Development of non-model bacteria as cell factories

Focusing on metabolic modelling and engineering of *Rhodothermus marinus* and *Lactobacillus reuteri*

Þórdís Kristjánsdóttir

Faculty of Industrial Engineering, Mechanical Engineering and Computer Science

University of Iceland

2022

Development of non-model bacteria as cell factories

Focusing on metabolic modelling and engineering of *Rhodothermus marinus* and *Lactobacillus reuteri*

Þórdís Kristjánsdóttir

Dissertation submitted in partial fulfillment of a
Philosophiae Doctor degree in Bioengineering

PhD Committee
Professor Steinn Guðmundsson
Professor Guðmundur Ó. Hreggviðsson
Dr. Ólafur H. Friðjónsson
Dr. Markus J. Herrgård

Opponents
Dr. Sheila Ingemann Jensen
Dr. Eduard Kerkhoven

University of Iceland
Faculty of Industrial Engineering, Mechanical Engineering and Computer Science
Dunhagi 5
Reykjavík, Iceland

Matís ohf.
Biotechnology Division
Vínlandsleið 12
Reykjavík, Iceland

Development of non-model bacteria as cell factories: Focusing on metabolic modelling and engineering of *Rhodothermus marinus* and *Lactobacillus reuteri*
Short title: Development of non-model bacteria as cell factories
Dissertation submitted in partial fulfillment of a *Philosophiae Doctor* degree in Bioengineering

ISBN: 978-9935-9647-3-1

Faculty of Industrial Engineering, Mechanical Engineering and Computer Science
School of Engineering and Natural Sciences
University of Iceland
Dunhagi 5
107 Reykjavik
Iceland

Telephone: 525 4000

Bibliographic information:
Þórdís Kristjánsdóttir, 2022, Development of non-model bacteria as cell factories: Focusing on metabolic modelling and engineering of *Rhodothermus marinus* and *Lactobacillus reuteri*, PhD dissertation, Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland, 167 pp.

Author ORCID: 0000-0003-2244-1723

Printing: Háskólaprent
Reykjavik, Iceland, January 2022

This work is dedicated to my mother, Dr. Sigríður Hjörleifsdóttir

fyrir stuðninginn, spjöllin og auðvitað Rhodothermus marinus

Abstract

The current climate crisis and human overpopulation demands more sustainable technologies for industrial chemical synthesis, such as biorefineries. The aim of this project was to study the potential of two non-model microorganisms to produce industrially relevant chemicals from renewable biomass, through metabolic modelling and engineering. Model organisms are often used for these applications, but due to extreme environments that can occur in biorefineries, this is not always optimal. *Rhodothermus marinus* is a thermophilic bacterium that can use polysaccharides from complex biomass and produces carotenoids. *Lactobacillus reuteri* is a bacterium that has a wide growth temperature range compared to other LAB and is a native producer of 1,3-propanediol and B12-vitamin.

Genome-scale metabolic models of the two species were reconstructed, manually curated and validated. The *L. reuteri* model was used to predict how a gene deletion and different carbon sources affected the flux of main metabolites in central metabolism, and to suggest strategies for 1-propanol production. Different aspects of the *R. marinus* metabolism were reviewed and subsequently, both cell densities and carotenoid production were investigated under different growth and genotype conditions.

Native *R. marinus* strains produce γ-carotenoids in large quantities. The carotenoid biosynthetic genes and pathway in *R. marinus* were studied with the aim of making a platform strain for production of carotenoids of industrial interest. Subsequently, a mutant was obtained that produced lycopene, which is a common precursor to the synthesis of a variety of carotenoids. This is the first *R. marinus* strain that produces an industrially relevant carotenoid, which in turn is a vital step towards obtaining a robust *R. marinus* strain that can be used in industry.

Genetic engineering of *R. marinus* is currently limited by marker selection, which are difficult to reuse for further genetic manipulation. A thermostable CRISPR-Cas9 system was developed, which gives hope for more advanced and easier genetic engineering of *R. marinus* in the future. This work strengthened our understanding on the metabolism of *R. marinus* and *L. reuteri*, it produced metabolic models that increase their utility as cell factories and it provided a thermophilic platform strain for industrial carotenoid production.

Útdráttur

Vegna loftlagsbreytinga og offjölgunar mannkyns er mikil þörf á sjálfbærum aðferðum í efnaiðnaði, s.s. notkun lífmassavera. Markmið þessa verkefnis var að rannsaka möguleikann á að nota tvær örverur, sem ekki eru skilgreindar sem tilraunalífverur, þ.e. eru tiltölulega lítt þekktar, til framleiðslu iðnaðarefna úr endurnýjanlegum lífmassa. Til þess voru efnaskiptalíkön og -verkfræði notuð. Yfirleitt eru tilraunalífverur notaðar sem frumuverksmiðjur, en það getur verið óákjósanlegt að nota þær í lífmassaverum þar sem umhverfið getur verið öfgakennt. *Rhodothermus marinus* er hitakær baktería sem getur notað fjölsykrur úr flóknum lífmassa og framleitt karótenóíða. *Lactobacillus reuteri* er baktería sem getur vaxið við breytt hitastigsbil miðað við aðrar LAB og getur búið náttúrulega til 1,3-própanedíól og B12 vítamín.

Efnaskiptalíkkön af báðum bakteríunum voru smíðuð, farið var yfir þau handvirkt og þau prófuð. *L. reuteri* líkanið var notað til að spá fyrir um hvernig útsláttur ákveðins gens og mismunandi kolefnisgjafar hafði áhrif á flæði helstu lífefna í miðlægu efnaskiptanetinu, og til að finna leiðir til að búa til 1-própanól. Þónokkrir hlutar af efnaskiptaneti *R. marinus* voru skoðaðir sérstaklega og í kjölfarið var frumuþéttni og karótenóíð framleiðsla rannsökuð við mismunandi vaxtar- og arfgerðaraðstæður.

R. marinus framleiðir náttúrulega γ-karótenóíða í miklu magni. Karótenóíð nýmyndunargenin og -ferillinn í *R. marinus* voru rannsökuð með það að markmiði að gera grunnstofn fyrir framleiðslu á iðnaðarkarótenóíðum. Í framhaldinu fékkst stökkbrigði sem býr til lycopene, sem er sameiginlegur forveri margra karótenóíða. Þetta er fyrsti *R. marinus* stofninn sem býr til markaðsvænt karótenóíð, og þetta er því mikilvægt skref í átt að stofni sem hægt er að nota í iðnaði.

Núverandi aðferðir sem notaðar eru til að erfðabreyta *R. marinus* eru takmarkaðar af fjölda valgena, sem erfitt er að endurnota fyrir frekari erfðabreytingar. Hitaþolið CRISPR-Cas9 kerfi var þróað, sem gefur vonir um víðtækari og auðveldari erfðabreytingar í *R. marinus* í framtíðinni. Afrakstur verkefnisins er aukinn skilningur á efnaskiptum *R. marinus* og *L. Reuteri*. Efnaskiptalíkönin sem smíðuð voru auka möguleika á hagnýtingu og *R. marinus* stökkbrigðið sem framleiðir lycopene er hægt að nota sem grunnstofn fyrir frekari stökkbreytingar til framleiðslu fleiri iðnaðarkarótenóíða.

Table of Contents

Abstract vi
Útdráttur vii
Table of Contents viii
List of Papers x
Acknowledgements xi
Abbreviations xii

1 Hypotheses and Objectives 1

2 Introduction 3
2.1 Microbial cell factories 5
2.1.1 Biorefinery 5
2.1.2 Non-model microorganisms as cell factories 6
2.1.3 Microbial genetic tools 8
2.2 Target organisms 10
2.2.1 *Lactobacillus reuteri* 10
2.2.2 *Rhodothermus marinus* 12
2.2.3 *Thermus thermophilus* 14
2.3 Genome-scale metabolic models 16
2.3.1 Reconstruction 16
2.3.2 Flux balance analysis 17
2.3.3 Applications 18
2.4 Carotenoids 20
2.4.1 Properties and functions 20
2.4.2 Industrial production 21
2.4.3 Structures and biosynthesis 22

3 Summary of the work 27
3.1 My contribution to the papers 29
3.2 A metabolic reconstruction of *Lactobacillus reuteri* JCM 1112 and analysis of its potential as a cell factory (Paper 1) 31
3.2.1 Reconstruction of the genome-scale metabolic model 31
3.2.2 Model validation 32
3.2.3 *In-silico* analysis of *L. reuteri* as a cell factory 35
3.3 Engineering the carotenoid biosynthetic pathway in *Rhodothermus marinus* for lycopene production (Paper 2) 37
3.3.1 Carotenoid biosynthetic genes 37
3.3.2 Genetic modifications to alter carotenoid production 38
3.3.3 Carotenoid biosynthetic pathway 39
3.4 A genome-scale metabolic reconstruction provides insight into the metabolism of the thermophilic bacterium *Rhodothermus marinus* (Paper 3) 41
3.4.1 Reconstruction of the genome-scale metabolic model 41
3.4.2 Comparing genomes of strains DSM 4252^T and ISCaR-493 43
3.4.3 Model validation 43
3.4.4 Effects of external factors on carotenoid production 44
3.5 Other papers 46
3.5.1 Characterization of carotenoids in *Rhodothermus marinus* (Paper 4) 46

3.5.2 Composition analysis and minimal treatments to solubilize polysaccharides from the brown seaweed *Laminaria digitata* for microbial growth of thermophiles (Paper 5) 47
3.5.3 Efficient genome editing of an extreme thermophile, Thermus thermophilus, using a thermostable Cas9 variant (Paper 6) 48
3.6 Conclusions 49

References 51

Papers 63
A metabolic reconstruction of Lactobacillus reuteri JCM 1112 and analysis of its potential as a cell factory 65
Abstract 68
1. Background 69
2. Materials and methods 71
3. Results and discussion 77
4. Conclusions 97
5. Declarations 98
References 99
Engineering the carotenoid biosynthetic pathway in Rhodothermus marinus for lycopene production 107
Abstract 108
1. Introduction 108
2. Methods and materials 110
3. Results 115
4. Discussion 123
References 127
A genome-scale metabolic reconstruction provides insight into the metabolism of the thermophilic bacterium Rhodothermus marinus 133
Abstract 134
Importance 134
1. Introduction 135
2. Methods 137
3. Results and discussion 142
4. Conclusions 159
References 161

List of Papers

1. **Kristjansdottir, T.**, Bosma, E. F., Branco dos Santos, F., Özdemir, E., Herrgård, M. J., Franca, L., Ferreira, B., Nielsen, A. T., Gudmundsson, S. "A metabolic reconstruction of *Lactobacillus reuteri* JCM 1112 and analysis of its potential as a cell factory". *Microb Cell Fact* 18, 186 (2019). doi: 10.1186/s12934-019-1229-3

2. **Kristjansdottir, T.**, Ron, E. Y. C., Delgado, D. M., Fridjonsson, O. H., Turner, C., Bjornsdottir, S. H., Gudmundsson, S., van Niel, E. W. J., Karlsson, E. N., Hreggvidsson, G. O. "Engineering the carotenoid biosynthetic pathway in *Rhodothermus marinus* for lycopene production". *Metab Eng Commun* 11, e00140 (2020). doi: 10.1016/j.mec.2020.e00140.

3. **Kristjansdottir, T.**, Hreggvidsson, G. O., Gudmundsdottir, E. E., Bjornsdottir, S. H., Stefansson, S. K., Fridjonsson, O. F., Karlsson, E. N., Vanhalst, J., Reynisson, B., Gudmundsson, S. "A genome-scale metabolic reconstruction provides insight into the metabolism of the thermophilic bacterium *Rhodothermus marinus*". Manuscript available on bioRxiv. doi: 10.1101/2021.05.17.444423.

4. Ron, E. Y. C., Plaza, M., **Kristjansdottir, T**., Sardari, R. R. R., Bjornsdottir, S. H., Gudmundsson, S., Hreggvidsson, G. O., Turner, C., van Niel, E. W. J., Karlsson, E. N. "Characterization of carotenoids in *Rhodothermus marinus*". *MicrobiologyOpen* 7, 1 (2018). doi: e00536. doi:10.1002/mbo3.536.

5. Allahgholi, L., Sardari, R. R. R., Ara, K. Z. G., **Kristjansdottir, T.**, Aasen, I. M., Fridjonsson, O. H., Brautaset, T., Hreggvidsson, G. O., Karlsson, E. N. "Composition analysis and minimal treatments to solubilize polysaccharides from the brown seaweed *Laminaria digitata* for microbial growth of thermophiles". *Journal of Applied Phycology* 32, 1933-1947 (2020). doi: 10.1007/s10811-020-02103-6.

6. Adalsteinsson, B. T., **Kristjansdottir, T.**, Merre, W., Helleux, A., Dusaucy, J., Tourigny, M., Fridjonsson, O., Hreggvidsson, G. O. "Efficient genome editing of an extreme thermophile, *Thermus thermophilus*, using a thermostable Cas9 variant". *Sci Rep* 11, 9586 (2021). doi: 10.1038/s41598-021-89029-2.

Acknowledgements

I want to thank those who helped and supported me through this work.

First, I thank my supervisor, Professor Steinn Guðmundsson, for the guidance, insights and all the chats about everything and nothing. It was truly great to learn from you. I also thank my co-supervisor, Professor Guðmundur Ó. Hreggviðsson, for the ideas, trust and optimism. I thank my PhD advisory committee, both Dr. Ólafur H. Friðjónsson, who is also my group leader at Matís, for always answering my questions, helping me interpret results and plan experiments, and Dr. Markus J. Herrgård for helpful suggestions.

I am very thankful that I was able to work on this research and on my thesis while working at Matís. I am especially thankful for all the wonderful people I work with at Matís, thank you for all the help, encouragement and friendships. I thank Professor Snædís H. Björnsdóttir for the good advice and valuable insight into *Rhodothermus marinus*. I thank my co-author, Dr. Elleke Bosma for great teamwork and for the *Lactobacillus* knowledge I gained from you. I also thank my colleagues at Lund University, especially Professor Eva Nordberg Karlsson for inviting me to visit your department and learn from you, and Dr. Emanuel Ron for the cooperation.

Finally, I thank my family for all the support and love that helped me get to the end of the rocky road of a PhD. My husband, Bjarki Ómarsson, thank you for standing by me and holding me up when I couldn´t. My mother, Dr. Sigríður Hjörleifsdóttir and father, Kristján G. Sveinsson, thank you for encouraging me and always having my back.

This work was supported by grants from the Marine Biotechnology ERA-NET, *ThermoFactories,* project grant number 5178–00003B; the Technology Development fund in Iceland, grant number 159004-0612; the Icelandic Research fund, *ThermoExplore*, project grant number 207088-051 and the Soroptimist International of Europe Scholarship.

Abbreviations

1,2-PDO	1,2-propanediol
3-HPA	3-hydroxypropionaldehyde
BOF	Biomass objective function
Cas	CRISPR associated protein
CDW	Cell dry weight
COBRA	Constraint-based reconstruction and analysis
CRISPR	Clustered, regularly interspaced, short palindromic repeats
CrtB	Phytoene synthase
CrtI	Phytoene desaturase
CrtO	γ-carotene ketolase
CrtZ	β-carotene hydroxylase
CrtY	Lycopene beta cyclase
CruC	Glycosyltransferase
CruD	Acyltransferase
CruF	1',2'-hydratase
DEH	4-deoxy-L-erythro-5-hexoseulose uronic acid
DMAPP	Dimethylallyl diphosphate
DXP	1-deoxy-D-xylose 5-phosphate
ED	Entner-Doudoroff
EMP	Embden-Meyerhof-Parnas
FBA	Flux balance analysis
FPP	Farnesyl-diphosphate
G-unit	α-L-guluronate unit
GAM	Growth associated maintenance
GEM	Genome-scale metabolic model
GGPP	Geranylgeranyl-diphosphate
GPP	Geranyl-diphosphate
GRAS	Generally recognized as safe
IPP	Isopentenyl-diphosphate
KDG	2-keto-3-deoxygluconate
KDPG	2-keto-3-deoxygluconate 6-phosphate
LAB	Lactic acid bacteria
LP	Linear programming
M-unit	β-D-mannuronate unit
MEP	2-C-methyl-D-erythritol 4-phosphate
MGS	Methylglyoxal synthase
MSW	Municipal solid waste

MVA	Mevalonate
NGAM	Non-growth associated maintenance
ORF	Open reading frame
PAM	Protospacer adjacent motifs
pFBA	Parsimonious FBA
PK	Phosphoketolase
WT	Wild type

1 HYPOTHESES AND OBJECTIVES

The overall aim of this project was to develop two non-model bacteria, *Rhodothermus marinus* and *Lactobacillus reuteri*, as cell factories. The main hypothesis (H1), with more specific hypotheses (H2-H4) and a hypothesis with a minor role in the thesis (H5), are listed below.

H1 The non-model bacteria *R. marinus* and *L. reuteri* can be used as cell factories.

H2 Genome-scale metabolic models can aid in strain design of these bacteria.

H3 *R. marinus* can degrade and utilize polysaccharides from complex biomass, e.g. brown algae, and produce marketable carotenoids, e.g. lycopene.

H4 *L. reuteri* can be used to produce 1,2-propanediol and 1-propanol.

H5 A thermostable CRISPR-Cas9 system can be used to genetically modify thermophiles, offering solutions to the limitations of existing methods.

The following objectives (O1-O4) were used to assess the hypotheses and guide the work.

O1 Reconstruct manually curated and validated genome-scale metabolic models of *R. marinus* and *L. reuteri,* and use the models to optimize production and utilization of target metabolites.

O2 Obtain *R. marinus* growth data for model reconstruction and validation, and for studying utilization and production of target metabolites.

O3 Characterize the carotenoid biosynthesis in *R. marinus* and alter the carotenoid production to produce lycopene instead of native carotenoids, through genetic engineering.

O4 Aid in developing a CRISPR-Cas9 system for the model thermophile *Thermus thermophilus*.

2 INTRODUCTION

Platform- and specialty chemicals are today largely produced by chemical synthesis using fossil fuel feedstocks. Fossil fuels are, however, being depleted and their use is also harmful to the environment. This makes underexploited natural biomass, agricultural residues and other waste, an attractive alternative. In biorefineries, biologically derived waste from agriculture, forestry, aquaculture, industry and the public would be used as feedstock for microbial production of fuels and chemicals for industry. In developing biomass as a feedstock for bioconversions, enzymes can be used to degrade biomass to fermentable sugars, which can subsequently be used as carbon sources for microbial production of various biochemicals. Microbes are a great source of various endogenous chemicals of potential use in industry. Additionally, genetic engineering can be used to optimize the production or to produce heterologous chemicals. Model organisms are often used as cell factories for industrial production of chemicals. These organisms are well studied, and established methods are available for their cultivation and genetic engineering. However, they are not always ideal production organisms, as they may have limited substrate utilization or ability to produce a compound of interest, and they often cannot endure the harsh environments that can be encountered in biorefineries. Non-model organisms, such as thermophilic bacteria, might make more suitable candidates.

Genome-scale metabolic models describe the biochemical reactions in the metabolic network of organisms and can be used to aid in strain design and help optimize cultivation and chemical production. They can be more difficult to reconstruct and use for non-model organisms, as their metabolism might be partially unknown. However, a significant part of the metabolic network is almost identical in different species and a lot of metabolic abilities can also be estimated from the genome. Additionally, the models can be used to help identify gaps in the knowledge of the specie´s metabolism.

Carotenoids are chemicals produced for industrial purposes. They are found in nature and are produced by many different organisms, including some bacteria and yeast. The two

major carotenoids on the market, astaxanthin and β-carotene, are mostly produced using chemical synthesis. However, other industrial carotenoids such as lycopene, lutein and zeaxanthin are obtained from plants and biotechnological processes. Microbial production of chemicals can be expensive compared to chemical synthesis, but due to environmental considerations and public awareness for the need of renewable resources, solutions and improvements in bioproduction are constantly being made.

2.1 Microbial cell factories

2.1.1 Biorefinery

Fossil fuels are commonly used for energy generation and as feedstock for oil-refineries in production of platform- and commodity chemicals. Although this is often economical, the disadvantages are becoming clear. Their extraction, logistics and subsequent utilization has harmful impacts on the environment and climate. With the growing population the demand increases while the availability is fast decreasing, resulting in fluctuating and at times inflated prices and less accessible sources being mined with higher environmental costs [1], [2]. Biorefineries are now emerging as an alternative, producing energy carriers and chemicals for industry, using sustainable biomass resources and environmentally benign biological processes for the production [3]. Carbohydrates are important substrates for bioproduction, as they are mainly used in living organisms as energy reserves and can be easily broken down into building blocks. Biomass is an attractive feedstock alternative, as it is a renewable resource and a lot of biomass waste is produced in agriculture, forestry, aquaculture, industry and by the public (municipal solid waste, MSW). Although agricultural crop plants are easily cultivated and economically optimal as raw material for biorefinery production of chemicals, competing with food sources is undesirable. 2nd and 3rd generation natural biomass, including waste material, are the preferred alternatives [4]. Today, they are largely underexploited but also challenging resources, and call for new biocatalytic refinery tools, enzymes and microbes.

A biorefinery refers to the overall process of using biomass as feedstock to produce food, feed, materials, chemicals and energy. Biomass can refer to crops, which compete with our food source (1st generation), biomass that is not used for food, but competes with food source for land (2nd generation), and biomass that does not compete with food sources at all (3rd generation). Biorefineries can be classified based on which feedstock is used, which intermediates or platforms are produced (sugars, syngas, biogas, etc.), which final products are produced (fuels, fine chemicals, etc.) and which processes are used (mechanical, biochemical, chemical and/or thermochemical). Pretreatment of 2nd and 3rd generation biomass, such as lignocellulose, MSW and algae, changes the structure of the biomass. This increases accessibility to the biomass components and therefore facilitates hydrolysis of

complex polymers into fermentable monomers, such as cellulose into glucose. Pretreatment can be carried out by physical methods (drying, milling, grinding, etc.), chemical methods, physiochemical methods (acids, alkali, etc.) and/or biological methods (enzymes, etc.). Pretreated biomass can then be used for fermentation by bacteria or yeast which in turn produce desired products [4].

2.1.2 Non-model microorganisms as cell factories

Microbes are suitable organisms for producing chemicals in biorefineries. They produce a variety of metabolites, many with industrial applications. Furthermore, microbes can often be easily genetically manipulated to produce new and even unnatural organic compounds based on pathways and metabolic intermediates available in the organism. Usually, mesophilic model organisms are used as chassis species for construction of cell factories, because they have been extensively studied and several efficient genetic engineering tools are available. *Escherichia coli, Lactococcus lactis, Streptomyces* sp., *Bacillus* sp., *Saccharomyces cerevisiae* and *Aspergillus* sp. are "classic" biotech chassis species for metabolic engineering and they have all been used for production of several biopharmaceutical compounds [5].

Many factors need to be considered when selecting a host strain for industrial bioproduction. Ideally, the selection should be based on the metabolic characteristics of the organisms, including the substrate utilization range required for the feedstock to be used and the synthetic capabilities or potential of the organism. The latter is based on the present catabolic or anabolic pathways and the available precursor pool in the organism. Anaerobic fermentative species are generally preferred for production of lower value commodity chemicals, such as simple organic acids and alcohols, which are typically waste products from catabolic metabolism. However, aerobes can carry a greater metabolic burden, usually have more extensive anabolic possibilities, and are therefore suitable for production of complex secondary metabolites or specialty chemicals.

Several species, including many model microorganisms, are not suitable production hosts in biorefineries using complex biomass. They are not able to utilize the available carbon sources in the feedstock and some are inhibited by present chemicals [6]. The processing conditions can also be harsh and fluctuating, with adverse effects on growth of the

organism. Desired traits in host organisms include the ability to use variety of carbon sources, both simple monosaccharides and complex polysaccharides, tolerance to high product concentrations and tolerance to extreme conditions such as pH, temperature, high substrate loadings and oxidative stress [7]. Other important factors are growth rate, cost of cultivation and downstream processes, purification of the product, secretion systems, genetic stability, codon usage, GRAS status, etc. The production strain can also be improved through metabolic engineering or laboratory adaptive evolution [8]. Product yield, productivity and titer often need to be optimized and the cost of cultivation and down-stream processes need to be taken into account for large-scale industrial production [9]. It can be concluded that the closer the metabolic potential of a species for production of a target product, and the more its physiological traits are compatible with processing conditions, the better choice it is as a chassis for metabolic engineering and subsequent industrial production. Non-model microorganisms may, therefore, be a better choice than available conventional model-microorganisms, although they are not as well-studied. So far, available genetic tools have been the major threshold for their use, but variety of tools are now being developed for many non-model species, and CRISPR based technologies appear to be especially promising.

Extremophiles are organisms that grow optimally under extreme conditions such as unusually high or low temperatures, pH, salinity and pressure. These organisms are interesting candidates for production in biorefineries and as production hosts for enzymes stable at extreme conditions (extremozymes) for pretreating biomass [10]. Thermophiles are an especially interesting option as biorefinery species for utilization of 2nd and 3rd generation biomass that are rich in recalcitrant polysaccharides. High temperature increases the solubility of polysaccharides and reduces the viscosity of fermentation broths. Consequently, it enables higher feedstock loads and facilitates enzymatic access to polysaccharides. In addition, growth at high temperatures in bioreactors mitigates scale-up problems of mixing and aeration, reduces costs of cooling, distillation and extraction, and minimizes the danger of contamination of mesophilic spoilage bacteria [11]. The disadvantages are that not all products are thermostable, and it can be challenging to find heterologous genes encoding thermostable enzymes for metabolic engineering of a synthetic pathway in a thermophile.

2.1.3 Microbial genetic tools

One of the most important aspects in choosing a suitable production organism for biorefineries or other industrial bioprocesses is the availability of tools to genetically modify the organism. For the conventional mesophilic model organisms these tools are established and efficient, including CRISPR-Cas9 systems for *E. coli* [12], [13] and *B. subtilis* [14]. These systems cannot always be easily applied to other bacteria, especially thermophiles, and lack of genetic tools places severe restrictions on the use of non-model microorganisms in industry. In the following, the main components in a genetic toolbox will be described. This includes a shuttle vector, selectable marker and an expression system.

A shuttle vector is a small DNA molecule that is used to deliver foreign DNA into the host for heterologous expression of genes and/or genome editing. This vector should contain a robust selection marker, it should have functional expression elements such as a promoter, it should be able to replicate in the host organism and in *E. coli* in a stable manner and it should have low homology to the host´s genome to prevent unwanted recombination. Robust and constitutive promoters that work in the host organism are essential for controlling gene expression. For model organisms, many well characterized promoters are known, but these promoters do not necessarily work in non-model organisms. Considerable work often goes into finding appropriate promoters. A variety of different selection markers can be used. Antibiotic resistance markers are common, but this can be problematic in thermophilic bacteria since some antibiotics and antibiotic resistance enzymes are inactive at high temperatures. Also, antibiotic systems are sometimes not favorable in certain industries, such as food. Other types of selective systems include dominant markers and complementation markers. Dominant markers introduce new abilities to the host such as heat-shock resistance or a sugar utilization pathway. Complementation systems are based on auxotrophy in a mutant strain, where a gene has been deleted from the genome and is subsequently used as a selective marker. This can for instance be a gene encoding an essential step in the biosynthesis of an amino acid or a nucleotide [15]. The number of selective markers is usually limited in thermophilic bacteria. Recycling marker systems are sometimes used, but their applications in genome editing can be time-consuming and labor intensive.

Prokaryotes have an adaptive immune system called clustered, regularly interspaced, short palindromic repeats (CRISPR). CRISPR consists of several spacer sequences derived from foreign elements, such as bacteriophages [16]. The bacterium uses these spacers to recognize the intruder and eliminate it by cleaving its DNA at the spacer site using an endonuclease called CRISPR associated protein (Cas). Cas is guided to the cleavage site by a CRISPR RNA (crRNA) which is generated from the spacer. The spacers are not randomly selected but depend on a sequence adjacent to the spacer called protospacer adjacent motifs (PAM), which varies between species. Several types of CRISPR-Cas systems exist, and they are classified based on their effector complex (crRNA and Cas). Scientists have successfully used type II systems, CRISPR-Cas9, to genetically modify their target organisms in a much easier way than before, which has resulted in a genetic engineering revolution in biology and medicine [12]. A few thermostable Cas9 enzymes have been characterized [17]–[21], but they cannot be used to efficiently edit the genomes of extreme thermophiles. An efficient thermostable CRISPR-Cas9 system would offer improved methods for genome engineering in many extreme thermophiles, which includes selective markers that are relatively easy to recycle.

2.2 Target organisms

2.2.1 *Lactobacillus reuteri*

Lactic acid bacteria (LAB) are a large and diverse group of bacteria that belong to the phylum *Firmicutes*. They are Gram-positive non-spore-forming anaerobes that can use a wide range of carbon sources and produce lactic acid as the main fermentation product. LAB have a long history of safe use and many species are generally recognized as safe (GRAS). They play an important role in our lives, they have been used in food and beverage fermentation for centuries, are used as pre- and probiotics and more recently have been used as production organisms for fuels and other chemicals [22]. *Lactococcus lactis* is a model organism, widely used in industry. However, there is considerable interest in using other LAB, including *Lactobacillus reuteri*, as cell factories [22].

L. reuteri is a heterofermentative LAB that is present in the human gut and is an important probiotic organism [23]. Although it is not a thermophile, it does possess several qualities compared to other LAB, which make it an interesting cell factory candidate. It has high growth and glycolytic rates, without the requirement for either aeration or strictly anaerobic conditions. It is tolerant to low pH, ethanol and salt, and has a wide growth temperature range. The species is known to produce 1,3-propanediol, reuterin, and other related industrially important compounds in high yields from glycerol. Whereas many LAB are auxotrophic for vitamin B12, it is produced natively in *L. reuteri*. Vitamin B12 is important as a cofactor in, for example the propanediol-utilization pathway, but is also of relevance for biotechnological and medical/health applications, *e.g.* when produced by probiotic strains. The vitamin B12 structure and biosynthetic genes in *L. reuteri* have been studied in detail [24], [25]. Tools for genetically engineering *L. reuteri* have been developed. ssDNA recombineering to obtain chromosomal mutations without selection has been described for *L. reuteri* and used to obtain decreased vancomycin resistance [26]. This method, together with a CRISPR-Cas9 system, was later used, which increased the mutation efficiency from 0.4 – 19% to 90 – 100% [27].

In this project, the compounds we were interested in producing in *L. reuteri* were 1,2-propanediol (1,2-PDO) and 1-propanol. They have many industrial applications, e.g. in the production of polyester resins (1,2-PDO) and as a solvent and potential biofuel (1-propanol)

[28]. Several different pathways have been described for 1-propanol production in microorganisms [29], including the 1,2-PDO pathway which produces 1-propanol from methylglyoxal, with 1,2-PDO as an intermediate (Figure 1). Other pathways include the citramalate, threonine, acetone, Wood-Werkman (methylmalonyl), acrylate and succinate pathways (Figure 1). For a recent overview of these pathways and engineered and non-engineered production organisms, the reader is referred to the review [29]. The thermodynamic maximum yield of 1-propanol from glucose in *L. reuteri* calculated based on the degree of reduction is 1.33 mol/mol (44,4% carbon yield). However, only the stoichiometry of the 1,2-PDO, succinate, acrylate and Wood-Werkman pathways allow this maximum yield – the others result in up to 25% less yield [29]. One of the advantages of using *L. reuteri* for 1-propanol production through the 1,2-PDO pathway is that it produces vitamin B12, which is needed as a co-factor for the B12-dependent diol dehydratase step and has been suggested to be a limiting factor in this pathway in for example *C. glutamicum* [30] and was added to the fermentations in an *E. coli* strain harboring this pathway [31].

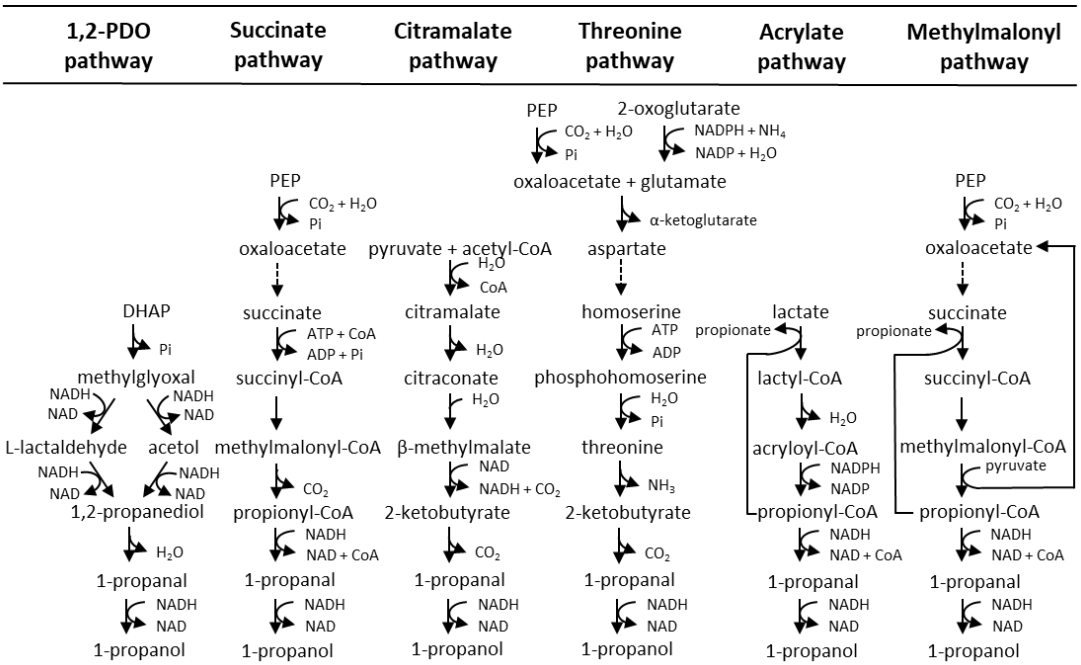


Figure 1. *1-propanol biosynthetic pathways. Names of substrates, products and co-factors are shown, some abbreviated: Dihydroxyacetone phosphate (DHAP) and phosphoenolpyruvic acid (PEP). Solid arrows indicate single reactions and dotted arrows indicate multiple reactions.*

2.2.2 *Rhodothermus marinus*

Rhodothermus is the type genus of the phylum *Rhodothermaeota* [32]. The aerobic bacterium *R. marinus* is thermophilic and moderately halophilic, growing optimally at 65°C and 1-2% NaCl [33]. It has been found in geothermal environments around the world but was first isolated in Ísafjarðardjúp, NW of Iceland (Figure 2) in 1988 [33]. The first characterized *Rhodothermus* bacteriophage, RM378, was isolated from the same area [34]. Several strains of *R. marinus* have been isolated [33], [35]–[37], but this work focuses on the type strain DSM 4252^T and strain ISCaR-493. DSM 4252^T is the most studied *R. marinus* strain until now [38] and the genome sequence has been published [39]. There is increasing interest in strain ISCaR-493 due to the development of genetic tools for the strain [40]. ISCaR-493 has other strain designations which are sometimes found in literature, including DSM 16675, PRI 493 and MAT 493.

Figure 2. *Reykjanes in Ísafjarðardjúp, NW of Iceland, were R. marinus was first isolated from. Picture taken in July 2020 by Sigríður Hjörleifsdóttir.*

R. marinus produces thermostable enzymes, many of which have been characterized. This includes a number of polysaccharide degrading enzymes, such as cellulase [41], laminarinase [42] and xylanase [43], [44]. For a more comprehensive list, the reader is referred to a review on *R. marinus* [38]. Most *R. marinus* strains are red-pigmented, due to carotenoid production. The carotenoids in *R. marinus* strain DSM 4253 were found to be two types of

carotenoid glucosides and their acyl derivatives [45]. Other interesting metabolites natively produced by *R. marinus* include exopolysaccharides [46], polyamines [47] and compatible solutes [48].

Genetic tools have been developed for *R. marinus* [49]. Strain ISCaR-493 was identified to be well suited for genetic manipulation [40] and mutant strain SB-62 (Δ*trpB*Δ*purA*) was subsequently obtained [50]. Tryptophan and adenine selection can be used to genetically modify strain SB-62. The shuttle vector pRM3000, which contains the *trpB* selective marker and sites for replication in both *R. marinus* and *E. coli* [51] was created for *R. marinus* cloning. This vector was used to introduce heterologous and homologous genes to *R. marinus* which were successfully expressed [51]. Chromosomal deletions have also been obtained in *R. marinus*, using double crossover homologous recombination using linear insertion cassettes [50].

Brown algae is an example of a 3rd generation biomass that could be used in biorefineries. *R. marinus* is a marine bacterium, so seaweed is common in its natural environment. Alginate and laminarin are major constituents of polysaccharides in brown algae and *R. marinus* possesses several genes encoding the corresponding degradation enzymes. In this project, we investigated the degradation and growth of *R. marinus* on alginate. Alginate is a structural component of brown algae and can comprise up to 40% of its dry matter [52]. It is a polyuronate that consists of β-D-mannuronate (M) and α-L-guluronate (G) units forming (1 → 4) linked G-, M- and mixed blocks in the polysaccharide chain. *R. marinus* DSM 4252^T has four genes encoding alginate lyases [53], [54] that, together, depolymerize alginate into the same unsaturated mono-uronate derivative of the M and G units. *R. marinus* also possesses the genes encoding the remaining enzymes of the alginate catabolic pathway for its utilization. The unsaturated monouronate is converted to 4-deoxy-L-erythro-5-hexoseulose uronic acid (DEH) by a spontaneous reaction and further catalyzed to 2-keto-3-deoxygluconate (KDG) by an aldose reductase. KDG enters the ED pathway, where it is catalyzed to 2-keto-3-deoxygluconate 6-phosphate (KDPG) by 2-keto-3-deoxygluconokinase (EC 2.7.1.45) and then further to pyruvate and glyceraldehyde 3-phosphate by 2-dehydro-3-deoxyphosphogluconate aldolase (EC 4.1.2.14), which enter central metabolism [55], [56] (Figure 3). *R. marinus* does not possess a key enzyme in the ED pathway, phosphogluconate dehydrogenase (EC 4.2.1.12), which is needed to metabolize glucose. This is why the ED

pathway is not active when *R. marinus* is grown on glucose [57], but the partial pathway is essential for utilization of alginate (Figure 3).

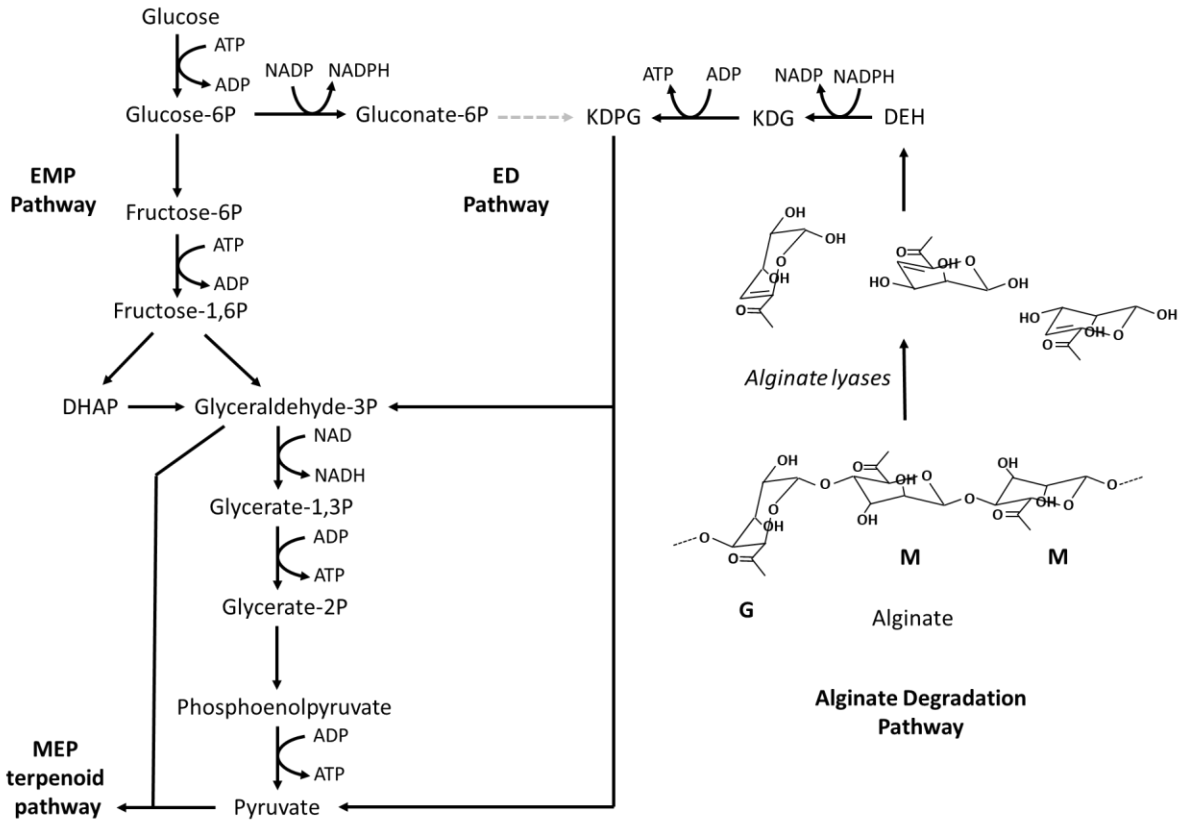


Figure 3 *Alginate degradation pathway and how it connects to the Embden-Meyerhof-Parnas (EMP) pathway, the partial Entner-Doudoroff (ED) pathway and the MEP terpenoid pathway in R. marinus. Names of substrates, products and energy molecules (ATP, NADH, NADPH) are shown, some abbreviated: 4-deoxy-L-erythro 5-hexoseulose uronic acid (DEH), 2-keto 3-deoxygluconate (KDG) and 2-keto 3-deoxygluconate 6-phosphate (KDPG). Molecular structure of a partial alginate molecule is shown: β-D-mannuronate (M) and α-L-guluronate (G). The missing reaction of the ED pathway in R. marinus is represented by a grey dotted arrow.*

2.2.3 *Thermus thermophilus*

Many thermophilic bacterial species which have been isolated from all around the world belong to the genus *Thermus*. This includes *T. thermophilus*, which is a gram negative aerobe that has a growth temperature range from 50°C to 82°C [58]. *Thermus* spp., including *T. thermophilus*, have high growth rates and can reach high cell densities. They can be genetically manipulated relatively easily, and they produce industrially relevant compounds such as carotenoids and polyamines. They also possess thermostable enzymes, and some are used for industrial applications such as DNA polymerases. *T. thermophilus* has been the

subject of many studies, focusing on genetic manipulation, structural genomics and systems biology and has become a model organism for studying thermophiles [59]. The work described in this thesis mostly focuses on non-model organisms and *T. thermophilus* only plays a minor role.

One of the properties that make *T. thermophilus* a good model organims is its natural competence [60], which makes genetic manipulation of the bacterium much simpler than for many others. Several tools have been developed, including shuttle vectors [61]–[63] and integration systems [64], [65]. A thermostable kanamycin resistance gene is commonly used as a selective marker in *T. thermophilus* [66], [67]. Counterselection systems have also been developed [64], [68], so that the thermostable kanamycin marker can be used again for further genetic modifications. Recycling of selectable markers in this way is however a time-consuming process involving a lot of screening.

2.3 Genome-scale metabolic models

The metabolic network of an organism can be reconstructed on the genome-scale by compiling all known metabolic reactions and metabolic genes of the organism. This is based on the annotated genome, general biochemical data on metabolism and organism-specific information obtained from literature. The metabolic reactions and associated genes are structured in a standardized way, which enables conversion to a computational model, often referred to as a genome-scale metabolic model (GEM). In addition to describing the metabolism, enzyme kinetics [69], protein-protein interactions [70], regulation [71] and signaling [72] networks can also be reconstructed, although it is very difficult to do on a genome scale. These models are based on large amount of data, such as data on fluxomics, protein structures, transcriptomics, regulatory genes and signaling networks. Such data has been generated for a few organisms/cell types, but for the non-model organisms studied in this project, this is not the case. Therefore, we focused solely on metabolic reconstructions, ignoring many of these relevant factors which is obviously a strong assumption and should be noted when interpreting model simulations. The first GEM was reconstructed of *Haemophilus influenzae* in 1999 [73] and with the rapid increase of publicly available genomes, the number of GEMs keeps growing. In early 2019, GEMs for 6239 organisms had been reconstructed, including 5897 bacteria [74]. Most of these models were automatically reconstructed and only 183 had been subjected to manual curation. Protocols to guide reconstruction of high quality GEMs have been published [75], [76] as well as tools to test and assess GEM quality, such as MEMOTE [77]. Here the main steps of the reconstruction process, the simulation methods and applications of the models will be described briefly.

2.3.1 Reconstruction

A draft reconstruction can be automatically obtained from the annotated genome sequence. Several tools for automatic reconstructions are available, which have different features [78]. Choosing a tool depends on the researcher, the target organism and the intended purpose of the model. Some tools are suitable if well curated model(s) of closely related organisms are available, which can serve as reference models in the reconstruction process. This includes MetaDraft [79] and RAVEN [80]. Some tools are implemented in Python, e.g. CarveMe [81], others in MATLAB, e.g. AutoKEGGRec [82], and yet others run in a web

browser, e.g. ModelSEED [83]. Whatever tool is chosen, the automatically obtained reconstruction should be carefully refined to obtain a high-quality model [76].

The manual refinement is a time consuming but critical step in the reconstruction process. It includes reviewing the entire draft reconstruction, evaluating which reactions should be present, if any are missing and which genes should be associated with each reaction. The process further entails reviewing substrate and cofactor usage of reactions, revising formulas and charges of metabolites in unbalanced reactions, checking reaction directionality and compartmentalization, inclusion of transport and exchange reactions, etc. The refinement is based on organism-specific data when it is available, the annotated genome sequence and general biochemical data obtained in databases such as BiGG [84], KEGG [85], MetaCyc [86] and Brenda [87]. Another important part of this step is constructing a biomass objective function (BOF), which describes the macromolecular composition (protein, DNA, RNA, lipids, etc.) of the biomass, the ratio of their corresponding building blocks (amino acids, nucleotides, fatty acids, etc.) and the energy required. Here, organism-specific data is highly important, as the BOF is used to simulate growth.

2.3.2 Flux balance analysis

The reconstruction can be converted to a computational model where the network can be tested, validated and ultimately used for simulations. In the reconstruction process, this is an important step as it helps find gaps and errors in the reconstruction, which in turn can be fixed and the model tested again. This iterative process continues until a functioning model is obtained. COnstraint-Based Reconstruction and Analysis (COBRA) methods are used to work with metabolic models and can be applied in different platforms, such as Python [88] and Matlab [89], [90].

Flux balance analysis (FBA) is a COBRA method which is commonly used to work with and analyze genome-scale metabolic models [91], [92]. It uses linear programming (LP) to identify a flow of metabolites through the network which optimize a specific biological objective, such as maximizing growth rate. Physiological limitations and assumptions introduce constraints in FBA, such as the steady-state conservation of metabolites in the system and flux capacity limitations of reactions. The reaction stoichiometry is specified in an m by n matrix S where n is the number of reactions in the model, m is the number of

metabolites and S_{ij} is the stoichiometric coefficient for metabolite *i* in reaction *j*. The LP problem can be formulated as follows:

$$\begin{aligned} &\text{maximize} && c^T v \\ &\text{subject to} && Sv = 0 \\ &\text{and} && v_{min} \leq v \leq v_{max} \end{aligned} \tag{1}$$

where **c** is an *n*-vector specifying the objective function, typically the biomass reaction, **v** is an *n*-vector of reaction fluxes, with positive values denoting flux in the forward direction and negative values denoting flux in the reverse direction. Vectors $\mathbf{v}_{min}$ and $\mathbf{v}_{max}$ specify the minimum and maximum flux in each reaction. The decision variables of the LP are the elements of **v** while vectors **c** and $\mathbf{v}_{min}$ and $\mathbf{v}_{max}$ are given constants. The elements in $\mathbf{v}_{min}$ and $\mathbf{v}_{max}$ corresponding to internal reactions are set to an arbitrary large number since experimental measurements of internal fluxes are generally unavailable, leaving the internal fluxes essentially unbounded. Experimental values of substrate uptake and secretion rates are used to define the elements corresponding to exchange reactions in the model. In addition, element *i* of $\mathbf{v}_{min}$ is zero if reaction *i* is irreversible. Thermodynamics are only included in GEMs to a limited extent, with reaction directionality. Because of this, the simulation solutions might not be thermodynamically feasible and frequently contain infeasible flux cycles.

A variant of FBA called parsimonious FBA (pFBA) [93] is sometimes used for simulations instead of the basic FBA. pFBA identifies flux values corresponding to maximum growth with the side constraint that the sum of absolute flux values is made as small as possible. The sum of fluxes is proxy for enzyme usage and the method can therefore be considered to simulate biological pressure for rapid and efficient growth using minimum amount of resources (enzymes). An advantage over FBA is that the resulting solution is likely to contain fewer infeasible flux cycles.

