Refactoring metabolic pathways for synthon production from renewable carbon sources

Claudio Jose Remedios Frazao

Université Fédérale
Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Institut National des Sciences Appliquées de Toulouse

Présentée et soutenue par

Cláudio José REMEDIOS FRAZÃO

Le 29 octobre 2018

Refactoring voie métabolique pour la production de synthon à partir de sources de carbone renouvelables

Ecole doctorale : **SEVAB - Sciences Ecologiques, Vétérinaires, Agronomiques et Bioingenieries**

Spécialité : **Ingénieries microbienne et enzymatique**

Unité de recherche :
LISBP - Laboratoire d'Ingénierie des Systèmes Biologiques et des Procédés

Thèse dirigée par
Jean-marie FRANCOIS

Jury

M. Pablo NIKEL, Rapporteur
Mme Cecília ROQUE, Rapporteur
M. Brian JESTER, Examinateur
M. Guy LIPPENS, Examinateur
M. Jean-marie FRANCOIS, Directeur de thèse

Refactoring metabolic pathways towards synthon production from renewable carbon sources

by

Cláudio José Remédios Frazão

Thesis submitted for the doctoral degree
in Microbial and Enzyme Engineering

Institut National des Sciences Appliqués, Toulouse

September 2018

Para os meus pais

Abstract

Metabolic engineering, defined as the rational engineering of organisms towards production goals, has greatly evolved since its conception over three decades ago. Once applied to overproduce cell endogenous metabolites, it is now a promising approach also for the biosynthesis of non-natural compounds through the expression of synthetic metabolic pathways. Improved over billions of years by evolution, enzymes are however less adapted to new catalytic functions as required by synthetic metabolism. The present work was aimed at the construction of artificial routes for the biosynthesis of commodity chemicals through the application of concepts of enzyme engineering.

(L)-2,4-dihydroxybutyrate (DHB) is a non-natural compound of industrial interest for the synthesis of methionine analogues, and whose biological production was previously demonstrated by the expression of a two-step pathway via homoserine in *Escherichia coli*. The pathway sequentially employs homoserine (HMS) transaminase and 2-keto-4-hydroxybutyrate (OHB) reductase activities. In this thesis, rational enzyme design was used as a strategy to improve the last catalytic step of the pathway. Simultaneous expression of the evolved OHB reductase Ec-Mdh-5Q and HMS transaminase Ec-AlaC A142P:Y275D variant in an engineered homoserine-overproducing *E. coli* strain resulted in the production of 89.0 mM DHB from glucose, the highest titer reported to date.

Of industrial relevance is also the synthesis of 1,3-propanediol (PDO), a metabolite generated from glycerol catabolism in various *Clostridia* species. Expanding the susbtrate range to sugars would render PDO production more flexible. Therefore, a six-step synthetic pathway yielding PDO from glucose via malate was conceived. While the three first reaction steps were previously demonstrated, the remaining DHB dehydrogenase, OHB decarboxylase and PDO oxidoreductase activities were identified in candidate enzymes acting on sterically cognate substrates. Improved enzyme activities were obtained based on sequence- and structure-based protein design. The feasibility of the PDO pathway was validated though expression of all required enyme activities in a single *E. coli* strain, while further improvements were achieved through co-cultivation of two *E. coli* strains each one expressing a partial segment of the completepathway (up to 3.8 mM PDO).

During the design and construction of the PDO pathway, OHB decarboxylases which release 3-hydroxypropanal as product of OHB decarboxylation, were found to be catalytically low efficient. To this end, a transcription factor-based metabolite sensor towards high-throughput detection of aldehydes in *E. coli* was developed. Optimization tasks of the metabolite sensor through 5'-UTR engineering rendered the sensor more sensitive to target compounds. In a proof of concept, simultaneous expression of the metabolite sensor and a segment of the PDO pathway in *E. coli* afforded the discrimination of two OHB decarboxylases with distinct kinetic properties. Therefore, the metabolite sensor can be implemented in directed evolution campaigns aiming at OHB decarboxylase development.

Keywords: metabolic engineering, synthetic pathways, enzyme engineering, 2,4-dihydroxybutyrate, 1,3-propanediol

Résumé

L’ingénierie métabolique utilise des techniques de clonage pour moduler directement les voies métaboliques de microorganismes dans le but de produire des molécules d’intérêts. Précédemment envisagée pour surproduire des métabolites endogènes, l’ingénierie métabolique est aussi considérée maintenant comme une approche prometteuse pour la biosynthèse de composés non naturels par l'expression de voies métaboliques synthétiques. Cependant, les enzymes sont moins adaptées aux nouvelles fonctions catalytiques requises par le métabolisme synthétique. Le but de cette thèse est donc la construction de voies artificielles pour la biosynthèse de molécules d’intérêts en appliquant des concepts d'ingénierie enzymatique.

Le (L)-2,4-dihydroxybutyrate (DHB) est un composé non naturel qui présente un intérêt industriel pour la synthèse d'analogues de la méthionine. Des travaux précédemment réalisés dans notre équipe ont démontré la biosynthèse de DHB par l'expression d'une voie à deux étapes via l'homosérine chez *Escherichia coli*. Cette voie emploie séquentiellement l’action d’enzymes ayant une activité homosérine (HMS) transaminase et 2-céto-4-hydroxybutyrate (OHB) réductase. Dans ce travail, l’optimisation de la dernière étape de la voie catalytique a été envisagé grâce à l’ingénierie rationnelle des enzymes. Ainsi une OHB réductase évoluée, Ec-Mdh-5Q, exprimée simultanémment avec la HMS transaminase Ec-AlaC A142P:Y275D dans *E. coli* a permis la production de 89.0 mM DHB à partir de glucose, le titre le plus élevé signalé à ce jour.

La synthèse du 1,3-propanediol (PDO), un métabolite généré par le catabolisme du glycérol chez diverses espèces de *Clostridia*, apporte un intérêt industriel. L’élargissement de la gamme de matières premières aux sucres rendrait la production des PDO plus flexible. Par conséquent, une voie de synthèse en six étapes produisant du PDO à partir du glucose via le malate a été conçue. Si les trois premières étapes de cette réaction ont été ingénierées dans des travaux précédents, les trois activités enzymatiques suivantes (DHB déshydrogénase, OHB décarboxylase et PDO-oxydoréductase) ont due être identifiées dans des enzymes candidates agissant sur des substrats similiaires. Des activités enzymatiques améliorées ont été obtenues par mutagénèse dirigée en se basant sur l’analyse de la séquence et de la structure de ces protéines. La faisabilité de la voie PDO a ensuite été validée par l'expression de ces six activités enzymatiques dans *E. coli* et des améliorations supplémentaires ont été obtenues grâce à la co-culture de deux

souches *E. coli* exprimant chacune un segment partiel de la voie complète (jusqu'à 3,8 mM PDO).

Au cours de la construction de la voie PDO, les OHB décarboxylases qui en décarboxylant l'OHB libérent l'aldéhyde 3-hydroxypropanal se sont révélées être faiblement efficaces sur le plan catalytique. Ainsi afin de cribler à haut débit des banques d'enzymes mutées enzymes, un biosenseur a été mis au point afin de détecter les aldéhydes libérés par cette bioconversion. Son optimisation via l'ingénierie de 5'-UTR a par la suite permis de le rendre plus sensible aux composés ciblés. Dans une preuve de concept, l'expression simultanée du biosenseur et d'une partie de la voie PDO dans *E. coli* a permis de distinguer les capacités catalytiques de deux OHB décarboxylases. Par conséquent, ce biosenseur pourra être mis en œuvre dans de futures campagnes d'évolution dirigée visant à optimiser l'OHB décarboxylase.

Mots clés: ingénierie métabolique, voies synthétiques, ingénierie enzymatique, 2,4-dihydroxybutyrate, 1,3-propanediol

Acknowledgements

Doing a PhD was an intense, yet enjoyable and marking period of my life. This thesis was only possible by the contributions of the many people that helped me during these almost three years, and to whom I would like to express my sincere gratitude:

My supervisor, Jean Marie François, for giving me the opportunity to work in the exciting field of synthetic biology. Thank you for the guidance, support, enthusiasm, and for the freedom given to also pursue my own ideas.

My co-supervisor, Thomas Walther, for always taking the time to discuss the many questions and challenges that synthetic biology frequently poses. Also thank you for welcoming me during my short stay in your research group at TU Dresden.

Christopher Topham, for all the help with enzyme computational design, and your valuable and always interesting comments on Chapters 2 and 3.

Victor, for your help and dedication put in the development of the metabolite sensor.

To the Synthacs and EcometBio teams at Toulouse White Biotechnology. Special thanks to Yoann Malbert for all the help with bioreactor experiments. Also to Yannick Malbert, who helped in giving my the first steps in the construction of metabolite sensors. Thanks for all the other members for the help whenever needed.

All partners involved in the Synpathic ANR project, thank you for the fruitful discussions. Special thanks to François Képès and Andrew Griffiths for taking part in the thesis committee.

To the various people I met during these years at the EAD5 team and who added a lot to my professional and personal experience. Thanks Ceren, for helping me getting started with bacterial metabolic engineering and the intensive protein purification tasks; all our scientific discussions were of unvaluable help. Cléa, for the always valuable input and

advices and for the company during the countless hours spent in the lab. Jian, Pablo: thanks for always having a good mood, ideas and the many adventures.

To other present and former members of the student's office EAD5 team: Lucie, Sevan, Matthias, Clara, Marine, Cathy, Luce, Agustina, Rebeca, Pierre-Luc, Pablulu, Adilia, Mélanie and so many others. Also the permanent staff: Jean-Luc Parrou, Jean Pascal, Didier Zerbib, Marion Schiavone and Hélène Martin-Yken.

To the Chair of Bioprocessing Technology at TU Dresden, where I spent only one but memorable month. Special thanks to Cristoph, Tim and Saskia.

To all the incredible people that crossed my path in Toulouse, you have made this a very enjoyable experience!

Aos meus pais, família e aos de sempre, pela força, motivação e apoio incondicional!

List of original publications

This thesis is based on the following scientific papers:

I. Frazão, C.J.R., Topham, C., Malbert, Y., François J.M., Walther, T. (in press). Rational engineering of a malate dehydrogenase for microbial production of 2,4-dihydroxybutyric acid via homoserine pathway. Biochem. J.

II. Frazão, C.J.R., Maton, V., François J.M., Walther, T. (2018). Development of a metabolite sensor for high-throughput detection of aldehydes in *Escherichia coli*. *Front. Bioeng. Biotechnol.* 6.

III. Frazão, C.J.R.*, Trichez, D.*, Spina, L., Lozano-Huguet, L., Serrano-Bataille, H., Dagkesamanskaya, A., Topham, C., François, J.M., Walther, T. Construction of a non-natural pathway for production of 1,3-propanediol from glucose. Manuscript in preparation.

* Equal contributions

List of abbreviations

^{13}C-NMR	^{13}C nuclear magnetic resonance
3-HPA	3-Hydroxypropanal
BCAA	Branched chain amino acids
CPD	Computational protein design
DHAP	Dihydroxyacetone phosphate
DHB	2,4-Dihydroxybutyrate
FACS	Fluorescence-activated cell sorting
FBA	Flux balance analysis
GC-MS	Gas chromatography – mass spectrometry
HPLC	High pressure liquid chromatography
k_{cat}	Turnover number
k_{cat}/K_m	Catalytic efficiency
K_m	Michaelis constant
MFA	Metabolic flux analysis
OHB	2-Keto-4-hydroxybutyrate
PDO	1,3-Propanediol
QMMM	Quantum mechanics/molecular mechanics
RBS	Ribosome binding site
V_{max}	Maximum velocity
X1P	Xylulose-1-phosphate
$\Delta_f G^0$	Standard Gibbs free energy of formation
$\Delta_r G^0$	Standard Gibbs free energy of a reaction

Table of contents

"Believe you can and you're halfway there."

—Theodore Roosevelt

Chapter 1. Introduction

1.1 Microbial cell factories

Developments in petrochemistry have greatly enhanced industrial and societal progress. The continuous increasing rate of oil reserve exploitation has served as an abundant and inexpensive raw material for the synthesis of transportation fuels and production of commodity chemicals. But triggered by a rising energy demand, geopolitical uncertainties, environmental effects and oil-depletion concerns, industries have in recent times been demonstrating interest in the utilization of alternative feedstocks, including plant biomass and inexpensive sugars[1].

Originally seen as a source of food and beverage products, the past few decades have disclosed the potential of microorganisms as cell factories for the production of industrially relevant molecules such as biofuels[2,3] and commodity chemicals[4,5]. Since the advent of metabolic engineering, a few number of bio-synthesized molecules (e.g. ethanol, succinate, lactate) successfully achieved commercialization thereby contributing to a shift towards a more sustainable economy[6,7]. However, the development of microbial-based processes at prices competitive to those derived from petroleum are in part responsible for the still limited success rates of process industrialization.

Cellular metabolisms are diverse and complex, composed of metabolic pathways from which thousands of unique chemicals are produced[6]. While metabolic engineering traditionally aims at channeling carbon flux towards overproduction of a target metabolite, recent advances in computational biology, molecular genetics and protein engineering are moving this young and exciting field to a new era[8]. By enabling the creation of non-natural metabolic pathways, the production of non-natural compounds can now be foreseen while the biosynthesis of natural metabolites can be achieved by multiple and diversified routes. Improved by evolution to perform metabolic functions, enzymes are however less adapted towards novel functions as required by synthetic metabolism.

1.2 Objective and outline of this thesis

The construction of synthetic metabolic pathways requires the discovery of new enzyme functions. The main objective of this doctoral thesis was to engineer enzymes with new catalytic activities to be implemented in non-natural pathways aiming at the production of commodity chemicals, in particular 1,3-propanediol (PDO) and 2,4-dihydroxybutyrate (DHB).

In Chapter 2, the field of metabolic engineering is reviewed. Starting by pathway design, the creation and implementation of metabolic pathways is approached, and optimization tasks reviewed. Particular attention is given to the construction of non-natural pathways and the domain of enzyme engineering, which permits (re)creating enzymes with new functions.

In Chapter 3, rational enzyme engineering was employed as a strategy to improve the efficiency of the last reaction step towards DHB production from glucose via homoserine. The kinetic properties of created enzyme variants were determined, and those displaying the highest catalytic efficiencies were further evaluated upon implementation in the synthetic pathway.

In Chapter 4, a synthetic pathway was designed and established in *E. coli* for the direct production of 1,3-propanediol from glucose via 2,4-dihydroxybutyrate. While those enzyme activities linking glucose to DHB were previously reported, the downstream pathway converting DHB into PDO employs two unknown enzyme functions. In this study, the discovery and engineering of those enzymes was reported and PDO production demonstrated.

In Chapter 5, an *E. coli* metabolite sensor was designed and optimized towards intracellular aldehyde detection. Coupled to a powerful screening/sorting system, it can be used in directed evolution campaigns for the detection and isolation of improved aldehyde-producing enzyme variants.

In Chapter 6, the main conclusions of this thesis are presented and the significance of current and future work discussed.

1.3 References

(1) Zhou, H. (2011) Metabolic engineering of yeast for xylose uptake and fermentation. Doctoral dissertation.

(2) Hahn-Hägerdal, B., Galbe, M., Gorwa-Grauslund, M. F., Lidén, G., and Zacchi, G. (2006) Bio-ethanol – the fuel of tomorrow from the residues of today. *Trends Biotechnol. 24*, 549–556.

(3) Kung, Y., Runguphan, W., and Keasling, J. D. (2012) From Fields to Fuels: Recent Advances in the Microbial Production of Biofuels. *ACS Synth. Biol. 1*, 498–513.

(4) Adkins, J., Pugh, S., McKenna, R., and Nielsen, D. R. (2012) Engineering microbial chemical factories to produce renewable biomonomers. *Front. Microbiol. 3*, 313.

(5) Schmidt-Dannert, C. (2017) The future of biologically inspired next-generation factories for chemicals. *Microb. Biotechnol. 10*, 1164–1166.

(6) Trinh, C. T., and Mendoza, B. (2016) Modular cell design for rapid, efficient strain engineering toward industrialization of biology. *Curr. Opin. Chem. Eng. 14*, 18–25.

(7) Choi, S., Song, C. W., Shin, J. H., and Lee, S. Y. (2015) Biorefineries for the production of top building block chemicals and their derivatives. *Metab. Eng. 28*, 223–239.

(8) Erb, T. J., Jones, P. R., and Bar-Even, A. (2017) Synthetic metabolism: metabolic engineering meets enzyme design. *Curr. Opin. Chem. Biol. 37*, 56–62.

"Science is simply the word we use to describe a method of organizing our curiosity."

—Tim Minchin

Chapter 2. Metabolic engineering meets enzyme design

The concept of engineering metabolic pathways has been widely used in the past. For instance, 500-fold improved productivities of the antibiotic penicillin were found after submitting the fungi *Penicillium chrysogenum* to repeated rounds of random mutation and selection[1]. With the introduction of recombinant DNA technology by Cohen and Boyer[2] started the development of techniques focused on rationally engineering microorganisms towards production goals, a field known today as metabolic engineering. First proposed by Bailey[3] over three decades ago, metabolic engineering aims ultimately at the development of cost-efficient processes for the microbial production of fuels, chemicals and pharmaceuticals. Three are the main parameters used to evaluate the success of metabolic engineering projects: titer, yield and productivity[4].

