

UNIVERSIDADE DE LISBOA
INSTITUTO SUPERIOR TÉCNICO

Synthetic biology approaches to foster Yeasts as hosts for the production of carboxylic acids and their derivatives: emphasis on levulinic and itaconic acids

Ana Isabel de Vila-Santa Braga Campos

Supervisor: Doctor Nuno Gonçalo Pereira Mira
Co-Supervisors: Doctor Frederico Castelo Alves Ferreira
Doctor Kristala Jones Prather

Thesis approved in public session to obtain the PhD Degree in
Bioengineering

Jury Final Classification: Pass with Distinction

2021

UNIVERSIDADE DE LISBOA
INSTITUTO SUPERIOR TÉCNICO

Synthetic biology approaches to foster yeasts as hosts for the
production of carboxylic acids and their derivatives: emphasis on
levulinic and itaconic acids

Ana Isabel de Vila-Santa Braga Campos

Supervisor: Doctor Nuno Gonçalo Pereira Mira

Co-Supervisors: Doctor Frederico Castelo Alves Ferreira
Doctor Kristala Jones Prather

Thesis approved in public session to obtain the PhD Degree in
Bioengineering
Jury Final Classification: Pass with Distinction

Jury

Chairperson: Doctor Duarte Miguel de França Teixeira dos Prazeres, Instituto Superior Técnico, Universidade de Lisboa;

Members of the Committee:

Doctor Duarte Miguel de França Teixeira dos Prazeres, Instituto Superior Técnico, Universidade de Lisboa;

Doctor Miguel Nobre Parreira Cacho Teixeira, Instituto Superior Técnico, Universidade de Lisboa;

Doctor Zengyi Shao, Iowa State University, EUA;

Doctor Isabel Cristina de Almeida Pereira da Rocha, Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa;

Doctor Nuno Gonçalo Pereira Mira, Instituto Superior Técnico, Universidade de Lisboa.

Funding Institution:

Fundação para a Ciência e Tecnologia

Resumo

O aumento da procura de moléculas plataforma, necessárias para a produção em massa de produtos como de roupas, peças de automóvel e cosméticos, entre outros, e o facto destas moléculas serem derivadas de processos petroquímicos tem incitado uma corrente de investigação focada no desenvolvimento de processos mais ecológicos de produção destas moléculas. Esta tese encontra-se dentro desse âmbito focando a sua atenção no desenvolvimento de processos que permitam a obtenção dos ácidos levulínico, itacónico e metacrílico, por fermentação microbiana em leveduras. Ao passo que no caso do ácido itacónico, a sua produção em micróbios já se encontrava descrita e o que se descreve nesta tese se centra no melhoramento do processo através da engenharia de novos hospedeiros microbianos, o ácido levulínico e o ácido metacrílico constituem moléculas denominadas *new-to-nature* uma vez que não fazem parte das redes metabólicas conhecidas de micróbios. O potencial dos ácidos levulínico, itacónico e metacrílico enquanto moléculas plataformas bem como as estratégias experimentais que podem ser usadas para implementar a produção por via microbiana de moléculas *new-to-nature*, essencialmente baseadas em abordagens de biologia sintética moderna suportada por análises computacionais, encontram-se descritas no capítulo 1 desta tese. No capítulo 2 é feita uma descrição geral do trabalho executado durante esta tese.

No terceiro capítulo descreve-se, pela primeira vez, a implementação de uma estratégia para seleccionar vias biosintéticas com maior potencial para a produção de ácido levulínico em micróbios, resultando na descrição de 4 novas vias biosintéticas de ácido levulínico, através dos precursores ácido δ -aminolevulínico, glutamato-1-semialdeído, ácido 3-oxoadípico e ácido 5-aminovalérico. Destas vias, a que assenta na utilização do precursor δ -aminolevulínico e das enzimas 8-amino-7-oxononanoato transaminase e diaminopropiónico amónia liase foi seleccionada para posterior validação experimental sendo esses resultados descritos no capítulo 4 desta tese. Para essa validação experimental tentou-se confirmar a capacidade das enzimas escolhidas na abordagem *in silico* catalisarem as reacções sintéticas identificadas, tendo sido possível confirmar que a enzima 8-amino-7-oxononanoato transaminase leva à conversão de ácido δ -aminolevulínico em ácido 4,5-diaminovalérico, embora em quantidades pouco significativas, sugerindo que

será necessário no futuro intervenções (por exemplo através de engenharia enzimática) que permitam aumentar a atividade desta enzima e, conseqüentemente, o aumento deste passo. Durante o tempo que decorreu o trabalho experimental desta tese não foi possível confirmar se o segundo passo de reação, que promove a conversão de ácido 4,5-diaminovalérico em ácido levulínico, poderia ser catalisado pela enzima identificada *in silico*, sendo este um aspeto que terá que ser analisado de futuro. Outra abordagem que foi usada para aumentar o rendimento da via seleccionada constituiu na engenharia de uma estirpe de *S. cerevisiae* que sobre produza ácido δ -aminolevulínico, tendo-se para isso tentado promover o silenciamento do gene essencial HEM2, que codifica a enzima ácido δ -aminolevulínico desidratase, envolvida na conversão deste precursor em porfobilinogénio, através de um sistema dependente de doxiciclina. Apesar da estirpe obtida apresentar uma produção total mais baixa que a selvagem, a produção específica de D-ALA é superior ao fim de 48 horas, embora não tenha sido possível confirmar de forma inequívoca que esse fenótipo se deve à redução da expressão do gene HEM2.

Uma estratégia semelhante à utilizada para fazer a prospeção de vias para a produção de ácido levulínico foi usada para encontrar vias candidatas à produção de ácido metacrílico, resultando na seleção de 6 potenciais vias (descritas no capítulo 5). Destas, 5 já tinham sido previamente descritas, iniciando-se pelo glutamato, valina ou timina, e uma não tinha sido descrita previamente, iniciando-se pelo glioxilato. De entre as várias possibilidades encontradas considerou-se mais promissora a possibilidade de obter ácido metacrílico por descarboxilação direta do ácido mesacónico através de duas descarboxíases, AtCad1 de *Aspergillus terreus* e ACMSD de *Pseudomonas fluorescens*. A realização de ensaios enzimáticos *in vitro* usando extratos enriquecidos nestas duas enzimas não permitiu demonstrar que as mesmas sejam capazes de promover a conversão desejada, embora seja de realçar que os métodos usados para a deteção de MAA podem não ter sido suficientemente sensíveis para detetar eventuais quantidades mínimas produzidas.

Finalmente, no capítulo 6 são descritos os resultados obtidos quanto ao desenho das partes genéticas que podem ser utilizadas para o desenho de um sensor de ácido itacónico em *S. cerevisiae*, utilizando a sua rede genética endógena, e resultando na descoberta de 3 genes responsivos a ácido itacónico (DAL5, MEP2 e TMT1) com uma curva dose-

resposta linear e 2 possíveis reguladores que podem estar a orquestrar a resposta (Leu3 e Dal81). Para além de permitir o *screening* rápido e imediato de estirpes de *S. cerevisiae* com capacidade aumentada para a produção de ácido itacónico, os resultados deste capítulo também forneceram pistas importantes acerca dos mecanismos de toxicidade exercidos pelo ácido itacónico nas células de levedura, um conhecimento que pode ser usado para melhorar a capacidade produtiva destas células.

Palavras-chave: biologia sintética, prospeção de vias, ácido levulínico, ácido metacrílico, ácido itacónico

Abstract

The increasing demand of bulk chemicals, required for the mass production of clothes, automobile parts and cosmetics (among other products), and the fact that these molecules are derived from petrochemical sources has initiated a stream of research focused on the development of greener processes to obtain bulk chemicals, including using fermentation. This thesis is framed in this context, with the main goal of using yeast fermentation to produce itaconic, levulinic and methacrylic acids. While itaconic acid is naturally produced by fungi and the efforts described in this thesis are aimed at improving the already established yeast-based production of this acid, levulinic and methacrylic acids are “new-to-nature” molecules, since these are not found in the metabolism of any organism. In the first chapter of this thesis it is reviewed the potential of the three acids as platform molecules and the main experimental approaches that have been used to establish their bio-based production. The contribution of synthetic biology methodologies and pathway prospecting tools to aid in these tasks is also described. In the second chapter there is an introduction to the work produced during this thesis.

In the third chapter a strategy is described to select potential biosynthetic pathways that can be introduced in microbes leading to levulinic acid production, resulting in the description of 4 new levulinic acid biosynthetic pathways, from the precursors δ -aminolevulinic acid, glutamate-1-semialdehyde, 3-oxoadipic acid and 5-aminovaleric acid. Of these, the pathway relying on the ubiquitous precursor δ -aminolevulinic acid and the enzymes 8-amino-7-oxononanoate transaminase and diaminopropionate ammonia lyase was selected for posterior experimental validation, described in chapter 4 of this thesis. The ability of these enzymes to catalyze the required synthetic reactions was analyzed *in vitro*, enabling the confirmation that the enzyme 8-amino-7-oxononanoate can catalyze the amination of δ -aminolevulinic acid to 4,5-diaminovaleric acid, although in low amounts, suggesting that future interventions will be required to increase the enzyme's activity with the non-native substrate, including enzyme engineering. At the time of this work, it was not possible to confirm that the second enzymatic step, entailing the deamination of 4,5-aminovaleric acid to levulinic acid could be catalyzed by the enzyme identified in the *in silico* work. Another goal in this chapter was improving the

pool of this pathway's precursor, δ -aminolevulinic acid, in *S. cerevisiae* through the doxycycline-dependent downregulation of HEM2, encoding δ -aminolevulinic acid dehydratase, the enzyme responsible for the conversion of the precursor in porphobilinogen. While the titer obtained with the engineered strain after 48 hours of fermentation is lower than the wild-type strain, the engineered strain presents a higher specific production, which could not be confirmed to be attributed to a lower HEM2 expression.

A similar strategy that was used to prospect for pathways for levulinic acid production was also applied to methacrylic acid, resulting in the selection of 6 potential pathways (described in chapter 5). Of these, 5 had been previously described, steaming from glutamate, valine or thymine and one, starting from glyoxylate, had not been previously described. From the pathways described the one entailing the direct decarboxylation of mesaconic acid was selected for further validation, through the action of two decarboxylases: AtCad1 from *Aspergillus terreus* and ACMSD from *Pseudomonas fluorescens*. The results obtained from *in vitro* enzymatic assays using enriched crude cell extracts could not confirm the decarboxylations, although the detection method used (HPLC) may not be sensitive to the concentrations used here.

Finally, in chapter 6 it is described a strategy to find genetic parts that can be used to develop an itaconic acid sensor in *S. cerevisiae* by using its endogenous regulatory network and it was uncovered 3 potential itaconic acid-responsive genes (DAL5, MEP2 and TMT1) with a linear dose-response curve and 2 possible regulators behind the response (Leu3 and Dal81). While these results may open the door for a fast screening methodology of itaconic acid producer strains, these also provide some clues on the mechanism of itaconic acid toxicity in yeast, required to improve the cellular tolerance to this acid.

Keywords: synthetic biology, pathway prospecting, levulinic acid, methacrylic acid, itaconic acid

Acknowledgments

I would like to thank my supervisors for the opportunity to work in this project and their continuous guidance. To Prof. Nuno Mira, thank you for the teaching, brainstorming sessions, opportunities to investigate new problems and motivation that were so much appreciated. To Prof. Kristala Prather I'm grateful for welcoming me in the group, for providing the opportunity to learn and for believing in the project. To Prof. Frederico Ferreira I would like to thank for the advices and curricular guidance that helped in having a bigger picture perspective of this thesis.

I would also like to thank Dr. Ahsan Islam for the guidance in the pathway prospecting work.

A thank you also to Professor Guo-Qiang Chen (Tsingua University), who shared the pET28a-HEM1 plasmid, and to by Prof. Dr. Aimin Liu (Department of Biochemistry, University of Texas Southwestern Medical Center), for providing the the pET16b-nbaD plasmid.

Thank you to FCT and the MIT Portugal Program for funding of the PhD grant (PD/BD/114121/2015). Research performed in this thesis was financially supported by FCT and by the EraNet in Industrial Biotechnology (6th edition) through funding of the TTRAFFIC (Toxicity and Transport for Fungal production of Industrial Compounds) research project (ERAIB2/6/003/2014), through funding to iBB (contracts UID/BIO/04565/2019, UID/BIO/04565/2013 and UIDB/04565/2020) and ERA-IB-2-6/0003/2014. Support received from Programa Operacional Regional de Lisboa 2020 (project no. 007317) is also acknowledged.

To the BSRG and Prather Lab members, thank you for the help in the lab and troubleshooting discussions. I would also like to thank Mafalda, Sara, Maria João, Ricardo, Joana and Melike for the companionship in the lab, making it a happier place. To all the NPM team, thank you for being such a great group to work with. I would also like to thank Fernão and Inês for their work in the project and their enthusiasm and Fernão for the contributions to Chapter 5.

Obrigada aos MITugas pela companhia, piadas, corridas, incursões à free food, viagens e partilha de hábitos portugueses no geral.

Aos meus amigos um grande agradecimento pela paciência para as nerdices de leveduras e momentos felizes sem os quais estes anos não teriam sido tão alegres. Caty, Sofia, Miru, Porto, Freire, João Luís, Cátia: obrigada pela vossa amizade. Ao Rodrigo,

obrigada pelo teu apoio e companheirismo. À minha mãe, que merece o maior agradecimento pela sua presença em tantos momentos e apoio incondicional. E ao meu irmão, a quem dedico esta tese, és o meu exemplo de determinação e generosidade.

Outputs

Papers:

- Vila-Santa, Ana; Islam, M. Ahsanul; Ferreira, Frederico C.; Prather, Kristala L.J.; Mira, Nuno P. “*In silico* prospecting of biochemical pathways for microbe-based production of levulinic acid” (under submission in ACS Synthetic Biology)
- Vila-Santa, Ana; Ferreira, Frederico C.; Prather, Kristala L.J.; Mira, Nuno P. Exploring synthetic biology to bridge the gap between microbe-based production of carboxylic acids and of their “new-to-nature” derivatives: a case study focused on itaconic acid and its derivatives metacrylic acid (MAA) and methyl-metacrylic acid (MMA) (in preparation for Submission to Frontiers in Bioengineering)
- Sun, W.; Vila-Santa, A.; Liu, N.; Prozorov, T.; Xie, Dongming; Faria, N.T.; Ferreira, F.C.; Mira, N.P.; Shao, Z. “Metabolic Engineering of an Acid-Tolerant Yeast Strain *Pichia kudriavzevii* for Itaconic Acid Production”. Metab Eng Commun. vol 10, June 2020.

Conference Posters:

- 1st Open Meeting EuroMicroPH (Lisbon, Feb 2020) - Vila-Santa, Ana; Islam, M. Ahsanul; Ferreira, Frederico C.; Prather, Kristala L.J.; Mira, Nuno P. “Pathway prospecting for the implementation of microbe-based production of levulinic acid: a computational approach.”
- 7th Conference on Physiology of Yeast and Filamentous Fungi (PYFF) (Milan, July 2020)
 - Vila-Santa, Ana.; Moita, Raquel M.; Rodrigues, Nicole; Brito, Tiago; Carvalho, Carla; Mira, Nuno P. “Chemogenomic and transcriptomic analysis of *Saccharomyces cerevisiae* response and tolerance to itaconic acid stress.”
 - Vila-Santa, Ana; Lúcio, João; Sun, Wan; Correia, Jéssica; Marques, Nuno; Faria, Nuno; Vinga, Susana; Ferreira, Frederico C.; Shao, Zenghy; Mira, Nuno P. “Towards sustainable Yeast-based production of itaconic acid”.
- 28th International Conference on Yeast Genetics and Molecular Biology (ICYGMB) (Prague, Aug 2017) – Moita, Maria R.; Rodrigues, Nicole; Vila-Santa, Ana; Lúcio, João Pedro; Carvalho, Carla; Henriques, Sílvia; Sá-Correia, Isabel; Mira, Nuno P. “Chemogenomic Analyses Unveils New Players Governing *Saccharomyces cerevisiae* Tolerance To The Add-Value Itaconic And Levulinic Carboxylic Acid.”

Abbreviations

2,4 - 2,4-Dichlorophenoxyacetic Acid

3-HPO – 3-Hydroxypropionic

5-FOA 5-FluoOrotic Acid

5-HAnt - 5-Hydroxyanthranilate

ACMS - Aminocarboxymuconate-Semialdehyde

ACMSD - Aminocarboxymuconate-Semialdehyde Decarboxylase

Amp - Ampicillin

An - Anthranilate

BDO - Butanediol

BEM - Bond Electron Matrix

BSA – Bovine Serum Albumin

CA – Carboxylic Acid

CAGR – Compound Annual Growth Rate

CCM – cis-cis-Muconic acid

CDS - Coding Sequence

CPEC – Circular Polymerase Extension Cloning

DOE – Department Of Energy (U.S.A)

Dox - Doxycycline

DTT - Dithiothreitol

E.C. – Enzyme Comission

EIC - Extracted Ion Chromatogram

FACS – Fluorescent Activated Cell Sorting

FDCA – 2,5-Furan Dicarboxylic Acid

FRET - Forster-Ressonance Energy Transfer

GABA - Gamma Aminobutyric Acid

GBL - Gamma Butyrolactone

gDNA - genomic DNA

GEMs – Genome Scale Metabolic models

GRAS – Generally Regarded As Safe

HEMA 2-Hydroxyethyl Methacrylic Acid

HMF – 5-Hydroxymethylfurfural

HP-ITA - 2-Hydroxyparaconate

HPLC – High Performance Liquid Chromatography

HPLC High Performance Liquid Chromatography

IPM – 3-Isopropylmalate

ITA – Itaconic acid

Kan - Kanamycin

KANA - 7-Keto-8-Aminononanoic Acid

KEGG - Kyoto Encyclopedia of Genes and Genomes

KO – Knock-Out

LA – Levulinic Acid

LC-MS – Liquid Chromatography – Mass Spectrometry

LFC - Log Fold-Changes

MAA – Methacrylic Acid

MAA-coA – Methacrylyl-coA

MAPK - Mitogen Activated Protein Kinase

MCS - Multiple Cloning Site

MEG – Monoethylene Glycol

MEK – Methylketone

MFS - Major Facilitator Superfamily

MMA – Methyl Methacrylic Acid

mtDNA - mitochondrial DNA

MTHF - 5,10-Methylenetetrahydrofolate

MW - Molecular Weight

NCR - Nitrogen Catabolite Repression

PLP - Pyridoxal Phosphate

RBS - Ribosome Binding Site
RFP - Red Fluorescent Protein
RT-qPCR Reverse Transcription Quantitative PCR
SAM - Adenosylmethionine
SMARTS - SMILES Arbitrary Target Specification
SMILES - Simplified Molecular Input Line Entry Specification
TCA - TriChloroacetic Acid
TF - Transcription Factor
THF - Tetrahydrofuran
THMPT - Methylene-Tetrahydromethanopterin
tTA - tetracyclin Transactivator Protein
UAS - Upstream Activatin Sequence
UAS_{NTR} - UAS Nitrogen rRegulated
UIS - Upstream Inducing Sequence
URS - Upstream Represison Sequence
WT – Wild-Type
YIP - Yeast Integrative Plasmid

Index

Chapter 1 – Introduction to the theme of the thesis	1
1.1. Thesis Outline	3
Chapter 2 – State of the Art	7
2.1. Overview	8
2.1.1. Main CA building blocks identified in the literature.....	10
2.1.2. Itaconic acid.....	20
2.1.3. Methacrylic acid	22
2.1.4. Levulinic acid	24
2.2. Approaches to implement bio-based production of LA, MAA and ITA	26
2.2.1. Itaconic Acid.....	26
2.2.2. Methacrylic Acid	31
2.2.3. Levulinic Acid	33
2.3. How can synthetic biology contribute for the production of “new-to-nature molecules” like MAA or LA.....	36
2.3.1. Overview of prospection of <i>new-to-nature</i> pathways	38
2.3.2. Design of synthetic pathways: approaches to the problem.....	39
2.3.3. Computational pathway prospection tools	43
2.3.4. Applications of pathway prospecting tools	49
2.4. References	51
Chapter 3 – <i>In silico</i> prospecting of metabolic pathways for the implementation of microbe-based production of levulinic acid	65
3.1. Abstract	66
3.2. Introduction	67
3.3. Methods.....	70
3.3.1. Precursor search.....	70

3.3.2. Enzyme assignment	71
3.3.3. Pathway yield analysis using <i>in silico</i> metabolic modelling	71
3.4. Results and Discussion.....	72
3.5. Conclusions	91
3.6. References	94
Chapter 4 – Strategies for the successful implementation of an <i>in silico</i> predicted LA-forming pathway using D-ALA as a precursor	99
4.1. Abstract	100
4.2. Introduction	101
4.3. Materials and Methods	105
4.3.1. Microorganisms and culture conditions	105
4.3.2. General molecular biology procedures.....	107
4.3.3. Genome integrations to obtain the H2 strain	107
4.3.4. Cloning in of LA pathway enzymes in pETDuet plasmids	108
4.3.5. Yeast fermentations to over-produce D-ALA	108
4.3.6. Assessment of HEM2 expression in W303 and H2 cells	110
4.3.7. <i>In vitro</i> assays of the BioA, BioK and YgeX enzymes	110
4.3.7.1. Preparation of crude cell extracts.....	110
4.3.7.2. Enzymatic assays	111
4.3.7.3. Assessment by LC-MS of the extract obtained after the enzymatic step	111
4.3.8. <i>In vivo</i> assays to obtain LA in <i>E. coli</i> cells	112
4.3.8.1. Using cells that over-produce D-ALA and that over-express BioA/BioK and YgeX	112
4.3.8.2. Using a two-step bioconversion exploring cells that over-produce D-ALA which is used as a growth supplement for cells over-expressing BioA/BioK and YgeX.....	113
4.4. Results and Discussion.....	114
4.4.1. Can the candidate enzymes identified in the <i>in silico</i> analysis promote the	

conversion of D-ALA into D-AVA and of this into LA?	114
4.4.2. In vivo assembly of the whole DALA - LA conversion pathway in <i>E. coli</i>	123
4.4.3. Improvement of D-ALA internal pool in <i>S. cerevisiae</i> cells	129
4.5. Conclusion and Future Work	139
4.6. References	140
Chapter 5 – Prospecting of metabolic pathways for implementation of microbe-based production of Methacrylic Acid (MAA)	149
5.1. Abstract	150
5.2. Introduction	151
5.3. Methods	154
5.3.1. <i>In silico</i> pathway prospecting	154
5.3.2. Experimental methods	154
5.3.2.1. Microorganisms and culture conditions	154
5.3.2.2. In vitro enzymatic assays to assess the capacity of <i>A. terreus</i> CadA to convert mesaconic acid into MAA	156
5.3.2.3. Preparation of <i>E. coli</i> crude cell extracts enriched in <i>Pseudomonas</i> <i>syringae</i> ACMSD	157
5.4. Results and Discussion	158
5.4.1. <i>In silico</i> prospecting of pathways for microbe-based production of MAA	158
5.4.1.1. Selection of precursors	158
5.4.1.2. Enzyme assignment	165
5.4.2. Exploration of mesaconic acid-decarboxylation as a putative route for synthesis of MAA	174
5.4.2.1. Exploring AtCadA as a possible decarboxylase of mesaconic acid	176
5.4.2.2. Exploring ACMSD as a possible decarboxylase of mesaconic acid	178

5.5. Conclusions and Future Work.....	181
5.6. References	183
Chapter 6 – Searching for the genetic parts to develop a yeast-based biosensor for itaconic acid.....	189
6.1. Abstract	190
6.2. Introduction	191
6.3. Methods.....	194
6.3.1. Microorganisms and culture conditions	194
6.3.2. Cell cultivation in the presence of ITA and RNA extraction	194
6.3.3. RNA sequencing.....	195
6.4. Results and Discussion.....	197
6.4.1. Short-term <i>S. cerevisiae</i> transcriptomic response to ITA.....	198
6.4.1.1. ITA down-regulated genes.....	204
6.4.2. Transcriptomic <i>S. cerevisiae</i> response to ITA in the exponential phase	209
6.4.3. Experimental confirmation of the results obtained in the RNA-seq of yeast ITA-stressed cells.....	212
6.4.4. Overall transcriptomic response to ITA exposure and physiological significance.....	217
6.5. Conclusions and Future Work.....	223
6.6. References	225
Chapter 7 – Summary of the work.....	231
Chapter 8 – Annex	239
8.1. Annex 1- Pathway prospecting for α -methylglutamic acid (MGU) biosynthesis	240
8.2. Annex 2 - Supplementary tables of the LA pathway prospecting work	242
8.3. Annex 3 – Primers used in this work	251
8.4. Annex 4 – Analysis of LA pathway expression and ScHEM1 expression	253
8.5. Annex 5 – ITA transcriptomic analysis.....	255

Table Index

Table 1.1: Main CA building blocks (list of molecules identified in [7], [9], [1], [15], [25], [26]). Alternative microbe-based production methods developed either at the industrial or research scale are listed only when the main industrial production method is not fermentation.....	14
Table 1.2: ITA production strategies in hosts other than <i>Aspergilli sp</i> and <i>Ustilago sp</i>	29
Table 1.3: Overview of most known <i>new-to-nature</i> pathway prospecting tools. Many tools don't perform ranking of the predicted pathways (indicated by N/A); however, this may be done afterwards, as is the case for many BNICE applications. The reaction prediction field details how a reaction is predicted to be catalyzed by a generic reaction rule or a specific enzyme. This information is not known for Sympheny, depicted by ?.	47
Table 3.1. Predicted new-to-nature pathways for LA biosynthesis using selected precursors. enzymes to the totaly of the steps are represented in red and yellow, respectively. Pathways to which it was possible to assign specific enzymes to all the reactions are depicted in green and are graphically represented in Figure 3.4. Pathways to which it was not posible to assign E.C. sub(sub)classes or specific enzymes to the totaly of the steps are depicted in red and yellow, respectively. Pathways to which it was possible to assign specific enzymes to all the reactions are depicted in green and are graphically represented in Figure 3.4.	79
Table 3.2. Advantages, disadvanatges and mitigation approaches for each LA pathway found in this work.	90
Table 4.1. List of strains used in chapter 4.	105
Table 4.2. List of plasmids used in chapter 4.	106
Table 5.1. Vitamins, aminoacids and trace elements used to supplement the MMF media	155
Table 5.2. List of strains used in Chapter 5.	155
Table 5.3. List of plasmids used in Chapter 5.....	156
Table 5.4. Predicted conversions in the MINE database where MAA is a product.	160

Table 5.5. Molecules found to be structurally similar to MAA in KEGG using the SIMCOMP tool. The Tanimoto coefficient threshold was 0.65.....	162
Table 5.6. Predicted conversions in the Atlas of Biochemistry database where Methacrylyl-CoA is the product. Only conversions with negative Gibbs free energy were considered. Conversions that had been already predicted in MINE were disregarded (HEMA, MMA and chronoacyl). MTHF: 5,10-methylenetetrahydrofolate; THMPT: Methylene-tetrahydromethanopterin.	164
Table 5.7. Precursor to MAA conversions in TRANSFORM-MINER with maximal 3 steps and minimal similarity score of 0.6. Only one route was selected for each precursor, being the first criteria the shortest path and the second criteria the sum of the similarity score of all the pathway reactions	169
Table 5.8. Precursor to Methacrylyl-CoA conversions and Methacrylyl-CoA to MAA conversion in TRANSFOR-MINER with maximal 3 steps and minimal similarity score of 0.6. Only one route was selected for each precursor, being the first criteria the shortest path and the second criteria the sum of the similarity score of all the pathway reactions.	172
Table 6.1. List of strains used in chapter 6.	194
Table 6.2. Top 20 ITA-inducible genes found in this work. The genes are sorted in descending order of the activation slope.	200
Table 6.3. TF regulations enriched in the top 20 genes activated by ITA in the lag phase dataset. Data obtained using the TF rank function of the Yeastract database [25]. It is shown the enrichment of the TF in this dataset in comparison to the enrichment of the TF in the overall Yeastract database.....	203
Table 6.4. Top 20 ITA-repressible genes found in this work. RSQ: Square of the Pearson Product-Moment Correlation Coefficient. The genes are sorted in ascending order of the log activation slope	206
Table 6.5. TF regulations enriched in the top 20 genes repressed by ITA in the lag phase dataset. Data obtained using the TF rank function of the Yeastract database [25]	208
Table 6.6. TF regulations enriched in the top 12 list of ITA-inducible genes absed on lag and exponential phase analysis. Data obtained using the TF rank function of the Yeastract database [30]. 8 out 11 TFs were also enriched in top 20 ITA-inducible gene list from the lag phase (last column).	211

Table 6.7. List of genes upregulated in all three conditions: the lag phase (175 mM and 290 mM ITA) and in the exponential phase (290 mM ITA). The genes were manually grouped by functional category.	218
Table 8.1. Reactions inserted in the iJO1366 and Yeast 5.0 models to simulate the heterologous pathways.	242
Table 8.2. Candidate precursors selected from the KEGG database according to the structural similarity search.....	243
Table 8.3. LA forming reactions predicted in MINE. The precursors were selected according to the ability to produce them in a microbial cell.	244
Table 8.4. LA forming pathways predicted in ReactPred and considered after the filtering criteria. The original 26 pathways were fused in 11 pathways, since many pathways were repeated in the results. A manual check of every pathway was required to remove predictions that included unrealistic reactions, unstable molecules or undesirable precursors . Reactions were considered unrealistic when they were too complex to occur enzymatically (ex: simultaneous deamination, isomerization and methylation).....	245
Table 8.5. Assigned enzymatic reactions. The desired synthetic reaction is depicted along with the main native reaction. The Tanimoto coefficient between the native and synthetic substrate was calculated with ChemMine Similarity workbench	247
Table 8.6. Additional information on the enzymatic steps comprised in the selected pathways	249
Table 8.7. Flux Coupling Analysis results for each pathway/organism combination. The reactions selected by F2C2 considered to be directionally coupled to the LA production reaction are listed. In all cases the reactions belonging to the heterologous pathways were directionally coupled to LA, as expected, and are not listed here. In all cases the reactions belonging to the heterologous pathways were directionally coupled to LA, as expected, and are not listed here.	250
Table 8.8. Primers used in Chapter 4	251
Table 8.9. Primers used in Chapter 6	252
Table 8.10. Genes upregulated in the lag phase in both ITA concentrations. Genes with a positive slope of activation are highlighted in grey.....	255
Table 8.11. Genes downregulated in the lag phase in both ITA concentrations	258
Table 8.12. Genes upregulated in the exponential phase with 290 mM ITA	261

Figure Index

Figure 2.1. Main concepts addressed in this thesis. The yeast platform for CA production requires an expansion of the molecules that can be produced and the development of high-throughput screening methodologies.	3
Figure 2.2. Thesis Outline. Chapters 3 to 5 are dedicated to the establishment of new production processes in yeast for LA and MAA, while chapter 6 is focused on the improvement of an already established yeast production of ITA.	5
Figure 1.1: Envisioned chemical conversions of ITA to industrially interesting derivatives and their corresponding main applications.	20
Figure 1.2. Production processes for MAA and MMA established in the industry. A) Conventional ACH route with the use of excess sulfuric acid established by Rohm & ICI; B) New ACH route established by Mitsubishi; C) Alpha process established by Lucite.	23
Figure 1.3. Envisioned chemical conversions of LA to industrially interesting derivatives and their corresponding main applications. The derivatives which are predicted to displace fuel-derived chemicals are depicted with a red arrow.	25
Figure 1.4. Metabolic and genetic elements in biosynthesis of ITA in <i>A. terreus</i> and <i>U. maydis</i> . A) ITA biosynthesis pathway in both organisms. The cypC-mediated conversion of ITA in HP-ITA and the Itp1-mediated export of HP-ITA are hypothetical and are depicted with (?). B) Gene clusters in both organisms. The genes encoding proteins represented in panel A are depicted in the same colors; the attributed function to each protein is described in panel B.	28
Figure 1.5. Published biobased routes to MAA production. Chemical catalysis and fermentation processes are depicted in blue and green squares, respectively. A) Lucite route through citric acid. B) Route through citric acid from [92]. C) Route through ITA from [93], [94]. D) Ascenix Technologies mixed fermentative/chemical process through syngas. E) Exclusively fermentative route for MAA production, patented by Genomatica. F) Bioconversion of isobutyric to MAA, patented by Lucite.	33
Figure 1.6. Published metabolic routes leading to LA. A) Envisioned route from pyruvic acid, patented in [84], with only two of five potential steps validated <i>in vitro</i> for specific enzymes (depicted in red). B) Route established in <i>E. coli</i> , starting from succinyl-coA and acetyl-coA; the overexpressed enzymes are depicted in red. [85]	36

Figure 1.7. Impact of biorefinery product diversification in the bioeconomy. Low value biomass is treated in a biorefinery to be converted in added value molecules which are further transformed in consumer goods, thus increasing the value. By diversifying the biochemical conversion routes in the biorefinery, a larger panoply of added value products can be obtained, enabling more types of consumer goods to be obtained from biomass-derived processes.38

Figure 1.8. Approaches to create *new-to-nature* pathways, based on [116]. Approaches that use natural enzymes can be based on mixing & matching natural reactions from different organisms (B) or can take advantage of natural enzymatic promiscuity by filling gaps in the host endogenous metabolism (A) or by supplying synthetic substrates to the natural pathway (C). Approaches where enzymes are engineered to accept new substrates can rely on the use of synthetic precursors (D) or endogenous ones (E).....41

Figure 1.9. Two examples where enzymes were engineered to accept new substrates. A) In [122] aspartate kinase was modified by rational engineering and combinatorial screening to accept malate as a substrate. B) In [123] the benzaldehyde lyase enzyme was modified to accept formaldehyde as the substrate and catalyze two consecutive rounds of carbon-carbon coupling. This was possible through 4 iterations of rational protein design and screening.42

Figure 1.10. Example of how reported enzymatic reactions contribute to the creation of reaction rule (here in BEM format, example from BNICE [130].), which enables the prediction of enzymatic reactions. The reported enzymatic reactions are of glutamine and phenylpyruvate transamination to glutamate and phenylalanine and of aspartate and 2-oxoglutarate transamination to oxaloacetate and glutamate. In the predicted enzymatic reaction pyruvate and 7,8-diaminononanoate are transaminated to alanine and 7-keto-8-aminononanoate.....44

Figure 1.11: Typical procedure followed by a synthetic pathway prospecting tool. In the first stage enzymatic data from various databases is compiled to create reaction rules (in this image an example from ReactPred reaction rules in SMART format). In the second stage the application of the reaction rules enables the search for possible pathways, which are ranked in the third stage.46

Figure 3.1. Approaches used in this study to identify “new-to-nature” pathways to implement microbial production of LA. The strategy is divided in three stages: search for potential precursors (based on compound similarity and retrosynthesis), enzymatic assignment to predicted reactions, and pathway evaluation.....72

Figure 3.2. Filtering criteria used to prune the list of possible pathways leading to LA unveiled by ReactPRED. The more than 50,000 pathways were filtered to a final 2 possible pathways.76

Figure 3.3. List of possible precursors for LA biosynthesis, based on the search for chemically similar molecules or from retrobiosynthesis tools. The search method that yielded each precursor is depicted in light color circles: 1) similarity search in KEGG (blue), 2) Retrobiosynthesis in MINE (orange), 3) Retrobiosynthesis in ReactPRED (green). The precursors in a red font were discarded and the ones in black font were selected for enzyme assignment.77

Figure 3.4. Selected pathways leading to LA production. The pathway letter is shown on top, where Pathway A has two variations (A1 and A2), depending on the KAPA transaminase that catalyses the first step. The native enzyme pathways leading to the precursor are depicted in black block arrows. Heterologous reactions that occur naturally in vivo are depicted in thin green arrows, thin red arrows represent synthetic reactions that were assigned an enzyme in this work based on enzyme promiscuity and thin purple arrows represent reactions that were assigned based on an enzyme that requires further engineering. In the case of the lysine degradation pathway, it is native to *S. cerevisiae* (black arrows) but not *E. coli* (yellow thin arrows). D-ALA (Delta-aminolevulinic acid), 4,5-DAVA (4,5-Diaminovaleric acid), Glut-1-Semiald (Glutamate-1-semialdehyde), 3,2-AMP (3-amino-2-methyl-propionic acid), Pyr (Pyruvic acid), Lys (Lysine), Orn (Ornithine), AceAce (Acetoacetic acid), OAD (3-Oxoadipic acid).83

Figure 3.5. Number of heterologous and synthetic reactions for each pathway. Considered for the import of each pathway in *E. coli* or in *S. cerevisiae*.85

Figure 3.6. Evaluation of the selected pathways A) Maximal theoretical yields of each pathway in the metabolic networks of *E. coli* and *S. cerevisiae*. B) Production envelopes for each pathway in *E. coli* and in *S. cerevisiae*.87

Figure 3.7. Outline of the steps taken in this work to select the 5 new LA routes. ..93

Figure 4.1. LA biosynthetic pathway explored in this chapter. The synthetic pathway is depicted in the top, while the native steps catalyzed by the selected enzymes are depicted in the colored boxes.102

Figure 4.2. C4 and C5 pathway for D-ALA biosynthesis. The steps depicted below the D-ALA box represent the tetrapyrrole biosynthetic pathway, which is common to all organisms.104

Figure 4.3. SDS-PAGE with 5 μ L of crude cell extracts of BL21 expressing pETDuet,

pETDuet-bioA, PETDuet-bioK and PETDuet-ygeX. The expected sizes of the proteins are: bioA 47 kDA, bioK 50 kDA and ygeX 43 kDA.115

Figure 4.4. Progression of absorbance at 335 nm and 420 nm during KANA transaminase enzymatic assays. A) Two-step mechanism of KANA transaminase, with the depiction of PLP and PMP-enzyme forms and their maximum absorption wavelengths. B-D) Ratio of absorbance Abs_{420nm}/Abs_{335nm} . E-G) Percentage ratio Abs_{420nm}/Abs_{335nm} : calculated by dividing the ratio at all time points with the initial ratio.117

Figure 4.5. LC-MS analysis of samples of KANA transaminase assays with KAPA as the substrate. A) Sample from assay with pDuet extract and KANA, taken after 1 h of reaction. B) Sample from assay with bioK extract and KANA, taken after 45 min of reaction. For each sample it is depicted the total ion chromatogram (TIC) and the extracted ion chromatograms (EICs) for caffeine (internal standard), KANA, DANA, PLP and lysine. Note the different scales for each compound.118

Figure 4.6. LC-MS analysis of samples of KANA transaminase assays with D-ALA as the substrate. A) Sample from assay with pDuet extract and D-ALA, taken after 1 h of reaction. B) Sample from assay with bioK extract and D-ALA, taken after 45 min of reaction. For each sample it is depicted the total ion chromatogram (TIC) and the extracted ion chromatograms (EICs) for caffeine (internal standard), D-ALA, DAVA, PLP and lysine. Note the different scales for each compound.119

Figure 4.7. DAPA ammonia lyase enzymatic assays with crude cell extracts prepared from BL21(DE3) cells carrying pDuet and pD-ygeX. A) Ratio of absorbance Abs_{423nm}/Abs_{412nm} ; inset depicts initial 10 minutes of reaction in larger magnification. B) Pyruvic acid concentration.120

Figure 4.8. Stack depiction of HPLC traces of supernatants from double enzymatic assays carried out with 3 combinations of crude cell extracts: pDuet, PD-bioA+pD-ygeX and pD-bioK+pD-ygeX. A) Samples were taken at the initial time and after 1h. B) Samples were taken at 1h and 24h. LA retention time was 15 min.121

Figure 4.9. Tentative *in vivo* assembly of LA biosynthetic pathway in *E. coli*. A) Outline of the strategy used to couple D-ALA overproduction with the pathway enzyme in two sets of compatible plasmids. B) Extracellular D-ALA concentration. C) HPLC trace of supernatants after 24h and 48h of fermentation with strains 0, 5 and 6. LA retention time in fermentation supernatants is depicted (15.2 min).124

Figure 4.10. Protein expression analysis of LA pathway *in vivo* assembly. A) SDS-

PAGE of cells harvested 24h after fermentation. B) SDS-PAGE of BL21(DE3) cells carrying the HEM1 expression plasmids. The induction was made by addition of 0.05 mM IPTG after 2 hours of culture and transfer to 30°C125

Figure 4.11. Two-step strategy for in vivo LA production based on bioconversion with growing cells. A) Experimental setup used. B) HPLC traces of supernatant samples taken at 24 hours and 5 days of bioconversion.....127

Figure 4.12. Two-step strategy for in vivo LA production based on bioconversion with whole resting cells. A) Experimental setup used. B) D-ALA extracellular concentration at 24h and 48h of bioconversion. C) HPLC traces of supernatant samples taken at 24 and 48 hours of bioconversion.128

Figure 4.13. D-ALA cultivations with BY4741 transformed with pGreg586 (empty circles) or pGREG-HEM1 (black circles). The extracellular concentrations of D-ALA (A) and succinic acid (B) are represented.130

Figure 4.14. Engineering of the H2 valve strain, where both tTA and HEM2 expression is controlled by tetO elements in the promoters. Dox addition prevents tTA recruitment to the tetO operator sites, repressing expression of tTa and HEM2. Design and figure based on [22].131

Figure 4.15. D-ALA cultivations with H2 strain and W303 strain (control) in MMB glucose. A pulse of Dox was added at the exponential phase (blue arrow, 8h). The OD_{600nm} (A) and extracellular D-ALA concentrations (B) are represented.132

Figure 4.16. D-ALA cultivations with H2 strain and W303 strain (control) carrying the pGREG-HEM1 plasmid in MMB galactose. A pulse of Dox was added at the exponential phase (23.5h for W303 and 27h for H2, blue arrows). The OD_{600nm} (A) and extracellular D-ALA concentrations (B) are represented. In C) it is represented the D-ALA concentrations normalized to OD_{600nm} of the culture (C).....133

Figure 4.17. Microplate D-ALA cultivations with H2 strain and W303 strain (control) carrying the pGREG-HEM1 plasmid in MMB galactose. A pulse of Dox was added at initial time of cultivation, at the exponential phase or from the inoculum. The OD_{600nm} (A) and extracellular D-ALA concentrations (B) are represented. In C) it is represented the D-ALA concentrations normalized to OD_{600nm} of the culture (C).....134

Figure 4.18. Analysis of HEM2 expression of W303 and H2 cells cultivated in MMB glucose during exponential and stationary phase. A) Cultures where no Dox pulse was added; the reference is the W303 cultures. B) H2 cultures where a Dox pulse was added at the exponential phase (250, 500 or 1000 ng/mL); the H2 culture with no Dox pulse is the

reference.135

Figure 4.19. Microplate fermentations with W303 pHem1 and H2 pHem1 in MMB raffinose + galactose, with the Dox pulse added after 24. OD600 after 48h (A), extracellular D-ALA concentrations after 48h (B) and D-ALA concentrations normalized to OD600 of the culture (C).....136

Figure 5.1 Candidate MAA biosynthetic pathways that have been published. A) The 8 MAA forming pathways published in [15], [16] are depicted. The 6 MAA and Methacrylyl-CoA precursors are depicted in circles. The reaction steps of each pathway that have been reported to be catalyzed by a specific enzyme are represented in black with the corresponding enzyme E.C. number. The enzymatic steps depicted in blue are hypothesized to be catalyzed by an enzyme belonging to E.C. sub (sub)class and the enzymatic steps depicted in orange have similar reactions catalyzed the enzymes represented by the orange EC numbers. Both blue and orange steps don't have experimental validation. The natural valine degradation pathway where MAA is an intermediate is pinpointed with green dashes. B) The bioconversion of isobutyric acid to MAA in *E. coli* has been reported in [17]. The heterologous enzymes expressed to assemble the pathway are annotated.....153

Figure 5.2. List of possible MAA, Methacrylyl-CoA and MMA precursors selected in this work. The Methacrylyl-CoA and MMA conversions were searched in the Atlas of Biochemistry database because these are KEGG metabolites. Since MAA is not present in the KEGG database, its corresponding conversions were searched in the MINE database and also structurally similar molecules were searched in KEGG.....165

Figure 5.3. MAA routes and precursors studied in this work. A) The six MAA routes found in this *in silico* work. The reactions depicted in black have experimental validation and reactions in red represent hypothetical enzymatic steps. The common microbial metabolites that are the source of each route are represented in green, which were valine, glutamate, glyoxylate and thymine. B) For each hypothetical enzymatic step (represented in red in panel A) it is described the enzymes assigned to it, the substrate for the synthetic reaction and the native substrate.....173

Figure 5.4. The MAA route studied in this work, the decarboxylation of MES(A), for which five enzymes were selected as potential candidates to catalyze the reaction (C-E). In the case of glutaconyl-coA decarboxylase (E), glutaryl-coA is also accepted as a substrate.175

Figure 5.5. Stacked HPLC traces of CAD activity assay supernatants after 10 min and

1 h of reaction with 30 mM cis-aconitic. The traces are shown for activity assays with extracts from cells carrying the empty plasmid or pGREG_AtCad1. ITA elutes at ~12 min. Aconitic acid elutes at ~7 and 9~min, corresponding to the trans and cis isomers, respectively. This is due to the spontaneous isomerization of cis-aconitic to trans in dilute aqueous solutions has been described before [52].177

Figure 5.6. Overlaid HPLC traces of CAD activity assay supernatants after 10 min and 1 h of reaction. The traces are shown for activity assays with extracts from cells carrying the empty plasmid or pGREG_AtCad1. MAA elutes at ~ 22 min (the MAA standard trace is represented) and mesaconic acid elutes at ~16 min.178

Figure 5.7. SDS-PAGE of BL21(DE3) cells carrying the pET16b-nbaD plasmid before and after 4h of induction with 1 mM IPTG and expression at 28°C.179

Figure 5.8. ACMSD activity assays. A and B) Overlaid HPLC traces of ACMSD activity assay supernatants after 1, 3 and 6 hours of reaction. Supernatants from activity assays with no substrate (A) or with 6 mM mesaconic acid (B). MAA elutes at ~22 min (red arrow) and mesaconic acid elutes at ~16 min. The unknown peak at ~9 min is highlighted in the orange box. C) Mesaconic acid concentration (mM) and area of the 9 min peak in ACMSD activity assay supernatants.180

Figure 6.1. Growth curve of By4741 cells in MMB with the supplementation of ITA (175 mM, red open circles, 290 mM black open circles) and without (black closed circles). Samples for RNA sequencing were harvested during the lag phase (1 hour) and exponential phase (5h for cultures with no ITA and 14h for cultures with 290 mM), represented in blue arrows.197

Figure 6.2. ITA-inducible genes. A) Approach to find the top 20 genes with an ITA induction: 385 genes were found to be upregulated in both ITA exposure conditions, where 295 of these had a positive slope between the fold-change at 175 mM and 290 mM of ITA. The genes with a higher dose-dependent activation were found by considering the slope of the fold change between the points with ITA=0, 175 and 290 mM. B) Dose-response curve of the top 20 ITA-inducible genes found in this work.199

Figure 6.3. Structures of ITA, the metabolite that binds and activates the Leu3 regulator (2-IPM) and the major yeast substrate of Tmt1, trans-aconitate methyltransferase (3-IPM).204

Figure 6.4. ITA-repressible genes. A) Approach to find the top 20 genes with an ITA repression: 426 genes were found to be down-regulated in both ITA exposure conditions, where 323 of these had a negative slope between the fold-change at 175 mM and 290 mM

of ITA. The genes with a higher dose-dependent repression were found by considering the slope of the log fold-change between the 0, 175 mM and 290 mM ITA conditions. B) Dose-response curve of the top 20 ITA-repressible genes found in this work.....205

Figure 6.5. Top12 ITA inducible genes found based on the analysis of lag and exponential phase. A) 43 genes were found to be upregulated in all 3 conditions and 12 of those were in the top 20 list of lag phase ITA inducible gene list. B) Fold change of the top 12 ITA inducible genes in the three conditions tested here.....210

Figure 6.6. Fold-change of 5 candidate ITA-inducible genes in the presence of 175 mM ITA. Experiments were performed in By4741 and By4741 Δ leu3. The $\Delta\Delta$ CT is normalized for the wt cultures in media without ITA (control).213

Figure 6.7. Fold-change of 3 candidate ITA-inducible genes in 175 mM and 290 mM ITA. Experiments were performed in By4741 and By4741 Δ leu3 and By4741 Δ dal81. The $\Delta\Delta$ CT is normalized for the wt cultures in media without ITA (control). Note the break in the Y axis of DAL% plot.....214

Figure 6.8. Dose-response curve of MEP2, TMT1 and DAL5 in the wt background. Data from RT-qPCR (blue) and QuantSeq (orange). The linear trendlines were calculated with slope and RSQ displayed in each chart in the corresponding color.215

Figure 6.9. Model of Leu3 regulation of leucine biosynthetic genes. The green arrows represent positive regulations during the leucine starvation response, where 2-IPM activates Leu3, which activates the transcription of the leucine biosynthetic genes. The positive interactions during ITA exposure suggested from the data in this work are represented in red arrows, where ITA activates LEU3 and results in the increase of Tmt1 transcription, either through Leu3 activation or another regulator. Tmt1 also methylates ITA to its ester.220

Figure 6.10. Allantoin catabolism genes induced by ITA. A) Allantoin and ammonia utilization pathway that is activated in nitrogen depletion conditions; genes upregulated in all ITA exposure conditions are marked in yellow. B) NCR-sensitive genes and their documented regulator, from the Yeastract database. Only TFs reported to act in the NCR and amino acid starvation response are represented here.222

Figure 7.1. Summary of the work performed during this thesis, with the main findings and proposed future work.237

Figure 8.1. MGU forming pathways found in this work.240

Figure 8.2. SDS-PAGE of BL21(DE3) cells carrying pColA-ScHEM1 and pD-ygeX

(top) or pColA-ScHEM1 and pD-bioK-ygeX (bottom). Induction was carried either 2h or 5h after culture start and proceeded at 25°C, 30°C or 37°C. The two lanes matching each temperature correspond to induction with 0.1 and 0.5 mM of IPTG. The bands corresponding to ygeX and bioK molecular weight are depicted in green (43 kDa) and yellow (50 kDa), respectively.....253

Figure 8.3. SDS-PAGE of cells harvested at 24h after induction with 0.1 mM IPTG at 30°C (induction after 5h of culture start). The bands corresponding to ygeX and bioA/bioK molecular weight are depicted in green (43 kDa) and yellow (47/50 kDa), respectively.....254

Figure 8.4. SDS-PAGE of cells carrying the pColA-ScHEM1 plasmid. The cells were harvested at 24h after induction with the LA pathway protocol (induction at 5h after culture start with 0.1 mM IPTG at 30°C).....254

Figure 8.5. Hierarchical clustering tree of the enriched functional categories in the list of 43 genes upregulated in the presence of ITA (in the three conditions studied here). Analysis was performed in ShinyGO.....262

Figure 8.6. Functional interaction network generated by STRING on the 43 upregulated genes. Only interactions based on experiments, databases and co-expression were considered. The line thickness of each interaction indicates the strength of data support.263

Chapter 1 – Introduction to the theme of the thesis

This thesis has Yeasts at its center by exploring them as potential hosts for production of value-added carboxylic acids (CA). *Saccharomyces cerevisiae* has been well explored as an industrial producer of the CAs succinic and lactic and also investigated for the potential production of other CAs. This results not only from the strong track record of utilization in other industrial processes (like production of fermented beverages or foods or production of ethanol) but also from its extreme genetic tractability that allows a fast engineering of strains. *S. cerevisiae* and, in fact, Yeasts in general, possess other traits that render them more attractive as hosts for the specific production of CAs of which the ability to produce these molecules at acidic pHs is the more relevant one since it addresses one of the more significant constraints of bio-based production of CAs which is cost. This work is largely focused on the development of Yeast-based synthetic biology, enabling the exploration of *S. cerevisiae* for the production of levulinic acid (LA) and itaconic acid (ITA), as well as of methacrylic acid (MAA), one of the most important derivative products of itaconic acid.

Although Yeasts are known natural producers of some CAs (like citric or succinic acids, for example) they are unable to produce any of the three CAs focused in this work, ITA, MAA and LA. However, the underlying biosynthetic pathways can be ported for the selected Yeast hosts and this was performed with success already in the case of *S. cerevisiae* which has been successfully engineered (including in the iBB-IST laboratory) to over-produce ITA. Concerning ITA production, this thesis had one main goal: promote the design of a metabolite-based biosensor that could be used to rapidly identify Yeast strains (either based on *S. cerevisiae* or not) over-producing ITA. Considering the extreme genetic tractability of Yeasts, the use of extensive genomic engineering to obtain large amounts of strains with different genomic repertoires is reasonably easier to obtain. In the case of *S. cerevisiae* one can easily think, for example, in screening the performance of the 5000 mutants devoid of all non-essential genes, the so-called disruptome (which can identify mutations that can result in improved ITA production by altering functions that intuitively cannot be linked with this process). However, screening of the capabilities of ITA production in a high number of strains is very difficult since ITA is an inconspicuous molecule and its quantification requires laborious fermentations and analytics based on 30 min HPLC runs. Thus, this work addressed two challenges in the exploitation of the Yeast platform for CA production: how to expand the molecule repertoire that can be produced by this host's metabolism and how to design new high-throughput screening tools for CA production (**Error! Reference source not found.**).

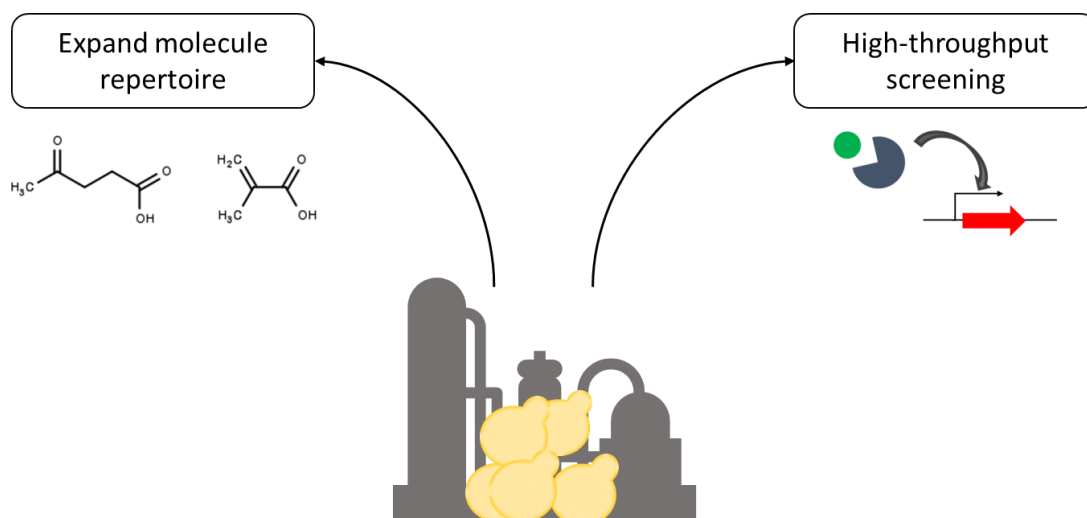


Figure 1.1. Main concepts addressed in this thesis. The yeast platform for CA production requires an expansion of the molecules that can be produced and the development of high-throughput screening methodologies.

1.1. Thesis Outline

This thesis objective is to improve the yeast CA production platform, by investigating two new potential molecules of industrial interest that can be produced in this host (LA and MAA) and by improving the already established yeast-based production of ITA. The thesis is divided in seven chapters. Chapter 1 provided a state-of-the-art review in bio-based production of CAs, with an emphasis on the exploration of Yeast as a platform host. The use of synthetic biology tools to implement new-to-nature metabolic pathways is also reviewed in Chapter 1. The main research questions and approaches used in chapters 3-6 are briefly summarized below:

Chapter 3: Which biochemical pathways could be used to implement microbial production of LA in industrially relevant hosts (including *S. cerevisiae*)?

In this chapter the results of a pipeline developed to prospect potential pathways leading to LA production, with the aid of prospecting and retrobiosynthesis tools, are described. A discussion concerning the ranking of the pathways obtained *in silico*, the performance of those considered more likely to be implemented *in vivo* based on metabolic modeling and aspects that might impact their assembly with success in *E. coli*

and *S. cerevisiae* are discussed.

Chapter 4: Can one of the *in silico* identified LA-production pathway result in LA production *in vivo* and/or *in vitro* and which strategies can be used to improve its performance?

In this chapter are described the efforts that had been undertaken to demonstrate if one of the pathways identified in Chapter 3, using D-ALA as a precursor for LA synthesis, does result in production of LA. The efforts undertaken to improve the pool of D-ALA in *S. cerevisiae*, which was suggested as a putative constraint for the success of the pathway in the metabolic modeling performed, are also described.

Chapter 5: Which biochemical pathways can be used to implement microbe-based production of MAA?

In this chapter a methodology similar to the one described in Chapter 3 was used to prospect possible pathways for the biosynthesis of MAA, also being described the efforts that were undertaken to confirm if one of the candidate pathways does produce MAA.

Chapter 6: Which genetic parts could be used with success to design and build an ITA sensor in yeast?

This chapter describes the experimental approach that was used to identify the genetic parts of an ITA biosensor based on the identification of a responsive transcription factor and its binding site.

Chapter 7: Final discussion of the results obtained in this thesis and future work.

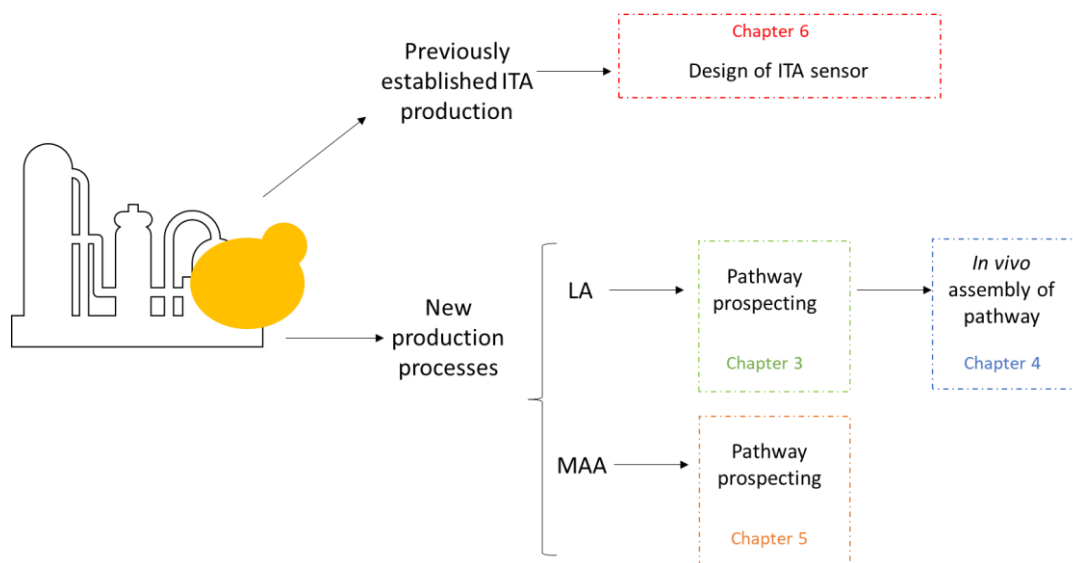


Figure 1.2. Thesis Outline. Chapters 3 to 5 are dedicated to the establishment of new production processes in yeast for LA and MAA, while chapter 6 is focused on the improvement of an already established yeast production of ITA.

During this thesis the following methodologies and model systems were used:

- Organisms: *E. coli* was used as a model organism for the overexpression of enzymes and assembly of a LA biosynthetic pathway, *S. cerevisiae* was used as potential hosts to produce ITA. Furthermore, *S. cerevisiae* was used to study the regulatory response to ITA in order to design an ITA sensor.
- Pathway prospecting methodologies: Promiscuous enzymatic activity prediction databases (Atlas of Biochemistry and MINEs) and a retrobiosynthesis tool (ReactPRED) were utilized to predict LA producing pathways. The BRENDA database was used to assign enzymes to desired synthetic enzymatic reactions.
- Genome-scale metabolic modeling: The COBRA toolbox and the genome-scale models of yeast and *E. coli* were used.
- Assessing promiscuous enzymatic activity: crude cell extracts of *E. coli* cells overexpressing the required enzymes were used to perform enzymatic assays.
- Assessing a biosynthetic pathway: *E. coli* was engineered to express the pathway enzymes and overproduce the pathway precursor
- A tetracyclin-based downregulation strategy was implemented to dynamically regulate production in yeast

Chapter 1– Introduction to the theme of the thesis

- A quant-seq methodology was used to unravel yeast transcriptional responses to ITA and identify potential parts to be used in the design of an ITA sensor
- General quantification methods used: High-Performance Liquid Chromatography (HPLC), Liquid-Chromatography Mass Spectrometry (LC-MS) and modified Ehrlich method
- General genetic engineering methods: cloning by restriction-ligations, Circular-Polymerase Extension Cloning (CPEC) and Gibson assembly and yeast genome integration by homologous recombination

Chapter 2 – State of the Art

Part of the review described in this chapter are part of a paper that is being prepared for submission:

Ana Vila Santa, Frederico Ferreira, Kristala Prather and Nuno P Mira, Exploring synthetic biology to bridge the gap between microbe-based production of carboxylic acids and of their “new-to-nature” derivatives: a case study focused on itaconic acid and its derivatives metacrylic acid (MAA) and methyl-metacrylic acid (MMA) (in preparation for Submission to Frontiers in Bioengineering)

2.1. Overview

The development of the consumer goods industry in the last centuries has allowed humans to have access to a large list of products that improve their comfort (ranging from shampoos, plastic bottles, colorful clothes to cars and plane trips) and has shaped modern society. However, it is becoming increasingly clear that the way these goods are manufactured and the rate at which they are sold is not environmentally sustainable and has to be changed. The problems come not only from the rapid exhaustion of natural resources used in the pathways explored to produce these goods (e.g., in agriculture), but also from problems that are related with the industrial implementation itself since many of these goods are produced in synthetic routes that are largely dependent of precursor molecules (also known as commodity chemicals) obtained from petrochemistry. A good example of a “problematic” commodity chemical is acrylic acid, a molecule that is produced in high volume and low price from propylene (derived from gasoline and ethylene) and is sold to various industries to be used in the production of adhesives, coatings and resins [1]. The need to shift the current economy from oil to environmentally friendlier options has also been recognized in other contexts, out of which, the reduction in the greenhouse effect caused by excessive accumulation of CO₂ in the atmosphere is the better known one. However, it is clear that this reduction of petrochemistry will only occur if alternatives to the use of precursors obtained from oil cracking are provided, either through the identification of other molecules with the same potential to be used as precursors in chemical synthesis or through the identification of “greener” routes for the production of the precursors used today.

A potential solution that has been studied to reduce the dependency from oil of commodity chemicals’ production is the exploration of biomass fermentations in the so-called bio-refineries. These are facilities that function in a circular bioeconomy perspective making use of biomass as ingoing material (e.g. food remains, agricultural residues or industrial waste) to produce value-added products [2]. Although initially the main goal of biorefineries has been the production of biofuels (out of which ethanol production stands out as the more relevant one), it soon became clear that this would not suffice to assure economical viability to these facilities nor it would solve the problem of identifying alternatives to the building block molecules provided by petrochemical industry [3]. Consequently, a more integrated perspective on biorefineries has been envisaged, where the raw material is used to produce biofuels but also solvents, building blocks or polymers, among others [4]. This conversion of biomass in the envisaged

products may occur through “chemically friendly” conversions (that is conversions similar to those currently used but exploring routes/molecules that have lower environmental impact) but also biological conversions or combinations of both. Biological conversions in this context concern the exploration of microbes as cellular factories to produce the compounds of interest this being fostered by the significant advances in the biology field that have enabled the development of fast and effective strategies for microbial engineering. These advances include the exponential availability of sequenced genomes, the development of genetic engineering methodologies that enable overexpression of enzymes and/or entire metabolic pathways in heterologous hosts, as well as good methods for optimization of the process, frequently guided by computational biology approaches. Despite this, although biorefineries appear as promising solutions to foster circular bio-economy and improve environmental sustainability of the modern industry, their implementation is being complicated by multiple factors. Among the more relevant barriers are the high costs associated with the production, which necessarily causes the price of the products not being competitive compared to their-fuel derived counterparts. The yield of products obtained using biomass as raw material is frequently low, most often due to a high toxicity over fermenting cells of the raw material itself or of the products that are obtained during pre-treatment of the biomass material [6]. The different availabilities of biomass products available across the Globe will also render very difficult to think in a universal biorefinery solution [7]. Finally, the fact that many of the molecules of higher economic value as building blocks correspond to “new-to-nature” molecules, that is, molecules that are not known to be produced by living systems, is another problem. Altogether the challenges identified point to the need of intensifying research focused on biorefineries, with emphasis on the microbial hosts that could be “engineered” to produce the envisaged products at higher titers (e.g., by overcoming the toxicity aspects) or using more accessible raw materials to produce new products with higher economic value. When addressing this challenge, the recently booming field of synthetic biology is particularly relevant since it has been proven successful in engineering microbes to create fully new-to-nature bioprocesses or improve old ones. A paradigmatic case is the one of artemisin, an anti-malarial drug whose biosynthetic pathway native of the plant *Artemisia annua* has fully been assembled in *Saccharomyces cerevisiae*, leading to the successful fermentative production of the drug precursor at industrially relevant titers[9]. The establishment of yeast as an artemisinic acid producer enabled the scale up of the process to enter commercial production, avoiding the market volatility that characterized plant extract-derived

artemisin. This study, along with others that followed it, are a clear demonstration of the power of synthetic biology in enabling new processes overcoming the identified “*lack of broad-based conversion technology coupled with a plethora of potential targets*” [8] linked with biorefineries and taking a step further into the necessary environmental sustainability that industrial (bio)-chemical industry needs to take.

This thesis is embedded in this spirit of using microbial cells for the production of three carboxylic acids, itaconic acid (IA), levulinic acid (LA) and methacrylic acid (MAA), that have a recognized high potential market value by being able to replace a number of building blocks explored today by the industry and that derive from oil. This subsection of this chapter reviews the panorama of industrial bio-based production of carboxylic acids and, in particular, of the 3 that are at the center of this work. In subsection 2.2 the main production strategies used today to obtain MAA, ITA and IA will be reviewed while in subsection 2.3 it is described the usefulness of synthetic biology in the implementation of pathways to produce “new-to-nature” molecules and in designing high-throughput screening methodologies to improve production titers in microbial systems.

2.1.1. Main CA building blocks identified in the literature

CAs are regarded as important building blocks for synthesis due to the functional groups (often more than one) that are frequently found in these molecules which include, besides the carboxylic group, keto or hydroxyl groups, that can be used to enable, for example, for the formation of polyesters. Dicarboxylic acids can be condensed, for example, for the synthesis of polyamides [12]. The industrial interest of CAs is reflected in their global market, valued at 19,91 Billion USD in 2018 with a Compound Annual Growth Rate (CAGR) of 8%. The current market application segments for CAs is the food and beverages sector, however, applications are also envisioned in animal feed, chemical industry, pharmaceuticals, personal care and agriculture [13]. The increasing preference of customers for bio-based production processes together with the increasing environmental regulation imposed to industrial processes are the main drives that support the development of microbe-based production of CAs. On the other hand, the main constraints concern the use of substrates that might be used as food (e.g. crops or derivatives) and the still high production costs [14].

In this section the previously identified CAs with industrial interest and potential to

be produced in a biorefinery will be reviewed based on literature surveys published [8], [10], [15]. A general overview on the information gathered about the identified CAs is summarized in **Table 2.1**, including the most relevant derivative molecules that can be obtained from each CA, as well the markets of application of each CA and the production method currently used to obtain the CAs in use.

The initial report published by the US DoE identified three C4 diacids as more promising: succinic, malic and fumaric acid [10]. Because these three CAs can be interconverted succinic, malic and fumaric are routinely grouped together and can also be used in the same derivative streams [16]. Succinic acid is currently obtained from maleic anhydride, a fossil derivative, although bio-based production of succinic acid is also implemented. In fact, production of succinic acid has been the rising star of the bio-based production of CAs attracting a lot of investment. However, this commercialization has been losing momentum, with now only one out of the original four companies still in the market (Succinity). One main challenge for the establishment of “bio-succinic acid” in the market has been the low oil prices in the last years, making it difficult to compete with the fuel-derived route [18]. An important advantage that has been associated to bio-based production of succinic acid is the expansion of possible market applications to include the production of γ -butyrolactone (GBL), tetrahydrofuran (THF) or polyesters [17]. Fumaric acid is used mostly as an acidulant in the food and feed industries [19]. Although the initial process of obtaining fumaric acid was through fermentation using the fungus *Rhizopus arrhizus*, later on this process was replaced to a petrochemical route due to the low cost of fuels [19]. In the case of malic acid, the largest applications are also in the food industry which accounts for 85-90% of its potential market [16]. However, it also has the potential to displace maleic anhydride and both malic acid and its derivatives (such as 2-pyrrolidone, N-methyl-2-pyrrolidone or succinonitrile) can be applied in the polymer industry as monomers or solvents [16]. 2,5-furan dicarboxylic acid (FDCA) was another dicarboxylic acid identified in the US DoE report, although in this case the implementation of microbe-based production of this CA was not as obvious of the other 3 mentioned above which are direct intermediates of central metabolism. The main application of FDCA are green polymers with the potential to displace the fuel-dependent polyethylene terephthalate. One of the bioprocesses for production of FDCA that is attracting investment nowadays involves the chemical oxidation of 5-hydroxymethylfural (HMF), a process from lignocellulosic biomass [20]. 3-Hydroxypropionic (3-HPO) is another CA with the potential to become a bulk chemical, due to its derivative routes leading to acrylates and 1,3-propanediol [10]. However, no large-scale production of this

chemical is implemented due to the high cost of the chemical routes. This has increased the interest in its microbial production from glycerol (a byproduct of biodiesel) or directly from sugars [21]. Aspartic acid, a component used in the synthesis of the artificial sweetener aspartame, is mainly obtained industrially with the enzymatic amination of fumarate, using immobilized cells containing aspartase [22]. Glucaric acid production has been traditionally performed through the chemical oxidation route of glucose but a microbial synthesis route has been shown [23]. Glutamic acid, the main amino acid responsible for the *umami* flavor and with a major use as a flavor enhancer, has been traditionally obtained in the industry through the fermentative route using *Corynebacterium glutamicum* and research efforts have been focused on understanding the mechanisms of production, strain optimization and usage of diverse raw materials [24]. ITA and LA are both C5 carboxylic acids with various derivatives routes, with ITA industrial production being fermentative and LA production being chemical. Further details on the relevance of these two CAs as well as a discussion on the industrial production processes, markets and applications of these two compounds are further explored in sections 2.2.1 and 2.2.3.

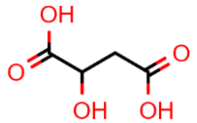
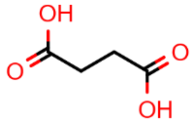
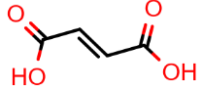
One relevant note about the promise of building block CAs is that they may serve as direct substitutes of petrochemicals in the industry (e.g malic acid can replace maleic anhydride) or may lead to the establishment of new derivatives routes leading to new market applications. This is the case of FDCA, which may create new green biopolymers to replace old fuel-derived threphthalates [20]. It is also relevant the fact that various building block molecules are linked by their derivative routes. As an example, acrylic is considered a building block by itself, with potential to be obtained from fermentation. However, this molecule and methyl-acrylic acid may also be obtained from the conversion of 3-HPO or ITA [10].

Following the publication of the DOE report with the top 15 bio-based building blocks that ignited the research efforts in the field, the list was revisited in 2010 to assess the progress in the production processes and commercial success and come up with a new “top chemical opportunities list” [8]. A significant alteration comparing with the previous list was the removal of aspartic acid, glutamic acid, ITA, malic acid, glucaric and fumaric acid, indicating that effervescent research ongoing in these compounds had yet to translate in industrial development. In the case of glutamic acid, the promise of derivative routes has not been observed in the industry, where it continues to be used as an end product instead of a platform. Lactic acid was inserted as a new potential platform molecule which can undergo esterification, catalytic reduction or dehydration routes, besides its primary

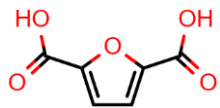
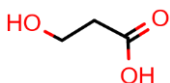
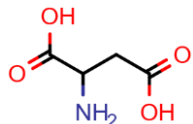
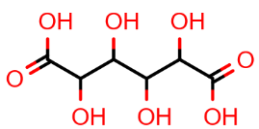
utilization in production of poly-lactic acid. Its potential to be produced commercially through the biological route is already achieved in the companies NatureWorks (USA), Purac (The Netherlands), Galactic (Belgium) and several Chinese companies [25].

In addition to the DoE-listed building block compounds, other CAs have been identified in literature reviews as having promising industrial applications and that can be obtained in a microbial process including acetic acid, citric acid, gluconic acid, acrylic acid, adipic acid, glycolic acid and muconic acid [1], [15], [25], [26]. Among all these, only citric and gluconic acids are currently produced at an industrial scale by fermentation with *Aspergillus niger*, the rest being derived from fuel precursors[27]. Citric acid is not considered a building block, since there are no major derivative routes that can be produced from it, however its use in the food industry is widespread [26]. Adipic acid is traditionally obtained from benzene with its first bio-based production being first reported in 1994 [28]. Since then various pathways were developed for the microbial synthesis of adipic acid and its immediate precursors, glucaric and muconic acids [31]. Acrylic acid presents a very large market (8 billion USD annually), has an established petrochemical production process and various companies are investigating its direct production through fermentation of sugars or through the conversion of the bio-based precursors 3-HPO and fumaric [1], [30], [31]. Likewise, glycolic acid industrial production has been traditionally chemical; however, in the case of glycolic acid, its microbe-based production has only been investigated at the level of academia by exploring the natural glyoxylate cycle of *E. coli* or engineering of yeast. [32]. Muconic acid biobased production is also not being pursued industrially yet, despite its derivative routes to nylons, plastics, resins and food ingredients [25]. However, its microbial production through bioconversion of aromatics or from glucose fermentation has been investigated [33].

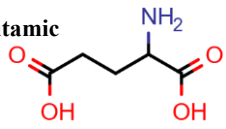
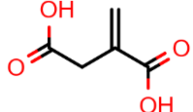
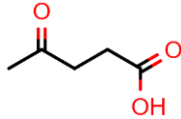
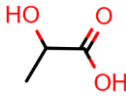
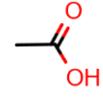
Table 2.1: Main CA building blocks (list of molecules identified in [7], [9], [1], [15], [25], [26]). Alternative microbe-based production methods developed either at the industrial or research scale are listed only when the main industrial production method is not fermentation.

CA	Applications	Important derivatives	Main industrial production method	Alternative microbe-based production method?	In original DOE?	In revisited DOE?
Malic 	Detergent, food additives, pharmaceutical, polyesters, solvents	Butanediol, tetrahydrofuran, GBL	Hydration of maleic anhydride	<i>A. oryzae</i> fermentation from renewable resources at the pilot scale (Novozymes)	Y	N
Succinic 			Hydrogenation of maleic anhydride, maleic or fumaric	<i>E. coli</i> fermentation of sugar (BioAmber and Myriant); <i>S. cerevisiae</i> fermentation of sugars (Reverdia); <i>Basfia succiniciproducens</i> fermentation of sugars and glycerol (Succinity)	Y	Y
Fumaric 			Synthesis from butane-derived maleic	No industrial process. Production by fermentation has been investigated (reviewed in [19]), where filamentous fungi are the most used hosts	Y	N

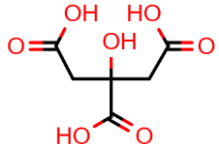
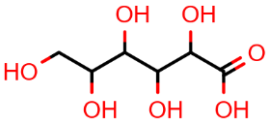
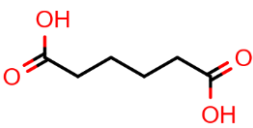
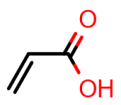
Chapter 2– State of the Art

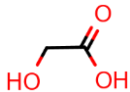
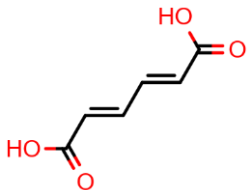
2,5-furan dicarboxylic (FDCA) 	Green polymers	Furans	Bio/ chemical conversion of HMF; catalytic transformation of furan derivatives; oxidation of 2,5 di-substituted furans; and dehydration of hexose derivatives	Microbial conversion of HMF to FDCA (Corbion). Some studies of whole cell catalysis using bacteria isolated from soil, <i>Cupravidus basilensis</i> or <i>Pleurotus ostreatus</i> (reviewed in [34])	Y	Y
3-Hydroxy-propionic 	Superabsorbent, adhesive, surface coating, and pain	Acrylates and 1,3-propanediol	Hydrolysis of 3-hydroxypropionitrile, hydrolysis of β -propiolactone and oxidation of 1,3-propanediol	No industrial process. Studies in microbial fermentation from glucose or glycerol, mainly in <i>E. coli</i> and <i>Klebsiella sp</i> (reviewed in [21]).	Y	N
Aspartic 	Nutritional supplement in food and animal feed, sweetener	Polyaspartic, aspartic anhydride, Amine butanediol, amine tetrahydrofuran, amine butyrolactone	Amination of fumaric acid (enzymatic or with immobilized cells), fermentation with <i>E. coli</i> or <i>C. glutamicum</i>	-	Y	N
Glucaric 	Nylons and polyesters	Lactone, polyglucaric esters and amides	Chemical glucose oxidation or nitric acid oxidation	<i>E. coli</i> fermentation on glucose (Kalion)	Y	N

Chapter 2– State of the Art

Glutamic 	Food additive, potential new polymers	1,5-propanediol, 1,5-propanediacid, 5-amino, 1-butanol	Glucose fermentation with <i>C. glutamicum</i>	-	Y	N
Itaconic 	Rubber, solvents, acrylates, detergents, superabsorbants, drug delivery polymers, dental materials	MAA, MMA, polyesters, poly-ITA and styrene-butadiene	Glucose fermentation with <i>A. terreus</i>	-	Y	N
Levulinic 	Solvents, polymers, acrylates, herbicides, photodynamic therapy	2-Methyl-THF, Levulinate esters, 1,4-Pentanediol, β-acetoacrylate, lactones, δ-aminolevulinic, diphenolic acid	Chemical treatment of pure sugars or sugars from lignocellulosic residues	No industrial process. One reported study with <i>E. coli</i> production from glycerol [35] and one reported study of mixed culture production from hemicellulose hydrolysates [107]	Y	Y
Lactic 	Biodegradable fibers in clothing, furniture and biomaterials	Lactate esters, propylene glycol, acrylates, poly-lactic acid	Glucose is extracted from corn, cassava, sugar cane and fermented with <i>Lactobacillus sp.</i>	-	N	Y
Acetic 	Food additive, solvent, fibers, filters, cellulose plastics and resins (used in paints, adhesives, coatings and textiles)	Vinyl acetate, acetic anhydride, acetate esters, monochloroacetic acid	Methanol carbonylation and liquid phase oxidation of aliphatic hydrocarbons, fermentation	Fermentative acetic acid processes mainly used in the vinegar industry using acetic acid bacteria	N	N

Chapter 2– State of the Art

Citric 	Acidulant, preservative, emulsifier, flavoring additive, sequestrant and buffering agent	-	Starch or glucose fermentation with <i>Aspergillus niger</i>	-	N	N
Gluconic 	Cleaning and construction industries, food additives including prebiotics	Glucono-lactone, sodium gluconate	Oxidation of glucose and fermentation with <i>Aspergillus niger</i>	-	N	N
Adipic 	Nylons and polyesters, plasticizers and lubricants	Esters for polymerization (PVC)	Synthesis from benzene	Pilot-scale yeast fermentation of fatty acid rich-feedstocks (Verdezyne) or glucose (BioAmber); patented adipic acid microbial process and ongoing research (Genomatica)	N	N
Acrylic 	Various coatings (decorative, industrial, drug tablets, cloathes), adhesives, polishes, carpet backing compounds	Methyl acrylate, ethyl acrylate, butyl acrylate and 2-ethylhexyl acrylate, polyacrylates	Oxidation of propene	Fermentation from renewable feedstocks (Arkema); pilot-scale <i>E. coli</i> fermentation from dextrose and sucrose-based feedstocks (OPXBio and Dow)	N	N

Glycolic 	Tanning and dyeing agent for textiles, packaging materials	Polyglycolate, polyglycoside, butyl-glycolate	Catalysis from CO ₂ and formaldehyde and hydrolysis of chloroacetic acid	No industrial process. <i>E. coli</i> (natural producer) [36] and the yeast <i>S. cerevisiae</i> and <i>Kluveromyces lactis</i> (engineered producers) [37] have been investigated	N	N
Muconic 	Plastics industry (automotive and packaging applications), synthetic fibers for textiles or industry (mainly nylon) and food acidifying agent	Adipic, terephthalic acid and trimellitic acid, caprolactam	Catalytic oxidation of cyclohexanol or a cyclohexanol/cyclohexanone mixtures	No industrial process. Biotransformation of aromatics (toluene) and glucose fermentation to muconic has been investigated in natural producers (<i>Pseudomonas spp.</i>) and engineered hosts such as <i>E. coli</i> and <i>S. cerevisiae</i> . Reviewed in [33].	N	N

The industrial production of the molecules describe above is largely variable. For malic and acetic acid, there is an already established petrochemical industrial process and the challenge of biobased processes is to become as cost-effective as the petrochemical competitors. For muconic and 3-HPO there is currently no large production industry and a biobased process presents as an opportunity to displace their derivatives such as adipates and terephthalates from fuel-dependency [32]. In the case of glucaric acid, the current chemical production capacity is low and biobased production is a possible strategy to increase it [38]. Finally, in cases as itaconic, citric and lactic acid, there is already a fermentative industrial process established and biotechnological efforts are mainly focused on improving it, either in terms of titer obtained, cost, productivity, etc. One particular factor to be considered in this context concerns the fact that most of the established processes for production of these CAs involves the use of filamentous fungi as hosts, which for some producing species raises safety issues for the consumer health and food supply, due to the emergence of hypervirulent isolates [39], as well as increased difficulties in terms of process due to the challenges that the filamentous morphology poses in terms of broth aeration. The following sections are dedicated to ITA, MAA and LA (the main target CAs focused in this work), emphasizing what are the different relevant intermediates that these building blocks can be converted to as well as their predicted economical market and level of industrialization already obtained.

2.1.2. Itaconic acid

ITA, or methylenesuccinic acid, is a C5 dicarboxylic acid which is nowadays mostly used in the industry as a copolymer with acrylic and styrene-butadiene in addition polymerization reactions. However, the presence of a methylene group and two carboxylic groups opens a lot of possibilities for ITA to be used as a monomer: it may not only be used in addition polymerization reactions using other unsaturated bonds, but it can also be used in polymerization condensation reactions with alcohols and amines. This versatility to be used with a large variety of copolymers creates a big catalogue of potential novel polymers with new and interesting characteristics such as the ability to cross-link, form dendrimers and chelate [40]. For example, ITA-derived polyesters been studied for the medical applications and ability to have a shape-memory [41]. In addition to a potential monomer, its similarity to maleic anhydride allowed the identification of potential derivative routes through chemical conversions that are routinely executed in the industry with the fuel derived anhydride (Figure 2.1). The transformation of ITA into THFs, GBLs and butanediol (BDO) may confer new abilities to these families of polymers, while ITA-derived pyrrolidines may have applications as polymer precursors and solvents [10].

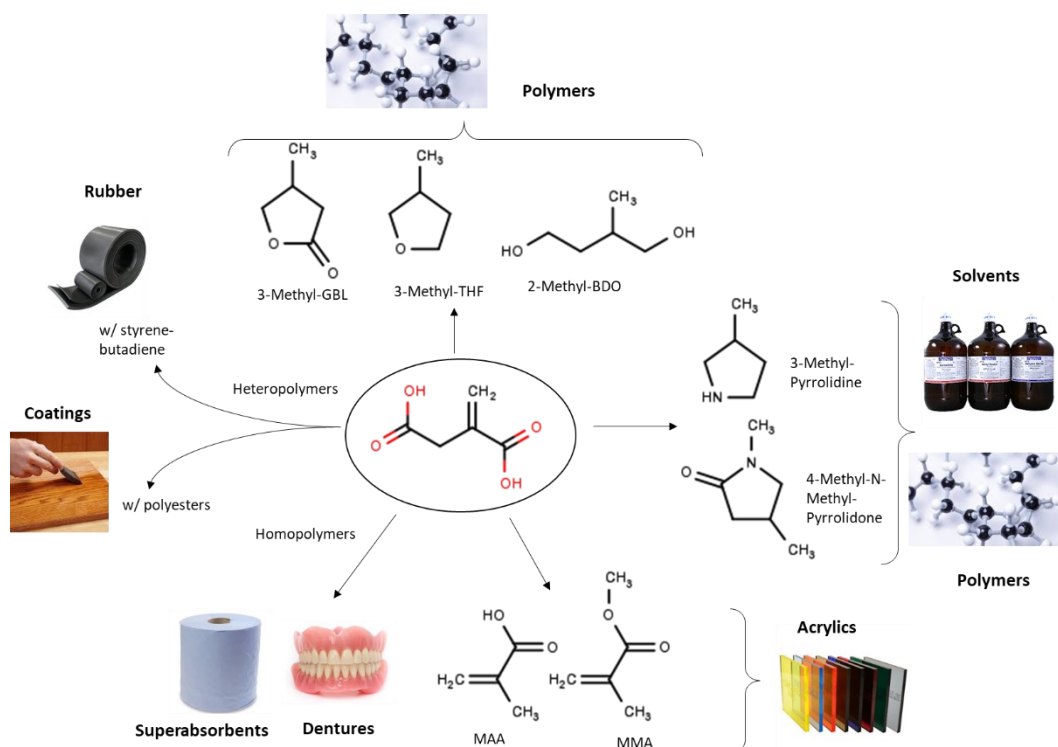


Figure 2.1: Envisioned chemical conversions of ITA to industrially interesting derivatives and their corresponding main applications.

The global market of ITA was estimated in 86.8 Million USD in 2018 and is projected to reach 117.1 Million USD by 2026, at a Compound Annual Growth Rate (CAGR) of 3.8% [44]. The price of ITA is 1.8-2 USD/kg, depending on the quality grade and supplier and the production is mostly located in China [45]. The largest market segment nowadays is the one of styrene-butadiene rubber and synthetic latex. However, the ITA market is still considered a niche and it is expected to expand as the cost of production decreases. As an example, it is estimated that the ITA market in 2020 could reach 500 million USD if the price drops low enough for ITA to enter the methyl acrylates' market and compete with MAA derived from acetone cyanohydrin [45]. Indeed, the conversion route of ITA into MAA and methyl-methacrylate (MMA) presents the largest expected growth of the ITA market, dependent on a large price drop. For this reason, MAA has also been targeted in this thesis as a molecule to be produced in yeast, with more information in section 2.1.3. Another projected expansion of the ITA market, into superabsorbants, relies on the use of cheaper raw materials, exemplified by attempts made by the company Itaconix to use xylan from wood pulp manufacturing process to produce poly-ITA, enabling its price to drop to 1.5 USD/kg, a price that is competitive with the one obtained when fuel-derived poly-acrylates are used [46]. Other potential markets that may be addressed by ITA in the future are poly-ITA detergents to replace sodium tripolyphosphate and poly-ITA dispersants to replace aminopolycarboxylic polymers [45].