2.3.3 Applications

GEMs have been used to predict biological phenotypes for various applications, including metabolic engineering, evolutionary processes and analyzing network properties and phenotypic behaviors [94]. Predicting quantitative phenotypes has proven challenging, but

feasible. This is done by optimizing a hypothesized objective, while constraining the network with biological limitations, using methods such as FBA. The resulting flux predictions can in turn be compared to experimental data, which either agree and therefore validate the hypothesized objective and constraints or disagree, which often can help guide new discoveries and understanding of the metabolism. Incorrect model predictions can for instance suggest lack of knowledge and therefore point researchers to new research opportunities, suggest that biological constraints that are not covered by the model, such as gene regulation, are involved, or suggest that the objective is not as hypothesized or that a mix of competing objectives might apply [95].

GEMs can be used in the design of cell factories. They can suggest environmental and genetic modifications that improve the desired function or production. Several algorithms have been developed that find strategies to genetically engineer or modify environmental factors to achieve a biological objective. OptStrain [96] identifies reactions to delete or heterologously express for overproduction of a target compound. OptStoic [97] aims to do the same thing, while taking the overall stoichiometry and co-metabolites into account. CosMos [98] finds deletions, downregulations and upregulations of fluxes to achieve production of a target metabolite. k-OptForce [99] integrates available data on kinetics of metabolic reactions with stoichiometric models for improving product production. These methods represent only a fraction of the methods available, but they give a general idea on how these tools can be used to aid strain design.

2.4 Carotenoids

2.4.1 Properties and functions

Carotenoids are pigments, ranging from red to yellow, naturally produced by many plants, algae, archaea, and bacteria. Currently, over 1100 carotenoid structures from more than 600 organisms are known [100]. Carotenoids produced by plants, algae and photosynthetic bacteria aid in photosynthesis by harvesting light energy and transferring it to the chlorophylls. Additionally, carotenoids have been shown to protect cells against harmful effects of light and oxygen, first suggested by Sistrom, Griffiths and Stanier in 1956 [101], which explains why they are also synthesized by some non-photosynthetic bacteria, such as *R. marinus*.

Carotenoids are hydrophobic compounds and are incorporated into the cell membrane. Their structure is very rigid due to a polyene chain containing multiple conjugated double bonds. The double bonds decrease the fluidity of the membrane and stabilize it. This can be important in extreme environments such as high temperatures. Carotenoids in complexes with proteins can also provide protection from extreme environment or proteases. Besides transferring energy to aid photosynthetic reactions, carotenoids also protect the photosynthetic reaction center in phototrophs from photodamage. UV radiation and oxidative stress also effects other cell components and the anti-oxidative function of carotenoids helps protect against this damage. Carotenoids play important roles in animals, even though they cannot synthesize these compounds themselves and must consume them through diet. Retinal, which is produced from provitamin A carotenoids, together with opsin form a complex called rhodopsin. Rhodopsin is found in the visual systems of mammals, it absorbs light at different wavelengths and is therefore essential for color vision. Animals and plants also use carotenoids for various communication platforms. They can for instance be used to indicate change in sex or social status as can often be observed in birds and fish, they can be used to disguise themselves as has been observed in chameleons, or they can be used to create attractive coloration to attract insects as commonly seen in flowers and fruits [102].

Carotenoids are widely used in the dietary supplement, food, feed, cosmetic and pharmaceutical industries. They are used for their coloring properties and for their

antioxidant and provitamin A activities. Studies have shown that carotenoids have potential health benefits, both in preventing and treating various diseases [103], [104]. Although positive effects of carotenoids on human health have been observed, it is important to keep in mind that the biological functions of carotenoids are complex and there are studies that have reported conflicting results [105]. This highlights the importance of continued research on the effects that carotenoids can have on human health and the importance of emphasizing both sides when discussing possible applications of carotenoids. Regardless, the global market for carotenoids has reached USD 1.5 billion and is predicted to continue growing [106]. The most abundant carotenoids on the market are astaxanthin, β-carotene, canthaxanthin, zeaxanthin, lycopene and lutein (Figure 4).

2.4.2 Industrial production

Carotenoids are produced commercially from plants and microalgae and while such natural production of carotenoids is generally preferred by consumers, chemically synthesized carotenoids are often more economical [107], [108]. Microbes are a potential source for both economical and sustainable production of carotenoids. Many species have been observed to produce carotenoids natively and some species can grow on cheap substrates. Unlike plants, cultivating bacteria for carotenoid production does not compete with human and animal food and feed sources and does not occupy large areas of land, especially if carbon sources are derived from 2nd and 3rd generation biomass. However, several issues are often observed during microbial production of carotenoids. This includes poor process control, high uptake of nutrients and low yields of carotenoids in both open pond cultivation and bioreactors. Problems with contaminations and with maintaining optimal growth and nutrient conditions can also arise. Microbes also often produce carotenoids which are not always preferred by the industry. Genetic engineering can be used to alter the structures, but this is not yet acceptable by many industry sectors and society in general. Many problems can arise in microbial production of carotenoids and a number of studies have been carried out to solve problems and increases yields, including optimizing growth conditions and control of temperature, pH, salinity, nutrient composition and levels, and light.

β-carotene is the major carotenoid on the market and is commonly used in the food and feed industries. The majority of β-carotene produced for industry today is chemically

synthesized, but it is also produced using the fungus *Blakeslea trispora*. Astaxanthin, the second major industrial carotenoid, is mainly used in aquaculture especially in the breeding of salmonoids, due to its strong orange-red color. Astaxanthin is also mostly produced through chemical synthesis. However, several microorganisms, such as the yeast *Xanthophyllomyces dendrorhous* and the microalgae *Haematococcus pluvialis*, are promising production hosts for astaxanthin. As an example, the production cost using large-scale cultivation of *H. pluvialis* was estimated to be even lower than the cost of chemical synthesis [109] and the achieved yields are between 1.5 - 3.0% of dry weight. Continued studies and improved cultivation conditions and strain improvement using metabolic engineering will increase the economic viability of natural production of astaxanthin and β-carotene. Lycopene is a common intermediate in anabolic synthesis of more complex carotenoids and various industry sectors are interested in using lycopene, both due to its beneficial link to health and as a coloring agent. Lycopene yields from chemical synthesis are generally low and it is therefore mostly extracted from plant products, such as tomatoes, and produced by biotechnological methods. Microbial production is an option and the fungus *B. trispora*, with defected lycopene cyclase activity, can produce lycopene in yields suitable for industrial production. Lutein and zeaxanthin are smaller players on the market than β-carotene, astaxanthin and lycopene. However, they play an important role in the human macula and neural system and are marketed as food and feed supplements. The cost of chemical synthesis of these carotenoids is high and they are mostly extracted from marigold flowers. Microbial production of lutein and zeaxanthin has been studied, for instance in the microalgae *Murielopsis* sp., *Chlorella* sp., *Scenedesmus* sp. and the bacteria of *Flavobacterium* spp. [110]

2.4.3 Structures and biosynthesis

The precursors for terpenoids, including carotenoids, can be synthesized through two independent pathways: The mevalonate (MVA) pathway and the non-mevalonate pathway, also known as the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway. The MVA pathway was first described in the 1960s [111], [112] and for a long time it was thought to be the sole terpenoid biosynthetic pathway, even though it could not be identified in several carotenoid producing bacteria. In the 1990s, the MEP pathway was discovered in bacteria [113]. In the MVA pathway, mevalonate is produced from three units of acetyl-CoA and then further

metabolized to isopentenyl-diphosphate (IPP) which can be converted to dimethylallyl diphosphate (DMAPP). This pathway uses in total seven enzymes, two NADPH and three ATPs. The MEP pathway produces 1-deoxy-D-xylose 5-phosphate (DXP) from pyruvate and glyceraldehyde 3-phosphate, which is further metabolized to 2-C-methyl-D-erythritol 4-phosphate (MEP). In total of seven enzymatic steps, using three ATP equivalents and three NADPH, the same products as obtained from MVA, DMAPP and IPP, are produced by the MEP pathway [114]. In *R. marinus*, the MEP pathway is used for terpenoid biosynthesis. Both terpenoid pathways (MEP and MVA), with the key metabolites and enzymes are shown in Figure 4. In both pathways, a polyprenyl pyrophosphate chain is built from the precursors DMAPP and IPP. First, geranyl-diphosphate (GPP) is synthesized, followed by farnesyl-diphosphate (FPP) and finally geranylgeranyl-diphosphate (GGPP) (Figure 4).

GGPP is the precursor for carotenoid biosynthesis. Two units of GGPP are converted to phytoene by phytoene synthase (CrtB). Phytoene is desaturated in several steps to the red colored lycopene by phytoene desaturase (CrtI). Several other enzymes can modify the carotene backbone. This includes lycopene beta cyclase (CrtY), which cyclizes either one or both ends of lycopene resulting in γ- or β-carotene, respectively; beta-carotene ketolase (CrtO), which introduces a keto group to the β-ring; beta-carotene hydroxylase (CrtZ) and 1',2'-hydratase (CruF), which both introduce hydroxyl groups to the backbone; and glycosyl- and acyltransferases (CruC and CruD), which add sugar and fatty acid groups to one or both ends of the backbone, as in the case of *R. marinus* [45] and *T. thermophilus* [115], respectively. The lower part of figure 4 shows the biosynthesis of carotenoids, including lycopene and other industrially relevant carotenoids, along with the carotenoids from *R. marinus*.

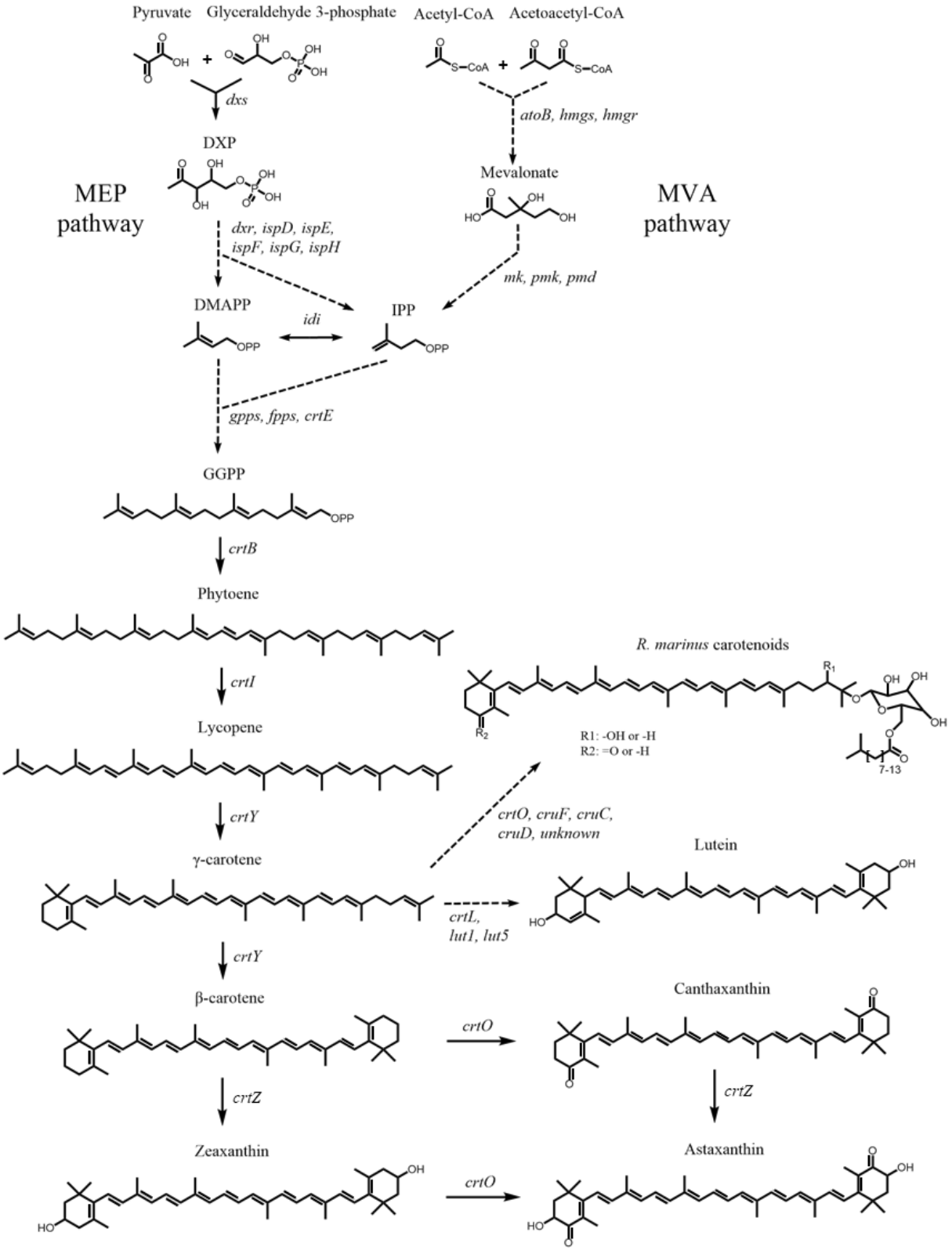


Figure 4 *Terpenoid and carotenoid biosynthetic pathways. The two terpenoid pathways, MEP and MVA, are shown in the upper half and biosynthesis of carotenoids natively produced by R. marinus and industrially relevant carotenoids are shown in the lower half. Selected metabolites and their names are shown, some abbreviated: 1-deoxy-D-xylose 5-phosphate (DXP), dimethylallyl diphosphate (DMAPP), isopentenyl-diphosphate (IPP), geranylgeranyl-diphosphate (GGPP). Abbreviations of genes known to encode pathway enzymes are also shown: 1-deoxy-D-xylose 5-phosphate synthase (dxs), 1-*

deoxy-D-xylulose 5-phosphate reductoisomerase (dxr), 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (ispD), 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (ispE), 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (ispF), (E)-4-hydroxy-3-methylbut-2-enyl-diphosphate synthase (ispG), 4-hydroxy-3-methylbut-2-en-1-yl diphosphate reductase (ispH), acetyl-CoA acetyltransferase (atoB), hydroxymethylglutaryl-CoA synthase (hmgs), 3-hydroxy-3-methyl-glutaryl-CoA reductase (hmgr), mevalonate kinase (mk), phosphomevalonate kinase (pmk), diphosphomevalonate decarboxylase (pmd), isopentenyl-diphosphate isomerase (idi), geranyl diphosphate synthase (GPPS), farnesyl diphosphate synthase (FPPS), geranylgeranyl pyrophosphate synthase (crtE), phytoene synthase (crtB), phytoene desaturase (crtI), lycopene beta cyclase (crtY), beta-carotene ketolase (crtO), 1',2'-hydratase (cruF), glycosyltransferase (cruC), acyltransferases (cruD), lycopene epsilon cyclase (crtL), epsilon-ring carotenoid hydroxylase (lut1), beta-ring carotenoid hydroxylase (lut5), beta-carotene hydroxylase (crtZ). Solid arrows indicate single reactions and dotted arrows indicate multiple reactions.

3 SUMMARY OF THE WORK

The aim of the study was to investigate biotechnological potential of non-model bacterial species using approaches of systems- and synthetic biology. *Lactobacillus reuteri* and *Rhodothermus marinus* possess several abilities which make them interesting candidates as cell factories and in biorefineries. As a LAB, *L. reuteri* can be used in different industries including health and food, and it is a native producer of compounds such as vitamin B-12 and 1,3-propanediol. As a thermophile, *R. marinus* grows optimally in harsh environments which are often encountered in biorefineries in the processing of complex biomass. It produces carotenoid compounds in high yields and grows on a variety of carbon sources, including several polysaccharides.

Genome-scale metabolic models were reconstructed for both species, based on available draft models. Metabolic engineering of *R. marinus* was also carried out, to construct a thermophilic platform for production of carotenoids of industrial interest. The carotenoid biosynthetic genes were evaluated, and a biosynthetic pathway proposed. Subsequently, *R. marinus* was engineered for production of lycopene, which is a common intermediate in biosynthesis of α-, β- and γ-carotenoids. Furthermore, growth and carotenoid production in *R. marinus* under different conditions were investigated, including using alginate as a carbon source. Alginate is one of the major polysaccharides in brown algae and the ability of *R. marinus* to degrade it contributes to the suitability of *R. marinus* as a biorefinery candidate.

As previously noted, the metabolic engineering of thermophiles suffers for the lack of efficient recyclable selective markers so only limited number of genome edits can be made per strain in a reasonable time. A CRISPR-Cas system would be a very useful addition to the genetic toolbox of *R. marinus.* Here, the function of a newly discovered thermostable CRISPR-Cas9 system by our group was tested for the thermophile *Thermus thermophilus*. The development of a similar system is underway for *R. marinus*.

Looking back on the objectives listed in Chapter 1:

O1 Reconstruct manually curated and validated genome-scale metabolic models of *R. marinus* and *L. reuteri* and use the models to optimize production and utilization of target metabolites.

O2 Obtain *R. marinus* growth data for model reconstruction and validation, and for studying utilization and production of target metabolites.

O3 Characterize the carotenoid biosynthesis in *R. marinus* and alter the carotenoid production to produce lycopene instead of native carotenoids, through genetic engineering.

O4 Aid in developing a CRISPR-Cas9 system for the model thermophile *Thermus thermophilus*.

Overall, the main goals of this project were reached. Genome-scale metabolic models were reconstructed, validated and are publicly available (O1, O2). The model of *L. reuteri* was used to simulate 1,2-propanediol and 1-propanol production in an ethanol deficient strain (O1). Lycopene overproduction simulations in the *R. marinus* model did not result in any directly realizable strategies (data not shown) (O1). However, during the reconstruction process the metabolism in *R. marinus* was carefully reviewed which led to several discoveries that can be exploited to increase production of carotenoids, such as good growth and high carotenoid production when grown on alginate, and light induced production of carotenoids (O2). The carotenoid biosynthetic pathway and corresponding enzyme coding genes in *R. marinus* were studied. This resulted in the discovery of a carotenoid gene which plays a key role in the downstream pathway from lycopene, and by knocking it out a lycopene producing strain was obtained (O3). This is the first time an industrially relevant carotenoid is produced in *R. marinus*, which is a vital step towards obtaining a robust *R. marinus* strain that can be used in industry. Finally, a thermostable CRISPR-Cas9 system was successfully used to genetically engineer the thermophile *T. thermophilus* (O4), giving hope to more advanced and easier genetic engineering of *R. marinus* in the future.

3.1 My contribution to the papers

Paper 1 Kristjansdottir et al. "A metabolic reconstruction of *Lactobacillus reuteri* JCM 1112 and analysis of its potential as a cell factory". *Microb Cell Fact* (2019).

E. Bosma and I contributed equally to this work. I curated and validated the metabolic reconstruction, performed numerical simulations and wrote the manuscript together with E. Bosma.

Paper 2 Kristjansdottir et al. "Engineering the carotenoid biosynthetic pathway in *Rhodothermus marinus* for lycopene production". *Metab Eng Commun* (2020).

E. Ron and I contributed equally to this work. I planned and performed all genetic modifications and analysis, absorbance spectra analysis and wrote the manuscript together with E. Ron.

Paper 3 Kristjansdottir et al. "A genome-scale metabolic reconstruction provides insight into the metabolism of the thermophilic bacterium *Rhodothermus marinus*". Manuscript available on bioRxiv.

I reconstructed, curated and validated the metabolic reconstruction, performed numerical simulations, planned and performed the cultivations and the carotenoid analysis and wrote the manuscript. A part of the metabolic reconstruction was performed during my MSc. thesis.

Paper 4 Ron et al. “Characterization of carotenoids in *Rhodothermus marinus*”. *MicrobiologyOpen* (2018).

E. Ron planned and performed most of this work. I analyzed the genome in respect to carotenoid biosynthetic genes and proposed a carotenoid biosynthetic pathway.

Paper 5 Allahgholi et al. “Composition analysis and minimal treatments to solubilize polysaccharides from the brown seaweed *Laminaria digitata* for microbial growth of thermophiles”. *Journal of Applied Phycology* (2020).

L. Allahgholi planned and performed most of this work. I performed cultivations of *R. marinus* in alginate minimal medium.

Paper 6 Adalsteinsson et al. “Efficient genome editing of an extreme thermophile, *Thermus thermophilus*, using a thermostable Cas9 variant”. *Sci Rep* (2021).

B. Adalsteinsson planned and performed most of this work. I contributed to the *in vivo* experiments: designed and constructed cloning vectors, performed genetic modifications, verified clones and analyzed data.

3.2 A metabolic reconstruction of *Lactobacillus reuteri* JCM 1112 and analysis of its potential as a cell factory (Paper 1)

This work describes the reconstruction, validation and implementation of a genome-scale metabolic model of *Lactobacillus reuteri*. The metabolism was thoroughly reviewed during the reconstruction process, several knowledge gaps were identified and resolved and experimentally obtained growth data was used to validate the model. Lastly, the model was used to evaluate the potential of *L. reuteri* as a cell factory, specifically for the production of the non-native compound 1-propanol.

3.2.1 Reconstruction of the genome-scale metabolic model

The metabolic model of *L. reuteri* JCM 1112 was reconstructed based on a draft model described in [116]. The model was extensively curated and validated using data obtained from laboratory experiments and available literature, resulting in the model Lreuteri_530. This included resequencing and comparing the two type strains obtained from two strain collections (DSM and JCM), analyzing the usage of the two glycolytic pathways present, the phosphofructokinase (PFK) and the Embden–Meyerhof–Parnas (EMP) pathways, and a thorough review of individual pathways and reactions in accordance with gene annotations and the literature. We emphasized on making the model publicly available and easy to use. For this purpose we 1) published the model in different formats (SBML, MATLAB, .json, EXCEL); 2) abbreviated and annotated most metabolites and reactions according to the BiGG database; 3) identified the genes according to the locus tags obtained in the JCM 1112 GenBank file; 4) provided a table to easily compare gene locus tags between the type strain from the two culture collections and to find the corresponding NCBI protein IDs, annotations and sequences; 5) constrained the model with experimentally obtained data so when optimized, by default it maximizes the biomass reaction and predicts realistically; 6) created metabolic maps of the central metabolism in *L. reuteri*, for use in the Escher [117] visualization tool, both a simplified map and a detailed map linking different sugar catabolic pathways to it; 7) provided all code used in the simulations as a Jupyter notebook and 8) created a GitHub page (https://github.com/steinng/reuteri) for easy access of the model, relevant data and updates.

The biomass objective function for *L. reuteri* was assembled from experimental data obtained in our study and information found in literature, *L. reuteri* specific data when available[118] [119] [120] but otherwise from *L. plantarum* [121] and *L. lactis* [122]. The sensitivity of the predicted growth rate to changes in biomass and energy components was investigated by varying the coefficient of each component, one at a time, by 50% while varying the glucose uptake rate. The analysis showed that predicted growth rate was sensitive to changes in the protein and GAM (energy required for growth) components of the biomass. Protein and amino acid contents were measured in our study for strain JCM 1112 and we used one of our experimental datasets to estimate the GAM value.

3.2.2 Model validation

Several different datasets were used to validate the model, using three strains: JCM 1112 (WT), SJ11774 (SJ (WT*)) which is a modified JCM 1112 stain enabling genetic engineering and SJΔ*adhE* which has the *adhE* gene (bifunctional aldehyde/alcohol dehydrogenase) deleted, resulting in impaired ethanol production. All the strains were grown in defined medium with only glucose as the carbon source on one hand and both glucose and glycerol on the other.

We used growth data from these experiments in two different *in silico* experiments to validate the model. In the first one, the model was constrained with the measured uptake- and secretion rates of glucose, glycerol, amino acids and organic by-products, optimized for maximum growth rates and compared to the experimentally obtained growth rates (Figure 5). This analysis showed that the predicted and experimental growth rates were in good agreement. One exception was the mutant strain grown on glucose (dataset E in Figure 5). The model predicted a slightly lower growth rate than observed *in vivo*, though both show a large decrease in growth compared to the WT. The experimental data showed that no glucose was taken up during the cultivation, while small amounts of lactate and acetate were produced. The amount of glucose needed to produce the measured acids is low and was in fact within the error of the assay. If the model was allowed to take up enough glucose to produce the measured acids, the predicted growth rate increased from 0.22 to 0.34 h^{-1}, compared to 0.30 h^{-1} *in vivo.*

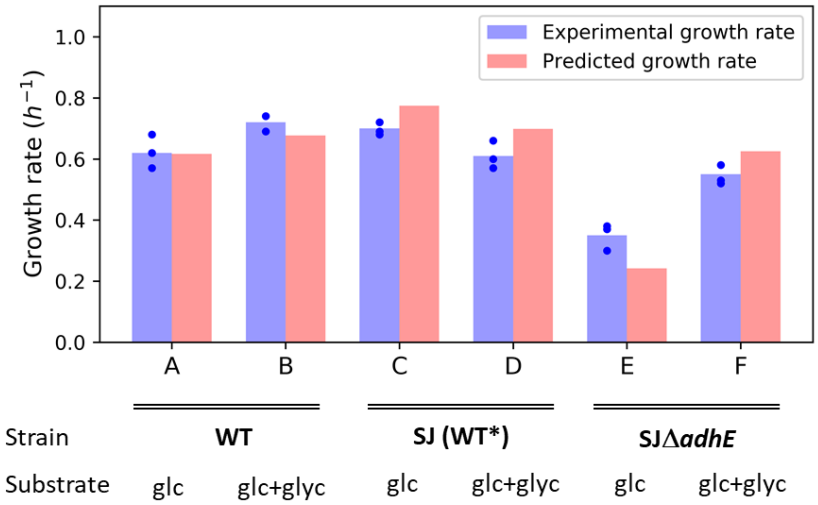


Figure 5 *Predicted and experimental growth rates. Experimentally measured growth rates for each of the six data sets are shown in blue, with blue dots denoting individual replicates and blue bars representing average values. For each dataset, the model was constrained with average experimental values for uptake and secretion rates of carbon sources, byproducts and selected amino acids, and optimized for growth. Predicted growth rates are represented by red bars. Different datasets used are indicated with letters - abbreviations: glc: glucose; glyc: glycerol.*

In the second *in silico* experiment, predicted flux distributions in the central metabolism, both with and without strain- and condition-specific experimentally determined constraints, were investigated. The resulting uptake- and secretion rates of main metabolites were compared to experimentally obtained values (Figure 6). One of the main goals of using a model like this is to probe the effects of genetic and media perturbations *in silico*, *i.e.* without having to do extensive condition-specific cultivations and measurements beforehand. We used the model to predict the effects of adding glycerol to the glucose-based medium and knocking out the *adhE* gene, which plays a critical role in ethanol production and redox balance. For this purpose, we constrained the model only with the uptake rates of glucose and selected amino acids measured for the WT* strain (Figure 6, case 1). This analysis showed that the model predicted the same changes in fluxes of key metabolites in and out of the cell as observed *in vivo*. By adding glycerol to the medium, acetate production increased at the expense of ethanol production and the glycerol was metabolized to 1,3-propanediol. By knocking out the *adhE* gene, the growth rate dropped significantly due to redox imbalance, which in turn was rescued by adding glycerol to the medium.

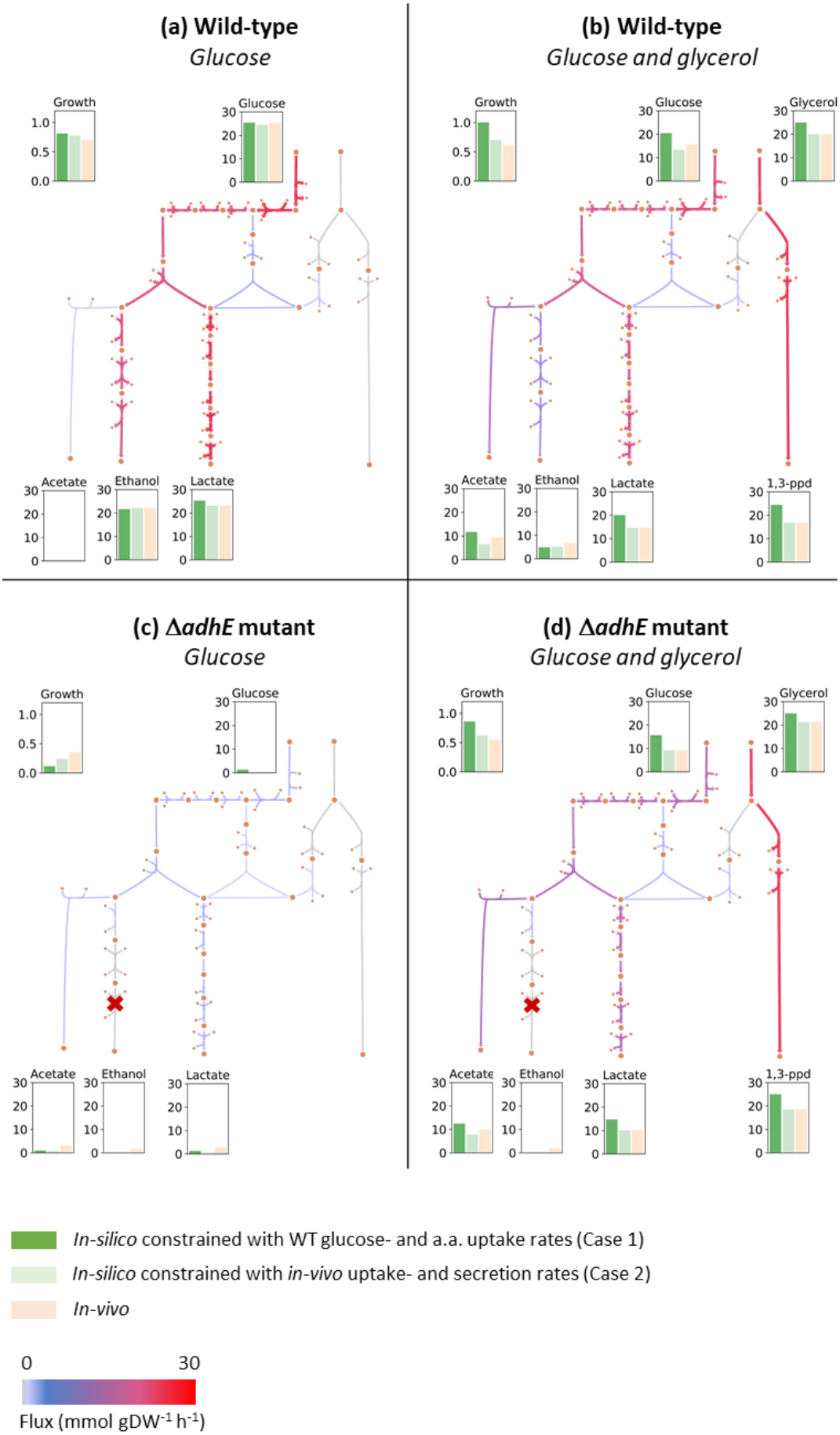

(a) Wild-type
Glucose
(b) Wild-type
Glucose and glycerol
(c) ΔadhE mutant
Glucose
(d) ΔadhE mutant
Glucose and glycerol
Growth
Glucose
Glycerol
Acetate
Ethanol
Lactate
1,3-ppd
In-silico constrained with WT glucose- and a.a. uptake rates (Case 1)
In-silico constrained with in-vivo uptake- and secretion rates (Case 2)
In-vivo
0
30
Flux (mmol gDW-1 h-1)

Figure 6 *Predicted and experimental fluxes of key metabolites in the wild-type strain (SJ) and the adhE mutant. The wild-type strain was grown on glucose (a) and glucose and glycerol (b), and the adhE mutant was also grown on glucose (c) and glucose and glycerol (d). Bar plots show the average measured rates from 3 replicates (light orange), predicted rates from model constrained with average experimental uptake rates of the WT grown on glucose, or case 1 (dark green), and predicted rates from model constrained with average experimental rates from the strain and condition under study, or case 2 (light green). Metabolic maps show predicted flux distributions for case 1. All units for uptake- and secretion rates are in mmol gDW^{-1} h^{-1} and for growth rates in h^{-1}.*

Although the model correctly predicted changes in flux distribution, it predicted higher uptake rates of glucose and glycerol than observed experimentally, which subsequently resulted in predicted values of secretion rates of by-products and growth rates being too high. The lower uptake rates observed *in vivo* are most likely due to metabolic regulations and enzyme kinetics, which are outside the scope of the model. In this second *in silico* experiment we also constrained the model with experimental measurements for each of the four conditions (Figure 6, case 2), by allowing, but not forcing, the model to take up and secrete metabolites according to the experimental data. Here, we wanted to see if the model, when imposed with realistic limitations, "chooses" a flux distribution which results in extracellular fluxes of metabolites in line with *in vivo* data. The model predictions and *in vivo* values were in excellent agreement.

3.2.3 *In-silico* analysis of *L. reuteri* as a cell factory

It has been shown that heterologous expression of the methylglyoxal synthase (*mgs*) gene in *L. reuteri* results in the production of 1,2-propanediol and 1-propanol [123]. Both these compounds have many industrial applications, e.g. production of polyester resins from 1,2-propanediol and the use of 1-propanol [28] as a solvent and biofuel. In addition to the 1,2-propanediol pathway, several different pathways have been described for 1-propanol production in other organisms [29]. The model was used to analyze the suitability of six different pathways for 1-propanol production in *L. reuteri*, the 1,2-propanediol, succinate, citramalate, threonine, acrylate and methylmalonyl pathways. These pathways have been described in literature [29], but in addition we used the minRxn algorithm [97] and the accompanying database of enzymatic reactions to search for heterologous pathways from glucose to 1-propanol. This search did not reveal pathways that differed significantly from the published pathways, neither qualitatively nor in terms of carbon yields. The highest carbon yields (40-45%) were observed in the 1,2-PDO, succinate, acrylate and methylmalonyl pathways. One of the advantages of using the 1,2-propanediol pathway in *L. reuteri* for 1-

propanol production is that *L. reuteri* produces vitamin B12, which is needed as a co-factor for the B12-dependent diol dehydratase step and has been suggested to be a limiting factor in this pathway [30] [31]. Another advantage is that only one gene (*mgs*) needs to be heterologously expressed to activate the pathway. For these reasons we decided to analyze 1-propanol production via the 1,2-propanediol pathway further.

The MGS reaction was added to the model, which was then maximized for growth while constrained with experimental uptake rates of glucose and selected amino acids from the WT grown on glucose (same as case 1 in Figure 6). The *adhE* mutant grows poorly on glucose due to redox imbalance (Figure 6c), but the synthesis of both 1,2-propanediol and 1-propanol consume NADH and therefore have the potential to restore the growth. The *adhE* gene was knocked out *in silico*, and flux predictions with (Figure 7) and without (Figure 6c) an active 1,2-propanediol pathway were compared. The active 1,2- propanediol pathway resulted in a high increase in growth rate (0.11 to 0.49 h^{-1}) as well as growth-coupled production of 1-propanol (14.7 mmol gDW^{-1} h^{-1}).

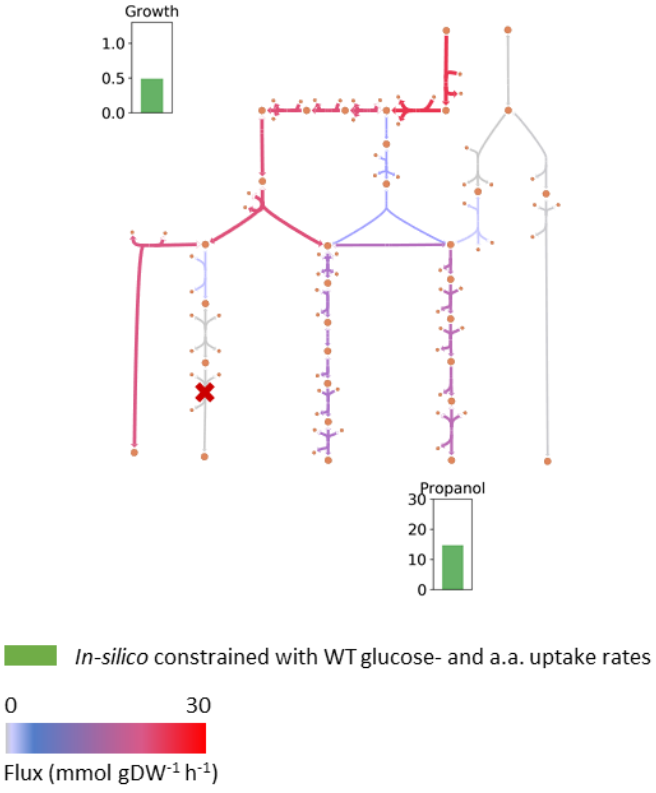


Figure 7 *Predicted flux distribution, growth rate and 1-propanol production of adhE mutant grown on glucose, with active 1,2-propanediol and 1-propanol pathways. The model was constrained with average experimental uptake rates of the WT grown on glucose and optimized for growth. Units for propanol secretion rate is in mmol gDW^{-1} h^{-1} and growth rate in h^{-1}.*

3.3 Engineering the carotenoid biosynthetic pathway in *Rhodothermus marinus* for lycopene production (Paper 2)

This work describes the genetic engineering of the carotenoid biosynthetic pathway in *Rhodothermus marinus*, with the aim of obtaining a lycopene producing mutant. Lycopene is an industrially relevant carotenoid as well as a precursor for various other important carotenoids. *R. marinus* has the potential to be well suited for biorefineries. It is an aerobic thermophile that produces thermostable enzymes and can utilize polysaccharides from different 2nd and 3rd generation biomass. It produces carotenoids natively, but since they are not in demand by the industry, we set out to alter them. Three different mutants were obtained, all modified in the carotenoid gene cluster. One mutant was shown to produce 0.49 g/kg CDW of lycopene, as the sole carotenoid. This is 1-2 orders of magnitude less lycopene compared to optimized commercial lycopene producing microorganisms [124], [125], but the production can potentially be optimized through further genetic engineering and culture conditions. Lastly, the carotenoid biosynthetic pathway that was proposed in **Paper 4** was revised, based on the results obtained here.

3.3.1 Carotenoid biosynthetic genes

Several strains of the thermophilic bacterium *R. marinus* produce carotenoids derived from γ-carotene. A gene cluster, present in the genomes of both the *R. marinus* type strain DSM 4252^T and strain ISCaR-493, was previously identified (Bjornsdottir et al., 2011, **Paper 4**). Here the analysis of the genome sequence was expanded, resulting in further identification of carotenoid biosynthetic genes. The carotenoid gene *cruF*, which likely encodes a 1',2'-hydratase [127], was identified in the gene cluster. In other organisms, this gene is usually found as a separate open reading frame (ORF), but in *R. marinus* it is fused together with the carotenoid gene *crtB*, which encodes a phytoene synthase. The *cruF* is without a stop codon, which makes it a single ORF together with the downstream *crtB* (Figure 8A). This gene is located in the smaller operon of the gene cluster and was one of the targets for genetic modifications in this study (Figure 8C and D). The other target is located on the larger operon and includes the carotenoid genes *crtY* and *crtO*, encoding lycopene cyclase and carotene ketolase, respectively (Figure 8B).

3.3.2 Genetic modifications to alter carotenoid production

Three different genetic modifications were performed in the carotenoid gene cluster of *R. marinus* strain SB-62 (Δ*trpB*Δ*purA*) (Figure 8A), resulting in mutant strains TK-1 (Δ*trpB*Δ*purA*Δ*crtYO::trpB*) (Figure 8B), TK-2 (Δ*trpB*Δ*purA*Δ*cruF::trpB*) (Figure 8C) and TK-3 (Δ*trpB*Δ*purA*Δ*cruFcrtB::trpBcrtB*$_{T.thermophilus}$) (Figure 8D). In TK-1 a 5890 bp region from the larger operon was deleted. This included two known carotenoid genes, *crtY* and *crtO*, and three genes with unclear functions in relation to the carotenoid pathway. One of these genes showed slight homology (36% similarity) to the carotenoid gene *crtI* (phytoene desaturase), which led us to hypothesize that it might encode a 3',4'-desaturase. Carotenoid structure analysis showed that the only difference between the carotenoids produced by TK-1 and SB-62 was the 4-keto β-ionone ring, which was not present in TK-1 (Figure 8B). This suggests that *crtY* and *crtO* are correctly annotated as lycopene cyclase and phytoene desaturase, and that they are the only genes in this region to be involved in the carotenoid biosynthetic pathway. In TK-2 the *cruF* part of the fused *cruF-crtB* gene was deleted leaving the *crtB* part intact. The *crtB* part of the gene includes a start codon, which raised the question if it could encode a functional enzyme without the *cruF* part. Cultures of this mutant were not visibly red colored, as seen for the other strains. Furthermore, extracted carotenoids from TK-2 showed no absorbance between 350 – 600 nm, confirming the absence of carotenoids. This data suggests that the fused *cruF-crtB* gene is transcribed as a whole, and without the *cruF* part, *crtB* does not encode a functioning enzyme (Figure 8C). In TK-3 the entire *cruF-crtB* gene was deleted and the *crtB* gene from *Thermus thermophilus* was inserted. Carotenoid structure analysis showed that this mutant produced lycopene as the sole carotenoid, indicating the *crtB* gene from *T. thermophilus* encodes a functioning enzyme in *R. marinus*.

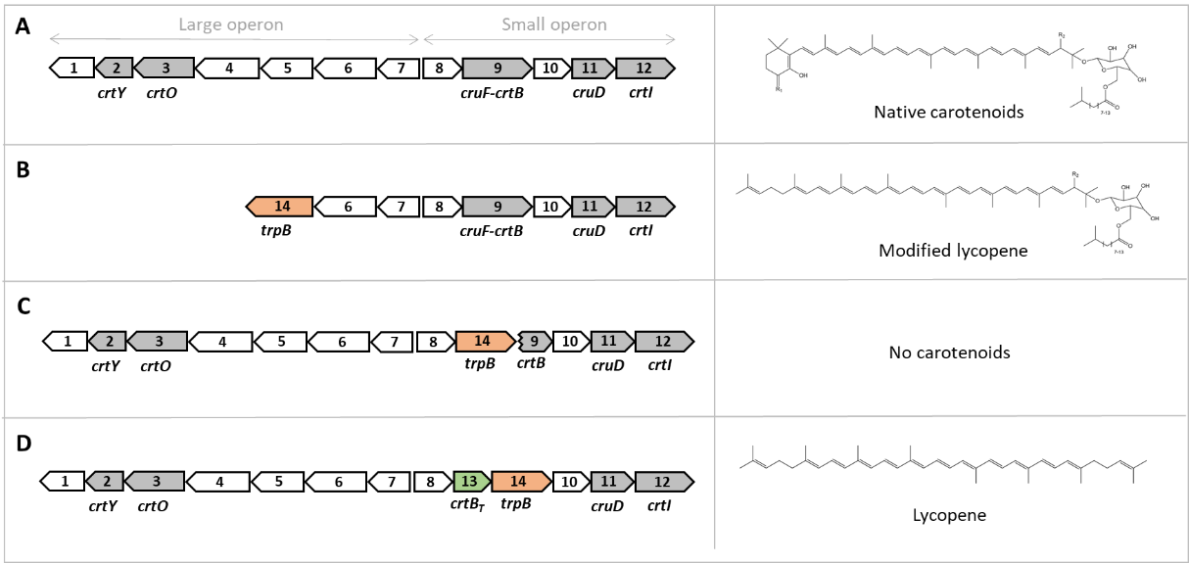


Figure 8 *The carotenoid gene cluster in the R. marinus stain SB-62 (ΔtrpBΔpurA) (**A**) and three different genetically modified strains: mutant TK-1 (ΔtrpBΔpurAΔcrtYO::trpB) (**B**), mutant TK-2 (ΔtrpBΔpurAΔcruF::trpB) (**C**) and mutant TK-3 (ΔtrpBΔpurAΔcruFcrtB::trpBcrtB$_{T.thermophilus}$) (**D**). Genes involved in the carotenoid biosynthesis are colored grey, the selective marker (trpB) used in the genetic modifications is colored orange and the crtB gene from T. thermophilus is colored green. The chemical structures of identified carotenoids of each strain are shown to the right side of the gene clusters (native carotenoids: 4-keto 2'-hydroxy β,ψ-carotene acyl glycoside, modified lycopene: 2'-hydroxy ψ,ψ-carotene acyl glycoside).*

3.3.3 Carotenoid biosynthetic pathway

In addition to the carotenoid producing mutants that were obtained in this study, the different genetic modifications and the corresponding carotenoid structure analysis helped clarify the biosynthetic pathway, which enzymes are involved and in what order they act. Since *cruF* and *crtB* are fused in one gene, it can be argued that their corresponding enzymes likely act together or successively. Whether the enzymes are fused into one polypeptide chain or not cannot be concluded from the data presented here. However, our results do show that CrtB is inactive without CruF, which suggests that the two enzymes might be acting as one entity. Two steps in the pathway are catalyzed by unknown enzymes in *R. marinus*. They add the C-2' hydroxyl group and the C-3',4' double bond to the carotene backbone. CrtI is a desaturase and has previously been reported to produce 3,4-didehydrolycopene in *Neurospora crassa* [128]. It is therefore possible that the CrtI in *R. marinus* is responsible for the 3',4'-desaturation, but while we do not have stronger evidence of that, this step remains unknown. The deletion of *crtY* and *crtO* only resulted in the absence of modifications catalyzed by their corresponding enzymes, suggesting the remaining enzymes of the pathway are not dependent of their activity. However, deleting

cruF resulted in the absence of modifications catalyzed by all other enzymes of the pathway, except for CrtB and CrtI. This suggests that CrtY and CrtO are dependent on the modifications on the right side of the backbone. Based on the results and conclusions from this study, we proposed a revised carotenoid biosynthetic pathway in *R. marinus* (Figure 9).

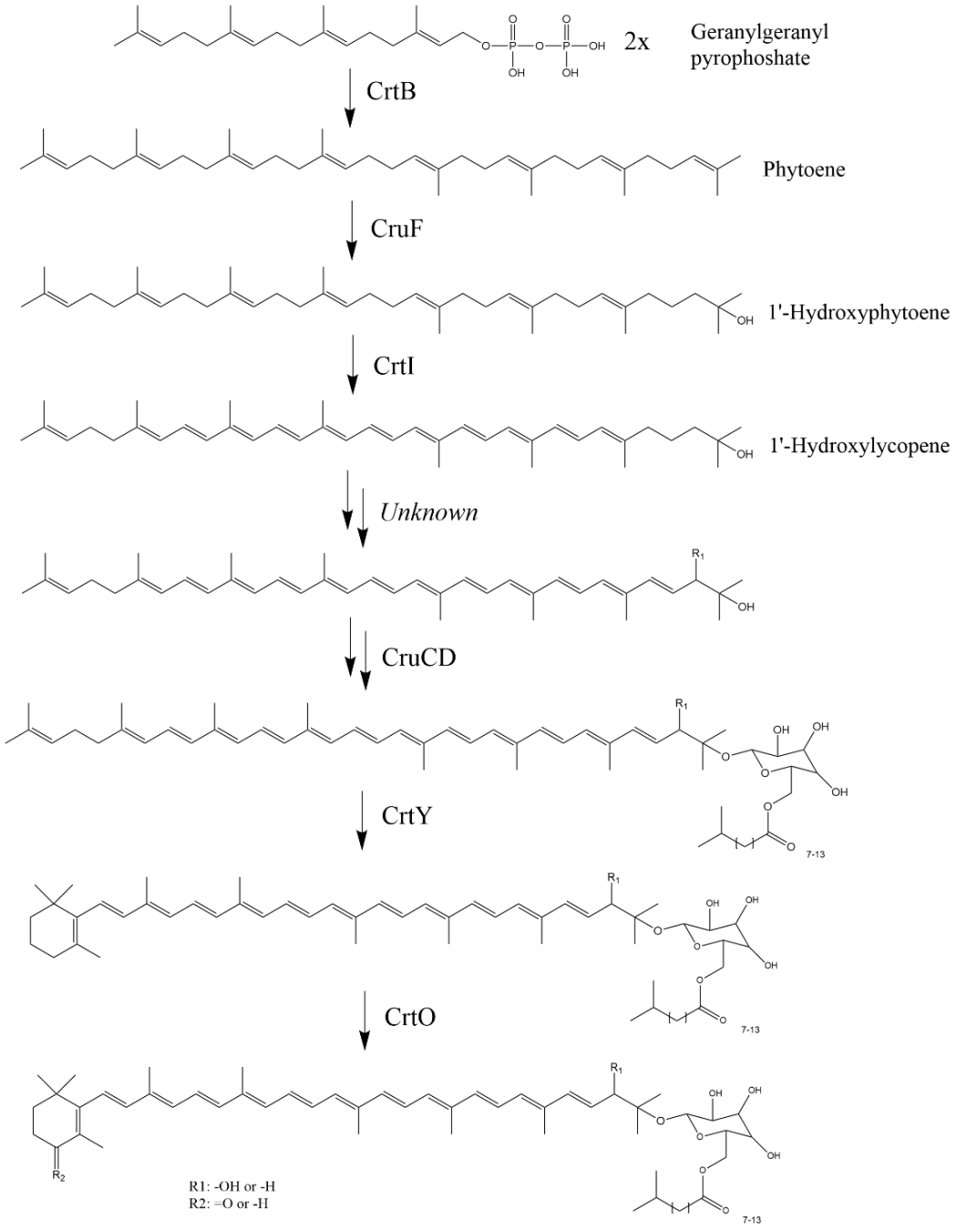


Figure 9 *Proposed carotenoid biosynthetic pathway in R. marinus, based on (Ron et al., 2018), with added information obtained in this study. Molecular structures and abbreviated enzyme names are shown, phytoene synthase (CrtB), 1',2'-hydratase (CruF), phytoene desaturase (CrtI), glycosyltransferase (CruC), acyltransferase (CruD), lycopene cyclase (CrtY) and carotene ketolase (CrtO).*

3.4 A genome-scale metabolic reconstruction provides insight into the metabolism of the thermophilic bacterium *Rhodothermus marinus* (Paper 3)

This work describes the reconstruction, curation and validation of a genome-scale metabolic model of *Rhodothermus marinus*. *R. marinus* is an interesting candidate for biorefineries and as a cell factory, due to its thermostability, native production of compounds such as carotenoids and exopolysaccharides, its ability to grow on a wide range of carbon sources including polysaccharides, and available genetic tools. Here we reviewed the metabolism of *R. marinus* and focused on biotechnologically relevant compounds, especially carotenoids. The genomes of two strains were compared and the analysis of the core genome indicated that the model could be used for both strains. Experimental growth data for both strains was obtained to validate model predications. Lastly, a cultivation experiment was performed where the effects of different growth conditions on cell density and carotenoid production were investigated.