The field of metabolic engineering assumes an industrial dimension. As such, it comprises not only the design and implementation of a functional pathway in a suitable host, but also its optimization so that a commercial process can be envisaged. Metabolic engineering can therefore be considered as an iterative design-build-test cycle. These elements encompass genetic engineering and molecular biology, but also components from graph theory, chemical reaction engineering, biochemistry, and optimization[4,5].

In this thesis, *E. coli* has been chosen as the host for metabolic engineering, reason for which only this organism is covered in this chapter. Considered as a prokaryotic model organism for which gene-editing tools and genome sequence information are readily available, *E. coli* constitutes the primary choice as host for metabolic engineering studies. In addition, the well-understood physiology, metabolism and genetics further aided in its development as a production organism. However, several other organisms have been explored for the successful biosynthesis of high-value metabolites[4]. While not covered in this thesis, the choice of an appropriate host is determinant for the success of metabolic engineering projects and should be carefully executed in each case scenario.

2.1 Pathway design

When aiming at the overproduction of a target compound, both metabolic engineers and synthetic organic chemists pose themselves a common question: what pathways can be used to produce the compound of interest? The concept of pathway design is well-established in organic chemistry, and aims at identifying the most favorable set of reactions towards the synthesis of a given compound. Metabolic engineering introduces three additional constraints in pathway design:

- Set of reactions must be in overall thermodynamically favorable under biological conditions;
- Intermediate molecules should not be toxic;
- Enzymes required for conversion must be expressible in the host organism.

Nature has evolved cell metabolism as a large and complex metabolic reaction network, in which through an extensive number of intermediates and enzymatic control mechanisms, the functional adaptability and stability required by life is maintained. The interconnectedness of metabolic reactions thus results in a very high number of potential pathways linking substrate to product. This means that pathways must be annotated, enumerated and assessed against other criteria[6–8]. The first complete method for pathway enumeration from a database of biochemical reactions was published in 1990 by Mavrovouniotis and colleagues[9] and was used to identify several routes yielding lysine. In particular, the method allowed the authors to conclude that oxaloacetate is a necessary precursor for the biosynthesis of lysine. Advances in genome sequencing further permitted the construction and refinement of genome-scale metabolic models by linking metabolic and genomic data. Two major extensive collections of metabolic networks for a large number of organisms which describe interactions between enzymes and substrates are provided by KEGG Pathway and MetaCyc, whose inspection represents today the first step of any metabolic engineering project[10]. Intuition and manual design have been the preferred strategy for postulating both natural and non-natural pathways enabling the synthesis of a desired metabolite. But during the last decade a range of computational tools to assist in the prediction and prioritization of new metabolic pathways yielding a target compound have been developed.

Early pathway finding algorithms (e.g. FMM, PathComp, Pathway Hunter Tool, MetaRoute) were developed based on known enzymatic reactions. These tools, also known as network-based path finding methods, all extract information from biochemical databases (e.g. KEGG, MetaCyc, BRENDA, MetRxn, Rhea) to identify those pathways that in a minimum set of reaction steps can link an input molecule to a target compound. On the other side, pathway prediction tools (e.g. BNICE, RetroPath, Route Designer) allow the user to design pathways that encompass novel reactions not found in metabolic databases. Those are mostly based on the concept of retrosynthesis, that by starting from a target product iteratively applies reversed biotransformations to reach a precursor present in natural metabolism[11–14]. Such strategy may for instance be applied in metabolic engineering projects which aim at overproducing non-natural compounds.

While pathway design tools often identify multiple routes, these are now associated with pathway prioritization which allows to rank predicted routes. The most common method to rank pathways is based on the number of reactions steps (e.g. FindPath), in which the shortest pathway is considered as more favorable since it implies minimal enzyme requirement and avoids metabolic burden. Thermodynamic feasibility, intermediate toxicity, number of known reactions and kinetic proficiency are lately combined with pathway length for a more reliable prioritization of predicted pathways (e.g. BNICE, Retropath, SimPheny)[11].

The selection of the most appropriate computational tool for pathway design very much depends on user specifications, but one should keep in mind that the combination of prediction with ranking may yield more satisfactory results. For example, aiming at demonstrating the production of the non-natural compound 1,4-butanediol in *E. coli*, Yim and co-workers[15] used the SymPheny Biopathway Predictor software to elucidate and evaluate all potential pathways linking host central metabolites to the target compound. Based on the transformation of chemical groups by known chemistry, the algorithm identified more than 10,000 possible pathways. Combined with the utilization of an in-house software for pathway sorting and ranking, the authors identified the 1,4-butanediol pathway through 4-hydroxybutyrate as the highest priority for construction and implementation in an *E. coli* host strain. Rounds of optimization further resulted in a process which today paves its way to commercialization. In another example, Fehér and colleagues[16] experimentally validated the RetroPath software in a study aiming at the design and implementation of a pathway yielding the flavonoid pinocembrin in an *E. coli*

strain. Starting with the desired compound as an input, the software searches and lists all possible pathways that link an endogenous metabolite in the host organism to the target molecule. Each of the listed pathways corresponded to a unique list of known enzymatic reactions, for which a set of criteria was pre-established towards pathway ranking. Of the 11 found pathways, the algorithm identified as expected the natural flavonoid synthesis route as the most favorable one, and proposed a series of genes encoding for desired enzyme activities. After assembling and expressing the metabolic pathway in *E. coli*, pinocembrin was produced up to 24.1 mg L^{-1}, a value in line with other studies targeting flavanone biosynthesis.

2.2 Pathway construction

Pathway engineering involves ultimately at finding the right combination of enzymes that permits linking in an efficient manner a given molecule (e.g. substrate, endogenous metabolites) to a target compound. For clarity, in this thesis enzyme development is considered as a stage in pathway construction rather than a pathway optimization task. According to the synthetic character and biochemical solution space, Erb and colleagues[17] have previously classified engineered metabolic pathways into five levels, which are described in the next subsections.

2.2.1 Endogenous pathways

Early stages of metabolic engineering focused on optimizing the endogenous metabolism, i.e. existing pathways within their natural hosts, towards production goals (**Figure 2.1a**). Based on genome sequencing and annotated pathways, individual gene deletion and/or over-expression can be predicted aiming at pathway optimization. For example, Zhang and co-workers[18] successfully redirected carbon flux towards overproduction of the Krebs cycle intermediate (L)-malate in *E. coli*. Starting with a previously engineered succinate-overproducing strain (*E. coli* Δ*ldhA* Δ*ackA* Δ*adhE* Δ*pflB*), step-wise gene deletion of carbon diverting and malate degradation encoded-enzymes afforded the creation of a strain with 11 gene deletions which was able to significantly accumulate the desired metabolite under anaerobic conditions. Optimization of cell cultivation conditions further enhanced malate titers and productivities (up to 34 g L^{-1} and 0.47 g L^{-1} h^{-1}).

2.2.2 Copy, paste and fine-tuning pathways

In a more advanced strategy named "copy, paste and fine-tuning", annotated pathways or sub-pathways are introduced to a heterologous host (**Figure 2.1b**). Expressing the Weimberg pathway from *Caulobacter crescentus* in *E. coli* allows for an alternative route for xylose assimilation[19] and illustrates the concept of copy-paste. While the engineered pathways keep unaltered their basic structure, minor changes can be a result of improved kinetics and/or thermodynamics. The replacement of individual enzymes is included in the latter case. For instance, in a pathway dedicated to the production of *n*-butanol

expressed in *E. coli*, substituting the thermodynamically limiting acetoacetyl-Coa synthetase with an irreversible acetyl-CoA:malonyl-CoA acetyltransferase (decarboxylating) substantially improved product titers, resulting in a fine-tuning of the production system[20].

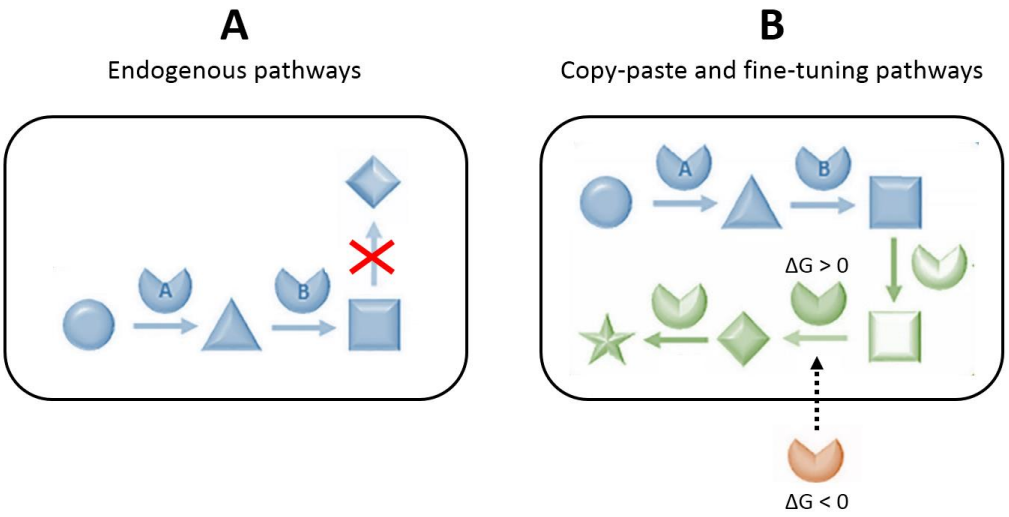

Figure 2.1. The process of engineering natural pathways can be divided in two strategies. **(A)** Engineering endogenous pathways (highlighted in blue) involves gene deletion and/or overexpression aiming at carbon flux channeling towards target metabolite (square). **(B)** The concept of copy-paste is based on the transfer of entire pathways or genes (green) to a heterologous host, while fine-tuning can be achieved by replacing rate-limiting enzymes (e.g. thermodynamically unfavorable, in which $\Delta G > 0$) by more favorable ones (e.g. thermodynamically favorable, in which $\Delta G < 0$) (orange). Adapted from Hossain et al[21].

While engineering existing pathways or "copy, paste and fine-tuning" pathways is confined to the manipulation of natural metabolism, the creation of non-natural or synthetic pathways came to revolutionize the field of metabolic engineering. In specific, non-natural routes as described next enlarge the metabolic solution space by enabling the creation of both natural and non-natural compounds. Scientists are therefore able to implement the so-called concept of "total synthesis" also in the domain of industrial biotechnology[17].

2.2.3 Mix and match synthetic pathways

Whilst the utilization of natural enzymes may enable the creation of synthetic pathways, the creation of new enzymes and engineering of natural ones may further expand the metabolic solution space. The simplest form of synthetic metabolism is built upon a "mix

and match" concept, in which non-natural pathways are created through combinatorial expression of natural enzymes from various sources (**Figure 2.2**). Provided that the host strain produces the necessary precursors from its own primary or secondary metabolism, subsequent conversion to the desired target metabolite can be accomplished through heterologous gene expression. Without the need to evolve novel enzymatic reactions, a wide range of previously unreported pathways can be identified.

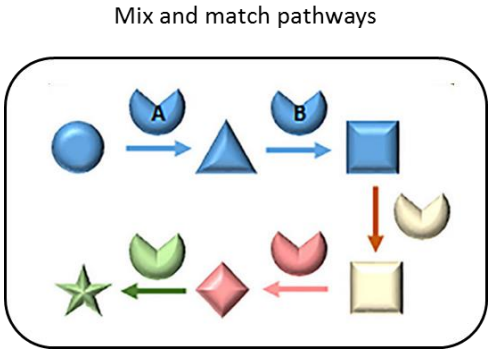

Figure 2.2. The process of engineering non-natural pathways based on a "mix and match" approach. Based on endogenous metabolism (blue) which results in the synthesis of a certain precursor (blue square), expression of natural enzymes from various sources (each color, each source) can link the referred precursor to a desired metabolite (star). From Hossain et al[21].

The biosynthesis of arbutin, a glycoside used as skin-lightering agent, from glucose illustrates well this concept. By expressing 4-hydroxybenzoate 1-hydroxylase from *Candida parapsilopsis* (MNX1) and arbutin synthase from *Rauvolfia serpentina* (AS) enzyme activities in an *E. coli* host strain, Shen and co-workers[22] were able to connect the endogenous precursor *p*-hydroxybenzoic acid to the desired metabolite. Other examples include production of pinocembrin and polyketide analogs, as summarized by Hossain and colleagues[21].

2.2.4 Synthetic pathways with novel reactions

Pathway design frequently results in the identification of routes involving reaction steps that natural enzymes are unable (or unknown) to catalyze. Those catalytic steps may involve new enzymatic functions based on new reactions (**Figure 2.3a**). Designing metabolic pathways involving new reactions relies on the fact that biocatalysts can

catalyze the conversion of new substrates whilst keeping unaltered the respective enzymatic mechanisms. Three different strategies have been attempted to accomplish this goal (**Figure 2.3b-d**).

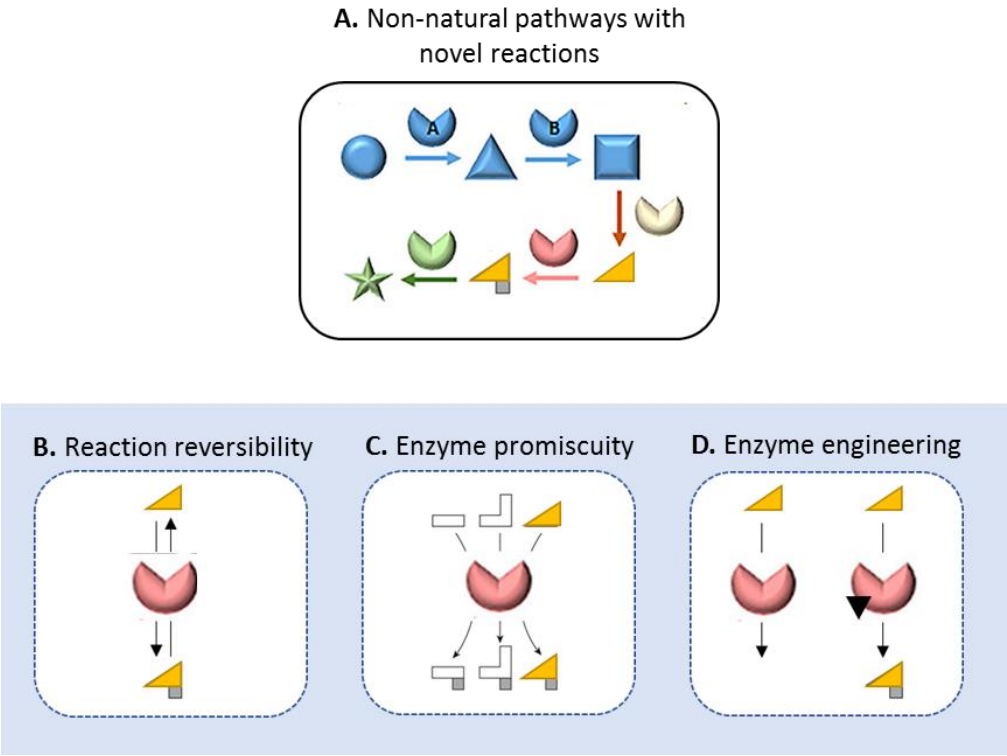

Figure 2.3. The process of engineering non-natural pathways based on novel reactions **(A)**. Based on endogenous metabolism (blue) which results in the synthesis of a certain precursor (blue square), expression of natural enzymes from various sources (each color, each source) can link the referred precursor to a desired metabolite (star). Novel reactions can be accomplished based on reaction reversibility **(B)**, enzyme promiscuity **(C)** and/or enzyme engineering **(D)**. Legend: mutation (black inverted triangle).

The first approach is built upon the concept of reversibility of (bio)chemical reactions (**Figure 2.3b**). While some enzymatic reactions have been experimentally shown to be reversible in a cellular context depending on reactant(s)/product(s) ratio (e.g. branched-chain amino acid transaminase), others are clearly irreversible (e.g. glycolytic enzymes hexokinase, phosphofructokinase). In some other cases, *in vivo* reversibility is still matter of debate. The case of the anaerobic pyruvate formate-lyase clearly shows how reversibility can enlarge the metabolic and enzyme solution spaces. Despite possessing a strong preference (standard Gibbs free energy, $\Delta_r G^0$ = -21 kJ mol^{-1}) for the CoA-dependent pyruvate cleavage yielding formate and acetyl-CoA, Zelcbuch and co-

workers[23] demonstrated this same enzyme to enable growth on acetate and formate in cells devoid of the glyoxylate shunt by catalyzing the reverse reaction *in vivo*.

