions so reactions whose incorporation makes L-lysine a non essential metabolites while maintaining the same accuracy of specific consumption/production rate predictions remains a pending challenge. This figure was made using Pathway Tools.....90

Figure 4-26: **Excel spreadsheet showing real and predicted consumption/production rates of different metabolites and real and predicted specific growth rates**. Percentage differences are also presented in order to visualize errors.....100

Figure B-1: **Logarithm of biomass concentration versus culture time**. Input data for this plot is found in Table 4-14. The slope of each trendline represents the growth rate for each of the batch culture conditions recorded in Table 4-14.....134

Figure B-2: **Growth and metabolic monitoring of *O. oeni* PSU-1 at pH 4.0 (squares) and pH 3.5 (circles), in the absence (solid lines and filled symbols) and presence of 10% ethanol (dashed lines and empty symbols)**. A: glucose, B: fructose, C: citrate, D: L-malic acid, E: acetic acid, F: L-lactic acid, G: D-lactic-acid, H: biomass, dry

weight. Red points represent the data extracted with WebPloAnalyzer. Figure modified from Olguín et al, (2009). 135

ABSTRACT

Oenococcus oeni is one of the most important lactic acid bacteria that develops during winemaking. It is responsible of carrying out malolactic fermentation, which significantly affects the final quality of wine. However, the completion of this endeavor is erratic and, so far, no conclusive explanation has been given to explain this behavior.

In this thesis, I constructed the first Genome-scale metabolic model (GEM) of *O. oeni* based on genome annotation, database information and primary literature. The model includes 914 reactions, 792 metabolites and 512 genes. An *in silico* minimal growth medium was determined and compared with experimental data. With a total of 44 growth/non growth experiments, the model showed a prediction accuracy of 86%, with an F-score value of 0.23, indicating a high overall performance. Furthermore, MATLAB scripts developed during this work to facilitate refinement and analysis of the *O. oeni* GEM proved to be efficient, useful and user friendly.

The model was employed to predict specific growth rates and consumption/production rates of glucose, fructose, citric acid, L-malic acid, L-lactic acid, D-lactic acid and acetate, under different medium conditions of pH and ethanol concentration. In the absence of ethanol, a mean difference of 38% was found between the experimental and predicted specific growth rates, while only a 0.1% mean difference was obtained for specific consumption/production rates. Unfortunately, less accurate predictions were achieved under ethanol growing conditions; the latter will be tackled in the future by determining gene expression and biomass composition of *O. oeni*, growing under increasing ethanol concentrations.

All in all, the construction and partial validation of this GEM of *O. oeni* represents a step forward in the comprehensive understanding, and the consequent control, of malolactic fermentation during winemaking.

Keywords: *O. oeni*, GEM, minimal medium, consumption/production rate prediction, winemaking, refinement tools.

RESUMEN

O. oeni es la bacteria láctica más importante en la vinificación dado que es responsable de llevar a cabo la fermentación maloláctica, proceso que afecta significativamente la calidad final del vino. El comportamiento de esta bacteria es, sin embargo, errático, y se desconocen las causas de dicho fenómeno .

En este trabajo, se construyó el primer modelo metabólico a escala genómica (GEM) de *O. oeni* basado en su anotación genómica, múltiples bases de datos y literatura primaria. El modelo incluye 914 reacciones, 792 metabolitos y 512 genes. Un medio mínimo *in silico* fue determinado y comparado con datos experimentales. Con un total de 44 experimentos de crecimiento/no crecimiento, el modelo muestra una exactitud de 86% en sus predicciones y un F-score de 0.82, sugiriendo su alto rendimiento. Por otro lado, los scripts desarrollados en MATLAB para facilitar el refinamiento y análisis del GEM de *O. oeni* han demostrado ser eficientes, útiles y amigables con el usuario.

El modelo fue empleado para predecir tasas de crecimiento específico y tasas de consumo/producción de glucosa, fructosa, ácido cítrico, ácido L-málico, ácido L-láctico, ácido D-láctico y acetato bajo diferentes condiciones de pH y toxicidad de etanol. En ausencia de etanol, se observó una diferencia porcentual promedio de 38% entre la tasa específica de crecimiento experimental y predicha; y una diferencia porcentual promedio de 0.1% para las tasas de consumo/producción. Desafortunadamente, las predicciones obtenidas para estas variables fueron menos precisas cuando *O. oeni* es crecido con etanol, lo cual podría mejorarse determinando expresión génica y composición de la biomasa de *O. oeni* crecida en presencia de etanol. La construcción y validación parcial de este GEM de *O. oeni* representa un paso importante para la mejor compresión y consecuente control de la fermentación maloláctica durante la vinificación

Palabras Claves: *O. oeni*, modelo metabólico a escala genómica, medio mínimo, predicción de tasas de consumo/producción, vinificación, herramientas de curación.

1 INTRODUCTION

1.1 Oenology-related lactic acid bacteria

It is common to associate yeasts with winemaking. They are responsible for the main change in this process, i.e. the transformation of the sugars present in grape must into ethanol, better known as primary fermentation (Swiegers, Bartowsky, Henschke, & Pretorius, 2005). However, the microbiology of wine involves much more than just yeasts. Lactic Acid Bacteria (LAB) are also present in grapes and must. They metabolize numerous substrates, playing an important role in the final quality of wine (Ribéreau-Gayon, Dubourdieu, Donèche, & Lonvaud, 1998).

Five genera of LAB can be found in grape must: *Lactobacillus, Pediococcus, Leuconostoc, Oenococcus* and *Weissella.* While *Pediococcus* are homofermentative, *Leuconostoc, Oenococcus* and *Weissella* are obligate heterofermentative. Interestingly, *Lactobacillus* could be homofermentative (e.g. *Lb. vini*), facultative heterofermentative (e.g. *Lb. plantarum*) or obligate heterofermentative (e.g. *Lb. brevis*) (König, Unden, & Jürgen, 2009; Lonvaud-Funel, 1999). Until 1995, *Leuconostoc* spp were classified as *Leuconostoc mesenteroides* and *Leuconostoc oenos*. In 1995, a study based on 16S r-DNA and 23S r_DNA proposed the creation of a new species, *Oeonoccus oeni*.

1.2 Impact of *Oenococcus oeni* on wine quality

Oenococcus oeni is the predominant species at the end of primary fermentation and after. Its remarkable metabolism makes it resistant to high ethanol concentrations (15% v/v), low pH (as low as 2.9), limited nutrient availability (Bartowsky, 2005) and high SO_2 concentrations (50 ppm) (Bauer & Dicks, 2004). These characteristics make *O. oeni* the main organism responsible of carrying out malolactic fermentation (MLF), although all wine-related LABs are capable of such.

The MLF (or secondary fermentation) is responsible for the decarboxylation of L-malic acid to L-lactic acid and carbon dioxide, which has three main consequences. Firstly, MLF results in the deacidification of wine, increasing pH in 0.1-0.2 units and decreasing titratable acidity. Secondly, MLF confers microbial stability to the wine through the removal of malic acid, which could act as a carbon source for other microorganisms. Finally, MLF impacts on sensory properties that could be perceived through aroma and palate (Bartowsky, 2005). MLF generally occurs once the primary fermentation has finished, but could also start earlier, or even several months after. The reasons for this not yet clear.

Which of the three consequences of MLF is more important to wine quality depends on the climate. This is because L-malic acid concentrations vary with the climate of the region from where grapes are sourced. Thus, in cool regions, the concentration of L-malic acid in the grape juice is between 2 and 5 g/L; whereas in warm regions, the concentrations often do not exceed more than 2 g/L (Sponholz, 1989, Zoecklein et al 1990). In cooler regions, the deacidification process is regarded as the most important modification associated with MLF, while in warmer regions, deacidification is not that important, and the changes in sensory profile become the most important feature (Lerm, Engelbrecht, & Toit, 2010).

MLF contributes directly to sensory changes by replacing the strong green apple taste of malic acid by the less aggressive taste of lactic acid. However, more subtle changes also occur. The major change in aroma during MLF is related with diacetyl biosynthesis. This compound imparts a "buttery" aroma and flavor to wine, which can easily be perceived by sensory panelists. Other compounds that significantly affect wine taste are acetic acid and acetoin compounds, all of which are products of citrate degradation (Bartowsky & Henschke, 2004; Lonvaud-Funel, 1999).

1.3 *O. oeni* PSU-1 and its genome

The *O. oeni* PSU-1 strain was isolated from Pennsylvania in 1972 and characterized for the first time in 1977. This strain was found to be quite similar in most biochemical characteristics to the previously characterized strain ML-34, a commonly studied strain isolated from California. The main advantage of PSU-1 over ML-34 strain is its capacity to induce malolactic fermentation more quickly in red wines (Beelman, III Gavin, & Keen, 1977). Since then, it has become a common strain to be as starter culture.

The physical map of the strain PSU-1 was studied intensively at the end of the 1990s (Zé-Zé, Tenreiro, Brito, Santos, & Paveia, 1998; Zé-Zé, Tenreiro, & Paveia, 2000). In 2005, the genome sequence and annotation of this strain was determined and released for public access (Mills, Rawsthorne, Parker, Tamir, & Makarova, 2005). Some general features of the genome of *O. oeni* PSU-1 are its single circular chromosome of 1.780.517 nucleotides (nt), and a G + C content of 38%. The precise annotation of the *O. oeni* genome allowed for the study of the metabolic pathways of this organism. It is worth noting that the PSU-1 strain was the first strain to be sequenced

More recently, Borneman et al. (2012) carried out a comparative analysis of the *O. oeni* pan genome, using the strain PSU-1 as a reference, reporting important differences between the strains. On average, each strain was predicted to contain 1800 ± 52 ORFs and 104 ± 25 potential pseudogenes. The *O. oeni* pan genome comprises 2846 ORFs, while the core genome contains 1165.

1.4 Genome-scale metabolic models as an opportunity for studying *O. oeni* metabolism

Metabolism is essentially a large network of chemical conversions catalyzed mostly by enzymes. In this process, nutrients are converted into building blocks, such as fatty acids, nucleotides, and amino acids, for the synthesis of macromolecules, such as lipids, DNA and proteins. The latter are fundamental pieces for the maintenance of cellular integrity and formation of new cells. The reactions occurring in the cell essentially consist of the transformation of substrates into products, while mass and charge are balanced in each side of the equation (Maarleveld, Khandelwal, Olivier, Teusink, & Bruggeman, 2013).

Metabolic networks at the genome scale (so called GENREs or genome-scale metabolic reconstructions) are built systematically using genome annotation, knowledge databases, "omics", and primary literature. Therefore, GENREs provide the best representation of the metabolic capabilities of the target organism at the time of reconstruction (Monk, Nogales, & Palsson, 2014). Integrating this information in a structured fashion has enabled its translation into computational models, called genome-scale metabolic models or GEMs, that can be used to calculate metabolic phenotypes (McCloskey, Palsson, & Feist, 2013). Thus, the release of the genome annotation of *O. oeni* PSU-1 represents an opportunity to generate a high-quality GEM that allows for the calculation of possible phenotypic states in order to achieve a comprehensive understanding of its metabolism.

1.5 Hypothesis

The development of a genome-scale metabolic model of *Oenococcus oeni* PSU-1 and its analysis through flux balance analysis will allow the identification and evaluation of nutritional requirements and specific metabolic rates of this bacterium

1.6 Objectives

Consistent with the proposed hypothesis, the overall objective of this thesis is to develop a genome-scale metabolic model of the malolactic bacterium *Oenococcus oeni* capable of properly identifying and evaluating the nutritional requirements and specific rates of this bacterium. The specific objectives, in agreement with the five stages of the Thiele and Palsson 's protocol for generating a high-quality genome-scale metabolic model, are the following:

1. Generation of an automatic draft reconstruction containing candidate metabolic reactions of *Oenococcus oeni* PSU-1.
2. Refinement by correcting errors in *Oenococcus oeni* PSU-1 draft reconstruction generated in objective one.
3. Conversion of refined reconstruction into a format suitable for modeling.
4. Network evaluation of *Oenococcus oeni* PSU-1 genome-scale metabolic model through comparison with reported nutritional requirements, specific growth rates and specific consumption/production rates.
5. Data assembly and dissemination of *Oenococcus oeni* PSU-1 genome-scale metabolic model.

2 MODELING AND SCOPE

2.1 Flux Balance Analysis.

Flux balance analysis (FBA) is a widely used mathematical modeling approach to quantitatively simulate microbial metabolism (Kauffman, Prakash, & Edwards, 2003). Since its validation as a predictive tool (Varma & Palsson, 1994), FBA has been used in a wide number of applications (Raman & Chandra, 2009), such as prediction of bacterial phenotype under different environmental conditions (Orth, Thiele, & Palsson, 2010) and industrial strain improvement by metabolic engineering (Koffas & Stephanopoulos, 2005). The principles behind the FBA technique are explained below.

2.1.1 Mathematical representation of metabolism

Metabolic reactions can be represented by a stoichiometric matrix **S**, in which every row represents one unique metabolite; and every column represents one reaction (Figure 2-1) The entries in the column j are the stoichiometric coefficients of the metabolites participating in the reaction j with negative coefficients for every metabolite consumed, positive coefficients for every metabolite that is produced and coefficients equal to zero for every metabolite that does not participate in that reaction. For a biochemical network containing *n* reactions and involving *m* metabolites, the size of matrix S will be *m* x *n* (Equation 1)

Equation 1. Stoichiometric matrix of a biochemical network involving n reactions and m metabolites. Each element S_{ij} represent the stoichiometric coefficient of metabolite i participating in reaction j.

$$S = \begin{pmatrix} s_{11} & \cdots & s_{1j} & \cdots & s_{1n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ s_{i1} & \cdots & s_{ij} & \cdots & s_{in} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ s_{m1} & \cdots & s_{mj} & \cdots & s_{mn} \end{pmatrix}_{mxn}$$

The metabolic flux through all the reactions in a network can be represented by a vector $\boldsymbol{v}$, which has a length of *n* (Equation 2) the entry in the row j is the flux in $\frac{mmol}{gDW\ hr}$ of the reaction j (Orth et al., 2010)

Equation 2. Vector representing metabolic fluxes in a biochemical network of n reactions. The element V_j represent the flux through reaction j.

$$v = \begin{pmatrix} v_1 \\ \dots \\ v_j \\ \dots \\ v_n \end{pmatrix}_n$$

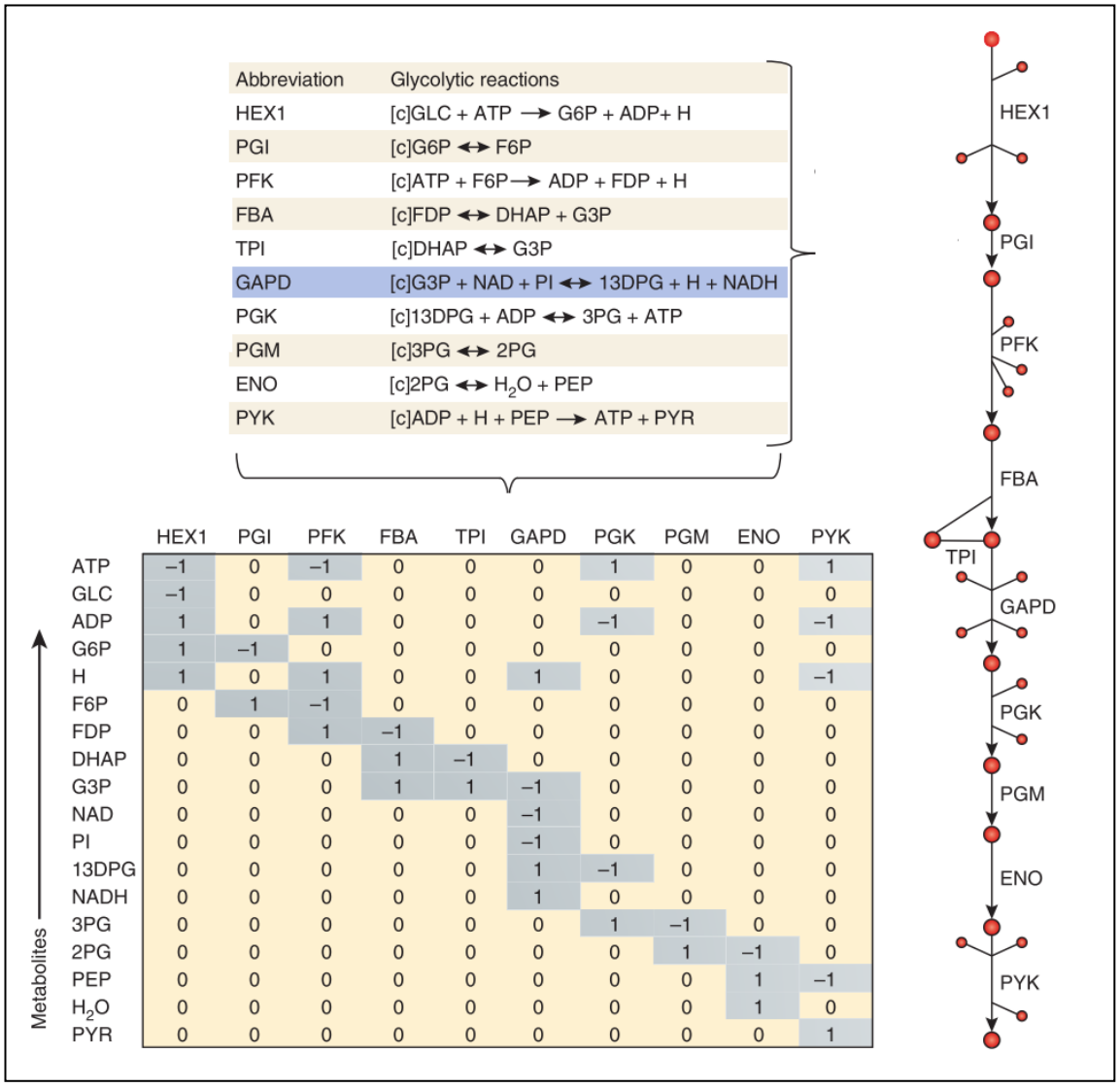

Figure 2-1. Example of mathematical representation of biochemical network. Here, the matrix represents the glycolysis pathway. Each reaction of the pathway (top) is represented in a determined column of the matrix (bottom). For example, the highlighted reaction glyceraldehyde 3-phosphate dehydrogenase (GAPD) is represented in column six. The reactants have a negative number in the matrix while products are positive. To the right, reactions of this pathway are showed as a graph where nodes represent metabolites and links represent reactions. Figure modified from Schellenberger et al (2011).

2.1.2 Steady state assumption

Concentrations of all metabolites can be represented by a vector $\boldsymbol{x}$, which has a length of *m.* Steady-state implies that variations of metabolites concentrations in time are equal to zero $\left(\frac{dx}{dt} = 0\right)$, i.e. the sum of fluxes consuming a particular metabolite must be equal to the sum of fluxes producing that metabolite (Equation 3)

Equation 3. Steady state equation

$$\boldsymbol{S} * \boldsymbol{v} = \begin{pmatrix} s_{11} & \dots & s_{1j} & \dots & s_{1n} \\ \dots & \dots & \dots & \dots & \dots \\ s_{i1} & \dots & s_{ij} & \dots & s_{in} \\ \dots & \dots & \dots & \dots & \dots \\ s_{m1} & \dots & s_{mj} & \dots & s_{mn} \end{pmatrix}_{mxn} * \begin{pmatrix} v_1 \\ \dots \\ v_j \\ \dots \\ v_n \end{pmatrix}_{n} = \mathbf{0}$$

Since the reactions of the biochemical network under study and their stoichiometric coefficients are known, matrix **S** can easily be constructed. Therefore, the only unknown variable in equation 3 is $\boldsymbol{v}$. Any $\boldsymbol{v}$ that satisfies this equation is said to be in the null space of **S.** This set of equations is also known as mass balance constraints, which ensure that the total amount of any compound being produced must be equal to the total amount being consumed at steady state.

In a GEM, there are usually more reactions than compounds (n > m) in the biochemical network, meaning that there are more unknown variables than equations. These systems are known as underdetermined, because there is no unique solution that satisfies these equations.

2.1.3 Capacity constraints

Besides the mass balance constraints given by equation 3, capacity constraints could be defined to further restrict the solution space and establish a context-specific condition. Capacity constraints are defined as the range of flux values that biochemical reactions can reach. There are two variables that determine capacity constraints: lower and upper

bounds. The entry in the row j of the lower bound vector represents the minimum flux value which can be reached by reaction j, while the entry in the row j of the upper bound vector represents its maximum value. For a biochemical network with n reactions, lower and upper bounds are vectors of size n, as shown by equation 4:

Equation 4. Lower and upper bounds vectors.

$$\begin{pmatrix} l_1 \\ \dots \\ l_j \\ \dots \\ l_n \end{pmatrix}_n \leq \begin{pmatrix} v_1 \\ \dots \\ v_j \\ \dots \\ v_n \end{pmatrix}_n \leq \begin{pmatrix} u_1 \\ \dots \\ u_j \\ \dots \\ u_n \end{pmatrix}_n$$

In cases of known experimental bounds for a particular reaction, setting lower and upper bounds for that reaction will restrict the solution space. Furthermore, if there is evidence that a reaction occurs from left to right, the lower bound must be set to zero and the upper bound must be set to infinity. In practice, an upper bound equal to 1000 is big enough to represent infinity. On the other hand, if there is evidence that a reaction occurs from right to left the lower bound must be set to negative infinity; and the upper bound to zero. In a similar way, a lower bound equal to – 1000 is enough to represent negative infinity. Additionally, in cases where the exact value of flux through the reaction j is known, the row j of both vectors must be set to that value (Orth et al., 2010).

2.1.4 Definition and optimization of an objective function

As the system of equations has infinite solutions, FBA seeks to optimize an objective function in order to obtain a unique solution. However, a unique solution cannot be guaranteed, as it is common for multiple solutions from the same value of the objective function, or even for no values to satisfy the balance or capacity constraints.

The objective function can be defined as $Z = c^T v$, where c is a vector of weights indicating how much each reaction (v) contributes to the objective function. When only

one reaction is desired for maximization, c is a vector of zeros with a one at the position of the reactions of interest. It is common to maximize the biomass reaction when simulating growth. In such a case, the vector c will have a one at the position of the biomass reaction (Orth et al., 2010). The different types of objective functions and the progress made in this topic has been reviewed by Feist & Palsson (2010)

Then the FBA problem can be written as

$$Maximize\ Z = c^T v$$
$$subject\ to:$$
$$S\ v = 0$$
$$lb\ \leq v \leq ub$$

The succession of steps mentioned above seek to constrain the solution space in order to find a context-specific unique flux distribution (Figure 2-2). With no mass balance constraints, fluxes are able to reach any point in the solution space. This solution obviously lacks biological interpretation, because organism-specific reactions are not incorporated into the problem. The application of mass balance constraints, given by the equation $S\ v = 0$, restricts the solution space. Nevertheless, the use of mass balances is not enough to describe a context-specific experimental condition, because reaction fluxes are allowed to reach values that are not achieved under real experimental conditions. The application of capacity constraints given by lower and upper bounds, allows for a restricted solution space in which a solution with biological meaning is found. This space is known as allowable solution space. The optimization of an objective function, restricted by the allowable solution space, results in a single flux distribution which lies on the edge of the allowable solution space, whih can be interpreted with a biological perspective.

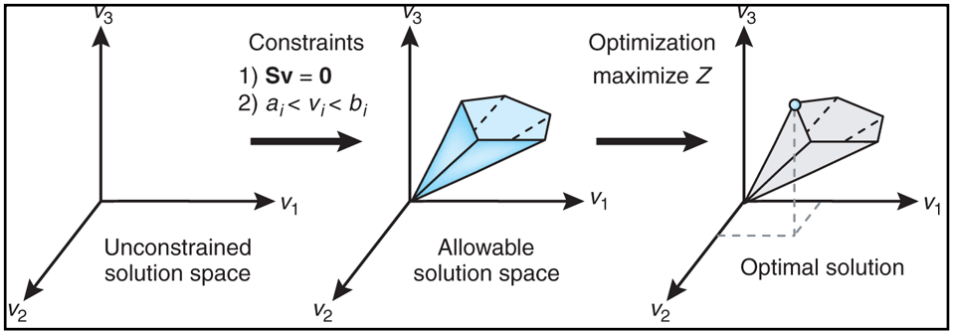

Figure 2-2. Conceptual basis of constraint-based modeling. With no constraints, the flux distribution of a biological network may lie at any point in a solution space. When mass balance constraints imposed by the stoichiometric matrix S and capacity constraints imposed by lower and upper bounds are applied to a network it defines an allowable solution space. The network may acquire any flux distribution within this space, but points outside this space are denied by the constraints. Through optimization of the objective function, FBA identifies a single optimal flux distribution that lies on the edge of the allowable solution space. Extracted from Orth et al. (2010)

2.2 Genome-scale metabolic models.

A genome-scale network reconstruction (GENRE) is an organism-specific collection of biochemical reactions and associated genes that describes its metabolism (Thiele & Palsson, 2010). Since the first GENRE, created in 1999 (Edwards & Palsson, 1999), 117 GENREs have been published (Monk et al., 2014) for diverse organisms, including important cell factories such as *Escherichia coli* (Orth et al., 2011) and *Saccharomyces cerevisiae* (Heavner, Smallbone, Barker, Mendes, & Walker, 2012). These GENREs are constructed from genome annotation, databases and primary literature (Feist, Herrgård, Thiele, Reed, & Palsson, 2009). At present, a rigorous protocol of 96 steps has been set up to generate a high quality GENRE, capable of correctly predicting the metabolism of a particular microorganism (Thiele & Palsson, 2010).

A genome-scale metabolic model (GEM) is a mathematical representation of GENREs that allows the phenotype of an organism to be studied *in silico.* In other words, GEMs are a structured format of different types of biological knowledge that can be used to perform computational and quantitative queries to answer questions about the capabilities of organisms and their likely phenotypic states.

2.2.1 Applications of genome-scale metabolic models

GEMs have been used to answer different types of questions. Most fall into one of the six categories described below (McCloskey et al., 2013).

2.2.1.1 Metabolic engineering

Metabolic engineering is the rational design to convert cell factories into highly efficient, focused machines capable of generating significant quantities of a molecule of interest (Lee et al., 2012). This is achieved by altering metabolic fluxes to increase the concentrations, yields and productivities of innate and non-innate bio-products (Na, Kim, & Lee, 2010). In this application, GEMs can be used to identify metabolic targets to optimize the cell factory, which can then be implemented *in vivo*. These strategies include gene deletions, gene over- and under-expression, mapping high throughput data onto network reconstruction to identify bottlenecks or competing pathways and integration of non-native pathways into standard microbial production hosts for production of compounds that are either natively found in, or only synthesized, in minute concentrations by the host (McCloskey et al., 2013).

For example, Sohn et al (2010) built a GEM for *Pichia pastoris*, a methylotrophic yeast that has gained much attention during the last decade as a platform for producing heterologous recombinant proteins of pharmaceutical importance (Caspeta & Nielsen, 2013; Caspeta, Shoaie, Agren, Nookaew, & Nielsen, 2012). They incorporated equations describing the production of two heterologous proteins, human serum albumin and human superoxide dismutase and investigated how the oxygen supply affected capabilities in producing these two proteins. As a result, they found that under oxygen limiting conditions the rate of protein production is maximized, without decreasing growth rate; therefore, fermentation strategies could be designed to optimize protein production and growth rate by limiting the oxygen supply.

2.2.1.2 Biological discovery

Several features of bacterial functions still remain uncharacterized. For example, even in *E. coli*, the most studied and well known bacterium, 34% of the genes have an unknown function (Orth et al., 2011). The function of uncharacterized open reading frames (ORFs) can be elucidated using GEMs by comparing growth phenotypes from *in silico* predictions to *in vivo* experimental data. Discrepancies between GEM prediction and experimental results can point to where current knowledge is missing or where there are discrepancies. Thus, GEMs allow one to systematically formulate testable hypotheses (McCloskey et al., 2013).

As an example of biological discovery, Kim *et al* (2011) reconstructed a GEM of *Vibrio vulnificus*, an opportunistic pathogen that causes primary septicemia, necrotized wound infections and gastroenteritis in humans. This pathogen is considered the major causative agent of death from ingestion of raw or undercooked seafood. The model developed was employed to predict essential metabolites and genes interacting with these metabolites. Based on these results, an antibiotic was developed by finding compounds with similar structure to the essential metabolites. Thus, this study showed a strategy for discovering novel antibiotics and drugs based on systems-level analysis of metabolic networks.

2.2.1.3 Phenotypic functions

Understanding and predicting the phenotypic potential of microorganisms is another GEM application (Hyduke, Lewis, & Palsson, 2013). Constraint-based modeling with GEMs has allowed to rapidly predict growth in various conditions, explore different objectives of microbial metabolism to examine the driving force behind cellular function, and understand the suboptimal behavior of cells following perturbation and latent pathway activation (McCloskey et al., 2013). O'Brien, Lerman, Chang, Hyduke, & Palsson (2013) released a new version of an *Escherichia coli* model, integrating

metabolic and gene product expression pathways. The model is able to accurately predict multi-scale phenotypes, ranging from coarse-grained (growth rate, nutrient uptake, by-product secretion) to fine-grained (metabolic fluxes, gene expression levels) phenotypes.

2.2.1.4 Biological network analysis

The metabolic reaction network is a highly complex, interwoven system that responds to environmental and genetic perturbations. In order to elucidate and understand the relationship between the network structure and function, researchers have turned to network analysis. This exercise is mathematical in nature. In network analysis, biochemical reactions are transformed into a graph, where the nodes and links take the form of metabolites and enzymatic reactions. Once formulated as a graph, the network can be sampled and explored to arrive at biologically insightful conclusions (McCloskey et al., 2013). Kun et al (2008) searched for autocatalytic components (compounds which are needed to its own synthesis) in the metabolic network of 10 organisms. They found that all metabolic networks have at least one universal autocatalytic molecule, ATP, supporting the view that a small, but important part of inheritance is provided by the set of autocatalytic compounds of intermediary metabolism.

2.2.1.5 Bacterial evolution

Bacterial species are constantly adapting to meet the demands of the imposed environmental conditions. Nevertheless, due to time limitations, the study of evolutionary processes, through computational frameworks, is perhaps the easiest way to elucidate principles governing evolution. GEM of *E. coli* was used to study the contribution of different genetic mechanisms to network growth and the selective forces driving network evolution (Pál, Papp, & Lercher, 2005). Results suggested that most changes to the *E. coli* metabolic network during the past 100 million years are due to horizontal gene transfer, with little contribution from gene duplicates. Additionally,

network growth resulted by acquiring genes involved in the transport and catalysis of external nutrients, driven by adaptations to changing environments.

2.2.1.6 Cell interactions

Few cells grow in pure cultures in nature. Therefore, the study of interspecies interactions allows for a better description of cells growing in their natural state. Furthermore, in many cases, the most interesting phenotypes emerge when particular species interact. GEMs are being applied to evaluate these multi-cell interaction challenges. Thus, a model of the syntrophic bacteria *Desulfovibrio vulgaris* and *Methanococcys maripaludis* has been developed for studying the mutualistic microbial community composed by these two species (Stolyar et al., 2007). This model includes the central metabolism of both bacteria and allows to accurately predict several ecologically relevant characteristics, including the ratio of *D. vulgaris* and *M. maripaludis* cells during co-culture (Oberhardt, Palsson, & Papin, 2009).

Another example in this field is the study of interactions between different types of human cells. Lewis et al. (2010) examined the interactions between astrocytes and cholinergic neurons in brain tissue to gain insights into Alzheimer's disease. They used three different models derived from RECON1, a human GENRE, to represent three different cell types and included exchange reactions between different cell types to simulate the interactions that these cells have within the brain tissue. Thus, new insights into the mechanisms of disease and strategies for treating the disease were formulated (T. Y. Kim, Sohn, Kim, Kim, & Lee, 2012).

2.2.2 Limitations of genome-scale metabolic models

GEMs also have limitations. For example, due to inherent mathematical representation of FBA, GEMs lack a representation for metabolite concentrations. This is a big issue as, in many cases, metabolite concentration is an important parameter for process optimization. For example, when optimizing a culture medium, it is of paramount

importance to find the exact concentration of nutrients that would lead to optimal growth. Too low a nutrient concentration could hamper growth; while too much could cause growth inhibition.

Another limitation is that GEMs cannot predict the expression of genes. Therefore, expensive and time-consuming high throughput data must be incorporated into the model to account for this type of information. Moreover, GEMs cannot predict the functional state of proteins (posttranscriptional modifications), another big issue because, in many cases, the production of recombinant proteins is dependent on posttranscriptional modifications.

3 MATERIALS AND METHODS

3.1 Procedure

The GEM of *Oenococcus oeni* was constructed following the protocol for generating a high quality genome-scale metabolic reconstructions proposed by Thiele & Palsson (2010). This protocol comprises 5 stages (Figure 3-1).