ITA is traditionally obtained in the industry via submerged fermentation with *A. terreus*, a natural producer, a process that has been optimized to the point that outcompetes the chemical synthesis route in terms of production costs [42]. In addition to this host, the natural producer *Ustilago maydis* as well as an extensive panoply of non-natural producers, have also been extensively researched as alternative industrial producers. The strategies used to engineer these hosts are reviewed in section 2.2.1. Some reasons to study hosts other than *A. terreus* for ITA production include the difficulties associated to the use of filamentous fungi in fermentative processes already mentioned above [43] and the high sensitivity of this fungus to metal concentrations which complicates the use of raw materials with a less defined composition [43]. The higher safety level required in the US and in the UE for the manipulation of *A. terreus* is another problem.

2.1.3. Methacrylic acid

MAA and its most used ester, Methyl Methacrylic acid (MMA), are important monomers in the chemical industry due to their high reactivity. This high reactivity occurs due to the polarization of the double bond that enables a subsequent nucleophilic addition; due to possible esterification of MAA, transesterification of MMA and, most importantly, polymerization reactions upon heating and addition of free radical initiators [47]. MMA is used to produce poly-MMA which is a transparent and UV-resistant polymer with the commercial name of Plexiglas®. It is vastly used in shelving, mirrors, impact shields, picture frames and road signals. Besides the homopolymer, MMA is also a co-monomer in various blends for use in construction, paints, coatings, automotive components and biomedical materials. In addition to poly-MMA, co-polymerization of the MAA and its derivative 2-Hydroxyethyl methacrylate (HEMA) with other monomers (such as ITA) creates polymers with interesting properties such as being biocompatible, stimuli-responsive (to pH and temperature) and creating a microbe-barrier, generating possibilities for utilization in drug delivery, wound dressings and as a “biological glue” [48]–[50].

MMA industrial production nowadays is chemical, mainly through the acetone cyanohydrin (ACH) route. In the conventional route proposed by ICI& Rohm, that is still widely used nowadays, ACH is reacted with excess sulfuric acid to give methacrylamide sulfate, which is either hydrolyzed with excess water to give MAA and ammonium sulfate or undergoes hydrolysis and esterification in the presence of methanol to give a mixture of MAA and MMA (Figure 2.2A) [51]. Mitsubishi proposed an alternative formulation to prevent sulfuric acid use, where ACH is hydrated in the presence of a manganese catalyst then esterified and dehydrated to yield MMA [52] (Figure 2.2B). Two major pitfalls pointed to ACH route is the formation of large quantities of ammonium bisulfate that are costly to deal with and the need of a large supply of hydrogen cyanide [52]. To circumvent these problems, Lucite International has developed and is currently using the Alpha 2-stage process to produce MMA from carbon monoxide, methanol and ethylene [53], [54]. In this process, ethylene, methanol and carbon monoxide are reacted together in the presence of palladium homogenous catalyst to form methyl propionate, which reacts with formaldehyde in the gaseous phase in the presence of methanol and a caesium/zirconia/silica catalyst to form MMA (Figure 2.2C). While the Alpha process has the potential to use substrates that may be from renewable resources, such as syngas (mixture of hydrogen, carbon monoxide and dioxide) and bioethanol, ethylene is currently

mainly produced at high volumes and low costs through steam cracking of fuel products. Furthermore, the Alpha process handles toxic compounds, such as methanol and formalin (used to increase process selectivity). Possibly for these reasons, research to implement bio-based production of MAA and MMA has been ongoing and these biobased processes are further explored in section 2.2.2.

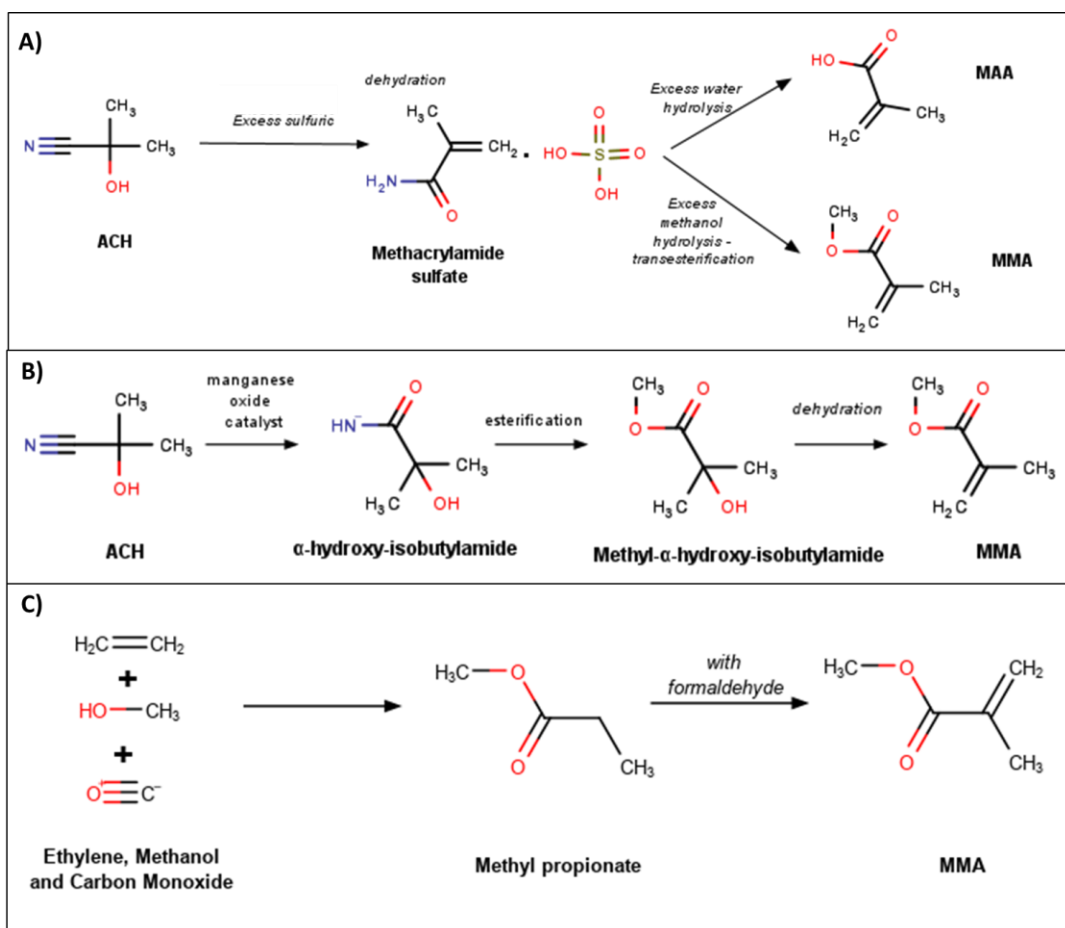


Figure 2.2. Production processes for MAA and MMA established in the industry. A) Conventional ACH route with the use of excess sulfuric acid established by Rohm & ICI; B) New ACH route established by Mitsubishi; C) Alpha process established by Lucite.

The market for MAA was valued at 1.13 Billion USD by 2019 with an expected CAGR of 3.7% [55]. Indeed, this market is so important and well-developed that ITA conversion into MAA and MMA is considered a major market driver of ITA [40], [56]. Lucite has over 35% of the global market share for MAA and MMA and is giving an example in the biobased production strategies by patenting MMA production routes via fermentative

ITA and further investigating the fully fermentative route to MAA. Evonik is also investing in biobased MAA production routes (fermentative and mixed) and Ascenix Biotechnologies has licensed a technology to produce sugar-based isobutyric, which can be converted into MMA.

2.1.4. Levulinic acid

LA is a platform compound that can be converted in various relevant intermediates as detailed in Figure 2.3. One of the more interesting LA derivatives that already has an interesting market as a solvent and that is considered of interest for production of synthetic latexes is γ -valerolactone [57]. This compound can be obtained upon a reduction and subsequent dehydration of LA. Methyl-THF and angelicalactone, other two LA derivatives (Methyl-THF being obtained from γ -valerolactone), also have applications as potential industrial solvents. 1,4-pentanediol can also be formed from LA, with expected applications as a building block in synthesis of polyesters [58]. On another derivative route, the amination of LA yields δ -aminolevulinic acid (D-ALA), which is used as an herbicide and is also being investigated as a potential pharmaceutical in photodynamic therapy [59]. It is also predicted that LA can be converted in acrylic and β -acetoacrylic acid, thus providing a sustainable acrylate source [10]. In the original DOE report it was also suggested that LA could be efficiently oxidized to succinic acid, which, as discussed above, is also a building block with a large number of applications [10]. LA may also be esterified with methanol or ethanol and the resulting esters can be used in biofuel blends and additives (an application shared with methyl-THF) [60]. Besides the drop-in applications of LA derivatives, both diphenolic acid and levulinic ketal have the potential to displace two fuel-derived industrial intermediates: bisphenol-A, an extremely toxic compound, currently used in the manufacture of carboxy resins and may be replaced by diphenolic acid; and LA-ketal esters that can substitute phthalates in plasticizing applications [8].

The older established production process for LA in the industry uses a petrochemical route, where furfuryl alcohol is synthesized from maleic anhydride and is subsequently hydrolysed to LA in the presence of hydrogen chloride [61]. Attempts to create LA production processes that do not depend on fuel-derived resources and are not as polluting have been performed, out of which the most successful is Biofine. In this process the cellulose in lignocellulosic biomass is hydrolysed by dilute sulfuric acid at high temperatures, obtaining the main product, LA, and two by-products, formic acid and

biochar [60]. The Biofine process reached the pilot scale with the establishment of a plant in Caserta (Italy), with a processing capability of 50 dry tons of feedstock and was latter acquired by GFB Biochemicals [62]. The market volume of LA produced via the petrochemical and biomass routes is unclear but there was an estimation in 2012 for LA production at 2.5 kT with a price of 5 USD/kg, mainly through the petrochemical route [62]. As with many other building blocks, a substantial drop in price and increase in production capacity is required for LA derivatives reach their high-volume low-price markets. In the biobased market, two Biofine-like processes where biomass is chemically treated to obtain LA have been developed by GFB Biochemicals (Atlas technology) and Avantium Chemical, with both at the commercial stage. GFB Biochemicals has established itself as a main player the market of biomass-derived LA and is now selling LA derivatives in the specialty market (esters and ketals) and approaching the large volume market (lactones, D-ALA and diphenolic acid) [63]. DSM is also working on LA production processes from papper sludge [62].

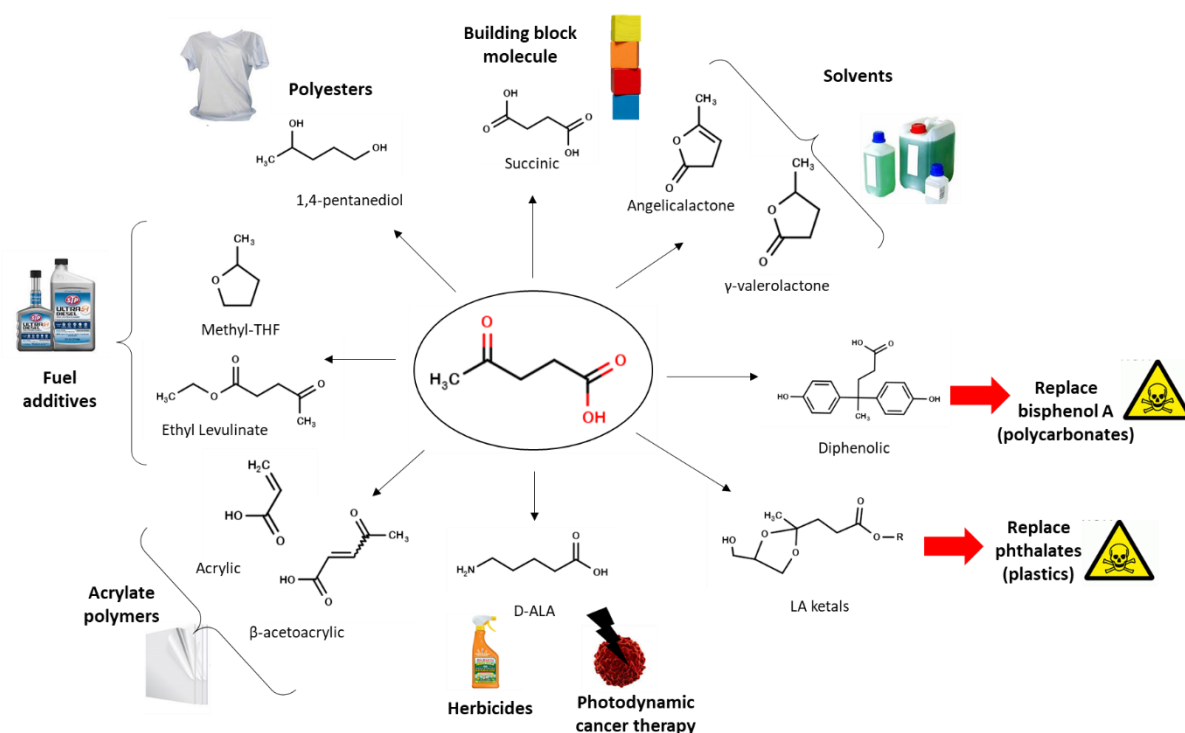


Figure 2.3. Envisioned chemical conversions of LA to industrially interesting derivatives and their corresponding main applications. The derivatives which are predicted to displace fuel-derived chemicals are depicted with a red arrow.

2.2. Approaches to implement bio-based production of LA, MAA and ITA

In the previous section the main derivate routes and markets of ITA, MAA and LA were discussed as well as the current stage of their industrial production processes. In this section the bio-based approaches explored for production of these building blocks from renewable feedstocks are reviewed, in some cases already implemented by the industry. These included routes that rely only on chemical catalysis, routes that rely on microbial fermentation and routes that involve a mixture of the two kinds of processes. A fundamental aspect that has to be taken into account is the fact that MAA and LA are molecules that are not described in any biochemical pathway described in living systems, which necessarily complicates the implementation of a microbe-based production process.

2.2.1. Itaconic Acid

Production of ITA has been reported in the filamentous fungi *A. terreus* and *U. maydis* and also in mammalian macrophages where this molecule plays an immunosupportive role [64]–[66]. Yeast species belonging to the *Candida* and *Pseudozyma* genera, were also described to produce ITA, although these were the sole descriptions made in the literature thus far [43]. The biosynthetic pathway leading to ITA and the genetic elements involved have been characterized essentially in *A. terreus* and, more recently, in *U. maydis* [64], [67] (Figure 2.4). In both cases the pathway involves compartmentalization: the precursor cis-aconitic is produced in the mitochondrial TCA cycle as the result of aconitase activity, being subsequently decarboxylated to ITA in the cytosol by cis-aconitate decarboxylases (AtcadA). In *U. maydis*, upon shuttling of cis-aconitic to the cytosol, isomerization by a aconitate-D-isomerase (Adi1) to trans-aconitate occurs, being this one subsequently decarboxylated to ITA by the trans-aconitate decarboxylase Tad1 [64] (Figure 2.4). In *A. terreus* a mitochondrial tricarboxylate transporter (mttA) catalyzes the extrusion of cis-aconitate to the cytosol [67], while in *U. maydis* the shuttling of cis-aconitate to the cytosol is attributed to Mtt1, a protein sharing 35% sequence similarity with AtMttA [64]. The export of ITA to the extracellular environment is attributed to the major facilitator superfamily transporters MfsA in *A. terreus* and Idp1 in *U. maydis* [64], [68]. One relevant ITA drain in these natural producers is ITA oxidation to 2-hydroxyparaconate (HP-ITA), catalyzed by a P450 cytochrome, which in *U. maydis* is

attributed to Cyp3 and in *A. terreus* is hypothesized to be encoded by cypC [69], [70].

Despite the problems associated with industrial production using filamentous fungi, the most used host in the industrial setting is *A. terreus* mainly due to the high titers of ITA produced by this species that can reach ~20 g/L in a batch fermentation in a wild-type (WT) strain without any genetic modifications, a titer that upon some process improvement (pH control, increased phosphate concentration and fed-batch mode) can increase to 160 g/L, already considered an industrially relevant titer [71], [72]. Various metabolic engineering strategies have been applied to further improve performance of this host. One of these relied on increasing the pool of TCA cycle intermediates by engineering phosphofructokinase to become overactive and citrate-independent, this resulting in production of two times more ITA than the one obtained with the original strain [71]. Additionally, overexpression of cadA has enabled a 10% increase in ITA titer, while random mutagenesis and selection of best producers allowed ITA production from corn starch and fruit waste to reach titers of 50 and 32 g/L [65], [68]. Due to the already mentioned above limitations of using *A. terreus*, attempts have been made to implement ITA production in the closely related species *A. niger*, a natural citric acid overproducer largely with a strong track record of use in industrial biotechnology. The production of ITA (enabled through heterologous expression of CadA) in a wild-type *A. niger* strain and culture media favoring stability of CadA was 0.57 g/L [73]. Expression of CadA in a good citric acid producer strain (ATCC 1015) resulted in titers of 50-135 mg/L [74], [75]. The impact of compartmentalization in ITA production was also investigated, by expressing both aconitase and cis-aconitate decarboxylase either in the cytosol or in the mitochondria and it was concluded that exclusive mitochondrial ITA production leads to a doubling in productivity, in comparison to with exclusive cytosolic production, with a final titer of 1.2 g/L [75].

The elucidation of the biosynthetic pathway in *U. maydis* has enabled the exploration of this species as new ITA production host. The deletion of cyp3, coupled with the overexpression of the cluster regulator ria1 (Figure 2.4B) resulted in an overproducer strain (Δ cyp3 OE-ria1) able to produce ITA to a titer 4-fold higher than the WT and 2.6-fold higher than a strain only over-expressing ria1 overexpression [70]. Upon further optimization of the feeding regime, the Δ cyp3 OE-ria1 strain led to titers of 54 g/L and a yield of 67% the maximal theoretical. The close relative *U. cynodontis* was recently reported to be a high ITA natural producer having as an advantage compared with *U. maydis* the fact that is more acid-tolerant and therefore production of ITA can undergo for a longer period of time. The pathway leading to ITA biosynthesis in *U. cynodontis*

was identical to the one described in *U. maydis* and it was found that on cultivation in 2-(N-morpholino)ethanesulfonic acid media in 24-well plates no ITA is produced by *U. maydis* after 96h, while *U. cynodontis* produced 2.6 g/L [76]. To further improve this host to enable bioreactor-scale fermentation, morphological engineering was attempted by creating the triple mutant $\Delta ras2\Delta fuz7\Delta ubc3$, devoid of three genes involved in the Mitogen Activated Protein Kinase (MAPK) pathway, which plays a role in filamentous growth. The morphological engineering of *U. cynodontis*, coupled with a similar metabolic engineering strategy to the one attempted in *U. maydis* (overexpression *ria1* and deletion of *cyp3*, Figure 2.4) and with the overexpression of *At_mttA*, the *A. terreus* mitochondrial cis-aconitate transporter, resulted in 22.3 g/L ITA titer [76].

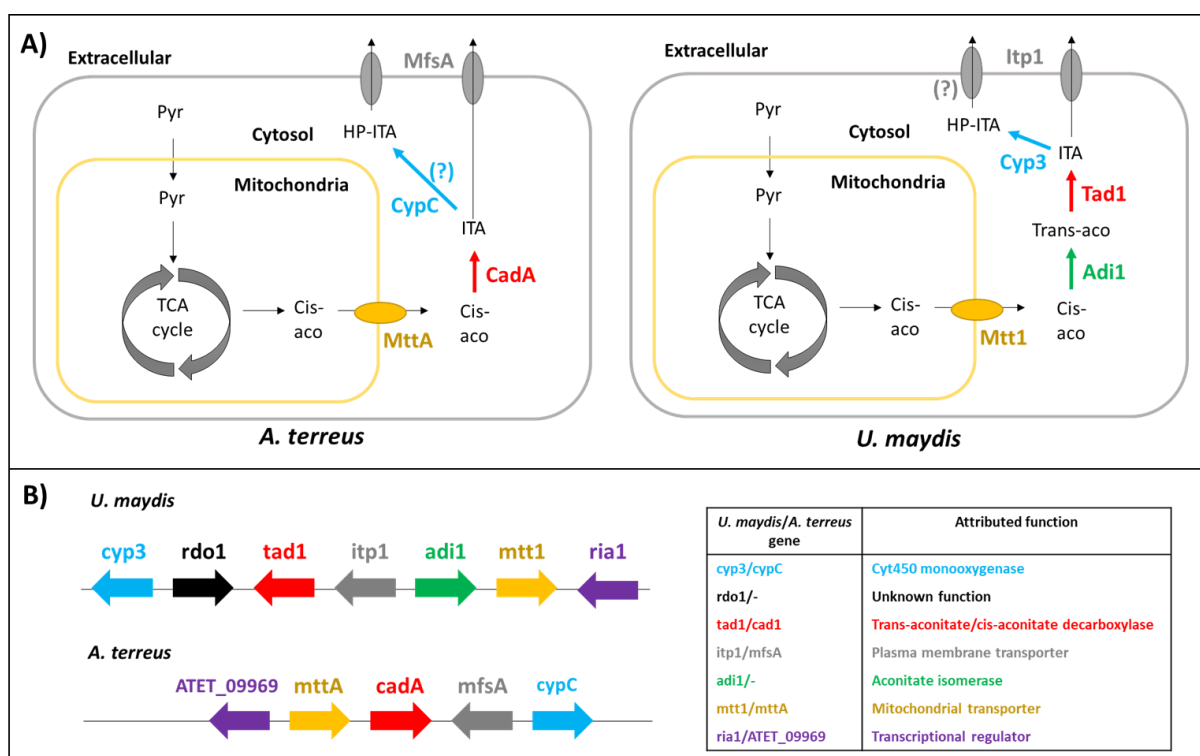


Figure 2.4. Metabolic and genetic elements in biosynthesis of ITA in *A. terreus* and *U. maydis*. A) ITA biosynthesis pathway in both organisms. The *cypC*-mediated conversion of ITA in HP-ITA and the *Itp1*-mediated export of HP-ITA are hypothetical and are depicted with (?). B) Gene clusters in both organisms. The genes encoding proteins represented in panel A are depicted in the same colors; the attributed function to each protein is described in panel B.

In addition to *Aspergilli sp* and *Ustilago sp*, a variation of hosts with no phylogenetic proximity to these organisms have also been explored as ITA producers through the heterologous expression of *A. terreus* cis-aconitate decarboxylase. Until this moment the tested hosts were the industrial amino acid producer *C. glutamicum*, the model organisms

E. coli and *S. cerevisiae*, the non-conventional yeast *Yarrowia lipolytica*, known for its lipid and organic accumulation; the photosynthetic cyanobacteria *Synechocystis sp* PCC6803 and the cellulose consumers *Neurospora crassa* or *Pichia stipitis*. The metabolic engineering strategies applied to improve the production titers in these hosts (reviewed in Table 2.2) range from rational strategies, such as overexpression of aconitase to increase the precursor pool or deletion of competing steps, to co-localization of the enzymes or *in silico* modeling tools to predict useful deletions. Random methodologies have also been attempted including the screening of a combinatorial promoter-terminator library to optimize pathway expression [77], [78]. In a more structural methodology, protein scaffolds were constructed in *E. coli* to channel the flux of metabolites directly from the citrate synthase step until the cis-aconitate decarboxylase step [79], [80].

Table 2.2: ITA production strategies in hosts other than *Aspergilli sp* and *Ustilago sp*.

Host	Strategy	Production	Ref
<i>C. glutamicum</i>	Culture media optimization and isocitrate dehydrogenase downregulation.	Titer of 7.8 g/L and productivity of 0.27 g/L/h	[81]
<i>E. coli</i>	Expression of citrate synthase and aconitase from <i>C. glutamicum</i> , deletion of phosphate acetyltransferase and lactate dehydrogenase.	Titer of 690 mg/L	[82]
	Model-guided knockouts; fed-batch culture.	Titer of 32 g/L and yield of 0.77 mol/mol	[77]
	Synonymous codon variant of <i>cadA</i> to enable expression in the soluble fraction, culture media optimization	Titer of 7.2 g/L	[83]
	Deletion of isocitrate dehydrogenase, overexpression pyruvate carboxylase from <i>C. glutamicum</i> and of citrate synthase and aconitase. Fed batch in optimized media with glycerol as carbon source.	Titer of 43 g/L and yield of 0.6 g/g	[84]

	Self-assembly of aconitase and cis-aconitate decarboxylase by use of protein-protein interactions of the PDZ and PDZ-ligand domains.	Titer of 222.15 mg/L	[80]
	Self assembly of citrate synthase, aconitase and cis-aconitate decarboxylase by interactions between GBD ligand, SH3 ligand and PDZ ligand. Temperature and pH optimization. Deletion of competing lactic and acetic acid production pathways.	Titer of 6.57 g/L	[79]
<i>S. cerevisiae</i>	Model-guided knockouts and high cell density fermentation.	Titer of 168 mg/L	[78]
	Alternative pathway encompassing acetylating acetaldehyde dehydrogenase to feed cytosolic acetyl-coA pool and a cytosolic citrate synthase from <i>Listeria innocua</i> . Expression optimization by using a combinatorial promoter-terminator library.	Titer of 815 mg/L	[85]
<i>Y. lipolytica</i>	Expression of aconitase without mitochondrial localization signal, culture media optimization and pH-controlled bioreactor.	Titer of 4.6 g/L, yield of 0.058 g/g and maximum productivity of 0.045 g/L/h.	[86]
	Expression of aconitase and MFS ITA transporter from <i>A. terreus</i> , media optimization, fed-batch at pH-controlled bioreactor	Titer of 22.03 g/L	[87]
<i>Synecho Cystis sp. PCC6803</i>	Photosynthetic ITA production from CO ₂	Titer of 14.5 mg/L and productivity of 42.8 µl/L/day	[88]
<i>P. stipitis</i>	Media optimization, overexpression of cytoplasmic aconitase, fed-batch bioreactor.	Titer of 1.5 g/L	[89]
<i>N. crassa</i>	Direct cellulose consumption from lignocellulosic material	Titer of 20.4 mg/L	[90]

2.2.2. Methacrylic Acid

Currently all MAA and MMA production processes in the industry rely on chemical catalysis with no bio-based production being already implemented in the industry. However, some attempts to this end are starting to be reported and these will be described below and are summarized in Figure 2.5. In order to improve environmental sustainability of the chemical processes used to produce MAA, an attempt to replace the use of fossil-derived intermediates for renewable ones has been performed. Another option has been to use a CA (like citric acid or ITA) obtained by fermentation of renewable sources which is afterwards converted into MAA or MMA. This is the concept of bio-chemocatalytic routes, where there is a first fermentative stage and a second chemical catalysis stage [91]. Lucite^R has patented a process based on citramalic one-stage conversion (with dehydration and decarboxylation) to MAA in the presence of a sodium hydroxide catalyst using high temperatures and pressures [92] (Figure 2.5B). Despite this, the Alpha process (using ethylene, carbon monoxide and methanol) is the main process used by Lucite to produce MAA nowadays. Citric acid may also yield MAA upon a series of dehydration and decarboxylation reactions occurring simultaneously in supercritical and near critical water systems [93]. The more common of such routes to produce MAA involves dehydration of citric acid to cis-aconitic acid, which is then decarboxylated to ITA and subsequently decarboxylated to MAA (Figure 2.5A). A second possibility involves decarboxylation of citric acid to citramalic, which is then decarboxylated to 2-hydroxyisobutyric, which is finally dehydrated to MAA (Figure 2.5B). Direct decarboxylation from ITA has been developed as a possibility to produce MAA with high selectivity (less byproduct formation), however, this requires extreme temperatures and pressures and it leads to the formation as byproducts of 2-hydroxybutyric acid and propylene [94] (Figure 2.5C). Another process to enable ITA decarboxylation into MAA under milder conditions involves heterogeneous catalysis, resulting in superior selectivities (84%) (less formation of the degradation products crotonic and pyruvic acid) [95]. Coupling of this route with the production of ITA by fermentation can result in a somehow hybrid bio-based process for production of MAA [96]. Another possibility is to produce the MMA precursor, 2-hydroxyisobutyric, by fermentation and subsequently convert it to MMA. A process as such has been investigated by Ascenix Technologies and is now reported to be successfully developed by Evonik, where bacteria use waste syngas to produce the 2-hydroxyisobutyric, which is subsequently converted to MMA by chemical catalysis [97] (Figure 2.5D). This fermentative process is likely dependent on acetogenic bacteria expressing a 2-hydroxyisobutyryl-coenzyme A mutase [98].

Finally, the direct fermentative production of MAA from glucose has also been investigated and it is of note that methacrylyl-coA (MAA-coA) is a natural intermediate in valine catabolism [91]. A patent has been filled by Genomatica to produce MAA and MMA by fermentation [99], [100]. The patent describes three potential metabolic pathways leading to MAA: deamination of 3-amino-2-methylpropionate, dehydration of 3-hydroxyisobutyrate, decarboxylation of mesaconate and coA transfer of methacrylyl-coA (Figure 2.5E). The possibility of obtaining MMA via the enzymatic action of alcohol acetyltransferases on methanol and methacrylyl-coA is also envisioned. However, in this patent no *in vivo* MAA production is reported. Additionally, a patent has also been filled by Lucite for the microbial production of MAA through MAA-coA, where this intermediate is obtained from isobutyryl-coA through the action of an acyl-coA oxidase from *Arabidopsis thaliana* (ACX4), followed by removal of the coA from the promiscuous activity of an hydroxybenzoyl coA thioesterase from *Arthrobacter sp.* SU (4HBT) [101] (Figure 2.5F). In this work, MAA production was reported with an *E. coli* strain expression ACX4 and 4HBT with media supplementation of isobutyric acid.

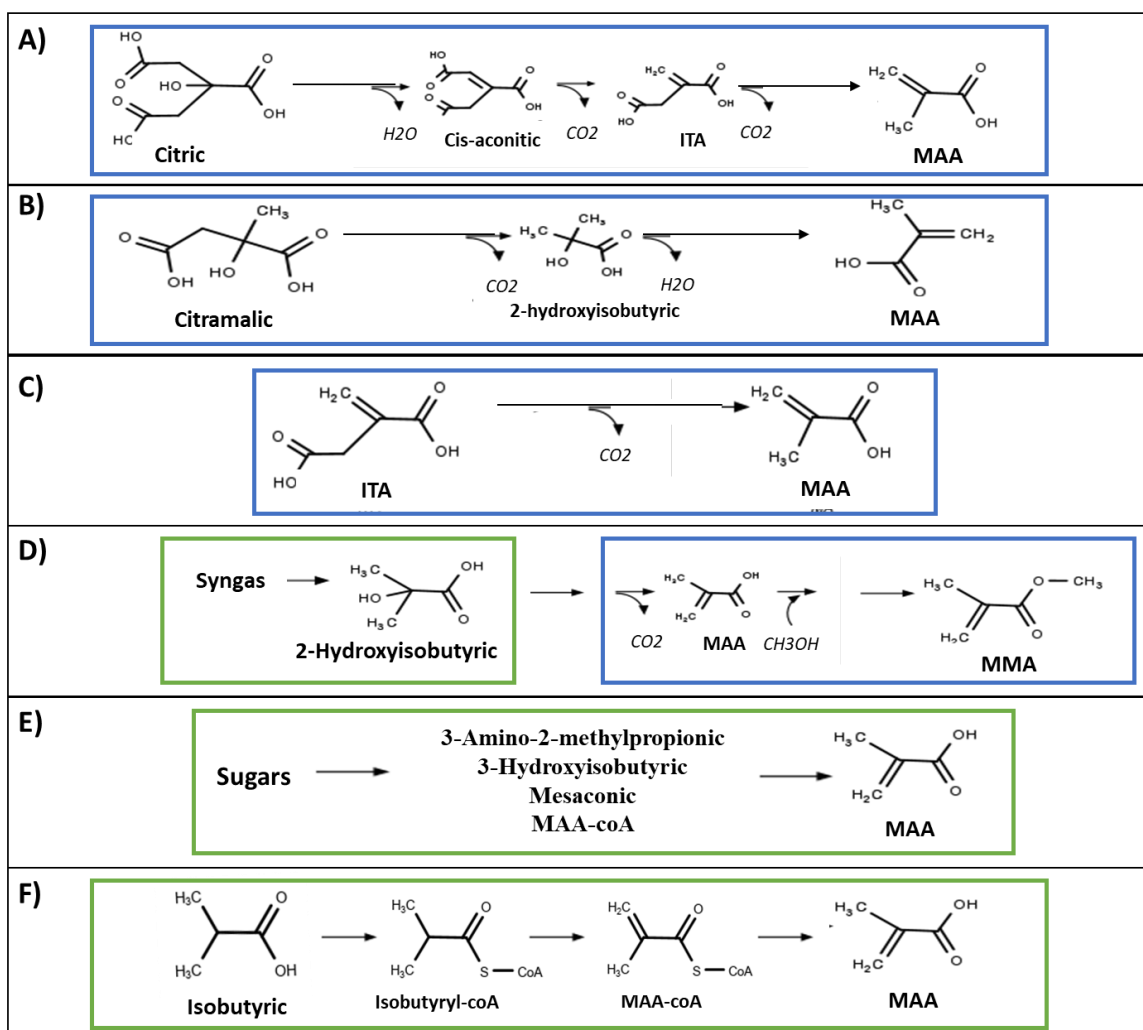


Figure 2.5. Published biobased routes to MAA production. Chemical catalysis and fermentation processes are depicted in blue and green squares, respectively. A) Lucite route through citric acid. B) Route through citric acid from [92]. C) Route through ITA from [93], [94]. D) Ascenix Technologies mixed fermentative/chemical process through syngas. E) Exclusively fermentative route for MAA production, patented by Genomatica. F) Bioconversion of isobutyric to MAA, patented by Lucite.

2.2.3. Levulinic Acid

In alternative to the conventional petrochemical route that involves maleic anhydride conversion to furfuryl alcohol and its subsequent hydrolysis, LA production from the hydrolysis of 5-hydroxymethylfurfural (HMF), which can be obtained from the dehydration of fructose, has been attempted. Pentoses (e.g., xylose and arabinose) can go

through acid catalyzed dehydration to yield furfural, which can then be reduced to furfuryl alcohol and further hydrolyzed to LA. These routes, mainly focused on sugar-containing residues, are the basis for the variations of LA production processes that have been developed so far. The most advanced technology for bio-based production of LA is the Biofine process, a 2-stage treatment of lignocellulosic biomass where hexoses are first converted to HMF (upon acid hydrolysis at high temperatures and pressures) which is subsequently hydrolyzed to LA [60]. This process co-produces formic acid and furfural, being LA obtained at a maximum yield close to 80%. Despite these interesting numbers, there are relevant disadvantages to the Biofine process including the formation of humins which clog the systems, the costs in recovery of the acid catalyst and extensive water and energy usage [102], [103]. Although homogenous acid catalysis approaches result in high LA formation, it leads to corrosion of equipments as well as to important environmental concerns due to the extensive use of toxic acids in the process [104]. Heterogenous catalysis with solid metals circumvents the problems of catalyst recovery and corrosion but the process has lower yield, presumably due to adsorption of LA to the catalyst and is less optimized. Other production processes developed for LA production from lignocellulosic biomass, cellulose and glucose include solvolysis (e.g. in butanol to allow catalyst precipitation and easy recovery), ionic liquids and supercritical fluids (e.g. supercritical water or acetone) but are under developed [103].

Although the aforementioned processes for LA production represent steps forward in the direction of higher environmental sustainability, they do not represent truly bio-based production process, likely due to the fact that LA is not an identified metabolite in described biochemical pathways in living systems. Only recently some studies have been performed concerning the attempt to implement a biosynthetic pathway that could lead to LA production from sugars. One of those studies is a patent submitted by Arzeda corp [105] that proposes a route starting with the aldol condensation of pyruvate and acetaldehyde is described, yielding 4-hydroxy-2-oxopentanoic acid (Figure 2.6A). The subsequent oxidation and dehydration of this molecule yield 4-oxo-2-hydroxypentanoic acid, which can then be dehydrated to 4-hydroxy-2-pentenoic acid. An alternative oxidative dehydration enzymatic step enables obtaining 4-hydroxy-2-pentenoic in one step. Finally, this intermediate is reduced to LA. However, although it was possible to identify enzymes to mediate the first aldol condensation reaction (*Pseudomonas putida* Hpa1 aldolase) and for the reduction of 4-oxo-2-pentenoic acid (*S. cerevisiae* OYE3 NADPH dehydrogenase or *Pseudomonas savastanoi* pc. *Glycinea* NAD(P)H-dependent 2-cyclohexen-1-one reductase), no enzymes have been assigned for the remaining steps

which is a major bottleneck for a successful implementation of this pathway. The other study that reports a biochemical pathway for production of LA explores the use of Claisen non-decarboxylative condensation reactions as a mean to initiate anabolic routes that may lead to compounds of industrial interest [35]. In a combinatorial effort that resulted in the successful synthesis of 18 products of 10 different classes, LA was synthesized *in vivo* in *E. coli* from succinyl-coA and acetyl-coA (Figure 2.6B). The condensation of these two molecules catalyzed by *E. coli* β -ketoacyl-CoA thiolase (PaaJ) yields 3-oxoadipyl-coA. The CoA group is removed by *P. putida* coA transferase PcaIJ resulting in 3-oxoadipic acid which is presumably decarboxylated to LA by an endogenous *E. coli* enzyme, resulting in the accumulation of 48 mg/L of LA in the broth. Expression of the decarboxylases methylketone synthase from *Solanum habrochaites* (Mks1) and acetoacetate decarboxylase from *Clostridium acetobutylicum* (Adc) increased the titers to 71 mg/L and 159 mg/L, respectively, indicating that these have higher activity for 3-oxoadipic than the endogenous decarboxylase. This work demonstrates that indeed producing LA in a microbial host is possible; however, there are challenges in this route to be addressed such as the low titers (50-150 mg/L) or the requirement of succinic acid supplementation in the culture media, which would imply extra economic costs for an industrialized process. In eukaryotic hosts succinyl-coA is synthesized in the mitochondria which can create additional compartmentalization problems that could be solved (e.g. by re-targeting the involved enzymes to the same compartment or by using a bacterial α -ketoglutarate dehydrogenase to enable a cytosolic pool of succinyl-coA, however, these are strategies that can also bring additional problems (for example the use of a bacterial α -ketoglutarate dehydrogenase requires a supply in the cytosol for the enzyme cofactor, lipoic acid) [106]. Finally, a very recent study reported the co-production of LA with Poly-hydroxyalkanoates (PHA) from hemicellulose hydrolysates, using a mixed microbial culture [107]. While the metabolic route leading to LA formation was not studied, it was hypothesized that the pentoses were converted in pyruvic acid through the pentose phosphate pathway, which was converted to LA through a pathway similar to the one described in the Arzeda patent [105], or that the pyruvic acid was oxidized to acetyl-coA, which, jointly with succinyl-coA could form LA in a route similar to the one described by Cheong *et al*, 2016 [35]. The LA-forming mixed microbial culture was obtained through the selection of organisms able to use hemicellulose hydrolysates in sequential batch reactors inoculated with activated sludge and its composition was predominated by bacteria of the *Paracoccus*, *Lactococcus*, *Enterococcus* and *Azospirillum* genera and by fungi of the *Haglerozyma* and *Gueomyces* genera.

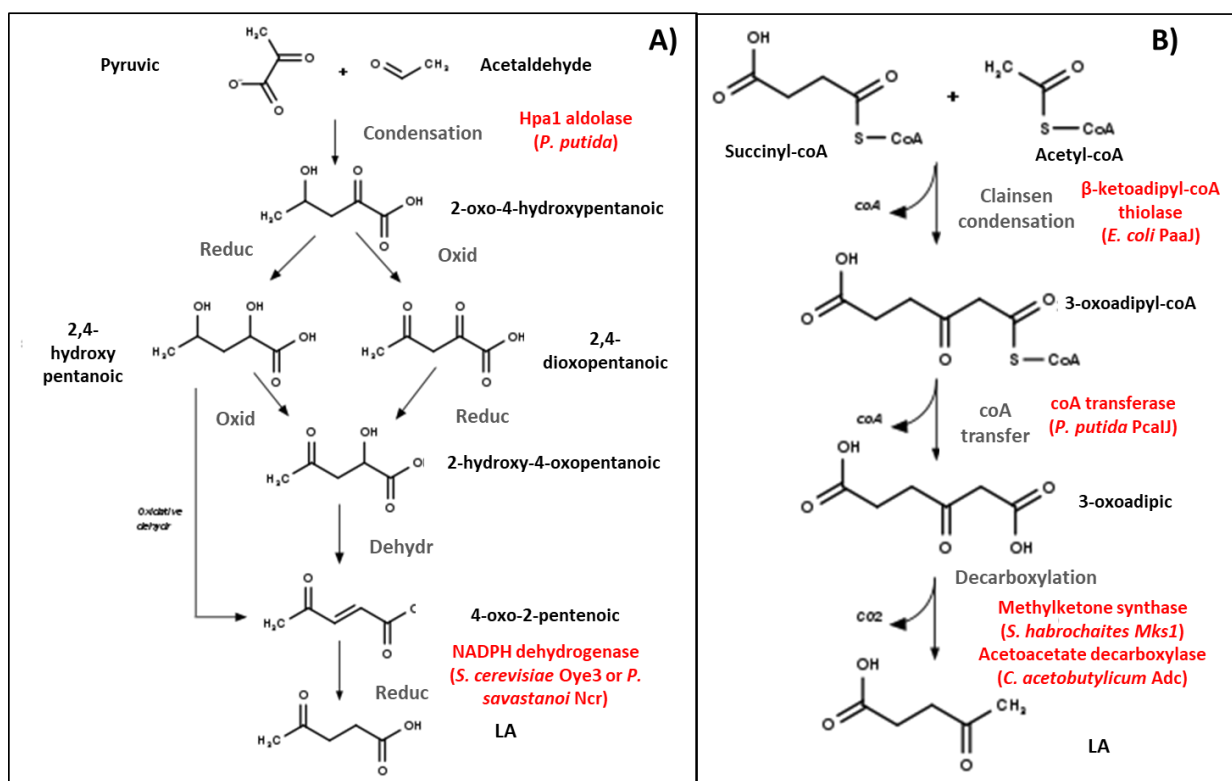


Figure 2.6. Published metabolic routes leading to LA. A) Envisioned route from pyruvic acid, patented in [84], with only two of five potential steps validated *in vitro* for specific enzymes (depicted in red). B) Route established in *E. coli*, starting from succinyl-coA and acetyl-coA; the overexpressed enzymes are depicted in red. [85]

2.3. How can synthetic biology contribute for the production of “new-to-nature molecules” like MAA or LA

In the section above it was reviewed the strategies that are currently in place as well as the studies under development aiming to implement bio-based production routes for LA, ITA and MAA. As it was clear, the fact that LA and MAA are molecules that fall outside of microbial metabolic repertoire is an important bottleneck since it becomes difficult to establish functional pathways that can produce these molecules. Synthetic biology can provide an essential contribution in this field by helping to implement these “new-to-nature” pathways. The objective of this section is thus to provide an overview on how

this is performed, detail the methodologies involved, as well as present the main challenges and bottlenecks to it.

The field of synthetic biology is diverse and a rigorous definition is lacking, however, it can be generally stated that this area is dedicated to the use of genetic engineering techniques to manipulate new-to-nature cellular behaviour. The ultimate goal may be considered the one of formalizing biology to the level of rigor similar to the one present in electrical engineering, a field that contributed greatly at the conceptual level to synthetic biology. In this formalization context, the cell is envisioned as a highly programmable machine (computer), consisting of biochemical pathways (circuits), which are a collection of biochemical reactions (logical gates), constituted by many genes and proteins (resistor and capacitor parts), which are made of nucleotides and amino acids (raw materials) [108]. Therefore, cells can be manipulated to perform desired tasks that solve current problems of our society. One impressive example of the power of synthetic biology is the 90 days pressure test imposed on a group of scientists to engineer organisms to produce 10 molecules, with the goal of evaluating the preparedness to address potential unexpected shortages of drugs and industrial chemicals [109]. In this test 6 out of 10 targets were successfully produced and advances were made to improve production titers, where 215 strains were built of 5 microbial species, 2 cell-free systems were devised and 690 new assays were developed. The promise of synthetic biology in the bioeconomy is of creating disruptive technologies that can solve problems in healthcare (e.g. produce affordable drugs), in agriculture (e.g. increase crop yield and resistance), in manufacturing (e.g. create sustainable production processes) and in the environment (e.g. bioremediation solutions). Within this context, it is an important contribution of synthetic biology to be able to provide predictions of what can be the possible transformations leading to the production of compounds that go beyond the described metabolic capabilities of microbes (as it is the case of levulinic acid) and that can contribute to improve the economic viability of biorefineries (Figure 2.7). This effort requires a precise combination of bioengineering, chemistry and computational biology in order to examine all cohorts of possible pathways while assigning the enzymes that could be responsible for the necessary steps. Furthermore, the need of having appropriate tools that can provide a fast and reliable response, in a high-throughput manner, of the performance of strains is also needed considering the velocity at which it becomes possible to engineer strains. Part of these screening methods involve the exploration of biosensors. The two sections below describe the advances in these two areas that are key for the successful implementation of biorefineries.

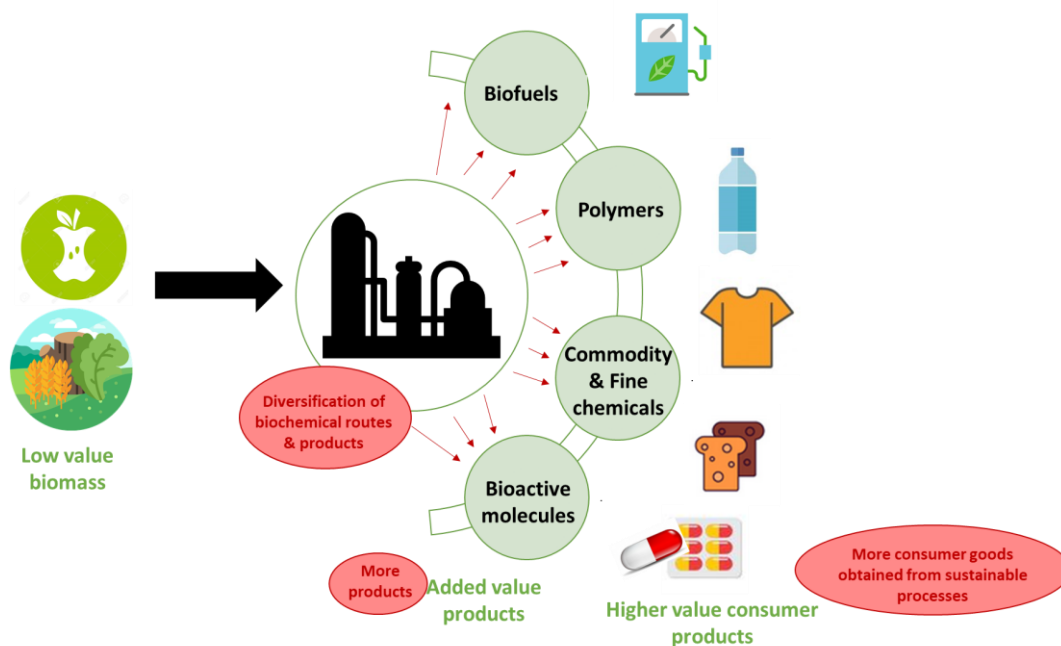


Figure 2.7. Impact of biorefinery product diversification in the bioeconomy. Low value biomass is treated in a biorefinery to be converted in added value molecules which are further transformed in consumer goods, thus increasing the value. By diversifying the biochemical conversion routes in the biorefinery, a larger panoply of added value products can be obtained, enabling more types of consumer goods to be obtained from biomass-derived processes.

2.3.1. Overview of prospection of *new-to-nature* pathways

Many efforts have been focused on enabling chassis organisms to produce molecules of high industrial interest, however, in the vast majority of the cases the targeted molecules are part of the microbial network (even if not of the selected organism of choice). It may be the case that the chassis organism already produces the target molecule (e.g. yeast and ethanol [110]) and so most of metabolic engineering is focused on optimizing the production. In other cases the targeted compounds are natural plant products, such as opioids or antioxidants, and there is interest to move away from plants into microorganism hosts to enable large scale production [111], [112]. However, it is not uncommon that in very large heterologous pathways some of the plant enzymes become non-functional upon heterologous expression due to problems in expression, folding or unexpected interactions. In cases like this the problem has been solved by replacing the

problematic steps with enzymes from different sources that can catalyze the desired reaction, creating new combination of enzymes that are not seen in nature. One recent case concerns the engineering of yeast cells to produce the 4 main cannabinoids including tetahydrocannabinol (THC), where only 5 enzymes were retrieved from the plant that naturally produces these chemicals (*Cannabis sativa*) the remaining 6 enzymes being selected from different bacteria [113]. It is of note that in this work not only naturally cannabinoids were synthesized by yeast cells, but also non-natural ones, these being THC alternatives with tailored C3 side chains that could be functionalized during the biosynthetic process, facilitating post-fermentation chemical derivatization, which can be useful to create new cannabinoid analogues. This was made possible by feeding the cannabinoid pathway with non-native substrates (functionalized fatty acids) that were still able to go through the pathway due to enzymatic promiscuity.

Non-natural compounds represent a new category of chemicals that can now be produced using biotechnology and that can be used as fuels, commodity chemicals, pharmaceuticals, etc. The problem posed by the fact that there is no natural biochemical route leading to their production can be relieved through the design of new synthetic pathways that explore either naturally occurring enzymes and their promiscuity or engineered designed enzymes. To achieve success in the design of new-to-nature pathways, systems biology gave an important contribution since large databases of enzymatic and genomic information have to be coupled with modeling methods (stoichiometric and cheminformatic) to identify what can be the more promising candidate pathways. The following section details what can be the approaches available to endeavor in that task.

2.3.2. Design of synthetic pathways: approaches to the problem

The problem of creating new pathways and/or new compounds has been approached by many different strategies and these have been summarized and categorized in different ways [114]–[116]. In this section the classification will be based on the one used in [117] to overview the efforts in creating new-to-nature pathways, which starts by dividing the potential strategies in ones that only use natural enzymes (that is, enzymes that are expressed by any organism and known to catalyze a given reaction(s)) and other strategies that also use engineered enzymes, that can catalyze new reactions that have not been described before. Within the first set, the strategies can be of “Gap Filling”, “Mix&match” and “Precursor-directed biosynthesis” (Figure 2.8). “Gap filling” consists

on adding heterologous enzymes to a chassis, which may connect parts of its endogenous metabolism that were not connected, enabling shunts and new metabolites to be formed (Figure 2.8A). An example of this is the introduction of the plant α -dioxygenase DoxR from *Oryza sativa* in *S. cerevisiae* that enabled the conversion of native even-chain fatty acids to odd-chain ones, which were then reduced to their corresponding alcohols by the native yeast metabolism [118]. The other strategy using natural enzymes, “Mix & Match”, relies on the combination of enzymes from different hosts to create new pathways and diversifying the molecules that can be produced from the combinations (Figure 2.8B). In *E. coli* the plant carotenoid synthetic pathway was engineered by introducing enzymes from soil bacterium and liverwort, enabling the production of a variant carotenoid, 4-ketoninoxanthin, with reported anti-tumor activity [119]. An important note on these two approaches is the fact the mix & match strategies often rely on enzyme promiscuity, since the heterologous enzymes will catalyze the conversions on substrates that are similar to their native ones, while gap filling approaches use only reactions that already occur in other microorganisms. Another strategy based on natural enzyme is precursor biosynthesis, where a natural pathway is modified only by supplying a synthetic precursor that is structurally similar to the natural precursor (Figure 2.8C). This was applied to the production of the non-natural aminoacid 5-tryptophan in *E. coli*, which was possible simply by feeding the cells that expressed the tryptophan biosynthetic pathway (TrpDCBA) with the synthetic analogue 5-hydroxyanthranilate (5-HAnt), instead of the natural one, anthranilate (Ant) [120]. However, the *ex vivo* supply of the synthetic precursor may be a limitation, which can be overcome if an upstream pathway is inserted leading to the production of the synthetic precursor (depicted in Figure 2.8C as “Or synthetic precursor is generated *in vivo*”). In the case of 5-tryptophan production in *E. coli*, this was possible by expressing a pathway that converted glucose to Ant (endogenous to *E. coli*), plus the salicylate 5-hydroxylase salABCD from *Ralstonia eutropha*, which converted the endogenous metabolite Ant to 5-HAnt [120]. While the production of 5-HAnt was carried out in a different *E. coli* strain than the conversion of 5-HAnt to 5-tryptophan, the establishment of a unique biosynthetic pathway was still achieved.

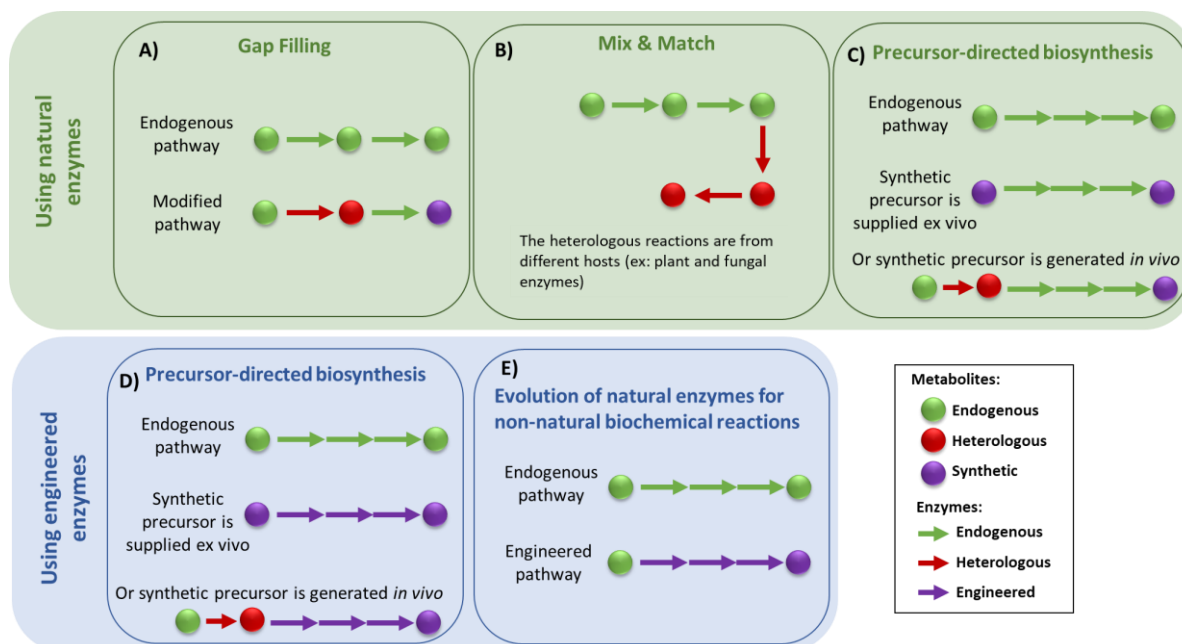


Figure 2.8. Approaches to create *new-to-nature* pathways, based on [116]. Approaches that use natural enzymes can be based on mixing & matching natural reactions from different organisms (B) or can take advantage of natural enzymatic promiscuity by filling gaps in the host endogenous metabolism (A) or by supplying synthetic substrates to the natural pathway (C). Approaches where enzymes are engineered to accept new substrates can rely on the use of synthetic precursors (D) or endogenous ones (E).

However, there are cases where natural enzymatic promiscuity is not enough to generate the substrate range required to produce the target molecule. In these cases, natural enzymatic activity can be used as a starting point to engineer new substrate specificities. In what is also “precursor-directed biosynthesis”, the enzymes of a natural pathway may be modified to accept structurally diverse analogs of natural substrates, instead of relying only on their natural substrate tolerance (Figure 2.8D). In an example of this, a new calcium-dependent antibiotic (CDA) was produced by modifying the nonribosomal peptide synthetase (NRPS) responsible for the incorporation of glutamate to accept the synthetic 3-methylglutamine (supplied *ex vivo*), resulting in the biosynthesis of a glutamine-containing CDA variant [121]. Another type of approach based on engineered enzymes is the “Evolution of natural enzymes for non-natural biochemical reactions”, where the enzyme is modified to catalyze reactions that did not occur in nature but still based on changing the substrate specificity of a natural enzyme (Figure 2.8E). An example of this is the modification of natural biosynthetic pathway of homoserine in

E. coli, so that it accepts malate instead of aspartate, which goes through the 3-step pathway and is converted in 2,4-dihydroxybutyric acid [122]. The three pathway enzymes were modified by rational engineering of the active site and screening of the mutants generated led to the identification of variants with improved affinity towards malate and its intermediates. The modification of the substrate specificity for the first pathway enzyme is depicted in Figure 2.9A. Yet, substrate engineering alone, while ambitious and with success cases, may still not be enough to broaden the repertoire of biological products so that it can meet the range of compounds that can be generated by chemical synthesis. This is because the number of natural enzymatic mechanisms is limited and thus, there is a need to create “designer enzymes” that can catalyze new reactions that differ substantially from natural ones, a strategy for which there are only a few cases yet described. Among these is the design of an enzyme able to catalyze the formolase reaction, a carboligation of three formate molecules into the central metabolite dihydroxyacetone phosphate, a reaction not seen in nature and that would enable carbon assimilation. This was made possible by taking benzaldehyde lyase as a starting point, since it catalyzes the coupling of two benzaldehyde molecules (benzoin reaction), and going through four rounds of computational design and testing to decrease the rate of benzoin reaction and increasing the rate of formose reaction [123] (Figure 2.9B).

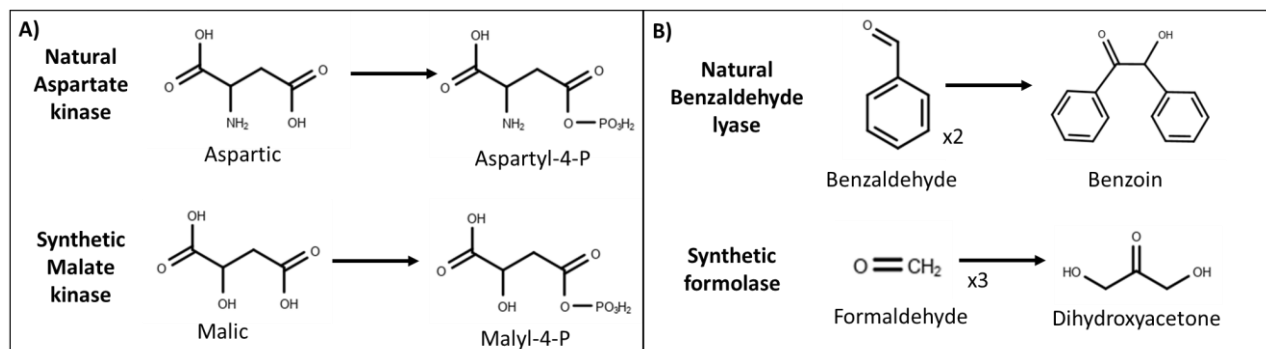


Figure 2.9. Two examples where enzymes were engineered to accept new substrates. A) In [122] aspartate kinase was modified by rational engineering and combinatorial screening to accept malate as a substrate. B) In [123] the benzaldehyde lyase enzyme was modified to accept formaldehyde as the substrate and catalyze two consecutive rounds of carbon-carbon coupling. This was made possible through 4 iterations of rational protein design and screening.

2.3.3. Computational pathway prospection tools

The basis for synthetic pathway prospecting tools strongly relies on enzymatic promiscuity, that is, the ability of enzymes to accept substrates that are not their “preferred” ones but that may be structurally similar. Just in *E. coli* it has been estimated that 37% of enzymes display promiscuous activity [124]. It is this ability to accept related substrates that bioengineers have been taking advantage of to promote desired synthetic reactions using natural enzymes. Synthetic pathway prospecting tools take into account this enzymatic promiscuity and organize all the possible biochemical conversions in a set of reaction rules [125]. These are generalizations of the reported enzymatic reactions described in biological databases such as Kyoto Encyclopedia of Genes and Genomes (KEGG) [99]. A reaction rule is generated by joining together very similar enzymatic reactions and finding the rearrangement of atoms and bonds that is common to all such enzymatic reactions. The bond rearrangement occurring during the reaction is described in the form of a Bond-Electron Matrix (BEM), a SMILES/SMARTS notation, a RDM pattern (defined by the reaction center, R, the matched region M, and the difference region D) or a reaction signature [128]. In this way, a reaction rule may be applied not only to the native substrates of the enzymes that contributed to that rule but also to similar substrates, generating new metabolites. In the example in Figure 2.10, the transamination reactions catalyzed by enzymes belonging to the sub-subclasses 2.6.1 are joined in a general reaction rule in a BEM; this rule may be applied not only to the native substrates (in the example these are phenylpyruvate/glutamine and aspartate/2-oxoglutarate) but also to the non-native substrates (in the example pyruvate/7,8-diaminonanoate).

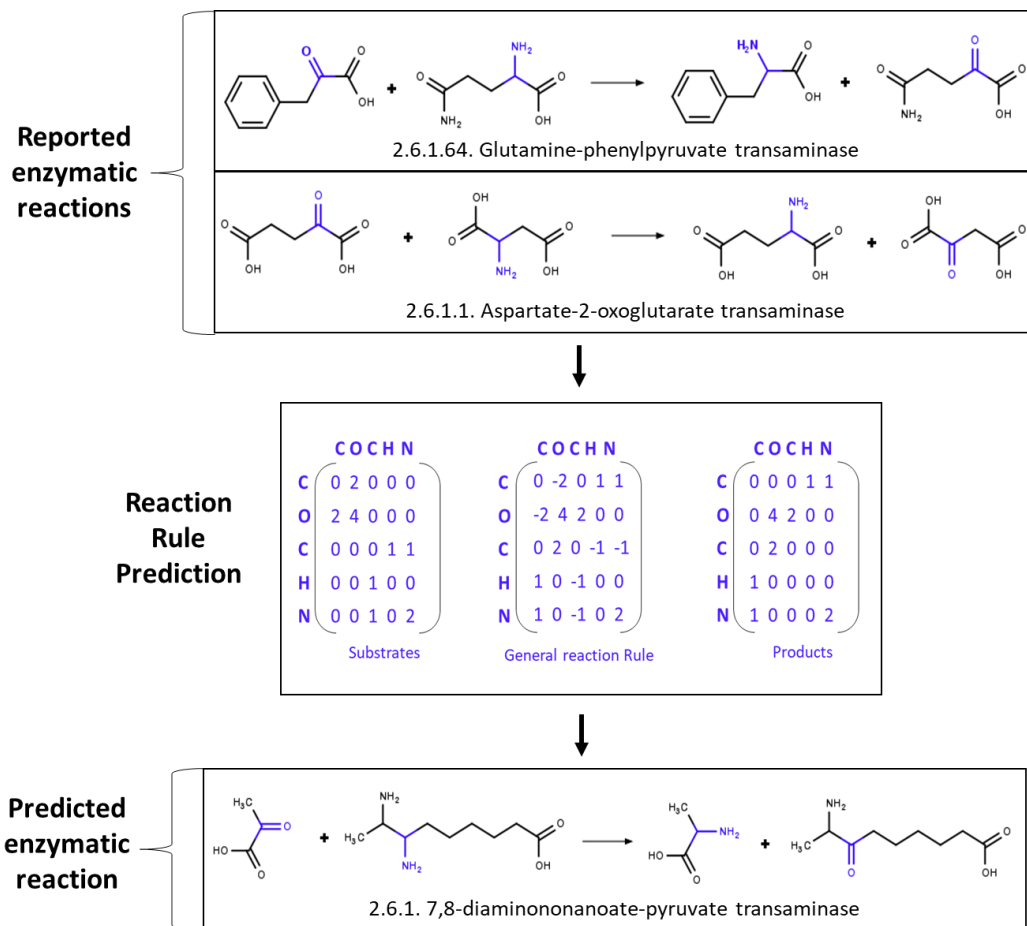


Figure 2.10. Example of how reported enzymatic reactions contribute to the creation of reaction rule (here in BEM format, example from BNICE [130].), which enables the prediction of enzymatic reactions. The reported enzymatic reactions are of glutamine and phenylpyruvate transamination to glutamate and phenylalanine and of aspartate and 2-oxoglutarate transamination to oxaloacetate and glutamate. In the predicted enzymatic reaction pyruvate and 7,8-diaminononanoate are transaminated to alanine and 7-keto-8-aminononanoate.

The first application of the concept of generalized reaction rules resulted in the development of BNICE, the most explored pathway prospecting tool until now [129], [130]. In this pioneering work the tool was used to study alternative aromatic amino acid biosynthetic pathways and the application of reaction rules to the precursor chorismate predicted 246 possible compounds in the phenylalanine pathway, 289 in the tyrosine pathway and 58 in the tryptophan pathway. With the newly generated biosynthetic

pathways it was possible to study the thermodynamics of each one as well as their length, including the uncovering of 7, 23 and 15 alternative pathways for the biosynthesis of phenylalanine, tyrosine and tryptophan, which, although less thermodynamically favorable than the native pathways, have the same number of reaction steps and may be of biotechnological interest in different production scenarios. The development of BNICE would lay the ground for many other pathway prospecting tools to follow, which can be said to have a more or less common way of working in 3 steps: 1) establishing the set of reaction rules; 2) application of the selected reaction rules to the selected molecule of choice and identify possible pathways leading to it/derived from it; 3) ranking (Figure 2.11).

The first step is the establishment of the reaction rules set, which, as said above, is compiled from a repository of reported enzymatic reactions described in public databases such as KEGG [126], BRENDA [127], MetaCyc [131], BIGG [132] or MetRxn[133]. The way the reactions rules are created will influence the number and quality of the predicted pathways. In this sense, one important parameter is how much of the molecule besides the transformation center is considered in the reaction rule. This may be defined in terms of “reaction rule degree” (in the case of ReactPred [134]), “distance” (in the case of RetroPath [128]) or “moiety size” (in the case of RePrime [135]). For any of the approaches/tools, specificity is increased (and consequently, less pathways are predicted) as more parts of the molecule are considered, besides just the transformation center. Recently a repository of reaction rules, RetroRules, compiled from the reaction databases BNICE, Simpheny, KEGG, Reactome, Rhea, MetaCyc, was made available, thus creating a centralized resource to be used by other tools [136].

Upon definition of the reaction rules set, it is possible to apply them to the target compound in a backwards manner to find all possible reactions that lead to its production. The way this is done defines the search algorithm, which can be categorized as heuristic, retrosynthetic or a double direction pathway search [137]. Heuristic searches approach the problem by searching the graph representation of the network layer by layer (breadth first) or path by path (depth first) [137]. Retrosynthesis searches work very simply by applying the reaction rules backwards to the target compound n -times, where n is the number of reaction steps allowed for the pathway; this is the most common search method. The double direction pathway search combines retrosynthesis and heuristic search from the substrate side. All the methodologies suffer from combinatorial explosion that is controlled by the number of reaction rules, with retrosynthetic searches creating the maximal number of combinations [128]. Some tools manage to decrease the

biochemical solution space by using filtering criteria to remove unfeasible pathways from the network during the search. Sympheny does this by setting a bound on molecule size along the biochemical pathway, while GEM-PATH verifies in every iteration (reaction step) of the search if the predicted reactions are thermodynamically feasible and if they have a candidate enzyme to catalyze them, based on a promiscuity analysis [97],[138]. Retropath applies a similar methodology by scoring every reaction according to the likelihood of finding an enzyme to catalyze the reaction [128]. *Pertusi et al, 2015* decreased the computational space by using a similarity measure (SimIndex) to guide search, thus prioritizing nodes that are more similar to the target product [139]. More recent algorithmic solutions to this problem were presented by RePrime, which solves a MILP optimization problem in every iteration, and RetroPath RL, which applies a Monte Carlo Tree Search [125], [135]. RetSynth also solves a MILP problem based on FBA but here the objective is the minimization of genetic interventions [140].

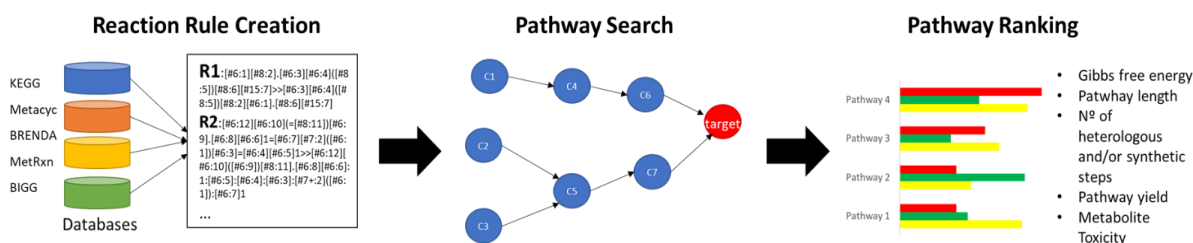


Figure 2.11: Typical procedure followed by a synthetic pathway prospecting tool. In the first stage enzymatic data from various databases is compiled to create reaction rules (in this image an example from ReactPred reaction rules in SMART format). In the second stage the application of the reaction rules enables the search for possible pathways, which are ranked in the third stage.

The last step of pathway prospecting tools comprises ranking all candidate pathways. Some tools have been developed to generate a large number of possible pathways and the user can decide its own ranking and filtering criteria, while others have them already incorporated, such as Sympheny, Retropath and Reactpred [128], [134], [141]. The most common criteria include pathway length, number of synthetic reactions and thermodynamic feasibility, computed by the route Gibbs free energy. Additionally, Genome Scale Metabolic models (GEMs) have been used in the ranking step, to calculate the maximal theoretical yield of the computed pathways and also determine if the synthetic pathways can be growth-coupled via deletion knockouts [141], [142].

Table 2.3: Overview of most known *new-to-nature* pathway prospecting tools. Many tools don't perform ranking of the predicted pathways (indicated by N/A); however, this may be done afterwards, as is the case for many BNICE applications. The reaction prediction field details how a reaction is predicted to be catalyzed by a generic reaction rule or a specific enzyme. This information is not known for Sympheny, depicted by ?.

Tool	Reaction rules	Ranking of pathways	Reaction prediction	Comments	Ref	Year
UM-BBD	From literature on functional group transformations	N/A	Specific enzyme	Applied to xenobiotic degradation; only allows KEGG compounds	[143]	2004
BNICE	From KEGG, BEM representation at the 3 rd EC level	N/A	Rule	-	[130]	2005
Biopath finder	Generated from all known classes of enzymes that operate on carbohydrates	Pathway length	Rule	Only applied to central carbon metabolism	[144]	2010
PathPred	RDM patterns from KEGG RPAIR	Similarity of synthetic substrates to native ones	KEGG Reaction	Applied to xenobiotic degradation, only allows KEGG compounds	[145]	2010
Cho et al 2010	From KEGG	Binding site covalence, chemical similarity, thermodynamics, pathway length, and organism specificity	Full EC number	-	[146]	2010
Simpheny	From general E.C. information	Yield, pathway length, number of non-native steps, number of novel steps and thermodynamics	?	Integration with GEMs	[147]	2011
GEM-Path	From general E.C. information. SMILES/SMARTS	Theoretical productivity	Full EC number	Integration w/ GEMs	[142]	2014
XTMS	Molecular signatures	Enzyme efficiency,	Similar	Search tool is	[148]	2014

Chapter 2– State of the Art

		toxicity, yield and thermodynamics	reactions	RetroPath		
SimIndex and SimZyme	Chemical fingerprints and BNICE rules	Similarity measure	Full EC number	Algorithmic improvement by Byers -Waterman search	[139]	2015
M-Path	Reaction feature vectors based on KEGG RPAIR	Similarity of synthetic substrates to native ones	Similar KEGG reactions	Only allows KEGG compounds (in the demo server)	[149]	2015
ReactPred	From Metacyc SMILES/SMARTS	Pathway length, thermodynamics, molecular size and substructure	Reaction rule	Reaction rules may be generated automatically from other libraries	[134]	2016
RetroPath 2.0	From Metacyc, SMILES/SMARTS	Can be done using an external tool (RP2paths)	Full EC number	-	[128]	2018
RING	From KEGG (small subset)	Pathway length	Rule	Created small manual networks	[150]	2018
RePrime	From metRxn, based on molecular signatures and metabolite moieties	Implicit constraints in the MILP optimization problem	Enzyme ID	Integration w/ GEMs	[135]	2018
RetSynth	RetroRules	The min of number of genes added is the MILP objective	Full EC number	Integration w/ GEMs	[140]	2019
Transfor-MinER	From KEGG based on reaction center (RC) and the molecular environment (RCME)	Similarity score	Similar KEGG reactions	-	[151]	2019
RetroPath RL	From RetroRules database, accepts user reaction rules	Separate procedure	EC number	-	[125]	2020

It is noteworthy that some of the pathway prospecting tools reviewed in Table 2.3 have been developed for purposes other than the design of non-natural compounds. This is the case of PathPred or UM-BBD, whose reaction rules are based on xenobiotic degradation, but can still be useful to create new-to-nature pathways leading to synthetic products [145], [143]. Additionally, these tools do not allow for the formation of compounds that aren't present in KEGG, thus conditioning their use for the design of

pathways leading only to molecules that are already present in some metabolic reaction.

In addition to pathway prospecting tools with the typical pipeline of Figure 2.11, two databases of predicted enzymatic conversions are now published, Atlas of Biochemistry and the Metabolic In silico Network Expansions (MINE). Atlas of Biochemistry was obtained by applying the BNICE reaction rules that connect two or more compounds in the KEGG database, thus obtaining the “*whole theoretical reactome from the known metabolome*” [152]. Complementary to Atlas, MINE database was created by applying the same BNICE reaction rules to KEGG compounds, allowing for the formation of molecules that haven’t been reported to occur *in vivo* yet, but that are likely to exist in a cell [153].

It is also relevant to mention other tools that can be useful in the design of new metabolic pathways, even if not following the entire procedure of Figure 2.10. In what concerns the choice of a possible precursor for a target molecule, PrecursorFinder can be used to conduct a search based on chemical similarity and maximum common substructure (MCS) in a manually curated database of 60 000 biosynthesized chemicals is used [154]. When a pathway is already envisioned and the task is choosing particular enzymes to catalyze the desired reactions, Selenzyme, a free online tool, may be of use [155]. The Bridgit method, developed to assign enzymes to orphan reactions, can also be used to select an enzyme based on substrate reactive sites [156].

2.3.4. Applications of pathway prospecting tools

While many tools have been developed to guide the design of new-to-nature pathways, a few cases have been published concerning their application to particular target molecules. BNICE has been applied to select metabolic routes for the production of the 3-hydroxypropionic acid, monoethylene glycol (MEG) and methyl ethyl ketone (MEK). In the case of 3-HPO, while there are already proposed biochemicals routes for its production, they involve the consumption of ATP, decreasing the yield. To address this problem, *Christopher et al, 2010* used BNICE to evaluate potential new 3-HPO routes, finding one alternative route that employs less enzymatic steps than the original one (2 compared to 4) and matches the maximal yield of the pathway currently used in commercial production of 3HP [157]. BNICE has also been used to assess the potential of devising a microbial process to convert syngas (mixture of hydrogen and carbon monoxide and dioxide) in MEG, and the potential pathways were evaluated according to

pathway length, maximal theoretical yield, thermodynamic feasibility [158]. Two potential MEG producing pathways were identified, both using glycolaldehyde as an intermediate with one starting from glycerate (obtained from gluconeogenesis) and the other one from ethanol (obtained from glycolysis). A similar strategy was applied to find alternative routes leading to MEK and to 5 of its derivatives whose production could be enabled in *E. coli* [159]. In this case, the large number of candidate pathways was clustered to find common patterns, enabling the identification of 11 core MEK precursors and common routes for the production of the targets.

The utility of RetroPath has been validated experimentally by assembly in *E. coli* of a new pathway leading to the plant flavonoid pinocembrin [160]. The candidate pathways were scored according to predicted enzyme performance in the host, toxicity of pathway metabolites and steady-state production fluxes (based on FBA) and one route was selected to be implemented *in vivo* through 12 possible enzyme combinations corresponding to the reaction steps in the pathway, enabling the selection of 2 enzyme combinations that results in pinocembrin production. Subsequent network-level optimization to improve the pool of the precursor malonyl-coA enabled the production the flavonoid. Most importantly, the accuracy of enzyme score predictions were confirmed by comparing the accumulation of the pathway intermediate trans-cinnamate with their predicted scores. These four examples are focused on molecules that have been already produced in a microbe (natural, in the case of 3-HPO or engineered, in the case of MEG, MEK and pinocembrin) and the task is mainly focused on finding alternative pathways from more desirable carbon sources or with higher yield. In opposition to these cases, Sympheny was designed to address the problem of 1,4-butanediol (1,4-BDO) production, a compound that is not described in any known metabolic route. One promising candidate pathway was selected from those predicted by Sympheny and the route was divided in two modules to validate experimentally, with the upstream module being dedicated to the conversion of succinate to 4-hydroxybutyrate and the downstream one to convert this precursor in 1,4-BDO, in the end obtaining a fully functional 1,4-BDO biosynthetic pathway in *E. coli* [147], that is now used at the industrial scale (Genomatica). This is a paradigmatic example of how pathway prospecting tools can aid in the task of engineering a microbial host to produce a new-to-nature compound and bring the process to the industrial level, presenting a promising sustainable alternative to fuel-derived molecules. This thesis is framed in this context and is outlined in the following chapter.

2.4. References

- [1] S. Alonso, M. Rendueles, and M. Díaz, “Microbial production of specialty organic acids from renewable and waste materials,” *Crit. Rev. Biotechnol.*, vol. 35, no. 4, pp. 497–513, 2015, doi: 10.3109/07388551.2014.904269.
- [2] F. Cherubini, “The biorefinery concept: Using biomass instead of oil for producing energy and chemicals,” *Energy Convers. Manag.*, vol. 51, no. 7, pp. 1412–1421, 2010, doi: 10.1016/j.enconman.2010.01.015.
- [3] V. Balan, “Current challenges in commercially producing biofuels from lignocellulosic biomass,” *ISRN Biotechnol.*, vol. 2014, p. 463074, Jan. 2014, doi: 10.1155/2014/463074.
- [4] R. Van Ree, J. Sanders, R. Bakker, R. Blaauw, R. Zwart, and B. Van Der Drift, “Biofuel-driven biorefineries for the co-production of transportation fuels and added-value products,” in *Handbook of Biofuels Production: Processes and Technologies*, Woodhead Publishing Limited, 2011, pp. 559–580.
- [5] BIO-TIC, “Overcoming hurdles for innovation in industrial biotechnology in Europe - Biobased Chemical Building Blocks,” 2014.
- [6] T. Willke and K.-D. Vorlop, “Industrial bioconversion of renewable resources as an alternative to conventional chemistry,” *Appl. Microbiol. Biotechnol.*, vol. 66, no. 2, pp. 131–42, Dec. 2004, doi: 10.1007/s00253-004-1733-0.
- [7] J. Lindorfer, M. Lettner, K. Fazeni, D. Rosenfeld, and Bert Ann, “Technical , Economic and Environmental Assessment of Biorefinery Concepts,” 2019.
- [8] J. J. Bozell and G. R. Petersen, “Technology development for the production of biobased products from biorefinery carbohydrates—the US Department of Energy’s ‘Top 10’ revisited,” *Green Chem.*, vol. 12, no. 4, p. 539, 2010, doi: 10.1039/b922014c.
- [9] P. C.J., P. J. Westfall, D. J. Pitera, and K. Benjamin, “High-level semi-synthetic production of the potent antimalarial artemisinin,” *Nature*, vol. 496, 2013, doi: 10.1038/nature12051.
- [10] T. Werpy and G. Petersen, “Top Value Added Chemicals from Biomass,” 2004.
- [11] D. a Abbott, R. M. Zelle, J. T. Pronk, and A. J. a van Maris, “Metabolic engineering of *Saccharomyces cerevisiae* for production of carboxylic acids: current status and challenges,” *FEMS Yeast Res.*, vol. 9, no. 8, pp. 1123–36, Dec. 2009, doi: 10.1111/j.1567-1364.2009.00537.x.
- [12] J. Becker and C. Wittmann, “Advanced Biotechnology: Metabolically Engineered Cells for the Bio-Based Production of Chemicals and Fuels, Materials, and Health-Care Products,” *Angew. Chemie Int. Ed.*, vol. 54, no. 11, pp. 3328–3350, Mar. 2015, doi: 10.1002/anie.201409033.
- [13] “Organic Acids Market Analysis & Share | Global Industry Report, 2019,” 2019.
- [14] P. M. Research, “Bio-based Organic Acids Market Size, Share,” 2014.

- [15] A. G. Sandström, H. Almqvist, D. Portugal-Nunes, D. Neves, G. Lidén, and M. F. Gorwa-Grauslund, “Saccharomyces cerevisiae: A potential host for carboxylic acid production from lignocellulosic feedstock?,” *Appl. Microbiol. Biotechnol.*, vol. 98, no. 17, pp. 7299–7318, 2014, doi: 10.1007/s00253-014-5866-5.
- [16] A. Kövilein, C. Kubisch, L. Cai, and K. Ochsenreither, “Malic acid production from renewables: a review,” *J. Chem. Technol. Biotechnol.*, no. October, 2019, doi: 10.1002/jctb.6269.
- [17] C. Pais *et al.*, *Production of Dicarboxylic Acid Platform Chemicals Using Yeasts: Focus on Succinic Acid*. Elsevier Inc., 2016.
- [18] M. McCoy, “Succinic acid, once a biobased chemical star, is barely being made,” *Chemical & Engineering News*, 2019.
- [19] V. Martin-Dominguez, J. Estevez, F. De Borja Ojembarrena, V. E. Santos, and M. Ladero, “Fumaric acid production: A biorefinery perspective,” *Fermentation*, vol. 4, no. 2, 2018, doi: 10.3390/fermentation4020033.
- [20] M. Sajid, X. Zhao, and D. Liu, “Production of 2,5-furandicarboxylic acid (FDCA) from 5-hydroxymethylfurfural (HMF): Recent progress focusing on the chemical-catalytic routes,” *Green Chem.*, vol. 20, no. 24, pp. 5427–5453, 2018, doi: 10.1039/c8gc02680g.
- [21] L. Matsakas, K. Hrůzová, U. Rova, and P. Christakopoulos, “Biological production of 3-hydroxypropionic acid: An update on the current status,” *Fermentation*, vol. 4, no. 1, pp. 1–21, 2018, doi: 10.3390/fermentation4010013.
- [22] Y. Li *et al.*, “Current status on metabolic engineering for the production of L-aspartate family amino acids and derivatives,” *Bioresource Technology*, vol. 245. Elsevier Ltd, pp. 1588–1602, 01-Dec-2017, doi: 10.1016/j.biortech.2017.05.145.
- [23] T. S. Moon, S. H. Yoon, A. M. Lanza, J. D. Roy-Mayhew, and K. L. Jones Prather, “Production of glucaric acid from a synthetic pathway in recombinant Escherichia coli,” *Appl. Environ. Microbiol.*, vol. 75, no. 3, pp. 589–595, Feb. 2009, doi: 10.1128/AEM.00973-08.
- [24] T. Hirasawa and H. Shimizu, “Glutamic acid fermentation: Discovery of glutamic acid-producing microorganisms, analysis of the production mechanism, metabolic engineering, and industrial production process,” *Ind. Biotechnol. Prod. Process.*, pp. 339–360, 2016, doi: 10.1002/9783527807833.
- [25] J. Becker, A. Lange, J. Fabarius, and C. Wittmann, “Top value platform chemicals: Bio-based production of organic acids,” *Curr. Opin. Biotechnol.*, vol. 36, no. Figure 1, pp. 168–175, 2015, doi: 10.1016/j.copbio.2015.08.022.
- [26] Y. Chen and J. Nielsen, “Biobased organic acids production by metabolically engineered microorganisms,” *Curr. Opin. Biotechnol.*, vol. 37, pp. 165–172, 2016, doi: 10.1016/j.copbio.2015.11.004.
- [27] O. V. Singh and R. Kumar, “Biotechnological production of gluconic acid: Future implications,” *Applied Microbiology and Biotechnology*, vol. 75, no. 4. Springer, pp. 713–722, 14-Jun-2007, doi: 10.1007/s00253-007-0851-x.