Another alternative to find new reactions consists in exploring the substrate repertoire of natural enzymes (**Figure 2.3c**). Notably, around 10% of bacterial and archaeal enzymes have been reported as promiscuous[24]. Identification of those enzymes accepting substrates structurally analogous to the prospective substrate may therefore constitute a feasible strategy for constructing a non-natural pathway. For instance, in this doctoral thesis the NADPH-dependent broad-range aldehyde reductase from *E. coli* (Ec-YqhD) was employed as part of an alternative synthetic PDO pathway starting from glucose due to its ability to reduce various aldehydes into corresponding alcohols (see **Chapter 4**). While in this specific case 3-hydroxypropanal (3-HPA) was the target substrate, others took advantage of the relaxed substrate specifity of Ec-Yqhd to engineer other non-natural pathways for synthesis of e.g. 1,2,4-butanetriol[25] and ethylene glycol[26] (in which YqhD substrates were D-3,4-dihydroxybutanal and glycolaldehyde, respectively). Another example of a non-natural route based on substrate promiscuity is the xylulose-1-phosphate pathway (addressed in **Chapter 5**), seen as an alternative route for xylose assimilation, that generates stoichiometrically equivalent amounts of glycolaldehyde and dihydroxyacetone phosphate (DHAP)[27].

Despite providing a convenient alternative towards the creation of synthetic pathways, the utilization of natural enzymes catalyzing reverse reactions and/or multiple substrates frequently results in poor product titers and yields. For example, Ec-YqhD has a catalytic efficiency (k_{cat} / K_m) for HPA equal to 0.13 s^{-1} M^{-1}, a value lower by 6-orders of magnitude when compared with an average enzyme (~10^5 s^{-1} M^{-1})[28,29]. Engineering enzymes displaying promiscuity and / or activity on sterically cognate substrates may in those cases be a strategy to further improve the efficiency of both, biocatalysts and pathways (**Figure 2.3d**). The exciting field of enzyme engineering and methods thereof will be reviewed later in this chapter. Two textbook examples of pathway construction based on engineered enzyme activities have been recently reported. In the first one, Walther and colleagues[30] successfully created a complete synthetic route that produces the non-natural compound DHB from the TCA cycle intermediate (L)-malate based on structure-guided rational protein design. In the second one, Schwander and co-workers[31] followed a similar approach to engineer a synthetic pathway composed of 17 enzyme activities for the *in vitro* fixation of CO_2 (CETCH cycle) at rates up to 5 mU mg $protein^{-1}$.

2.2.5 Synthetic pathways with novel chemistries

Empowered by the rapid growth of computational tools for engineering proteins, *de novo* enzyme design towards new biochemistries can now also be envisaged, i.e. new reactions with new mechanisms (e.g. using artificial co-factors). Out of the scope of this thesis, new enzyme biochemistries may enable the creation of fully artificial pathways that represent the maximum exponent of synthetic metabolism. The most remarkable example is perhaps the development of a formolase enzyme that allows the previously unreported carboligation of three formate molecules into a single molecule of DHAP, involved in central metabolism[32].

2.3 Pathway optimization

Once the possible pathways have been established in the host organism, the next step in engineering a production strain is to analyze the respective phenotype. Conventional chromatography techniques (e.g. HPLC, GC-MS) enable the user to quantify product formation, substrate consumption and accumulation of commonly released by-products (e.g. acetate, formate) at different stages of cell cultivation. This strategy enables a rapid phenotypic evaluation and allows the user to predict the most obvious genetic modifications that need to be introduced in the product strain aiming at improving product formation. A more comprehensive and systemic overview of cell metabolism as provided by omics data analysis is however necessary at later stages of metabolic engineering projects when all obvious genetic modifications have been attempted. This strategy may provide for a wide knowledge on network topography, kinetics and regulation of a metabolic pathway, identification of kinetic bottlenecks and competing pathways, and an assessment of dysfunction in pathway operation. Alternatively or in combination with rational-based approaches, adaptive laboratory evolution is not uncommon in the field on metabolic engineering, while pathway co-localization has recently emerged as an alternative but effective strategy for metabolism optimization[5].

2.3.1 Metabolic flux analysis (MFA)

MFA aims at determining intracellular reaction rates (i.e. *in vivo* fluxes) which can be compared between different conditions (e.g. two engineered strains). More important than the flux values *per se* are the deviations from control conditions which can provide insights into metabolic and regulatory responses, e.g. identification of key branch point flux distributions, detect carbon loss through competing pathways, evaluate engineered strains. While MFA has mostly been used to quantify fluxes in central metabolism, efforts have been made towards its application on a genome-scale. Typically, a ^{13}C-labeled substrate is fed with cells and used to label downstream metabolites whose isotopic distributions are measured by GC-MS or ^{13}C-NMR techniques. Intracellular fluxes are then determined by using a stoichiometric model for the major intracellular reactions and applying mass balances around intracellular metabolites. A set of measured extracellular fluxes, normally substrate uptake rates and product secretion rates, is used as input to computational analysis platforms (e.g. 13CFLUX2, OpenFlux, FiatFlux). The output of

MFA is a metabolic flux map showing a diagram of biochemical reactions included in calculations along with estimated reaction flux[4,5,33]. For example, MFA analysis of an engineered *E. coli* strain for the biosynthesis of 1,4-butanediol aided in the identification of the two last enzyme-reaction steps as rate-limiting, as their intracellular flux values were considerably lower than those of the upstream reaction steps[34].

2.3.2 Metabolomics

Metabolomics analysis can point for metabolic bottlenecks, redox co-factor ratios and energy charge by looking at the instantaneous accumulation or depletion of metabolites, or by monitoring dynamic labeling of metabolites by using a labeled tracer. The combination of chromatography separation techniques with mass spectrometry detection affords the identification and quantification (relative or absolute) of a wide range of compounds. Comparison between different conditions associated with statistical analysis methods (e.g. PCA) may afford discrimination between samples[34].

2.3.3 Transcriptomics and proteomics

Transcriptomics (e.g. RNAseq, microarray, qPCR) and proteomics (e.g. iTRAQ, MRM) elucidate the user on the expression levels of endogenous, heterologous and manipulated genes or proteins. They may enable the user to assess the physiology and overall cell health, based on the expression of regulons associated with global stress responses[5,34].

2.3.4 Genomics

As costs associated with next-generation sequencing technologies keep decreasing, genomics emerges as a rapid method which permits to analyze engineered strains at a DNA scale during the different stages of a metabolic engineering project. In specific, genome and/or plasmid sequences allow to verify that the desired genetic modifications were correctly introduced. In addition, DNA sequence analysis enables the detection of unwanted introduction of mutations during cloning processes[34].

2.3.5 Computational modelling

A range of computational tools have been developed in an attempt to reduce the experimental workload required in pathway optimization tasks.

Flux balance analysis

A common problem associated with MFA is the cost and expertise that experimental procedures require. In this line, purely computational stoichiometric-methods to estimate fluxes have been developed, such as flux balance analysis (FBA). Both FBA and MFA make use of genome-scale metabolic models and mass balances. However, FBA is an undetermined system in which equations are formulated as a constrained optimization problem with a specific objective function to be maximized or minimized. By taking advantage of genome-scale models in simulating cell metabolism behavior, cell designs for maximizing chemical production can in principle be predicted. A common objective function are the maximization of cell growth and the minimization of metabolic adjustment. There are three key advantages of this methodology over the experimental approach:

- Minimal experimental data are needed;
- Perturbations (e.g. gene knockouts) can be made by changing a 1 to a 0 *in silico*;
- Models can be continuously refined as new omics data become available.

However, the major disadvantage of FBA-derived fluxes is that they are not the real fluxes of the system but fluxes yielding maximum biomass. A number of FBA-based algorithms have been developed to identify groups of gene knockouts that are predicted to change the fermentation profile of a cell when growing at a maximum growth rate, allowing simultaneously for the maximization of metabolite synthesis and growth rate (e.g. COBRA 2.0, OptKnock). OptKnock derivatives may also enable the simultaneous application of multiple up- or down-regulations and gene deletion (e.g. OptForce)[4,5,10,33,35].

Protein expression levels

Cells do not always express proteins at levels sufficient to provide enzyme activities at velocities compatible with the well-functioning of expressed metabolic pathways. While transcriptomics, proteomics and/or enzymatic assays may provide with valuable information on protein expression in its various levels (mRNA transcript, translation, activity), computational methods may allow to fine-tune their expression in cells. In particular, the ribosomal binding site (RBS) of an mRNA transcript has been shown to affect translation initiation rates and presents therefore an alternative approach for controlling enzyme production at the RNA level. The RBS Calculator tool is a web-based application that correlates translation initiation rates with RBS sequences. A protein coding sequence, together with desired translation initiation rate, and host organism are required as inputs so that RBS calculator can suggest a DNA sequence[10,36]. For example, RBS modification was previously used as a strategy to improve the phenotype of production strains[37].

2.3.6 Adaptive laboratory evolution

Adaptive laboratory evolution, also known as evolutionary engineering, aims at the random introduction of mutations in the host genome at rapid rates through serial passaging that successively improves phenotype fitness by imposing a growth dependent-selective pressure. Alternatively, the process can be executed in well-controlled reactors under steady state (i.e. chemostat) which can last from several weeks to several months. The strategy is particularly useful in cases where sugar assimilation is rate limiting or the cells are sensitive to toxic metabolites. In the scope of metabolic engineering, this strategy is normally applied when all possible rational genetic modifications have been exhausted[38].

2.3.7 Pathway co-localization

The spatial proximity of enzymes has previously been reported to allow for metabolic flux fine-tuning[8]. For example, inspired by those enzymes exhibiting substrate channeling (e.g. tryptophan synthase), Dueber and colleagues[39] constructed synthetic protein

scaffolds towards the recruitment of target enzymes involved in the synthesis of melavonate in *E. coli*. By providing a modular control over metabolic flux, product titers were improved up to 77-fold when compared to spatially unorganized enzymes. In another study, Lewicka and co-workers[40] fused pyruvate decarboxylase and alcohol dehydrogenase enzymes from *Zymomonas mobilis*. Expression of the resulting chimeric protein in *E. coli* JM109 wild-type cells resulted in the production of more than 400 mM ethanol after 72 h of cultivation, which corresponded to an improvement of ~2-fold if compared to cells expressing the two non-fused proteins.

2.4 Enzymes as key biocatalysts

Most biochemical reactions cannot occur at rates compatible with life without the presence of catalysts[41]. While the first evidence of biological catalysis was found by the end of the 18th century, only in 1877 was the term enzyme employed by Kühne in reference to the ferments isolated from living organisms[42]. Enzymes are the product of billions of years of evolution, having evolved substrate and function specificity to improve the overall fitness of organisms in response to environmental needs[43]. All enzymes are by definition proteins, but other types of molecule such as RNA (i.e. ribozymes) may act to accelerate reactions, and are thus also catalysts.

2.4.1 Structure and catalytic function

Enzymes are mostly globular entities of variable size, with subunits ranging from approximately 60 to several hundreds of amino acids. They catalyze reactions where substrates are converted into products, which occur in a small region of the enzyme usually referred to as the active site. Often consisting of a partially or completely buried cavity or cleft constituting at most 10-20% of the total volume of an enzyme, the side chains of active site residues orient the substrate (binding site) for its subsequent conversion into a product (catalytic site)[44]. In **Figure 2.6**, the three-dimensional structure of the malate dehydrogenase enzyme from *E. coli* (Ec-Mdh) and corresponding active site are shown.

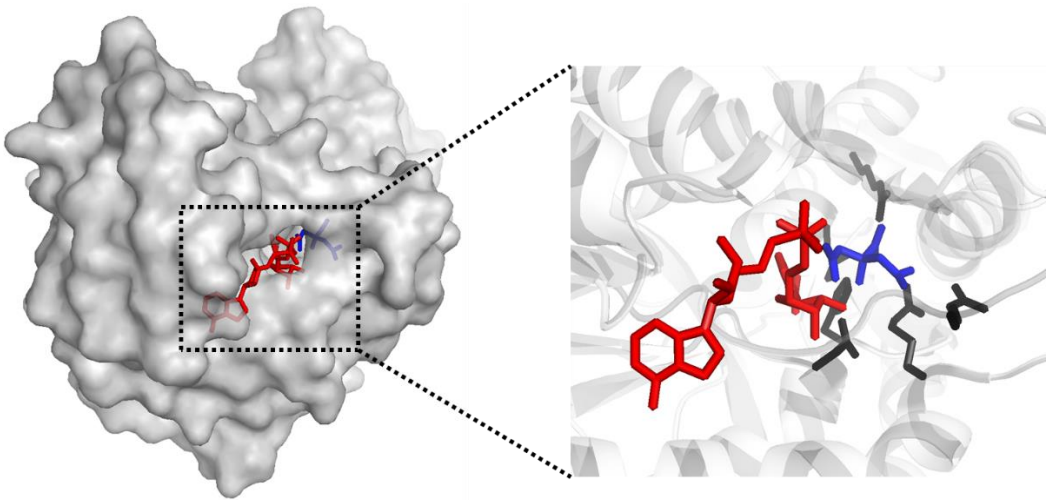

Figure 2.4. Three-dimensional X-ray crystal structure of Ec-Mdh (left) and its active site region (right) bound with NAD^+ (red sticks) and citrate (blue sticks) (PDB code 1emd). Side-chains of active site residues are shown in a dark stick representation.

Enzymes engage in both intra- and inter-molecular interactions (electrostatic, van der Waals interactions) which are important in catalytic functions and thermal stability. Substrate and/or product interactions should not be too tight to avoid enzyme inhibition, but exceptions do exist in which enzyme feedback inhibition has been reported as a mechanism of pathway regulation (e.g. phosphoenol pyruvate carboxylase)[45,46].

Enzymes are extraordinary catalysts able to accelerate the rates of biological reactions under mild conditions of temperature and pressure up to 10^{17}-fold[47]. Like any other catalyst, enzymes enhance chemical reaction rates by following two main principles. First, they are not consumed or permanently altered by the reaction and second, the chemical equilibrium between reactants and products remains unchanged in accordance with the laws of thermodynamics. By decreasing the activation energy required to form an unstable transition-state complex, enzymes are able to convert substrates into products at higher rates when compared with uncatalyzed reactions. Among the factors responsible for the enhanced catalytic activities of enzymes are:

- Approximation of reactants;
- Covalent catalysis;
- Acid-base catalysis;
- Conformational distortion;
- Pre-organization of the active site for transition state complementarity.

Most enzymes utilize several of the referred effects to ensure efficient catalysis[48].

2.4.2 Enzyme specificity

The substrate specificity of an enzyme depends on the arrangement of atoms in the enzyme active site pocket such that they complement the transition-state structure of the enzyme-bound substrate molecule, as opposed to its ground-state structure[48]. Enzyme specificity can be measured experimentally as (k_{cat}/K_m), corresponding to the difference in free energy of the enzyme-bound substrate transition-state relative to that of the substrate and enzyme free in solution. The (k_{cat}/K_m) parameter therefore provides a direct measure of the catalytic efficiency of the enzyme (relative to the uncatalyzed reaction in solution with a much higher transition-state activation energy barrier). This kinetic parameter is also used to evaluate the degree of enzyme stringency towards substrate

acceptance. For instance, enzymes involved in central metabolism (e.g. malate dehydrogenases) tend to be highly specific, while detoxification enzymes like Ec-YqhD are less specific since they act on a broad range of substrates. In the latter case, enzymes are called as promiscuous (or generalist). It is noteworthy to mention that promiscuity is a generic term, that can be based on: the ability of enzymes to utilize alternative co-factors (e.g. farnesyl diphosphate synthase), active site plasticity (e.g. β-lactamases) and ambiguous substrate (e.g. cytochrome P450)[49]. In the scope of this thesis, promiscuity is used for those enzymes with activity on a large spectrum of substrates.

According to the neutral drift theory of evolution, highly specific-modern-day enzymes are thought to have evolved from promiscuous but more thermally stable primitive ancestral enzymes through an iterative trajectory of random gene duplication and mutation events[49]. In a similar fashion, enzyme promiscuity can be exploited in protein engineering to create new functions by mimicking natural evolution in a laboratory environment.

2.4.3 Enzyme selectivity

Selectivity refers to the ability of enzymes to discriminate between closely related substrates. It can be quantified as the ratio of the (k_{cat} / K_m) values for the individual substrates concerned. The concept of selectivity can be categorized into three distinct classes: chemoselectivity, regioselectivity and stereoselectivity. Chemoselectivity refers to the ability of an enzyme to catalyze the transformation of a single functional group in the presence of others in a substrate molecule. On the other hand, a regioselective enzyme can distinguish between one or more identical functional groups located in different sites of the substrate molecules. As a result, only one group participates in the reaction yielding a selective product. Finally, stereoselectivity refers to the capability of an enzyme to distinguish between a pair of stereoisomers, resulting in an 100% optically pure product in the case of fully selective catalysis[50]. For example, (L)-malate dehydrogenases are highly selective as they are able to only catalyze the NAD-dependent oxidation of (L)-enantiomer of malate.