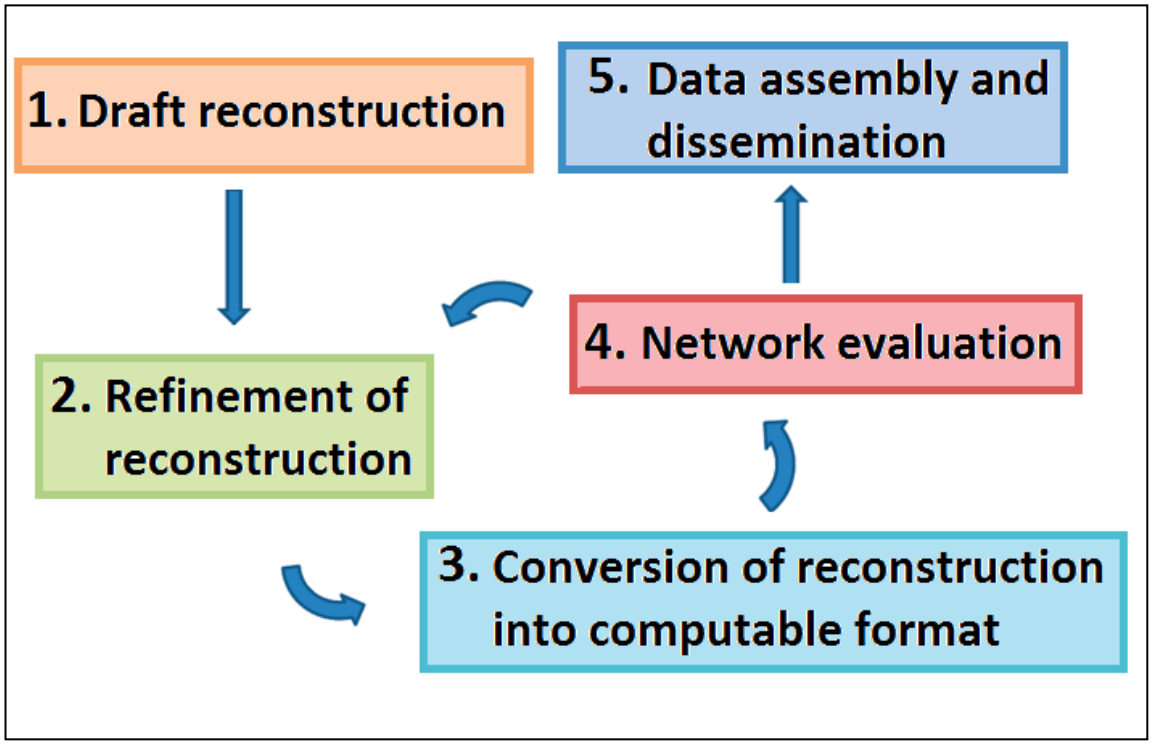

Figure 3-1. Overview of the procedure to iteratively reconstruct a high-quality metabolic network. The protocol comprises 5 stages involving a total of 96 steps. In the first stage a draft reconstruction is built. This draft reconstruction is often generated automatically therefore could contain incorrect information. In the second stage each reaction in this draft reconstruction is revised manually to repair possible autogenerated errors. Additionally missing reactions, not included initially in the draft reconstruction, are added from literature and databases. In the third stage, the reconstruction is converted into a computable format in order to be evaluated in the fourth stage. In the fourth stage, *in silico* simulations of phenotypes are tested. If there are inconsistencies between *in silico* results and literature, stage 2 must be carried out again in order to further refine the network. This iterative process ends when *in silico* results are considered to agree with experimental data. Figure modified from Thiele & Palsson (2010).

The first stage consists of generating a draft reconstruction, based on the genome annotation of the target organism and biochemical databases. This draft reconstruction is

an automatically created collection of genome-encoded metabolic functions and some of these may be falsely included, while others may be missing.

In the second stage, the entire draft reconstruction is re-evaluated and refined. The metabolic functions and reactions collected in the draft reconstruction are individually evaluated against organism-specific literature and expert opinion. Information about biomass composition, maintenance parameters and growth conditions are collected at this stage, providing a basis for simulation.

In the third stage, the reconstruction is converted into a mathematical format and condition-specific models are defined. This stage can mostly be automated. Moreover, systems boundaries are defined, converting the general reconstruction into a condition-specific model.

The fourth stage in the reconstruction process consists of network verification, evaluation and validation. The metabolic model created in the third stage is tested, among others things, for its ability to synthesize biomass precursors such as amino acids and lipids. This evaluation generally leads to the identification of missing metabolic functions in the reconstruction (the so-called network gaps) and the consequent reparation of these gaps.

In the fifth stage, the final reconstruction is made available to the research community in two formats: as a spreadsheet containing all information collected during the reconstruction process; and as a SBML file, a transportable format of the models which can be used with other modeling tools.

3.2 Sensitivity Analysis

3.2.1 By varying all coefficients at the same time

First, a sensitivity analysis was performed respecting that the sum of biomass component masses, expressed as weight/weight, results 100%. To do this, we calculated - using supplemental files from Oliveira, Nielsen, & Förster, (2005) - the mass of each component per 100 grams of biomass (hereafter called mass fraction). In this first analysis, the mass fraction of a particular biomass component was increased by a fixed percentage (1%, 10% and 50%) and mass fraction of other components was reduced, thus respecting sum of mass fractions gives 1. In our case, biomass is conformed for 40 components. By increasing the mass fraction of any component, call it A component, the other 39 components decreased their mass fraction in a thirty ninth of the mass increased, thereby maintaining constant total biomass. Then, we calculated each of the stoichiometric coefficients from the modified mass fractions, being incorporated into the model to replace the original coefficients. In each case, a model optimization was performed.

3.2.2 By varying one coefficient at time

Second, individual variations in the stoichiometric coefficients were analyzed. To this end, each of the coefficient was increased in fixed percentage units (1%, 10% and 50%). In each case, a model optimization was performed obtaining a specific growth rate that was compared with the one obtained using the model without changing the stoichiometric coefficient.

3.3 Softwares and Databases

3.3.1 The Pathway Tools software

Pathway Tools v. 16.5 (Karp et al., 2010; Karp, Paley, & Romero, 2002) was employed to create a draft reconstruction of the *O. oeni* PSU-1 strain, as well as a platform to

refine this first reconstruction. One of the most important components of Pathway Tools is the PathoLogic program. This program creates a new Pathway Genome Database (PGDB) from an input file that describes the annotated genome of an organism. This input file is usually in GenBank format and can be easily downloaded from the National Center for Biotechnology Information. PathoLogic performs both, a conversion and an inference process. The conversion process transforms flat file descriptions of genes and gene products into a PGDB representation of that information. The PathoLogic inference process predicts the metabolic pathway complement of the organism from its genome by comparison to the MetaCyc pathway DB. The resulting reconstruction of *O. oeni* was exported to an SBML file that serves as input to the COBRA Toolbox.

3.3.2 COBRA Toolbox

The Cobra Toolbox (Becker et al., 2007; Schellenberger et al., 2011) was used to simulate, analyze and predict metabolic phenotypes using the genome-scale model of *O. oeni*. Specifically, we carried out Flux Balance Analysis (FBA) based on linear optimization of the biomass formation under different sets of constraints.

3.3.3 Marvin Suite

An academic license was requested in the web site of ChemAxom in order to use Marvin Suite. We employed Marvin Sketch, an application contained in Marvin Suite, to predict the pK_a of those metabolites for which chemical formula is established.

3.3.4 Scripts for refinement and analysis

We developed several scripts in MATLAB to refine and analyze the GEM of *O. oeni*, which are described in the chapter of results.

3.3.5 TransportDB

TransportDB (Ren, Chen, & Paulsen, 2007) is a comprehensive database resource of information on cytoplasmic membrane transporters and outer membrane channels in organisms whose complete genome sequence is available. Several types of transporters, including ATP-dependent transporters, ion channels, phosphotransferase systems and secondary transporters were exported from transportDB into the metabolic model of *O. oeni.*

3.3.6 The MetaCyc Database

The MetaCyc Database was employed towards the use of Pathway Tools. The former is a database containing metabolic data (pathways, enzymes, reactions and substrate compounds) for many different organisms (Karp, Riley, Paley, & Pellegrini-Toole, 2002). MetaCyc is a review-level database in which a given entry often integrates information from multiple sources. The reactions included in MetaCyc were determined experimentally and are labeled with the species in which they were known to occur, based on literature references.

4 RESULTS AND DISCUSSION

4.1 Generation of draft reconstruction

The generation of a draft reconstruction corresponds to stage number in the Thiele & Palsson's protocol (2010). In this stage an initial automatic genome reconstruction is obtained. The main steps of this stage are the following:

4.1.1 Obtaining the genome sequence and annotation of *O. oeni*

Files containing genome sequence and genome annotation of *O. oeni* PSU-1 were downloaded from the National Center for Biotechnology Information[1] (NCBI) website in gene bank (.gbk) and fasta (.fna) formats, respectively. The genome annotation was uploaded to NCBI by Mills et al. (2005) and contains the following information for each gene (Figure 4-1):

1. Genome position
2. Coding region
3. Strand
4. Locus name
5. Gene function
6. Protein classification

[1] U.S. government-funded national resource for molecular biology information. Web site: http://www.ncbi.nlm.nih.gov/

```
     gene            complement(292470..293648)
                     /locus_tag="OEOE_0304"
                     /note="LactoCOG number LaCOG01324"
                     /db_xref="GeneID:4416230"
     CDS             complement(292470..293648)
                     /locus_tag="OEOE_0304"
                     /EC_number="2.7.1.6"
                     /note="catalyzes the formation of alpha-D-
galactose
                     1-phosphate from D-galactose in galactose
metabolism"
                     /product="galactokinase"
                     /protein_id="YP_809945.1"
                     /db_xref="GI:116490401"
                     /db_xref="GeneID:4416230"
```

Figure 4-1. Example of information contained in genome annotation. The gene OEOE_0304 is located between the pair bases 292.470 and 293.648 in the complementary strand and encodes for a galactokinase which catalyzes the formation of alpha-D-galactose 1-phosphate from D-galactose (Enzyme Commission number 2.7.1.6)

4.1.2 Identifying candidate metabolic functions and obtaining of candidate metabolic reactions

Identification of metabolic functions refers to finding functions for each gene product; meanwhile, obtaining candidate metabolic reactions refers to finding the reactions catalyzed by each gene product. It is worth to noting that identifying gene functions for each gene product is a straightforward step since these functions are explicitly mentioned in genome annotation. These two steps were carried out with Pathway Tools software. The total candidate metabolic reactions shape the draft reconstruction of *O. oeni*.

Pathway Tools allows one to explore draft reconstruction, enabling the proper visualization of each candidate metabolic function and each candidate metabolic reaction. Different types of information can be found for each reaction: compounds participating in the reaction, E.C. number, gene-reaction scheme and the pathways it belongs to (Figure 4-2).

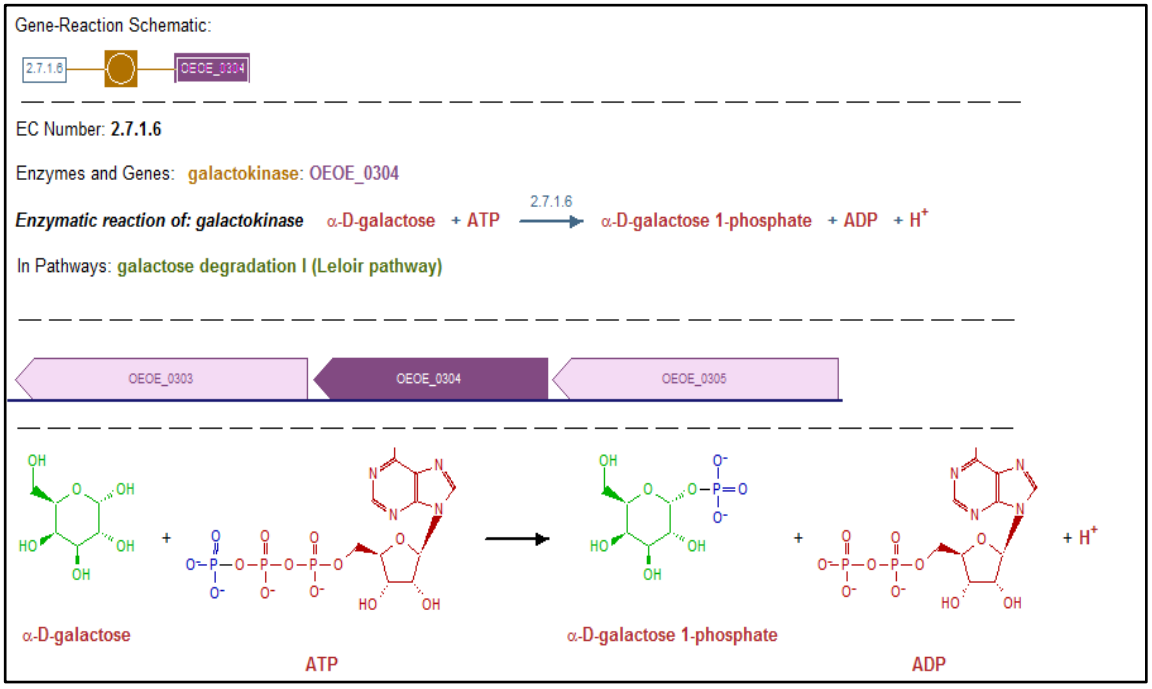

Figure 4-2. Overview of information visualized with Pathway Tools for gene OEOE_0304 of *Oenococcus oeni* PSU-1. Gene OEOE_0304 encodes for a galactokinase (candidate metabolic function), which transforms α-D-galactose to α-D-galactose 1-phosphate (candidate metabolic reaction). Additionally, this figure shows the gene-reaction scheme representing the relationship between gene OEOE_0304, the gene product and the reaction catalyzed. Moreover the E.C number of the reaction is specified as well as the pathway it belongs to. Finally gene context is also showed. This figure was extracted from the Pathway Tools platform.

4.1.3 Assembly of draft reconstruction and collection of experimental data

We employed Pathway Tools as a platform for assembling the draft reconstruction. In this platform, we saved all the changes made to the draft reconstruction. The collection of experimental data from literature was carried out during the entire reconstruction process. We mainly focused on collecting experimental data about nutritional requirements, specific growth rates and specific consumption/production rates, gene functions and known *O. oeni* capabilities. The features of this draft reconstruction are shown in (Table 4-1).

4.1.4 General analysis of draft reconstruction

The initial draft reconstruction lacked important exchange reactions supplying nutrients and exporting end products. Furthermore, it lacked a reaction that produced biomass, therefore, it was not functional, i.e. no description or prediction of metabolism could be obtained. Moreover, it contained more metabolites than reactions (Table 4-1), meaning that the model was overdetermined. A refined model is supposed to have a higher number of reactions than metabolites (Orth et al., 2010), because no solutions can be found in an overdetermined system. This issue resulted from the fact that the draft reconstruction contained 202 reactions that comprised 417 dead-ends, in total. The presence of these reactions in the model adds more equations than variables to the system, converting it to an overdetermined system.

Therefore, refining the draft model was imperative to achieve a representative model of the *O. oeni* metabolism.

Table 4-1. General features of the draft reconstruction of Oenococcus oeni. The draft reconstruction contains 1040 reactions and 1219 metabolites resulting in an overdetermined system with no solution. This results from the presence of 202 internal reactions that contain more than one dead-end, causing an important increase in the number of metabolites. Many of these reactions describe metabolic processes that are not of interest, or contain unspecific metabolites. Despite that 84 exchange reactions were included in this reconstruction, several others exchange reactions were missing for important nutrients such L-malate.

Genes	531
Pathways	168
Total reactions	1040
Internal reactions	956
External reactions	84
Exchange reactions	0
Total metabolites	1219
Internal metabolites	1152
External metabolites	67
Reactions	
Internal reactions associated to genes	56%
Reactions with one dead-end	214
Reactions with at least two dead-ends	202
Proteins	
Transporters	84
Protein complexes	151
Metabolites	
Dead-Ends	631
Metabolites missing chemical formula	481

4.2 Refinement of *O. oeni* GEM

In this stage, all reactions and metabolites that made up the draft reconstruction were evaluated by searching information supporting their existence. This stage comprised the following steps, which are explained in detail below.

4.2.1 Determination and verification of substrate and cofactor usage

We found that many reactions contained different names for the same metabolite due to different classification levels of metabolites. To repair this, a unique metabolite name was given for all of the reactions in which the metabolite was believed to participate. For example, *O. oeni* draft reconstruction contains the reaction represented by equation 5, which is the first step of the heterolactic fermentation. Nevertheless, the metabolite β Dfructoruranose[c][2] was found to participate only in that reaction. While no supply of β Dfructoruranose[c] was provided to the model, it was impossible to have a flux downstream of heterolactic fermentation reactions.

Equation 5. Reaction catalyzed by fructokinase.

$$\beta\ Dfructofuranose[c] + ATP[c] \rightarrow D\ fructose\ 6\ phosphate[c] + ADP[c] + H\hat{}\ + [c]\ (EC\ 2.7.1.4)$$

Equation 6. Exchange reaction for sugar

$$ATP[c] +\ a\ sugar[e] +\ H2O[c] \rightarrow ADP[c] + a\ sugar[c] +\ phosphate[c]$$

Equation 7. Hydrolysis of sucrose

$$sucrose[c] + H2O[c] \rightarrow D\ fructose[c] + D\ glucose[c]\ (E.C.3.2.1.10)$$

[2] The letter between brackets refers to the compartment in which the metabolites is located. [c] refers to the cytoplasm while [e] refers to extracellular space

Equation 8. Hydrolysis of sucrose

$$sucrose[c] + H2O[c] + phosphate[c] \rightarrow$$
$$D\ fructose[c] + D\ glucose[c]\ (E.C.2.4.1.7)$$

Furhermore, the draft reconstruction contained reactions given by equations 6, 7 and 8 which have the potential of producing β Dfructoruranose[c], since β Dfructofuranose[c] is the major natural isomer of D fructose[c]; and in turn, D fructose[c] is a type of sugar[c]. However, the way these reactions were written did not allow the model to produce "β Dfructofuranose[c]" but "a sugar[c]" and "D fructose[c]", metabolites that strictly do not participate in the heterolactic fermentation. Therefore, it was necessary to create instances of these reactions by replacing "a sugar[c]" and "D fructose[c]" with "β Dfructofuranose[c]" (Equations 9, 10 and 11).

Equation 9. Reaction catalyzed by fructokinase

$$ATP[c] + \beta\ Dfructoruranose[e] + H2O[c] \rightarrow ADP[c] + \beta\ Dfructoruranose[c] + phosphate[c]$$

Equation 10. Hydrolysis of sucrose

$$sucrose[c] + H2O[c] \rightarrow \beta\ Dfructoruranose[c] + D\ glucose[c]$$

Equation 11. Hydrolysis of sucrose

$$sucrose[c] + H2O[c] + phosphate[c] \rightarrow$$
$$\beta\ Dfructoruranose[c] + D\ glucose\ 1\ phosphate[c]$$

As a result, the model now contains three reactions that produce β Dfructoruranose, a substrate for heterolactic fermentation, and allows reaction fluxes in this pathway.

4.2.2 Obtaining charged formula for each metabolite

Metabolites were automatically protonated to a reference pH value of 7.3 by Pathway Tools. Since internal pH of *O. oeni* is 5.8 at external pH between 3.0 and 4.0 (pH of wine) (Ramos et al., 1995; Salema, Lolkema, Romão, & Dias, 1996), we used Marvin Sketch for predicting pK_a values for every metabolite having a defined chemical formula (Supplementary spreadsheet 1). In total, we compiled pK_a for 676 metabolites and found that only 14% of metabolites had a pK_a between 5.8 and 7.3. Since such a small fraction of metabolites would change its chemical formula by accepting protons, we considered that original formula of metabolites at pH 7.3 was sufficiently accurate for performing flux balance analysis.

4.2.3 Determination of reaction stoichiometry and directionality

All reactions were reviewed manually to balance equations. A total of 27 chemical equations were found to be unbalanced and were consequently repaired. Additionally, all new added chemical equations were carefully balanced.

The directionality of reactions found in the *O. oeni* draft reconstruction was assigned by Pathway Tools based on MetaCyc database. They were not modified, unless experimental evidence or expert opinion supported the modifications. As indicated by Thiele & Palsson (2010), reactions involving transfer of phosphate from ATP to an acceptor molecule were generally considered to be irreversible, except if those reactions are known to be reversible.

4.2.4 Reconciliation of draft reconstruction with the KEGG database

The KEGG database (Ogata, Goto, Sato, Fujibuchi, & Bono, 1999) contains 71 pathway maps for *O. oeni* metabolism. These maps were manually reviewed and each of the reactions was compared with those of the *O. oeni* draft reconstruction.

94 reactions were found to be present in the KEGG database but not in the *O. oeni* draft reconstruction. The reactions that improved network connectivity upon addition were incorporated into the model. On the contrary, the reactions whose addition worsened the network connectivity, i.e. those that created more than two new dead-end metabolites, were not added (Table A-1).

4.2.5 Adding information for gene and reaction localization

O. oeni is a Gram positive bacterium (Bartowsky, 2005) and therefore lacks periplasmic space. This implies that the reconstruction only contained two compartments: cytosol and extracellular space. All the genes and reactions were located in the cytosol, except for the intramembrane and exchange reactions.

4.2.6 Drawing of the metabolic map

The resulting reconstruction is shown below as a metabolic map.

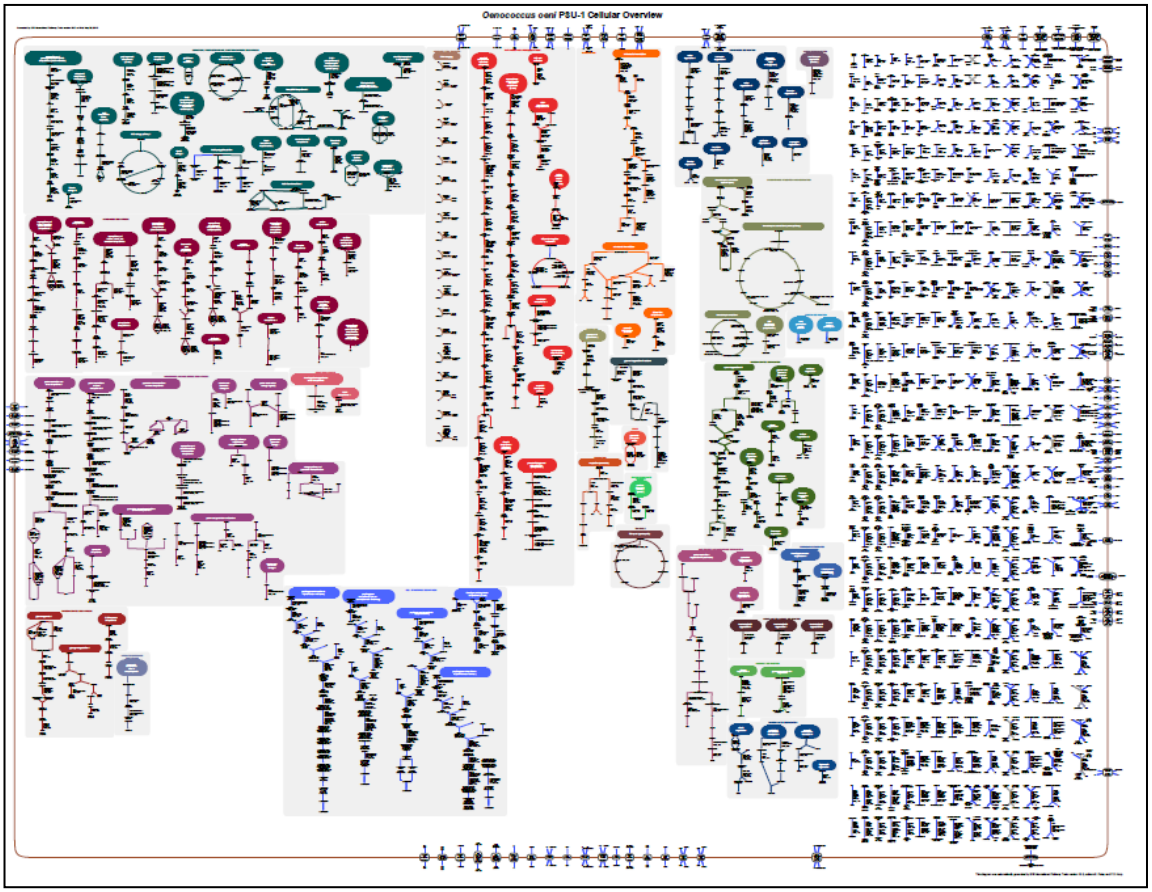

Figure 4-3. Overview of genome-scale metabolic reconstruction of *Oenococcus oeni* PSU-1. Each color represents a specific metabolic pathway. Reactions to the right are reactions that do not belong to any metabolic pathway. Nevertheless, these reactions are not necessarily disconnected from the network.

4.2.7 Determination of biomass composition

Since *O. oeni* biomass composition has not been determined yet, the biomass formation formula was incorporated from the GEM of *Lactococcus lactis,* due to its phylogenetic closeness. Despite *O. oeni* being phylogenetically close to *Lactobacillus plantarum*, the equation of *L. lactis*'s equation was preferred because of its simplicity. The formula consists of the sum of moles of the macromolecules that shape biomass: proteins, deoxyribonucleic acids, ribonucleic acids, lipoteichoic acids, lipids, peptidoglycan and polysaccharides (Equation 12).

Equation 12. Biomass production formula

$$0.421\ PROT + 0.0074\ DNA + 0.329\ RNA + 0.015\ LTA + 0.032\ LIP + 0.119\ PEPTIDOGLYCAN + 0.064\ POLYS + 18.15\ ATP \rightarrow biomass + 18.15\ ADP + 18.15\ phosphate$$

The equations related to macromolecular assembly from the corresponding building blocks were also incorporated. Thus, proteins (Equation 13) are constituted by the amino acids that contribute to its mass; DNA (Equation 14) is represented by the sum of deoxyadenosine-triphosphate, deoxycitidine-triphosphate, deoxyguanosine-triphosphate and thymidine triphosphate. The same logic is applied to RNA (Equation 15), lipoteichoic acids (Equation 16), lipids (Equation 17), peptidoglycan (Equation 18) and polysaccharides (Equation 19).

Equation 13. Protein Biosynthesis

$$8.6\ L\ alanine + 4.1\ L\ arginine + 3.1\ L\ aspartate + 5.9\ L\ asparagine + 3.4\ L\ cysteine + 3.6\ L\ glutamate + 6.4\ L\ glutamine + 9.2\ glycine + 1.5\ L\ histidine + 6.1\ L\ isoleucine + 8.7\ L\ leucine + 7.2\ L\ lysine + 2.5\ L\ methionine + 3.8\ L\ phenylanaline + 3.5\ L\ proline + 5.1\ L\ serine + 5.6\ L\ threonine + 1.7\ L\ tryptophan + 2.7\ L\ tyrosine + 7.2\ L\ valine + 430.6\ ATP \rightarrow 100\ PROT + 430.6\ ADP + 430.6\ phosphate$$

Equation 14. DNA Biosynthesis

$$32.3\ dATP + 17.7\ dCTP + 17.7\ dGTP + 32.3\ dTTP + 137.2\ ATP + 137.2\ H2O \rightarrow 100\ DNA + 137.2\ \text{ADP} + 100\ \text{diphosphate} + 137.2\ \text{H}^{+}$$

Equation 15. RNA Biosynthesis

$$26.2\ AMP + 20\ CMP + 32.2\ GMP + 21.6\ uridine\ 5'phosphate + 240\ ATP \rightarrow 100\ RNA + 240\ ADP + 240\ phosphate$$

Equation 16. Lipoteichoic acid Biosynthesis

$$D\ alanyl\ D\ galactosyl\ poly(glycerol\ phosphate) + diglucosyl\ diacylglycerol \rightarrow LTA + H2O$$

Equation 17. Lipids Biosynthesis

$$18.9\ L\_1\_phosphatidyl\_glycerol + 42.5\ cardiolipin + 4.3\ lysophosphatidylglycerol + 30.3\ diglucosyl\ diacylglycerol + 3\_D\_glucosyl\_1,2\_diacylglicerol \rightarrow 100\ LIP$$

Equation 18. Peptidoglycan Biosynthesis

$$N\ acetylmuramoyl\ L\ alanyl\ D\ glutamyl\ L\ lysyl\ D\ alanyl\ D\ alanine\ diphosphoundecaprenyl - N\ acetylglucosamine + D\ aspartate \rightarrow L\ alanine + di\ trans\ octa\ cis\ undecaprenyl\ diphosphate + peptidoglycan$$

Equation 19. Polysaccharides Biosynthesis

$$5.5\ UDP\ D\ glucose + UDP \propto D\ galactose + 5.6\ dTDP\ 6\ deoxy\ \beta\ L\ mannose \rightarrow POLYS + 6.5\ UDP + 5.6\ dTDP$$

4.2.8 Adding of ATP-maintenance

The reaction given by equation 20 was added to represent non-growth associated maintenance (NGAM) requirements. Zhang & Lovitt (2006) determined that NGAM was $0.6\frac{mmol\ ATP}{gDW\ h}$ for strain *O. oeni* 11648 growing in a continuous culture at pH 4.5 with glucose and fructose. On the other hand, Salou, Loubiere, & Pareilleux (1994) performed a batch culture at pH 5.0 using a strain isolated from a Burgundy red wine (France), and found that NGAM was $1.6, 2.3$ and $4.2\frac{mmol\ ATP}{gDW\ h}$ when *O. oeni* was fed with glucose, fructose, and fructose + glucose, respectively. We used the value determined by Zhang & Lovitt (2006) because FBA assumes a steady state which is accomplish in a continuous culture. Finally, it is worth mentioning that neither of these studies used a medium containing ethanol. Nevertheless NGAM is expected to increase in this condition since ethanol is highly toxic for bacteria.

As a reference point, *Escherichia coli* posses a NGAM of $8.36\ \frac{mmol\ ATP}{gDW\ h}$ in optimal growth conditions. As 8.36 is more than ten times the value determined by Zhang &

Lovitt (2006), we set NGAM to 0.6 $\frac{mmol\ ATP}{gDW\ h}$ in simulations without ethanol and up to 8.36 $\frac{mmol\ ATP}{gDW\ h}$ in simulation with ethanol.

Equation 20. Non-growth associated maintenance

$$ATP + H_2O \rightarrow ADP + P_i + H$$

4.2.9 Membrane transporters

There are four main types of transporters in *O. oeni*: ATP-dependent transporters, secondary transporters, phosphotransferase systems and ion channels. ATP-dependent transporters, also known as primary transport systems, use the free energy that is released upon the hydrolysis of ATP. Secondary (active) transporters require the free energy that is stored in the electrochemical gradients of protons, sodium ions or other solutes across the membrane. Phosphotransferase system is a form of transport in which the substance transported is chemically modified during its uptake across the membrane. Energy for the phosphotransferase system comes from the energy-rich compound phosphoenolpyruvate, which is dephosphorylated into pyruvate for phosphorylating the incoming substrate. Finally, ions channels are membrane proteins that allow particular ions to pass through them from one side of the membrane to the other (Broome-Smith et all, 1999, Brock, Ion channels).

Some transporters were predicted by Pathway Tools, while others were added from TransportDB and literature (Table 4-2). The most relevant transporters incorporated to the model from literature were the transporters related to malolactic reaction and heterolactic fermentation. Transporters for L-lactate, D-lactate, diacetyl, 2,3 butanediol and acetate (Figure 4-4) were incorporated to the model from literature, since none were included by Pathway Tools or identified by TransportDB.

L-lactate is the product of the malolactic reaction and its transport through the cell membrane involves the excretion of one proton (Bartowsky, 2005; Loubiere, Salou, Leroy, Lindley, & Pareilleux, 1992). This system represents an evolutionary advantage because, thanks to this mechanism, *O. oeni* is able to generate the proton motive force necessary to produce ATP. Also, a D-lactate transporter was incorporated taking into account the same principle. It is worthy to mention that D-lactate can only be produced from pyruvate in *O. oeni*, while L-lactate can only be produced by malolactic reaction. Therefore, it is easy to track down the process involved in the bacterial metabolism, by performing a flux balance analysis and studying the production rates of L-lactate and D-lactate. Finally, transporters that allow facilitated diffusion were also incorporated for acetate, diacetyl and 2,3 butanediol, the main end products of the heterolactic fermentation.