- [28] K. M. Draths and J. W. Frost, “Environmentally compatible synthesis of adipic acid from D-glucose,” *J. Am. Chem. Soc.*, vol. 116, no. 1, pp. 399–400, Jan. 1994, doi: 10.1021/ja00080a057.
- [29] N. S. Kruyer and P. Peralta-yahya, “Metabolic engineering strategies to bio-adipic acid production,” *Curr. Opin. Biotechnol.*, vol. 45, pp. 136–143, 2017, doi: 10.1016/j.copbio.2017.03.006.
- [30] M. Burk, P. Priti, V. D. Stephen, A. Burgard, C. Schilling, and CHRISTOPHER, “Methods for the synthesis of acrylic acid and derivatives from fumaric acid,” EP2188243, 2008.
- [31] M. D. Lynch, R. T. Gill, and T. E. W. Lipscomb, “Methods for producing 3-hydroxypropionic acid and other products,” US8883464B2, 2009.
- [32] A. Vidra and Á. Németh, “Bio-produced acetic acid: A review,” *Period. Polytech. Chem. Eng.*, vol. 62, no. 3, pp. 245–256, 2018, doi: 10.3311/PPch.11004.
- [33] N. Z. Xie, H. Liang, R. B. Huang, and P. Xu, “Biotechnological production of muconic acid: Current status and future prospects,” *Biotechnol. Adv.*, vol. 32, no. 3, pp. 615–622, 2014, doi: 10.1016/j.biotechadv.2014.04.001.
- [34] R. O. Rajesh, T. K. Godan, R. Sindhu, A. Pandey, and P. Binod, “Bioengineering advancements, innovations and challenges on green synthesis of 2, 5-furan dicarboxylic acid,” *Bioengineered*, vol. 11, no. 1, pp. 19–38, Jan. 2020, doi: 10.1080/21655979.2019.1700093.
- [35] S. Cheong, J. M. Clomburg, and R. Gonzalez, “Energy-and carbon-efficient synthesis of functionalized small molecules in bacteria using non-decarboxylative Claisen condensation reactions,” *Nat. Biotechnol.*, vol. 34, no. 5, pp. 556–561, 2016, doi: 10.1038/nbt.3505.
- [36] Fernando, “(12) United States Patent,” vol. 2, no. 12, 2019.
- [37] O. M. Koivistoinen *et al.*, “Glycolic acid production in the engineered yeasts *Saccharomyces cerevisiae* and *Kluyveromyces lactis*,” *Microb. Cell Fact.*, vol. 12, no. 1, p. 82, Sep. 2013, doi: 10.1186/1475-2859-12-82.
- [38] T. Mehtiö, M. Toivari, M. G. Wiebe, A. Harlin, M. Penttilä, and A. Koivula, “Production and applications of carbohydrate-derived sugar acids as generic biobased chemicals,” *Crit. Rev. Biotechnol.*, vol. 36, no. 5, pp. 904–916, Sep. 2016, doi: 10.3109/07388551.2015.1060189.
- [39] T. Mehtiö, M. Toivari, M. G. Wiebe, A. Harlin, M. Penttilä, and A. Koivula, “Production and applications of carbohydrate-derived sugar acids as generic biobased chemicals,” *Crit. Rev. Biotechnol.*, vol. 36, no. 5, pp. 904–916, 2016, doi: 10.3109/07388551.2015.1060189.
- [40] J. C. De Carvalho, A. I. Magalhães, and C. R. Soccol, “Biobased itaconic acid market and research trends-is it really a promising chemical?,” *Chim. Oggi/Chemistry Today*, vol. 36, no. 4, pp. 56–58, 2018.
- [41] T. Robert and S. Friebe, “Itaconic acid-a versatile building block for renewable

- polyesters with enhanced functionality,” *Green Chem.*, vol. 18, no. 10, pp. 2922–2934, 2016, doi: 10.1039/c6gc00605a.
- [42] A. Kuenz and S. Krull, “Biotechnological production of itaconic acid—things you have to know,” *Appl. Microbiol. Biotechnol.*, vol. 102, no. 9, pp. 3901–3914, 2018, doi: 10.1007/s00253-018-8895-7.
 - [43] H. Hajian, W. Mohtar, and W. Yusoff, “Itaconic Acid Production by Microorganisms: A Review,” *Curr. Res. J. Biol. Sci.*, vol. 7, no. 2, p. 3742, 2015.
 - [44] “Itaconic Acid Market Size, Share, Trends, Opportunities & Forecast,” 2018.
 - [45] Weastra, “Determination of market potential for selected platform chemicals,” *BioConSepT*, 2012.
 - [46] Y. Durant, “Development of Integrated Production of Polyitaconic Acid from Northeast Hardwood Biomass - ITACONIX, LLC,” 2012.
 - [47] W. Bauer, “Methacrylic Acid and Derivatives,” *Ullmann’s Encycl. Ind. Chem.*, 2000, doi: 10.1002/14356007.a16_441.
 - [48] S. L. Tomić, E. H. Suljovrujić, and J. M. Filipović, “Biocompatible and bioadhesive hydrogels based on 2-hydroxyethyl methacrylate, monofunctional poly(alkylene glycol)s and itaconic acid,” *Polym. Bull.*, vol. 57, no. 5, pp. 691–702, Jun. 2006, doi: 10.1007/s00289-006-0606-3.
 - [49] S. L. Tomić, M. M. Mićić, S. N. Dobić, J. M. Filipović, and E. H. Suljovrujić, “Smart poly(2-hydroxyethyl methacrylate/itaconic acid) hydrogels for biomedical application,” *Radiat. Phys. Chem.*, vol. 79, no. 5, pp. 643–649, May 2010, doi: 10.1016/j.radphyschem.2009.11.015.
 - [50] C. Bell and N. A. Peppas, “Poly(methacrylic acid-g-ethylene glycol) hydrogels as pH responsive biomedical materials,” in *Materials Research Society Symposium Proceedings*, 1994, vol. 331, pp. 199–204, doi: 10.1557/proc-331-199.
 - [51] M. J. Darabi Mahboub, J. L. Dubois, F. Cavani, M. Rostamizadeh, and G. S. Patience, “Catalysis for the synthesis of methacrylic acid and methyl methacrylate,” *Chem. Soc. Rev.*, vol. 47, no. 20, pp. 7703–7738, 2018, doi: 10.1039/c8cs00117k.
 - [52] K. Nagai, “New developments in the production of methyl methacrylate,” *Appl. Catal. A Gen.*, vol. 221, no. 1–2, pp. 367–377, 2001, doi: 10.1016/S0926-860X(01)00810-9.
 - [53] A. H. Tullo, “In with the new,” *Chem Engineering news*, pp. 22–23, 2009.
 - [54] X. L. Wang, R. P. Tooze, K. Whiston, and G. R. Eastham, “Process for the carbonylation of ethylene and catalyst systems for use therein,” US6348621, 2002.
 - [55] A. Reports, “Global Methacrylic Acid Market,” 2019.
 - [56] “Itaconic Acid Market Size, Share, Price,” 2015.
 - [57] D. Esposito and M. Antonietti, “Redefining biorefinery: the search for unconventional building blocks for materials,” *Chem. Soc. Rev.*, vol. 44, no. 16,

- p. 5821–5835, 2015, doi: 10.1039/C4CS00368C.
- [58] B. M. Stadler, A. Brandt, A. Kux, H. Beck, and J. G. de Vries, “Properties of Novel Polyesters Made from Renewable 1,4-Pentanediol,” *ChemSusChem*, vol. 13, no. 3, pp. 556–563, Feb. 2020, doi: 10.1002/CSSC.201902988.
 - [59] S. Sansaloni-Pastor, J. Bouilloux, and N. Lange, “The dark side: Photosensitizer prodrugs,” *Pharmaceuticals*, vol. 12, no. 4. MDPI AG, 2019, doi: 10.3390/ph12040148.
 - [60] D. J. Hayes, J. Ross, M. H. B. Hayes, and S. Fitzpatrick, “The Biofine process: production of levulinic acid, furfural and formic acid from lignocellulosic feedstocks,” *Biorefineries-Industrial Process. Prod.*, vol. 1, 2005, doi: 10.1002/9783527619849.
 - [61] V. Isoni, D. Kumbang, P. N. Sharratt, and H. H. Khoo, “Biomass to levulinic acid : A techno-economic analysis and sustainability of biore fi nery processes in Southeast Asia,” *J. Environ. Manage.*, vol. 214, pp. 267–275, 2018, doi: 10.1016/j.jenvman.2018.03.012.
 - [62] J. C. Van Der Waal and E. De Jong, “Avantium Chemicals: The High Potential for the Levulinic Product Tree,” *Ind. Biorenewables A Pract. Viewp.*, pp. 97–120, 2016, doi: 10.1002/9781118843796.ch4.
 - [63] “Products | GFBiochemicals.” [Online]. Available: <http://www.gfbiochemicals.com/products/#existing-products>. [Accessed: 16-Jun-2020].
 - [64] E. Geiser *et al.*, “Ustilago maydis produces itaconic acid via the unusual intermediate trans-aconitate,” *Microb. Biotechnol.*, vol. 9, no. 1, pp. 116–126, Jan. 2016, doi: 10.1111/1751-7915.12329.
 - [65] C. S. K. Reddy and R. P. Singh, “Enhanced production of itaconic acid from corn starch and market refuse fruits by genetically manipulated *Aspergillus terreus* SKR10,” *Bioresour. Technol.*, vol. 85, no. 1, pp. 69–71, Oct. 2002, doi: 10.1016/S0960-8524(02)00075-5.
 - [66] C. L. Strelko *et al.*, “Itaconic acid is a mammalian metabolite induced during macrophage activation,” *J. Am. Chem. Soc.*, vol. 133, no. 41, pp. 16386–9, Oct. 2011, doi: 10.1021/ja2070889.
 - [67] M. G. Steiger, P. J. Punt, A. F. J. Ram, D. Mattanovich, and M. Sauer, “Characterizing MttA as a mitochondrial cis-aconitic acid transporter by metabolic engineering,” *Metab. Eng.*, vol. 35, pp. 95–104, May 2016, doi: 10.1016/j.ymben.2016.02.003.
 - [68] X. Huang, X. Lu, Y. Li, X. Li, and J. J. Li, “Improving itaconic acid production through genetic engineering of an industrial *Aspergillus terreus* strain,” *Microb. Cell Fact.*, vol. 13, no. 1, p. 119, Aug. 2014, doi: 10.1186/s12934-014-0119-y.
 - [69] N. Wierckx, G. Agrimi, P. S. Lübeck, M. G. Steiger, N. P. Mira, and P. J. Punt, “Metabolic specialization in itaconic acid production: a tale of two fungi,” *Curr. Opin. Biotechnol.*, vol. 62, pp. 153–159, 2020, doi: 10.1016/j.copbio.2019.09.014.

- [70] E. Geiser *et al.*, “Genetic and biochemical insights into the itaconate pathway of *Ustilago maydis* enable enhanced production,” *Metab. Eng.*, vol. 38, pp. 427–435, Nov. 2016, doi: 10.1016/j.ymben.2016.10.006.
- [71] G. Tevž, M. Benčina, and M. Legiša, “Enhancing itaconic acid production by *Aspergillus terreus*,” *Appl. Microbiol. Biotechnol.*, vol. 87, no. 5, pp. 1657–1664, Aug. 2010, doi: 10.1007/s00253-010-2642-z.
- [72] S. Krull, A. Hevekerl, A. Kuenz, and U. Prübe, “Process development of itaconic acid production by a natural wild type strain of *Aspergillus terreus* to reach industrially relevant final titers,” *Appl. Microbiol. Biotechnol.*, vol. 101, no. 10, pp. 4063–4072, May 2017, doi: 10.1007/s00253-017-8192-x.
- [73] A. H. Hossain *et al.*, “Rewiring a secondary metabolite pathway towards itaconic acid production in *Aspergillus niger*,” *Microb. Cell Fact.*, vol. 15, no. 1, pp. 1–15, 2016, doi: 10.1186/s12934-016-0527-2.
- [74] A. Li, N. Pfelzer, R. Zuijderwijk, A. Brickwedde, C. Van Zeijl, and P. Punt, “Reduced by-product formation and modified oxygen availability improve itaconic acid production in *Aspergillus niger*,” *Appl. Microbiol. Biotechnol.*, vol. 97, no. 9, pp. 3901–3911, 2013, doi: 10.1007/s00253-012-4684-x.
- [75] M. Blumhoff, M. G. Steiger, D. Mattanovich, and M. Sauer, “Targeting enzymes to the right compartment: Metabolic engineering for itaconic acid production by *Aspergillus niger*,” *Metab. Eng.*, vol. 19, pp. 26–32, 2013, doi: 10.1016/J.YMBEN.2013.05.003.
- [76] H. Hosseinpour Tehrani, A. Tharmasothirajan, E. Track, L. M. Blank, and N. Wierckx, “Engineering the morphology and metabolism of pH tolerant *Ustilago cynodontis* for efficient itaconic acid production,” *Metab. Eng.*, vol. 54, no. April, pp. 293–300, 2019, doi: 10.1016/j.ymben.2019.05.004.
- [77] B. J. Harder, K. Bettenbrock, and S. Klamt, “Model-based metabolic engineering enables high yield itaconic acid production by *Escherichia coli*,” *Metab. Eng.*, vol. 38, pp. 29–37, 2016, doi: 10.1016/j.ymben.2016.05.008.
- [78] J. Blazeck *et al.*, “Metabolic engineering of *Saccharomyces cerevisiae* for itaconic acid production,” *Appl. Microbiol. Biotechnol.*, vol. 98, no. 8155, Jul. 2014, doi: 10.1007/s00253-014-5895-0.
- [79] K. N. T. Tran, S. Somasundaram, G. T. Eom, and S. H. Hong, “Efficient Itaconic acid production via protein–protein scaffold introduction between GltA, AcnA, and CadA in recombinant *Escherichia coli*,” *Biotechnol. Prog.*, vol. 35, no. 3, 2019, doi: 10.1002/btpr.2799.
- [80] Z. Yang, X. Gao, H. Xie, F. Wang, Y. Ren, and D. Wei, “Enhanced itaconic acid production by self-assembly of two biosynthetic enzymes in *Escherichia coli*,” *Biotechnol. Bioeng.*, vol. 114, no. 2, pp. 457–462, 2017, doi: 10.1002/bit.26081.
- [81] A. Otten, M. Brocker, and M. Bott, “Metabolic engineering of *Corynebacterium glutamicum* for the production of itaconate,” *Metab. Eng.*, vol. 30, pp. 156–165, 2015, doi: 10.1016/j.ymben.2015.06.003.

- [82] K. S. Vuoristo *et al.*, “Metabolic engineering of itaconate production in *Escherichia coli*,” *Appl. Microbiol. Biotechnol.*, vol. 99, pp. 221–228, 2014, doi: 10.1007/s00253-014-6092-x.
- [83] H.-G. Jeon, D.-E. Cheong, Y. Han, J. J. Song, and J. H. Choi, “Itaconic acid production from glycerol using *Escherichia coli* harboring a random synonymous codon-substituted 5'-coding region variant of the *cad A* gene,” *Biotechnol. Bioeng.*, vol. 113, no. 7, pp. 1504–1510, Jul. 2016, doi: 10.1002/bit.25914.
- [84] P. Chang, G. S. Chen, H. Y. Chu, K. W. Lu, and C. R. Shen, “Engineering efficient production of itaconic acid from diverse substrates in *Escherichia coli*,” *J. Biotechnol.*, vol. 249, no. March, pp. 73–81, 2017, doi: 10.1016/j.jbiotec.2017.03.026.
- [85] E. M. Young *et al.*, “Iterative algorithm-guided design of massive strain libraries, applied to itaconic acid production in yeast,” *Metab. Eng.*, vol. 48, pp. 33–43, Jul. 2018, doi: 10.1016/j.ymben.2018.05.002.
- [86] J. Blazeck, A. Hill, M. Jamoussi, A. Pan, J. Miller, and H. S. Alper, “Metabolic engineering of *Yarrowia lipolytica* for itaconic acid production,” *Metab. Eng.*, vol. 32, pp. 66–73, 2015, doi: 10.1016/j.ymben.2015.09.005.
- [87] C. Zhao *et al.*, “Enhanced itaconic acid production in *Yarrowia lipolytica* via heterologous expression of a mitochondrial transporter MTT,” *Appl. Microbiol. Biotechnol.*, vol. 103, no. 5, pp. 2181–2192, Mar. 2019, doi: 10.1007/s00253-019-09627-z.
- [88] T. Chin, M. Sano, T. Takahashi, H. Ohara, and Y. Aso, “Photosynthetic production of itaconic acid in *Synechocystis* sp. PCC6803,” *J. Biotechnol.*, vol. 195, pp. 43–45, 2015, doi: 10.1016/j.jbiotec.2014.12.016.
- [89] H. Qi *et al.*, “Engineering a new metabolic pathway for itaconate production in *Pichia stipitis* from xylose,” *Biochem. Eng. J.*, vol. 126, pp. 101–108, 2017, doi: 10.1016/j.bej.2017.06.011.
- [90] C. Zhao, S. Chen, and H. Fang, “Consolidated bioprocessing of lignocellulosic biomass to itaconic acid by metabolically engineering *Neurospora crassa*,” *Appl. Microbiol. Biotechnol.*, vol. 102, no. 22, pp. 9577–9584, Nov. 2018, doi: 10.1007/s00253-018-9362-1.
- [91] J. Lebeau, J. P. Efromson, and M. D. Lynch, “A Review of the Biotechnological Production of Methacrylic Acid,” *Front. Bioeng. Biotechnol.*, vol. 8, no. March, pp. 1–10, 2020, doi: 10.3389/fbioe.2020.00207.
- [92] D. Johnson, G. Eastham, and M. Poliakoff, “Method of producing acrylic and methacrylic acid,” WO2011/077140A2, 2011.
- [93] M. Carlsson *et al.*, “Study of the Sequential Conversion of Citric to Itaconic to Methacrylic Acid in Near-critical and Supercritical Water,” *Ind. Eng. Chem. Res.*, vol. 33, no. 8, pp. 1989–1996, 1994, doi: 10.1021/ie00032a014.
- [94] D. W. Johnson, G. R. Eastham, M. Poliakoff, and T. A. Huddle, “A process for the production of methacrylic acid and its derivatives and polymers produced

therefrom,” EP2643283A1, 18-Nov-2011.

- [95] J. Le Nôtre, S. C. M. Witte-van Dijk, J. van Haveren, E. L. Scott, and J. P. M. Sanders, “Synthesis of Bio-Based Methacrylic Acid by Decarboxylation of Itaconic Acid and Citric Acid Catalyzed by Solid Transition-Metal Catalysts,” *ChemSusChem*, vol. 7, no. 9, pp. 2712–2720, 2014, doi: 10.1002/cssc.201402117.
- [96] A. Bohre, U. Novak, M. Grilc, and B. Likozar, “Synthesis of bio-based methacrylic acid from biomass-derived itaconic acid over barium hexa-aluminate catalyst by selective decarboxylation reaction,” *Mol. Catal.*, vol. 476, no. July, 2019, doi: 10.1016/j.mcat.2019.110520.
- [97] C. Press, “Bacteria like the taste of syngas,” 2013. [Online]. Available: <https://corporate.evonik.com/en/pages/article.aspx?articleId=103071>. [Accessed: 17-Feb-2020].
- [98] T. Haas, Y. Schiemann, D. Przybylski, T. Rohwerder, R. H. Mueller, and H. Harms, “Biotechnological 2-hydroxyisobutyric acid production,” US 2015 / 0290184A1, 2015.
- [99] A. P. Burgard, M. J. Burk, R. E. Osterhout, and P. Pharkya, “Microorganisms for the production of methacrylic acid,” US8241877B2, 2008.
- [100] M. J. Burk, A. P. Burgard, R. E. Osterhout, J. Sun, and P. Pharkya, “Microorganisms for producing methacrylic acid and methacrylate esters and methods related thereto,” US9133487B2, 2015.
- [101] G. R. Eastham, G. Stephens, and A. Yiakoumetti, “Process for the biological production of methacrylic acid and derivatives thereof,” US 2018 / 0171368 A1, 2018.
- [102] R. Weingarten, W. C. Conner, and G. W. Huber, “Production of levulinic acid from cellulose by hydrothermal decomposition combined with aqueous phase dehydration with a solid acid catalyst,” *Energy Environ. Sci.*, vol. 5, no. 6, pp. 7559–7574, 2012, doi: 10.1039/c2ee21593d.
- [103] A. Morone, M. Apte, and R. A. Pandey, “Levulinic acid production from renewable waste resources: Bottlenecks, potential remedies, advancements and applications,” *Renew. Sustain. Energy Rev.*, vol. 51, pp. 548–565, 2015, doi: 10.1016/Zj/rser.2015.06.032.
- [104] F. D. Pileidis and M. M. Titirici, “Levulinic Acid Biorefineries: New Challenges for Efficient Utilization of Biomass,” *ChemSusChem*, vol. 9, no. 6, pp. 562–582, 2016, doi: 10.1002/cssc.201501405.
- [105] A. L. Zanghellini, “Fermentation route for the production of levulinic acid, levulinate esters and valerolactone and derivatives thereof,” US10246727 B2, 2019.
- [106] N. Baldi *et al.*, “Functional expression of a bacterial α -ketoglutarate dehydrogenase in the cytosol of *Saccharomyces cerevisiae*,” *Metab. Eng.*, vol. 56, no. August, pp. 190–197, 2019, doi: 10.1016/j.ymben.2019.10.001.

- [107] M. Cea, F. Cabrera, M. Abanto, F. E. Felissia, M. C. Area, and G. Ciudad, “Strategy for biological co-production of Levulinic Acid and Polyhydroxyalkanoates by using mixed microbial cultures fed with synthetic hemicellulose hydrolysate,” *Bioresour. Technol.*, p. 123323, 2020, doi: 10.1016/j.biortech.2020.123323.
- [108] D. E. Cameron, C. J. Bashor, and J. J. Collins, “A brief history of synthetic biology,” *Nat. Rev. Microbiol.*, vol. 12, no. 5, pp. 381–390, 2014, doi: 10.1038/nrmicro3239.
- [109] A. Casini *et al.*, “A Pressure Test to Make 10 Molecules in 90 Days: External Evaluation of Methods to Engineer Biology,” *J. Am. Chem. Soc.*, vol. 140, no. 12, pp. 4302–4316, Mar. 2018, doi: 10.1021/jacs.7b13292.
- [110] D. Dikicioglu, P. Pir, Z. I. Onsan, K. O. Ulgen, B. Kirdar, and S. G. Oliver, “Integration of Metabolic Modeling and Phenotypic Data in Evaluation and Improvement of Ethanol Production Using Respiration-Deficient Mutants of *Saccharomyces cerevisiae*,” *Appl. Environ. Microbiol.*, vol. 74, no. 18, pp. 5809–5816, 2008, doi: 10.1128/AEM.00009-08.
- [111] S. Galanie, K. Thodey, I. J. Trenchard, M. Filsinger Interrante, and C. D. Smolke, “Complete biosynthesis of opioids in yeast,” *Science (80-.)*, vol. 349, no. 6252, 2015.
- [112] H. Seki, K. Tamura, and T. Muranaka, “Plant-derived isoprenoid sweeteners: Recent progress in biosynthetic gene discovery and perspectives on microbial production,” *Biosci. Biotechnol. Biochem.*, vol. 82, no. 6, pp. 927–934, 2018, doi: 10.1080/09168451.2017.1387514.
- [113] X. Luo *et al.*, “Complete biosynthesis of cannabinoids and their unnatural analogues in yeast,” *Nature*, vol. 567, no. 7746, pp. 123–126, 2019, doi: 10.1038/s41586-019-0978-9.
- [114] T. J. Erb, P. R. Jones, and A. Bar-Even, “Synthetic metabolism: metabolic engineering meets enzyme design,” *Curr. Opin. Chem. Biol.*, vol. 37, pp. 56–62, 2017, doi: 10.1016/j.cbpa.2016.12.023.
- [115] G. S. Hossain *et al.*, “Rewriting the metabolic blueprint: Advances in pathway diversification in microorganisms,” *Frontiers in Microbiology*, vol. 9, no. FEB. Frontiers Media S.A., 12-Feb-2018, doi: 10.3389/fmicb.2018.00155.
- [116] K. L. J. Prather and C. H. Martin, “De novo biosynthetic pathways: rational design of microbial chemical factories,” *Current Opinion in Biotechnology*, vol. 19, no. 5, pp. 468–474, 2008, doi: 10.1016/j.copbio.2008.07.009.
- [117] G. S. Hossain *et al.*, “Rewriting the Metabolic Blueprint: Advances in Pathway Diversification in Microorganisms,” *Front. Microbiol.*, vol. 9, Feb. 2018, doi: 10.3389/fmicb.2018.00155.
- [118] Z. Jin *et al.*, “Engineering *Saccharomyces cerevisiae* to produce odd chain-length fatty alcohols,” *Biotechnol. Bioeng.*, vol. 113, no. 4, pp. 842–851, Apr. 2016, doi: 10.1002/bit.25856.

- [119] T. Maoka, M. Takemura, H. Tokuda, N. Suzuki, and N. Misawa, “4-Ketozeinoxanthin, a novel carotenoid produced in *Escherichia coli* through metabolic engineering using carotenogenic genes of bacterium and liverwort,” *Tetrahedron Lett.*, vol. 55, no. 49, pp. 6708–6710, Dec. 2014, doi: 10.1016/j.tetlet.2014.10.033.
- [120] X. Sun, Y. Lin, Q. Yuan, and Y. Yan, “Precursor-Directed Biosynthesis of 5-Hydroxytryptophan Using Metabolically Engineered *E. coli*,” *ACS Synth. Biol.*, vol. 4, no. 5, pp. 554–558, 2015, doi: 10.1021/sb500303q.
- [121] J. Thirlway *et al.*, “Introduction of a non-natural amino acid into a nonribosomal peptide antibiotic by modification of adenylation domain specificity,” *Angew. Chemie - Int. Ed.*, vol. 51, no. 29, pp. 7181–7184, 2012, doi: 10.1002/anie.201202043.
- [122] T. Walther *et al.*, “Construction of a synthetic metabolic pathway for biosynthesis of the non-natural methionine precursor 2,4-dihydroxybutyric acid,” *Nat. Commun.*, 2017, doi: 10.1038/ncomms15828.
- [123] J. B. Siegel *et al.*, “Computational protein design enables a novel one-carbon assimilation pathway,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 12, pp. 3704–3709, 2015, doi: 10.1073/pnas.1500545112.
- [124] H. Nam *et al.*, “Network context and selection in the evolution to enzyme specificity,” *Science (80-.)*, vol. 337, no. 6098, pp. 1101–1104, Aug. 2012, doi: 10.1126/science.1216861.
- [125] M. Koch, T. Duigou, and J. L. Faulon, “Reinforcement Learning for Bioretrosynthesis,” *ACS Synth. Biol.*, 2020, doi: 10.1021/acssynbio.9b00447.
- [126] M. Kanehisa, “KEGG: Kyoto Encyclopedia of Genes and Genomes,” *Nucleic Acids Res.*, vol. 28, no. 1, pp. 27–30, Jan. 2000, doi: 10.1093/nar/28.1.27.
- [127] L. Jeske, S. Placzek, I. Schomburg, A. Chang, and D. Schomburg, “BRENDA in 2019: a European ELIXIR core data resource,” *Nucleic Acids Res.*, vol. 47, pp. 542–549, Jan. 2019, doi: 10.1093/nar/gky1048.
- [128] B. Delépine, T. Duigou, P. Carbonell, and J. L. Faulon, “RetroPath2.0: A retrosynthesis workflow for metabolic engineers,” *Metab. Eng.*, vol. 45, no. November 2017, pp. 158–170, 2018, doi: 10.1016/j.ymben.2017.12.002.
- [129] C. Li, C. S. Henry, M. D. Jankowski, J. A. Ionita, V. Hatzimanikatis, and L. J. Broadbelt, “Computational discovery of biochemical routes to specialty chemicals,” *Chem. Eng. Sci.*, vol. 59, no. 22–23, pp. 5051–5060, 2004, doi: 10.1016/j.ces.2004.09.021.
- [130] V. Hatzimanikatis, C. Li, J. A. Ionita, C. S. Henry, M. D. Jankowski, and L. J. Broadbelt, “Exploring the diversity of complex metabolic networks,” *Bioinformatics*, vol. 21, no. 8, pp. 1603–1609, 2005, doi: 10.1093/bioinformatics/bti213.
- [131] R. Caspi *et al.*, “The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases,” *Nucleic Acids Res.*, vol.

- 42, pp. 459–471, Jan. 2014, doi: 10.1093/nar/gkt1103.
- [132] J. Schellenberger, J. O. Park, T. M. Conrad, and B. T. Palsson, “BiGG: A Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions,” *BMC Bioinformatics*, vol. 11, Apr. 2010, doi: 10.1186/1471-2105-11-213.
 - [133] A. Kumar, P. F. Suthers, and C. D. Maranas, “MetRxn: A knowledgebase of metabolites and reactions spanning metabolic models and databases,” *BMC Bioinformatics*, vol. 13, no. 1, p. 6, Jan. 2012, doi: 10.1186/1471-2105-13-6.
 - [134] T. V. Sivakumar, V. Giri, J. H. Park, T. Y. Kim, and A. Bhaduri, “ReactPRED: a tool to predict and analyze biochemical reactions,” *Bioinformatics*, vol. 32, no. 22, pp. 3522–3524, 2016, doi: 10.1093/bioinformatics/btw491.
 - [135] A. Kumar, L. Wang, C. Y. Ng, and C. D. Maranas, “Pathway design using de novo steps through uncharted biochemical spaces,” *Nat. Commun.*, vol. 9, no. 1, 2018, doi: 10.1038/s41467-017-02362-x.
 - [136] T. Duigou, M. Du Lac, P. Carbonell, and J. L. Faulon, “Retrorules: A database of reaction rules for engineering biology,” *Nucleic Acids Res.*, vol. 47, no. D1, pp. D1229–D1235, 2019, doi: 10.1093/nar/gky940.
 - [137] A. Biz, S. Proulx, Z. Xu, K. Siddartha, A. Mulet Indrayanti, and R. Mahadevan, “Systems biology based metabolic engineering for non-natural chemicals,” *Biotechnol. Adv.*, vol. 37, no. 6, p. 107379, 2019, doi: 10.1016/j.biotechadv.2019.04.001.
 - [138] M. A. Campodonico, B. A. Andrews, J. A. Asenjo, B. O. Palsson, and A. M. Feist, “Generation of an atlas for commodity chemical production in *Escherichia coli* and a novel pathway prediction algorithm, GEM-Path,” *Metab. Eng.*, vol. 25, pp. 140–158, 2014, doi: 10.1016/j.ymben.2014.07.009.
 - [139] D. A. Pertusi, A. E. Stine, L. J. Broadbelt, and K. E. J. Tyo, “Efficient searching and annotation of metabolic networks using chemical similarity,” *Bioinformatics*, vol. 31, no. 7, pp. 1016–1024, 2015, doi: 10.1093/bioinformatics/btu760.
 - [140] L. S. Whitmore, B. Nguyen, A. Pinar, A. George, and C. M. Hudson, “RetSynth: Determining all optimal and sub-optimal synthetic pathways that facilitate synthesis of target compounds in chassis organisms,” *BMC Bioinformatics*, vol. 20, no. 1, pp. 1–14, 2019, doi: 10.1186/s12859-019-3025-9.
 - [141] Z. Bodor, L. Tompos, A. C. Nechifor, and K. Bodor, “In silico Analysis of 1, 4-butanediol Heterologous Pathway Impact on *Escherichia coli* Metabolism,” no. October, pp. 3448–3455, 2011, doi: 10.37358/RC.19.10.7574.
 - [142] M. A. Campodonico, B. A. Andrews, J. A. Asenjo, B. O. Palsson, and A. M. Feist, “Generation of an atlas for commodity chemical production in *Escherichia coli* and a novel pathway prediction algorithm, GEM-Path,” *Metab. Eng.*, vol. 25, pp. 140–158, 2014, doi: 10.1016/j.ymben.2014.07.009.
 - [143] K. H. Bo, L. B. M. Ellis, and L. P. Wackett, “Encoding microbial metabolic logic: Predicting biodegradation,” *J. Ind. Microbiol. Biotechnol.*, vol. 31, no. 6, pp. 261–

- 272, Jul. 2004, doi: 10.1007/s10295-004-0144-7.
- [144] E. Noor, E. Eden, R. Milo, and U. Alon, “Central Carbon Metabolism as a Minimal Biochemical Walk between Precursors for Biomass and Energy,” *Mol. Cell*, vol. 39, no. 5, pp. 809–820, 2010, doi: 10.1016/j.molcel.2010.08.031.
 - [145] Y. Moriya *et al.*, “PathPred: an enzyme-catalyzed metabolic pathway prediction server,” *Nucleic Acids Res.*, vol. 38, no. Web Server issue, pp. W138–43, Jul. 2010, doi: 10.1093/nar/gkq318.
 - [146] A. Cho, H. Yun, J. H. Park, S. Y. Lee, and S. Park, “Prediction of novel synthetic pathways for the production of desired chemicals,” *BMC Syst. Biol.*, vol. 4, 2010, doi: 10.1186/1752-0509-4-35.
 - [147] H. Yim *et al.*, “Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol,” *Nat. Chem. Biol.*, vol. 7, no. 7, pp. 445–452, 2011, doi: 10.1038/nchembio.580.
 - [148] B. Delépine, P. Carbonell, and J. L. Faulon, “XTMS in action: Retrosynthetic design in the extended metabolic space of heterologous pathways for high-value compounds,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8859, pp. 256–259, 2014, doi: 10.1007/978-3-319-12982-2_21.
 - [149] M. Araki *et al.*, “M-path: A compass for navigating potential metabolic pathways,” *Bioinformatics*, vol. 31, no. 6, pp. 905–911, 2015, doi: 10.1093/bioinformatics/btu750.
 - [150] U. Gupta, T. Le, W. S. Hu, A. Bhan, and P. Daoutidis, “Automated network generation and analysis of biochemical reaction pathways using RING,” *Metab. Eng.*, vol. 49, no. March, pp. 84–93, 2018, doi: 10.1016/j.ymben.2018.07.009.
 - [151] J. D. Tyzack, A. J. M. Ribeiro, N. Borkakoti, and J. M. Thornton, “Exploring Chemical Biosynthetic Design Space with Transform-MinER,” *ACS Synth. Biol.*, vol. 8, pp. 2494–2506, 2019, doi: 10.1021/acssynbio.9b00105.
 - [152] N. Hadadi, J. Hafner, A. Shajkofci, A. Zisaki, and V. Hatzimanikatis, “ATLAS of Biochemistry: A Repository of All Possible Biochemical Reactions for Synthetic Biology and Metabolic Engineering Studies,” *ACS Synth. Biol.*, vol. 5, no. 10, pp. 1155–1166, 2016, doi: 10.1021/acssynbio.6b00054.
 - [153] J. G. Jeffries *et al.*, “MINEs: open access databases of computationally predicted enzyme promiscuity products for untargeted metabolomics,” *J. Cheminform.*, vol. 7, no. 1, p. 44, Dec. 2015, doi: 10.1186/s13321-015-0087-1.
 - [154] L. Yuan *et al.*, “PrecursorFinder: A customized biosynthetic precursor explorer,” *Bioinformatics*, vol. 35, no. 9, pp. 1603–1604, 2019, doi: 10.1093/bioinformatics/bty838.
 - [155] P. Carbonell *et al.*, “Selenzyme: Enzyme selection tool for pathway design,” *Bioinformatics*, vol. 34, no. 12, pp. 2153–2154, 2018, doi: 10.1093/bioinformatics/bty065.

- [156] N. Hadadi, H. MohammadiPeyhani, L. Miskovic, M. Seijo, and V. Hatzimanikatis, “Enzyme annotation for orphan and novel reactions using knowledge of substrate reactive sites,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 116, no. 15, pp. 7298–7307, 2019, doi: 10.1073/pnas.1818877116.
- [157] C. S. Henry, L. J. Broadbelt, and V. Hatzimanikatis, “Discovery and analysis of novel metabolic pathways for the biosynthesis of industrial chemicals: 3-hydroxypropanoate,” *Biotechnol. Bioeng.*, vol. 106, no. 3, pp. 462–473, 2010, doi: 10.1002/bit.22673.
- [158] M. A. Islam, N. Hadadi, M. Ataman, V. Hatzimanikatis, and G. Stephanopoulos, “Exploring biochemical pathways for mono-ethylene glycol (MEG) synthesis from synthesis gas,” *Metab. Eng.*, vol. 41, pp. 173–181, 2017, doi: 10.1016/j.ymben.2017.04.005.
- [159] M. Tokic *et al.*, “Discovery and Evaluation of Biosynthetic Pathways for the Production of Five Methyl Ethyl Ketone Precursors,” *ACS Synth. Biol.*, vol. 7, no. 8, pp. 1858–1873, 2018, doi: 10.1021/acssynbio.8b00049.
- [160] T. Fehér *et al.*, “Validation of RetroPath, a computer-aided design tool for metabolic pathway engineering,” *Biotechnol. J.*, vol. 9, no. 11, pp. 1446–1457, 2014, doi: 10.1002/biot.201400055.

Chapter 3 – *In silico* prospecting of metabolic pathways for the implementation of microbe-based production of levulinic acid

This chapter has been submitted for publication:

Vila-Santa, A.; Islam, A.; Ferreira, F. C.; Prather, K. L. J.; Mira, N. P. “In silico prospecting of biochemical pathways for microbe-based production of levulinic acid” (submitted to ACS Synthetic Biology).

3.1. Abstract

Synthetic biology has opened the possibility of full-scale design of biological systems through the assembly of new-to-nature pathways, i.e., biochemical pathways that explore new combinations of enzymes (usually not present *in vivo*) to produce value-added chemicals. Levulinic acid (LA) is a platform molecule with potential to be used as an intermediate in the synthesis of many value-added products with different applications, ranging from cosmetics to fuels. While there were two instances where LA was produced in a biological system, there is to date no in-depth analysis of the metabolic pathways that may lead to its production. Resorting to a combined approach involving different complimentary computational tools and extensive manual curation for prospecting of possible biosynthetic pathways for LA production in *E. coli* or *S. cerevisiae*, five candidate pathways were assembled and their performance in the selected hosts was compared using genome-scale metabolic modelling. These analyses identified 3-oxoadipic acid as the most amenable substrate for LA production, but a possible use of D-aminolevulinic acid or glutamate semialdehyde was also uncovered for the first time. Not only does the herein described approach offer a platform for future implementation of microbial production of LA, but it also provides an organized research strategy that can be used as a framework for the implementation of other new-to-nature biosynthetic pathways for the production of value-added chemicals, thus fostering the emerging field of synthetic industrial biotechnology.

3.2.Introduction

The utilization of fossil fuels to meet the energetic demands of modern societies (e.g., in transportation or in commodity production) is an increasing concern as these resources are becoming scarce, and their extensive use is associated with great environmental concerns of global warming and climate change. In this context, efforts have been made towards finding more sustainable alternatives to obtain products that are currently manufactured by the petrochemical industry. The exploration of “green” molecules (i.e., sustainable produced, e.g., via microbial fermentation from renewable feed stocks) to be used as building blocks in synthetic production routes, that presently use fossil fuel derivatives, is one of the most promising approaches to successfully transition from a petrochemical to a bio-based economy [1]. Among the “green molecules” considered, carboxylic acids have emerged as the most promising ones to replace fossil oil derivatives. These compounds are ubiquitous across microbial metabolism (and therefore can be produced via fermentation), and have multiple chemical groups with relevant properties for use as reactants in chemical synthesis [2]. This study is focused on levulinic acid (LA), which has been consistently pinpointed as having a large market, of 2.5 kT in 2012 at 5 USD/Kg [3] and expected to expand to 3.8 kT by this year [4]. The possible applications of LA include its transformation into polycarbonates or plastics, replacing currently used toxic intermediates such as bisphenol or phthalates [5], [6], as well as in renewable jet fuels, epoxy resins, herbicides, pharmaceuticals or flavouring agents [7].

LA can be chemically obtained from furfuryl alcohol or through the reaction of maleic anhydride with acetone under acidic conditions [8]. The main bottleneck of these routes is their dependency on fossil oil, which not only has a negative impact in terms of pollutant emissions but also contributes to an increased product price due to volatility of feedstock prices [9]. A more environmentally friendly alternative to obtain LA involves the acid-catalysed dehydration of pentoses, hexoses, or polysaccharides via a process known as Biofine [9][3]. In this process, lignocellulosic biomass such as cellulose-rich paper sludge is treated with dilute sulfuric acid, resulting in the production of LA as the main product, and furfural and formic acid as co-products [5]. Although this process is promising, there are recognized challenges such as the recovery of the acid catalyst, the complications associated with downstream processing to purify LA, and the high energy and water requirements [10]. The Biofine process also results in the production of formic acid from the C6 sugars and furfural from the C5 sugars, creating the challenge of how to

valorise these co-products in the waste stream of the process [3], [5]. In this context, the implementation of a microbe-based pathway for the synthesis of LA from fermentable sugars would be advantageous; however, no microbial species has been described to naturally produce this acid to date.

Recent advances in synthetic biology have rendered possible the production of a panoply of molecules not naturally produced by microbes, resorting to what is generally known as “new-to-nature” pathways [11]. Paradigmatic examples of these synthetic pathways include the implementation of odd-chain fatty alcohols production in *S. cerevisiae*, or the production of 2,4-dihydroxybutyric acid and chiral organoboranes in *E. coli* [12]–[14]. Successful production of these “new-to-nature” compounds in a selected host of choice is largely dependent on the identification of suitable precursor molecules and enzymes that could implement the identified steps.

The identification of promising enzyme-substrate matches can be achieved using computational tools such as Biochemical Network Integrated Computational Explorer (BNICE) [15]. BNICE uses the KEGG database [16] to create a set of generalised reaction rules based on mapping the biologically relevant sites of each biotransformation, whereby atoms and bonds of the substrate(s) rearrange to form the product [15]. The resulting set of reaction rules can then be applied in a retrosynthetic manner to identify which reactions and substrates could lead to a given molecule of interest. Using this approach, in silico prospecting of pathways for the production of value-added chemicals, including monoethylene glycol, methyl ethyl ketone, and 3-hydroxypropanoate have been proposed [17]–[19]. In addition, a repository of the “whole theoretical reactome from the known metabolome” was established in the Atlas of Biochemistry database (ATLAS) [20] using BNICE. However, a major limitation of this repository is that it includes metabolites that are present only in KEGG, and therefore, its use for the prospecting of pathways to produce LA or other chemicals not found in KEGG is limited. Several other computational tools and databases including the Metabolic In silico Network Expansions (MINE) database have been developed to overcome this bottleneck. MINE extends the possibility of analysing molecules beyond those that are not present in KEGG, but whose occurrence is considered possible based on the BNICE-generated reaction rules and also on expert curated-analysis of the enzymatic reactions predicted in the enzyme classification system [21]. Other examples include ReactPRED [22], Retropath [23], and Sympheny Biopath Predictor [24] that differ from BNICE on the methodologies used to create the set of reaction rules, as well as used for enzyme assignment and pathway

ranking. The exploration of Sympheny enabled the successful implementation of a biocatalytic route in *E. coli* for the conversion of renewable carbohydrates to 1,4-butanediol, a “new-to-nature” commodity chemical used in the synthesis of polymers [25].

Following the path of previous studies that have successfully explored computational tools for in silico prospecting of “new-to-nature” pathways for the production of value-added chemicals [17]–[19], [25], we aimed at identifying routes for the production of LA from carbohydrates. The synthesis of LA from glycerol and succinate in *E. coli* has been shown in a study that focused on the development of an orthogonal platform to promote non-decarboxylative Claisen condensation reactions followed by corresponding β -reductions [26]. Although this example clearly confirms the feasibility of LA production in a microbial host, several bottlenecks were identified: the obtained low product titers (in the range of 50-150 mg/L, depending on the decarboxylase used for the last step) and the need to have a continuous supplementation of succinate as a co-substrate, which constrains the economic viability of the process. Furthermore, in eukaryotic microbial hosts, although more favorable for the production of organic acids [27], the synthesis of succinyl-CoA takes place in mitochondria, which can create compartmentalization problems. This is because the production of LA would either have to take place in mitochondria (requiring an additional transport step between mitochondria and cytosol), or a supply of cytosolic succinyl-coA and lipoic acid, the cofactor of α -ketoglutarate dehydrogenases, would have to be engineered [28]. More recently, a study reported the co-production of LA with poly-hydroxyalkanoates from hemicellulose hydrolysates, using a mixed microbial culture [29]. However, the metabolic route leading to LA formation in this setting was not elucidated.

To identify the most feasible alternative biosynthetic pathways for the production of LA, we used BNICE-related tools and ReactPRED [22] in this study. In addition to identify suitable precursors, the search of specific enzymes that could catalyze the selected reactions was also performed. Aiming for a stronger integration in chosen hosts, we have also performed a downstream pathway yield analysis, using constraint-based metabolic modeling and identification of reactions in the metabolism that are limiting to LA production. *S. cerevisiae* and *E. coli* were selected as hosts due to their central role as experimental models and microbial chassis in industrial biotechnology; specifically, for the production of carboxylic acids [27], [30]. Altogether our study provides not only an integrated view of the best performing pathways to produce LA from renewable sources,

but also a framework for analyzing pathway design strategies that can be implemented for microbe-based production of value-added compounds that fall outside the natural microbial metabolic repertoire.

3.3.Methods

Computational prospecting of LA production pathways. The pathway prospecting procedure includes three steps: the search for possible precursors, enzyme assignment, and estimation of the pathway yield by metabolic modelling. The details of the computational workflow performed in different steps are described below:

3.3.1. Precursor search

The search for possible precursors for LA synthesis was performed using either a chemical structure similarity to LA approach or retrobiosynthetic tools. For the similarity search approach, all KEGG compounds were compared with LA using the SIMCOMP tool [31] and those having a Tanimoto coefficient (a popular metric for the similarity of two molecules [32]) above 0.8 were considered as possible candidates. For retrobiosynthetic searches, MINE and ReactPRED were used. In the case of MINE, all reactions leading to the production of LA were considered. For ReactPRED, the search was performed using a maximum of 2 possible irreversible reaction steps, a prediction tolerance of 0.5, and reaction rule degree of 1. The output file with the total set of predicted pathways was pruned using MATLAB to apply the following selection filters: i) reactions with a negative Gibbs free energy; ii) pathways involving substrate and/or co-substrate in PubChem database; iii) pathways with all substrates found in KEGG; iv) pathways with a correct reaction stoichiometry. The application of these filters resulted in a final set of candidate pathways that was further manually curated to remove repeated or highly similar pathways, check feasibility of the predicted reactions (i.e., if they are biochemically possible), and assess whether bio-based production of the selected precursors was feasible.

3.3.2.Enzyme assignment

The compiled list of possible LA precursors was searched in the MINE and ATLAS databases to find the possible conversions (reactions) resulting in the transformation of these precursors into LA, as well as the associated reaction rule which is defined at the third level of the E.C. classification system (e.g., the reaction rule 4.1.1.- entails a decarboxylation is a decarboxylation). For each predicted enzymatic conversion, the E.C. number information was manually mined in the BRENDA database [33] to find enzyme(s) involving substrates similar to those in the predicted synthetic reactions. Relevant literature was also searched to verify the reaction mechanism of each candidate enzyme, as well as their activities in a species-specific context. The Maximal Common Substructure Tanimoto coefficient between native and synthetic substrates was calculated with the ChemMine Similarity workbench using the PubChem fingerprints [34].

3.3.3.Pathway yield analysis using in silico metabolic modelling

All simulations were performed using the Constraint-Based Reconstruction and Analysis (COBRA) MATLAB Toolbox v3.0 [35], the Gurobi Solver [36], and the genome-scale Yeast model 5.05 [37] and E. coli model iJO1366 [38], with all MATLAB scripts and models used are provided in the SI2. The E. coli model was constrained to simulate continuous aerobic growth using data from [39]. Specifically, the reactions “catalase” (CAT), “dihydropteridine reductase” (DHPTDN), “dihydropteridine reductase (NADH)” (DHPTDNRN), “Formate-hydrogen lyase” (FHL), “superoxide dismutase” (SPODM), “succinate: aspartate antiporter (periplasm)” (SUCASPtp), “succinate: fumarate antiporter (periplasm)” (SUCFUMtp), “succinate: malate antiporter (periplasm)” (SUCASPtp), “succinate: D-tartrate antiporter (periplasm)” (SUCTARTtp) were constrained to zero and the lower bounds of “Cob(I)alamin exchange” and “ATP maintenance requirement” were set to -0.01 and 3.15 mmol/gDCW/h. Glucose uptake was set at 3.008 and oxygen at 5.703 mmol/gDCW/h [39]. The Yeast 5.0 model was used to simulate the production of ethanol and glycerol by constraining the following reactions: “Ferrocyclochrome-c: oxygen oxidoreductase” (0,10), “Triose-phosphate isomerase” (-1000,6) and glucose exchange (-10) [40]. 6

heterologous models of iJO1366 and of Yeast 5.0 were created, one for each candidate synthetic pathway selected for LA production. The reactions added to each heterologous model are summarized in Table 8.1 – Annex 8.2 and the resulting xml models are also available in the SI2. All heterologous reactions were cytoplasmic and the sink reactions for new heterologous metabolites, including LA, were added to all models. For pathway B in the Yeast model, a sink reaction for tRNA-Glu was added with an upper bound of 1000 because the availability of tRNAs is unaccounted for in the model, and it was assumed that glutamate, not tRNA-glu, would be the limiting substrate in the glutamyl-synthetase reaction. The maximal yield for each pathway was calculated by setting the cellular objective to the LA exchange and accounting for glucose uptake. The production envelopes were computed using the COBRA toolbox function, productionEnvelope, and Flux Coupling Analysis (FCA) was performed using the F2C2 tool [41].

3.4. Results and Discussion

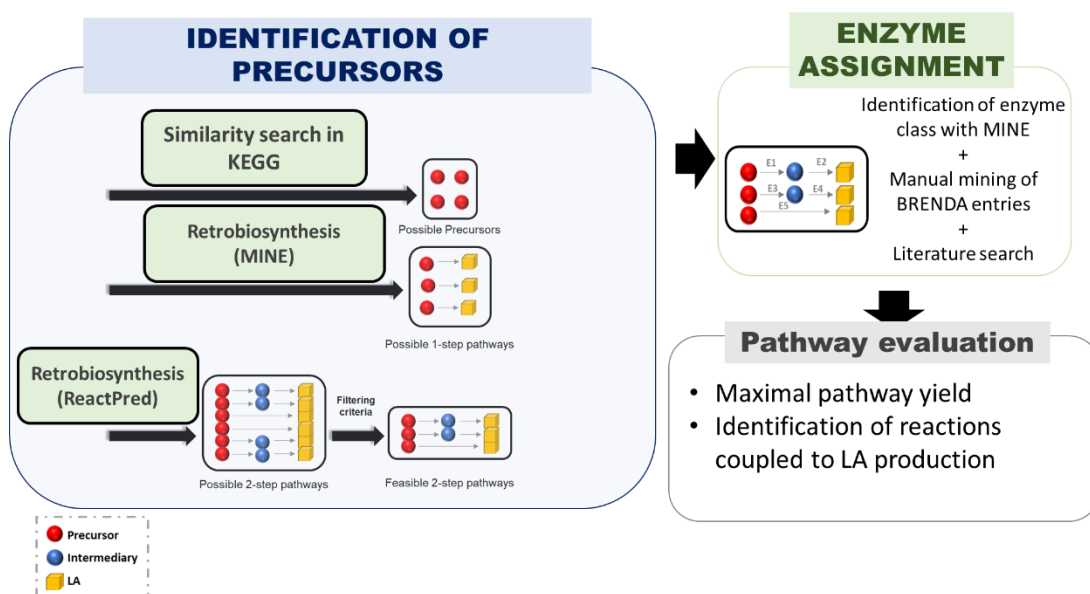


Figure 3.1. Approaches used in this study to identify “new-to-nature” pathways to implement microbial production of LA. The strategy is divided in three stages: search for potential precursors (based on compound similarity and retrobiosynthesis), enzymatic assignment to predicted reactions, and pathway evaluation.

Figure 3.1 schematically represents the approach used in this study to develop an integrated pathway prospecting and analysis workflow for the identification of promising microbial routes for LA production in the two selected hosts, *E. coli* and *S. cerevisiae*. We started by searching possible precursors for the synthesis of LA. The first step involved the identification of metabolites that were chemically similar to LA and, in the second stage, we resorted to a retro-biosynthetic approach (using ReactPRED and MINE) to identify all compounds that could be converted into LA. These ReactPRED approach resulted in the enumeration of thousands of possible pathways that were subsequently pruned and curated based on the thermodynamics of the reactions involved, the stability of the reaction intermediates, and the availability of a specific enzyme that could catalyse the predicted reaction (Figure 3.2 and Methods). This systematic analysis resulted in the selection of a set of 4 complete promising pathways (i.e., with substrates identified and enzymes assigned), which performance in *E. coli* and *S. cerevisiae* was examined using *in silico* metabolic modelling, further shedding light into possible implementation strategies to improve their efficacy in these host systems. A detailed description of the results obtained, along with different steps performed is provided below.

1. Identification of suitable precursors for the production of LA

To identify possible precursors for LA production, we started by surveying the information available in BRENDA and the literature to assess whether this molecule was previously reported as a product or co-product in the described enzymatic reactions. Only the decarboxylation of α -methyl-glutamic acid by the *E. coli* glutamate decarboxylase (GadA) was described to produce LA [1]. However, α -methyl-glutamic acid is not naturally produced by microbial metabolism, and after a thorough analysis resorting to a similar strategy to the one herein described for LA, we found that production of this α -methyl amino-acid from carbohydrates would be very challenging, since the only pathway that could be devised for its biosynthesis would encompass 5 synthetic steps with intermediate compounds that are not commercially available, making it difficult to experimentally validate the pathway (details in the Annex 8.1). Consequently, we

discarded the use of α -methyl-glutamic acid as a reasonable precursor for LA synthesis.

The next step was to search the collection of metabolites described in KEGG for molecules similar to LA. We used the Tanimoto coefficient for making the quantitative comparison between molecules, where their structural information is represented in the form of molecular fingerprints, considering this parameter's previously demonstrated usefulness for the same purpose [2]. Seven metabolites similar to LA (having a Tanimoto coefficient above 0.8) were identified (Table 8.2 – Annex 8.2) with cis-acetylacrylate, an intermediate in the degradation pathway of chlorocyclohexane and chlorobenzene [3], being the most similar and valeric acid, an intermediate in the degradation of valeramide [4], the most dissimilar (Table 8.2 – Annex 8.2). Cis-acetylacrylate and valeric acid were discarded as possible precursors for LA production, since it was anticipated to be difficult to implement production of these molecules from sugar fermentation considering that they are intermediates produced during the degradation of the xenobiotic's chlorobenzene and valeramide, respectively. The next molecules found to be more similar to LA were δ -aminolevulinic acid (D-ALA) and β -aminolevulinic acid (β -ALA) (Table 8.2 – Annex 8.2). D-ALA is the first committed precursor in the synthesis of the tetrapyrrole moiety of heme being produced by both *E. coli* and *S. cerevisiae*, although through different routes [5]. β -ALA is an intermediate in the oxidative degradation of ornithine in *Clostridia* [6], but is not produced by *E. coli* or *S. cerevisiae*. The remaining molecules showing similarity to LA were: 4,5-dioxovaleric acid, an intermediate of alternative pathways for D-ALA synthesis found in plants and yeasts (including *S. cerevisiae*) [7]; 2,4-dioxovaleric acid, an intermediate present in the catabolism of orcinol and produced by insects and lichens but not described in microbes [8]; and succinate semialdehyde, an intermediate generated in the degradation of γ -aminobutyric acid (GABA) [9] (Table 8.2 – Annex 8.2). Altogether, at the end of this similarity analysis, we considered five metabolites as possible precursors for biosynthesis of LA in *E. coli* and *S. cerevisiae*: D-ALA, β -ALA, 2,4-dioxovaleric acid, 4,5-dioxovaleric acid and succinate semialdehyde (Figure 3.3).

To expand the range of possible precursors of LA, we resorted to the utilization of

MINE and the retrobiosynthesis tool, ReactPRED. From the reactions predicted to produce LA by MINE (Table 8.3– Annex 8.2), succinate semialdehyde and 3-oxoadipic acid (3-OAD) were considered more relevant, since these molecules are already described to be produced by microbes, while cis-acetylacrylate and mycophenolic acid are not (Figure 3). Indeed, 3-oxoadipic acid was reported to produce LA *in vivo* [10]. ReactPRED identified 50,768 possible routes (with a maximum of 2 steps) to produce LA. However, only 26,163 of these were found to be thermodynamically favourable (Figure 3.2). The much higher number of possibilities uncovered by ReactPRED, as compared to those provided by MINE, is attributed to the 2-step reactions allowed (MINE only allows 1), but is also a consequence of the known effect on amplification of combinatorial solutions provided by retrobiosynthesis tools [11]. In part, this is also due to the fact that ReactPRED's reaction rules are being automatically created, resulting in the algorithm to consider substrates and intermediates that have chemical structures unlikely to occur in nature. Indeed, out of the 26,163 thermodynamically favourable pathways, only 1,804 included compounds classified in the PubChem database (which includes all known chemical compounds, regardless of whether they are of biological origin or not). In this context, it was considered essential to develop a pipeline that may introduce some biological feasibility analysis to further prune the pathways identified by ReactPRED (see Methods and Figure 3.2 for details). Besides, to identify thermodynamically favourable pathways, we also pruned ReactPred results for pathways that include substrates found in PubChem, involve stoichiometrically correct reactions, and that use substrates and co-substrates exclusively found in the KEEG database (Figure 3.2). This last criterion was used to select pathways that could be more easily connected to the metabolism of *E. coli* or *S. cerevisiae*. Using these criteria, 26 possible routes were identified, which was further pruned down to 12 candidate routes by removing highly similar pathways (Table 8.4 – Annex 8.2). Implementation of additional pruning criteria, including having a consistent reaction mechanism and precursors with a simpler biosynthetic pathway led to 2 final pathways for LA synthesis that we consider to have a higher biosynthetic feasibility: one starting from GABA and the other starting from 5-aminovaleric acid (pathway R8 and R12 in Table 8.4 – Annex 8.2). Both GABA and 5-aminovaleric acid are intermediates

of the yeast endometabolome [12] [13], and GABA is also produced by *E. coli*.

Altogether, the search for substrates performed with ReactPRED, MINE, or looking for similar molecules as LA resulted in the identification of 8 possible precursors for LA biosynthesis: D-ALA, β -ALA, 2,4-dioxovaleric, 4,5-dioxovaleric, succinate semialdehyde, 3-oxoadipic, GABA, and 5-aminovaleric acid (Figure 3.3).

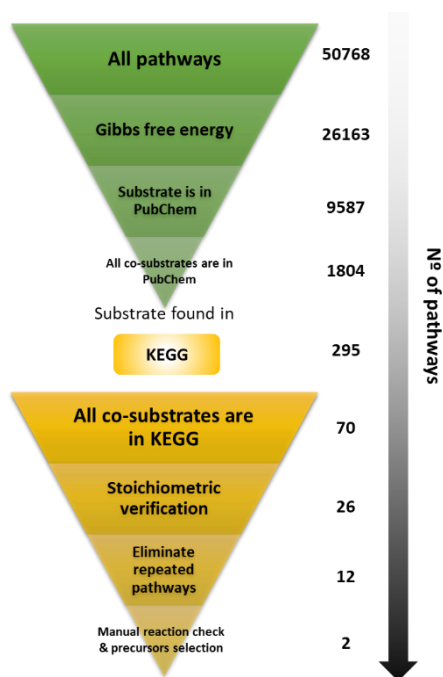


Figure 3.2. Filtering criteria used to prune the list of possible pathways leading to LA unveiled by ReactPRED. The more than 50,000 pathways were filtered to a final 2 possible pathways.

2. Enzyme assignment to the identified synthetic routes

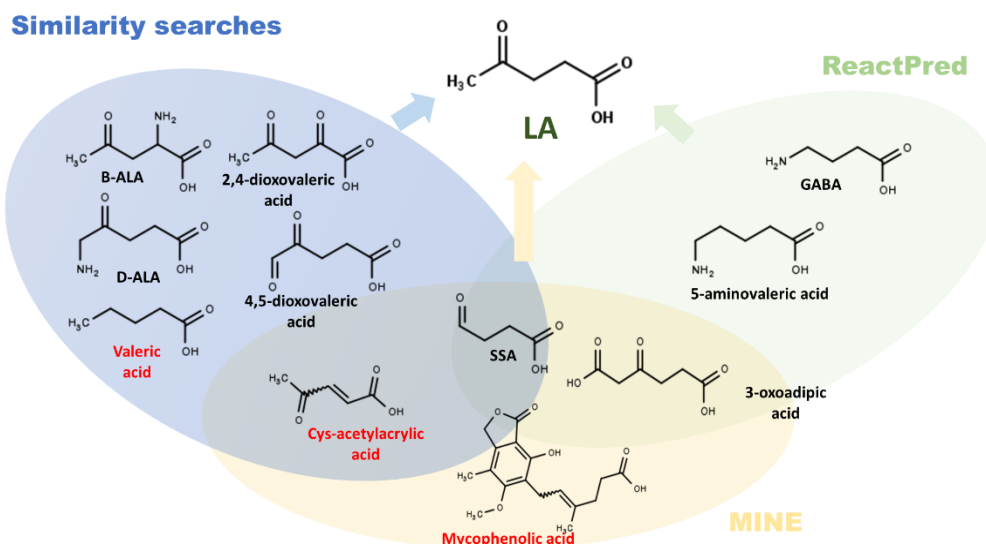


Figure 3.3. List of possible precursors for LA biosynthesis, based on the search for chemically similar molecules or from retrobiosynthesis tools. The search method that yielded each precursor is depicted in light color circles: 1) similarity search in KEGG (blue), 2) Retrobiosynthesis in MINE (orange), 3) Retrobiosynthesis in ReactPred (green). The precursors in a red font were discarded and the ones in black font were selected for enzyme assignment.

After the identification of the most suitable precursors for LA production, we searched for candidate enzymes in the MINE and ATLAS databases that could catalyze the envisaged reactions. Out of the total of the 9 precursors to LA routes, the two tools allowed the assignment of an enzyme class to 7 routes; however, only for 4 of the proposed pathways it was possible to identify a specific enzyme that could catalyze the associated reactions, and therefore, the remaining 3 pathways were discarded (Table 3.1). This difficulty in assigning a specific enzyme reflects the generalist nature of reaction rules used by computational tools, which may serve as bioconversion predictors but do not guarantee the existence of enzyme(s) that could accept the substrates envisaged, at least not without subsequent protein engineering. Indeed, previous studies have shown that there are still around 1000 “orphan activities” (i.e., activities that although described have no associated protein sequence) classified in the EC classification system, which corresponds to around 22% of the total entries of the EC system [14]. This proportion

dramatically increases to nearly 50% when all reactions described in biochemical databases are taken into account [14]. Other problems in enzyme assignment come from the difficulty in picking up specific features of enzyme activity, including specificities associated with the substrates whose absence could be detrimental for catalytic activity.

The conversion of succinic acid into LA in a single step via methylation of succinate semialdehyde by a C-methyltransferase (Table 3.1), is an example of this substrate specificity issue since methyltransferases accepting linear substrates only have shown catalytic activity when the substrate is an α -ketoacid (so that the carbon of the carbonyl group can serve as a nucleophile for the subsequent transfer of the methyl group) [15]. This is not the case for succinate semialdehyde which is a γ -ketoacid. Another example is the impracticality of identifying reductases that could accept substrates similar to 2-hydroxy-levulinic acid (Table 3.1). The identified routes to which it was possible to perform a full enzyme assignment were those starting from D-ALA, 5-aminovaleric acid, and 3-oxoadipic acid, as detailed in Table 3.1 and Figure 3.4.

Chapter 3– In silico prospecting of metabolic pathways for the implementation of microbe-based production of levulinic acid

Table 3.1. Predicted new-to-nature pathways for LA biosynthesis using selected precursors. enzymes to the totality of the steps are represented in red and yellow, respectively. Pathways to which it was possible to assign specific enzymes to all the reactions are depicted in green and are graphically represented in Figure 3.4. Pathways to which it was not possible to assign E.C. sub(sub)classes or specific enzymes to the totality of the steps are depicted in red and yellow, respectively. Pathways to which it was possible to assign specific enzymes to all the reactions are depicted in green and are graphically represented in Figure 3.4.

Method	Precursor	Pathway Description	Predicted reaction rules and specific enzymes assigned
Structural Similarity	Delta-aminolevulinic acid (D-ALA)	D-ALA → 4,5-Diaminovaleric acid → LA	First step was predicted to be 2.6.1.-/1.4.1.- (assigned to 2.6.1.62 and 2.6.1.105); second step was directly assigned to EC 4.3.1.15
	Beta-aminolevulinic acid (B-ALA)	D-ALA → 2,4-Diaminovaleric acid → LA	First step was predicted to be 2.6.1.-/1.4.1.- (with no specific enzymes that could be assigned) , no predicted EC number/reaction rule for the second step
	2,4-dioxovaleric acid	2,4-dioxovaleric acid → 2-hydroxylevulinic acid → cis-acetylacrylic acid → LA	1.1.1.- (first step), 4.2.1.- (second step), 1.3.1.- (third step). No specific enzymes could be assigned.
	4,5-dioxovaleric acid	4,5-dioxovaleric acid → 5-hydroxylevulinic acid → LA	First step is 1.1.1.- , no predicted EC number/reaction rule for the second step
	Glutamate semialdehyde	Glutamate semialdehyde → 4,5-diaminovaleric acid → LA	First step was directly assigned to EC 5.4.3.8, second step was directly assigned to EC 4.3.1.15
Structural Similarity and MINE	Succinate semialdehyde	Succinate semialdehyde → LA	2.1.1.- (no specific enzyme could be assigned)
MINE	3-oxoadipic acid	3-oxoadipic acid → LA	4.1.1.- (assigned to 4.1.1.112 or 4.1.1.4 or methylketone synthase*)
ReactPRED	Gamma-aminobutyric acid (GABA)	GABA → 4-aminovaleric acid → LA	First step is 2.1.1.- (no specific enzyme could be assigned) , second step is 2.6.1.- (assigned to 2.6.1.18 and 2.6.1.40)
	5-aminovaleric acid	5-aminovaleric acid → 4-aminovaleric acid → LA	First step is 5.4.3.- (assigned to 5.4.3.3 or 5.4.3.5), second step is 2.6.1.- (assigned to 2.6.1.18 and 2.6.1.40)

*Methylketone has no attributed E. C. number

The conversion of D-ALA or β -ALA to LA was predicted to involve first a transamination of these precursors into 4,5-diaminovaleric acid (DAVA) (or 2,4-DAVA), which is then transaminated to LA (Table 3.1). Although it might seem counter-intuitive to make a second amination of D-ALA (or β -ALA) when the objective is to remove the amino group present in this precursor (because this is not present in LA), deaminases reactions result on a carbonyl group, and therefore, the single deamination of D-ALA (or β -ALA) would result in 4,5-dioxovaleric acid (or 2,4-dioxovaleric acid) but not LA. No enzymes accepting substrates similar to 4,5-DAVA (or 2,4-DAVA) were identified. However, a di-aminopropionate ammonia lyase (EC 4.3.1.15) that catalyses the α,β -elimination of adjacent ammonia groups in di-aminopropionate [16] and structurally similar to 4,5-DAVA was identified (see Table 8.5 – Annex 8.2 for further details on the similarity between native and synthetic substrates). Because we could not identify an enzyme recognizing a substrate structurally similar to 2,4-DAVA, we discarded the possibility of using β -ALA as a substrate for LA production. 4,5-DAVA is not reported to be produced by *E. coli* or *S. cerevisiae*; however, *in vitro* assays have shown that 4,5-DAVA is transiently formed during the isomerization of glutamate 1-semialdehyde to D-ALA by glutamate semialdehyde mutase (GSAM, EC 5.4.3.8) [17]. As such, one possibility to obtain 4,5-DAVA could be through the engineering of GSAM to release this intermediate from the enzyme pocket, this being facilitated by the previous identification of residues mediating the catalytic conversion of 4,5-DAVA to D-ALA[18]. Similar enzyme gate engineering has been performed in the past to control the size of product formed (polyketides and polymers) by controlling the number of reactions occurring in the active site of two sequential enzymes [68], [69]. This example supports the possibility of using glutamate semialdehyde as a new potential precursor for LA. Another possibility to produce 4,5-DAVA is through oxidative deamination (EC 1.4.1) or transamination (EC 2.6.1) of D-ALA, as suggested by MINE (Table 3.1). Mining of the enzymes classified in these classes that could catalyse these reactions led to the identification of two 7-keto-8-pelargonic acid (KAPA) transaminases, adenosylmethionine-8-amino-7-oxononanoate transaminase (EC 2.6.1.62), and lysine-8-amino-7-oxononanoate transaminase (EC 2.6.1.105), as possible candidates since their

native substrate (8-amino-7-oxopelargonic acid) is similar to D-ALA [19][20] (Table 8.5 – Annex 8.2). At the end of this analysis, we assembled three pathways (A1, A2, and B) for the production of LA: two starting from D-ALA, using an amino donor S-adenosyl methionine (SAM) or lysine, and the other from glutamate semialdehyde (Figure 3.4).

The results obtained with ReactPRED predicted that the conversion of 5-aminovaleric acid to LA required an isomerization to 4-aminovaleric acid followed by a deamination (Table 3.1 and Figure 3.4). Enzyme assignment (based on MINE) for the isomerization step led to the identification of two enzymes that could catalyse this reaction: lysine 5,6-aminomutase (E.C. 5.4.3.3) and ornithine 4,5-aminomutase (E.C. 5.4.3.5) from *Clostridium sticklandii* [21], working on substrates similar to D-ALA (further details on similarity between synthetic and native substrates in Table S5 in SI5). Similarly, three other enzymes, β -alanine-pyruvate transaminase (EC 2.6.1.18) and 3-amino-2-methylpropionate-pyruvate transaminase (EC 2.6.1.40) were also identified as possible candidates to promote the deamination of 4-aminovaleric acid into LA (Table 8.5 – Annex 8.2). This analysis resulted in the assembly of pathway D (Table 3.1 and Figure 3.4).

For the utilization of 3-oxoadipic acid, MINE predicted the use of a decarboxylase to obtain LA, which was assigned to oxaloacetate decarboxylase (4.1.1.112) in this work, due to the similarity of the native substrate to 3-oxoadipic acid (Table 8.5 – Annex 8.2). Furthermore, Cheong et al have demonstrated that Methylketone synthase (no attributed E.C. number) and acetoacetate decarboxylase (4.1.1.4) can also decarboxylated 3-oxoadipic acid to LA [10] (Table 3.1 and Figure 3.4). In this strategy to obtain LA *in vivo* a Claisen condensation reaction catalysed by a β -ketoadipyl-CoA synthetase used succinyl-CoA and acetyl-CoA as precursors. Because the supply of succinyl-CoA can be a limitation, we have searched for other possibilities to produce 3-oxoadipic acid in *E. coli* or *S. cerevisiae*. This survey led to the possibility of producing 3-oxoadipic acid from cis-cis-muconic acid using a 3-oxoadipate enol-lactonase [22], [23]. The production of cis-cis-muconic acid from fermentable sugars was previously implemented *in vivo* [10], [22], [23], thereby increasing interest in the exploration of this pathway as a means to

produce 3-oxoadipic acid. Overall, we considered two pathways starting from 3-oxoadipic acid: one which has already implemented in vivo (pathway E) and the other one is using the shikimate pathway (pathway C) (Table 3.1 and Figure 3.4).

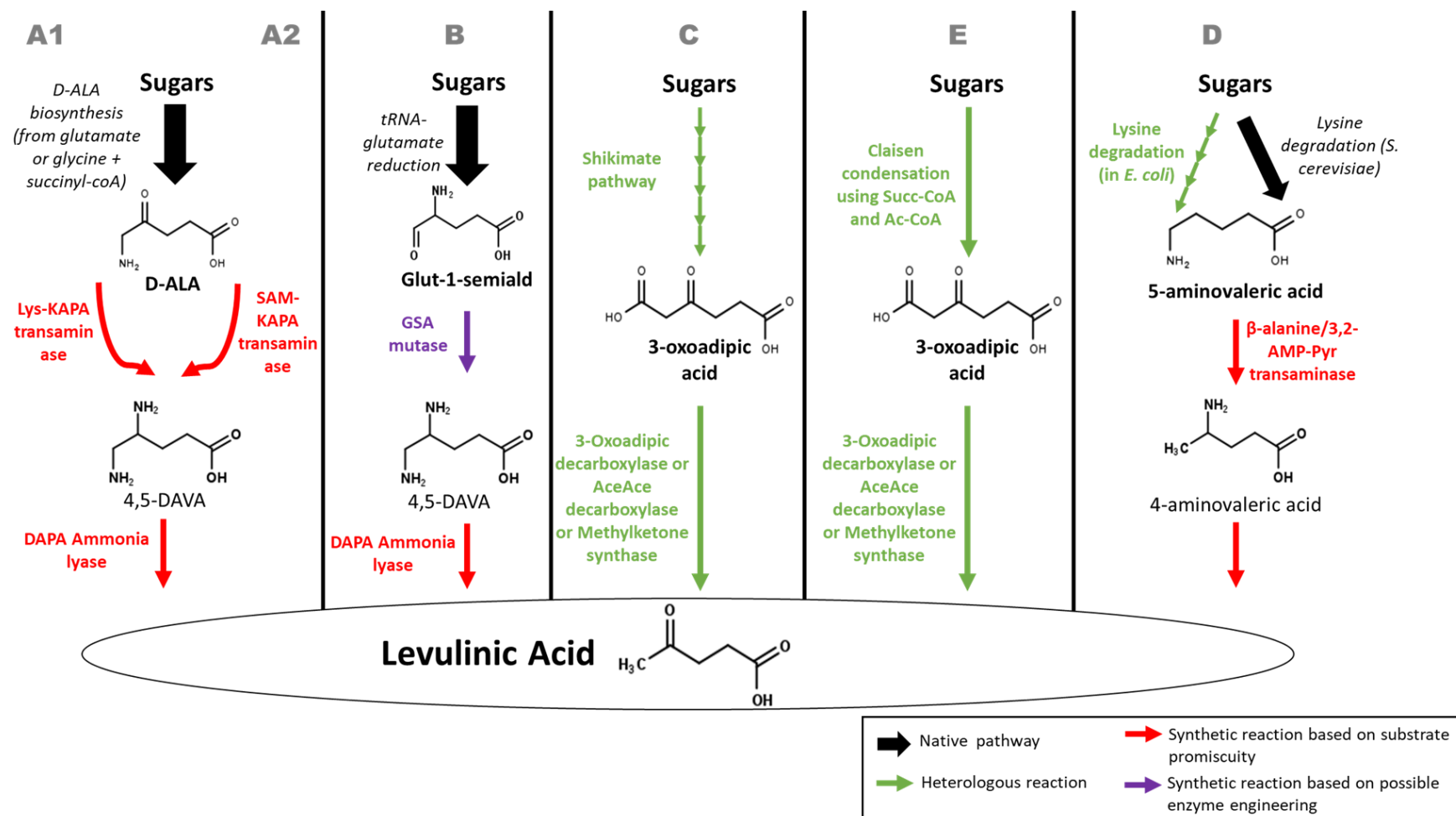


Figure 3.4. Selected pathways leading to LA production. The pathway letter is shown on top, where Pathway A has two variations (A1 and A2), depending on the KAPA transaminase that catalyses the first step. The native enzyme pathways leading to the precursor are depicted in black block arrows. Heterologous reactions that occur naturally in vivo are depicted in thin green arrows, thin red arrows represent synthetic reactions that were assigned an enzyme in this work based on enzyme promiscuity and thin purple arrows represent reactions that were assigned based on an enzyme that requires further engineering. In the case of the lysine degradation pathway, it is native to *S. cerevisiae* (black arrows) but not *E. coli* (yellow thin arrows). D-ALA (Delta-aminolevulinic acid), 4,5-DAVA (4,5-Diaminovaleric acid), Glut-1-Semiald (Glutamate-1-semialdehyde), 3,2-AMP (3-amino-2-methyl-propionic acid), Pyr (Pyruvic acid), Lys (Lysine), Orn (Ornithine), AceAce (Acetoacetic acid), OAD (3-Oxoadipic acid).

3. Analysis of in vivo implementation of the candidate pathways

The last step of this analysis focused on comparing the ease of implementing the 5 assembled pathways in the two selected hosts of choice. For that, we looked into the number of heterologous steps that have to be assembled *in vivo*, the existence of a prior description of the enzymes catalysing the envisaged conversions, the maximum theoretical yield of the pathways, as well as their predicted impact on the physiology of the hosts. For the last two criteria, we have resorted to the use of computational metabolic modelling, a strategy that has been used before with success for examining the performance of synthetic pathways in microbial hosts [24]–[26].

Although a large body of tools for genetic and genomic engineering have been developed to allow simultaneous expression of dozens of heterologous genes in *E. coli* and *S. cerevisiae*, frequently the introduction of these steps that are foreign to the cell interfere with its native metabolism, causing toxic effects such as reduced availability of cofactors [27], whose alleviation requires implementation of dedicated metabolic engineering strategies. In this context, the number of heterologous reactions steps that need to be inserted in the host organism and the number of which are synthetic reaction steps (for which there is no experimental proof that the enzyme can catalyze the reaction) are relevant criteria in considering different pathways (Figure 3.5). The fact that pathway C requires 6 heterologous enzymes can be an important limitation for the success of its implementation (Table 8.6 – Annex 8.2), although it has the advantage that all involved enzymes have described activity for the purpose they are being used in the pathway (Figure 3.4). Pathway E also stems from 3-oxoadipic acid and requires only one heterologous step; however, it is limited by the need to have available cytosolic succinyl-CoA which requires a constant supply of succinate in the fermentation broth (Figure 3.4). Pathways starting from D-ALA or from glutamate semialdehyde (through pathways A and B, respectively) involve a lower number of heterologous steps; however, it remains to be determined whether the enzymes identified are indeed able to catalyse the synthetic reactions. Assembly of pathway D in *E. coli* is more difficult due to the requirement of 4 heterologous steps to produce 5-aminovaleric acid, in addition to the other 2 steps

required to convert this precursor into LA (Figure 3.4 and Table 8.6 – Annex 8.2). Because *S. cerevisiae* already produces 5-aminovaleric acid, only 2 heterologous steps are required to implement pathway D in this system (Figure 3.4).

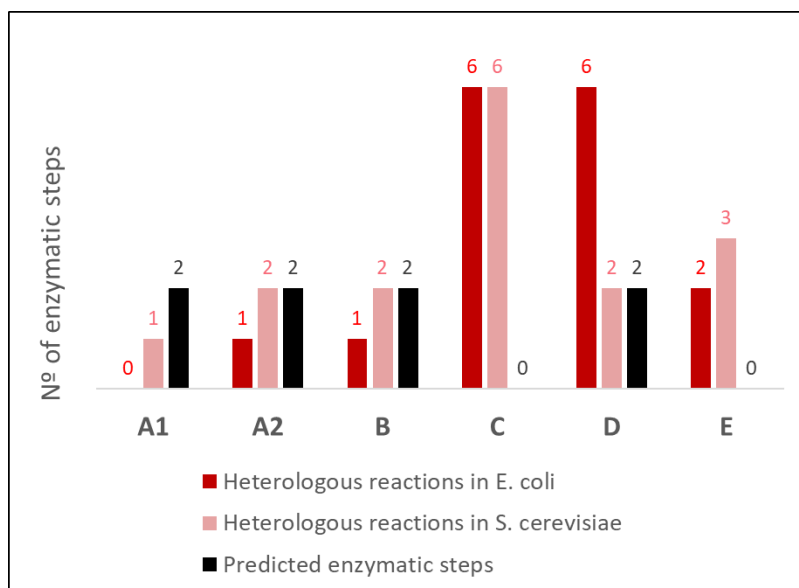


Figure 3.5. Number of heterologous and synthetic reactions for each pathway. Considered for the import of each pathway in *E. coli* or in *S. cerevisiae*.

To estimate the maximum yield of different pathways and their impact on the physiology of the two hosts, we explored constraint-based metabolic modelling using the extensively curated models iJO1366 [28] for *E. coli*, and Yeast 5.05 [29] for *S. cerevisiae* (Figure 3.6). In both cases, we used experimentally validated constraints to simulate continuous aerobic growth in both systems [30] [31] (as detailed in Materials and Methods). One important modification of the Yeast 5.0 model to enable simulation of the pyridoxal phosphate (PLP)-dependent pathways A1 and A2 was to introduce a PLP synthase reaction (4.3.3.6), enabling *de novo* PLP production (although this pathway is known to exist in yeast and an enzyme with pyridoxal synthase function has been recently identified [32], this reaction was not included in the published 5.0 model). In the case of

pathway E, succinyl-CoA is only produced in the mitochondria (using the TCA cycle) but there is no reported transport across the mitochondrial membrane. To account for the need to have cytosolic succinyl-CoA to have a functional pathway, a bacterial alpha-ketoglutarate cytosolic reaction was included in the model; this being supported by previous demonstrations of the success in the use of this enzyme in vivo in yeast cells as a means to provide cytosolic succinyl-CoA [33]. The maximal pathway yields predicted for different assembled pathways in *E. coli* or *S. cerevisiae* are shown in Figure 3.6A, while panels B and C show the corresponding production envelopes (representing the trade-off between production and growth [34]). In terms of estimated yields, pathways C and E achieved higher yields in both *E. coli* and *S. cerevisiae*, with pathway E achieving yields close to the maximal theoretical yield (1.091 mol LA/mol glucose), estimated based on the energy difference of the two molecules, in terms of reducing equivalents [27] (Figure 3.6). These two pathways use 3-oxoadipate as the precursor, and the similar yields obtained in the two hosts likely to arise from similar organization of the two metabolic networks to produce this precursor. Although pathways A and B both use 4,5-DAVA as a precursor for LA synthesis, the maximal predicted yield of pathway B is higher in both *E. coli* and *S. cerevisiae* (Figure 3.6). Indeed, pathway B in *E. coli* is where the second highest yield is achieved, with a significant difference between this host and yeast, which may be a reflection of the pathways that these two hosts use for glutamate production. Another contributing factor for the higher efficacy of pathway B is the lower formation of by-products, as the amination of glutamate semialdehyde uses pyridoxamine phosphate (PMP) that can be recycled endogenously (via PLP amination in the amine salvage pathway [35],[36]) while in pathways A1 and A2, the donor groups are consumed through the 4,5-DAVA forming reaction to yield S-adenosyl-4-(methylsulfanyl)-2-oxobutanoate (when SAM is used as the amino donor, pathway A1) or 2-amino-6-hexanoate (when lysine is used as the amino donor, pathway A2), that cannot be recycled. It is also interesting to note that in both variations of pathway A, the yields are higher in *E. coli* than in *S. cerevisiae*, likely reflecting the different routes by which D-ALA is obtained in these hosts: *E. coli* utilizes the C5-pathway from glutamyl-tRNA, while *S. cerevisiae* utilizes the C4-pathway from succinyl-coA and glycine.

Moreover, the differences observed in *E. coli* concerning the yield of pathways A1 and A2 are clearly pointing to lysine being a better candidate for the amination of D-ALA than SAM (Figure 3.6).

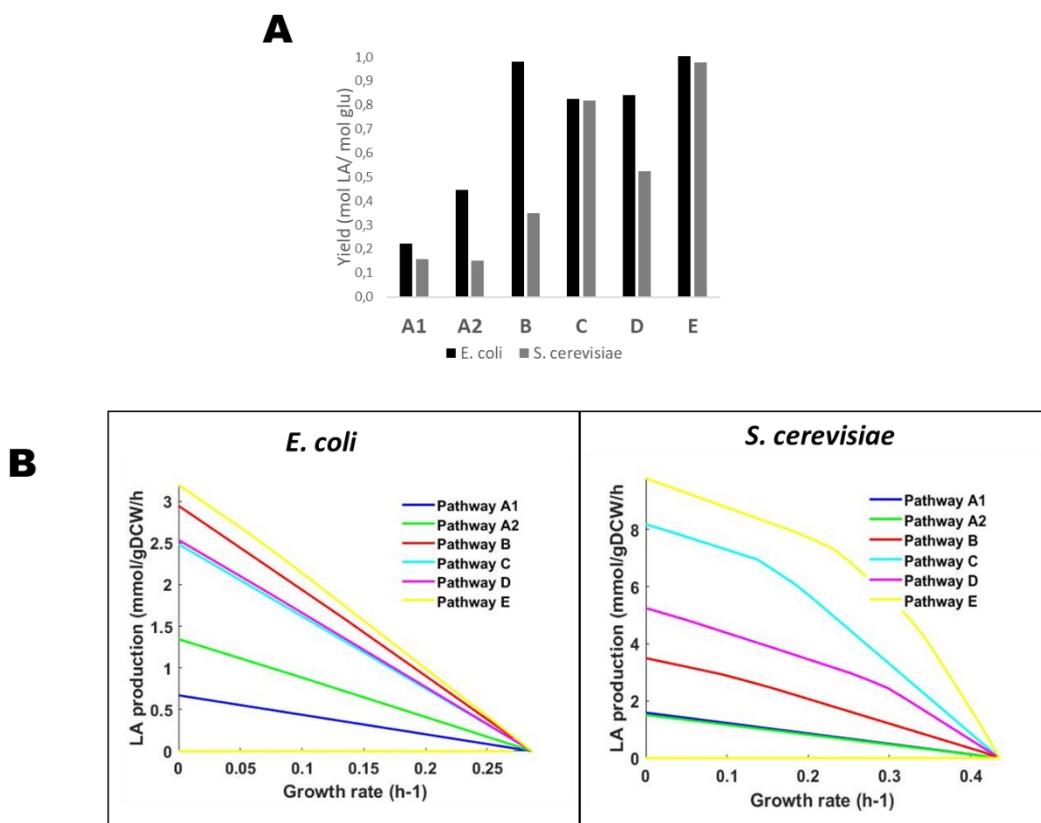


Figure 3.6. Evaluation of the selected pathways A) Maximal theoretical yields of each pathway in the metabolic networks of *E. coli* and *S. cerevisiae*. B) Production envelopes for each pathway in *E. coli* and in *S. cerevisiae*.

The yield differences of different pathways in the same host also reflect the very different maximum fluxes predicted, and these differences can be visualized with the production envelopes (Figure 3.6B). Clearly, the much higher fluxes of LA production are predicted by pathways C and E, especially in the case of *S. cerevisiae*, while pathways A1 and A2, starting from D-ALA, show much lower LA production fluxes. This result strongly suggests that the availability of D-ALA can be a limiting step for a successful

implementation of pathway A in both systems. Previous experimental studies have identified and addressed this issue, and strategies have been envisaged to improve the production of D-ALA both in *E. coli* and in *S. cerevisiae* [37], [38]. Thus, this knowledge can be incorporated to improve the production of LA from this precursor. Pathway D presents as a middle candidate in what concerns the maximal yield and production envelope, both for *E. coli* and *S. cerevisiae*. To get further insights into other factors that could be limiting the efficacy of different assembled pathways, we used FCA to identify the reactions whose flux is associated to chemical production of LA (i.e. the reactions which are required to carry a non-zero flux so that the LA production flux can be different than zero) [39]. For pathways B and E, we could not identify such reactions (besides the expected heterologous reactions added to the model and the uptake of glucose); however, it was possible to identify bottlenecks for the remaining routes (detailed in Table 8.7–Annex 8.2). Expectedly, the synthesis of D-ALA, as well as the activity of aconitase was suggested to limit LA production in yeast, consistent with in vivo observations that the supply of citrate and isocitrate limit the production of D-ALA [38]. The use of SAM for the synthesis of 4,5-DAVA from D-ALA does not appear to be a bottleneck; however, when lysine is used as the amino donor, the biosynthesis of this amino acid is suggested to limit the production of D-ALA by the modelling analysis (Table 8.7–Annex 8.2). The comparison of production envelopes of pathways A1 and A2 in *E. coli* suggests that SAM biosynthesis might be limiting the production of LA; however, our FCA analysis did not identify a reaction directly linked to the biosynthesis of this precursor as having a flux limiting LA production. Differently, reactions linked to the biosynthesis of purines and glutamine were found to limit LA production both in *S. cerevisiae* and *E. coli*, suggesting that the internal availability of these precursors can limit the production of LA. The production envelopes also render clear the expected competition between growth and LA production in both hosts (Figure 3.6B), as LA is not consumed by the hosts and therefore, has no other connections in the two metabolic networks.

While all pathways show a similar competitive behaviour in *E. coli*, this is not the case for *S. cerevisiae*, in which pathways E, C, D and, less significantly, pathway B, show that the competition between production and growth is accelerated after a given biomass

threshold (translating into a higher slope of the graph and happening much earlier for pathway C than for the others). This dual behaviour indicates that at higher growth rates, the production potential is further decreased. These types of production envelopes have been described for other end-metabolites such as product methylation [40] and even for endogenous metabolites [41]; however, the reason for this acceleration of competition between growth and biomass is not clear. Either way, the strong competition with growth suggested for pathways A and B, especially in the case of *S. cerevisiae*, emphasises that the production of LA through these routes should not be designed using exponentially growing cells, for example, by using inducible systems that are only activated when the cells cease growth or high-density fermentations in continuous reactors.

Taking into consideration all the aspects that have been discussed above, we have summarised the pros and cons of each pathway for both hosts (Table 2).

Chapter 3– In silico prospecting of metabolic pathways for the implementation of microbe-based production of levulinic acid

Table 3.2. Advantages, disadvantages and mitigation approaches for each LA pathway found in this work.