2.4.4 Enzyme kinetics

When aiming at creating metabolic pathways towards production goals, the evaluation of enzyme performance plays a vital role by allowing possible bottlenecks to be predicted. The field of enzyme kinetics aims at studying the reaction rates of enzyme-catalyzed reactions, which in many cases can be conveniently described by a simplified Michaelis-Menten model. According to the Michaelis-Menten model, enzyme activity increases hyperbolically with the increase of substrate concentration up to a maximum level at which enzyme saturation is reached. The Michaelis-Menten equation is given by

$$v = \frac{V_{max} \times [S]}{K_m + [S]}$$

where v is the initial enzyme velocity, V_{max} is the maximum velocity, [S] the substrate concentration and K_m the Michaelis constant.

The performance of a given enzyme is evaluated based on three fundamental kinetic parameters, of which only any two are independent: (1) the catalytic constant or turnover number (k_{cat}) which is defined as the maximum number of substrate molecules that a single enzyme copy converts to product per unit of time, and is a function of V_{max}

$$k_{cat} = \frac{V_{max}}{E_0}$$

where E_0 is the total enzyme concentration; (2) the Michaelis constant (K_m) is the substrate concentration at half-maximal velocity, which under conditions of quasi-equilibrium substrate binding approximates to the affinity of the enzyme for its substrate, but which in general is defined by a more or less complicated function of kinetic rate constants describing the enzyme mechanism, and (3) ratio of k_{cat} and K_m (k_{cat}/K_m) which measures the catalytic efficiency at which an enzyme converts a substrate into the corresponding product, and allows information concerning substrate specificity to be inferred[51]. Among the external factors influencing the performance of an enzyme are pH, temperature and ionic strenght. Enzymes are usually experimentally characterized *in vitro* following protein purification based on affinity chromatography techniques.

Although many enzymes follow the Michaelis-Menten kinetics, others show deviations. Many factors can cause such behavioral deviations, including product/substrate

inhibition, kinetic and binding cooperativity, or the action of an allosteric effector at physically remote binding site. In these cases, other more complex models of the enzyme kinetics have been proposed[48].

2.5 Creating new enzymes

Although a considerable number of enzymes has been characterized (104,000 enzyme names available in BRENDA database by 2016), it corresponds only to a small fraction of the earth's biodiversity[52–54]. The creation of new enzymes performing new reactions and/or chemistries remains therefore as a major goal in the successful implementation of non-natural metabolic pathways.

In the last decades, protein engineering emerged as a popular approach for enzyme discovery largely contributing to the creation of the concept of "made-to-order" enzymes. Protein engineering towards a novel function can be accomplished by two main strategies, (i) site-directed mutagenesis based on rational protein design and (ii) random mutagenesis coupled with directed evolution[55]. Both approaches have their advantages and challenges. Lately, the combination of the two strategies in an iterative approach has been attempted. In this section, enzyme engineering is presented with special emphasis on engineering template specialist enzymes acting on sterically cognate substrates or promiscuous generalist enzymes. The shift towards gain in new function most typically follows a weak negative trade-off in which engineered enzymes display activity on multiple substrates (**Figure 2.5**). The creation of enzymes truly specialized in a new function still remains a holy grail in the field of synthetic biology, and is limited to very few examples.

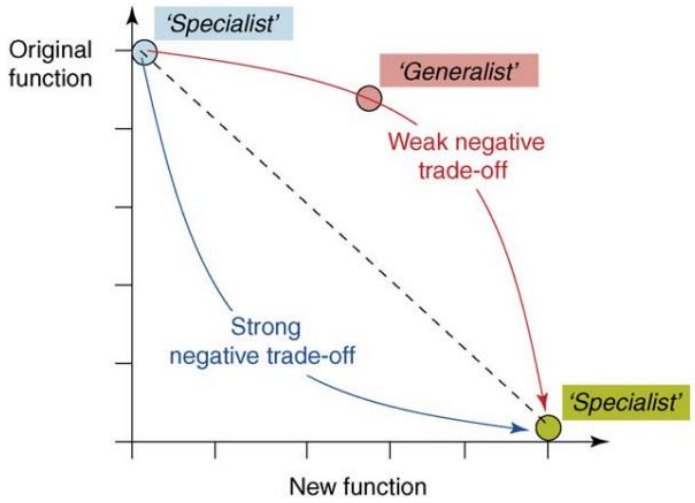

Figure 2.5. Possible routes for acquisition of new enzymatic function. The gain–loss of the new versus old function, and the conversion of one 'specialist' protein into another, may trade-off linearly (dashed line), or follow either concave or convex routes. From Khersonsky et al[56].

2.5.1 Rational design

In the process of rationally engineering enzymes, available information on protein sequence, structure and function may be exploited, and computational tools used to identify target mutation sites likely to improve a desired trait (e.g. new function, activity, affinity, inhibition tolerance, thermostability). By focusing on specific amino acid positions, a small number of variants is generated resulting in reduced screening efforts when compared to enzyme directed evolution-based approaches. Rationally engineered enzymes are usually assayed in low-throughput assays, e.g. by traditional chromatographic and NMR techniques, or microplate-based assays. However, although a rational design strategy is particularly powerful in the prediction of the molecular factors influencing catalytic activity, only a small number of examples of the successful engineering of stereospecificity or enantioselectivity have been reported in the literature[57,58].

Sequence-based design

The primary structure (i.e. amino acid sequence) of enzymes provides the most abundant source of information concerning natural diversity which may be exploited in rational design. In particular, the analysis of multiple sequence alignments (MSA) of homologous proteins is used in the identification of functional- and specificity-related residues, spatial constraints in 3D protein modeling and protein function prediction. While those residues determining catalytic activity show high degree of conservation, more variability can be found in other regions among homologous proteins. Several multiple sequence aligners are publicly available (e.g. ClustalW, MAFFT, Muscle)[59]. Position-dependent residue profiling and mutation correlation analysis can be performed using software such as EVfold, CCMpred or pySCA. The applicability of this strategy has been demonstrated e.g. in a design study aiming at converting Ec-Mdh into an enzyme with (L)-lactate dehydrogenase activity, in which target mutations were predicted on the basis of multiple sequence alignments between proteins belonging to the Ldh/Mdh superfamily of enzymes[60].

Structure-based design

The function of a given enzyme is intrinsically linked to its 3D structure, as well as its sequence. With the increasing number of solved protein crystal structures with bound substrate(s) and/or inhibitor(s) available in the PDB (Protein Data Bank), protein design is expected to be more accurate compared to sequence-based methods. However, experimentally solved structures are not available for all enzymes. Enzyme structure can often be predicted in these cases by homology or comparative modeling methods given the existence of a known structure of at least one homologous protein (sequence identity, > 30%). A number of structural bioinformatics tools and web-severs can be used for this purpose (e.g. I-Tasser, Modeller, SwissProt). Several studies report enzyme engineering strategies based on the multiple overlay of protein and protein-ligand crystal structures. Root mean square deviation (RMSD) measures allow the identification of locally variable protein chain conformations resulting from residue substitution, and the clustering of bound substrate configurations within the active sites of homologous enzyme families. Structure-based design approaches can be used in conjunction with complementary MSA analysis methods[59].

Computational structure-based enzyme design

Besides structure comparison, enzyme structures can be used as starting templates in computationally more intensive automated combinatorial exploration of sequence and protein conformational space. Computational protein design (CPD) software such as ROSETTA[61] can be used to perform side-chain rotamer conformational searching with either fixed, or more computationally demanding, flexible representations of the protein main-chain backbone. The introduction of protein backbone flexibility permitted a > 18-fold increase in Diels-Alderase activity[62] in an artificial enzyme previously designed using a fixed backbone[63]. CPD atom-based energy functions[64], used to score relative folding free energies and ligand-protein binding interaction energies, have been primarily constructed on the basis of their computational efficiency and amenability to pair-wise residue interaction energy decomposition. The computational overheads of automated enzyme design typically limit the search of substrate/ligand configurational space to six external translational and rotational degrees of freedom of pre-calculated conformers of low internal energy. Automated computational methods do not require any prior

knowledge of natural sequence variation, although appropriate constraints on the types of permitted residue substitutions can be applied.

Classical ligand docking methods alone can be used to predicting the preferred binding mode of (non-) natural ligands within the binding pocket of individual mutant or wild-type protein structures. Scoring functions predict ligand-protein binding energies, in which top ranked scores correspond to the most likely ligand configuration. Key residue positions involved in substrate binding and catalysis may be located through reasoned analysis of experimental or model structural data. Docking applications range from enzyme drug design to engineer enzyme promiscuity to enable conversion of synthetic substrates. For example, by manually docking the synthetic OHB substrate in to the crystal structure of (L)-lactate dehydrogenase from *Geobacillus stearothermophilus*, Walther and colleagues[65] observed a steric clash between the glutamine residue at position 85 and the synthetic substrate (**Figure 2.6**). Saturation mutagenesis at this position revealed the variant Q85C from *E. coli* to confer enhanced enzyme activity on OHB when compared to the wild-type counterpart.

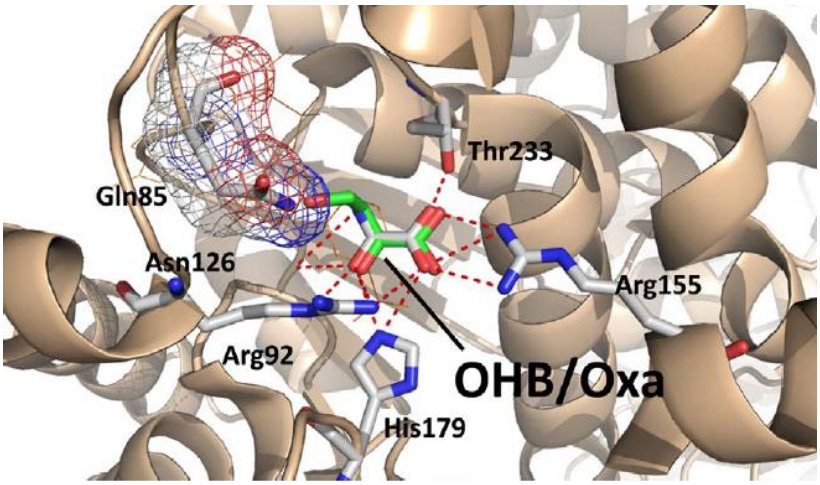

Figure 2.6. Active site region in X-ray crystal structure of the (L)-lactate dehydrogenase from *Geobacillus stearothermophilus*. The synthetic OHB substrate (green sticks) was manually docked into the enzyme complex bound with NAD^+ and the substrate analogue oxamate (PDB code 1ldn). Side-chains of active site residues are shown in a grey stick representation. From Walther et al[65].

While early ligand docking tools focused primarily on the assumption of a static nature of the receptor protein, changes observed in protein conformation upon binding of a small molecule suggested that both protein backbone and side-chain and ligand flexibility were

important to correctly model protein/ligand interactions. Numerous tools have been developed to better address the ligand docking over the past decades (e.g. Gold, FlexX, Dock, AutoDock, Glide, RosettaLigand). These tools utilize a wide range of protein representations, sampling algorithms and scoring functions to predict protein/ligand binding geometry[66,67]. While many predicted residue mutations are typically confined to the active site, rational engineering should ideally aim for a complete description of the enzymatic process, in which substrate migration and recognition (studied by biophysical techniques), and the biochemical reaction (modeled by combined quantum mechanics/molecular mechanics (QMMM) methods) are addressed[68].

2.5.2 Directed evolution

As opposed to enzyme rational design, the strategy of directed evolution mimics the theory of Darwinian evolution (mutation, recombination and selection) to evolve template enzymes towards an impaired user-specified phenotype. The workflow for a standard directed evolution assay can be divided into two consecutive stages: (a) *in vitro* diversity generation and transformation, and (b) screening and selection. Each step can impose significant technical limitations on the efficient exploration of sequence space[69].

Diversity generation and transformation

Next to the selection of an appropriate template enzyme, genotypic variability is introduced *in vitro* into the protein-coding sequence. Although knowledge of protein structure or respective catalytic mechanism is not necessary for diversity generation, a thoughtful analysis is required so that a mutant library can significantly cover the protein sequence space. For example, in a protein composed of 100 amino acids random mutagenesis could generate up to 1.3×10^{130} variants. While ensuring a full coverage of the sequence space, the overwhelming majority of residue substitutions will result in unfolded or misfolded proteins, and neutral or deleterious effects on the thermal stability and biological function of correctly folded proteins. In addition, screening and selection methods are limited in throughput (as later discussed). To this end, diversity generation techniques are commonly limited to a few possible substitutions in the sequence space (e.g. two substitutions per 100 amino acids yield up to 2.0×10^{6} possible variants) thereby increasing the average quality of mutant libraries. Random mutagenesis of a given

protein-coding sequence is normally performed by error-prone PCR with the aid of low-fidelity DNA polymerases, followed by ligation of resulting DNA fragments to a plasmid backbone. The obtained plasmid libraries are typically composed of 10^{12} molecules per mg plasmid.

Subsequently, plasmid libraries are introduced in a host strain so that protein-encoding genes can gain *in vivo* functionality. *E. coli* and *S. cerevisiae* are the most common host strains, but other hosts can be used to meet the specifications of the protein of interest. Transformation of an *in vitro* library into an *in vivo* screening context can entail significant losses in library size. Enabling screening and selection in an *in vitro* environment through compartmentalization, the utilization of cell-free transcription-translation systems has lately gained particular attention as it bypasses the process of library transformation. Out of the scope of this thesis, an excellent review[70] on enzyme directed evolution through cell-free systems is available. Another alternative would be the utilization of bacterial mutator strains (e.g. *E. coli* XL1-Red) to generate *in vivo* mutant libraries thereby avoiding the low efficiencies imposed by plasmid transformation[71].

Screening and selection

Last but not least in a directed evolution campaign stand the screening and selection of strains expressing best-performing enzyme variants. Screening and selection processes represent the most significant bottleneck in directed evolution as they need to be optimized in each case scenario considering three main factors: throughput, sensitivity and selectivity. A variety of screening assays ranging in their throughput are presented in **Figure 2.7**.

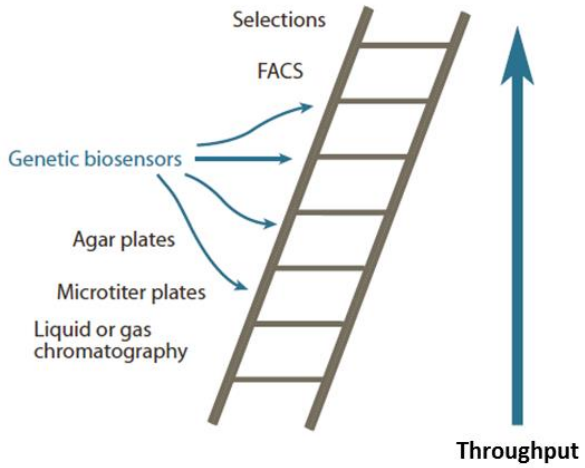

Figure 2.7. Flowchart of a directed evolution campaign, typically based on two main stages: diversity generation and transformation, followed by screening and selection of improved variants. From Dietrich et al[69].

In the bottom of the throughput scale are the traditional chromatographic and NMR techniques. Despite enabling a precise identification and quantification of metabolites from *in vivo* or *in vitro* enzyme bioconversion studies, their throughput is not compatible with a strategy of directed evolution. Microplate-based assays may therefore represent a more adequate strategy for screening enzyme mutant libraries (~10^4 variants per day)[69]. By miniaturizing test tubes to multiple wells, high density formats are available (up to 9600-well) albeit a 96-well format is the most widely used. After cell culture and enzyme production, traditional enzymatic reactions can be performed in a microtiter plate by adding reaction components together with manually added crude extracts of purified proteins. Using a plate reader, production formation or substrate consumption can be followed by UV-vis absorbance or fluorescence[72]. This strategy is particularly convenient for engineering NAD(P)(H)-dependent enzymes or NAD(P)(H)-coupled enzyme activities. For example, Iragues and colleagues[73] developed a general coupled assay for screening poorly active kinase enzymes, in which a NAD^+ alkali derivate is used as a measurable output signal at 360 nm. The method was successfully validated by screening an aspartate kinase mutant library towards improved unnatural malate kinase activities.

Higher in the throughput scale are the agar plate-based methods (~10^5 mutants per day). By shifting from an *in vitro* to an *in vivo* environment, colonies are incubated with the enzyme substrate whose conversion yields a (in)direct output signal (e.g. fluorescence, color). While agar plate screenings are easy to operate and outstanding in identifying active variants, they do not allow the inspection of different enzyme catalytic rates. For

this reason, they are commonly coupled with a second screening system aiming at estimating enzyme activities[74,75]. In an ingenuous example, an agar plate assay based on pH indicators was established in which color change upon substrate hydrolysis allowed the detection of improved esterase variants acting on 3-hydroxyester[58].

Growth complementation assays further extend the throughput scale. In this strategy, the utilization of a strain auxotrophic for an essential molecule is required (e.g. amino acids, key precursors). Growth complementation is a robust and accurate assay that enables selection on the basis of growth rates, which means that an auxotrophy will only be relieved when the lost enzymatic function is regained upon transformation with an improved enzyme variant. This technique is however limited in the enzyme sequence space, and it is frequently associated with detection in microtiter plates (growth rates) or agar plates (colony size). Another drawback is the limited growth dynamic range, as active enzymes cannot be distinguished from very active enzymes in which wild-type growth is restored. Fine-tuning of protein expression levels may solve this problem[69].