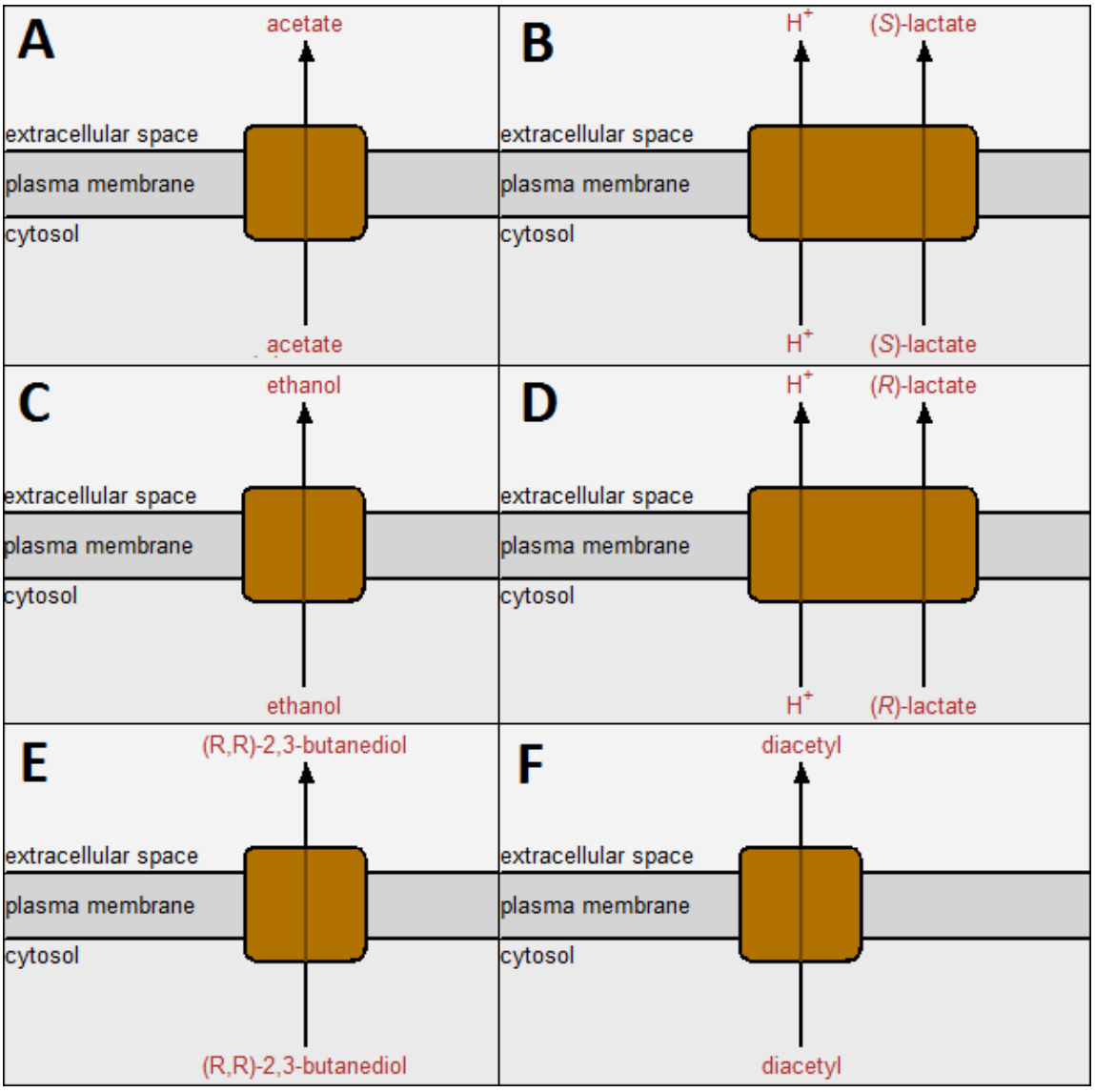

Figure 4-4. Representation of transporters incorporated into *Oenococcus oeni* reconstruction for acetate (A), (S)-lactate (B), ethanol (C), (R)-lactate (D), 2,3-butanediol (E) and diacetyl (F). Created with Pathways Tools.

Table 4-2. Exchange metabolites between cytoplasm and extracellular space in *O. oeni* reconstruction after manual refinement. The mechanism of transport is described in the second column for each metabolite transported. In case of multiple mechanisms of transport, all transporters are mentioned with the corresponding reference in the last column. Also, the direction of transport is described in the third column

Metabolite	Mechanism	Direction	Reference
Carbon sources			
Citrate	Secondary transporter	Consumed	MetaCyc
S-malate	Secondary transporter	Consumed	TransportDB
α-D-galactose	Na+/lactose/H+ symporter, ATP-dependent	Consumed	Blast *L. plantarum,* MetaCyc
B-D-galactose	Na+/lactose/H+ symporter, ATP-dependent	Consumed	Blast *L. plantarum,* MetaCyc
raffinose	Na+/lactose/H+ symporter	Consumed	Blast *L. plantarum*
cellobiose	Facilitated diffusion, PTS	Consumed	TransportDB, MetaCyc
D-gluconate	H+ symporter	Consumed	Metacyc
lactose	H+/Na+/raffinose/melibiose/ α-D-galactose/ B-D-galactose symporter	Consumed	Blast *L. plantarum*
B -glucoside	PTS	Consumed	TransportDB
α -D-xylopyranose	ATP-dependent, H+symporter	Consumed	Metacyc
α -L-arabinopyranose	ATP-dependent, H+symporter	Consumed	Metacyc
B -D-fructofuranose	ATP-dependent, PTS	Consumed	TransportDB
B -D-glucose	ATP-dependent, H+ symporter, PTS	Consumed	Metacyc

arabinose	ATP-dependent, H+symporter	Consumed	MetaCyc
α -D-glucose	ATP-dependent, H+ symporter, PTS	Consumed	TransportDB
B -D-ribofuranose	ATP-dependent	Consumed	Metacyc
B -D-ribopyranose	ATP-dependent	Consumed	Metacyc
D-xylose	ATP-dependent, H+symporter	Consumed	Metacyc
Melibiose	Na+/lactose/H+ symporter	Consumed	Blast *L. plantarum*
D-galactitol	ATP-dependent, PTS	Consumed	TransportDB
D-mannose	ATP-dependent, PTS	Consumed	TransportDB, MetaCyc
Salicin	PTS	Consumed	MetaCyc
Trehalose	PTS	Consumed	MetaCyc
hydroquinone-O-β–D-glucopyranoside	PTS	Consumed	MetaCyc
Nitrogen sources			
L-glutamate	ATP-dependent, facilitated diffusion and 4-aminobutanoate antiporter	Consumed	TransportDB
L-isoleucine	ATP-dependent, facilitated diffusion	Consumed	TransportDB
L-cysteine	ATP-dependent, facilitated diffusion	Consumed	TransportDB
L-phenylalanine	ATP-dependent, facilitated diffusion	Consumed	TransportDB

L-asparagine	ATP-dependent, facilitated diffusion	Consumed	TransportDB
L-proline	ATP-dependent, facilitated diffusion	Consumed	TransportDB
L-glutamine	ATP-dependent, facilitated diffusion	Consumed	TransportDB
L-methionine	ATP-dependent, facilitated diffusion	Consumed	TransportDB
L-aspartate	ATP-dependent, facilitated diffusion	Consumed	TransportDB
L-lysine	ATP-dependent, facilitated diffusion	Consumed	TransportDB
L-histidine	ATP-dependent, facilitated diffusion	Consumed	TransportDB
L-tyrosine	ATP-dependent, facilitated diffusion	Consumed	TransportDB
glycine	ATP-dependent, facilitated diffusion	Consumed	TransportDB
L-arginine	ATP-dependent, facilitated diffusion	Consumed	TransportDB
L-leucine	ATP-dependent, facilitated diffusion	Consumed	TransportDB
L-alanine	ATP-dependent, facilitated diffusion	Consumed	TransportDB
L-valine	ATP-dependent, facilitated diffusion	Consumed	TransportDB
L-tryptophan	ATP-dependent, facilitated diffusion	Consumed	TransportDB
L-threonine	ATP-dependent, facilitated diffusion	Consumed	TransportDB

L-serine	ATP-dependent, facilitated diffusion	Consumed	TransportDB
Ammonia	H+ symporter	Consumed	MetaCyc
4-aminobutanoate	H+ symporter and L-glutamate antiporter	Consumed/Produced	MetaCyc
Vitamins			
R-pantothenate	H+ symporter	Consumed	(Richter, Vlad, & Unden, 2001)
riboflavin	H+ symporter	Consumed	
L-ascorbate	H+ symporter	Consumed	Inferred conputationally
thiamin	ATP-dependent	Consumed	
Secretion products			
R-lactate	H+ symporter	Produced	(Olguín et al., 2009)
S-lactate	H+ symporter	Produced	(Olguín et al., 2009)
acetate	Facilitated diffusion	Produced	(Olguín et al., 2009)
Ethanol	Facilitated diffusion	Produced	(Ramos & Santos, 1996)
Diacetyl	Facilitated diffusion	Produced	(Ramos & Santos, 1996)
(R,R)-2,3-butanediol	Facilitated diffusion	Produced	(Ramos & Santos, 1996)
D-mannitol	PTS	Produced	Metacyc
Ions			
Cd2+	Co/Zn/Cd efflux system	Produced	MetaCyc
K+	Ion channel, ATP-dependent	Consumed	Metacyc, TransportDB
Mg+2	Ion channel	Consumed	TransportDB

Fe+2	H+ symporter	Consumed	MetaCyc
Na+	H+/xyloside/raffinose/ melibiose/ α-D-galactose/ B-D-galactose/lactose symporter, ATP-dependent	Consumed	Transport DB, Metacyc
Ni2+	ATP-dependent	Consumed	Metacyc
Co2+	Co/Zn/Cd efflux system, ATP-dependent, Ion channel	Both	TransportDB
Mn2+	ATP-dependent, Facilitated diffusion, H+ symporter	Consumed	TransportDB, Metacyc
Zn2+	ATP-depedent and Co/Zn/Cd efflux system	Both	TransportDB
Cu2+	ATP-dependent	Consumed	TransportDB
H+	?	Consumed	For modeling purposes
phosphate	ATP-dependent	Consumed	TransportDB
Nitrate	ATP-dependent	Consumed	MetaCyc
Bicarbonate	ATP-dependent	Consumed	MetaCyc
Bases			
xanthine	Uracil/H+ symporter	Consumed	Inferred computatitoally
uracil	xanthine	Consumed	Inferred computationally
Cytosine	Facilitated diffusion	Consumed	
adenine	Facilitated diffusion	Consumed	For modeling purposes
Others			
glycerol	ATP-dependent, H+symporter	Consumed	Inferred computactionally
3,5-dimethoxytoluene		Consumed	MetaCyc

Cyanate	H+/symporter	Consumed	MetaCyc
Carbon monoxide	Facilitated diffusion	Consumed	MetaCyc
H20			For modeling purposes
Spermidine	ATP-dependent	Consumed	TransportDB, Metacyc
Putrescine	ATP-dependent	Consumed	TransportDB, Metacyc
Gases		Consumed	
Carbon dioxide (CO2)	Facilitated diffusion	Produced	For modeling purposes
Oxygen (O2)	Facilitated diffusion	Consumed	For modeling purposes
Cofactors			
[FeS] iron-sulfur cluster	ATP-dependent	Consumed	TransportDB

4.2.10 Malolactic fermentation

Surprisingly, *Oenococcus oeni* annotation does not contain a gene associated to the malolactic reaction. The gene OEOE_1563 encodes a malate permease, while the gene OEOE_1565 encodes a malolactic transcription activator for the fermentarion system. As the malolactic enzyme, the malate permease and the regulatory protein are in the same operon (Bartowsky, 2005), the malolactic enzyme was considered to be encoded by the gene OEOE_1564.

The malolactic reaction is represented by the equation $(S) - malate + H^+ \rightarrow (S) - lactate + CO_2$ in the model, (Figure 4-5). Even though this reaction needs Mn^{++} and NAD^+, these cofactors are not consumed and, therefore, were not included in the equation (Naouri, Chagnaud, Arnaud, & Galzy, 1990). The proton on the left side of the equation was added to balance the mass

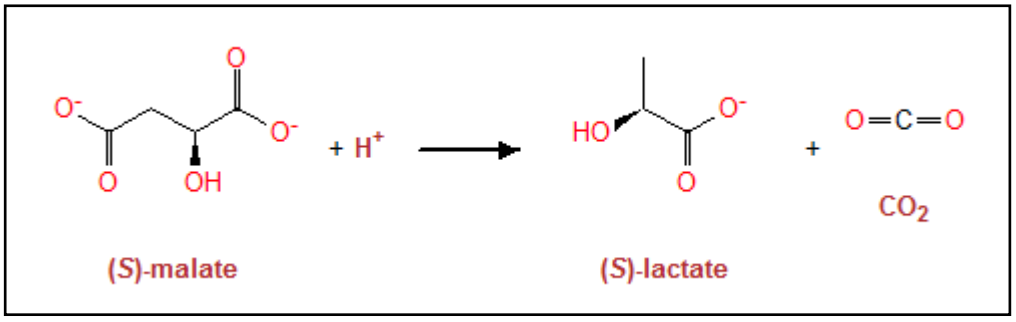

Figure 4-5. Malolactic stoichiometric equation included in the genome-scale metabolic model of *Oenococcus oeni*. (S)-malate (or L-malate) is converted to (S)-lactate (or L-lactate) and CO_2. Made with Pathways Tools

4.2.11 TCA cycle

All the reactions related to the TCA cycle were included in the model by Pathway Tools, even though the genes for most reactions were not present in the genome annotation (Figure 4-6). *O. oeni* lacks a functional TCA cycle (Garcia, Blancato, Repizo, Magni, & López, 2008), therefore, the only enzymes that were conserved in the model were those that have existing bioinformatic evidence. Malate oxidoreductase (E.C. 4.2.1.2) is

predicted to be associated to four independent genes: OEOE_0418, OEOE_0553, OEOE_1325 and OEOE_1552; while fumarase (E.C. 1.1.1.37) is associated to OEOE_0029. Hence, these reactions were conserved in the model and all the other TCA reactions were removed, since neither bioinformatic nor experimental evidence has been provided to claim their presence.

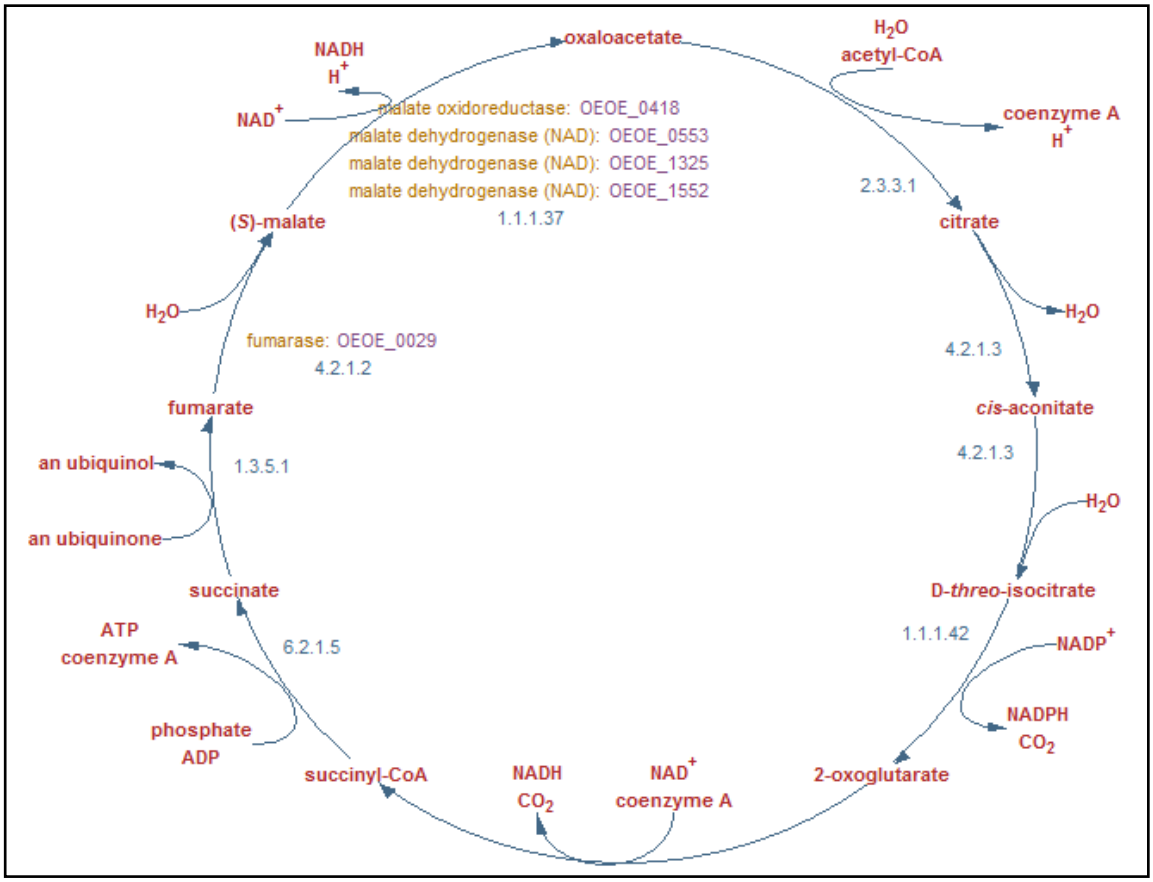

Figure 4-6. Defective TCA cycle of *Oenococcus oeni.* Genes associated to malate dehydrogenase (E.C. 1.1.1.37) and fumarase (E.C. 4.2.1.2) were the only ones found. All the other TCA reactions lack associations with genes.

4.2.12 Heterolactic fermentation

O. oeni metabolizes glucose, fructose and citrate through heterolactic fermentation to form different end products, such as acetate, ethanol, diacetyl, 2,3 butanediol and D-lactate (Lerm et al., 2010). Nevertheless, the heterolactic pathway was incomplete in the draft reconstruction because two reactions were missing. To complete this pathway, the transformation of diacetyl to acetoin (Figure 4-7 A) and the transformation of acetoin to 2,3-butanediol (Figure 4-7 B) were included.

Additionally, one falsely included reaction related to heterolactic fermentation was removed from the draft reconstruction. The reaction catalyzing the transformation of pyruvate to L-lactate was removed since *O. oeni* only produces D-lactate, and not L-lactate, through heterolactic fermentation (Wagner et al., 2005); L-lactate is only produced from L-malic acid by the malolactic enzyme (Bartowsky, 2005).

Finally, a map of this pathway was constructed in Pathway Tools to properly visualize all the reactions involved. Here, we include the uptake of glucose, fructose and citrate (Figure 4-8) and the consequent reactions needed to form acetate, ethanol, diacetyl, 2,3-butanediol and D-lactate (Figure 4-9).

It is worth mentioning that *O. oeni* is able to obtain energy from heterolactic fermentation. Acetate production through acetate kinase involves ATP generation. This does not occur with acetate production through citrate lyase. Ethanol, acetoin, 2,3 butanediol and R-lactate production involves the regeneration of the redox cofactor NAD^+; and diacetyl, (R)-acetoin and (S)-2-acetolactate production involves CO_2 generation.

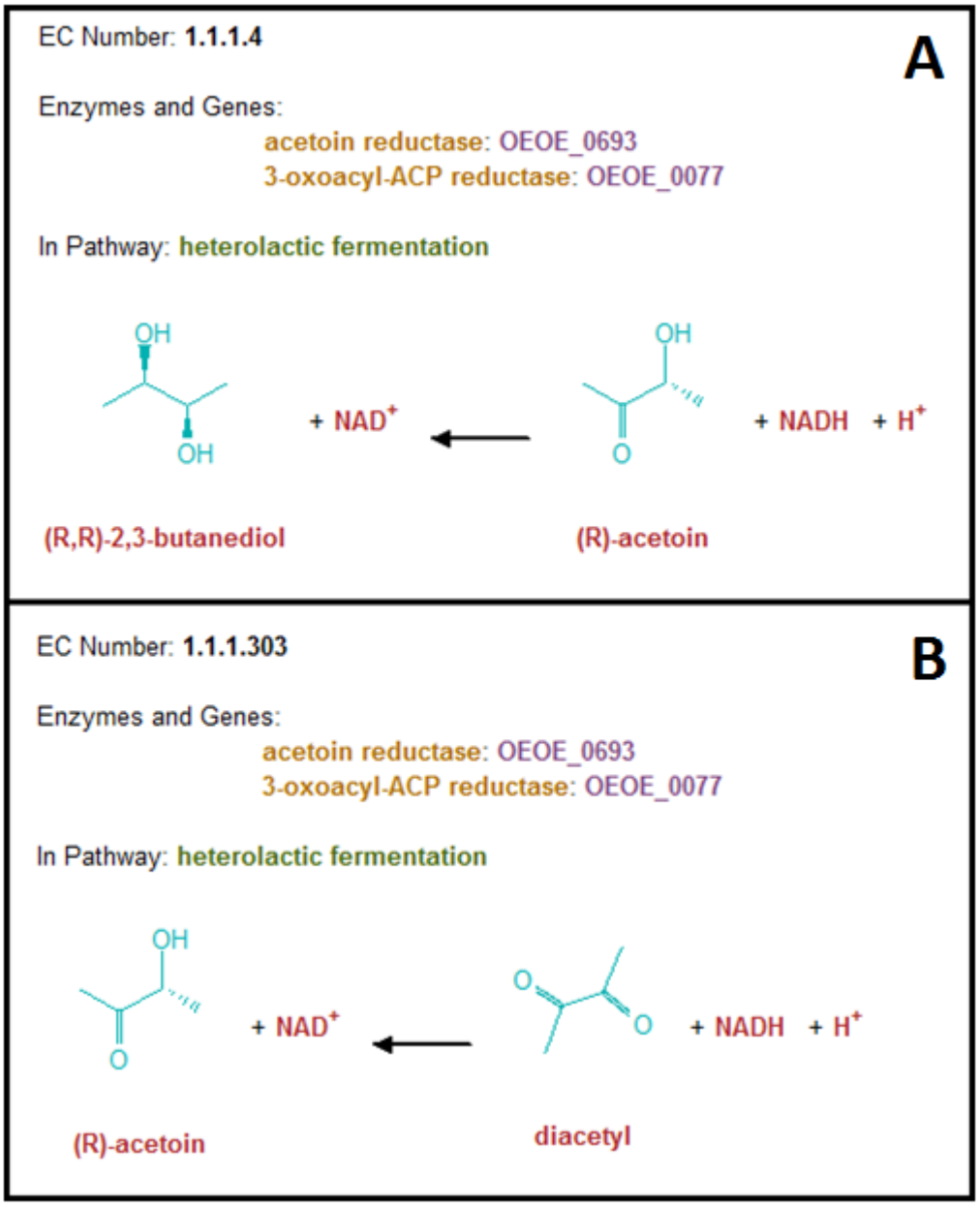

Figure 4-7. Reactions added to the *Oenococcus oeni* reconstruction to complete the heterolactic fermentation pathway. The transformation of acetoin to 2-3 butanediol (A) and the transformation of diacetyl to acetoin (B) were added to the reconstruction to complete this pathway. Two isoenzymes are responsible of carrying out these reactions. Genes encoding these reactions were found in KEGG database.

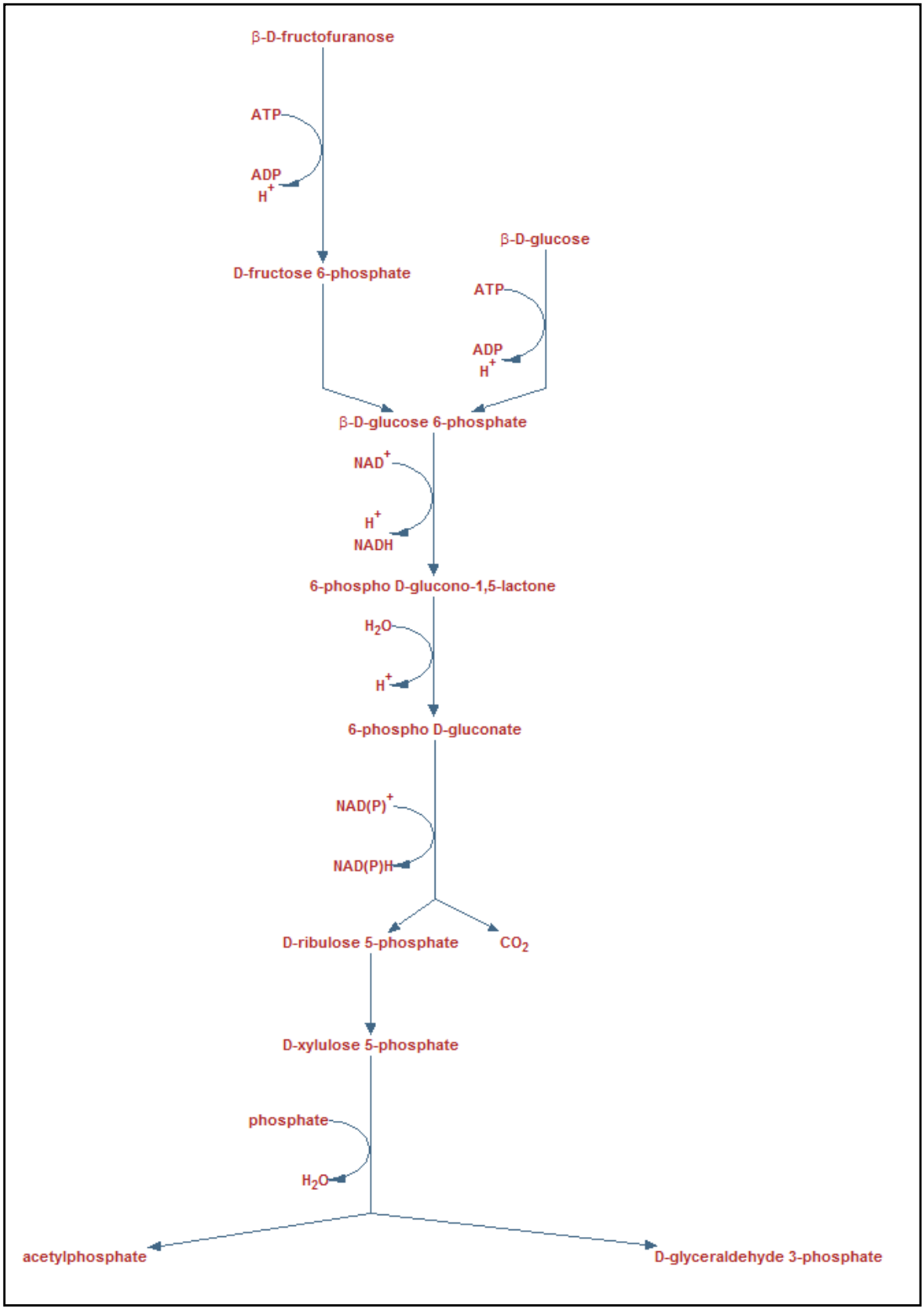

Figure 4-8. First half of the heterolactic fermentation. Hexoses (β-D-fructofuranose and β-D-glucose) are converted to two compounds with 3 carbons: acetylphosphate and D-glyceraldehyde. Made with Pathways Tools.

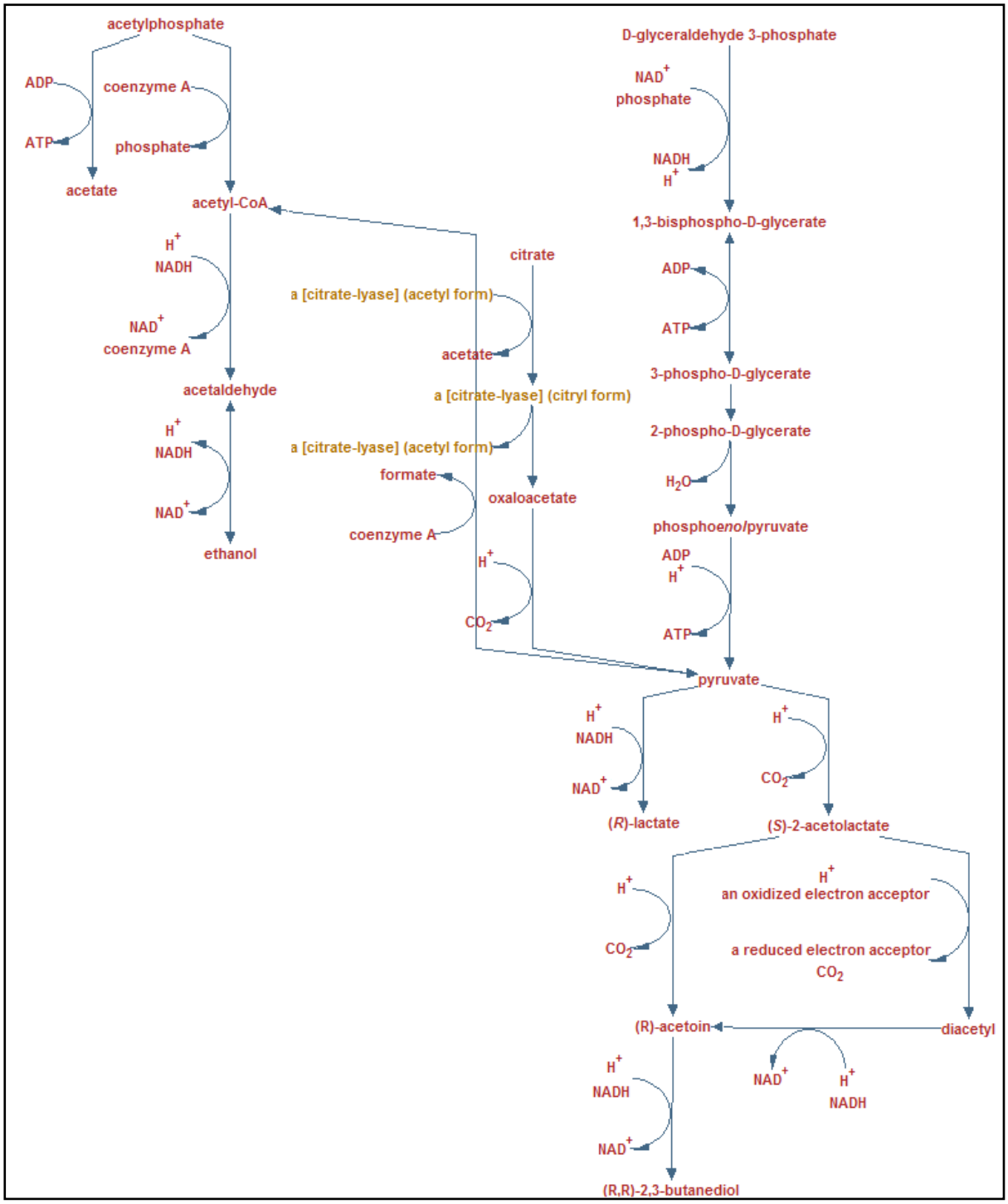

Figure 4-9. Second half of the heterolactic fermentation. Citrate is converted to oxaloacetate through citrate lyase. The main end products of this pathway are generated from pyruvate: D-lactate, diacetyl and 2,3-butanodiol. Made with Pathways Tools.

4.2.13 General features of the GEM of *O. oeni*

An improvement in connectivity was achieved after refinement. Connectivity refers to the number of reactions in which a given metabolite participates (Orth et al., 2010). It is expected that in a high quality GENRE, most metabolites only participate in a few reactions; and just a few metabolites participate in a large number of reactions. Many metabolites having a very low connectivity could be detrimental to the functionality of the model because, in such case, the model contains many gaps avoiding flux through reactions. This is the case of most draft reconstructions in which connectivity must be improved to achieve production of biomass. Even though having a high connectivity does not ensure a high quality GEM, it is a prerequisite; and the refinement process allows the network's connectivity to increase.

The *O. oeni* draft reconstruction contained 631 metabolites participating in just one reaction (Figure 4-10 A). This large degree of disconnection was significantly reduced in our refined model, down to 224 metabolites participating in one reaction (Figure 4-10 B). However, comparing with the most recent reconstruction of *Escherichia coli*, in which only 87 metabolites participate in just one reaction (Orth et al., 2011) (Figure 4-11), there is still a lot of room for improvement.