	Pros	Cons	Mitigation approaches to improve efficacy
A	Only 1 or 0 heterologous steps in <i>E. coli</i> (depending on which pathway variant) and only 2 heterologous steps in <i>S. cerevisiae</i> . Production of the precursor D-ALA has been carried out in a variety of hosts. Metabolic engineering strategies for lysine overproduction can be applied here to increase the cofactor pool (in pathway A2) extensively optimized	Lower pathway yield. Strong competition with growth. The cofactors in the first reaction step (SAM and lysine) are required in stoichiometric amounts, likely contributing to a low yield	Improve availability of D-ALA, glutamine and purines. Use in experimental setups that do not require exponentially growing cells such as high-density fermentations or growth-limited chemostats
B	Only 1 heterologous step in <i>E. coli</i> and 2 in <i>S. cerevisiae</i> . High pathway yield in <i>E. coli</i> .	Necessary enzyme gate engineering for one of the two heterologous enzymes.	
C	Complete pathway has not been assembled but all enzymatic steps are proven <i>in vivo</i> (0 predicted enzymatic steps). Intermediate pathway yield. Metabolic engineering strategies for 3-dehydroshikimate overproduction can be applied here to increase the precursor pool.	High number of heterologous steps (6) for both <i>E. coli</i> and <i>S. cerevisiae</i>	Improve availability of 3-dehydroshikimate.
D	Metabolic engineering strategies for lysine overproduction can be applied here to increase the precursor pool	High number of heterologous steps in <i>E. coli</i> (6)	Improve availability of lysine.
E	Pathway has been assembled <i>in vivo</i> (0 predicted enzymatic steps). Highest theoretical yield (considering the expression of a bacterial alpha-ketoglutarate dehydrogenase in yeast cytosol). High pathway yield in both <i>E. coli</i> and <i>S. cerevisiae</i> .	Need for a cytosolic supply of succinyl-coA in <i>S. cerevisiae</i>	Consider the use of pathway C to avoid succinyl-coA supply limitations

3.5. Conclusions

This work provided a systematic framework to design, analyse, and implement new-to-nature pathways that can result in the production of the building block chemical LA in *E. coli* or *S. cerevisiae* (Figure 3.7). One challenging step in pathway prospecting is enzyme assignment, even when computational tools are used, since this identification is made at the 3rd level of the enzyme class number (EC), which often contains dozens of enzymes from different sources and with different specificities. To circumvent this, expertise in biology and laborious literature mining are necessary to identify whether are suitable candidates among the enzymes assigned by the algorithms. Resorting to a combined approach involving the use of different complimentary computational tools and extensive expert manual curation for exhaustive prospecting of suitable LA biosynthetic pathways, a set of five candidate pathways was identified in this work, and their performance in the selected hosts was compared using the hosts' genome-scale metabolic models and various constraint-based modelling techniques. These analyses highlighted that 3-oxoadipic acid is the most amenable substrate for LA production at high yields. However, the success of 3-oxoadipic acid-based pathways is dependent on the need to bypass the requirement for a constant supply of succinate in the fermentation broth, or successfully implementing a 5-step heterologous pathway to produce 3-oxoadipic acid through implementing the shikimate pathway that bypasses the need for succinyl-CoA. The survey performed also proposed two routes for the use of D-ALA to produce LA in *S. cerevisiae* with only one or two heterologous steps (zero in the case of route A1 in *E. coli*); however, further studies are required to determine if the identified enzymes do have the ability to mediate the proposed conversions. Besides, the success of these pathways also appear to be dependent on the need to use strategies that increase the available internal pool of D-ALA and of experimental settings that resort to the use of non-exponentially growing cells to maximize the use of the precursors available for the production of LA. For the first time, it is also proposed to use glutamate semialdehyde and 5-aminovaleric acid as alternative precursor for LA production. Overall, not only does this work open the door for further investigations that might result in unlocking the

Chapter 3– In silico prospecting of metabolic pathways for the implementation of microbe-based production of levulinic acid

potential of microbial LA production at an industrial scale, but it also provides a framework that can be used for pathway prospecting of other value-added chemicals.

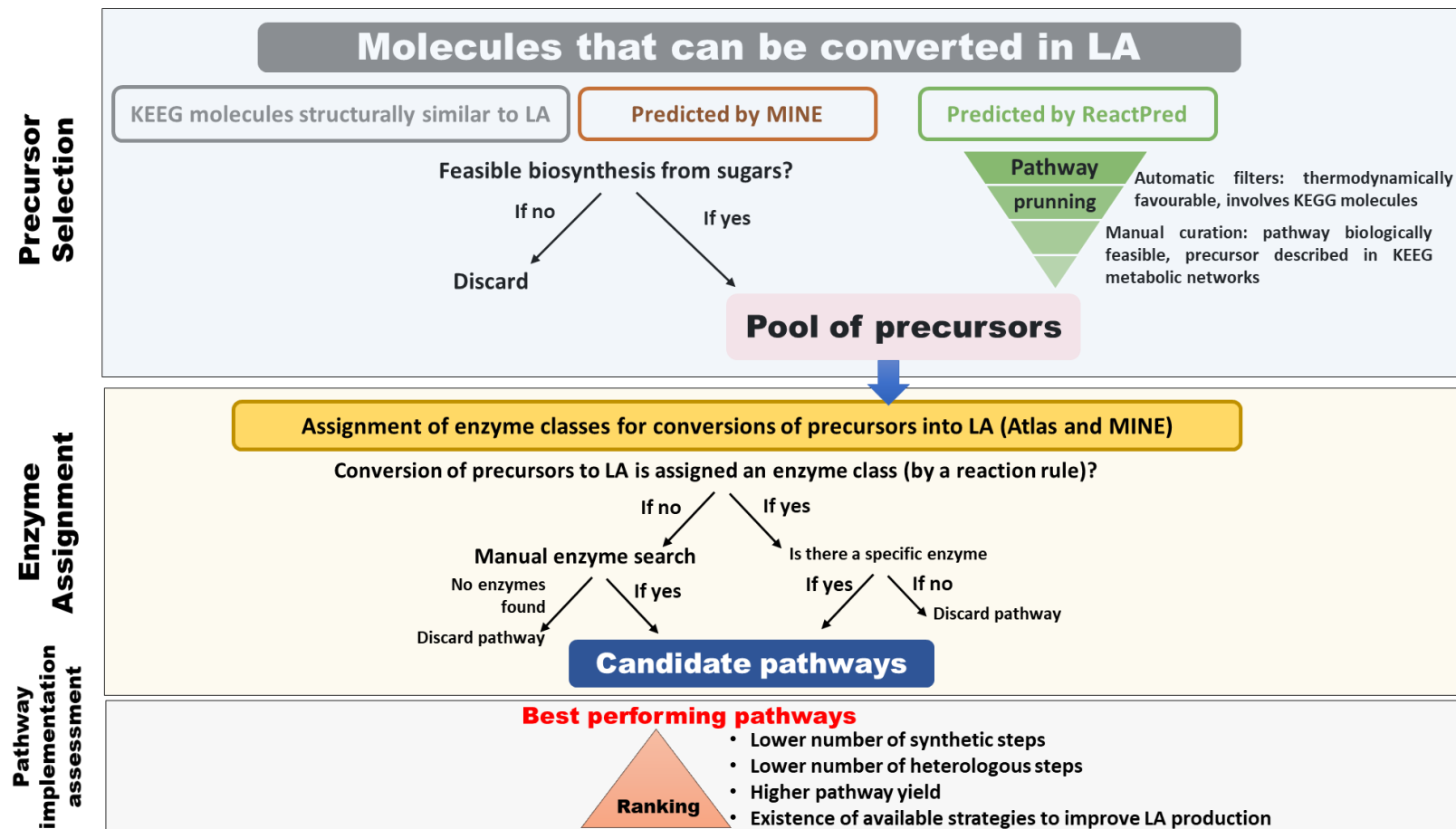


Figure 3.7. Outline of the steps taken in this work to select the 5 new LA routes.

3.6. References

- [1] M. Bertoldi *et al.*, “Ornithine and glutamate decarboxylases catalyse and oxidative deamination of their alpha-methyl substrates,” *Biochem. J.*, vol. 512, pp. 509–512, 1999, doi: 10.1042/0264-6021:3420509.
- [2] D. Bajusz, A. RÁCZ, and K. Héberger, “Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations?,” *J. Cheminform.*, vol. 7, no. 1, p. 20, Dec. 2015, doi: 10.1186/s13321-015-0069-3.
- [3] M. Brückmann, R. Blasco, K. N. Timmis, and D. H. Pieper, “Detoxification of protoanemonin by dienelactone hydrolase,” *J. Bacteriol.*, vol. 180, no. 2, pp. 400–2, Jan. 1998, Accessed: Mar. 03, 2020. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/9440530>.
- [4] M. Maestracei, K. Bui, A. Arnaud, P. Galzy, and P. Viala, “Activity and regulation of an amidase (acylamide amidohydrolase, EC 3.5.1.4) with a wide substrate spectrum from a *Brevibacterium* sp.,” pp. 315–320, 1984.
- [5] K. Sasaki, M. Watanabe, and T. Tanaka, “Biosynthesis, biotechnological production and applications of 5-aminolevulinic acid,” *Appl. Microbiol. Biotechnol.*, vol. 58, no. 1, pp. 23–29, Jan. 2002, doi: 10.1007/s00253-001-0858-7.
- [6] Gerhard Gottschalk and Louis-Moreau, *Bacterial Metabolism*. Springer Science & Business Media, 1986.
- [7] K. Hoare and K. Datta, “Characteristics of l-alanine:4,5-dioxovaleric acid transaminase: An alternate pathway of heme biosynthesis in yeast,” *Arch. Biochem. Biophys.*, vol. 277, no. 1, pp. 122–129, 1990, doi: 10.1016/0003-9861(90)90559-H.
- [8] P. J. Chapman and D. W. Ribbons, “Metabolism of resorcinyl compounds by bacteria: orcinol pathway in *Pseudomonas putida*,” *J. Bacteriol.*, vol. 125, no. 3, pp. 975–84, Mar. 1976, Accessed: Aug. 12, 2019. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/1254564>.
- [9] E. H. Jang, S. A. Park, Y. M. Chi, and K. S. Lee, “Kinetic and structural characterization for cofactor preference of succinic semialdehyde dehydrogenase from *Streptococcus pyogenes*,” *Mol. Cells*, vol. 37, no. 10, pp. 719–726, 2014, doi: 10.14348/molcells.2014.0162.
- [10] S. Cheong, J. M. Clomburg, and R. Gonzalez, “Energy- and carbon-efficient synthesis of functionalized small molecules in bacteria using non-decarboxylative Claisen condensation reactions,” *Nat. Biotechnol.*, vol. 34, no. 5, pp. 556–561, 2016, doi: 10.1038/nbt.3505.
- [11] G. M. Lin, R. Warden-Rothman, and C. A. Voigt, “Retrosynthetic design of metabolic pathways to chemicals not found in nature,” *Curr. Opin. Syst. Biol.*, vol. 14, pp. 82–107, 2019, doi: 10.1016/j.coisb.2019.04.004.
- [12] M. Ramirez-Gaona *et al.*, “YMDB 2.0: A significantly expanded version of the

- yeast metabolome database,” *Nucleic Acids Res.*, vol. 45, no. D1, pp. D440–D445, Jan. 2017, doi: 10.1093/nar/gkw1058.
- [13] F. P. Guengerich and H. P. Broquist, “Lysine catabolism in *Rhizoctonia leguminicola* and related fungi,” *J. Bacteriol.*, vol. 126, no. 1, pp. 338–47, Apr. 1976, Accessed: Aug. 29, 2019. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/131119>.
 - [14] M. Sorokina, M. Stam, C. Médigue, O. Lespinet, and D. Vallenet, “Profiling the orphan enzymes,” *Biol. Direct*, vol. 9, no. 1, pp. 1–16, 2014, doi: 10.1186/1745-6150-9-10.
 - [15] A. W. Struck, M. L. Thompson, L. S. Wong, and J. Micklefield, “S-Adenosyl-Methionine-Dependent Methyltransferases: Highly Versatile Enzymes in Biocatalysis, Biosynthesis and Other Biotechnological Applications,” *ChemBioChem*, vol. 13, no. 18, pp. 2642–2655, 2012, doi: 10.1002/cbic.201200556.
 - [16] F. Khan, V. R. Jala, N. A. Rao, and H. . Savithri, “Characterization of recombinant diaminopropionate ammonia-lyase from *Escherichia coli* and *Salmonella typhimurium*,” *Biochem. Biophys. Res. Commun.*, vol. 306, no. 4, pp. 1083–1088, Jul. 2003, doi: 10.1016/S0006-291X(03)01100-8.
 - [17] R. Contestabile, S. Angelaccio, R. Maytum, F. Bossa, and R. A. John, “The contribution of a conformationally mobile, active site loop to the reaction catalyzed by glutamate semialdehyde aminomutase,” *J. Biol. Chem.*, vol. 275, no. 6, pp. 3879–86, Feb. 2000, doi: 10.1074/JBC.275.6.3879.
 - [18] J. Stetefeld, M. Jenny, and P. Burkhard, “Intersubunit signaling in glutamate-1-semialdehyde-aminomutase,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 37, pp. 13688–13693, Sep. 2006, doi: 10.1073/pnas.0600306103.
 - [19] S. W. Van Arsdell *et al.*, “Removing a bottleneck in the *Bacillus subtilis* biotin pathway: BioA utilizes lysine rather than S-adenosylmethionine as the amino donor in the KAPA-to-DAPA reaction,” *Biotechnol. Bioeng.*, vol. 91, no. 1, pp. 75–83, 2005, doi: 10.1002/bit.20488.
 - [20] R. S. Breen, D. J. Campopiano, S. Webster, M. Brunton, R. Watt, and R. L. Baxter, “The mechanism of 7,8-diaminopelargonate synthase; the role of S-adenosylmethionine as the amino donor,” *Org. Biomol. Chem.*, vol. 1, no. 20, pp. 3498–9, 2003, doi: 10.1039/b310443p.
 - [21] C. H. Tseng, C. H. Yang, H. J. Lin, C. Wu, and H. P. Chen, “The S subunit of D-ornithine aminomutase from *Clostridium sticklandii* is responsible for the allosteric regulation in D- α -lysine aminomutase,” *FEMS Microbiol. Lett.*, vol. 274, no. 1, pp. 148–153, Sep. 2007, doi: 10.1111/j.1574-6968.2007.00820.x.
 - [22] C. Weber, C. Brückner, S. Weinreb, C. Lehr, C. Essl, and E. Boles, “Biosynthesis of cis,cis-muconic acid and its aromatic precursors, catechol and protocatechuic acid, from renewable feedstocks by *Saccharomyces cerevisiae*,” *Appl. Environ. Microbiol.*, vol. 78, no. 23, pp. 8421–30, Dec. 2012, doi: 10.1128/AEM.01983-12.

- [23] W. Niu, K. M. Draths, and J. W. Frost, “Benzene-Free Synthesis of Adipic Acid,” *Biotechnol. Prog.*, vol. 18, no. 2, pp. 201–211, Apr. 2002, doi: 10.1021/bp010179x.
- [24] H. Yim *et al.*, “Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol,” *Nat. Chem. Biol.*, vol. 7, no. 7, pp. 445–452, 2011, doi: 10.1038/nchembio.580.
- [25] M. A. Islam, N. Hadadi, M. Ataman, V. Hatzimanikatis, and G. Stephanopoulos, “Exploring biochemical pathways for mono-ethylene glycol (MEG) synthesis from synthesis gas,” *Metab. Eng.*, vol. 41, pp. 173–181, 2017, doi: 10.1016/j.ymben.2017.04.005.
- [26] M. Tokic *et al.*, “Discovery and Evaluation of Biosynthetic Pathways for the Production of Five Methyl Ethyl Ketone Precursors,” *ACS Synth. Biol.*, vol. 7, no. 8, pp. 1858–1873, 2018, doi: 10.1021/acssynbio.8b00049.
- [27] D. Dugar and G. Stephanopoulos, “Relative potential of biosynthetic pathways for biofuels and bio-based products,” *Nat. Biotechnol.*, vol. 29, no. 12, pp. 1074–1078, 2011, doi: 10.1038/nbt.2055.
- [28] J. D. Orth *et al.*, “A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism,” *Mol. Syst. Biol.*, vol. 7, no. 535, Jan. 2011, doi: 10.1038/msb.2011.65.
- [29] B. D. Heavner, K. Smallbone, B. Barker, P. Mendes, and L. P. Walker, “Yeast 5 - an expanded reconstruction of the *Saccharomyces cerevisiae* metabolic network.,” *BMC Syst. Biol.*, vol. 6, no. 55, Jan. 2012, doi: 10.1186/1752-0509-6-55.
- [30] D. S. Weaver, I. M. Keseler, A. Mackie, I. T. Paulsen, and P. D. Karp, “A genome-scale metabolic flux model of *Escherichia coli* K – 12 derived from the EcoCyc database,” pp. 1–24, 2014.
- [31] R. Andrade, M. Doostmohammadi, J. L. Santos, M.-F. Sagot, N. P. Mira, and S. Vinga, “MOMO - multi-objective metabolic mixed integer optimization: application to yeast strain engineering,” *BMC Bioinformatics*, vol. 21, no. 1, p. 69, Dec. 2020, doi: 10.1186/s12859-020-3377-1.
- [32] M. D. Paxhia and D. M. Downs, “SNZ3 encodes a PLP synthase involved in thiamine synthesis in *saccharomyces cerevisiae*,” *G3 Genes, Genomes, Genet.*, vol. 9, no. 2, pp. 335–344, Feb. 2019, doi: 10.1534/g3.118.200831.
- [33] N. Baldi *et al.*, “Functional expression of a bacterial α -ketoglutarate dehydrogenase in the cytosol of *Saccharomyces cerevisiae*,” *Metab. Eng.*, vol. 56, no. August, pp. 190–197, 2019, doi: 10.1016/j.ymben.2019.10.001.
- [34] J. D. Orth, I. Thiele, and B. Ø. Palsson, “What is flux balance analysis?,” *Nat. Biotechnol.*, vol. 28, no. 3, pp. 245–8, Mar. 2010, doi: 10.1038/nbt.1614.
- [35] T. B. Fitzpatrick, N. Amrhein, B. Kappes, P. Macheroux, I. Tews, and T. Raschle, “Two independent routes of de novo vitamin B6 biosynthesis: Not that different after all,” *Biochem. J.*, vol. 407, no. 1, pp. 1–13, 2007, doi: 10.1042/BJ20070765.
- [36] U. Schell, R. Wohlgemuth, and J. M. Ward, “Synthesis of pyridoxamine 5’-

- phosphate using an MBA:pyruvate transaminase as biocatalyst,” *J. Mol. Catal. B Enzym.*, vol. 59, no. 4, pp. 279–285, Aug. 2009, doi: 10.1016/j.molcatb.2008.10.005.
- [37] Z. Kang, W. Ding, X. Gong, Q. Liu, G. Du, and J. Chen, “Recent advances in production of 5-aminolevulinic acid using biological strategies,” *World J. Microbiol. Biotechnol.*, vol. 33, no. 11, p. 200, Nov. 2017, doi: 10.1007/s11274-017-2366-7.
- [38] K. Y. Hara *et al.*, “5-Aminolevulinic acid fermentation using engineered *Saccharomyces cerevisiae*,” *Microb. Cell Fact.*, vol. 18, no. 1, pp. 1–8, 2019, doi: 10.1186/s12934-019-1242-6.
- [39] A. Larhlimi, L. David, J. Selbig, and A. Bockmayr, “F2C2: a fast tool for the computation of flux coupling in genome-scale metabolic networks,” *BMC Bioinformatics*, vol. 13, no. 1, 2012, doi: 10.1186/1471-2105-13-57.
- [40] K. Jensen, V. Broeken, A. S. L. Hansen, N. Sonnenschein, and M. J. Herrgård, “OptCouple: Joint simulation of gene knockouts, insertions and medium modifications for prediction of growth-coupled strain designs,” *Metab. Eng. Commun.*, vol. 8, p. e00087, Jun. 2019, doi: 10.1016/j.mec.2019.e00087.
- [41] A. M. Feist, D. C. Zielinski, J. D. Orth, J. Schellenberger, M. J. Herrgård, and B. O. Palsson, “Model-driven evaluation of the production potential for growth-coupled products of *Escherichia coli*,” *Metab. Eng.*, vol. 12, no. 3, pp. 173–186, May 2010, doi: 10.1016/j.ymben.2009.10.003.

Chapter 3– In silico prospecting of metabolic pathways for the implementation of microbe-based production of levulinic acid

**Chapter 4 – Strategies for the
successful implementation of an *in
silico* predicted LA-forming
pathway using D-ALA as a
precursor**

Statement about research contributions:

The LC-MS analysis presented in this chapter were performed by Maria Conceição Oliveira from Centro de Química Estrutural.

4.1. Abstract

The interest in having levulinic acid (LA) produced via microbial fermentation has fostered an *insilico* search for a new-to-nature pathway that could result in the biosynthesis of this carboxylic acid from fermentable sugars. The results of this *in silico* search, described in Chapter 3, led to the identification of 5 possible pathways for production of LA in *E. coli* or in *S. cerevisiae*. One of the candidate pathways that seemed more promising starts from the use of the ubiquitous heme precursor D-ALA and requires two “synthetic” steps (that is two conversions for which it was possible to identify a candidate enzyme) in which D-ALA is converted into the 4,5-DAVA precursor (through the action of two candidate transaminases) which is afterwards converted into LA (by a DAPA ammonia lyase). This chapter describes the efforts that were taken to confirm whether or not the candidate enzymes are able to catalyse the predicted reactions. Two approaches were used to test whether or not the *in silico* identified pathway could result in LA production: i) test *in vitro* (through enzymatic assays) the capacity of the identified enzymes to catalyze the envisaged conversions of D-ALA into LA; ii) make the *in vivo* assembly of the pathway in *E. coli* to see whether or not LA can be produced upon over-expression of the enzymes. The results obtained with the *in vitro* enzymatic assays confirmed that KANA transaminase can catalyze the envisaged conversion of D-ALA into DAVA, albeit at low levels. During the time of this thesis no confirmation could be obtained concerning the possibility of DAPA ammonia lyase converting DAVA into LA.

In vivo assembly of the selected pathway in a D-ALA over-producer *E. coli* strain did not result in identification of LA in the fermentation broth, demonstrating that, even if it is confirmed the possibility of the enzymes to promote the envisaged synthetic reactions, there are still challenges to be addressed to create a functional LA biosynthetic pathway in this host. Since one of the bottlenecks that were identified with the implementation of this pathway in the two selected hosts was the availability of D-ALA (see results from Chapter 3), in this chapter we also describe the efforts that were undertaken to obtain an *S. cerevisiae* over-producing this precursor (in *E. coli* this was not pursued since strains with this phenotype had been previously described).

4.2.Introduction

LA is a C5 carboxylic acid with industrial interest, due to its potential to be used as a platform chemical and possibility of feeding various industries, including latexes, solvents, herbicides, acrylics and biofuel additives [47]. LA can be obtained from lignocellulosic biomass using various chemical processes, the most developed being the Biofine [48]. This process resorts to the use of dilute sulfuric acid to release 5-HMF which is then converted to LA. Production of LA via microbial fermentation has been very little explored with the more relevant description of this “bio-based production” being the recently reported production in *E. coli*. For this, acetyl-coA and succinyl-coA undergo a non-decarboxylate condensation (catalyzed by *E. coli* β -ketoadipyl-CoA synthetase PaaJ) forming 3-oxoadipyl-coA, from which the coA group is afterwards removed (by *P. putida* CoA transferase PcaIJ) to yield 3-oxoadipate which is then decarboxylated by a methylketone synthase (Mks1 from *Solanum habrochaites* or by acetoacetate decarboxylase Adc from *Clostridium acetobutylicum*) to yield LA [22]. Although successful, this pathway led to very low titers of LA and it also required a continuous supplementation of the medium with succinic acid which will pose important problems for larger-scale implementation. As such, with the goal of envisioning a diversification of other possible pathways that could be used to promote the production of LA, a prospection of possible pathways for the synthesis of LA was performed, the results of this being described in Chapter 3. Resorting to a combination of tools it was possible to propose 5 candidate pathways, starting from different precursors, for LA production in the bacterium *E. coli* or in the yeast *S. cerevisiae*, selected based on their extensive track-record of use in industrial biotechnology. In this chapter are described the efforts that were undertaken to explore one of the identified pathways as a possible mean to implement LA production in *E. coli* or *S. cerevisiae*.

The candidate pathway that was selected for subsequent validation (for reasons described above) starts with the ubiquitous precursor D-ALA, which is aminated to form DAVA and subsequently deaminated to obtain LA (Figure 4.1). For the first enzymatic step two enzymes were proposed as being able to catalyze the amination of D-ALA into DAVA this enzyme assignment being based on the similarity between D-ALA and the native substrates of these enzymes: i) an adenosylmethionine-KANA transaminase (2.6.1.62) that aminates 7-keto-8-aminononanoate (KANA) and that is present in *E. coli* and in *S. cerevisiae* (among other organisms) and that uses S-Adenosylmethionine (SAM) as the amino donor; ii) a lysine-KANA transaminase (2.6.1.105), present in *B. subtilis*

and that uses lysine as the amino donor (Figure 4.1). The use of adenosylmethionine-KANA transaminase (SAM-KANA transaminase) for the purpose of amination represents the only known case where SAM, usually a methyl donor, donates the amino group for the reaction, resulting in its transamination product 4-(S-adenosyl)-2-oxobutanoate [32]. The enzyme is PLP-dependent and its catalytic mechanism involves the reception of the amino group by PLP in the first half-reaction to yield PMP. In the second half-reaction PMP is cycled back to PLP by donating the amino group to KANA [49]. The *B. subtilis* lysine-KANA transaminase (Lys-KANA transaminase) shares 33% sequence similarity with *E. coli* SAM-KANA, is also PLP dependent and its structure has been elucidated [50]. The second step of the LA route was predicted to be catalyzed by diaminopropionate ammonia lyase (DAPAL, EC 4.3.1.15), that natively catalyses the α,β -elimination of adjacent ammonia groups in diaminopropionate, a substrate structurally similar to 4,5-DAVA. This enzyme is present in *E. coli*, *Pseudomonas sp*, *Salmonella sp* and *Devosia riboflavina* and is classified as a fold type II PLP-dependent [51], [52].

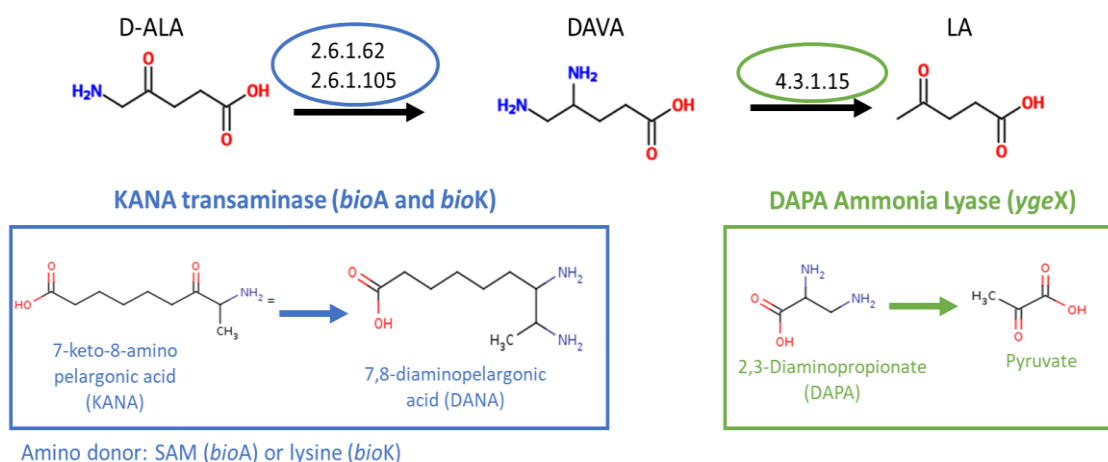


Figure 4.1. LA biosynthetic pathway explored in this chapter. The synthetic pathway is depicted in the top, while the native steps catalyzed by the selected enzymes are depicted in the colored boxes.

The precursor of the alternative LA route explored in this chapter is D-ALA, which is the first committed precursor for tetrapyrrole synthesis, a group of molecules required for the building of heme and vitamin B12. Bio-based production of D-ALA has been explored, mainly due to the possible applications of this chemical as an herbicide or in photodynamic therapy [53]. Two pathways exist for biosynthesis of D-ALA biosynthesis, the C4 and the C5 pathways (Figure 4.2). The C4 pathway involves a condensation of glycine and succinyl-coA, catalyzed by the D-ALA synthase HemA, present in

proteobacteria, yeast or in mammalian cells [54]. The C5 pathway encompasses three enzymatic steps: glutamyl-tRNA synthase, glutamyl-tRNA reductase and glutamate-1-semialdehyde aminotransferase, this being the pathway used by bacteria (other than proteobacteria) or algae [17]. Independently of using the C4 or the C5 pathway for D-ALA synthesis, the synthesis of tetrapyrroles proceeds in all organisms by condensing 2 D-ALA molecules to form porphobilinogen, catalyzed by D-ALA dehydratase (hemB) and 4 molecules of porphobilinogen will form a tetrapyrrole. Thus, 8 D-ALA molecules are required to form a tetrapyrrole. Biotechnological production of D-ALA has been mostly focused in *E. coli*, either by using the native C5 route of *Salmonella arizona* [55]–[57] or by importing the C4 pathway via heterologous expression of D-ALA synthase from *Rhodobacter spheroids* [54] or *S. cerevisiae* [58]. The GRAS amino acid producer species *C. glutamicum* has also been explored as a host for D-ALA production through heterologous expression of the D-ALA synthase from *R. spheroids* [59] or from *R. capsulatus* [60], or by expressing HemA from *S. arizona* [61]. Recently, the endogenous production of D-ALA was improved in *S. cerevisiae* by overexpression of D-ALA synthase (*HEM1*), catalytic inhibition of D-ALA dehydratase and overexpression of aconitase (*ACO2*), which was found to be the rate-limiting enzyme in conditions where glycine is supplied [44].

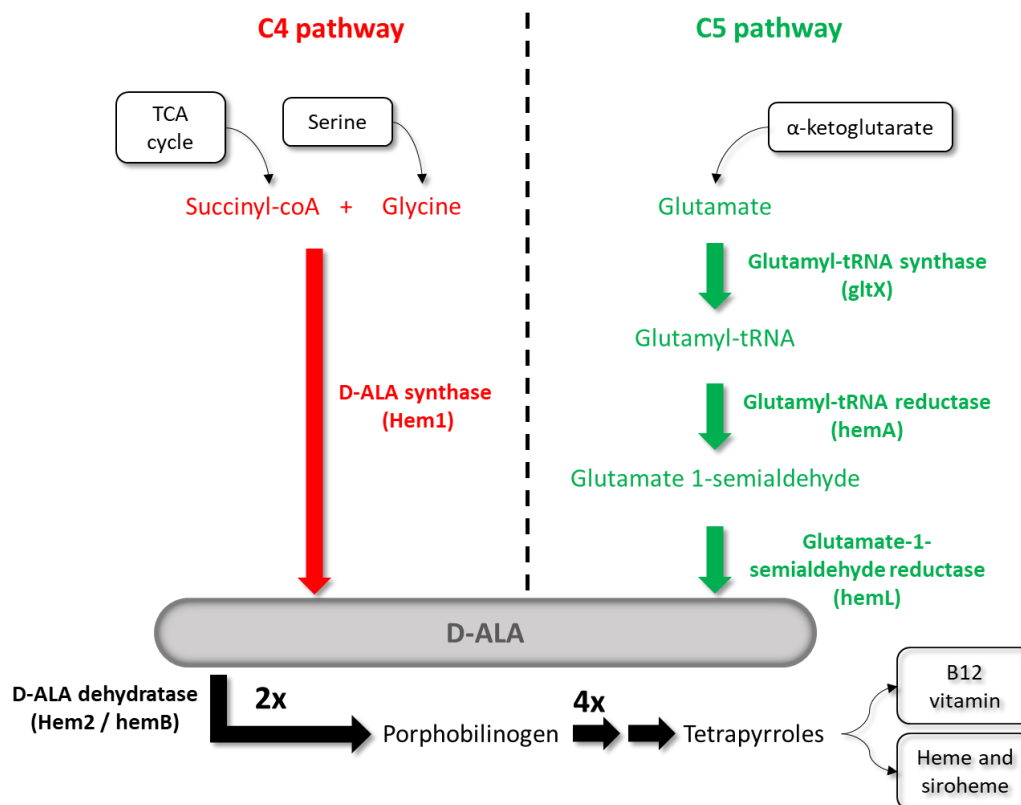


Figure 4.2. C4 and C5 pathway for D-ALA biosynthesis. The steps depicted below the D-ALA box represent the tetrapyrrole biosynthetic pathway, which is common to all organisms.

In this chapter the feasibility of implementing an alternative LA route based on D-ALA in *S. cerevisiae* is explored. Based on the prospecting work described in chapter 3, the route stemming from D-ALA was chosen since it only requires two enzymatic steps and candidate enzymes were identified to mediate these envisaged conversions. Furthermore, the implementation of this pathway in *E. coli* is simplified by the fact that the two enzymes, SAM-KANA transaminase and DAPA ammonia lyase are endogenous to this host (bioA and ygeX, respectively). A variation of the pathway was also attempted using Lys-KANA transaminase from *B. subtilis* (bioK) (instead of the *E. coli* SAM-KANA transaminase). The *in vivo* implementation of this pathway was attempted in *E. coli* at the same time that it was tested *in vitro* the ability of the candidate enzymes to catalyse the envisaged conversions of D-ALA into DAVA and of this molecule into LA. Since the metabolic modelling performed when assessing the performance of the pathway (described in chapter 3) suggested that the pathway could be constrained in *S. cerevisiae* due to D-ALA depletion, the attempt performed to overcome this issue by creating a strain

devoid of the *HEM2* gene, encoding an essential enzyme that uses D-ALA as a substrate, will also be described.

4.3.Materials and Methods

4.3.1.Microorganisms and culture conditions

The *S. cerevisiae* strains were maintained in YPD medium and in MMB (when required to maintain the selection). The composition of YPD medium is 2% glucose (Nzytech), 2% peptone (Nzytech) and 1% yeast extract (Nzytech). The composition of MMB is 2% glucose/galactose (Nzytech/Merck) 0.267% ammonium sulfate and 0.17% Yeast Nitrogen Base without aminoacids (Difco). MMB media is supplemented with the required aminoacids and nucleobases (all from Sigma) to maintain growth and selection in the following concentrations: 20 mg/L histidine, 60 mg/L leucine, 20 mg/L methionine, 20 mg/L tryptophan, 20 mg/L uracil. *E. coli* strains were maintained and cultured in LB medium (Nzytech). Antibiotics were added when necessary, to maintain the selection: 150 mg/L ampicillin (Amp, Nzytech) and 50 mg/L kanamycin (Kan, Nzytech). The lists of strains and plasmids used in this chapter are available in Table 4.1 and Table 4.2, respectively. The pET28a-HEM1 plasmid was kindly shared by Professor Guo-Qiang Chen (Tsingua University)

Table 4.1. List of strains used in chapter 4.

	Name	Description	Source
<i>S. cerevisiae</i>	By4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	Euroscarf
	W303	MATA ura3-1 leu2-3,112 trp1-1 his3-11,5 can1-100 GAL+	Euroscarf
	W303 ΔHem2	W303 Hem2::TAP-KIURA3	Euroscarf
	H1	W303 prHem2 :: 6xtetO-CYC1	This work
	H2	H1 his3-11,5 :: his3-2xtetO-CYC1-tTA	This work
<i>E. coli</i>	DH5	F ⁻ <i>endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20</i> φ80dlacZΔM15 Δ(<i>lacZYA-argF</i>)U169, hsdR17(<i>r_K⁻m_K⁺</i>), λ ⁻	Invitrogen
	BL21(DE3)	F ⁻ <i>ompT gal dcm lon hsdS_B(r_B⁻m_B⁻)</i> λ(DE3 [<i>lacI lacUV5-T7p07 ind1</i>	Invitrogen

Chapter 4– Strategies for the successful implementation of an in silico predicted LA-forming pathway using D-ALA as a precursor

		<i>sam7 nin5</i>) [<i>malB</i> ⁺] _{K-12} (λ ^S)	
BL21 (DE3) pLysS	F ⁻ <i>ompT gal dcm lon hsdS_B(r_B⁻m_B⁻)</i> λ(DE3 [<i>lacI lacUV5-T7p07 ind1 sam7 nin5</i>) [<i>malB</i> ⁺] _{K-12} (λ ^S)	pLysS[<i>T7p20 ori_{p15A}</i>](Cm ^R)	Invitrogen
Rosetta TM BL21(DE3)	F ⁻ <i>ompT gal dcm lon? hsdS_B(r_B⁻m_B⁻)</i> λ(DE3 [<i>lacI lacUV5-T7p07 ind1 sam7 nin5</i>) [<i>malB</i> ⁺] _{K-12} (λ ^S)	pLysSRARE[<i>T7p20 ileX argU thrU tyrU glyT thrT argW metT leuW proL ori_{p15A}</i>](Cm ^R)	Novagen

Table 4.2. List of plasmids used in chapter 4.

Name	Description	Source
pGreg586	CEN plasmid with URA3 marker, expression of N-terminal 6x His tagged protein under control of GAL1 promote	[62]
pGREG-HEM1	The <i>S. cerevisiae</i> HEM1 gene was cloned under the control of the GAL1 promoter	[63]
pBS-KIURA3miniblaster-7xtetOpr	Yeast integrative plasmid (YIP) with 7xtetO-CYC1 TATA with loopable URA3 marker	[64]
pRS304-7xtetO-tTA	YIP with cassette to be inserted in the HIS3 locus (7xtetO CYC1 TATA tTA)	[64]
pETDuet (pDuet)	Bacterial expression vector, ColE1 replicon, lacI gene, Amp. Contains 2 MCS with T7/lac promoters.	Merck
pD-bioA	bioA from <i>E. coli</i> DH5 cloned in MCS1 (EcoRI and HindIII) of pETDuet	This work
pD-bioK	bioK from <i>B. subtilis</i> cloned in MCS1 (EcoRI and HindIII) of pETDuet	This work
pD-ygeX	ygeX from <i>E. coli</i> DH5 cloned in MCS2 of pETDuet (NdeI and XhoI)	This work
pD-bioA-ygeX	ygeX from <i>E. coli</i> DH5 was cloned in MCS2 of pD-bioA (NdeI and XhoI)	This work

pD-bioK-ygeX	ygeX from <i>E. coli</i> DH5 was cloned in MCS2 of pD-bioA (NdeI and XhoI)	This work
pET28a	Bacterial expression vector with T7 lac promoter, KanR	Merck
pET28a-HEM1	Codon-optimized HEM1 from <i>S. cerevisiae</i> was cloned in pET28a vector	[58]
pColA-Duet	Bacterial expression vector with T7 lac promoter, ColA replicon, KanR. Contains 2 MCS with T7/lac promoters.	Merck
pColA-HEM1	pColA inserted with codon optimized ScHEM1 (BamHI and HindIII)	This work

4.3.2. General molecular biology procedures

Plasmids were extracted using the QiaPrep Spin Miniprep Kit (Quiagen) or the Nzy Miniprep Kit (Nzytech). PCR amplification of fragments for cloning or genome integration was performed using Q5 polymerase or Phusion polymerase (NEB). Restriction enzymes used were from Nzytech, except when stated otherwise. Gel, PCR and digestion purifications were performed using the QIAquick PCR & Gel Cleanup Kit (Quiagen) or the Nzy Gelpure kit (Nzytech). Ligations were carried out overnight at 16°C using T4 Ligase (NEB). Yeast transformations were carried out using the Alkali-Cation Yeast Transformation Kit (MP Biomedicals). *E. coli* transformations were carried out using the heat-shock method. Quick yeast genome extractions for confirmation PCRs were performed using the GC prep method [65]. Confirmation PCRs (colony or from GC preps) were performed using OneTaq (Neb) or HorsePower Taq Polymerase (Canvax). The sequence for the primers used in this work can be found in Annex 8.3- Table 8.8.

4.3.3. Genome integrations to obtain the H2 strain

To insert the 7xtetO-CYC1 cassette in the genomic HEM2 promoter the cassette 7xtetO-CYC1-pBS-URA3-BS was PCR amplified from pBS-KIURA3miniblaster-7xtetOpr, the reaction was purified, and it was transformed into W303, where the transformants were selected in MMB without uracil. Genomic integration of the complete

cassette was confirmed by PCR and the positive colonies were grown overnight on MMB media supplemented with uracil and plated on MMB medium supplemented with 5-fluorotic acid (5-FOA) to select candidates that had lost the URA3 loopable fragment. The correct PBS-URA3-PBS looped-out recombination was confirmed by PCR, followed by sequencing of the purified amplicon, obtaining strain H1. To insert the 7xtetO-CYC1-tTA cassette in the HIS3 locus, the pRS303-7xtetO-tTA plasmid was digested with NheI, gel-purified and transformed into the H1 strain, where the transformants were selected in MMB His- media. Correct integration at the HIS3 locus was confirmed by PCR, followed by sequencing of the amplicon, obtaining strain H2. The integration was confirmed to be stable in non-selective media by passaging in YPD and surveying the region of the HIS3 locus and the HEM2 promoter by confirmation PCR and sequencing.

4.3.4. Cloning in of LA pathway enzymes in pETDuet plasmids

For cloning in pDuet plasmids, the bioA and bioK coding regions were amplified from genomic DNA (gDNA) of *E. coli* DH5 α and *B. subtilis* Py79 and cloned into pDuet by restriction-ligation with EcoRI (NEB) and HindIII (NEB). The ygeX genomic fragment was amplified from gDNA of *E. coli* DH5 α and cloned into pDuet, pD-bioA and pD-bioK by restriction-ligation with NdeI and XhoI. For cloning in pColADuet plasmid, the HEM1 codon optimized fragment was shuttled from pET28a-HEM1 to pColADuet plasmid by restriction ligation with BamHI/HindIII.

4.3.5. Yeast fermentations to over-produce D-ALA

By4741 cells transformed with PGreg586 or pGreg586-HEM1 were pre-cultured overnight in MMB glucose and then re-inoculated into this same medium until mid-exponential phase (OD_{600nm} of 0.5-0.8). Afterwards, an appropriate volume of this inoculum was inoculated (at an initial OD_{600nm} of 0.1) in 50 mL MMB with 2% of galactose in 250 mL shake flasks. The cultures were maintained at 37° and 250 rpms, and OD_{600nm} measurements and supernatant samples were taken after 24h, 48h or 72h. The same procedure was used for the fermentations that were performed with the W303 and H2 strain, with the exception of using MMB 2% glucose and that 250, 500 and 1000 ng/mL Doxycycline (Dox) were added to the growing cultures when they reached mid exponential phase (after 8h). The same procedure was followed in MMB 2% galactose when W303 and H2 cells transformed with the pGreg586-HEM1 plasmid were used for

fermentations, in which case the Dox pulse in the exponential phase was carried out approximately after 24 hours.

Fermentations to test different Dox pulse times and concentrations with the W303 and H2 strains transformed with the pGreg586-HEM1 plasmid were carried out in microplates. Three repression conditions were tested: repression at time of fermentation start (1), repression at the exponential phase (2) and pre-repression in the inoculum (3). For condition 1 the microplate was prepared with MMB galactose containing the corresponding double concentrations of dox: 2000, 1000, 500 and 0 ng/mL. Overnight pre-inocula and 8h inocula were carried out in 10 mL capped glass tubes with 5 mL of culture in MMB glucose, at 37°C 250 rpms. When the glass tube inocula reached OD_{600nm} of 0.5 to 0.8, cellular suspensions were prepared with OD_{600nm}=0.2 in MMB galactose and were used to inoculate the microplate. For condition 2 the same pre-inoculum and inoculum procedure were carried out. When the inoculum reached ODs of 0.5 to 0.8, a cellular suspension of OD_{600nm}=0.1 in MMB galactose was prepared and used to inoculate the microplate directly. After 24h the corresponding volumes of dox stock and DMSO were applied to the plate wells to obtain the final dox concentrations of 0, 250, 500 and 1000 ng/mL. For condition 3 overnight pre-inocula were carried out in glass tubes and the inoculum was carried out in a microplate with MMB glucose and the corresponding dox concentrations. After 8h hours of inoculum growth, 50 µL of each well was used to inoculate a new microplate with MMB galactose and the corresponding dox concentrations. The fermentations in MMB galactose+raffinose were carried out in microplates using the same procedure described for condition 2.

D-ALA in the supernatants was quantified by the modified Ehrlich method [66] that was miniaturized. The modified Ehrlich reagent was prepared fresh by dissolving 1 g of p-dimethylaminobenzaldehyde (Sigma) in 42 mL glacial acetic acid and adding 8 mL of perchloric acid (Milipore). The assay was carried out in PCR microtubes, to which it was added 3.5 µL of acetylacetone (Sigma), 40 µL of acetate buffer 2M pH=4.6 and 30 µL of sample. The tubes were heated in a PCR thermocycler (Biorad) to 100°C 10 min and let cool down to 22°C 10 min. After cooling, 70 µL of modified Ehrlich reagent was added to the tubes and 100 µL of the mixture was transferred to a microplate to read the absorbance at 556 nm in a microplate reader (BMG Labtech). For each assay two calibration curves were prepared, with D-ALA solutions at concentrations of 5, 10 and 20 mg/L.

4.3.6.Assessment of HEM2 expression in W303 and H2 cells

W303 and H2 cells were cultured overnight in MMB glucose and the cells were used to inoculate flasks with new fresh media at an $OD_{600nm}=0.1$. When OD_{600nm} reached 0.5 the appropriate volume of cellular culture was filtered to inoculate 50 mL of MMB glucose with an initial $OD_{600nm}=0.1$. After 24h of growth the OD_{600nm} was registered to verify the cells had reached stationary phase and Dox was added to the media to obtain concentrations of 250, 500 and 1000 ng/mL. 10 mL of each culture was harvested after 2h and 24h of the Dox pulse and the samples were centrifuged at 5000 rpm for 5 min at 4°C. The cells were stored at -80°C for future RNA extraction. The MasterPure™ Yeast RNA Purification Kit (Lucigen) was used to extract RNA from the samples and the cDNAs were prepared using the QuantiTect Reverse Transcription Kit (Qiagen). The obtained cDNAs were used for Reverse Transcription Quantitative PCR (RT-qPCR) with the Nzy qPCR Green Master Mix kit (Nzytech) in an Applied Biosystems 7500 RT-qPCR machine. The results were analyzed with the $\Delta\Delta C_T$ method, where ALG9 was the housekeeping gene-

4.3.7.In vitro assays of the BioA, BioK and YgeX enzymes

4.3.7.1.Preparation of crude cell extracts

E. coli BL21(DE3) cells were transformed with plasmids pDuet, pD-bioA, pD-bioK and pD-ygeX and the transformants inoculated overnight in LB medium. In the next day an appropriate volume was used to inoculate 20 mL of LB (in a shake flask) at an initial OD_{640} of 0.05. The cells were left to grow at 37°C 250 rpms until an OD_{640} of 0.5 was reached. At this point 0.3 mM IPTG was added to the culture and after 5h of incubation under the same conditions (37°C, 250 rpm), 5ml of culture were harvested and centrifuged at 5000G, 10 min and 4°C. The pelleted cells were stored at -80°C until further use. To obtain the crude cell extracts, presumably enriched in BioA, BioK or in YgeX, the cells were resuspended in 500 μ L lysis buffer (50 mM Tris / HCl, pH 8.0, 10 mM 2-mercaptoethanol, 0.2 mM PLP), sonicated (7 cycles of 10 second pulses with 1 min cooling in between cycles) and centrifuged (10 000 g; 20 min; 4°C). The clarified supernatants were recovered and the total protein content quantified using the Bradford assay with Bovine Serum Albumin (BSA) standard solutions for calibration curves (Biorad).

The overexpression of bioA, bioK and ygeX was confirmed by SDS-PAGE gels

prepared using a 12.5% and a 4% concentration gel. The samples were diluted 1:2 in Laemmli buffer (20% glycerol, 4%SDS, 100 mM Tris-HCl pH=6.8, 0.2% bromophenol blue, 200 mM DTT) before application. The running buffer contained 0.25M Tris-base, 1.92M glycine and 1% SDS.

4.3.7.2.Enzymatic assays

The experimental conditions used for the enzymatic assays of BioA, BioK and YgeX (alone or in combination) were based on those previously described for BioA [67]–[69], BioK [31] and YgeX [52]. The assays were carried out at 37°C in 50 mM potassium phosphate buffer at pH 8 supplemented with 5 mM dithiothreitol (DTT) and 0.2 mM PLP in a total volume of 1200 µL. For reaction assays with BioA and BioK, 5 mM of the amino donor (SAM or lysine) and 10 mM of the substrate (D-ALA or KANA) were added to the reaction mixture. In the case of the YgeX assay 10 mM of DAPA were used. For all assays the reaction vessel containing the buffer with the cofactors and substrates was pre-warmed at 37°C and the reaction was started by adding the crude cell extract to the reaction buffer in a volume ratio of 1:5. For the double enzymatic assays (that is assays that used crude cell extracts enriched in BioA/BioK and YgeX) a 1:1 mix of the extracts was previously prepared. BioA and BioK reactions were followed by accompanying the ratio of absorbance at 420 nm and 335 nm, while in YgeX reactions the ratio of absorbance at 432 nm and 412 nm was followed. The reactions were stopped by addition of ¼ of volume of 15% trichloroacetic acid (TCA) and centrifuged to recover the supernatant and remove the precipitated protein.

The supernatants from the enzymatic assays were analyzed by High Performance Liquid Chromatography (HPLC) by separating 10 µL of the culture supernatant in an Aminex HPX- 87H column (Biorad) eluted with 0.05% sulfuric acid at a flow rate of 0.6 mL/min with a UV detector set at 210 nm.

4.3.7.3.Assessment by LC-MS of the extract obtained after the enzymatic step

For the KANA transaminase assay with BioK prepared for LC-MS analysis the buffer was exchanged to 10 mM ammonium bicarbonate pH=8.3 to become MS-compatible.

The LC-MS analysis was performed by Maria Conceição Oliveira from Centro de Química Estrutural. Enzymatic assay supernatants were analyzed by LC (UHPLC Elute) interfaced with a QqTOF Impact II mass spectrometer equipped with an ESI source

(Bruker Daltonics). Chromatographic separation was performed on a HILIC XBridge BEH column 100 Å (150 mm x 2.1 mm, 2.5 µm particle size; Waters). Mobile phase consisted in ammonium acetate 10 mM pH 8.6 (A) and ammonium acetate 10 mM pH 8.6 in acetonitrile with 2% of water (B). The used elution gradient was as follows: 0 min 95 %; at 2 min; 80 % (B); 5 to 10 min 40 % (B); and 12-18 min 95% (B). The injected volume was 10 µL, the flow rate was 300 µL min⁻¹, and the temperature of the column and autosampler were maintained at 40 °C and 8 °C, respectively.

The high-resolution mass spectra were acquired in the ESI in positive mode, the optimized parameters were set as follows: ion spray voltage, + 4.5 kV; end plate offset, 500 V, nebulizer gas (N₂), 2.8 bars; dry gas (N₂), 8 Lmin⁻¹; dry heater, 200 °C. Internal calibration was performed on the high-precision calibration mode (HPC) with a solution of sodium formate 10 mM introduced to the ion source via a 20 µL loop, at the beginning of each analysis using a six-port valve. Acquisitions were performed in the full scan mode with a range between m/z 100-1000, and a rate of 1 Hz. To evaluate the variation of the signal intensity across the entire measurement intensity of the raw files, samples were spiked with a standard solution of caffeine 10⁻⁵ M, and three replicate injections were analyzed. Data acquisition and processing were performed using DataAnalysis 4.4 software (Bruker Daltonics).

4.3.8. In vivo assays to obtain LA in *E. coli* cells

4.3.8.1. Using cells that over-produce D-ALA and that over-express BioA/BioK and YgeX

To enable D-ALA overproduction and expression of the LA pathway enzymes in *E. coli*, BL21(DE3) cells were transformed with pColADuet-HEM1 and pD-bioA-ygeX or pD-bioK-ygeX. The experimental conditions used here for enzyme overexpression were based on a previous work where HEM1 from *S. cerevisiae* was also expressed in *E. coli* BL21 (DE3) cells [58]. The D-ALA production media was LB adjusted with phosphate buffer (27.8 mM K₂HPO₄·3H₂O, 72.2 mM KH₂PO₄), 3 g/L glycine and 6 g/L succinic. The concentrations of Kan and Amp were doubled to account for the increased antibiotic resistance in high phosphate concentrations: 100 mg/L and 300 mg/L, respectively. For one-step LA fermentations, 15 mL cultures were prepared with an initial OD=0.05 from an overnight inoculum and were allowed to grow at 37°C for 5 hours, after which induction was carried out by addition of 0.05 mM IPTG and transfer of the culture flasks to 30°C. After 18h the cultures were fed with 1% glucose, 5 g/L glycine and 10 g/L

succinic acid. Samples were collected for SDS-PAGE, HPLC and Ehrlich quantification after 24 and 48 hours. The induction protocol for LA pathway enzymes was subsequently modified to addition of 0.1 mM of IPTG at 5 hours after culture start and subsequent transfer to 30°C. The overexpression of BioA, BioK, YgeX and Hem1 was confirmed by SDS-PAGE.

4.3.8.2. Using a two-step bioconversion exploring cells that over-produce D-ALA which is used as a growth supplement for cells over-expressing BioA/BioK and YgeX

For the over-production of D-ALA, *E. coli* BL21(DE3) cells transformed with plasmid pET28-HEM1 were cultivated overnight, inoculated (at an initial OD₆₄₀ of 0.05) in D-ALA production media (described in 4.3.8.1) and incubated at 37°C 250 rpm. After 5 hours, 0.05 mM IPTG were added to the broth and the cells were incubated at 30°C for. After 24 hours of induction the culture was fed with 1% glucose, 5 g/L glycine and 10 g/L succinic acid. The fermentation broth was collected by centrifugation after 48 hours of induction with IPTG and filtered. The amount of D-ALA in the fermentation broth at this point was quantified using the Ehrlich reaction method, described in 4.3.5.

For the second step, *E. coli* BL21(DE3) cells transformed with pColA and pduet, pD-bioA-ygeX or pD-bioK-ygeX were cultivated overnight in LB and then: i) directly inoculated in the fermentation broth enriched in D-ALA obtained from the fermentation undertaken with cells over-expressing *HEM1* supplemented with 0.1 mM IPTG and Amp; ii) or the cells were inoculated (at an OD₆₄₀ of 0.05) in fresh LB medium and incubated at 37°C and 250 rpm for 5 h. After this time 0.1 mM IPTG were added to the broth and after 5 h the cells were collected by centrifugation and finally used to inoculate the D-ALA enriched fermentation broth obtained from the first step using *E. coli* cells transformed with pHEM1. In both situations *i* and *ii*, the cultures were incubated for 5 days at 30°C and samples were taken regularly to quantify the amount of LA eventually accumulated in the broth (by HPLC) or the consumption of D-ALA (based on Ehrlich quantification method, described above).

4.4.Results and Discussion

In the first subsection of this chapter the LA pathway it will be described the results obtained when the candidate enzymes were tested *in vitro* for their ability to catalyze the desired synthetic reactions. Afterwards it will be described the results obtained when the entire pathway was assembled *in vivo* in *E. coli* and the results obtained in the engineering of a *S. cerevisiae* strain with an improved D-ALA intracellular pool.

4.4.1.Can the candidate enzymes identified in the *in silico* analysis promote the conversion of D-ALA into D-AVA and of this into LA?

The enzymes predicted to catalyze the two reactions of the LA forming pathway (Figure 4.1) were tested for their ability to accept the synthetic substrates using *in vitro* assays performed with whole-protein extracts obtained from *E. coli*. For that the KANA transaminase from *E. coli* (encoded by the *bioA* gene, SAM-KANA transaminase) and from *B. subtilis* (encoded by *bioK* gene, Lys-KANA) and DAPA ammonia lyase from *E. coli* (encoded by *ygeX*) were cloned in the IPTG-inducible plasmid pETDuet. Overexpression was carried out in BL21(DE3) cells using 0.3 mM IPTG, at 37°C for 5 h, to obtain whole-cell extracts, presumably enriched in BioA, BioK or YgeX. The success of the overexpression of the enzymes was confirmed by SDS-PAGE (Figure 4.3).

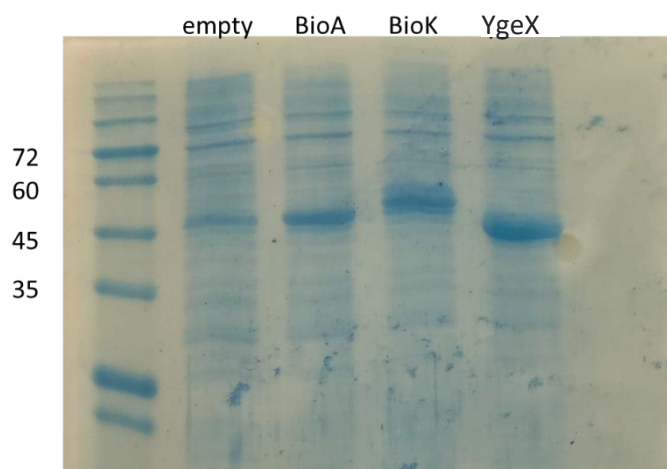


Figure 4.3. SDS-PAGE with 5 μ L of crude cell extracts of BL21 expressing pETDuet, pETDuet-bioA, PETDuet-bioK and PETDuet-ygeX. The expected sizes of the proteins are: bioA 47 kDA, bioK 50 kDA and ygeX 43 kDA.

To test if the KANA transaminases BioA and BioK could catalyze the conversion of D-ALA in DAVA, crude cell extracts enriched in these proteins were used in an enzymatic assay that had D-ALA as a substrate. To control whether the conditions that were being used for the enzymatic assays were favourable for the enzymes, these same protein extracts were put in the presence of KANA expecting to obtain the expected DANA product. The progress of the reaction was followed by monitoring the variations in absorbance at 420 and 335 nm, considering that KANA transaminase displays an absorption spectrum typical of PLP-dependent enzymes reaching a maximum at 420 nm and a shoulder at 335 nm (these peaks are attributable to the PLP and PMP form, respectively [70] (Figure 4.4A). The PLP and PMP forms are expected to cycle during the KANA transaminase reaction (as detailed in the introduction) since in the first half reaction the PLP transamination is responsible for the formation of enzyme-PMP intermediate (aldimine, absorption peak at 335 nm) and the SAM keto-acid. In the second half reaction the amino group is donated to the substrate KANA, regenerating the enzyme-PLP complex (ketimine, absorption peak at 420 nm) and forming DANA [32]. Having this in mind the absorption at 335 and 420 nm was recorded, with the anticipation that recognition of the substrate by KANA transaminase could create a shift in the chemical equilibrium of the enzyme-bound PLP/PMP forms that would be detectable in the 420/335 absorbance ratio. The results obtained for the different enzymatic assays performed are shown in Figure 4.4. For all reactions, independent of the substrate present,

it was observed that the initial 420/335 ratio is lower in the reactions that have extracts enriched in BioA than in the extracts obtained from *E. coli* control cells (transformed with the cloning vector pDuet) or enriched in BioK, attributable to the different amounts of enzyme present in each extract in the initial moment of the reaction (Figure 4.4B-D). To take this effect into account, in Figure 4.4E-G are shown the changes in the A420nm/A335nm ratio using as a reference the initial ratio obtained for each extract. With this visualization mode, it is possible to observe that in the reactions where no substrate is added the percentage ratio displays an initial change until an equilibrium is reached, which is maintained throughout the remaining of the reaction. The initial change for the BioA extracts lasts longer than for the pDuet and BioK extracts (10 min vs 2 min), although the significance of this different kinetics is unknown (Figure 4.4E). No major differences can be observed in the overall 120 minutes time-course between the reactions with no substrate or with KANA, however, quick oscillations of the ratio are observed in the reactions that use KANA and BioK (Figure 4.4F). The fact there are other PLP-dependent enzymes in the extract used which can affect the oscillations observed in the Abs_{420nm}/Abs_{335nm} ratio complicates the establishment of definitive conclusions, however, it is interesting that the changes are much more visible when KANA is used in the presence of BioK suggesting that at least in this case the BioK is being able to catalyse the conversion of KANA into DANA. In the reactions undertaken with KANA and BioA-enriched extracts no significant changes were observed in the Abs_{420nm}/Abs_{335nm} ratio, suggesting that there is little or no activity of this enzyme BioA (Figure 4.4F). In all reactions in which D-ALA was used as a substrate there is a steady decrease of the Abs_{420nm}/Abs_{335nm} ratio over time, independent of the extract added to the reaction mixture. This decrease, which is not mediated by the presence of the enzymes, may result from some chemical degradation of D-ALA during the incubation time (Figure 4.4F). On the overall the results obtained with the monitoring of the Abs_{420nm}/Abs_{335nm} ratio did not allowed a clear indication whether or not the reactions were occurring in the presence of D-ALA, however, the fact that there are oscillations observed in the BioK enriched extract and not in the others (including in BioA enriched extracts) is interesting.

Chapter 4– Strategies for the successful implementation of an in silico predicted LA-forming pathway using D-ALA as a precursor

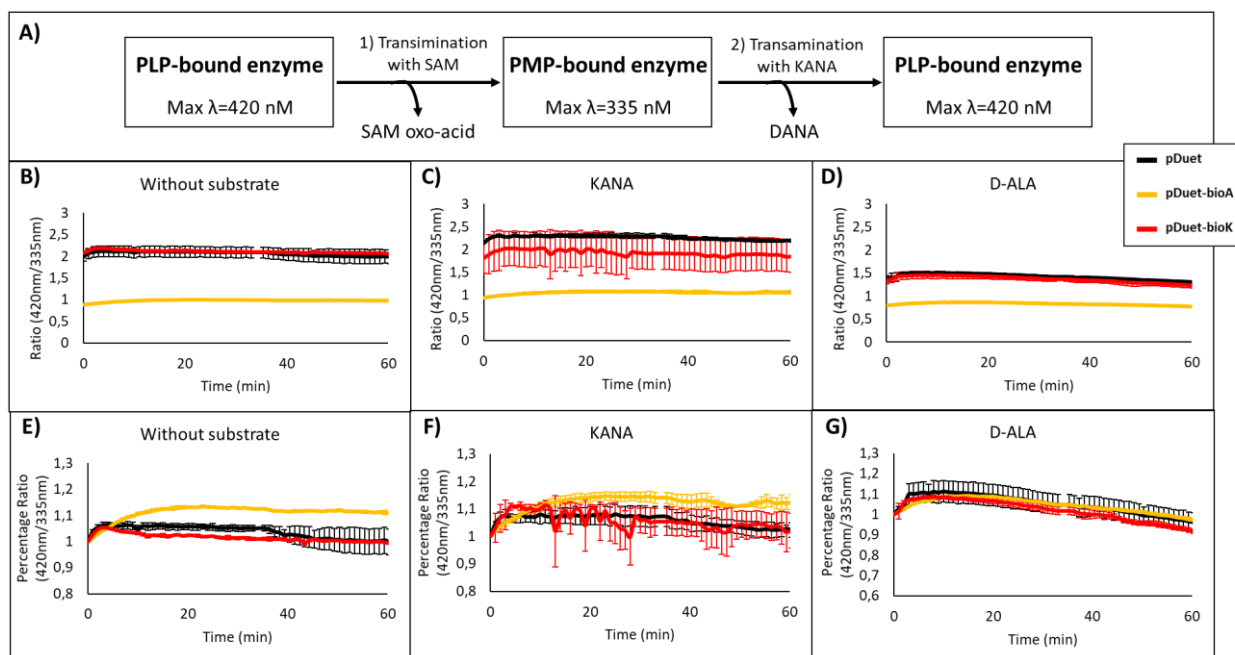


Figure 4.4. Progression of absorbance at 335 nm and 420 nm during KANA transaminase enzymatic assays. A) Two-step mechanism of KANA transaminase, with the depiction of PLP and PMP-enzyme forms and their maximum absorption wavelengths. B-D) Ratio of absorbance Abs_{420nm}/Abs_{335nm} . E-G) Percentage ratio Abs_{420nm}/Abs_{335nm} : calculated by dividing the ratio at all time points with the initial ratio.

To further determine whether BioA and BioK catalyze the transformation of D-ALA into DAVA, we analyzed the enzymatic extracts by LC-MS with the objective of identifying the presence of DAVA. For that, the enzymatic reactions were performed as detailed above with the difference that the time was extended to 45 minutes and 1h of reaction. LC-MS provides an important advantage here since there are no available standards for DANA and DAVA and therefore the use of an experimental methodology that might provide evidences concerning the identity of the compounds is essential. During the time of this work, it was only possible to obtain LC-MS data for the supernatants obtained in the enzymatic assays enriched in BioK and in the corresponding control assays (that used protein extracts obtained from cells harboring only the cloning vector pDuet). The LC-MS spectra obtained confirmed the presence of DANA in assays undertaken with KANA confirming that under the conditions used for the assay the BioK enzyme is active (Figure 4.5). These results are in agreement with the suggestive results obtained with monitoring the ration in absorbances at 430 and 330 nm described above. A small signal of DANA was also detected in the reactions that were undertaken with the cloning vector pDuet (Figure 4.6) likely resulting from the endogenous activity of *E. coli* endogenous enzyme BioA.

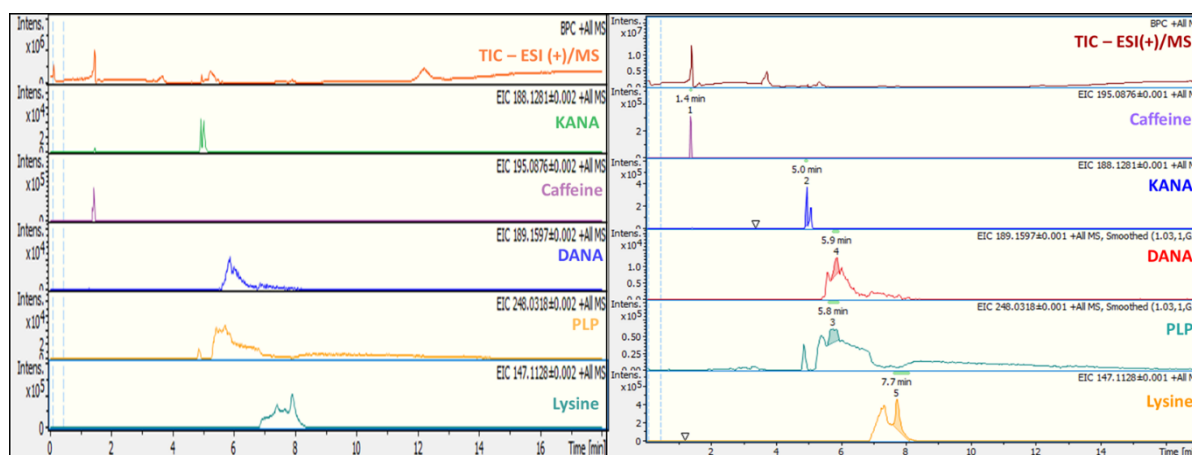


Figure 4.5. LC-MS analysis of samples of KANA transaminase assays with KAPA as the substrate. A) Sample from assay with pDuet extract and KANA, taken after 1 h of reaction. B) Sample from assay with bioK extract and KANA, taken after 45 min of reaction. For each sample it is depicted the total ion chromatogram (TIC) and the extracted ion chromatograms (EICs) for caffeine (internal standard), KANA, DANA, PLP and lysine. Note the different scales for each compound.

We have afterwards proceeded with the analysis of the extracts obtained using D-ALA as the substrate (Figure 4.6). The results obtained clearly confirmed the presence of DAVA in the supernatants of the reactions that were performed with the BioK-enriched extract confirming that indeed this enzyme can catalyze the envisaged synthetic conversion of D-ALA into DAVA. The presence of DAVA was also registered in the control extracts, albeit at a much lower extent than the one detected for BioK (likely reflecting the much lower abundance of BioA in the non-induced pDuet extracts, compared to the induced BioK extracts). Although a more quantitative analysis will be required to determine exactly the reaction rates of BioK (and eventually also of BioA), the results obtained confirm the possibility of these enzymes catalyzing the envisaged conversion of D-ALA into DAVA.

Chapter 4– Strategies for the successful implementation of an in silico predicted LA-forming pathway using D-ALA as a precursor

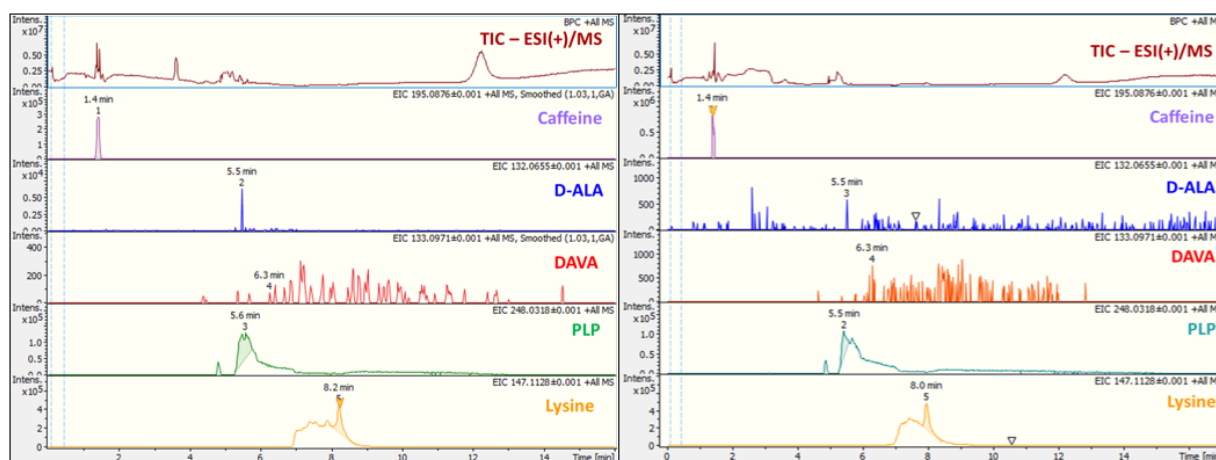


Figure 4.6. LC-MS analysis of samples of KANA transaminase assays with D-ALA as the substrate. A) Sample from assay with pDuet extract and D-ALA, taken after 1 h of reaction. B) Sample from assay with bioK extract and D-ALA, taken after 45 min of reaction. For each sample it is depicted the total ion chromatogram (TIC) and the extracted ion chromatograms (EICs) for caffeine (internal standard), D-ALA, DAVA, PLP and lysine. Note the different scales for each compound.

Considering the results obtained with the LC-MS assays that confirmed the production of DAVA by the BioK (and eventually by the BioA) enzyme it would be optimal to confirm that DAPA ammonia lyase (encoded by *E. coli* ygeX) can catalyze the second pathway reaction, where DAVA is converted to LA. However, since a DAVA standard was not available, assays performed with protein extracts enriched in BioK (or BioA) and YgeX were devised with the idea that the activity of BioK over D-ALA could provide the DAVA needed for YgeX to work with. To demonstrate that YgeX is active under the conditions used in the assays, protein extracts enriched in YgeX were put in contact with the native substrate, DAPA, being expected to obtain pyruvic acid as a deamination product. Indeed, it is confirmed that pyruvic acid is only formed in the reactions where DAPA was provided as a substrate and more significantly when ygeX extracts initiate the reaction (Figure 4.7B). Another evidence of YgeX functionality were the changes observed in Abs_{423nm}/Abs_{412nm} since DAPAL has been demonstrated to have an absorption maximum at 412 nm due to PLP bound enzyme, suffering a red shift to 423 nm upon addition of the substrate [51]. In the reactions where DAPA was present there was an initial perturbation of the Abs_{423nm}/Abs_{412nm} ratio, followed by a steady maintenance of the ratio after the first 5 minutes of reaction, the perturbation being more pronounced in the reactions where the ygeX extract was added than in the ones where the pDuet extract was added (Figure 4.7A). This observation, coupled with a steady maintenance of the Abs_{423nm}/Abs_{412nm} ratio in the reactions where no substrate was provided, indicates that the absorbance changes occurring in the first 5 minutes of reaction

are due to the ygeX-mediated deamination of DAPA. The reaction rates calculated at 1 minute of reaction indicate a 29-fold increase in the ygeX extracts-lead reaction when compared to the pDuet extracts (10.4 ± 0.617 mM pyruvic/mg/min vs 0.355 ± 0.0655). Thus, it can be concluded that the overexpressed ygeX in the crude cell extracts maintains the enzymatic activity in the conditions used in this assay.

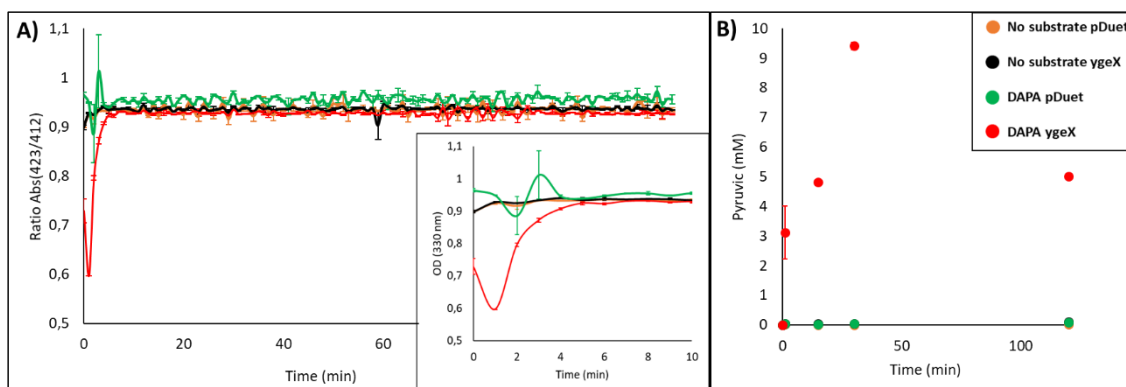


Figure 4.7. DAPA ammonia lyase enzymatic assays with crude cell extracts prepared from BL21(DE3) cells carrying pDuet and pD-ygeX. A) Ratio of absorbance Abs_{423nm}/Abs_{412nm}; inset depicts initial 10 minutes of reaction in larger magnification. B) Pyruvic acid concentration.

With the confirmation that both BioK (and possibly BioA) and YgeX maintain enzymatic activity in the conditions used in this work, it was possible to devise a double double enzymatic assay combining extracts enriched in BioA/BioK and also in YgeX and search the supernatant for the presence of LA. The supernatants at the beginning of the reaction (t_0), after 1 hour and 24h of reaction were analyzed by HPLC (Figure 4.8). The longer reaction time of 24h was chosen in order to take into account a presumed low affinity of the enzymes for their “synthetic” substrates. The chromatograms obtained did not allowed us to visualize a peak that could correspond to LA in the supernatants after 24h of reactions (Figure 4.8). It is possible that this absence results from a low concentration of LA in the supernatants, below the HPLC limit of detection. To take more definitive conclusions it would be required to analyze the supernatants by LC-MS. As future possibilities of work, it is suggested to perform the assays with a higher concentration of D-ALA, increase the enzyme concentration in the assay, for example, by using purified protein extracts containing BioK (or BioA) and YgeX.

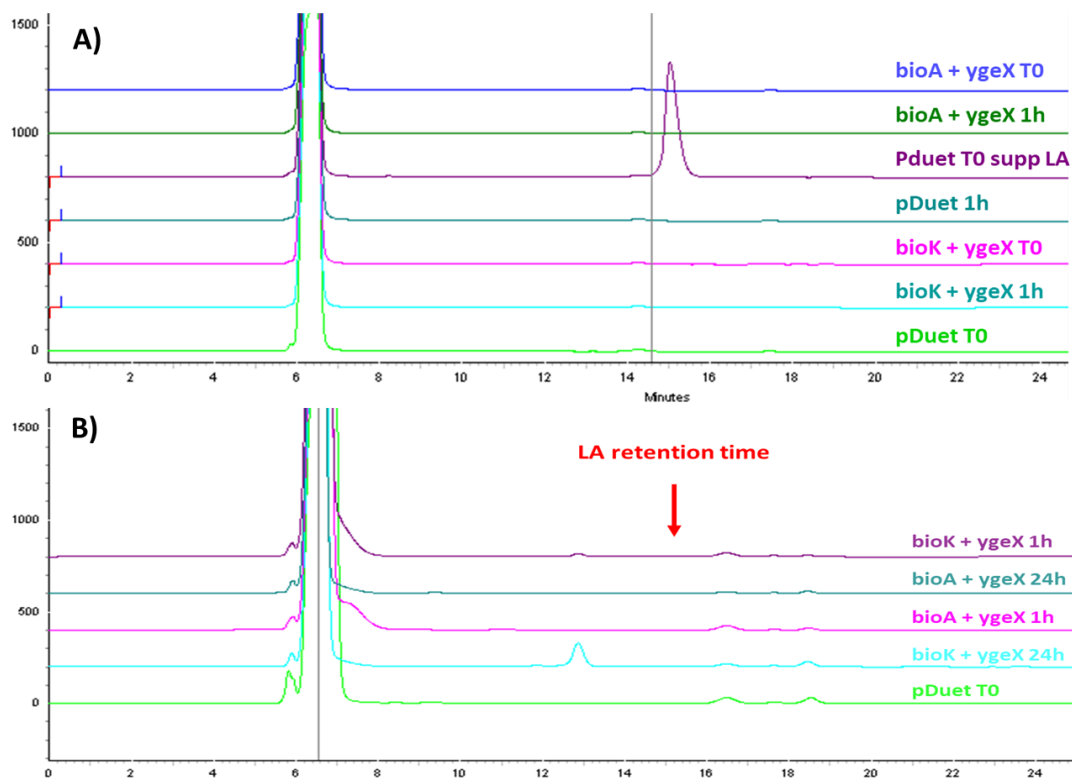


Figure 4.8. Stack depiction of HPLC traces of supernatants from double enzymatic assays carried out with 3 combinations of crude cell extracts: pDuet, PD-bioA+pD-ygeX and pD-bioK+pD-ygeX. A) Samples were taken at the initial time and after 1h. B) Samples were taken at 1h and 24h. LA retention time was 15 min.

Overview

In case is confirmed that DAPA cannot use DAVA as a substrate, engineering of the YgeX enzyme could be attempted, for example targeting residues that are important for substrate recognition, or by directed enzyme evolution. Even in the case of BioA and BioK, the affinity for the synthetic substrate D-ALA can be improved by rational or directed evolution of the enzymes. In what concerns rational modification, the structure of DAPA ammonia lyase has been elucidated and a carboxy binding loop has been identified in the active site, likely involved in substrate recognition [52]. One large difference between DAPA and DAVA is the fact that amino groups in the native substrate are located alpha and beta to the carboxy group, in opposition to the synthetic substrate, where these are located gamma and delta to carboxy group. Considering this, modifications in the enzyme that may enable a larger distance between the carboxy

binding loop and the PLP-interacting loop may improve DAPA ammonia lyase specificity to substrates with a longer carbon chain such as DAVA. In the case of *E. coli* KANA transaminase BioA, various studies have resulted in the identification of at least one residue in binding to the carboxy group of KANA and three residues involved in substrate recognition via a network of hydrogen bonds that include the 8-amino group in KANA [71], [72]. Enzymatic activity and crystal structures of mutants in two of the three residues suggests that indeed these residues are important for recognition of the KANA substrate but not the SAM one, making them potential candidates to be “moved around” in the active site and enable a correct placement of a smaller substrate such as D-ALA [49]. For both KANA transaminase and DAPA ammonia lyase, one main difference between the synthetic substrates and the native ones is the different length of the carbon skeleton, implying that the substrate binding pocket of the enzymes is not accommodated to the size of the synthetic substrate. Engineering of substrate binding pocket to create steric hindrance has been used to switch the specificity of an enzyme to smaller substrates, such as the case of the aldehyde deformilating enzyme mutants that were engineered to have higher specificity to low chain aldehydes [73]. For any case of rational enzyme engineering to improve substrate specificity, molecular docking tools can help in finding the residues that are most determinant for substrate specificity and simulate binding of mutant enzymes to synthetic substrates. PocketOptimizer, as an example, was developed to use an enzyme structure and a small binding molecule as a starting point and to predict mutations in the binding pockets that will increase affinity to the new ligands [74].

An alternative to rational engineering of enzymes is enzyme directed evolution, a process by which a library of enzyme variants is screened for the new enzymatic capabilities and a series of iterations of this process enables the selection of an enzyme with new properties, including activity in unusual environment, improved thermostability, new enantioselectivity and improved activity to new substrates, the property that is most interesting in this context [75]. One important factor for the application of directed enzyme evolution is the existence of a high-throughput screening methodology, be it based on fluorescence, colorimetry or microbial complementation, which can be screened through agar plates, microtiter plates, microfluids devices or FACS [76]. While for the first synthetic substrate, D-ALA, there is a colorimetric assay available, the modified Ehrlich method, the concentrations to be used in a screening methodology would have to be greatly increased to be captured by the method. For the second synthetic step, the deamination of DAVA to LA, the product has been routinely quantified by HPLC and no other method that is high throughput has been developed. The

lack of high-throughput screening methodologies for the individual reaction steps can be surpassed by focusing on whole pathway directed evolution and using microbes as natural sensors for LA presence in the fermentation broth, a hypothesis that is approached in the discussion of the following subchapter, as it is applicable to *in vivo* settings.

4.4.2. *In vivo* assembly of the whole DALA - LA conversion pathway in *E. coli*

In this subsection we describe the attempts made concerning the *in vivo* of the whole D-ALA – LA conversion pathway. This *in vivo* assembly was performed in *E. coli* cells harboring plasmids pD-bioA/K-ygeX and pColA-*HEM1*, which allows over-expression of the BioK/BioA and YgeX proteins and also the over-expression of ScHem1 to improve the D-ALA intracellular pool [58]. Coupling of D-ALA overproduction and over-expression of the BioA/BioK/YgeX enzymes was performed simultaneously or in a subsequent manner (two-step strategy), as it will be detailed below.

In the first approach the enzymes BioA/BioK and YgeX and the codon optimized ScHem1 were overexpressed in BL1(DE3) cells under the control of the lac/T7 promoters. The different combination of pathway enzymes and D-ALA overproduction resulted in 7 different experimental setups to be tested as detailed in (Figure 4.9A). To induce the over-expression of the enzymes the broth was supplemented with 0.05 mM IPTG at 30°C after 5h of inoculation of the *E. coli* transformants in the medium. This approach was used since previous reports have shown that these are favorable for *ScHEM1* over-production in *E. coli* and consequently result in accumulation of D-ALA in the broth [58]. HPLC analysis of the culture broth could not detect LA in any of the supernatants examined, not even in those obtained with cells over-producing the overall pathway (settings 5 and 6 in Figure 4.9C) suggesting that the pathway was not successfully assembled. A noticeable aspect was the observation that D-ALA was not significantly accumulated in the broth not even in the cells that were only supposedly over-expressing Hem1 (1 in Figure 4.9B).

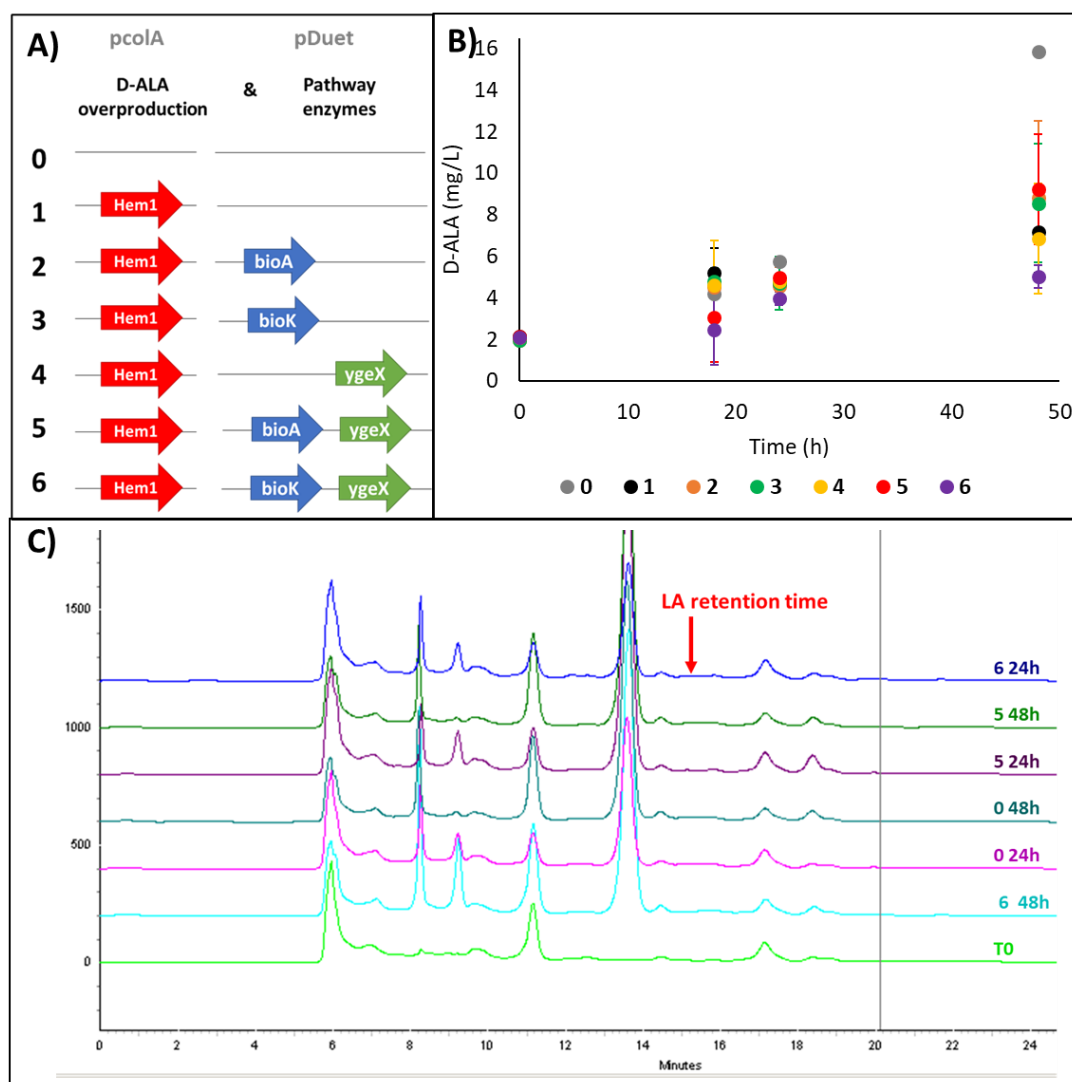


Figure 4.9. Tentative *in vivo* assembly of LA biosynthetic pathway in *E. coli*. A) Outline of the strategy used to couple D-ALA overproduction with the pathway enzyme in two sets of compatible plasmids. B) Extracellular D-ALA concentration. C) HPLC trace of supernatants after 24h and 48h of fermentation with strains 0, 5 and 6. LA retention time in fermentation supernatants is depicted (15.2 min).

To understand what could be preventing the assembly of the LA biosynthetic pathway and the overproduction of D-ALA, the protein expression profile of the 7 experimental settings was analyzed by SDS-PAGE (Figure 4.10A). The results obtained showed that simultaneous expression of BioA/BioK with YgeX does not occur (that is, it is not possible to clearly distinguish two bands) suggesting there is an imbalance of

protein expression. Furthermore, the expected band corresponding to ScHem1 was also not visible (consistent with the absence of D-ALA production) thus pointing to a possible problem with the pColA induction system. It has been reported that in a pDuet system the expression from lower-copy number plasmids is lowered by co-expression of a high copy number plasmid [77], which can be attributable to the competition between T7/lac promoters. To remove the effects of co-expression of two plasmids and verify if ScHEM1 induction in the pColA plasmid can be similar to the one obtained in pET28 (previously used in [58]), the protein expression in BL21(DE3) cells carrying only the pET28-HEM1 or pColA-HEM1 was analyzed (Figure 4.10B). Indeed, a band corresponding to ScHem1 molecular weight is visible upon induction of cells carrying pET28a-ScHEM1 but not in cells with pColA-ScHEM1, indicating that the lack of protein expression is inherent to the pColA plasmid in the induction conditions used here (Figure 4.10B).