Close to the threshold of plasmid transformation efficiency and on the top of the throughput scale (~10^9 variants per day) is fluorescence activated cell sorting (FACS). Based on the concept of flow cytometry, FACS technology allows a rapid evaluation of size and fluorescence levels of each single cell with additional sorting. Specifically, libraries are first passed through a preliminary screen, in which top-fluorescent cells (~0.5-1 % of cell population) are sorted and enriched. Post-sorting viable cells are then submitted to a second screen in which the desired phenotype is confirmed. Further enrichments can be performed to reduce cell population and decrease false positive rates. The best performing enzyme variants (around 10-10^3) are further characterized with suitable low-throughput techniques. The direct detection of fluorescent metabolites (e.g. carotenoids, lipids) or proteins is well explored in FACS. In the last few years, particular attention has been paid to the development of genetic biosensors, which can detect a molecule of interest and output a fluorescent reporter protein. Living cells dispose of an extensive repertoire of natural sensor devices that by detecting a metabolite (e.g. riboswitches, transcription factors) are able to regulate downstream gene expression. Taking advantage of this mechanism, replacement of downstream gene by a fluorescent reporter protein ensures product detection. The co-expression of desired enzyme activities and an appropriate genetic sensor in an *in vivo* context, in which product formation rate regulates the reporter output signal, enables the creation of metabolite and/or whole-cell

biosensors. While FACS technology empowers throughput to unprecedent levels, genetic biosensors must be highly sensitive and selective towards the presence of a molecule of interest in a crowded intracellular environment. The creation of genetic sensors for detection of small molecules is further discussed in **Chapter 5**.

2.7 References

(1) Nielsen, J. (1998) Metabolic engineering: techniques for analysis of targets for genetic manipulations. *Biotechnol. Bioeng. 58*, 125–132.

(2) Cohen, S. N., Chang, A. C., Boyer, H. W., and Helling, R. B. (1973) Construction of biologically functional bacterial plasmids in vitro. *Proc. Natl. Acad. Sci. 70*, 3240–3244.

(3) Bailey, J. E. (1991) Toward a science of metabolic engineering. *Science 252*, 1668–1675.

(4) Woolston, B. M., Edgar, S., and Stephanopoulos, G. (2013) Metabolic Engineering: Past and Future. *Annu. Rev. Chem. Biomol. Eng. 4*, 259–288.

(5) Adamczyk, P. A., and Reed, J. L. (2017) *Escherichia coli* as a model organism for systems metabolic engineering. *Curr. Opin. Syst. Biol. 6*, 80–88.

(6) Feng, F., Lai, L., and Pei, J. (2018) Computational chemical synthesis analysis and pathway design. *Front. Chem. 6*.

(7) Seressiotis, A., and Bailey, J. E. (1988) MPS: An artificially intelligent software system for the analysis and synthesis of metabolic pathways. *Biotechnol. Bioeng. 31*, 587–602.

(8) Woolston, B. M., Edgar, S., and Stephanopoulos, G. (2013) Metabolic engineering: past and future. *Annu. Rev. Chem. Biomol. Eng. 4*, 259–288.

(9) Mavrovouniotis, M. L., Stephanopoulos, G., and Stephanopoulos, G. (1990) Computer-aided synthesis of biochemical pathways. *Biotechnol. Bioeng. 36*, 1119–1132.

(10) Copeland, W. B., Bartley, B. A., Chandran, D., Galdzicki, M., Kim, K. H., Sleight, S. C., Maranas, C. D., and Sauro, H. M. (2012) Computational tools for metabolic engineering. *Metab. Eng. 14*, 270–280.

(11) Wang, L., Dash, S., Ng, C. Y., and Maranas, C. D. (2017) A review of computational tools for design and reconstruction of metabolic pathways. *Synth. Syst. Biotechnol. 2*, 243–252.

(12) Delépine, B., Duigou, T., Carbonell, P., and Faulon, J.-L. (2018) RetroPath2.0: a retrosynthesis workflow for metabolic engineers. *Metab. Eng. 45*, 158–170.

(13) Kumar, A., Wang, L., Ng, C. Y., and Maranas, C. D. (2018) Pathway design using de novo steps through uncharted biochemical spaces. *Nat. Commun. 9*.

(14) Medema, M. H., van Raaphorst, R., Takano, E., and Breitling, R. (2012) Computational tools for the synthetic design of biochemical pathways. *Nat. Rev. Microbiol. 10*, 191–202.

(15) Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., Khandurina, J., Trawick, J. D., Osterhout, R. E., Stephen, R., Estadilla, J., Teisan, S., Schreyer, H. B., Andrae, S., Yang, T. H., Lee, S. Y., Burk, M. J., and Van Dien, S. (2011) Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nat. Chem. Biol. 7*, 445–452.

(16) Fehér, T., Planson, A.-G., Carbonell, P., Fernández-Castané, A., Grigoras, I., Dariy, E., Perret, A., and Faulon, J.-L. (2014) Validation of RetroPath, a computer-aided design tool for metabolic pathway engineering. *Biotechnol. J. 9*, 1446–1457.

(17) Erb, T. J., Jones, P. R., and Bar-Even, A. (2017) Synthetic metabolism: metabolic engineering meets enzyme design. *Curr. Opin. Chem. Biol. 37*, 56–62.

(18) Zhang, X., Wang, X., Shanmugam, K. T., and Ingram, L. O. (2011) L-Malate production by metabolically engineered *Escherichia coli*. *Appl. Environ. Microbiol. 77*, 427–434.

(19) Rossoni, L., Carr, R., Baxter, S., Cortis, R., Thorpe, T., Eastham, G., and Stephens, G. Engineering *Escherichia coli* to grow constitutively on D-xylose using the carbon-efficient Weimberg pathway. *Microbiol.* 164, 287–298.

(20) Shen, C. R., Lan, E. I., Dekishima, Y., Baez, A., Cho, K. M., and Liao, J. C. (2011) Driving forces enable high-titer anaerobic 1-butanol synthesis in *Escherichia coli*. *Appl. Environ. Microbiol. 77*, 2905–15.

(21) Hossain, G. S., Nadarajan, S. P., Zhang, L., Ng, T.-K., Foo, J. L., Ling, H., Choi, W. J., and Chang, M. W. (2018) Rewriting the metabolic blueprint: advances in pathway diversification in microorganisms. *Front. Microbiol. 9*, 155.

(22) Shen, X., Wang, J., Wang, J., Chen, Z., Yuan, Q., and Yan, Y. (2017) High-level de novo biosynthesis of arbutin in engineered *Escherichia coli*. *Metab. Eng. 42*, 52–58.

(23) Zelcbuch, L., Lindner, S. N., Zegman, Y., Vainberg Slutskin, I., Antonovsky, N.,

Gleizer, S., Milo, R., and Bar-Even, A. (2016) Pyruvate Formate-lyase enables efficient growth of *Escherichia coli* on acetate and formate. *Biochemistry 55*, 2423–2426.

(24) Alberto Martínez-Núñez, M., Rodríguez-Escamilla, Z., Rodríguez-Vázquez, K., and Pérez-Rueda, E. (2017) Tracing the repertoire of promiscuous enzymes along the metabolic pathways in archaeal organisms. *Life (Basel)* 7.

(25) Valdehuesa, K. N. G., Lee, W.-K., Ramos, K. R. M., Cabulong, R. B., Choi, J., Liu, H., Nisola, G. M., and Chung, W.-J. (2015) Identification of aldehyde reductase catalyzing the terminal step for conversion of xylose to butanetriol in engineered *Escherichia coli*. *Bioprocess Biosyst. Eng. 38*, 1761–1772.

(26) Alkim, C., Cam, Y., Trichez, D., Auriol, C., Spina, L., Vax, A., Bartolo, F., Besse, P., François, J. M., and Walther, T. (2015) Optimization of ethylene glycol production from (D)-xylose via a synthetic pathway implemented in *Escherichia coli*. *Microb. Cell Fact. 14*, 127.

(27) Cam, Y., Alkim, C., Trichez, D., Trebosc, V., Vax, A., Bartolo, F., Besse, P., François, J. M., and Walther, T. (2016) Engineering of a synthetic metabolic pathway for the assimilation of (D)-xylose into value-added chemicals. *ACS Synth. Biol. 5*, 607–618.

(28) Jarboe, L. R. (2011) YqhD: a broad-substrate range aldehyde reductase with various applications in production of biorenewable fuels and chemicals. *Appl. Microbiol. Biotechnol. 89*, 249–257.

(29) Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D. S., and Milo, R. (2011) The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry 50*, 4402–4410.

(30) Walther, T., Topham, C. M., Irague, R., Auriol, C., Baylac, A., Cordier, H., Dressaire, C., Lozano-Huguet, L., Tarrat, N., Martineau, N., Stodel, M., Malbert, Y., Maestracci, M., Huet, R., André, I., Remaud-Siméon, M., and François, J. M. (2017) Construction of a synthetic metabolic pathway for biosynthesis of the non-natural methionine precursor 2,4-dihydroxybutyric acid. *Nat. Commun. 8*, 15828.

(31) Schwander, T., Schada von Borzyskowski, L., Burgener, S., Cortina, N. S., and Erb, T. J. (2016) A synthetic pathway for the fixation of carbon dioxide in vitro. *Science*

(80-.). 354, 900–904.

(32) Siegel, J. B., Smith, A. L., Poust, S., Wargacki, A. J., Bar-Even, A., Louw, C., Shen, B. W., Eiben, C. B., Tran, H. M., Noor, E., Gallaher, J. L., Bale, J., Yoshikuni, Y., Gelb, M. H., Keasling, J. D., Stoddard, B. L., Lidstrom, M. E., and Baker, D. (2015) Computational protein design enables a novel one-carbon assimilation pathway. *Proc. Natl. Acad. Sci. 112*.

(33) Lee, S. Y., Park, J. M., and Kim, T. Y. (2011) Application of Metabolic Flux Analysis in Metabolic Engineering, in *Methods in enzymology*, 67–93.

(34) Barton, N. R., Burgard, A. P., Burk, M. J., Crater, J. S., Osterhout, R. E., Pharkya, P., Steer, B. A., Sun, J., Trawick, J. D., Van Dien, S. J., Yang, T. H., and Yim, H. (2015) An integrated biotechnology platform for developing sustainable chemical processes. *J. Ind. Microbiol. Biotechnol. 42*, 349–360.

(35) Orth, J. D., Thiele, I., and Palsson, B. Ø. (2010) What is flux balance analysis? *Nat. Biotechnol. 28*, 245–248.

(36) Salis, H. M., Mirsky, E. A., and Voigt, C. A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol. 27*, 946–950.

(37) Nowroozi, F. F., Baidoo, E. E. K., Ermakov, S., Redding-Johanson, A. M., Batth, T. S., Petzold, C. J., and Keasling, J. D. Metabolic pathway optimization using ribosome binding site variants and combinatorial gene assembly. *Appl. Microbiol. Biotechnol.* 98, 1567–1581.

(38) Portnoy, V. A., Bezdan, D., and Zengler, K. (2011) Adaptive laboratory evolution—harnessing the power of biology for metabolic engineering. *Curr. Opin. Biotechnol. 22*, 590–594.

(39) Dueber, J. E., Wu, G. C., Malmirchegini, G. R., Moon, T. S., Petzold, C. J., Ullal, A. V, Prather, K. L. J., and Keasling, J. D. (2009) Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol. 27*, 753–759.

(40) Lewicka, A. J., Lyczakowski, J. J., Blackhurst, G., Pashkuleva, C., Rothschild-Mancinelli, K., Tautvaišas, D., Thornton, H., Villanueva, H., Xiao, W., Slikas, J., Horsfall, L., Elfick, A., and French, C. (2014) Fusion of pyruvate decarboxylase and alcohol dehydrogenase increases ethanol production in *Escherichia coli*. *ACS Synth. Biol. 3*, 976–978.

(41) Cooper, G. M., and Hausman, R. E. (2013) The cell. a molecular approach. ASM Press.

(42) Shanmugam, S., and Sathishkumar, T. (2009) Enzyme technology. I.K. International Pub. House.

(43) Tyzack, J. D., Furnham, N., Sillitoe, I., Orengo, C. M., and Thornton, J. M. (2017) Understanding enzyme function evolution from a computational perspective. *Curr. Opin. Struct. Biol. 47*, 131–139.

(44) Norrgård, M. A. (2011) Modulating enzyme functions by semi-rational redesign and chemical modifications: a study on Mu-class glutathione transferases. Doctoral dissertation.

(45) Gerhart, J. C., and Pardee, A. B. (1962) The Enzymology of control by feedback inhibition. *J. Biol. Chem.* 237, 891–896.

(46) Xu, D., Zhao, J., Cao, G., Wang, J., Li, Q., Zheng, P., Zhao, S., and Sun, J. (2018) Removal of feedback inhibition of corynebacterium glutamicum phosphoenolpyruvate carboxylase by addition of a short terminal peptide. *Biotechnol. Bioprocess Eng. 23*, 72–78.

(47) Dalby, P. A. (2007) Engineering enzymes for biocatalysis. *Recent. Pat. Biotechnol.* 1, 1–9.

(48) Copeland, R. A. (2004) Enzymes: a practical introduction to structure, mechanism, and data analysis, 2nd Edition. John Wiley & Sons, Ltd, Chichester, UK.

(49) Gupta, R. D. (2016) Recent advances in enzyme promiscuity. *Sustain. Chem. Process. 4*.

(50) Hedstrom, L. (2010) Enzyme specificity and selectivity, in *Encyclopedia of Life Sciences*. John Wiley & Sons, Ltd, Chichester, UK.

(51) Choi, B., Rempala, G. A., and Kim, J. K. (2017) Beyond the Michaelis-Menten equation: Accurate and efficient estimation of enzyme kinetic parameters. *Sci. Rep. 7*.

(52) Scheer, M., Grote, A., Chang, A., Schomburg, I., Munaretto, C., Rother, M., Söhngen, C., Stelzer, M., Thiele, J., and Schomburg, D. (2011) BRENDA, the enzyme information system in 2011. *Nucleic Acids Res. 39*, 670-676.

(53) Martin, C. H., Nielsen, D. R., Solomon, K. V., and Prather, K. L. J. (2009) Synthetic metabolism: engineering biology at the protein and pathway Scales. *Chem. Biol. 16*, 277–286.

(54) Schomburg, I., Jeske, L., Ulbrich, M., Placzek, S., Chang, A., and Schomburg, D. (2017) The BRENDA enzyme information system–From a database to an expert system. *J. Biotechnol. 261*, 194–206.

(55) Chen, R. (2001) Enzyme engineering: rational redesign versus directed evolution. *Trends Biotechnol. 19*, 13–14.

(56) Khersonsky, O., Roodveldt, C., and Tawfik, D. (2006) Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr. Opin. Chem. Biol. 10*, 498–508.

(57) Lutz, S. (2010) Beyond directed evolution: semi-rational protein engineering and design. *Curr. Opin. Biotechnol. 21*, 734–743.

(58) Bornscheuer, U. T., and Pohl, M. (2001) Improved biocatalysts by directed evolution and rational protein design. *Curr. Opin. Chem. Biol. 5*, 137–143.

(59) Li, X., Zhang, Z., and Song, J. (2012) Computational enzyme design approaches with significant biological outcomes: progress and challenges. *Comput. Struct. Biotechnol. J. 2*.

(60) Yin, Y., and Kirsch, J. F. (2007) Identification of functional paralog shift mutations: conversion of *Escherichia coli* malate dehydrogenase to a lactate dehydrogenase. *Proc. Natl. Acad. Sci. 104*, 17353–7.

(61) Leaver-Fay, A., Tyka, M., Lewis, S. M., Lange, O. F., Thompson, J., Jacak, R., Kaufman, K. W., Renfrew, P. D., Smith, C. A., Sheffler, W., Davis, I. W., Cooper, S., Treuille, A., Mandell, D. J., Richter, F., Ban, Y.-E. A., Fleishman, S. J., Corn, J. E., Kim, D. E., Lyskov, S., Berrondo, M., Mentzer, S., Popović, Z., Havranek, J. J., Karanicolas, J., Das, R., Meiler, J., Kortemme, T., Gray, J. J., Kuhlman, B., Baker, D., and Bradley, P. (2011) Rosetta3, in *Methods in enzymology*, 545–574.

(62) Eiben, C. B., Siegel, J. B., Bale, J. B., Cooper, S., Khatib, F., Shen, B. W., Players, F., Stoddard, B. L., Popovic, Z., and Baker, D. (2012) Increased Diels-Alderase activity through backbone remodeling guided by Foldit players. *Nat. Biotechnol. 30*, 190–192.

(63) Siegel, J. B., Zanghellini, A., Lovick, H. M., Kiss, G., Lambert, A. R., St Clair, J. L.,

Gallaher, J. L., Hilvert, D., Gelb, M. H., Stoddard, B. L., Houk, K. N., Michael, F. E., and Baker, D. (2010) Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science 329*, 309–313.