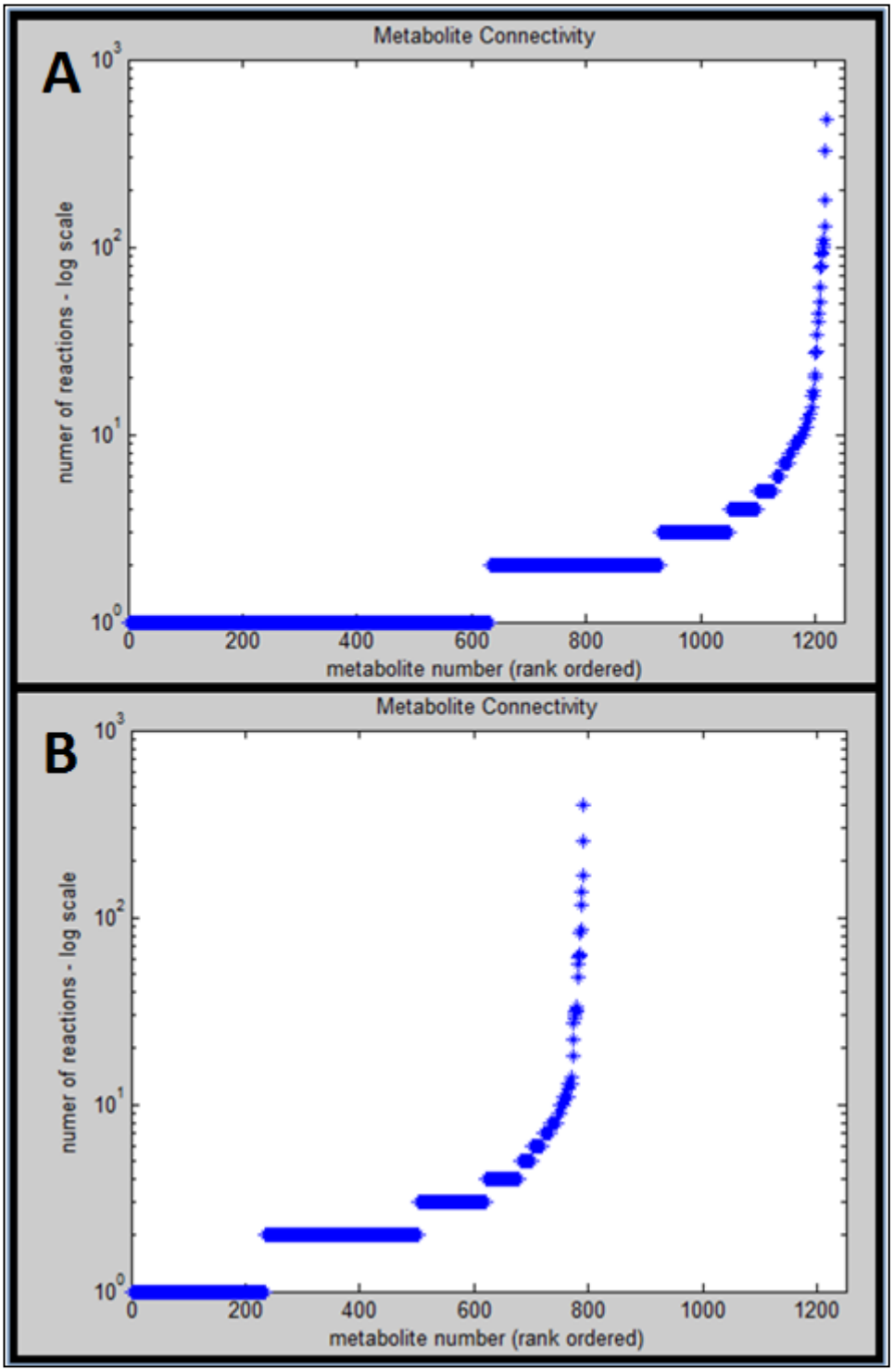

Figure 4-10. Comparison of metabolite connectivity of *Oenococcus oeni* draft reconstruction (A) and *Oenococcus oeni* reconstruction after manual refinement (B). The number of metabolites participating in just one reaction diminishes from 631 to 234 after manual refinement. This resulted from elimination of non specific and disconnected reactions that do not contribute to describe the *Oenococcus oeni* metabolism.

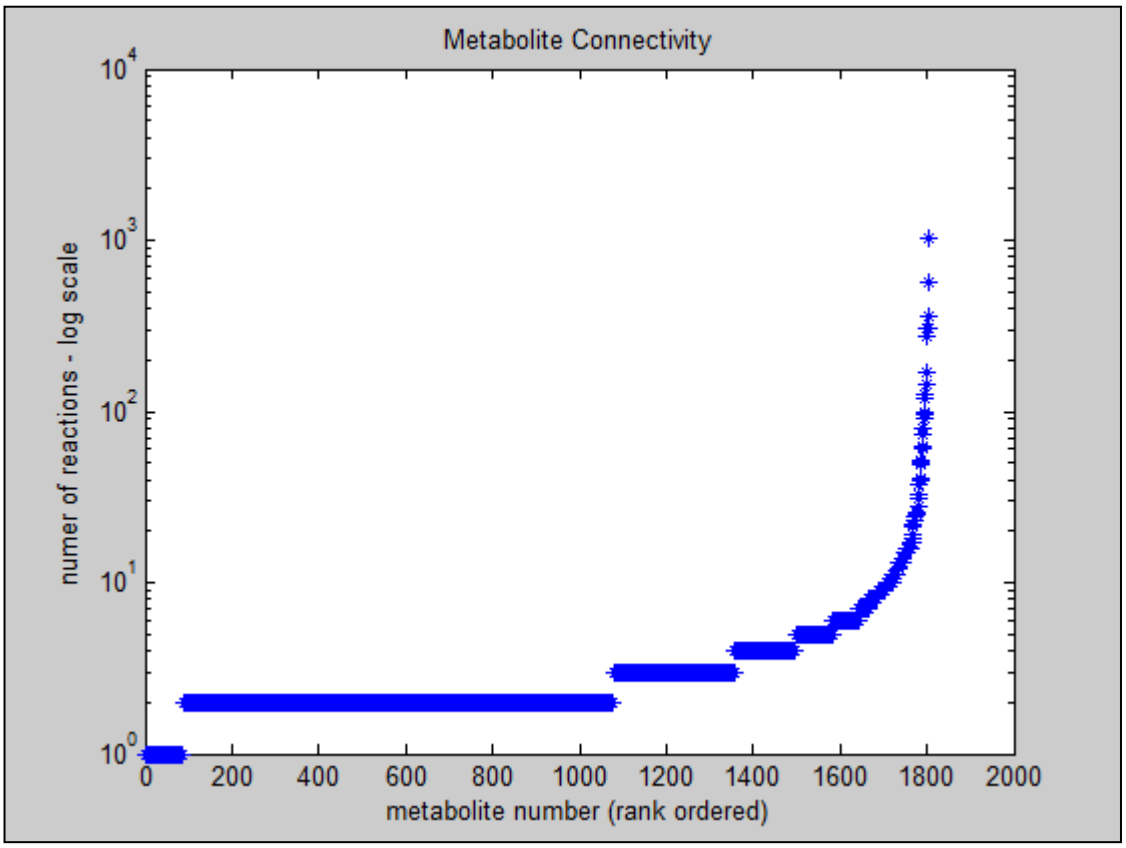

Figure 4-11. Metabolite connectivity in GEM of *Escherichia coli.* It can be appreciated that many metabolites participate in a small number of reactions while few metabolites participate in a large number of reactions which is the general behavior of high quality GEMs. There are 87 metabolites participating in just one reaction and 991 metabolites participating in two reactions. The reason for the large number of metabolites participating in two reactions is because 480 metabolites are involved in exchange reactions, including transport between extracellular space, periplasm and cytoplasm. On the other hand, there are few metabolites participating in a large number of reactions such as the proton that participates in 1031 reactions.

The resulting genome-scale metabolic model of *O. oeni* after refinement is a subdetermined model containing 914 reactions, 792 metabolites and 512 genes (Table 4-3).

Table 4-3. Comparison between the initial draft reconstruction and the final reconstruction after manual refinement.

	After automatic reconstruction	After manual refinement
Genes	531	512
Pathways	168	171
Total reactions	1040	914
Internal reactions	956	690
External reactions	84	134
Exchange reactions	0	90
Total metabolites	1219	792
Internal metabolites	1152	702
External metabolites	67	90
Reactions		
Internal reactions associated to genes	56%	76%
Number of reactions with one dead-end	214	191
Number of reactions with at least two dead-ends	202	21
Proteins		
Transporters	84	134
Protein complexes	151	236
Metabolites		
Dead-Ends	631	234
Metabolites missing chemical formula	481	227

4.3 Conversion of *O. oeni* GENRE to a computable format

Conversion of *O. oeni* GENRE corresponds to stage three in the protocol of Thiele & Palsson (2010). In this stage, the constructed GENRE is converted into a mathematical format. Moreover, system boundaries are defined, converting the general reconstruction into a condition-specific model.

4.3.1 Initialization of COBRA Toolbox.

COBRA Toolbox was run in MATLAB 7.10.0 (R2010a) and initialized using glpk as LP solver.

4.3.2 Loading of reconstruction into MATLB

O. oeni GENRE was exported from Pathway Tools to a SBML[3] format file. This was done automatically through Pathway Tools. Then, the SBML file was read with the COBRA script:

model = ReadCbModel('modelName.xml');

This loaded the variable *model* which contains descriptive and numeric fields for reactions, metabolites and genes (Figure 4-12). Below, we give a brief description of the most relevant fields that can be found once the model is read with COBRA Toolbox:

1) *rxns*: vector containing unique IDs for each of the reactions present in the model
2) *mets:* vector containing unique IDs for each of the metabolites present in the model
3) *rxnNames*: vector containing reaction names for each of the reactions present in the model. These match with the reaction IDs given by Pathway Tools

[3] Systems Biology Markup Language (SBML) a free and open interchange format for computer models of biological processes. SBML is useful

4) *metNames:* vector containing metabolite names for each of the metabolites present in the model
5) *S:* stoichiometric matrix. Entries of each column are the stoichiometric coefficients of the metabolites participating in the corresponding reactions.
6) *lb:* vector containing lower bounds for each of the reactions present in the model
7) *ub:* vector containing upper bounds for each of the reactions present in the model
8) *c:* vector containing weights for each reaction in the objective function.

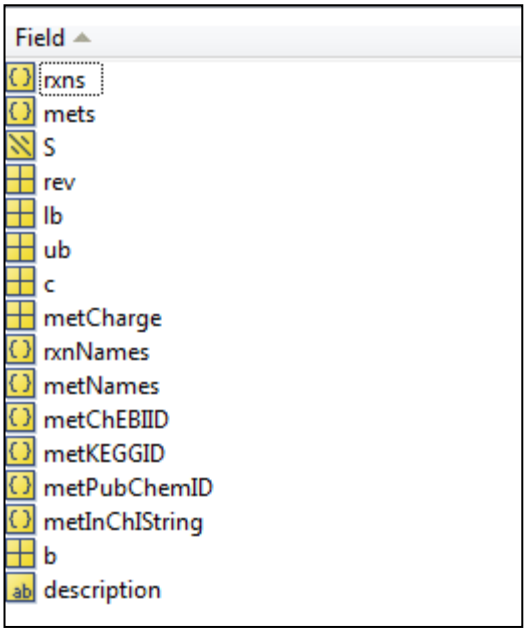

Figure 4-12. Fields contained in the *Oenococcus oeni* model once it is loaded using COBRA Toolbox. The model contains several fields describing reactions and metabolites present in the model but lacks important fields such as genes, rules, grRules and rxnGeneMat.

Once constructed, we determined that the model lacked the field *genes*, *rules*, *grRules* and *rxnGeneMat*, which are key elements of any high quality GEM. Also, the model contained many repeated metabolite names, due to the lack of references to the compartments. Both issues were solved by developing new Matlab scripts, which are explained below.

4.3.2.1 Adding new fields.

Pathway Tools can export GENREs to a SBML format, allowing easier access to several softwares, such as COBRA (Becker et al., 2007) or Omix (Droste, Nöh, & Wiechert, 2013). However, the SBML file does not contain information about gene reactions associations. Therefore, the model loaded through COBRA Toolbox lacks important fields such as *Genes, GrRules, Rules and RxnGeneMat*[4], all of which are necessary to run gene deletion analysis. We developed the scripts *GenerateGeneFields.m, GenerateRulesFields.m and GenerateRxnGenMat.m* to read the information contained in Pathway Tools and create the corresponding fields. After this, we were able to run COBRA Toolbox for any gene deletion analysis.

These scripts were of great help to execute the next steps of Thiele & Palsson´s protocol after "loading a GEM into MATLAB" (step 39). The resulting algorithm required 13 seconds to create the fields *Genes, GrRules, Rules and RxnGeneMat* (Figure 4-13) in the *O. oeni* GEM, which contained 914 reactions and 512 genes.

[4] Gene field is a MATLAB array containing the names of all model genes. GrRules field is a MATLAB array containing the logic description (OR/AND) of what genes are associated to each reaction in the model. Rules field is a MATLAB array containing the logic description (|/&) of what genes are associated to each reaction in the model. RxnGeneMat is a binary matrix where the presence of the number one in the position i,j indicates gene j encodes a enzyme catalyzing reaction i

	model2.rxnNames <914x1 cell>	model2.rules <914x1 cell>	model2.grRules <914x1 cell>
	1	1	1
331	GLUTAMIN-RXN	OEOE_0260 | OEOE_0261 | OEOE_1036	OEOE_0260 or OEOE_0261 or OEOE_1036
332	RXN-9591	OEOE_1025	OEOE_1025
333	RXN0-5021	OEOE_0995	OEOE_0995
334	2.7.4.22-RXN	OEOE_1008 | OEOE_0977	OEOE_1008 or OEOE_0977
335	RXN-1623	OEOE_0973	OEOE_0973
336	3.4.11.9-RXN	OEOE_0959	OEOE_0959
337	RXN0-6274	OEOE_0950	OEOE_0950
338	GLUTATHIONE-PEROXIDASE-RXN	OEOE_0888	OEOE_0888
339	THIAMIN-PYROPHOSPHOKINASE-RXN	OEOE_0791	OEOE_0791
340	5-FORMYL-THF-CYCLO-LIGASE-RXN	OEOE_0779	OEOE_0779
341	DIHYDRODIPICSYN-RXN	OEOE_0774	OEOE_0774
342	GLUCOSAMINE-6-P-DEAMIN-RXN	OEOE_1797 | OEOE_0644	OEOE_1797 or OEOE_0644
343	RXN0-1342	OEOE_0579	OEOE_0579
344	GLYC3PDEHYDROG-RXN	OEOE_0564	OEOE_0564
345	1.1.1.8-RXN	OEOE_0564	OEOE_0564
346	GLYC3PDEHYDROGBIOSYN-RXN	OEOE_0564	OEOE_0564

Figure 4-13. Extract of rxnNames (left), grRules (center) and rules (right) MATLAB arrays corresponding to the fields of *O. oeni* GEM.

4.3.2.2 Generation of meaningful names for metabolites.

The compartments in which the metabolites could be found are a key aspect for the correct development of a model. In this regard, it is relevant that the names of the metabolites loaded with COBRA Toolbox contain information about the compartments. For example, if metabolite names contain information about compartments, the user will be able to differentiate between an extracellular glucose, *glucose[e]*, and a cytoplasmic glucose, *glucose[c]*. Performing this step manually is tedious; therefore, an automated tool that could automate this task allows a reproducible and effective process. Incorporating this information into metabolite names permits one to easily visualize reaction localization.

For this purpose, we developed the script *compartimentalize.m* that assigns at the end of the metabolite´s name a letter representing the compartment where the metabolite is located, saving the information in the COBRA model. The information needs to go through a “cleaning process” before being added to the model. This process consists of separating names from chemical formulae, since the information is contained in a single string (e.g. "*Glucose: C6H12O6*"). We developed a script to perform this task named

separateNamesAndFormulas.m, which is automatically called by *compartimentalize.m,* separating the name and chemical formula for each of the metabolites. It takes approx. 0.3 seconds to run this script.

The new fields added to the model (Figure 4-14) can be summarized as follows:

1) *metsFormulas*: vector containing metabolite chemical formula for each of the metabolites present in the model
2) *genes*: vector containing gene names for each of the genes present in the model.
3) *rules:* vector containing the logic description (|/&) of which genes are associated to each reaction in the model
4) *grRules:* vector containing the logic description (AND/OR) of which genes are associated to each reaction in the model
5) *rxnGeneMat:* Binary matrix in which the presence of the number one in the position i,j indicates a gene j encodes an enzyme catalyzing reaction i

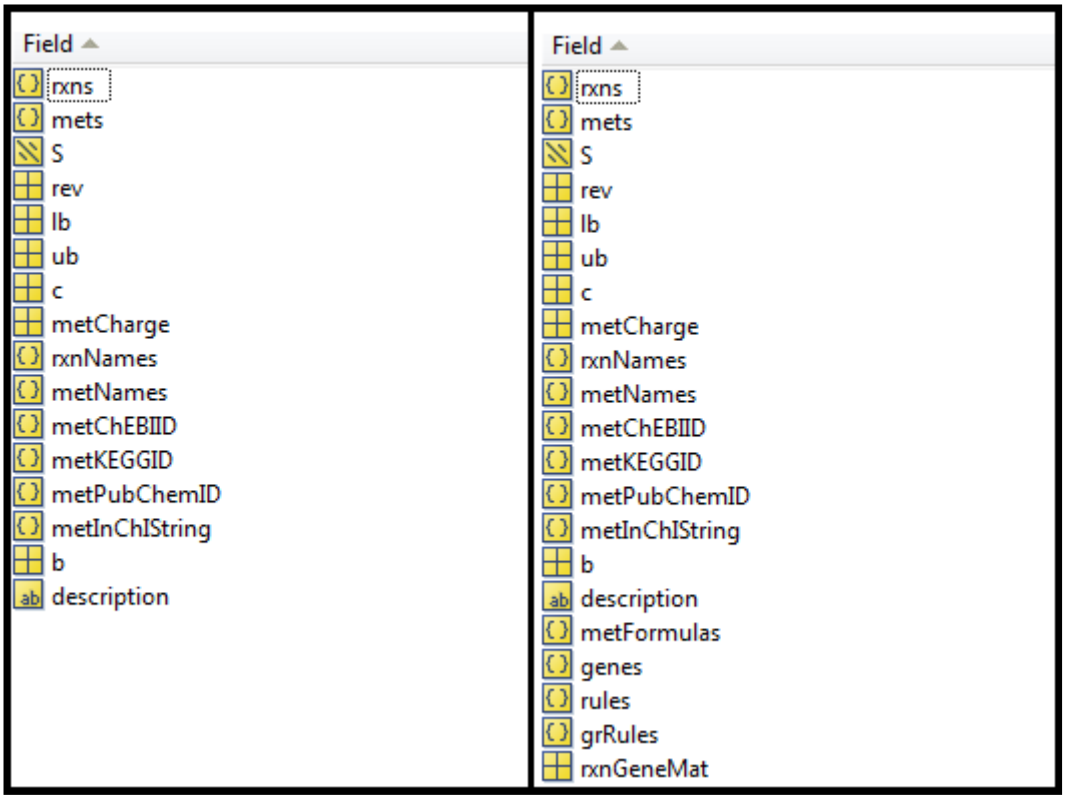

Figure 4-14. Comparison of fields contained in the *Oenococcus oeni* model before and after running the scripts to add the fields metFormulas, genes, rules, grRules and rxnGeneMat. The model loaded into Matlab without any later change lacked important fields such as metFormulas, genes, rules, grRules and rxnGeneMat (left). These fields were added to the model running the scripts *GenerateGeneField.m, GenerateRulesFields.m and GenerateRxnGenMat.m* and *SeparateNamesAndFormulas.m* (right).

4.3.3 Setting of objective function and simulations constraints

Once the model was loaded into a Matlab variable, a demand reaction was added for biomass. The c vector (see description above) was filled with zeros, except for the position of biomass demand reaction which was filled with a number one, indicating that biomass demand reaction is the objective function. On the other hand, simulations constraints were set according to *in silico* experiments performed.

4.4 Network Evaluation

4.4.1 Troubleshooting of non-produced biomass precursors

4.4.1.1 Detecting non-produced biomass precursors.

A critical step in the achievement of a GEM is to make it functional, i.e. that the optimization of the flux balance analysis provides a positive (in the case of maximizing the cost function) or negative (in the case of minimizing the cost function) solution. Since biomass synthesis is generally the objective function (Feist & Palsson, 2010), achieving a functional model is equivalent to ensuring that the flux through the biomass production equation is greater than zero. However, it often occurs that the constructed model is unable to produce biomass, once the biomass equation is introduced. Since databases that containing information about reactions, enzymes and genes are limited, as well as algorithms to annotate the genome, the GENREs created from this data generally contain gaps and orphan reactions.

Gaps are a result of dead-end metabolites. These metabolites appear in the model as only being produced or consumed by reactions and, therefore, will never participate in a feasible solution. They will in turn block any reaction in which they are involved in. There are two classes of dead-end metabolites: i) Root-Non Produced metabolites (RNP) *i.e.* metabolites that are only consumed by the system's reactions and ii) Root-Non-Consumed metabolites (RNC), which include those only produced by the network, but never consumed (Ponce-de-León, Montero, & Peretó, 2013; Satish Kumar, Dasika, & Maranas, 2007).

Several methods to repair gaps and orphan reactions have been developed. A few, such as GapFind/GapFill (Satish Kumar et al., 2007), have become very successful, and have been incorporated in well-known platforms, like COBRA (Becker et al., 2007), Pathways Tools (Karp, Paley, et al., 2002), RAVEN (Agren et al., 2013) and Model

SEED (Henry et al., 2010). Bautista et al (2013) recently proposed a methodology that integrates a genetic algorithm and flux balance analysis to find problematic metabolites whose analysis can led to the addition of new functionalities. Even though these algorithms may be useful in adding new functionalities to the network, their inclusion does not ensure that the generated model will be functional.

To solve this issue, Latendresse, Krummenacker, Trupp, & Karp, 2012 developed a method associated to Pathway Tools, called "multiple gap filling", which suggests corrections to different model components (reactions, metabolites, nutrients or secretions) to allow the model to run. Nevertheless, this method requires that the user previously identify a set of fixed reactions forming a functional model, which does not make it user friendly. On the other hand, one of the limitations of this method is that the visualization of reactions may be confusing, due to the inherent operation of the "Cellular Overview" in Pathway Tools. With the same purpose, Brooks, Burns, Fong, Gowen, & Roberts, 2012 developed an algorithm called FBA-GAP that identifies gaps that avoid the flux through an objective function, suggesting changes to the network in order to obtain a functional model. However, its automatic nature leaves little space for integrating the user's expert knowledge.

We developed a script called *TestPrecursors.m* that finds problematic model metabolites, guiding the user in an iterative process, allowing to decide which elements to repair, with the possibility to integrate their expert knowledge; and not in finding a set of reactions or metabolites that are added to fulfill a certain criterion (e.g. a set that minimizes the cost of adding elements to the model). Generally, to solve just one dead-end biomass precursor, it is necessary to add more than one reaction. It is common that each time a reaction is added to the model, a new dead-end biomass precursor is generated and therefore, it is necessary to repair this new dead-end biomass precursor. To the best of our knowledge, there is currently no method that does this in an iterative

gap-filling process in order to solve, at each step, the new dead-end biomass precursors that are generated.

In our method, the user must first identify the metabolite whose production he or she wants to maximize, generally the biomass. The first step that the algorithm performs is add a demand reaction for that metabolite. Next, the demand reaction is maximized. If there is a positive flux through that reaction, the algorithm finishes because the model is functional. Otherwise, each of the precursors of the metabolite is analyzed. For each precursor, one at a time, a demand reaction is added and then maximized. If the flux through that reaction is non-zero, that means that the precursor can be produced and is labeled as a produced essential metabolite (PEM). Then, the algorithm continues on to the next precursor; if the flux through this next reaction is zero, the algorithm analyzes if there is an existing reaction that could produce this precursor. If no reaction exists, that precursor is labeled as “root non-produced essential metabolite”. On the contrary, if the reaction exists, the precursor is labeled as “downstream non-produced essential metabolite” and the precursor of these precursors are then analyzed in an iterative process. The whole process ends either when all the RNPEM are found, or when the algorithm reaches the depth threshold entered by the user (Figure 4-15).

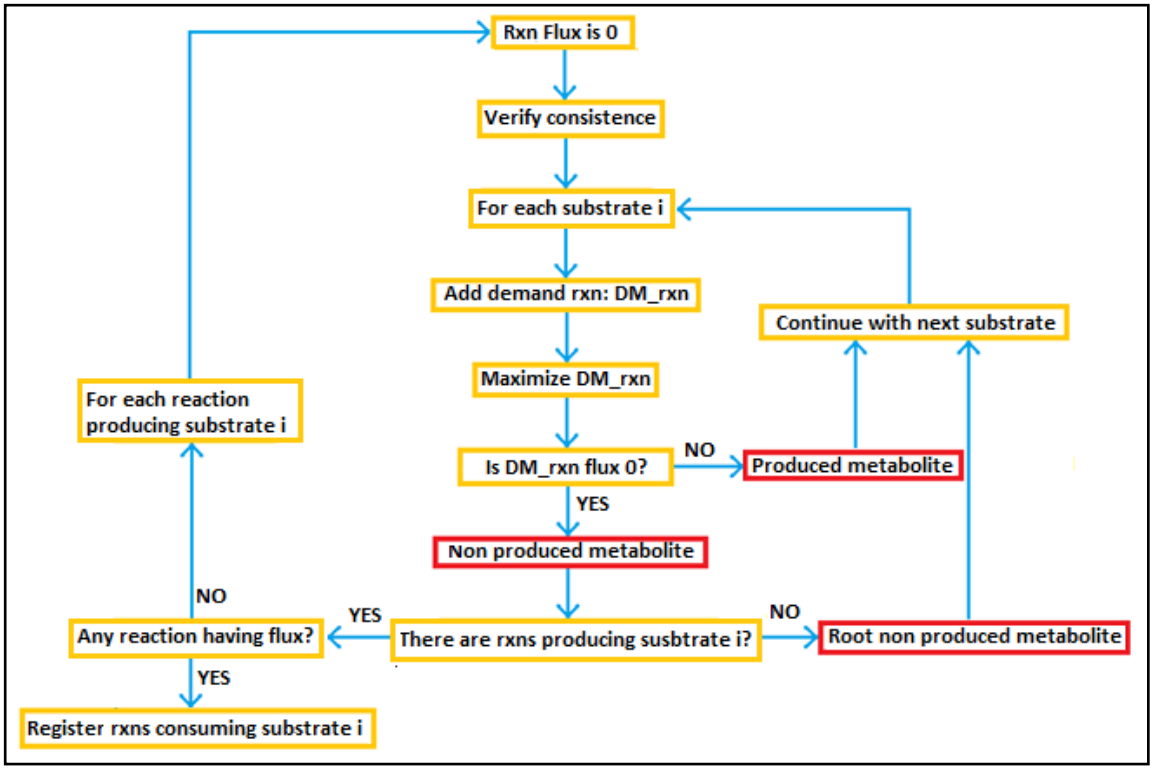

Figure 4-15. Algorithm used for identifying non-produced metabolites in case of having a model with biomass equation but not functional. First a demand reaction is created for the metabolite of interest. Then, the flux for that reaction is maximized. If the flux for the demand reaction is zero, for each substrate in the equation producing the metabolite of interest, a demand reaction is added and maximized. If the flux for the later demand reaction is not zero for the substrate *i*, the substrate *i* is labeled as produced metabolite. Otherwise, the substrate *i* is labeled as non produced metabolite and is analyzed if there are reactions producing the substrate *i*. If there are not reactions producing the substrate *i*, the substrate is labeled as root non produced metabolite. Otherwise, it is analyzed if there is flux through any of the reactions producing the substrate *i*. If there is no reaction having flux, the algorithm analyzes the substrate *i* in a recursive way. The algorithm returns the produced metabolites, the non produced metabolites and the root non produced metabolites that must be reviewed in order to make the model functional.

Our method is one of the few that focuses on achieving a model that could be functional (Brooks et al., 2012; Latendresse et al., 2012), evidencing just the problems that must be solved in the model to make it run. The main advantages of this method, over the methods currently available, are that it is implemented in a widely used platform, it is user friendly and it allows the detection of problems in an iterative manner. Additionally, this method does not require any previous process made by the user - which is the case of MetaFlux for example - making it comparatively even more user-friendly. The execution time is around 150 seconds, faster than other algorithms.

However, once the result is obtained it requires more user intervention, because he/she must decide what action to take to solve the problem encountered in the model, which may slower the whole process. Another disadvantage is that it only allows the analysis of objective functions that produce only one metabolite. Finally, we found that in some cases, despite the existence of reactions producing and consuming a particular metabolite, the metabolite could not be produced because of the mass balancing requirements. For example, Figure 4-16 illustrates that metabolite D cannot be produced even when a feasible consuming reaction exists in the model, even though feasible reactions exists for D precursors. Here, the lack of a feasible reaction consuming C is avoiding flux through the reaction and hence the production of D. This is due to the steady state assumption stating that every produced metabolite must be consumed in the same quantity. In this case, only manual inspection could fix the production of the target metabolite.

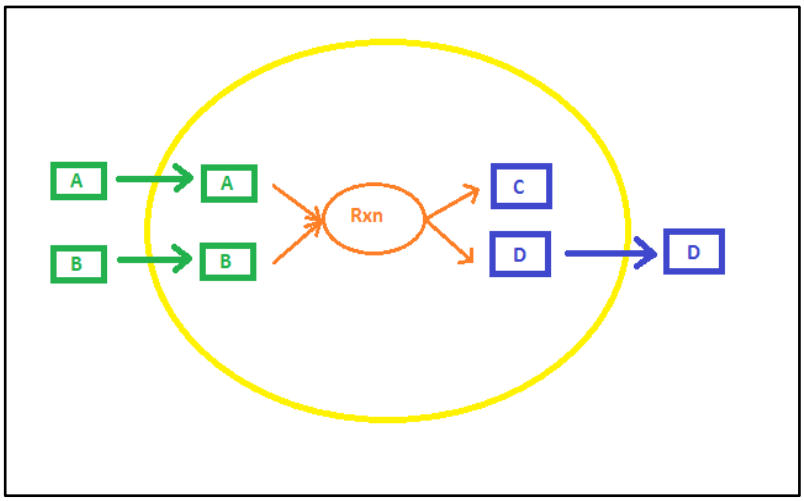

Figure 4-16. Example of an unfeasible metabolic network caused by the lack of a consuming reaction. In this example, the reaction labeled as *Rxn* consumes metabolites A and B and produces metabolites C and D. Despite the reactions supplying substrates A and B being present, if the production of D is maximized, there will be no flux through the reaction *Rxn* because there is no reaction that consumes C. In this case, the network is unfeasible and there will be no solution when FBA is applied because mass balances cannot be accomplished.

The algorithm identified methionine, lysine, histidine, asparagine, phenylalanine, proline, threonine, tryptophan, tyrosine, arginine, leucine, isoleucine, cysteine, valine as non-produced metabolites. Transporters for these amino acids were also incorporated in

order to allow these compounds to be available for metabolism. Oxygen was also identified as a non-produced metabolite. Since *O. oeni* is a microaerophylic bacterium (Mills et al., 2005), a transporter for oxygen was also included for this element.

4.4.1.2 Incorporating non-produced biomass precursors

Once the biomass precursors unable to be synthesized by the metabolic network were identified, the next step is to ensure that these metabolites are actually produced. Consider the case of a reaction consuming metabolites A and B and producing C and D. To ensure that C is produced, it is necessary to fulfill two requirements. First, C and D precursors, namely A and B, must be produced. Second, C and D have to be consumed by the network. Otherwise, the steady state restriction stating that every metabolite produced must be consumed is not fulfilled.

For those identified metabolites to be produced, reactions generating them are introduced. This step is tedious because inspecting the MetaCyc database – or others - for possible reactions generating the target metabolite could be slow. It is common to spend too much time on manual examination of the metabolic pathways producing the target compound. Therefore, faster inspection and identification of these reactions results in an obvious advantage.

To this end, we developed the algorithm *SearchMetaCycReactions.m* that finds all the reaction sets producing the desired compound in the MetaCyc database, and illustrates them in a MATLAB nodes network (Figure 4-17). This illustration allows a rapid visualization of the different alternatives. Additionally, the graph highlights which compounds are in the model and which are not. Thus, the user is able to know what reactions are more suitable for the model's functionality, thus facilitating search for reactions producing the target compound.