3.4.1 Reconstruction of the genome-scale metabolic model

A genome-scale metabolic model of *R. marinus* DSM 4252^T was reconstructed based on the annotated genome using the automatic reconstruction tool Model SEED [83]. This draft reconstruction was revised using other well curated models [129], [130], and experimental data obtained from laboratory experiments, literature and databases. Several interesting features in the metabolism of *R. marinus* were reviewed and included in the reconstruction. We highlight some of them here and related them to a growth and carotenoid study in section 3.4.4.

Estimation of flux in the central metabolism of three thermophiles, including *R. marinus*, using 13C metabolic flux analysis [57] demonstrated that two central metabolic pathways in *R. marinus* are highly active, the EMP pathway and the TCA cycle. *R. marinus* can grow on many carbon sources, such as the di- and monosaccharides lactose and galactose. The genes encoding the enzymes needed to metabolize these sugars are present in the genome, except for phosphoglucomutase (EC 5.4.2.2), which catalyzes the conversion of glucose-1-

phosphate to glucose-6-phosphate. A homology search showed similarity of known phosphoglucomutase genes from other bacteria to gene RMAR_RS08875 (E value 1e-25) which is annotated as a phosphoglucosamine mutase (EC 5.4.2.10). Phosphoglucosamine mutase in *E. coli* has been shown to be able to catalyze the conversion of glucose-1-phosphate to glucose-6-phosphate at a high rate if the Ser100 residue in the polypeptide chain is replaced by a threonine residue [131]. The corresponding residue in gene RMAR_RS08875 in *R. marinus* was found to be a threonine residue, which indicates that the enzyme encoded by RMAR_RS08875 is a likely candidate for the interconversion of glucose-6-phosphate and glucose-1-phosphate in *R. marinus*.

R. marinus can grow on different polysaccharides, including the major components of brown algae, laminarin and alginate, as seen in **paper 5**. *R. marinus* has four genes encoding alginate lyases that, together, depolymerize alginate into the same unsaturated mono-uronate derivative of the M and G units. These monouronates are metabolized into pyruvate and glyceraldehyde 3-phosphate in several steps, which then enter the central metabolism. These steps include a part of the Entner-Doudoroff (ED) pathway. The ED pathway is only partially present in *R. marinus*, not enough to metabolize glucose through it [57], but enough to metabolize alginate.

The precursors for carotenoid production in *R. marinus* are produced via the non-mevalonate pathway, also known as the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway. The substrates for this pathway are the same as the products from the alginate degradation pathway, pyruvate and glyceraldehyde 3-phosphate. The gene encoding the first step in the MEP pathway, 1-deoxy-D-xylulose-5-phosphate synthase (*dxs*), was not found in the *R. marinus* genome. Different hypotheses of alternative routes to by-pass the first step include a promiscuous pyruvate dehydrogenase complex [132] and a by-product from an alternative methionine salvage pathway [133]. These hypotheses have however not been conclusively studied in *R. marinus*. Carotenoid production in many non-phototrophic bacteria is light induced [134]–[137], regulated by a MerR family transcriptional regulator, LitR. A homologue for this gene was found in the carotenoid gene cluster of *R. marinus*, as shown in **paper 2**.

The biomass reaction of *R. marinus* was assembled from *R. marinus* specific data obtained in this study and from literature [45]–[47], [57], [138], [139] and data from *E. coli* [140] when *R. marinus* data was not available. Sensitivity analysis, which showed how much variation in each macromolecule affects the predicted growth rate, was performed. This showed that the predicted growth rate was most sensitive to changes in the protein component, followed by the lipid component.

3.4.2 Comparing genomes of strains DSM 4252^T and ISCaR-493

The metabolic model was reconstructed based on genomic information from *R. marinus* DSM 4252^T. This stain is not amenable to genetic manipulation, while genetic tools are available for *R. marinus* ISCaR-493. The genome of ISCaR-493 was sequenced and the resulting contigs were annotated and compared to the genome of strain DSM 4252^T using the in-house pan-genomic software Genset. The strains are the same species, but some differences were observed. Most differences were observed in genes with unknown or poorly characterized functions, while most similarity was observed in metabolic genes. A total of 578 genes were used in the metabolic network modelling of *R. marinus* 4252^T and the great majority of them belonged to the core genome. Only seven model genes were absent in strain ISCaR-493 and four of them were replaced by genes encoding isozymes that exhibited high similarity to the DSM 4252^T enzymes. The remaining three genes are involved in EPS formation and xylan- and alginate degradation. EPSs of both strains have been previously studied [46] and shown to be of similar structures. It was also observed that strain ISCaR-493 grows well in defined medium with xylan and alginate as the sole carbon sources. This analysis suggests that the model is applicable for both strains DSM 4252^T and ISCaR-493.

3.4.3 Model validation

Experimental growth data for strains DSM 4252 and ISCaR-493, obtained in bioreactors, was used to validate model predictions (Figure 10). Measured uptake of glucose and pyruvate and secretion of acetate and lactate during growth was used to constrain the model for simulations. The experimental growth data showed that the growth of *R. marinus* did not follow the typical batch growth curve of bacteria. Instead of following a steady exponential phase followed by a stationary phase, the specific growth rate decreased over time and the secretion of organic acids increased. Oxygen limitation in the cells could be the reason for this trend in the growth phase. However, oxygen levels were carefully monitored in the

medium during growth. A possible explanation for why the cells experience oxygen limitation in medium with access oxygen levels is cell aggregation [141]. Aggregation of several *R. marinus* strains have been reported previously [40]. *R. marinus* produces considerable amounts of exopolysaccharides [46], which can cause cells to aggregate [142]. In the beginning of the growth phase, each strain followed a similar exponential phase before the growth rates started decreasing (Figure 10b) (time points 3-6 for strain DSM 4252 and 1-5 for strain ISCaR-493). Data from these timepoints was used here. This analysis showed that the model accurately predicts growth for both strains (Figure 10a).

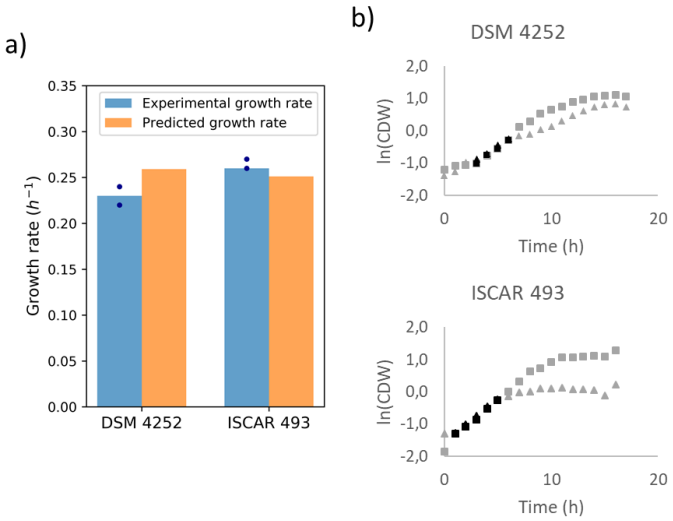


Figure 10 *Experimental growth data for strains DSM 4252^T and ISCaR-493 used to validate model growth predictions. The model was constrained with uptake (glucose and pyruvate) and secretion (lactate and acetate) rates observed in vivo and optimized for growth. Predicted growth rates were compared to growth rates observed in vivo (**a**). Experimental data from the exponential growth phase of two replicates (▲ and ■), obtained early in the growth phase (time points 3-6 for strain DSM 4252 and 1-5 for strain ISCaR-493), was used (**b**).*

3.4.4 Effects of external factors on carotenoid production

To better understand carotenoid production in *R. marinus,* a cultivation experiment comparing different culture conditions was performed where cell density and carotenoid production was observed following 24 hours of growth (Figure 11). The different conditions included glucose, pyruvate and alginate as carbon sources in the medium, light as a stimuli and heterologous expression of the missing *dxs* gene, which encodes 1-deoxy-D-xylulose-5-phosphate synthase, the first step in the MEP pathway. This experiment showed that the highest cell density was obtained in glucose medium supplemented with pyruvate, but high carotenoid production was obtained when it was grown on alginate, when pyruvate was added to a glucose based medium and in light conditions. It also showed that the production

increased under starvation conditions, indicating that yields can potentially be increased by either allowing the culture to reach and stay in stationary phase or transfer the cells after growth to new medium with limited or no carbon source. The motivation for the latter is that after the growth phase, the medium might not be optimal, e.g. due to accumulation of by-products that alter pH, and the cells might stay alive and produce carotenoids longer in fresh medium. Finally, cloning the *dxs* gene from *T. thermophilus* in *R. marinus* resulted in the highest yields of carotenoids, but much lower cell density compared to strain ISCaR-493.

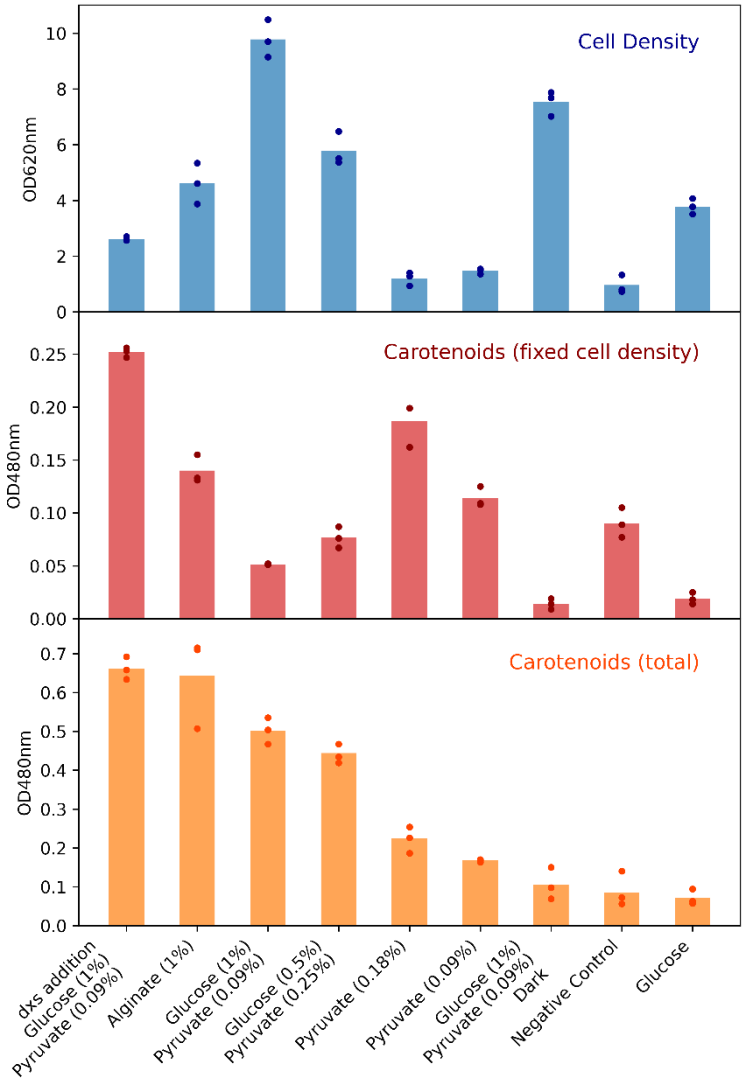


Figure 11 *Cell density (OD620nm) and carotenoid (OD480nm) production following growth of R. marinus strain ISCaR-493 for 24 hours on glucose; mixture of glucose and pyruvate under light and dark conditions; alginate; pyruvate and without any carbon sources. Additionally, the mutant strain TK-4 (ΔtrpBΔpurA::trpBdxs$_{T.thermophilus}$) was grown on a mixture of glucose and pyruvate. The carotenoids were always extracted from a fixed number of cells (OD620nm = 1). The total amount of carotenoids was calculated by multiplying the cell density by the measured carotenoids. The average cell density of each culture condition is represented by a blue bar, the average measured carotenoid value as a red bar and the average total carotenoids by an orange bar. Dots represent individual replicates.*

3.5 Other papers

The following papers describe work that was mostly planned and performed by my colleagues, with some contribution from me. I briefly describe the main results of each study, followed by a more detailed description of my contribution.

3.5.1 Characterization of carotenoids in *Rhodothermus marinus* (Paper 4)

In this study the carotenoids of *R. marinus* strains DSM 4252^T, DSM 4253 and ISCaR-493 (or PRI 493 as it was referred to in this study) were characterized. The strains were cultured in bioreactors, the carotenoids separated from the cells using pressurized liquid extraction and the carotenoids analyzed using ultra-high performance supercritical fluid chromatography with diode array and quadropole time-of flight mass spectrometry detection. The analysis showed that Salinixanthin, the carotenoid first found in *Salinibacter ruber*, was detected in hydroxylated and non-hydroxylated form in all strains. Additionally, the same structures were also found without the 4-keto group on the β-ionone ring.

My contribution to this work was to analyze the genome in regard to carotenoid biosynthetic genes and to suggest a biosynthetic pathway for the carotenoids in *R. marinus*. The proposed pathway was based on the analyzed structures of the carotenoids, the analysis of the genes found in the genome and pathways observed in other bacteria producing structurally similar carotenoids. We identified six carotenoid biosynthetic genes in the genome, most of them located in a single gene cluster. The identified genes were phytoene synthase (*crtB*), phytoene desaturase (*crtI*), lycopene cyclase (*crtY*), carotene ketolase (*crtO*), glycosyltransferase (*cruC*) and acyltransferase (*cruD*). Three additional genes encoding enzymes that were suggested to partake in the pathway based on the carotenoid structures were not identified in the genome. They are 1',2'-hydratase (*cruF* or *crtC*), 3',4'-desaturase (*crtD*) and monooxygenase (*crtA*). These steps in the pathway were studied in more detail in **paper 2**.

3.5.2 Composition analysis and minimal treatments to solubilize polysaccharides from the brown seaweed *Laminaria digitata* for microbial growth of thermophiles (Paper 5)

In this study two methods for pre-treating the brown algae *Laminaria digitata* to release polysaccharides were investigated: hot water treatment by autoclaving and two-step extraction with mild acid followed by alkaline treatment. Additionally, cultivations were performed of two thermophilic bacteria, *R. marinus* strain ISCaR-493 (referred to as DSM 16675 in this study) and *Bacillus methanolicus* strain MGA3, in medium supplemented with the pre-treated algae on one hand and the major polysaccharides found in brown algae on the other (commercial compounds).

My contribution to this work was to cultivate *R. marinus* in defined medium with sodium alginate as the carbon source. Two types of alginate were tested, high and low viscosity. Additionally, alginate in combination with low concentration of pyruvate was also tested (Figure 12). This was done in case *R. marinus* would struggle using alginate as the sole carbon source. This however was not the case, as it grew well in alginate medium, reaching OD (620nm) values of 4.7 on high viscosity alginate and 2.4 on low viscosity alginate. Pyruvate addition did not increase the OD. The color of *R. marinus* cultured in alginate medium was deep red and was investigated and discussed further in **paper 3**.

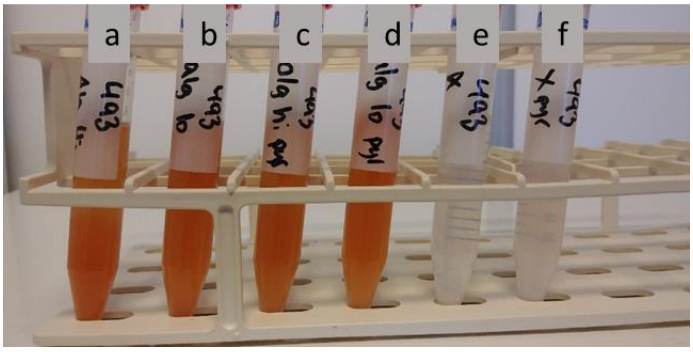


Figure 12 *R. marinus grown for 24 hours at 65°C in defined medium with* ***a)*** *1% high viscosity sodium alginate,* ***b)*** *1% low viscosity sodium alginate,* ***c)*** *1% high viscosity sodium alginate and 0.02% pyruvate,* ***d)*** *1% low viscosity sodium alginate and 0.02% pyruvate,* ***e)*** *no carbon source and* ***f)*** *0.02% pyruvate.*

3.5.3 Efficient genome editing of an extreme thermophile, Thermus thermophilus, using a thermostable Cas9 variant (Paper 6)

In this study a thermostable CRISPR-Cas9 system was developed, for genome editing in thermophilic bacteria. CaldoCas9 was identified and an associated guide RNA was designed, which together had targetable nuclease activity *in vitro* at temperatures up to 65°C. Additionally, the protospacer adjacent motif (PAM) specificity of CaldoCas9 was characterized.

My contribution to this work was to help establish the function of the system *in vivo* for *Thermus thermophilus*. A plasmid vector with the CaldoCas9 was constructed and used to carry out gene deletions, both from the chromosome and the mega plasmid. The carotenoid biosynthetic gene *crtI*, encoding phytoene desaturase, was deleted from the mega plasmid and the adenine biosynthetic gene *purA*, encoding adenylosuccinate synthase, was deleted from the chromosome. We were able to verify the gene deletions in obtained clones by analyzing both the phenotype and the genotype. The Δ*crtI* clones did not produce the native yellow pigment and appeared white. The Δ*purA* clones were only able to grow in defined medium when adenine was supplemented, different from the WT. The genotype of obtained clones was evaluated using PCR and sequencing. This analysis showed that correct clones were obtained at a frequency of about 90%. Lastly, we were able to cure the vector from the obtained clones, making it possible to easily reuse the system for further genome editing.

3.6 Conclusions

Metabolic modelling and engineering were used to study the potential of two non-model organisms, *L. reuteri* and *R. marinus*, as cell factories. A genome-scale metabolic model of *L. reuteri* JCM 1112 was reconstructed and validated. The model correctly predicted the growth rate, as well as uptake- and secretion rates of the main metabolites of two strains: the WT and an ethanol (Δ*adhE*) mutant. This was tested in defined medium with glucose as the sole carbon source on one hand and both glucose and glycerol as carbon sources on the other. The main results from this analysis were that the growth was highly reduced in the ethanol mutant compared to WT, but growth was rescued by adding glycerol to the glucose-based medium. Lastly, the model was used to predict the suitability of using *L. reuteri* as a cell factory for the production of the non-native compound 1-propanol. This analysis showed that with minimum genetic engineering, 1-propanol could likely be produced through the 1,2-propanediol pathway at a theoretical carbon yield of 40 - 45% and if done in the ethanol mutant, the 1-propanol production would be growth coupled. These experiments indicate that the model simulations are accurate, and the model can be used for future metabolic engineering studies in the future.

The carotenoid biosynthetic pathway in *R. marinus* was proposed, using structural and genomic information as well as the literature. Three different genetic modifications were performed in the carotenoid gene cluster, resulting in a lycopene producing mutant. This is the first *R. marinus* strain that produces an industrially relevant carotenoid as a final product, which strengthens the hypothesis that *R. marinus* can be used as a cell factory and is a vital step towards obtaining a production strain for carotenoids that can be used in industry. A genome-scale metabolic model of *R. marinus* DSM 4252^T was reconstructed and validated. The genome of strain ISCaR-493, which is amenable to genetic manipulation, was subsequently sequenced and the genomes of the two strains compared. This analysis, along with experimental growth data, showed that the model can be used for both strains and correctly predicts growth rates. During the reconstruction process, several interesting features in the metabolism were highlighted. Subsequently, a growth study was performed where cell density and carotenoid production of strain ISCaR-493 under different growth conditions was performed. This showed that the highest cell density was obtained by adding pyruvate to glucose-based medium. Also, carotenoid yields increased when *R. marinus* was

grown in light, compared to dark; when alginate was used as carbon source compared to glucose; when pyruvate was added to the medium; under starving conditions; and when the *dxs* gene from *T. thermophilus*, which encodes the enzyme that catalyzes the first step in the MEP terpenoid pathway and is missing in *R. marinus*, was cloned into *R. marinus*.

The currently available genetic engineering method for *R. marinus* only has two selective markers (*trpB* and *purA*), so the number of possible genome edits are limited. A thermostable CRISPR-Cas9 system has been developed and used in *T. thermophilus*. This system would offer a great improvement in genetic engineering in *R. marinus*, as the same selective marker could be recycled. The work on adapting this system to *R. marinus* is now underway. With the possibility of multiple genome edits, the opportunities for further metabolic engineering arise. Astaxanthin and β-carotene are in high demand by the industry and it would be an interesting engineering strategy to produce these compounds in *R. marinus*. This would require several gene deletions and insertions in the carotenoid gene cluster, which could include insertion of genes from *T. thermophilus*. Furthermore, attempts to increase the carotenoid yields through genetic engineering could be performed, such as overexpression of key terpenoid and carotenoid genes. Another interesting engineering strategy could be to improve the capability of *R. marinus* to utilize cellulose. *R. marinus* can utilize many polysaccharides of 2nd and 3rd generation biomass, including the hemicellulose xylan, but it can only poorly degrade cellulose. Expressing heterologous cellulases in *R. marinus* might make it a better candidate to break down wood-based biomass.

Replacing fossil fueled chemical synthesis with sustainable technologies such as biorefineries, where renewable biomass is used as feedstock instead of fossil fuels, offers solutions in the current climate crisis. Furthermore, underutilized biomass and waste can be used for this purpose, instead of using edible crops. Robust non-model organisms, that are evolved to endure extreme environments, are great candidates as cell factories in biorefineries. Efficient genetic engineering methods and GEMs are important tools to study the potentials and limitations of these organisms, which in turn is crucial for exploiting them in industry. This work contributed to the field by developing two bacteria, which we had limited knowledge on but showed great promise as cell factories. Increased knowledge on their metabolism, and their abilities and limitations as cell factories, along with available metabolic models, improved genetic engineering methods and engineered strains, play an important part in paving the way to a more sustainable future.

REFERENCES

[1] T. R. Karl, J. M. Melillo, T. C. Peterson, and S. J. Hassol, *Global climate change impacts in the United States*. Cambridge University Press, 2009.

[2] M. Höök and X. Tang, "Depletion of fossil fuels and anthropogenic climate change—A review," *Energy Policy*, vol. 52, pp. 797–809, 2013.

[3] D. M. Alonso, J. Q. Bond, and J. A. Dumesic, "Catalytic conversion of biomass to biofuels," *Green Chem.*, vol. 12, no. 9, pp. 1493–1513, 2010.

[4] S. Takkellapati, T. Li, and M. A. Gonzalez, "An Overview of Biorefinery Derived Platform Chemicals from a Cellulose and Hemicellulose Biorefinery," *Clean Technol. Environ. policy*, vol. 20, no. 7, pp. 1615–1630, Sep. 2018.

[5] J. V Pham *et al.*, "A Review of the Microbial Production of Bioactive Natural Products and Biologics ," *Frontiers in Microbiology* , vol. 10. p. 1404, 2019.

[6] A. Dornau, J. F. Robson, G. H. Thomas, and S. J. McQueen-Mason, "Robust microorganisms for biofuel and chemical production from municipal solid waste," *Microb. Cell Fact.*, vol. 19, no. 1, p. 68, 2020.

[7] Z. Fatma, J. C. Schultz, and H. Zhao, "Recent advances in domesticating non-model microorganisms," *Biotechnol. Prog.*, 2020.

[8] T. E. Sandberg, M. J. Salazar, L. L. Weng, B. O. Palsson, and A. M. Feist, "The emergence of adaptive laboratory evolution as an efficient tool for biological discovery and industrial biotechnology," *Metab. Eng.*, vol. 56, pp. 1–16, 2019.

[9] M. Gustavsson and S. Y. Lee, "Prospects of microbial cell factories developed through systems metabolic engineering," *Microb. Biotechnol.*, vol. 9, no. 5, pp. 610–617, Sep. 2016.

[10] D. Zhu, W. A. Adebisi, F. Ahmad, S. Sethupathy, B. Danso, and J. Sun, "Recent Development of Extremophilic Bacteria and Their Application in Biorefinery," *Front. Bioeng. Biotechnol.*, vol. 8, p. 483, 2020.

[11] A. López-Contreras *et al.*, "Biorefinery Approach to the Use of Macroalgae as Feedstock for Biofuels," in *Algal Biofuels*, L. Pereira, Ed. CRC Press, 2017.

[12] W. Jiang, D. Bikard, D. Cox, F. Zhang, and L. A. Marraffini, "RNA-guided editing of bacterial genomes using CRISPR-Cas systems," *Nat. Biotechnol.*, vol. 31, no. 3, pp. 233–239, 2013.

[13] Y. Jiang, B. Chen, C. Duan, B. Sun, J. Yang, and S. Yang, “Multigene Editing in the Escherichia coli Genome via the CRISPR-Cas9 System,” *Appl. Environ. Microbiol.*, vol. 81, no. 7, pp. 2506–2514, Apr. 2015.

[14] A. W. Westbrook, M. Moo-Young, and C. P. Chou, “Development of a CRISPR-Cas9 Tool Kit for Comprehensive Engineering of Bacillus subtilis,” *Appl. Environ. Microbiol.*, vol. 82, no. 16, pp. 4876 LP – 4895, Aug. 2016.

[15] Q. Yan and S. S. Fong, “Challenges and Advances for Genetic Engineering of Non-model Bacteria and Uses in Consolidated Bioprocessing,” *Frontiers in Microbiology*, vol. 8. p. 2060, 2017.

[16] R. Barrangou *et al.*, “CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes,” *Science (80-.).*, vol. 315, no. 5819, pp. 1709 LP – 1712, Mar. 2007.

[17] I. Mougiakos *et al.*, “Characterizing a thermostable Cas9 for bacterial genome editing and silencing,” *Nat. Commun.*, vol. 8, no. 1, p. 1647, 2017.

[18] L. B. Harrington *et al.*, “A thermostable Cas9 with increased lifetime in human plasma,” *Nat. Commun.*, vol. 8, no. 1, p. 1424, 2017.

[19] T. K. M. Tsui, T. H. Hand, E. C. Duboy, and H. Li, “The Impact of DNA Topology and Guide Length on Target Selection by a Cytosine-Specific Cas9.,” *ACS Synth. Biol.*, vol. 6, no. 6, pp. 1103–1113, Jun. 2017.

[20] S. T. Schmidt, F. B. Yu, P. C. Blainey, A. P. May, and S. R. Quake, “Nucleic acid cleavage with a hyperthermophilic Cas9 from an uncultured Ignavibacterium,” *Proc. Natl. Acad. Sci.*, vol. 116, no. 46, pp. 23100 LP – 23105, Nov. 2019.

[21] G. Gasiunas *et al.*, “Biochemically diverse CRISPR-Cas9 orthologs,” *bioRxiv*, p. 2020.04.29.066654, Jan. 2020.

[22] E. F. Bosma, J. Forster, and A. T. Nielsen, “Lactobacilli and pediococci as versatile cell factories – Evaluation of strain properties and genetic tools,” *Biotechnol. Adv.*, vol. 35, no. 4, pp. 419–442, 2017.

[23] D. M. Saulnier *et al.*, “Exploring Metabolic Pathway Reconstruction and Genome-Wide Expression Profiling in Lactobacillus reuteri to Define Functional Probiotic Features,” *PLoS One*, vol. 6, no. 4, p. e18783, Jan. 2011.

[24] F. Santos *et al.*, “Pseudovitamin B12 is the corrinoid produced by Lactobacillus reuteri CRL1098 under anaerobic conditions,” *FEBS Lett.*, vol. 581, no. 25, pp. 4865–4870, 2007.

[25] F. Santos *et al.*, “The complete coenzyme B12biosynthesis gene cluster of Lactobacillus reuteri CRL 1098,” *Microbiology*, vol. 154, no. 1, pp. 81–93, 2008.

[26] J.-P. van Pijkeren and R. A. Britton, “High efficiency recombineering in lactic acid bacteria,” *Nucleic Acids Res.*, vol. 40, no. 10, pp. e76–e76, Feb. 2012.

[27] J.-H. Oh and J.-P. van Pijkeren, “CRISPR–Cas9-assisted recombineering in Lactobacillus reuteri,” *Nucleic Acids Res.*, vol. 42, no. 17, pp. e131–e131, Jul. 2014.

[28] Y.-S. Jang *et al.*, “Bio-based production of C2–C6 platform chemicals,” *Biotechnol. Bioeng.*, vol. 109, no. 10, pp. 2437–2459, 2012.

[29] T. Walther and J. M. François, “Microbial production of propanol,” *Biotechnol. Adv.*, vol. 34, no. 5, pp. 984–996, 2016.

[30] D. Siebert and V. F. Wendisch, “Metabolic pathway engineering for production of 1,2-propanediol and 1-propanol by Corynebacterium glutamicum,” *Biotechnol. Biofuels*, vol. 8, no. 1, pp. 1–13, 2015.

[31] R. Jain, X. Sun, Q. Yuan, and Y. Yan, “Systematically Engineering Escherichia coli for Enhanced Production of 1,2-Propanediol and 1-Propanol,” *ACS Synth. Biol.*, vol. 4, no. 6, pp. 746–756, 2015.

[32] R. Munoz, R. Rosselló-Móra, and R. Amann, “Revised phylogeny of Bacteroidetes and proposal of sixteen new taxa and two new combinations including Rhodothermaeota phyl. nov.,” *Syst. Appl. Microbiol.*, vol. 39, no. 5, pp. 281–296, 2016.

[33] G. A. Alfredsson, J. K. Kristjansson, S. Hjorleifsdottir, and K. O. Stetter, “Rhodothermus marinus, gen.nov., sp. nov., a Thermophilic, Halophilic Bacterium from Submarine Hot Springs in Iceland,” *J. Gen. Microbiol.*, no. 134, pp. 299–306, 1988.

[34] S. Hjorleifsdottir, A. Aevarsson, G. O. Hreggvidsson, O. H. Fridjonsson, and J. K. Kristjansson, “Isolation, growth and genome of the Rhodothermus RM378 thermophilic bacteriophage,” *Extremophiles*, vol. 18, no. 2, pp. 261–270, 2014.

[35] S. K. Petursdottir, G. O. Hreggvidsson, M. S. Da Costa, and J. K. Kristjansson, “Genetic diversity analysis of Rhodothermus reflects geographical origin of the isolates.,” *Extremophiles*, vol. 4, no. 5, pp. 267–274, Oct. 2000.

[36] O. C. Nunes, M. M. Donato, and M. S. Da Costa, “Isolation and Characterization of Rhodothermus Strains from S. Miguel, Azores,” *Syst. Appl. Microbiol.*, vol. 15, no. 1, pp. 92–97, 1992.

[37] Y. Sako, K. Takai, Y. Ishida, A. Uchida, and Y. Katayama, “Rhodothermus obamensis sp. nov., a modern lineage of extremely thermophilic marine bacteria.,” *Int. J. Syst. Bacteriol.*, vol. 46, no. 4, pp. 1099–1104, Oct. 1996.

[38] S. H. Bjornsdottir *et al.*, “Rhodothermus marinus : physiology and molecular biology,” *Extremophiles*, vol. 10, pp. 1–16, 2006.

[39] M. Nolan *et al.*, “Complete genome sequence of Rhodothermus marinus type strain (R-10 T),” *Stand. Genomic Sci.*, vol. 1, no. 3, pp. 283–291, 2009.

[40] S. H. Bjornsdottir, S. H. Thorbjarnardottir, and G. Eggertsson, “Establishment of a gene transfer system for Rhodothermus marinus,” *Appl. Microbiol. Biotechnol.*, vol. 66, no. 6, pp. 675–682, 2005.

[41] S. Halldórsdóttir *et al.*, “Cloning, sequencing and overexpression of a Rhodothermus marinus gene encoding a thermostable cellulase of glycosyl hydrolase family 12,” *Appl. Microbiol. Biotechnol.*, vol. 49, no. 3, pp. 277–284, 1998.

[42] M. Krah, R. Misselwitz, O. Politz, K. K. Thomsen, H. Welfle, and R. Borriss, “The laminarinase from thermophilic eubacterium Rhodothermus marinus,” *Eur. J. Biochem.*, vol. 257, no. 1, pp. 101–111, 1998.

[43] E. Nordberg Karlsson, E. Bartonek-Roxå, and O. Holst, “Evidence for substrate binding of a recombinant thermostable xylanase originating from Rhodothermus marinus,” *FEMS Microbiol. Lett.*, vol. 168, no. 1, pp. 1–7, Nov. 1998.

[44] E. Nordberg Karlsson, E. Bartonek-Roxå, and O. Holst, “Cloning and sequence of a thermostable multidomain xylanase from the bacterium Rhodothermus marinus.,” *Biochim. Biophys. Acta*, vol. 1353, no. 2, pp. 118–124, Aug. 1997.

[45] B. F. Lutnaes, Å. Strand, S. K. Pétursdóttir, and S. Liaaen-Jensen, “Carotenoids of thermophilic bacteria - Rhodothermus marinus from submarine Icelandic hot springs,” *Biochem. Syst. Ecol.*, vol. 32, no. 5, pp. 455–468, 2004.

[46] R. R. R. Sardari *et al.*, “Evaluation of the production of exopolysaccharides by two strains of the thermophilic bacterium Rhodothermus marinus,” *Carbohydr. Polym.*, vol. 156, pp. 1–8, 2017.

[47] K. Hamana, H. Hamana, M. Niitsu, K. Samejima, and S. Matsuzaki, “Distribution of unusual long and branched polyamines in thermophilic eubacteria belonging to "Rhodothermus, " Thermus and Thermonema,” *J. Gen. Appl. Microbiol.*, vol. 38, no. 6, pp. 575–584, 1992.

[48] O. C. Nunes, C. M. Manaia, M. S. Da Costa, and H. Santos, “Compatible Solutes in the Thermophilic Bacteria Rhodothermus marinus and ‘Thermus thermophilus,’” *Appl. Environ. Microbiol.*, vol. 61, no. 6, pp. 2351–2357, Jun. 1995.

[49] S. H. Bjornsdottir, “Genetic engineering of Rhodothermus marinus,” University of Iceland, 2010.

[50] S. H. Bjornsdottir, O. H. Fridjonsson, G. O. Hreggvidsson, and G. Eggertsson, “Generation of targeted deletions in the genome of Rhodothermus marinus,” *Appl. Environ. Microbiol.*, vol. 77, no. 15, pp. 5505–5512, 2011.

[51] S. H. Bjornsdottir, O. H. Fridjonsson, J. K. Kristjansson, and G. Eggertsson, “Cloning and expression of heterologous genes in Rhodothermus marinus,” *Extremophiles*, vol. 11, no. 2, pp. 283–293, 2007.

[52] Y. Qin, “Alginate fibres: an overview of the production processes and applications in wound management,” *Polym. Int.*, vol. 57, no. 2, pp. 171–180, Feb. 2008.

[53] G. O. Hreggvidsson *et al.*, “International Publication Number EP 3092247 A1.” 2016.

[54] G. O. Hreggvidsson and O. H. Fridjonsson, “International Publication Number WO

2020/044379 A1." 2020.

[55] R. Takase, A. Ochiai, B. Mikami, W. Hashimoto, and K. Murata, "Molecular identification of unsaturated uronate reductase prerequisite for alginate metabolism in Sphingomonas sp. A1.," *Biochim. Biophys. Acta*, vol. 1804, no. 9, pp. 1925–1936, Sep. 2010.

[56] J. Preiss and G. Ashwell, "Alginic Acid Metabolism in Bacteria. The Enzymatic Reduction of 4-deoxy-L-erythro-5-hexoseulose uronic acid to 2-keto-3-deoxy-D-guconic acid," *J. Biol. Chem.*, vol. 237, no. 2, pp. 317–321, 1962.

[57] L. T. Cordova, R. M. Cipolla, A. Swarup, C. P. Long, and M. R. Antoniewicz, "13C metabolic flux analysis of three divergent extremely thermophilic bacteria: Geobacillus sp. LC300, Thermus thermophilus HB8, and Rhodothermus marinus DSM 4252," *Metab. Eng.*, vol. 44, no. October, pp. 182–190, 2017.

[58] T. OSHIMA and K. IMAHORI, "ISOLATION OF AN EXTREME THERMOPHILE AND THERMOSTABILITY OF ITS TRANSFER RIBONUCLEIC ACID AND RIBOSOMES," *J. Gen. Appl. Microbiol.*, vol. 17, no. 6, pp. 513–517, 1971.

[59] F. Cava, A. Hidalgo, and J. Berenguer, "Thermus thermophilus as biological model," *Extremophiles*, vol. 13, no. 2, p. 213, 2009.

[60] Y. Koyama, T. Hoshino, N. Tomizuka, and K. Furukawa, "Genetic transformation of the extreme thermophile Thermus thermophilus and of other Thermus spp.," *J. Bacteriol.*, vol. 166, no. 1, pp. 338 LP – 340, Apr. 1986.

[61] H. Maseda and T. Hoshino, "Development of expression vectors for Thermus thermophilus," *J. Ferment. Bioeng.*, vol. 86, no. 1, pp. 121–124, 1998.

[62] M. de Grado, P. Castán, and J. Berenguer, "A high-transformation-efficiency cloning vector for Thermus thermophilus.," *Plasmid*, vol. 42, no. 3, pp. 241–245, Nov. 1999.

[63] A. Fujita, Y. Misumi, and Y. Koyama, "Two versatile shuttle vectors for Thermus thermophilus-Escherichia coli containing multiple cloning sites, lacZα gene and kanamycin or hygromycin resistance marker.," *Plasmid*, vol. 67, no. 3, pp. 272–275, May 2012.

[64] M. Tamakoshi, M. Uchida, K. Tanabe, S. Fukuyama, A. Yamagishi, and T. Oshima, "A new Thermus-Escherichia coli shuttle integration vector system.," *J. Bacteriol.*, vol. 179, no. 15, pp. 4811 LP – 4814, Aug. 1997.

[65] M. Tamakoshi, T. Yaoi, T. Oshima, and A. Yamagishi, "An efficient gene replacement and deletion system for an extreme thermophile, Thermus thermophilus.," *FEMS Microbiol. Lett.*, vol. 173, no. 2, pp. 431–437, Apr. 1999.

[66] J. Hoseki, T. Yano, Y. Koyama, S. Kuramitsu, and H. Kagamiyama, "Directed evolution of thermostable kanamycin-resistance gene: a convenient selection marker for Thermus thermophilus.," *J. Biochem.*, vol. 126, no. 5, pp. 951–956, Nov. 1999.

[67] Y. Hashimoto, T. Yano, S. Kuramitsu, and H. Kagamiyama, “Disruption of Thermus thermophilus genes by homologous recombination using a thermostable kanamycin-resistant marker.,” *FEBS Lett.*, vol. 506, no. 3, pp. 231–234, Oct. 2001.

[68] J. F. Carr, M. E. Danziger, A. L. Huang, A. E. Dahlberg, and S. T. Gregory, “Engineering the Genome of Thermus thermophilus Using a Counterselectable Marker,” *J. Bacteriol.*, vol. 197, no. 6, pp. 1135 LP – 1144, Mar. 2015.

[69] A. Khodayari, A. R. Zomorrodi, J. C. Liao, and C. D. Maranas, “A kinetic model of Escherichia coli core metabolism satisfying multiple sets of mutant flux data,” *Metab. Eng.*, vol. 25, pp. 50–62, 2014.

[70] Y. Zhang *et al.*, “Three-dimensional structural view of the central metabolic network of Thermotoga maritima.,” *Science*, vol. 325, no. 5947, pp. 1544–1549, Sep. 2009.

[71] M. W. Covert, E. M. Knight, J. L. Reed, M. J. Herrgard, and B. O. Palsson, “Integrating high-throughput and computational data elucidates bacterial networks.,” *Nature*, vol. 429, no. 6987, pp. 92–96, May 2004.

[72] J. A. Papin and B. O. Palsson, “The JAK-STAT signaling network in the human B-cell: an extreme signaling pathway analysis.,” *Biophys. J.*, vol. 87, no. 1, pp. 37–46, Jul. 2004.

[73] J. S. Edwards and B. O. Palsson, “Systems properties of the Haemophilus influenzae Rd metabolic genotype.,” *J. Biol. Chem.*, vol. 274, no. 25, pp. 17410–17416, Jun. 1999.

[74] C. Gu, G. B. Kim, W. J. Kim, H. U. Kim, and S. Y. Lee, “Current status and applications of genome-scale metabolic models,” *Genome Biol.*, vol. 20, no. 1, p. 121, 2019.

[75] J. Nogales, “A Practical Protocol for Genome-Scale Metabolic Reconstructions BT - Hydrocarbon and Lipid Microbiology Protocols: Genetic, Genomic and System Analyses of Pure Cultures,” T. J. McGenity, K. N. Timmis, and B. Nogales, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2017, pp. 197–221.

[76] I. Thiele and B. Ø. Palsson, “A protocol for generating a high-quality genome-scale metabolic reconstruction,” *Nat. Protoc.*, vol. 5, no. 1, pp. 93–121, 2010.

[77] C. Lieven *et al.*, “MEMOTE for standardized genome-scale metabolic model testing,” *Nat. Biotechnol.*, vol. 38, no. 3, pp. 272–276, 2020.

[78] S. N. Mendoza, B. G. Olivier, D. Molenaar, and B. Teusink, “A systematic assessment of current genome-scale metabolic reconstruction tools,” *Genome Biol.*, vol. 20, no. 1, p. 158, 2019.

[79] M. Hanemaaijer, B. G. Olivier, W. F. M. Röling, F. J. Bruggeman, and B. Teusink, “Model-based quantification of metabolic interactions from dynamic microbial-community data,” *PLoS One*, vol. 12, no. 3, p. e0173183, Mar. 2017.

[80] H. Wang *et al.*, “RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on Streptomyces coelicolor,” *PLOS Comput. Biol.*, vol. 14, no. 10, p. e1006541, Oct. 2018.

[81] D. Machado, S. Andrejev, M. Tramontano, and K. R. Patil, “Fast automated reconstruction of genome-scale metabolic models for microbial species and communities,” *Nucleic Acids Res.*, vol. 46, no. 15, pp. 7542–7553, Jun. 2018.

[82] E. Karlsen, C. Schulz, and E. Almaas, “Automated generation of genome-scale metabolic draft reconstructions based on KEGG,” *BMC Bioinformatics*, vol. 19, no. 1, p. 467, 2018.

[83] C. S. Henry, M. Dejongh, A. A. Best, P. M. Frybarger, B. Linsay, and R. L. Stevens, “High-throughput generation , optimization and analysis of genome-scale metabolic models,” *Nat. Biotechnol.*, vol. 28, no. 9, pp. 969–974, 2010.

[84] Z. A. King *et al.*, “BiGG Models: A platform for integrating, standardizing and sharing genome-scale models,” *Nucleic Acids Res.*, vol. 44, no. D1, pp. D515--D522, 2016.

[85] M. Kanehisa and S. Goto, “KEGG: kyoto encyclopedia of genes and genomes.,” *Nucleic Acids Res.*, vol. 28, no. 1, pp. 27–30, Jan. 2000.

[86] R. Caspi *et al.*, “The MetaCyc database of metabolic pathways and enzymes,” *Nucleic Acids Res.*, vol. 46, no. D1, pp. D633–D639, Oct. 2017.

[87] 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, no. D1, pp. D542–D549, Nov. 2018.

[88] A. Ebrahim, J. A. Lerman, B. O. Palsson, and D. R. Hyduke, “COBRApy: COnstraints-Based Reconstruction and Analysis for Python,” *BMC Syst. Biol.*, vol. 7, 2013.

[89] 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, 2019.

[90] N. Vlassis, M. P. Pacheco, and T. Sauter, “Fast Reconstruction of Compact Context-Specific Metabolic Network Models,” *PLOS Comput. Biol.*, vol. 10, no. 1, p. e1003424, Jan. 2014.

[91] D. A. Fell and J. R. Small, “Fat synthesis in adipose tissue. An examination of stoichiometric constraints,” *Biochem. J.*, vol. 238, pp. 781–786, 1986.

[92] J. M. Savinell and B. O. Palsson, “Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism,” *J. Theor. Biol.*, vol. 154, no. 4, pp. 421–454, 1992.

[93] N. E. Lewis *et al.*, “Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models,” *Mol. Syst. Biol.*, vol. 6, no. 390, 2010.

[94] A. M. Feist and B. Ø. Palsson, “The growing scope of applications of genome-scale metabolic reconstructions using Escherichia coli,” *Nat. Biotechnol.*, vol. 26, no. 6, pp. 659–667, Jun. 2008.

[95] E. J. O’Brien, J. M. Monk, and B. O. Palsson, “Using Genome-scale Models to Predict

Biological Capabilities.," *Cell*, vol. 161, no. 5, pp. 971–987, May 2015.

[96] P. Pharkya, A. P. Burgard, and C. D. Maranas, "OptStrain: a computational framework for redesign of microbial production systems.," *Genome Res.*, vol. 14, no. 11, pp. 2367–2376, Nov. 2004.

[97] A. Chowdhury and C. D. Maranas, "Designing overall stoichiometric conversions and intervening metabolic reactions," *Sci. Rep.*, vol. 5, pp. 1–20, 2015.

[98] C. Cotten and J. L. Reed, "Constraint-based strain design using continuous modifications (CosMos) of flux bounds finds new strategies for metabolic engineering.," *Biotechnol. J.*, vol. 8, no. 5, pp. 595–604, May 2013.

[99] A. Chowdhury, A. R. Zomorrodi, and C. D. Maranas, "k-OptForce: integrating kinetics with flux balance analysis for strain design.," *PLoS Comput. Biol.*, vol. 10, no. 2, p. e1003487, Feb. 2014.

[100] J. Yabuzaki, "Carotenoids Database: structures, chemical fingerprints and distribution among organisms," *Database (Oxford).*, vol. 2017, no. 1, p. bax004, Jan. 2017.

[101] W. R. Sistrom, M. Griffiths, and R. Y. Stanier, "The biology of a photosynthetic bacterium which lacks colored carotenoids," *J. Cell. Comp. Physiol.*, vol. 48, no. 3, pp. 473–515, Dec. 1956.

[102] A. Vershinin, "Biological functions of carotenoids - Diversity and evolution," *BioFactors*, vol. 10, no. 2–3, pp. 99–104, 1999.

[103] J. R. Mein, F. Lian, and X.-D. Wang, "Biological activity of lycopene metabolites: implications for cancer prevention," *Nutr. Rev.*, vol. 66, no. 12, pp. 667–683, Dec. 2008.

[104] R. Sathasivam and J.-S. Ki, "A Review of the Biological Activities of Microalgal Carotenoids and Their Potential Use in Healthcare and Cosmetic Industries," *Mar. Drugs*, vol. 16, no. 1, p. 26, Jan. 2018.

[105] A. J. Young and G. L. Lowe, "Carotenoids-Antioxidant Properties," *Antioxidants (Basel, Switzerland)*, vol. 7, no. 2, p. 28, Feb. 2018.

[106] Markets And Markets, "Carotenoids Market," 2020. [Online]. Available: https://www.marketsandmarkets.com/Market-Reports/carotenoid-market-158421566.html?gclid=CjwKCAiA-vLyBRBWEiwAzOkGVOx0oK25469b9LfMwRcRsTARZR8-HP8b4Jx6AzZzR615k466eqQBJxoCY0kQAvD_BwE.

[107] L. Novoveská, M. E. Ross, M. S. Stanley, R. Pradelles, V. Wasiolek, and J.-F. Sassi, "Microalgal Carotenoids: A Review of Production, Current Markets, Regulations, and Future Direction," *Mar. Drugs*, vol. 17, no. 11, p. 640, Nov. 2019.