(64) Li, Z., Yang, Y., Zhan, J., Dai, L., and Zhou, Y. (2013) Energy functions in de novo protein design: current challenges and future prospects. *Annu. Rev. Biophys. 42*, 315–335.

(65) Walther, T., Calvayrac, F., Malbert, Y., Alkim, C., Dressaire, C., Cordier, H., and François, J. M. (2018) Construction of a synthetic metabolic pathway for the production of 2,4-dihydroxybutyric acid from homoserine. *Metab. Eng. 45*, 237–245.

(66) DeLuca, S., Khar, K., and Meiler, J. (2015) Fully flexible docking of medium sized ligand libraries with rosettaligand. *PLoS One 10*.

(67) Combs, S. A., Deluca, S. L., Deluca, S. H., Lemmon, G. H., Nannemann, D. P., Nguyen, E. D., Willis, J. R., Sheehan, J. H., and Meiler, J. (2013) Small-molecule ligand docking into comparative models with Rosetta HHS public access. *Nat Protoc 8*, 1277–1298.

(68) Acebes, S., Fernandez-Fueyo, E., Monza, E., Lucas, M. F., Almendral, D., Ruiz-Dueñas, F. J., Lund, H., Martinez, A. T., and Guallar, V. (2016) Rational enzyme engineering through biophysical and biochemical modeling. *ACS Catal. 6*, 1624–1629.

(69) Dietrich, J. A., McKee, A. E., and Keasling, J. D. (2010) High-throughput metabolic engineering: advances in small-molecule screening and selection. *Annu. Rev. Biochem. 79*, 563–590.

(70) Dodevski, I., Markou, G. C., and Sarkar, C. A. (2015) Conceptual and methodological advances in cell-free directed evolution. *Curr. Opin. Struct. Biol. 33*, 1–7.

(71) Muteeb, G., and Sen, R. (2010) Random mutagenesis using a mutator strain, in *Methods in molecular biology (Clifton, N.J.)*, 411–419.

(72) Xiao, H., Bao, Z., and Zhao, H. (2015) High throughput screening and selection methods for directed enzyme evolution. *Ind. Eng. Chem. Res. 54*, 4011–4020.

(73) Irague, R., Topham, C. M., Martineau, N., Baylac, A., Auriol, C., Walther, T.,

François, J.-M., André, I., and Remaud-Siméon, M. (2018) A generic HTS assay for kinase screening: validation for the isolation of an engineered malate kinase. *PLoS One 13*.

(74) Leemhuis, H., Kelly, R. M., and Dijkhuizen, L. (2009) Directed evolution of enzymes: library screening strategies. *IUBMB Life* 61, 222–228.

(75) Rigoldi, F., Donini, S., Redaelli, A., Parisini, E., and Gautieri, A. (2018) Review: engineering of thermostable enzymes for industrial applications. *APL Bioeng. 2*.

"Knowing is not enough; we must apply. Wishing is not enough; we must do."

—Johann Wolfgang Von Goethe

Chapter 3. Rational enzyme design towards improved microbial production of 2,4-dihydroxybutyric acid

3.1 Introduction

A central challenge in the creation of a resource-efficient and sustainable bioeconomy lies in the reduction of our dependence on fossil resources and a shift in focus towards the development of efficient biochemical pathways for the production of fuels and bio-based chemicals[1]. (L)-2,4-dihydroxybutyrate (DHB), an α-hydroxy acid whose industrial interest has been rapidly growing, can serve as a precursor of methionine analogues for animal nutrition[2]. Petrochemical synthesis of DHB is however not cost-competitive and no natural biosynthetic pathways leading to its production are currently known, despite its occurrence at trace levels in patients with succinic semialdehyde dehydrogenase deficiency[3]. Most recent developments in synthetic biology domain have enabled the development of non-natural biochemical pathways, leading to the production of economically relevant compounds[4–6] and in particular DHB. Previous work from our group has demonstrated the feasibility of DHB biosynthesis from glucose through the engineering of glyoxylate[7], homoserine[8] or artificial malyl-phosphate[9] pathways in *E. coli* (**Figure 3.1a-c**). More recently, Dischert and colleagues have disclosed a new artificial DHB route through the extension of a synthetic 1,2,4-butanetriol pathway (**Figure 3.1d**), thus enabling product formation from xylose, a pentose sugar and a major constituent of the hemicellulose fraction in lignocellulosic biomass[10].

In particular, DHB production from glucose via extension of the homoserine pathway is enabled by the introduction of a two-step pathway composed of two hitherto unknown enzymatic activities (**Figure 3.1a**). Homoserine is first converted into 2-keto-4-hydroxybutyrate (OHB) by the action of an engineered homoserine transaminase. This compound is then reduced to DHB by an improved OHB reductase variant. In an ongoing effort to improve DHB production, strategies to further increase the efficiency of the pathway are highly desired. Pathway optimization strategies[11] (*e.g.* gene disruption/overexpression, 5'-UTR engineering, plasmid copy number variation, codon optimization) to redirect carbon flux towards a molecule of interest are today well established in the scientific community, with numerous successful examples reported in the

literature[12,13]. However, those do not address situations in which poor performance of a given pathway is hindered by enzyme-associated limitations. Protein design has been described in these cases as an indispensable tool to engineer an optimal pathway[14].

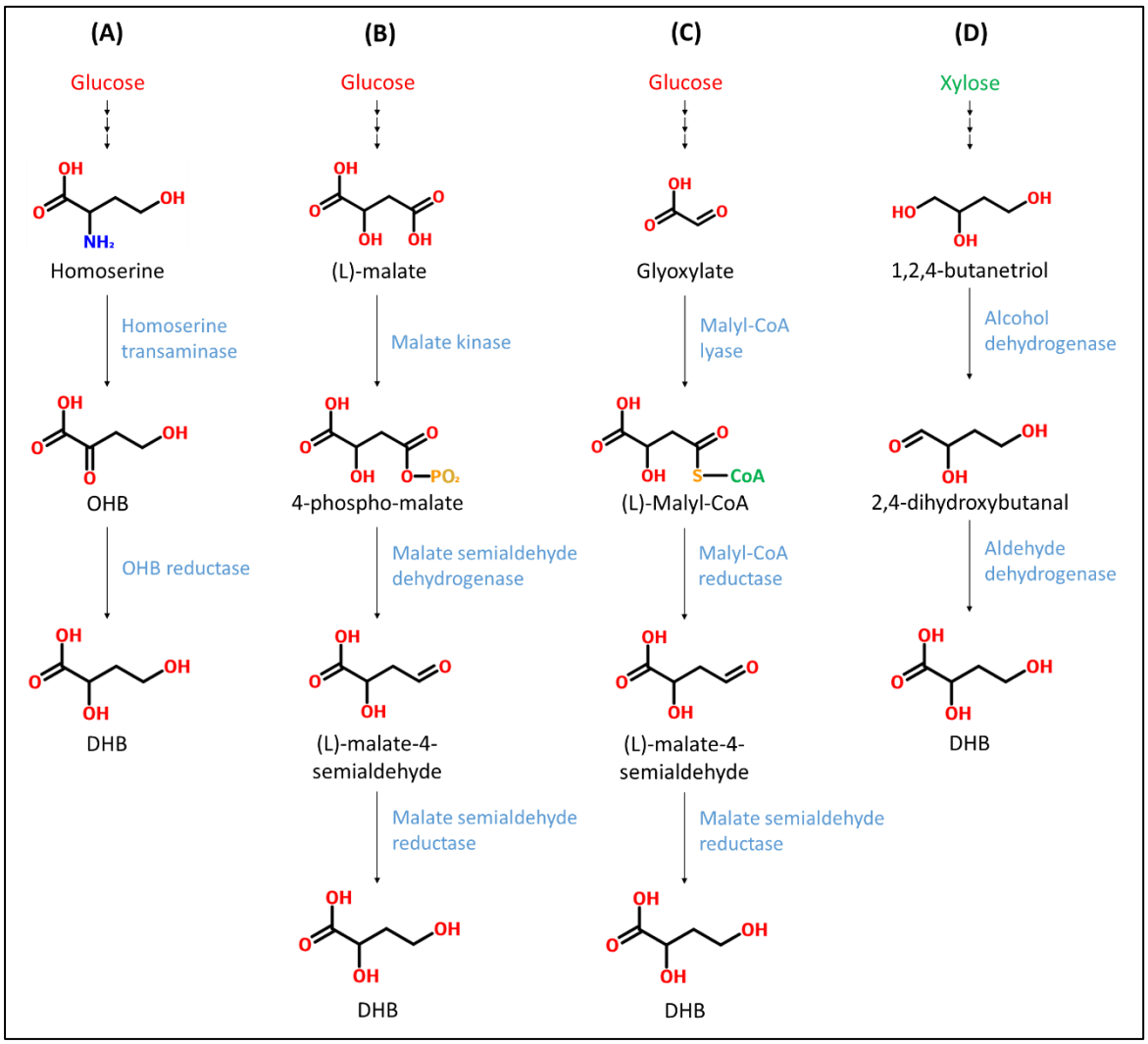

Figure 3.1. Disclosed synthetic (L)-2,4-dihydroxybutyrate (DHB) pathways starting from pentose and hexose sugars. Adapted from Dischert and colleagues[10]. Legend: OHB – 2-keto-4-hydroxybutyric acid.

Our group has previously identified enzymes displaying homoserine transaminase and OHB reductase activities, but which however possessed low affinities towards the corresponding substrates. In particular, studies carried out by our group have revealed the (L)-lactate dehydrogenases from *Lactococcus lactis* (Ll-LdhA)[2] and rabbit muscle (Rm-Ldh) as candidate OHB reductases[8,15]. (L)-lactate dehydrogenases (Ldh) are NAD-dependent dehydrogenases and one the most well-studied enzyme families, comprising over 100 fully

biochemically characterized members that also include (L)-malate dehydrogenases (Mdh). Ldh and Mdh notably both act on 2-keto acids: pyruvate and oxaloacetate, respectively. Additionally, they share a similar tertiary fold in which the nucleotide co-enzyme binding pocket and catalytic residues are conserved[16]. On the basis of these findings, a strategy to rationally engineer OHB reductase activity into the (L)-malate dehydrogenase from *E. coli* (Ec-Mdh) was employed based on a comparative structural analysis with Ldh enzymes complexed with sterically similar cognate substrates. The best-performing mutants as assessed *in vitro* were introduced into the homoserine DHB pathway and product titers evaluated.

3.2 Strategy for OHB reductase design

The dimeric Ec-Mdh enzyme was previously shown to possess only residual activity towards OHB (reductive reaction) and DHB (oxidative reaction). Malate dehydrogenases are highly selective and specific for dicarboxylic acids, while the OHB/DHB target molecules are monocarboxylic acids. Increased activity towards the OHB/DHB substrate/product couple was therefore envisaged using a rational design approach to guide site-directed mutagenesis of Ec-Mdh.

As a first step, the possibility of using a homologous protein template for OHB reductase design was investigated. Both Ldh and Mdh, catalyze the interconversion between 2-hydroxyacids and their corresponding 2-keto acids using the NAD/NADH co-enzyme system[17]. Ldh and Mdh share a similar tertiary structural fold in which the nucleotide binding pocket and catalytic residues are conserved, but yet maintain strong selectivity for mono- (pyruvate/lactate) and dicarboxylic acids (oxaloacetate/malate), respectively[16,18]. Key differences between the enzymes are found in the sequences of a mobile loop region (positions 79 through 91 in Ec-Mdh) covering the active-site. The loop is known to exist in two distinct conformations in Mdh: an open form, in which the enzyme is inactive, and a closed form, in which the enzyme is active (**Figure 3.2**).

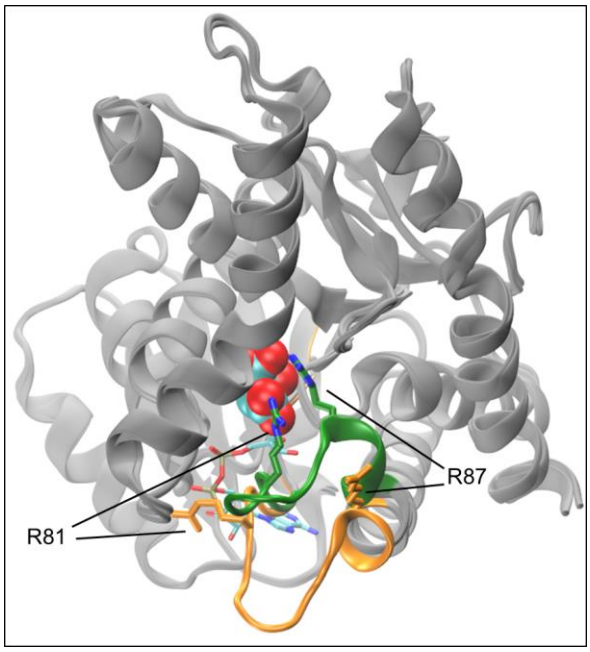

Figure 3.2. Superposition of X-ray structures of an Ec-Mdh complex with bound NAD^+ (stick representation), showing the "closed" form of the mobile active-site loop in green (PDB code 1ib6), and the "open" (gold colored) loop form of the *apo* enzyme (PDB code 1ie3). An overlay of the malate substrate from the modelled ternary complex with coenzyme (PDB code 1cme) is shown as van der Waals spheres. Side-chain positions Arg81 and Arg87 involved in binding of the malate substrate are indicated in the two loop conformers.

Conformational change is triggered by the binding of the β-carboxylate group of the substrate to Arg81 and/or Arg87, and involves a change in the local secondary structure of the loop at positions 86 through 89 in Ec-Mdh[19,20]. In the homologous Ldh system, ordered binding of NADH and pyruvate triggers incremental conformational changes in the active site and the expulsion of solvent necessary for catalysis to occur. The existence of a common kinetic mechanism in Ldh and Mdh enzymes implies that a similar conformational transition of the active-site loop operates in both enzymes. Crucially however, Ldh active-site loop closure occurs in the absence of the favorable energy contributions of a stabilizing salt bridge formed between a second substrate carboxylate group, absent in pyruvate, and a second arginine guanidinium group (replaced by Glu81 in Ldh, **Table 3.1**). Since pyruvate and OHB are both monocarboxylic acids, the identification of compensatory stabilizing loop interactions in Ldh afforded a design strategy for the engineering of OHB reductase activity into Ec-Mdh.

Table 3.1. Comparison of Mdh/Ldh active-site loop sequences in selected species. Entirely conserved residue positions are highlighted in red. Main-chain residue conformation key: H, helix; P, +ve ϕ main-chain dihedral angle; C, random coil.

Enzyme / Source		Residue Position (*E. coli* MDH)													
		78	**79**	**80**	**81**	**82**	**83**	**84**	**85**	**86**	**87**	**88**	**89**	**90**	**91**
Conformation		P	C	C	C	C	C	P	C	C	C	H	H	H	H
E. coli	MDH	G	V	A	R	K	P	G	M	D	R	S	D	L	F
B. subtilis	MDH	G	I	A	R	K	P	G	M	S	R	D	D	L	V
B.longum	LDH	G	P	R	Q	K	P	G	Q	S	R	L	E	L	V
B. stearothermophilus	LDH	G	A	N	Q	K	P	G	E	T	R	L	D	L	V
Dogfish	LDH	G	A	R	Q	Q	E	G	E	S	R	L	N	L	V
Human heart	LDH	G	V	R	Q	Q	E	G	E	S	R	L	N	L	V
Pig muscle	LDH	G	A	R	Q	Q	E	G	E	S	R	L	N	L	V
T. thermophilus	LDH	G	V	A	Q	R	P	G	E	T	R	L	Q	L	L

3.3 Identification of target mutation sites in Ec-Mdh

Comparison of active-site loop sequences in representative Mdh and Ldh enzymes shows complete conservation of three structurally important residues at positions Gly78, Gly84 and Leu90, as well as the conservation of Arg87 directly involved in substrate binding (**Table 3.1**). Analysis of larger multiple sequence alignments of Mdh and Ldh enzymes further confirmed the heavily conserved nature of residue types at these positions. These common "key" or "canonical" residues[21] thus appear to be necessary for the stabilization of the active-site loop in the same overall conformation in Mdh and Ldh enzymes. On the other hand, comparison of overlaid closed loop conformers in experimental structures of Ldh and the binary complex of Ec-Mdh/NAD^+ (**Figure 3.3a**) revealed the presence of two internal loop stabilizing interactions specific to Ldh. The first is the replacement of Ala80 in Ec-Mdh by an arginine, which would be expected to assist in the orientation of the side-chain towards the solvent and stabilize the co-enzyme binding through favorable electrostatic interactions with phosphate groups bridging adenine and nicotinamide nucleotides. The second loop stabilizing interaction involves the substitution of Met85 by either glutamate or glutamine. Stabilization occurs via hydrogen bond formation between the main-chain -NH group at position 82 with either the carboxylate oxygen of a glutamate, or the side-chain carbamoyl oxygen of a glutamine. The mobile active-site loop in Ldh appears to be further

stabilized by a third specific set of non-covalent interactions at position 86 on the solvent-exposed surface (**Figure 3.3b**). In principle, these interactions can be reproduced in Ec-Mdh via the following residue exchanges: D86S (in the mobile loop), G179D (in the Gly176-Thr181 loop) and T211Y (in the Gly210-Ala216 helix). The mutually stabilizing side-chain interactions at the juncture of the three structural elements in Ldh fill the space otherwise occupied by the (A317) sulphate ion in the X-ray structure of Ec-Mdh.