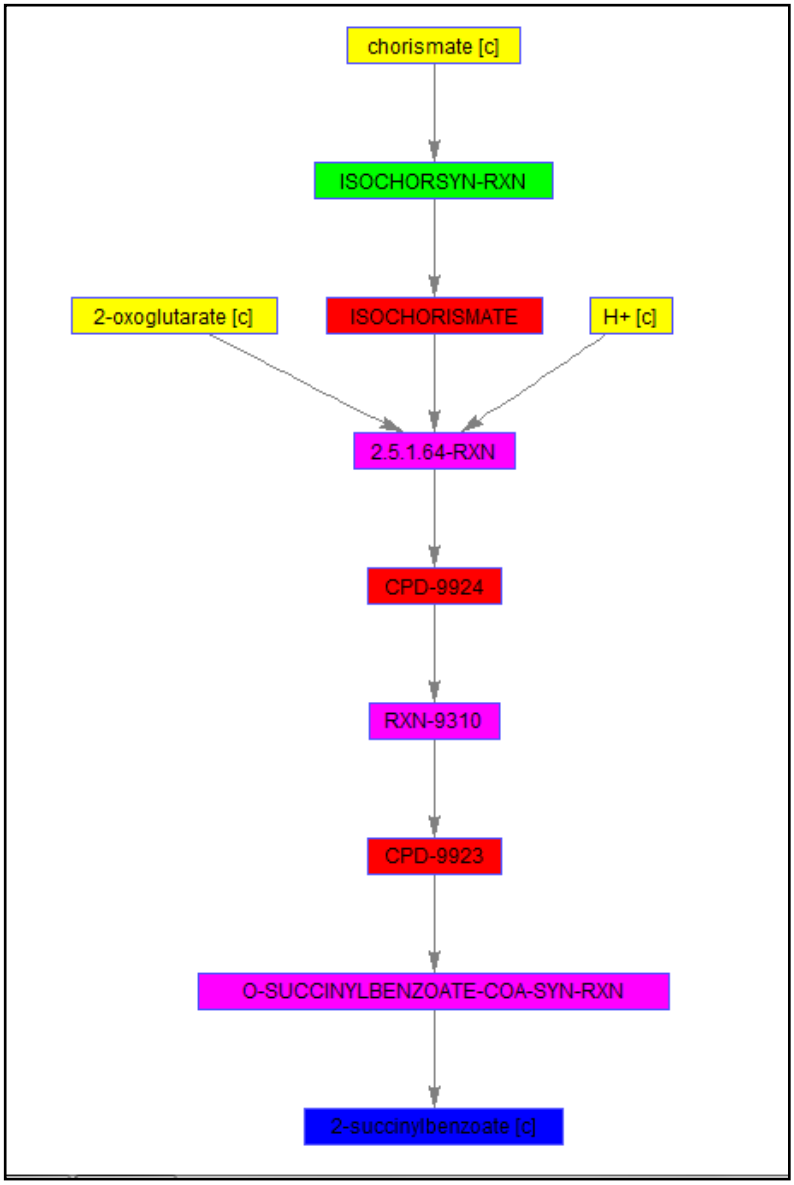

Figure 4-17. Reactions found in MetaCyc leading to the production of 2-succinylbenzoate. The algorithm *SearchMetaCycReactions.m* search in MetaCyc for reactions producing the metabolism of interest and then the results are showed with a node graph. Blue indicates the target metabolite, in this case 2-succinylbenzoate. Red indicates metabolites not present in the model (the names of these nodes correspond to the MetaCyc IDs). Purple indicates reactions utilizing metabolites not present in the model. Yellow indicates metabolites present in the model. Green indicates reactions utilizing metabolites present in the model.

4.4.2 Elimination of reactions for which no evidence exists

During the draft reconstruction process, Pathway Tools automatically added many orphan reactions in order to complete pathways. This is based on the principle that if most reactions of a particular pathway are present in the reconstruction, it is likely that the rest of them should also be present. Nevertheless, there is no bioinformatic evidence supporting their existence. Therefore, all reactions not required for growth and for which neither bioinformatic nor experimental evidence exists to support their presence in the model, were removed.

4.4.3 Genes, reactions and compound deletion analysis

One of the most useful analyses of GEMs is the analysis of single gene deletion. This analysis is particularly interesting in the refining process of GEMs, since it allows the reconciliation of data obtained through simulations *in silico* and *in vivo*. Another useful analysis for model refinement is reaction deletion analysis, which identifies candidate reactions that might be producing false microbial capabilities. These two analyses correspond to steps 79 and 82 of Thiele and Palsson´s protocol.

COBRA Toolbox has scripts for both gene deletion analysis and reaction deletion analysis (Becker et al., 2007; Schellenberger et al., 2011). However, information relating deleted items (genes or reactions) with metabolic pathways is not included in these analyses. Incorporating this information is quite useful as it allows the user to know which metabolic process corresponds to each deleted element (Chavali, Whittemore, Eddy, Williams, & Papin, 2008). For example, it identifies the metabolic pathways whose associated genes, when removed, do not allow growth of the microorganism (hereinafter called essential metabolic pathways). Furthermore, COBRA Toolbox's tools classify results according to whether the microorganism is capable of growing or not, instead of classifying them according to the ratio between microorganism growth with and without the deletion. A classification based on the growth rate ratio has proved

beneficial in the analysis because, with this system, the user is able to disaggregate reactions according to their impact on growth; thus, the user can easily determine those reactions which hamper growth, those that have no effect on growth, as well as those that have a partial impact.

We implemented three scripts in MATLAB, *"GeneDeletionAnalysis" "ReactionDeletionAnalysis" and "CompoundDeletionAnalysis"*, which performed deletion analysis of genes, reactions and compounds, respectively. The main innovation regarding the analysis that can be performed in COBRA Toolbox is that these scripts incorporate information about the relationship between each of these elements and the metabolic pathways in which they participate, allowing a more complete analysis. Incorporating this information is useful as it allows the user to know which metabolic processes are affected by the deletion of elements that are contained in the GEM. Moreover, to the best of our knowledge, software that could provide support for compound deletion analysis has not been developed yet. The analysis provided by this script will be very useful for applications such as drug discovery and targeting.

The script "*geneDeletionAnalysis"* deletes each gene present in the GEM, analyzing what happens to the growth of the microorganism in each case. Unlike the deletion analysis of genes running in COBRA, our script classifies the results in four categories rather than two, allowing more detailed analysis. The classification is based on the resulting ratio between specific growth rate with the deleted gene over specific growth rate without deletion. The resulting value indicates the category that the deleted gene belongs to:

- Essential: $0 \leq \frac{\mu_1}{\mu_2} < 0.1$
- Reduces growth: $0.1 \leq \frac{\mu_1}{\mu_2} < 0.9$
- Slightly reduces growth: $0.9 \leq \frac{\mu_1}{\mu_2} < 1$

- No effect: $\frac{\mu_1}{\mu_2} = 1$

Once the genes are sorted in these categories, the associated reactions in each of the categories are found. Thus, for example, reactions associated to essential genes can be found and then the algorithm quantifies how many reactions are involved in each of the metabolic pathways, resulting in a distributionthat can be graphed as a pie chart.

The script *CompoundDeletionAnalysis* performs the same procedure as above, but compounds are deleted instead of genes. In the script *ReaccionDeletionAnalysis* once reactions are classified in the four categories, the algorithm proceeds directly to quantify how many reactions are involved in each of the metabolic pathways. Then, it generates a pie chart with the distribution.

Applying the script *GeneDeletionAnalysis* to the GEM of *O. oeni PSU-1* revealed that single deletion of most of the genes has no effect on growth (Figure 4-20), which is consistent with other results obtained for other microorganisms. It is worth noting that a higher proportion of genes was found in the category in which genes slightly reduce growth $(0.9 \leq \frac{\mu_1}{\mu_2} < 1)$ than in the category of growth reduction $(0.1 \leq \frac{\mu_1}{\mu_2} < 0.9)$, although the second category covers a broader spectrum nine-fold. This was an unexpected, non trivial result, which confirmed the benefit of including four categories instead of two.

Once the reactions associated to essential genes are found, the algorithm counts the reactions involved in each pathway. This distribution shows that the majority of essential genes fall into three metabolic processes: lipid, peptidoglycan and nucleic acids metabolism, highlighting how critical these processes are to the overall function of the microorganism.

Reaction deletion analysis (Figure 4-18) showed that the removal of most reactions (81%) in the model had no effect on growth. A similar result was found for gene

deletion analysis (Figure 4-20), which showed that the removal of most genes (96%) does not affect growth. Both results are consistent with the general concept of metabolic network robustness, stating that in most cases the biochemical network present in an organism would be robust enough to maintain cell growth even if a particular reaction or gene is removed.

On the other hand, 15% of model reactions were considered to be essential. 38% of the essential reactions are related to fatty acid biosynthesis: 20% are related to palmitate, 7% to palmitoleate, 6% to cis-dodecenyl, 5% to cis-vaccenate and 4% to stereate biosynthesis. Other essential reactions are related to peptidoglycan biosynthesis (15%), purine (5%) and pyrimidine (5%) biosynthesis (Figure 4-19). The algorithm groups all of the metabolic processes that represent less than a 3% in the group "other metabolic processes" in order to visualize which processes are more sensitive to the reaction deletion analysis. The same principle is applied to the gene deletion analysis. The reactions that fall into the category of "other metabolic processes" are related to the biomass production equations. It is expected that if one of the biomass related reactions is removed, the model lacks its functionality.

A similar situation occurs with the gene deletion analysis (Figure 4-21) 34% of reactions associated to essential genes are related to fatty acid biosynthesis: 21% are related to palmitate, 6% to palmitoleate, 4% to stearate and 3% to cis-dodecenoyl biosynthesis. 12% of reactions are related to peptidoglycan biosynthesis. Surprisingly, here in the gene deletion analysis, heterolactic fermentation emerges as a sensitive metabolic process susceptible to be disrupted by deletion of genes.

There are 3% of the reactions that reduce growth. The most affecting reactions are related to CO2 which reduce growth down to 45% of the original value. This could be explained by the fact that *O. oeni* is a heterofermentative bacterium, mechanism whereby the organism is able to obtain energy and regenerate redox cofactors. If reactions associated to CO_2 are removed, some important reactions of the heterolactic

fermentation will not be active, therefore it is expected that the cell will not be able to generate the same energy levels.

For *O. oeni* model, the reaction deletion analysis required 86 seconds to run, the deletion gene analysis 110 seconds, and the compound deletion analysis 74 seconds. Each algorithm creates an excel spreadsheet containing, in separate sheets, the elements (genes, reactions and compounds, respectively) that are essential; elements that reduce growth; elements that slightly reduce, and have no effect on growth; and, eventually, the elements that promote growth when removed.

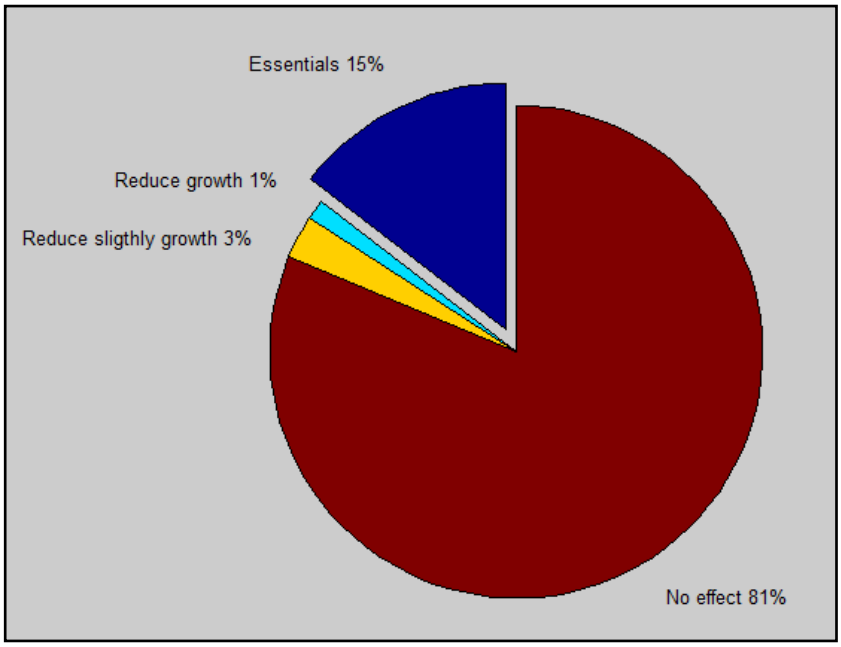

Figure 4-18. Reaction deletion analysis for genome-scale metabolic model of *Oenococcus oeni*. Most reactions (81%) have no effect on growth as was expected based on the fact that metabolic network of microorganisms are robust enough to maintain growth even when a particular metabolic function is deleted. On the other hand, 15% of the reactions are considered to be essential as their individual deletion causes no growth. In the middle, only 3% of reactions slightly decrease growth (90-100%) and 1% of reactions strongly decrease growth (0-90%). These two last groups of reactions are interesting as it could provide insight in order adjust predicted growth rate.

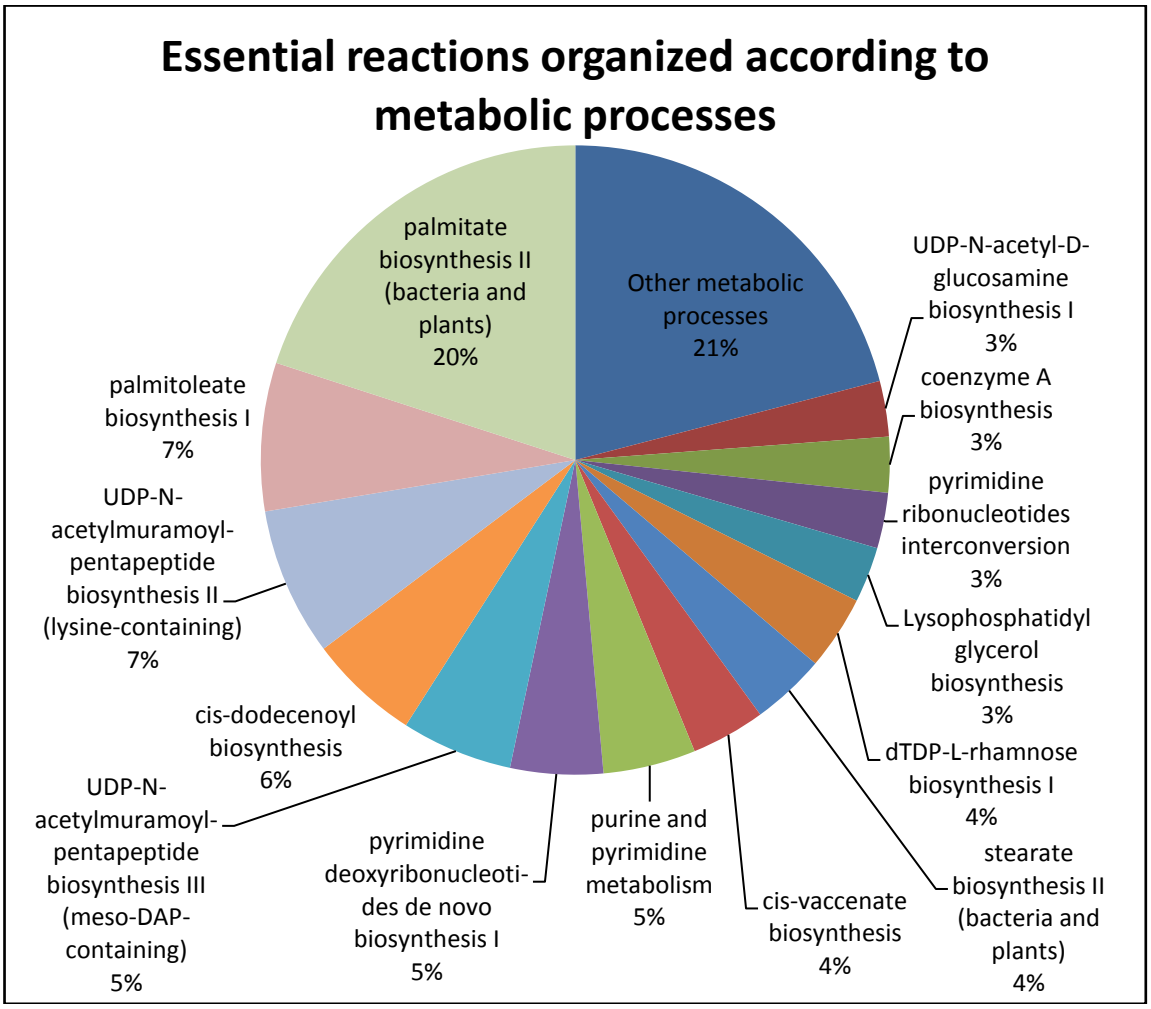

Figure 4-19. Essential reactions organized according to metabolic processes. The most fragile metabolic processes susceptible to disruption by removal of a particular reaction are those related to fatty acid biosynthesis

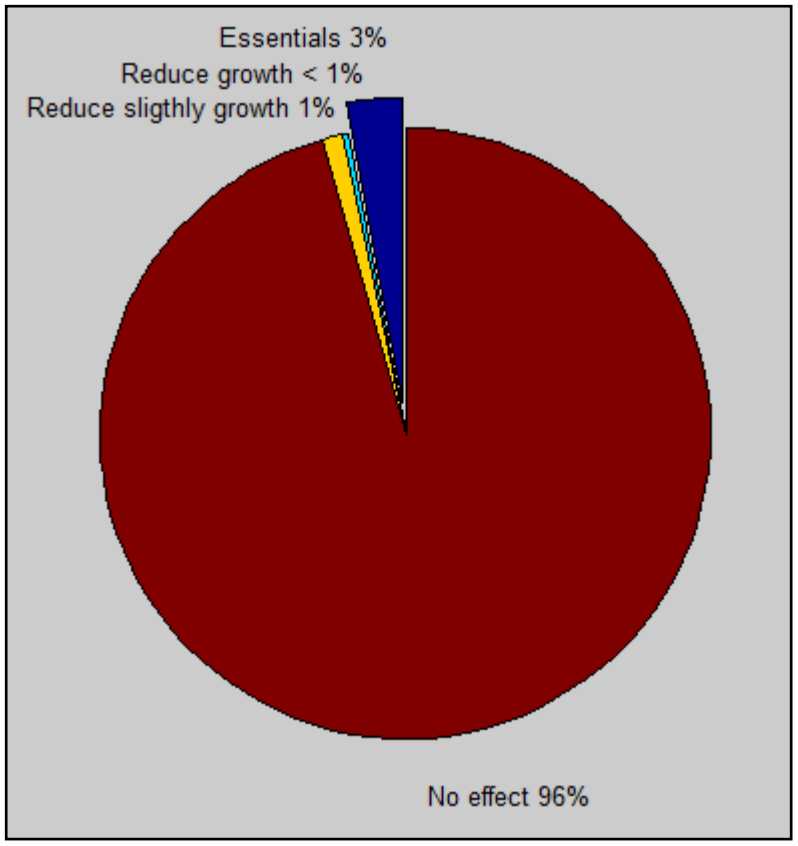

Figure 4-20. Gene deletion analysis for genome-scale metabolic model of *Oenococcus oeni*. Most genes (96%) have no effect on growth as was expected based on the fact that the metabolic network of microorganisms is robust enough to maintain growth even when a particular metabolic function is deleted. On the other hand, 3% of the genes are considered to be essential as their individual deletion causes no growth. In the middle, only 1% of genes slightly decrease growth (90-100%) and less than 1% of genes strongly decrease growth (0-90%). These two last groups of reactions are interesting as they could provide insights on adjusting predicted growth rate.

.

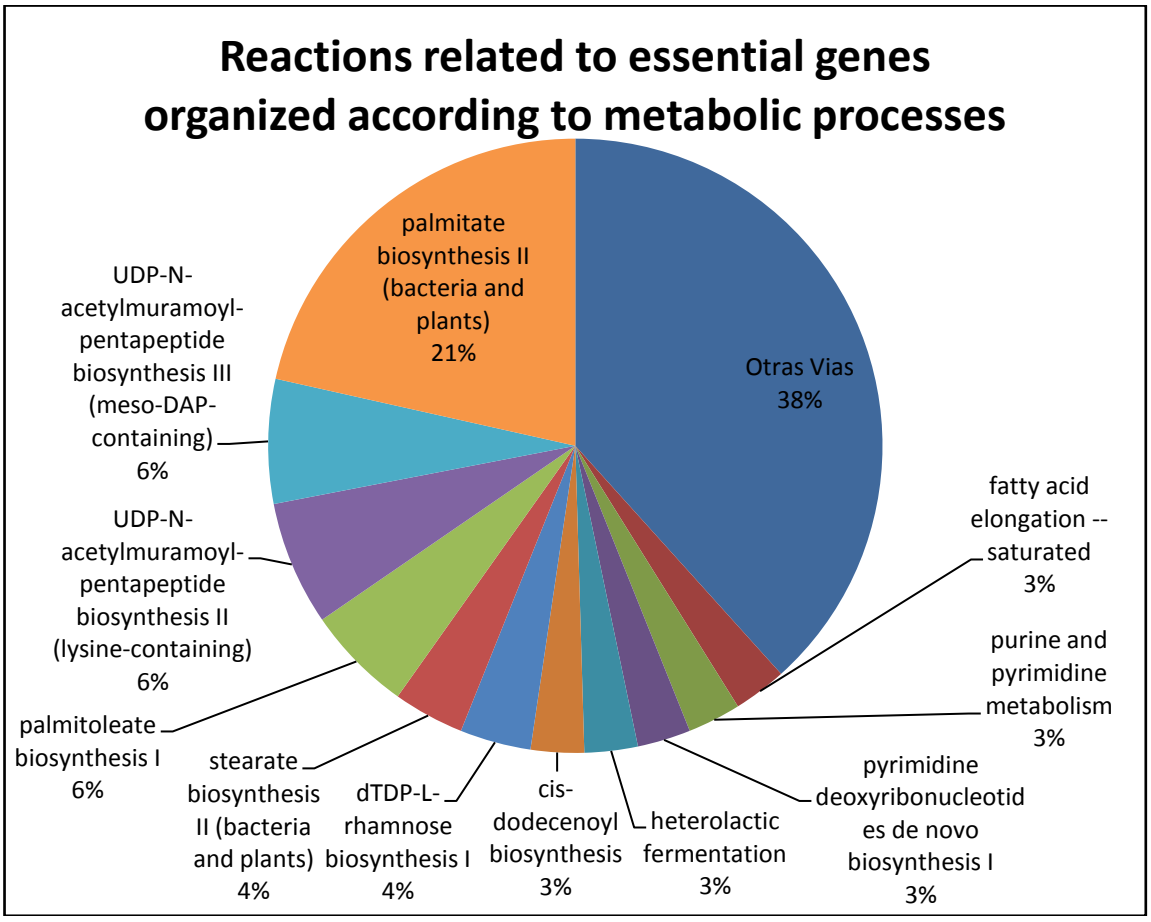

Figure 4-21. Essential reactions organized according to metabolic processes. The most fragile metabolic processes susceptible to be disrupted through removal of a particular reaction are those related to fatty acid biosynthesis. Heterolactic fermentation appears as a sensitive metabolic process that could be affected by the deletion of genes.

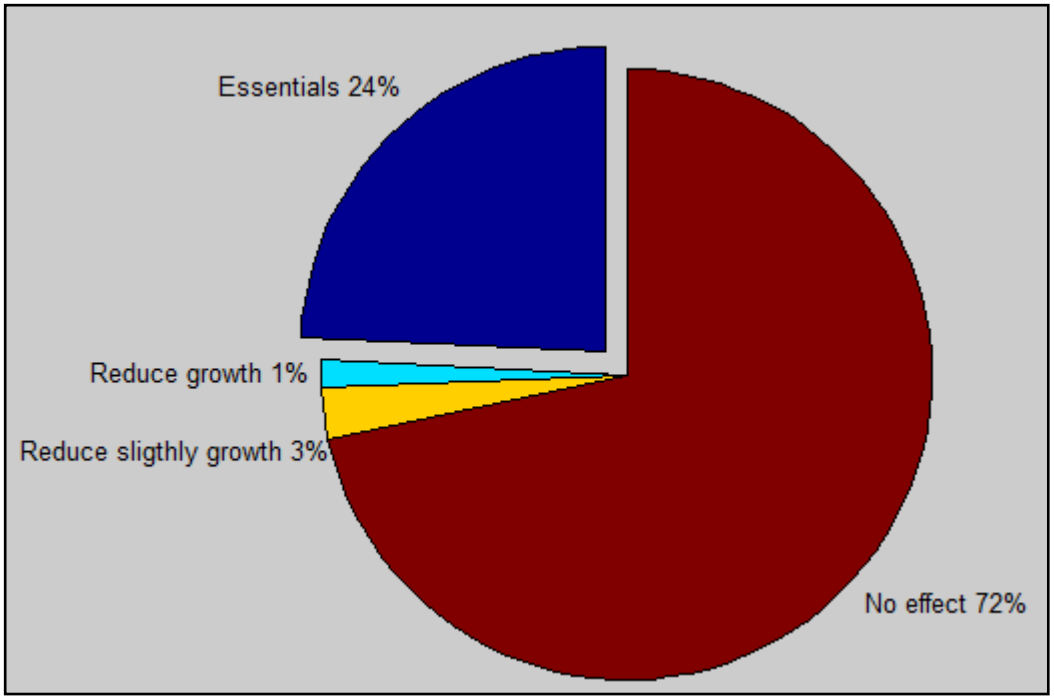

Figure 4-22. Compound deletion analysis for genome-scale metabolic model of *Oenococcus oeni*. Most compounds (72%) have no effect on growth as was expected based on the fact that metabolic networks of microorganisms are robust enough to maintain growth even when a particular component is deleted. In this case the percentage of essential compounds is smaller than the percentage of essential reactions or genes because in most causes a particular compound is involved in more than one reaction and its deletion causes the elimination of all the reactions that participate. On the other hand 24% of the compounds are considered essential as their individual deletion causes no growth. In the middle, only 3% of compounds slightly decrease growth (90-100%) and 1% of compounds strongly decrease growth (0-90%). These two last groups of reactions are interesting as they could provide insight adjusting predicted growth rate.

4.4.4 Analysis of minimal requirements

A minimal medium is a medium which the removal of any of the medium components abolishes the growth of the microorganism completely (Wegkamp, Teusink, de Vos, & Smid, 2010). Determining a minimal medium is useful to achieve a comprehensive understanding of the nutritional requirements of a microorganism, which is scientifically interesting per se but would also lead to the optimization of the culture medium conditions (Fan et al., 2014). If the minimal medium is known, only the necessary nutrients could be used to generate a culture medium, instead of using the large list of nutrients found in typical culture mediums. This saves laboratory supplies and culture medium preparation time.

If the *in vivo* minimal medium is not known, the *in silico* determination of a minimal medium using a GEM is of paramount importance because it would lead to the driven generation of hypotheses to design a minimal medium which could be tested experimentally. On the other hand, if the *in vivo* minimal medium is known and is consistent with the *in silico* minimal medium, this is a clear signal that a comprehensive understanding of nutritional requirements has been achieved (Becker & Palsson, 2005; Fan et al., 2014; Schilling et al., 2002; Wegkamp et al., 2010). It is because of this reason that the determination of *in silico* minimal requirements could be used to validate the quality of a constructed GEM. (Teusink et al., 2005).

Additionally, analyzing the minimal medium components is particularly useful during the refinement stage (step 37 in Palsson & Thiele´s protocol), because it allows for the comparison of the minimal nutrients required to sustain growth *in vivo* versus those that sustain growth *in silico*. Inconsistencies between *in vivo* and *in silico* results must be fixed during the refinement process. For example, if a certain metabolite is essential for growth *in vivo*, and the minimal medium determined *in silico* does not include it, it is necessary to fix the metabolic network so that the model would correctly describe the experimental behavior.

On the other hand, for poorly studied organisms - or those for which it is difficult to obtain a significant biomass in the laboratory -, determining an *in silico* minimal medium can provide insights on the nutritional requirements of the organism *in vivo*. The latter will be very valuable to optimize the culture medium in order to maximize biomass production or a product of interest.

The essential nutritional requirements in different strains of *O. oeni* have been studied for at least four decades (Fourcassie, Belarbi, & Maujean, 1992; Garvie, 1967; Remize et al., 2006; Terrade & Mira de Orduña, 2009). Therefore this data will be used to validate the developed GEM of *O. oeni.* Since no study determining nutritional requirements has been performed systematically in PSU-1, we considered that a nutrient is essential if it is essential for most strains of *O. oeni*. Under this criterion, 10 amino acids resulted to be essential: arginine, cysteine, glutamic acid, histidine, isoleucine, methionine, phenylanaline, tryptophan, tyrosine and valine (Table 4-4).

On the other hand, others types of nutrients such as carbon sources, vitamins, nucleotides and minerals have also been studied for different strains of *O oeni* (Garvie, 1967; Terrade & Mira de Orduña, 2009). Acordance between the two studies. Nevertheless some differences have been noted for folic acid and riboflavin. In these cases the results of Terrade & Mira de Orduña were considered to be more reliable because of the careful methodology used. Specifically, that sudy used a chemical medium providing significant growth (Terrade, Noël, Couillaud, & de Orduña, 2009) thus allowing for large dose responses upon omission of medium consituents. Aditionally, that work implemented three washing steps between all transfers as well as small inoculation sizes (initial OD 0.03-0.05 equivalent to 5-8 x 10^4 CFU ml^{-1}) avoiding the risk of carrying over of medium constituents into deficient media which could lead to false negative classification of nutrients as non-essential. Under this criterion, D-

ribose, nicotinic acid, pantothenate, manganse sulfate and phosphate are considered to be essential for *O. oeni* (Table 4-5).

Table 4-4. *In vivo* amino acid requirements for different strains of *Oenococcus oeni* and comparison with *in silico* nutritional requirements. Columns two to five show the number of strains that require a particular amino acid for each of the four studies analyzed. The sixth column shows the total number of strains that require a particular amino acid. If more than half of the strains require a particular amino acid, that amino acid is considered to be essential in the seventh column. Last column shows *in silico* nutritional requirements for each amino acid.

Amino acid	(Garvie, 1967)	**(Fourcassie et al, 1992)**	**(Remize et al, 2006)**	**(Terrade et al, 2009)**	**Total**	**Essential *in vivo***	**Essential *in silico***
Alanine	0/9	0/6	0/5	0/2	0/22	No	Yes
Arginine	9/9	6/6	3/5	2/2	20/22	Yes	Yes
Asparagine	-	-	-	1/2	1/2	?	Yes
Aspartic acid	0/9	0/6	1/5	0/2	1/22	No	No
Cysteine	9/9	3/6	1/5	2/2	15/22	Yes	Yes
Glutamic acid	9/9	6/6	5/5	1/2	21/22	Yes	Yes
Glutamine	-	-	-	0/2	0/2	No	No
Glycine	2/9	0/6	0/5	2/2	4/22	No	No
Histidine	6/9	1/6	3/5	2/2	12/22	Yes	Yes
Isoleucine	9/9	6/6	3/5	2/2	20/22	Yes	Yes
Leucine	1/9	3/6	4/5	2/2	10/22	No	Yes
Lysine	0/9	1/6	1/5	0/2	2/22	No	Yes
Methionine	4/9	5/6	5/5	2/2	16/22	Yes	Yes
Phenylalanine	7/9	1/6	5/5	2/2	15/22	Yes	Yes
Proline	0/9	0/6	0/5	2/2	2/22	No	No
Serine	2/9	0/6	5/5	1/2	8/22	No	No

Threonine	1/9	0/6	1/5	2/2	3/22	No	Yes
Tryptophan	7/9	6/6	4/5	2/2	19/22	Yes	Yes
Tyrosine	9/9	1/6	5/5	2/2	17/22	Yes	Yes
Valine	9/9	1/6	4/5	2/2	16/22	Yes	Yes

Table 4-5. *In vivo* nutritional requirements for different strains of *Oenococcus oeni* and comparison with *in silico* nutritional requirements. Columns two to four show the number of strains that require a particular nutrient for each of the 2 studies analyzed. If discrepancies exists between the two studies, results from Terrade et al., 2009. Last column shows *in silico* nutritional requirements for each amino acid.

Nutrient	**(Terrade et al, 2009)**	(Garvie, 1967)	**Essential *In vivo***	**Essential *in silico***
Carbon sources				
D-ribose	2/2	-	Yes	Yes
Vitamins				
Aminobenzoic acid	0/2	0/5	No	No
Biotin	0/2	-	No	No
Choline chloride	0/2	-	No	No
Cobalamin	0/2	0/5	No	No
Folic acid	0/2	5/5	No	No
Nicotinic acid	2/2	-	Yes	No
Pantothenate	2/2	-	Yes	Yes
Pyridoxine	0/2	0/5	No	No
Riboflavin	0/2	5/5	No	No
Thiamine	0/2	-	No	No
Nucleotides				

Adenine	0/2	0/5	No	No
Guanine	0/2	0/5	No	No
Xanthine	0/2	0/5	No	No
Cytosine	0/2	-	No	No
Thymine	0/2	-	No	No
Uracil	0/2	1/5	No	No
Minerals				
Manganse sulfate	2/2	-	Yes	No
Magnesium sulfate	0/2	-	No	No
Potassium dihydrogenphosphate	2/2	-	Yes	Yes
Calcium chloride	0/2	-	No	No
Copper sulfate	0/2	-	No	No
Ferrous sulfate	0/2	-	No	No
Zinc sulfate	0/2	-	No	No
Others				
Guanine + adenine + Xanthine + uracil	-	5/5	Yes	Yes
Guanine + adenine + Xanthine + uracil + Cytosine + thymine	0/5	-	No	Yes
Nicotinic acid + thiamine + Pantothenate + biotin	-	5/5	Yes	Yes

To be able of compare *in vivo* and *in silico* nutritional requirements it is necessary to determine an *in silico* minimal medium. The first algorithm for such a purpose was developed by Schilling et al. (2002) to determine the minimum nutrients of *Helicobacter pylori,* a microaerophilic bacterium responsible for gastric ulcers. Beginning with all of

the extracellular metabolites available to the metabolic network, the algorithm individually removes each of the metabolites to determine which are required for producing biomass constituents. This determination is accomplished by constraining the exchange flux or uptake reaction for the metabolite to zero and optimizing for the biomass objective reaction. After thorough examination, a set of metabolites is defined for which their individual removal renders the network unable to produce the biomass demands. This set of metabolites constitutes a defined minimal medium required by the *in silico* model to support growth.