[108] R. Ciriminna, A. Fidalgo, F. Meneguzzo, L. M. Ilharco, and M. Pagliaro, "Lycopene: Emerging Production Methods and Applications of a Valued Carotenoid," *ACS Sustain.*

Chem. Eng., vol. 4, no. 3, pp. 643–650, Mar. 2016.

[109] J. Li, D. Zhu, J. Niu, S. Shen, and G. Wang, "An economic assessment of astaxanthin production by large scale cultivation of Haematococcus pluvialis," *Biotechnol. Adv.*, vol. 29, no. 6, pp. 568–574, 2011.

[110] R. M. Schweiggert and R. Carle, "Carotenoid Production by Bacteria, Microalgae, and Fungi," in *Carotenoids Nutrition, Analysis and Technology*, A. Kaczor and M. Baranska, Eds. Wiley-Blackwell, 2016.

[111] H. Katsuki and K. Bloch, "Studies on the biosynthesis of ergosterol in yeast. Formation of methylated intermediates.," *J. Biol. Chem.*, vol. 242, no. 2, pp. 222–227, Jan. 1967.

[112] F. Lynen, "Biosynthetic pathways from acetate to natural products.," *Pure Appl. Chem.*, vol. 14, no. 1, pp. 137–167, 1967.

[113] M. Rohmer, M. Knani, P. Simonin, B. Sutter, and H. Sahm, "Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate," *Biochem. J.*, vol. 295 (Pt 2, no. Pt 2, pp. 517–524, Oct. 1993.

[114] J. Paniagua-Michel, J. Olmos-Soto, and M. A. Ruiz, "Pathways of Carotenoid Biosynthesis in Bacteria and Microalgae," in *Microbial Carotenoids from Bacteria and Microalgae. Methods in Molecular Biology (Methods and Protocols)*, J.-L. Barredo, Ed. Totowa, NJ: Humana Press, 2012, pp. 1–12.

[115] A. Yokoyama, G. Sandmann, T. Hoshino, K. Adachi, M. Sakai, and Y. Shizuri, "Thermozeaxanthins, new carotenoid-glycoside-esters from thermophilic eubacterium thermus thermophilus," *Tetrahedron Lett.*, vol. 36, no. 27, pp. 4901–4904, 1995.

[116] F. Santos, "Vitamin B12 synthesis in Lactobacillus reuteri," Wageningen University, 2008.

[117] Z. A. King, A. Dräger, A. Ebrahim, N. Sonnenschein, N. E. Lewis, and B. O. Palsson, "Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways," *PLoS Comput. Biol.*, vol. 11, no. 8, pp. 1–13, 2015.

[118] X. T. Liu, C. L. Hou, J. Zhang, X. F. Zeng, and S. Y. Qiao, "Fermentation conditions influence the fatty acid composition of the membranes of Lactobacillus reuteri I5007 and its survival following freeze-drying," *Lett. Appl. Microbiol.*, vol. 59, pp. 398–403, 2014.

[119] J. Walter *et al.*, "D-Alanyl ester depletion of teichoic acids in Lactobacillus reuteri 100-23 results in impaired colonization of the mouse gastrointestinal tract," *Environ. Microbiol.*, vol. 9, no. 7, pp. 1750–1760, 2007.

[120] P. Ksonzeková *et al.*, "Exopolysaccharides of Lactobacillus reuteri : Their influence on adherence of E . coli to epithelial cells and inflammatory response," *Carbohydr. Polym.*, vol. 141, pp. 10–19, 2016.

[121] B. Teusink *et al.*, “Analysis of Growth of Lactobacillus plantarum WCFS1 on a Complex Medium Using a Genome-scale Metabolic Model,” *J. Biol. Chem.*, vol. 281, no. 52, pp. 40041–40048, 2006.

[122] A. P. Oliveira, J. Nielsen, and J. Förster, “Modeling Lactococcus lactis using a genome-scale flux model,” *BMC Microbiol.*, vol. 5, no. 1, 2005.

[123] B. Christensen, P. B. Olsen, T. B. Regueira, B. Koebmann, S. T. Joergensen, and T. I. Dehli, “International Publication Number WO 2014/102180 AI,” no. 12. 2014.

[124] F. Zhu *et al.*, “Targeted engineering and scale up of lycopene overproduction in Escherichia coli,” *Process Biochem.*, vol. 50, no. 3, pp. 341–346, 2015.

[125] Y. Chen, W. Xiao, Y. Wang, H. Liu, X. Li, and Y. Yuan, “Lycopene overproduction in Saccharomyces cerevisiae through combining pathway engineering with host engineering,” *Microb. Cell Fact.*, vol. 15, no. 1, p. 113, Jun. 2016.

[126] E. Y. C. Ron *et al.*, “Characterization of carotenoids in Rhodothermus marinus,” *Microbiologyopen*, vol. 7, no. 1, 2018.

[127] Z. Sun *et al.*, “A novel carotenoid 1,2-hydratase (CruF) from two species of the non-photosynthetic bacterium Deinococcus,” *Microbiology*, vol. 155, no. 8, pp. 2775–2783, 2009.

[128] A. Hausmann and G. Sandmann, “A Single Five-Step Desaturase Is Involved in the Carotenoid Biosynthesis Pathway to β-Carotene and Torulene in Neurospora crassa,” *Fungal Genet. Biol.*, vol. 30, no. 2, pp. 147–153, 2000.

[129] J. D. Orth *et al.*, “A comprehensive genome-scale reconstruction of Escherichia coli metabolism-2011,” *Mol. Syst. Biol.*, vol. 7, no. 535, pp. 1–9, 2011.

[130] A. Heinken, S. Sahoo, R. M. T. Fleming, and I. Thiele, “Systems-level characterization of a host-microbe metabolic symbiosis in the mammalian gut,” *Gut Microbes*, vol. 4, no. 1, pp. 28–40, 2013.

[131] L. Jolly, P. Ferrari, D. Blanot, J. Van Heijenoort, F. Fassy, and D. Mengin-Lecreulx, “Reaction mechanism of phosphoglucosamine mutase from Escherichia coli,” *Eur. J. Biochem.*, vol. 262, no. 1, pp. 202–210, 1999.

[132] S. Sauret-Güeto, E. M. Urós, E. Ibáñez, A. Boronat, and M. Rodríguez-Concepción, “A mutant pyruvate dehydrogenase E1 subunit allows survival of Escherichia coli strains defective in 1-deoxy-d-xylulose 5-phosphate synthase,” *FEBS Lett.*, vol. 580, no. 3, pp. 736–740, 2006.

[133] T. J. Erb *et al.*, “A RubisCO like protein links SAM metabolism with isoprenoid biosynthesis,” *Nat Chem Biol.*, vol. 8, no. 11, pp. 926–932, 2013.

[134] M. C. Pérez-Marín, S. Padmanabhan, M. C. Polanco, F. J. Murillo, and M. Elías-Arnanz, “Vitamin B12 partners the CarH repressor to downregulate a photoinducible promoter in Myxococcus xanthus,” *Mol. Microbiol.*, vol. 67, no. 4, pp. 804–819, 2008.

[135] H. Takano *et al.*, “Involvement of CarA/LitR and CRP/FNR family transcriptional regulators in light-induced carotenoid production in Thermus thermophilus,” *J. Bacteriol.*, vol. 193, no. 10, pp. 2451–2459, 2011.

[136] H. Takano, S. Obitsu, T. Beppu, and K. Ueda, “Light-induced carotenogenesis in Streptomyces coelicolor A3(2): identification of an extracytoplasmic function sigma factor that directs photodependent transcription of the carotenoid biosynthesis gene cluster,” *J. Bacteriol.*, vol. 187, no. 5, pp. 1825–1832, Mar. 2005.

[137] H. Takano *et al.*, “Role and function of LitR, an adenosyl B12-bound light-sensitive regulator of Bacillus megaterium QM B1551, in regulation of carotenoid production,” *J. Bacteriol.*, vol. 197, no. 14, pp. 2301–2315, 2015.

[138] L. Moreira, M. F. Nobre, I. Sa-correia, and M. S. Da Costa, “Genomic Typing and Fatty Acid Composition of Rhodothermus marinus,” *Syst. Appl. Microbiol.*, vol. 19, no. 1, pp. 83–90, 1996.

[139] O. C. Nunes, M. M. Donato, C. M. Manaia, and M. S. Da Costa, “The Polar Lipid and Fatty Acid Composition of Rhodothermus Strains,” *Syst. Appl. Microbiol.*, vol. 15, no. 1, pp. 59–62, 1992.

[140] A. M. Feist *et al.*, “A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information.,” *Mol. Syst. Biol.*, vol. 3, p. 121, 2007.

[141] A. K. Wessel *et al.*, “Oxygen limitation within a bacterial aggregate,” *MBio*, vol. 5, no. 2, pp. e00992–e00992, Apr. 2014.

[142] U. U. Nwodo, E. Green, and A. I. Okoh, “Bacterial exopolysaccharides: functionality and prospects,” *Int. J. Mol. Sci.*, vol. 13, no. 11, pp. 14002–14015, Oct. 2012.

PAPERS

Paper 1

A metabolic reconstruction of *Lactobacillus reuteri* JCM 1112 and analysis of its potential as a cell factory

Thordis Kristjansdottir(1,2,^), Elleke F. Bosma(3,^,#), Filipe Branco dos Santos(4), Emre Özdemir(3), Markus J. Herrgård(3), Lucas França(5), Bruno Sommer Ferreira(5), Alex T. Nielsen(3), Steinn Gudmundsson(1,*)

(1) Center for Systems Biology, School of Engineering and Natural Sciences, University of Iceland, Dunhagi 5, 107 Reykjavik, Iceland

(2) Matis, Vinlandsleid 12, 113 Reykjavik, Iceland

(3) The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Building 220, Kemitorvet, 2800 Kgs. Lyngby, Denmark

(4) Molecular Microbial Physiology Group of the Swammerdam Institute for Life Sciences, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, The Netherlands

(5) Biotrend SA – Biocant Park, Núcleo 04, Lote 2, 3060-197 Cantanhede, Portugal

(#) Present address: Chr. Hansen A/S, Bøge Allé 10-12, 2970 Hørsholm, Denmark

(*) Corresponding author, steinng@hi.is

^ These authors contributed equally to this work.

Abstract

Background

Lactobacillus reuteri is a heterofermentative Lactic Acid Bacterium (LAB) that is commonly used for food fermentations and probiotic purposes. Due to its robust properties, it is also increasingly considered for use as a cell factory. It produces several industrially important compounds such as 1,3-propanediol and reuterin natively, but for cell factory purposes, developing improved strategies for engineering and fermentation optimization is crucial. Genome-scale metabolic models can be highly beneficial in guiding rational metabolic engineering. Reconstructing a reliable and a quantitatively accurate metabolic model requires extensive manual curation and incorporation of experimental data.

Results

A genome-scale metabolic model of *L. reuteri* JCM 1112^T was reconstructed and the resulting model, Lreuteri_530, was validated and tested with experimental data. Several knowledge gaps in the metabolism were identified and resolved during this process, including presence/absence of glycolytic genes. Flux distribution between the two glycolytic pathways, the phosphoketolase and Embden-Meyerhof-Parnas pathways, varies considerably between LAB species and strains. As these pathways result in different energy yields, it is important to include strain-specific utilization of these pathways in the model. We determined experimentally that the Embden-Meyerhof-Parnas pathway carried at most 7% of the total glycolytic flux. Predicted growth rates from Lreuteri_530 were in good agreement with experimentally determined values. To further validate the prediction accuracy of Lreuteri_530, the predicted effects of glycerol addition and *adhE* gene knock-out, which results in impaired ethanol production, were compared to *in vivo* data. Examination of both growth rates and uptake- and secretion rates of the main metabolites in central metabolism demonstrated that the model was able to accurately predict the experimentally observed effects. Lastly, the potential of *L. reuteri* as a cell factory was investigated, resulting in a number of general metabolic engineering strategies.

Conclusion

We have constructed a manually curated genome-scale metabolic model of *L. reuteri* JCM 1112^T that has been experimentally parameterized and validated and can accurately predict metabolic behavior of this important platform cell factory.

Keywords: *Lactobacillus reuteri*, Genome-scale metabolic model, Cell factory.

1. Background

Lactobacillus reuteri is a heterofermentative Lactic Acid Bacterium (LAB) that is present in the human gut and is an important probiotic organism [1]. There is an increasing interest in using it as a cell factory for the production of green chemicals and fuels in a biorefinery [2], [3], due to its robustness properties. It has high growth and glycolytic rates, without the requirement for either aeration or strictly anaerobic conditions. It is tolerant to low pH, ethanol and salt, and has a wide growth temperature range. Moreover, it is genetically accessible, enabling metabolic engineering for cell factory optimization [4]. The species is known to produce 1,3-propanediol, reuterin, and other related industrially important compounds in high yields from glycerol [2], of which reuterin has also since long been known as antimicrobial [5]. *L. reuteri* also has most of the genes encoding for the enzymes needed for biosynthesis of 1,2-propanediol and 1-propanol, both of which are industrially relevant chemicals. These compounds are, however, not produced under normal conditions by *L. reuteri*, requiring improved engineering- and optimization strategies to achieve commercial level cell factories and production processes [6].

Genome-scale metabolic models are highly useful for directing metabolic engineering strategies, as well as to improve understanding of the physiology and metabolism of the target organism [1], [7]. So far, highly curated and experimentally validated metabolic models have been primarily developed for model organisms such as *Escherichia coli* and *Saccharomyces cerevisiae,* but models for several LAB species are also available, including *Lactobacillus plantarum* [8], *Lactobacillus casei* [9], *Lactococcus lactis* [10], [11] and *Streptococcus thermophilus* [12] (Table 1). These LAB are homofermentative or facultatively heterofermentative organisms and have substantial differences in metabolism compared to strict heterofermenters such as *L. reuteri* [4]. Metabolic models for the heterofermenter *Leuconostoc mesenteroides* [13], [14] are available (Table 1), but this is only distantly related to *L. reuteri* [4] and shows different metabolic features such as malolactic fermentation and a limited ability to use amino acids as energy source [13]. Models for two probiotic strains of *L. reuteri* have been previously published [1] (Table 1). They were automatically reconstructed from the same draft model we started with here [15]. The two previously published *L. reuteri* models were used along with transcriptomics data to identify qualitative metabolic differences between the two strains as well as to analyze their probiotic properties [1]. However, these previous models were not manually curated and were not

used to quantitatively predict metabolic behavior. The construction of a genome-scale metabolic model that can be reliably used in basic research and cell factory design is a time-consuming process, requiring significant amount of manual curation and availability of strain-specific phenotypic data. At present, models obtained using automated tools or models that do not include experimental data are generally of limited use for quantitative predictions.

Here, we set out to reconstruct the metabolic network of *L. reuteri* JCM 1112, specifically for use in metabolic engineering applications, which requires collection of phenotypic data under several different conditions. We first performed an in-depth analysis of the genome to evaluate conflicting reports about metabolic pathways compared to strain DSM 20016. We then performed experiments to collect phenotypic data for the wild-type strain as well as for an alcohol dehydrogenase (*adhE)* knockout strain to constrain, validate, and test the model. Lastly, we use the model to test predictions for metabolic engineering strategies. The model as well as the experimental data are available in supplementary files.

Table 1. *LAB species with available genome-scale metabolic models.*

Species	**Hetero-/ homo-fermenter**	**Main applications & distinguishing characteristics**	**Genome size (Mb)**	**References**
Lactobacillus reuteri	Hetero-	Produces Vitamin B12, 1,3-PDO, 3-HPA; used as probiotic and potential cell factory	2.0	[1], (This study)
Leuconostoc mesenteroides	Hetero-	Used in food fermentations (many non-dairy); malolactic fermentation; aroma production in foods	2.0	[13], [14]
Lactobacillus plantarum	Facultatively hetero-	Used in food fermentations, probiotics and potential cell factory	3.3	[8]
Lactobacillus casei	Facultatively hetero-	Used in food fermentations, probiotics and potential cell factory	2.9	[9]
Lactococcus lactis	Homo-	Used in dairy fermentations and potential cell factory	2.4	[10], [11]
Streptococcus thermophilus	Homo-	Used in dairy fermentations; fewer amino acid auxotrophies than other LAB, missing PPP genes	1.9	[12]

2. Materials and methods

2.1 Strains, media and culture conditions

Strains used in this study are listed in Table 2 and an overview of the experimental datasets in Table 3. All experiments were performed in triplicate except the one used for determining biomass composition and energy requirements as well as dataset B (Table 3). Apart from the growth mode, the dataset used for determining biomass and energy components and dataset A are identical, and the resulting data is in good agreement (Additional file 1). Dataset B was included, as it was available from the previous work that this paper builds upon [15], and is in good agreement with dataset D (Additional file 1).

De Mann Rosa Sharp (MRS) medium (incl. 20 g/L glucose) was obtained from VWR and prepared according to the manufacturer's instructions.

Chemically Defined Medium (CDM) was used as described in [15] / [16] with the following modifications: arginine 5 g/L, tween-80 1 mL/L. Substrates were 111 mM glucose and 20 mM glycerol as indicated. The CDM was filter-sterilized and the final pH after mixing all components was 5.6.

All flask cultivations were performed in a stationary incubator at 37°C. A 5 mm inoculation loop of culture was inoculated from -80°C glycerol stocks into 1 mL MRS with or without glycerol in a 1.5 mL Eppendorf tube and grown overnight (16h). Next morning, cultures were washed 3x with sterile 0.9% NaCl, after which OD_{600} was measured and cells were transferred to 12 mL CDM with or without glycerol in a 15 mL Falcon tube to a starting OD_{600} of 0.08. After 4h of growth, OD_{600} was measured and cultures were transferred to a starting OD_{600} of 0.05 in 100 mL pre-warmed CDM with or without glycerol in a 100 mL Schott flask. Samples for OD_{600} measurement and HPLC were taken directly after inoculation (t=0h) and at 2, 3, 4, 5, and 6h; cultures were swirled for mixing prior to taking samples. The 6h samples were also used for protein and amino acid determinations. The time points used were all during exponential growth, ensuring a pseudo steady state (Additional file 1).

All bioreactor cultivations were performed in batch mode and samples were taken during exponential/pseudo-steady state (Additional file 1). One of the fermentations was performed in CDM at 37°C in 3.0 L bioreactors (BioFlo 115, New Brunswick Scientific/Eppendorf) with a 2.2 L working volume, 50 rpm agitation without gas sparging.

The pH was controlled at 5.7±0.1 using 5N NaOH. Pre-cultures were performed similarly as for the flask cultures described above, with the pre-culture in CDM in 100 mL medium in 100 mL flasks, and reactors inoculated to an OD_{600} of 0.1. The other two reactor cultivations were performed in CDM, with and without glycerol, at 37°C in 0.4 L reactors with a 0.5 L working volume, 50 rpm agitation and sparged with N_2 at 15 mL/min for 1h prior to inoculation. The pH was controlled at 5.8 using 5M NaOH. Fermenters were inoculated to an initial OD_{600} of 0.05 from an exponentially growing culture on CDM without glycerol. As can be seen in Additional file 1, there is no difference between the cultures in the reactors that were sparged with N2 prior to fermentations and those that were not and hence we decided to treat these as replicates.

The correlation factor between cell dry weight (gDW) and OD_{600} was experimentally determined to be 0.4007 gDW/OD_{600} in CDM and used for calculating gDW from OD_{600} in all experiments.

Table 2. *Lactobacillus reuteri strains used in this study.*

Strain name	Description/genotype	Origin/reference
JCM 1112 (DSM 20016, 'WT')	Wild-type	DSMZ[1]
SJ11774 ('SJ (WT*)')	Strain JCM 1112 (DSM 20016) with two inactivated restriction-modification systems (ΔLAR_RS04635 ΔLAR_RS07680::cat)	Novozymes; patent WO2014102180 A1
SJΔadhE	Strain SJ11774 with a clean and full in-frame deletion of the bifunctional aldehyde/alcohol dehydrogenase adhE (LAR_RS01690)	Unpublished (manuscript in preparation)

[1] DSMZ = Deutsche Sammlung von Mikroorganismen und Zellkulturen.

Table 3. *Experimental datasets used for the model reconstruction.*

Strain	Substrate	Growth mode	Dataset name (Figure 4+5)	Used for:
WT	Glucose	Flask	-	Determining biomass composition and energy requirements (section 2.4 and 3.1.2)
WT	Glucose	Reactor	A	Model validation
WT	Glucose + glycerol	Reactor	B	Model validation
SJ (WT*)	Glucose	Flask	C	Model validation and model predictions
SJ (WT*)	Glucose + glycerol	Flask	D	Model validation and model predictions
SJΔadhE	Glucose	Flask	E	Model validation and model predictions
SJΔadhE	Glucose + glycerol	Flask	F	Model validation and model predictions

Growth curves and uptake and secretion data for all datasets can be found in Additional file 1.

2.2 Analytical methods

Protein concentration of the cells was determined in the 6h samples as described above, via a BCA protein assay (Merck-Millipore cat. 71285) according to the manufacturer's protocol. Prior to the BCA assay, cell pellets were washed once in 0.9% NaCl and resuspended in 0.25 mM Tris-HCl pH 7.5 and sonicated on ice with an Ultrasonic Homogenizer 300VT (BioLogics) for 3x 30s at 40% power, with 30s breaks on ice.

Amino acid composition of the cells was determined by Ansynth BV (The Netherlands) on washed cell pellets of a 6h CDM culture as described above.

Substrates, products and amino acids secreted and taken up during the cultivations were quantified using HPLC. Glucose, glycerol, ethanol, lactate, acetate, citrate, 1,2-propanediol, 1,3-propanediol, 1-propanol, 2-propanol, pyruvate, succinate and malate were quantified with either one of two HPLCs: 1) a Dionex Ultimate 3000 (Thermo Scientific) containing an LPG-3400SD pump, a WPS-3000 autosampler, a UV-visible (UV-Vis) DAD-3000 detector, and an RI-101 refraction index detector. Injection volume was 20 µL. An Aminex HPx87 ion exclusion 125-0140 column was used with a mobile phase of 5 mM H_2SO_4, a flow rate of 0.6 mL/min and an oven temperature of 60°C; 2) a Shimadzu LC-20AD equipped with refractive

index and UV (210 nm) detectors, with an injection volume of 20 µL. A Shodex SH1011 8.0mmIDx300mm column was used with a mobile phase of 5 mM H_2SO_4, a flow rate of 0.6 mL/min and an oven temperature of 50°C. All amino acids, ornithine and GABA were quantified using a Dionex Ultimate 3000 (Thermo Scientific), for which the procedure is as follows: 20 µg/mL 2-aminobutanoic acid and sarcosine were used as internal standards for dilution of the samples; derivatization was performed in the autosampler. 0.5 µL sample was added into 2.5 µL of (v/v) 3-mercaptopropionic acid in borate buffer (0.4 M, pH 10.2), mixed and incubated for 20 s at 4°C to reduce free cystines. Then 1 µL of 120 mM iodoacetic acid in 140 mM NaOH was added, mixed and incubated for 20 s at 4°C to alkylate reduced cysteines. 1.5 µL of OPA reagent (10 mg o-pthalaldehyde/mL in 3-mercaptopropionic acid) was then added to derivatize primary amino acids. The reaction was mixed and incubated for 20s at 4°C. 1 µL of FMOC reagent (2.5 mg 9-fluorenylmethyl chloroformate/mL in acetonitrile) was added, mixed and incubated for 20 s at 4°C to derivatize other amino acids. 50 µL of Buffer A (Buffer A: 40 mM Na_2HPO_4, 0.02% NaN_3 (w/v) at pH 7.8) at pH 7 was added to lower the pH of the reaction prior to injecting the 56.5 µL reaction onto a Gemini C18 column (3 um, 4.6 x 150 mm, Phenomenex PN: 00F-4439-E0) with a guard column (SecurityGuard Gemini C18, Phenomenex PN: AJO-7597). The column temperature was kept at 37°C in a thermostatic column compartment. The mobile phase had the following composition: Buffer A: see above, pH 7.8; Buffer B: 45% (v/v) acetonitrile, 45% (v/v) methanol and 10% (v/v) water; flow rate 1 mL/min. Derivatized amino acids were monitored using a fluorescence detector. OPA-derivatized amino acids were detected at 340_{ex} and 450_{em} nm and FMOC-derivatized amino acids at 266_{ex} and 305_{em} nm. Quantifications were based on standard curves derived from dilutions of a mixed amino acid standard (250 µg/mL). The upper and lower limits of quantification were 100 and 0.5 µg/mL, respectively.

2.3 Genome sequencing and analysis

For genomic DNA (gDNA) isolation, overnight cultures of DSM 20016 and SJ 11774 were grown in MRS and the pellet was used for gDNA isolation using the Epicentre MasterPure™ Gram Positive DNA Purification kit according to the manufacturer's protocol. Subsequent genome sequencing was performed at the sequencing facility at the NNF Center for Biosustainability. Library preparation was performed using KAPA HyperPlus Library Prep Kit (ROCHE) with Illumina-compatible dual-indexed PentAdapters (PentaBase). The average size of the library pool was 317 bp. Sequencing was performed on MiSeq (Illumina) using the

MiSeq Reagent Kit v2, 300 Cycles (Illumina). The libraries were loaded to the flow cell at 10 pM and sequenced using paired-end reads of 150 bp. Read quality check was performed with FastQC version 0.11.5. Mutations relative to reference (*L. reuteri* JCM 1112, GenBank accession nr AP007281, annotated with Prokka version 1.11) were identified using Breseq (version 0.31.0) [17]. Mean coverage was 143.7x (SJ 11774) and 129.5x (DSM 20016). All runs were performed at Danish national supercomputer for life sciences (Computerome), Technical University of Denmark. For this work, the annotated genome of *L. reuteri* JCM 1112 from NCBI was used. During the reconstruction, several genes were re-annotated, based on BLAST and physiological data. A list of all genes in the JCM 1112 genome can be found in Additional fie 2, along with annotations from the GenBank file and which model reactions are associated with each gene.

2.4 Metabolic reconstruction

The *L. reuteri* JCM 1112 metabolic reconstruction was based on an unpublished, automatically generated draft reconstruction of JCM 1112 [15]. We performed extensive manual curation, including: gap filling, updating and adding gene-protein-reaction (GPR) associations, updating gene IDs, updating metabolite- and reaction abbreviations, in line with the BiGG database [18], updating and adding missing formulas and/or charges to metabolites, fixing unbalanced reactions, adding annotation to metabolites, reactions and genes and detailed review and integration of organism specific data. A biomass objective function was formulated based on available data on *L. reuteri* and related strains. The ATP cost of growth-associated maintenance (GAM) was estimated using one of the data sets (Table 3) by adjusting the GAM parameter so that growth predictions matched *in vivo* growth. This data set was then excluded from subsequent validation and prediction steps.

2.5 Flux balance analysis

Flux balance analysis (FBA) was used to analyze the genome-scale metabolic model [19], [20] by constraining exchange reactions in the model with experimental values of substrate uptake and secretion rates. To take into account that the Embden–Meyerhof–Parnas (EMPP) pathway is a minor glycolytic pathway in *L. reuteri* compared to the phosphoketolase pathway (PKP) (section 3.1.1), an additional flux constraint was added to the model

$$\frac{v_{PFK}}{v_{PFK} + v_{G6PDH2r}} \leq r,$$

where *r* is an empirically determined flux ratio, $\mathbf{v}_{PFK}$ denotes flux in the rate limiting step of the EMPP and $\mathbf{v}_{G6PDH2r}$ is the flux in the first reaction branching into the PKP.

We used a variant of FBA called parsimonious FBA [21] which identifies flux values corresponding to maximum growth with the side constraint that the sum of absolute flux values is made as small as possible. The sum of fluxes is proxy for enzyme usage and the method can therefore be considered to simulate biological pressure for rapid and efficient growth using minimum amount of resources (enzymes). An advantage over FBA is that the resulting solution is likely to contain fewer infeasible flux cycles. Model simulations were carried out in Python with the CobraPy toolbox [22] and GLPK solver. All code used in the simulations is provided in the form of a Jupyter notebook in Additional file 3 and on https://github.com/steinng/reuteri. The Escher package [23] was used for visualization of flux predictions. Escher maps of *L. reuteri*´s central metabolism are provided in Additional file 4, both simplified maps as shown in sections 3.2.2 and 3.2.3 as well as a detailed map linking different sugar utilization pathways to the central metabolism.

To predict growth rates the model was constrained with uptake rates of glucose, glycerol and five amino acids (Arg, Ser, Asn, Asp and Glu), and with the secretion rates of ethanol, lactate, acetate and 1,3-propanediol. Effects of knocking out the *adhE* gene were predicted by temporarily deleting it from the network. Where the effects of an active 1,2-propanediol pathway were predicted, a methylglyoxal synthase (MGS) was added to the model and optimized for growth.

To predict the theoretical maximum yields of selected target compounds, a reaction enabling the secretion of the corresponding metabolite was added to the model, unless an exchange reaction already existed, and flux through the reaction maximized. The glucose uptake rate was 25.2 mmol gDW^{-1} h^{-1}, based on experimental data, and free secretion of by-products was allowed. For the production of L-alanine, an L-alanine dehydrogenase was added to the model. The production of ethyl lactate required the addition of a lactate acyl transferase and a reaction for the condensation of lactoyl-CoA with ethanol [24]. To produce 1-propanol, a methylglyoxal synthase (MGS) was added to the model. The presence of a complete 1-propanol pathway enables more efficient regeneration of NAD and the flux

predictions were therefore repeated in the presence of an active MGS. To simulate a non-limiting phosphofructokinase, the flux constraint involving $\mathbf{v}_{PFK}$ above was omitted.

3. Results and discussion

3.1 Metabolic network reconstruction

To reconstruct a genome-scale metabolic model of *L. reuteri* suitable for use in cell factory design and optimization, we built upon a draft metabolic model of *L. reuteri* JCM 1112 described in [15] that we in turn extensively curated. The Memote tool [25] was used to assess the quality of the reconstruction and to guide the curation process (Additional file 5). The main characteristics of the resulting Lreuteri_530 model (Additional file 6) are listed in Table 4.

Table 4. *Main characteristics of Lreuteri_530 - the L. reuteri JCM 1112 genome-scale metabolic reconstruction.*

Genome characteristics	
Genome size	*2.04 Mb*
Total protein coding sequences	*1943*
Model characteristics	
Genes	*530*
Percentage of genome	*27%*
Reactions (with GPR)	*710 (690)*
Metabolites (unique)	*658 (551)*
Memote total score	*62%*

3.1.1 Curation process

Reactions and metabolites were abbreviated according to the BiGG database nomenclature where applicable and annotations with links to external databases included. Genes from the JCM 1112 genome were identified with locus tags from the GenBank file, and annotations were included which contain: the old locus tag which is often found in older literature, the NCBI protein ID, gene annotation and the protein sequence. Apart from general network curation, organism-specific information obtained from laboratory experiments and from

available literature was integrated by reviewing reactions, genes and gene-protein-reaction (GPR) rules.

Resequencing reveals inconsistencies between the "same" strains L. reuteri DSM 20016 and JCM 1112 - implications for glycolytic genes

The two most well-known strain names and origins for the type strain are DSM 20016 and JCM 1112 from the DSMZ and JCM culture collections, respectively. These two are derived from the same original human faeces isolate *L. reuteri* F275 [26], which was grown and stocked in two different laboratories [27]. Both genomes have been sequenced previously and a comparison showed that they are identical except for two regions that were missing in DSM 20016 [28], which were most likely lost during the 20 years of separate laboratory cultivation [27]. The first region (8,435 bp, flanked by IS4 insertion sequences on each end) contains genes for glycolysis, namely glyceraldehyde-3-P dehydrogenase, phosphoglycerate kinase, triosephosphate isomerase, and enolase. The second region (30,237 bp, flanked by two different insertion sequence elements) contains for example a gene cluster for nitrate reductases and molybdopterin biosynthesis [28]. As the first island consists of glycolytic genes, the implications of its presence or absence are profound. This island is absent in DSM 20016, but we could identify homologs of all this island's genes except glyceraldehyde-3-P dehydrogenase elsewhere in its genome based on annotation and/or BLAST.

During the preparation of our model, it became clear that there are inconsistencies in naming and hence gene content of the *L. reuteri* type strain. We sequenced the DSM 20016 strain that we obtained from DSMZ and this showed that its genome is identical to that of JCM 1112 instead, meaning it contained the two islands missing in DSM 20016. A similar result of these strains being 'swapped' was obtained by others based on whole genome sequencing [29] and PCR of part of the largest missing region in DSM 20016 in a study looking at cell-surface proteins in the different strains [30]. This inconsistency between the two strains does not seem to be commonly known and taken into account, and we suspect that some papers referring to either the DSM or the JCM strain might in fact be working with the other strain. For example, the DSM 20016 strain used by Sun et al., sequenced in 2015 (accession nr AZDD00000000), contains the islands as indicated by the presence of all glycolytic genes and hence is actually the JCM 1112 strain [31]. Contrarily, the DSM 20016

referred to by Morita et al., sequenced in 2007 by JGI (accession nr CP000705), was shown to be DSM 20016, missing the islands [28]. Both strains were obtained from DSMZ. This highlights the importance of re-sequencing of strains ordered from culture collections or lab strains present in the laboratory before using them for engineering or characterization studies. We strongly suggest that studies working with any *L. reuteri* type strain perform PCR on the two islands or perform resequencing to validate the presence or absence of the genes in the two islands.

Based on our sequencing results, we have included all genes in the two islands in our metabolic reconstruction and model. The genes in the model are identified with the locus tags obtained from the JCM 1112 strain's genome in NCBI (NC_010609). As many other publications refer to genes in the DSM 20016 strain or use the old locus tags from the JCM 1112 genome, we have included a table (Additional file 2) which lists: the locus tags used in the model (gene numbers prefixed by LAR_RS), the old locus tags (gene numbers prefixed by LAR_), the annotations obtained from the NCBI GenBank file, the NCBI protein IDs (WP numbers), the locus tags of the corresponding genes in the DSM 20016 strain, when applicable (gene numbers prefixed by Lreu_), and finally the reaction(s) in the metabolic model associated with the genes.

Phosphofructokinase (PFK) and the distribution between EMP and PK pathway usage

Obligately heterofermentative lactobacilli like L. reuteri are often considered to solely use the phosphoketolase pathway (PKP) instead of the Embden-Meyerhof-Parnas pathway (EMPP) for glucose consumption [4] (Figure 1). Both pathways result in the glycolytic intermediate glyceraldehyde-3-phosphate but use different redox cofactors (Figure 1). As the PKP yields one and the EMPP two molecules of glyceraldehyde-3-phosphate, the PKP has a lower energy yield than the EMPP (Figure 1). The PKP generally results in the production of one molecule of lactate and one molecule of ethanol or acetate for one glucose molecule while the EMPP generally yields two lactate molecules. Key enzymes of the EMPP are fructokinase (FK), glucose-6-phosphate isomerase (PGI), phosphofructokinase (PFK), fructose-bis-phosphate aldolase (FBA), and triosephosphate isomerase (TPI). In line with the idea that heterofermenters use the PKP, Sun et al. showed in a comparison of 213 LAB genomes that pfk was lacking from a distinct monophyletic group formed by mainly (87%)

obligatively and otherwise facultatively heterofermentative Lactobacillus spp., including L. reuteri DSM 20016 and L. panis DSM 6035 [31]. Contrary to most other species in the same group, these two species did contain fba, which has traditionally been linked to the presence of the EMPP. Despite the absence of pfk, EMPP activity has been observed in several L. reuteri strains and in some strains it appears to play a major role compared to the PKP, depending on the growth phase, and showing strain-specific differences [32], [33]. For modeling and engineering purposes, it is crucial to understand the presence and activity of the PKP vs the EMPP.

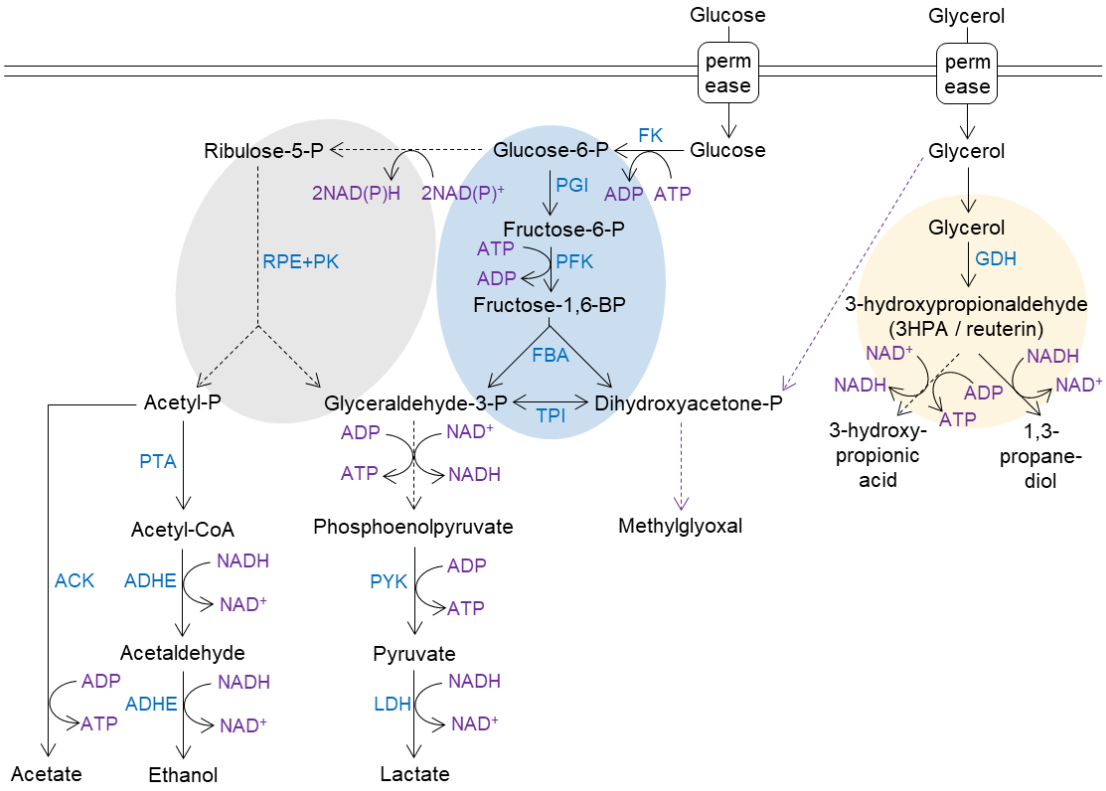


Figure 1. *Condensed overview of the central metabolism in L. reuteri. Dotted purple arrows indicate pathways for which genes or homologs are present but likely not active in L. reuteri JCM 1112. Dotted black arrows indicate multiple enzymatic steps. Yellow background circle indicates microcompartment; blue background indicates the EMP pathway; grey background indicates the phosphoketolase pathway. Abbreviations: FK: fructokinase/glucokinase; PGI: glucose-6-phosphate isomerase; PFK: phosphofructokinase; FBA: fructose-bis-phosphate aldolase; TPI: triosephosphate isomerase; PGM: phosphoglucomutase; SP: sucrose phosphorylase; M2DH: mannitol-2-dehydrogenase; RPE+PK: ribulose epimerase + phosphoketolase; GDH: glycerol dehydratase I. Adapted from* [4].

Årsköld et al. (2008) compared the genomic organization of 13 sequenced Lactobacillales and showed that L. reuteri (strains ATCC 55730 and DSM 20016) is one of the four exceptions that do not have a pfkA gene where this is located in all other species.

Nevertheless, they detect PFK and EMPP activity in strain ATCC 55730 and subsequently identify two genes (GenBank accession nrs EF547651 and EF547653) for orthologues of pfkB, a minor PFK-variant in E. coli [33]. In analogy with Årsköld et al. in L. reuteri, Kang et al. [34] identified a ribokinase in the obligately heterofermentative L. panis PM1 with 82% similarity to the pfkB gene identified in L. reuteri ATCC 55730 from Årsköld et al. (74% in our own BLAST search).

A BLAST comparison of the *pfkB* protein sequence of *L. panis* PM1 (GenBank accession nr AGU90228.1) and *L. reuteri* ATCC 55730 (GenBank accession nr ABQ23677.1) against *L. reuteri* JCM 1112 resulted in 81% and 99% identity, respectively, to JCM 1112 gene number LAR_RS02150, which is annotated as ribokinase rbsK_2. On a gene level, this gene shares 97% identity with *L. reuteri* ATCC 55730 and 73% with *L. panis* PM1. The same identities were found in *L. reuteri* DSM 20016 for gene LREU_RS02105 (previously Lreu_0404, GenBank protein KRK49592.1). A second gene annotated as "ribokinase rbsK_3" (locus tag LAR_RS06895) showed only limited query coverage and identity and hence rbsK_2 is the most likely homolog of *pfkB*. The growth experiments conducted in the present study with JCM 1112 are in line with the findings of Burgé et al. and indicate minor though detectable usage of the EMPP in this strain with a peak in the early growth stage (Figure 2), in which this rbsK_2 likely fulfills the role of *pfkB*. The average flux through the EMPP in all cultures was 7.0% (Figure 2) and was used to define the corresponding flux split ratio in the model (section 2.5).

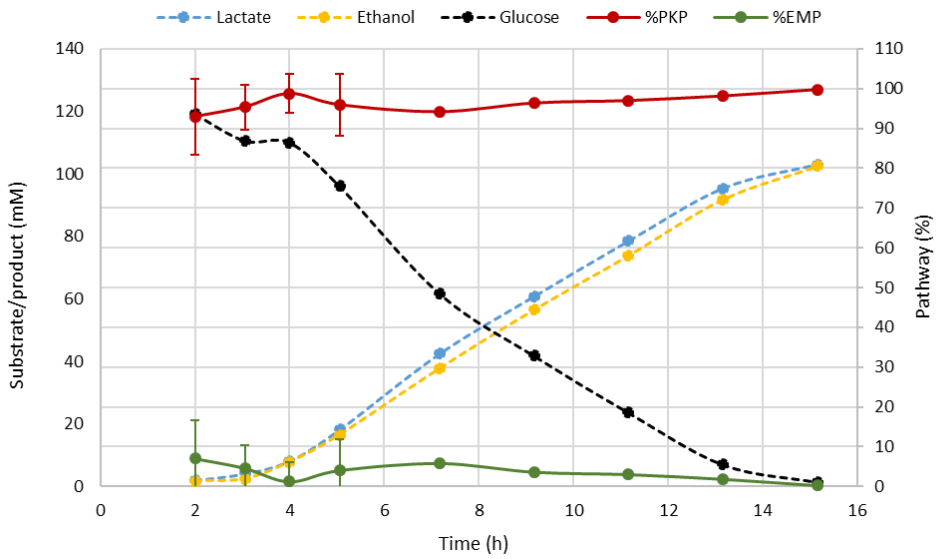


Figure 2. *Typical fermentation profile and distribution between the EMP and PK pathways in L. reuteri JCM 1112 in chemically defined medium with glucose as the sole carbon source. Data are averages of the all the datasets used to constrain and validate the model, with error bars representing standard deviation. The percentage of PKP usage was defined as in Burgé et al., i.e. as the ethanol concentration divided by the sum of lactate and ethanol concentrations divided by 2.*

Sugar transport

Transport of carbohydrates can be mediated by ATP-Binding Cassette (ABC) transporters, phosphotransferase systems (PTS), or secondary transporters (permeases of the Major Facilitator Superfamily, MFS) [35]. PTS systems mediate hexose mono- or dimer transport and phosphorylation simultaneously – mostly by using PEP to pyruvate conversion as phosphate donor, whereas ABC-transporters (mostly used for pentoses) and permeases (both pentoses and hexoses) perform only transport, and a separate ATP-utilizing kinase step is needed for sugar phosphorylation. Moreover, in Gram positives, PTS systems have an important role in carbon catabolite repression via phosphorylation cascades and direct interaction with the carbon catabolite repression protein A (ccpA) [36], [37]. Heterofermentative LAB contain fewer PTS system components than homofermentative LAB, which is thought to be the result of gene loss [38]. In general, organisms using the EMPP are believed to use PTS systems, and organism using the PKP to use secondary carriers [39]. Likely as a result of the lack of full PTS systems, glucose utilization is not constitutive but substrate-induced in heterofermenters, and utilization of several other sugars is not repressed by glucose [36]. Sugar transport in heterofermenters is poorly characterized, and only recently a study was dedicated to the genomic and phenotypic characterization of

carbohydrate transport and metabolism in *L. reuteri*, as representative of heterofermentative LAB [40]. This showed that *L. reuteri* completely lacks PTS systems and ABC-transporters and solely relies on secondary transporters of the MFS superfamily, which use the proton motive force (PMF) as energy source for transport [40]. In *L. reuteri* JCM 1112, we could identify the two common proteins of the PTS system, Enzyme I (Lreu_1324) and HPr (Lreu_1325). Some sugar-specific parts were present, but no complete PTS was identified. As a result, all sugar transport in the model takes place via the PMF.

Glycerol utilization

L. reuteri, like many lactobacilli, is known to be unable to grow on glycerol as a sole carbon source, but can use it as an alternative electron acceptor, providing a means to gain energy on a variety of carbon sources [41], [42]. *L. reuteri* is among the best native producers of large amounts of 3-hydroxypropionaldehyde (reuterin, 3-HPA) from glycerol that are currently known [43]. This is an intermediate in the pathway to 1,3-propanediol (1,3-PDO, also produced by *L. reuteri*, depending on the conditions used) that is known to be toxic and produced in a microcompartment [44]. The reason why it cannot grow on glycerol as sole carbon source is currently not fully clear, although it is likely related to gene regulation. All the genes that are necessary to convert glycerol to dihydroxyacetone phosphate via either dihydroxyacetone (DHA) or glycerol-3-phosphate and hence shuttle it into glycolysis are present in the *L. reuteri* genome [44]. However, several of these genes have been shown to be downregulated in the presence of glycerol [44], [45]. Furthermore, the *L. reuteri* glycerol dehydrogenase also has activity as 1,3-PDO:NAD-oxidoreductase, whereas in for example *Klebsiella pneumoniae*, which does produce glycolytic end products from glycerol, these are two different enzymes [41]. It seems that the physiological role of this enzyme in *L. reuteri* is the reduction of 3-HPA to 1,3-PDO, rather than glycerol to DHA conversion, explaining the lack of growth on glycerol [41].

Other pathways

Most heterofermentative LAB possess a malolactic enzyme but no malic enzymes [46], which is also the case for our *L. reuteri* strain, based on sequence comparisons with the *L. casei* strain used by Landete et al. [46]. Based on BLAST analysis and in line with literature, *L.*

reuteri JCM 1112 possesses a malate dehydrogenase and PEP carboxykinase, and cannot utilize citrate; malate (and fumarate) is converted to succinate [47].

From a biotechnological perspective, an interesting branch point of central carbon metabolism is the conversion from methylglyoxal (MG) to 1,2-propanediol (1,2-PDO), which can then be further metabolized into 1-propanol and propanoate. *L. reuteri* possesses all enzymes needed for these pathways, except methylglyoxal synthase (MGS), the step of the pathway, converting dihydroxyacetone phosphate into MG [42], [48]. It has been shown that when MG is added to *L. reuteri* JCM 1112 cultures or when a heterologous *mgs* is expressed, all the subsequent metabolites are formed [6]. Although we identified a potential distant homolog of *mgs* in the *L. reuteri* genome, this homolog is clearly not active under normal conditions since no 1,2-PDO was observed in our experiments. Hence, all the genes in these pathways except *mgs* were included in the reconstruction. For methylglyoxal reductase, *mgr*, we also identified several aldo/keto reductases as possible homologs, based BLAST comparison to genes identified in [48]. However, verification of these hypothetical activities would need extensive enzyme assays, and it is also likely that this reaction is performed by LAR_RS09730 (Glycerol dehydrogenase) [49], [50], this has been added to the reconstruction for the MGR reaction. Alternatively, MG might be converted directly to lactate by a glyoxalase [48].

Whereas many LAB are auxotrophic for vitamin B12, *L. reuteri* is a native producer. Vitamin B12 is important as a cofactor in, for example the 3-HPA pathway, but is also of relevance for biotechnological and medical/health applications (*e.g.* when produced by probiotic strains). The structure and biosynthetic genes have been studied in detail [45], [51]. The corresponding pathway is present in the reconstruction and is active during growth predictions.

3.1.2 Biomass reaction and energy requirements

A biomass objective function (BOF), which contains all necessary components for biomass biosynthesis, is commonly used to predict growth rate in metabolic models. Ideally, the BOF should be constructed based on organism-specific experimental data, mainly the fractional composition of the macromolecules (proteins, DNA, RNA, lipids, etc.) and their individual building blocks (amino acids, nucleotides, fatty acids, etc.), as well as the energy necessary

for their biosynthesis [52]. The protein fraction is a significant fraction of the biomass and was therefore measured experimentally. The ratio of individual amino acids in the *L. reuteri* biomass was also measured experimentally. The remaining macromolecular fractions were derived from *L. plantarum* [8] and *L. lactis* [10]. Nucleotide composition was estimated from the genome, which in the case of RNA is not ideal since it assumes equal transcription of all genes. We however preferred to use this approximation instead of using experimental data from another organism. Fatty acid composition of *L. reuteri* was obtained from literature [53], while phospholipid composition was adopted from *L. plantarum*. The composition of lipoteichoic acid [54] and exopolysaccharides [55] in *L . reuteri* were obtained from literature. Peptidoglycan composition was adopted from *L. plantarum* and glycogen was assumed to be negligible [56], [57].