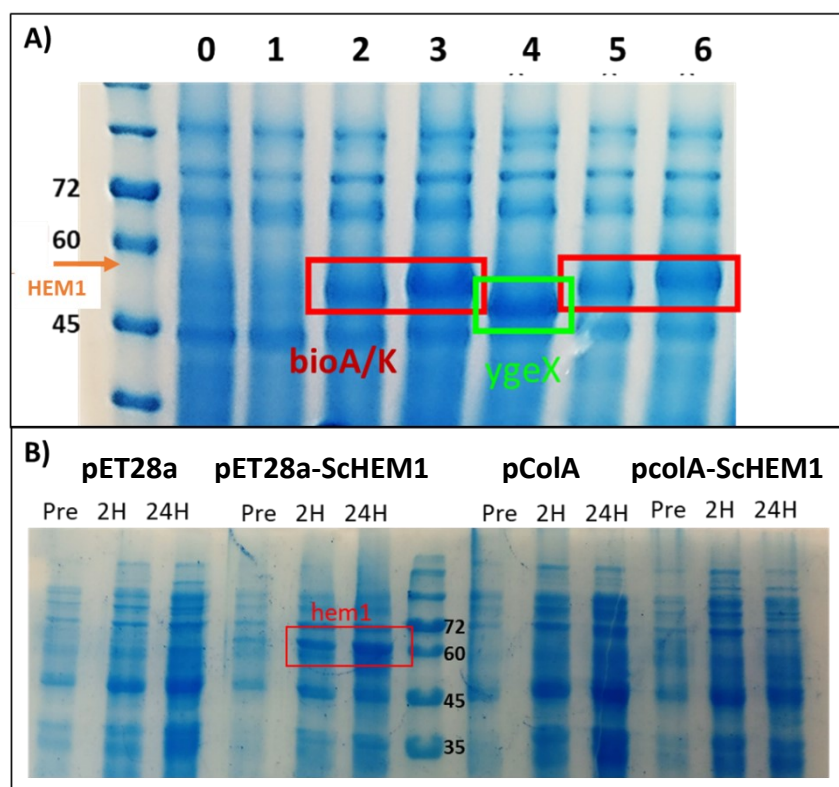


Figure 4.10. Protein expression analysis of LA pathway *in vivo* assembly. A) SDS-PAGE of cells harvested 24h after fermentation. B) SDS-PAGE of BL21(DE3) cells carrying the HEM1 expression plasmids. The induction was made by addition of 0.05 mM IPTG after 2 hours of culture and transfer to 30°C

To address the issue of pathway balancing an optimization of the conditions used to over-express the two sets of proteins was attempted. This was started with setting 1

(pColA-ScHEM1 & pD-ygeX) and 7 (pcolA-ScHEM1 & pD-bioK-ygeX). Three variable conditions were tested: different time of induction (2 or 5 hours after culture start), different concentration of IPTG (0.1 or 0.5 mM of IPTG) and different temperatures to cultivate the cells during the induction phase (25°C, 30°C or 37°C). The results (Annex 2 - Figure 8.2) indicate that the best conditions to enable co-expression of BioK and YgeX is by inducing after 5 hours of culture inoculation in fresh LB medium, with 0.1 mM IPTG and maintaining the culture during that time at 30°C. The conditions were further validated with the 7 strains harboring the combinations of BioA/BioK and YgeX co-expression (Annex 2 -Figure 8.3).

To study the second problem of enabling ScHem1 expression, 2 different expression strains with the B121(DE3) background were tested, to verify if ScHem1 expression could be improved in different genetic backgrounds. The *E. coli* strains tested included Rosetta, well suited for the expression of eukaryotic protein since it expresses tRNAs that correspond to codons rarely used in *E. coli*, even though the ScHEM1 has been codon optimized for *E. coli*; and pLysS, which expresses T7 lysozyme, preventing basal expression of the proteins to be induced. The results obtained indicated that it was not possible to express ScHem1 from the pColA plasmid (Annex 2 - Figure 8.4). Based on this, a closer look into the cloning strategy used to clone the ScHEM1 fragment in the pColA plasmid was taken, being detected that the engineered plasmid has the gene out of frame. Therefore, for a correct assembly of the entire an LA pathway in a single *E. coli* strain it will be required to re-clone the ScHEM1 gene in the pColA plasmid.

Given the detected problem with the pColA-ScHEM1 construct it was decided to attempt a two-step approach exploring the available pET28-HEM1 plasmid (which can't be used for the strategy described above because its pBBR1 origin is in the same compatibility group as the colE1 origin in the pET plasmids used to over-express BioA/BioK/YgeX and therefore it would not be possible to stably maintain strains harboring the two plasmids). In this approach we first used cells over-producing D-ALA from the over-expression of ScHEM1 and then the supernatant of this culture (enriched in D-ALA) was used to cultivate *E. coli* cells harboring the plasmids driving expression of BioA/BioK and YgeX and pColA empty plasmid to confer resistance to Kan (since the supernatant from the first stage contains Kan). Induction of the LA pathway enzymes was carried out as optimized in section 4.4.3.1: 0.1 mM IPTG after 5 hours of culture start and transfer to 30°C (Figure 4.11A). Supernatants samples were collected after 1 and 5 days and analyzed by HPLC, but under these conditions we could not detect the presence of LA in the culture broth (Figure 4.11B).

Chapter 4– Strategies for the successful implementation of an *in silico* predicted LA-forming pathway using D-ALA as a precursor

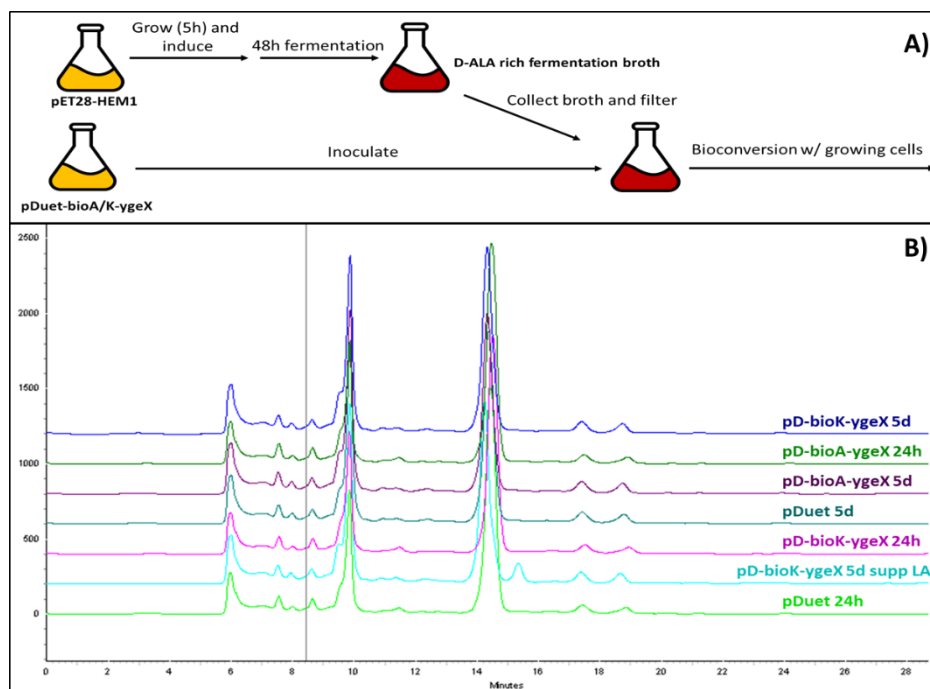


Figure 4.11. Two-step strategy for *in vivo* LA production based on bioconversion with growing cells. A) Experimental setup used. B) HPLC traces of supernatant samples taken at 24 hours and 5 days of bioconversion.

A second approach was attempted in which the cells carrying the LA pathway plasmid are separately allowed to grow for 5 hours, the expression of pathway enzymes is induced and after 5 hours of induction the cells are collected and inoculated with a high initial cellular density ($OD_{600}=2$) in the D-ALA enriched-broth (Figure 4.12A). Also in this case we could not detect a peak corresponding to LA (Figure 4.12C) and the extracellular D-ALA concentration after 24h and 48h was not significantly reduced indicating that there was no consumption of D-ALA by these cells (Figure 4.12C). On the overall, the data collected from the *in vivo* assembly of the hypothesized LA pathway thus indicates that the KANA transaminase and/or DAPA ammonia lyase were not able to catalyze the amination of D-ALA to DAVA and DAVA double deamination to LA, respectively, in a degree that would allow enough LA production to be detectable in the fermentation broth.

Chapter 4– Strategies for the successful implementation of an in silico predicted LA-forming pathway using D-ALA as a precursor

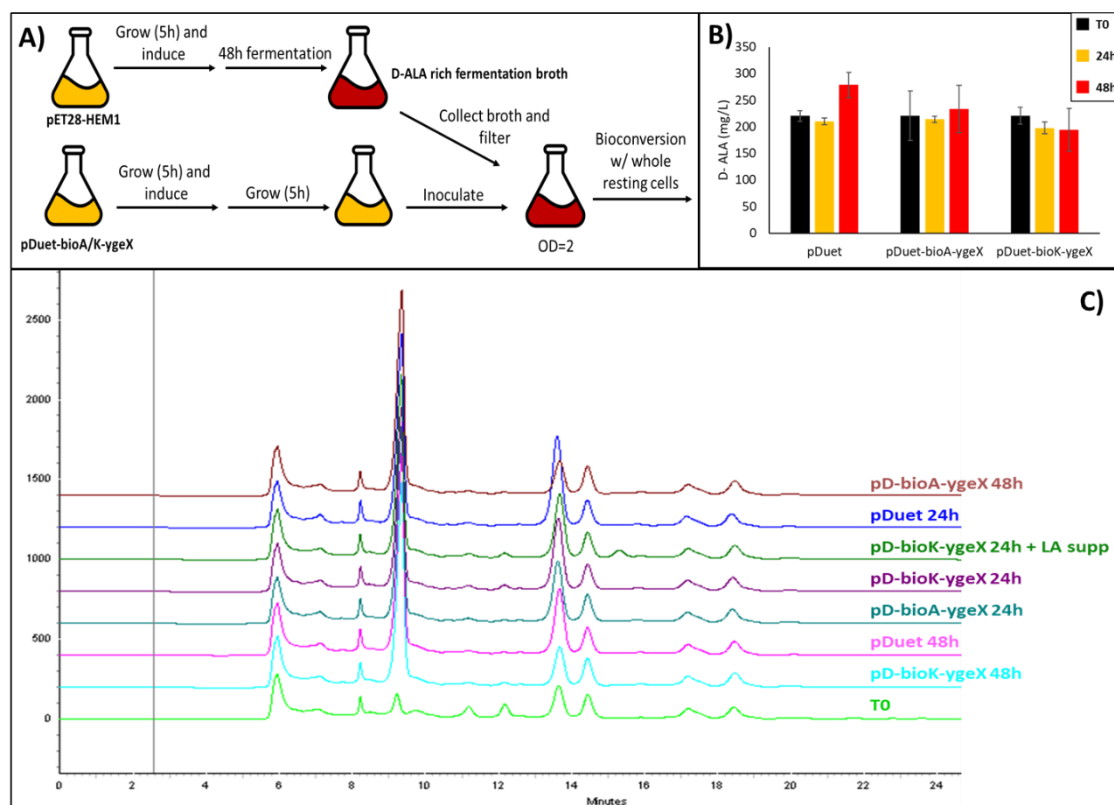


Figure 4.12. Two-step strategy for in vivo LA production based on bioconversion with whole resting cells. A) Experimental setup used. B) D-ALA extracellular concentration at 24h and 48h of bioconversion. C) HPLC traces of supernatant samples taken at 24 and 48 hours of bioconversion.

Overview

In this subsection the selected LA biosynthetic pathway was tentatively assembled in *E. coli*. Two main strategies were used: i) over-produce D-ALA (by over-expressing *ScHEM1*) and promote its conversion to LA through the over-expression of BioA/BioK and YgeX. Due to the problem identified in the plasmid that harbored the *ScHEM1* gene this was not completed in time but is considered a priority in the future, as it may enable the correct assembly of a functional LA pathway in *E. coli*. As an alternative a two-step strategy was devised, aiming to first over-produce D-ALA and then promote the bioconversion of the accumulated D-ALA in the broth to LA. Unfortunately, with the methodologies that were explored in this work we could not successfully detect LA in the fermentation broth, despite successful over-production of D-ALA has been confirmed to

occur. It is possible that this inability to produce LA results from difficulties in the import of D-ALA from the broth to the cytosol of *E. coli*. D-ALA has been reported to be imported into the cell through the dipeptide inner membrane transporter dppBCDF and the periplasmic binding proteins mppA and dppA, with a potential redundant contribution of other permeases [78]. It remains to be established whether or not these proteins are expressed under the conditions that we have used in this work and if they are if their activity promotes the entrance of a sufficient amount of D-ALA that results in a quantifiable amount of LA. Necessarily, the lack of LA may also result from the lack of activity of one of the enzymes involved, namely of YgeX which we could not confirm *in vitro* to be able to undertake the DAVA-LA conversion.

4.4.3.Improvement of D-ALA internal pool in *S. cerevisiae* cells

S. cerevisiae naturally produces D-ALA and this is advantageous for implementation of the selected pathway that was examined. However, the results from *in silico* simulation show that the internal availability of this metabolite can constrain LA production and therefore a strategy was devised in order to improve the internal pool of D-ALA inside *S. cerevisiae* cells. This approach is also interesting in the perspective of over-producing D-ALA since this metabolite is itself interesting as an add-value chemical.

As a first approach to increase D-ALA production, the overexpression of D-ALA synthase (encoded by the *HEM1* gene) was attempted. For this we cloned this gene under the control of the galactose-inducible promoter *GAL1* in the pGreg586-prGal1-Hem1 plasmid (from now pGREG-HEM1) [63]. Cultures undertaken with *S. cerevisiae* BY4741 cells transformed with the pGREG-HEM1 plasmid were carried out in MMB medium supplemented with 50 mM glycine and 50 mM succinic acid as these are precursors of D-ALA biosynthesis (Figure 4.2). Besides following the accumulation of D-ALA in the broth, it was also accompanied the rate of succinic acid consumption. The extracellular concentration of D-ALA obtained in the fermentation broth undertaken with *S. cerevisiae* cells transformed with the cloning vector pGreg586 was constant and close to 1.74 mg/L (Figure 4.13A). This concentration is within the background signal that was obtained with the modified Ehrlich method that was used to quantify D-ALA (observed at the initial time point). On the other hand, the over-expression of *HEM1* resulted in a continuous increase in the concentration of D-ALA accumulated in the fermentation broth and achieved a final titer of 11.6 mg/L after 72h (Figure 4.13A). No significant reduction in the concentration of succinic acid was observed, not even in the cells harboring the

pGREG-HEM1 plasmid (Figure 4.13B), which suggests that the cells were producing D-ALA at the expense of internally available succinic acid. This apparent inability to direct the provided succinic acid into D-ALA may result from yeast cells not having a cytosolic enzyme that can catalyze the formation of the thioester CoA bond. Either way, it was clear that the over-expression of HEM1 resulted in accumulation of D-ALA in the broth.

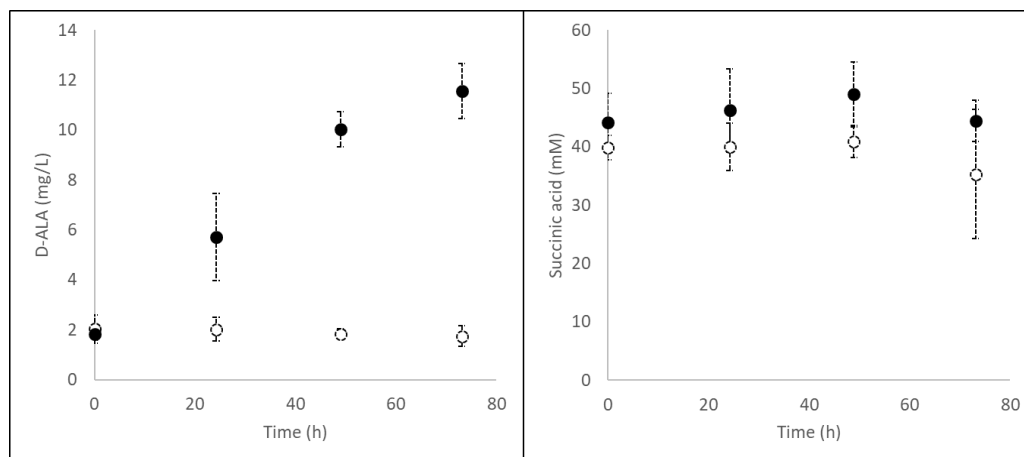


Figure 4.13. D-ALA cultivations with BY4741 transformed with pGreg586 (empty circles) or pGREG-HEM1 (black circles). The extracellular concentrations of D-ALA (A) and succinic acid (B) are represented.

Considering that D-ALA utilization is likely to contribute to decrease titers, we then decided to delete the Hem2 D-ALA dehydratase, the enzyme that uses D-ALA as a substrate in the tetrapyrrole biosynthetic pathway (Figure 4.2). However, previous studies resulting from the genome-wide functional analysis in *S. cerevisiae* revealed that the deletion of *HEM2* resulted in non-viable cells in the BY4741 background [79]. Strikingly, in the W303 background a $\Delta hem2$ mutant was described, however this mutant displayed an auxotrophy for cysteine or heme, due to the requirement of siroheme to synthesize this aminoacid [80]. Heme deficiency in $\Delta Hem2$ mutants has also been linked to auxotrophy to ergosterol/tween and methionine in D587-4B derived strains [81] and in the W303 background [80].

Since the use of an auxotrophic strain would imply additional costs by adding the required compounds to the culture media, we envisaged the possibility of downregulating HEM2 instead of deleting it. For that, we made use of a Tet-OFF system, a well-established genetic control tool in yeast that makes use of tetracycline-controlled transactivator (tTA) protein to control the activity of promoters via tetO operator sites [82]. The system devised here is based on a previous work in which tTa was used to downregulate the expression of *HXK1* as a mean to improve gluconate production [64]. A cassette having the tTA element under the control of a CYC1 TATA 7xtetO promoter was inserted in the *HIS3* locus of the W303 strain, and cassette with the CYC1 TATA 7xtetO promoter was inserted in the native HEM2 promoter, resulting in the H2 valve strain. Upon the genetic modifications performed the *HEM2* expression is expected to be under the control of tTA and therefore upon the use of doxycycline (Dox) the activity of tTA will be inhibited and, consequently, HEM2 repressed (Figure 4.14). Due to polymerase slippage occurring in repetitive regions during the PCR amplification steps of the cassettes to be integrated in the genome, the *HEM2* promoter was replaced by 6xtetO-TATA and the tTA expression was controlled by a 2xtetO-TATA in the H2 strain, instead of the 7xtetO elements that were initially predicted. Either way, previous studies have shown that a sole tetO operator sequence is already strong enough to produce a visible effect [83], for which we have decided to move forward with the engineered H2 strain and perform fermentations to produce D-ALA using conditions similar to those used above for the W303 strain. With this in mind, the obtained H2 strain was used for the following fermentations and it was confirmed that the integrations were stable after various passages in YPD (non selective) media.

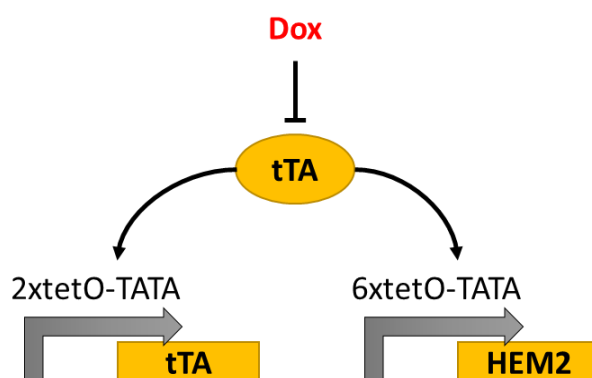


Figure 4.14. Engineering of the H2 valve strain, where both tTA and HEM2 expression is controlled by tetO elements in the promoters. Dox addition prevents tTA recruitment to the tetO operator sites, repressing expression of tTa and HEM2. Design and figure based on [22].

An initial attempt to test to the capacity of the H2 valve strain ability to produce D-ALA was devised providing a pulse of Dox when the cells reached the exponential phase. This approach allowed an initial stage of normal *HEM2* expression, after which the expression of this gene is repressed and, hopefully, D-ALA is produced. Different Dox concentrations were tested to assess the level of repression, based on the concentrations used previously with this system [64]. The results obtained show that the production titers of D-ALA produced by the W303 wild-type strain and by the engineered H2 strain were almost negligible (~2 mg/L, very close to the noise level obtained with the quantification methods used) (Figure 4.13). This indicates that accumulation of D-ALA in the fermentation broth requires overexpression of *HEM1*.

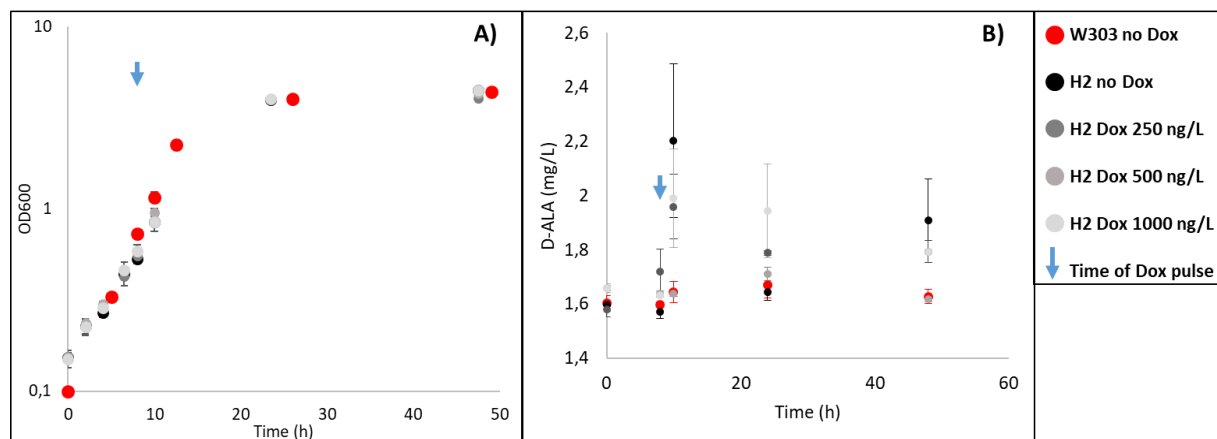


Figure 4.15. D-ALA cultivations with H2 strain and W303 strain (control) in MMB glucose. A pulse of Dox was added at the exponential phase (blue arrow, 8h). The OD_{600nm} (A) and extracellular D-ALA concentrations (B) are represented.

As such, the next step was to transform the H2 strain with the pGREG-*HEM1* plasmid to over-express *HEM1* while also downregulating the expression of *HEM2*. To promote the downregulation of *HEM2* dox was added to the galactose-induced culture at the exponential phase. Over-expression of *HEM1* resulted in a clear production of D-ALA by W303 cells, reaching 4.5 mg/L of D-ALA after 72h. This titer was lower than the previously obtained for the By4741 strain (11.6 mg/L, Figure 4.13), suggesting that the By4741 background could be more advantageous for D-ALA overproduction. Unexpectedly, the levels of D-ALA accumulated in the broth of H2 cells were lower than those obtained for the wild-type counterpart cells (W303) and this was observed for all the concentrations of Dox added. However, the H2 culture reached a significantly lower biomass than W303 cells, suggesting a potential growth defect of this strain when growing in galactose (no significant differences between growth rate of W303 and H2

were observed when the cells were growing on glucose, Figure 4.15). Consequently, the D-ALA titers normalized for cell biomass present in the fermentation shows a higher production capacity of H2 cells (Figure 4.16C), suggesting that if higher growth rates and/or shorter lag periods can be enabled in this strain this can result in improved production of D-ALA.

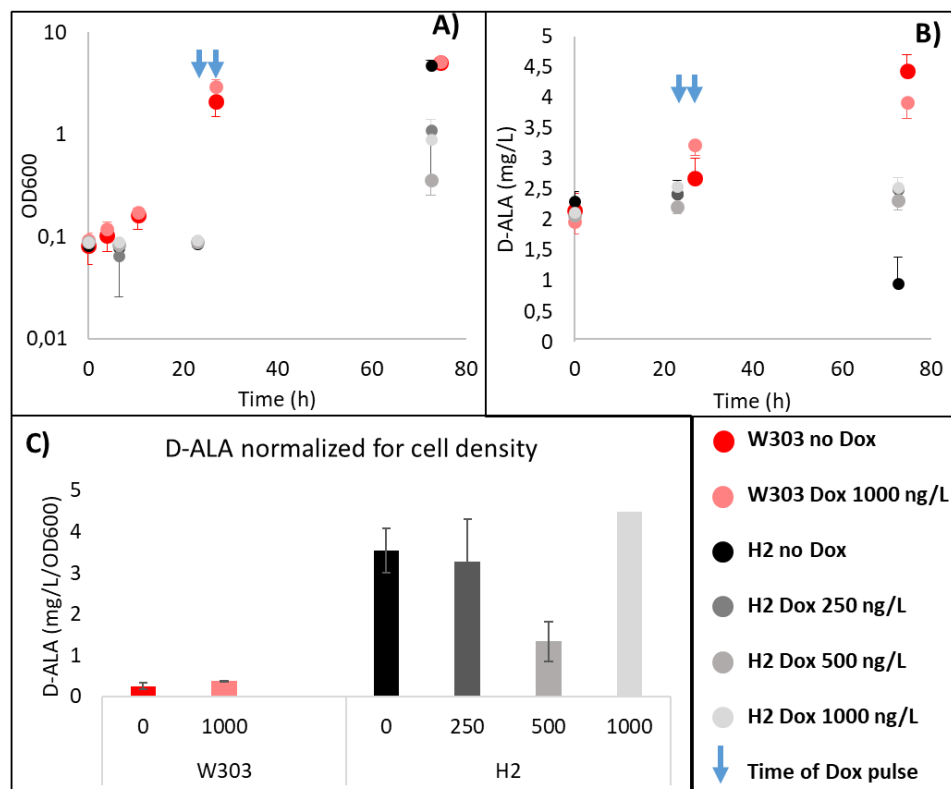


Figure 4.16. D-ALA cultivations with H2 strain and W303 strain (control) carrying the pGREG-HEM1 plasmid in MMB galactose. A pulse of Dox was added at the exponential phase (23.5h for W303 and 27h for H2, blue arrows). The OD_{600nm} (A) and extracellular D-ALA concentrations (B) are represented. In C) it is represented the D-ALA concentrations normalized to OD_{600nm} of the culture (C).

To verify if the repression conditions with Dox can affect both the growth of H2 and its D-ALA productivity, three different timings of Dox addition to the culture were tested: at the initial time of fermentation, at the exponential phase (the strategy that was used until this point) and a pre-repression of the cultures in the pre-inoculum stage that is sustained during the fermentation stage by also adding a second pulse at the initial time of fermentation. The fermentations were carried out in microplates to increase throughput, considering the number of conditions to be tested. The same trend that was previously

observed in shake-flask fermentations occurred in microplate fermentations, where the H2 strain produced lower titers of D-ALA when compared to the wild-type W303 and also reached lower OD_{600nm} after 48h in a manner independent of the Dox concentration and the timing of Dox pulse (Figure 4.17A and B). Again, when production titers are normalized for cellular density, it is observed that H2 is a better producer (Figure 4.17C). This data is again indicative of a growth defect of the H2 strain in galactose.

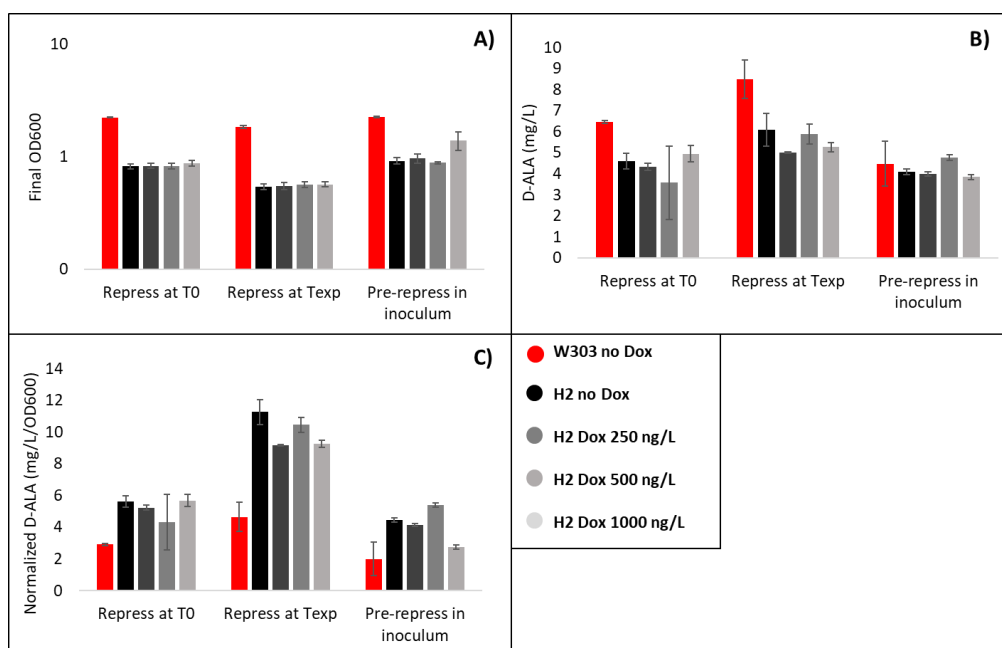


Figure 4.17. Microplate D-ALA cultivations with H2 strain and W303 strain (control) carrying the pGREG-HEM1 plasmid in MMB galactose. A pulse of Dox was added at initial time of cultivation, at the exponential phase or from the inoculum. The OD_{600nm} (A) and extracellular D-ALA concentrations (B) are represented. In C) it is represented the D-ALA concentrations normalized to OD_{600nm} of the culture (C).

The growth defect in galactose of the H2 strain was not expected since this was not described previously to occur with strains impaired for heme biosynthesis. Additionally, the D-ALA production data indicates no effect of varying the Dox concentration added to the media in growth or D-ALA production, suggesting that *HEM2* expression is not in fact controlled by Dox. To confirm this the effect of Dox addition on *HEM2* expression was quantified by Real Time Quantitative PCR (RT-qPCR). Since the H2 strain presents a growth defect in galactose, it was not possible to carry out the assays in the same conditions as the fermentations, which take place in galactose. For the Dox repression assays the cells were grown to exponential phase in MMB glucose, a Dox pulse was added

and samples for RT-qPCR were taken after 2h and 24h of exposure. A first analysis comparing only the HEM2 expression between the W303 and H2 strain when no Dox pulse is added reveals that there is no difference between these two strain's expressions, as expected (Figure 4.18A). On a second analysis focused only on the H2 cultures and comparing the expression between the cultures where a Dox pulse was added at the exponential phase and the cultures where no Dox pulse was added, reveals that there is no significant difference in expression either at the exponential phase (2h after the pulse) or stationary phase (24h after the pulse) (Figure 4.18B). Thus, the H2 strain does not appear to have a Dox-repressible HEM2 expression system as desired. It is possible that only 2xtetO sites controlling tTA expression were not enough in this case to obtain a strong enough tTA expression to drive Dox-dependency on HEM2 expression, although previous data have shown activity with 2xtetO (and even with only 1xtetO) [83]. It is also clear that the H2 strain displays a clear growth defect on galactose, with a lag phase of at least 48h and that isn't abrogated by supplementation of the medium with hemin or methionine (data not shown). The possibility of the tTA system creating a burden on the cell that decreases growth has also been previously ruled out [64], thus creating the question of what is the source of this phenotype in the H2 strain and if it can explain the lack of Dox-dependency on HEM2 expression.

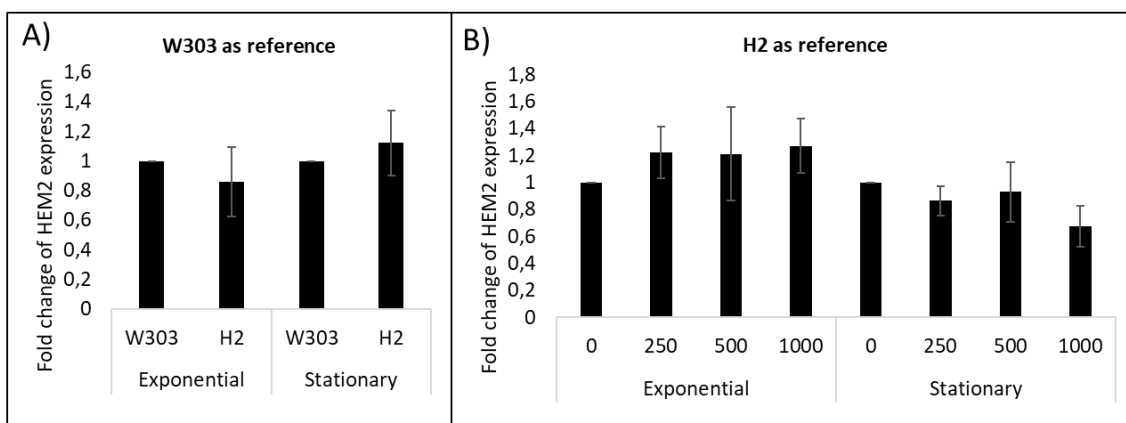


Figure 4.18. Analysis of HEM2 expression of W303 and H2 cells cultivated in MMB glucose during exponential and stationary phase. A) Cultures where no Dox pulse was added; the reference is the W303 cultures. B) H2 cultures where a Dox pulse was added at the exponential phase (250, 500 or 1000 ng/mL); the H2 culture with no Dox pulse is the reference.

Although the RT-qPCR results in this chapter demonstrate that the H2 strain does not display the desired Dox-repressible HEM2 expression, the fermentation data indicates that H2 has a higher specific D-ALA production that could not be explored due to the

long lag phase in galactose. For this reason, to explore the H2 strain capabilities to produce D-ALA it would be required to perform fermentations in a media other than MMB galactose. To this end, a new experimental setup was devised using MMB raffinose+galactose, where the trisaccharide is the carbon source and the galactose present in the media induces HEM1 expression. The results obtained showed that the H2 strain reached a lower OD_{600nm} after 48h and lower D-ALA production, indicating that the phenotype does not occur exclusively due to galactose being the carbon source (Figure 4.19A and B). However, unlike the fermentations in MMB galactose, when raffinose was added as the carbon source the H2 strain has a similar specific production as the wild-type counterpart, with a small decrease in specific production with higher Dox concentrations (Figure 4.19C). This data is suggestive of the galactose growth defect observed in the H2 strain being related to higher specific D-ALA productions, although the mechanism could not be understood and is not dependent on Dox.

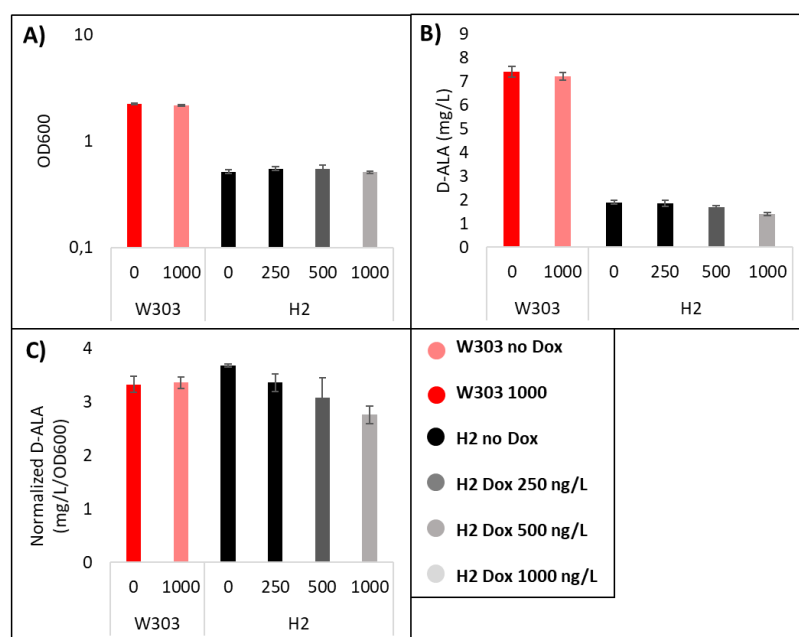


Figure 4.19. Microplate fermentations with W303 pHem1 and H2 pHem1 in MMB raffinose + galactose, with the Dox pulse added after 24. OD₆₀₀ after 48h (A), extracellular D-ALA concentrations after 48h (B) and D-ALA concentrations normalized to OD₆₀₀ of the culture (C).

Overview

In this subchapter the ability of *S. cerevisiae* to produce D-ALA, the chosen precursor for the LA forming pathway in thesis, was studied. Two main approaches to increase natural production were sought, the first being overexpression of HEM1. With this

approach it was possible to obtain titers of 11.6 mg/L in BY4741 and 4.5 mg/L in the W303 background. A recent study reported titers of 0.25 mg/L in a YPH499 strain overexpressing HEM1 under the control of the prGAL1 promoter [84]. The different background may account for the observed titers. However, it is noteworthy that the D-ALA quantification methods used here differ from the mentioned study, where GC-MS was used. In this work the modified Ehrlich method was used [66], a method that has been routinely used in fermentative D-ALA quantifications. Furthermore, in this work, as with previous ones in *E. coli* and *C. glutamicum* [54], [59], the media was supplemented with 50 mM glycine, thus increasing the precursor pool. The second strategy used in this subchapter was to downregulate HEM2 to decrease D-ALA consumption. Indeed, decreasing the rate of the D-ALA dehydratase steps has been pinpointed as a major focus in D-ALA biotechnological production, not only to decrease D-ALA consumption but also to prevent accumulation of downstream toxic compounds [43]. In the past, this has been attempted in *E. coli* by direct engineering of the hemB open reading frame [56], by artificially overexpressed the small noncoding RNA RyhB involved in iron metabolism [57] and by overexpression of downstream steps to enable the native negative feedback loop where the intermediate protoporphyrinogen IX inhibits D-ALA dehydratase [55]. In *S. cerevisiae*'s exploration as a D-ALA cell factory, the rate of D-ALA dehydratase was decrease by enzymatic inhibition of D-ALA dehydratase with the addition of LA to the culture media [44]. Additionally, *S. cerevisiae* strains with HEM2 alleles that have decrease function have been reported to have increased D-ALA accumulation and slower growth rates [85]. The results with the H2 valve strain do not demonstrate the desirable Dox-dependent repression of HEM2, but only that the engineered strain displays higher specific D-ALA productivity when galactose is the carbon source. While the correct integrations have been confirmed, no hypothesis could be established for the source of the galactose phenotype and the lack of HEM2 Dox-repression in the H2 strain. Effects of the genome integrations on nearby genes are a possibility, being of note that the integration of the 7xtetO-CYC1-tTA cassette was made via replacement of the native non-functional HIS3 locus with the YIP plasmid pRS303-7xtetO-CYC1-tTA and that ~250 bps upstream of the HIS3 locus is located the MRM1 locus in the opposing strand, which encodes a mitochondrial site-specific ribose methyltransferase. The pleiotropic effects of the HIS3 selectable marker have been studied in the past, with background strains that have a the his3-200 allele, with a disruption of the MRM1 bidirectional promoter having an increase in mtDNA point mutations and lower respiratory competency [86]. Considering the link between respiratory competency and galactose consumption in *S. cerevisiae* [87] and that respiratory mutants have been reported to be

unable to grow on galactose [88], a disruption of MRM1 transcription is a possible explanation for the observed phenotype in the H2 valve strain.

The results obtained from the efforts to improve D-ALA production in *S. cerevisiae* point to importance of overexpressing HEM1. However, the use of a promoter that can be induced regardless of the carbon source would be desirable, considering the interaction between galactose consumption, respiration and heme availability, which is the end product of D-ALA utilization. A potential alternative to be pursued is to use the DDI2 promoter (cyanamide hydratase), which can be induced by cyanamide or methyl methanesulfonate in sub-inhibitory concentrations and in a dose-dependent manner [89]. The establishment of a repressible system for HEM2 expression is also pinpointed here as a priority for improved D-ALA production, which in the case of this work may be achieved by integration of the 7xtetO-CYC1-tTA in a locus other than HIS3. Alternatively, copper may be used in *S. cerevisiae* to simultaneously induce and repress transcription by use of the CUP1 and CTR2/3 promoters, respectively, presenting an alternative to the tet-based systems for the control of HEM1 and HEM2 expression and decreasing the number of genome integration to avoid potential unpredicted effects [90]. The strategies presented in this work were focused on regulating the expression of D-ALA production and consumption; however, improving the precursor pool is also a relevant aspect of increased D-ALA biosynthesis. While it has been reported that aconitase is the limiting step in D-ALA production in *S. cerevisiae* [44], improving glycine and succinyl-coA has not been explored yet in the context of yeast-based D-ALA production. It has, however, been shown that the expression of a bacterial α -ketoglutarate coupled with the expression of *R. spheroids hemA* in the cytosol of *S. cerevisiae*, enables a functional cytosolic D-ALA production route [91], which would likely be beneficial in a scenario where the LA biosynthetic pathway explored in this chapter is established in *S. cerevisiae*, in opposition to the native D-ALA production route in the mitochondria. Finally, it is also relevant to note that LA is a catalytic inhibitor ALA dehydratase and its presence in the culture medium has been reported to be an effective strategy to increase D-ALA production in *S. cerevisiae* [44], [92]. In a future scenario where a functional LA biosynthetic is established in *S. cerevisiae* it is expected that the feedback mechanism will become advantageous for LA production.

4.5. Conclusion and Future Work

In this chapter the potential introduction of a LA biosynthetic pathway in *S. cerevisiae* based on the natural heme precursor D-ALA is explored. A first attempt at improving the yeast natural D-ALA production demonstrated that the overexpression of HEM1 is a successful strategy, although the effects of HEM2 downregulation could not be fully explored, which is pinpointed here as main limitation to improve D-ALA production in *S. cerevisiae*. Future work in this area may include the downregulation of HEM2 through an alternative system, such as the copper-repressible one and exploration of strategies to increase succinyl-coA and glycine availability. The alternative LA production pathway was explored through the ability of the natural KANA transaminase from *E. coli* and *B. subtilis* and DAPA ammonia lyase from *E. coli* to catalyze the synthetic amination of D-ALA and deamination of DAVA to LA, respectively. While the pathway could not be successfully assembled *in vivo* in the easy-to-use *E. coli* model, the first enzymatic synthetic step of the pathway was validated *in vitro*. The purification of bioA, bioK and ygeX to enable quantitative measurements of the reaction rates for the synthetic substrates and, in particular, the calculation of the K_m will be of use to understand the affinity difference between the synthetic and native substrates. It is likely that enzyme evolution may be required to enable a functional LA producing pathway, which here is considered to be a possibility to put in to practice by performing whole pathway directed evolution in an *in vivo* setting. However, a main step to be taken before pathway evolution is to correctly express ScHEM1 in *E. coli*, enabling the simultaneous D-ALA overproduction and conversion to LA, the lack of which is identified here as a limitation of the work.

4.6. References

- [1] M. Bertoldi *et al.*, “Ornithine and glutamate decarboxylases catalyse and oxidative deamination of their alpha-methyl substrates,” *Biochem. J.*, vol. 512, pp. 509–512, 1999, doi: 10.1042/0264-6021:3420509.
- [2] J. G. Jeffryes *et al.*, “MINEs: open access databases of computationally predicted enzyme promiscuity products for untargeted metabolomics,” *J. Cheminform.*, vol. 7, no. 1, p. 44, Dec. 2015, doi: 10.1186/s13321-015-0087-1.
- [3] T. V. Sivakumar, V. Giri, J. H. Park, T. Y. Kim, and A. Bhaduri, “ReactPRED: a tool to predict and analyze biochemical reactions,” *Bioinformatics*, vol. 32, no. 22, pp. 3522–3524, 2016, doi: 10.1093/bioinformatics/btw491.
- [4] L. Heirendt *et al.*, “Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0,” *Nat. Protoc.*, vol. 14, no. 3, pp. 639–702, Mar. 2019, doi: 10.1038/s41596-018-0098-2.
- [5] L. Gurobi Optimization, “Gurobi Reference Manual.” 2020, [Online]. Available: <http://www.gurobi.com>.
- [6] B. D. Heavner, K. Smallbone, B. Barker, P. Mendes, and L. P. Walker, “Yeast 5 - an expanded reconstruction of the *Saccharomyces cerevisiae* metabolic network,” *BMC Syst. Biol.*, vol. 6, no. 55, Jan. 2012, doi: 10.1186/1752-0509-6-55.
- [7] J. D. Orth *et al.*, “A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism,” *Mol. Syst. Biol.*, vol. 7, no. 535, Jan. 2011, doi: 10.1038/msb.2011.65.
- [8] D. S. Weaver, I. M. Keseler, A. Mackie, I. T. Paulsen, and P. D. Karp, “A genome-scale metabolic flux model of *Escherichia coli* K – 12 derived from the EcoCyc database,” pp. 1–24, 2014.
- [9] R. Andrade, M. Doostmohammadi, J. L. Santos, M.-F. Sagot, N. P. Mira, and S. Vinga, “MOMO - multi-objective metabolic mixed integer optimization: application to yeast strain engineering,” *BMC Bioinformatics*, vol. 21, no. 1, p. 69, Dec. 2020, doi: 10.1186/s12859-020-3377-1.
- [10] M. D. Paxhia and D. M. Downs, “SNZ3 encodes a PLP synthase involved in thiamine synthesis in *saccharomyces cerevisiae*,” *G3 Genes, Genomes, Genet.*, vol. 9, no. 2, pp. 335–344, Feb. 2019, doi: 10.1534/g3.118.200831.
- [11] N. Baldi *et al.*, “Functional expression of a bacterial α -ketoglutarate dehydrogenase in the cytosol of *Saccharomyces cerevisiae*,” *Metab. Eng.*, vol. 56, no. August, pp. 190–197, 2019, doi: 10.1016/j.ymben.2019.10.001.
- [12] A. Larhlimi, L. David, J. Selbig, and A. Bockmayr, “F2C2: a fast tool for the computation of flux coupling in genome-scale metabolic networks,” *BMC Bioinformatics*, vol. 13, no. 1, 2012, doi: 10.1186/1471-2105-13-57.
- [13] M. Kanehisa, “KEGG: Kyoto Encyclopedia of Genes and Genomes,” *Nucleic Acids Res.*, vol. 28, no. 1, pp. 27–30, Jan. 2000, doi: 10.1093/nar/28.1.27.

- [14] D. Bajusz, A. Rácz, and K. Héberger, “Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations?,” *J. Cheminform.*, vol. 7, no. 1, p. 20, Dec. 2015, doi: 10.1186/s13321-015-0069-3.
- [15] M. Brückmann, R. Blasco, K. N. Timmis, and D. H. Pieper, “Detoxification of protoanemonin by diene lactone hydrolase,” *J. Bacteriol.*, vol. 180, no. 2, pp. 400–2, Jan. 1998, Accessed: Mar. 03, 2020. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/9440530>.
- [16] M. Maestrci, K. Bui, A. Arnaud, P. Galzy, and P. Viala, “Activity and regulation of an amidase (acylamide amidohydrolase, EC 3.5.1.4) with a wide substrate spectrum from a *Brevibacterium* sp.,” pp. 315–320, 1984.
- [17] K. Sasaki, M. Watanabe, and T. Tanaka, “Biosynthesis, biotechnological production and applications of 5-aminolevulinic acid,” *Appl. Microbiol. Biotechnol.*, vol. 58, no. 1, pp. 23–29, Jan. 2002, doi: 10.1007/s00253-001-0858-7.
- [18] Gerhard Gottschalk and Louis-Moreau, *Bacterial Metabolism*. Springer Science & Business Media, 1986.
- [19] K. Hoare and K. Datta, “Characteristics of l-alanine:4,5-dioxovaleric acid transaminase: An alternate pathway of heme biosynthesis in yeast,” *Arch. Biochem. Biophys.*, vol. 277, no. 1, pp. 122–129, 1990, doi: 10.1016/0003-9861(90)90559-H.
- [20] P. J. Chapman and D. W. Ribbons, “Metabolism of resorcinolic compounds by bacteria: orcinol pathway in *Pseudomonas putida*,” *J. Bacteriol.*, vol. 125, no. 3, pp. 975–84, Mar. 1976, Accessed: Aug. 12, 2019. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/1254564>.
- [21] E. H. Jang, S. A. Park, Y. M. Chi, and K. S. Lee, “Kinetic and structural characterization for cofactor preference of succinic semialdehyde dehydrogenase from *Streptococcus pyogenes*,” *Mol. Cells*, vol. 37, no. 10, pp. 719–726, 2014, doi: 10.14348/molcells.2014.0162.
- [22] S. Cheong, J. M. Clomburg, and R. Gonzalez, “Energy- and carbon-efficient synthesis of functionalized small molecules in bacteria using non-decarboxylative Claisen condensation reactions,” *Nat. Biotechnol.*, vol. 34, no. 5, pp. 556–561, 2016, doi: 10.1038/nbt.3505.
- [23] G. M. Lin, R. Warden-Rothman, and C. A. Voigt, “Retrosynthetic design of metabolic pathways to chemicals not found in nature,” *Curr. Opin. Syst. Biol.*, vol. 14, pp. 82–107, 2019, doi: 10.1016/j.coisb.2019.04.004.
- [24] M. Ramirez-Gaona *et al.*, “YMDB 2.0: A significantly expanded version of the yeast metabolome database,” *Nucleic Acids Res.*, vol. 45, no. D1, pp. D440–D445, Jan. 2017, doi: 10.1093/nar/gkw1058.
- [25] F. P. Guengerich and H. P. Broquist, “Lysine catabolism in *Rhizoctonia leguminicola* and related fungi,” *J. Bacteriol.*, vol. 126, no. 1, pp. 338–47, Apr. 1976, Accessed: Aug. 29, 2019. [Online]. Available:

<http://www.ncbi.nlm.nih.gov/pubmed/131119>.

- [26] M. Sorokina, M. Stam, C. Médigue, O. Lespinet, and D. Vallenet, “Profiling the orphan enzymes,” *Biol. Direct*, vol. 9, no. 1, pp. 1–16, 2014, doi: 10.1186/1745-6150-9-10.
- [27] A. W. Struck, M. L. Thompson, L. S. Wong, and J. Micklefield, “S-Adenosyl-Methionine-Dependent Methyltransferases: Highly Versatile Enzymes in Biocatalysis, Biosynthesis and Other Biotechnological Applications,” *ChemBioChem*, vol. 13, no. 18, pp. 2642–2655, 2012, doi: 10.1002/cbic.201200556.
- [28] F. Khan, V. R. Jala, N. A. Rao, and H. . Savithri, “Characterization of recombinant diaminopropionate ammonia-lyase from *Escherichia coli* and *Salmonella typhimurium*,” *Biochem. Biophys. Res. Commun.*, vol. 306, no. 4, pp. 1083–1088, Jul. 2003, doi: 10.1016/S0006-291X(03)01100-8.
- [29] R. Contestabile, S. Angelaccio, R. Maytum, F. Bossa, and R. A. John, “The contribution of a conformationally mobile, active site loop to the reaction catalyzed by glutamate semialdehyde aminomutase,” *J. Biol. Chem.*, vol. 275, no. 6, pp. 3879–86, Feb. 2000, doi: 10.1074/JBC.275.6.3879.
- [30] J. Stetefeld, M. Jenny, and P. Burkhard, “Intersubunit signaling in glutamate-1-semialdehyde-aminomutase,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 37, pp. 13688–13693, Sep. 2006, doi: 10.1073/pnas.0600306103.
- [31] S. W. Van Arsdell *et al.*, “Removing a bottleneck in the *Bacillus subtilis* biotin pathway: BioA utilizes lysine rather than S-adenosylmethionine as the amino donor in the KAPA-to-DAPA reaction,” *Biotechnol. Bioeng.*, vol. 91, no. 1, pp. 75–83, 2005, doi: 10.1002/bit.20488.
- [32] R. S. Breen, D. J. Campopiano, S. Webster, M. Brunton, R. Watt, and R. L. Baxter, “The mechanism of 7,8-diaminopelargonate synthase; the role of S-adenosylmethionine as the amino donor,” *Org. Biomol. Chem.*, vol. 1, no. 20, pp. 3498–9, 2003, doi: 10.1039/b310443p.
- [33] C. H. Tseng, C. H. Yang, H. J. Lin, C. Wu, and H. P. Chen, “The S subunit of D-ornithine aminomutase from *Clostridium sticklandii* is responsible for the allosteric regulation in D- α -lysine aminomutase,” *FEMS Microbiol. Lett.*, vol. 274, no. 1, pp. 148–153, Sep. 2007, doi: 10.1111/j.1574-6968.2007.00820.x.
- [34] C. Weber, C. Brückner, S. Weinreb, C. Lehr, C. Essl, and E. Boles, “Biosynthesis of cis,cis-muconic acid and its aromatic precursors, catechol and protocatechuic acid, from renewable feedstocks by *Saccharomyces cerevisiae*,” *Appl. Environ. Microbiol.*, vol. 78, no. 23, pp. 8421–30, Dec. 2012, doi: 10.1128/AEM.01983-12.
- [35] W. Niu, K. M. Draths, and J. W. Frost, “Benzene-Free Synthesis of Adipic Acid,” *Biotechnol. Prog.*, vol. 18, no. 2, pp. 201–211, Apr. 2002, doi: 10.1021/bp010179x.
- [36] H. Yim *et al.*, “Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol,” *Nat. Chem. Biol.*, vol. 7, no. 7, pp. 445–452, 2011, doi:

10.1038/nchembio.580.

- [37] M. A. Islam, N. Hadadi, M. Ataman, V. Hatzimanikatis, and G. Stephanopoulos, “Exploring biochemical pathways for mono-ethylene glycol (MEG) synthesis from synthesis gas,” *Metab. Eng.*, vol. 41, pp. 173–181, 2017, doi: 10.1016/j.ymben.2017.04.005.
- [38] M. Tokic *et al.*, “Discovery and Evaluation of Biosynthetic Pathways for the Production of Five Methyl Ethyl Ketone Precursors,” *ACS Synth. Biol.*, vol. 7, no. 8, pp. 1858–1873, 2018, doi: 10.1021/acssynbio.8b00049.
- [39] D. Dugar and G. Stephanopoulos, “Relative potential of biosynthetic pathways for biofuels and bio-based products,” *Nat. Biotechnol.*, vol. 29, no. 12, pp. 1074–1078, 2011, doi: 10.1038/nbt.2055.
- [40] J. D. Orth, I. Thiele, and B. Ø. Palsson, “What is flux balance analysis?,” *Nat. Biotechnol.*, vol. 28, no. 3, pp. 245–8, Mar. 2010, doi: 10.1038/nbt.1614.
- [41] T. B. Fitzpatrick, N. Amrhein, B. Kappes, P. Macheroux, I. Tews, and T. Raschle, “Two independent routes of de novo vitamin B6 biosynthesis: Not that different after all,” *Biochem. J.*, vol. 407, no. 1, pp. 1–13, 2007, doi: 10.1042/BJ20070765.
- [42] U. Schell, R. Wohlgemuth, and J. M. Ward, “Synthesis of pyridoxamine 5'-phosphate using an MBA:pyruvate transaminase as biocatalyst,” *J. Mol. Catal. B Enzym.*, vol. 59, no. 4, pp. 279–285, Aug. 2009, doi: 10.1016/j.molcatb.2008.10.005.
- [43] Z. Kang, W. Ding, X. Gong, Q. Liu, G. Du, and J. Chen, “Recent advances in production of 5-aminolevulinic acid using biological strategies,” *World J. Microbiol. Biotechnol.*, vol. 33, no. 11, p. 200, Nov. 2017, doi: 10.1007/s11274-017-2366-7.
- [44] K. Y. Hara *et al.*, “5-Aminolevulinic acid fermentation using engineered *Saccharomyces cerevisiae*,” *Microb. Cell Fact.*, vol. 18, no. 1, pp. 1–8, 2019, doi: 10.1186/s12934-019-1242-6.
- [45] K. Jensen, V. Broeken, A. S. L. Hansen, N. Sonnenschein, and M. J. Herrgård, “OptCouple: Joint simulation of gene knockouts, insertions and medium modifications for prediction of growth-coupled strain designs,” *Metab. Eng. Commun.*, vol. 8, p. e00087, Jun. 2019, doi: 10.1016/j.mec.2019.e00087.
- [46] A. M. Feist, D. C. Zielinski, J. D. Orth, J. Schellenberger, M. J. Herrgård, and B. O. Palsson, “Model-driven evaluation of the production potential for growth-coupled products of *Escherichia coli*,” *Metab. Eng.*, vol. 12, no. 3, pp. 173–186, May 2010, doi: 10.1016/j.ymben.2009.10.003.
- [47] T. Werpy and G. Petersen, “Top Value Added Chemicals from Biomass,” 2004. doi: 10.2172/926125.
- [48] D. J. Hayes, J. Ross, M. H. B. Hayes, and S. Fitzpatrick, “The Biofine process: production of levulinic acid, furfural and formic acid from lignocellulosic feedstocks,” *Biorefineries-Industrial Process. Prod.*, vol. 1, 2005, doi: 10.1002/9783527619849.

- [49] J. Sandmark, A. C. Eliot, K. Famm, G. Schneider, and J. F. Kirsch, “Conserved and Nonconserved Residues in the Substrate Binding Site of 7,8-Diaminopelargonic Acid Synthase from *Escherichia coli* Are Essential for Catalysis,” *Biochemistry*, vol. 43, no. 5, pp. 1213–1222, 2004, doi: 10.1021/bi0358059.
- [50] S. Dey, J. M. Lane, R. E. Lee, E. J. Rubin, and J. C. Sacchettini, “Structural characterization of the mycobacterium tuberculosis biotin biosynthesis enzymes 7,8-diaminopelargonic acid synthase and dethiobiotin synthetase,” *Biochemistry*, vol. 49, no. 31, pp. 6746–6760, Aug. 2010, doi: 10.1021/bi902097j.
- [51] F. Khan, V. R. Jala, N. A. Rao, and H. S. Savithri, “Characterization of recombinant diaminopropionate ammonia-lyase from *Escherichia coli* and *Salmonella typhimurium*,” *Biochem. Biophys. Res. Commun.*, vol. 306, no. 4, pp. 1083–1088, 2003, doi: 10.1016/S0006-291X(03)01100-8.
- [52] S. Bisht *et al.*, “Crystal structure of *Escherichia coli* diaminopropionate ammonia-lyase reveals mechanism of enzyme activation and catalysis,” *J. Biol. Chem.*, vol. 287, no. 24, pp. 20369–81, Jun. 2012, doi: 10.1074/jbc.M112.351809.
- [53] S. Sansaloni-Pastor, J. Bouilloux, and N. Lange, “The dark side: Photosensitizer prodrugs,” *Pharmaceuticals*, vol. 12, no. 4. MDPI AG, 2019, doi: 10.3390/ph12040148.
- [54] M. J. van der Werf and J. G. Zeikus, “5-Aminolevulinate production by *Escherichia coli* containing the *Rhodobacter sphaeroides* hema gene,” *Appl. Environ. Microbiol.*, vol. 62, no. 10, pp. 3560–3566, 1996, [Online]. Available: <http://aem.asm.org/content/62/10/3560.abstract>.
- [55] J. Zhang, Z. Kang, W. Ding, J. Chen, and G. Du, “Integrated Optimization of the In Vivo Heme Biosynthesis Pathway and the In Vitro Iron Concentration for 5-Aminolevulinate Production,” *Appl. Biochem. Biotechnol.*, vol. 178, no. 6, pp. 1252–1262, Mar. 2016, doi: 10.1007/s12010-015-1942-2.
- [56] Z. Kang, Y. Wang, P. Gu, Q. Wang, and Q. Qi, “Engineering *Escherichia coli* for efficient production of 5-aminolevulinic acid from glucose,” *Metab. Eng.*, vol. 13, no. 5, pp. 492–498, 2011, doi: 10.1016/j.ymben.2011.05.003.
- [57] F. Li, Y. Wang, K. Gong, Q. Wang, Q. Liang, and Q. Qi, “Constitutive expression of RyhB regulates the heme biosynthesis pathway and increases the 5-aminolevulinic acid accumulation in *Escherichia coli*,” *FEMS Microbiol. Lett.*, vol. 350, no. 2, pp. 209–215, Jan. 2014, doi: 10.1111/1574-6968.12322.
- [58] T. Li, Y.-Y. Guo, G.-Q. Qiao, and G.-Q. Chen, “Microbial Synthesis of 5-Aminolevulinic Acid and Its Coproduction with Polyhydroxybutyrate,” *ACS Synth. Biol.*, vol. 5, 2016, doi: 10.1021/acssynbio.6b00105.
- [59] L. Feng *et al.*, “Metabolic engineering of *Corynebacterium glutamicum* for efficient production of 5-aminolevulinic acid,” *Biotechnol. Bioeng.*, vol. 113, no. 6, pp. 1284–1293, Jun. 2016, doi: 10.1002/bit.25886.
- [60] P. Yang, W. Liu, X. Cheng, J. Wang, Q. Wang, and Q. Qi, “A New Strategy for

- Production of 5-Aminolevulinic Acid in Recombinant *Corynebacterium glutamicum* with High Yield.,” *Appl. Environ. Microbiol.*, vol. 82, no. 9, pp. 2709–17, May 2016, doi: 10.1128/AEM.00224-16.
- [61] X. Yu, H. Jin, W. Liu, Q. Wang, and Q. Qi, “Engineering *Corynebacterium glutamicum* to produce 5-aminolevulinic acid from glucose,” *Microb. Cell Fact.*, vol. 14, no. 1, p. 183, Dec. 2015, doi: 10.1186/s12934-015-0364-8.
- [62] G. Jansen, C. Wu, B. Schade, D. Y. Thomas, and M. Whiteway, “Drag&Drop cloning in yeast.,” *Gene*, vol. 344, pp. 43–51, Jan. 2005, doi: 10.1016/j.gene.2004.10.016.
- [63] M. R. Moita, “MSc Thesis: Towards the production of levulinic and itaconic acids in *Saccharomyces cerevisiae*: a contribution for understanding the molecular mechanisms of toxicity of these acids in producing cells,” Instituto Superior Técnico, 2013.
- [64] S. Z. Tan, S. Manchester, and K. L. J. Prather, “Controlling Central Carbon Metabolism for Improved Pathway Yields in *Saccharomyces cerevisiae*,” *ACS Synth. Biol.*, vol. 5, no. 2, pp. 116–124, 2016, doi: 10.1021/acssynbio.5b00164.
- [65] B. A. Blount, M. R. M. Driessen, and T. Ellis, “GC preps: Fast and easy extraction of stable yeast genomic DNA,” *Sci. Rep.*, vol. 6, pp. 1–4, 2016, doi: 10.1038/srep26863.
- [66] K. Tomokuni and M. Ogata, “Simple method for determination of urinary -aminolevulinic acid as an index of lead exposure,” *Clin Chem*, vol. 18, no. 12, pp. 1534–1538, 1972, [Online]. Available: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=4639865%5Cnpapers2://publication/uuid/EF42970E-9C27-4958-A1CD-0B715B2E82F0.
- [67] S. Mann and O. Ploux, “7,8-Diaminopelargonic acid aminotransferase from *Mycobacterium tuberculosis*, a potential therapeutic target: Characterization and inhibition studies,” *FEBS J.*, vol. 273, no. 20, pp. 4778–4789, 2006, doi: 10.1111/j.1742-4658.2006.05479.x.
- [68] M. A. Eisenberg and G. L. Stoner, “Biosynthesis of 7,8-diaminopelargonic acid, a biotin intermediate, from 7-keto-8-aminopelargonic acid and S-adenosyl-L-methionine.,” *J. Bacteriol.*, vol. 108, no. 3, pp. 1135–40, Dec. 1971, Accessed: Jun. 26, 2018. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/4945185>.
- [69] S. Mann, L. Colliandre, G. Labesse, and O. Ploux, “Inhibition of 7,8-diaminopelargonic acid aminotransferase from *Mycobacterium tuberculosis* by chiral and achiral analogs of its substrate: Biological implications,” *Biochimie*, vol. 91, no. 7, pp. 826–834, 2009, doi: 10.1016/j.biochi.2009.03.019.
- [70] V. M. Bhor, S. Dev, G. R. Vasanthakumar, and A. Surolia, “Spectral and kinetic characterization of 7,8-diaminopelargonic acid synthase from *Mycobacterium tuberculosis*,” *IUBMB Life*, vol. 58, no. 4, pp. 225–233, 2006, doi: 10.1080/15216540600746997.

- [71] A. C. Eliot, J. Sandmark, G. Schneider, and J. F. Kirsch, “The dual-specific active site of 7,8-diaminopelargonic acid synthase and the effect of the R391A mutation,” *Biochemistry*, vol. 41, no. 42, pp. 12582–12589, Oct. 2002, doi: 10.1021/bi026339a.
- [72] H. Käck, J. Sandmark, K. Gibson, G. Schneider, and Y. Lindqvist, “Crystal structure of diaminopelargonic acid synthase: Evolutionary relationships between pyridoxal-5'-phosphate-dependent enzymes,” *J. Mol. Biol.*, vol. 291, no. 4, pp. 857–876, Aug. 1999, doi: 10.1006/jmbi.1999.2997.
- [73] L. Bao, J. J. Li, C. Jia, M. Li, and X. Lu, “Structure - oriented substrate specificity engineering of aldehyde - deformylating oxygenase towards aldehydes carbon chain length,” *Biotechnol. Biofuels*, pp. 1–14, 2016, doi: 10.1186/s13068-016-0596-9.
- [74] A. C. Stiel, M. Nellen, and B. Höcker, “PocketOptimizer and the Design of Ligand Binding Sites,” in *Methods in Molecular Biology: Design and Creation of Ligand Binding Proteins*, vol. 1414, 2016, pp. 63–75.
- [75] F. H. Arnold, “Directed Evolution: Bringing New Chemistry to Life,” *Angew. Chemie - Int. Ed.*, vol. 57, no. 16, pp. 4143–4148, 2018, doi: 10.1002/anie.201708408.
- [76] U. Markel, K. D. Essani, V. Besirlioglu, J. Schiffels, W. R. Streit, and U. Schwaneberg, “Advances in ultrahigh-throughput screening for directed enzyme evolution,” *Chem. Soc. Rev.*, vol. 49, no. 1, pp. 233–262, 2020, doi: 10.1039/c8cs00981c.
- [77] D. HELD, K. Yaeger, R. Novy, and D. Held, “New coexpression vectors for expanded compatibilities in E. coli,” *Innovations*, vol. 18, 2003.
- [78] S. Létoffé, P. Delepelaire, and C. Wandersman, “The housekeeping dipeptide permease is the Escherichia coli heme transporter and functions with two optional peptide binding proteins,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 34, pp. 12891–12896, 2006, doi: 10.1073/pnas.0605440103.
- [79] G. Giaever *et al.*, “Functional profiling of the Saccharomyces cerevisiae genome,” *Nature*, vol. 418, no. 6896, pp. 387–391, Jul. 2002, doi: 10.1038/nature00935.
- [80] K. N. Maclean, M. Janošík, J. Oliveriusová, V. Kery, and J. P. Kraus, “Transsulfuration in Saccharomyces cerevisiae is not dependent on heme: purification and characterization of recombinant yeast cystathionine β -synthase,” *J. Inorg. Biochem.*, vol. 81, no. 3, pp. 161–171, Aug. 2000, doi: 10.1016/S0162-0134(00)00100-8.
- [81] E. G. Gollub, K. P. Liu, J. Dayan, M. Adlersberg, and D. B. Sprinson, “Yeast mutants deficient in heme biosynthesis and a heme mutant additionally blocked in cyclization of 2,3 oxidosqualene,” *J. Biol. Chem.*, vol. 252, no. 9, pp. 2846–2854, 1977.
- [82] J. T. Cuperus, R. S. Lo, L. Shumaker, J. Proctor, and S. Fields, “A tetO Toolkit to Alter Expression of Genes in Saccharomyces cerevisiae,” *ACS Synth. Biol.*, vol. 4,

- no. 7, pp. 842–852, Jul. 2015, doi: 10.1021/sb500363y.
- [83] E. Garí, L. Piedrafita, M. Aldea, and E. Herrero, “A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in *Saccharomyces cerevisiae*,” *Yeast*, vol. 13, no. 9, pp. 837–848, 1997, doi: 10.1002/(SICI)1097-0061(199707)13:9<837::AID-YEA145>3.0.CO;2-T.
 - [84] O. Protchenko, M. Shakoury-Elizeh, P. Keane, J. Storey, R. Androphy, and C. C. Philpott, “Role of PUG1 in inducible porphyrin and heme transport in *Saccharomyces cerevisiae*,” *Eukaryot. Cell*, vol. 7, no. 5, pp. 859–871, May 2008, doi: 10.1128/EC.00414-07.
 - [85] W. E. Schauer and J. R. Mattoon, “Heterologous expression of human 5-aminolevulinate dehydratase in *Saccharomyces cerevisiae*,” *Curr. Genet.*, vol. 17, no. 1, pp. 1–6, Jan. 1990, doi: 10.1007/BF00313241.
 - [86] M. J. Young and D. A. Court, “Effects of the S288c genetic background and common auxotrophic markers on mitochondrial DNA function in *Saccharomyces cerevisiae*,” *Yeast*, vol. 25, no. 12, pp. 903–912, Dec. 2008, doi: 10.1002/yea.1644.
 - [87] S. M. Fendt and U. Sauer, “Transcriptional regulation of respiration in yeast metabolizing differently repressive carbon substrates,” *BMC Syst. Biol.*, vol. 4, pp. 6–8, 2010, doi: 10.1186/1752-0509-4-12.
 - [88] J. Quarterman *et al.*, “Rapid and efficient galactose fermentation by engineered *Saccharomyces cerevisiae*,” *J. Biotechnol.*, vol. 229, pp. 13–21, Jul. 2016, doi: 10.1016/j.jbiotec.2016.04.041.
 - [89] A. Lin *et al.*, “Utilization of a Strongly Inducible DD12 Promoter to Control Gene Expression in *Saccharomyces cerevisiae*,” *Front. Microbiol.*, vol. 9, p. 2736, Nov. 2018, doi: 10.3389/fmicb.2018.02736.
 - [90] S. Labbé and D. J. Thiele, “Copper ion inducible and repressible promoter systems in yeast,” *Methods Enzymol.*, vol. 306, pp. 145–153, Jan. 1999, doi: 10.1016/S0076-6879(99)06010-3.
 - [91] N. Baldi *et al.*, “Functional expression of a bacterial α -ketoglutarate dehydrogenase in the cytosol of *Saccharomyces cerevisiae*,” *Metab. Eng.*, vol. 56, pp. 190–197, Dec. 2019, doi: 10.1016/j.ymben.2019.10.001.
 - [92] K. Sasaki, M. Watanabe, and N. Nishio, “Inhibition of 5-aminolevulinic acid (ALA) dehydratase by undissociated levulinic acid during ALA extracellular formation by *Rhodobacter sphaeroides*,” *Biotechnol. Lett.*, vol. 19, no. 5, pp. 421–424, 1997, doi: 10.1023/A:1018331824331.

Chapter 4– Strategies for the successful implementation of an in silico predicted LA-forming pathway using D-ALA as a precursor

**Chapter 5 – Prospecting of
metabolic pathways for
implementation of microbe-based
production of Methacrylic Acid
(MAA)**

Statement of research contributions:

The MSc student Fernão Mendes performed the enzymatic assays with ACMSD and part of the enzymatic assays with CAD (in MMF media)

5.1. Abstract

This chapter is dedicated to the establishment of putative microbial production routes of the ITA-derivative methacrylic acid (MAA). For this, *in silico* prospecting of possible MAA forming routes was performed resorting to a strategy similar to the one described in chapter 3 for levulinic acid. Compared to what had been reported in the field, this analysis expanded the range of possible precursors that could be used to include 3-ureidoisobutyric acid, mesaconyl-C4-coA, acrylic acid, 2-methylbutyric acid, propanoic acid, methylmalonic semialdehyde, mesaconyl-C1-coA, tiglyl-coA and acrylyl-coA. Out of this analysis, 5 possible routes were devised using as possible precursors mesaconyl-C4-CoA, mesaconic acid, isobutyryl-coA, 3-hydroxyisobutyric acid, methylmalonic semialdehyde and 3-aminoisobutyric acid. For the enzyme assignment step three pathway prospecting algorithms were used (Atlas of Biochemistry, MINE and Transform-Miner), resulting in the assembly of 6 possible pathways that were, in essence, identical to those that had been previously suggested since it was not possible to assign an enzyme for the conversions involving the herein identified new precursors. In order to start experimental validation of the proposed pathways, it was tested the capacity of the *Aspergillus terreus* cis-acconitic acid decarboxylase enzyme CadA and of *Pseudomonas syringae* aminocarboxymuconate semialdehyde decarboxylase (ACMSD) to undertake the decarboxylation of mesaconic acid into MAA. Under the experimental conditions that were used for the *in vitro* enzymatic assays (that, among other aspects involved the use of HPLC to detect the presence of MAA) it was not possible to detect MAA, suggesting that further engineering of these enzymes may be required in order to render them able to mediate the envisaged conversion. A noticeable aspect concerned the identification of an unknown compound in the extracts enriched in ACMSD, which was presumed to be 2-hydroxy-2-methylsuccinic acid, based on the previous demonstration that besides undertaking decarboxylations these enzymes are also able to mediate hydrations.

5.2. Introduction

MAA is a bulk molecule with high industrial interest, being a widely used monomer with a projected market close to 1.12 billion USD in 2019 [1]. The main use of MAA, as well as of its more famous ester, methyl-methacrylate (MMA), is to produce poly-MMA, a polymer used in acrylic glass largely explored in the manufacture of electronics, car components, lights, signs and displays [2]. Additionally, MAA and its derivative 2-hydroxyethyl methacrylic acid (HEMA) are also raising interest as possible co-monomers in the formation of polymers for biomedical applications like drug delivery or wound dressings [3]–[5]. Nowadays, both MAA and MMA are produced resorting to the oil derivative acetone cyanohydrin using a route that poses serious environmental challenges [6]. While a potential “greener” alternative has been presented by the development of the Alpha process (detailed in the introduction of this thesis), in which ACH is replaced by syngas (methanol, carbon monoxide) and ethylene, it still presents the problem of requiring manipulation of potentially hazardous compounds such as methanol and formalin [7], [8]. In this sense, aiming for the sustainability of MAA production, a process in which this bulk chemical can be produced entirely from fermentation using renewable substrates is envisaged. However, a biological production route of MAA is hindered by this compound not being reported to be naturally produced by any organism, although its CoA ester (methacrylyl-CoA) is a natural intermediate in the valine degradation pathway used in some bacterial and fungal species.

One strategy that has been envisaged to produce MAA involves a biocatalytic process, where the synthesis of a precursor by fermentation is followed by its chemical conversion to MAA. The precursors considered for this mixed process are citramalic acid [9], ITA [10] [11] [12], 2-hydroxyisobutyric [13] and citric acid [14]; all compounds whose microbial production has been established. In what concerns the development of a process involving only production of MAA by fermentation, the sole literature available are two patents filed by the companies Genomatica and Lucite. In the case of Genomatica’s patent, three main potential metabolites are identified as possible precursors for MAA biosynthesis: 3-aminoisobutyrate, 3-hydroxyisobutyrate and mesaconic acid [15], [16]. The precursors 3-hydroxyisobutyric and isobutyric acids can be obtained from multiple intracellular sources while mesaconic acid can be obtained, for example, from the alpha-ketoglutarate via the glutamate degradation pathway described in *Clostridium tetanomorphum*. It is also mentioned the possibility of converting methacrylyl-CoA into MAA through the removal of the CoA thioester. In this line,

strategies alternative to the use of valine degradation pathway to produce methacrylyl-CoA have been identified in the Genomatica patent, resulting in the identification of isobutyryl-coA, 2- and 3-hydroxyisobutyryl-coA as possible precursors. In Figure 5.1A is compiled the full list of 8 MAA-forming pathways described in the patent. An important drawback is the fact that for all the predicted pathways there is at least one reaction step with no assigned enzyme, which can hamper the successful implementation of these pathways.

In the case of the patent filed by Lucite the route approached to produce MAA involves the conversion of isobutyric acid into MAA, via methacrylyl-CoA [17]. In this conversion route, isobutyryl-coA is reduced to methacrylyl-CoA through the action of an acyl-coA oxidase from *Arabidopsis thaliana* (ACX4), and the coA thioester is subsequently removed from Methacrylyl-CoA by an hydroxybenzoyl coA thioesterase from *Arthrobacter sp.* SU (4HBT) (Figure 5.1B). A noticeable aspect is that the activity of these enzymes was validated *in vivo* (in *E. coli*) resulting in production of MAA, although supplementation of the medium with isobutyric acid was required [17].

In this chapter the proposed pathways for production of MAA were further revised mainly with the objective of trying to perform an assignment on enzymatic steps still lacking that essential feature, and also with the objective of trying to identify new possible substrates. In addition, it was decided to test *in vivo* the possibility of using two identified decarboxylases to promote the conversion of mesaconic acid into MAA, this being one of the pathways that had been identified in the patent filled by Genomatica [15], [16].

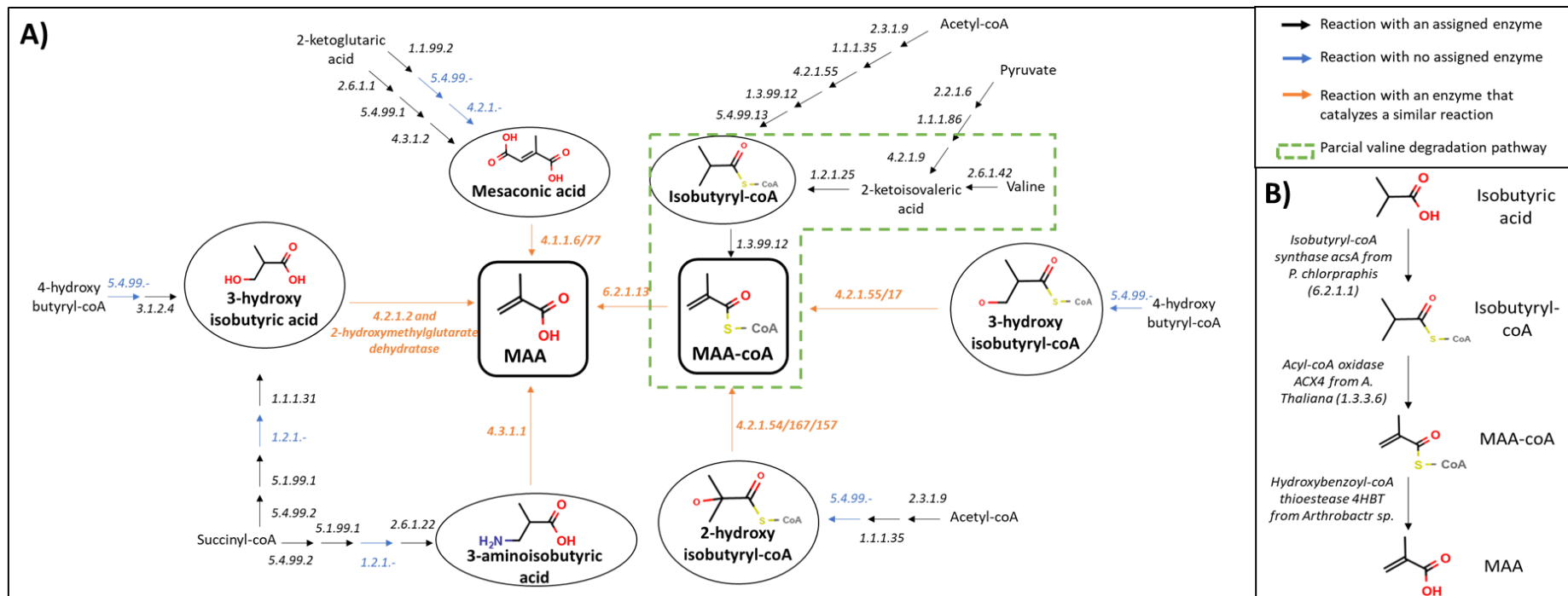


Figure 5.1 Candidate MAA biosynthetic pathways that have been published. A) The 8 MAA forming pathways published in [15], [16] are depicted. The 6 MAA and Methacrylyl-CoA precursors are depicted in circles. The reaction steps of each pathway that have been reported to be catalyzed by a specific enzyme are represented in black with the corresponding enzyme E.C. number. The enzymatic steps depicted in blue are hypothesized to be catalyzed by an enzyme belonging to E.C. sub (sub)class and the enzymatic steps depicted in orange have similar reactions catalyzed the enzymes represented by the orange EC numbers. Both blue and orange steps don't have experimental validation. The natural valine degradation pathway where MAA is an intermediate is pinpointed with green dashes. B) The bioconversion of isobutyric acid to MAA in *E. coli* has been reported in [17]. The heterologous enzymes expressed to assemble the pathway are annotated.

5.3. Methods

5.3.1. *In silico* pathway prospecting

The precursors were selected based on their predicted conversion to MAA and Methacrylyl-CoA according with the information available in the “Atlas of Biochemistry” and “MINE” databases. In Atlas of Biochemistry only reactions with a negative Gibbs free energy were considered. Precursors were also selected based on survey of the entire library of KEGG compounds selecting those showing similarity with MAA (Tanimoto coefficient above 0.65) as possible candidates. The SIMCOMP tool (<https://www.genome.jp/tools/simcomp/>) was used to make the comparative analysis between MAA and the KEEG compounds. To search for possible conversion routes between the precursors and target molecules the Transform-Miner web server was used [18]. Pathways with up to 3 reaction steps and a minimal similarity score of 0.6 were searched. When multiple pathways were predicted by Transform-Miner, the selection was based on two criteria which were applied subsequently: 1) pathway involving a lower number of steps; 2) pathways showing a higher cumulative similarity score. To assign an enzyme to a reaction, the top matches provided by Transform-Miner were analyzed. If the similar reaction is in the reverse direction of the predicted reaction, this enzyme was discarded. When Transform-Miner provided similar reactions that can be catalyzed by multiple enzymes (such as succinate ligase), all enzymes able to catalyze the required reaction were manually mined to find the best enzyme candidate. The selection considered substrate range and similarity of the native substrate to the synthetic one.

5.3.2. Experimental methods

5.3.2.1. Microorganisms and culture conditions

The *S. cerevisiae* strains were maintained in YPD medium and in MMB (when required to maintain the selection). The composition of YPD medium is 2% glucose (Nzytech), 2% peptone (Nzytech) and 1% yeast extract (Nzytech). The composition of MMB is 2% glucose/galactose (Nzytech/Merck) 0.267% ammonium sulfate and 0.17% Yeast Nitrogen Base without aminoacids (Difco). MMB media is supplemented with the required aminoacids and nucleobase (all from Sigma) to maintain growth and selection in the following concentrations: 20 mg/L histidine, 60 mg/L leucine, 20 mg/L methionine,

Chapter 5– Prospecting of metabolic pathways for implementation of microbe-based production of Methacrylic Acid (MAA)

20 mg/L tryptophan, 20 mg/L uracil. MMF media was used to express *cadI*, the composition being 20 g/L carbon source, 1.7 g/L Yeast Nitrogen Base without aminoacids and ammonium sulphate, 2.67 g/L ammonium sulphate, 2 g/L potassium sulphate and a supplementation solution of aminoacids, vitamins and trace elements to the final solution represented in Table 5.1. *E. coli* strains were maintained and cultured in LB medium (Nzytech), to which 150 mg/L Amp (Nzytech) was added to maintain selection (LBA). The lists of strains and plasmids used in this chapter are available in Table 5.2 and Table 5.3, respectively. The pET16b-nbaD plasmid was kindly provided by Prof. Dr. Aimin Liu (Department of Biochemistry, UniVersity ofTexas Southwestern Medical Center,).

Table 5.1. Vitamins, aminoacids and trace elements used to supplement the MMF media

CSM U-		VITAMINS		TRACE ELEMENTS	
Aminoacid	Concentration (mg/L)	Chemical compound	Concentration (mg/L)	Chemical compound	Concentration (mg/L)
Adenine	2,5	D-biotin	0,05	Na ₂ EDTA	1.5
Arginine Hcl	12,5	Panthothenic acid calcium salt	1	ZnSO ₄ ·7H ₂ O	0.45
L-Aspartic Acid	20	Nicotinic acid	1	MnCl ₂ ·2H ₂ O	
L-Histidine Hcl	5	Myo-inositol	25	CoCl ₂ ·6H ₂ O	
L-Isoleucine	12,5	Thiamine chloride hydrochloride	1	CuSO ₄ ·5H ₂ O	0.03
L-Leucine	25	Pyridixol hydrochloride	1	Na ₂ MoO ₄ ·2H ₂ O	0.03
L-Lysine Hcl	12,5	4-Aminobenzoic acid	0,2	CaCl ₂ ·2H ₂ O	0.45
L-Methionine	5			FeSO ₄ ·7H ₂ O	0.30
L-Phenylalanine	12,5			H ₃ BO ₃	0.1
L-Threonine	25			KI	0.01
L-Tryptophane	12,5				
Uracil	5				
L-Tyrosine	12,5				
L-Valine	35				

Table 5.2. List of strains used in Chapter 5.

Name	Description	Source
------	-------------	--------

Chapter 5– Prospecting of metabolic pathways for implementation of microbe-based production of Methacrylic Acid (MAA)

<i>S. cerevisiae</i>	BY4741	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0	Euroscarf
	Y258	MATa pep4-3, his4-580, ura3-53, leu2-3,112	BSRG lab
<i>E. coli</i>	DH5	F ⁻ <i>endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20</i> ϕ80dlacZΔM15 Δ(<i>lacZYA-argF</i>)U169, hsdR17(<i>r_K⁻m_K⁺</i>), λ ⁻	Invitrogen
	BL21(DE3)	F ⁻ <i>ompT gal dcm lon hsdS_B(r_B⁻m_B⁻)</i> λ(DE3 [<i>lacI lacUV5-T7p07 ind1 sam7 nin5</i>]) [<i>malB</i> ⁺] _{K-12} (λ ^S)	Invitrogen

Table 5.3. List of plasmids used in Chapter 5.

Name	Description	Source
pGreg586	CEN plasmid with URA3 marker, expression of N-terminal 6x His tagged protein under control of GAL1 promote	[19]
pGREG_AtCad1	pGreg586-pGal1-Atca1 (codon optimized for <i>S. cerevisiae</i>)	[20]
pET16b-nbaD	Bacterial expression vector with T7 lac promoter, AmpR	[21]

5.3.2.2. In vitro enzymatic assays to assess the capacity of *A. terreus* CadA to convert mesaconic acid into MAA

To obtain the *A. terreus* CadA protein to be used in the envisaged conversion of mesaconic acid into MAA, *S. cerevisiae* BY4741 cells were transformed with the cloning vector pGreg586 or with pGREG_AtCad1 (which drives expression of AtCadA from the galactose-inducible plasmid *GALI*) using the lithium acetate method. The transformants obtained were grown overnight (at 30°C and 250 rpm) in mineral medium MMF 2% raffinose and then inoculum (at an OD_{600nm} of 0.1) in fresh MMF 2% raffinose medium. The cells were incubated until the culture reached the mid-exponential phase (OD_{600nm}~0.5-0.8) after which they were re-inoculated (at OD_{600nm} of 0.1) in 150 mL of the inducing medium MMF Galactose. After 24 hours of growth the cells were harvested by centrifugation (5 min at 5000 rpm 4°C), the pellets washed with 25 mL of water and 25 mL of potassium phosphate buffer (25 mM KPO₄ pH=8) and stored at -80°C. For

preparation of crude cell extracts, the pellets were resuspended in 500 μ L of ice-cold lysis buffer (25 mM Tris-HCl (pH 7.5), 1 mM EDTA, 100 mM NaCl, 1 mM DTT, 1 mM PSMF, NZYTech protease inhibitor cocktail for mammalian cells) and the cellular suspension was transferred to tubes with pre-chilled beads (1:1 ratio of bead and cellular suspension). The cells were disrupted by doing 10 cycles of 2 min vortex and 1 min of cooling on ice, and the bead/lysate mixture was centrifuged at full speed at 4°C for 15 min to remove cell debris. The protein concentration in each lysate was determined by the Bradford method, using BSA standards for the calibration curve. To confirm protein expression, SDS-PAGE was performed in whole-yeast cells cultivated under the same conditions used above. For that, 1.5 mL of culture were centrifuged, and the pellet resuspended in 100 μ L of lithium acetate 2M. After 5 min incubation on ice, the suspension was centrifuged, the pellet was resuspended in 100 μ L sodium hydroxide 0.4 M and incubated on ice for 5 min. The suspension was then centrifuged, resuspended in 100 μ L Laemmli buffer, incubated at 100°C for 5 min and centrifuged. The supernatant was collected as the soluble fraction and pellet was resuspended in 100 μ L Laemmli buffer as the insoluble fraction. 5 μ L of each sample were applied in each lane of the gel.