Despite the usefulness of such an algorithm, it was not implemented in any platform; therefore, it cannot be easily employed by others. This is an important issue because the usability of an algorithm depends directly on its availability in a well known and user friendly platform. Another issue of this algorithm is that it does not detect groups of essential metabolites which is the case of nucleotides for *O. oeni*. Garvie (1967) determined that nucleotides are not essential for *O. oeni* when they are omitted separately. However, they cause no growth when they are omitted together.

Another algorithm aiming to find *in silico* minimal medium was recently published by Eker et al (2013). The authors reported a complex algorithm able to find minimum nutritional nutrients that support growth of a microorganism, making assumptions that correctly simulate the consumption and production of metabolites. However, the complexity of the algorithm results in a very long execution time (3 days). When refining a GEM and reconciling the experimental data with the simulations, an algorithm that is able to deliver results in minutes is fundamental, because it is necessary to rapidly assess if modifications of the metabolic network results in better congruence between experimental and computational data. Furthermore, the algorithm was not implemented in any widely used platform; therefore, its use is cumbersome.

We developed a script in MATLAB, "*FindMinimalMedium*" (Figure 4-23) to find the minimal medium supporting growth of a microorganism *in silico.* Our script finds the minimal components necessary for the organism to generate biomass by simulating successive transfers. First, the microorganism is grown in a culture medium with N defined nutrients. Then, the first nutrient is removed from the simulated medium. If the organism is still able to grow in this new environment, the nutrient removed is classified as non-essential and it is removed from the simulated medium for the following iterations. On the contrary, if the microorganism is unable to grow in the new medium, the nutrient removed is classified as essential and kept in the medium for the next iterations, becoming part of the minimal medium. This process continues until all the N nutrients are evaluated. Additionally, for each nutrient, alternative nutrients are determined (Figure 4-24), *i.e.* nutrients that when replacing the original one, it is still possible to obtain a minimal medium in which the microorganism is able to produce biomass. For example, if the organism is capable of growing on glucose, fructose or ribose independently, the alternative sources for glucose identified by the algorithm are fructose and ribose. The advantage of knowing what substrates can be replaced is that an alternative defined culture medium could be designed.

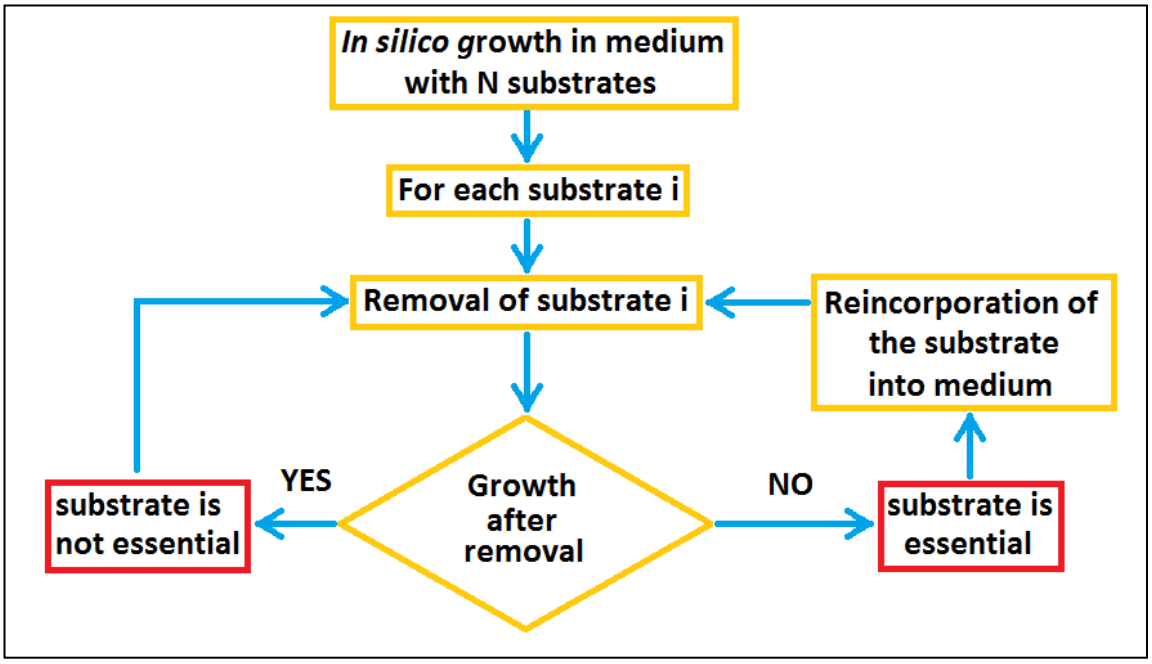

Figure 4-23. Algorithm implemented in MATLAB to determine minimal medium through the simulation of successive transfers. Before running the algorithm the model must be able to produce biomass in a chemically defined medium. The algorithm consists in removing each substrate one at a time and in turn. Once a particular substrate is removed the model is optimized and the growth is analyzed. If the model is able to produce biomass the substrate is considered to be a non-essential nutrient and the next substrate is removed. Otherwise, the substrate is considered to be an essential nutrient and the substrate is reincorporated into the medium before removing the next substrate. The algorithm finishes when all the substrates are analyzed and it returns the minimal medium for which the suppression of growth is achieved if one of the nutrients is removed

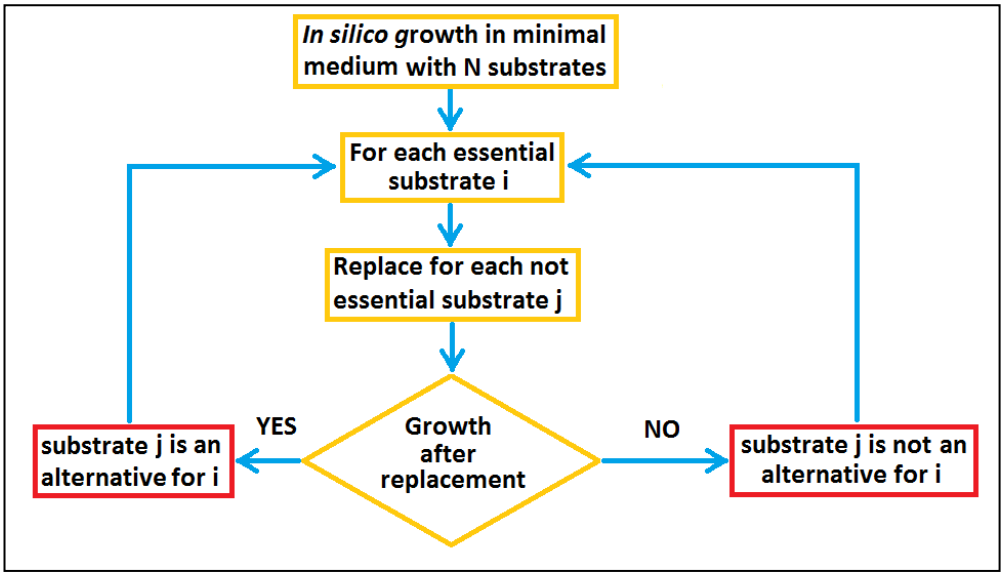

Figure 4-24. Algorithm implemented in MATLAB to determine alternatives nutrients of minimal nutrient requirements. Because the *in silico* minimal medium determined with the script *FindMinimalMedium* depends on the order in which the nutrients are omitted, alternative nutrients are determined replacing each of the essential

substrates for each of the nutrients that were not considered to be essential. If the model is able to produce biomass replacing substrate i for substrate j, the substrate j is an alternative for substrate i. The algorithm finishes when all substrates are analyzed and it returns the alternatives for each nutrient present in the minimal medium.

Unlike the algorithm of Schilling et al. (2002), our algorithm is implemented in Matlab, a well known platform. Additionally, it is able to identify essential groups of metabolites as well as interchangeable metabolites. Moreover, while our algorithm is much simpler than the one presented by Eker et al. (2013), primarily by assuming that the cells begin as empty bags of metabolites, it allows rapid (within minutes) estimates of the minimal medium, which is valuable in the refining process.

The determined *in silico* minimal medium is composed of 21 nutrients: 17 amino acids, 1 vitamin, 1 carbon source, 2 nucleotides and 1 mineral (Table 4-6). Only four of these nutrients could be replaced by others while maintaining the minimal medium (Table 4-7). What remains to be experimentally probed is that this medium is in fact a minimal medium for *O. oeni* PSU-1.

Table 4-6. *In silico m*inimal culture medium determined using the script *FindMinimalMedium.m*. The script *FindMinimalMedium.m* simulates an experiment of successive transfers, returning an *in silico* minimal medium, i.e. a medium for which if one of the nutrients is removed, *Oenococcus oeni* is not able to grow. Different groups of nutrients can be found in this minimal medium such as vitamins, amino acids, carbon sources, nucleotides and minerals.

Minimal nutrients	Minimal transport reactions
(R)-pantothenate	Exchange_(R)-pantothenate
L-alanine	Exchange_L-alanine
L-arginine	Exchange_L-arginine
L-asparagine	Exchange_L-asparagine
L-aspartate	Exchange_L-aspartate

L-cysteine	Exchange_L-cysteine
L-glutamate	Exchange_L-glutamate
L-histidine	Exchange_L-histidine
L-isoleucine	Exchange_L-isoleucine
L-leucine	Exchange_L-leucine
L-lysine	Exchange_L-lysine
L-methionine	Exchange_L-methionine
L-phenylalanine	Exchange_L-phenylalanine
L-threonine	Exchange_L-threonine
L-tryptophan	Exchange_L-tryptophan
L-tyrosine	Exchange_L-tyrosine
L-valine	Exchange_L-valine
α -D-glucose	Exchange_alpha-D-glucose
Glycine	Exchange_glycine
Guanine	Exchange_guanine
Phosphate	Exchange_phosphate

Table 4-7. Alternatives for nutrients present in the *in silico* minimal medium. Only four nutrients present in the *in silico* minimal medium could be replaced by others in order to maintain a minimal medium. Other 9 carbon sources could replace α-D-glucose

Nutrient	Alternative nutrients
L-aspartate	L-serine
Glycine	L-serine
Guanine	Adenine
α -D-glucose	β-D-fructofuranose, β-D-galactose, β-D-glucose, β-D-ribofuranose, cellobiose, lactose, melibiose, raffinose, trehalose.

In order to compare the *in silico* nutritional requirements with experimental data, an *in silico* single omission experiment using the chemically defined medium determined by (Terrade et al., 2009) was performed. This experiment consisted in removing each nutrient from the medium separately and analyzing if the microorganism is able to grow after its removal. Nutrients that inhibit growth when removed are considered to be essential.

Thus, the model predicts that arginine, cysteine, histidine, isoleucine, methionine, phenylalanine, tryptophan, tyrosine and valine are essential aminoacids (Table 4-8), which fully concurs with what has been shown experimentally in most strains of *O. oeni* (Fourcassie et al., 1992; Garvie, 1967; Remize et al., 2006; Terrade & Mira de Orduña, 2009). Interestingly, the model predicts that glutamic acid is not essential in the medium used but becomes essential when glutamine is removed, which also can be observed *in vivo*. Fourcassie et al. (1992), Garvie (1967), Remize et al. (2006) showed that glutamic

acid was essential using a medium lacking glutamine and Terrade & Mira de Orduña (2009) found than when glutamine is present in the medium *O. oeni* is able to grow, suggesting that glutamic acid can be synthesized from glutamine in *O. oeni*.

On the other hand, alanine, aspartic acid, glutamine, glycine, leucine, lysine, proline, threonine and serine have been shown to be non-essential nutrients *in vivo* (Fourcassie et al., 1992; Garvie, 1967; Remize et al., 2006). Among these, aspartic acid, glutamine, glycine, proline, and serine were also non essential nutrients *in silico* while alanine, lysine, leucine and threonine were erroneously classified as essential nutrients. It is not clear if asparagine is an essential nutrient or not, although Mills et al. (2005) suggested that this amino acid is essential for growth, which would be consistent with the *in silico* analysis.

Table 4-8. *In silico* essential nutrients for Oenococcus oeni. A nutrient is essential if its omission from the chemically defined medium resulted in no growth of the microorganism

Essential nutrients	Essential transport reactions
(R)-pantothenate	Exchange_(R)-pantothenate
L-alanine	Exchange_L-alanine
L-arginine	Exchange_L-arginine
L-asparagine	Exchange_L-asparagine
L-cysteine	Exchange_L-cysteine
L-glutamate*	Exchange_L-glutamate
L-histidine	Exchange_L-histidine
L-isoleucine	Exchange_L-isoleucine
L-leucine	Exchange_L-leucine
L-lysine	Exchange_L-lysine
L-methionine	Exchange_L-methionine
L-phenylalanine	Exchange_L-phenylalanine
L-threonine	Exchange_L-threonine
L-tryptophan	Exchange_L-tryptophan
L-tyrosine	Exchange_L-tyrosine
L-valine	Exchange_L-valine
phosphate	Exchange_phosphate
β-D-ribofuranose	Exchange_beta-D-ribofuranose

* L-glutamate was considered to be essential because it is required for growth in mediums lacking L-glutamine, which are commonly used. Nevertheless, we must consider that glutamate is not an essential

nutrient for *O. oeni* since Terrade & Mira de Orduña (2009) have demonstrated that some strains of *O. oeni* are able to grow in a medium lacking glutamate yet containing glutamine.

Alanine could not be synthesized by the model because the reaction that synthesizes it from L glutamate (E.C. 2.6.1.2) requires another reaction that consumes 2-oxoglurate in order to satisfy mass balances, and the model lacked this reaction. In fact, adding a demand reaction for 2-oxoglutarate allowed the model to synthesize L-alanine, making it a non-essential amino acid. 2-oxoglutarate is a compound that participates in the TCA cycle but, as *O. oeni* lacks a fully functional TCA cycle, it made sense to search for other mechanisms that consume it.

Lysine could not be synthesized because of three reasons: 1) The reaction that synthesizes N-succinyl-2-amino-ketopimelate (E.C. 2.3.1.117), an indispensable metabolite for lysine biosynthesis, requires the production of succinyl-CoA in order to satisfy mass balances 2) The same reaction requires the consumption of coenzyme A in order to satisfy mass balances. 3) The next step (E.C. 2.6.1.17) in the lysine biosynthesis pathway (Figure 4-25) also requires a reaction that consumes 2-oxoglutarate. Once again, adding a demand reaction for 2-oxoglutarate and coenzyme A and a sink reaction for succinyl-CoA allows the model to synthesize lysine, transforming it in a non-essential amino acid. Nevertheless, the incorporation of the previously analyzed sink and demand reactions affected the accuracy of specific consumption/production rate predictions. Therefore, further refinement is necessary to find reactions whose incorporation makes L-alanine and lysine non essential metabolites while maintaining the same accuracy of specific consumption/production rate predictions.

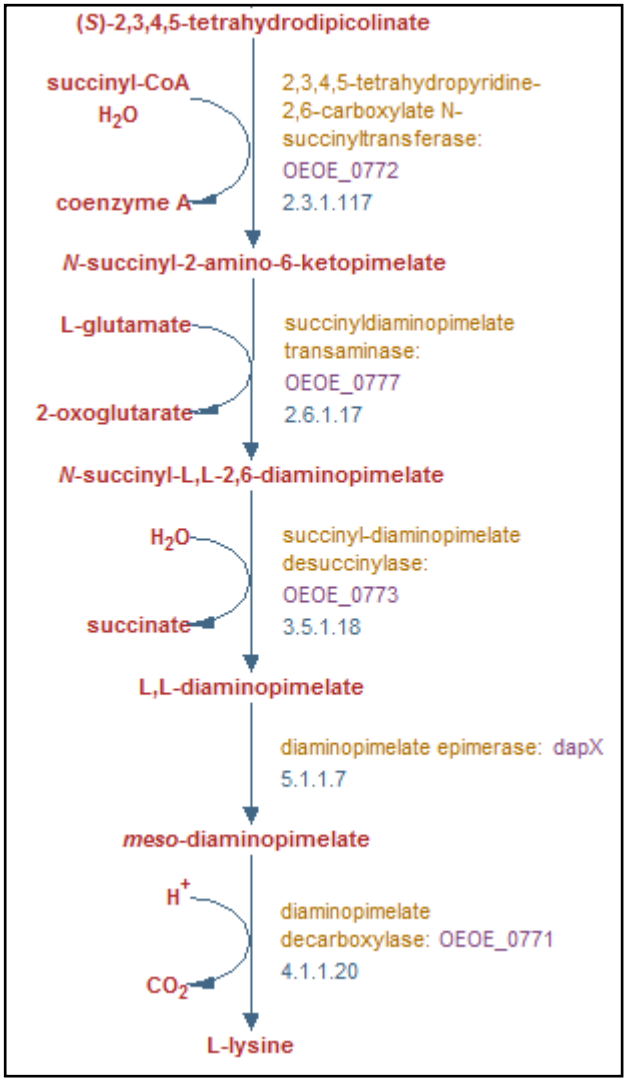

Figure 4-25. Partial lysine biosynthesis pathway. Here, the last five reactions of the lysine biosynthesis pathway are shown. None of these reactions are able to carry flux when the model is optimized because the model lacks feasible reactions producing succinyl-CoA and consuming conenzyme A and 2-oxoglutarate. Sink and demand reactions were tested for these metabolites showing that their incorporation allows the biosynthesis of L-lysine. Unfortunately their incorporation also diminishes the accuracy of specific consumption/production rates predictions so reactions whose incorporation makes L-lysine a non essential metabolites while maintaining the same accuracy of specific consumption/production rate predictions remains a pending challenge. This figure was made using Pathway Tools.

Most strains of *O. oeni* are able to synthesize L-leucine; however the strain PSU-1 contains a pseudogene (OEOE_1724) for the 3-isopropylmalate dehydrogenase (E.C. 1.1.1.85), an enzyme required for the biosynthesis of this amino acid; therefore, bioinformatic evidence suggests that this strain is not able to synthesize it. On the other

hand, there is only one strain that has been reported to be unable to synthesize threonine (Garvie, 1967). However, *O. oeni* PSU-1 strain contains a pseudogene for the homoserine kinase (E.C. 2.7.1.39), an enzyme required to synthesize threonine; therefore, this strain would not be able to synthesize this amino acid.

Concerning vitamins, the *in silico* analysis of minimal requirements shows that pantothenate is essential for growth which is consistent with experimental data (Garvie, 1967). Richter et al (2001) found that *O. oeni* requires pantothenic acid for growth because, under pantothenate limitation, phosphotransacetylase, and in particular acetaldehyde dehydrogenase activities, became limited due to low levels of the co-substrates HS-CoA and acetyl-CoA. Garvie (1967) showed that other vitamins, such as riboflavin, folate, nicotinic acid, thiamine and biotin, were also essential for *O. oeni* growth, while a more recent study showed that, among these vitamins, only nicotinic acid was essential, and that riboflavin, folate, thiamine and biotin only decreased growth (Terrade & Mira de Orduña, 2009). Our *in silico* analysis of minimal requirements showed that nicotinic acid, riboflavin, folate, thiamine and biotin were not required for growth. Moreover, neither cobalamin, p-amino-benzoic acid nor pyridoxal were neither essential for *O. oeni* growth (Garvie, 1967; Terrade & Mira de Orduña, 2009), which is coherent with the analysis of minimal requirements.

Surprisingly, magnesium sulfate, copper sulfate, ferrous sulfate, zinc sulfate, calcium chloride and choline chloride are not required for growth *in vivo* (Terrade & Mira de Orduña, 2009). Because of the small quantities of metals generally required for growth of microorganisms, these results may be due to possible small contaminations from other medium compounds, as well as glassware, which would be enough to sustain growth even if individual metals were omitted. Nevertheless, the results of Terrade & Mira de Orduña coincide with model predictions. On the other hand, although manganese sulfate is required (Terrade & Mira de Orduña, 2009), it was not detected by the *in silico* analysis. Manganese dependence has been studied before, showing that this metal has

great impact on growth (Theobald, Pfeiffer, & König, 2005). Probably, the most relevant role of this metal in *O. oeni* metabolism is the activation of the malolactic enzyme (Saadt, 1982; Spettoli, Nuti, & Zamorani, 1984), allowing the generation of a proton gradient which produces ATP and leads to growth stimulation (Cox & Henick-kling, 1989).

Garvie (1967) and Terrade & Mira de Orduña (2009) showed by single omission that none of the nucleotides were essential for *O. oeni*. Nevertheless, differences in the results between both studies were found when nucleotides were omitted simultaneously. Garvie (1967) detected no growth when nucleotides were omitted simultaneously, while Terrade & Mira de Orduña (2009) showed the opposite. The latter makes sense since LAB could use intracellular precursors for nucleotide biosynthesis such as demonstrated for arginine in diary LAB (Bringel & Hubert, 2003). The model predicts that *O. oeni* is able to grow even when nucleotides are omitted by single omission. However, when adenine and guanine are both omitted no growth could be achieved. Further investigation is needed in order to understand the nucleotide biosynthesis and its integration with the whole metabolic model.

The *in silico* analysis of nutritional requirements suggested that at least one carbon source is required for growth (Table 4-7). Terrade & Mira de Orduña (2009) showed that *O. oeni* was unable to grow when the only carbon source present in the medium (D-ribose) was removed. The model also predicted that phosphate is required for growth (Table 4-8). This is also consistent with the nutritional requirements determined by Terrade & Mira de Orduña (2009), who showed that potassium dihydrogenophosphate is essential (Table 4-5).

Table 4-9. Summary of classification results. True positives are essential metabolites which were classified as essential metabolites. True negatives are non essential metabolites which were classified as non essential metabolites. False positives are non-essential metabolites which were classified as essential metabolites. False negatives are essential metabolites which were classified as non essential metabolites. Most metabolites were classified correctly with the exception of alanine, lysine, leucine and threonine which were classified as essential metabolites but were not essential *in vivo,* and nicotinate and manganse sulfate, that were classified as not essential but were essential *in vivo*

True positives	True negatives	False positives	False negatives
Arginine	Proline	Alanine	Nicotinate
Cysteine	Aspartic acid	Lysine	Manganese sulfate
Glutamic acid	Glutamine	Leucine	
Isoleucine	Serine	Threonine	
Tyrosine	Glycine		
Valine	Guanine		
Histidine	Adenine		
Phenylalanine	Uracil		
Tryptophan	Xanthine		
Methionine	Cytosine		
Asparagine	Thymine		
Phosphate	Cobalamin		
D-ribose	p-aminobenzoic acid		
Pantothenate	Pyridoxal		
	Riboflavin		
	Folate		
	Thiamine		
	biotin		

	Choline chloride
	Magnesium sulfate
	Copper sulfate
	Ferrous sulfate
	Zinc sulfate
	Calcium chloride

Table 4-10. Confusion matrix of minimal medium requirement predictions. 14 essential metabolites were predicted as such (true positives) while 24 non-essential metabolites were predicted as such (true negatives). On the other hand, four non-essential metabolites were predicted as essential (false positives) and two essential metabolites were predicted as non-essential (false negatives)

	Essential	**Non-essential**
Predicted as essential	14	4
Predicted as non-essential	2	24

Based on the confusion matrix (Table 4-10), we calculated statistical measurements of performance (Table 4-11). The model identifies 88% of the essential metabolites, as such (sensitivity); meanwhile, it identifies 86% of the non-essential metabolites as such (specificity). Moreover, 78% of the metabolites predicted as essential were actually essential (precision), while the accuracy of the model, i.e. the proportion of correct results to the entire of predictions, was 86%. F1 score, a measurement of the accuracy that can be interpreted as a weighted average of the sensitivity and the precision, reaches 0.82, indicating that the model has a very good performance overall.

Table 4-11. Statistical measures of performance. Sensitivity (proportion of essential metabolites predicted as such), specificity (proportion of non-essential metabolites predicted as such), precision (proportion of metabolites predicted as essential being essential *in vivo*) and accuracy (proportion of correct results) are indicated. The model F-1 score indicates a high overall performance

Sensitivity	**Specificity**	**Precision**	**Accuracy**	**F1 score**
0,88	0,86	0,78	0,86	0,82

4.4.5 Analysis of specific consumption/production rates

4.4.5.1 Calculation of experimental specific rates

The validation of *in silico* predictions required a comparison with experimental data. For this purpose, all papers performing batch and continuous cultures, which registered substrate uptake and product production as well as growth rates for *O. oeni*, were explored. Below, we give a brief review of the literature found.

Only one continuous culture could be found for *O. oeni* (Zhang & Lovitt, 2006). In this study, the growth performance of the strain NCIMB 11648 was assessed at dilution rates between 0.007 and 0.052 $\frac{1}{h}$. The authors determined dry cell weight, substrate uptake (fructose and glucose) and product formation (lactate, acetate, mannitol and ethanol) rates, as well as yields (maximum growth rate, maximun cell productivity, biomass yield and maintenance coefficient). Nevertheless, this study was carried out in the absence of ethanol, malic acid and citric acid, which could have significantly affected the physiology of the microorganism. Additionally, considering the high genotypic and phenotypic variation between strains (Borneman et al., 2012), we preferred to employ experimental data from the strain PSU-1 to refine our model and not use this data.

Several batch cultures have been performed. Campos et al (2009) performed a batch culture to study the influence of phenolic acids on glucose and organic acid metabolism of *O. oeni* VF, a commercial starter culture strain. Every 25 hours the authors monitored the concentrations of citric acid, glucose, malic acid, lactic acid and acetic acid of *O. oeni* batch cultures growing in culture media supplemented with different phenolic acids (caffeic acid, ferulic acid, p-coumaric acid, gallic acid and protocatechuic acid).

We employed the data from Olguín et al., (2009) to calculate the specific consumption/production rates of L-malic acid, fructose, glucose, citric acid, L-lactic acid, D-lactic acid and acetic acid in the exponential phase of a batch culture of *O. oeni* PSU-1 (Olguín et al., 2009). Since no continuous culture has been carried out with the *O. oeni* PSU-1 strain yet, we used the data from the batch culture, assuming that mass balances were conserved during the exponential phase.

Specific consumption and production rates were obtained by calculating the difference between final and initial concentrations among two experimental points of exponential phase, then dividing this difference by the corresponding fermentation time and the mean biomass concentration. Final concentrations correspond to one day after inoculation, while initial concentrations correspond to those found at the time of inoculation. We decided to employ just two experimental points because citric and malic acids were completely consumed after one day of cultivation.

For example, we calculated the consumption rate of L-malic by *O. oeni* growing in a medium of pH 4.0, without ethanol. The final concentration of L-malic acid under these conditions was $0\ \frac{mmol}{L}$ and the initial concentration, $37.7\ \frac{mmol}{L}$. The difference between these values is $37.7\ \frac{mmol}{L}$. Dividing by 24 hours, the fermentation time, and by the biomass generated during this period, $X = 0.18\ \frac{g}{L}$, we obtain $r_s = 8.57\ \frac{mmol}{g\ hr}$, which represents the specific rate of consumption of L-malic acid (Table 4-12).

The specific growth rate μ was calculated using the slope of the graph of logarithm of biomass concentration (Table 4-13) - in grams per liter - versus time, in hours (Figure 4-27).

Table 4-12. Specific consumption rates of L-malic acid, fructose, glucose, citric acid and specific production rates of L-lactate, D-lactate and acetate determined from a batch culture of *O. oeni* growing at different pH levels and ethanol concentrations. Initial and final concentrations of each compound were extracted from four batch cultures using different pH levels (3.5 and 4.0) and ethanol concentrations (0% and 10% v/v). The specific rate was determined dividing the difference between final and initial concentrations by the corresponding fermentation time (final concentration was measured 24 hrs after measurement of initial concentration) and the mean biomass concentration.

pH	Ethanol	Initial Concentration [mmol/L]	Final Concentration [mmol/L]	Difference [mmol/L]	Specific Rate [mmol/g hr]
L-malic acid					
4	Absent	37.7	0	37.7	8.6
3.5	Absent	40.9	0	40.9	13.3
4	Present	44.7	6.4	38.3	63.8
3.5	Present	33.9	7.4	26.5	50.1
Fructose					
4	Absent	18.3	13.9	4.4	1,0
3.5	Absent	22.4	21.2	1.2	0,4
4	Present	20.8	22.3	1.6	2.6
3.5	Present	17.4	18.9	1.6	3.0
Glucose					
4	Absent	26.4	22.0	4.3	1.0
3.5	Absent	32.3	32.1	0.2	0.6
4	Present	29.9	29.5	0.4	0.6
3.5	Present	27.2	27.7	0	0
Citric acid					
4	Absent	0.8	0	0.8	0.2
3.5	Absent	1.0	0	1.0	0.3

4	Present	1.4	0.8	0.6	1.0
3.5	Present	1.4	0.8	0.6	1.1
L-lactic acid					
4	Absent	15.5	38.1	6.0	5.1
3.5	Absent	6.9	40.9	34.0	11.1
4	Present	8.1	30.6	22.5	37.6
3.5	Present	9.8	29.5	19.8	37.3
D-lactic acid					
4	Absent	6.991	13.022	6.031	1.370
3.5	Absent	5.139	10.167	5.028	1.640
4	Present	6.706	7.412	0.706	1.176
3.5	Present	6.471	6.706	0.235	0.444
acetic acid					
4	Absent	5.3	12.6	7.3	1.7
3.5	Absent	6.1	10.3	4.2	1.4
4	Present	6.2	7.4	1.1	1.9
3.5	Present	5.6	10.0	4.4	8.3

Table 4-13. Specific growth rates determined from a batch culture of *O. oeni* growing at different pH levels and ethanol concentrations. Initial and final biomass concentration was extracted from 4 batch cultures using different pH levels (3.5 and 4.0) and ethanol concentrations (0% and 10% v/v). Final concentration was measured 24 hrs after measurement of initial concentration. Growth rates were determined calculating the slope of the logarithm of biomass concentration versus culture time.

pH	Ethanol	Initial Concentration [g/L]	Final Concentration [g/L]	Difference [g/L]	Specific growth rate [1/hr]
biomass					
4	Absent	0.04	0.23	0.18	0.03
3.5	Absent	0.08	0.21	0.13	0.02
4	Present	0.05	0.08	0.03	0.007
3.5	Present	0.07	0.09	0.02	0.005

4.4.5.2 Determination of *in silico* specific rates

Cross validation consisted in, having N+1 experimental rates, fix N rates in the GEM and predict the missing one. This experiment is useful in the refinement process, because if there is a significant error in the prediction of one rate, the user has a guide to identify where the error might be found in the metabolic model. Thus, once the probable error in the metabolic model is identified, the user can fix it and run the analysis again. Typical errors in the metabolic model are related to additional or missing reactions, directionality reaction mistakes, poor connectivity, etc. On the other hand, predictions consistentwith experimental data validate the model.

We developed a MATLAB script that carries out this experiment. Additionally, the script calculates the percentage differences between the experimental rates and the predicted ones and it writes these differences into an excel spreadsheet (Figure 4-26) It took 122 seconds to carry out these *in silico* predictions for the *O. oeni* model.

	A	B	C	D
1	Reaction ID	Experimental value	Predicted value	Difference (%)
2	Demand_Reaction_Biomass	0,0303	0,0408	35,0 %
3	Exchange_citrate	-0,23	-0,24	4,6 %
4	Exchange_acetate	1,65	1,63	1,5 %
5	Exchange_(R)-lactate	1,37	1,36	0,7 %
6	Exchange_beta-D-fructofuranose	-1,00	-1,01	0,1 %
7	Exchange_alpha-D-glucose	-0,99	-1,00	0,1 %
8	Exchange_(S)-lactate	5,14	5,13	0,3 %
9	Exchange_(S)-malate	-8,57	-8,59	0,2 %

Figure 4-26. Excel spreadsheet showing real and predicted consumption/production rates of different metabolites and real and predicted specific growth rates. Percentage differences are also presented in order to visualize errors.

Fructose and glucose consumption rates were dependent on CO_2 production under all conditions, i.e. while more CO_2 was produced, more fructose and glucose were consumed. To set a value of CO_2 production that accurately simulates sugar consumption, we constrained all the rates and determined the CO_2 production rate. Then, we used this value as given information for each condition (Table 4-14).

Table 4-14. I*n silico* CO_2 specific production rates used to adjust fructose and glucose consumption. These rates were determined by optimizing the genome-scale metabolic model of *O. oeni* constrained with four sets of restrictions corresponding to the four batch culture conditions (two different pH leves: 3.5 and 4.0 and two different ethanol concentration: 0% and 10% v/v).

Specific Rate at pH 4 without ethanol [mmol/g hr]	Specific Rate at pH 3.5 without ethanol [mmol/g hr]	Specific Rate at pH 4 with ethanol [mmol/g hr]	Specific Rate at pH 3.5 with ethanol [mmol/g hr]
18,69	18,40	103,91	79,86

Most of the CO_2 produced in the model arises from the transformations of 6-phospho D-gluconate to D-ribulose 5 phosphate (pentose phosphate pathway), (S)-malate to pyruvate (malic enzyme), (S)-malate to (S)-lactate (malolactic enzyme), coenzyme A to acetyl-CoA and several reactions associated to fatty acid biosynthesis.