Energy required for growth- (GAM) and cell maintenance (NGAM) are important parameters in metabolic models, and can be estimated from ATP production rates, which can be calculated from experimental data obtained at different dilution rates [58]. Unfortunately, this data is not publicly available for *L. reuteri*. These parameters have been estimated from experimental data for several other LAB, including *L. plantarum*, and reported in literature [8]. Even though *L. reuteri* and *L. plantarum* are relatively closely related, adopting these parameters from *L. plantarum* can negatively affect the quality of model predictions. When the differences in physiologies of *L. plantarum* and *L. reuteri* are considered, it is possible that *L. reuteri* requires less energy: (1) The genome is only ~2 Mb, while *L. plantarum´*s genome is 3.3 Mb. (2) *L. reuteri* is an obligate heterofermenter, which means it uses almost solely the PKP (Fig. 2) to break down glucose, resulting in one ATP per glucose, while a facultative heterofermenter like *L. plantarum* uses the EMPP when grown on glucose, resulting in two ATPs. (3) LAB in general have low catabolic capabilities, and for *L. reuteri* this includes auxotrophy for several amino acids. This, combined with the fact that macromolecular biosynthesis is already accounted for in the model reactions, supports the claim that adopting energy parameters from *L. plantarum* can negatively affect model predictions, as we also observed when evaluating this in our model. We decided to use one of our experimental datasets (Table 3) to estimate the GAM value, while using the NGAM value from *L. plantarum* (section 2.4). In general, NGAM represents only a small portion of the total energy requirements of the cell and therefore has much smaller effect on model predictions than GAM. This resulted in a GAM value of 10.2 mmol gDW^{-1} h^{-1}. Detailed

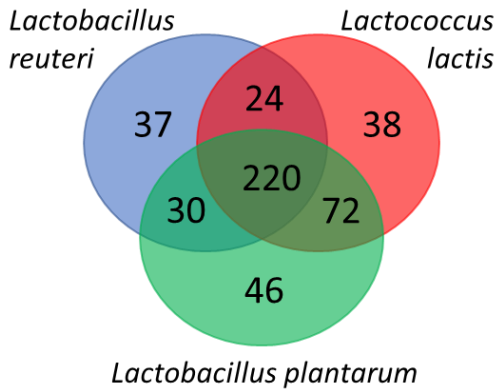
description of the biomass reaction, relevant data and calculations can be found in Additional file 7.

The sensitivity of the predicted growth rate to changes in biomass and energy components was investigated by varying the coefficient of each component, one at a time, by 50% while varying the glucose uptake rate. The components tested were protein, polysaccharide, DNA, RNA, lipid, GAM and NGAM. The analysis showed that predicted growth rate was sensitive to changes in the protein and GAM components of the biomass, compared to the other components (see figure in Additional file 7). As described earlier, these two particular components of the biomass are based on *L. reuteri* specific experimental data obtained in this study.

3.1.3 *L. reuteri* model compared to models of *L. lactis* and *L. plantarum*

The model was compared to genome-scale metabolic models of two other LAB, *L. lactis* and *L. plantarum* (Table 1). Common and unique metabolic reactions were analyzed based on EC-numbers (Figure 3A). Unique reactions in *L. reuteri* included reactions belonging to: cofactor and prosthetic group biosynthesis, most of which related to B-12 vitamin synthesis; alternative carbon metabolism, such as glycerol; amino acid metabolism, which can be explained by the different amino acid auxotrophies among the three strains; methylglyoxal metabolism (see Additional file 8 for more details). Basic model statistics and biomass composition from the three models are presented in Figure 3B. Comparing biomass ratios shows that *L. plantarum´*s biomass contains less protein then the others and more teichoic acid. Model predictions reflected the well-established and previously discussed differences in glycolytic pathways between the strains, namely how *L. lactis* and *L. plantarum*, as homofermenter and facultative heterofermenter, use the EMP pathway resulting in higher energy compared to *L. reuteri* which, as a strict heterofermenter, mostly uses the PKP.

A



B

		Model statistics		Biomass composition (%)						
Species	Genes	Reactions	Metabolites	Protein	DNA	RNA	Lipids	TCA	PS	PG
L. reuteri	530	710	658	43,0	2,3	10,9	7,7	9,7	12,0	14,3
L. lactis	516	754	650	46,0	2,3	10,7	3,4	8,0	12,0	11,8
L. plantarum	724	778	662	29,9	1,9	9,0	6,3	17,9	9,9	14,5

Figure 3. *Comparison of genome-scale metabolic models of L. reuteri, L. lactis and L. plantarum. A) Venn diagram showing number of EC-numbers that represent common and unique reactions in the three models. B) Model statistics for the three strains, along with biomass ratio.*

3.2 Model applications

3.2.1 Model validation using experimental data: Growth rate comparisons

To validate the model, several different datasets (Table 3) with measured uptake- and secretion rates of carbon sources, amino acids and organic byproducts were used to constrain exchange fluxes in the model. The predicted growth rates were compared with observed experimental growth rates (Figure 4). In all cases, flux through the EMPP was set to maximally 7% based on the experimentally determined value (Figure 1). The chemically defined culture medium used in the growth experiments contained all 20 amino acids, except for L-glutamine. Subsequently, all these amino acids were quantified during growth and the model was constrained with the resulting uptake rates. Of all the amino acids, only arginine was depleted at the end of the exponential phases in data sets A, B and C (Additional file 1). Due to auxotrophy for several amino acids (Glu, His, Thr, Arg, Tyr, Val, Met, Try, Phe, Leu), the model is highly sensitive to uncertainties in measurements, as well as in determined protein- and amino acid fractions of the biomass reaction. To accurately represent amino acids in the biomass reaction, both the protein content and the amino acid ratio were measured (Additional file 7). By enabling unrestricted uptake of amino acids in the model, we noticed that only 5 amino acids (Arg, Ser, Asn, Asp, Glu) needed to be

constrained with measured uptake rates for accurate growth predictions, for both the wild-type and the mutant. This is due to their role in energy- and cofactor metabolism, not only in biomass biosynthesis. Hence, only this minimum number of amino acids was used to constrain the model in the following. The remainder were assumed to be non-limiting by allowing unrestricted uptake. This has twofold advantage. First, it limits the effects of uncertainties in amino acid uptake rate measurements on model predictions, a problem exacerbated by the amino acid auxotrophy. Second, it simplifies future applications of the model by reducing the number of measurements needed.

In most cases, model predictions and *in vivo* data were in good agreement (Figure 4). Datasets C and D in Figure 4 show a variant of the WT strain (marked SJ (WT*)), which lacks two restriction modification (RM) systems for easier genetic manipulation (Table 2). Datasets E and F show a mutant derived of the SJ strain with a clean and in-frame deletion of the *adhE* gene (bifunctional aldehyde/alcohol dehydrogenase). The model predicts slightly higher growth rates than observed *in vivo* for the SJ strain (datasets C and D in Figure 4) and the mutant strain grown on glucose and glycerol (F in Figure 4). Unexpectedly, the RM-modifications in the SJ strain seem to slightly alter its behavior on CDM with glucose and glycerol compared to the WT (Additional file 1). For the mutant strain grown on glucose (dataset E in Figure 4), the model predicts a slightly lower growth rate than observed *in vivo*, though both show a large decrease in growth, compared to the WT. The most likely explanation for this is that some glucose is being taken up *in vivo*, even though the measurements did not show this (the likely amount consumed between two samples is within the error of the assay). Secretion of 2.6 mmol gDW^{-1} h^{-1} of lactate and 2.7 mmol gDW^{-1} h^{-1} of acetate was observed *in vivo*. The model, however, does not predict lactate and acetate secretion unless some glucose uptake is allowed. If a glucose uptake of 2.6 mmol gDW^{-1} h^{-1} is allowed, the growth rate increases from 0.22 to 0.34 h^{-1}, compared to 0.30 h^{-1} *in vivo.* Amino acid measurements showed that the mutant in dataset E used L-arginine to a greater extent than the WT, which the model predicts is used to generate energy via the arginine deiminase pathway, resulting in increased growth.

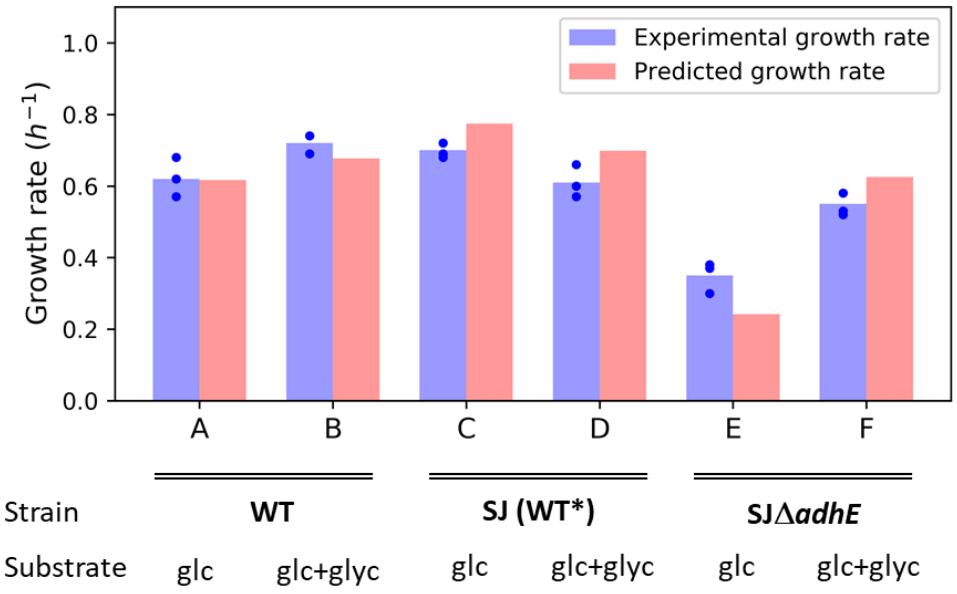


Figure 4. *Predicted and experimental growth rates. Experimentally measured growth rates for each of the six data sets are shown in blue, with blue dots denoting individual replicates and blue bars representing average values. For each dataset, the model was constrained with average experimental values for uptake and secretion rates of carbon sources, byproducts and selected amino acids, and optimized for growth. Predicted growth rates are represented by red bars. Different datasets used are indicated with letters - abbreviations: glc: glucose; glyc: glycerol.*

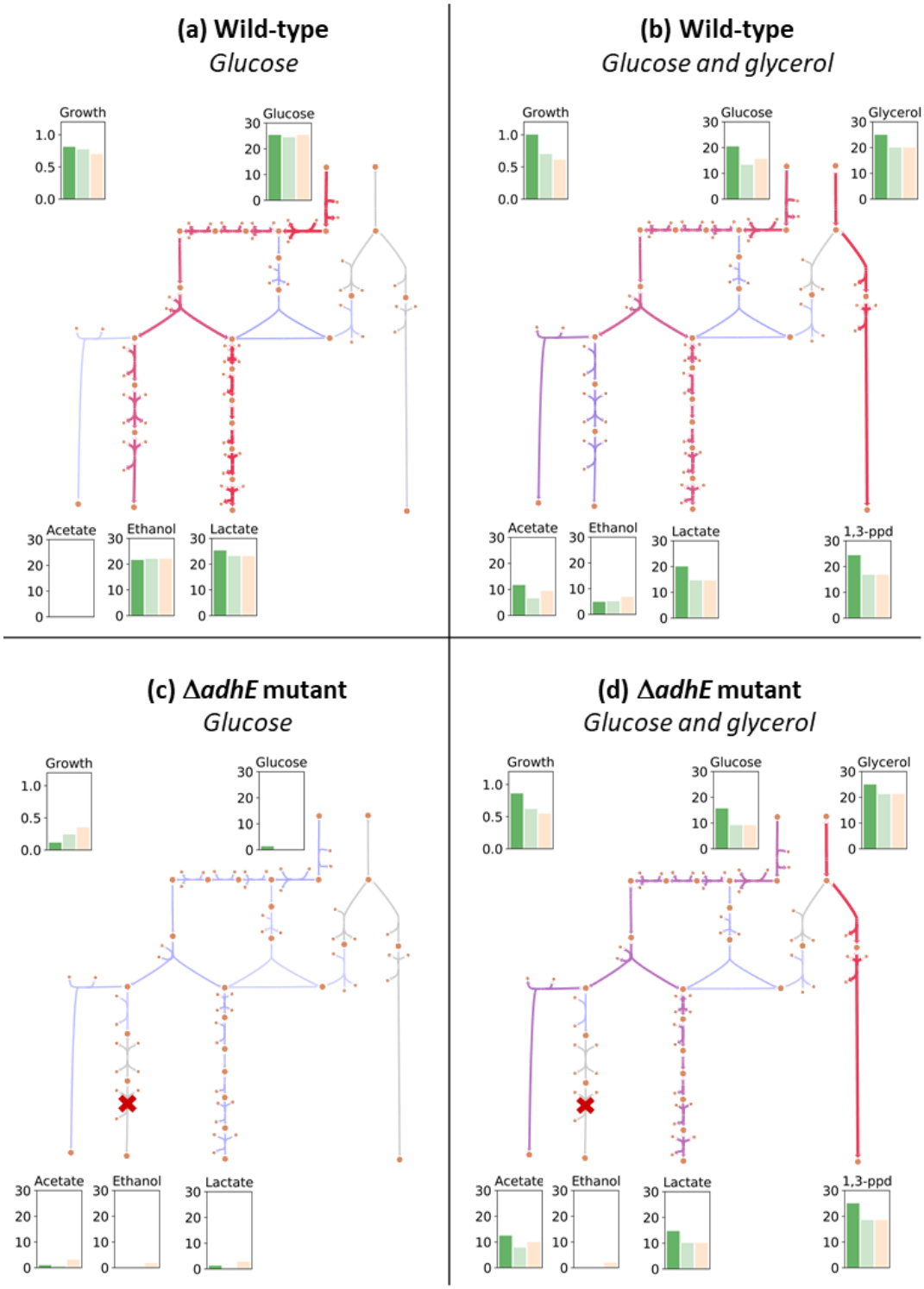
3.2.2 Effects of adding glycerol and deleting *adhE*

To investigate the applicability of the model for cell factory design, it was used to predict the effects of adding glycerol to the glucose-based culture medium, as well as knocking out the *adhE* gene, which plays a critical role in ethanol production and redox balance (Figure 1). The datasets used here are the same as in section 3.2.1 (datasets C - F in Figure 4). There, the aim was to validate the model by means of comparing predicted growth rates to experimentally determined growth rates. In this section, we look more specifically at predicted flux distributions in central metabolism, both with and without strain- and condition-specific experimentally determined constraints. For this purpose, we studied two cases in order to answer the following questions: (1) If the model is constrained only with experimentally determined glucose- and five amino acid uptake rates from the WT strain grown on glucose, how do the predicted effects of glycerol addition and/or *adhE* knock-out (dark green bars in figure) compare to *in vivo* growth rate and uptake- and secretion measurements (light orange bars in figure 5)? This was tested to evaluate the applicability of the model in a practical setting. One of the main goals of using a model like this should be to probe the effects of genetic and media perturbations *in silico*, *i.e.* without having to do extensive condition-specific cultivations and measurements beforehand. (2) If the model is

constrained with uptake- and secretion rates of carbon source(s), amino acids and byproducts of the strain and condition under study, how well do the model predictions (light green bars in figure 5) compare to *in vivo* results? Here the model was allowed, but not forced, to take up (lower bound constrained, upper bound unconstrained) and secrete (lower bound unconstrained, upper bound constrained) metabolites according to the experimental data. This tells us if the model, when imposed with realistic limitations, "chooses" a flux distribution which results in extracellular fluxes of metabolites in line with *in vivo* data. In both cases, the constrained amino acids only included Arg, Ser, Asn, Asp and Glu as before (see section 3.2.1) and in case 1 the allowed glycerol uptake rate was arbitrarily limited to 25 mmol gDW^{-1} h^{-1}, when glycerol effects were being predicted.

The flux maps in Figure 5 show results for case 1 (dark green bars). The predicted uptake of glucose and glycerol (dark green bars in figure 5b) is higher than observed *in-vivo* (light orange bars in figure 5b), resulting in higher secretion of by-products and a higher growth rate as well. However, the distribution of secreted by-products is very similar. The effect of glycerol can be predicted quite well with the model as ethanol secretion decreases and acetate secretion increases, relative to glucose uptake, and 1,3-propanediol is secreted in large amounts (compared to graphs in Figure 5a). Several studies have described an increased growth rate in *L. reuteri* when glycerol is added to a glucose-based medium (in flasks and bioreactors), which is to be expected based on inspection of redox balance [15], [41], [44] and this is also what we observed *in silico* in case 1. *L. reuteri* uses practically only the PKP and not the EMP for glucose fermentation. In the PKP, two extra NAD(P)H molecules are formed compared to the EMP, which are regenerated to $NAD(P)^+$ by *adhE* through the formation of ethanol (Figure 1). When glycerol is added, it is used as an alternative electron acceptor via the production of 1,3-PDO, which generates one NAD^+. As a result, one of the actetyl-phosphates that is normally converted to ethanol can now be converted to acetate. This does not yield NAD^+ (which is now regenerated in 1,3-PDO production) but does yield one ATP, enabling a higher growth rate [15], [44].

Figure 5. *Predicted and experimental fluxes of key metabolites in the wild-type strain (SJ) and the adhE mutant. The wild-type strain was grown on glucose (a) and glucose and glycerol (b), and the adhE mutant was also grown on glucose (c) and glucose and glycerol (d). Bar plots show the average measured rates from 3 replicates (light orange), predicted rates from model constrained with average experimental uptake rates of the WT grown on glucose, or case 1 (dark green), and predicted rates from model constrained with average experimental rates from the strain and condition under study, or case 2 (light green). Metabolic maps show predicted flux distributions for case 1. All units for uptake- and secretion rates are in mmol gDW^{-1} h^{-1} and for growth rates in h^{-1}. Metabolic maps correspond to Figure 1.*


(a) Wild-type
Glucose
Growth
Glucose
Acetate
Ethanol
Lactate
(b) Wild-type
Glucose and glycerol
Growth
Glucose
Glycerol
Acetate
Ethanol
Lactate
1,3-ppd
(c) ΔadhE mutant
Glucose
Growth
Glucose
Acetate
Ethanol
Lactate
(d) ΔadhE mutant
Glucose and glycerol
Growth
Glucose
Glycerol
Acetate
Ethanol
Lactate
1,3-ppd
In-silico constrained with WT glucose- and a.a. uptake rates (Case 1)
In-silico constrained with in-vivo uptake- and secretion rates (Case 2)
In-vivo
0
30
Flux (mmol gDW-1 h-1)

Along these lines of reasoning and in line with existing literature [44], knocking out the *adhE* gene has dramatic effects on the metabolism when glucose is the sole carbon source, both *in vivo* and *in silico* (Figure 5c). When ethanol production was inactive, the growth decreased which also led to reduction in lactate production. This is due to redox imbalance since AdhE no longer recycles the NADH generated in glycolysis. The predictions in case 1 show highly decreased uptake of glucose, yet a small amount of glucose is still taken up, resulting in acetate and lactate production. As discussed in 3.2.1, it is possible that glucose is being taken up *in vivo,* even though this is not detected by measurements, which is in line with model predictions and would also explain the lower growth rate observed *in silico* in case 2 compared to *in vivo*. The higher growth rate *in vivo* compared to *in silico* in case 1 is due to a much higher arginine uptake than measured in the WT. Also in line with published studies and the redox balance explained above [44], addition of glycerol to the *adhE* mutant increases the growth rate to almost WT levels (Figure 5d). Similarly to the WT predictions, the model in case 1 predicts slightly higher growth rate and uptake rates of glucose and glycerol, resulting in higher secretion of by-products. But as before, the flux distribution is very similar to the one measured *in vivo*.

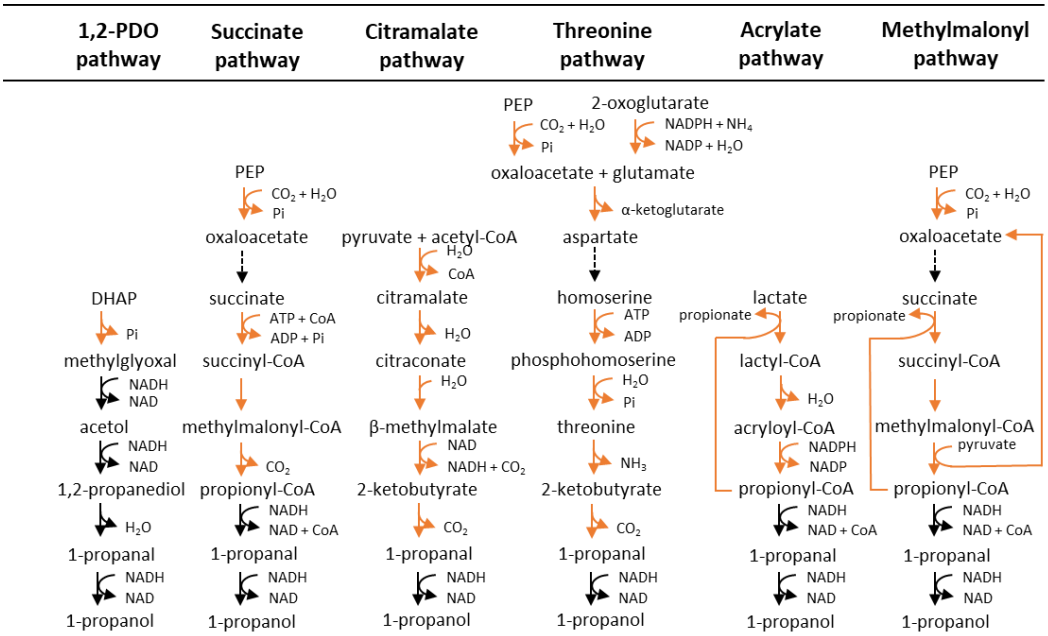
In all four conditions in figure 5 the *in silico* predictions in case 2 and the *in vivo* data are almost identical, with the exception of the few instances described above. In few cases discrepancies can be explained by carbon imbalance *in vivo*, which is most likely due to measurement uncertainties. Taken together, these results show that the model can be used to accurately predict metabolic behavior, without requiring extensive experimental data.

3.2.3 Model-based analysis of 1-propanol production in *L. reuteri*

In the two previous sections we used experimental data to validate model predictions. In this section we focus on *in silico* predictions involving the production of 1-propanol. It has been shown that heterologous expression of methylglyoxal synthase (*mgs)* in *L. reuteri* can activate the pathway to 1,2-PDO and 1-propanol production [6]. Both these compounds have many applications, *e.g.* in the production of polyester resins for 1,2-PDO and as a solvent and potential biofuel for 1-propanol [59]. In addition to the 1,2-PDO pathway, several different pathways have been described for 1-propanol production [60]. The 1,2-PDO pathway towards 1-propanol has not frequently been reported for 1-propanol production;

the most frequently used pathways are the citramalate and threonine pathways (Figure 6). Other options are the acetone, Wood-Werkman (or methylmalonyl), acrylate and succinate pathways [60]. For a recent extensive overview of these pathways and engineered and non-engineered organisms, the reader is referred to the review by Walther & Francois [60]. The thermodynamic maximum yield of 1-propanol from glucose calculated based on the degree of reduction is 1.33 mol/mol (or 44,4% carbon yield). However, only the stoichiometry of the 1,2-PDO, succinate, acrylate and Wood-Werkman pathways allow this maximum yield – the others result in up to 25% less yield [60]. Here, we evaluated the citramalate, threonine, succinate, acrylate, methylmalonyl, and 1,2-PDO pathway. One of the advantages of *L. reuteri* for using the 1,2-PDO pathway towards 1-propanol is that *L. reuteri* produces vitamin B12, which is needed as a co-factor for the B12-dependent diol dehydratase step and has been suggested to be a limiting factor in this pathway in for example *C. glutamicum* [61] and was added to the fermentations in an *E. coli* strain harboring this pathway [62]. The model was used to analyze the suitability of six different pathways for 1-propanol production in *L. reuteri* (Figure 6). These were the known pathways as described in literature, but in addition we used the minRxn algorithm [63] and the accompanying database of enzymatic reactions to search for heterologous pathways from glucose to 1-propanol. This search did not reveal pathways that differed significantly from the already-known pathways as shown in Figure 6, neither qualitatively nor in terms of carbon yields (data not shown). The model was maximized for 1-propanol production, while constrained with the experimentally determined uptake rate of glucose and additional 20 mmol/gDW/h of glycerol when applicable, no uptake of amino acids and free secretion of by-products. Of the six pathways in Figure 6, the citramalate pathway performed the worst, with no production on glucose and only 9.5% maximum carbon yield when glycerol was added. This was due to redox imbalance, which was due to lack of production of the precursors pyruvate and acetyl-CoA, and was partly fixed by adding glycerol. Adding this pathway in combination with the threonine pathway did not increase maximum carbon yield compared to threonine pathway alone. The highest carbon yields, 40.0% and 45.7% on glucose and glucose+glycerol, respectively, were observed in four different pathways, namely the 1,2-PDO, succinate, acrylate and methylmalonyl pathways. Since the maximum theoretical carbon yield of this pathway is 44.4% on glucose and 50.8% on glucose+glycerol, these values represent around 90% of the theoretical maximum, which can be achieved with minimal metabolic

engineering (*i.e.* only heterologous expression of *mgs*). Since for the 1,2-PDO pathway only one heterologous enzyme is needed, we decided to analyze 1-propanol production via the 1,2-PDO pathway further.



	1,2-PDO pathway	Succinate pathway	Citramalate pathway	Threonine pathway	Acrylate pathway	Methylmalonyl pathway
Glc	40.0%	40.0%	0%	25.2%	40.0%	40.0%
Glc+ Glyc	45.7%	45.7%	9.5%	29.5%	45.7%	45.7%

Figure 6. *1-propanol biosynthetic pathways tested in silico. Black reactions are natively present and yellow are heterologous. Theoretical carbon yields, obtained by maximizing for 1-propanol production on either glucose or glucose and glycerol, are listed under each pathway.*

While optimizing *in silico* for 1-propanol production as described above is a good way to assess suitability of different engineering strategies, optimizing for growth gives more biologically realistic results. Hence, next the model was used to maximize growth while constrained with experimental uptake rates of glucose and the 5 amino acids from the WT grown on glucose (same as case 1 in section 3.2.2). As previously discussed (section 3.2.2), the *adhE* mutant grows poorly on glucose due to redox imbalance. The synthesis of both 1,2-propanediol and 1-propanol consumes NADH and activating these pathways therefore has the potential to restore growth. The *adhE* gene was knocked out *in silico*, and flux predictions with (Figure 7) and without (Figure 5c) an active 1,2-PDO pathway were compared. The active 1,2-PDO pathway resulted in a high increase in growth rate (0.11 to 0.49 h^{-1}) as well as growth-coupled production of 1-propanol (14.7 mmol gDW^{-1} h^{-1}). Given the good agreement between *in silico* predictions and *in vivo* measurements in section 3.2.2,

the expression of the missing *mgs* gene in the 1,2-PDO pathway at a sufficiently high level *in vivo* is expected to result in a relatively fast-growing 1-propanol-producing cell factory, which is also in agreement with existing literature (International Publication Number WO 2014/102180 AI, 2014).

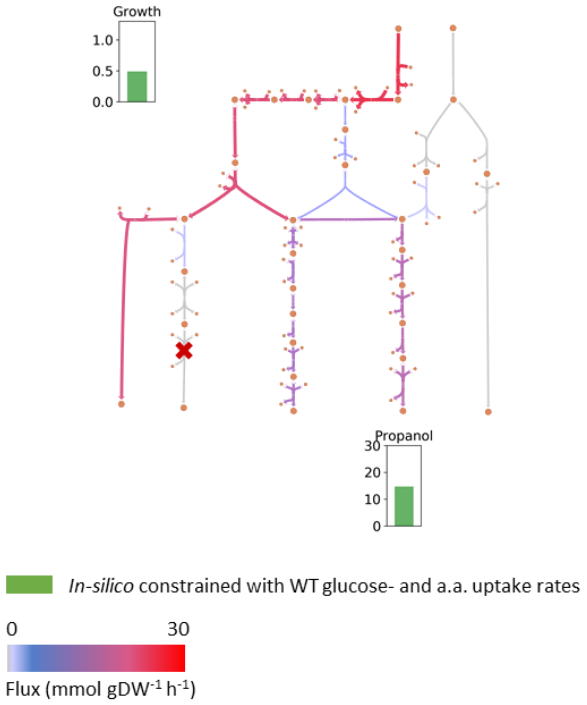


Figure 7. *Predicted flux distribution, growth rate and 1-propanol production of adhE mutant grown on glucose, with active 1,2-propanediol and 1-propanol pathways. The model was constrained with average experimental uptake rates of the WT grown on glucose and optimized for growth. Units for propanol secretion rate is in mmol gDW^{-1} h^{-1} and growth rate in h^{-1}.*

3.2.4 Model-based analysis of *L. reuteri* as a cell factory

LAB are natural producers of several chemicals of industrial interest [4], [64], [65]. They possess high sugar uptake rates and, in many species, the central metabolism is only weakly coupled to biomass formation because of their adaptation to nutrient rich environments. As a result, the carbon source is mostly used for energy gain and is converted to fermentation products in high yields. Combined with high tolerance to environmental stress, these properties have led to significant interest in using LAB as cell factories.

The heterofermentative nature of *L. reuteri* and the dominance of the phosphoketolase over the Embden-Meyerhof-Parnas pathway make some target compounds less suitable than others, with lactic acid being an obvious example. On the other hand, these properties can

also be used to an advantage as is demonstrated here We used our newly established *L. reuteri* metabolic model to study the feasibility of this organism to produce some of the compounds that have been the subject of recently published LAB metabolic engineering experiments. These native and non-native compounds include a flavoring compound (acetoin), a food additive (L-alanine), biofuels (1-propanol and ethanol), chemical building blocks (acetaldehyde and 2,3-butanediol) and an environmentally friendly solvent (ethyl lactate). The last compound has recently been produced in an engineered *E. coli* strain [24] and is an interesting target in *L. reuteri* since it is a condensation product of the two major products of glucose fermentation via the phosphoketolase pathway, lactate and ethanol.

The suitability of *L. reuteri* for producing a particular compound was assessed in terms of the maximum carbon yield, using a fixed glucose uptake rate (Table 5). This gives an overly optimistic estimate of product yields in most cases since it completely ignores variations in enzyme efficiency, compound toxicity, regulation and other issues outside the scope of the model. The maximum flux is still useful to identify products that appear to be ill suited for a particular metabolism as well as products that may be suitable.

The predicted flux for acetaldehyde, acetoin and 2,3-butanediol, which are all derived from acetyl-CoA, was low, suggesting that the metabolism in the wild type is not well suited for overproducing these compounds. The flux increased significantly upon addition of methylglyoxal synthase, suggesting the importance of the 1-propanol pathway in cofactor balancing (section 3.2.3). Addition of glycerol to the medium served the same purpose and increased the predicted flux in all cases (data not shown), which is in line with glycerol being known and used as an external electron sink in *L. reuteri* [2]. For all the compounds except ethanol and 1-propanol, the addition of a fully functional phosphofructokinase was predicted to increase the yields even further (Table 5). Such a strategy has been shown successful for mannitol production [66].

Taken together, the model suggests that *L. reuteri* is better suited for producing compounds derived from pyruvate than compounds derived from acetyl-CoA and that the simultaneous expression of heterologous MGSA and PFK enzymes is a general metabolic engineering strategy for increasing product yields in *L. reuteri*.

Table 5. *Model predictions of the maximum theoretical yields of selected target compounds.*

Compound	***Maximum flux [mmol gDW^{-1} h^{-1}]***			***Maximum carbon yield***
		MGSA	*MGSA, ↑PFK*	
Ethanol	*50.4*	*50.4*	*50.4*	*67%*
Acetaldehyde	*0*	*31.5*	*37.8*	*50%*
1-propanol (n-n)	*20.2*	*20.2*	*20.2*	*40%*
L-alanine (n-n)	*27.0*	*27.0*	*50.4*	*100%*
Acetoin	*0*	*10.1*	*18.9*	*50%*
2,3-butanediol	*0*	*11.6*	*21.6*	*57%*
Ethyl lactate (n-n)	*20.4*	*20.4*	*25.2*	*83%*

A maximum glucose uptake rate of 25.2 mmol gDW^{-1} h^{-1} was assumed. MGSA indicates the presence of methylglyoxal synthase in the model, ↑PFK indicates the presence of a phosphofructokinase that is not flux-limiting. Non-native compounds are indicated with (n-n).

4. Conclusions

In this study, we have established a manually curated genome-scale metabolic model of *L. reuteri* JCM 1112, referred to as Lreuteri_530, and validated it with experimental data. We identified several knowledge gaps in the metabolism of this organism that we resolved with a combination of experimentation and modeling. The distribution of flux between the PKP and EMPP pathways is strain-specific and in line with other studies, we found that the EMPP activity is maximally around 7% of total glycolytic flux during early exponential phase. The predictive accuracy of the model was estimated by comparing predictions with experimental data. Several scenarios were tested both *in vivo* and *in silico*, including addition of glycerol to a glucose-based growth medium and the deletion of the *adhE* gene, which encodes a bifunctional aldehyde/alcohol dehydrogenase. The results showed that the model gives accurate predictions, both with respect to growth rate and uptake- and secretion rates of main metabolites in the central metabolism. This indicates that the model can be useful for predicting metabolic engineering strategies, such as growth-coupled production of 1-propanol. The model also serves as a starting point for the modeling of other L. *reuteri* strains and related species. The model is available in SBML, Matlab and JSON formats at https://github.com/steinng/reuteri as well as in Additional file 6. Metabolic maps in Escher format are provided in Additional file 4. The Escher maps together with the model in JSON

format can be used directly with the Escher-FBA online tool [67] as well as the Caffeine cell factory design and analysis platform (https://caffeine.dd-decaf.eu/).

5. Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

The model, experimental data, code and other relevant material are available from github.com/steinng/reuteri and Additional files.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by the Marine Biotechnology ERA-NET *ThermoFactories* project grant number 5178–00003B; the Technology Development fund in Iceland grant number 159004-0612; The Novo Nordisk Foundation in Denmark; and the European Union's Horizon 2020 research and innovation programme under grant agreement No 686070 (DD-DeCaF).

Authors' contributions

TK and SG curated and validated the metabolic reconstruction, performed numerical simulations and wrote the manuscript. EFB performed the experimental work except the bioreactor cultivations, performed data processing and analysis, curated the metabolic reconstruction and wrote the manuscript. FBdS constructed the draft model, performed bioreactor cultivations, analyzed the resulting data and revised the manuscript. EÖ curated the original draft metabolic reconstruction, processed and analyzed the genome sequencing data and revised the manuscript. ATN and MJH were involved in the metabolic reconstruction and revised the manuscript. BSF and LF performed a bioreactor cultivation, analyzed the resulting data and revised the manuscript. EFB, TK, SG and ATN conceived and coordinated this study. All authors read and approved the final manuscript.

Acknowledgements

The authors would like to thank Bjarke Krysel Christensen, Steen Troels Jørgensen and Brian Kobmann from Novozymes for providing strain SJ11774; Anna Koza from DTU Biosustain for performing the genome sequencing; Amalie Melton Axelsen from DTU Biosustain for technical support with construction and analysis of the *adhE* mutant strain.

References

[1] D. M. Saulnier *et al.*, "Exploring Metabolic Pathway Reconstruction and Genome-Wide Expression Profiling in Lactobacillus reuteri to Define Functional Probiotic Features," *PLoS One*, vol. 6, no. 4, p. e18783, Jan. 2011.

[2] T. Dishisha, L. P. Pereyra, S.-H. Pyo, R. A. Britton, and R. Hatti-Kaul, "Flux analysis of the Lactobacillus reuteri propanediol-utilization pathway for production of 3-hydroxypropionaldehyde, 3-hydroxypropionic acid and 1,3-propanediol from glycerol.," *Microb. Cell Fact.*, vol. 13, p. 76, 2014.

[3] M. A. Ricci, A. Russo, I. Pisano, L. Palmieri, M. de Angelis, and G. Agrimi, "Improved 1,3-Propanediol Synthesis from Glycerol by the Robust Lactobacillus reuteri Strain DSM 20016," *J. Microbiol. Biotechnol.*, vol. 25, no. 6, pp. 893–902, 2015.

[4] E. F. Bosma, J. Forster, and A. T. Nielsen, "Lactobacilli and pediococci as versatile cell factories – Evaluation of strain properties and genetic tools," *Biotechnol. Adv.*, vol. 35, no. 4, pp. 419–442, 2017.

[5] T. L. Talarico and W. J. Dobrogosz, "Chemical characterization of an antimicrobial substance produced by Lactobacillus reuteri," *Antimicrob. Agents Chemother.*, vol. 33, no. 5, pp. 674–679, 1989.

[6] B. Christensen, P. B. Olsen, T. B. Regueira, B. Koebmann, S. T. Joergensen, and T. I. Dehli, "International Publication Number WO 2014/102180 AI," no. 12. 2014.

[7] M. H. Rau and A. A. Zeidan, "Constraint-based modeling in microbial food biotechnology.," *Biochem. Soc. Trans.*, vol. 46, no. 2, pp. 249–260, 2018.

[8] B. Teusink *et al.*, "Analysis of Growth of Lactobacillus plantarum WCFS1 on a Complex Medium Using a Genome-scale Metabolic Model," *J. Biol. Chem.*, vol. 281, no. 52, pp. 40041–40048, 2006.

[9] E. Vinay-Lara, J. J. Hamilton, B. Stahl, J. R. Broadbent, J. L. Reed, and J. L. Steele, "Genome -scale reconstruction of metabolic networks of lactobacillus casei ATCC 334 and 12A," *PLoS One*, vol. 9, no. 11, 2014.

[10] A. P. Oliveira, J. Nielsen, and J. Förster, "Modeling Lactococcus lactis using a genome-scale flux model," *BMC Microbiol.*, vol. 5, no. 1, 2005.

[11] N. A. L. Flahaut *et al.*, "Genome-scale metabolic model for Lactococcus lactis MG1363 and its application to the analysis of flavor formation," *Appl. Microbiol. Biotechnol.*, vol. 97, no. 19, pp. 8729–8739, 2013.

[12] M. I. Pastink, B. Teusink, P. Hols, S. Visser, W. M. De Vos, and J. Hugenholtz, “Genome-scale model of Streptococcus thermophilus LMG18311 for metabolic comparison of lactic acid bacteria,” *Appl. Environ. Microbiol.*, vol. 75, no. 11, pp. 3627–3633, 2009.

[13] L. Koduru, Y. Kim, J. Bang, M. Lakshmanan, N. S. Han, and D.-Y. Lee, “Genome-scale modeling and transcriptome analysis of Leuconostoc mesenteroides unravel the redox governed metabolic states in obligate heterofermentative lactic acid bacteria,” *Sci. Rep.*, vol. 7, no. 1, p. 15721, 2017.

[14] E. Özcan, S. S. Selvi, E. Nikerel, B. Teusink, E. Toksoy Öner, and T. Çakır, “A genome-scale metabolic network of the aroma bacterium Leuconostoc mesenteroides subsp. cremoris,” *Appl. Microbiol. Biotechnol.*, vol. 103, no. 7, pp. 3153–3165, 2019.

[15] F. Santos, “Vitamin B12 synthesis in Lactobacillus reuteri,” Wageningen University, 2008.

[16] B. Teusink *et al.*, “In silico reconstruction of the metabolic pathways of *Lactobacillus plantarum*: comparing predictions of nutrient requirements with those from growth experiments,” *Appl. Environ. Microbiol.*, vol. 71, no. 11, pp. 7253–7262, 2005.

[17] D. E. Deatherage and J. E. Barrick, “Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq.,” *Methods Mol. Biol.*, vol. 1151, pp. 165–188, 2014.

[18] Z. A. King *et al.*, “BiGG Models: A platform for integrating, standardizing and sharing genome-scale models,” *Nucleic Acids Res.*, vol. 44, no. D1, pp. D515--D522, 2016.

[19] D. A. Fell and J. R. Small, “Fat synthesis in adipose tissue. An examination of stoichiometric constraints,” *Biochem. J.*, vol. 238, pp. 781–786, 1986.

[20] J. M. Savinell and B. O. Palsson, “Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism,” *J. Theor. Biol.*, vol. 154, no. 4, pp. 421–454, 1992.

[21] N. E. Lewis *et al.*, “Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models,” *Mol. Syst. Biol.*, vol. 6, no. 390, 2010.

[22] A. Ebrahim, J. A. Lerman, B. O. Palsson, and D. R. Hyduke, “COBRApy: COnstraints-Based Reconstruction and Analysis for Python,” *BMC Syst. Biol.*, vol. 7, 2013.

[23] Z. A. King, A. Dräger, A. Ebrahim, N. Sonnenschein, N. E. Lewis, and B. O. Palsson, “Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways,” *PLoS Comput. Biol.*, vol. 11, no. 8, pp. 1–13, 2015.

[24] J.-W. Lee and C. T. Trinh, “De novo Microbial Biosynthesis of a Lactate Ester Platform,” *bioRxiv*, p. 498576, Jan. 2018.

[25] C. Lieven *et al.*, “Memote: A community-driven effort towards a standardized genome-scale metabolic model test suite,” *bioRxiv*, p. 350991, Jan. 2018.

[26] O. Kandler, K.-O. Stetter, and R. Köhl, “Lactobacillus reuteri sp. nov., a New Species of Heterofermentative Lactobacilli,” *Zentralblatt für Bakteriol. I. Abt. Orig. C Allg. Angew.*

und ökologische Mikrobiol., vol. 1, no. 3, pp. 264–269, Sep. 1980.

[27] S. A. Frese *et al.*, "The Evolution of Host Specialization in the Vertebrate Gut Symbiont Lactobacillus reuteri," *PLoS Genet.*, vol. 7, no. 2, p. e1001314, Feb. 2011.

[28] H. Morita *et al.*, "Comparative Genome Analysis of Lactobacillus reuteri and Lactobacillus fermentum Reveal a Genomic Island for Reuterin and Cobalamin Production," *DNA Res.*, vol. 15, no. 3, pp. 151–161, 2008.

[29] S. T. Joergensen, T. B. Regueira, B. Kobmann, P. B. Olsen, and B. Christensen, "Bacterial mutants with improved transformation efficiency." p. US 2015O125959A1, 2015.

[30] S. Etzold *et al.*, "Structural and molecular insights into novel surface-exposed mucus adhesins from *L actobacillus reuteri* human strains," *Mol. Microbiol.*, vol. 92, no. 3, pp. 543–556, May 2014.

[31] Z. Sun *et al.*, "Expanding the biotechnology potential of lactobacilli through comparative genomics of 213 strains and associated genera," *Nat. Commun.*, vol. 6, p. 8322, Jan. 2015.

[32] G. Burgé, C. Saulou-Bérion, M. Moussa, F. Allais, V. Athes, and H.-E. Spinnler, "Relationships between the use of Embden Meyerhof pathway (EMP) or Phosphoketolase pathway (PKP) and lactate production capabilities of diverse Lactobacillus reuteri strains," *J. Microbiol.*, vol. 53, no. 10, pp. 702–710, 2015.

[33] E. Årsköld, E. Lohmeier-Vogel, R. Cao, S. Roos, P. Rådström, and E. W. J. Van Niel, "Phosphoketolase pathway dominates in Lactobacillus reuteri ATCC 55730 containing dual pathways for glycolysis," *J. Bacteriol.*, vol. 190, no. 1, pp. 206–212, 2008.

[34] T. S. Kang, D. R. Korber, and T. Tanaka, "Regulation of dual glycolytic pathways for fructose metabolism in heterofermentative Lactobacillus panis PM1.," *Appl. Environ. Microbiol.*, vol. 79, no. 24, pp. 7818–7826, 2013.

[35] M. H. Saier, "Families of transmembrane sugar transport proteins," *Mol. Microbiol.*, vol. 35, no. 4, pp. 699–710, Feb. 2000.

[36] A. Galinier and J. Deutscher, "Sophisticated Regulation of Transcriptional Factors by the Bacterial Phosphoenolpyruvate: Sugar Phosphotransferase System," *J. Mol. Biol.*, vol. 429, no. 6, pp. 773–789, Mar. 2017.

[37] B. Görke and J. Stülke, "Carbon catabolite repression in bacteria: many ways to make the most out of nutrients," *Nat. Rev. Microbiol.*, vol. 6, no. 8, pp. 613–624, 2008.

[38] J. Zheng, L. Ruan, M. Sun, and M. Gänzle, "A Genomic View of Lactobacilli and Pediococci Demonstrates that Phylogeny Matches Ecology and Physiology.," *Appl. Environ. Microbiol.*, vol. 81, no. 20, pp. 7233–7243, 2015.

[39] A. H. Romano, J. D. Trifone, and M. Brustolon, "Distribution of the phosphoenolpyruvate:glucose phosphotransferase system in fermentative bacteria," *J. Bacteriol.*, vol. 139, no. 1, pp. 93–97, 1979.

[40] X. Zhao and M. G. Gänzle, "Genetic and phenotypic analysis of carbohydrate

metabolism and transport in Lactobacillus reuteri," *Int. J. Food Microbiol.*, vol. 272, no. August 2017, pp. 12–21, 2018.

[41] T. L. Talarico, L. T. Axelsson, J. Novotny, M. Fiuzat, and W. J. Dobrogosz, "Utilization of Glycerol as a Hydrogen Acceptor by Lactobacillus reuteri: Purification of 1,3-Propanediol:NAD Oxidoreductase.," *Appl. Environ. Microbiol.*, vol. 56, no. 4, pp. 943–948, Apr. 1990.

[42] D. D. Sriramulu *et al.*, "Lactobacillus reuteri DSM 20016 produces cobalamin-dependent diol dehydratase in metabolosomes and metabolizes 1,2-propanediol by disproportionation.," *J. Bacteriol.*, vol. 190, no. 13, pp. 4559–4567, 2008.

[43] K. A. Lindlbauer, H. Marx, and M. Sauer, "3-Hydroxypropionaldehyde production from crude glycerol by Lactobacillus diolivorans with enhanced glycerol uptake," *Biotechnol. Biofuels*, vol. 10, no. 1, pp. 1–11, 2017.

[44] L. Chen, P. D. Bromberger, G. Nieuwenhuiys, and R. Hatti-Kaul, "Redox balance in Lactobacillus reuteri DSM20016: Roles of iron-dependent alcohol dehydrogenases in glucose/glycerol metabolism," *PLoS One*, vol. 11, no. 12, pp. 1–20, 2016.

[45] F. Santos *et al.*, "The complete coenzyme B12biosynthesis gene cluster of Lactobacillus reuteri CRL 1098," *Microbiology*, vol. 154, no. 1, pp. 81–93, 2008.

[46] J. M. Landete, S. Ferrer, V. Monedero, and M. Zúñiga, "Malic enzyme and malolactic enzyme pathways are functionally linked but independently regulated in Lactobacillus casei BL23.," *Appl. Environ. Microbiol.*, vol. 79, no. 18, pp. 5509–5518, Sep. 2013.

[47] M. G. Gänzle, N. Vermeulen, and R. F. Vogel, "Carbohydrate, peptide and lipid metabolism of lactic acid bacteria in sourdough," *Food Microbiol.*, vol. 24, no. 2, pp. 128–138, Apr. 2007.

[48] N. N. Gandhi, P. F. Cobra, J. L. Steele, J. L. Markley, and S. A. Rankin, "Lactobacillus demonstrate thiol-independent metabolism of methylglyoxal: Implications toward browning prevention in Parmesan cheese.," *J. Dairy Sci.*, vol. 101, no. 2, pp. 968–978, Feb. 2018.

[49] N. E. Altaras and D. C. Cameron, "Metabolic engineering of a 1,2-propanediol pathway in Escherichia coli," *Appl. Environ. Microbiol.*, vol. 65, no. 3, pp. 1180–1185, 1999.

[50] K. Yamada and Y. Tani, "Glycerol dehydrogenase and dihydroxyacetone reductase of a methylotrophic yeast, Hansenula ofunaensis.," *Agric. Biol. Chem.*, vol. 52, no. 3, pp. 711–719, 2011.

[51] F. Santos *et al.*, "Pseudovitamin B12 is the corrinoid produced by Lactobacillus reuteri CRL1098 under anaerobic conditions," *FEBS Lett.*, vol. 581, no. 25, pp. 4865–4870, 2007.

[52] A. M. Feist and B. O. Palsson, "The Biomass Objective Function," *Curr Opin Microbiol.*, vol. 13, no. 3, pp. 344–349, 2010.