For the activity assay, a protocol was devised based on the native reaction in which CadA is involved which is the conversion of cis-aconitic acid into ITA (Dwiarti, Yamane, Yamatani, Kahar, & Okabe, 2002). Specifically, 100 μ L of the crude cell extract was incubated at 37°C in the presence 30 mM of substrate (cis-aconitic or MES) and 1.5 mL of sodium phosphate buffer 0.2 M (pH=6.2) for 10 min. The reaction was stopped by the addition of 100 μ L of 12 M HCl. Afterwards, 10 μ L of the reaction mixture were separated in an HPLC equipped with an Aminex HPX- 87H column (Biorad) and eluted with 0.05% sulfuric acid at a flow rate of 0.6 mL/min. A UV detector set at 210 nm was used for detection mesaconic acid, cis-aconitic acid, ITA and MAA with elutions at 16 min, 6 min, 14 min and 22 min, respectively.

5.3.2.3. Preparation of *E. coli* crude cell extracts enriched in *Pseudomonas syringae* ACMSD

To obtain the *Pseudomonas syringae* ACMSD protein to be used in the envisaged conversion of mesaconic acid into MAA, *E. coli* BL21(DE3) cells were transformed with the pET16b-nbaD plasmid and then inoculated in 150 mL LBA supplemented with 5 mM iron sulphate and 5 mM zinc chloride. The cells were incubated at 37°C, 250 rpms and when the OD_{600nm} of the culture reached 0.6, 1 mM IPTG was added to the broth to induce

expression. Afterwards, the culture was incubated at 28°C and 250 rpm for 4 hours. After that time, cells were centrifuged and the cell pellets were store at -80°C for future use. The pellets were disrupted by sonication (7 cycles of 10 min of sonication with 1 min of interval for cooling on ice) and centrifuged for 45 minutes at 4°C at 2000 rpms. The concentration in the crude cell extracts was measured by Bradford. To confirm protein expression, SDS-PAGE was performed in whole-*E. coli* lysates ressuspending for that 1.5 mL of the culture (prior the final harvesting step) in Laemmli buffer, followed by an incubation at 100°C for 5 min and subsequent application in each lane of the gel.

The in vitro assay was performed in a total volume of 2 mL in 25 mM HEPES buffer pH=7, 5% glycerol and 6 mM mesaconic acid. The reaction was initiated by addition of 300 or 500 µL crude cell extracts and incubated at 22°C for 6 hours. Samples were taken after 1h, 3h, and 6h for HPLC analysis (as described for the CadA assays).

5.4. Results and Discussion

5.4.1. *In silico* prospecting of pathways for microbe-based production of MAA

This pathway prospecting analysis started with the identification of suitable compounds that could serve as precursors for the synthesis of MAA. For that, the databases Atlas of Biochemistry, MINES and KEEG compounds, were searched for reactions involving metabolites structurally similar to MAA but also to methacrylyl-CoA since it can be enzymatically converted to MAA, as reported in the Lucite patent [17]. From the list of precursors, a compilation of conversions leading to the formation of MAA or methacrylyl-CoA was searched, to find possible enzymes that can catalyze the formation of these two molecules.

5.4.1.1. Selection of precursors

To find the potential MAA precursors in the MINE database (a collection of the whole theoretical reactome based on BNICE reaction rules and KEGG metabolites [22]) was searched for predicted enzymatic reactions having MAA as a product. 29 compounds were predicted by the database to be converted into MAA and each substrate was analyzed to understand if there is already a known biological metabolic pathway to produce it

(Table 5.4). Out of the 29 potential substrates (listed in Table 5.4), eight were selected as more interesting candidates to be used as MAA precursors with the main reason for that being the fact that their production in microbes has already been achieved: ITA, isobutyric acid, acrylic acid, mesaconic acid, methacrylyl-CoA, mesaconyl-C4-CoA, 3-ureidoisobutyric acid and 3-aminoisobutyric acid (their structures are represented in Figure 5.2). As already well detailed in this thesis, ITA is a secondary metabolic product of *A. terreus* and *U. maydis* with well elucidated biosynthetic routes [23]. Isobutyric and acrylic acids, although not “natural” microbial metabolites, have been successfully produced in *E. coli* through the assembly of synthetic pathways using as precursors isobuturaldehyde and 3-hydroxypropionic, respectively [24],[25]. Mesaconic acid is an isomer of ITA and is an intermediate in the *Methylobacterium extorwuens* ethylmalonyl pathway, where the precursor is acetyl-coA [26]. Additionally, mesaconic acid is also a product of glutamate degradation in *Clostridium sp* [27] and the pathway has been ported to *E. coli*, enabling heterologous production of mesaconic acid [28], [29]. Methacrylyl-coA and 3-hydroxyisobutyric are known intermediates in valine catabolism and mesaconyl-C4-coA is an intermediate in the glyoxylate assimilation cycle in phototrophic bacteria. Mesaconyl-C4-CoA can also be obtained from isomerization of mesaconyl-C1-coA which is found in the ethylmalonyl-coA cycle in *Rhodobacter sp* and the methylaspartate cycle in *Halobacteria sp*. 3-ureidoisobutyric and 3-aminoisobutyric acids are found in the thymine reductive degradation pathway in various bacteria, including in *E. coli* as well as in the valine degradation pathway. The remaining 20 precursors were disregarded for various reasons out of which the fact that their implementation inside microbial cells seemed very difficult due to the big size of these molecules (e.g., glaucolide A or molephantin), consequently significantly increasing the resources the cells would need to allocate in this task). In other cases, microbial production was not even described (as it is the case of bisphenol A or glycidylmethacrylate (Table 5.4).

Chapter 5– Prospecting of metabolic pathways for implementation of microbe-based production of Methacrylic Acid (MAA)

Table 5.4. Predicted conversions in the MINE database where MAA is a product.

	Substrate	Reaction	Metabolic pathways involved or information of the substrate	Passed the filter?
1	Isobutyric	1.3.1	Not naturally produced; synthetic pathway assembled in E. coli, from isobutyraldehyde [24]	Yes
2	Mesaconic	4.1.1	Glyoxylate and dicarboxylate metabolism C5 branched dibasic acid metabolism	Yes
3	Methacrylyl-coA	3.1.2	Valine degradation	(considered a product, explored in table 3)
4	3-hydroxyisobutyric	4.2.1		Yes
5	Mesaconyl-C4-coA	3.7.1	Glyoxylate assimilation and isomerization of Mesaconyl-C1-coA	Yes
6	Acrylic	2.1.1	Synthetic pathway assembled in E. coli through 3-hydroxypropionic [25]	Yes
7	Itaconic	4.1.1	C5-branched dibasic metabolism	Yes
8	3-ureidoisobutyric	4.3.1	Thymine reductive degradation	Yes
9	3-aminoisobutanoic	4.3.1		Yes
10	MMA	3.1.1	MAA derivative, no known pathway to produce	No (not a biological product)
11	2-Hydroxyethyl methacrylic (HEMA)	3.1.1		
12	4-methylene-2-oxoglutaric	3.7.1.c	Benzoate metabolism C5 branched dibasic acid metabolism	No (not a biological product)
13	Penicillic acid	3.7.1	Fungal toxin	No (large MW molecule)
14-25	Molephantin, Erioflorin methacrylate, Glaucolide A, Zexbrevin B, Vernolide, Chloranocryl, Calaxin, Phantomolin, Deoxyelephantopin, Elephantopin, Goyazensolide, Radiatin	3.1.1	Phytochemical compound	No (large MW molecule)
26-28	2-Methacryloyloxyethyl phenyl phosphate, Bisphenol A glycidylmethacrylate, Trimethylolpropane trimethacrylate (TMPT)	3.1.1.	None	No (not a biological product)
29	Bisphenol A dimethylacrylate	1.14.13		

Another approach that was used to identify possible precursors for MAA biosynthesis included the search of the KEGG SIMCOMP tool (<https://www.genome.jp/tools/simcomp/>) for molecules structurally similar to MAA. For this analysis only the compounds showing a Tanimoto similarity coefficient above 0.65 were considered, resulting in 16 possible precursors (Table 5.5), seven of which (isobutyric, propanoic, acrylic, 3-hydroxyisobutyric, 3-aminoisobutyric, 2-methylbutyric acids and methylmalonic semialdehyde) were described to be produced in microbes (Table 5.5). Although the tautomers 2-aminoacrylate and 2-iminopropanoate, are intermediates in serine degradation and methionine biosynthesis, both are unstable and undergo spontaneous hydrolysis of the C-N bond [30] and for this reason they were disregarded. The metabolites found to be structurally similar to MAA include 4 precursors also predicted in Atlas of Biochemistry (isobutyric, acrylic, 3-hydroxyisobutyric and 3-aminoisobutyric) but three (propanoic and methylbutyric acids and malonate semialdehyde) were only identified using this approach. Propanoic acid is found in the archeal pathway for 2-oxobutanoic acid degradation and is also a byproduct of nicotinate degradation in *Eubacterium barkeri* and in the synthesis of heme d1 in denitrifying bacteria. 2-methylbutanoate is a major product of the parasitic nematode *Ascaris sp.*, which ferments it from propanoate [25] and methyl malonate semialdehyde is an intermediate in the widely distributed valine degradation pathway. Altogether, 11 molecules were identified as possible precursors for MAA biosynthesis these being compiled in the schematic representation shown in Figure 5.2. Notably, three of these precursors, 3-hydroxyisobutyric acid, isobutyric acid and mesaconic acid, have already been suggested as interesting candidates in the patents filed by Genomatica and Lucite described above [15], [16] [17].

Chapter 5– Prospecting of metabolic pathways for implementation of microbe-based production of Methacrylic Acid (MAA)

Table 5.5. Molecules found to be structurally similar to MAA in KEGG using the SIMCOMP tool. The Tanimoto coefficient threshold was 0.65.

	Molecule	Tanimoto coefficient	Metabolic pathways involved or information of the substrate	Passed the filter?
1	Isobutyric	0.85	Already in Table 1	Yes
2	Tiglic	0.8	No	No (not a biological product)
3	MMA	0.77	MAA derivative, no known pathway to produce	No (not a biological product)
4	P-MMA	0.77	MAA derivative, no known pathway to produce	No (not a biological product)
5	Propanoic	0.70	2-oxobutanoate and nicotinate catabolism, heme d1 synthesis	Yes
6	Acrylic	0.70	Already in Table 1	Yes
7	PolyAcrylic	0.7	No	No (not a biological product)
8	3-hydroxyisobutyric	0.69	Already in Table 1	Yes
9	3-aminoisobutyric	0.69	Already in Table 1	Yes
10	Cyclopropanecarboxylic	0.69	Aminobenzoate metabolism, no known pathway to produce	No (not a biological product)
11	2-methylbutyric	0.69	Fermentation from propanoate	Yes
12	Methylmalonic semialdehyde	0.67	Valine, leucine and isoleucine degradation	Yes
13	Propynoic	0.67	Can be converted to malonated semialdehyde but not reverse	No (the synthesis pathway is not known)
14	2-aminoacrylic	0.67	Serine degradation and methionine biosynthesis	No (unstable intermediate)
15	2-chloroacrylic	0.67	Chloroacrylate degradation	No (the synthesis pathway is not known)
16	2-iminopropanoic	0.67	Serine degradation and methionine biosynthesis	No (unstable intermediate)

Another approach that was envisaged for production of MAA involves the use of its ester, methacrylyl-CoA, as a possible substrate [15]–[17]. As said above, methacrylyl-CoA is an intermediate of valine catabolism (resulting from isobutyryl-coA degradation), however, we have hypothesized whether methacrylyl-CoA could be obtained with a different route, this also having in mind to avoid the need of supplying the broth with isobutyric acid. Again, the Atlas of Biochemistry database was used, searching for reactions (having a negative Gibbs free energy) that result in methacrylyl-CoA formation (Table 5.6). Three new precursors are predicted to have the potential to be converted to methacrylyl-CoA: mesaconyl-C1-coA, acrylyl-coA and tiglyl-coA. Strikingly, these substrates had been identified before (in their “CoA-free” version) in the search for substrates for MAA that we have conducted before, with the main difference of tiglyl-coA (but not tiglic acid) being a described intermediate in isoleucine catabolism. Mesaconyl-C1-coA can be obtained from the enzymatic isomerization of mesaconyl-C4-coA and is an intermediate in the ethylmalonyl pathway in *R. spheroides* and in the methylaspartate cycle in *Halobacterium* sp. Acrylyl-coA is an intermediate of pyruvate fermentation to propanoate and β -alanine biosynthesis. However, the most relevant from a metabolic engineering perspective is the 3-hydroxypropionic acid pathway that has been introduced in *E. coli*, where acrylyl-coA is an intermediate, derived from endogenous propionic acid [31]. It is noteworthy that Atlas of Biochemistry pinpoints KEGG reactions that are most similar with the predicted ones, which could indicate potential enzymes to catalyze the conversion of these precursors in methacrylyl-CoA. However, in the case of the transformations herein under analysis the predicted reactions are quite different from the predicted ones (Table 5.6). One example is the predicted decarboxylation of mesaconyl-C1-coA to methacrylyl-CoA, which in Atlas is predicted to be catalyzed by enzymes belonging to the 1.14.13 sub(sub)class (monooxygenases) and was categorized as similar to the reaction catalyzed by peroxyaminoacrylate reductase. In this case, the predicted reaction and the KEGG reaction are very different in mechanism (decarboxylation and reduction) and in substrates (aminoacrylate has a terminal N group). For this reason, the KEGG similar reactions pinpointed in Atlas of Biochemistry were not taken into account for the enzyme assignment step, which is described in the next subsection.

Chapter 5– Prospecting of metabolic pathways for implementation of microbe-based production of Methacrylic Acid (MAA)

Table 5.6. Predicted conversions in the Atlas of Biochemistry database where Methacrylyl-CoA is the product. Only conversions with negative Gibbs free energy were considered. Conversions that had been already predicted in MINE were disregarded (HEMA, MMA and chronoacyl). MTHF: 5,10-methylenetetrahydrofolate; THMPT: Methylene-tetrahydromethanopterin.

Substrate	Reaction	Metabolic pathways involved or information of the substrate	Passed the filter?
Mesaconyl-C1-coA	1.14.13 Most similar KEGG reaction is peroxyaminoacrylate reductase	Ethylmalonyl pathway, methylaspartate cycle, isomerization of mesaconyl-C4-coA and coA transfer to MES	Yes
Isobutyryl-coA*	Most similar KEGG reaction is 2-methylpropanoyl-CoA:(acceptor) 2,3-oxidoreductase (associated to 1.3.8.1/7 and 1.3.99.12)	Valine degradation	Yes
Tyglyl-coA	2.1.1 (via SAM) Most similar KEGG reaction is SAM:6-methylsalicylyl-CoA 3-C-methyltransferase	Isoleucine catabolism	Yes
Acrylyl-coA	2.1.1 (via SAM) Most similar KEGG reaction is SAM:6-methylsalicylyl-CoA 3-C-methyltransferase		Yes
	2.1.1. (via MTHF) Most similar KEGG reaction is MTHF: dUMP C-methyltransferase	Pyruvate fermentation to propanoate, β -alanine biosynthesis and 3-hydroxypropionic acid pathway	Yes
	2.1.1 (via THMPT) Most similar KEGG reaction is N-methylation of) L-alpha-Acetyl-N-normethadol		Yes

* This conversion had a calculated positive Gibbs free energy. However, the reaction similarity score (Bridgit) is 1, since this is a KEGG reaction, proven to occur in vivo and with associated enzymes.

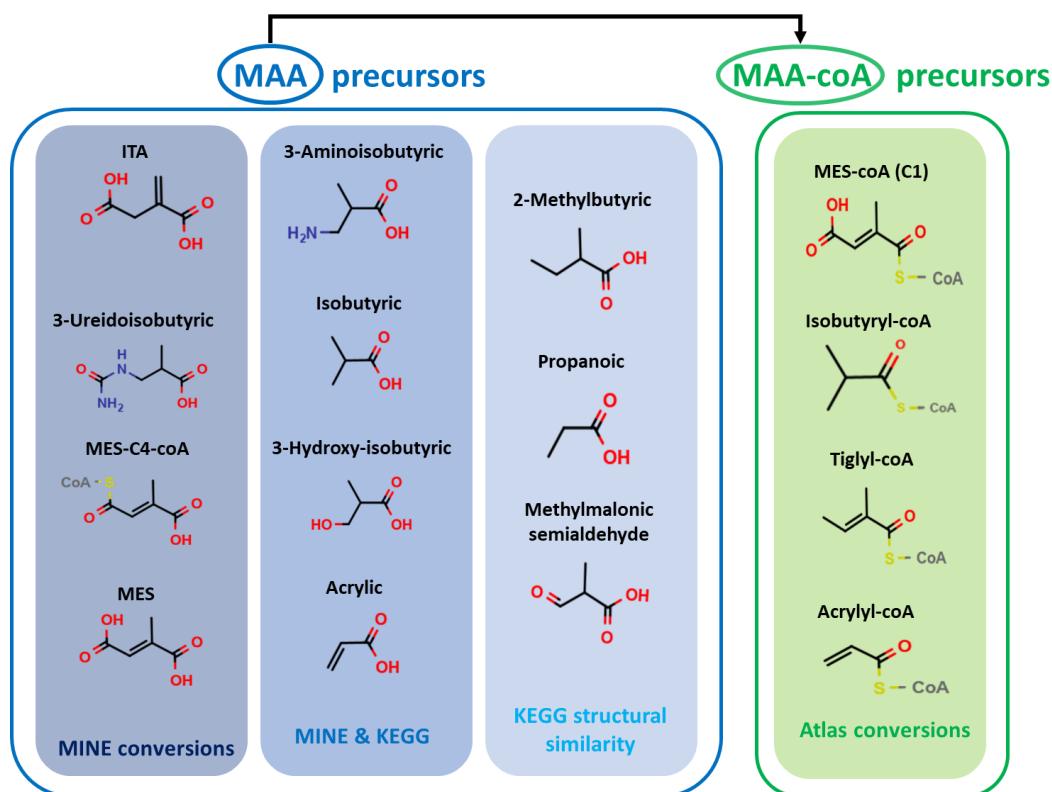


Figure 5.2. List of possible MAA, Methacrylyl-CoA and MMA precursors selected in this work. The Methacrylyl-CoA and MMA conversions were searched in the Atlas of Biochemistry database because these are KEGG metabolites. Since MAA is not present in the KEGG database, its corresponding conversions were searched in the MINE database and also structurally similar molecules were searched in KEGG.

5.4.1.2. Enzyme assignment

After the search for substrates, 15 molecules were considered as candidates for the implementation of a MAA biosynthetic production pathway, 11 being MAA precursors and four being methacrylyl-CoA precursors (as shown in Figure 5.2). In order to make an assignment of enzymes (or enzyme classes) that could mediate the envisaged conversions, the Transform-Miner web-server was used, a tool that identifies the possible reaction paths between a source and target molecule and also the transformations described in KEGG with the highest similarity to the input provided by the user [32]. One key feature of Transform-Miner is the methodology used to quantify the similarity between query and native reactions, which enables a prioritization of the search for paths that move quickly

toward the target, that is, shortest paths (and most efficient ones) are preferentially selected by this tool. Considering this specificity, the first criteria to choose between the different paths that Transform-Miner proposes between a source and a target molecule was the shortest path. After this, the results were ranked based on the cumulative similarity scores registered for the entire pathway. From the 15 possible MAA precursors, 1 route was identified (Table 5.7 and Table 5.8). These routes are described below and the ones that were considered likely to lead to a successful implementation *in vivo* are represented in Figure 5.3A. The main consideration for the successful implementation of each pathway was the availability of enzymes that can catalyze the required reactions (enzymes assigned to synthetic reaction steps are described in Figure 5.3B).

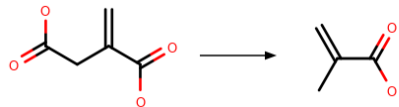
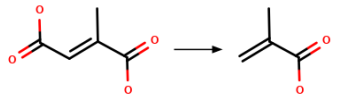
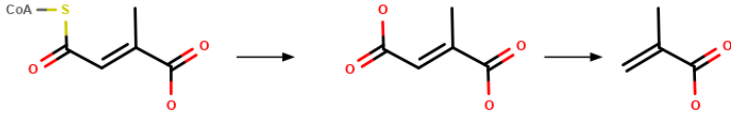
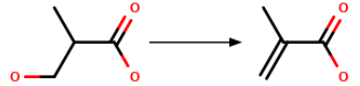
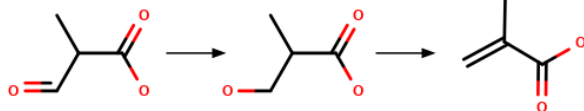
In what concerns the transformation of ITA into MAA, the decarboxylation was considered the shortest path with the most similar enzyme-catalyzed decarboxylations in KEGG being the ones of aspartate (4.1.1.12) and malate (4.1.1.101). However, while the α carbon in aspartate and malate is bound to an amino or hydroxyl group, in ITA the α carbon is involved in a double bond. Since the mechanisms underlying a decarboxylation usually involve the formation of an enolate that requires an alpha hydrogen, which is not present in ITA, it is unlikely that this molecule can be decarboxylated through this route. In what concerns mesaconic acid, Transform-Miner predicts it to be directly decarboxylated to MAA with two candidate enzymes: cis-aconitate or trans-aconitate decarboxylases (4.1.1.6 and 4.1.1.113). Cis-aconitate decarboxylase is the *A. terreus* enzyme that is being explored in this thesis for the production of ITA [34], [32], while trans-aconitate decarboxylase is involved in the same biochemical pathway but in *U. maydis* [37]. Considering the similarity between the native substrates (cis and trans-aconitic) to the synthetic one (mesaconic acid), this route was considered as a promising candidate (route 1 in Figure 5.3A, substrates compared in panel B). In what concerns the use of mesaconyl-C4-coA for MAA synthesis, the shortest route was based the removal of CoA to yield mesaconic acid, which is decarboxylated to MAA. Since mesaconyl-C4-coA can be derived from a different pathway than mesaconic acid (glyoxylate cycle in prototrophic bacteria vs glutamate degradation in *Clostridium sp*), this was considered as an alternative MES-based MAA route (route 2 in Figure 5.3). Although 29 KEGG reactions were identified as being similar to the herein envisaged coA removal from mesaconyl-C4-coA, the top 5 hits provided by Transform-Miner corresponded actually to reactions in which the CoA moiety is transferred to mesaconyl-C4 which is the reverse reaction. The most similar KEGG reaction identified by the tool in which CoA thioester group is removed from a substrate similar to mesaconyl-C4-coA is the reaction catalyzed

by the enzyme succinyl-CoA hydrolase (3.1.2.3). Mining of the BRENDA database to find enzymes within this class that may have substrates similar to mesaconyl-C1-coA identified the human and rat succinate-hydroxymethylglutarate CoA-transferase (2.8.3.13), which also accepts 3-hydroxyglutaryl-coA and methylmalonyl-coA, two substrates with some similarity with mesaconyl-C4-CoA (the native and synthetic substrates are represented in Figure 5.3B). Concerning the use of 3-hydroxyisobutyric acid it was found that it can be directly dehydrated to MAA and a native similar KEGG reaction was found: the dehydration of (S)-2-Hydroxymethylglutarate to 2-Methyleneglutarate, two substrates highly similar to 3-hydroxyisobutyric acid, specially because 3-hydroxyisobutyric is a substructure of (S)-2-Hydroxymethylglutarate (also represented in Figure 5.3B). However, there is no specific enzyme associated to this EC class in the databases, although it is known to participate in the anaerobic degradation of nicotinate in *Eubacterium barkeri* and is likely to be catalyzed by the gene product *hmd*, based on the predicted protein function and sequence similarity to bacterial serine dehydrogenases [38]. This route is represented as pathway 3 in Figure 5.3. The shortest MAA conversion route from methylmalonic semialdehyde involved reduction of the later to 3-hydroxyisobutyric acid, which is then converted to MAA. Since methylmalonic semialdehyde is derived from 3-hydroxyisobutyric acid in the valine degradation pathway, a route starting from this precursor would be circular and for this reason, it was discarded.

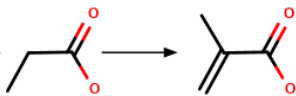
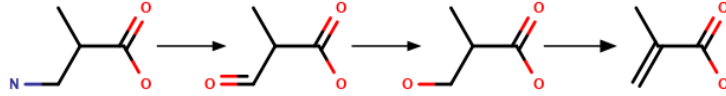
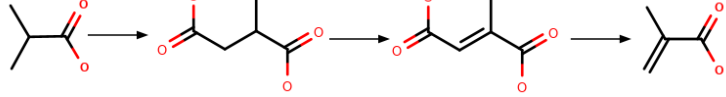
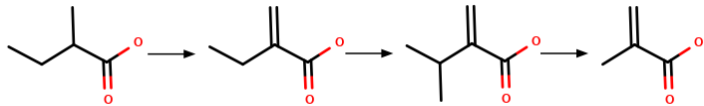
As for propanoic acid, the shortest path to convert it to MAA involved the direct addition of a double-bonded carbon. The most similar reactions found in the Transport-Miner tool were the coA transfer to acrylic (associated to glutaconyl-coA transferase 2.8.3.12, where acetyl-coA is the donor) and to ITA (associated to succinyl-coA: itaconate transferase 2.8.3.7, where succinyl-coA is the donor). Since both reactions are of coA transfer and not of methyl transfer, this prediction is an artifact of the search methodology likely due to the existence of two co-substrates in these reactions, which leads to the erroneous classification of these reactions as double-bonded carbon additions, instead of coA transfers. In consequence, the route from propionic to MAA was discarded. The shortest paths to produce MAA from ureidoisobutyric, 3-aminoisobutyric, isobutyric, acrylic and 2-methylbutyric acids to MAA all involved a final step in which propanoic acid is converted to MAA. This demonstrates that the Transform-Miner search methodology that prioritizes more efficient paths (with faster transformations in the direction of the target molecule), while useful to make the search more efficient, can also introduce a preference towards paths that use a particular conversion that will not be

successful *in vivo*, such as the case found here of propanoic acid conversion to MAA. This preference may prevent finding other routes that are more likely to be successful *in vivo*. Thus, the manual analysis of each route and potential enzymes to catalyze it is required. Based on what was said above concerning the impossibility of that final conversion, these shortest routes were not considered and other options not resorting to the propanoic acid-MAA conversion were searched for. A suitable alternative was only found for 3-aminoisobutyric since in the other cases the alternatives were considered very difficult to implement (e.g., by involving CO₂ carboxylation or using non-natural intermediates). The alternative route found to use 3-aminoisobutyric acid for MAA production involved a deamination reaction, followed by reduction by native enzymes in the valine degradation pathway (3-amino-2-methylpropanoate: pyruvate aminotransferase, 2.6.1.4, and 3-hydroxy-2-methylpropanoate: NAD⁺ oxidoreductase, 1.1.1.31) to form 3-hydroxyisobutyric acid, which is then dehydrated to MAA (as described before). In this alternative route, 3-hydroxyisobutyric acid may be produced from valine but also from thymine to be finally converted to MAA (route 4 in Figure 5.3). However, the direct deamination of 3-aminoisobutyric acid to MAA is also an easy possibility that was not pinpointed by Transform-Miner. Mining of the 4.3.1.- reactions in BRENDA resulted in the selection of six ammonia lyases where the product is unsaturated instead of an aldehyde. Of these, only aspartate and methylaspartate ammonia lyase (4.3.1.1 and 2) utilize similar substrates and it is relevant that aspartate ammonia lyase belongs to the fumarase family [39], making it a good candidate for removal of the amino group in 3-aminoisobutyric acid concomitant with the dehydration of the bond, generating MAA. Additionally, it is interesting that while the main substrate of phenylalanine ammonia lyase (4.3.1.24) is aromatic, it has a broad substrate range, including aliphatic molecules such as 3-methoxy-alanine. The native substrates of aspartate and phenylalanine ammonia-lyase are compared to 3-aminoisobutyric acid in Figure 5.3B. Thus, the direct deamination of 3-aminoisobutyric acid to MAA is here considered as a possible route (route 5 in Figure 5.3).

Table 5.7. Precursor to MAA conversions in TRANSFORM-MINER with maximal 3 steps and minimal similarity score of 0.6. Only one route was selected for each precursor, being the first criteria the shortest path and the second criteria the sum of the similarity score of all the pathway reactions

Precursors	Conversions	Most similar reactions and associated enzymes	Similarity score	Passed the filter?
ITA		L-aspartate 4-carboxy-lyase (L-alanine-forming) (4.1.1.12) <hr/> (S)-malate carboxy-lyase (4.1.1.101)	0.64 <hr/> 0.64	No (native reactions not similar enough)
Mesaconic acid		Cis-aconitate decarboxylase (4.1.1.6) <hr/> Trans-aconitate decarboxylase (4.1.1.113)	0.75 <hr/> 0.75	Yes
Mesaconyl-C4-coA		Succinyl-coA hydrolase (3.1.2.3) ** <hr/> The decarboxylation of mesaconic acid to MAA is described above	0.77 <hr/>	Yes
3-Hydroxyisobutyric acid		(S)-2-Hydroxymethylglutarate <=> 2-Methyleneglutarate + H2O (no associated E.C. but likely catalyzed by <i>E. barkeri</i> gene product <i>hmd</i>) *	0.64	Yes
Methylmalonic semialdehyde		To convert methylmalonic semialdehyde (which is derived from 3-hydroxyisobutyric) again to 3-hydroxyisobutyric would be futile.		No (pathways with futile cycle)

Chapter 5– Prospecting of metabolic pathways for implementation of microbe-based production of Methacrylic Acid (MAA)

Propanoic acid		Acetyl-CoA + Acrylic acid <=> Acetate + Propenoyl-CoA (associated to glutacetyl-coA transferase 2.8.3.12)	0.8	No (native reactions not similar enough)
		Succinyl-CoA:itaconate CoA-transferase (2.8.3.7)	0.63	
3-Aminoisobutyric acid		3-Amino-2-methylpropanoate:pyruvate aminotransferase (2.6.1.4)		Yes
		3-Hydroxy-2-methylpropanoate:NAD+ oxidoreductase (1.1.1.31)		
		The conversion of 3-hydroxyisobutyric acid to MAA is described above.		
3-Ureidoisobutyric acid	None (All routes pass through propanoic acid)			
Acrylic acid				
Isobutyric acid		The first reaction step (carboxylation of isobutyric acid to methylsuccinic acid) is not viable since it requires carbon dioxide fixation		No
2-Methylbutyric acid		Cyclohexanone:(acceptor) 2-oxidoreductase (1.3.99.14)	0.7	No (native reactions not similar enough)
		L-alanine:2-oxobutanoate aminotransferase(2.6.1.44)	0.63	
		L-Valine:pyruvate aminotransferase (2.6.1.66)	0.6	

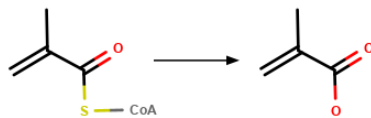
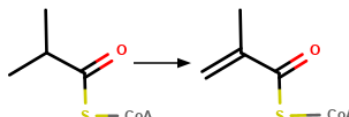
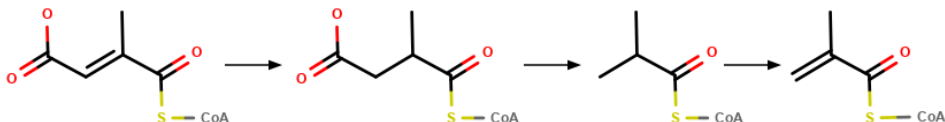
* When the most similar reaction is not the main reaction catalyzed by the associated enzyme, the reaction is depicted in “substrate → product” form. Otherwise, only the name and EC number of the reaction is represented.

** This enzyme was not considered the most suitable one and further searches in BRENDA for enzymes catalyzing the same reaction unraveled a better suited enzyme candidate: succinate-hydroxymethylglutarate CoA-transferase (2.8.3.13)

On a second approach, the enzymes that could lead to the formation of methacrylyl-CoA from the four precursors identified before were searched in the Transform-Miner tool (Table 5.8). The only path found for mesaconyl-C1-CoA conversion to methacrylyl-CoA entails the unsaturation of C2-C3 bond to methylsuccinyl-coA, with the most similar KEGG reaction being the reverse reaction, catalyzed by (2S)-methylsuccinyl-CoA dehydrogenase. This enzyme is not annotated as reversible, which might hamper the use of this enzyme for the desired reaction. The possibility of directly decarboxylating mesaconyl-C1-coA to methacrylyl-CoA was not predicted by Transform-Miner, indicating that there are no KEGG reactions similar to that conversion. Indeed, the most similar reactions found correspond to the reactions catalyzed by cis and trans-aconitate decarboxylases, with a similarity score of only 0.38. Thus, it is unlikely that mesaconyl-C1-coA can be decarboxylated to methacrylyl-CoA. Overall, the results obtained with the Transform Miner could lead only to the proposal of a pathway for production of methacrylyl-CoA that had been already established. Another possibility concerned the removal of the CoA group that would automatically result in MAA. Upon search on Tranform-Miner, 36 KEGG reactions showing some similarity (similarity score above or equal to 0.6) to this envisaged removal of the CoA group from methacrylyl-CoA were identified. Of these, only 13 reactions were in the direction of removal of CoA thioester (the others corresponded to the opposite) with the top matches being 3-hydroxyisobutyryl-CoA hydrolase (3.1.2.4) and acetoacetyl-CoA hydrolase (3.1.2.11). The first enzyme has a broader substrate specificity since it also accepts malyl-coA and aromatic substrates and its native substrate is branched (like methacrylyl-CoA) rendering this enzyme an interesting candidate to mediate the conversion of methacrylyl-CoA into MAA (Figure 5.3B). Thus a 6th route was added to the list (route 6 in Figure 5.3).

In the end of the pathway prospecting work, it was concluded that from a set of 15 potential precursors it was possible to find six suitable conversion routes leading to MAA from five precursors: mesaconic acid, mesaconyl-C4-CoA, 3-hydroxyisobutyric acid, 3-aminoisobutyric acid and isobutyryl-CoA (Figure 5.3). With the methodology used in this work, while it was possible to select new precursors, no new conversion route to MAA or methacrylyl-CoA could be found since the results found here were identical to the ones reported in the patents filled by Genomatica [15], [16] and Lucite [17], mentioned in the introduction of this chapter. From the six conversion routes in this work to which enzymes could be assigned, route 1 was selected for experimental validation of mesaconic acid decarboxylation. The main motivation for the selection of this route was the fact that mesaconic acid production from glucose has been successfully established in *E. coli*, through the expression of glutamate mutase and 3-methylaspartate ammonia lyase from *C. tetanomorphum* [28].

Table 5.8. Precursor to Methacrylyl-CoA conversions and Methacrylyl-CoA to MAA conversion in TRANSFOR-MINER with maximal 3 steps and minimal similarity score of 0.6. Only one route was selected for each precursor, being the first criteria the shortest path and the second criteria the sum of the similarity score of all the pathway reactions.

Precursors	Conversions	Most similar reactions and associated enzymes	Similarity score	Passed the filter?
Methacrylyl-CoA→ MAA		3-Hydroxyisobutyryl-CoA hydrolase (3.1.2.4)	0.75	Yes
		Acetoacetyl-CoA hydrolase (3.1.2.11).	0.75	
Isobutyryl-coA		2-methylpropanoyl-CoA:(acceptor) 2,3-oxidoreductase (associated to 1.3.8.1/7 and 1.3.99.12)	1	Yes
Mesaconyl-C1-coA		(2S)-methylsuccinyl-CoA dehydrogenase (1.3.8.12)	1	No (native reactions are not similar enough)
		propionyl-CoA:succinate CoA transferase (not associated to specific E.C. number)	0.94	
		2-methylpropanoyl-CoA:(acceptor) 2,3-oxidoreductase (associated to 1.3.8.1/7 and 1.3.99.12)	1	
Tiglyl-coA	No path found			
Acrylyl-coA	No path found			

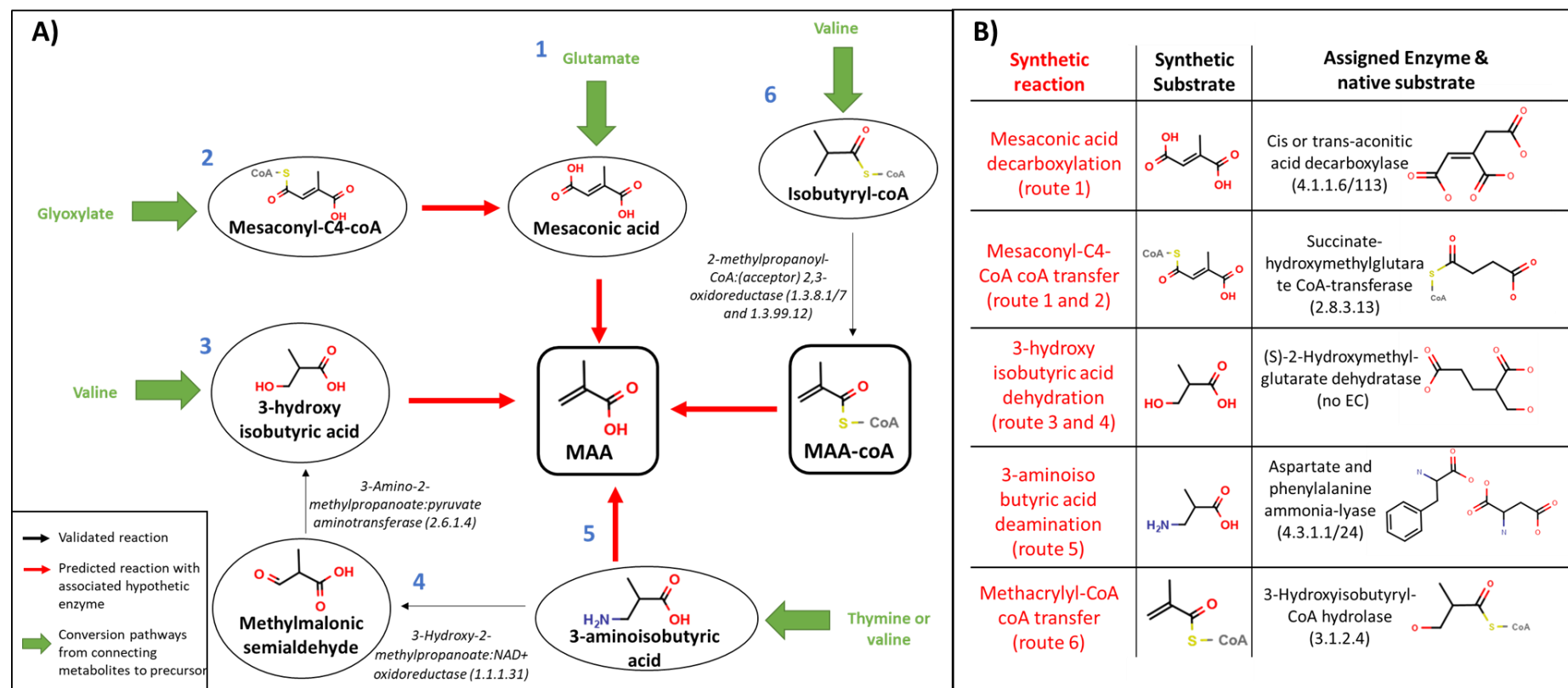


Figure 5.3. MAA routes and precursors studied in this work. A) The six MAA routes found in this *in silico* work. The reactions depicted in black have experimental validation and reactions in red represent hypothetical enzymatic steps. The common microbial metabolites that are the source of each route are represented in green, which were valine, glutamate, glyoxylate and thymine. B) For each hypothetical enzymatic step (represented in red in panel A) it is described the enzymes assigned to it, the substrate for the synthetic reaction and the native substrate.

5.4.2. Exploration of mesaconic acid-decarboxylation as a putative route for synthesis of MAA

In the previous subsection six MAA forming routes were identified among which was the decarboxylation of mesaconic acid prompted by decarboxylases. In this section it was decided to investigate this further since this is a simple pathway (only 1 step) and at least two of the enzymes described as candidates to mediate the decarboxylation of mesaconic acid (AtCadA or UmTad1) have been the topic of work of the laboratory. The ability of AtCadA or UmTad1 to accept the two isomers of aconitic acid (respectively cis- and trans- aconitic acid) is likely to result from divergent evolutive pathways, since the predicted fold of TAD is very different from the structure of CAD and is more similar to bacterial fumarate lyase (42% sequence identity) [40]. It is also believed that the mechanism of catalysis might be different in the decarboxylations mediated by the two enzymes [40]. Although aconitic acid is a tricarboxylic acid and mesaconic acid is dicarboxylic, there is a similarity between the decarboxylation of both substrates (reaction represented in Figure 5.4A and B). This similarity is more evident when considering that it was determined early-on that CAD catalyzed the removal of the primary C5 carboxyl group, and not the secondary C6 group [41].

In order to find other decarboxylases that could also mediate the decarboxylation of mesaconic acid, all decarboxylation reactions in BRENDA were manually mined and three potential enzyme candidates were selected (the reactions catalyzed by the three new enzyme candidates, along with TAD and CAD are represented in Figure 5.4):

i) aminocarboxymuconate-semialdehyde decarboxylase (ACMSD, 4.1.1.45); an enzyme that participates in the kynurenine pathway. Although the native substrate of this enzyme aminocarboxymuconate-semialdehyde (ACMS) is larger than mesaconic acid, this is almost a substructure of ACMS with both being dicarboxylic acids having a double bond in a α - β position that can be removed in the decarboxylation reaction (Figure 5.4C). The main differences between the two molecules are the presence of a 3-carbon side chain in ACMS and an amino group adjacent to the unsaturated bond, instead of the methyl group in MES. The different electronegativities of these two groups may be relevant for the enzymatic mechanism and to the substrate interaction with the enzyme's binding pocket. The structure of ACMS decarboxylase has been elucidated and its inclusion in the amidohydrolase superfamily points to a mechanism where the addition of a zinc-bound hydroxy group to a carbon to create an tetrahedral intermediate is likely key to the reaction mechanism [42]. Considering the two carboxyl groups in the molecule and the

structure of the enzyme's active site, two mechanisms have been proposed for the reaction where in the first one the C2 carboxyl group binds next to the metal center and in the second mechanism it is the C3 leaving carboxyl group that binds next to the enzyme's metal center. In this second mechanism the presence of the amino group is important to create an imine bond and stabilize the tetrahedral intermediate. Thus, if the second mechanism is the one through which ACMS decarboxylation occurs, it is likely that mesaconic acid cannot be accepted as a substrate by the enzyme.

ii) 2-oxo-3-hexenedioic acid decarboxylase (4.1.1.77), a bacterial enzyme that participates in the catechol meta fission degradation pathway and that has been mostly studied in *P. putida* [43]. The native substrate of this enzyme is 2-oxo-3-hexenedioic acid, also an unsaturated dicarboxylic acid but that is α -keto acid (Figure 5.4D), which can be problematic since this group is proposed to form a complex with Mg^{2+} to act as an electron sink [44]. This enzyme occurs *in vivo* in a complex with vinylpyruvate hydratase, which would complicate the use of 2-oxo-3-hexenedioic acid decarboxylase in a pathway to produce MAA. Despite this, the structural similarity of the substrates may enable the enzyme to catalyze mesaconic acid decarboxylation, as it also been suggested before [45].

iii) glutaconyl-CoA decarboxylase (4.1.1.70), a bacterial biotin-dependent Na^+ pump that uses the energy derived from substrate decarboxylation to drive transport of Na^+ from the cytoplasm to the periplasm [46]. Like mesaconic acid, glutaconyl-coA has an unsaturated bond that is α - β to the carboxyl group (Figure 5.4E), however, this enzyme would likely be a better candidate for Mesoconyl-C4-coA decarboxylation, since it accepts CoA substrates. Furthermore, the enzyme is also able to accept glutaryl-coA, a saturated dicarboxylic acid, as a substrate.

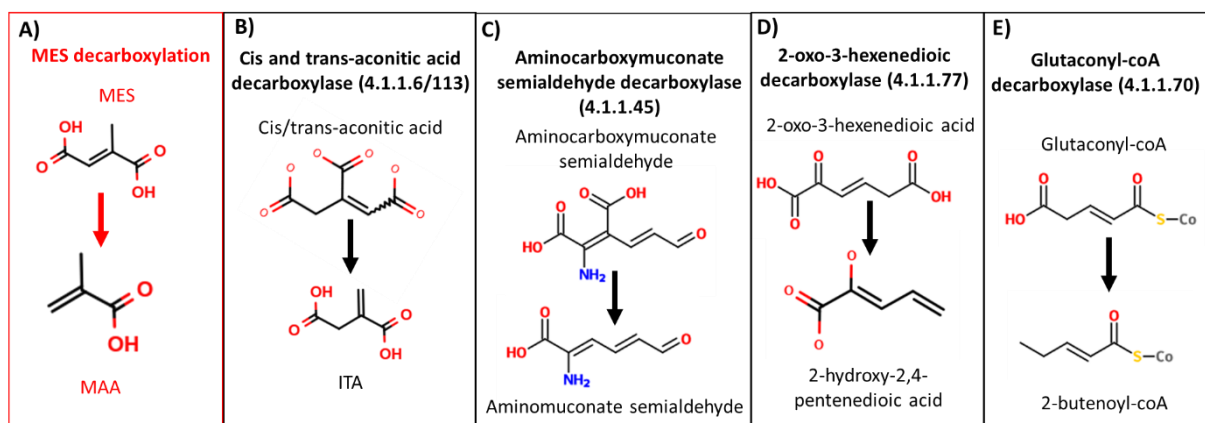


Figure 5.4. The MAA route studied in this work, the decarboxylation of MES(A), for which five enzymes were selected as potential candidates to catalyze the reaction (C-E). In the case of glutaconyl-coA decarboxylase (E), glutaryl-coA is also accepted as a substrate.

Considering the difficulties associated with the successful expression of 2-oxo-3-hexenedioic acid decarboxylase that requires a complex with vinylpyruvate hydratase, and the fact the glutaconyl-coA decarboxylase is only described to accept coA substrates, these two enzymes were not considered for further work. Thus, cis-aconitate decarboxylase and ACMSD were considered for further experimental validation of the decarboxylation of mesaconic acid.

5.4.2.1.Exploring AtCadA as a possible decarboxylase of mesaconic acid

To test the ability of AtCadA to accept mesaconic acid as a substrate, an *S. cerevisiae*-based expression system for *A. terreus* CAD1 that was already available in the laboratory was used. This is based on the use of the pGREG_AtCad1 plasmid which has the AtCAD1 gene cloned under the control of the galactose inducible promoter *GAL1*. This system has been successfully used at BSRG to produce ITA in *S. cerevisiae* strains [20], [47]–[51], however, until thus far it had not been performed its use to perform in vitro enzymatic assays. As such, crude cell extracts were prepared from exponential *S. cerevisiae* BY4741 cells carrying pGREG_AtCad1 and harvested after 1h of incubation in the presence of galactose. As a control, cells carrying only the cloning vector (pGREG586) were used. A first attempt to test to CAD activity in the extracts was made by performing enzymatic assays with 8 mM of cis-aconitic acid, the native substrate. However, no CAD activity could be detected, since no ITA was formed or cis-aconitic acid was consumed. Further modifications in the cell growth protocol and enzymatic assay setting were performed. The optimal setting was found to be to pre-culture By4741 cells carrying the pGREG_AtCad1 plasmid in MMF raffinose, induce in MMF galactose and harvest 150 mL of cell culture after 24 hours of induction, followed by carrying out enzymatic assays with 30 mM of cis-aconitic acid. With this setting it was possible to obtain a CAD activity of 73.61 μ M ITA/mg protein (HPLC traces represented in Figure 5.5).

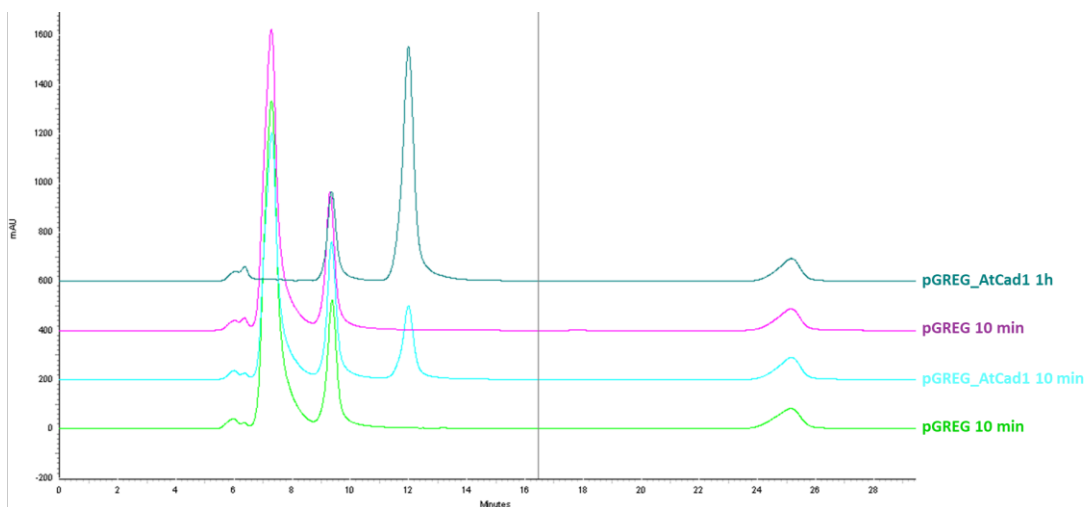


Figure 5.5. Stacked HPLC traces of CAD activity assay supernatants after 10 min and 1 h of reaction with 30 mM cis-aconitic. The traces are shown for activity assays with extracts from cells carrying the empty plasmid or pGREG_AtCad1. ITA elutes at ~12 min. Aconitic acid elutes at ~7 and 9~min, corresponding to the trans and cis isomers, respectively. This is due to the spontaneous isomerization of cis-aconitic to trans in dilute aqueous solutions has been described before [52].

After confirmation of *in vitro* activity of the expressed AtCadA, the assays were repeated using mesaconic acid as a substrate (Figure 5.6). Under the conditions that were tested (that included to increase reaction time for a maximum of 6h, use of 40 mM substrate or change in the pH of the reaction) it was not possible to observe the presence of MAA in the chromatograms. While this may indicate that AtCadA cannot accept mesaconic acid as a substrate, it is also possibility that an expected low affinity for the synthetic substrate may result in low MAA formation rates, with final MAA concentrations under the limit of detection for the HPLC UV-VIS detector used here. Thus, it would be desirable to analyze the same supernatants using a more sensible technique, such as LC-MS.

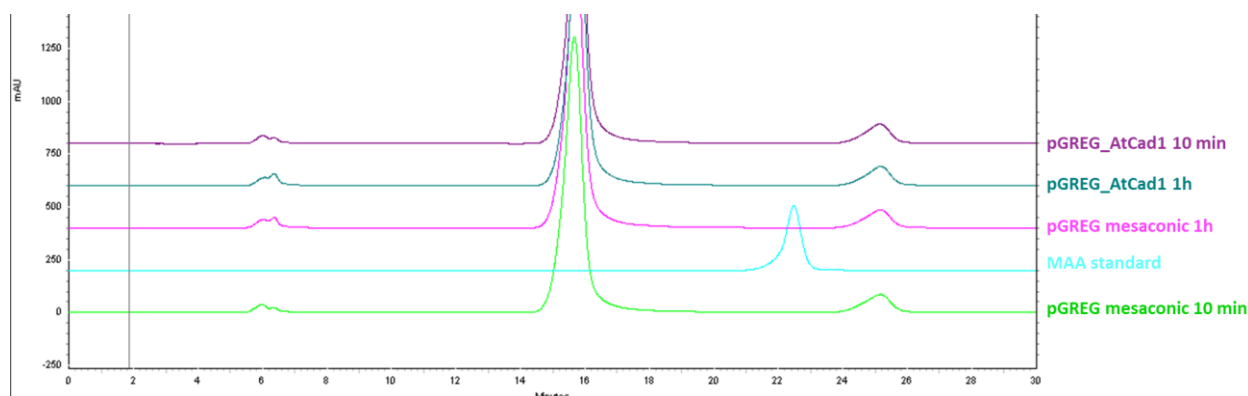


Figure 5.6. Overlaid HPLC traces of CAD activity assay supernatants after 10 min and 1 h of reaction. The traces are shown for activity assays with extracts from cells carrying the empty plasmid or pGREG_AtCad1. MAA elutes at ~ 22 min (the MAA standard trace is represented) and mesaconic acid elutes at ~16 min.

5.4.2.2. Exploring ACMSD as a possible decarboxylase of mesaconic acid

Besides CAD, the enzyme ACMSD was also selected as a potential candidate to catalyze mesaconic acid decarboxylation to MAA, due to the similarity between the native substrate of this enzyme (ACMS) and mesaconic acid (Figure 5.4). To experimentally test the enzyme ability to decarboxylate MES, the ACMSD from *Pseudomonas fluorescens* was studied making use of the IPTG-inducible pET16b-nbaD (kindly provided by Prof. Dr. Aimin Liu), a system that has been successfully used to over-express this protein in *E. coli* cells [52]. Analysis by SDS-PAGE of whole *E. coli* cell extracts, confirmed a strong up-regulation of ACMSD expression after cultivation at 28°C of the cells in LB supplemented with 1 mM IPTG (Figure 5.7).

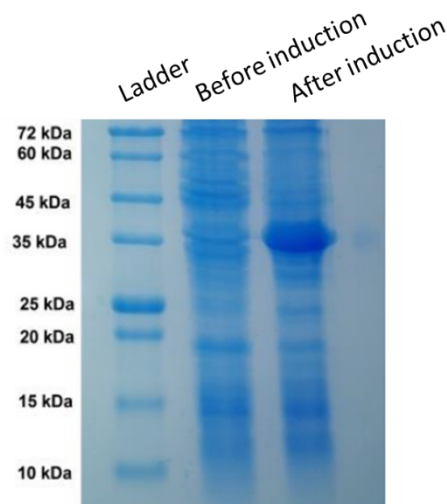


Figure 5.7. SDS-PAGE of BL21(DE3) cells carrying the pET16b-nbaD plasmid before and after 4h of induction with 1 mM IPTG and expression at 28°C.

Enzymatic assays were performed with protein extracts recovered from *E. coli* cells cultivated under the conditions described above using 6 mM MES. In this case it was not possible to make a control reaction to assess whether or not the enzyme was active since ACMS is not commercially available. HPLC analysis of the supernatants obtained after the enzymatic reactions could not reveal the presence of MAA in the supernatant, however, a new peak (at ~9 min) was identified only in samples where mesaconic acid was added (but not in controls that were performed without the addition of mesaconic acid) (Figure 5.8A and B). In line with this, a slight consumption of mesaconic acid was observed in the samples where we detect the formation of this additional peak (Figure 5.8C). Since this peak does not have the MAA retention time, it is possible that the decarboxylation has taken place at the C1 group, forming 2-butenic acid, instead of the C4 group (which would result in MAA formation). Another possibility is the formation of the mesaconic hydration product, 2-hydroxy-2-methylsuccinic acid. This is possible since ACMSD belongs to the “ACMSD-related protein family” which includes proteins with hydratase activity and the mechanism ACMSD resembles the one of a hydration [53]. Despite the absence of MAA production with the crude cell extracts, strong conclusions could not be taken about the ability of ACMSD ability to catalyze the mesaconic acid decarboxylation due to the lack of a positive control with the native ACMSD substrate, to confirm that the enzyme is active in the conditions employed here.

Chapter 5– Prospecting of metabolic pathways for implementation of microbe-based production of Methacrylic Acid (MAA)

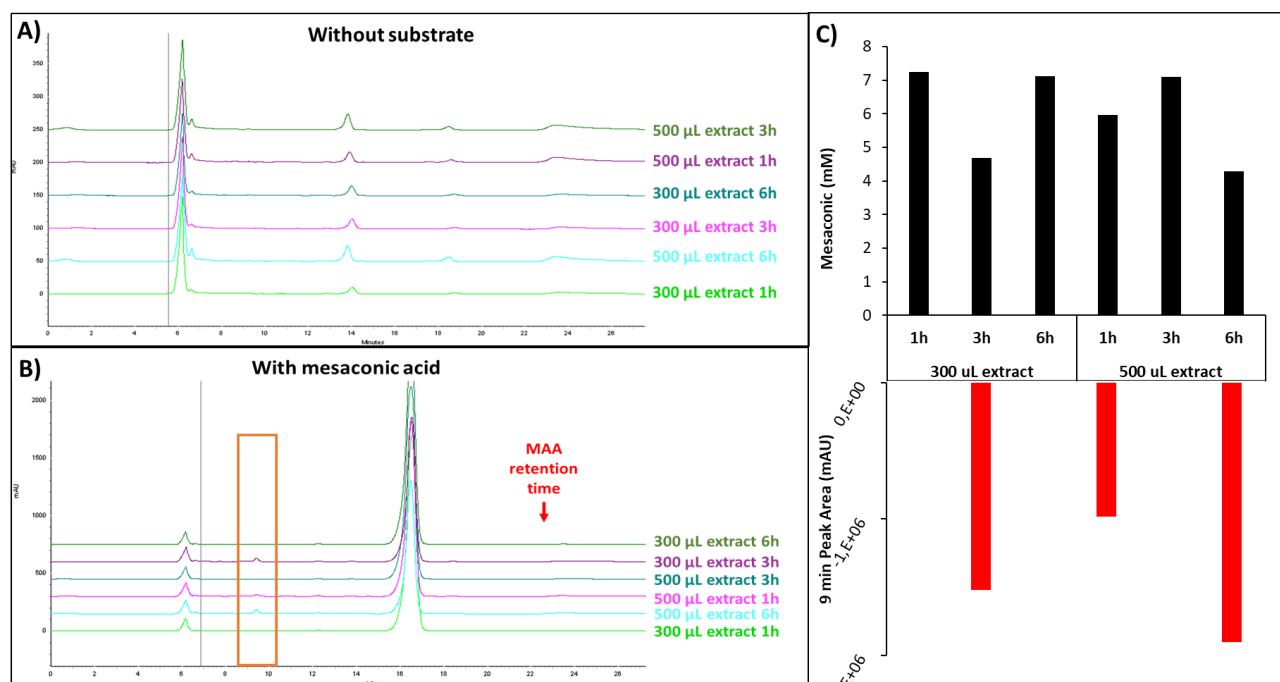


Figure 5.8. ACMSD activity assays. A and B) Overlaid HPLC traces of ACMSD activity assay supernatants after 1, 3 and 6 hours of reaction. Supernatants from activity assays with no substrate (A) or with 6 mM mesaconic acid (B). MAA elutes at ~22 min (red arrow) and mesaconic acid elutes at ~16 min. The unknown peak at ~9 min is highlighted in the orange box. C) Mesaconic acid concentration (mM) and area of the 9 min peak in ACMSD activity assay supernatants.

The results obtained from the CAD and ACMSD enzymatic assays suggest that these enzymes cannot accept mesaconic acid as substrate without further modification. However, CAD and ACMSD may be a starting point for future enzyme modification to enable an exchanged substrate specificity. The modification may be rationally motivated or may be performed by directed-enzyme evolution. In the last option, a high-throughput screening methodology either for mesaconic acid consumption or MAA production is required, which haven't been developed at this point. The first option, of rational modification of CAD or ACMSD to accept mesaconic acid as a substrate, is made easier by the availability of the corresponding enzyme structures. In the case of CAD, the structure of AtCadA has not been obtained directly; however, the human and murine CAD structures have been recently elucidated and enabled the superimposition of fungal CAD, shedding light on the enzyme's evolution and mechanism [53]. The structural similarity of CAD to prokaryotic 2-methylcitrate and iminidissuccinate epimerase, which have a tricarboxylic and tetracarboxylic substrates, respectively, suggests that CAD has evolved

from prokaryotic enzymes that accept carboxylic-rich substrates and that it has acquired the decarboxylase feature [40]. If this is the case, it may be required to engineer the binding pocket of CAD to accept a dicarboxylic substrate. Indeed, the suggested binding mode of cis-aconitic acid to the CAD binding pocket indicates that the C1 and C6 carboxylic groups of the substrate are key factors in the reaction. The orientation of cis-aconitic in the active site so that the C5 carboxyl group (the one being removed) is accommodated in hydrophobic binding pocket was suggested to be a driving force for the formation of the intermediate enolate, in addition to the formation of hydrogen bonds between the C1 and C6 carboxyl groups to the surrounding residues, which would create an extra polarization [40]. Considering the potential orientation of mesaconic acid in the same binding pocket, the exposure to the hydrophobic pocket could be hampered by the smaller length of the main chain of MES, causing the substrate to be too far away from this nonpolar region. Also, in this presumptive orientation mesaconic acid could present a methyl group instead of the C6 carboxyl group of cis-aconitic acid to the binding pocket, not being able to form the hydrogen bridges. Both these situations may be engineered so that steric hindrance is created in the binding pocket to enable mesaconic acid to come into close contact with the hydrophobic region and to exchange negatively charged carboxyl-interacting residues to ones that are more amenable to interaction with the methyl group [54]. Finally, the rational engineering of both CAD and ACMSD to accept mesaconic acid as substrate could be facilitated by molecular docking simulations, pinpointed the key residues that may affect substrate recognition. Indeed, docking has been used in the past to motivate residue modification that broaden the substrate specificity of enzymes [55], [56].

5.5. Conclusions and Future Work

In this chapter both *in silico* and *in vitro* prospecting methodologies were pursued in order to implement microbe-based production of metacrylic acid. With the *in silico* methodologies six routes of MAA formation that had been previously hypothesized were explored here, with the addition of potential enzyme candidates to catalyze predicted conversions in these routes. Also *in silico*, seven additional precursors were studied in their potential to be converted to MAA; however, it was not possible to assign enzymes likely to catalyze the required conversions, suggesting that these are unlikely to be mediated by a natural enzyme. *In vitro* methodologies were applied to test one MAA route in which the key step is mesaconic acid decarboxylation, in this case predicted to be

Chapter 5– Prospecting of metabolic pathways for implementation of microbe-based production of Methacrylic Acid (MAA)

catalyze by CAD or ACMSD. While the results obtained here suggest that the WT enzyme can't decarboxylate MES, they open the door for rational engineering or directed evolution of these enzymes to accept the mesaconic acid substrate analog. In particular, the recent publication of the crystal structure of CAD may be of use to motivate rational engineering of the substrate binding pocket.

5.6. References

- [1] A. Reports, “Global Methacrylic Acid Market,” 2019.
- [2] J. Lebeau, J. P. Efromson, and M. D. Lynch, “A Review of the Biotechnological Production of Methacrylic Acid,” *Front. Bioeng. Biotechnol.*, vol. 8, no. March, pp. 1–10, 2020, doi: 10.3389/fbioe.2020.00207.
- [3] S. L. Tomić, E. H. Suljovrujić, and J. M. Filipović, “Biocompatible and bioadhesive hydrogels based on 2-hydroxyethyl methacrylate, monofunctional poly(alkylene glycol)s and itaconic acid,” *Polym. Bull.*, vol. 57, no. 5, pp. 691–702, Jun. 2006, doi: 10.1007/s00289-006-0606-3.
- [4] S. L. Tomić, M. M. Mićić, S. N. Dobić, J. M. Filipović, and E. H. Suljovrujić, “Smart poly(2-hydroxyethyl methacrylate/itaconic acid) hydrogels for biomedical application,” *Radiat. Phys. Chem.*, vol. 79, no. 5, pp. 643–649, May 2010, doi: 10.1016/j.radphyschem.2009.11.015.
- [5] C. Bell and N. A. Peppas, “Poly(methacrylic acid-g-ethylene glycol) hydrogels as pH responsive biomedical materials,” in *Materials Research Society Symposium Proceedings*, 1994, vol. 331, pp. 199–204, doi: 10.1557/proc-331-199.
- [6] M. J. Darabi Mahboub, J. L. Dubois, F. Cavani, M. Rostamizadeh, and G. S. Patience, “Catalysis for the synthesis of methacrylic acid and methyl methacrylate,” *Chem. Soc. Rev.*, vol. 47, no. 20, pp. 7703–7738, 2018, doi: 10.1039/c8cs00117k.
- [7] A. H. Tullo, “In with the new,” *Chem Engineering news*, pp. 22–23, 2009.
- [8] X. L. Wang, R. P. Tooze, K. Whiston, and G. R. Eastham, “Process for the carbonylation of ethylene and catalyst systems for use therein,” US6348621, 2002.
- [9] D. Johnson, G. Eastham, and M. Poliakoff, “Method of producing acrylic and methacrylic acid,” WO2011/077140A2, 2011.
- [10] D. W. Johnson, G. R. Eastham, M. Poliakoff, and T. A. Huddle, “A process for the production of methacrylic acid and its derivatives and polymers produced therefrom,” EP2643283A1, 18-Nov-2011.
- [11] J. Le Nôtre, S. C. M. Witte-van Dijk, J. van Haveren, E. L. Scott, and J. P. M. Sanders, “Synthesis of Bio-Based Methacrylic Acid by Decarboxylation of Itaconic Acid and Citric Acid Catalyzed by Solid Transition-Metal Catalysts,” *ChemSusChem*, vol. 7, no. 9, pp. 2712–2720, 2014, doi: 10.1002/cssc.201402117.
- [12] A. Bohre, U. Novak, M. Grilc, and B. Likozar, “Synthesis of bio-based methacrylic acid from biomass-derived itaconic acid over barium hexa-aluminate catalyst by selective decarboxylation reaction,” *Mol. Catal.*, vol. 476, no. July, 2019, doi: 10.1016/j.mcat.2019.110520.
- [13] C. Press, “Bacteria like the taste of syngas,” 2013. [Online]. Available: <https://corporate.evonik.com/en/pages/article.aspx?articleId=103071>. [Accessed: 17-Feb-2020].

- [14] M. Carlsson *et al.*, “Study of the Sequential Conversion of Citric to Itaconic to Methacrylic Acid in Near-critical and Supercritical Water,” *Ind. Eng. Chem. Res.*, vol. 33, no. 8, pp. 1989–1996, 1994, doi: 10.1021/ie00032a014.
- [15] A. P. Burgard, M. J. Burk, R. E. Osterhout, and P. Pharkya, “Microorganisms for the production of methacrylic acid,” US8241877B2, 2008.
- [16] M. J. Burk, A. P. Burgard, R. E. Osterhout, J. Sun, and P. Pharkya, “Microorganisms for producing methacrylic acid and methacrylate esters and methods related thereto,” US9133487B2, 2015.
- [17] G. R. Eastham, G. Stephens, and A. Yiakoumetti, “Process for the biological production of methacrylic acid and derivatives thereof,” US 2018 / 0171368 A1, 2018.
- [18] J. D. Tyzack, A. J. M. Ribeiro, N. Borkakoti, and J. M. Thornton, “Exploring Chemical Biosynthetic Design Space with Transform-MinER,” *ACS Synth. Biol.*, vol. 8, pp. 2494–2506, 2019, doi: 10.1021/acssynbio.9b00105.
- [19] G. Jansen, C. Wu, B. Schade, D. Y. Thomas, and M. Whiteway, “Drag&Drop cloning in yeast,” *Gene*, vol. 344, pp. 43–51, Jan. 2005, doi: 10.1016/j.gene.2004.10.016.
- [20] N. M. Rodrigues, “Metabolic and genetic engineering strategies to explore *Saccharomyces cerevisiae* as a cell factory for the production of itaconic acid,” IST, 2014.
- [21] J. M. Friedman and C. S. Mantzoros, “20 years of leptin: From the discovery of the leptin gene to leptin in our therapeutic armamentarium,” *Metabolism*, vol. 64, no. 1, pp. 1–4, 2015, doi: 10.1016/j.metabol.2014.10.023.
- [22] J. G. Jeffryes *et al.*, “MINEs: open access databases of computationally predicted enzyme promiscuity products for untargeted metabolomics,” *J. Cheminform.*, vol. 7, no. 1, p. 44, Dec. 2015, doi: 10.1186/s13321-015-0087-1.
- [23] N. Wierckx, G. Agrimi, P. S. Lübeck, M. G. Steiger, N. P. Mira, and P. J. Punt, “Metabolic specialization in itaconic acid production: a tale of two fungi,” *Curr. Opin. Biotechnol.*, vol. 62, pp. 153–159, 2020, doi: 10.1016/j.copbio.2019.09.014.
- [24] K. Zhang, A. P. Woodruff, M. Xiong, J. Zhou, and Y. K. Dhande, “A synthetic metabolic pathway for production of the platform chemical isobutyric acid,” *ChemSusChem*, vol. 4, no. 8, pp. 1068–1070, 2011, doi: 10.1002/cssc.201100045.
- [25] H. S. Chu, J. Ahn, J. Yun, I. S. Choi, T. Nam, and K. Myung, “Direct fermentation route for the production of acrylic acid,” *Metab. Eng.*, vol. 32, pp. 23–29, 2015, doi: 10.1016/j.ymben.2015.08.005.
- [26] F. Sonntag, J. E. N. Müller, P. Kiefer, J. A. Vorholt, J. Schrader, and M. Buchhaupt, “High-level production of ethylmalonyl-CoA pathway-derived dicarboxylic acids by *Methylobacterium extorquens* under cobalt-deficient conditions and by polyhydroxybutyrate negative strains,” *Appl. Microbiol. Biotechnol.*, vol. 99, no. 8, pp. 3407–3419, Apr. 2015, doi: 10.1007/s00253-015-6418-3.
- [27] W. Buckel and H. A. Barker, “Two pathways of glutamate fermentation by

- anaerobic bacteria,” *J. Bacteriol.*, vol. 117, no. 3, pp. 1248–1260, Mar. 1974, doi: 10.1128/jb.117.3.1248-1260.1974.
- [28] J. Wang and K. Zhang, “Production of mesaconate in *Escherichia coli* by engineered glutamate mutase pathway,” *Metab. Eng.*, vol. 30, pp. 190–196, 2015, doi: 10.1016/j.ymben.2015.06.001.
- [29] W. Bai *et al.*, “Engineering nonphosphorylative metabolism to synthesize mesaconate from lignocellulosic sugars in *Escherichia coli*,” *Metab. Eng.*, vol. 38, no. September, pp. 285–292, 2016, doi: 10.1016/j.ymben.2016.09.007.
- [30] D. C. Ernst, J. A. Lambrecht, R. A. Schomer, and D. M. Downs, “Endogenous synthesis of 2-aminoacrylate contributes to cysteine sensitivity in *Salmonella enterica*,” *J. Bacteriol.*, vol. 196, no. 18, pp. 3335–3342, Sep. 2014, doi: 10.1128/JB.01960-14.
- [31] H. Luo, D. Zhou, X. Liu, Z. Nie, D. L. Quiroga-Sánchez, and Y. Chang, “Production of 3-Hydroxypropionic Acid via the Propionyl-CoA Pathway Using Recombinant *Escherichia coli* Strains,” *PLoS One*, vol. 11, no. 5, p. e0156286, May 2016, doi: 10.1371/journal.pone.0156286.
- [32] J. D. Tyzack, A. J. M. Ribeiro, N. Borkakoti, and J. M. Thornton, “Exploring Chemical Biosynthetic Design Space with Transform-MinER,” *ACS Synth. Biol.*, vol. 8, pp. 2494–2506, 2019, doi: 10.1021/acssynbio.9b00105.
- [33] J. C. Lansing, R. E. Murray, and B. R. Moser, “Biobased Methacrylic Acid via Selective Catalytic Decarboxylation of Itaconic Acid,” *ACS Sustain. Chem. Eng.*, vol. 5, no. 4, pp. 3132–3140, 2017, doi: 10.1021/acssuschemeng.6b02926.
- [34] A. Li, N. van Luijk, M. ter Beek, M. Caspers, P. Punt, and M. van der Werf, “A clone-based transcriptomics approach for the identification of genes relevant for itaconic acid production in *Aspergillus*,” *Fungal Genet. Biol.*, vol. 48, no. 6, pp. 602–611, Jun. 2011, doi: 10.1016/J.FGB.2011.01.013.
- [35] G. Tevž, M. Benčina, and M. Legiša, “Enhancing itaconic acid production by *Aspergillus terreus*,” *Appl. Microbiol. Biotechnol.*, vol. 87, no. 5, pp. 1657–1664, Aug. 2010, doi: 10.1007/s00253-010-2642-z.
- [36] C. L. Strelko *et al.*, “Itaconic acid is a mammalian metabolite induced during macrophage activation,” *J. Am. Chem. Soc.*, vol. 133, no. 41, pp. 16386–9, Oct. 2011, doi: 10.1021/ja2070889.
- [37] E. Geiser *et al.*, “*Ustilago maydis* produces itaconic acid via the unusual intermediate trans-aconitate,” *Microb. Biotechnol.*, vol. 9, no. 1, pp. 116–126, Jan. 2016, doi: 10.1111/1751-7915.12329.
- [38] A. Alhapel, D. J. Darley, N. Wagener, E. Eckel, N. Elsner, and A. J. Pierik, “Molecular and functional analysis of nicotinate catabolism in *Eubacterium barkeri*,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 33, pp. 12341–12346, Aug. 2006, doi: 10.1073/pnas.0601635103.
- [39] G. Fibriansah, V. P. Veetil, G. J. Poelarends, and A. M. W. H. Thunnissen, “Structural basis for the catalytic mechanism of aspartate ammonia lyase,”

- Biochemistry*, vol. 50, no. 27, pp. 6053–6062, Jul. 2011, doi: 10.1021/bi200497y.
- [40] F. Chen *et al.*, “Crystal structure of cis-aconitate decarboxylase reveals the impact of naturally occurring human mutations on itaconate synthesis,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 116, no. 41, pp. 20644–20654, 2019, doi: 10.1073/pnas.1908770116.
 - [41] R. Bentley and C. P. Thiessen, “Biosynthesis of itaconic acid in *Aspergillus terreus*: the properties and reaction mechanism of cis-aconitic decarboxylase,” *J. Mol. Biochem.*, vol. 226, pp. 703–720, 1956.
 - [42] § Dariusz Martynowski, § Yvonne Eyobo, || Tingfeng Li, § Kun Yang, *,|| and Aimin Liu, and § Hong Zhang*, “Crystal Structure of α -Amino- β -carboxymuconate- ϵ -semialdehyde Decarboxylase: Insight into the Active Site and Catalytic Mechanism of a Novel Decarboxylation Reaction^{†,‡},” 2006, doi: 10.1021/BI060903Q.
 - [43] T. M. Stanley, W. H. Johnson, E. A. Burks, C. P. Whitman, C. C. Hwang, and P. F. Cook, “Expression and stereochemical and isotope effect studies of active 4-oxalocrotonate decarboxylase,” *Biochemistry*, vol. 39, no. 4, pp. 718–726, 2000, doi: 10.1021/bi9918902.
 - [44] S. L. Guimarães, J. B. Coitinho, D. M. A. Costa, S. S. Araújo, C. P. Whitman, and R. A. P. Nagem, “Crystal Structures of Apo and Liganded 4-Oxalocrotonate Decarboxylase Uncover a Structural Basis for the Metal-Assisted Decarboxylation of a Vinylogous β -Keto Acid,” *Biochemistry*, vol. 55, no. 18, pp. 2632–2645, 2016, doi: 10.1021/acs.biochem.6b00050.
 - [45] A. P. Burgard, M. K. Burk, R. E. Osterhout, and P. Pharkya, “Microorganisms for the production of methacrylic acid,” US 8241877 B2, 14-Aug-2012.
 - [46] D. Reaction *et al.*, “Crystal Structure of R -Amino- -carboxymuconate- -semialdehyde Decarboxylase: Insight into the Active Site and Catalytic Mechanism of a Novel,” *Biochemistry*, no. Scheme 1, pp. 10412–10421, 2006.
 - [47] J. P. L. Santos, “Improvement of yeast-based production of itaconic acid guided by in silico metabolic modelling,” IST, 2016.
 - [48] N. M. S. Marques, “Suspension bioreactor strategies for itaconic acid production using engineered microorganisms,” IST, 2017.
 - [49] F. Mendes, “Synthetic Biology Approaches to Implement Microbe-based Production of Economically Relevant Itaconic Acid Derivatives,” IST, 2019.
 - [50] J. Correia, “Strategies to improve yeast-based production of itaconic acid: focus on mitochondrial to cytosol transport,” IST, 2018.
 - [51] A. Vila-Santa, “Metabolic engineering strategies to improve yeast- based production of itaconic acid guided by in silico metabolic modelling,” IST, 2015.
 - [52] T. Li, H. Iwaki, R. Fu, Y. Hasegawa, H. Zhang, and A. Liu, “ α -amino- β -carboxymuconic- ϵ -semialdehyde decarboxylase (ACMSD) is a new member of the amidohydrolase superfamily,” *Biochemistry*, vol. 45, no. 21, pp. 6628–6634, 2006, doi: 10.1021/bi060108c.

Chapter 5– Prospecting of metabolic pathways for implementation of microbe-based production of Methacrylic Acid (MAA)

- [53] A. Liu and H. Zhang, “Transition metal-catalyzed nonoxidative decarboxylation reactions,” *Biochemistry*, vol. 45, no. 35, pp. 10407–10411, 2006, doi: 10.1021/bi061031v.
- [54] S. Horowitz and R. C. Trievel, “Carbon-oxygen hydrogen bonding in biological structure and function,” *J. Biol. Chem.*, vol. 287, no. 50, pp. 41576–41582, 2012, doi: 10.1074/jbc.R112.418574.
- [55] P. B. Juhl, K. Doderer, F. Hollmann, O. Thum, and J. Pleiss, “Engineering of *Candida antarctica* lipase B for hydrolysis of bulky carboxylic acid esters,” *J. Biotechnol.*, vol. 150, no. 4, pp. 474–480, 2010, doi: 10.1016/j.jbiotec.2010.09.951.
- [56] S. Sacchi, S. Lorenzi, G. Molla, M. S. Pilone, C. Rossetti, and L. Pollegioni, “Engineering the substrate specificity of D-amino-acid oxidase,” *J. Biol. Chem.*, vol. 277, no. 30, pp. 27510–27516, 2002, doi: 10.1074/jbc.M203946200.

Chapter 5– Prospecting of metabolic pathways for implementation of microbe-based production of Methacrylic Acid (MAA)

Chapter 6 – Searching for the genetic parts to develop a yeast-based biosensor for itaconic acid

6.1. Abstract

The main goal of this chapter is to take the first steps into the design of a biosensor endogenous to *S. cerevisiae*, based on a responsive transcription factor and its respective binding site, that could be used to detect itaconic acid. To identify whether *S. cerevisiae* was equipped with an ITA-responsive regulator, a transcriptomic analysis, based on RNA-sequencing, was undertaken. For that yeast cells were exposed to two inhibitory concentrations of ITA (at pH 3.5) and the transcriptomes were analyzed during the acute phase of response (1h prior exposure) or after adaptation (when cells resumed exponential growth in the presence of the acid). From the results obtained it was possible to obtain a list of 12 genes that were found to be induced by the two different ITA concentrations (sub and inhibitory) in a linear manner, with *DAL2*, *DAL5*, *MEP2* and *TMT1* being more promising. Analysis of the regulatory associations of these 12 candidate genes with the corresponding regulators, based on the information described in the YEASTRACT database, showed that the majority is under the regulation of Leu3, previously described to be a transcription factor whose activity is modulated upon 2-malate (2-IPM) binding. The experimental work that was performed afterwards in order to confirm whether Leu3 and some of their regulated genes may serve as suitable genetic parts in the design of an ITA biosensor will be described.

6.2. Introduction

ITA is a C5 dicarboxylic acid that has attracted interest due to its potential to be converted in various molecules with wide industrial applications, from rubber and acrylic polymers to superabsorbents and solvents [1]. ITA is naturally produced by *A. terreus*, *U. maydis*, human macrophages and other non-conventional yeasts [2]–[5]. Industrial production of ITA has been undertaken by submerged fermentation of *A. terreus*; however, problems due to broth rheology and lack of reproducibility complicate the process. On the other hand, the production of ITA in alternative hosts has been widely explored in the last few years, including in the natural host *U. maydis* [6] and in the engineered hosts *A. niger* [7], *E. coli* [8], *S. cerevisiae* [9], *Y. lipolytica* [10] and *P. kudriavzevi* [11]. As the number of potential microbial hosts for ITA production increases, it also increases the need of developing suitable methods that could be used to rapidly screen candidate producing strains, either from different species or even from the same species (mutants, for example). In the case of *S. cerevisiae*, for example, there are available various collections of mutants that could be phenotyped for their ability to produce ITA, provided that there are methods that could speed up the processing of ITA detection since these collections comprise more than 5,000 strains and the current methods for ITA detection require, at least, 30 min of an HPLC run. This is a problem that has been solved for other biomolecules by exploring the development of metabolite sensors, which couple the detection of the concentration of a given bioanalyte to a high-throughput readout (fluorescence, cell growth or cell resistance)[12].

As said before, the main motivation for the development of metabolite-sensors is the ability to rapidly screen large libraries of strains, although this is also useful to test in a rapid manner different metabolic engineering strategy that can be used to improve pathway flux. A good example undertaken in yeast cells was the use of a malonyl-coA biosensor to screen a genomic cDNA library to select strains exhibiting higher internal pools of malonyl-coA, these being strains with high potential to produce the bulk chemical 3-hydroxypropionic acid [13]. Metabolite sensors have also been used to optimize pathway expression [14] and to enable adaptive laboratory evolution of overproducer strains [15]. While in most applications the sensor system is encoded in a producer strain, it is also possible to use a sensing strain to screen fermentation supernatants in a separate stage from production, as demonstrated by the use of a War1 yeast sensor to screen of medium and short-chain fatty acid producers [16]. In addition to

high-throughput screening, biosensors can be further modified to enable dynamic pathway regulation, by switching the reporter genes with pathway genes [17], [18].

Currently, metabolite biosensors can be divided in three main categories, based on how design of the sensor module: Förster-Resonance Energy Transfer (FRET) -based, RNA-based and Transcription Factor (TF)-based. FRET-based sensors consist on a single polypeptide chain of two fluorophores flanking an analyte recognition domain, which undergoes a conformational change upon analyte binding that enables FRET to occur and, consequently, produces a fluorescent readout from the acceptor [19]. RNA-based sensors are complex structures formed by single stranded RNA, which can be riboswitches or ribozymes, and that interfere with the expression of a reporter gene in a function of the metabolite binding to the RNA[20]. TF-based sensors utilize the natural ability of the analyte to bind to a regulator and modulate its activity in increasing or decreasing the expression of a reporter gene [21]. The accumulated knowledge on transcriptional regulators enable the design of various metabolite-responsive TFs and, in particular, many bacterial TFs that are naturally responsive to metabolites have been explored to create sensors, including AraC for lycopene, benM for benzoate, DcuR for succinate, FadR for fatty acids and FapR for malonyl-coA [22]. In the case of ITA, a bacterial sensor has been developed based on the IctR regulator found in *Y. pseudotuberculosis* and *P. aeruginosa* that regulates the expression of the ITA degradation operon in these species [23]. The two sensor systems were implemented in *E. coli* and *C. necator* by expressing the regulator IctR and the reporter gene Red Fluorescent Protein (RFP) under the control of a promoter from an ITA degradation gene (Pccl from *Y. pseudotuberculosis* and Pich from *P. aeruginosa*). While the *Y. pseudotuberculosis* system enabled ITA-dependent fluorescence in both hosts, it was observed in *C. necator* that induction could be obtained without expressing IctR, suggesting that this host has a IctR homolog that can activate the Pccl promoter. In the case of the *P. aeruginosa* system, it was only possible to observe ITA-induction of fluorescence in *C. necator*, which may be attributed to a lack of correct IctR translation in *E. coli* (the gene was not codon optimized), which in *C. necator* can be compensated by an endogenous IctR homolog. The *Y. pseudotuberculosis* sensor system in *E. coli* was further characterized and the linear range was found to be between 0.07 and 0.7 mM ITA, which will become problematic in settings where the ITA production titer is higher, such as in *P. kudriavzevii* [11]. Furthermore, the sensor was also demonstrated to have cross-reactivity to succinic, methylsuccinic, mesaconic, α -ketoglutaric, propionic, butyric, 3-butenic, acrylic, methacrylic, cis-aconitic and trans-aconitic acids [23].

In this work it was aimed to develop a new biosensor for ITA that could only be dependent of the *S. cerevisiae* endogenous network. The simple porting of the ItcR regulator and its operator sites into *S. cerevisiae*, a strategy used with success for malonyl-coA and cis-cis-muconic acid [13], [24], could be a possibility, however, the precise binding site of the ItcR regulator was not determined (the authors have used the entire promoter region of the itaconic acid-degradation operon) and cloning of a promoter with around 400 bp would be required which could cause interference with the endogenous yeast transcriptional regulatory network. Furthermore, often the porting of prokaryotic sensors into eukaryotic hosts requires extensive engineering to tune the response of the sensor. Additionally, considering that the most explored ITA hosts are eukaryotic, the design of an ITA sensor that is based on a eukaryotic regulator and promoter would likely have higher chances to be ported to hosts like *Aspergilli sp*, *U. maydis* or *P. kudriavzevii*. In the case of *S. cerevisiae*, the ability to directly screen collections of mutant strains for ITA production in one step would be dependent on the ability to co-transform the collections with a plasmid driving CAD1 expression and thus, enabling ITA production, and a plasmid expressing the sensor system components, enabling ITA sensing.

Since almost nothing has been investigated concerning the response of *S. cerevisiae* to ITA, let alone the existence of regulators reported to interact with ITA, the first step of this approach was to devise an experimental strategy that could lead to the identification of the more suitable genetic parts required for the engineering of a suitable ITA biosensor: a regulator whose activity is responsive to ITA and its corresponding operating sequences (or binding sites). For such identification, we have undertaken a transcriptomics analysis in which the cells were exposed to ITA, expecting that the cellular stress response to intracellular accumulation could result in the identification of activated TFs, a strategy that was pursued before to find TF-promoter pairs to be used for dynamic regulation purposes [17]. In specific, the alterations in the transcriptome of *S. cerevisiae* to ITA to two different ITA concentrations was assessed aiming to identify which could be the genes more strongly up-regulated (preferably in a linear manner) and what could be their regulators.

6.3. Methods

6.3.1. Microorganisms and culture conditions

The *S. cerevisiae* strains used in this work were maintained in YPD medium and cultivated in MMB. The composition of YPD medium is 2% glucose (Nzytech), 2% peptone (Nzytech) and 1% yeast extract (Nzytech). The composition of MMB is 2% glucose/galactose (Nzytech/Merck) 0.267% ammonium sulfate and 0.17% Yeast Nitrogen Base without aminoacids (Difco). MMB media was supplemented with the required amino acids and nucleobase (all from Sigma) to maintain growth of the BY4741 genetic background as well as to assure selection of plasmids. The following concentrations were used: 20 mg/L histidine, 60 mg/L leucine, 20 mg/L methionine and 20 mg/L uracil. Whenever needed MMB medium was acidified at pH 3.5 using HCl as the acidulant. The lists of strains used in this chapter is available in Table 6.1.