It is worth noting that the increase of CO_2 production under ethanol growing conditions is due to the increase in the flux value of some specific reactions. To exemplify this, we analyzed the malolactic reaction, the main source of CO_2 in the simulations (28% and 36% at pH 4.0 in the absence and presence of ethanol, respectively), to predict the biomass function. The flux through the malolactic reaction increases from 5,1 $\frac{mmol}{g\ hr}$ to 37,6 $\frac{mmol}{g\ hr}$ at pH 4.0 in the absence and presence of ethanol, respectively (Table 4-15) This could be explained by the fact that, under stress conditions, *O. oeni* triggers several anti-stress mechanisms (Bourdineaud, Nehmé, Tesse, & Lonvaud-Funel, 2004; Maitre et al., 2014; Mateo, Medina, Mateo, & Jiménez, 2010) that require additional ATP. Carrying out the malolactic reaction allows the bacterial cell to export lactate and protons, ensuring that the proton motive force provides the extra ATP required. A similar situation occurs at pH 3.5 (data not shown).

Table 4-15. Main *in silico* CO_2 sources at pH 4.0, in absence (0%) and presence (10% v/v) of ethanol. In parenthesis the percentages of these contributions to total CO_2 are shown. At pH 4.0, the model predicts that the main CO_2 source is malolactic reaction (28% and 36% of the total CO_2, without and with ethanol respectively). Interestingly the reaction producing diacetyl is activated only when ethanol is present in the medium.

Reaction	Flux through reaction [mmol/g hr] at pH 4 without ethanol	Flux through reaction [mmol/g hr] at pH 4 with ethanol
6-phosphogluconate dehydrogenase	3,1 (16,6%)	9,0 (8,6%)
Malic enzyme	3,7 (19,5%)	27,2 (26,2%)
Pyruvate dehydrogenase	4,2 (22,5%)	10,9 (10,5%)
Malolactic enzyme	5,1 (27,5%)	37,6 (36,2%)
Phosphopantothenoylcysteine decarboxlylase	0,5 (2,6%)	1,2 (1,1%)
Diacetyl producing reaction	0 (0,0%)	9,0 (8,7%)
Total	**13,5 (72,2%)**	**94,9 (91,3%)**

Values in parenthesis are the percentage of the contribution corresponding to the CO_2 produced by the specific reaction over the total amount of CO_2 produced.

4.4.5.3 Comparison of *in vivo* and *in silico* specific rates

Accordance between experimental and predicted data was found when *O. oeni* was grown in the absence of ethanol at both pH 4.0 and 3.5. A percentage difference of 35% (Table 4-16) and 40% (Table 4-17) was obtained between the real biomass produced and the *in silico* produced biomass, respectively. This difference most likely results from the utilization of the biomass equation of *Lactococcus lactis*. Thus, determining the macromolecular content of *O. oeni* PSU-1 cells should improve this prediction.

Additionally, a remarkable adjustment between the real specific consumption/production rates and the predicted ones was found when *O. oeni* is grown in the absence of ethanol in both pH 4.0 and 3.5. At pH 4.0, the highest difference was 4.5% for the consumption of citrate, while the mean of the percentage difference was 1.3% for all metabolites. Even better is the case at pH 3.5, where the maximum difference was 1.45% for fructose consumption, while the mean was 0.59% difference for all metabolites.

Table 4-16. Specific rates predicted by the *Oenococcus oeni* genome-scale metabolic model using constraints corresponding to a medium at pH 4.0 without ethanol.

Metabolite	Experimental rate	Predicted	Difference (%)
Biomass	0,0303	0,0409	35,0%
Citrate	-0,2285	-0,2388	4,5%
Acetate	1,6586	1,6330	1,5%
(R)-lactate	1,3700	1,3609	0,7%
Beta-D-fructofuranose	-1,0003	-1,0110	1,1%
Alpha-D-glucose	-0,9856	-0,9962	1,1%
(S)-lactate	5,1454	5,1283	0,3%
(S)-malate	-8,5714	-8,5873	0,2%

Table 4-17. Specific rates predicted by *Oenococcus oeni* genome-scale metabolic model using constraints corresponding to a medium at pH 3.5 without ethanol.

Metabolite	Experimental rate	Predicted	Difference (%)
Biomass	0,017	0,010	39,8%
Citrate	-0,3282	-0,3284	0,1%
Acetate	1,3697	1,3559	1,0%
(R)-lactate	1,6404	1,6307	0,6%
Beta-D-fructofuranose	-0,3863	-0,3919	1,5%
Alpha-D-glucose	-0,6263	-0,6319	1,9%
(S)-lactate	11,0974	11,0876	0,1%
(S)-malate	-13,3357	-13,3360	0,003%

Unfortunately, for *O. oeni* grown in a medium containing ethanol, biomass predictions were not accurate. The percentage difference between the real biomass production value and the predicted one is 1370 % and 1680% at pH 4.0 and 3.5, respectively. This poorly accurate prediction was probably due to errors in the model. The most likely cause is that model cannot correctly simulate the changes that occur in the metabolism of *Oenoccocus* oeni under stress conditions caused by ethanol. Ethanol produces significant changes in the lipid composition of the cell membrane (Grandvalet et al., 2008; Silveira et al., 2003), as well as other macromolecular assembly proportions. Therefore, it is expected that, while stoichiometric coefficients related to fatty acids synthesis and other macromolecules remained constant under *in silico* growth in the ethanol-containing medium, good biomass predictions would not be achieved.

Additionally, the specific consumption/production rates under ethanol growing conditions are less accurate than in the medium without ethanol. At pH 4.0, the mean difference is 11%, with the maximum difference found for the acetate production rate with a value of 46% (Table 4-18). At pH 3.5, the percentage difference is 72%, with the maximum difference for citrate uptake, with a value of 155% (Table 4-19).

To solve the inaccurate biomass prediction and the imprecise prediction of consumption/production rates of *O. oeni* growing under stress conditions, the determination of the biomass composition and the incorporation of high throughput data to the model, both obtained under stress conditions caused by ethanol, are proposed. Biomass composition determined under ethanol-growing conditions ensures that correct stoichiometric coefficients are used for biomass production, while integrating transcriptomic data will lead to better predictions of the specific consumption/production rates, because the genes that are expressed under these harsh conditions will be quantified and, consequently so will the reactions that are carried out under stress conditions.

Table 4-18. Specific rates predicted by the *Oenococcus oeni* genome-scale metabolic model using constraints corresponding to a medium at pH 4.0 with ethanol.

Metabolite	Experimental rate	Predicted	Difference (%)
Biomass	0,0068	0,1	1370%
Citrate	-1,0105	-0,7081	30%
Acetate	1,8667	1,0173	45 %
(R)-lactate	1,1764	1,1798	0,3%
Beta-D-fructofuranose	-2,6113	-2,6114	4,6 e-11%
Alpha-D-glucose	-0,6391	-0,6391	1,8 e-10%
(S)-lactate	37,5610	37,5610	9,3 e-12%

(S)-malate	-63,7981	-63,7979	0,0002%

Table 4-19. Specific rates predicted by *Oenococcus oeni* genome-scale metabolic model using constraints corresponding to a medium at pH 4.0 with ethanol.

Metabolite	**Experimental rate**	**Predicted**	**Difference (%)**
Biomass	0,005	0,0890	1680,0%
Citrate	-1,0966	-2,7918	154,6%
Acetate	8,3111	16,9607	104,1%
(R)-lactate	0,4444	0	100%
Beta-D-fructofuranose	-2,9595	3,8074	28,7%
Alpha-D-glucose	0	-0,8484	-
(S)-lactate	37,3074	22,1141	40,7%
(S)-malate	-50,0737	-51,7696	3,4%

4.4.6 Sensitivity analysis

During model refinement, it is often necessary to incorporate reactions from other completed GEMs to the developing GEM. However, the stoichiometric coefficients of some reactions are organism-specific, which is the case of all of the reactions related to biomass composition. If these kinds of reactions are incorporated to the developing GEM from other completed GEMs, care must be taken to analyze the sensitivity of the objective function to changes in the values of the stoichiometric coefficients.

Following the methodology 1 explained in materials and methods, the model was optimized to yield surprisingly specific growth rates equal to zero in each of the 40 cases analyzed. This is probably because the stoichiometric coefficients resulted in ratios of biomass components that the metabolic network could not accomplish. Given this issue, methodology 2 was followed. Thus, it was determined that the specific growth rate is more sensitive to changes in coefficients related to lipids and maintenance associated to the assembly of proteins (Table 4-20) In particular, we observed that the specific growth rate was strongly decreased when the stoichiometric coefficient of cardiolipin increased (decreasing 0.4% in specific growth rate when increasing the stoichiometric coefficient by 1%).

On the other hand, the specific growth rate was also reduced mostly by increasing the ATP needed to assemble proteins (0.3%). This could partly explain the over-estimation of the specific growth rate in ethanol growing conditions since the ATP needed to generate biomass was maintained constant in all *in silico* experimental conditions. When the culture medium contains ethanol, *O. oeni* is growing under critical stress conditions, therefore it is reasonable to think that the ATP required to generate biomass is increased in this situation. Even so, the over-estimation of the growth rate is much greater than the compensation by an increase in the growth associated maintenance. It is possible that the

model is capable of generating ATP in ways that do not occur naturally in the metabolism of *O. oeni*.

Table 4-20. Sensitivity analysis of the biomass formation equation imported from the *Lactococcus lactis* genome-scale metabolic model.

Metabolite	**μ Variation (%) (1%)**	**μ Variation (%) (10%)**	**μ Variation (%) (50%)**
Protein			
L-histidine	-0,0003	-0,04	-1,2
L-lysine	-0,0003	-0,04	-1,2
L-phenylalanine	-0,0003	-0,04	-1,2
L-arginine	-0,0003	-0,04	-1,2
L-leucine	-0,0003	-0,04	-1,2
L-valine	-0,0003	-0,04	-1,2
L-threonine	-0,0003	-0,04	-1,2
Glycine	0,0005	-0,04	-1,1
L-alanine	-0,0004	-0,03	-1,2
L-serine	-0,0004	-0,04	-1,2
L-cysteine	-0,0004	-0,04	-1,2
L-glutamine	0,0007	-0,03	-1,4
L-tryptophan	-0,0004	-0,04	-1,3
L-asparagine	-0,0004	-0,04	-1,3
L-aspartate	-0,0004	-0,04	-1,3
L-tyrosine	-0,0004	-0,04	-1,3
L-isoleucine	-0,0004	-0,04	-1,3
L-glutamate	-0,0004	-0,04	-1,3
L-proline	-0,0004	-0,04	-1,3
L-methionine	-0,0004	-0,04	-1,3

ATP	-0,3	-3,3	-19,3
DNA			
dATP	-0,0004	-0,06	-0,07
dCTP	-0,0002	-0,003	-0,04
dGTP	-0,0003	-0,03	-0,04
dTTP	-0,0006	-0,006	-0,05
ATP	-0,001	-0,02	-0,1
H_2O	0,0002	0,002	0,02
RNA			
CMP	-0,0006	-0,01	-0,2
GMP	-0,001	-0,02	-0,3
Uridine-5'-phosphate	-0,0009	-0,01	-0,2
AMP	-0,001	-0,01	-0,2
ATP	-0,01	-0,1	-0,9
Lipids			
Lysophosphatidyl-glycerol	-0,01	-0,1	-0,5
Cardiolipin	-0,4	-4,6	-33,1
L-1-phosphatidyl-glycerol	-0,09	-1,2	-12,5
3-D-glucosyl-1,2-diacylglycerol	-0,006	-0,2	-5,1
Diglucosyl diacylglycerol	-0,14	-1,6	-12,8
Peptidoglycan			
D-aspartate	-0,00005	-0,007	-0,2
Lipid II[a]	-0,03	-0,35	-2,5

Lipoteichoic acid			
Diglucosyl diacylglycerol	-0,23	-2,5	-18,6
D-alanyl-D-galactosyl-poly(glycerol phosphate)	-0,004	-0,05	-0,4
Polysaccharides			
dTDP-6-deoxy-beta-L-mannose	-0,01	-0,2	-1,2
UDP-D-glucose	-0,01	-0,1	-0,9
UDP-alpha-D-galactose	-0,002	-0,01	0,02

4.5 Data assembly and dissemination

GENREs builders usually require printing the model information in an Excel sheet, so it can be easily inspected by any person, including those who are not experts in the field. In fact, to manually order the data in an Excel sheet is a significantly time-consuming task, so automatizing this step is essential. We developed a set of scripts in MATLAB that reads the information contained in a GENRE developed in Pathway Tools, pre-process and print it automatically into an Excel file.

The scripts *extractReactionsInfo.m, extractMetabolitesInfo.m, extractGenesInfo.m, extractProteinsInfo.m, extractRNAsInfo.m, extractPathwaysInfo.m* extract information of different elements of the GENRE from Pathway Tools. This information can be used by other scripts, such as *printMetabolites.m* and *printReactions.m*, to print the information relative to metabolites or reactions, respectively, in an Excel sheet.

Printing the GENRE reactions and sorting them by metabolic pathways in an Excel sheet is of particular interest. The script *printReactions.m*, allow the user to inspect in a simple manner the features associated with each one of the reactions, such as the stoichiometric equation, EC number, or associated enzyme names. Furthermore, the script *findGeneReactionsAssociations.m* allows the user to visualize the association gene-protein-reaction in the form of gene-rules and print them.

GENRE information contained in Pathway Tools must be first exported to a plaintext, then it must be opened in an Excel sheet and, finally, it must be saved having removed all of the explanation text in order to allow MATLAB to read it. The algorithm runtime is 0.3 seconds

The script *printReactions.m* was used to export the model reactions into a spreadsheet (Suppementary Spreadsheet 2).

5 GENERAL CONCLUSION AND PERPECTIVES

In this thesis, we constructed a GEM of the wine malolactic bacterium *O. oeni* containing 914 reactions, 792 metabolites and 512 genes. This model was used to predict a minimal growth medium, which was compared with literature. With a total of 44 growth/non growth experiments, the model showed an accuracy of 86% in its predictions and a F-score value of 0.82, suggesting a high overall performance.

Additionally, the model was employed to predict specific growth rates and consumption/production rates of substrates and products under different medium conditions of pH and ethanol toxicity. In absence of ethanol, a mean percentage difference of 38% was observed between the experimental and predicted specific growth rates, while a mean percentage difference of 0.1% was found for the consumption/production rates. Less accurate predictions were achieved under ethanol growing conditions, probably due to the limited information concerning activation/deactivation of genes/reactions of *O. oeni* growing in a medium containing ethanol. Further model refinement will require determining biomass composition and whole gene expression of *O. oeni* under ethanol growing conditions.

In addition, reaction and gene deletion analyses showed that most fragile metabolic processes are those related to fatty acid, peptidoglycan, purine and pyrimidine biosyntheses and heterolactic fermentation. Removal of some of the reactions related to CO_2 production caused the most significant reduction of growth, probably by their implication in the heterolactic fermentation. Compound deletion analysis showed that removal of most compounds do not affect growth, with only 26% of the metabolites being essential for growth.

We developed a set of tools implemented in MATLAB that facilitate refinement and analysis of GEMs. Some of these tools are associated to Pathways tools, therefore it requires the development of the GEM through this platform. These tools were used to support the construction and posterior analysis of the genome-scale metabolic model of *O. oeni*.

The scripts developed to support refinement are useful for facilitating metabolite reading in the MATLAB GEM, printing and reading information contained in the GEM, creation of fields necessary to run gene deletion analysis and to make the model functional. All these scripts represent support for the protocol developed by Thiele and Palsson in order to generate high quality GEMs. Additionally, these scripts were very efficient, showing runtimes of 0.3, 0.3, 13 and 143 seconds respectively, when they were used in the *O. oeni* model.

On the other hand, the scripts developed to support analysis were used for determining minimum medium requirements; performing the cross validation experiment; carrying out reaction, compound and gene deletion analysis; and performing sensitivity analysis. These scripts represent new tools that were previously unavailable. Reaction, compound and gene deletion analysis are unique because they incorporate information about pathways in which reactions are present having the user being able to analyze which metabolic processes tend to be affected by the removal of a particular reaction/gene. The scripts were efficient, showing runtimes of 69, 86, 74,110 seconds, respectively, when used for analysis of the *O. oeni* GEM.

Finally, this model is the first step towards the generation of a comprehensive understanding of *O. oeni* metabolism and the consequent control of malolactic fermentation during winemaking. This model could be useful to predict the rate at which malolactic fermentation occurs in wineries, helping winemakers to predict the putative

duration of this process. Challenges remain in improving consumption/production predictions of *O. oeni* under ethanol growing conditions and in incorporating high throughput data and regulatory mechanisms.

ABBREVIATIONS

ATP	Adenosine triphosphate
COBRA	Constraint-based reconstruction and analysis
EC	Enzyme Commision
FBA	Flux Balance Analysis
GENRE	Genome-Scale Network Reconstruction
GEM	Genome-Scale Model
NGAM	Non Growth Associated Maintenance
ORF	Open Reading Frame
KEGG	Encyclopedia of Genes and Genomes
LAB	Lactic Acid Bacteria
SBML	Systems Biology Markup Language

REFERENCES

Agren, R., Liu, L., Shoaie, S., Vongsangnak, W., Nookaew, I., & Nielsen, J. (2013). The RAVEN toolbox and its use for generating a genome-scale metabolic model for Penicillium chrysogenum. *PLoS Computational Biology*, *9*(3), e1002980. doi:10.1371/journal.pcbi.1002980

Bartowsky, E. J. (2005). Oenococcus oeni and malolactic fermentation – moving into the molecular arena. *Australian Journal of Grape and Wine Research*, *11*(2), 174–187. doi:10.1111/j.1755-0238.2005.tb00286.x

Bartowsky, E. J., & Henschke, P. a. (2004). The "buttery" attribute of wine--diacetyl--desirability, spoilage and beyond. *International Journal of Food Microbiology*, *96*(3), 235–52. doi:10.1016/j.ijfoodmicro.2004.05.013

Bauer, R., & Dicks, L. M. T. (2004). Control of Malolactic Fermentation in Wine . A Review. *S. Afr. J. Enol. Vitic.*, *25*(2), 74–88.

Bautista, E. J., Zinski, J., Szczepanek, S. M., Johnson, E. L., Tulman, E. R., Ching, W.-M., … Srivastava, R. (2013). Semi-automated curation of metabolic models via flux balance analysis: a case study with Mycoplasma gallisepticum. *PLoS Computational Biology*, *9*(9), e1003208. doi:10.1371/journal.pcbi.1003208

Becker, S. a, Feist, A. M., Mo, M. L., Hannum, G., Palsson, B. Ø., & Herrgard, M. J. (2007). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nature Protocols*, *2*(3), 727–38. doi:10.1038/nprot.2007.99

Becker, S. a, & Palsson, B. Ø. (2005). Genome-scale reconstruction of the metabolic network in Staphylococcus aureus N315: an initial draft to the two-dimensional annotation. *BMC Microbiology*, *5*, 8. doi:10.1186/1471-2180-5-8

Beelman, R. B., III Gavin, A., & Keen, R. M. (1977). A new strain of Leuconostoc Oenos for induced malo-lactic fermentation in eastern wines. *Am. J. Enol. Vitic*, *28*(3), 159–165.

Borneman, A. R., McCarthy, J. M., Chambers, P. J., & Bartowsky, E. J. (2012). Comparative analysis of the Oenococcus oeni pan genome reveals genetic diversity in industrially-relevant pathways. *BMC Genomics*, *13*(1), 373. doi:10.1186/1471-2164-13-373

Bourdineaud, J.-P., Nehmé, B., Tesse, S., & Lonvaud-Funel, a. (2004). A bacterial gene homologous to ABC transporters protect Oenococcus oeni from ethanol and other stress

factors in wine. *International Journal of Food Microbiology*, *92*(1), 1–14. doi:10.1016/S0168-1605(03)00162-4

Bringel, F., & Hubert, J. (2003). Extent of Genetic Lesions of the Arginine and Pyrimidine Biosynthetic Pathways in Lactobacillus plantarum , L . paraplantarum , L . pentosus , and L . casei : Prevalence of CO 2 -Dependent Auxotrophs and Characterization of Deficient arg Genes in Extent o. doi:10.1128/AEM.69.5.2674

Brooks, J. P., Burns, W. P., Fong, S. S., Gowen, C. M., & Roberts, S. B. (2012). Gap detection for genome-scale constraint-based models. *Advances in Bioinformatics*, *2012*, 323472. doi:10.1155/2012/323472

Campos, F. M., Figueiredo, A. R., Hogg, T. a, & Couto, J. a. (2009). Effect of phenolic acids on glucose and organic acid metabolism by lactic acid bacteria from wine. *Food Microbiology*, *26*(4), 409–14. doi:10.1016/j.fm.2009.01.006

Caspeta, L., & Nielsen, J. (2013). Toward systems metabolic engineering of Aspergillus and Pichia species for the production of chemicals and biofuels. *Biotechnology Journal*, *8*(5), 534–44. doi:10.1002/biot.201200345

Caspeta, L., Shoaie, S., Agren, R., Nookaew, I., & Nielsen, J. (2012). Genome-scale metabolic reconstructions of Pichia stipitis and Pichia pastoris and in silico evaluation of their potentials. *BMC Systems Biology*, *6*(1), 24. doi:10.1186/1752-0509-6-24

Chavali, A. K., Whittemore, J. D., Eddy, J. a, Williams, K. T., & Papin, J. a. (2008). Systems analysis of metabolism in the pathogenic trypanosomatid Leishmania major. *Molecular Systems Biology*, *4*(177), 177. doi:10.1038/msb.2008.15

Cox, D. J., & Henick-kling, T. (1989). Chemiosmotic Energy from Malolactic Fermentation I-, *171*(10), 5750–5752.

Droste, P., Nöh, K., & Wiechert, W. (2013). Omix - A Visualization Tool for Metabolic Networks with Highest Usability and Customizability in Focus. *Chemie Ingenieur Technik*, *85*(6), 849–862. doi:10.1002/cite.201200234

Edwards, J. S., & Palsson, B. O. (1999). Systems Properties of the Haemophilus influenzae Rd Metabolic Genotype. *J. Biol Chem.*, *274*(25), 17410–17416.

Eker, S., Krummenacker, M., Shearer, A. G., Tiwari, A., Keseler, I. M., Talcott, C., & Karp, P. D. (2013). Computing minimal nutrient sets from metabolic networks via linear constraint solving. *BMC Bioinformatics*, *14*, 114. doi:10.1186/1471-2105-14-114

Fan, S., Zhang, Z., Zou, W., Huang, Z., Liu, J., & Liu, L. (2014). Development of a minimal chemically defined medium for Ketogulonicigenium vulgare WSH001 based on its genome-scale metabolic model. *Journal of Biotechnology*, *169*, 15–22. doi:10.1016/j.jbiotec.2013.10.027

Feist, A. M., Herrgård, M. J., Thiele, I., Reed, J. L., & Palsson, B. Ø. (2009). Reconstruction of biochemical networks in microorganisms. *Nature Reviews. Microbiology*, *7*(2), 129–43. doi:10.1038/nrmicro1949

Feist, A. M., & Palsson, B. O. (2010). The biomass objective function. *Current Opinion in Microbiology*, *13*(3), 344–9. doi:10.1016/j.mib.2010.03.003

Fourcassie, P., Belarbi, A., & Maujean, A. (1992). Growth , D-glucose utilization and malolactic fermentation by Leuconostoc aenos strains in 18 media deficient in one amino acid.

Garcia, N., Blancato, V., Repizo, G., Magni, C., & López, P. (2008). Citrate metabolism and aroma compound production in lactic acid bacteria. *Molecular Aspects of Lactic Acid Bacteria for Traditional and New Applications, Ed. by Mayo B, López P and Pérez-Martínez G. Research Signpost, Kerala*, *661*(2), 65–88.

Garvie, E. (1967). The Growth Factor and Amino Acid Requirements of Species of the Genus Leuconostoc ,. *J. Gen. Microbiol*, *48*, 439–447.

Grandvalet, C., Assad-García, J. S., Chu-Ky, S., Tollot, M., Guzzo, J., Gresti, J., & Tourdot-Maréchal, R. (2008). Changes in membrane lipid composition in ethanol- and acid-adapted Oenococcus oeni cells: characterization of the cfa gene by heterologous complementation. *Microbiology (Reading, England)*, *154*(Pt 9), 2611–9. doi:10.1099/mic.0.2007/016238-0

Heavner, B. D., Smallbone, K., Barker, B., Mendes, P., & Walker, L. P. (2012). Yeast 5 - an expanded reconstruction of the Saccharomyces cerevisiae metabolic network. *BMC Systems Biology*, *6*(1), 55. doi:10.1186/1752-0509-6-55

Henry, C. S., DeJongh, M., Best, A. a, Frybarger, P. M., Linsay, B., & Stevens, R. L. (2010). High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nature Biotechnology*, *28*(9), 977–82. doi:10.1038/nbt.1672

Hyduke, D. R., Lewis, N. E., & Palsson, B. Ø. (2013). Analysis of omics data with genome-scale models of metabolism. *Molecular bioSystems*, *9*(2), 167–74. doi:10.1039/c2mb25453k

Karp, P. D., Paley, S. M., Krummenacker, M., Latendresse, M., Dale, J. M., Lee, T. J., … Caspi, R. (2010). Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Briefings in Bioinformatics*, *11*(1), 40–79. doi:10.1093/bib/bbp043

Karp, P. D., Paley, S., & Romero, P. (2002). The pathway tools software. *BIOINFORMATICS*, *18*, 1–8.

Karp, P. D., Riley, M., Paley, S. M., & Pellegrini-Toole, A. (2002). The MetaCyc Database. *Nucleic Acids Research*, *30*(1), 59–61. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3964957&tool=pmcentrez&rendertype=abstract

Kauffman, K. J., Prakash, P., & Edwards, J. S. (2003). Advances in flux balance analysis. *Current Opinion in Biotechnology*, *14*(5), 491–496. doi:10.1016/j.copbio.2003.08.001

Kim, H. U., Kim, S. Y., Jeong, H., Kim, T. Y., Kim, J. J., Choy, H. E., … Lee, S. Y. (2011). Integrative genome-scale metabolic analysis of Vibrio vulnificus for drug targeting and discovery. *Molecular Systems Biology*, *7*(460), 460. doi:10.1038/msb.2010.115

Kim, T. Y., Sohn, S. B., Kim, Y. Bin, Kim, W. J., & Lee, S. Y. (2012). Recent advances in reconstruction and applications of genome-scale metabolic models. *Current Opinion in Biotechnology*, *23*(4), 617–23. doi:10.1016/j.copbio.2011.10.007

Koffas, M., & Stephanopoulos, G. (2005). Strain improvement by metabolic engineering: lysine production as a case study for systems biology. *Current Opinion in Biotechnology*, *16*(3), 361–6. doi:10.1016/j.copbio.2005.04.010

König, H., Unden, G., & Jürgen, F. (2009). *Biology of Microorganisms on Grapes, in Must and in Wine*.

Kun, A., Papp, B., & Szathmáry, E. (2008). Computational identification of obligatorily autocatalytic replicators embedded in metabolic networks. *Genome Biology*, *9*(3), R51. doi:10.1186/gb-2008-9-3-r51

Latendresse, M., Krummenacker, M., Trupp, M., & Karp, P. D. (2012). Construction and completion of flux balance models from pathway databases. *Bioinformatics (Oxford, England)*, *28*(3), 388–96. doi:10.1093/bioinformatics/btr681

Lee, J. W., Na, D., Park, J. M., Lee, J., Choi, S., & Lee, S. Y. (2012). Systems metabolic engineering of microorganisms for natural and non-natural chemicals. *Nature Chemical Biology*, *8*(6), 536–46. doi:10.1038/nchembio.970

Lerm, E., Engelbrecht, L., & Toit, M. (2010). Malolactic Fermentation : The ABC ' s of MLF, *31*(2), 186–212.

Lewis, N. E., Schramm, G., Bordbar, A., Schellenberger, J., Andersen, M. P., Cheng, J. K., … Palsson, B. Ø. (2010). Large-scale in silico modeling of metabolic interactions between cell types in the human brain. *Nature Biotechnology*, *28*(12), 1279–85. doi:10.1038/nbt.1711

Lonvaud-Funel, a. (1999). Lactic acid bacteria in the quality improvement and depreciation of wine. *Antonie van Leeuwenhoek*, *76*(1-4), 317–31. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10532386

Loubiere, P., Salou, P., Leroy, M. J., Lindley, N. D., & Pareilleux, a. (1992). Electrogenic malate uptake and improved growth energetics of the malolactic bacterium Leuconostoc oenos grown on glucose-malate mixtures. *Journal of Bacteriology*, *174*(16), 5302–8. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=206366&tool=pmcentrez&rendertype=abstract

Maarleveld, T. R., Khandelwal, R. a, Olivier, B. G., Teusink, B., & Bruggeman, F. J. (2013). Basic concepts and principles of stoichiometric modeling of metabolic networks. *Biotechnology Journal*, *8*(9), 997–1008. doi:10.1002/biot.201200291

Maitre, M., Weidmann, S., Dubois-Brissonnet, F., David, V., Covès, J., & Guzzo, J. (2014). Adaptation of the Wine Bacterium Oenococcus oeni to Ethanol Stress: Role of the Small Heat Shock Protein Lo18 in Membrane Integrity. *Applied and Environmental Microbiology*, *80*(10), 2973–80. doi:10.1128/AEM.04178-13

Mateo, E. M., Medina, Á., Mateo, R., & Jiménez, M. (2010). Effect of ethanol on the ability of Oenococcus oeni to remove ochratoxin A in synthetic wine-like media. *Food Control*, *21*(6), 935–941. doi:10.1016/j.foodcont.2009.12.015

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

Mills, D. a, Rawsthorne, H., Parker, C., Tamir, D., & Makarova, K. (2005). Genomic analysis of Oenococcus oeni PSU-1 and its relevance to winemaking. *FEMS Microbiology Reviews*, *29*(3), 465–75. doi:10.1016/j.femsre.2005.04.011

Monk, J., Nogales, J., & Palsson, B. O. (2014). Optimizing genome-scale network reconstructions. *Nature Biotechnology*, *32*(5), 447–452. doi:10.1038/nbt.2870

Na, D., Kim, T. Y., & Lee, S. Y. (2010). Construction and optimization of synthetic pathways in metabolic engineering. *Current Opinion in Microbiology*, *13*(3), 363–70. doi:10.1016/j.mib.2010.02.004

Naouri, P., Chagnaud, P., Arnaud, A., & Galzy, P. (1990). Purification and properties of a malolactic enzyme from Leuconostoc oenos ATCC 23278. *J. Basic Microbiol.*, *30*(8), 577–585.

O'Brien, E. J., Lerman, J. a, Chang, R. L., Hyduke, D. R., & Palsson, B. Ø. (2013). Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular Systems Biology*, *9*(693), 693. doi:10.1038/msb.2013.52

Oberhardt, M. A., Palsson, B. Ø., & Papin, J. A. (2009). Applications of genome-scale metabolic reconstructions, (320), 1–15. doi:10.1038/msb.2009.77

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., & Bono, H. (1999). KEGG : Kyoto Encyclopedia of Genes and Genomes, *27*(1), 29–34.

Olguín, N., Bordons, A., & Reguant, C. (2009). Influence of ethanol and pH on the gene expression of the citrate pathway in Oenococcus oeni. *Food Microbiology*, *26*(2), 197–203. doi:10.1016/j.fm.2008.09.004

Oliveira, A. P., Nielsen, J., & Förster, J. (2005). Modeling Lactococcus lactis using a genome-scale flux model, *15*, 1–15. doi:10.1186/1471-2180-5-39

Orth, J. D., Conrad, T. M., Na, J., Lerman, J. a, Nam, H., Feist, A. M., & Palsson, B. Ø. (2011). A comprehensive genome-scale reconstruction of Escherichia coli metabolism--2011. *Molecular Systems Biology*, *7*(535), 535. doi:10.1038/msb.2011.65

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

Pál, C., Papp, B., & Lercher, M. J. (2005). Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nature Genetics*, *37*(12), 1372–5. doi:10.1038/ng1686

Ponce-de-León, M., Montero, F., & Peretó, J. (2013). Solving gap metabolites and blocked reactions in genome-scale models: application to the metabolic network of Blattabacterium cuenoti. *BMC Systems Biology*, *7*, 114. doi:10.1186/1752-0509-7-114

Raman, K., & Chandra, N. (2009). Flux balance analysis of biological systems: applications and challenges. *Briefings in Bioinformatics*, *10*(4), 435–49. doi:10.1093/bib/bbp011

Ramos, A., Lolkema, J. S., Konings, W. N., Santos, H., Ramos, A. N. A., Lolkema, J. S., … Qui, I. D. T. (1995). Enzyme Basis for pH Regulation of Citrate and Pyruvate Metabolism by Leuconostoc oenos . These include : Enzyme Basis for pH Regulation of Citrate and Pyruvate Metabolism by Leuconostoc oenos, *61*(4).