[53] X. T. Liu, C. L. Hou, J. Zhang, X. F. Zeng, and S. Y. Qiao, "Fermentation conditions influence the fatty acid composition of the membranes of Lactobacillus reuteri I5007

and its survival following freeze-drying," *Lett. Appl. Microbiol.*, vol. 59, pp. 398–403, 2014.

[54] J. Walter *et al.*, "D-Alanyl ester depletion of teichoic acids in Lactobacillus reuteri 100-23 results in impaired colonization of the mouse gastrointestinal tract," *Environ. Microbiol.*, vol. 9, no. 7, pp. 1750–1760, 2007.

[55] P. Ksonzeková *et al.*, "Exopolysaccharides of Lactobacillus reuteri : Their influence on adherence of E . coli to epithelial cells and inflammatory response," *Carbohydr. Polym.*, vol. 141, pp. 10–19, 2016.

[56] M. Dauner, T. Storni, and U. W. E. Sauer, "Bacillus subtilis Metabolism and Energetics in Carbon-Limited and Excess-Carbon Chemostat Culture," *J. Bacteriol.*, vol. 183, no. 24, pp. 7308–7317, 2001.

[57] M. Dauner and U. Sauer, "Stoichiometric Growth Model for Riboflavin-Producing Bacillus subtilis," *Biotechnol. Bioeng.*, vol. 76, no. 2, pp. 132–143, 2001.

[58] D. W. Tempest and O. Neijssel, "The status of YATP and maintenance energy as biologically interpretable phenomena," *Annu. Rev. Microbiol.*, vol. 38, pp. 459–486, 1984.

[59] Y.-S. Jang *et al.*, "Bio-based production of C2–C6 platform chemicals," *Biotechnol. Bioeng.*, vol. 109, no. 10, pp. 2437–2459, 2012.

[60] T. Walther and J. M. François, "Microbial production of propanol," *Biotechnol. Adv.*, vol. 34, no. 5, pp. 984–996, 2016.

[61] D. Siebert and V. F. Wendisch, "Metabolic pathway engineering for production of 1,2-propanediol and 1-propanol by Corynebacterium glutamicum," *Biotechnol. Biofuels*, vol. 8, no. 1, pp. 1–13, 2015.

[62] R. Jain, X. Sun, Q. Yuan, and Y. Yan, "Systematically Engineering Escherichia coli for Enhanced Production of 1,2-Propanediol and 1-Propanol," *ACS Synth. Biol.*, vol. 4, no. 6, pp. 746–756, 2015.

[63] A. Chowdhury and C. D. Maranas, "Designing overall stoichiometric conversions and intervening metabolic reactions," *Sci. Rep.*, vol. 5, pp. 1–20, 2015.

[64] M. Papagianni, "Metabolic engineering of lactic acid bacteria for the production of industrially important compounds," *Comput Struct Biotechnol J.*, vol. 3, no. October, pp. 1–8, 2012.

[65] M. Sauer, H. Russmayer, R. Grabherr, C. K. Peterbauer, and H. Marx, "The Efficient Clade: Lactic Acid Bacteria for Industrial Chemical Production," *Trends Biotechnol.*, vol. 35, no. 8, pp. 756–769, 2017.

[66] M. Papagianni and M. Legiša, "Increased mannitol production in Lactobacillus reuteri ATCC 55730 production strain with a modified 6-phosphofructo-1-kinase," *J. Biotechnol.*, vol. 181, pp. 20–26, 2014.

[67] E. Rowe, B. O. Palsson, and Z. A. King, "Escher-FBA : a web application for interactive flux balance analysis," *BMC Syst. Biol.*, vol. 12, no. 84, pp. 1–7, 2018.

Paper 2

Engineering the carotenoid biosynthetic pathway in *Rhodothermus marinus* for lycopene production

Thordis Kristjansdottir(1,2,^,*), Emanuel Y. C. Ron(3,^), Daniel Molins-Delgado(4), Olafur H. Fridjonsson(1), Charlotta Turner(4), Snaedis H. Bjornsdottir(5), Steinn Gudmundsson(2), Ed W.J. van Niel(6), Eva Nordberg Karlsson(3), Gudmundur O. Hreggvidsson(1,5)

(1) Matis, Vinlandsleid 12, 113. Reykjavik, Iceland

(2) University of Iceland, Department of Computer Science, School of Engineering and Natural Sciences, Dunhagi 5, 107 Reykjavik, Iceland

(3) Lund University, Department of Chemistry, Division of Biotechnology, Box 124, 221 00 Lund, Sweden

(4) Lund University, Department of Chemistry, Centre for Analysis and Synthesis, Box 124, 221 00 Lund, Sweden

(5) University of Iceland, Department of Biology, School of Engineering and Natural Sciences, Sturlugata 7, 102 Reykjavík, Iceland

(6) Lund University, Department of Chemistry, Division of Applied Microbiology, Box 124, 221 00 Lund, Sweden

(*) Corresponding author, thordis@matis.is

(^) These authors contributed equally to this work.

Abstract

Rhodothermus marinus has the potential to be well suited for biorefineries, as an aerobic thermophile that produces thermostable enzymes and is able to utilize polysaccharides from different 2nd and 3rd generation biomass. The bacterium produces valuable chemicals such as carotenoids. However, the native carotenoids are not established for industrial production and *R. marinus* needs to be genetically modified to produce higher value carotenoids. Here we genetically modified the carotenoid biosynthetic gene cluster resulting in three different mutants, most importantly the lycopene producing mutant TK-3 ($\Delta trpB\Delta purA\Delta cruFcrtB::trpBcrtB_{T.thermophilus}$). The genetic modifications and subsequent structural analysis of carotenoids helped clarify the carotenoid biosynthetic pathway in *R. marinus*. The nucleotide sequences encoding the enzymes phytoene synthase (CrtB) and the previously unidentified 1',2'-hydratase (CruF) were found fused together and encoded by a single gene in *R. marinus*. Deleting only the *cruF* part of the gene did not result in an active CrtB enzyme. However, by deleting the entire gene and inserting the *crtB* gene from *Thermus thermophilus*, a mutant strain was obtained, producing lycopene as the sole carotenoid. The lycopene produced by TK-3 was quantified as 0.49 g/kg CDW (cell dry weight).

Keywords: *Rhodothermus marinus*, Carotenoids, Lycopene, Genetic engineering.

1. Introduction

Rhodothermus is the type genus of the phylum *Rhodothermaeota* [1]. The aerobic bacterium *R. marinus* is thermophilic and moderately halophilic, growing optimally at 65°C and 1-2% NaCl [2]. The bacterium has potential for biorefinery applications, as it possesses many features of importance for industrial bioconversion of recalcitrant 2nd and 3rd generation feedstock. This includes various biomass-degrading enzymes, as well as the capacity to produce interesting primary and secondary anabolic compounds, such as carotenoids [3] and exopolysaccharides (EPS) [4]. In addition, as a robust extremophile, *R. marinus* is adapted to growth at high temperatures which may be beneficial for biorefining of 2nd and 3rd generation feedstocks. High temperature increases the solubility of polysaccharides used as carbon sources and reduces the viscosity of fermentation broths. Consequently, it enables higher feedstock loads and facilitates enzymatic access to polysaccharides. In addition, growth at high temperatures in bioreactors mitigates scale-up problems of mixing and

aeration, reduces costs of cooling, distillation and extraction and minimizes the danger of contamination of spoilage bacteria [5].

To generate efficient biorefinery production strains, metabolic engineering is often required to modify the production profile of the microorganism. A few thermophiles have to date been the subject of metabolic engineering efforts, mainly anaerobic thermophiles for production of biofuels (and/or commodities), with encouraging results [6]. Tools for genetic engineering of the thermophilic aerobe *R. marinus* have been developed, and genes have both been heterologously expressed [7] and deleted from the genome [8].

Selection of a chassis species for metabolic engineering of pathways for production of compounds of industrial interest should be based on the metabolic characteristics of the organism, including both the substrate utilization range required for a particular feedstock, and the synthetic capabilities or potential. Anaerobic fermentative species are generally preferred for production of lower value commodity chemicals, such as simple organic acids and alcohols, which are typically waste products from catabolic metabolism. However, aerobes can carry a greater metabolic burden and are the organisms of choice for the heterotrophic production of complex secondary metabolites. This includes carotenoids, which are pigments, naturally produced by many plants, algae and bacteria. Currently, over 1100 carotenoid structures from more than 600 organisms are known [9]. They are widely used as colorants and additives in the food, feed and cosmetic industries. Numerous studies have shown that carotenoids have potential health benefits, both in preventing and treating various diseases [10], [11]. However, the biological functions of carotenoids are complex and other studies have reported conflicting results [12]. Nevertheless, the global market for carotenoids has reached USD 1.5 billion and is predicted to continue growing [13].

R. marinus natively produces γ-carotenoids and their structures have been characterized [3], [14]. Four variations of carotenoid glucoside esters were demonstrated, including that of salinixanthin. These native carotenoids are derived from γ-carotene, while most of the carotenoids of industrial interest are β-carotenoids. Astaxanthin, β-carotene, canthaxanthin, zeaxanthin, lycopene and lutein are in highest demand [13].

We previously carried out a bioinformatic analysis of the carotenoid biosynthetic pathway in *R. marinus* [3]. The feasibility of modifying the pathway has also been demonstrated in a

prior study, where the genes encoding the pytoene desaturase (CrtI) and phytoene synthase (CrtB) enzymes of the pathway were deleted, resulting in the colorless *R. marinus* strain SB-71 [8]. The present study aims to clarify the function of the genes involved in the *R. marinus* carotenoid biosynthesis, and to engineer the pathway for lycopene production. Lycopene is one of the more demanded carotenoids on the market, used in cosmetics, pharmaceuticals and as a food coloring agent, and it is also the common precursor of γ- and β-carotenoids. For this purpose and to illustrate the potential *of R. marinus* as a biorefinery strain, lycopene production is an ideal first target for metabolic engineering.

2. Methods and materials

2.1 Strains, plasmids, media and culture conditions

The strains and plasmids used and generated in this study are listed in Table 1. For genetic modifications of *R. marinus*, strain SB-62 (Δ*trpB*Δ*purA*) [8] was used. Tryptophan selection was used in all cases, leaving the possibility of adenine complementation for further genetic modifications of the strains in the future. Genomic DNA from *R. marinus* ISCAR-493 and *Thermus thermophilus* HB8 as well as the plasmid vector pUC19 [15] were used for generating recombinant molecules. NEB Stable *E. coli* cells (New England BioLabs) were used for molecular cloning. The pRM3000 plasmid [7] was used as a control in the *R. marinus* transformation experiments. All *R. marinus* cells were cultured at 65°C and the liquid cultures set to shaking at 200 rpm (New Brunswick Innova 4400 Incubator Shaker). Medium 162 [16] was used, with modifications (2 mM $MgSO_4$ and 0.2 mM $CaCl_2$ in final volume) and additions of 1% NaCl and 0.053% NH_4Cl. Two variations of the medium were used: A rich non-selective medium (R-medium) which contained 0.25% tryptone and 0.25% yeast extract and a selective agar medium (RS-medium), which contained 0.2% soluble starch, 0.2% casamino acids, vitamin solution [16], 0.25% adenine and 2.5% agar. *E. coli* cells were cultured at 37°C in LB-medium [17] with 100 μg/mL ampicillin.

Table 1. *Stains and plasmids used in this study.*

Strain or plasmid	Details	Origin or reference
NEB Stable *E. coli*	Competent, High Efficiency	New England BioLabs
T. thermophilus HB8	WT strain	Complete genome sequence available in GenBank (NC_006461.1)
R. marinus ISCAR-493	WT strain. Alternative names: PRI 493, MAT 493, DSM 16675	Matís, [18]
R. marinus SB-62	Trp^-Ade^-, ISCAR-493 derivative (Δ*trpB*Δ*purA*)	Matís, [8]
R. marinus SB-71	Trp^+Ade^-, colorless, SB-62 derivative (Δ*trpB*Δ*purAcrtBI'::trpB*)	Matís, [8]
R. marinus TK-1	Trp^+Ade^-, SB-62 derivative (Δ*trpB*Δ*purA*Δ*crtYO::trpB*)	This work
R. marinus TK-2	Trp^+Ade^-, SB-62 derivative (Δ*trpB*Δ*purA*Δ*cruF::trpB*)	This work
R. marinus TK-3	Trp^+Ade^-, SB-62 derivative (Δ*trpB*Δ*purA*Δ*cruFcrtB::trpBcrtB*$_{T.thermophilus}$)	This work
pUC19	Amp^R general cloning vector for *E. coli*	[15]
pRM3000	*trpB*$^+$ *R. marinus-E. coli* shuttle vector	Matís, [7]

For the analysis of carotenoids, *R. marinus* strains TK-1, TK-2, TK-3 and SB-71 were grown on agar plates with modified medium 162 supplemented with 1% NaCl and 0.025 g/L adenine. After 48 h at 65 °C, the colonies were transferred to 5 mL liquid LB-medium, supplemented with 0.025 % adenine, and cultivated at 65 °C for 24 h in 50 mL falcon tubes. The cells were then cultivated in baffled shake flasks at 200 rpm increasing the cultivation volume from 50 mL to 100 mL sequential cultivations, maintaining 10 % (v/v) inoculation volume. The harvested cells were then split and transferred to two 500 mL bioreactors (Multifors 2, Infors) for cultivation with the following parameters: 65 °C, pH 7, 1 VVM aeration, 200 rpm stirrer rate cascaded with dissolved oxygen tension (DOT) that was set to 40%. The cultivation was terminated at the onset of the stationary phase, indicated by a sudden increase of DOT, at which the cell culture was set to cool at 10 °C. The cells were harvested

by centrifugation at 4 °C at 10,000 g for 10 min, after which the pellets were frozen at -80 °C and lyophilized until constant weights were reached.

2.2 Design and generation of cloning molecules

Recombinant DNA molecules were designed using the plasmid vector pUC19 and amplified regions of the genome sequences of *R. marinus* ISCAR-493 (Matís, unpublished) and *T. thermophilus* HB8 (NC_006461.1). DNA was isolated from the strains using the MasterPure Complete DNA purification Kit (Lucigen). Primers (Table S1) were designed for the amplification of ~1500 bp 5' and 3' flanking regions of the genes targeted for deletion as well as the *trpB* gene (selection marker) and the *crtB* gene from *T. thermophilus*. The primers were designed to support HiFi DNA assembly (NEBuilder HiFi DNA Assembly Master Mix, New England BioLabs). The primers Gene5_5_F, Gene5_5_R (5' flanking region), Gene1_3_F, Gene1_3_R (3' flanking region), Gene5_Gene14_F and Gene5_Gene14_R (*trpB* gene) were used to amplify the fragments for the gene deletion cassette used to obtain strain TK-1 (Δ*trpB*Δ*purA*Δ*crtYO::trpB*). The primers Gene9a_5_F, Gene9a_5_R (5' flanking region), Gene9a_3_F, Gene9a_3_R (3' flanking region), Gene9a_Gene14_F and Gene9a_Gene14_R (*trpB* gene) were used to amplify the fragments for the gene deletion cassette used to obtain strain TK-2 (Δ*trpB*Δ*purA*Δ*cruF::trpB*). The primers Gene9_5_F, Gene9_5_R (5' flanking region), Gene9_3_F, Gene9_3_R (3' flanking region), Gene9_Gene13_F, Gene9_Gene13_R (*crtB* gene from *T. thermophilus*), Gene9_Gene14_F and Gene9_Gene14_R (*trpB* gene) were used to amplify the fragments for the gene deletion/insertion cassette used to obtain strain TK-3 (Δ*trpB*Δ*purA*Δ*cruFcrtB::trpBcrtB*$_{T.thermophilus}$). All following enzymes, kits and cells in this section were obtained from New England BioLabs. Q5 High-Fidelity DNA polymerase was used in all amplifications according to the manufacturer´s instructions. The correct sizes of the resulting fragments were verified by electrophoresis and they were subsequently purified from gel using the Monarch DNA Gel Extraction Kit. The fragments were assembled into *Sma*I restricted pUC19 vector and introduced into competent *E. coli* cells (NEB stable) by chemical transformation. Positive clones were confirmed by verifying the presence of the *trpB* gene by PCR, using *Taq* DNA polymerase and the primers Gene14_Verify_F and Gene14_Verify_R (Table S1). Vectors were isolated from positive clones using the Monarch Plasmid Miniprep Kit. Linear inserts for *R. marinus* transformation for the construction of strains TK-1 and TK-2 were obtained by PCR, using the Q5 High-Fidelity DNA polymerase and

the primers Gene5_5_F and Gene1_3_R, and TK2_Total_F and TK2_Total_R (Table S1), respectively. A linear insert for *R. marinus* transformation for the construction of strain TK-3 was obtained by digesting the plasmid with *Xba*I and *Kpn*I-HF.

2.3 Transformation of *R. marinus*

R. marinus cells were prepared for electroporation as described elsewhere [18]. The electroporation protocol was followed, using the GenePulser Xcell electroporation system (Bio-Rad) with pulses delivered at 20 kV/cm. 1 µg of DNA was used per transformation, in ≤ 5 µL and mixed carefully with 40 µL of washed cells. Negative (sterile MilliQ water) and positive (pRM3000) controls were included. Transformed cells were grown on selective agar plates (without tryptophan) for 3-5 days.

2.4 Verification of genotypes using PCR and sequencing

Positive *R. marinus* clones were verified by PCR, using *Taq* DNA polymerase. Modifications of strain TK-1 were verified using primers Gene7_Verify_F, Gene7_Verify_R, Gene5_Verify_F, Gene5_Verify_R, Gene4_Verify_F, Gene4_Verify_R, Gene14_Verify_F and Gene14_Verify_R (Table S1). Modifications in strains TK-2 and TK-3 were verified using the primers Gene9a_Verify_F, Gene9a_Verify_R, Gene9b_Verify_F, Gene9b_Verify_R, Gene13_Verify_F, Gene13_Verify_R, Gene14_Verify_F and Gene14_Verify_R (Table S1). Additionally, the modifications were confirmed by sequencing. Several PCR reactions were performed for strains SB-62 and TK-1 using the primers Gene1_Seq_F, Gene1_Seq_R, Gene4_Seq_F, Gene4_Seq_R, Gene5_Seq_F and Gene5_Seq_R (Table S1) and Q5 DNA polymerase. The resulting amplicons were sequenced using the ABI3730 system (Applied Biosystems, Thermo Fisher Scientific). For strain TK-2, Q5 DNA polymerase and primers Gene9a_Seq_F and Gene9a_Seq_R (Table S1) were used to amplify a 5188 bp region. For strain TK-3, Q5 DNA polymerase and primers Gene9_Seq_F and Gene9_Seq_R (Table S1) were used to amplify a 4686 bp region. Sequencing libraries were made from the amplicons using the Nextera DNA (TK-2) and Nextera Flex (TK-3) methods (Illumina) and libraries were sequenced on the Illumina MiSeq sequencing platform using the V3 2x 300 cycle chemistry. Obtained fragments were assembled using Geneious.

2.5 Carotenoid extraction

Carotenoids were extracted from *R. marinus* strains TK-1, TK-2, TK-3, SB-71 and ISCAR-493 using two methods. In the first method, aqueous cell suspensions were sonicated in an ice bath for 5 × 2 min, with one-minute rests in-between in order to keep cold sample conditions. After sonication the samples were mixed with ethyl acetate (1:1). In the second method, lyophilized cells were powdered with a glass rod before mixing with dichloromethane (25 mL solvent per g freeze-dried cells). All organic phase extracts were vacuum filtered before being dried by rotary evaporation (Heidolph instruments) at 80 rpm at 40 °C. The extracts were reconstituted in 3 mL of dichloromethane, filtered through a 0.2 µm PTFE syringe filter, flushed with N_2 gas and stored at -80 °C until mass spectrometry analysis.

Carotenoids were additionally extracted from strains TK-1, TK-2, TK-3, SB-71 and ISCAR-493, for absorbance spectra analysis, by mixing full loops of cells from agar plates with hexane:acetone (1:1) and sonicating in a bath for 20 min. This was done in triplicates.

2.6 Carotenoid analysis

The carotenoid extracts were analyzed by means of supercritical fluid chromatography – mass spectrometry (MS) in an Ultra Performance Convergence Chromatography system (UPC^2) coupled to a quadrupole – orthogonal acceleration time-of-flight tandem mass spectrometer (XEVO-G2 Q-TOF) with an electrospray ion source, both instruments from Waters (Mildford, MA, USA). Both systems were controlled, and all data was analyzed, with MassLynx™ (v 4.1, SCN 77; Waters). The chromatographic method used was based on a modification of the method described in [19]. Briefly, the chromatographic separation was achieved using Acquity UPC^2 Torus 1-Aminoanthracene (100 mm x 3 mm, 1.7 µm) column from Waters. The mobile phase consisted of (A) CO_2 and (B) methanol in a gradient elution analysis programmed as follows: 0-0.5 min, 5% (B); 0.5-2.5 min, 5-15% (B); 2.5-8 min, 15% (B); 8-9 min, 15-5% (B); and 9-10 min, 5% (B). The flow rate was 1.5 ml min^{-1}, column temperature was kept at 56 °C and backpressure at 160 bar. Ammonium formate (10 mM) in methanol was used as a make-up solvent at a flowrate of 0.5 ml min^{-1}. The total run-time of the program was 10 min.

UV-Vis spectra were recorded in the range of 200-500 nm by the diode array detector (DAD). The electrospray ionization (ESI) ion source was operated in positive mode and full-scan MS spectra were obtained by scanning the range m/z 50 - 1000. The mass spectrometer was calibrated with sodium formate. Centroid mode data was collected after mass correction during acquisition using an external reference comprising of 10 µL/min solution of leucine-enkephalin (2 ng/µL). The capillary and cone voltage were set at 3 kV and 40 V, respectively. Nitrogen was used as both cone gas (50 L/h) and desolvation gas (800 L/h). The source and desolvation temperature were set at 150 and 300°C, respectively. Simultaneous acquisition of exact mass at high and low collision energy, MS^E (where E represents collision energy), was used to obtain full scan accurate mass fragment, precursor ion, and neutral loss information. The collision energy in function 1 (low energy) was off while in function 2 (high energy) and the collision energy ranged between 15 to 60 V. MS/MS analysis of the carotenoids was performed with the quadropole set at m/z 536 in MS^E mode.

The absorbance spectra (350-600 nm) of the carotenoids of strains TK-1, TK-2, TK-3, SB-71 and ISCAR-493, extracted with hexane:acetone (1:1), were analyzed in a benchtop spectrophotometer (MULTISKAN Sky, Thermo Scientific), using 1 cm cuvette. This was done to investigate the profile of the spectra and A_{max} values of obtained peaks.

Lyophilized TK-3 cells from shake flask cultivations were weighed before organic solvent extraction and reconstitution in chloroform. The extracts were then analyzed by a spectrophotometer at 485 nm. The lycopene concentration was calculated using Beer's law with the molar attenuation coefficient of 150855 L mol^{-1} cm^{-1} for lycopene in chloroform [20]. Lycopene, α-, β- and γ-carotene standards were of analytical grade quality (Supelco).

3. Results

3.1 Bioinformatic analysis of genes encoding carotenoid biosynthetic enzymes

Several strains of the thermophilic bacterium *R. marinus* produce carotenoids derived from γ-carotenoid. Previous gene homology searches for known carotenoid biosynthetic genes using the genomic data of the type strain *R. marinus* DSM 4252^T, resulted in the discovery of genes encoding homologues to enzymes in the carotenoid biosynthesis pathways from other species. This allowed the identification of a gene cluster, with two adjacent operons, one small and one larger [3], [8]. In this work, the bioinformatic analysis of genomic sequence

data was expanded, enabling further identification of the enzymes in the biosynthetic pathway for carotenoid production in *R. marinus* (Figure 1).

Here, we report the identification of the *cruF* gene in the gene cluster (Figure 2), which likely acts as a 1',2'-hydratase [21]. This gene was not identified in the previous work because it is fused together with the phytoene synthase (*crtB*) gene (Figure 2A, gene 9) and together they were annotated as a phytoene synthase. These genes are usually separate, while *cruF* in *R. marinus* is without a stop codon and fused with the downstream *crtB* in one open reading frame (ORF). The fused genes, along with the carotenoid genes acyltransferase (*cruD*) and phytoene desaturase (*crtI*) (Figure 2A, genes 9, 11 and 12, respectively) are located in the smaller operon in the gene cluster. Two additional genes are also present in the smaller operon, encoding products showing homology to hypothetical proteins (gene 10) and a MerR family transcriptional regulator from *T. thermophilus* (31% similarity) (gene 8).

Two genes in the larger operon were also identified based on sequence homology, with high sequence similarity to the known carotenoid genes encoding lycopene cyclase (*crtY*) (50%) (Figure 2A, gene 2) and carotene ketolase (*crtO*) (60%) (Figure 2A, gene 3). Five other genes (Figure 2A, genes 1, 4-7) were found in this larger operon, but since they lack sequence homology to known genes in carotenoid biosynthetic pathways, their role in the pathway remains unknown. They were annotated as hypothetical protein (gene 1), NAD dependent epimerase (gene 4), FAD dependent oxidoreductase (gene 5), which showed slight homology (36% similarity) to *crtI*, deoxyribodipyrimidine photo-lyase (gene 6) and short-chain dehydrogenase/reductase (gene 7). Several known carotenoid biosynthetic genes belong to the *crtI* gene family, including *crtO* (gene 3, above) and also 3',4'-desaturase, an enzyme for which the corresponding gene is not identified in *R. marinus*. The native structures of the carotenoids in *R. marinus* suggest that a 3',4'-desaturase is a part of the biosynthetic pathway, but since the *crtI* homology of gene 5 is low and there is no other evidence so far that it encodes a 3',4'-desaturase, this step in the pathway is still labeled as unknown (Figure 1).

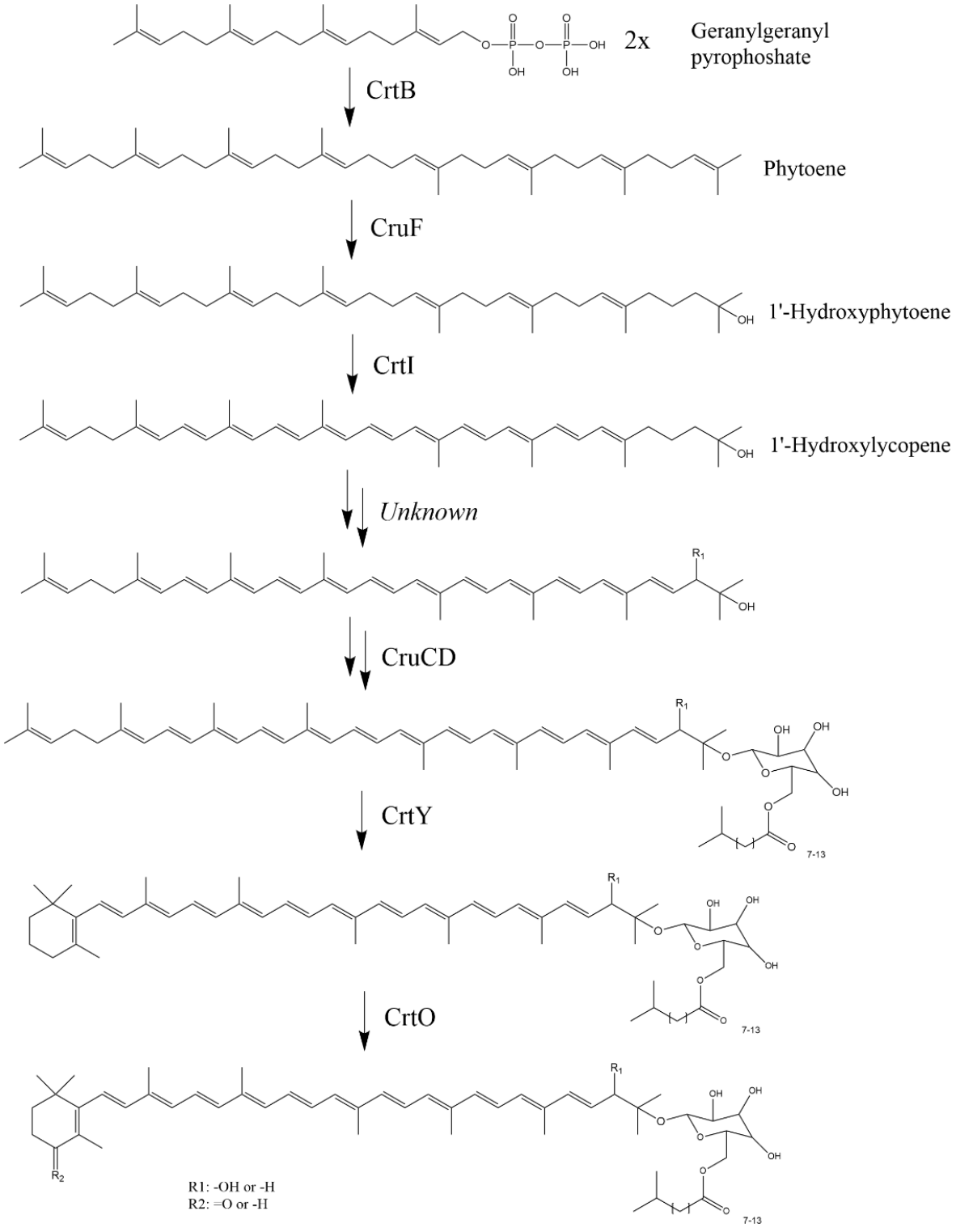


Figure 1. *Proposed carotenoid biosynthetic pathway in R. marinus, based on [3], with added information obtained in this study. Molecular structures and abbreviated enzyme names are shown, phytoene synthase (CrtB), 1',2'-hydratase (CruF), phytoene desaturase (CrtI), glycosyltransferase (CruC), acyltransferase (CruD), lycopene cyclase (CrtY) and carotene ketolase (CrtO).*

3.2 Genetic modifications to alter the carotenoid production

Three different genetic modifications were performed in the carotenoid gene cluster of *R. marinus* strain SB-62 (Δ*trpB*Δ*purA*) (Figure 2). The target genes were deleted and replaced with the selective marker *trpB* (encoding the tryptophan synthase beta chain), using double crossover homologous recombination using linear insertion cassettes. In the first modification, a 5890 bp region was deleted from the larger operon, resulting in the *R. marinus* mutant TK-1 (Δ*trpB*Δ*purA*Δ*crtYO::trpB*) (Figure 2B). This region includes the genes *crtY* (gene 2) and *crtO* (gene 3) which encode the enzymes responsible for the 4-keto β-ionone ring modifications displayed on the left side of the carotene backbone (Figure 1), as well as gene 5, where the deduced amino acid sequence is showing slight homology to *crtI* (see section 3.1). The other two modifications involved knocking out the *cruF* gene (1',2'-hydratase), which encodes one of the enzymes that modify the right side of the carotene backbone (Figure 1). In the first *cruF* modification, only the *cruF* part of the gene was deleted, leaving the *crtB* part intact, resulting in the *R. marinus* mutant TK-2 (Δ*trpB*Δ*purA*Δ*cruF::trpB*) (Figure 2C). The *crtB* part of the gene includes a start codon, which raises the question if it could encode a functional enzyme without the *cruF* part. In the second *cruF* modification, both the *cruF* and *crtB* parts of the gene were deleted and the *crtB* gene from *Thermus thermophilus* strain HB8 was inserted, resulting in the *R. marinus* mutant TK-3 (Δ*trpB*Δ*purA*Δ*cruFcrtB::trpBcrtB*$_{T.thermophilus}$) (Figure 2D). This modification was designed in case the native CrtB enzyme in TK-2 would be non- or low functioning after the deletion of the *cruF* part.

Successful genetic modifications were verified by amplifying target regions from the genomes of putative mutants, and sequencing. PCR reactions amplifying several genes in the cluster (genes 1, 4, 5, 7, 9a (*cruF*), 9b (*crtB*), 13 and 14 in Figure 2), using primers designed to bind inside each gene, were performed for all three mutants and strain SB-62. For strain SB-62 all reactions resulted in the expected sized amplicons, except for the *crtB* gene from *T. thermophilus* and the *trpB* gene (Figure 2A). The presence of an amplicon using primers for amplifying the *crtB* gene from *T. thermophilus* in *R. marinus* was not expected. However, in all *R. marinus* strains the same sized unknown amplicon (~300 bp) was obtained and was subsequently sequenced. This showed that the amplification occurred from a gene on the chromosome distant from the carotenoid gene cluster (Figure S9). This gene did not show

homology to *crtB* from *T. thermophilus*. The primers used to amplify the *crtB* gene from *T. thermophilus* did however show homology to this region, which explains the unexpected amplification. The *trpB* gene was not amplified in strain SB-62 which was to be expected since it has the *trpB* gene deleted, enabling us to use it as a selective marker.

For strain TK-1 (Δ*trpB*Δ*purA*Δ*crtYO::trpB*), the PCR results showed two main differences compared to strain SB-62 (Figure 2B). First, genes 1, 4 and 5 were not amplified, which indicated the successful deletion of the region. Second, the reaction for the *trpB* gene (gene 14) did result in the expected sized amplicon, indicating the insertion of the selection marker. To further verify this, additional PCR reactions (a – f in Figure 2) were performed for strains SB-62 and TK-1 and the products were sequenced (Figures S5 and S6).

For strains TK-2 (Δ*trpB*Δ*purA*Δ*cruF::trpB*) and TK-3 (Δ*trpB*Δ*purA*Δ*cruFcrtB::trpBcrtB*$_{T.thermophilus}$), the *cruF* part of gene 9 was not amplified, while gene 14 was, indicating the expected substitution of *cruF* for *trpB*. Additionally, an amplicon of the expected size for gene 13 (*crtB* from *T. thermophilus*) was observed in strain TK-3. An identical amplicon was observed from *T. thermophilus* (Figure S1). The native *crtB* gene (second half of gene 9) was deleted from strain TK-3 and the amplicon seen in strain SB-62 (~600 bp) was not amplified in strain TK-3. However, a smaller unknown amplicon (~400 bp) was obtained and was subsequently sequenced. The results showed that the amplification did not occur from the carotenoid gene cluster (Figure S8), but in a gene on the chromosome distant from the cluster. This gene did not show homology to the *crtB* gene, but the primers used to amplify *crtB* in *R. marinus* did show homology to this region, which explains the unexpected amplification. To further verify the successful modifications in strains TK-2 and TK-3, additional PCR reactions were performed (g in Figure 2) and the resulting products were sequenced (Figures S10 and S11).

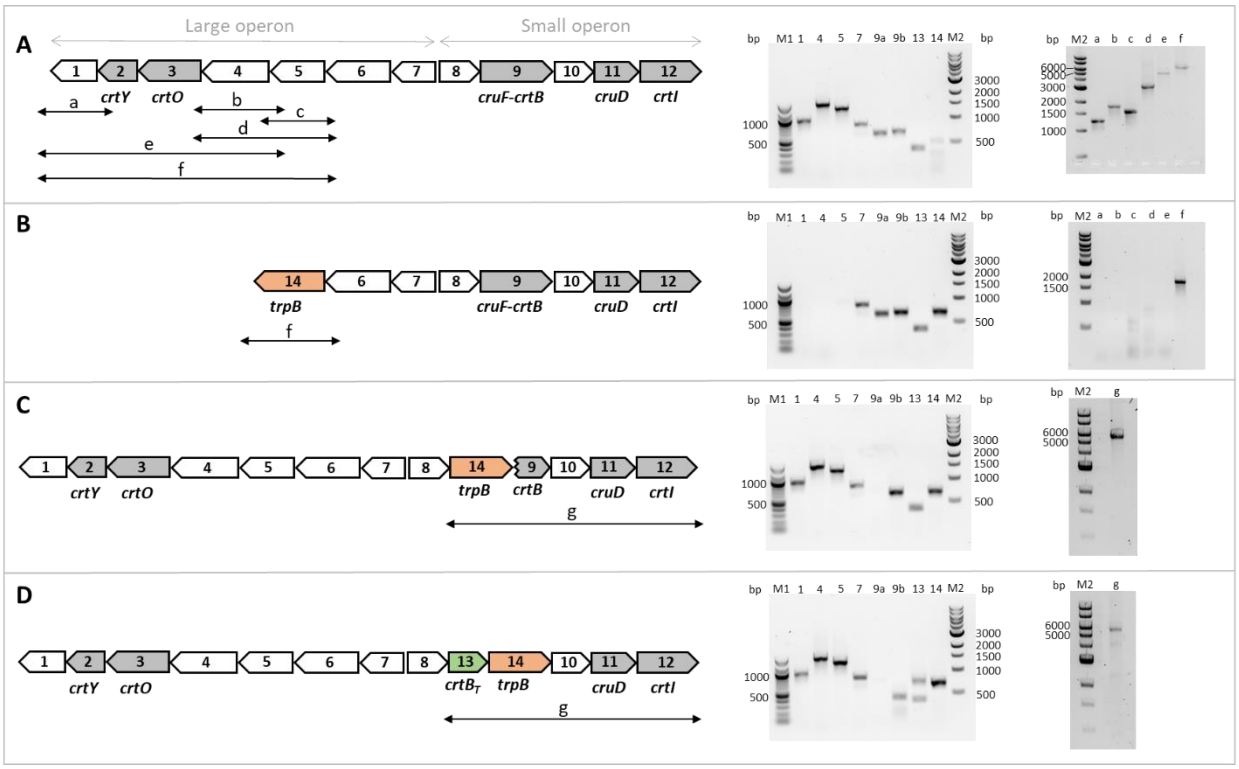


Figure 2. *The carotenoid gene cluster in the R. marinus stain SB-62 (ΔtrpBΔpurA)* (***A***) *and three different genetically modified strains: mutant TK-1 (ΔtrpBΔpurAΔcrtYO::trpB)* (***B***)*, mutant TK-2 (ΔtrpBΔpurAΔcruF::trpB)* (***C***) *and mutant TK-3 (ΔtrpBΔpurAΔcruFcrtB::trpBcrtB$_{T.thermophilus}$)* (***D***)*. Genes involved in the carotenoid biosynthesis are colored grey. Gene deletions and insertions were verified using PCR (1-14) and the sequencing of PCR products (a-g). Pictures showing gel electrophoresis following PCR are shown to the right side of the gene clusters. Lane M1: 100 bp DNA ladder. Lane M2: 1 kb DNA ladder. Unedited electrophoresis pictures and sequencing results are shown in the supplementary file (Figures S1-S11).*

3.3 Analysis of carotenoids by UHPSFC-DAD-QTOF/MS and spectrophotometry

Absorbance spectra of carotenoids from strains TK-1 (Δ*trpB*Δ*purA*Δ*crtYO::trpB*), TK-3 (Δ*trpB*Δ*purA*Δ*cruFcrtB::trpBcrtB*$_{T.thermophilus}$) and ISCAR-493 extracted with hexane:acetone (1:1), showed A_{max} for strain ISCAR-493 at 478 nm, for TK-1 at 488 nm and for TK-3 at 472 nm, with an additional peak at 502 nm (Figure 3 and Figures S17-S19). The spectra for TK-3 is identical to previously published spectra for lycopene [22]. Absorbance between 350 – 600 nm was neither observed for strain TK-2 (Δ*trpB*Δ*purA*Δ*cruF::trpB*) nor SB-71 (Δ*trpB*Δ*purAcrtBI'::trpB*).

Extracts from *R. marinus* strains TK-1, TK-2, TK-3, SB-71 and ISCAR-493 were analyzed by UHPSFC-DAD-QTOF/MS. By DAD, it was not possible to detect any peaks between 400 and 600 nm for strains TK-2 and SB-71. Moreover, the mass spectrometer did not detect any masses corresponding to the native carotenoids nor masses corresponding to common C40

carotenoids. It was therefore concluded that TK-2 (Δ*trpB*Δ*purA*Δ*cruF::trpB*), like SB-71 (Δ*trpB*Δ*purAcrtBI'::trpB*), does not produce carotenoids or produces them in very low concentrations. This was expected for SB-71 since the genes *crtB* and *crtI* (Figure 1) were knocked-out [8]. For TK-2, these results indicated that the *crtB* part of the *cruF-crtB* gene (Figure 2, gene 9) does not encode an active enzyme without the *cruF* part.

Two of the native carotenoids were detected in the TK-1 (Δ*trpB*Δ*purA*Δ*crtYO::trpB*) extract. The first detected carotenoid (t_R 2.48) had a mass (m/z 910.672, Δ3.2 mDa, Δ3.5 ppm) matching the β,ψ-carotenoid acyl glucoside without the 2'-hydroxyl group previously identified in *R. marinus.* MS/MS fragmentation confirmed this by detection of the carotenoid fragment ($C_{40}H_{55}$) but not the 4-keto β-ionone ring fragment (m/z 203.1436) found in salinixanthin [3]. Moreover, the characteristic m/z 28 pattern, resulting from acyl group length variation, was detected for C11 (m/z 882.637, Δ2.4 mDa), C13 (m/z 910.672, Δ3.2 mDa), C15 (m/z 938.703, Δ2.9 mDa), and C17 (m/z 966.732, Δ0.8 mDa). The other detected carotenoid (t_R 2.83 min) in the TK-1 extract had a mass identical to that of the native 2'-hydroxyl version of the above mentioned β,ψ-carotenoid acyl glucoside (m/z 926.660 Δ4.1 mDa). This carotenoid followed a similar pattern of acyl length loss and MS/MS fragmentation as the native non-hydroxylated β,ψ-carotenoid acyl glucoside identified in a previous study [3]. Salinixanthin in both the 2'-hydroxyl and non-hydroxylated forms was detected in TK-1. These results suggest that all the native modifications displayed on the right side of the carotene backbone in Figure 1, including the C-3',4' double bond, are still present in TK-1. This refutes the hypothesis that gene 5 encodes a 3',4'-desaturase (see section 3.1). The results also showed that no keto group was present on the β-ring on the left side of the carotene backbone (Figure 1), confirming the deletion of the *crtO* gene. Carotenoids with a linear ψ-end and a β-ring have the exact same mass and the deletion of the *crtY* gene could therefore not be confirmed by mass spectrometry alone. However, the change from a β-ring and ψ-end should increase the wavelength of the absorption spectrum, due to the lower contribution of the β-ring to the bathochromic shift [23]. The absorption spectra of the carotenoids extracted from ISCAR-493 and TK-1 (Figure 3) show a clear shift towards higher wavelengths for the TK-1 carotenoids, suggesting the presence of a ψ-end and therefore the deletion of the *crtY* gene.

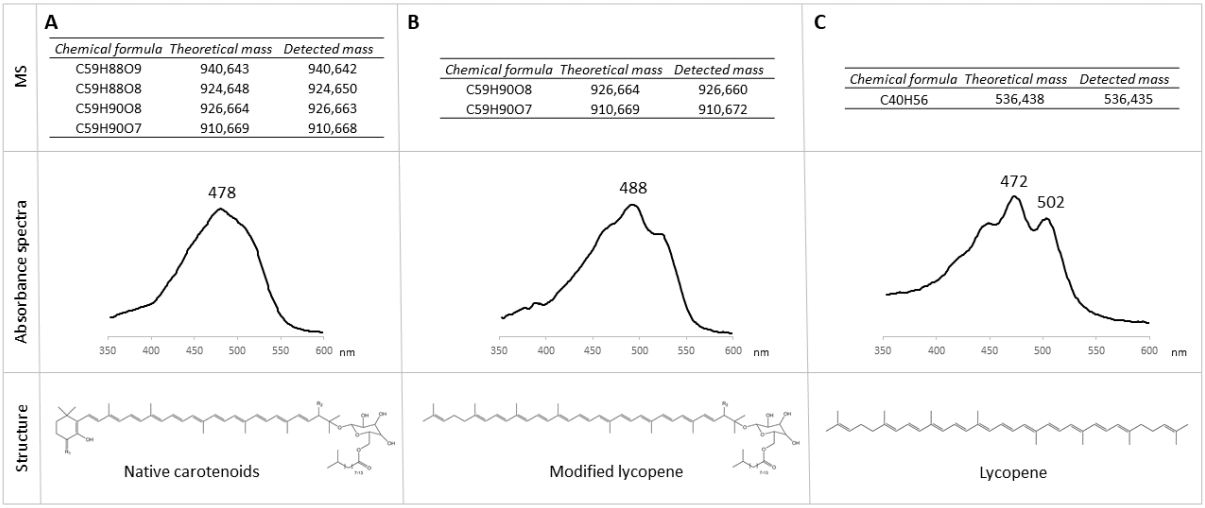


***Figure 3**. Carotenoid structures identified in R. marinus strain ISCAR-493 (**A**), mutant TK-1 (ΔtrpBΔpurAΔcrtYO::trpB) (**B**) and mutant TK-3 (ΔtrpBΔpurAΔcruFcrtB::trpBcrtB$_{T.thermophilus}$) (**C**). The top row shows chemical formulas of identified carotenoids, corresponding theoretical masses and detected masses by MS. The middle row shows absorbance spectra from 350-600 nm of carotenoids isolated with acetone:hexane (1:1) and A_{max} values of obtained peaks. Graphs with both axes are shown in Figures S17-S19. The bottom row shows the chemical structures of identified carotenoids: Native carotenoids (4-keto 2'-hydroxy β,ψ-carotene acyl glycoside), modified lycopene (2'-hydroxy ψ,ψ-carotene acyl glycoside) and lycopene. For more information on R_1 and R_2, refer to Figure 1.*

The extract of *R. marinus* strain TK-3 (Δ*trpB*Δ*purA*Δ*cruFcrtB::trpBcrtB*$_{T.thermophilus}$) showed a sole peak (t_R 2.75 min) by DAD ($\lambda_{450\ nm}$). This peak had a mass (m/z 536.428 ± 0.010) corresponding to $C_{40}H_{56}$ carotenoids, such as α-carotene, β-carotene, γ-carotene and lycopene. By comparing the t_R of the peak of TK-3 to these carotene standards, TK-3 had the same t_R as lycopene (Figures S12 and S13). In addition, MS/MS fragmentation spectra were analyzed. The carotenes γ-carotene and lycopene have acyclic ψ-ends, which can be determined by removal of an isoprene unit, resulting in a $[M-69]^{+\bullet}$ fragment of m/z 467.368 [24]. This fragment could be detected in the standards γ-carotene and lycopene but also in the TK-3 MS/MS spectrum at t_R 2.75 min. Moreover, absorption spectra of the TK-3 extract (Figure 3) shows an identical wavelength profile as established spectra for lycopene [22]. From these results it can be concluded that lycopene is produced by TK-3 as the sole product from the carotenoid biosynthetic pathway. Lycopene was quantified spectrophotometrically using Beer's law in TK-3 extracts in a separate shake flask experiment. The result was a yield of 0.49±0.01 g/kg cell dry weight (CDW), which corresponded to 0.14±0.004 mg/L cultivation volume. In conclusion, the results suggest that the *cruF-crtB* gene was successfully deleted from the genome of *R. marinus* and that the heterologous *crtB* gene from *T. thermophilus* encodes an active enzyme in strain TK-3. Additionally, and somewhat surprisingly, the only

active carotenoid enzymes in TK-3 seem to be CrtB and CrtI, leaving the other enzymes, such as CrtY, non-functional.

4. Discussion

Three successful modifications were performed in the carotenoid gene cluster of *R. marinus*, resulting in two carotenoid producing mutant strains. Based on the results obtained here, we propose the carotenoid biosynthesis pathway in *R. marinus*, which enzymes are involved and in what order they act. The mutant TK-1 (Δ*trpB*Δ*purA*Δ*crtYO::trpB*) produces a lycopene backbone with all native modifications (displayed on the right side in Figure 1) still present, but without the 4-keto β-ionone ring. The carotenoid genes *crtY* and *crtO*, which encode the enzymes that catalyze the keto ionone ring formation, were a part of the 5890 bp region from the larger operon that was deleted in TK-1. The remaining genes of the region apparently do not play a role in the biosynthetic pathway. The enzymes responsible for the modifications displayed on the right side of the backbone are mostly encoded by genes located in the smaller operon of the gene cluster (Figure 2). They are active without the keto ionone ring according to the structure analysis of the carotenoids obtained from TK-1. This means they are not dependent on the activity of CrtY and CrtO.