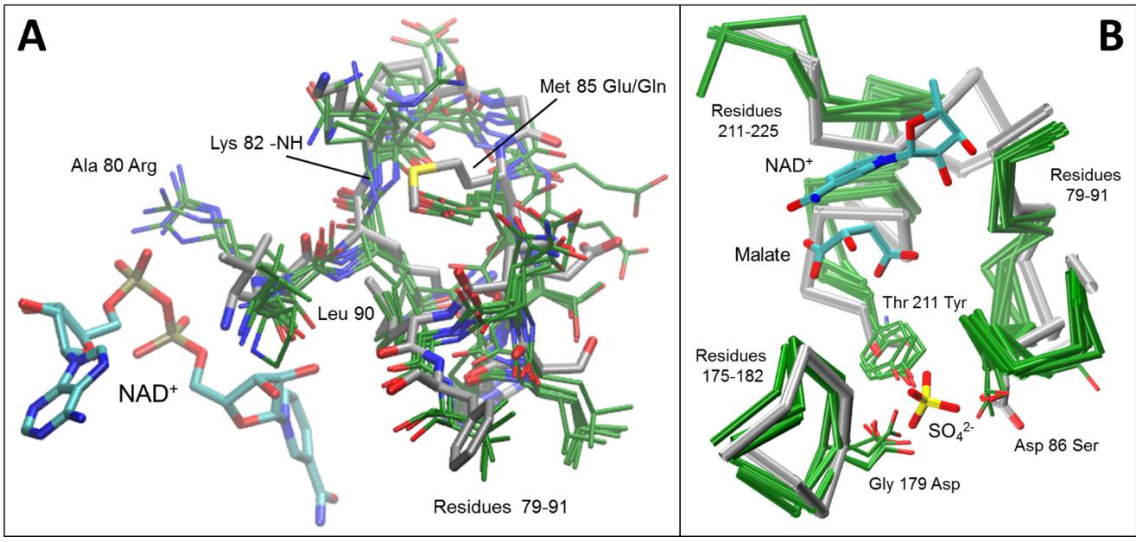

Figure 3.3. Structural alignment of Ldh X-ray structures from multiple organisms (PDB codes 1lth, 1ldn, 2v7p, 3ldh, 1i0z, 9ldt) depicted with green carbon atoms, with the binary complex of Ec-Mdh and NAD^+ (PDB code 1ib6). **(A)** Overlay of closed loop conformations (positions 79-91, Ec-Mdh). **(B)** Overlay of three structural elements in the solvent-exposed region surrounding protein active-site. The malate substrate, extracted from an Ec-Mdh model ternary complex with co-enzyme (PDB code 1cme), replaces the (A316) sulphate ion bound in the active site. The second sulphate (A317) ion which stabilizes the mobile loop (residues 79 to 81) is shown as a stick representation. Overlays were obtained by pairwise structural superposition at all aligned protein chain residue positions with respect to Ec-Mdh subject to a 2Å cut-off (C^{α} - C^{α}) separation distance.

Visual inspection of an Ec-Mdh ternary complex model construct with malate (reaction product) and NAD^+ (PDB code 1cme), and multiple sequence alignment analysis were used to identify putative target positions for mutation in the substrate and co-enzyme binding sites. Residue positions closest to the C_4 atom in bound malate, and by implication to the 4-OH group in the DHB/OHB couple, are Arg81, Val214 and Ser222. Estimation of the relative Shannon entropy (H_X, see **Methods**) of amino acid frequency distributions revealed that all three positions are fully conserved ($H_X = 0$) amongst Mdh enzymes, with more variation being observed in the wider Ldh/Mdh superfamily: H_X values for the Pfam

(Ldh_1_N and Ldh_1_C) protein family alignments at positions 81, 214 and 222 were 0.34, 0.43 and 0.54, respectively. Putative residue side-chain replacements at these positions in the active-site of Ec-Mdh were further explored by flexible docking of OHB (**Figure 3.4**), suggesting that mutation of residue Arg81 is likely to be essential for OHB reductase activity.

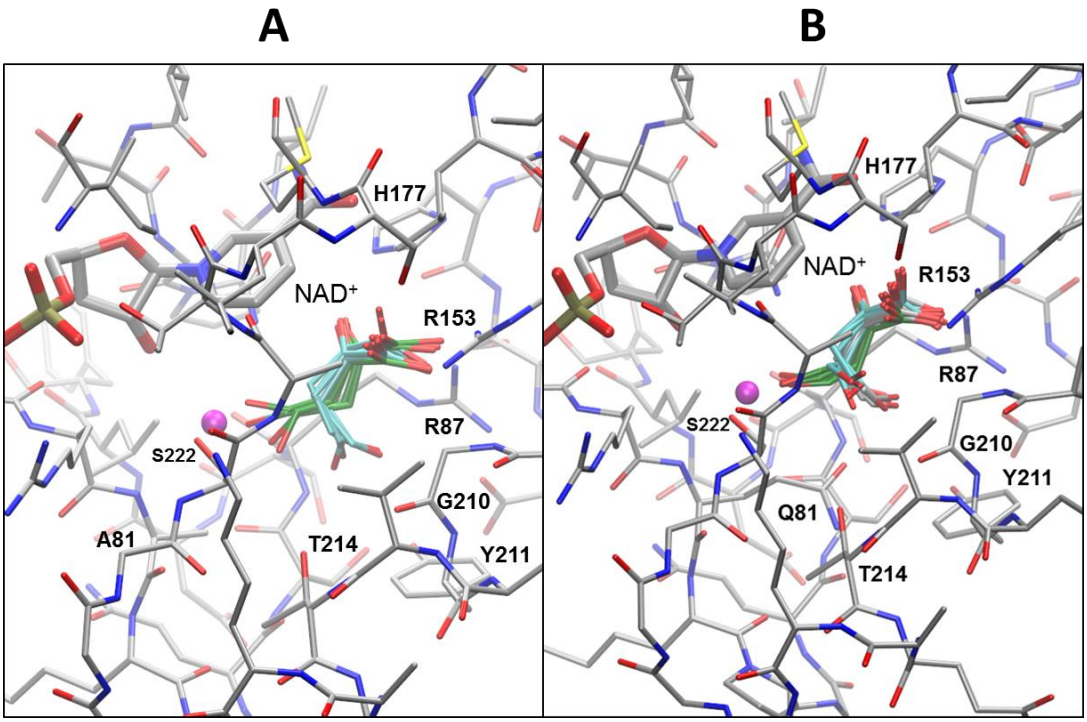

Figure 3.4. Active-site regions in molecular models of (A) A80R:R81A:M85Q:D86S:G179D:T211Y:V214T and (B) A80R:R81Q:M85Q:D86S:G179D:T211Y:V214T mutant Ec-Mdh with NAD^+ and flexibly docked OHB. The OHB carbons in five members of the two top-ranked clusters with equal interaction energies are shown in green and cyan in (A). Members of an additional high-ranking energy interaction OHB cluster are depicted in (B) with grey carbons. The OHB cluster with the most favorable interaction energy in (B) is that shown with green colored carbon atoms. A potentially displaceable crystallographic water molecule is represented as a magenta colored sphere. Each of the models is overlaid on the X-ray structure of an Ec-Mdh model ternary complex with co-enzyme (PDB code 1cme) from which it was derived as described in Methods.

In **Figure 3.4a**, OHB docking into an R81A-containing mutant resulted in two ligand conformational clusters (green and blue sticks) with equal Boltzmann averaged interaction energies. On the other hand, replacement of Arg81 by glutamine (as in Ldh, **Figure 3.4b**) revealed a clear preference (of 2.3 kcal mol^{-1}) for the green conformer. Whilst the latter can interact favorably with Gln81, it is at the same time forced into making close contact with the α-carbon of Gly210. The R81Q mutation did not therefore appear to confer any real

improvement in substrate/product binding interactions, and may indeed hinder access to the binding site. In this regard, the replacement of Arg81 by alanine seemed more preferable since it should allow for unimpeded access to the active-site by the substrate. As shown in **Figure 3.4b**, favorable hydrogen bonding interactions of the substrate/product 4-OH group can be made in two of the three bound ligand conformations (shown with green or grey carbons) with the side-chain of Ser222. Thus, there did not seem to be any obvious advantage to be gained from the modification of this residue. In contrast, isosteric mutation of Val214 to threonine might be expected to aid in the orientation of the Ser222 side-chain through intra-protein hydrogen bonding. However, the closest approach distance of the substrate/product 4-OH group to the Thr214 oxygen was observed to be too long (at approximately 4.0 to 4.6 Å) for the formation a good hydrogen bond. The effects of amino acid changes introduced at this position were investigated experimentally and the results are described in the next section.

Further analysis of the Ldh multiple sequence alignment notably revealed the presence of smaller residues close to the co-enzyme nicotinamide ring at equivalent positions Ile12, Ile117 and Met227 in Ec-Mdh (**Figure 3.5**). Substitution of smaller residue types at these positions could potentially influence the positioning of the nicotinamide ring, and thereby exert a direct effect on catalysis (*e.g.* on the rate of hydride transfer). Candidate residue type replacements at these positions were identified from frequency analysis of the Ldh multiple sequence alignment: position 12: Val; position 117: Ala or Val; position 227: Ile or Val. The introduction of mutations at these three residue positions was by inference expected to shift the substrate specifity of Ec-Mdh towards the OHB/DHB couple.

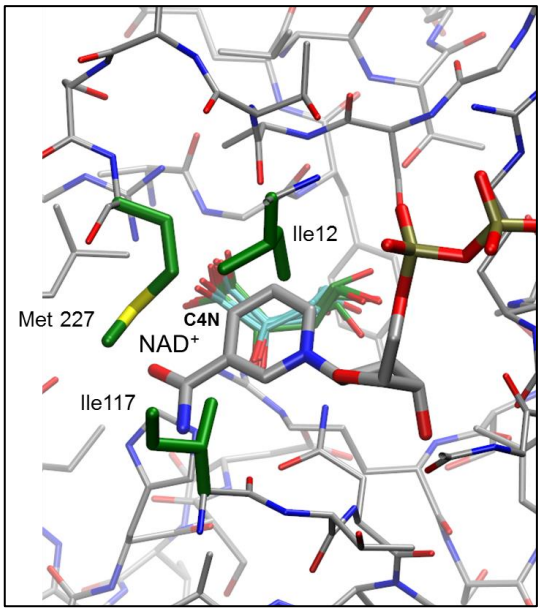

Figure 3.5. Positions of variable residue positions (highlighted with green carbon side-chain atoms) in Ldh/Mdh superfamily in the vicinity of the co-enzyme nicotinamide ring in a modelled ternary complex of Ec-Mdh mutant (A80R:R81A:M85Q:D86S:G179D:T211Y:V214T) NAD^+ and flexibly docked OHB conformer representatives of the top-ranked ligand pose (shown in stick representation with cyan carbon atoms). The model is overlaid on the X-ray structure of an Ec-Mdh model ternary complex with co-enzyme (PDB code 1cme) from which it was derived as described in Methods.

3.4 *In vitro* analysis of site-directed variants

In order to experimentally assess the rational engineering approach, both wild-type and site-directed enzyme variants were characterized in the reductive (biosynthetic) reaction using OHB as the substrate (**Figure 3.6**). As expected, the wild-type enzyme Ec-Mdh displayed very little OHB reductase activity (k_{cat} = 0.03 s^{-1}). Point mutations were then introduced in a step-wise manner into the substrate-binding region of Ec-Mdh, either individually or as groups. Substitution of Arg81 by alanine resulted in a dramatic 287-fold increase in maximum enzyme activity on OHB (k_{cat} = 8.6 s^{-1}). In contrast, insertion of an additional mutation at position 214 in the R81A:V214T variant resulted in lower activity (k_{cat} = 3.1 s^{-1}).

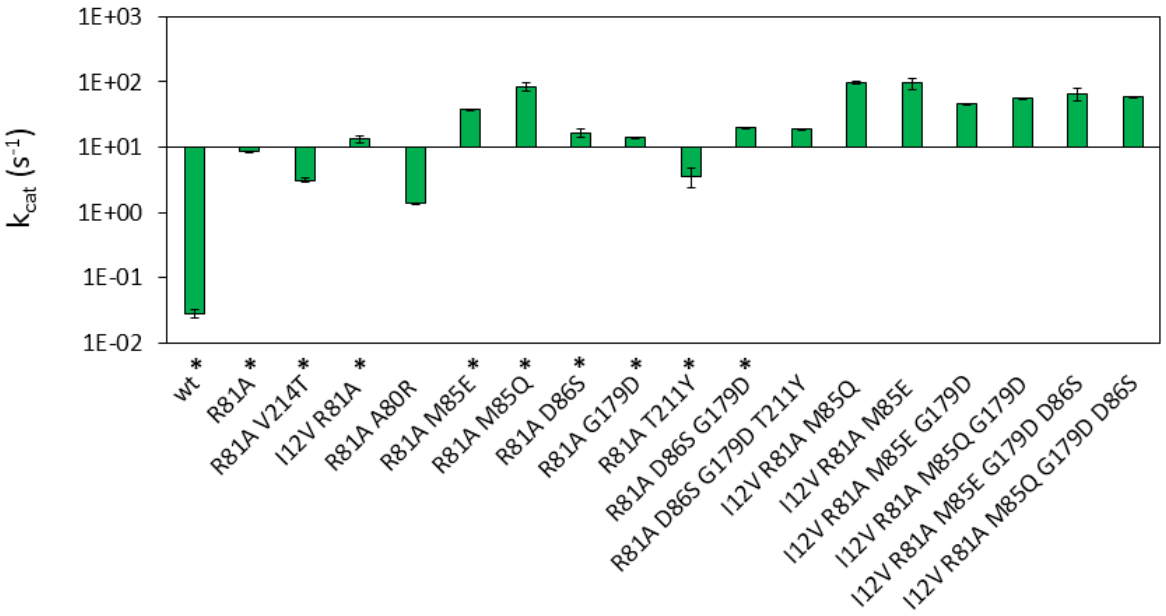

Figure 3.6. Activity of wild-type and Ec-Mdh enzymes on OHB, expressed in s^{-1}, and shown on a logarithmic-scale. The abscissa positioning allows the relative effects of each enzyme mutant to be readily discerned. Enzymes highlighted with an asterisk could not be saturated under the test conditions, and the reported activities were obtained at a substrate concentration of 50 mM. Activities of the other mutant enzymes correspond to calculated turnover numbers. The results are the mean of at least two biological replicate experiments. Error bars correspond to the standard deviation of the mean.

The R81A mutation was then combined with amino acid changes either at residue position Ile12 proximal to the coenzyme nicotinamide ring, residue positions in the mobile loop covering the active site (Ala80, Met85 and Asp86) or at positions Gly179 and Thr211 contacting the mobile loop. Increased OHB reductase activity (k_{cat} = 13.4 s^{-1}) was observed in the I12V:R81A double mutant, as compared to the R81A variant, when residue changes were simultaneously introduced in the substrate binding site and the coenzyme nicotinamide ring binding pocket. Grouping amino acid changes in the substrate binding site with those elsewhere in the active-site resulted in distinct underlying patterns of response. Thus, while changes in loop residue positions Met85, Asp86 and Gly179 enhanced OHB reductase activities yielding k_{cat} values of up to 85.4 s^{-1} (variant R81A:M85Q), substitutions at the Ala80 and Thr211 target sites showed the opposite tendency. It is however noteworthy that the R81A:M85Q mutant could not be saturated within the tested OHB concentration range, and that the reported activity value was obtained at a substrate concentration of 50 mM. The results suggest that residue alterations at different locations within the active-site region elicit synergistic effects on both binding and catalysis.

To investigate whether OHB reductase activity could be further improved, the simultaneously introduction of amino acid changes in the substrate binding site, coenzyme binding pocket and at a structural contact region on the external surface of the mobile loop covering the active-site during catalysis was attempted. Variants from this group were found to be highly active on OHB, but more importantly could be saturated at lower substrate concentrations (< 35 mM). The catalytic efficiencies (k_{cat}/K_m) of the wild-type and best mutant enzymes were then evaluated for comparison on both OHB and the natural substrate oxaloacetate (**Figure 3.7**). The results indicate that Ec-Mdh engineering not only significantly improved enzyme efficiency for OHB, but also considerably decreased that towards oxaloacetate by at least 50-fold. Despite observing increased catalytic efficiencies towards OHB, the cumulation of amino acid changes in the best mutants resulted in variants with lower specificities (i.e. $(k_{cat}/K_m)_{OHB}$ / $(k_{cat}/K_m)_{Oxaloacetate}$). However, intracellular concentrations of oxaloacetate are known to be very low, and for this reason we decided to fully kinetically characterize the four variants with the highest catalytic efficiency towards OHB notwithstanding their lowered specificity[22,23].