Ramos, A., & Santos, H. (1996). Citrate and Sugar Cofermentation in Leuconostoc oenos , a (sup13) C Nuclear Magnetic Resonance Study . These include : Citrate and Sugar Cofermentation in Leuconostoc oenos , a 13 C Nuclear Magnetic Resonance Study, *62*(7).

Remize, F., Gaudin, A., Kong, Y., Guzzo, J., Alexandre, H., Krieger, S., & Guilloux-Benatier, M. (2006). Oenococcus oeni preference for peptides: qualitative and quantitative analysis of nitrogen assimilation. *Archives of Microbiology*, *185*(6), 459–69. doi:10.1007/s00203-006-0116-6

Ren, Q., Chen, K., & Paulsen, I. T. (2007). TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. *Nucleic Acids Research*, *35*(Database issue), D274–9. doi:10.1093/nar/gkl925

Ribéreau-Gayon, P., Dubourdieu, D., Donèche, B., & Lonvaud, A. (1998). *Handbook of Enology Volume 1 The Microbiology of Wine and Vinifications 2nd Edition*.

Richter, H., Vlad, D., & Unden, G. (2001). Significance of pantothenate for glucose fermentation by Oenococcus oeni and for suppression of the erythritol and acetate production. *Archives of Microbiology*, *175*(1), 26–31. doi:10.1007/s002030000233

Saadt, A. M. S. D. E. (1982). Purification and Properties of a Malolactic Enzyme from a Strain of Leuconostoc mesenteroides Isolated from Grapes, *43*(2), 357–361.

Salema, M., Lolkema, J. S., Romão, M. V. S., & Dias, M. C. L. (1996). The proton motive force generated in Leuconostoc oenos by L-malate fermentation . These include : The Proton Motive Force Generated in Leuconostoc oenos by L -Malate Fermentation, *178*(11).

Salou, P., Loubiere, P., & Pareilleux, A. (1994). Growth and energetics of Leuconostoc oenos during cometabolism of glucose with citrate or Growth and Energetics of Leuconostoc oenos during Cometabolism of Glucose with Citrate or Fructose, *60*(5).

Satish Kumar, V., Dasika, M. S., & Maranas, C. D. (2007). Optimization based automated curation of metabolic reconstructions. *BMC Bioinformatics*, *8*, 212. doi:10.1186/1471-2105-8-212

Schellenberger, J., Que, R., Fleming, R. M. T., Thiele, I., Orth, J. D., Feist, A. M., … Palsson, B. Ø. (2011). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature Protocols*, *6*(9), 1290–307. doi:10.1038/nprot.2011.308

Schilling, C. H., Covert, M. W., Famili, I., Church, G. M., Edwards, J. S., & Palsson, B. O. (2002). Genome-Scale Metabolic Model of Helicobacter pylori 26695, *184*(16), 4582–4593. doi:10.1128/JB.184.16.4582

Silveira, M. G. Da, Golovina, E. A., Folkert, A., Rombouts, F. M., Abee, T., Silveira, D., & Hoekstra, F. A. (2003). Membrane Fluidity Adjustments in Ethanol-Stressed Oenococcus oeni Cells Membrane Fluidity Adjustments in Ethanol-Stressed Oenococcus oeni Cells, *69*(10), 1–8. doi:10.1128/AEM.69.10.5826

Sohn, S. B., Graf, A. B., Kim, T. Y., Gasser, B., Maurer, M., Ferrer, P., … Lee, S. Y. (2010). Genome-scale metabolic model of methylotrophic yeast Pichia pastoris and its use for in silico analysis of heterologous protein production. *Biotechnology Journal*, *5*(7), 705–15. doi:10.1002/biot.201000078

Spettoli, P., Nuti, M. P., & Zamorani, A. (1984). Properties of Malolactic Activity Purified from Leuconostoc ML34 by Affinity Chromatography, *48*(4), 900–901.

Stolyar, S., Van Dien, S., Hillesland, K. L., Pinel, N., Lie, T. J., Leigh, J. a, & Stahl, D. a. (2007). Metabolic modeling of a mutualistic microbial community. *Molecular Systems Biology*, *3*(92), 92. doi:10.1038/msb4100131

Swiegers, J. H., Bartowsky, E. J., Henschke, P. a., & Pretorius, I. S. (2005). Yeast and bacterial modulation of wine aroma and flavour. *Australian Journal of Grape and Wine Research*, *11*(2), 139–173. doi:10.1111/j.1755-0238.2005.tb00285.x

Terrade, N., & Mira de Orduña, R. (2009). Determination of the essential nutrient requirements of wine-related bacteria from the genera Oenococcus and Lactobacillus. *International Journal of Food Microbiology*, *133*(1-2), 8–13. doi:10.1016/j.ijfoodmicro.2009.03.020

Terrade, N., Noël, R., Couillaud, R., & de Orduña, R. M. (2009). A new chemically defined medium for wine lactic acid bacteria. *Food Research International*, *42*(3), 363–367. doi:10.1016/j.foodres.2008.12.011

Teusink, B., Enckevort, F. H. J. Van, Francke, C., Wiersma, A., Wegkamp, A., Smid, E. J., … Siezen, R. J. (2005). In Silico Reconstruction of the Metabolic Pathways of Lactobacillus plantarum : Comparing Predictions of Nutrient Requirements with Those from Growth Experiments In Silico Reconstruction of the Metabolic Pathways of Lactobacillus plantarum : Comparing Pre. doi:10.1128/AEM.71.11.7253

Theobald, S., Pfeiffer, P., & König, H. (2005). Manganese-dependent growth of Oenococci. *Journal of Wine Research*, *16*(2), 171–178. doi:10.1080/09571260500317078

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

Varma, a, & Palsson, B. O. (1994). Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. *Applied and Environmental Microbiology*, *60*(10), 3724–31. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=201879&tool=pmcentrez&rendertype=abstract

Wagner, N., Tran, Q. H., Richter, H., Paul, M., Unden, G., & Selzer, P. M. (2005). Pyruvate Fermentation by Oenococcus oeni and Leuconostoc mesenteroides and Role of Pyruvate Dehydrogenase in Anaerobic Fermentation Pyruvate Fermentation by Oenococcus oeni and Leuconostoc mesenteroides and Role of Pyruvate Dehydrogenase in Anaerobic Ferm. doi:10.1128/AEM.71.9.4966

Wegkamp, a, Teusink, B., de Vos, W. M., & Smid, E. J. (2010). Development of a minimal growth medium for Lactobacillus plantarum. *Letters in Applied Microbiology*, *50*(1), 57–64. doi:10.1111/j.1472-765X.2009.02752.x

Zé-Zé, L., Tenreiro, R., Brito, L., Santos, M. a, & Paveia, H. (1998). Physical map of the genome of Oenococcus oeni PSU-1 and localization of genetic markers. *Microbiology (Reading, England)*, *144 (Pt 5*(1 998), 1145–56. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9611789

Zé-Zé, L., Tenreiro, R., & Paveia, H. (2000). The Oenococcus oeni genome: physical and genetic mapping of strain GM and comparison with the genome of a “divergent” strain, PSU-1. *Microbiology (Reading, England)*, *146 Pt 12*, 3195–204. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11101677

Zhang, D., & Lovitt, R. W. (2006). Performance assessment of malolactic fermenting bacteria Oenococcus oeni and Lactobacillus brevis in continuous culture. *Applied Microbiology and Biotechnology*, *69*(6), 658–64. doi:10.1007/s00253-005-0021-y

APPENDIX

APPENDIXA: SUPPLEMENTARY TABLES

Table A-1. Reactions found in *O. oeni* KEGG metabolic maps but not in the *O. oeni* draft reconstruction. Reactions that were added to the model were classified as accepted reactions while reactions that were not were classified as rejected reactions.

Reaction	KEGG ID	Pathway Tools ID
Accepted reactions		
1 2-dehydro-3-deoxy-D-gluconate + 1 NAD+ <-> 1 3-deoxy-D-glycero-2,5-hexodiulosonate + 1 NADH + 1 H+	R01542	1.1.1.127-RXN
1 L-alanyl-tRNAala + 1 UDP-N-acetylmuramoyl-L-alanyl-gamma-D-glutamyl-L-lysyl-D-alanyl-D-alanine <-> 1 UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-N6-(L-alanyl)-L-lysyl-D-alanyl-D-alanine + 1 tRNAala	ooe00550	2.3.2.10-RXN
2 all-trans-geranyl-geranyl diphosphate <-> 1 prephytoene diphosphate + 1 diphosphate	R02065	2.5.1.32-RXN
1 a protein-Npi-phospho-L-histidine -> 1 a sugar phosphate + 1 a [protein]-L-histidine	R03076	2.7.1.69-RXN
1 beta-nicotinamide D-ribonucleotide + 1 ATP + 1 H+ -> 1 diphosphate + 1 NAD+	R00137	2.7.7.1-RXN
1 isomaltose + 1 H2O <-> 1 alpha-D-glucose	R01718	3.2.1.10-RXN
1 a L-prolyl peptide + 1 H2O -> 1 a peptide + 1 L-proline	R00135	3.4.11.5-RXN
1 a holo-[acp] + 1 acetyl-CoA -> 1 an acetyl-[acp] + 1 coenzyme A	R01624	ACP-S-ACETYLTRANSFER-RXN
1 (R)-amygdalin + 1 H2O <-> 1 (R)-Prunasin + 1 beta-D-glucose	R02985	AMYGDALIN-BETA-GLUCOSIDASE-RXN
1 taurochenodeoxycholate + 1 H2O <-> 1 taurine + 1 chenodeoxycholate	R03977	CHENODEOXYCHOLOYLTAURINE-HYDROLASE-RXN

1 glycocholate + 1 H2O <-> 1 cholate + 1 glycine	R05835	CHOLOYLGLYCINE-HYDROLASE-RXN
1 hydroxypyruvate + 1 NADH + 1 H+ -> 1 D-glycerate + 1 NAD+	R01388	GLYCERATE-DEHYDROGENASE-RXN
1 glycolaldehyde + 1 NAD+ + 1 H2O <-> 1 glycolate + 1 NADH + 2 H+	R01333	GLYCOLALD-DEHYDROG-RXN
1 glycolate + 1 NAD+ <-> 1 glyoxylate + 1 NADH + 1 H+	R00717	GLYCOLATE-REDUCTASE-RXN
1 methylglyoxal + 1 NAD+ + 1 H2O -> 1 NADH + 1 pyruvate + 2 H+	R00203	METHYL-GLYOXAL-DEHYDROG-RXN
1 2-succinylbenzoate + 1 coenzyme A + 1 ATP -> 1 2-succinylbenzoyl-CoA + 1 diphosphate + 1 AMP	R04030	O-SUCCINYLBENZOATE-COA-LIG-RXN
1 hydroxymethylpyrimidine + 1 ATP -> 1 4-amino-2-methyl-5-phosphomethylpyrimidine + 1 ADP + 1 H+	R03471	OHMETPYRKIN-RXN
1 pantetheine + 1 ATP -> 1 4'-phosphopantetheine + 1 ADP + 1 H+	R02971	PANTETHEINE-KINASE-RXN
1 (R)-Prunasin + 1 H2O <-> 1 mandelonitrile + 1 beta-D-glucose	R02558	PRUNASIN-BETA-GLUCOSIDASE-RXN
1 propanoyl-CoA + 1 phosphate <-> 1 propanoyl phosphate + 1 coenzyme A	R00921	PTAALT-RXN
1 4-amino-2-methyl-5-phosphomethylpyrimidine + 1 ATP -> 1 4-amino-2-methyl-5-diphosphomethylpyrimidine + 1 ADP	R04509	PYRIMSYN3-RXN
1 (R)-citramalate <-> 1 citraconate + 1 H2O	R03896	R-2-

		METHYLMALATE-DEHYDRATASE-RXN
1 (R)-acetoin + 1 NADH + 1 H+ -> 1 (R,R)-2,3-butanediol + 1 NAD+	R03707	RR-BUTANEDIOL-DEHYDROGENASE-RXN
1 putrescine + 1 acetyl-CoA <-> 1 N-acetylputrescine + 1 coenzyme A + 1 H+	R01154	RXN-0
1 cellobiose + 1 H2O <-> 2 beta-D-glucose	R00026	RXN-10773
1 diacetyl + 1 NADH + 1 H+ -> 1 (R)-acetoin + 1 NAD+	R09078	RXN-11036
3-oxo-glutaryl1 malonyl-CoA methyl ester + 1 a malonyl-[acp] <-> 1 a -[acp] methyl ester + 1 CO2 + 1 coenzyme A	R10115	RXN-11474
1 2-phenylacetamide + 1 H2O <-> 1 phenylacetate + 1 ammonia + 1 H+	R02540	RXN-12492
1 butanoyl-CoA + 1 acetyl-CoA <-> 1 3-oxohexanoyl-CoA + 1 coenzyme A	R01177	RXN-12565
1 cysteinylglycine + 1 H2O -> 1 glycine + 1 L-cysteine	R00899	RXN-6622
1 citraconate + 1 H2O <-> 1 beta-methyl-D-malate	R03898	RXN-7744
1 an L-1-phosphatidyl-glycerol + 1 a CDP-diacylglycerol <-> 1 a cardiolipin + 1 CMP + 1 H+	ooe00564	RXN-8141
1 1-(beta-D ribofuranosyl)nicotinamide + 1 H2O <-> 1 beta-D-ribofuranose + 1 nicotinamide + 1 H+	R01273	RXN-8441
1 prephytoene diphosphate <-> 1 15-cis-phytoene + 1 diphosphate	R04218	RXNARA-8002

1 2-deoxy-D-ribose-5-phosphate + 1 ADP + 1 H+ <-> 1 deoxyribose + 1 ATP	R02750	RXNK9E-143
1 beta-D-fructofuranose + 1 a protein-Npi-phospho-L-histidine -> 1 fructose-1-phosphate + 1 a [protein]-L-histidine	R03232	RXNK9E-144
1 D-mannose + 1 a protein-Npi-phospho-L-histidine -> 1 D-mannose 6-phosphate + 1 a [protein]-L-histidine	ooe00051	RXNK9E-145
1 sucrose + 1 H2O <-> 1 beta-D-fructofuranose + 1 alpha-D-glucose	R00801	RXNK9E-146
1 D-galactitol + 1 a protein-Npi-phospho-L-histidine -> 1 galactitol-1-phosphate + 1 a [protein]-L-histidine	R05570	RXNK9E-147
1 epimelibiose + 1 H2O <-> 1 D-mannose + 1 D-galactose	R01329	RXNK9E-151
1 L-ascorbate + 1 a protein-Npi-phospho-L-histidine -> 1 L-ascorbate-6-phosphate + 1 a [protein]-L-histidine	R07671	RXNK9E-152
1 alpha-maltose + 1 a protein-Npi-phospho-L-histidine -> 1 maltose 6-phosphate + 1 a [protein]-L-histidine	R04111	RXNK9E-154
1 a beta-D glucoside + 1 H2O <-> 1 an alcohol + 1 alpha-D-glucose	R03527	RXNK9E-156
1 cellulose + 1 H2O <-> 1 cellulose + 1 beta-D-glucose	R02887	RXNK9E-157
1 a dextrin + 1 H2O <-> 1 a dextrin + 1 alpha-D-glucose	R01791	RXNK9E-158
1 N-acetylmuramate + 1 a protein-Npi-phospho-L-histidine + 2 H+ -> 1 N-Acetylmuramic acid 6-phosphate + 1 a [protein]-L-histidine	R08559	RXNK9E-159
1 thiamin diphosphate + 1 (S)-2-acetolactate <-> 1 2-(alpha-hydroxyethyl)thiamine diphosphate + 1 pyruvate	R03050	RXNK9E-160
1 thiosulfate + 1 a reduced thioredoxin + 1 O-acetyl-L-serine <-> 1 an oxidized thioredoxin + 1 sulfite + 1 acetate + 1 L-cysteine + 2 H+	R04859	RXNK9E-161
1 methylselenic acid + 2 NADPH + 2 H+ <-> 1	R09372	RXNK9E-165

methaneselenol + 2 NADP+ + 2 H2O		
1 D-4'-phosphopantothenate + 1 L-cysteine + 1 ATP -> 1 R-4'-phosphopantothenoyl-L-cysteine + 1 diphosphate + 1 AMP + 1 H+	R04230	RXNK9E-169
1 N-((R)-Pantothenoyl)-L-cysteine + 1 ATP -> 1 R-4'-phosphopantothenoyl-L-cysteine + 1 ADP + 1 H+	R04391	RXNK9E-170
1 tetrahydrofolate + 1 NAD+ <-> 1 7,8-dihydrofolate + 1 NADH + 1 H+	R00936	RXNK9E-171
1 folate + 2 NADH + 2 H+ -> 1 tetrahydrofolate + 2 NAD+	R00937	RXNK9E-172
1 tetrahydrofolate + 2 NADP+ <-> 1 folate + 2 NADPH + 2 H+	R00940	RXNK9E-173
1 7,8-dihydrofolate + 1 NAD+ <-> 1 folate + 1 NADH + 1 H+	R02235	RXNK9E-174
1 7,8-dihydrofolate + 1 NADP+ <-> 1 folate + 1 NADPH + 1 H+	R02236	RXNK9E-175
1 taurocholate + 1 H2O <-> 1 taurine + 1 cholate	R02797	RXNK9E-178
1 glycochenodeoxycholate + 1 H2O <-> 1 chenodeoxycholate + 1 glycine	R03975	RXNK9E-179
1 thiamin diphosphate + 1 ATP -> 1 thiamin triphosphate + 1 ADP	R00616	THIAMIN-DIPHOSPHATE-KINASE-RXN
4 isopentenyl diphosphate + 1 (2E,6E)-farnesyl diphosphate <-> 1 all-trans-heptaprenyl diphosphate + 4 diphosphate	R09247	TRANS-HEXAPRENYLTRANSTRANSFERASE-RXN
Rejected reactions		
1 L-1-glycero-3-phosphocholine [c] + 1 H2O [c] <-> 1 choline [c] + 1 sn-glycerol-3-phosphate [c] + 1 H+ [c]	R01030	3.1.4.2-RXN

1 L-cystine [c] + 1 H2O [c] <-> 1 thiocysteine [c] + 1 ammonium [c] + 1 pyruvate [c]	R02408	CYSTHIOCYS-RXN
1 L-2-acetamido-6-oxoheptanedioate [c] + 1 L-glutamate [c] <-> 1 N-acetyl-L,L-2,6-diaminopimelate [c] + 1 2-oxoglutarate [c]	ooe00300	RXN-4822
1 linamarin [c] + 1 H2O [c] <-> 1 acetone cyanohydrin [c] + 1 beta-D-glucose [c]	R10040	RXN-5341
1 a [Cys-Gly]-S-conjugate [c] + 1 H2O [c] <-> 1 an L-cysteine-S-conjugate [c] + 1 glycine [c]	R04951	RXN-6642
1 (2E,4Z)-2-hydroxymuconate [c] <-> 1 (3Z)-2-oxohex-3-enedioate [c]	R03966	RXN-8844
1 (3R)-3-hydroxy-stearoyl-CoA [c] + 1 NADP+ [c] <-> 1 3-oxo-stearoyl-CoA [c] + 1 NADPH [c] + 1 H+ [c]	R07763	RXN-9544
1 dhurrin [c] + 1 H2O [c] <-> 1 beta-D-glucose [c] + 1 (S)-4-hydroxymandelonitrile [c]	R10035	RXN-9588
1 lotaustralin [c] + 1 H2O [c] <-> 1 (2R)-2-hydroxy-2-methylbutanenitrile [c] + 1 beta-D-glucose [c]	R10039	RXN-9674
1 alpha-D-glucose [c] + 1 ATP [c] -> 1 alpha-D-glucose 6-phosphate [c] + 1 ADP [c] + 1 H+ [c]	R01786	RXNK9E-141
1 galactinol [c] + 1 H2O [c] <-> 1 myo-inositol [c] + 1 alpha-D-galactose [c]	R01194	RXNK9E-149
1 melibiitol [c] + 1 H2O [c] <-> 1 D-sorbitol [c] + 1 D-galactose [c]	R02926	RXNK9E-150
1 sucrose [c] <-> 1 a dextran [c] + 1 D-fructose [c]	R02120	RXNK9E-153
1 sucrose [c] + 1 phosphate [c] <-> 1 a D-glucose-1-phosphate [c] + 1 D-fructose [c]	R00803	RXNK9E-155
1 trans-4-hydroxy-L-proline [c] + 1 NAD+ [c] <-> 1 pyrroline-hydroxy-carboxylate [c] + 1 NADH [c] + 2 H+ [c]	R03291	RXNK9E-162
1 ATP [c] + 1 selenomethionine [c] + 1 tRNAmet [c] <-> 1 AMP [c] + 1 diphosphate [c] + 1 selenomethionyl-	R04773	RXNK9E-164

tRNA(Met) [c]		
1 hydrogen selenide [c] + 3 NADP+ [c] + 3 H2O [c] <-> 1 selenite [c] + 3 NADPH [c] + 5 H+ [c]	R03596	RXNK9E-167
1 4-phospho-hydroxy-L-threonine [c] + 1 H2O [c] <-> 1 4-hydroxy-L-threonine [c] + 1 phosphate [c]	R05086	RXNK9E-168
1 NADH [c] + 1 H+ [c] + 1 a menaquinone [c] <-> 1 NAD+ [c] + 1 a menaquinol [c]	R02964	RXNK9E-176
1 phylloquinone [c] + 1 NADH [c] + 3 H+ [c] <-> 1 phylloquinol [c] + 1 NAD+ [c]	R03816	RXNK9E-177
1 an alpha,beta-digalactosyldiacylglycerol [c] + 1 H2O [c] <-> 1 a 1,2-diacyl-3-beta-D-galactosyl-sn-glycerol [c] + 1 D-galactose [c]	R04470	RXNK9E-180
1 L-1-glycerophosphorylethanolamine [c] + 1 H2O [c] <-> 1 sn-glycerol-3-phosphate [c] + 1 ethanolamine [c] + 1 H+ [c]	R01470	RXNK9E-181
1 a beta-D-galactosyl-(1->4)-beta-D-glucosyl-(1<->1)-ceramide [c] + 1 H2O [c] + 2 H+ [c] <-> 1 a D-glucosyl-N-acylsphingosine [c] + 1 D-galactose [c]	R03355	RXNK9E-182
1 digalactosylceramide [c] + 1 H2O [c] <-> 1 a cerebroside [c] + 1 D-galactose [c]	R04019	RXNK9E-183
1 5(S)-HPETE [c] + 2 glutathione [c] + 1 H+ [c] <-> 1 5(S)-HETE [c] + 1 glutathione disulfide [c] + 1 H2O [c]	R07034	RXNK9E-184
1 15(S)-HPETE [c] + 2 glutathione [c] <-> 1 15(S)-HETE [c] + 1 glutathione disulfide [c] + 1 H2O [c]	R07035	RXNK9E-185
1 3-ketopimeloyl-[acp] methyl ester [c] + 1 NADPH [c] + 1 H+ [c] <-> 1 3-hydroxypimeloyl-[acp] methyl ester [c] + 1 NADP+ [c]	R10120	RXNK9E-186
1 (3R)-3-Hydroxybutanoyl-[acyl-carrier protein] [c] + 1 NADP+ [c] <-> 1 acetoacetyl-[acp] [c] + 1 NADPH [c] + 1 H+ [c]	R04333	RXNK9E-187
1 trans-3-Chloro-2-propene-1-ol [c] + 1 NAD+ [c] <-> 1	R05233	RXNK9E-188

trans-3-Chloroallyl aldehyde [c] + 1 NADH [c] + 1 H+ [c]		
1 cis-3-Chloro-2-propene-1-ol [c] + 1 NAD+ [c] <-> 1 cis-3-Chloroallyl aldehyde [c] + 1 NADH [c] + 1 H+ [c]	R05234	RXNK9E-189
1 2-hydroxy-5-methyl-cis,cis-muconate [c] <-> 1 2-oxo-5-methyl-cis-muconate [c]	R05389	RXNK9E-190
1 1-hydroxymethylnaphthalene [c] + 1 NAD+ [c] <-> 1 1-naphthaldehyde [c] + 1 NADH [c] + 1 H+ [c]	R06917	RXNK9E-191
1 (2-naphthyl)methanol [c] + 1 NAD+ [c] <-> 1 2-naphthaldehyde [c] + 1 NADH [c] + 1 H+ [c]	R06927	RXNK9E-192
1 selenocysteine [c] + 1 a reduced electron acceptor [c] <-> 1 selenide [c] + 1 an oxidized electron acceptor [c] + 1 L-alanine [c] + 1 H+ [c]	R03599	SELENOCYSTEINE-LYASE-RXN
1 5,10-methylene-tetrahydromethanopterin [c] + 1 glycine [c] + 1 H2O [c] <-> 1 tetrahydromethanopterin [c] + 1 L-serine [c]	R09099	THMPT-SER-RXN

APPENDIX B: SUPPLEMENTARY FIGURES

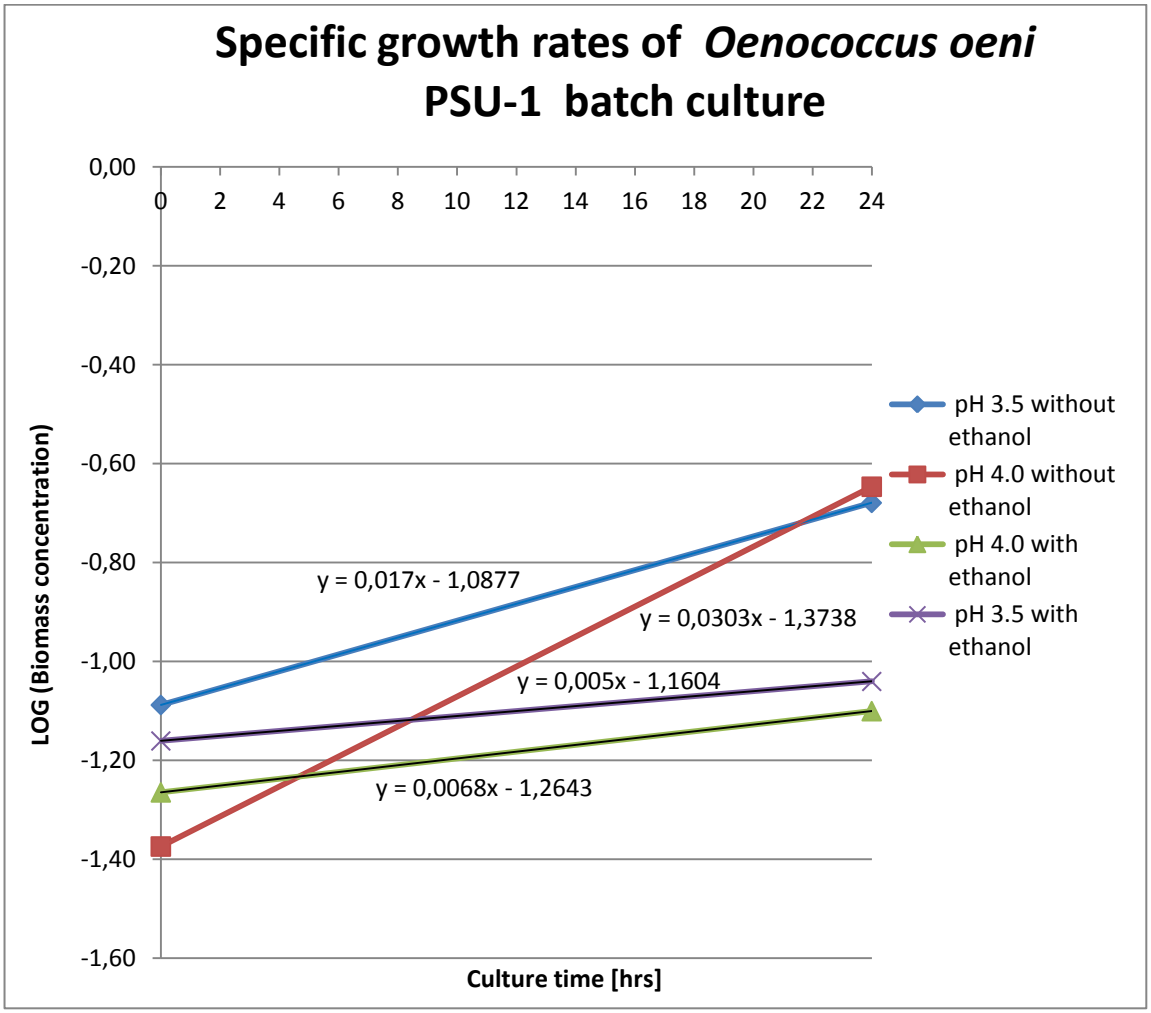

Figure B-13. Logarithm of biomass concentration versus culture time. Input data for this plot is found in Table 4-14. The slope of each trendline represents the growth rate for each of the batch culture conditions recorded in Table 4-14.

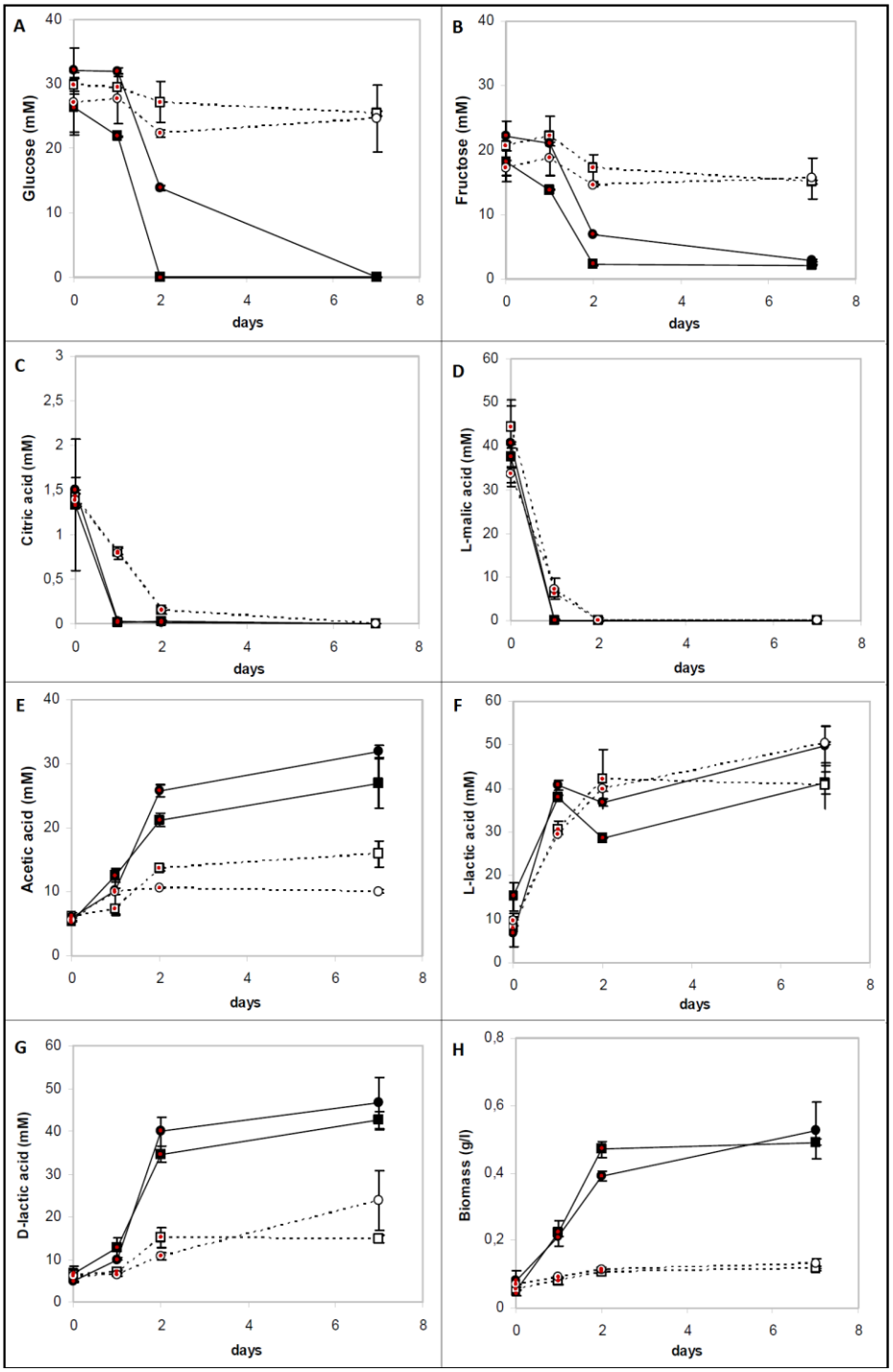

Figure B-2. Growth and metabolic monitoring of *O. oeni* PSU-1 at pH 4.0 (squares) and pH 3.5 (circles), in the absence (solid lines and filled symbols) and presence of 10% ethanol (dashed lines and empty symbols). A: glucose, B: fructose, C: citrate, D: L-malic acid, E: acetic acid, F: L-lactic acid, G: D-lactic-acid, H: biomass, dry weight. Red points represent the data extracted with WebPloAnalyzer. Figure modified from Olguín et al, (2009)