The results indicate that the mutant strain TK-3 (Δ*trpB*Δ*purA*Δ*cruFcrtB::trpBcrtB*$_{T.thermophilus}$) produces lycopene as the sole carotenoid. This mutant has the *cruF-crtB* gene (fused *cruF* and *crtB* genes) deleted from its genome and the *crtB* gene from *T. thermophilus* inserted. Deleting only the *cruF* part of the gene did not result in a carotenoid producing strain (TK-2). Apparently, the *crtB* part by itself does not result in the corresponding enzyme activity that is high enough to produce carotenoids at detectable levels. Fused carotenoid genes have been observed in other species, such as the *crtYB* gene in *Xanthophyllomyces dendrorhous*. Deletions in the lycopene cyclase domain of this gene did not result in an active phytoene synthase either [25]. TK-3, however, produces lycopene, implying that the *crtB* gene from *T. thermophilus* encodes an active enzyme in *R. marinus* and also that without CruF activity, most of the remaining enzymes downstream in the pathway cannot act on the carotene backbone. The CrtI enzyme is clearly still active, as it is essential for producing lycopene. This means that CrtI is not dependent on the modification done by CruF and that the two enzymes might act simultaneously in the native pathway. Since *cruF* and *crtB* are fused in one gene, it can be argued that their corresponding enzymes likely act together or

successively. Whether the enzymes are fused into one polypeptide chain or not cannot be concluded from the data presented here. However, our results do show that CrtB is inactive without CruF, which suggests that the two enzymes might be acting as one entity. Also, coupled CruF and CrtB activities could explain how the organism is able to produce asymmetric carotenoids. While we can only speculate on the order of which these enzymes act, further studies on substrate specificity of the enzymes could elucidate their respective functions. The CruF enzyme is likely a 1',2'-hydratase. Without its activity, it was to be expected that the CruC and CruD enzymes, which add a glycosyl and an acyl group to the C-1' hydroxyl group, would not modify the backbone. Our results confirmed this. Two additional native modifications on the right side of the backbone were absent in the carotenoid from TK-3, which are catalyzed by unknown enzymes in *R. marinus*. They add the C-2' hydroxyl group and the C-3',4' double bond. CrtI is a desaturase and has previously been reported to produce 3,4-didehydrolycopene in *Neurospora crassa* [26]. It is therefore possible that the CrtI in *R. marinus* is responsible for the 3',4'-desaturation. However, since the C-3',4' double bond is not present in TK-3, this putative activity of CrtI must be dependent on the hydration of the C-1',2' double bond. While we do not have stronger evidence of CrtI catalyzing the 3',4'-desaturation in *R. marinus*, this step in the pathway remains unknown (Figure 1). On the left side of the native carotene backbone is a 4-keto β-ionone ring, which is not present in the carotenoid produced in TK-3. The *crtY* gene, which encodes a lycopene cyclase, was not disrupted during the genetic modifications, which means that the corresponding enzyme simply cannot act on the backbone without the modifications on the right side of it. Based on this, the enzymes responsible for the modifications on the right side are proposed to modify the backbone before CrtY and CrtO (Figure 1). To confirm this hypothesis, substrate specificity of the enzymes would have to be tested.

The objective of this work was to engineer the carotenoid biosynthetic pathway in *R.* marinus to produce lycopene instead of its native carotenoids. We successfully obtained the *R. marinus* mutant strain TK-3 (Δ*trpB*Δ*purA*Δ*cruFcrtB::trpBcrtB*$_{T.thermophilus}$), that produced 0.49 g/kg CDW of lycopene. This can be compared to optimized commercial lycopene producing microorganisms, such as *Escherichia coli*, *Saccharomyces cerevisiae* and *Blakeslea trispora,* which have been reported to produce 43.7 [27], 55.6 [28] and 15 [29], [30] g/kg CDW of lycopene, respectively. Currently, *R. marinus* produces 1-2 orders of magnitude less

lycopene than these strains. However, there is considerable room for improvement. Further optimization of the carotenoid production through genetic engineering and culture conditions is likely to improve yields significantly. For instance, in previous work on cultivation of *R. marinus,* a 28-fold increase of 450 nm absorption was observed in native carotenoid extracts from sequential batch cultivation with cell recycling, as compared to that of shake flask cultivations [31]. Taking the higher cell densities into consideration, carotenoid absorption per cell density was still 11-fold higher than shake flask cultivations.

R. marinus is a robust versatile organism and in many aspects a preadapted production organism for utilization of recalcitrant polysaccharides in 2nd and 3rd generation biomass, an important task of biorefinery development. In the present study we have demonstrated that *R. marinus* is amenable to metabolic engineering, comprising both gene deletions and insertions, leading to efficient production of a metabolite of industrial interest. The work also revealed the potential of the methodology to help unravel complex pathways.

Declaration of competing interest

None

CRediT author statement

Thordis Kristjansdottir: Conceptualization, Methodology, Investigation, Writing – Original Draft, Writing – Review and Editing, Visualization. **Emanuel Y. C. Ron**: Conceptualization, Methodology, Investigation, Writing – Original Draft, Writing – Review and Editing. **Daniel Molins-Delgado**: Methodology, Writing – Review and Editing. **Olafur H. Fridjonsson**: Conceptualization, Writing – Review and Editing. **Charlotta Turner**: Methodology, Writing – Review and Editing. **Snaedis H Bjornsdottir**: Methodology, Writing – Review and Editing. **Steinn Gudmundsson**: Conceptualization, Writing – Review and Editing. **Ed van Niel**: Writing – Review and Editing. **Eva Nordberg-Karlsson**: Conceptualization, Project administration, Funding acquisition, Supervision, Writing – Review and Editing. **Gudmundur O. Hreggvidsson**: Conceptualization, Supervision, Project administration, Funding acquisition, Writing – Review and Editing.

Funding

This work was supported by the Marine Biotechnology ERA-NET, *ThermoFactories,* project grant number 5178–00003B; the Technology Development fund in Iceland, grant number 159004-0612; the Icelandic Research fund, *ThermoExplore*, project grant number 207088-051 and the Novo Nordisk Foundation (NNF18OC0034792). ENK, CT, EYCR and DM-D would like to acknowledge the Swedish Research Council Formas (2018-01863).

Acknowledgements

The authors would like to thank Louna Maignien and Madelaine Stoll for their work on genetic engineering of *Rhodothermus marinus*; Elisabet Eik Gudmundsdottir, Steinunn Magnusdottir and Agusta K. Sigfusdottir from Matis for their help performing the gene sequencing.

References

[1] R. Munoz, R. Rosselló-Móra, and R. Amann, “Revised phylogeny of Bacteroidetes and proposal of sixteen new taxa and two new combinations including Rhodothermaeota phyl. nov.,” *Syst. Appl. Microbiol.*, vol. 39, no. 5, pp. 281–296, 2016.

[2] G. A. Alfredsson, J. K. Kristjansson, S. Hjorleifsdottir, and K. O. Stetter, “Rhodothermus marinus, gen.nov., sp. nov., a Thermophilic, Halophilic Bacterium from Submarine Hot Springs in Iceland,” *J. Gen. Microbiol.*, no. 134, pp. 299–306, 1988.

[3] E. Y. C. Ron *et al.*, “Characterization of carotenoids in Rhodothermus marinus,” *Microbiologyopen*, vol. 7, no. 1, 2018.

[4] R. R. R. Sardari *et al.*, “Evaluation of the production of exopolysaccharides by two strains of the thermophilic bacterium Rhodothermus marinus,” *Carbohydr. Polym.*, vol. 156, pp. 1–8, 2017.

[5] A. López-Contreras *et al.*, “Biorefinery Approach to the Use of Macroalgae as Feedstock for Biofuels,” in *Algal Biofuels*, L. Pereira, Ed. CRC Press, 2017.

[6] E. Nordberg Karlsson *et al.*, “Metabolic engineering of thermophilic bacteria for production of biotechnologically interesting compounds,” in *Biotechnological applications of extremophilic microorganisms. Volume 6 of Life in Extreme Environments*, N. M. Lee, Ed. Berlin, Boston: De Gruyter, 2020.

[7] S. H. Bjornsdottir, O. H. Fridjonsson, J. K. Kristjansson, and G. Eggertsson, “Cloning and expression of heterologous genes in Rhodothermus marinus,” *Extremophiles*, vol. 11, no. 2, pp. 283–293, 2007.

[8] S. H. Bjornsdottir, O. H. Fridjonsson, G. O. Hreggvidsson, and G. Eggertsson, “Generation of targeted deletions in the genome of Rhodothermus marinus,” *Appl. Environ. Microbiol.*, vol. 77, no. 15, pp. 5505–5512, 2011.

[9] J. Yabuzaki, “Carotenoids Database: structures, chemical fingerprints and distribution among organisms,” *Database (Oxford).*, vol. 2017, no. 1, p. bax004, Jan. 2017.

[10] R. Sathasivam and J.-S. Ki, “A Review of the Biological Activities of Microalgal

Carotenoids and Their Potential Use in Healthcare and Cosmetic Industries," *Mar. Drugs*, vol. 16, no. 1, p. 26, Jan. 2018.

[11] J. R. Mein, F. Lian, and X.-D. Wang, "Biological activity of lycopene metabolites: implications for cancer prevention," *Nutr. Rev.*, vol. 66, no. 12, pp. 667–683, Dec. 2008.

[12] A. J. Young and G. L. Lowe, "Carotenoids-Antioxidant Properties," *Antioxidants (Basel, Switzerland)*, vol. 7, no. 2, p. 28, Feb. 2018.

[13] Markets And Markets, "Carotenoids Market," 2020. [Online]. Available: https://www.marketsandmarkets.com/Market-Reports/carotenoid-market-158421566.html?gclid=CjwKCAiA-vLyBRBWEiwAzOkGVOx0oK25469b9LfMwRcRsTARZR8-HP8b4Jx6AzZzR615k466eqQBJxoCY0kQAvD_BwE.

[14] B. F. Lutnaes, Å. Strand, S. K. Pétursdóttir, and S. Liaaen-Jensen, "Carotenoids of thermophilic bacteria - Rhodothermus marinus from submarine Icelandic hot springs," *Biochem. Syst. Ecol.*, vol. 32, no. 5, pp. 455–468, 2004.

[15] C. Yanisch-Perron, J. Vieira, and J. Messing, "Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mpl8 and pUC19 vectors," *Gene*, vol. 33, no. 1, pp. 103–119, 1985.

[16] E. Degryse, N. Glansdorff, and A. Piérard, "A comparative analysis of extreme thermophilic bacteria belonging to the genus Thermus," *Arch. Microbiol.*, vol. 117, no. 2, pp. 189–196, 1978.

[17] J. H. Miller, *Experiments in molecular genetics*. Cold Spring Harbor Laboratory, 1972.

[18] S. H. Bjornsdottir, S. H. Thorbjarnardottir, and G. Eggertsson, "Establishment of a gene transfer system for Rhodothermus marinus," *Appl. Microbiol. Biotechnol.*, vol. 66, no. 6, pp. 675–682, 2005.

[19] F. Jumaah, M. Plaza, V. Abrahamsson, C. Turner, and M. Sandahl, "A fast and sensitive method for the separation of carotenoids using ultra-high performance supercritical

fluid chromatography-mass spectrometry," *Anal. Bioanal. Chem.*, vol. 408, no. 21, pp. 5883–5894, 2016.

[20] D. Naviglio, F. Pizzolongo, L. Ferrara, A. Aragòn, and A. Santini, "Extraction of pure lycopene from industrial tomato by-products in water using a new high-pressure process," *J. Sci. Food Agric.*, vol. 88, no. 14, pp. 2414–2420, 2008.

[21] Z. Sun *et al.*, "A novel carotenoid 1,2-hydratase (CruF) from two species of the non-photosynthetic bacterium Deinococcus," *Microbiology*, vol. 155, no. 8, pp. 2775–2783, 2009.

[22] G. Britton, S. Liaaen-Jensen, and H. Pfander, *Carotenoids: Handbook*. Birkhäuser Basel, 2004.

[23] N. I. Krinsky, S. T. Mayne, and H. Sies, *Carotenoids in Health and Disease*. CRC Press, 2004.

[24] R. B. van Breemen, L. Dong, and N. D. Pajkovic, "Atmospheric Pressure Chemical Ionization Tandem Mass Spectrometry of Carotenoids," *Int. J. Mass Spectrom.*, vol. 312, pp. 163–172, Feb. 2012.

[25] W. Xie, X. Lv, L. Ye, P. Zhou, and H. Yu, "Construction of lycopene-overproducing Saccharomyces cerevisiae by combining directed evolution and metabolic engineering," *Metab. Eng.*, vol. 30, pp. 69–78, 2015.

[26] A. Hausmann and G. Sandmann, "A Single Five-Step Desaturase Is Involved in the Carotenoid Biosynthesis Pathway to β-Carotene and Torulene in Neurospora crassa," *Fungal Genet. Biol.*, vol. 30, no. 2, pp. 147–153, 2000.

[27] F. Zhu *et al.*, "Targeted engineering and scale up of lycopene overproduction in Escherichia coli," *Process Biochem.*, vol. 50, no. 3, pp. 341–346, 2015.

[28] Y. Chen, W. Xiao, Y. Wang, H. Liu, X. Li, and Y. Yuan, "Lycopene overproduction in Saccharomyces cerevisiae through combining pathway engineering with host engineering," *Microb. Cell Fact.*, vol. 15, no. 1, p. 113, Jun. 2016.

[29] V. M. Tereshina, A. S. Memorskaya, G. A. Kochkina, and E. P. Feofilova, "Dormant Cells

in the Developmental Cycle of Blakeslea trispora: Distinct Patterns of the Lipid and Carbohydrate Composition," *Microbiology*, vol. 71, no. 6, pp. 684–689, 2002.

[30] W. Hu, D. Dai, and W. Li, "Anti-aging effect of Blakeslea trispora powder on adult mice," *Biotechnol. Lett.*, vol. 35, no. 8, pp. 1309–1315, 2013.

[31] E. Y. C. Ron, R. R. R. Sardari, R. Anthony, E. W. J. van Niel, G. O. Hreggvidsson, and E. Nordberg-Karlsson, "Cultivation technology development of Rhodothermus marinus DSM 16675," *Extremophiles*, vol. 23, no. 6, pp. 735–745, 2019.

Paper 3

A genome-scale metabolic reconstruction provides insight into the metabolism of the thermophilic bacterium *Rhodothermus marinus*

Thordis Kristjansdottir(1,2,*), Gudmundur O. Hreggvidsson(1,3), Sigmar Karl Stefansson(3), Elisabet Eik Gudmundsdottir(1), Snaedis H. Bjornsdottir(3), Olafur H. Fridjonsson(1), Eva Nordberg Karlsson(4), Justine Vanhalst(1), Birkir Reynisson(1), Steinn Gudmundsson(2)

(1) Matis, Vinlandsleid 12, 113. Reykjavik, Iceland

(2) University of Iceland, Department of Computer Science, School of Engineering and Natural Sciences, Dunhagi 5, 107 Reykjavik, Iceland

(3) University of Iceland, Department of Biology, School of Engineering and Natural Sciences, Sturlugata 7, 102 Reykjavík, Iceland

(4) Lund University, Department of Chemistry, Division of Biotechnology, Box 124, 221 00 Lund, Sweden

(*) Corresponding author, thordis@matis.is

Abstract

The thermophilic bacterium *Rhodothermus marinus* has mainly been studied for its thermostable enzymes. More recently, the potential of using the species as a cell factory and in biorefinery platforms has been explored, due to the elevated growth temperature, native production of compounds such as carotenoids and EPSs, the ability to grow on a wide range of carbon sources including polysaccharides, and available genetic tools. A comprehensive understanding of the metabolism of production organisms is crucial. Here, we report a genome-scale metabolic model of *R. marinus* DSM 4252^T. Moreover, the genome of the genetically amenable *R. marinus* ISCaR-493 was sequenced and the analysis of the core genome indicated that the model could be used for both strains. Bioreactor growth data was obtained, used for constraining the model and the predicted and experimental growth rates were compared. The model correctly predicted the growth rates of both strains. During the reconstruction process, different aspects of the *R. marinus* metabolism were reviewed and subsequently, both cell densities and carotenoid production were investigated for strain ISCaR-493 under different growth conditions. Additionally, the *dxs* gene, which was not found in the *R. marinus* genomes, from *Thermus thermophilus* was cloned on a shuttle vector into strain ISCaR-493 resulting in a higher yield of carotenoids.

Keywords: *Rhodothermus marinus*, Genome-scale metabolic model, Carotenoids

Importance

A biorefinery converting biomass into fuels and value-added chemicals is a sustainable alternative to fossil fuel-based chemical synthesis. *Rhodothermus marinus* is a bacterium that is potentially well suited for biorefineries. It possesses various enzymes that degrade biomass, such as macroalgae and parts of plants (e.g. starch and xylan) and grows at high temperatures (55-77°C) which is beneficial in biorefinery processes. In this study, we reviewed the metabolism of *R. marinus* and constructed a metabolic model. Such a model can be used to predict phenotypes, e.g. growth under different environmental and genetic conditions. We focused specifically on metabolic features that are of interest in biotechnology, including carotenoid pigments which are used in many different industries. We described cultivations of *R. marinus* and the resulting carotenoid production in different growth conditions, which aids in understanding how carotenoid yields can be increased in the bacterium.

1. Introduction

Rhodothermus marinus is an aerobic bacterium that belongs to the phylum *Rhodothermaeota* [1]. It grows optimally at 65°C on various proteinaceous and carbohydrate substrates and was first isolated from a submarine hot spring in Iceland [2]. The *R. marinus* genome encompasses several gene clusters encoding pathways for utilization and cellular import of diverse carbohydrates. Several enzymes from *R. marinus* have been characterized, and many have biotechnological potential, including a number of polysaccharide degrading enzymes, such as cellulase [3], laminarinase [4] and xylanase [5], [6] (see [7] for a review). *R. marinus* grows on a wide range of sugars from second and third generation biomass and can be cultivated to relatively high yields [8]. It has anabolic pathways and precursor pools for production of various biotechnologically interesting primary and secondary compounds, such as polyamines [9], exopolysaccharides [10], carotenoids [11], compatible solutes [12] and lipids [13], [14].

Anaerobic fermentative organisms are generally preferred to produce low-value commodity chemicals, simple organic acids and alcohols that are typically catabolic waste products from incomplete oxidation of substrates. Conversely, heterotrophic aerobes such as *R. marinus*, typically oxidize their carbon substrates completely under optimized conditions, and therefore utilize organic substrates more efficiently for both energy and carbon. Consequently, aerobes can carry a greater metabolic burden and are the preferred organisms for the anabolic production of complex secondary metabolites.

Cultivation at high temperatures (60-70°C) may be beneficial in bioreactors as it reduces the cost of cooling, and higher temperatures protect the cultures from mesophilic spoilage bacteria. High temperature also increases the solubility of polysaccharides and leads to reduced viscosity of the fermentation broth. This may alleviate scale up problems of mixing and aeration to a significant extent and enable greater substrate loadings. Elevated temperatures may also enable cost-effective recovery of volatile products by distillation or gas stripping, reducing product inhibition and prolonging the production phase of the culture [15].

R. marinus has the potential to serve as a robust production organism in the emerging biorefinery industry and as a chassis species that can be metabolically engineered for the

production of novel chemical compounds of industrial interest. For this purpose, comprehensive knowledge and understanding of its metabolism are needed. A genome-scale metabolic reconstruction describes the metabolic network of a given organism based on both genomic information and available physiological data. A well-curated reconstruction contains a comprehensive overview of the metabolism of the organism in question. A reconstruction can be converted to a computational model, where phenotypic features are predicted under different conditions. The model can for instance be used to predict growth capabilities on different substrates and guide genetic engineering efforts [16]. Reconstructing the network of a poorly studied organism can result in gaps in the model. While this may skew predictions, it can nevertheless be useful for focusing future research efforts. Although *R. marinus* is not as well studied as common model organisms, it has been the subject of several studies, including the development and application of genetic tools [17], particularly for the engineering of the carotenoid pathway [18].

Carotenoids are pigments produced by many plants, fungi, algae and bacteria. Some non-photosynthetic bacteria produce carotenoids to stabilize cell membrane fluidity in response to extreme environments (high/low temperatures, pH, salinity, etc.) and to protect themselves against UV radiation and oxidative stress [19]. Carotenoids are in demand for different applications, such as the food, feed and cosmetic industries. In a previous study we engineered the carotenoid biosynthetic pathway in *R. marinus* to produce the industrially relevant carotenoid lycopene, instead of native γ-carotenoids [18]. In another, sequential batch cultivation resulted in higher carotenoid production than shake flask cultivation [8]. Drawing upon the metabolic reconstruction of the current study, we further investigated the effects of culture conditions on carotenoid production and growth of *R. marinus*.

Here we reconstructed the genome-scale metabolic network of the *R. marinus* type strain DSM 4252^T, which has been the subject of most of the published studies so far and for which an annotated genome sequence is available [20]. However, the type strain is not amenable to genetic manipulation as it aggregates in liquid cultures and harbors a highly active DNA restriction enzyme with a 4 base recognition site [21]. Therefore, existing genetic tools were developed for another *R. marinus* strain, ISCaR-493 (DSM 16675), which was selected after screening of numerous *R. marinus* strains for a restriction-deficient phenotype [22]. Here, the genome of strain *R. marinus* ISCaR-493 was sequenced and the genomes of the two

strains were compared. This analysis was used to find if any model genes from DSM 4252^T could not be found in ISCaR-493, and subsequently if the DSM 4252^T model could be extrapolated to strain ISCaR-493. Growth curves and uptake and secretion rates of the main metabolites from bioreactor cultivations were obtained for both strains and the data used to validate the model. During the reconstruction process, several interesting features related to carotenoid production were identified and investigated further. This included heterologous expression of a gene from *Thermus thermophilus* encoding the terpenoid biosynthetic enzyme 1-deoxy D-xylulose 5-phosphate (DXP) synthase, which was not identified in the genomes of *R. marinus*.

2. Methods

2.1 Strains, media and culture conditions

Three *R. marinus* strains were used in this study, DSM 4252^T, ISCaR-493 (DSM 16675) and the mutant strain TK-4 (ISCaR-493 derivative, Δ*trpB*Δ*purA::trpBdxs*$_{T.thermophilus}$, section 2.3). All cultivations were at 65°C and liquid cultures were set to shaking at 200 rpm. For each cultivation, *R. marinus* was first streaked on an agar plate containing rich medium, 10% medium 162 [23], with modifications (2 mM $MgSO_4$ and 0.2 mM $CaCl_2$ in final volume) and addition of 1% NaCl, 0.03% K_2HPO_4, 0.1% yeast extract, 0.1% tryptone, 0.1% peptone, 0.05% glucose, 0.05% starch, 0.06% pyruvate and 0.018% Na_2CO_3. Utilization of different carbon sources (section 3.1) was examined on defined medium agar plates containing 10% medium 162 [23] with addition of 8 mM phosphate buffer (KH_2PO_4, Na_2HPO_4, pH=7.2), 0.1% vitamin solution [23] and 0.4% of each carbon source, except the amino acids which were 0.2%. *R. marinus* strain DSM 4252^T from a rich medium agar plate (see above) was resuspended in a drop of 0.9% NaCl solution and subsequently streaked on agar plates containing the different carbon sources. The plates were incubated at 65°C and growth was examined after 1, 3, 5 and 7 days.

The cultivations used to validate the model (section 3.3) were carried out in bioreactors (Labfors 5, Infors HT, Bottmingen, Switzerland), where pO_2 was kept at 40% with stirrer speed at 200-500 rpm and airflow as needed, pH at 7.2 with addition of 16.5% NH_4OH and the temperature at 65°C. Defined medium, which contained 10% modified medium 162 with the addition of 1% NaCl, 8 mM phosphate buffer (KH_2PO_4, Na_2HPO_4, pH=7.2), 10 mM NH_4Cl,

0.02% asparagine, 0.02% glutamine, 0.1% vitamin solution [23] and the carbon sources 1% glucose and 0.09% pyruvate, was used. The cultures were inoculated with 10% of pre-culture and were performed in duplicates. Samples were taken for OD (620 nm) and HPLC measurements every hour, until a stationary phase was reached (18 – 32h).

The cultivations for examining cell density and carotenoid production under different conditions (section 3.4) were performed in shake flasks under light and dark conditions. All cultures were exposed to day light and additional light from a halogen lamp, except when grown under dark conditions where the flasks were covered from the light. A defined medium (same as in section 3.3, except Wolf´s vitamin- and trace elements solution from [24] were used), was used for these cultivations, but with different carbon sources: 1% glucose, 1% glucose and 0.09% pyruvate, 1% alginate, 0.5% glucose and 0.25% pyruvate, 0.09% pyruvate, 0.18% pyruvate, and without a carbon source for a negative control. For cultivation of strain TK-4, adenine (0.0025%) was also supplemented. Two pre-cultures were prepared for each liquid culture to be monitored. First, cells were transferred from a fresh rich medium agar plate (see above) to defined liquid medium containing 0.5% glucose and 0.018% pyruvate and grown overnight (16h). This culture was used to inoculate (10%) a fresh defined liquid medium containing 1% glucose and 0.09% pyruvate, which was also grown overnight (16h). All the monitored cultures were inoculated (10%) with the second pre-culture. The cultures were stopped after 24h and cell density was estimated by measuring OD at 620nm in a spectrophotometer (Novaspec III^{+}, Biochrom, Harvard Bioscience Inc., Holliston, Massachusetts, US).

All media components were autoclaved, except for the vitamin solution and the trace element solution, which were filter sterilized. The alginate was not sterilized due to browning and degradation when autoclaved. The probability of contamination in the alginate was low, because of the high growth temperature and because alginate is not a trivial carbon source for most bacteria. The alginate cultures were plated to verify that contamination had not occurred during growth. Individual colonies were obtained, which all had the characteristic red color, and were subsequently identified as *R. marinus* using MALDI-TOF (Microflex, Bruker, Billerica, Massachusetts, US). MALDI-TOF was used according to the manufacturer´s instructions.

2.2 Analytical methods

The DNA content in *R. marinus* cells was estimated from exponential bioreactor cultures using the fluorometric Quant-iT™ PicoGreen™ dsDNA Assay Kit (ThermoFisher, Waltham, Massachusetts, US). A sample of freeze-dried biomass (approx. 10^7 cells/ml) was dissolved in water and sonicated for 10 sec to lyse the cells. A standard curve was obtained from PicoGreen measurements of λ-dsDNA and used to estimate DNA concentration in *R. marinus* cells.

Glucose, lactate and acetate concentrations in the samples taken during bioreactor cultivations were measured using high-performance liquid chromatography (HPLC). The samples were filtered through 0.2 µm filters (Phenomenex) and metabolites subsequently quantified using the Dionex 2000 HPLC system (Dionex, Idstein, Germany) with a Rezex ROA-Organic Acid H + (8%, Phenomenex, Aschaffenburg, Germany) and a RI-101 detector (Shodex, München, Germany). Chromeleon evaluation software version 6.80 (Dionex, Idstein, Germany) was used. Separation was obtained using 60°C column temperature with 0.2 mM sulfuric acid (Carl Roth, Karlsruhe, Germany) as the eluent at a flow rate of 600 µl/min for 30 min. External standards of HPLC grade (Merck, Darmstadt, Germany, Sigma-Aldrich, St. Louis, USA) were used. The pyruvate concentration was estimated using a pyruvic acid kit (Megazyme, Bray, Ireland). The two amino acids in the defined medium, asparagine and glutamine, were not measured. Their concentration in the medium was low and did therefore not influence the model predictions to a significant extent. OD (620 nm) values from bioreactor cultivations were converted to CDW (g/l) based on data obtained elsewhere [25], [26] (CDW = 0.75 x OD).

For the carotenoid measurements, the cultures were diluted to OD = 1 at 620 nm. The cell pellets from 1 mL of the diluted cultures were mixed well with 1 mL of acetone and incubated in a sonication bath for 20 min. The samples were centrifuged at 16,000 g for 7 min, resulting in a colorless cell pellet and acetone supernatant containing the extracted carotenoids. The *R. marinus* carotenoids in acetone have maximum absorbance at ~480 nm (data not shown). The OD at 480 nm of the acetone extracts was measured in a 1 cm quarts cuvette in a spectrophotometer (Novaspec III$^+$, Biochrom).

2.3 Genetic modification of *R. marinus*

The genomic DNA from *T. thermophilus* HB8 was isolated using the MasterPure Complete DNA purification Kit (Lucigen, Middleton, Wisconsin, US) and used as template for the amplification of the *dxs* (1-deoxy D-xylulose 5-phosphate synthase) gene (TTHA0006) (1848bp) by PCR. The amplification was performed using the Q5 High-Fidelity DNA polymerase (New England BioLabs, Ipswich, Massachusetts, US), according to the manufacturer´s instructions. The primers were designed to support HiFi DNA assembly (NEBuilder HiFi DNA Assembly Master Mix, New England BioLabs): dxs_thermus_F (5'-AAACCATGGAGGTGCGCGATATGATCTTGGACAAGGTGAAC-3') and dxs_thermus_R (5'-AGCAGTTCGGTCTCGGCGGTCGACATTCAGGCCCGTTCATGCACCAAG-3'), with the underlined bases matching the *dxs* gene and the rest matching a *Sal*I (New England BioLabs) digested *R. marinus* shuttle vector pRM3000.0. The pRM3000.0 vector is the pRM3000 vector [17] with an ampicillin resistance gene added, for selection in *E. coli.* The pRM3000.0-*dxs* vector was introduced into chemically competent *E. coli* (NEB 5-alpha) and plated on L-medium [27] with 100 µg/mL ampicillin. After incubation overnight at 37°C, positive clones were identified by amplifying the *dxs* gene, using the *Taq* DNA polymerase (New England BioLabs) according to the manufacturer´s instructions. The primers dxs_thermus_F and R were used. The pRM3000.0-*dxs* vector was isolated from positive *E. coli* clones using the Monarch Plasmid Miniprep Kit (New England BioLabs).

The pRM3000.0-*dxs* vector was introduced into *R. marinus* SB-62 (ISCaR-493 derivative, Δ*trpB*Δ*purA*) by transformation [28], using the *trpB* gene for Trp+ selection. Competent *R. marinus* cells were prepared and transformed as described elsewhere [22]. The GenePulser Xcell electroporation system (Bio-Rad) was used for the transformation, with pulses delivered at 20 kV/cm. For each transformation, 40 µl of washed cells and 1 µg of DNA in ≤ 5 µL volume were used. The pRM3000.0 vector was used as a positive control and sterile MilliQ water as a negative control. Transformed cells were spread on agar plates containing medium 162 [23], with modifications (2 mM $MgSO_4$ and 0.2 mM $CaCl_2$ in final volume) and additions of 1% NaCl, 0.053% NH_4Cl, 0.2% soluble starch, 0.2% casamino acids, vitamin solution [23], 0.0025% adenine and 2.5% agar. Plates were incubated at 65°C for 5 days. Positive *R. marinus* clones were verified using the same PCR as positive *E. coli* clones.

2.4 Genome sequencing and analysis

R. marinus strain ISCaR-493 was cultured in rich liquid medium (see section 2.1). For sequencing by short-read technology, genomic DNA was extracted using the MasterPure Complete DNA purification Kit (Lucigen) and sequencing libraries made by both the Nextera XT (FC-131-1024) and Nextera Mate Pair (FC-132-1001) methods. The two resulting libraries were sequenced on the MiSeq sequencing platform using V3 2x 300bp and V2 2x 250bp chemistry, respectively. The sequence reads were quality assessed using FastQC (v0.11.7) [29] and trimmed using Trimmomatic (v0.39) [30]. A genome was assembled using SPAdes (v3.14.0) [31] with flag for isolates. Genome polishing was done by mapping all shotgun short-reads (Nextera XT) to the largest contig from the *de novo* assembly and generating a consensus sequence using the highest quality bases, using Geneious (v9.1.4). The genome was annotated using the PGAP pipeline from NCBI [32].

2.5 Metabolic reconstruction and simulations

A protocol for reconstructing genome-scale metabolic models [33] was used to guide the reconstruction process. A draft reconstruction, based solely on the annotated genome sequence, was obtained using the Model SEED [34]. Since no well-curated model of a closely related bacterium was available at the time of the reconstruction, the obtained draft was neither comprehensive nor functional. It was therefore used as a reference model, along with the well-curated model of *E. coli* (iJO1366) [35]. The *R. marinus* model was manually reconstructed based on these two reference models and *R. marinus* experimental data, obtained here and from the literature. The BiGG [36], BRENDA [37], KEGG [38] and MetaCyc [39] databases and the BLAST tool [40] were extensively used.

Cobrapy [41] was used in all model simulations, along with the GLPK solver. The corresponding code can be found as a Jupyter notebook on Github (https://github.com/steinng/rmarinus). For all simulations, flux balance analysis (FBA) was used [42], [43]. Exchange reactions corresponding to metabolites taken up from the media (glucose and pyruvate) and secreted (lactate and acetate) during growth, were constrained with experimentally obtained rates (Supplementary file 1). FBA was subsequently used to optimize for growth by maximizing flux through the biomass reaction.

For accurate growth rate predictions, the biomass reaction should ideally be based on data obtained for the target organism. Here, the biomass reaction was formulated based mostly on available data on *R. marinus* (Supplementary file 2). Separate biosynthetic reactions for each group of macromolecules (protein, lipid, DNA, etc.) were formulated, describing the ratio of the building blocks (amino acids, fatty acids, nucleotides, etc.) and the energy required. Sensitivity analysis, which shows how much variation in each macromolecule affects the predicted growth rate, was performed. This analysis helps to identify which biomass components most urgently need to be accurately measured.

3. Results and discussion

3.1 Reconstruction of a genome-scale metabolic model of *R. marinus* DSM 4252^T

A genome-scale metabolic model of *R. marinus* DSM 4252^T was reconstructed, named Rmarinus_578 (https://github.com/steinng/rmarinus). The main features of the reconstruction are listed in Table 1. The Memote tool [44] was used to help guide the reconstruction process, by verifying stoichiometric consistency, mass and charge balance and annotation quality (Supplementary file 3). Reactions and metabolites were usually abbreviated in accordance with the BiGG database and annotations with links to external databases are included. The genome sequence for strain DSM 4252^T was obtained from GenBank (accession nr: NC_013501). The genes in the reconstruction were identified with the locus tags from the GenBank file. They were annotated with the old gene locus tag from the GenBank file, the protein ID, protein annotation and protein sequence. Experimental data on *R. marinus* obtained in this study and the available literature was used to curate reactions, genes and gene-protein-reaction (GPR) rules. Several metabolic features were reviewed during the reconstruction process. In the following, we highlight a few, which are of interest for biotechnological application of *R. marinus*.

Table 1. *Main features of the genome-scale metabolic model of R. marinus DSM 4252^T, Rmarinus_578.*

Genome features	
Genome size	3.26 Mb
Protein coding sequences	2889
Model features	
Genes	578
Metabolites (unique)	871 (784)
Reactions (with GPR)	929 (771)
Memote total score	55%

Sugar metabolism

R. marinus produces pyruvate from glucose through the Embden-Meyerhof-Parnas (EMP) pathway. A 13C metabolic flux study of the central metabolism in *R. marinus* [45] showed that the EMP pathway and the TCA cycle are both highly active while metabolizing glucose. The oxidative pentose phosphate pathway and the glyoxylate shunt had very low activity and the Entner-Doudoroff (ED) pathway, malic enzyme and phosphoenolpyruvate carboxykinase were inactive.

Growth of *R. marinus* strain DSM 4252^T was tested on many different carbon sources, both *in vivo* and *in silico* (Table 2). Growth has been shown on several mono-, di- and polysaccharides, which was also observed *in silico*. However, growth on cellulose was predicted *in silico* while not observed *in vivo*. *R. marinus* does contain genes encoding a endocellulase (EC 3.2.1.4) [3], [46], which can degrade cellulose into differently sized cellooligosaccharides and cellobiose. However, low specific activities were reported compared to benchmark enzymes. The strain does not contain genes encoding specific exocellulases (EC 3.2.1.91), which degrade glucans into β-cellobioses. However, it does contain several genes encoding glycoside hydrolases, that can degrade different oligosaccharides, including cellobiose and cellooligosaccharides, to monosaccharides [47]. Although *R. marinus* might be able to use some of the cellulose partially degraded by the cellulase and glycoside hydrolases, as the model predicts, the cellulose-specific activity of the enzymes is probably not high enough for it to grow solely on cellulose. *R. marinus* was

unable to grow *in vivo* on several of the tested carbon sources, especially organic- and amino acids (Table 2). The model predictions usually showed the same results. However, when transport reactions were added to the model, growth was often observed *in silico* for carbon sources that did not result in growth *in vivo*. These carbon sources might be good targets for adaptive evolution experiments. Promiscuous enzymes might adapt to transport these compounds into the cell, where they can presumably be used for growth [48].

Lactose and galactose are examples of the di- and monosaccharides that *R. marinus* can use for growth (Table 2). The gene encoding a β-galactosidase (EC 3.2.1.23), which hydrolyzes lactose into glucose and galactose, was found in the genome. Three steps are needed to convert galactose to glucose-6-phosphate, which then enters the glycolysis EMP pathway. The genes encoding galactokinase and galactose-1-phosphate uridylyltransferase, catalyzing the first two steps (galactose -> galactose-1-phosphate -> glucose-1-phosphate), were found in the genome. The third step, where glucose-1-phosphate is turned into glucose-6-phosphate, is usually performed by the enzyme phosphoglucomutase (EC 5.4.2.2). The gene for this enzyme was not found in the genome. However, a homology search showed similarity between known phosphoglucomutase genes from other bacteria and genes RMAR_RS01880 (E value 2e-37) and RMAR_RS08875 (E value 1e-25) which are annotated as phosphomannomutase (EC 5.4.2.8) and phosphoglucosamine mutase (EC 5.4.2.10), respectively. The enzyme phosphoglucosamine mutase in *E. coli*, which usually catalyzes the interconversion of glucosamine-6-phosphate and glucosamine-1-phosphate, was also shown to be able to catalyze the interconversion of glucose-6-phosphate and glucose-1-phosphate, at a lower rate [49]. The phosphorylation site of this enzyme in *E. coli* is Ser102 and a mutational change of Ser100 to a threonine residue increased the phosphoglucomutase activity significantly. Gene RMAR_RS08875 from *R. marinus* was investigated and the serine residue responsible for the phosphorylation was found to be residue number 103. The corresponding residue to Ser100 in the *E. coli* enzyme was found to be a threonine, which indicates that this *R. marinus* enzyme may be responsible for the interconversion of glucose-6-phosphate and glucose-1-phosphate in *R. marinus*.

Table 2: *Growth of R. marinus strain DSM 4252^T on different carbon sources investigated in silico and in vivo. Green indicates growth, light green indicates weak growth, 'w' indicates white colonies (instead of the characteristic red), orange indicates no growth in silico except if a transport reaction was added to the model, red indicates no growth and no color indicates no data available.*

	Metabolite	*In Silico*	*In Vivo*
Mono- and Disaccharides	Glucose[1]		
	Galactose[1]		
	Sucrose[1]		
	Lactose[1]		
	Raffinose[1]		
	Maltose[1]		
	Fructose[1]		
	Mannose[1]		
	Ribulose[1]		
	Xylose[1]		
Other	Pyruvate[1]		w
	Acetate[1]		w
	Lactate		
	Formate		
Polysaccharides	Starch[1]		w
	Xylan[1]		
	Laminarin[1]		
	Alginate[1]		
	Cellulose (CMC)[1]		
Citric- and Urea cycles	Malate[1]		
	Fumarate[2]		
	Succinate[2]		
	Citrate[2]		
	Ornithine		
	Oxaloacetate		
	2-oxoglutarate[2]		

	Metabolite	*In Silico*	*In Vivo*
Amino acids	Glutamate[1]		
	Aspartate[2]		
	Glutamine[1]		
	Asparagine[2]		
	Arginine[2]		
	Leucine[2]		
	Phenylalanine[2]		
	Proline[2]		
	Serine[2]		
	Threonine[2]		
	Valine[2]		
	Alanine		
	Histidine		
	Glycine		
	Lysine		
	Tryptophan		
	Tyrosine		
	Cysteine		
	Methionine		
	Isoleucine		

[1]Data from this study.
[2]Data from [2].

R. marinus possesses several genes encoding polysaccharide degrading enzymes. As a marine bacterium, seaweed is common in its natural environment. Alginate and laminarin are major polysaccharides of brown algae, which *R. marinus* can break down and use as sole carbon sources for growth [50]. Alginate is a structural component of brown algae and can

comprise up to 40% of its dry matter [51]. It is a polyuronate that consists of β-D-mannuronate (M) and α-L-guluronate (G) units forming (1 → 4) linked G-, M- and mixed blocks in the polysaccharide chain. The *R. marinus* genome has four genes encoding alginate lyases [52], [53] that, together, depolymerize alginate into the same unsaturated mono-uronate derivative of the M and G units. The *R. marinus* genome also possesses the genes encoding for the remaining enzymes of the alginate catabolic pathway enabling its utilization. The unsaturated monouronate is converted to 4-deoxy l-erythro 5-hexoseulose uronic acid (DEH) y a spontaneous reaction and further catalyzed to 2-keto 3-deoxygluconate (KDG) by an aldose reductase [54]. KDG enters the partial ED pathway in *R. marinus*, where it is catalyzed to 2-keto-3-deoxygluconate 6-phosphate (KDPG) by 2-keto 3-deoxygluconokinase (EC 2.7.1.45) and then by 2-dehydro-3-deoxyphosphogluconate aldolase (EC 4.1.2.14), to pyruvate and glyceraldehyde 3-phosphate which enter central metabolism [54], [55] (Figure 1). *R. marinus* does not possess a key enzyme in the ED pathway, phosphogluconate dehydrogenase (EC 4.2.1.12), which is needed to metabolize glucose. This explains why the ED pathway is not active when *R. marinus* is grown on glucose [45], while the partial pathway is essential for utilization of alginate (Figure 1).

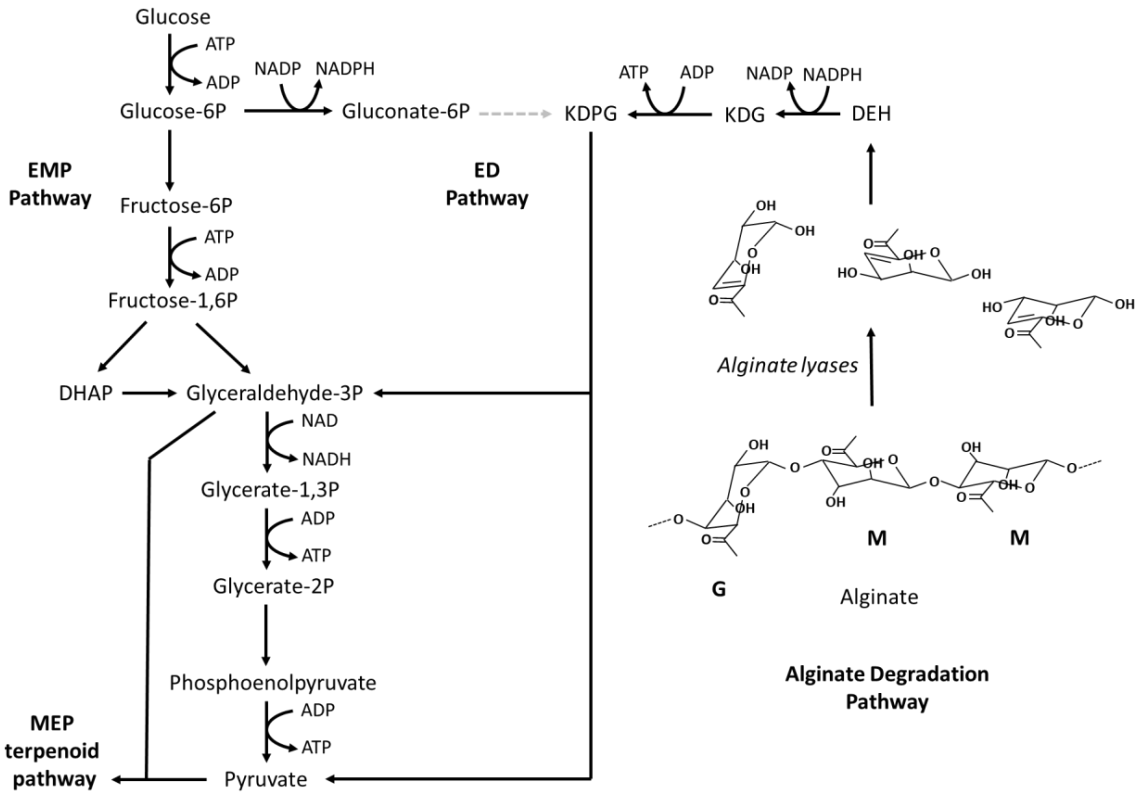


Figure 1. *The alginate degradation pathway of R. marinus and its connections to the Embden-Meyerhof-Parnas (EMP) pathway, the partial Entner-Doudoroff (ED) pathway and the MEP terpenoid pathway. Names of substrates, products and energy molecules (ATP, NADH, NADPH) are shown, some abbreviated: 4-deoxy-L-erythro 5-hexoseulose uronic acid (DEH), 2-keto 3-deoxygluconate (KDG) and 2-keto 3-deoxygluconate 6-phosphate (KDPG). The molecular structure of a partial alginate molecule is shown: β-D-mannuronate (M) and α-L-guluronate (G). The missing reaction of the ED pathway in R. marinus is represented by a grey dotted arrow.*

Terpenoid and carotenoid metabolism

The carotenoids produced by several *R. marinus* strains have been characterized [11], [56]. A carotenoid pathway was proposed in [11] using structural and bioinformatic data, and later refined [18]. This pathway is included in the reconstruction along with candidate genes. Terpenoids, which serve as precursors to carotenoids, can be produced through two different pathways, the mevalonate [57], [58] and the non-mevalonate (MEP) [59] pathways. In the latter, which is used by most bacteria including *R. marinus*, geranylgeranyl diphosphate (GGDP) is produced from pyruvate and glyceraldehyde 3-phosphate in multiple steps. *R. marinus* possesses genes coding for all the enzymes in the MEP pathway, except for 1-deoxy D-xylulose 5-phosphate (DXP) synthase (DXS, EC 2.2.1.7), which catalyzes the first step in the pathway. Studies of the MEP pathway in other bacteria have shown that DXS is not strictly necessary for the synthesis of DXP. Examples include a mutated form of pyruvate

dehydrogenase that is known to rescue *E*. *coli* cells defective in DXS [60] and a mutated RibB protein and a YajO protein that synthesized DXP from ribulose 5-phosphate, also in *E. coli* [61]. Another possibility to bypass DXS is via the MTA-isoprenoid shunt, as has been shown in *Rhodospirillum rubrum* [62]. Here, the dead-end metabolite of polyamine biosynthesis, 5-methylthioadenosine (MTA) is metabolized by an alternative methionine salvage pathway, which produces DXP as a side-product. Phylogenetic analysis showed that the genes in this pathway are partially present in the *R. marinus* genome. At present, it is not known how DXP is produced in *R. marinus* and without further evidence of alternative pathways, the DXP synthase reaction is present in the reconstruction without any gene candidates assigned. The absence of DXS in *R. marinus* directly suggests heterologous expression of a thermostable DXS as means to increase flux through the terpenoid pathway.

Light-inducible carotenoid production has been observed in many organisms, including non-photosynthetic bacteria, and the regulatory mechanisms have been studied in some of them, including *Myxococcus xanthus* [63], *Thermus thermophilus* [64], *Streptomyces coelicolor* [65] and *Bacillus megaterium* [66]. The MerR family transcriptional regulator, LitR, acts as a repressor in the carotenoid gene cluster. Its activity is dependent on the binding with adenosyl B_{12} and the LitR-AdoB_{12} complex becomes inactivated when illuminated. This causes cell cultures to become colorless under dark conditions while producing carotenoids in light. A homologue for the *T. thermophilus litR* gene was previously identified in the carotenoid gene cluster in *R. marinus* [18], located upstream of the carotenoid gene *crtB* (phytoene synthase). This suggests that light might help increase carotenoid yields in *R. marinus*.

Other pathways

The respiratory chain in *R. marinus* has been extensively studied. The first two complexes, NADH dehydrogenase and succinate dehydrogenase, have been characterized [67], [68] and found to be similar to those of other bacteria. The third complex, cytochrome dehydrogenase, is not the typical bc_1 but an alternative complex [69]. It has an entirely different structure but carries out the same function, oxidizing reduced menaquinones-7 [70] and reducing High Potential iron-sulfur Protein (HiPIP) and cytochrome *c*. Finally, three

different types of the fourth complex have been characterized in *R.* marinus, *cbb3* [71], *caa3* [72] and *ba3* [73]. These reactions and associated genes are included in the reconstruction.

Polyamines are alkaline organic compounds with at least two primary amino groups. They can be found in most forms of life and have diverse functions. Some polyamines are biotechnologically interesting because they can be used to produce plastics and are used as curing agents in polymer applications [74]. In thermophilic bacteria unusual, long and branched polyamines have been observed [75]. They are believed to have protective effects on nucleic acids and proteins under high-temperature conditions. Seven different polyamines have been characterized in *R. marinus* to date, putrescine, spermidine, cadaverine, spermine, thermopentamine, N4-aminopropylspermidine and N4-bis(aminopropyl)spermidine [9], and their biosynthesis is included in the reconstruction. *R. marinus* produces compatible solutes to protect the cell against sudden osmotic changes. They include amino acids, monosaccharides such as trehalose, small peptides and, most abundantly, mannosylglycerate [76]. Mannosylglycerate in *R. marinus* is synthesized via two pathways [77], which have been studied in detail, along with the corresponding enzymes and genes [14], [77], [78]. These pathways are present in the metabolic reconstruction, but as they are produced in response to stress, they are not included in the biomass reaction and the pathways therefore not active during growth simulations where the biomass is maximized.