Table 6.1. List of strains used in chapter 6.

Name	Description	Source
By4741	MATa his3 Δ 1 leu2 Δ 0 met15 Δ 0 ura3 Δ 0	Euroscarf
By4741 Δ leu3	By4741 leu3::kanMX4	Euroscarf
By4741 Δ dal81	By4741 dal81::kanMX4	Euroscarf

6.3.2. Cell cultivation in the presence of ITA and RNA extraction

S. cerevisiae BY4741 cells were cultivated overnight at 30°C and 250 rpm in MMB (at pH=3.5). On the next day these cells were inoculated in fresh MBB medium at an initial OD_{600nm} of 0.1. These cells were cultivated at 30°C and 250 rpm until an OD_{600nm} of 0.5-0.8 was reached. At this point an appropriate volume of cells was filtered in order to inoculate 150 mL of MMB glucose (at pH 3.5), either or not supplemented with 175 mM or 290 mM ITA at pH=3.5, at an OD of 0.1. Cells were harvested (by centrifugation at 5000 rpm for 5 min at 4°C) after 1h of cultivation in the presence or absence of the acid, washed with PBS and stored at -80°C until RNA extraction. To obtain the points

corresponding to mid-exponential phase, cells were harvested after 5h in the presence of 170 mM ITA and after 14h when growing in the presence of 290 mM ITA. For RNA extraction, the MasterPure™ Yeast RNA Purification Kit (Lucigen) was used following the manufacturer's instructions.

6.3.3.RNA sequencing

Analysis of the effect of ITA in the transcriptome of yeast cells was assessed using RNA-sequencing based on the QuantSeq platform. This was performed as a paid service at Instituto Gulbenkian de Ciência- Gene Expression Unit. Prior to RNA-seq, the RNA samples obtained were subjected to tested for integrity in a Bioanalyzer, Only samples exhibiting a RIN (RNA Integrity Number) above 8.5 were selected for subsequent sequencing in the QuantSeq platform. The analysis of the data obtained after the sequencing was performed in the Bluebee® (Lexogen) platform. Briefly, the fastq files were uploaded to Bluebee and a series of steps were followed: 1) run the pipeline FASTQ Merging Pipeline for merging multiple FASTQ files from the same sample; 2) run the QuantSeq FWD-UMI pipeline to trim and align the reads against *S. cerevisiae*'s genome; 3) run the DESeq2 pipeline to analyze differential gene expression. After this procedure a list of genes with a significant difference of expression in the control and in the two ITA samples (175 mM and 290 mM) was obtained. Only genes whose expression in the presence of ITA was at least 1.5-fold different from the one registered in control cells (and with an associated p-value below 0.05) were considered to be differentially expressed and selected for downstream analysis. To select the best ITA-inducible genes, only genes found to be up-regulated in the presence of 175 mM and 290 mM of ITA were considered. To quantify that up-regulation a slope of induction was calculated (as detailed in equation i) being selected the genes showing a slope above zero. The resulting list of genes was ordered by activation slope (ii) to obtain the top-20 ITA inducible genes (by considering the zero ITA concentration point in the activation slope it was expected to find the genes that display a higher upregulation both between 0 and 175 mM ITA and between 175 mM and 290 mM ITA). The RSQ (Square of the Pearson Product-Moment Correlation Coefficient) was also calculate for the top 20 ITA inducible genes to evaluate the linearity of the upregulation. The same procedure was performed to obtain the top20-down-regulated genes.

$$\text{Slope of fold change} = \frac{\text{Fold change 290} - \text{Fold change 175}}{(290 - 175)} \quad (\text{i})$$

$$\text{Activation slope} = \text{Slope of linear regression of } \frac{(1, \text{Fold change 175}, \text{Fold change 290})}{(0, 175, 290)} \quad (\text{ii})$$

To find the candidate regulators of the top 20 genes up- or down- regulated by ITA, the Yeasttract database (<http://yeastract-plus.org/>) [25] was used, by use of the TF rank tool. The ShinyGO online tool was used to find enriched GO terms in the datasets (<http://bioinformatics.sdstate.edu/go/>) [26]. A functional protein interaction network was generated by use of the online server of STRING (<https://string-db.org/>) [27].

To confirm some of the results indicated by the RNA-seq, RT-qPCR was performed. For this the cells were cultivated under the same conditions as described above for RNA-seq. cDNAs were prepared using the QuantiTect Reverse Transcription Kit (Quiagen) and used for Reverse Transcription Quantitative PCR (RT-qPCR) with the Nzy qPCR Green Master Mix kit (Nzytech) in an Applied Biosystems 7500 RT-qPCR machine. The RT-qPCR results were analyzed using the $\Delta\Delta\text{CT}$ method, using ALG9 as the housekeeping gene. The primers used can be found in Annex 8.3-Table 8.9.

6.4. Results and Discussion

To find the genetic elements (responsive promoter and TF) necessary for the engineering of an ITA biosensor using the endogenous machinery of *S. cerevisiae*, a transcriptomics approach was explored. For that, the alterations in the transcriptome of yeast cells when suddenly exposed to two concentrations of ITA, 175 mM and 290 mM, at pH 3.5, was monitored. Under the experimental conditions used, the concentration of 175 mM had no deleterious effect in growth of yeast cells, being observed even a slight stimulation of the growth rate when compared with the one obtained for cells growing in the absence of the acid. Differently, the cells exposed to 290 mM ITA exhibited a lag phase of around 6h, after which they resumed exponential growth at a growth rate similar to the one observed in unstressed cells (Fig.6.1). For the transcriptomic analysis it was decided to sample the cells after 1h of cultivation in the presence of the acid, this corresponding to an early period of adaptation where most transcriptome-wide changes are reported to occur. In order to have an idea on how the adaptation shapes the transcriptomic response of the cells a second time-point was selected, corresponding to the mid-exponential phase which was reached after around 5h in the cells that were cultivated in the presence of 175 mM ITA and after 14h in the cells that were being cultivated in the presence of 290 mM ITA. As a control cells unexposed to ITA were harvested after 1 and 5 hours of culture start, corresponding to lag and exponential phase, respectively. In the following sections the responses obtained in these two points will be described with more detail.

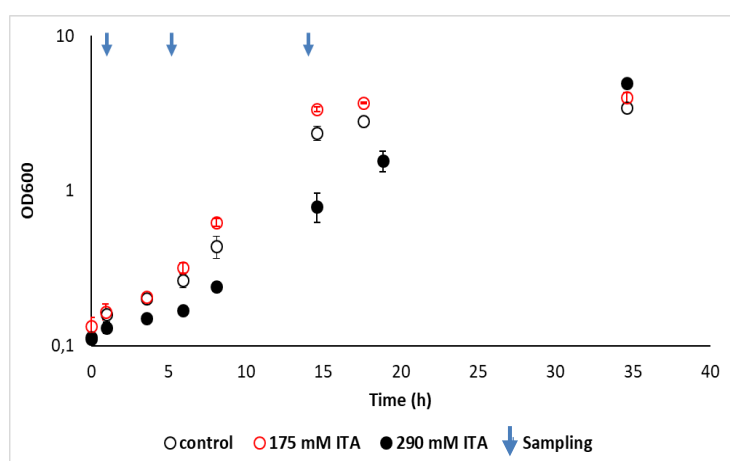


Figure 6.1. Growth curve of By4741 cells in MMB with the supplementation of ITA (175 mM, red open circles, 290 mM black open circles) and without (black closed circles). Samples for RNA sequencing were harvested during the lag phase (1 hour) and exponential phase (5h for cultures with no ITA and 14h for cultures with 290 mM), represented in blue arrows.

6.4.1. Short-term *S. cerevisiae* transcriptomic response to ITA

After the short exposure (of 1h) of unadapted *S. cerevisiae* cells to 175 and 290 mM of ITA (at pH 3.5), 985 and 1176 genes, respectively, were found to be up-regulated (above 1.5-fold), with 385 genes being in common in the two datasets. The list of the genes up-regulated in both conditions can be found in Annex 8.5-Table 8.10. Since linearity is an important feature of a biosensor, attention was focused on genes whose induction promoted by ITA was higher at 290 mM than the one registered for 175 mM. To identify those genes, a slope in the activation ratio was calculated (as detailed in materials and methods) and only genes having an associated induction with a positive slope (meaning it was higher at 290 than at 175) were selected for downstream analysis. A total of 295 genes were found to meet this criterion, this being highlighted in the tables shown in Annex 8.5-Table 8.10. To get further insights into the magnitude by which the transcription of these genes was impacted by ITA, a second slope was calculated, the “activation slope”, which compared the expression not only in the 175 and 290 mM conditions, but also in the control sample (0h) (Figure 6.2A). The genes were sorted by descending activation slope, with the top genes being those that displayed a higher ITA-dependent activation. In Figure 6.2B it is represented the fold-change of the top 20 genes showing a higher ITA-dependent activation, with *DAL5* being the gene with the highest activation slope (induced 36-fold when cells were cultivated in the presence of 175 mM ITA and 93-fold when the cells were cultivated in the presence of 290 mM ITA). Despite this, the *DAL5* dose-response curve does not appear to be as linear as those observed for the *MEP2*, *DAL2* and *DAL3* genes that were less strongly expressed but whose expression increased in a more linear manner (Fig.6B).

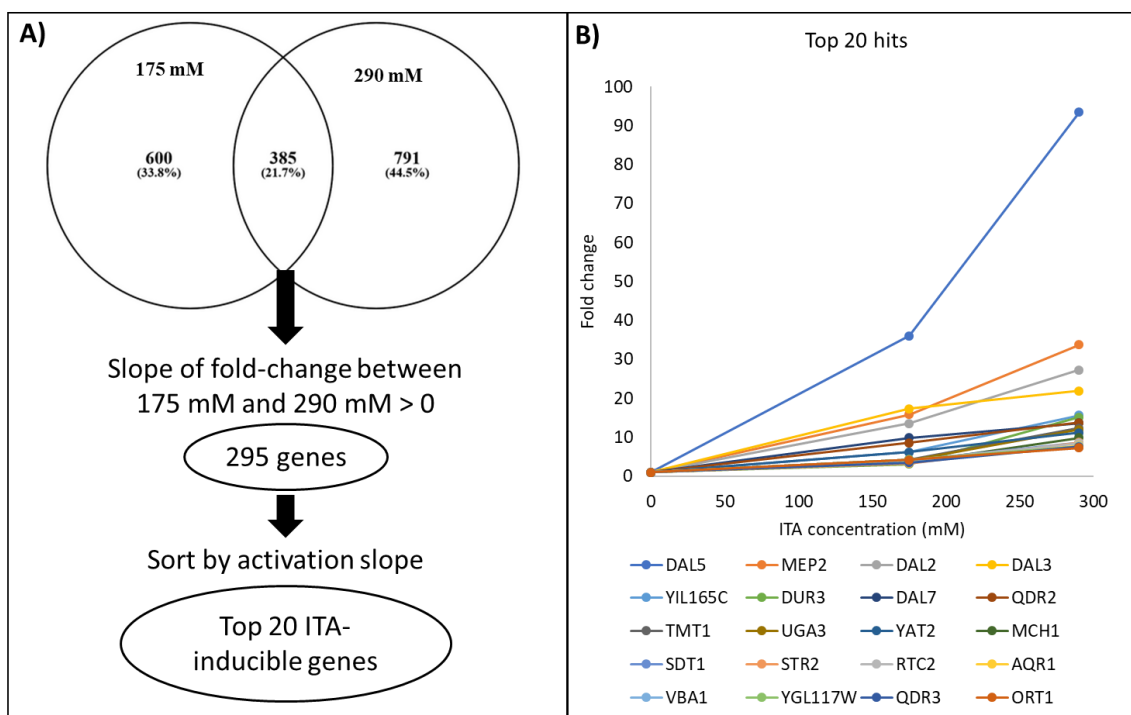


Figure 6.2. ITA-inducible genes. A) Approach to find the top 20 genes with an ITA induction: 385 genes were found to be upregulated in both ITA exposure conditions, where 295 of these had a positive slope between the fold-change at 175 mM and 290 mM of ITA. The genes with a higher dose-dependent activation were found by considering the slope of the fold change between the points with ITA=0, 175 and 290 mM. B) Dose-response curve of the top 20 ITA-inducible genes found in this work.

A brief description of the top 20 ITA-inducible genes that were selected with the above-described approach is provided in Table 6.2, including also the activation slope and the linearity of the dose-response curve (evaluated by the RSQ, Square of the Pearson Product-Moment Correlation Coefficient), two factors that were used here to evaluate the ITA expression profile that could be advantageous for the design of an ITA sensor. From the functional point of view, it was interesting to observe a high prevalence of genes involved in nitrogen metabolism including in the degradation of allantoate (*DAL2*, *DAL3* and *DAL7*) and in transport of allantoate (*DAL5*), ammonia (*MEP2*), urea or polyamines (*DUR3*). The list also included the major facilitator superfamily transporters *QDR2* and *AQR1*, and also *MCH1*, a protein with a similarity to monocarboxylate permeases. The inclusion of *Qdr2* in this list is particularly interesting considering that previous work from our laboratory has shown that this transporter, as well as its homologue *Qdr3*, are involved in the export of itaconate thus being critical determinants of *S. cerevisiae*

Chapter 6– Searching for the genetic parts to develop a yeast-based biosensor for itaconic acid

tolerance to this organic acid [28]. Another gene that also appeared relevant was *TMT1*, encoding a “moonlighting” methyltransferase involved in methylation of 3-isopropylmalate (3-IPM) (that serves as a signal to induce invasive growth in conditions of amino acid starvation [29]) and in methylation of trans-aconitate, which is a toxic product formed by spontaneous isomerization of cis-aconitate, the precursor of ITA [30]. Interestingly, Tmt1 was also found to methylate ITA to its ester and its *A. niger* homologue was up-regulated in ITA-producing conditions, although the physiological relevance of this observation has not been elucidated [31].

Table 6.2. Top 20 ITA-inducible genes found in this work. The genes are sorted in descending order of the activation slope.

Gene	Activation slope	RSQ	Gene description
<i>DAL5</i>	0.309	0.935	Allantoate permease; ureidosuccinate permease; also transports dipeptides.
<i>MEP2</i>	0.111	0.970	Ammonium permease involved in regulation of pseudohyphal growth.
<i>DAL2</i>	0.089	0.978	Allantoicase, converts allantoate to urea and ureidoglycolate in the second step of allantoin degradation.
<i>DAL3</i>	0.074	0.962	Ureidoglycolate hydrolase, converts ureidoglycolate to glyoxylate and urea in the third step of allantoin degradation.
<i>YIL165C</i>	0.049	0.921	Putative protein of unknown function; this and the adjacent YIL164C likely constitute a single ORF encoding a nitrilase gene
<i>DUR3</i>	0.046	0.768	Plasma membrane transporter for both urea and polyamines.
<i>DAL7</i>	0.044	0.989	Malate synthase, role in allantoin degradation unknown.
<i>QDR2</i>	0.044	1.000	Multidrug transporter of the major facilitator superfamily
<i>TMT1</i>	0.037	0.875	Trans-aconitate methyltransferase. cytosolic enzyme that catalyzes the methyl esterification of 3-IPM. an intermediate of the leucine biosynthetic pathway. and trans-aconitate

Chapter 6– Searching for the genetic parts to develop a yeast-based biosensor for itaconic acid

<i>UGA3</i>	0.036	0.873	Transcriptional activator necessary for GABA-dependent induction of GABA genes (such as <i>UGA1</i> , <i>UGA2</i> , <i>UGA4</i>).
<i>YAT2</i>	0.035	0.988	Carnitine acetyltransferase
<i>MCH1</i>	0.029	0.875	Protein with similarity to mammalian monocarboxylate permeases.
<i>SDT1</i>	0.026	0.948	Pyrimidine nucleotidase
<i>STR2</i>	0.025	0.881	Cystathionine gamma-synthase, converts cysteine into cystathionine
<i>RTC2</i>	0.025	0.947	Protein of unknown function; similar to a G-protein coupled receptor from <i>S. pombe</i>
<i>AQR1</i>	0.023	0.942	Plasma membrane multidrug transporter of the major facilitator superfamily.
<i>VBA1</i>	0.022	0.971	Permease of basic amino acids in the vacuolar membrane
<i>YGL117W</i>	0.022	0.987	Putative protein of unknown function
<i>QDR3</i>	0.022	0.937	Multidrug transporter of the major facilitator superfamily.
<i>ORT1</i>	0.021	0.986	Ornithine transporter of the mitochondrial inner membrane; exports ornithine from mitochondria as part of arginine biosynthesis.

The promoters of genes listed in the table above are candidates to harbor ITA-responsive DNA regions, however, it is also necessary to figure out what could be the regulator responsible for recognition and binding to those binding sites. In other words, to find what could be the ITA-responsive regulator. To find TFs that could be mediating the activation by ITA of the top 20 genes listed in Table 6.2, the Yeastract database [25], a repository of all described associations between yeast TFs and their target genes, was used. When considering only regulations in which it is demonstrated binding of the TF to the promoter region of the target gene only Dal81 was found to be significantly enriched in the dataset, regulating 8 (*UGA3*, *DAL2*, *DAL7*, *DAL3*, *DAL5*, *VBA1*, *MEP2* and *QDR3*) out of the 20 genes. Dal81 is a positive regulator of genes in nitrogen catabolic pathways, including the degradation of allantoin and GABA, with a relevant role in Nitrogen Catabolite Repression (NCR) [32]. Although Dal81 has a helix-turn-helix motif and a GAL4-like Zn²/Cys₆ DNA binding domain, the deletion of the Zn²/Cys₆ domain does

not affect the induction of urea amidolyase activity (encoded by *DURI*,2), suggesting that this domain is not responsible for DNA binding and raising the possibility that its main mechanism of action may be other (e.g. binding to other regulators)[33]. Thus, although Dal81 has a binding domain, it is unclear if it indeed binds its target promoters or if it regulates their expression indirectly.

If the criterium is broaden to include also associations in which it is described the effect of the regulator in the target gene expression (but without a demonstration of *in vivo* binding to the promoter region of that target), eight other TFs appear as likely candidates to regulate the top 20 genes (Table 6.3). Three of these TFs are described to regulate all the top 20 genes: Gcn4, a regulator of the amino acid starvation response; Gln3, also a regulator of utilization of alternative nitrogen sources; and Yap1, a TF involved in *S. cerevisiae* response to oxidative stress tolerance. The Dal80 positive regulator of nitrogen degradation pathways and its homolog Gzf3 were also found in the set of candidate regulators, as well as Leu3, a co-repressor/activator of the leucine biosynthetic pathway. Notably, among these TFs, only Leu3 has a mechanism of activation that depends on the direct binding of a metabolite, 2-IPM, an intermediate in the leucine biosynthetic pathway [29]. In specific, it was demonstrated that binding of 2-IPM to Leu3 unmask the activation domain, enabling the transcriptional activation of Leu3-mediated genes [34]. There is some structural similarity between ITA and 2-IPM, it may be possible that ITA could also directly activate Leu3 resulting in the induced expression of the target genes listed in table 6.2. The fact that Tmt1, a methyltransferase, is able to accept both ITA and the 2-IPM isomer, 3-IPM, as substrates further supports the biochemical similarity of these two molecules [35] (Figure 6.3). One important feature of Leu3 regulation is that the degree of activation by Leu3 is Leu3 concentration-dependent and it has been hypothesized that is because elevated levels of Leu3 may titrate out a co-repressor protein (possibly Mot1) [34]. Since Leu3's own expression is regulated by Gcn4, this Leu3 regulatory mechanism that is dependent of the TF's concentration may be a way of the cell to control multiple pathways through only Gcn4-activation and control of Leu3 concentration in the cell [34]. In the search for a TF-promoter pair to design a metabolite-sensor, this setup is extremely desirable, since it would enable tuning the dose-response curve by changing the promoter strength controlling Leu3 expression and thus, the concentration of the TF [21].

Chapter 6– Searching for the genetic parts to develop a yeast-based biosensor for itaconic acid

Table 6.3. TF regulations enriched in the top 20 genes activated by ITA in the lag phase dataset. Data obtained using the TF rank function of the Yeastract database [25]. It is shown the enrichment of the TF in this dataset in comparison to the enrichment of the TF in the overall Yeastract database.

TF	Enrichment (%)		TF description
	in Dataset	in Yeastract database	
Gcn4p	100	0.33	Activator of amino acid biosynthetic genes in response to amino acid starvation; expression is tightly regulated at both the transcriptional and translational level
Gln3p	100	0.83	Activator of genes regulated by NCR, localization and activity regulated by quality of nitrogen source
Yap1p	100	0.33	TF required for oxidative stress tolerance; activated by H ₂ O ₂
Gzf3p	50	2.13	Dal80p homolog that negatively regulates NCR gene expression by competing with Gat1p for GATA site binding
Dal81p	45	2.39	Positive regulator of genes in multiple nitrogen degradation pathways; contains DNA binding domain but does not appear to bind the dodecanucleotide sequence present in the promoter region of many genes involved in allantoin catabolism
Dal80p	35	3.50	Negative regulator of genes in multiple nitrogen degradation pathways; expression is regulated by nitrogen levels and by Gln3p
Leu3p	50	1.87	TF that regulates genes involved in branched chain amino acid biosynthesis and ammonia assimilation
Hac1p	70	0.96	TF (ATF/CREB1 homolog) that regulates the unfolded protein response, via UPRE binding, and membrane biogenesis
Bas1p	90	0.60	Myb-related TF involved in regulating basal and induced expression of genes of the purine and histidine biosynthesis pathways

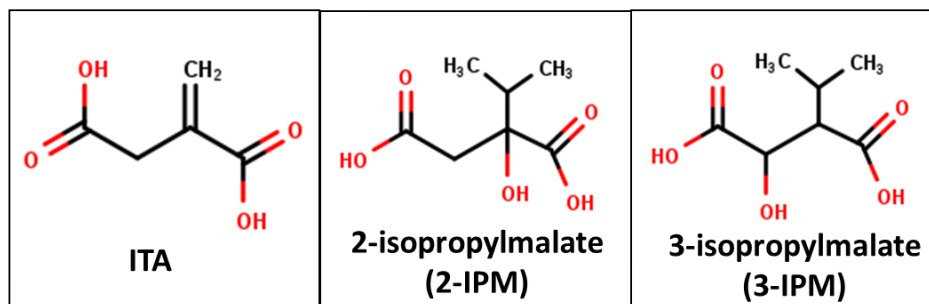


Figure 6.3. Structures of ITA, the metabolite that binds and activates the Leu3 regulator (2-IPM) and the major yeast substrate of Tmt1, trans-aconitate methyltransferase (3-IPM).

6.4.1.1. ITA down-regulated genes

While the analysis of genes up-regulated in the presence of ITA unraveled ITA-inducible genes that may be starting points of an activator-based ITA sensor, the analysis of down-regulated genes may uncover ITA-repressible genes, which can also be used to design a repressor-based ITA biosensor. As such, 926 genes were found to be down-regulated when cells were exposed for 1 hour to 175 mM ITA (at pH 3.5), this number increasing to 1180 when the cells were incubated for the same time in the presence of 290 mM ITA. 426 genes were found to be down-regulated in both conditions (complete list in Annex 8.5-Table 8.11). A similar procedure as the one described for the up-regulated genes was used here to find ITA-repressible genes. One difference was that to find the top 20 ITA-repressible genes the criteria was to sort by the slope of log fold change and not the slope of fold change (Figure 6.4A). This decision was made because the activation slope of the top 20 genes was very similar among them, making this measure not significant for a downregulation analysis. Thus, the genes were sorted by the log activation slope and the dose-response curve of the top 20 ITA-repressible genes are represent in Figure 6.4B. While the level of repression imposed by 290 mM ITA was similar for the top 20 genes (between 0.16 and 0.28), a wider dispersion was observed with 175 mM necessarily affecting the linearity of the dose-response curve. Nonetheless, it was possible to identify eight genes displaying a linear expression behavior ($RSQ > 0.98$): *CGR1*, *SPB1*, *AAH1*, *DPH5*, *IMD4*, *YMR304C-A*, *ELO3*, *EFM4*, *YFR052C-A*, *IMD3* and *KRE33*. Of these, it stands out the YFR052C-A ORF by showing the higher linearity and slope.

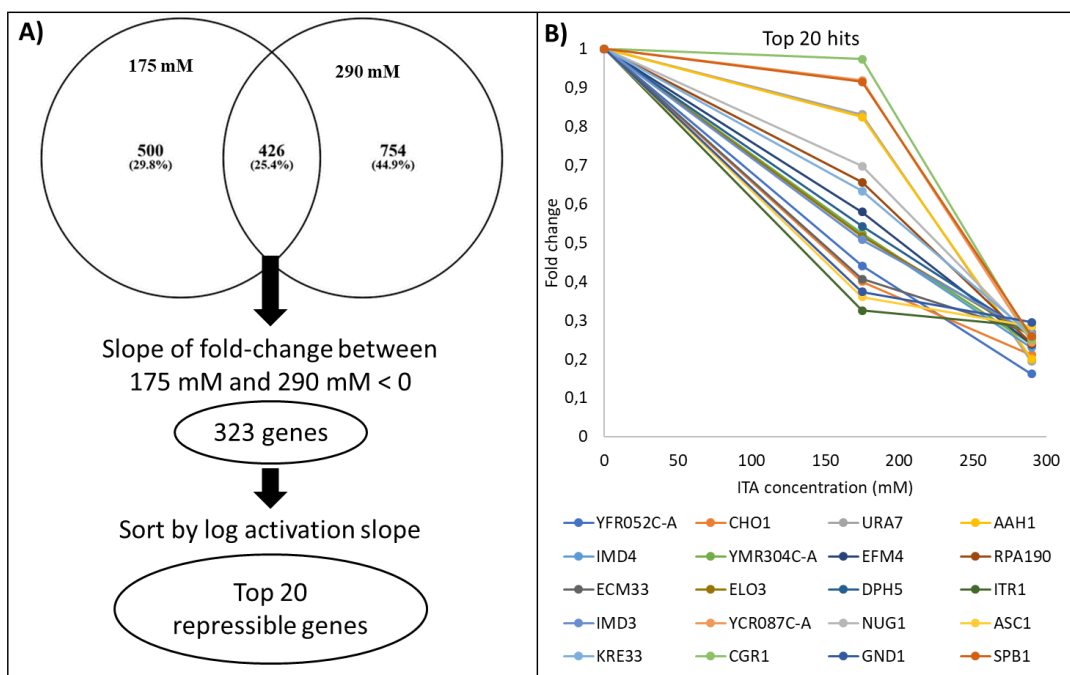


Figure 6.4. ITA-repressible genes. A) Approach to find the top 20 genes with an ITA repression: 426 genes were found to be down-regulated in both ITA exposure conditions, where 323 of these had a negative slope between the fold-change at 175 mM and 290 mM of ITA. The genes with a higher dose-dependent repression were found by considering the slope of the log fold-change between the 0, 175 mM and 290 mM ITA conditions. B) Dose-response curve of the top 20 ITA-repressible genes found in this work.

The description of the top 20 ITA-repressed genes is shown in Table 6.4, as well as the log activation slope and the linearity of the dose-response curve. In terms of functional categorization, it is possible to see that the list of 20 candidate genes includes five genes involved in ribosomal function (*NUG1*, *ASC1*, *KRE33*, *CGR1* and *SPB1*), two genes responsible for the methylation of translation elongation factors (*EFM4* and *DPH5*), four genes involved in purine synthesis (*IMD3* and *IMD4*) and salvage (*AAH1*) and pyrimidine synthesis (*URA7*), the phosphatidylserine synthase *CHO1*, a myo-inositol transporter (*ITR1*), a fatty acid elongase (*ELO3*), a protein involved in glucose uptake (*ECM33*) and a NADPH regenerating enzyme that participates in the pentose phosphate pathway (*GND1*). Finally, three ORFs with no clearly attributed function are also in the top 20 ITA-repressible gene list: YFR052C-A, YMR304C-A (the 3' end partially overlaps with *SWC10*, encoding a cell wall protein) and YCR087C-A, with no putative function. Among the list the more relevant functional trend observed is the repression of genes required for ribosomal function and translation machinery, which is a known feature of yeast response to environmental stress as a mean to constrain growth [36].

Chapter 6– Searching for the genetic parts to develop a yeast-based biosensor for itaconic acid

Table 6.4. Top 20 ITA-repressible genes found in this work. RSQ: Square of the Pearson Product-Moment Correlation Coefficient. The genes are sorted in ascending order of the log activation slope

Gene	Log Activation Slope	RSQ	Gene description
YFR052C-A	-0,00291	0,995	Dubious open reading frame
CHO1	-0,00278	0,970	Phosphatidylserine synthase, functions in phospholipid biosynthesis
URA7	-0,00263	0,817	Major CTP synthase isozyme, catalyzes the final step in de novo biosynthesis of pyrimidines
AAH1	-0,00261	0,825	Adenine deaminase, involved in purine salvage
IMD4	-0,00266	1,000	Inosine monophosphate dehydrogenase, catalyzes the first step of GMP biosynthesis
YMR304C-A	-0,00263	0,999	Dubious open reading frame
EFM4	-0,00261	0,997	Probable lysine methyltransferase involved in the dimethylation of eEF1A (Tef1p/Tef2p)
RPA190	-0,00257	0,970	RNA polymerase I largest subunit A190
ECM33	-0,00260	0,950	GPI-anchored protein involved in efficient glucose uptake
ELO3	-0,00257	0,997	Elongase, involved in fatty acid and sphingolipid biosynthesis
DPH5	-0,00255	1,000	Methyltransferase required for synthesis of diphthamide, which is a modified histidine residue of translation elongation factor 2 (Eft1p or Eft2p)
ITR1	-0,00258	0,880	Myo-inositol transporter
IMD3	-0,00255	0,994	Inosine monophosphate dehydrogenase, catalyzes the first step of GMP biosynthesis
YCR087C-A	-0,00243	0,732	Putative protein of unknown function

Chapter 6– Searching for the genetic parts to develop a yeast-based biosensor for itaconic acid

NUG1	-0,00250	0,949	GTPase that associates with nuclear 60S pre-ribosomes
ASC1	-0,00256	0,908	G-protein beta subunit and guanine dissociation inhibitor for Gpa2p; core component of the small (40S) ribosomal subunit; regulates P-body formation induced by replication stress; represses Gcn4p in the absence of amino acid starvation; controls phosphorylation of multiple proteins
KRE33	-0,00251	0,985	Essential protein, required for biogenesis of the small ribosomal subunit
CGR1	-0,00238	0,671	Protein involved in nucleolar integrity and processing of the pre-rRNA for the 60S ribosome subunit
GND1	-0,00252	0,911	6-phosphogluconate dehydrogenase (decarboxylating), catalyzes an NADPH regenerating reaction in the pentose phosphate pathway
SPB1	-0,00239	0,738	AdoMet-dependent methyltransferase involved in rRNA processing and 60S ribosomal subunit maturation

A TF-enrichment analysis of the top 20 ITA-repressible dataset was performed to unravel what regulators could be behind the response. When using only associations that require DNA binding evidence, no TF that could be involved in co-regulation of these genes could be identified. When the criterium was broaden to also include TFs described to affect expression of target genes (without no clear demonstration of DNA binding evidence), four TFs were identified (Table 6.5). Yap1, an transcriptional activator required for oxidative stress tolerance, and Pdr3, another activator that regulates the PDR network [37], were found to be associated with all the 20 ITA-repressible genes. It has been hypothesized that Yap1 may be involved in the general weak acid transcriptional response in yeast, considering its enrichment in the upregulated genes yeast cells in response to acetic, sorbic, propionic, 2,4-dichlorophenoxyacetic and artemisinic acids [38]. Similarly, the Pdr3 regulon has also been implied in the yeast transcriptional response to 2,4-dichlorophenoxyacetic and artemisinic acids and the TF is a determinant of tolerance to artemisinic acid, artesunic acid, and 2,4-D [38]. Two other TFs were found enriched in the dataset: Aft1, a transcriptional activator that regulates iron utilization [39]

and Aca1, a TF involved in the transcription of genes required for the utilization of non-optimal carbon sources [40].

Considering that the four regulators found by enrichment analysis of the top 20 ITA repressible genes are reported as transcriptional activators, their role in the downregulation of these genes must be indirect, via interaction with other regulators. Indeed, Pdr3 has been hypothesized to create heterodimers with Rdr1, acting as repressor [37][41]; and Yap1 can act as a negative regulator when its expression is near physiological levels (presumably through indirect effects on other regulators) [42]. With this in mind, these 4 four activators do not appear as good candidates to design a ITA sensors, since there would be more regulators involved in the expression of the target promoter. Additionally, the analysis of the lag phase transcriptome response revealed a more concerted and less general response for the upregulation, with the top ITA-inducible genes being clustered around allantoin catabolism. Furthermore, at least one regulator was hypothesized to mediate the ITA-induction of these genes (Dal81). Consequently, the remainder of the analysis was focused in upregulation.

Table 6.5. TF regulations enriched in the top 20 genes repressed by ITA in the lag phase dataset. Data obtained using the TF rank function of the Yeastract database [25]

TF	Enrichment (%)		TF description
	Dataset	Yeastract	
Yap1p	100	0.31	TF required for oxidative stress tolerance; activated by H2O2
Pdr3p	100	0.32	Activator of the pleiotropic drug resistance (PDR) network, regulates expression of ATP-binding cassette (ABC) transporters through binding to cis-acting sites known as PDREs (PDR responsive elements)
Aft1p	52.63	0.85	TF involved in iron utilization and homeostasis; activates the expression of target genes in response to changes in iron availability
Aca1p	26.32	1.75	TF of the ATF/CREB family, may regulate transcription of genes involved in utilization of non-optimal carbon sources

6.4.2. Transcriptomic *S. cerevisiae* response to ITA in the exponential phase

To study the transcriptomic response *S. cerevisiae* cells to the presence of ITA after adaptation has occurred and when exponential growth is resumed, the transcriptome of cells cultivated in the presence of 290 mM ITA for around 14h were examined (this time-point corresponding to a period when the cells have resumed growth in the presence of the acid as seen in Figure 6.1). Besides contributing to better elucidate the responses of yeast cells to stress after acute adaptation, this analysis also renders possible to find genes up-regulated not only in the lag phase but they can maintain increased expression during exponential phase. This persistent up-regulation feature would be desirable for a metabolite sensor, since activation of stress response genes in the lag phase is expected but it should be transient and in some sensing applications there is a need for continuous activation.

When we compared the list of genes up-regulated in the lag phase and in the exponential phase upon exposure to 290 mM ITA, 100 genes were found to be upregulated in the lag phase when exposed to 290 mM ITA (complete list in Annex 8.5-Table 8.12), and of these, 43 had also been up-regulated during the early response to the 175 mM and 290 mM ITA (Figure 6.5A). Interestingly, among the genes up-regulated in lag phase and in exponential phase of ITA-stressed cells we could find 12 (Figure 6.5B) that were selected as good candidates to be considered ITA-inducible genes. This observation suggests that ITA induction of these genes is biologically significant and not only a result of a stress response. Even though the expression of these 12 genes is up-regulated in the exponential phase when compared to the untreated control, the level of induction is below the one registered during the lag phase (Figure 6.5B).

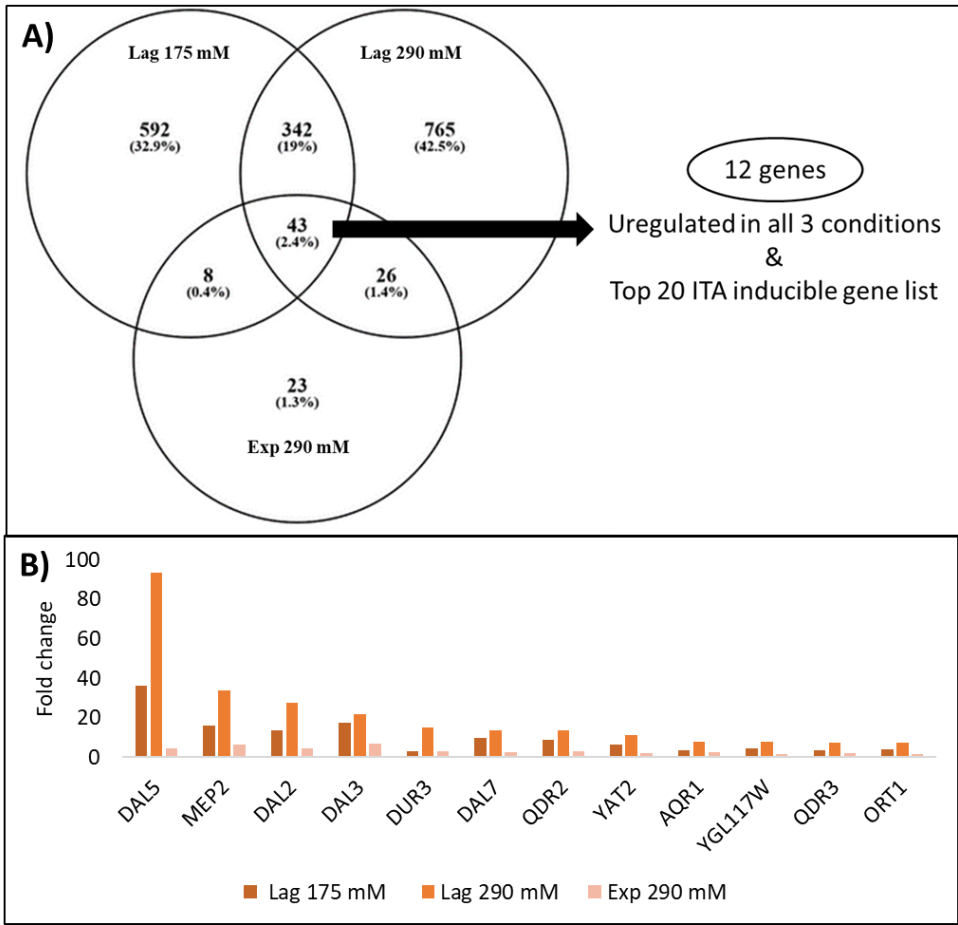


Figure 6.5. Top12 ITA inducible genes found based on the analysis of lag and exponential phase. A) 43 genes were found to be upregulated in all 3 conditions and 12 of those were in the top 20 list of lag phase ITA inducible gene list. B) Fold change of the top 12 ITA inducible genes in the three conditions tested here.

After analyzing the several datasets, this set of 12 genes listed in Figure 6.5 was considered as the more promising pool of candidates for an ITA-responsive promoter and an ITA-responsive regulator. TF clustering with mandatory direct DNA binding evidence, performed as detailed before, highlighted Dal81 as the only possible regulator of these genes, making this regulator a good candidate for a sensor design.

If the criterium is broaden to include evidence of the TF influencing the expression of the target (but without no mandatory DNA binding evidence), 10 other TFs are identified as putative regulators of these candidate genes under ITA stress, 8 of them previously described as having a role in the lag phase response (Table 6.6). The list contains one main orchestrator of the aminoacid starvation response (Gcn4) and two NCR regulators (Gln3 and Gzf3), which are not good candidates for a biosensor application, since so

many documented regulations could result in endogenous crosstalk that would likely compromise ITA induction and specificity. It also includes Yap1, which has been hypothesized to regulate a general weak acid stress response acids [38], thus making it an uninteresting candidate for an ITA sensor, due to the high number of regulations. Of the list, the Leu3 regulator stands out since it is a metabolite responsive regulator and due to the possibility of its direct interaction with ITA (Figure 6.3), making it a good candidate for an ITA sensor.

Table 6.6. TF regulations enriched in the top 12 list of ITA-inducible genes absed on lag and exponential phase analysis. Data obtained using the TF rank function of the Yeastract database [30]. 8 out 11 TFs were also enriched in top 20 ITA-inducible gene list from the lag phase (last column).

TF	Enrichment (%)		Regulated genes	In lag phase list?
	Dataset	Yeastract		
Gcn4p	100	0.20		Y
Gln3p	100	0.50		Y
Met28p	100	0.25	<i>YAT2 YGL117W DUR3 QDR2 DAL2 DAL7 DAL3 DAL5 MEP2 AQR1 ORT1 QDR3</i>	N
Cbf1p	100	0.23		N
Yap1p	100	0.20		Y
Dal80p	58	3.50	<i>YAT2 DUR3 DAL2 DAL7 DAL3 DAL5 MEP2</i>	Y
Gzf3p	66	1.70	<i>YAT2 DUR3 DAL2 DAL7 DAL3 DAL5 MEP2 AQR1</i>	Y
Dal81p	58	1.86	<i>DAL2 DAL7 DAL3 DAL5 MEP2 AQR1 QDR3</i>	Y
Leu3p	58	1.31	<i>DUR3 QDR2 DAL2 DAL7 DAL3 MEP2 ORT1</i>	Y
Bas1p	91	0.37	<i>YAT2 YGL117W DUR3 QDR2 DAL7 DAL3 DAL5 MEP2 AQR1 ORT1 QDR3</i>	Y
Ixr1p	83.3	0.41	<i>YAT2 YGL117W DUR3 QDR2 DAL2 DAL5 MEP2 AQR1 ORT1 QDR3</i>	N

6.4.3. Experimental confirmation of the results obtained in the RNA-seq of yeast ITA-stressed cells

The results obtained in the RNA-seq concerning *S. cerevisiae* transcriptomic response to ITA resulted in the identification of a set of genes that were considered as being more interesting candidates to be used in the design of an ITA biosensor. Five genes (*MEP2*, *DAL5*, *DAL2*, *DAL3* and *TMT1*) were considered more interesting to be used as potential sensors of ITA because they are up-regulated in the two times points analysed (lag and exponential), were induced in a linear manner in the two concentrations of ITA tested and are the top 4 genes with the highest activation slopes (Table 6.2). Although the *TMT1* gene did not show the best slope or linearity and was not up-regulated in exponentially-growing cells in the presence of ITA, it was also studied because of the interest in characterizing the transcriptional regulation of this gene, which is less studied and may impact ITA production in *S. cerevisiae*, as occurred in *A. niger* [31]. Further validation of these results by RT-qPCR was performed. Besides confirming the induction in the expression of these genes by ITA it was also examined whether their up-regulation was prompted by Leu3, considering the hypothesized activation of this regulator by ITA. As such, the expression of these genes was monitored in By4741 cells and in a derived Δ leu3 mutant strain. At the sub-inhibitory ITA concentration of 175 mM ITA only *DAL3* was not up-regulated, while *DAL2*, *DAL5*, *MEP2* and *TMT1* expression increased 2-, 4-, 7- and 3.5- fold, compared with the transcript levels registered in unstressed cells (Figure 6.6). In the case of *DAL2* and *DAL5* the expression of these genes in the Δ leu3 mutant strain was lower than in the wt, indicating that Leu3 may regulate the basal expression. There is still an upregulation of *DAL5* and *DAL2* in the Δ leu3 mutant strain when exposed to ITA (although their expression is higher in the wt cells exposed to ITA), suggesting that another regulator may mediate ITA induction. The expression of *MEP2* and *TMT1* appears to be LEU3-independent, since the fold-change induced by ITA is the same in the wt and Δ leu3 strain.

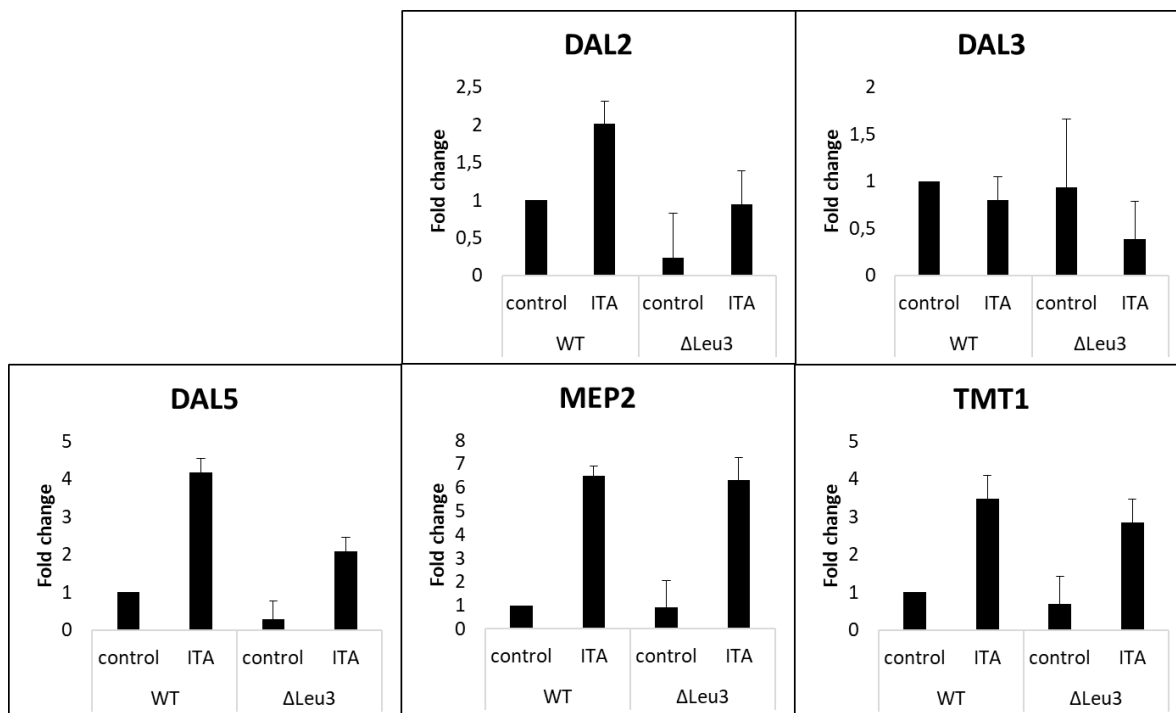


Figure 6.6. Fold-change of 5 candidate ITA-inducible genes in the presence of 175 mM ITA. Experiments were performed in By4741 and By4741Δleu3. The $\Delta\Delta CT$ is normalized for the wt cultures in media without ITA (control).

From this data two genes were discarded for further study: *DAL3*, since it was not upregulated in 175 mM of ITA in the wt or $\Delta leu3$ strain and *DAL2*, since the data suggests Leu3 regulates the basal expression but not ITA induction. While *DAL5* also meets this criterium, this gene had the highest activation slope according to the RNA-seq data (see Table 6.2 in section above), making it an interesting candidate to be pursued. Thus, a subsequent RT-qPCR analysis of the expression of *DAL5*, *MEP2* and *TMT1* was made to account also for the inhibitory 290 mM ITA concentration and to consider another potential regulator that may be behind ITA induction of these genes. Since the both *DAL5* and *MEP2* are annotated to be regulated by Dal81, one of the TFs regulating NCR and allantoin catabolism, RT-qPCRs for these genes were performed not only in a wt and $\Delta leu3$ background, but also in a $\Delta dal81$ background. The expression of *TMT1* was also studied to verify if Dal81 may also be regulating the expression of this gene.

At the higher concentration of 290 mM ITA the ITA-induced activation of *MEP2*, *TMT1* and *DAL5* was considerably higher than the one registered at 175 mM (Figure 6.7). Notably, the induction of these genes prompted by 290 mM ITA was significantly reduced in the $\Delta dal81$ and $\Delta leu3$ mutants, confirming their hypothesized role as activators of these genes. The fact that disproportionally high fold-changes are achieved at 290 mM

than at 175 mM for *MEP2*, *DAL5* and *TMT1* explains why the Leu3 dependence could not be observed before when screening only for the 175 mM ITA concentration. In this case, it appears that the lack of TF only has an effect on the expression of these three genes at high inducer concentrations. This data suggests that indeed Leu3 and Dal81 are mediating the ITA-induction of these genes but may be acting synergistically among themselves or with other regulators. This is relevant to understand how a future ITA sensing system based on, for example the MEP2 promoter and the Leu3 regulator might be susceptible to cross activation by Dal81. Considering the complexity of the NCR-mediated regulation of the DAL and MEP genes, it is likely that other TFs are binding the cis-elements of these genes and affecting gene expression [43].

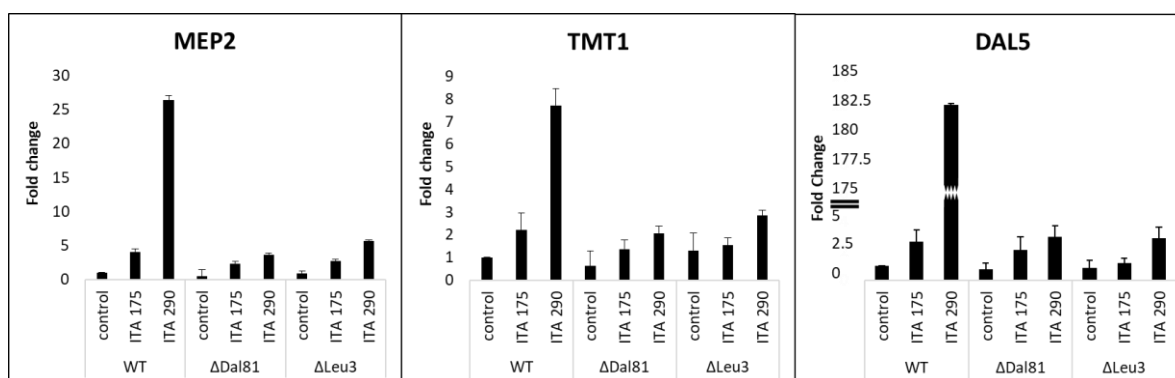


Figure 6.7. Fold-change of 3 candidate ITA-inducible genes in 175 mM and 290 mM ITA. Experiments were performed in By4741 and By4741Δleu3 and By4741Δdal81. The $\Delta\Delta CT$ is normalized for the wt cultures in media without ITA (control). Note the break in the Y axis of DAL% plot.

The dose-response curve for the expression of MEP2, TMT1 and DAL5 genes is shown in Figure 6.8, using the data obtained in the RNA-seq and in the RT-qPCR experiments performed. For the three genes it is evident the linearity of the dose-response curve, although the fit is better for the RNA-seq data than for the RT-qPCR validation studies. This is particularly evident for DAL5 expression (RSQ=0.611), which according to the RT-qPCR results can reach up-regulations of 180 on 290 mM ITA. For TMT1 and MEP2 the RSQ of linear fitting is of 0.791 and 0.728. The 175 mM and 290 mM concentration of ITA for these studies were chosen to guarantee that they covered a sub-inhibitory and inhibitory scenario. While the linearity of a sensor system is a measure of sensor performance, it is defined within an operational range; it is possible that these concentrations are outside of the operational range for these endogenous ITA sensors, which is very likely since the operational range of a metabolite sensor usually doesn't include null concentrations of the metabolite. Further studies using other concentrations

of ITA will be required in order to have a better assessment on what is the dose-response curve for the expression of these genes and to find the points where linearity is no longer observed.

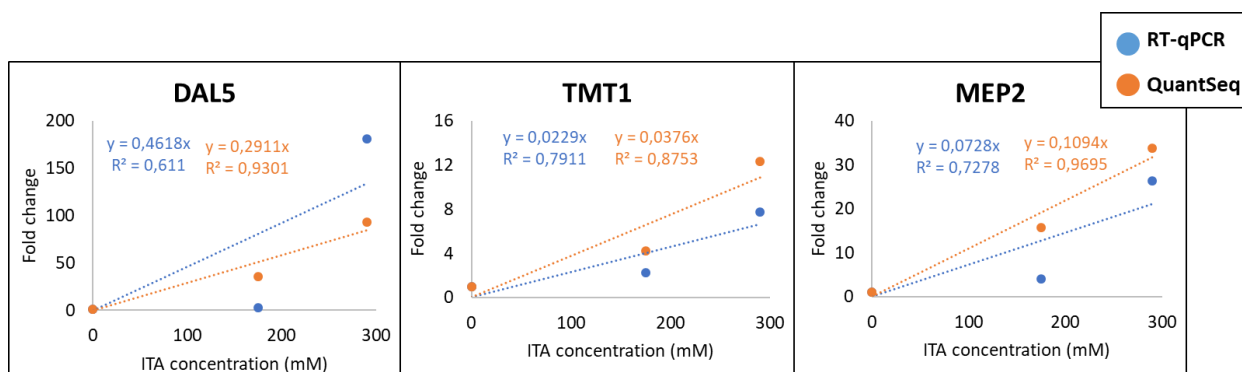


Figure 6.8. Dose-response curve of *MEP2*, *TMT1* and *DAL5* in the wt background. Data from RT-qPCR (blue) and QuantSeq (orange). The linear trendlines were calculated with slope and RSQ displayed in each chart in the corresponding color.

Considering the end goal of designing an ITA sensor and after having confirmed the positive insights of the possibility of using three genes (*DAL5*, *TMT1* and *MEP2*) as ITA responsive as well as *Leu3* or *Dal81* as putative sensing regulators, the next step would be to clone the promoters of these three genes upstream of a reporter gene (e.g. GFP, *lacZ*) to obtain a reporter system. This would facilitate the study of different ITA concentrations, time dynamics and even the study of the impact of different regulators, simply by inserting the reporter plasmid promoter (e.g., *prMEP2*-GFP) in KO mutants for the regulators that remain to be studied. It will also be required to study whether other TFs identified as putative regulators of the candidate genes, *Gln3*, *Gzf3*, *Dal80*, *Dal82*, also play a role in the observed ITA-induced activation.

One concern when designing TF-based sensors that rely on the native regulatory network is the interference of other host regulators in the response, which might render the reporter system unspecific, since it would be not only responsive to ITA but also to other cellular stimuli. In this case, the most likely source of concern would be that conditions of nitrogen starvation might activate the sensor system. In large scale screenings of mutant collections, a ITA sensor with this design is still useful to come up with a smaller list of potential ITA overproducers (based on a higher sensor output of these strains), although this list of strains would need to be validated to confirm that higher output was due to higher ITA production and not higher 2-IPM accumulation, due to nitrogen starvation. Another possibility to overcome this limitation is to evolve the specificity of the TF in question. The data from this work points to the possibility of a

direct interaction of ITA with Leu3, activating the regulator. If this is true, it would be possible to evolve Leu3 to become more sensitive to ITA and less sensitive to its native inducer, 2-IPM. With the aid of a fluorescent reporter system by creating a library of Leu3 variants (for example by error-prone PCR) it would be possible to perform a screening of the regulator variants in a microplate screening setting or a flow cytometer one to find variants with increased specificity for ITA and decreased 2-IPM activation. In the past strategies like this have been used to evolve regulator recognition, such as the case of the evolution of the *P. putida* regulator NahR to become sensitive to benzoic acid [44].

However, before planning the evolution of a regulator, it would be necessary to confirm the role of the regulator in ITA induction of the target genes. In the case of Leu3, the effect of ITA on Leu3 binding to its target genes or UAS^{LEU} could be tested by use of an electromobility shift assay, as previously reported for 2-IPM, where an 2-IPM dose-dependent response curve was obtained [45]. Additionally, the study of characterized Leu3 mutants that have constitutive masking of the activation domain (are constitutive repressors) [46] or of mutants that have inducer-independent unmasking of the activation domain (are constitutive activators)[34] and their activation responses to ITA might also provide indirect proofs of ITA interaction. In this context, it is relevant to note that LEU3 acts as an activator in the presence of 2-IPM and as repressor in its absence [47] so a similar mechanism would likely apply to ITA, if there is in fact a direct interaction.

In the case no direct regulator-ITA interaction can be found, the specificity of the ITA sensing system is likely compromised. Although this might become incompatible with sensor applications where the sensing strain and the ITA producing strain are the same (such as directed evolution of strains to improve ITA production), it still raises the possibility of using the system in setups where a sensing strain is used to screen a collection of supernatants from ITA fermentations. This approach, previously called “one pot two strains system” has been used to screen yeast and bacterial variant libraries for improved glucaric acid production, by use of an *E. coli* glucaric acid sensing strain [48]. A similar approach was employed to sense small to medium chain fatty acid in fermentation supernatants by use of yeast strains with a War1-based fluorescence sensor [16]. In the case of ITA production, this might be applied to the screening of the yeast disruptome collection for improved ITA production KO mutants or even to the screening of a *P. kudriavzevii* library, namely to find variants for the mttA mitochondrial tricarboxylate transporter with improved cis-aconitate transport to the cytosol.

6.4.4. Overall transcriptomic response to ITA exposure and physiological significance

While the inspection of ITA tolerance mechanisms in yeast is out of the scope of this work, it would be interesting to understand the overall physiological response to ITA and unravel the role of nitrogen starvation in ITA exposure. This may aid the design and understanding of an ITA sensor based on yeast regulators and specifically help understand the crosstalk that can be expected in such sensor. Furthermore, the study of *S. cerevisiae*'s response to ITA is particularly interesting considering the doubts that are involved in ITA production and degradation in fungi and bacteria and why these pathways exist in nature.

The analysis in the previous subsections had the objective of finding the genes with the highest inductions to ITA. In this subsection the overall response is analyzed to find the main processes that yeast activates when exposed to ITA. To obtain a wider perspective of the activation of genes in the presence of ITA, the 43 genes that were upregulated in the three conditions studied here were first manually matched to functional categories (an automatized GO enrichment term analysis is depicted in Annex 8.5 - Figure 8.5). Note that the results from lag phase and exponential phase are overlapped to discard changes that are expected to occur exclusively during the stress adaptation phase. The complete list of the 43 genes is present in Table 6.7. Two categories that are more specific can be seen in this dataset: genes that are involved in the NCR response and genes that encode proteins in the metabolism of methionine and sulfur. Concerning the eight NCR-sensitive genes in the dataset, these encode enzymes and transporters required for allantoin and urea utilization (*DAL2/3/7* and *DUR1,2/3*), an ammonium permease (*MEP2*) and a vacuolar carboxypeptidase (*CPS1*). The upregulation of these genes even in the exponential phase of cells grown in the presence of ITA indicates that the nitrogen starvation response is not transient. Furthermore, the upregulation of genes required for the synthesis of methionine and other aminoacids and for aminoacid transport further strengthens the hypothesis that ITA induces also an aminoacid starvation response in yeast. When analyzing the STRING network generated for the 43 gene dataset (Annex 8.5- Figure 8.6), two main interaction clusters are observed: one is centered around the citrate synthase encoding *CIT1/2*, with interactions with the NCR-sensitive genes (*DAL2/3/7/5* and *MEP2*, *DUR1,2/3*) and the other cluster is centered around the genes involved in methionine biosynthesis (*MET*). This highlights the importance of understanding the role of nitrogen starvation and methionine synthesis in the ITA response.

Chapter 6– Searching for the genetic parts to develop a yeast-based biosensor for itaconic acid

Table 6.7. List of genes upregulated in all three conditions: the lag phase (175 mM and 290 mM ITA) and in the exponential phase (290 mM ITA). The genes were manually grouped by functional category.

Category	Gene	Description	Category	Gene	Description
NCR sensitive	MEP2	Ammonium permease	Mitochondrial processes	PDH1	Putative 2-methylcitrate dehydratase
	DUR1,2	Urea amidolyase		IDP1	Mitochondrial NADP-specific isocitrate dehydrogenase
	DAL5	Allantoate permease		CIT1	Citrate synthase
	CPS1	Vacuolar carboxypeptidase		CIT2	Citrate synthase
	DAL2	Allantoicase		LSC2	Beta subunit of succinyl-CoA ligase
	DAL7	Malate synthase	YAT2	Carnitine acetyltransferase	
	DAL3	Ureidoglycolate hydrolase	Aminoacid synthesis (serine, glycine and arginine)	SER33	3-phosphoglycerate dehydrogenase
	DUR3	Plasma membrane transporter for both urea and polyamines	SER3	3-phosphoglycerate dehydrogenase	
			ORT1	Ornithine transporter of the mitochondrial inner membrane	
Methionine & sulfur metabolism	MET16	3'-phosphoadenylsulfate reductase	Transporters (plasma membrane and mitochondrial)	QDR2	MFS Multidrug transporter
	MET10	Subunit alpha of assimilatory sulfite reductase		QDR3	MFS Multidrug transporter of the major facilitator superfamily
	SUL1	High affinity sulfate permease		AQR1	Plasma membrane MFS multidrug transporter
	SAM3	High-affinity S-adenosylmethionine permease		DIC1	Mitochondrial dicarboxylate carrier
	MET1	S-adenosyl-L-methionine uroporphyrinogen III transmethylase		ZRT1	High-affinity zinc transporter of the plasma membrane
	MET5	Sulfite reductase beta subunit		APE1	Vacuolar aminopeptidase
	MET2	L-homoserine-O-acetyltransferase		GLO2	Cytoplasmic glyoxalase II
	Aminoacid transport	DIP5	Dicarboxylic amino acid permease	Others	SNR31
BAP3		Amino acid permease	CMC4		Protein that localizes to the mitochondrial intermembrane space via the Mia40p-Erv1p system
GAP1		General amino acid permease;	RAD59		Protein involved in the repair of double-strand breaks in DNA during vegetative growth
PTR2		Integral membrane peptide transporter	DLD3		2-hydroxyglutarate transhydrogenase, and minor D-lactate dehydrogenase
Unknown function	YGL117W, ICY1, SPO24, YGR125W				

In the 43 gene set one category that was also found was that of transport, with three genes encoding MFS transporters (*QDR2*, *QDR3* and *AQR1*), one encoding a mitochondrial dicarboxylate transporter (*DIC1*) and one encoding zinc transporter (*ZRT1*). *QDR2* and *QDR3* had been previously identified as determinants of *S. cerevisiae* tolerance to ITA, with *QDR3* likely catalyzing the extrusion of ITA from the cytoplasm [28], [49]. The *AQR1* transporter has been found to be required for tolerance to short-chain monocarboxylic acids [50]. These findings point to the hypothesis of the three MFS transporters being upregulated in a cellular attempt to export ITA, since this is a common strategy for the tolerance to weak acids [38]. In this work, it is demonstrated that the upregulation is not transient, since it is also observed during the exponential phase. The finding of the upregulation of the Dic1 mitochondrial transporter here is also interesting, considering that no ITA mitochondrial transporter has been reported so far. However, it is likely that it exists at least in *Aspergilli*, since the degradation pathway of ITA in *A. niger* and *A. terreus* is predicted to be mitochondrial and its biosynthesis in *A. terreus* is cytosolic [30], [31]. The Dic1 transporter is reported to catalyze the import of dicarboxylic acids into the mitochondrial matrix, namely malonate, at the cost of export of phosphate, malate and succinate [51].

In the mitochondrial processes' category, four of the upregulated genes participate in the TCA cycle: *LSC2*, encoding a subunit of succinyl-coA, *IDP1*, encoding an isocitrate dehydrogenase, and *CIT1* and *CIT2*, encoding citrate synthases. Although *S. cerevisiae* Tmt1 methyltransferase has a higher affinity for endogenous trans-aconitate than for ITA, it is possible that ITA competition for this enzyme leads to the accumulation of toxic trans-aconitate, causing inhibition of aconitase. Furthermore, ITA is an inhibitor of microbial isocitrate lyase, although it is unknown if the *S. cerevisiae* enzyme is sensitive to ITA. The inhibition of these two enzymes participating in the TCA and glyoxylate cycle may be relevant for the ITA-induced upregulation of three TCA cycle enzymes. However, the upregulation of *TMT1* expression might counterbalance competition between ITA and trans-aconitate.

The category of aminoacid transport is also significant in the 43 gene dataset, including 3 genes encoding aminoacid permeases (*DIP5*, *BAP3* and *GAP1*) and a peptide transporter (*PTR2*). These genes are known to be activated in response to extracellular amino acid signals, mediated by the Ssy1 pathway, in a manner that is dependent on Gcn4. Furthermore, Dal81 and Leu3 also regulate the transcription of *BAP3* and *GAT1*, since these genes are also nitrogen-regulated. Interestingly, it has been shown that leucine is a potent repressor of the allantoin catabolic genes, in a manner that is dependent on the Ssy

signaling pathway and on Leu3 [51]. Thus, the ITA induced activation of these genes is complex and may be mediated not only by Leu3, but also by Ssy and the NCR regulators.

In what concern the Leu3 hypothesis, an activation of this regulator would expectedly cause the upregulation of the leucine biosynthetic genes: *LEU4* and *LEU9* (encoding the major and minor IPM synthase), *LEU1* (encoding 2-IPM isomerase), *LEU2* (encoding 3-IPM dehydrogenase) and *BAT1* and *BAT2* (encoding branched-chain amino acid aminotransferases). Indeed, this was observed in this dataset for all leucine biosynthesis genes except *LEU2*, since the strain used in this work is knocked-out for this gene. A model of the possible interactions between Leu3, the leucine biosynthetic genes, ITA and Tmt1 is presented in Figure 6.9.

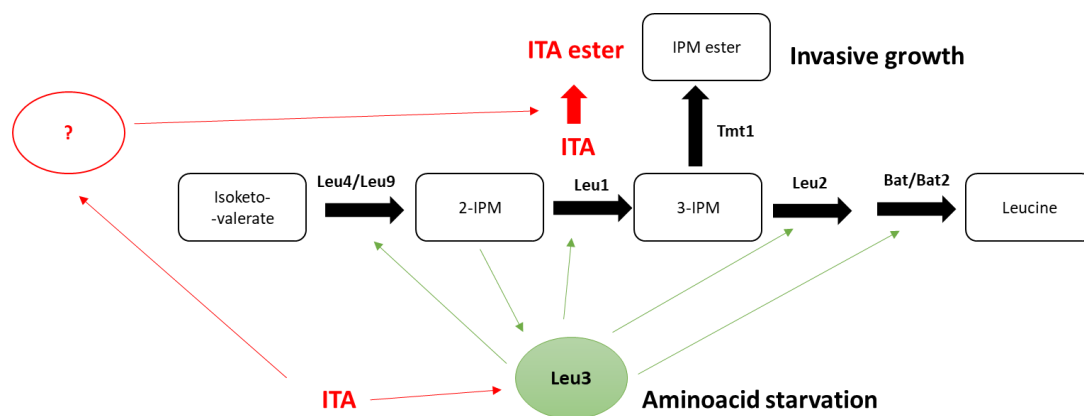


Figure 6.9. Model of Leu3 regulation of leucine biosynthetic genes. The green arrows represent positive regulations during the leucine starvation response, where 2-IPM activates Leu3, which activates the transcription of the leucine biosynthetic genes. The positive interactions during ITA exposure suggested from the data in this work are represented in red arrows, where ITA activates LEU3 and results in the increase of Tmt1 transcription, either through Leu3 activation or another regulator. Tmt1 also methylates ITA to its ester.

In this work it was found that ITA induces a de-repression of NCR in *S. cerevisiae*, where the main features are the upregulation of ammonium, urea and allantoin permeases and of enzymes required for allantoin utilization. Allantoin is a nitrogen reserve molecule, stored in the vacuole and in conditions where there is depletion of good nitrogen sources (ammonia, glutamine and asparagine), the *DAL* and *DUR* genes are upregulated to enable the utilization of this nitrogen reserve molecule and of environmental urea [43] (Figure 6.10A). The upregulation of amino acid uptake and biosynthesis genes (or increase in the

expression of the corresponding enzymes) has been reported in yeast cells exposed to acetic acid [52], sorbic acid [53] and 2,4-D [54],[55], presumably through the activation of the TOR pathway (Target-of-Rapamycin). Indeed, yeast cells exposed to acetic and 2,4-dichlorophenoxyacetic acid (2,4-D) have lower intracellular amino acid concentrations than unexposed cells[52], [54], implying that weak acid exposure may cause amino acid starvation. However, while the upregulation of aminoacid metabolism observed for acetic [52] and sorbic acid [53] is general, the response observed in this work is highly specific for genes regulated by the NCR, including allantoin/allantoate utilization genes, which has only been reported for 2,4-D (where the expression of *DAL2/3/7/5* and *UGA2/4* was upregulated during the lag phase). As one of the mechanisms of weak acid toxicity is the dissipation of the proton gradient across the membrane, it has been hypothesized that this may result in an inability to uptake amino acids, thus resulting in amino acid and/or nitrogen starvation [55]. In this section it was analyzed a group of genes that was upregulated both during the exponential and the lag phase, to rule out general stress adaptation responses (visible in the lag phase). However, it is possible that the nitrogen starvation response observed for ITA (and for other weak acids[54]) is not transient.

If ITA exposed cells actually experience nitrogen depletion due to the acid's toxicity mechanism or if they upregulate an NCR response due to other regulatory mechanisms is unknown. One hypothesis raised here is the direct interaction of ITA with Leu3, activating the *DAL* and *DUR* genes. However, LEU3 direct gene upregulation only explains part of the results obtained here (Figure 6.10B). Indeed, other nitrogen and aminoacid-regulating TFs are likely to be at play. It is also interesting to verify that of the complete NCR set upregulated during nitrogen starvation, *DAL1* and *DAL4*, encoding an allantoinase and an allantoin transporter, were only upregulated in the lag phase with 290 mM ITA.

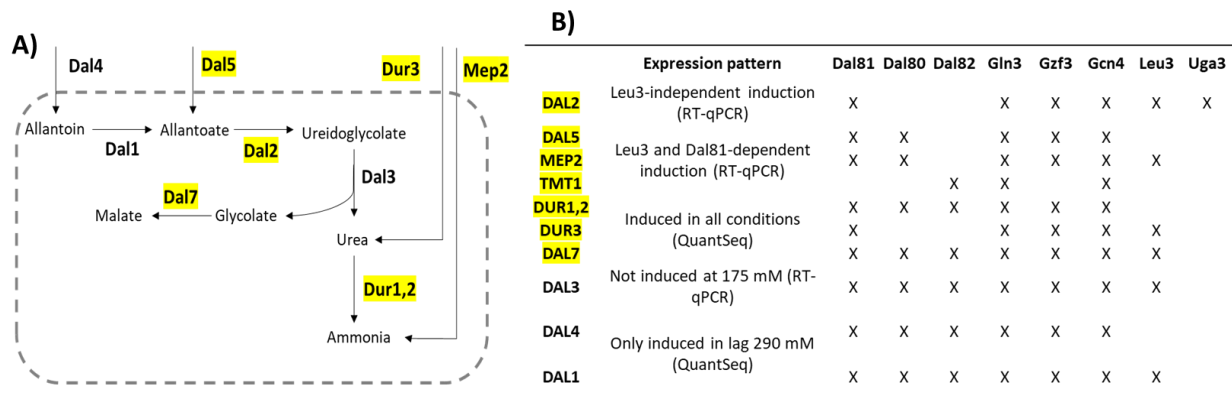


Figure 6.10. Allantoin catabolism genes induced by ITA. A) Allantoin and ammonia utilization pathway that is activated in nitrogen depletion conditions; genes upregulated in all ITA exposure conditions are marked in yellow. B) NCR-sensitive genes and their documented regulator, from the Yeasttract database. Only TFs reported to act in the NCR and amino acid starvation response are represented here.

While the link between ITA and nitrogen starvation in *S. cerevisiae* is not clear, there is an established relationship in the literature between ITA synthesis in natural producers and nitrogen depletion. In both *U. maydis* and *P. antarctica* ITA production starts when nitrogen is depleted [56],[57]. This has been mainly attributed to the competition between ITA production and growth, causing the production phase to start only when the growth-limiting nutrient is depleted, in this case, nitrogen. Indeed, conditions for ITA production in *A. terreus* are usually phosphate-limiting and not nitrogen-limiting. However, in *U. maydis* the promoters of the ITA clusters are activated upon nitrogen depletion, including the promoters of *MTT1* and *TAD1* [58]. This finding demonstrates that at least in this host the signal for ITA production is embedded in the regulatory circuit and is not only a metabolic one.

Another host where the regulatory response to ITA has been studied is *A. niger*, where the analysis of the transcriptomic response enabled the identification of the genes encoding the ITA degradation pathway in this host [31]. In this fungus it was found that in high ITA producing conditions there is a down-regulation of genes involved in ammonium transport and utilization, in opposition to what was found here. In this host the most relevant ITA response was the up-regulation of genes required for the degradation of ITA. In *A. niger* the main degradation pathway encompasses three enzymatic steps: IctA mediated formation of itaconyl-coA, which is dehydrated to citramalyl-coA by IchA and Ccl catalyzed conversion of citramalyl-coA to acetyl-coA and pyruvate. This pathway is functionally identical to the ones identified in *A. terreus*,

Y. pestis and *P. putida* [30], [59]. The up-regulation of an ITA degradation pathway in *S. cerevisiae* is unlikely, since this yeast cannot use ITA (data not shown). Additionally, the up-regulation of *A. niger* TmtA in response to ITA, a homologue of *S. cerevisiae* Tmt1, was identified as another bioconversion of ITA in this host, where it is methylated. In *S. cerevisiae* it is expected that upon ITA exposure there is also an accumulation of methyl-ITA, since *TMT1* is up-regulated; however, the importance of this methyl product in *S. cerevisiae* and *A. niger* is unknown. The induction of *TMT1* might also be relevant since it would expectedly result in a higher conversion of 3-IPM to its methyl ester, thus signaling invasive growth [29].

If indeed ITA interacts with Leu3, the interaction may be coincidental, or it may be a conserved response in this yeast to ITA. In *A. niger* the ITA response is characterized by the upregulation of genes for ITA catabolism, which likely also occurs in pathogenic bacteria. In *Y. pestis* and *P. putida* these genes are pathogenicity determinants, since ITA is produced in the macrophages and its degradation is hypothesized to aid these bacteria in evading the hosts defensive mechanisms. While in *S. cerevisiae* there would be no evident function of an embedded response to ITA, this is an organism with response mechanisms to a large number of environmental perturbations due to its nomadic lifestyle and diverse niches [60]. Thus, it is possible that ITA is one of the organic acids that *S. cerevisiae* finds in its natural habitats (*A. terreus* is found mainly in soil, compost and dust). How the concerted upregulation of nitrogen and aminoacid producing pathways would aid in this ITA response is unclear. One possibility is that the induction of *TMT1*, which is expected to cause an accumulation of methyl-3-IPM, a signaling molecule for invasive growth during nitrogen depletion conditions [29], may aid yeast in evading ITA-rich environments. Thus, it would be interesting to verify if ITA in the media can induce invasive growth in *S. cerevisiae* strain that maintains this ability, which could not be observed here since the By4741 is devoid of the *FLO8* gene.

6.5. Conclusions and Future Work

In this chapter the transcriptomic response to ITA in the lag and exponential phase was analyzed to find genetic parts belonging to the native yeast regulatory network that may be used in the design of a TF-based ITA sensor. The analysis resulted in a set of 12 candidate genes that display an ITA-inducible behavior. Validation studies with RT-qPCR resulted in the selection of three candidate genes for an ITA biosensor design:

DAL5, *MEP2* and *TMT1*. For these three genes, the ITA-induction displayed a degree of dependence on both Dal81 and Leu3, raising the possibility of these two regulators and possibly others, namely NCR ones, contributing to the expression of these genes in the presence of ITA. Thus, further studies are required to better understand the ITA induction of these genes. It is here determined as a priority the utilization of a reporter gene system for the candidate ITA-inducible promoters to pursue further studies, facilitating the characterization of ITA-based upregulation in terms of time dynamics, range and involved regulators. It would also be relevant for the understanding of the sensing system to understand if ITA affects directly the ability of Leu3 to bind directly to the four target promoters selected here. If this is the case, the specificity of the regulator may be evolved to decrease the native crosstalk derived from 2-IPM activation.

The analysis of the upregulated genes in the exponential and lag phase points to ITA inducing an NCR response in *S. cerevisiae*, where the main identified TFs are Dal81, Dal80, Gzf3 and Gln3. While the significance of this ITA response could not be understood, the upregulation of *TMT1* in the context of NCR may mediate the induction of invasive growth.

6.6. References

- [1] T. Werpy and G. Petersen, “Top Value Added Chemicals from Biomass,” 2004. doi: 10.2172/926125.
- [2] E. Geiser *et al.*, “*Ustilago maydis* produces itaconic acid via the unusual intermediate trans-aconitate,” *Microb. Biotechnol.*, vol. 9, no. 1, pp. 116–126, Jan. 2016, doi: 10.1111/1751-7915.12329.
- [3] C. S. K. Reddy and R. P. Singh, “Enhanced production of itaconic acid from corn starch and market refuse fruits by genetically manipulated *Aspergillus terreus* SKR10,” *Bioresour. Technol.*, vol. 85, no. 1, pp. 69–71, Oct. 2002, doi: 10.1016/S0960-8524(02)00075-5.
- [4] C. L. Strelko *et al.*, “Itaconic acid is a mammalian metabolite induced during macrophage activation,” *J. Am. Chem. Soc.*, vol. 133, no. 41, pp. 16386–9, Oct. 2011, doi: 10.1021/ja2070889.
- [5] H. Hajian, W. Mohtar, and W. Yusoff, “Itaconic Acid Production by Microorganisms: A Review,” *Curr. Res. J. Biol. Sci.*, vol. 7, no. 2, p. 3742, 2015.
- [6] E. Geiser *et al.*, “Genetic and biochemical insights into the itaconate pathway of *Ustilago maydis* enable enhanced production,” *Metab. Eng.*, vol. 38, pp. 427–435, Nov. 2016, doi: 10.1016/j.ymben.2016.10.006.
- [7] M. Blumhoff, M. G. Steiger, D. Mattanovich, and M. Sauer, “Targeting enzymes to the right compartment: Metabolic engineering for itaconic acid production by *Aspergillus niger*,” *Metab. Eng.*, vol. 19, pp. 26–32, 2013, doi: 10.1016/J.YMBEN.2013.05.003.
- [8] K. N. T. Tran, S. Somasundaram, G. T. Eom, and S. H. Hong, “Efficient Itaconic acid production via protein–protein scaffold introduction between GltA, AcnA, and CadA in recombinant *Escherichia coli*,” *Biotechnol. Prog.*, vol. 35, no. 3, 2019, doi: 10.1002/btpr.2799.
- [9] E. M. Young *et al.*, “Iterative algorithm-guided design of massive strain libraries, applied to itaconic acid production in yeast,” *Metab. Eng.*, vol. 48, pp. 33–43, Jul. 2018, doi: 10.1016/j.ymben.2018.05.002.
- [10] C. Zhao *et al.*, “Enhanced itaconic acid production in *Yarrowia lipolytica* via heterologous expression of a mitochondrial transporter MTT,” *Appl. Microbiol. Biotechnol.*, vol. 103, no. 5, pp. 2181–2192, Mar. 2019, doi: 10.1007/s00253-019-09627-z.
- [11] W. Sun *et al.*, “Metabolic engineering of an acid-tolerant yeast strain *Pichia kudriavzevii* for itaconic acid production,” *Metab. Eng. Commun.*, vol. 10, no. October 2019, p. e00124, 2020, doi: 10.1016/j.mec.2020.e00124.
- [12] L. Eggeling, M. Bott, and J. Marienhagen, “Novel screening methods-biosensors,” *Curr. Opin. Biotechnol.*, vol. 35, no. ii, pp. 30–36, 2015, doi: 10.1016/j.copbio.2014.12.021.
- [13] S. Li, T. Si, M. Wang, and H. Zhao, “Development of a Synthetic Malonyl-CoA Sensor in *Saccharomyces cerevisiae* for Intracellular Metabolite Monitoring and Genetic Screening,” *ACS Synth. Biol.*, vol. 4, no. 12, pp. 1308–1315, 2015, doi: 10.1021/acssynbio.5b00069.
- [14] S.-Y. Tang and P. C. Cirino, “Design and Application of a Mevalonate-Responsive Regulatory Protein,” *Angew. Chemie Int. Ed.*, vol. 50, no. 5, pp. 1084–1086, Feb.

- 2011, doi: 10.1002/anie.201006083.
- [15] J. Feng *et al.*, “A general strategy to construct small molecule biosensors in eukaryotes,” *Elife*, vol. 4, no. DECEMBER2015, Dec. 2015, doi: 10.7554/eLife.10606.
- [16] L. Baumann, A. S. Rajkumar, J. P. Morrissey, E. Boles, and M. Oreb, “A Yeast-Based Biosensor for Screening of Short- and Medium-Chain Fatty Acid Production,” *ACS Synth. Biol.*, 2018, doi: 10.1021/acssynbio.8b00309.
- [17] R. H. Dahl *et al.*, “Engineering dynamic pathway regulation using stress-response promoters,” *Nat. Biotechnol.*, vol. 31, no. 11, pp. 1039–46, 2013, doi: 10.1038/nbt.2689.
- [18] S. J. Doong, A. Gupta, and K. L. J. Prather, “Layered dynamic regulation for improving metabolic pathway productivity in *Escherichia coli*,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 115, no. 12, pp. 2964–2969, Mar. 2018, doi: 10.1073/pnas.1716920115.
- [19] L. Lindenburg and M. Merckx, “Engineering genetically encoded FRET sensors,” *Sensors (Switzerland)*, vol. 14, no. 7. MDPI AG, pp. 11691–11713, Jul. 02, 2014, doi: 10.3390/s140711691.
- [20] J. A. Gredell, C. S. Frei, and P. C. Cirino, “Protein and RNA engineering to customize microbial molecular reporting,” *Biotechnol. J.*, vol. 7, no. 4, pp. 477–499, 2012, doi: 10.1002/biot.201100266.
- [21] X. Wan, M. Marsafari, and P. Xu, “Engineering metabolite - responsive transcriptional factors to sense small molecules in eukaryotes : current state and perspectives,” *Microb. Cell Fact.*, pp. 1–13, 2019, doi: 10.1186/s12934-019-1111-3.
- [22] R. Mahr and J. Frunzke, “Transcription factor-based biosensors in biotechnology : current state and future prospects,” pp. 79–90, 2016, doi: 10.1007/s00253-015-7090-3.
- [23] E. K. R. Hanco, N. P. Minton, and N. Malys, “A Transcription Factor-Based Biosensor for Detection of Itaconic Acid,” *ACS Synth. Biol.*, vol. 7, no. 5, pp. 1436–1446, 2018, doi: 10.1021/acssynbio.8b00057.
- [24] M. L. Skjoedt *et al.*, “Engineering prokaryotic transcriptional activators as metabolite biosensors in yeast,” *Nat. Chem. Biol.*, vol. 12, no. 11, pp. 951–958, 2016, doi: 10.1038/nchembio.2177.
- [25] P. T. Monteiro *et al.*, “YEASTRACT+: A portal for cross-species comparative genomics of transcription regulation in yeasts,” *Nucleic Acids Res.*, vol. 48, no. D1, pp. D642–D649, 2020, doi: 10.1093/nar/gkz859.
- [26] S. X. Ge, D. Jung, and R. Yao, “ShinyGO: a graphical gene-set enrichment tool for animals and plants,” *Bioinformatics*, vol. 36, no. 8, pp. 2628–2629, 2020, doi: 10.1093/bioinformatics/btz931.
- [27] D. Szklarczyk *et al.*, “STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets,” *Nucleic Acids Res.*, vol. 47, no. D1, pp. D607–D613, 2019, doi: 10.1093/nar/gky1131.
- [28] M. R. Moita, “Towards the production of levulinic and itaconic acids in *Saccharomyces cerevisiae*: a contribution for understanding the molecular mechanisms of toxicity of these acids in producing cells (Master Thesis in

- Biotechnology),” IST, 2013.
- [29] D. S. Dumlao, N. Hertz, and S. Clarke, “Secreted 3-isopropylmalate methyl ester signals invasive growth during amino acid starvation in *Saccharomyces cerevisiae*,” *Biochemistry*, vol. 47, no. 2, pp. 698–709, 2008, doi: 10.1021/bi7018157.
 - [30] M. Chen, X. Huang, C. Zhong, J. Li, and X. Lu, “Identification of an itaconic acid degrading pathway in itaconic acid producing *Aspergillus terreus*,” *Appl. Microbiol. Biotechnol.*, vol. 100, no. 17, pp. 7541–7548, Sep. 2016, doi: 10.1007/s00253-016-7554-0.
 - [31] A. H. Hossain, A. Ter Beek, and P. J. Punt, “Itaconic acid degradation in *Aspergillus niger*: the role of unexpected bioconversion pathways,” *Fungal Biol. Biotechnol.*, vol. 6, no. 1, pp. 1–16, 2019, doi: 10.1186/s40694-018-0062-5.
 - [32] G. A. Marzluf, “Genetic regulation of nitrogen metabolism in the fungi,” *Microbiol. Mol. Biol. Rev.*, vol. 61, no. 1, pp. 17–32, 1997, doi: 10.1128/.61.1.17-32.1997.
 - [33] P. A. Bricmont, J. R. Daugherty, and T. G. Cooper, “The DAL81 gene product is required for induced expression of two differently regulated nitrogen catabolic genes in *Saccharomyces cerevisiae*,” *Mol. Cell. Biol.*, vol. 11, no. 2, pp. 1161–1166, 1991, doi: 10.1128/mcb.11.2.1161.
 - [34] D. Wang, F. Zheng, S. Holmberg, G. B. Kohlhaw, and G. B. J. B. Chem, “Yeast Transcriptional Regulator Leu3p: self-masking, specificity of masking and evidence for regulation by the intracellular level of Leu3p,” *J. Biol. Chem.*, vol. 274, no. 27, pp. 19017–19024, 1999.
 - [35] H. Cai, D. Dumlao, J. E. Katz, and S. Clarke, “Identification of the gene and characterization of the activity of the trans-aconitate methyltransferase from *saccharomyces cerevisiae*,” *Biochemistry*, vol. 40, no. 45, pp. 13699–13709, 2001, doi: 10.1021/bi011380j.
 - [36] R. A. Crawford and G. D. Pavitt, “Translational regulation in response to stress in *Saccharomyces cerevisiae*,” *Yeast*, vol. 36, no. 1, pp. 5–21, Jan. 2019, doi: 10.1002/yea.3349.
 - [37] M. Sidorova, E. Drobna, V. Dzugasova, I. Hikkel, and J. Subik, “Loss-of-function *pdr3* mutations convert the Pdr3p transcription activator to a protein suppressing multidrug resistance in *Saccharomyces cerevisiae*,” *FEMS Yeast Res.*, vol. 7, no. 2, pp. 254–264, Mar. 2007, doi: 10.1111/j.1567-1364.2006.00174.x.
 - [38] N. P. Mira, M. C. Teixeira, and I. Sa, “Adaptive Response and Tolerance to Weak Acids in *Saccharomyces cerevisiae* : A Genome-Wide View,” vol. 14, no. 5, 2010, doi: 10.1089/omi.2010.0072.
 - [39] L. Ojeda, G. Keller, U. Muhlenhoff, J. C. Rutherford, R. Lill, and D. R. Winge, “Role of glutaredoxin-3 and glutaredoxin-4 in the iron regulation of the Aft1 transcriptional activator in *Saccharomyces cerevisiae*,” *J. Biol. Chem.*, vol. 281, no. 26, pp. 17661–17669, Jun. 2006, doi: 10.1074/jbc.M602165200.
 - [40] M. A. Garcia-Gimeno and K. Struhl, “Aca1 and Aca2, ATF/CREB Activators in *Saccharomyces cerevisiae*, Are Important for Carbon Source Utilization but Not the Response to Stress,” *Mol. Cell. Biol.*, vol. 20, no. 12, pp. 4340–4349, Jun. 2000, doi: 10.1128/mcb.20.12.4340-4349.2000.
 - [41] K. Hellauer, B. Akache, S. MacPherson, E. Sirard, and B. Turcotte, “Zinc cluster

- protein Rdr1p is a transcriptional repressor of the PDR5 gene encoding a multidrug transporter,” *J. Biol. Chem.*, vol. 277, no. 20, pp. 17671–17676, May 2002, doi: 10.1074/jbc.M201637200.
- [42] H. Dumond, N. Danielou, M. Pinto, and M. Bolotin-Fukuhara, “A large-scale study of Yap1p-dependent genes in normal aerobic and H₂O₂-stress conditions: The role of Yap1p in cell proliferation control in yeast,” *Mol. Microbiol.*, vol. 36, no. 4, pp. 830–845, 2000, doi: 10.1046/j.1365-2958.2000.01845.x.
 - [43] T. G. Cooper, “Regulation of Allantoin Catabolism in *Saccharomyces cerevisiae*,” in *The mycota*, Springer, Ed. 1996, pp. 139–171.
 - [44] A. Cebolla, C. Sousa, and V. De Lorenzo, “Effector specificity mutants of the transcriptional activator NahR of naphthalene degrading *Pseudomonas* define protein sites involved in binding of aromatic inducers,” *J. Biol. Chem.*, vol. 272, no. 7, pp. 3986–3992, 1997, doi: 10.1074/jbc.272.7.3986.
 - [45] J. Sze, M. Woontner, J. A. Jaehning, and G. B. Kohlhaw, “In Vitro Transcriptional Activation by a Metabolic Intermediate : Activation by Leu3 Depends on ac-isopropylmalate,” *Science (80-.)*, vol. 1751, no. November, pp. 1143–1145, 1992.
 - [46] D. Wang, Y. Hu, F. Zheng, K. Zhou, and G. B. Kohlhaw, “Evidence that intramolecular interactions are involved in masking the activation domain of transcriptional activator Leu3p,” *J. Biol. Chem.*, vol. 272, no. 31, pp. 19383–19392, Aug. 1997, doi: 10.1074/jbc.272.31.19383.
 - [47] J. Y. Sze, E. Remboutsika, and G. B. Kohlhaw, “Transcriptional regulator Leu3 of *Saccharomyces cerevisiae*: separation of activator and repressor functions.,” *Mol. Cell. Biol.*, vol. 13, no. 9, pp. 5702–5709, Sep. 1993, doi: 10.1128/mcb.13.9.5702.
 - [48] S. Zheng *et al.*, “One-pot two-strain system based on glucaric acid biosensor for rapid screening of myo-inositol oxygenase mutations and glucaric acid production in recombinant cells,” *Metab. Eng.*, vol. 49, no. May, pp. 212–219, 2018, doi: 10.1016/j.ymben.2018.08.005.
 - [49] N. M. Rodrigues, “Metabolic and genetic engineering strategies to explore *Saccharomyces cerevisiae* as a cell factory for the production of itaconic acid,” IST, 2014.
 - [50] S. Tenreiro *et al.*, “AQR1 gene (ORF YNL065w) encodes a plasma membrane transporter of the major facilitator superfamily that confers resistance to short-chain monocarboxylic acids and quinidine in *saccharomyces cerevisiae*,” *Biochem. Biophys. Res. Commun.*, vol. 292, no. 3, pp. 741–748, Apr. 2002, doi: 10.1006/bbrc.2002.6703.
 - [51] D. Kakhniashvili, J. A. Mayor, D. A. Gremse, Y. Xu, and R. S. Kaplan, “Identification of a novel gene encoding the yeast mitochondrial dicarboxylate transport protein via overexpression, purification, and characterization of its protein product,” *J. Biol. Chem.*, vol. 272, no. 7, pp. 4516–4521, Feb. 1997, doi: 10.1074/jbc.272.7.4516.
 - [52] B. Almeida *et al.*, “Yeast protein expression profile during acetic acid-induced apoptosis indicates causal involvement of the TOR pathway,” *Proteomics*, vol. 9, no. 3, pp. 720–732, 2009, doi: 10.1002/pmic.200700816.
 - [53] H. De Nobel *et al.*, “Parallel and comparative analysis of the proteome and transcriptome of sorbic acid-stressed *Saccharomyces cerevisiae*,” *Yeast*, vol. 18, no. 15, pp. 1413–1428, 2001, doi: 10.1002/yea.793.

- [54] M. C. Teixeira, P. M. Santos, A. R. Fernandes, and I. Sá-Correia, “A proteome analysis of the yeast response to the herbicide 2,4-dichlorophenoxyacetic acid,” *Proteomics*, vol. 5, no. 7, pp. 1889–1901, 2005, doi: 10.1002/pmic.200401085.
- [55] M. C. Teixeira, A. R. Fernandes, N. P. Mira, J. D. Becker, and I. Sá-Correia, “Early transcriptional response of *Saccharomyces cerevisiae* to stress imposed by the herbicide 2,4-dichlorophenoxyacetic acid,” *FEMS Yeast Res.*, vol. 6, no. 2, pp. 230–248, 2006, doi: 10.1111/j.1567-1364.2006.00041.x.
- [56] N. Maassen *et al.*, “Influence of carbon and nitrogen concentration on itaconic acid production by the smut fungus *Ustilago maydis*,” *Eng. Life Sci.*, vol. 14, no. 2, pp. 129–134, Mar. 2014, doi: 10.1002/elsc.201300043.
- [57] W. E. Levinson, C. P. Kurtzman, and T. M. Kuo, “Production of itaconic acid by *Pseudozyma antarctica* NRRL Y-7808 under nitrogen-limited growth conditions,” *Enzyme Microb. Technol.*, vol. 39, no. 4, pp. 824–827, Aug. 2006, doi: 10.1016/j.enzmictec.2006.01.005.
- [58] T. Zambanini *et al.*, “Promoters from the itaconate cluster of *Ustilago maydis* are induced by nitrogen depletion,” *Fungal Biol. Biotechnol.*, vol. 4, no. 1, pp. 1–9, 2017, doi: 10.1186/s40694-017-0040-3.
- [59] J. Sasikaran, M. Ziemski, P. K. Zadora, A. Fleig, and I. A. Berg, “Bacterial itaconate degradation promotes pathogenicity,” *Nat. Chem. Biol.*, vol. 10, no. 5, pp. 371–377, 2014, doi: 10.1038/nchembio.1482.
- [60] P. Jouhten, O. Ponomarova, R. Gonzalez, and K. R. Patil, “*Saccharomyces cerevisiae* metabolism in ecological context,” *FEMS Yeast Res.*, vol. 16, no. 7, p. 80, 2016, doi: 10.1093/femsyr/fow080.

Chapter 6– Searching for the genetic parts to develop a yeast-based biosensor for itaconic acid

Chapter 7 – Summary of the work

The goal of this work was to use synthetic biology methodologies to foster Yeasts and, in particular, *S. cerevisiae*, as biotechnological hosts for production of carboxylic acids, including to those for which no biochemical production pathways has been established. LA and MAA are two examples of such CAs that are not naturally produced, while ITA was selected as an example for which a biochemical production pathway has been implemented but that still requires improvements namely in the design of high-throughput methods for a fast screening of natural/engineered strains/species. For the new-to-nature CAs LA and MAA, the main challenge was to devise suitable strategies to find the better candidate metabolic pathways that can be introduced in microbial cells (and in yeasts, in particular) to enable production. In what concerns ITA, the task tackled in this work aiming to contribute for the development of a handy tool for a rapid strain phenotyping based on the design of an ITA-responsive sensor using genetic parts from the yeast endogenous transcriptional machinery.

In this work four chapters of practical work were developed that have the objective of answering the challenges identified in the paragraph above. In specific, the work described in this PhD thesis aimed to answers the following research questions and with following findings:

What strategy could be used to identify candidate biochemical pathways to implement LA production in microbial hosts, in particular in *S. cerevisiae* ? (Chapter 3)

The strategy put in place, which resorted to the use of a combination of bioinformatics tools and expert manual curation for the identification of the more suitable substrates that could be converted into LA as well as the subsequent identification of enzymes that could catalyse the envisaged conversions, resulted in the identification of four new possible routes for LA production from fermentable carbon sources. These routes were categorized based on their potential to be successfully assembled with success in microbial hosts and using metabolic modeling their performance in *E. coli* and *S. cerevisiae* was assessed. On the overall the main findings obtained in this chapter were:

- Identification of possible D-ALA, glutamate semialdehyde, 3-oxoadipic acid and 5-aminovaleric acid as possible substrates for LA biosynthesis through five different biochemical routes (2 routes were identified for 3-oxoadipic acid, one of them being already reported);
- The shortest LA forming pathways, involving only two reactions to produce LA from the substrate, were those stemming from D-ALA and glutamate semialdehyde, which are converted to DAVA in the first reaction which is,

subsequently, deaminated to LA. The anticipated need to engineer glutamate semialdehyde mutase, the identified enzyme to catalyze the conversion of glutamate semialdehyde into DAVA, was an identified bottleneck when considering the use of glutamate semialdehyde as a substrate.

- The metabolic modeling performed indicated that the pathway with the highest theoretical yield in both *S. cerevisiae* and *E. coli* is the one starting with succinyl-CoA and acetyl-CoA condensation to 3-oxoadipic acid, which is then decarboxylated to LA. The use of this pathway to implement production of LA in *E. coli* has been published among a cohort of other products that could be synthesized through non-reductive Claisen condensation reactions. Despite this, the need for supplementation of the medium with succinate is a problem that needs further addressing since this is not sustainable at an industrial scale. Porting of this pathway to *S. cerevisiae* could help in providing a constant cytosolic supply of succinyl-CoA, since in past this was performed by expressing a multi-subunit bacterial α -ketoglutarate dehydrogenase in the cytosol. Necessarily the heterologous expression of such a high number of proteins (those necessary to produce succinyl-CoA and the enzymes necessary to produce LA) can become problematic.
- The pathway starting from 5-aminovaleric acid appears to be one displaying the best trade-off between number of required heterologous steps and predicted yield. In *S. cerevisiae*, where 5-aminovaleric is an intermediate in the lysine degradation pathway, only two enzymes would be needed to be expressed to convert this precursor in LA.
- The pathway starting from D- ALA, although having a lower maximal yield, is short (only two enzymatic steps), with the possibility of none of them being heterologous in *E. coli*. The fact that two possible enzymes were identified as candidates to catalyze the D-ALA amination step (first reaction: BioA from *E. coli* or BioK from *B. subtilis*) is also an advantage, since it increases the likelihood of obtaining a functional pathway. For this reason, this pathway was further studied *in vitro* and *in vivo* in the next chapter.

Of the potential routes identified previously, can one of these be successfully implemented *in vivo* and *in vitro*, leading to LA production? (Chapter 4)

Among the pathways identified in Chapter 3 the one stemming from D-ALA were selected to be explored *in vitro* and *in vivo*. This pathway includes a first step in which D-ALA is aminated to DAVA through the action of KANA transaminase (two candidate enzymes were identified, one from *B. subtilis* and another from *E. coli*, using as amino donors lysine or SAM, respectively). In the second step DAVA is deaminated to LA in a reaction catalyzed by DAPA ammonia lyase (from *E. coli*). The first step was to confirm *in vitro* that the transaminases could catalyze the D-ALA to DAVA conversions. The results obtained allowed us to conclude that:

- KANA transaminase BioK from *B. subtilis* can accept D-ALA as a substrate and catalyze its conversion to DAVA, as indicated by LC-MS analysis of the supernatants obtained from assays where this enzyme was put in contact with D-ALA. The results obtained also rendered likely that BioA from *E. coli* could catalyze this reaction. It is likely that further enzyme engineering of these enzymes is required to increase enzymatic activity towards the non-native substrate D-ALA.
- The HPLC analysis of double enzymatic assays with the combination BioA/K + YgeX did not reveal the formation of LA when these two enzymes are active (as the experimental conditions were validated in single enzymatic assays with the native substrates) and the substrate D-ALA is provided. Thus, it is expected that engineering of YgeX to modify substrate specificity may be necessary to enable DAVA deamination to LA.
- The *in vivo* assembly of the pathway in *E. coli* was attempted, however, under the conditions used (which included different experimental setups to produce D-ALA) the over-expression of the two selected enzymes did not result in detection of LA in the broth.
- Considering that the results from the *in silico* modeling showed that implementation of this pathway in *S. cerevisiae* is constrained by availability of D-ALA, a strategy to improve the internal availability of this metabolite was devised. First, the *HEM1* gene which encodes the D-ALA synthesizing enzyme D-ALA synthase, was over-expressed enabling the titers of D-ALA accumulated in the broth up to 11.6 mg/L. Subsequently, down-regulation of the essential gene *HEM2*, encoding the D-ALA consuming enzyme D-ALA dehydratase, was attempted. Although a strain having the aimed genetic modification was obtained,

the results obtained could not demonstrate that this strain does indeed exhibit lower levels of *HEM2* when compared with the corresponding wild-type strain.

Which biochemical pathways could be used to implement microbe-based production of MAA? (Chapter 5)

Following the strategy implemented in chapter 3 for the prospection of pathways that could lead to biosynthesis of LA, a prospecting work was undertaken to find metabolic routes leading to MAA. Among the alternatives identified, the route stemming from mesaconic acid was further explored to find enzymes that could mediate the decarboxylation of this metabolite into MAA. Two of the more promising candidate enzymes were tested *in vitro* for their ability to catalyse the envisaged mesaconic acid-MAA conversion. Altogether the results obtained in this chapter led to the following conclusions:

- Six MAA possible formation routes were identified, starting from five different precursors: mesaconic acid, mesaconyl-C4-CoA, 3-hydroxybutyric acid, 3-aminobutyric acid and isobutyryl-CoA; ~~the~~ the herein identified route that started from isobutyryl-CoA was previously validated *in vivo*.
- With the exception of the pathway stemming from mesaconyl-C4-CoA and isobutyryl-CoA, all the routes entail at least one reaction step for which it was not possible to assign an enzyme with demonstrated capability to catalyze the envisaged reaction.
- The route that uses mesaconic acid as a substrate involves a single decarboxylation step, for which two candidate enzymes were identified: the *A. terreus* cis-aconitic acid decarboxylase (AtCadA) and the *P. fluorescens* aminocarboxymuconate semialdehyde decarboxylase (NbaD). This route was selected for experimental validation since it only entails one synthetic step and the precursor mesaconic acid has already an established production pathway in *E. coli*.
- Enzymatic assays using crude cell extracts of *S. cerevisiae* cells over-expressing AtCadA and mesaconic acid could not detect formation of MAA; under these conditions, the enzyme AtCadA was proven functional since formation of ITA was observed when cis-aconitic acid was used as a substrate;
- Enzymatic assays with mesaconic were carried out from crude cell extracts of *E. coli* cells expressing *nbaD*. No formation of MAA was observed in the HPLC

analysis of the supernatants; however, mesaconic acid consumption was observed in half of the conditions tested, with the appearance of an unidentified HPLC peak which is presumed to be either the product of mesaconic acid decarboxylation at the C1 group (2-butenic acid) or the product of mesaconic hydration (2-hydroxy-2-methylsuccinic acid). The lack of the native enzyme substrate prevented the confirmation that the enzyme is active in the conditions employed here.

How to design and build an ITA sensor in yeast? (Chapter 6)

With the aim of developing a biosensor responsive to itaconic acid that could be used to rapidly screen in a high-throughput manner eukaryotic strains amenable for the production of this carboxylic acid, a strategy was devised to identify the necessary genetic parts (transcription factor and respective binding site) among the *S. cerevisiae* endogenous machinery. This strategy was settled on a transcriptomic analysis of *S. cerevisiae* cells when suddenly exposed to two concentrations of ITA and after a short (still in the lag phase of adaptation) and in mid-exponential growth in the presence of the acid. The results of this transcriptomic analysis allowed the following conclusions:

- 12 genes (*DAL5*, *MEP2*, *DAL2*, *DAL3*, *DUR3*, *DAL7*, *QDR2*, *YAT2*, *AQR1*, *YGL117W*, *QDR3* and *ORT1*) were identified to have an ITA-inducible behavior that is dose-dependent; this behavior was validated by independent RT-qPCR experiments for the *DAL2*, *DAL5*, *MEP2* and *TMT1* genes
- Three possible transcription factor-promoters pairs are suggested: *DAL5*, *TMT1* and *MEP2* with Leu3 being the regulator identified as to mediate ITA induction at high ITA concentrations. Although the induction of these genes was also dependent on Dal81, Leu3 is here identified as a potential regulator be used in a sensor design, since it is metabolite responsive and due to the structural similarity of ITA to its cognate inducer, 2-isopropylmalate.

The main findings in this thesis and the proposed future work are graphically summarized in Figure 7.1.

Chapter 7– Summary of the work

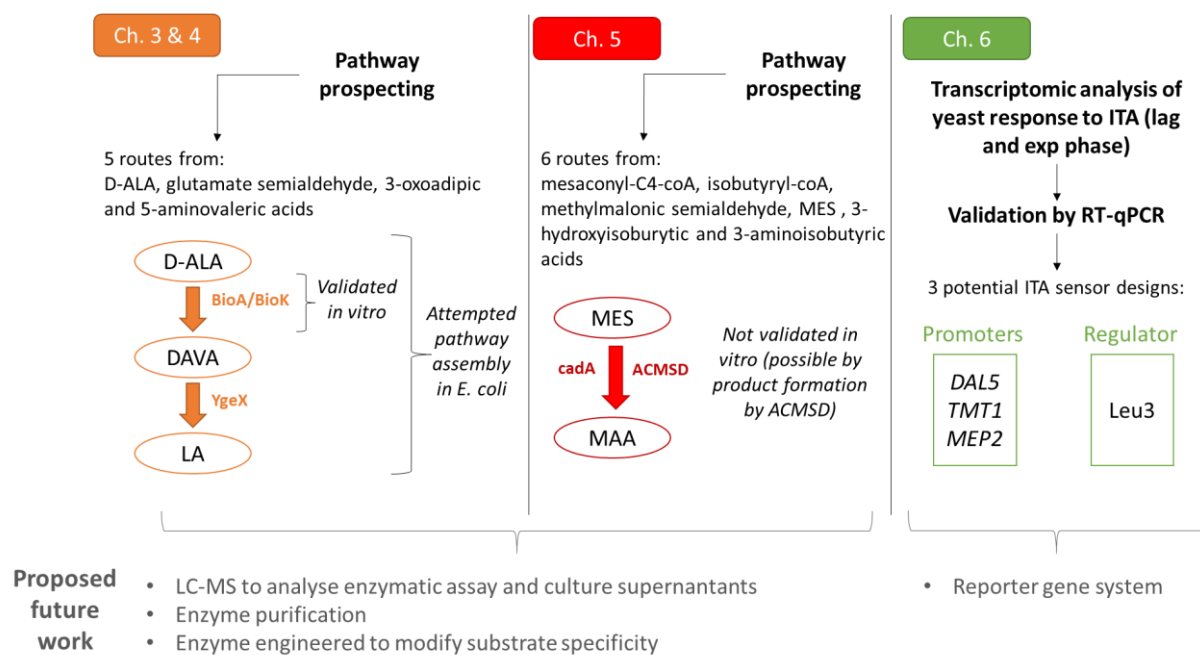


Figure 7.1. Summary of the work performed during this thesis, with the main findings and proposed future work.

Chapter 8 – Annex

8.1. Annex 1- Pathway prospecting for α -methylglutamic acid (MGU) biosynthesis

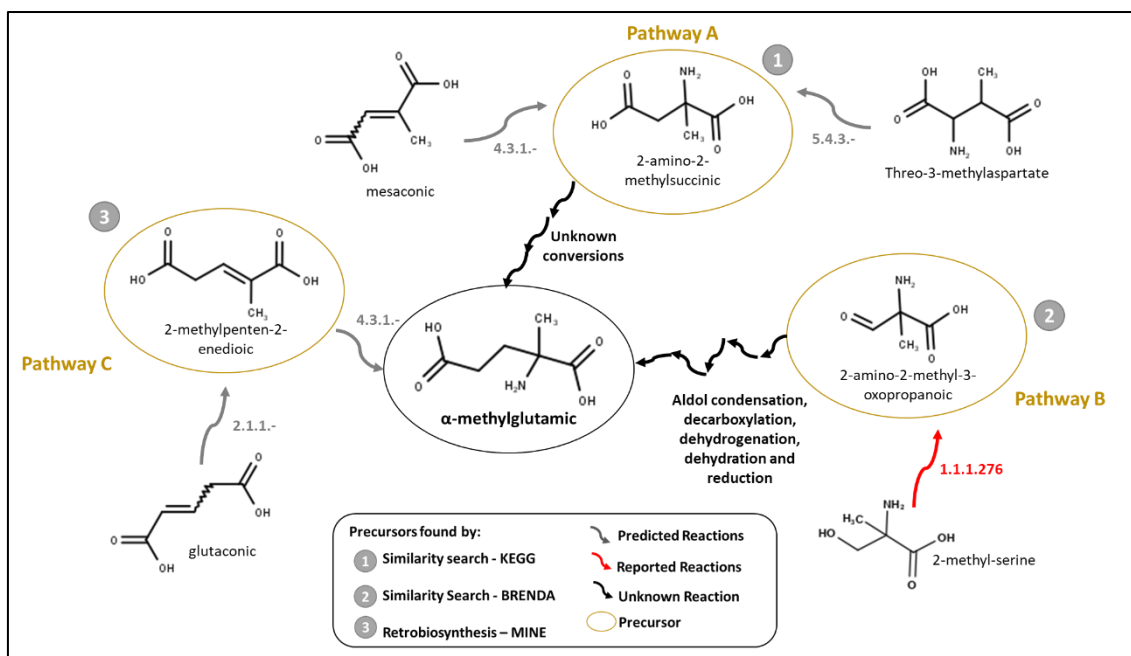


Figure 8.1. MGU forming pathways found in this work.

A similarity search in KEGG Compound yielded the possible precursor 2-amino-2-methylsuccinic. This precursor was selected for further pathway search, as it is predicted to be the product of enzymatic conversion of mesaconic or thero-3methylaspartate. A search performed in BRENDA, for compounds that contain the unusual α -methyl- α -aminoacid group, resulted in the selection of 2-amino-2-methyl-3-oxopropanoic acid as a possible precursor, which can be obtained from 2-methylserine. Another potential precursor was found in a search in MINE : 2-methylpenten-2-dioic acid, which was predicted by MINES to be formed by the action of C-methyltransferases (2.1.1) on glutaconic. From the precursor list 3, possible pathways were devised by searching in MINES or in general reaction chemistry (in the case of pathway 2) for possible conversions to obtain MGU. After this, it was attempted to assign enzymes to each enzymatic reaction.

Pathway A) For 2-amino-2methylsuccinic acid no possible conversions were found to extend the carbon chain and obtain MGU.

Pathway B) For 2-amino-2-methyl-3-oxopropanoic acid, a possible 5-step enzymatic

conversion was devised: aldol condensation of this precursor with pyruvate would yield 5-amino-5-methyl-4-hydroxy-2-oxoadipic acid, which could be decarboxylated to 5-oxo-4-hydroxy-2-methylpentanoic acid. Subsequent dehydrogenation of the carboxy acid would result in 4-hydroxy-2-methylglutamic acid, which upon dehydration would yield 2-methyl-2-aminoglutaconic acid, which could finally be reduced to MGU. This pathway was disregarded, due to the requirement of 5 predicted enzymatic steps and the fact that intermediate compounds are not available commercially, which would complicate the task of *in vitro* testing each of the predicted enzymatic steps, since only step 1 (aldol condensation) can be *tested in vitro* in this scenario.

Pathway C) For 2-methylpenten-2-dioic it was predicted in MINE that an ammonia lyase (4.3.1.-) could catalyze the addition of the 2-amino group to yield MGU. However, no ammonia lyase likely to catalyze this reaction was found.

8.2. Annex 2 - Supplementary tables of the LA pathway prospecting work

Table 8.1. Reactions inserted in the iJO1366 and Yeast 5.0 models to simulate the heterologous pathways.

Pathway	Reaction name	Reaction formula
A1	KAPA transaminase	D-ALA + SAM à DAVA + SAM_N (S-adenosyl-4-(methylsulfanyl)-2-oxobutanoate)
	DAVA ammonia lyase	DAVA + H ₂ O à LA + 2x ammonia
A2	KAPA transaminase	D-ALA + Lysine à DAVA + 2-amino6-oxohexanoate
	DAVA ammonia lyase	DAVA + H ₂ O à LA + 2x ammonia
B	Pyridoxal synthase	Ribose-5-P + Glyceraldehyde-3-P + Glutamine à PLP + 3xH ₂ O + P
	Glutamyl-tRNA reductase	Glutamyl-tRNA + NADPH + H à Glutamate-1-semialdehyde + NADP
	GSA mutase	Glutamate 1-semialdehyde à DAVA
	DAVA ammonia lyase	DAVA + H ₂ O à LA + 2x NH ₃
C	3-DHS dehydratase	3-dehydroshikimate à Protocatechuate + H ₂ O
	PCA decarboxylase	PCA à Catechol + CO ₂
	Catechol 1,2-dioxygenase	Catechol + O ₂ à Cis-cis-muconate
	Muconate cycloisomerase	Cis-cis-muconate à Muconolactone
	Muconolactone delta-isomerase	Muconolactone à 3-oxoadipate-enol-lactone
	3-oxoadipate enol-lactonase	3-oxoadipate-enol-lactone + H ₂ O à 3-oxoadipate
	Oxaloacetate decarboxylase	3-oxoadipate à LA + CO ₂
D	Lysine N-acetyltransferase	Lysine + acetyl-coA à N6-Acetyl-Lysine + coA + H+
	N-Acetyl-lysine aminotransferase	N6-Acetyl-Lysine + 2-oxoglutarate à Glutamate + 2-oxo-6-acetamidocaproate
	Oxidative decarboxylation	2-oxo-6-acetamidocaproate + NAD + H ₂ O à 2-acetoamidovalerate + NADH + H + CO ₂
	Deacetylation	Acetoamidovalerate + H ₂ O → 5-aminopentanoate + acetate
	5-aminopentanoate aminomutase	5-aminopentanoate à 4-aminopentanoate
	4-aminopentanoate transaminase	4-aminopentanoate à LA + NH ₃
E	2-oxoglutarate	2-oxoglutarate + coA + NAD → CO ₂ + NADH + Succinyl-coA
	Beta-ketoadipyl-coA thiolase	Succinyl-coA + Acetyl-coA à 3-oxoadipyl-coA + coa
	β-ketoadipate:succinyl CoA transferase	3-oxoadipyl-coA + succinate à 3-oxoadipic + succinyl-coA
	3-oxoadipate decarboxylase	3-oxodipate à LA + CO ₂

Reactions in orange were only added to the *E. coli* model, since these enzyme are present in yeast.

Reactions in red were only added to the Yeast 5.0 model, since they were not present in the model (although they are present in yeast metabolism).

The reaction in green was only added to yeast model, since it is a bacterial enzyme

Table 8.2. Candidate precursors selected from the KEGG database according to the structural similarity search.

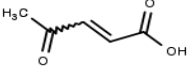
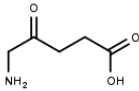
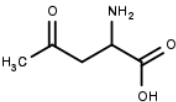
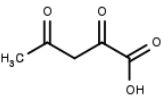
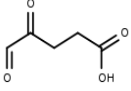
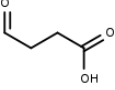
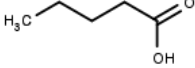
Molecule	KEGG Code	Name	Tanimoto coefficient	Relevant Pathways
	C07091	Cys-acetylacrylate	0.89	Chlorocyclohexane and chlorobenzene degradation
	C00430	Delta-Aminolevulinic	0.84	Heme biosynthesis (plants and bacteria, glutamate --> heme) (animals and fungi, glycine --> heme)
	C03341	Beta-aminolevulinic	0.85	Arginine and ornithine metabolism
	C02132	2,4-dioxovaleric (acetylpyruvic)	0.83	N/A
	C02800	4,5-dioxovaleric (4-Oxoglutarate semialdehyde)	0.83	Porphyrin and chlorophyll metabolism
	C00232	Succinate semialdehyde	0.81	GABA shunt, dicarboxylate/hydroxypropionate-hydroxybutyrate cycle, nicotine degradation
	C00803	Valeric	0.81	N/A

Table 8.3. LA forming reactions predicted in MINE. The precursors were selected according to the ability to produce them in a microbial cell.

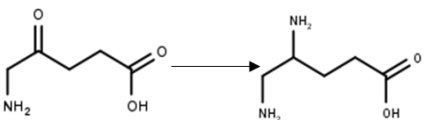
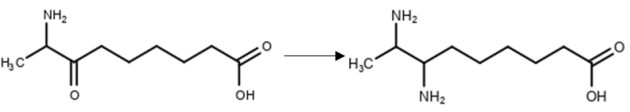
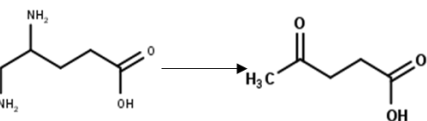
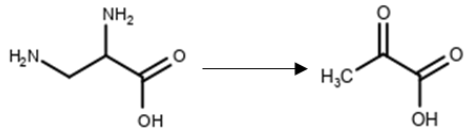
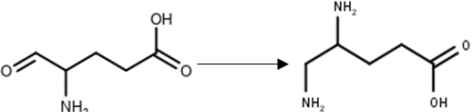
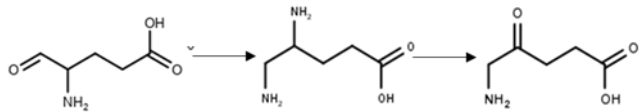
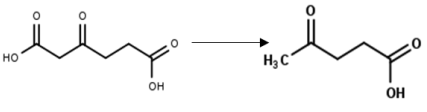
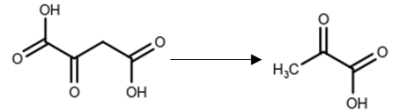
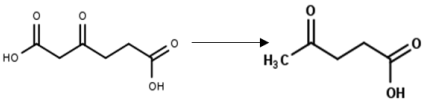
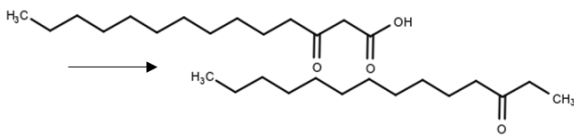
Precursor	Reaction	Reaction Rule	Precursor was selected?
Succinic semialdehyde		2.1.1.-	Yes
3-oxoadipic		4.1.1.-	Yes
Cis-acetylacrylic		1.3.1.-	No
Mycophenolic acid		1.3.1.-	No

Table 8.4. LA forming pathways predicted in ReactPred and considered after the filtering criteria. The original 26 pathways were fused in 11 pathways, since many pathways were repeated in the results. A manual check of every pathway was required to remove predictions that included unrealistic reactions, unstable molecules or undesirable precursors. Reactions were considered unrealistic when they were too complex to occur enzymatically (ex: simultaneous deamination, isomerization and methylation).

Pathway Number	Precursor	Pathway	Pathway passed manual check?	Reason for failing manual check
R1	2-hydroxy-3-oxoadipic	<p>2-hydroxy-3-oxoadipic → 3-oxoadipic → LA</p>	No	Unrealistic 1st reaction (a dehydrogenation would yield a double bond)
R2	4-hydroxy-4-methyl-2-oxoadipic	<p>propionic + 2-oxo-4-hydroxy-4-methyladipic → 2,4-dioxo-4-methylheptanoic → LA</p>	No	The precursor is present in KEGG but not associated to any enzymatic reaction
R3	2-amino-5-oxohexanoic	<p>2-amino-5-oxohexanoic + Isoleucine → 2-amino-3-methyl-5-oxohexanoic → LA</p>	No	Unrealistic reactions
R4	Glutamic	<p>Glutamic + Isoleucine → 3-Methylglutamic → LA</p>	No	Unrealistic reactions
R5	Glutamic	<p>Glutamic + Isoleucine → 2-amino-2,3-dimethylbutanoic → LA</p>	No	Unrealistic reactions
R6	4-hydroxy-2-oxoheptanoic	<p>4-hydroxy-2-oxoheptanoic + Isoleucine → 2-oxo-4-hydroxy-4-methylheptanoic → LA</p>	No	Unrealistic 1st reaction
R7	Valine	<p>Valine + Isoleucine → 4-oxobutanoic → LA</p>	No	The 2nd reaction (methylation of succinic semialdehyde) is disconnected from the 1st reaction.

R8	GABA	<p>GABA Isoleucine 4-aminovaleric 2-ketoglutaric LA</p>	Yes	-
R9	2-amino-5-oxohexanoic	<p>2-amino-5-oxohexanoic Isoleucine 2-amino-3-methyl-5-oxohexanoic Propanoic LA</p>	No	Unrealistic 1st reaction
R10	2-amino-2-methylbutyric	<p>2-amino-2-methylbutanoic Isoleucine 2-amino-2,3-dimethylbutanoic 4-oxobutanoic LA</p>	No	The 2nd reaction (methylation of succinic semialdehyde) is disconnected from the 1st reaction.
R11	3-methyl-2-oxovaleric	<p>Glutamic 2-oxo-3-methylvaleric 4-oxobutanoic Isoleucine LA</p>	No	The 2nd reaction (methylation of succinic semialdehyde) is disconnected from the 1st reaction.
R12	5-aminovaleric	<p>5-aminovaleric 4-aminovaleric 2-oxoglutaric LA</p>	Yes	-

Table 8.5. Assigned enzymatic reactions. The desired synthetic reaction is depicted along with the main native reaction. The Tanimoto coefficient between the native and synthetic substrate was calculated with ChemMine Similarity workbench

Pathway Number	Synthetic Reaction	Assigned Enzyme	Native reaction	Tanimoto coefficient
A		Ketoaminonanoate transaminase (2.6.1.62/105)		0.4545
		Diaminopropionate ammonia lyase (4.3.1.15)		0.4667
B		Glutamate semialdehyde mutase (5.4.3.8)		-
C		Oxaloacetate decarboxylase (4.1.1.1112)		0.5385
		Methylketone synthase (N/A)		0.4286

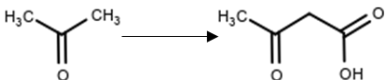
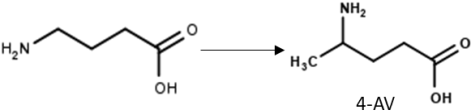
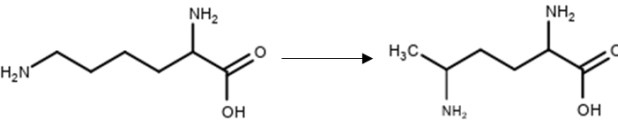
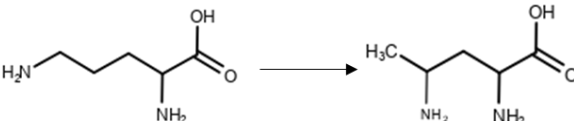
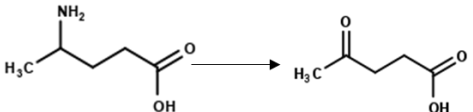
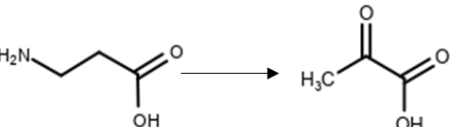
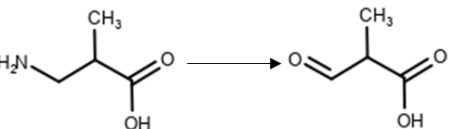
		Acetoacetate decarboxylase (4.1.1.4)		0.6364
D		Lysine aminomutase (5.4.3.3)		0.6364
		Ornithine aminomutase (5.4.3.5)		0.889
		Beta-alanine-pyruvate transaminase (2.6.1.18)		0.5556
		3-amino-2-methylpropionate-pyruvate transaminase (2.6.40)		0.5000

Table 8.6. Additional information on the enzymatic steps comprised in the selected pathways

Pathway	EC number	Enzyme	Native Pathway of enzyme	Presence in metabolism	
				<i>E. coli</i>	<i>S. cerevisiae</i>
A1	2.6.1.62	KAPA transaminase (SAM)	Biotin synthesis	Y	Y
	4.3.1.15	DAPA ammonia lyase	Elimination of DAPA	Y	N
A2	2.6.1.105	KAPA transaminase (lysine)	Biotin synthesis	N	N
	4.3.1.15	DAPA ammonia lyase	Elimination of DAPA	Y	N
B	1.2.1.70	Glutamyl-tRNA reductase	D-ALA biosynthesis (C5 pathway)	Y	Y
	5.4.3.8	GSA mutase		Synthetic (requires engineering)	
	4.3.1.15	DAPA ammonia lyase	Elimination of DAPA	Y	N
C	4.2.1.118	3-DHS dehydratase	Cis-cis-muconic acid synthetic pathway	N	N
	4.1.1.63	PCA decarboxylase		N	N
	1.13.11.1	Catechol 1,2-dioxygenase		N	N
	5.5.1.1	Muconate cycloisomerase	Aromatic degradation	N	N
	5.3.3.4	Muconolactone δ -isomerase		N	N
	3.1.1.24	3-oxoadipate enol-lactonase		N	N
	4.1.1.112	Oxaloacetate decarboxylase	Pyruvate metabolism	Y*	Y*
D	2.3.1.- ****	Lysine N-acetyltransferase	Lysine degradation (acetylating)	N	Y
	2.6.1.- ****	N-Acetyl-lysine aminotransferase		N	Y
	1.1.1.- ****	Oxidative decarboxylation		N	Y
	3.5.1.- *****	Deacetylation		N	Y
	5.4.3.3	Lysine 5,6-aminomutase *	Lysine degradation to acetate and butanoate	N	N
	2.6.1.18	β -alanine-pyruvate transaminase**	β -alanine synthesis	N	N
E	2.3.1.16	Beta-ketoadipyl-coA thiolase	Fatty acid β -oxidation	Y	Y
	2.8.3.6.	β -ketoadipate:succinyl CoA transferase	3-oxoadipate degradation	N	N
	4.1.1.4	Acetoacetate decarboxylase***	Acetone degradation	N	N

* May also be ornithine 4,5-aminomutase (5.4.3.5) but the stoichiometry of the reaction is the same.

** May also be 3-amino-2-methylpropionate-pyruvate transaminase (2.6.1.40) or taurine-2-oxoglutarate transaminase (2.6.1.55) but the stoichiometry of the reaction is equivalent

*** May also be methylketone synthase 1 (4.1.1.56) but the stoichiometry of the reaction is the same.

**** There is experimental evidence for the pathway. However, the exact enzymes catalyzing the reactions haven't been assigned yet.

Table 8.7. Flux Coupling Analysis results for each pathway/organism combination. The reactions selected by F2C2 considered to be directionally coupled to the LA production reaction are listed. In all cases the reactions belonging to the heterologous pathways were directionally coupled to LA, as expected, and are not listed here. In all cases the reactions belonging to the heterologous pathways were directionally coupled to LA, as expected, and are not listed here.

Pathway	Host	Reactions coupled to LA production
A1	<i>E. coli</i>	Phosphoribosyl(PPR)-aminoimidazolecarboxamide formyltransferase, Glutamine synthase, Glutamine PPR-diphosphate amidotransferase, IMP cyclohydrolase, PPR-aminoimidazole synthase, PPR-formylglycinamide synthase
	<i>S. cerevisiae</i>	5-aminolevulinate synthase, PPR-5-amino-4-(N-succinocarboxamide)-imidazole AMP-lyase, adenylosuccinate lyase, adenylosuccinate synthase, adenylyl-sulfate kinase, aspartate kinase, aspartate-semialdehyde dehydrogenase, cis-aconitate to isocitrate, citrate to cis-aconitate, glutamine synthase
A2	<i>E. coli</i>	-
	<i>S. cerevisiae</i>	2-aminoadipate transaminase, 2-methylcitrate dehydratase, 5-phosphoribosylformyl glycinamide synthase, 5,10-MTHF reductase (NADPH), adenylosuccinate lyase, adenylosuccinate synthase, adenylyl-sulfate kinase, aspartate kinase, aspartate-semialdehyde dehydrogenase, aminoimidazole synthase, phosphoadenylyl-sulfate reductase, PPR-amino imidazolesuccinocarboxamide synthase, PPR-aminoimidazole carboxylase, PPR-aminoimidazole-carboxylase, PPR-aminoimidazolecarboxamide formyltransferase, PPR-glycinamide synthase, PPR-pyrophosphate synthase, saccharopine dehydrogenase (NAD and NADP), thioredoxin reductase (NADPH), homocitrate synthase
B	<i>E. coli</i>	-
	<i>S. cerevisiae</i>	-
C	<i>E. coli</i>	-
	<i>S. cerevisiae</i>	3-dehydroquinate dehydratase, 3-dehydroquinate synthase, ribose 5-phosphate isomerase
D	<i>E. coli</i>	Aspartate-semialdehyde dehydrogenase, aspartate kinase, aspartate transaminase, diaminopimelate decarboxylase, diaminopimelate epimerase, dihydrodipicolinate synthase, succinyl-diaminopimelate desuccinylase, succinyl-diaminopimelate transaminase, tetrahydrodipicolinate succinylase
	<i>S. cerevisiae</i>	2-aminoadipate transaminase, 2-methylcitrate dehydrogenase, aconitase, homoaconitate dehydrogenase, homoisocitrate dehydrogenase, saccharopine dehydrogenase (NAD and NADP), homocitrate synthase
E	<i>E. coli</i>	-
	<i>S. cerevisiae</i>	-

8.3. Annex 3 – Primers used in this work

Table 8.8. Primers used in Chapter 4

Name	Purpose	Sequence
PBS-HEM2 (+)	Replacing Hem2 promoter with URA3 cassette from pBS-KIURA3miniblaster	CTCATATTATTGCTTACTCCAGTAATTCATACCTAGA
HEM2-CYC1 (-)		AAATTGACAAAACCCATAAGCACTAAATCGGAACC
prHem2_confirm_rev	Confirming miniblaster insertion upstream of Hem2 promoter	AACAGATGAGATTTCTGTTGGTTCTGTTTCAAAAA
prHem2_confirm_fwd		TTCAGCTGTATGCATCGAATTGATCCGGTAATTTAG
His3_fwd	To confirm integrations in His3 locus	CACTGTCTCAGCAGTGGGTG
His3_rev		GCTACAAGCCTGTGTCACCG
pDuet_bioA_fwd	For EcoRI-HindIII cloning of bioA (E. coli K12) into pETDuet-1 (MCS1)	TGCGATCTCTTTAAAGGGTGG
pDuet_bioA_rev		TTCAGTGGTGTGATGGTCGT
pDuet_bioK_fwd	For EcoRI-HindIII cloning of bioK (B. subtilis) into pETDuet-1 (MCS1)	ATGCGAATTCTATGACAACGGACGATCTTGC
pDuet_bioK_rev		TGATAAGCTTTTATTGGCAAAAAATGTTTCATCC
pDuet_confirm_fwd	Confirm insertion on MCS1 in pETDuet-1	ATGCGAATTCTATGACTCATGATTTGATAGAAAAA AGT
pDuet_confirm_rev		TGATAAGCTTTCAATCTTCAAGGCTCGTAACC
pDuet-ygex-fwd	For NdeI-XhoI cloning of ygeX (E. coli) into pETDuet-1 (MCS2)	ATGCGTCCGGCGTAGA
pDuet-ygex-rev		GATTATGCGGCCGTGTACAA
pDuet-confirm-rev2	For confirmation of insertion in pDuet (MCS2) w/ pDuet-confirm-fwd	TCGTCATATGATGTCCGTTTTCTCATTGAAGA
pColA_MCS2_fwd	To use with T7_fwd to confirm cloning in pColADuet MCS 2	GATCGCTCGAGTTAAGGTGCTACAGCGTGTT
alg9-fwd	RT-qPCR	CTGCGCTAGTAGACGAGTCC
alg9-rev		GTTCTGACTTAAGCATTATGC
hem2-fwd		TGCAGGCCAGGCAATGT
hem2-rev		GGCTATAGGAGCTGTGTAATTGTTCA

Table 8.9. Primers used in Chapter 6

Name	Purpose	Sequence
alg9-fwd	for RT-PCR	TGCAGGCCAGGCAATGT
alg9-rev		GGCTATAGGAGCTGTGTAATTGTTCA
tmt1-fwd		AGTGGAGTTTGCAGATTATGTCAGA
tmt1-rev		TTGGATCCTGCTTCCACTGAT
mep2-fwd		AGCCGCCGGGTTCTGT
mep2-rev		CTGCACCAGTAACCACACCAA
dal2-fwd		TCATTACCGTGGAAGGTTGCT
dal2-rev		CCCATGTTCCCTTCACCACTGT
dal3-fwd		AGGCTCCAATCAGGCGATT
dal3-rev		CGTACTACACGGATGCTTTTCG
dal5-fwd		TGGCAGGACCACAAACCTTT
dal5-rev		CGACCTTAGCGCCATGATATTT

8.4. Annex 4 – Analysis of LA pathway expression and ScHEM1 expression

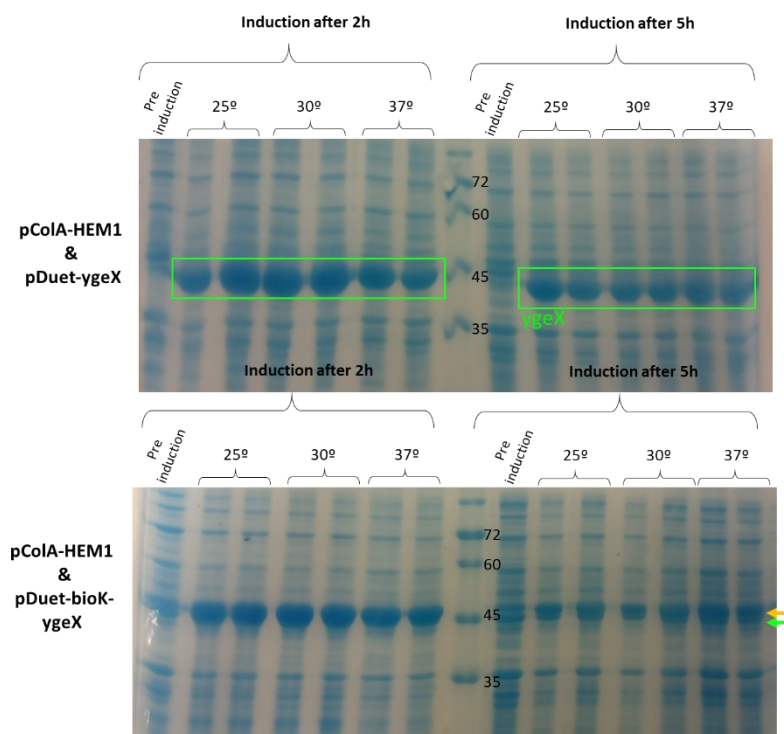


Figure 8.2. SDS-PAGE of BL21(DE3) cells carrying pColA-ScHEM1 and pD-ygeX (top) or pColA-ScHEM1 and pD-bioK-ygeX (bottom). Induction was carried either 2h or 5h after culture start and proceeded at 25°C, 30°C or 37°C. The two lanes matching each temperature correspond to induction with 0.1 and 0.5 mM of IPTG. The bands corresponding to ygeX and bioK molecular weight are depicted in green (43 kDa) and yellow (50 kDa), respectively.

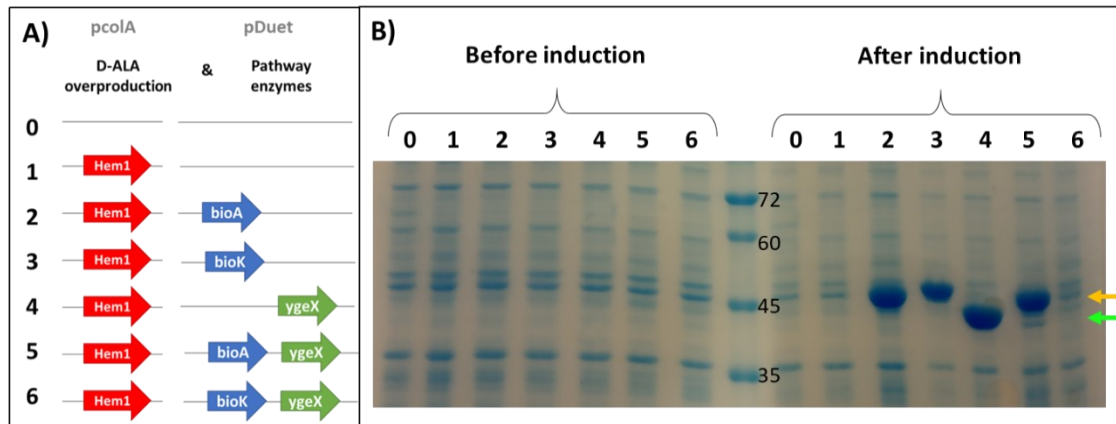


Figure 8.3. SDS-PAGE of cells harvested at 24h after induction with 0.1 mM IPTG at 30°C (induction after 5h of culture start). The bands corresponding to ygeX and bioA/bioK molecular weight are depicted in green (43 kDa) and yellow (47/50 kDa), respectively.

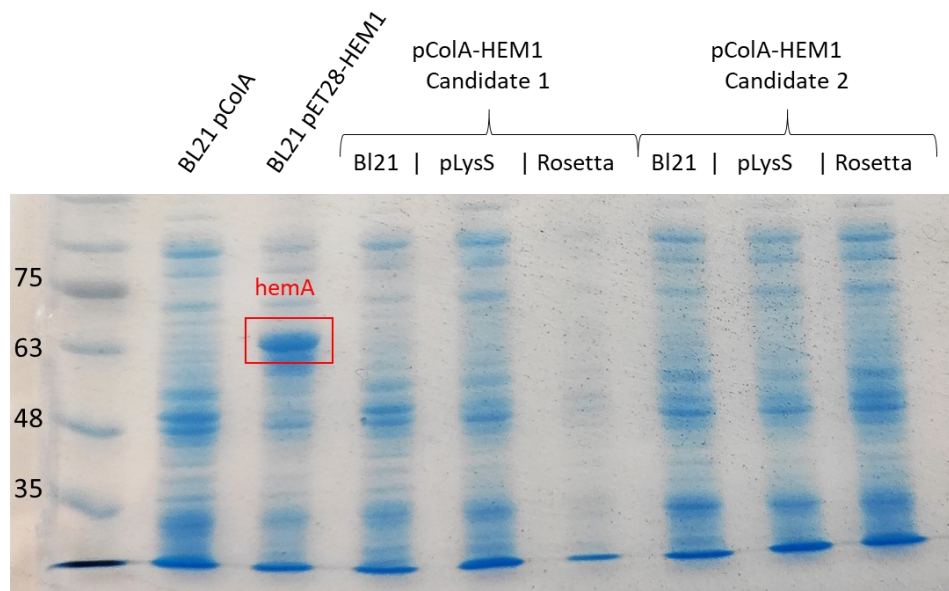


Figure 8.4. SDS-PAGE of cells carrying the pColA-ScHEM1 plasmid. The cells were harvested at 24h after induction with the LA pathway protocol (induction at 5h after cultura start with 0.1 mM IPTGa at 30°C).

8.5. Annex 5 – ITA transcriptomic analysis

Table 8.10. Genes upregulated in the lag phase in both ITA concentrations. Genes with a positive slope of activation are highlighted in grey.

Gene/ORF	logFC (175)	logFC (290)	Gene/ORF	logFC (175)	logFC (290)	Gene/ORF	logFC (175)	logFC (290)	Gene/ORF	logFC (175)	logFC (290)
DAL5	5,17	6,55	MPC2	1,61	2,54	PYC2	0,21	1,31	PEX12	1,06	1,59
MEP2	3,98	5,08	YCR023C	1,46	2,40	GZF3	0,81	1,61	ASI3	0,63	1,31
DAL2	3,75	4,77	FMS1	0,25	1,86	SFT2	1,20	1,85	AUS1	1,24	1,72
DUR3	1,64	3,92	SNZ1	1,64	2,48	PDR8	1,75	2,22	SSU1	1,48	1,89
YIL165C	2,63	3,97	MET30	0,93	2,12	VHT1	1,54	2,06	CAB1	1,50	1,91
TMT1	2,08	3,62	PXP2	2,22	2,82	CAF16	1,32	1,91	MRM1	0,86	1,45
UGA3	2,04	3,58	BUD16	1,54	2,39	LEU4	1,76	2,21	ARO3	1,44	1,85
MCH1	1,82	3,30	MEP1	0,78	1,96	YDL180W	0,71	1,53	PMU1	1,11	1,61
STR2	1,70	3,11	SPG4	2,25	2,78	NPR1	0,95	1,66	ATG33	1,02	1,54
QDR2	3,10	3,78	ARG1	1,28	2,15	IDH2	1,72	2,17	HOM2	0,75	1,36
YAT2	2,63	3,48	CIA2	1,73	2,41	SER1	1,10	1,74	CLG1	0,85	1,42
SDT1	2,02	3,11	HIS2	1,22	2,11	PDX3	1,47	1,98	CPS1	2,08	2,35
DAL3	4,12	4,45	AGP1	1,39	2,18	BRR1	1,03	1,69	ZRT2	0,39	1,13
RTC2	2,00	3,08	HIS4	0,77	1,81	CIT2	1,04	1,69	ERV1	0,99	1,51
STB4	1,48	2,83	YJL133C-A	1,25	2,06	MUP3	1,12	1,74	ALD5	0,63	1,26
AQR1	1,88	2,97	YGL114W	0,90	1,86	YJR112W-A	0,62	1,43	HFA1	0,55	1,21
QDR3	1,81	2,91	RIB5	1,45	2,16	SAM4	0,62	1,42	ILV1	0,50	1,17
HSP78	0,76	2,47	HIS7	1,02	1,90	OAC1	2,18	2,50	RHO5	0,26	1,03
DAL7	3,29	3,77	ARO1	0,72	1,69	ADH5	0,91	1,57	BNA3	1,26	1,68
ICY1	1,68	2,78	MAE1	0,84	1,75	ISA1	0,65	1,40	LYS14	0,85	1,38
VBA1	2,01	2,93	ARG3	0,99	1,82	CPA1	1,00	1,61	YMC1	1,57	1,90
YGL117W	2,11	2,92	YMR265C	0,39	1,52	ECM5	0,70	1,42	WWM1	0,30	1,00
ARG2	1,59	2,65	MEP3	0,66	1,62	YPR1	0,11	1,07	YPL071C	0,42	1,07
BOP2	1,77	2,72	VID24	1,89	2,38	ASN1	1,80	2,17	TRP5	0,57	1,17
ORT1	2,05	2,86	MET22	2,04	2,48	ARO8	0,71	1,40	PGD1	0,15	0,90
ARG4	1,61	2,62	BNA6	0,34	1,44	GPB2	0,56	1,29	PRP5	0,55	1,15
RFA3	0,54	2,17	SRY1	0,75	1,64	BAP3	1,14	1,66	CPT1	0,08	0,85
YBL048W	0,37	2,11	HRB1	1,24	1,92	HIS3	1,19	1,69	RIB3	0,62	1,19
ARG7	1,54	2,57	YLR152C	0,02	1,29	PYC1	1,41	1,85	TRP2	0,82	1,33
ECM32	0,43	2,12	SER3	1,00	1,77	YOL036W	0,16	1,05	ADP1	0,50	1,10
SKP2	1,28	2,42	NAR1	0,79	1,62	AIM46	0,12	1,02	HTB2	0,01	0,79
ZRT1	1,41	2,46	IDP1	1,91	2,35	HHT2	0,92	1,50	SNQ2	0,29	0,96
MCH4	1,52	2,52	PDR10	0,44	1,43	IML1	0,68	1,34	ERI1	0,48	1,08

Chapter 8– Annex

Gene/ORF	logFC (175)	logFC (290)	Gene/ORF	logFC (175)	logFC (290)	Gene/ORF	logFC (175)	logFC (290)	Gene/ORF	logFC (175)	logFC (290)
UBC13	0,48	1,08	PCL5	1,07	1,40	RIB1	0,66	1,00	GGC1	1,17	1,36
VPH2	0,03	0,80	PEX7	0,47	0,94	YMR027W	0,43	0,81	GIP3	0,73	0,98
TFC3	0,30	0,96	APS2	0,53	0,98	BAP2	0,54	0,90	YOR203W	1,45	1,60
AVT7	0,95	1,40	MDM38	0,24	0,77	VPS41	0,83	1,13	PDB1	0,60	0,86
TMS1	0,38	1,00	ADD66	0,54	0,99	PET10	1,16	1,40	ADE12	0,69	0,93
YLR179C	1,02	1,44	YGR125W	0,89	1,25	YBT1	0,49	0,85	ARO2	0,35	0,65
DAK1	0,29	0,93	VMA9	0,16	0,71	GSH2	0,30	0,71	YNL040W	0,85	1,07
YMR196W	0,27	0,92	AFG1	0,01	0,61	LDB19	0,07	0,53	MMF1	0,33	0,63
YKL033W-A	1,46	1,78	CWH41	0,56	0,99	PRP3	0,15	0,58	FMP42	0,80	1,02
VPS9	0,73	1,22	SRB6	0,50	0,94	TOR2	0,12	0,56	SFB2	0,44	0,72
LYS20	0,36	0,96	ALG1	0,23	0,75	IES6	0,24	0,64	TVP38	0,22	0,54
FCY2	1,31	1,65	FUM1	0,77	1,14	HIS1	0,34	0,72	YSA1	0,20	0,52
HOL1	0,08	0,78	RPL15b	0,16	0,69	FYV6	0,52	0,85	NAS6	0,13	0,46
TYR1	0,99	1,40	ALG6	0,00	0,59	BOL1	0,49	0,82	AAT2	0,80	1,01
CPA2	1,17	1,54	REH1	0,31	0,79	VCX1	0,22	0,61	ILS1	0,15	0,46
BRE4	1,26	1,61	SNR33	0,40	0,86	LSC2	1,78	1,93	GON7	0,63	0,86
MET7	0,31	0,90	QCR6	0,16	0,69	YGR182C	0,90	1,16	UBA4	0,24	0,54
ADE3	0,23	0,84	PUT3	0,67	1,06	DUR1,2	0,83	1,09	VPH1	0,45	0,70
MET4	0,39	0,95	DAP2	0,82	1,17	PRO2	0,04	0,47	APN1	0,45	0,70
PEX21	1,52	1,80	FOL2	0,95	1,27	RTC1	0,26	0,64	YPL199C	0,29	0,57
YLR001C	0,65	1,12	HTS1	0,32	0,79	YMR130W	0,04	0,47	BET3	0,38	0,64
APE1	0,58	1,07	APQ12	0,66	1,04	PEX2	0,83	1,08	YCT1	1,59	1,71
CWC15	0,41	0,94	TAF14	0,07	0,60	BAT1	1,08	1,30	YPQ1	0,67	0,88
TRP4	1,60	1,86	YDR341C	0,48	0,89	EAF7	0,21	0,58	NPY1	0,15	0,45
GLR1	0,11	0,74	GFD1	0,11	0,62	PBP4	0,34	0,68	TYS1	0,07	0,38
COX5b	1,18	1,51	VMA13	0,30	0,75	AVT1	0,29	0,64	GLT1	0,41	0,65
HNT1	0,22	0,81	GDH1	1,59	1,79	NCR1	0,61	0,89	PAF1	0,23	0,50
ATX1	0,24	0,82	TPN1	0,52	0,91	QNS1	0,66	0,94	AVT3	0,16	0,44
YDR262W	0,50	1,00	YGR017W	0,05	0,56	TOR1	0,38	0,70	UGA1	0,29	0,54
URE2	0,08	0,71	RMD8	0,36	0,78	YPP1	0,72	0,97	GYP8	1,08	1,23
ATG26	1,00	1,36	NUP116	0,04	0,54	DSL1	0,24	0,58	SNC1	0,26	0,52
GRX5	0,57	1,02	AIM7	0,34	0,75	SPT7	0,18	0,53	QCR7	0,54	0,75
OCA5	0,60	1,04	VPS45	0,34	0,75	INM2	0,51	0,80	ERG8	0,65	0,84

Chapter 8– Annex

Gene/ORF	logFC (175)	logFC (290)	Gene/ORF	logFC (175)	logFC (290)	Gene/ORF	logFC (175)	logFC (290)	Gene/ORF	logFC (175)	logFC (290)
LEU9	0,34	0,58	YPR146C	0,46	0,62	NSG2	0,70	0,69	OAZ1	0,96	0,60
ARG5,6	1,88	1,96	MIC27	0,54	0,69	VMA5	0,64	0,62	ARG8	1,51	1,27
PFY1	0,15	0,42	LST4	0,44	0,58	NIF3	0,44	0,41	GAP1	2,13	1,96
YBR027C	0,73	0,91	MTC5	0,43	0,56	DLD3	1,73	1,71	CCC1	1,05	0,61
FAA1	0,83	1,00	HMRA1	0,27	0,42	MIC12	0,74	0,71	SAM3	2,55	2,38
CDC36	0,33	0,56	KRS1	0,59	0,70	ATP5	0,62	0,58	YFR006W	0,99	0,40
CUE1	0,25	0,50	CIT1	1,46	1,53	PUS1	0,67	0,62	BAT2	1,88	1,45
VMA8	0,49	0,70	LAP3	1,56	1,61	RBL2	1,05	1,01	YJR149W	1,89	1,44
VTC1	0,27	0,51	BAS1	0,91	1,00	DIP5	2,69	2,68	CAB2	1,73	1,23
LYS1	0,70	0,88	SPO14	0,36	0,47	YMR226C	0,97	0,92	PTR2	2,36	1,99
SGT1	0,51	0,71	YMC2	1,33	1,38	KAE1	0,54	0,47	NIT3	1,82	1,16
ZRT3	0,33	0,55	RSM18	0,45	0,55	CSL4	0,46	0,37	SER33	2,12	1,34
UBC4	0,35	0,57	OM14	0,94	1,00	AAP1	1,13	1,08	MET16	2,84	2,38
PMP3	0,50	0,70	PGC1	0,35	0,45	GCN4	0,69	0,60	RAD59	1,98	0,80
CMD1	0,21	0,45	ACP1	0,41	0,49	SOK1	0,62	0,51	MET1	3,22	2,77
YJL055W	0,77	0,94	AIR1	1,01	1,06	PDR16	0,71	0,61	MET2	3,19	2,72
ATP7	0,32	0,54	TAF9	0,95	1,00	SPO24	2,04	2,00	MET10	3,07	2,25
MCX1	0,38	0,59	SNR31	0,64	0,71	BOL3	0,70	0,59	DIC1	2,66	1,23
THR4	0,37	0,58	VTC3	0,37	0,45	ALT1	1,76	1,70	YHI9	3,33	2,20
MRPL10	0,93	1,07	YPL225W	0,48	0,55	DUG1	0,88	0,75	PDH1	3,65	1,71
YPT52	0,64	0,82	YGL140C	0,75	0,81	LYP1	0,77	0,63	MET5	4,00	2,16
YDR306C	0,18	0,41	HAA1	0,46	0,52	ILV2	1,34	1,24	SUL1	4,83	2,25
PMT6	0,46	0,65	RPB4	0,70	0,75	IDH1	0,88	0,72			
GLO2	0,52	0,70	LEU1	2,12	2,13	GPN2	1,04	0,89			
GLE1	0,44	0,63	ELO1	0,66	0,70	SLX9	0,75	0,56			
BNA1	1,41	1,51	PAN5	1,15	1,18	CMC4	1,89	1,80			
YDL086W	0,24	0,45	EMC6	0,61	0,64	YHR045W	1,35	1,21			
MRP7	0,38	0,57	TFA2	0,54	0,57	UTR4	1,33	1,18			
YRA2	0,35	0,54	VMA10	0,48	0,50	YGR266W	0,90	0,69			
EMC3	0,53	0,69	VTC2	0,72	0,73	GNP1	1,17	1,00			
YPR098C	0,21	0,41	FUB1	0,47	0,48	LSC1	0,84	0,56			
YIL156W-B	0,38	0,56	POP4	0,64	0,65	RRP43	0,79	0,47			
ATP20	0,60	0,75	DRE2	1,37	1,36	VMA11	1,04	0,76			

Table 8.11. Genes downregulated in the lag phase in both ITA concentrations

Gene/ORF	logFC (175)	logFC (290)	Gene/ORF	logFC (175)	logFC (290)	Gene/ORF	logFC (175)	logFC (290)	Gene/ORF	logFC (175)	logFC (290)
CGR1	-0,04	-1,99	ALB1	-0,22	-1,51	SNR40	-0,05	-0,81	MNN11	-0,11	-0,76
YCR087C-A	-0,12	-2,03	DBP10	-0,09	-1,20	TSC10	-0,02	-0,75	SUP35	-0,18	-0,86
SPB1	-0,13	-1,95	NIP7	-0,32	-1,77	MEU1	-0,12	-0,92	TOM20	-0,03	-0,61
SAS10	-0,03	-1,59	RPA43	-0,24	-1,54	LHP1	-0,26	-1,16	NUP145	-0,07	-0,66
URA7	-0,27	-2,36	MAK21	-0,15	-1,31	LSG1	-0,44	-1,52	NHA1	-0,06	-0,64
NOP13	-0,07	-1,61	NOP7	-0,38	-1,83	YBL081W	-0,28	-1,19	IZH2	-0,33	-1,07
AAH1	-0,28	-2,32	MTC7	-0,20	-1,38	POR1	-0,28	-1,19	RPO31	-0,18	-0,82
PWP1	-0,10	-1,62	NHP2	-0,03	-1,01	EFM1	-0,03	-0,76	FPR4	-0,74	-1,81
YTM1	-0,01	-1,37	NPA3	-0,08	-1,11	RDI1	-0,06	-0,79	PRY2	-0,57	-1,47
HCA4	-0,01	-1,36	NOP58	-0,19	-1,32	TIF4631	-0,42	-1,45	DBP5	-0,13	-0,73
NOC3	-0,11	-1,62	XPT1	-0,04	-1,00	RPG1	-0,21	-1,05	YRB1	-0,08	-0,65
YBL028C	-0,02	-1,38	SGD1	-0,18	-1,26	CLU1	-0,26	-1,14	RPB8	-0,07	-0,63
NOP6	-0,06	-1,48	GAR1	-0,34	-1,62	CAF20	-0,25	-1,10	SBA1	-0,30	-0,97
NOC2	-0,02	-1,37	ELP2	-0,04	-0,95	SNU13	-0,33	-1,26	RAT1	-0,06	-0,60
NOP4	-0,06	-1,45	EMG1	-0,27	-1,41	KRE33	-0,66	-1,93	OLE1	-0,30	-0,96
DIM1	-0,17	-1,76	TMA46	-0,02	-0,89	PSA1	-0,19	-0,98	ADH6	-0,24	-0,87
MTR4	-0,09	-1,52	TIF3	-0,05	-0,94	PRS3	-0,15	-0,91	BLM10	-0,22	-0,82
PRP43	-0,01	-1,29	TIF35	-0,10	-1,03	NOP56	-0,41	-1,37	MOG1	-0,48	-1,23
ESF2	-0,01	-1,26	SUI1	-0,06	-0,95	POL5	-0,38	-1,31	IMD4	-0,94	-2,12
RLP7	-0,06	-1,34	NUG1	-0,52	-1,97	ARB1	-0,24	-1,06	APE3	-0,07	-0,60
HPT1	-0,02	-1,25	RVB1	-0,13	-1,08	KEL3	-0,52	-1,59	CCT4	-0,10	-0,64
UTP21	-0,27	-1,89	CBF5	-0,22	-1,25	GLN4	-0,32	-1,19	SGT2	-0,03	-0,52
RLP24	-0,13	-1,45	UTP10	-0,12	-1,04	TRM3	-0,20	-0,97	YMR304C-A	-0,93	-2,06
ENP2	-0,16	-1,51	NEW1	-0,19	-1,16	URB1	-0,32	-1,19	SWR1	-0,22	-0,80
DBP3	-0,16	-1,51	GET3	-0,04	-0,88	RRP5	-0,60	-1,74	UTR2	-0,79	-1,76
NSA1	-0,14	-1,41	COQ2	-0,16	-1,10	SUP45	-0,05	-0,72	PRE9	-0,15	-0,70
NOP2	-0,14	-1,40	HXK2	-0,08	-0,92	RMD1	-0,03	-0,69	DPH5	-0,88	-1,93
KRR1	-0,27	-1,72	RPC53	-0,03	-0,83	GCD11	-0,29	-1,10	ARF1	-0,21	-0,77
ECM16	-0,26	-1,68	RPA190	-0,61	-2,06	UTP20	-0,49	-1,48	YFR052C-A	-1,18	-2,63
ERB1	-0,35	-1,91	CCT2	-0,03	-0,82	EIS1	-0,17	-0,87	IKI3	-0,05	-0,55
YOR309C	-0,08	-1,22	TRM8	-0,35	-1,41	RPN1	-0,11	-0,77	LIA1	-0,74	-1,63
BUD27	-0,02	-1,09	HYP2	-0,41	-1,52	EFM4	-0,79	-2,07	TCB2	-0,27	-0,85
RRP15	-0,03	-1,11	ETT1	-0,07	-0,87	EMW1	-0,16	-0,85	LYS9	-0,79	-1,71
LOC1	-0,14	-1,34	SMB1	-0,08	-0,88	MVD1	-0,09	-0,73	STH1	-0,03	-0,50

Gene/ORF	logFC (175)	logFC (290)	Gene/ORF	logFC (175)	logFC (290)	Gene/ORF	logFC (175)	logFC (290)	Gene/ORF	logFC (175)	logFC (290)
SUI3	-0,14	-0,65	RVB2	-0,34	-0,81	CTR9	-0,14	-0,43	ASP1	-0,57	-0,86
YMR295C	-0,46	-1,12	SSZ1	-0,47	-0,98	GCV3	-0,44	-0,80	AIM45	-0,33	-0,57
ERG3	-0,66	-1,45	UFD1	-0,10	-0,49	PRE2	-0,40	-0,75	ATG5	-0,66	-0,96
SAH1	-0,72	-1,54	PIS1	-0,13	-0,52	EMC4	-0,17	-0,47	RPL3	-0,71	-1,02
KAP123	-0,84	-1,75	GAS1	-0,87	-1,58	YDL157C	-0,41	-0,76	MIR1	-0,54	-0,80
DUS3	-0,29	-0,85	ERG11	-0,30	-0,73	RPL32	-1,00	-1,57	PAP2	-0,45	-0,69
PSE1	-0,13	-0,61	KIN2	-0,21	-0,61	GCD2	-0,21	-0,50	SSH1	-0,43	-0,66
ERG25	-0,80	-1,67	PKR1	-0,24	-0,65	SER2	-0,42	-0,77	GDA1	-1,00	-1,37
SRM1	-0,16	-0,66	MTC1	-0,57	-1,11	RPL38	-0,93	-1,44	ERG4	-0,36	-0,58
CPR2	-0,16	-0,66	TIF1	-0,93	-1,66	SEC63	-0,28	-0,59	ZUO1	-0,81	-1,11
KAP104	-0,39	-0,98	NOP10	-0,18	-0,57	YNL247W	-0,31	-0,61	CHS7	-1,05	-1,42
TIF34	-0,44	-1,05	DEF1	-0,59	-1,13	SOD1	-0,25	-0,54	ERG26	-0,35	-0,56
ELO3	-0,95	-1,94	RDL1	-0,22	-0,62	SBP1	-0,25	-0,54	DYS1	-0,27	-0,46
NCL1	-0,67	-1,42	GRS1	-0,49	-0,97	TEF1	-0,65	-1,04	RQC2	-0,33	-0,53
IPT1	-0,54	-1,19	GCD1	-0,15	-0,52	GUK1	-0,58	-0,93	VPS1	-0,55	-0,79
FUN12	-0,55	-1,19	EMP70	-0,12	-0,47	RPL30	-0,80	-1,22	TOM70	-0,52	-0,75
SES1	-0,41	-0,98	CYK3	-0,12	-0,47	SCS2	-0,40	-0,71	IPP1	-0,23	-0,42
TMA20	-0,24	-0,74	TIM10	-0,26	-0,66	VPS21	-0,33	-0,62	BFR1	-0,61	-0,86
TMA22	-0,38	-0,93	OLA1	-0,84	-1,46	WTM1	-0,18	-0,43	SHM1	-0,30	-0,49
PRT1	-0,21	-0,69	EGD1	-0,63	-1,15	ECM33	-1,30	-1,90	GDT1	-0,25	-0,44
URA5	-0,24	-0,72	RIM1	-0,14	-0,49	SUB2	-0,41	-0,71	SSO1	-0,34	-0,53
CDC60	-0,04	-0,46	CDC48	-0,33	-0,73	AUR1	-0,24	-0,49	SPT16	-0,42	-0,61
IMD3	-0,98	-1,91	SSS1	-0,16	-0,51	HSL1	-0,83	-1,23	NOP1	-0,84	-1,11
NAB2	-0,45	-1,02	MIS1	-1,00	-1,69	SEN1	-0,30	-0,57	YSY6	-0,22	-0,39
SEC62	-0,10	-0,52	CHO1	-1,32	-2,24	TFG2	-0,31	-0,57	RPB11	-0,70	-0,94
HNM1	-0,59	-1,20	RPN2	-0,25	-0,61	ERG6	-0,74	-1,10	ISW1	-0,73	-0,97
MSB2	-0,39	-0,90	YEF3	-0,74	-1,27	RPS3	-0,90	-1,30	RPS15	-1,31	-1,68
EFB1	-0,64	-1,28	YER156C	-0,48	-0,90	SAC6	-0,86	-1,25	SSB2	-1,23	-1,58
EGD2	-0,29	-0,76	VAN1	-0,25	-0,61	ADO1	-0,67	-1,00	PUB1	-0,57	-0,78
BUG1	-0,39	-0,91	HSP60	-0,56	-1,01	SSB1	-1,24	-1,76	BGL2	-0,29	-0,45
ERG10	-0,02	-0,39	HYR1	-0,30	-0,66	FTR1	-1,01	-1,44	SUI2	-0,38	-0,54
THS1	-0,24	-0,69	FPR3	-0,38	-0,76	CRP1	-0,53	-0,82	TMA108	-0,34	-0,50
ABP140	-0,52	-1,07	YLR257W	-0,35	-0,71	HRP1	-0,64	-0,95	GND1	-1,42	-1,76
PHS1	-0,22	-0,64	TIM50	-0,44	-0,82	FPR1	-0,34	-0,59	RPS13	-1,29	-1,59

Chapter 8– Annex

Gene/ORF	logFC (175)	logFC (290)	Gene/ORF	logFC (175)	logFC (290)	Gene/ORF	logFC (175)	logFC (290)	Gene/ORF	logFC (175)	logFC (290)
ASC1	-1,47	-1,81	TSA1	-0,64	-0,70	YBR056W	-0,83	-0,71	GCV1	-1,06	-0,66
RPS12	-0,99	-1,21	DFG5	-0,84	-0,90	EMI2	-0,90	-0,78	PUN1	-1,48	-0,90
ACB1	-0,68	-0,86	EFT2	-1,33	-1,40	SEC61	-0,77	-0,64	EHT1	-1,16	-0,66
RTN1	-0,49	-0,64	EEB1	-0,75	-0,80	PDI1	-1,08	-0,92	SED1	-2,13	-1,25
DUT1	-0,38	-0,52	OYE2	-1,06	-1,12	TUB1	-0,71	-0,58	PEP12	-1,26	-0,71
STM1	-1,36	-1,65	CHS5	-0,66	-0,70	GSP1	-0,70	-0,55	PSD1	-1,19	-0,64
YGR210C	-0,40	-0,54	YET3	-0,66	-0,69	TRX2	-0,70	-0,52	YDL012C	-1,06	-0,55
UBP12	-0,32	-0,45	TMA19	-1,45	-1,50	SCP160	-1,51	-1,22	SHM2	-1,86	-1,06
TRR1	-0,59	-0,74	RPL29	-1,11	-1,14	MYO5	-0,66	-0,48	HSP31	-2,16	-1,16
CCW14	-0,91	-1,10	ADK1	-0,74	-0,76	NCP1	-1,19	-0,93	CHS1	-1,76	-0,92
XRN1	-0,41	-0,54	ADE1	-0,83	-0,85	MCA1	-0,71	-0,52	STF2	-1,90	-0,59
FKS1	-1,16	-1,37	SEC53	-1,26	-1,30	PAB1	-1,05	-0,81	SPI1	-2,29	-0,70
MYO2	-0,38	-0,49	RHO1	-0,65	-0,66	AIM17	-1,15	-0,87	YDL158C	-0,56	-0,99
RPL28	-1,14	-1,34	YIP3	-1,59	-1,60	OPI3	-1,71	-1,32	MIA40	-0,23	-0,91
TOP2	-0,70	-0,84	PMT4	-0,73	-0,73	ARC1	-0,62	-0,42	TSR3	-0,18	-0,79
RPL5	-1,35	-1,58	MAP1	-0,60	-0,61	ADH1	-1,19	-0,90	SXM1	-0,20	-0,61
ERP1	-0,63	-0,76	RPS2	-1,32	-1,32	YKL066W	-1,36	-1,02	SOL3	-0,45	-0,66
YLR342W-A	-1,05	-1,22	DCW1	-0,93	-0,92	ADE5,7	-1,49	-1,12	SEC66	-0,19	-0,54
ADE6	-0,90	-1,05	MDH1	-0,63	-0,63	YHR138C	-1,46	-1,08	UTP22	-0,35	-1,32
KAR2	-0,43	-0,53	DOA1	-0,63	-0,62	ACC1	-1,03	-0,74	SSC1	-0,42	-0,92
LAP2	-0,46	-0,56	SCW4	-0,76	-0,74	ADE17	-2,33	-1,70	SMF2	-0,19	-0,57
TMA7	-0,58	-0,68	NCW2	-1,90	-1,86	PMI40	-1,22	-0,90	FYV7	-0,01	-1,44
GEA2	-0,65	-0,76	URA1	-1,69	-1,65	PMC1	-1,71	-1,27	GAS3	-1,17	-1,26
RPO21	-0,73	-0,84	CAP1	-0,57	-0,53	FAS2	-1,60	-1,18	HXT4	-1,28	-2,74
RPS5	-1,43	-1,60	TEF2	-1,21	-1,14	DIA1	-1,61	-1,19	TRK1	-0,13	-0,64
RPS31	-1,08	-1,21	DLD1	-1,33	-1,24	YNL208W	-1,44	-1,06	POL2	-0,93	-0,85
ITR1	-1,62	-1,81	EFT1	-1,16	-1,06	FAS1	-1,98	-1,43	SIN3	-0,91	-0,78
TUB2	-0,80	-0,90	PMT1	-0,88	-0,80	RPP0	-1,70	-1,22	SFH1	-0,13	-0,59
PRO3	-0,48	-0,56	YOR385W	-1,87	-1,71	PDC1	-2,67	-1,85	CRH1	-0,93	-0,88
TEF4	-1,49	-1,63	CIS3	-1,54	-1,41	PEP1	-0,70	-0,43	SPB4	-0,39	-1,01
BDF2	-0,90	-0,98	ADE13	-0,99	-0,89	HAL5	-1,14	-0,78	VRG4	-0,63	-0,75
YDR154C	-1,15	-1,24	DPM1	-1,19	-1,06	PRB1	-2,34	-1,61	YLR202C	-0,29	-0,72
URA4	-0,98	-1,05	VMA1	-0,49	-0,40	NPT1	-0,79	-0,48	RRP17	-0,24	-1,19
RPE1	-0,65	-0,71	YNL134C	-2,00	-1,77	FIS1	-0,95	-0,60	ROK1	-0,07	-1,14
YOR062C	-1,13	-0,78	COP1	-0,66	-0,77	POL30	-2,24	-2,74	TOM40	-0,77	-0,87
ENP1	-0,17	-1,61	PST1	-2,17	-0,86	PMR1	-0,97	-0,57	YPS3	-2,76	-1,98
PWP2	-0,01	-1,87	NCB2	-0,01	-0,51	YPL245W	-0,40	-1,19	CYM1	-0,33	-0,66
IST2	-0,39	-0,53	ERG24	-0,27	-0,95	UBP13	-0,33	-0,82	TOH1	-1,13	-1,35
TIF5	-0,05	-0,84	RET1	-0,17	-1,00						

Table 8.12. Genes upregulated in the exponential phase with 290 mM ITA

Gene	log2FC	Gene	log2FC	Gene	log2FC
DAL3	2,79	PTR2	1,17	DOG1	0,8778594
MEP2	2,63	DUR1,2	1,16	NRE1	0,8772405
DAL80	2,61	YIR042C	1,15	CPS1	0,8741983
MET5	2,39	INO1	1,15	URA1	0,8685624
DAL2	2,18	YHB1	1,1384419	OPT1	0,8665519
DAL5	2,14	TPO2	1,1333766	UGA2	0,8557151
MET10	1,96	ZRT1	1,1279753	TDA6	0,8473244
MET16	1,90	DIC1	1,1093869	CIT1	0,8382347
YLR053C	1,83	SER3	1,0978811	BAP3	0,8356096
SUL2	1,83	YNL284C-B	1,0969717	SER33	0,8303486
SUL1	1,83	CRC1	1,0958462	IDP1	0,8264279
MET1	1,72	YAT2	1,0923633	DLD3	0,8235869
MET14	1,72	DPI8	1,0491792	HPA3	0,7920616
AMD2	1,68	ICY1	1,028895	TDH1	0,7900039
MET28	1,63	CMC4	1,0276592	ZEO1	0,7882932
MET2	1,63	YHR112C	1,0143059	GPP2	0,7838462
MHT1	1,60	DIP5	1,0056589	FUI1	0,7777101
MET32	1,58	PDH1	0,9843991	YIL001W	0,7739273
YBL029W	1,53	DCG1	0,9827071	SAM3	0,7708115
DUR3	1,50	YHR140W	0,9681406	YGL117W	0,7647672
QDR2	1,50	UGA4	0,9548654	CHO2	0,7227793
YRO2	1,33	PRX1	0,9539433	MNN4	0,7139412
GPM2	1,32	HXT4	0,9418561	APE1	0,6880404
STR3	1,30	COA2	0,9400404	SOM1	0,6790821
GAP1	1,26	WSC4	0,9385933	YJR011C	0,674265
AQR1	1,25	MET3	0,9327181	YGR125W	0,6000589
ATG41	1,25	RAD59	0,910093	LOG1	0,5984935
SPO11	1,25	GTO1	0,9050691	SNR31	0,5870957
CIT2	1,24	SPO24	0,9039725	ORT1	0,5684764
YDR210C-D	1,24	YGL185C	0,8921907	SIP3	0,5404728
LSC2	1,22	FMO1	0,8896994	GLO2	0,5382593
DAL7	1,22	AAD10	0,8838395	MET6	1,18
SAM2	1,22	RAD34	0,882698		
MXR1	1,21	QDR3	0,8817965		

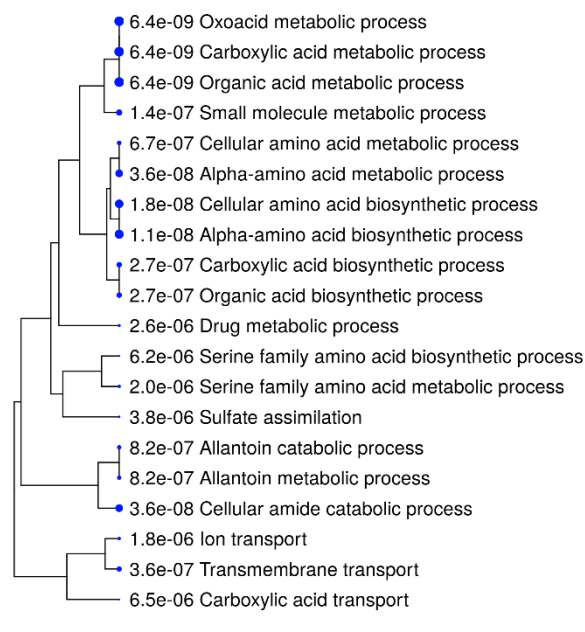


Figure 8.5. Hierarchical clustering tree of the enriched functional categories in the list of 43 genes upregulated in the presence of ITA (in the three conditions studied here). Analysis was performed in ShinyGO.

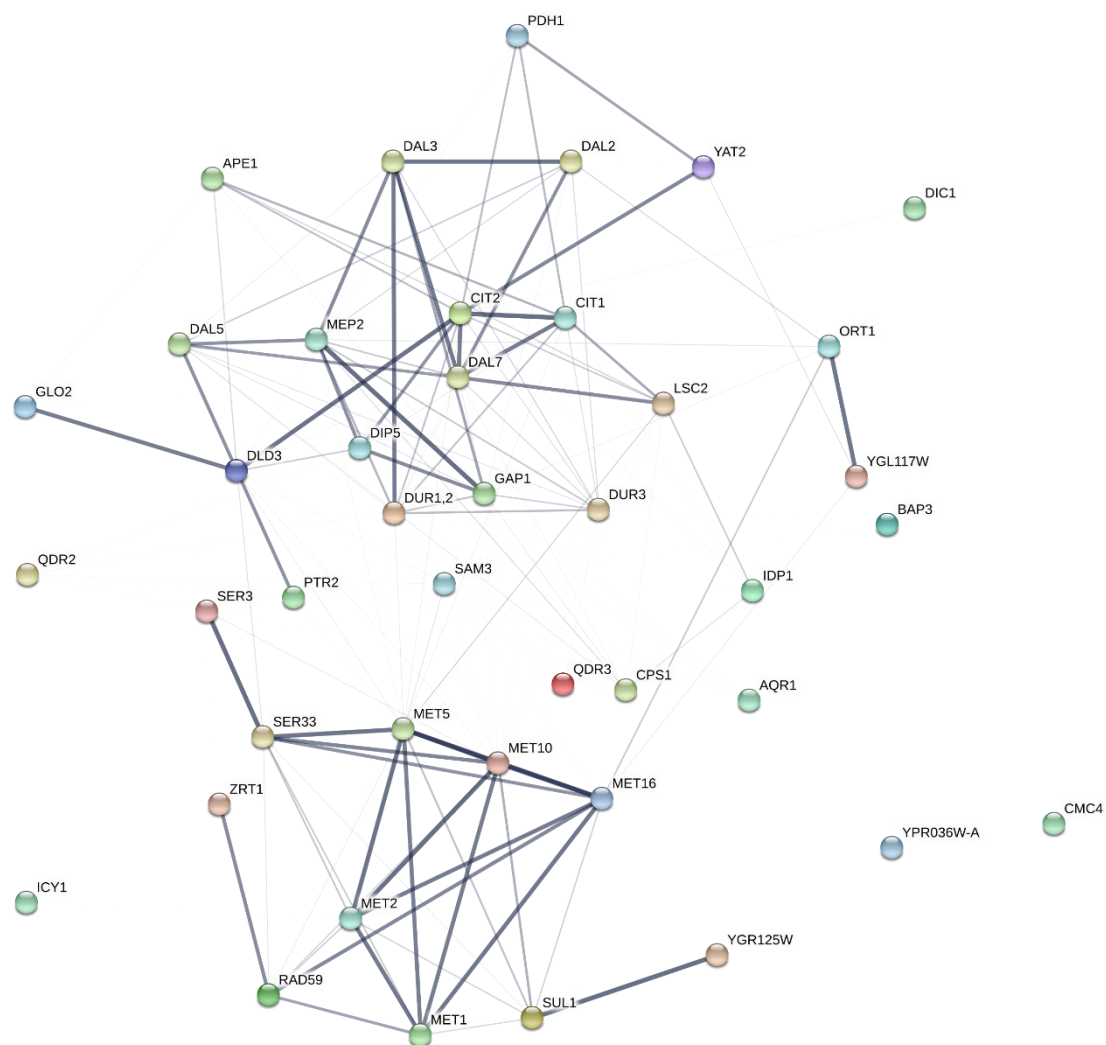


Figure 8.6. Functional interaction network generated by STRING on the 43 upregulated genes. Only interactions based on experiments, databases and co-expression were considered. The line thickness of each interaction indicates the strength of data support.