Genomic information indicates that *R. marinus* can synthesize all the amino acids needed for protein synthesis. This is supported by growth experiments which show that *R. marinus* can grow in defined medium, without any addition of amino acids [79]. Biosynthetic pathways for all the 20 amino acids are included in the reconstruction. The fatty acid and lipid composition of *R. marinus* have been characterized [13], [14]. The dominating fatty acids are iso- and anteiso-C15 and iso- and anteiso-C17, with iso-C16 and iso-C18 as minor components. The major polar lipids are phosphatydilethanolamine, diphosphatydilglycerol and one unidentified lipid and phosphatidylglycerol was identified as a minor lipid [14], [70]. The biosynthesis of the fatty acids and lipids are also included in the reconstruction.

The biomass objective function

In genome-scale metabolic models a biomass objective function (BOF) is used to simulate growth. The BOF describes the ratio between the macromolecules (protein, DNA, RNA, lipids, etc.), the composition of each macromolecule (amino acids, nucleotides, fatty acids, etc.) and the energy that the cell requires to grow and maintain itself. We collected data on several studies that describe different components of the biomass in *R. marinus*. Protein, RNA, lipid and glycogen content and amino acid ratio in the biomass were based on [45], carotenoid composition on [56], EPS on [10], lipids on [13], [14] and polyamines on [9]. Quantification of DNA in the biomass was measured in this study (Supplementary file 2). Nucleotide compositions of DNA and RNA were derived from the DSM 4252^T genome. The RNA estimate assumes equal transcription of all genes and is therefore expected to be somewhat inaccurate. Growth associated maintenance (GAM) was estimated using experimental data obtained here (supplementary file 1). The remainder of the biomass components was adopted from *E. coli* [80]. A detailed overview of the biomass, sources of information and calculations can be found in supplementary file 2.

Sensitivity analysis was performed to investigate how sensitive the predicted growth rate was to changes in biomass and energy components. The growth rate was predicted while varying each component by 50%, for multiple glucose uptake rates. The components tested were protein, DNA, RNA, lipid, exopolysaccharide, lipopolysaccharide, glycogen, peptidoglycan, GAM and non-growth associated maintenance (NGAM) (Supplementary file 2). The predicted growth rate was most sensitive to changes in the protein component, followed by the lipid component.

3.2 Genome comparison of strains DSM 4252^T and ISCaR-493

The metabolic model was reconstructed based on genomic information from *R. marinus* DSM 4252^T. This strain is, however, not amenable to genetic manipulation as it aggregates in liquid cultures and shows high DNA degrading activity due to the presence of a 4 cutter restriction enzyme, *Rma*I [21]. Therefore, genetic tools were developed for a derivative of the strain ISCaR-493 [22]. The 16S rRNA sequence similarity between the two strains is 97.92%.

Core genome

The genome of ISCaR-493 was sequenced using Illumina MiSeq and assembled using SPAdes. Subsequently, the 60 resulting contigs were annotated using the NCBI annotation software PGAP. The resulting draft genome was compared to the genome of strain DSM 4252^T using the in-house pan-genomic software Genset. The genome of ISCaR-493 had a calculated G+C% of 64.6% compared to 64.3% for strain DSM 4252^T. The Average Nucleotide Similarity (ANI value) [81] was 95.5%, indicating a relatively high sequence divergence while the genome metrics were very similar for the two strains (Supplementary file 4).

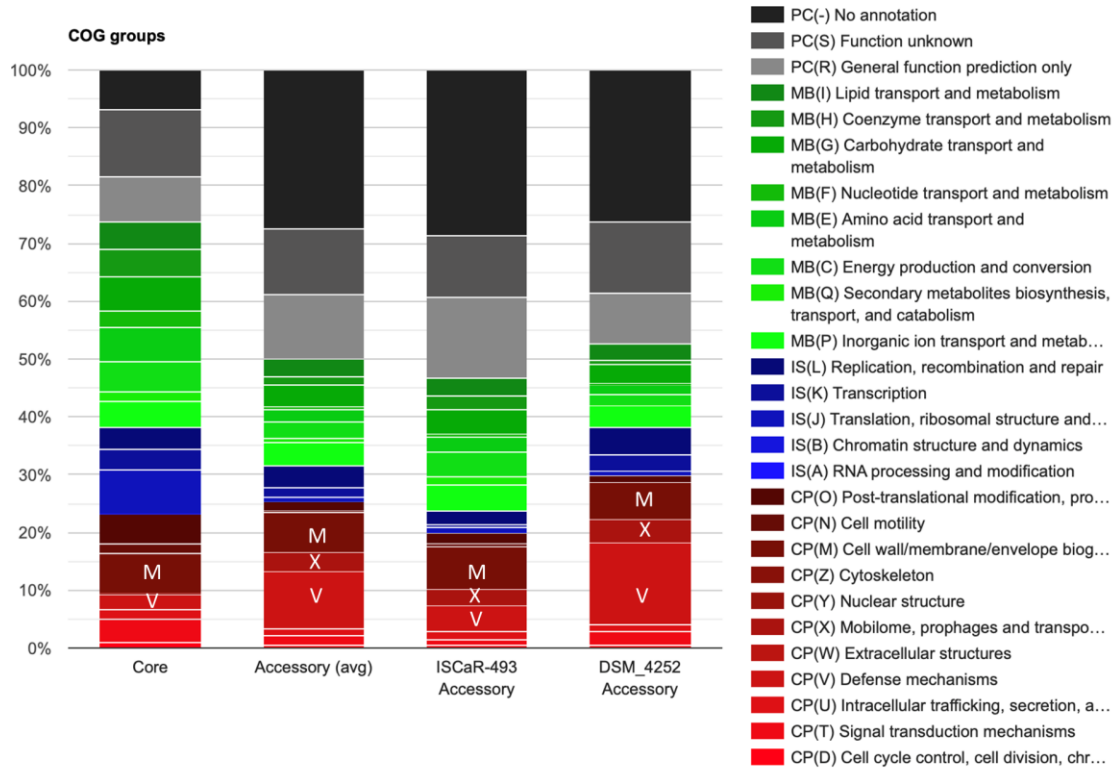


Figure 2. *Functional annotations of core and accessory genes in the R. marinus strains ISCIaR-493 and DSM 4252^T. The main functional categories are four: Cellular processes and signalling (CP) red stacks; Information storage and processing (IS) blue stacks; (MB) green stacks; poorly characterized (PC), grey stacks: Single letters designate the COG subcategories. The core genome consists of the predicted genes that are common in both strains. Each category or subcategory is graphed as a percentage of the total number of genes in the core or accessory genomes.*

The number of total genes was 2890 and 2937 for ISCaR-493 and DSM 4252^T, respectively and protein-coding genes were about 98.3% thereof for each strain. 2609 protein-coding genes belonged to the common core having 50% identity across at least 50% of the protein.

74% of the core protein-coding genes could be assigned to COG functional categories [82]. The remaining protein-coding core genes (26%) did not get COG IDs and had unknown (S) or poorly characterized functions (R). 230 protein-coding genes were unique to ISCaR-493 and 275 genes were unique to DSM 4252^T, a total of 505 genes that comprised the peripheric gene fraction. About 50% of the protein-coding genes in the peripheric fractions of both strains could not be assigned function compared with approximately 26% in the common core (Figure 2).

Accessory genome

Compared with the core genome the peripheric fractions of both strains were proportionally enriched in the following COG subcategories: V (Defense), M (Cell wall/membrane/envelop biogenesis), and X (Mobilome: Prophages and transposons) (Figure 2). Two restriction enzymes were shown to be only in the peripheric genome of DSM 4252^T, including the *Rma*I 4 cutter. This reflects the different restriction phenotypes observed [21] and the probably explains the facilitated uptake of foreign DNA in ISCaR-493 compared to DSM 4252^T.

A total of 578 genes were used in the metabolic network modelling of *R. marinus* 4252^T and the great majority of them belonged to the core genome. Only 7 model genes were not found in the genome of ISCaR-493 and only three of them did not have isozymes in the genome that showed high similarity to DSM 4252^T genes (Table 3).

Strain DSM 4252^T contains two genes encoding xylanases while strain ISCaR-493 contains only one homologue. Both strains grow on xylan as the sole carbon source (Table 2 for DSM 4252^T, data not shown for ISCaR-493). Strain DSM 4252^T contains four genes encoding alginate lyases and one of them is missing in ISCaR-493. The latter strain grows well in a medium with alginate as the sole carbon source [50], suggesting that the three alginate lyases are sufficient to degrade alginate for utilization.

Many enzymes take part in EPS (the envelope polysaccharides) biosynthesis and assembly, and their corresponding genes were all found in the core genome. A putative o-antigen polymerase (wzy) found in the accessory genome of strain ISCaR-493 showed a low similarity to a functionally corresponding protein in DSM 4252^T (E-value 0.004). However, the encoded gene showed high similarity to genes in more distantly related bacteria annotated as o-

antigen ligase and o-antigen polymerase. This enzyme activity is essential for EPS synthesis and must be present in ISCaR-493 as EPS is produced [10]. The relatively high content of genes in the accessory genomes belonging to subcategory M (Cell wall/membrane/envelop biogenesis), including glycosyl transferases and sulfotransferases, indicates that the EPS of the two strains may be different in sugar composition and sulfatation patterns and this may explain the differently aggregating phenotypes of the strains.

Table 3. *The model genes that are not present in R. marinus strain ISCaR-493. The model gene IDs from DSM 4252^T are listed, with corresponding annotations and the gene IDs of the isozymes, when applicable.*

DSM 4252^T gene ID	Annotation	Isozyme present
RMAR_RS02620	Dihydroneopterin triphosphate pyrophosphatase	Yes (RMAR_RS06180)
RMAR_RS14780	O-antigen polymerase (wzy)	No
RMAR_RS05305	Xylanase	No
RMAR_RS14585	Alginate lyase	No
RMAR_RS11910	Asparagine synthase (glutamine-hydrolyzing)	Yes (RMAR_RS10310)
RMAR_RS12230	Beta-ketoacyl-ACP synthase	Yes (RMAR_RS14045)
RMAR_RS12885	N-acetylglutamate synthase	Yes (RMAR_RS13420)

The above analysis suggests that the Rmarinus_578 model can be used for both strains DSM 4252^T and ISCaR-493.

3.3 Model validation for strains DSM 4252^T and ISCaR-493

Experimental growth data was obtained for strains DSM 4252^T and ISCaR-493 in bioreactors, along with measured uptake of glucose and pyruvate and secretion of acetate and lactate (Supplementary file 1). The average rates of two replicates (Supplementary file 1) was used to constrain the model to validate the accuracy of growth predictions (Figure 3a). The experimental growth data showed that the growth of *R. marinus* DSM 4252^T did not follow the typical batch growth curve of bacteria. A true exponential phase was not observed throughout the growth phase. Instead, the apparent specific growth rate decreased over time until stationary phase was reached (Figure 3b). The specific growth rate in strain ISCaR-493 was closer to being constant, with exponential growth during a longer period (Figure 3b). Data from time points 3-6 for strain DSM 4252^T and 1-5 for strain ISCaR-493 was used here. This analysis showed that the model accurately predicts growth for both strains (Figure 3a).

For both strains, but more so for DSM 4252^T, secretion rates of lactate and acetate increased during the growth phase (Supplementary file 1). A decrease in growth rate during batch cultivations has been observed in other bacteria, such as *E. coli* [83] where the main reason was oxygen limitation that could also lead to an increase in organic acid secretion. The cultivations here were carried out with high aeration as oxygen levels were kept fixed at 40% pO_2. A plausible explanation for why the cells would experience oxygen limitation in a medium with excess oxygen levels is local limitation due to cell aggregation [84]. Aggregation of several *R. marinus* strains has been reported previously [22], especially in DSM 4252^T and *R. marinus* is also shown to produce exopolysaccharides [10], which can cause cells to aggregate [85].

When the model was optimized for growth, without oxygen limitation and free secretion of acids, it did not predict any acid production and the predicted growth rate was slightly higher than observed *in vivo.* When oxygen was limited in the model, the predicted growth rate decreased, and the model predicted lactate secretion (data not shown). Experimental data showed that lactate was first secreted, followed by acetate (Supplementary file 1). The model predicted slightly higher growth rate when lactate was the sole acid produced, opposed to when it was forced to also produce acetate.

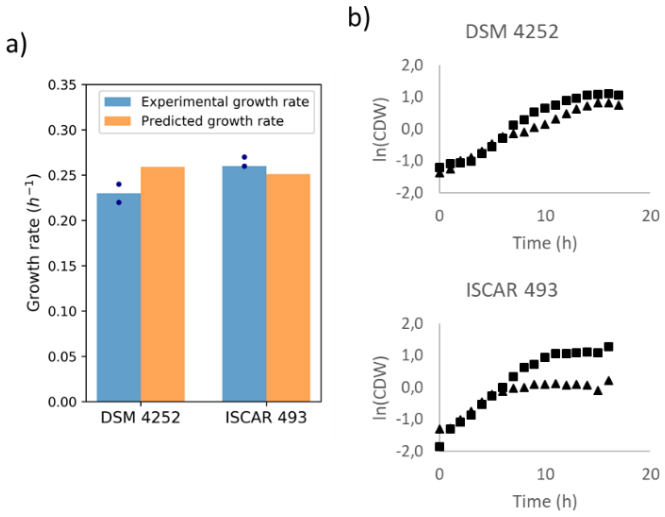


Figure 3. *Experimental growth data for strains DSM 4252^T and ISCaR-493 was used to validate model growth predictions. The model was constrained with uptake (glucose and pyruvate) and secretion (lactate and acetate) rates observed in vivo (Supplementary file 1) and optimized for growth. Predicted growth rates were compared to growth rates observed in vivo (**a**). Experimental data from the exponential growth phase of two replicates (▲ and ■), obtained early in the growth phase (time points 3-6 for strain DSM 4252 and 1-5 for strain ISCaR-493), was used (**b**).*

3.4 Carotenoid production and growth of *R. marinus* ISCaR-493

To better understand the carotenoid production in *R. marinus,* a cultivation experiment comparing different conditions was performed. Besides obtaining high yields of carotenoids per cell, high cell density is important for achieving high yields of carotenoids. Therefore, both extracted carotenoids (from 1 mL of cells diluted to OD620 nm = 1) and cell densities were measured from cultivations after 24 hours (Figure 4). The ISCaR-493 strain was used in this experiment, as it can be genetically modified and thus likely to be used for future cell factory designs

Glucose and pyruvate

R. marinus can grow on several monosaccharides, as predicted by the model. However, we have often observed better growth on oligo- and polysaccharides (data not shown). Growth of strain ISCaR-493 in defined medium with glucose (1%) as the sole carbon source resulted neither in high cell density nor high carotenoid production (Figure 4). *R. marinus* can utilize pyruvate as the sole carbon source (Table 2). Pyruvate is used in several pathways essential for growth and is the substrate, together with glyceraldehyde 3-phosphate, in the first step of the MEP terpenoid pathway (Figure 1 To increase both cell density and carotenoid production, pyruvate (0.09%) was added to the glucose-based medium. This resulted in increased carotenoid production and highly increased cell density (Figure 4). Visually, these cultures exhibited much stronger red color than the glucose cultures, which can be explained by both increased carotenoids yields and higher cell densities.

Impact of light

During the model reconstruction, a homolog for a light dependent regulatory gene was found in the carotenoid gene cluster in *R. marinus*. This indicates that carotenoid production in *R. marinus* is light induced (section 3.1). This was investigated here by cultivating ISCaR-493 in glucose (1%) and pyruvate (0.09%) medium in the dark and compared to the corresponding cultures grown in the light. The former cultures were colorless and the lack of carotenoids was confirmed by measurements (Figure 4).

Alginate

R. marinus can grow on many different polysaccharides (Table 2), making it an interesting candidate for processing 2nd or 3rd generation biomass, such as seaweed. Alginate is one of the major polysaccharides of brown algae. The products from alginate degradation are pyruvate and glyceraldehyde 3-phosphate (Figure 1), which are the same metabolites as used in the first step of the MEP terpenoid pathway. This raised the question whether *R. marinus* produces more carotenoids when grown on alginate, since it produces the two metabolites needed for the biosynthesis concurrently and in equal amounts. Cultivations in defined medium with alginate (1%) as the sole carbon source showed less cell density compared to glucose and pyruvate, but highly increased carotenoid production (Figure 4).

Glucose and pyruvate in equal quantities

To further examine if the availability of glyceraldehyde 3-phosphate and pyruvate in equal amounts results in higher carotenoid production, cultivation in defined medium with glucose (0.5%) and pyruvate (0,25%) was investigated. These cultivations showed lower cell density and higher carotenoid production compared to growth on glucose (1%) and pyruvate (0.09%) (Figure 4). The increased carotenoid production could be due to the equal availability of the two metabolites. Another possibility is that increased concentration of pyruvate alone in the medium caused higher carotenoid production.

Pyruvate

To examine if pyruvate alone affects the carotenoid production, two additional cultivations were set up, with pyruvate (0.09% and 0.18%) as the sole carbon source. The cell density in these cultures was low, only increased slightly after inoculation. This indicated that ISCaR-493 struggles to grow in liquid defined medium with pyruvate as the sole carbon source, which was surprising as growth was observed on agar medium (Table 2). The carotenoids per fixed cell density in the pyruvate cultures were much higher compared to cultures on glucose (1%) and pyruvate (0,09%). Additionally, increased pyruvate concentration resulted in increased carotenoid production (Figure 3). This suggests that the pyruvate is used for carotenoid production. Producing glyceraldehyde 3-phosphate from pyruvate costs energy (gluconeogenesis) and it cannot be determined from this data if this is the case for the

observed growth. However, glycogen is an alternative source of glyceraldehyde 3-phosphate. The amount of glycogen in the biomass of *R. marinus* has been estimated as 14% [45] and is relatively high compared to other bacteria. Inclusion that could possibly contain glycogen can be discerned on electron micrographs of *R. marinus* [7]. Considering the natural habitat of *R. marinus* in coastal hot springs, it is not unreasonable to assume that it accumulates high levels of glycogen. Due to tides, the availability of nutrients in the surroundings of *R. marinus* varies widely and it is likely that the bacterium stores carbohydrates when they are in abundance in the environment. Since little or no growth was observed on pyruvate in liquid cultures it is likely that the cells experienced starvation and therefore started the breakdown of glycogen and carotenoid production. This was also seen for the negative control cultures without a carbon source (Figure 4). The cell density did not increase from inoculation, while the carotenoid production did.

Addition of the *dxs* gene from *T. thermophilus*

In an effort to increase carotenoid yields, the *dxs* gene from *T. thermophilus* was cloned on a shuttle vector into *R. marinus* strain SB-62 (ISCaR-493 derivative, Δ*trpB*Δ*purA*), resulting in the mutant strain TK-4 (Δ*trpB*Δ*purA::trpBdxs*$_{T.thermophilus}$) (Supplementary file 5). The *dxs* gene encodes 1-deoxy-D-xylulose-5-phosphate synthase (DXS), which catalyzes the first step in the MEP terpenoid pathway (section 3.1) and could not be identified in the genomes of *R. marinus*. Compared to ISCaR-493, cultivation of TK-4 resulted in lower cell density but highly increased carotenoid production. Presumably the added *dxs* gene resulted in a higher flux of carbons through the terpenoid and carotenoid pathways. However, it is also possible that this strain struggles to grow and responds by producing carotenoids. The dramatically lower cell density compared to ISCaR-493 can most likely be explained by the metabolic burden caused by the replication of the shuttle vector and the expression of its genes. Inserting the *dxs* gene in to the chromosome could reduce such effects.

In summary, these experiments showed that the highest cell density was obtained in glucose medium supplemented with pyruvate, while higher carotenoid production was observed during growth on alginate, with pyruvate added to a glucose-based medium and in the presence of light. It also showed that the carotenoid production per cell increased during starvation, indicating that yields can potentially be increased by either allowing the culture to reach and stay in stationary phase or transfer the cells after growth to new medium with

limited or no carbon source. The motivation for the latter is that after the growth phase, the medium might not be optimal, e.g. due to accumulation of by-products that alter the pH, and the cells might stay alive and produce carotenoids longer in fresh medium. Finally, cloning the *dxs* gene from *T. thermophilus* in *R. marinus* resulted in the highest yields of carotenoids, but much lower cell density than the wild type strain ISCaR-493.

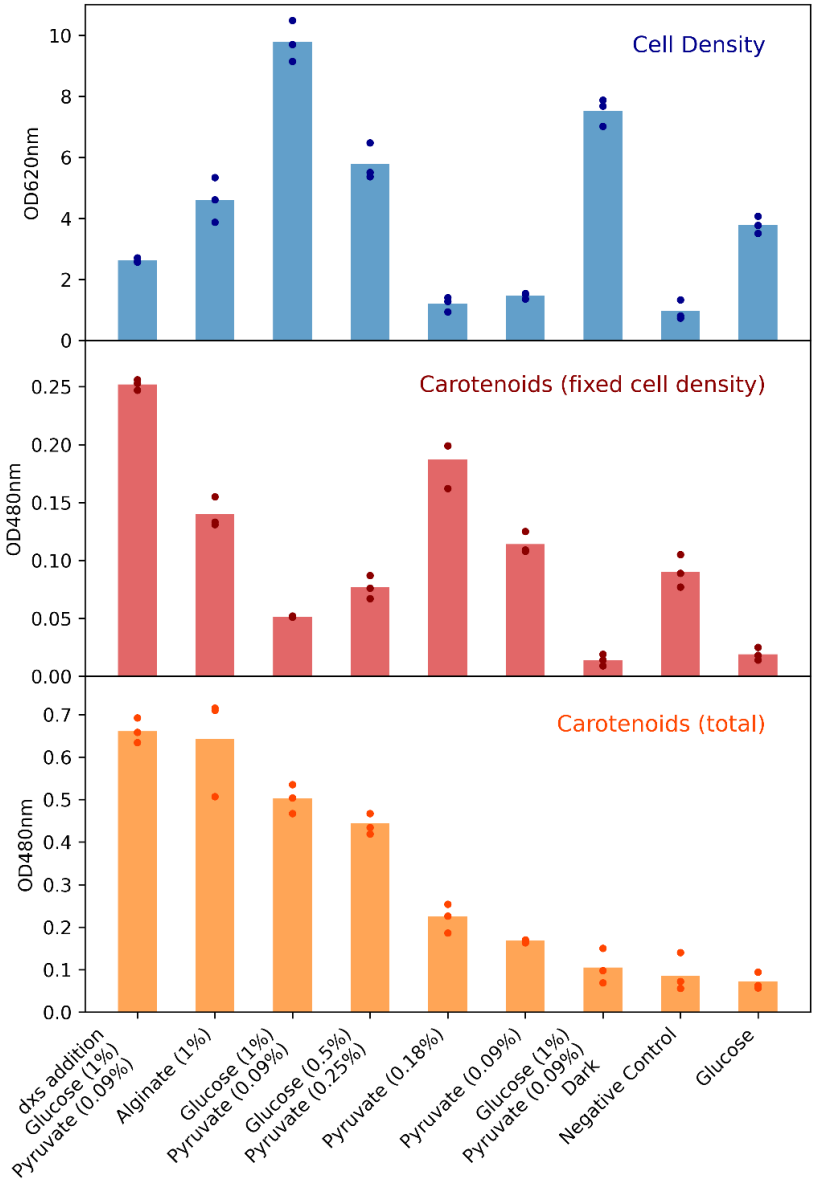


Figure 4. *Cell density (OD620nm) and carotenoid (OD480nm) production following growth of R. marinus strain ISCaR-493 for 24 hours on glucose; mixture of glucose and pyruvate under light and dark conditions; alginate; pyruvate and without any carbon sources. Additionally, the modified strain TK-4 (ΔtrpBΔpurA::trpBdxs$_{T.thermophilus}$) was grown on a mixture of glucose and pyruvate. The carotenoids were always extracted from a fixed number of cells (1 mL of OD620nm = 1). The total amount of carotenoids was calculated by multiplying the cell density by the measured carotenoids. The average cell density of each culture condition is represented by a blue bar, the average measured carotenoid value as a red bar and the average total carotenoids by an orange bar. Dots represent individual replicates.*

4. Conclusions

A manually curated genome-scale metabolic model of *R. marinus* DSM 4252^T was reconstructed and made publicly available (https://github.com/steinng/rmarinus). Experimental data from the literature and from this study was used to curate and validate the model. This includes growth data on various carbon sources, bioreactor cultivations and HPLC measurements of main metabolites, used for model validation, multiple studies on different metabolic pathways, components, genes and enzymes, and data on biomass components, which was used to formulate a species-specific biomass objective function.

The model was also evaluated for use with *R. marinus* ISCaR-493, from which the genetically modified SB-62 (Δ*trpB*Δ*purA*) was derived. The genome of strain ISCaR-493 was sequenced and the resulting draft genome was compared to that of strain DSM 4252^T. This analysis showed that only seven model genes were absent in strain ISCaR-493 and four of them were replaced by genes encoding isozymes that exhibited high similarity to the DSM 4252^T enzymes. The remaining three genes are involved in EPS formation and xylan- and alginate degradation. EPSs of both strains have been previously studied [10] and shown to be of similar structures. It was also observed that strain ISCaR-493 grows well in defined medium with xylan and alginate as the sole carbon sources. In conclusion, this analysis suggests that the model is applicable for both strains DSM 4252^T and ISCaR-493. Both strains should be considered when any future changes or additions to the model reconstruction are made. Data on growth and metabolites was used to constrain the model and compare the experimental and simulated growth rates. This revealed that the model predicts correct growth rates for both strains.

Different aspects of the metabolism of *R. marinus* were reviewed during the reconstruction process. Here, an emphasis was on those with a potential biotechnological aspect, carotenoids in particular. Cell density and carotenoid production of strain ISCaR-493 grown at different conditions were investigated. Pyruvate addition to a glucose-based medium, highly increased cell density was obtained. Carotenoid production varied considerably under different growth conditions. Higher carotenoid yields were observed when pyruvate was present in the growth medium, alginate was used as the sole carbon source, cultivating the cells in light conditions and the cells experienced starvation. Additionally, we cloned the *dxs*

gene from *T. thermophilus* on a shuttle vector into *R. marinus* and cultivation of the resulting mutant showed low cell density compared to ISCaR-493, but higher carotenoid production.

With its thermostable enzymes, wide range of potential carbon sources for growth and marketable products, *R. marinus* application potential is highly relevant in biotechnology and biorefineries. A genome-scale metabolic model helps us to understand its metabolism and should be useful in future strain designs.

Declaration of competing interest

None

CRediT author statement

Thordis Kristjansdottir: Conceptualization, Methodology, Investigation, Writing – Original Draft, Writing – Review and Editing, Visualization. **Gudmundur O. Hreggvidsson:** Conceptualization, Investigation, Supervision, Funding acquisition, Writing – Original Draft, Writing – Review and Editing. **Sigmar Karl Stefansson:** Methodology, Investigation, Writing – Review and Editing, Visualization. **Elisabet Eik Gudmundsdottir:** Methodology, Investigation, Writing – Review and Editing. **Snaedis H. Bjornsdottir:** Methodology, Investigation, Writing – Review and Editing. **Olafur H. Fridjonsson:** Conceptualization, Writing – Review and Editing. **Eva Nordberg Karlsson:** Conceptualization, Writing – Review and Editing. **Justine Vanhalst:** Investigation, Writing – Review and Editing. **Birkir Reynisson:** Investigation, Writing – Review and Editing. **Steinn Gudmundsson:** Conceptualization, Supervision, Project administration, Funding acquisition, Writing – Review and Editing.

Funding

This work was supported by the Marine Biotechnology ERA-NET, *ThermoFactories,* project grant number 5178–00003B; the Technology Development fund in Iceland, grant number 159004-0612; the Icelandic Research fund, *ThermoExplore*, project grant number 207088-051 and the Novo Nordisk Foundation (NNF18OC0034792).

Acknowledgements

The authors would like to thank Emanuel Y. C. Ron and Roya R. R. Sardari for the advice on how to cultivate *R. marinus* in bioreactors and to extract carotenoids, respectively, and Louna Maignien for assisting with the bioreactor cultivations.

References

[1] R. Munoz, R. Rosselló-Móra, and R. Amann, "Revised phylogeny of Bacteroidetes and proposal of sixteen new taxa and two new combinations including Rhodothermaeota phyl. nov.," *Syst. Appl. Microbiol.*, vol. 39, no. 5, pp. 281–296, 2016.

[2] G. A. Alfredsson, J. K. Kristjansson, S. Hjorleifsdottir, and K. O. Stetter, "Rhodothermus marinus, gen.nov., sp. nov., a Thermophilic, Halophilic Bacterium from Submarine Hot Springs in Iceland," *J. Gen. Microbiol.*, no. 134, pp. 299–306, 1988.

[3] S. Halldórsdóttir *et al.*, "Cloning, sequencing and overexpression of a Rhodothermus marinus gene encoding a thermostable cellulase of glycosyl hydrolase family 12," *Appl. Microbiol. Biotechnol.*, vol. 49, no. 3, pp. 277–284, 1998.

[4] M. Krah, R. Misselwitz, O. Politz, K. K. Thomsen, H. Welfle, and R. Borriss, "The laminarinase from thermophilic eubacterium Rhodothermus marinus," *Eur. J. Biochem.*, vol. 257, no. 1, pp. 101–111, 1998.

[5] E. Nordberg Karlsson, E. Bartonek-Roxå, and O. Holst, "Cloning and sequence of a thermostable multidomain xylanase from the bacterium Rhodothermus marinus.," *Biochim. Biophys. Acta*, vol. 1353, no. 2, pp. 118–124, Aug. 1997.

[6] E. Nordberg Karlsson, E. Bartonek-Roxå, and O. Holst, "Evidence for substrate binding of a recombinant thermostable xylanase originating from Rhodothermus marinus," *FEMS Microbiol. Lett.*, vol. 168, no. 1, pp. 1–7, Nov. 1998.

[7] S. H. Bjornsdottir *et al.*, "Rhodothermus marinus : physiology and molecular biology," *Extremophiles*, vol. 10, pp. 1–16, 2006.

[8] E. Y. C. Ron, R. R. R. Sardari, R. Anthony, E. W. J. van Niel, G. O. Hreggvidsson, and E. Nordberg-Karlsson, "Cultivation technology development of Rhodothermus marinus DSM 16675," *Extremophiles*, vol. 23, no. 6, pp. 735–745, 2019.

[9] K. Hamana, H. Hamana, M. Niitsu, K. Samejima, and S. Matsuzaki, "Distribution of unusual long and branched polyamines in thermophilic eubacteria belonging to "Rhodothermus, " Thermus and Thermonema," *J. Gen. Appl. Microbiol.*, vol. 38, no. 6, pp. 575–584, 1992.

[10] R. R. R. Sardari *et al.*, "Evaluation of the production of exopolysaccharides by two strains of the thermophilic bacterium Rhodothermus marinus," *Carbohydr. Polym.*, vol. 156, pp. 1–8, 2017.

[11] E. Y. C. Ron *et al.*, “Characterization of carotenoids in Rhodothermus marinus,” *Microbiologyopen*, vol. 7, no. 1, 2018.

[12] O. C. Nunes, C. M. Manaia, M. S. Da Costa, and H. Santos, “Compatible Solutes in the Thermophilic Bacteria Rhodothermus marinus and ‘Thermus thermophilus,’” *Appl. Environ. Microbiol.*, vol. 61, no. 6, pp. 2351–2357, Jun. 1995.

[13] L. Moreira, M. F. Nobre, I. Sa-correia, and M. S. Da Costa, “Genomic Typing and Fatty Acid Composition of Rhodothermus marinus,” *Syst. Appl. Microbiol.*, vol. 19, no. 1, pp. 83–90, 1996.

[14] O. C. Nunes, M. M. Donato, C. M. Manaia, and M. S. Da Costa, “The Polar Lipid and Fatty Acid Composition of Rhodothermus Strains,” *Syst. Appl. Microbiol.*, vol. 15, no. 1, pp. 59–62, 1992.

[15] A. López-Contreras *et al.*, “Biorefinery Approach to the Use of Macroalgae as Feedstock for Biofuels,” in *Algal Biofuels*, L. Pereira, Ed. CRC Press, 2017.

[16] E. J. O’Brien, J. M. Monk, and B. O. Palsson, “Using Genome-scale Models to Predict Biological Capabilities.,” *Cell*, vol. 161, no. 5, pp. 971–987, May 2015.

[17] S. H. Bjornsdottir, O. H. Fridjonsson, J. K. Kristjansson, and G. Eggertsson, “Cloning and expression of heterologous genes in Rhodothermus marinus,” *Extremophiles*, vol. 11, no. 2, pp. 283–293, 2007.

[18] T. Kristjansdottir *et al.*, “Engineering the carotenoid biosynthetic pathway in Rhodothermus marinus for lycopene production,” *Metab. Eng. Commun.*, p. e00140, 2020.

[19] A. Vershinin, “Biological functions of carotenoids - Diversity and evolution,” *BioFactors*, vol. 10, no. 2–3, pp. 99–104, 1999.

[20] M. Nolan *et al.*, “Complete genome sequence of Rhodothermus marinus type strain (R-10 T),” *Stand. Genomic Sci.*, vol. 1, no. 3, pp. 283–291, 2009.

[21] J. Rönkä, S. Hjörleifsdóttir, T. Tenkanen, K. Pitkänen, P. Mattila, and J. K. Kristjansson, “RmaI, a type II restriction endonuclease from Rhodothermus marinus which recognizes 5’ CTAG 3’,” *Nucleic Acids Res.*, vol. 19, no. 10, p. 2789, May 1991.

[22] S. H. Bjornsdottir, S. H. Thorbjarnardottir, and G. Eggertsson, “Establishment of a gene transfer system for Rhodothermus marinus,” *Appl. Microbiol. Biotechnol.*, vol. 66, no. 6, pp. 675–682, 2005.

[23] E. Degryse, N. Glansdorff, and A. Piérard, “A comparative analysis of extreme thermophilic bacteria belonging to the genus Thermus,” *Arch. Microbiol.*, vol. 117, no. 2, pp. 189–196, 1978.

[24] E. A. Wolin, M. J. Wolin, and R. S. Wolfe, “Formation of Methane by Bacterial Extracts,” *J. Biol. Chem.*, vol. 238, no. 8, pp. 2882–2886, 1963.

[25] A. Blücher, E. N. Karlsson, and O. Holst, “Substrate-dependent production and some properties of a thermostable, α-galactosidase from Rhodothermus marinus,” *Biotechnol. Lett.*, vol. 22, no. 8, pp. 663–669, 2000.

[26] L. Dahlberg, O. Holst, and J. K. Kristjansson, “Thermostable xylanolytic enzymes from Rhodothermus marinus grown on xylan,” *Appl. Microbiol. Biotechnol.*, vol. 40, no. 1, pp. 63–68, 1993.

[27] J. H. Miller, *Experiments in molecular genetics*. Cold Spring Harbor Laboratory, 1972.

[28] S. H. Bjornsdottir, O. H. Fridjonsson, G. O. Hreggvidsson, and G. Eggertsson, “Generation of targeted deletions in the genome of Rhodothermus marinus,” *Appl. Environ. Microbiol.*, vol. 77, no. 15, pp. 5505–5512, 2011.

[29] S. Andrews, “FastQC, Babraham Bioinformatics,” 2010. [Online]. Available: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

[30] A. M. Bolger, M. Lohse, and B. Usadel, “Trimmomatic: a flexible trimmer for Illumina sequence data.,” *Bioinformatics*, vol. 30, no. 15, pp. 2114–2120, Aug. 2014.

[31] S. Nurk *et al.*, “Assembling Genomes and Mini-metagenomes from Highly Chimeric Reads BT - Research in Computational Molecular Biology,” 2013, pp. 158–170.

[32] T. Tatusova *et al.*, “NCBI prokaryotic genome annotation pipeline,” *Nucleic Acids Res.*, vol. 44, no. 14, pp. 6614–6624, Aug. 2016.

[33] I. Thiele and B. Ø. Palsson, “A protocol for generating a high-quality genome-scale metabolic reconstruction,” *Nat. Protoc.*, vol. 5, no. 1, pp. 93–121, 2010.

[34] C. S. Henry, M. Dejongh, A. A. Best, P. M. Frybarger, B. Linsay, and R. L. Stevens, “High-throughput generation , optimization and analysis of genome-scale metabolic models,” *Nat. Biotechnol.*, vol. 28, no. 9, pp. 969–974, 2010.

[35] J. D. Orth *et al.*, “A comprehensive genome-scale reconstruction of Escherichia coli metabolism-2011,” *Mol. Syst. Biol.*, vol. 7, no. 535, pp. 1–9, 2011.

[36] Z. A. King *et al.*, “BiGG Models: A platform for integrating, standardizing and sharing genome-scale models,” *Nucleic Acids Res.*, vol. 44, no. D1, pp. D515–D522, Oct. 2015.

[37] 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, no. D1, pp. D542–D549, Nov. 2018.

[38] M. Kanehisa and S. Goto, “KEGG: kyoto encyclopedia of genes and genomes.,” *Nucleic Acids Res.*, vol. 28, no. 1, pp. 27–30, Jan. 2000.

[39] R. Caspi *et al.*, “The MetaCyc database of metabolic pathways and enzymes,” *Nucleic Acids Res.*, vol. 46, no. D1, pp. D633–D639, Oct. 2017.

[40] S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, “Basic local alignment search tool.,” *J. Mol. Biol.*, vol. 215, no. 3, pp. 403–410, Oct. 1990.

[41] A. Ebrahim, J. A. Lerman, B. O. Palsson, and D. R. Hyduke, “COBRApy: COnstraints-Based Reconstruction and Analysis for Python,” *BMC Syst. Biol.*, vol. 7, 2013.

[42] D. A. Fell and J. R. Small, “Fat synthesis in adipose tissue. An examination of stoichiometric constraints,” *Biochem. J.*, vol. 238, pp. 781–786, 1986.

[43] J. M. Savinell and B. O. Palsson, “Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism,” *J. Theor. Biol.*, vol. 154, no. 4, pp. 421–454, 1992.

[44] C. Lieven *et al.*, “MEMOTE for standardized genome-scale metabolic model testing,” *Nat. Biotechnol.*, vol. 38, no. 3, pp. 272–276, 2020.

[45] L. T. Cordova, R. M. Cipolla, A. Swarup, C. P. Long, and M. R. Antoniewicz, “13C metabolic flux analysis of three divergent extremely thermophilic bacteria: Geobacillus sp. LC300, Thermus thermophilus HB8, and Rhodothermus marinus DSM 4252,” *Metab. Eng.*, vol. 44, no. October, pp. 182–190, 2017.

[46] G. O. Hreggvidsson, E. Kaiste, O. Holst, G. Eggertsson, A. Palsdottir, and J. K. Kristjansson, “An extremely thermostable cellulase from the thermophilic eubacterium Rhodothermus marinus,” *Appl. Environ. Microbiol.*, vol. 62, no. 8, pp. 3047–3049, 1996.

[47] K. Z. G. Ara *et al.*, “Characterization and diversity of the complete set of GH family 3 enzymes from Rhodothermus marinus DSM 4253,” *Sci. Rep.*, vol. 10, no. 1, p. 1329, 2020.

[48] G. I. Guzmán *et al.*, “Enzyme promiscuity shapes adaptation to novel growth substrates,” *Mol. Syst. Biol.*, vol. 15, no. 4, p. e8462, Apr. 2019.

[49] L. Jolly, P. Ferrari, D. Blanot, J. Van Heijenoort, F. Fassy, and D. Mengin-Lecreulx, “Reaction mechanism of phosphoglucosamine mutase from Escherichia coli,” *Eur. J. Biochem.*, vol. 262, no. 1, pp. 202–210, 1999.

[50] L. Allahgholi *et al.*, “Composition analysis and minimal treatments to solubilize polysaccharides from the brown seaweed Laminaria digitata for microbial growth of thermophiles,” *J. Appl. Phycol.*, 2020.

[51] Y. Qin, “Alginate fibres: an overview of the production processes and applications in wound management,” *Polym. Int.*, vol. 57, no. 2, pp. 171–180, Feb. 2008.

[52] G. O. Hreggvidsson *et al.*, “International Publication Number EP 3092247 A1.” 2016.

[53] G. O. Hreggvidsson and O. H. Fridjonsson, “International Publication Number WO 2020/044379 A1.” 2020.

[54] J. Preiss and G. Ashwell, “Alginic Acid Metabolism in Bacteria. The Enzymatic Reduction of 4-deoxy-L-erythro-5-hexoseulose uronic acid to 2-keto-3-deoxy-D-guconic acid,” *J. Biol. Chem.*, vol. 237, no. 2, pp. 317–321, 1962.

[55] R. Takase, A. Ochiai, B. Mikami, W. Hashimoto, and K. Murata, “Molecular identification of unsaturated uronate reductase prerequisite for alginate metabolism in Sphingomonas sp. A1.,” *Biochim. Biophys. Acta*, vol. 1804, no. 9, pp. 1925–1936, Sep. 2010.

[56] B. F. Lutnaes, Å. Strand, S. K. Pétursdóttir, and S. Liaaen-Jensen, “Carotenoids of thermophilic bacteria - Rhodothermus marinus from submarine Icelandic hot springs,” *Biochem. Syst. Ecol.*, vol. 32, no. 5, pp. 455–468, 2004.

[57] H. Katsuki and K. Bloch, “Studies on the biosynthesis of ergosterol in yeast. Formation of methylated intermediates.,” *J. Biol. Chem.*, vol. 242, no. 2, pp. 222–227, Jan. 1967.

[58] F. Lynen, “Biosynthetic pathways from acetate to natural products.,” *Pure Appl. Chem.*, vol. 14, no. 1, pp. 137–167, 1967.

[59] M. Rohmer, M. Knani, P. Simonin, B. Sutter, and H. Sahm, “Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate,” *Biochem. J.*, vol. 295 (Pt 2, no. Pt 2, pp. 517–524, Oct. 1993.

[60] S. Sauret-Güeto, E. M. Urós, E. Ibáñez, A. Boronat, and M. Rodríguez-Concepción, “A mutant pyruvate dehydrogenase E1 subunit allows survival of Escherichia coli strains defective in 1-deoxy-d-xylulose 5-phosphate synthase,” *FEBS Lett.*, vol. 580, no. 3, pp. 736–740, 2006.

[61] J. Kirby *et al.*, “Enhancing Terpene Yield from Sugars via Novel Routes to 1-Deoxy-d-Xylulose 5-Phosphate,” *Appl. Environ. Microbiol.*, vol. 81, no. 1, pp. 130 LP – 138, Jan. 2015.

[62] T. J. Erb *et al.*, “A RubisCO like protein links SAM metabolism with isoprenoid biosynthesis,” *Nat Chem Biol.*, vol. 8, no. 11, pp. 926–932, 2013.

[63] M. C. Pérez-Marín, S. Padmanabhan, M. C. Polanco, F. J. Murillo, and M. Elías-Arnanz, “Vitamin B12 partners the CarH repressor to downregulate a photoinducible promoter in Myxococcus xanthus,” *Mol. Microbiol.*, vol. 67, no. 4, pp. 804–819, 2008.

[64] H. Takano *et al.*, “Involvement of CarA/LitR and CRP/FNR family transcriptional regulators in light-induced carotenoid production in Thermus thermophilus,” *J. Bacteriol.*, vol. 193, no. 10, pp. 2451–2459, 2011.

[65] H. Takano, S. Obitsu, T. Beppu, and K. Ueda, “Light-induced carotenogenesis in Streptomyces coelicolor A3(2): identification of an extracytoplasmic function sigma factor that directs photodependent transcription of the carotenoid biosynthesis gene cluster,” *J. Bacteriol.*, vol. 187, no. 5, pp. 1825–1832, Mar. 2005.

[66] H. Takano *et al.*, “Role and function of LitR, an adenosyl B12-bound light-sensitive regulator of Bacillus megaterium QM B1551, in regulation of carotenoid production,” *J. Bacteriol.*, vol. 197, no. 14, pp. 2301–2315, 2015.

[67] A. S. Fernandes, M. M. Pereira, and M. Teixeira, “The succinate dehydrogenase from the thermohalophilic bacterium Rhodothermus marinus: Redox-Bohr effect on heme

bL1," *J. Bioenerg. Biomembr.*, vol. 33, no. 4, pp. 343–352, 2001.

[68] A. S. Fernandes, M. M. Pereira, and M. Teixeira, "Purification and characterization of the complex I from the respiratory chain of Rhodothermus marinus," *J. Bioenerg. Biomembr.*, vol. 34, no. 6, pp. 413–421, 2002.

[69] M. M. Pereira, P. N. Refojo, G. O. Hreggvidsson, S. Hjorleifsdottir, and M. Teixeira, "The alternative complex III from Rhodothermus marinus - A prototype of a new family of quinol:electron acceptor oxidoreductases," *FEBS Lett.*, vol. 581, no. 25, pp. 4831–4835, 2007.

[70] B. J. Tindall, "Lipid composition of Rhodothermus marinus," *FEMS Microbiol. Lett.*, vol. 80, pp. 65–68, 1991.

[71] M. M. Pereira, J. N. Carita, R. Anglin, M. Saraste, and M. Teixeira, "Heme centers of Rhodothermus marinus respiratory chain. Characterization of its cbb3 oxidase.," *J. Bioenerg. Biomembr.*, vol. 32, no. 2, pp. 143–152, Apr. 2000.

[72] M. M. Pereira *et al.*, "The caa3 terminal oxidase of the thermohalophilic bacterium Rhodothermus marinus: a HiPIP:oxygen oxidoreductase lacking the key glutamate of the D-channel," *Biochim. Biophys. Acta - Bioenerg.*, vol. 1413, no. 1, pp. 1–13, 1999.

[73] A. F. Veríssimo, M. M. Pereira, A. M. P. Melo, G. O. Hreggvidsson, J. K. Kristjansson, and M. Teixeira, "A ba3 oxygen reductase from the thermohalophilic bacterium Rhodothermus marinus.," *FEMS Microbiol. Lett.*, vol. 269, no. 1, pp. 41–47, Apr. 2007.

[74] J. Schneider and V. F. Wendisch, "Biotechnological production of polyamines by bacteria: recent achievements and future perspectives.," *Appl. Microbiol. Biotechnol.*, vol. 91, no. 1, pp. 17–30, Jul. 2011.

[75] W. Fukuda, R. Hidese, and S. Fujiwara, "Long-Chain and Branched Polyamines in Thermophilic Microbes BT - Polyamines: A Universal Molecular Nexus for Growth, Survival, and Specialized Metabolism," T. Kusano and H. Suzuki, Eds. Tokyo: Springer Japan, 2015, pp. 15–25.

[76] Z. Silva, N. Borges, L. O. Martins, R. Wait, M. S. da Costa, and H. Santos, "Combined effect of the growth temperature and salinity of the medium on the accumulation of compatible solutes by Rhodothermus marinus and Rhodothermus obamensis.," *Extremophiles*, vol. 3, no. 2, pp. 163–172, May 1999.

[77] N. Borges, J. D. Marugg, N. Empadinhas, M. S. da Costa, and H. Santos, "Specialized roles of the two pathways for the synthesis of mannosylglycerate in osmoadaptation and thermoadaptation of Rhodothermus marinus.," *J. Biol. Chem.*, vol. 279, no. 11, pp. 9892–9898, Mar. 2004.

[78] M. M. Nielsen *et al.*, "Substrate and metal ion promiscuity in mannosylglycerate synthase," *J. Biol. Chem.*, vol. 286, no. 17, pp. 15155–15164, Apr. 2011.

[79] I. J. Mukti, R. R. R. Sardari, T. Kristjansdottir, G. O. Hreggvidsson, and E. Nordberg-Karlsson, "Growth and production of carotenoids and exopolysaccharides by the

extremophile Rhodothermus marinus DSM16675 in defined media," 2021.

[80] A. M. Feist *et al.*, "A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information.," *Mol. Syst. Biol.*, vol. 3, p. 121, 2007.

[81] M. Richter and R. Rosselló-Móra, "Shifting the genomic gold standard for the prokaryotic species definition," *Proc. Natl. Acad. Sci.*, vol. 106, no. 45, pp. 19126 LP – 19131, Nov. 2009.

[82] R. L. Tatusov, M. Y. Galperin, D. A. Natale, and E. V Koonin, "The COG database: a tool for genome-scale analysis of protein functions and evolution," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 33–36, Jan. 2000.

[83] M. Berney, H.-U. Weilenmann, J. Ihssen, C. Bassin, and T. Egli, "Specific growth rate determines the sensitivity of Escherichia coli to thermal, UVA, and solar disinfection," *Appl. Environ. Microbiol.*, vol. 72, no. 4, pp. 2586–2593, Apr. 2006.

[84] A. K. Wessel *et al.*, "Oxygen limitation within a bacterial aggregate," *MBio*, vol. 5, no. 2, pp. e00992–e00992, Apr. 2014.

[85] U. U. Nwodo, E. Green, and A. I. Okoh, "Bacterial exopolysaccharides: functionality and prospects," *Int. J. Mol. Sci.*, vol. 13, no. 11, pp. 14002–14015, Oct. 2012.