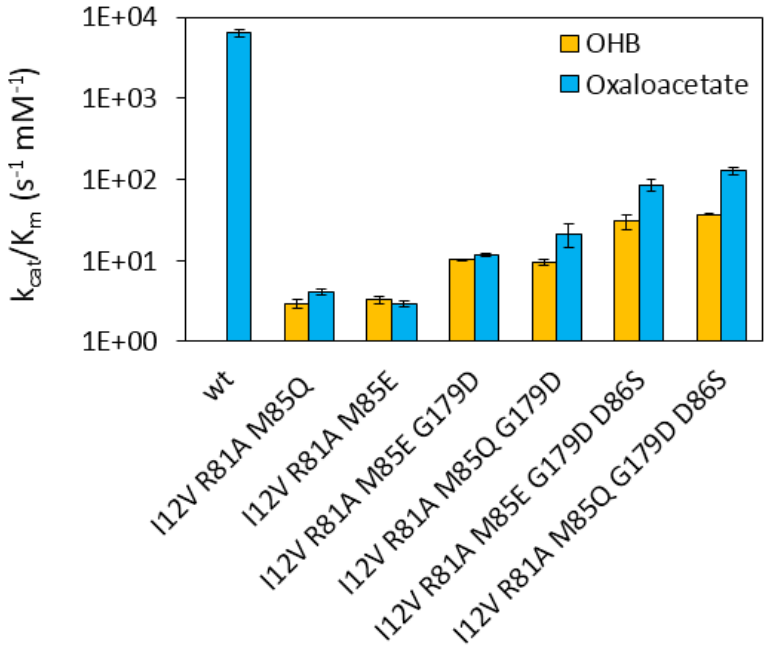

Figure 3.7. Catalytic efficiency (k_{cat}/K_m) in units of s^{-1} mM^{-1} of the wild-type and best Ec-Mdh mutant enzymes on OHB and oxaloacetate. The results are presented on a logarithmic scale as the mean of at least two biological replicate experiments. Error bars correspond to the standard deviation of the mean.

Table 3.2 shows the kinetic parameters for the reductive (biosynthetic) reaction of the wild-type enzyme and the best mutants, for brevity henceforth denoted as Ec-Mdh-4E (I12V:R81A:M85E:G179D), Ec-Mdh-4Q (I12V:R81A:M85Q:G179D), Ec-Mdh-5E (I12V:R81A:M85E:D86S:G179D) and Ec-Mdh-5Q (I12V:R81A:M85Q:D86S:G179D). In view of its structural similarity to OHB, enzyme kinetic analysis was also carried out using pyruvate as substrate. All four mutants exhibited affinity for oxaloacetate, OHB and pyruvate, in contrast to the wild-type enzyme that possesses high stringency for its natural substrate. Although the Ec-Mdh-5Q variant showed the highest affinity and efficiency towards OHB (K_m = 1.6 mM, k_{cat}/K_m = 37 s^{-1} mM^{-1}), Ec-Mdh-4E was the most specific enzyme towards OHB (($k_{cat}/K_m)_{OHB}$ / ($k_{cat}/K_m)_{Oxaloacetate}$ = 0.87). Of particularly important note, with the exception of Ec-Mdh-4E variant all four mutants displayed substrate-inhibition kinetics towards OHB (K_i ~30 mM) and oxaloacetate. Lower catalytic efficiencies were recorded when using pyruvate as substrate (k_{cat}/K_m < 0.5 s^{-1} mM^{-1}), and consequently Ec-Mdh mutants were more specific towards OHB (by up to two-orders of magnitude). The Ec-Mdh wild-type and variants were also characterized in the oxidative reaction using (L)-malate, (D/L)-DHB and (L)-lactate as substrates (**Table 3.3**). As expected, the catalytic efficiency of wild-type enzyme was much higher (by two orders of magnitude) in the reductive reaction (k_{cat}/K_m = 6,387 s^{-1} mM^{-1}) than on (L)-malate (k_{cat}/K_m = 40 s^{-1} mM^{-1}). The same trend was observed for Ec-Mdh mutants with respect to both natural and synthetic substrate/product couples. Additionally, the four best mutants displayed much higher DHB dehydrogenase activities when compared to the wild-type enzyme (by up to four orders of magnitude), but protein engineering resulted in considerably low affinities for the racemic mixture of DHB (Ec-Mdh-5Q, K_m = 64.1 mM L-DHB). Finally, none of the mutants demonstrated saturation kinetics when using (L)-lactate as the substrate.

Table 3.2. Comparison of kinetic parameters for reductive (biosynthetic) reaction catalyzed by Ec-Mdh wild-type and best variants.

Enzyme	**Ec-Mdh**	**Ec-Mdh-4E**	**Ec-Mdh-4Q**	**Ec-Mdh-5E**	**Ec-Mdh-5Q**
Mutated sites	wt	I12V, R81A, M85E, G179D	I12V, R81A, M85Q, G179D	I12V, R81A, M85E, D86S, G179D	I12V, R81A, M85Q, D86S, G179D
Oxaloacetate					
k_{cat} (s^{-1})	91.4 (±35.4)	6.7 (±0.1)	5.7 (±0.2)	8.5 (±0.5)	8.6 (±3.7)
K_m (mM)	0.01 (±0.004)	0.6 (±0.02)	0.3 (±0.09)	0.1 (±0.01)	0.1 (±0.04)
K_i (mM)	1.8 (±0.8)	18.8 (±2.5)	4.5 (±0.1)	0.9 (±0.03)	0.4 (±0.2)
k_{cat}/K_m (s^{-1} mM^{-1})	6,387 (±743)	12 (±0.6)	21 (±7)	84 (±14)	127 (±12)
OHB					
k_{cat} (s^{-1})	0.03 (±0.004)**	47.1 (±0.1)	55.4 (±2.3)	65.0 (±14.9)	59.4 (±0.1)
K_m (mM)	-	4.6 (±0.1)	5.9 (±0.7)	2.2 (±0.9)	1.6 (±0.01)
K_i (mM)	-	-	37.1 (±6.7)	31.8 (±16.5)	31.9 (±5.6)
k_{cat}/K_m (s^{-1} mM^{-1})	-	10 (±0.3)	10 (±0.7)	31 (±6)	37 (±0.3)
Pyruvate					
k_{cat} (s^{-1})	0.02 (±0.006)	3.8 (±0.7)	3.6 (±0.2)	6.6 (±0.1)	11.0 (±1.2)
K_m (mM)	46.3 (±13.3)	41.3 (±15.4)	29.7 (±1.5)	16.4 (±0.9)	28.6 (±2.8)
K_i (mM)	-	-	-	-	98.3 (±29.0)
k_{cat}/K_m (s^{-1} mM^{-1})	0.0005 (±0.00001)	0.1 (±0.02)	0.1 (±0.01)	0.4 (±0.02)	0.4 (±0.004)
*Specificity**					
$(k_{cat}/K_m)_{OHB}$ / $(k_{cat}/K_m)_{Oxaloacetate}$	-	0.87	0.44	0.36	0.29
$(k_{cat}/K_m)_{OHB}$ / $(k_{cat}/K_m)_{Pyruvate}$	-	105.78	77.34	76.03	97.30

* Specificity corresponds to the ratio of catalytic efficiencies between OHB and competitive substrates (oxaloacetate, pyruvate).

** Enzyme not saturated under test assay conditions. Turnover numbers correspond to enzyme activity at 50 mM substrate concentrations.

Table 3.3. Comparison of kinetic parameters for oxidative reaction catalyzed by Ec-Mdh wild-type and best variants.

Enzyme	**Ec-Mdh**	**Ec-Mdh-4E**	**Ec-Mdh-4Q**	**Ec-Mdh-5E**	**Ec-Mdh-5Q**
Mutated sites	wt	I12V, R81A, M85E, G179D	I12V, R81A, M85Q, G179D	I12V, R81A, M85E, D86S, G179D	I12V, R81A, M85Q, D86S, G179D
(L)-malate					
k_{cat} (s^{-1})	35.6 (±3.9)	1.7 (±0.08)*	28.2 (±6.5)	32.0 (±2.4)	34.7 (±1.2)
K_m (mM)	0.9 (±0.2)	-	388.3 (±119.5)	90.4 (±5.2)	54.8 (±2.5)
K_i (mM)	-	-	-	-	-
K_{cat}/K_m (s^{-1} mM^{-1})	40 (±13)	-	0.07 (±0.006)	0.4 (±0.006)	0.6 (±0.05)
*(L)-DHB***					
k_{cat} (s^{-1})	0.004 (±0.0003)*	5.5 (±0.6)*	41.2 (±11.5)	41.8 (±0.1)	39.1 (±0.6)
K_m (mM)	-	-	212.1 (±74.2)	113.6 (±5.7)	64.1 (±5.5)
K_i (mM)	-	-	-	-	-
K_{cat}/K_m (s^{-1} mM^{-1})	-	-	0.2 (±0.01)	0.4 (±0.02)	0.6 (±0.04)
(L)-lactate					
k_{cat} (s^{-1})	n.d. ***	0.2 (±0.02)	0.3 (±0.01)	0.7 (±0.03)	1.2 (±0.02)
K_m (mM)	-	-	-	-	-
K_i (mM)	-	-	-	-	-
K_{cat}/K_m (s^{-1} mM^{-1})	-	-	-	-	-
*Specificity****					
$(k_{cat}/K_m)_{DHB} / (k_{cat}/K_m)_{Malate}$	-	-	1.34	0.52	0.48
$(k_{cat}/K_m)_{DHB} / (k_{cat}/K_m)_{Lactate}$	-	-	-	-	-

* Enzyme not saturated under test assay conditions. Turnover numbers correspond to enzyme activity at substrate concentrations of 50 mM (L-malate, L-lactate, L-DHB).

** Not determined.

** Specificity corresponds to the ratio of catalytic efficiencies between (L)-DHB and competitive substrates (L-malate, L-lactate).

3.5 Optimization of DHB production using site-directed variants

In a previous study, Walther and colleagues[2] have demonstrated the *E. coli* parent strain ECO4 (MG1655 Δ*adhE* Δ*ldhA* Δ*thrB* Δ*metA*) to functionally express the homoserine DHB pathway when transformed with the medium-copy number plasmid pZA23-*thrA*$_{S345F}$-*aspC*-*ldhA*$_{Q85C}$, that encodes for the threonine-insensitive bifunctional aspartate kinase/homoserine dehydrogenase mutant ThrA S345F, the homoserine (HMS) transaminase Ec-AspC and the OHB reductase Ll-LdhA Q85C. The resulting strain ECO18 was able to produce 3.7 mM DHB after 24 h of culture in M9 mineral medium supplemented with 20 g L^{-1} glucose[9]. To improve DHB production, the increase in the availability of homoserine pathway precursors was first attempted. Since overexpression of the aspartate/malate insensitive phosphoenol pyruvate (PEP) carboxylase mutant Ec-Ppc K620S has previously been shown to increase the intracellular concentration of oxaloacetate[9,24], the corresponding gene was cloned downstream of the DHB operon and the obtained pECO2-ppc* plasmid was transformed into the ECO4 parent strain. After 24 h of culture, the resulting strain CF216 led to slightly improved DHB production (3.9 mM), demonstrating a moderate positive effect of the increase in PEP carboxylase activity (**Table 3.4**). Further investigations were therefore carried out using the pECO2-ppc* plasmid as a backbone for further modification.

The identified four highly active Ec-Mdh variants on OHB (Ec-Mdh-4E, Ec-Mdh-4Q, Ec-Mdh-5E and Ec-Mdh-5Q) were employed as OHB reductases in the DHB operon, leading to the generation of strains CF132-135 (**Table 3.4**). In all cases, DHB titers and productivities were improved (by up to two- and three-fold, respectively) in comparison to the use of Ll-Ldh Q85C (strain CF216), demonstrating the superior performance of the characterized Ec-Mdh variants. The DHB production levels were similar for all the tested Ec-Mdh variants, despite their previously described distinctive kinetic properties. The highest DHB titers were obtained from the expression of either Ec-Mdh-5E (CF134, 7.9 mM) or Ec-Mdh-5Q (CF135, 7.6 mM) with similar product yields (0.10 mol mol^{-1} glucose). The OHB reductase Ec-Mdh-5Q expressed from the pECO2-ppc*(Ec-Mdh-5Q) plasmid was therefore selected for further studies.

Table 3.4. DHB titer (mM) and yield (mol mol^{-1} glucose) after 24h culture of *E. coli* engineered strains. Cells were cultivated in 250 mL baffled shake flasks on mineral medium containing 20 g L^{-1} glucose. The error bars represent the standard deviation from the mean of two biological replicate experiments. All strains are derived from the parent strain ECO4 (*E. coli* MG1655 Δ*adhE* Δ*ldhA* Δ*thrB* Δ*metA*). All plasmids are derived from the pZA23 medium-copy number plasmid (Expressys, Germany). The vector pZA23::Ec.*thrA*$_{S345F}$::2::3::4 expresses, by this order, the genes encoding the enzymes 1-4, in which 1: Ec-ThrA S345F, 2: homoserine transaminase, 3: OHB reductase, 4: phosphoenol pyruvate carboxylase.

Strain	Plasmid architecture: pZA23-Ec.*thrA*$_{S345F}$-2-3-4			DHB titer at 24h (mM)	DHB yield at 24h (mol mol^{-1})	Ref.
	Gene 2	Gene 3	Gene 4			
ECO18	Ec-AspC	Ll-Ldh Q85C	-	3.7 (±0.2)	-	Walther et al, 2017[2]
CF216	Ec-AspC	Ll-Ldh Q85C	Ec-Ppc K620S	3.9 (±0.1)	0.04 (±0.0001)	This work
CF132	Ec-AspC	Ec-Mdh-4E[1]	Ec-Ppc K620S	6.9 (±0.8)	0.09 (±0.02)	This work
CF133	Ec-AspC	Ec-Mdh-4Q[2]	Ec-Ppc K620S	7.1 (±0.7)	0.09 (±0.01)	This work
CF134	Ec-AspC	Ec-Mdh-5E[3]	Ec-Ppc K620S	7.9 (±0.05)	0.10 (±0.01)	This work
CF135	Ec-AspC	Ec-Mdh-5Q[4]	Ec-Ppc K620S	7.6 (±0.05)	0.10 (±0.01)	This work
CF217	Ec-AlaC A142P:Y275D	Ec-Mdh-5Q[4]	Ec-Ppc K620S	19.8 (±1.7)	0.19 (±0.01)	This work
CF268	Ec-AlaC A142P:Y275D	Ll-Ldh Q85C	Ec-Ppc K620S	9.7 (±0.1)	0.15 (±0.02)	This work

In an ongoing effort to achieve higher DHB titers and productivities, also the replacement of the currently used HMS transaminase Ec-AspC was envisaged, since it has been shown to be a particularly inefficient enzyme when (L)-homoserine is used as the substrate[2]. In this regard, Bouzon and co-workers[25] have discovered a new HMS transaminase in a recent study of *E. coli* one-carbon metabolism following long-term cell cultivation. The newly identified HMS transaminase is a mutant variant of the alanine aminotransferase from *E. coli* (Ec-AlaC) with substitutions at positions Ala142 and Tyr275. The resulting variant Ec-AlaC A142P:Y275D displayed a high affinity for homoserine (K_m, 1.7 mM), in contrast to Ec-AspC that was not saturated at homoserine concentrations of up to 50 mM[2,25]. As such, Ec-AspC was replaced by the HMS transaminase Ec-AlaC A142P:Y275D variant in the pECO2-ppc*(Ec-Mdh-5Q) plasmid. When the vector obtained was expressed in the parent strain, resulting in the strain CF217, increased levels of DHB accumulation (19.8 mM) in higher yields (0.19 mol mol^{-1}) were observed, suggesting that the homoserine transamination step is rate-limiting (**Table 3.4**). To further confirm this, the improved Ec-AlaC variant was expressed together with the Ll-Ldh Q85C OHB reductase variant (strain CF268). After 24 h of cell cultivation, DHB production more than doubled (to 9.7 mM) compared to the CF216 strain expressing Ec-AspC and Ll-Ldh Q85C. Overall, our results show that redirection of the metabolic flux towards homoserine when combined with the improved (Ec-AlaC A142P:Y275D) HMS transaminase and (Ec-Mdh-5Q) OHB reductase variants led to an up to five-fold improvement in the DHB titer and productivity as compared to the original ECO18 strain.

3.6 Fed-batch cultivation of DHB-producing strains

The best DHB-producing strain CF217 (which expresses Ec-Mdh-5Q and Ec-AlaC A142P:Y275D enzymes) was further studied in a 1.5 L medium-containing fed-batch reactor (**Figure 3.8a**). Under fully aerobic conditions and non-limiting glucose concentrations, 89.0 mM DHB were produced from glucose after 48h of culture. Considering this period, DHB was produced at a rate of 1.85 mmol L^{-1} h^{-1} with a yield equal to 0.18 mol mol^{-1}. Accumulation of DHB was accompanied by the production of lactate and acetate (up to 21.3 mM and 20.8 mM, respectively). To further confirm the role of HMS transaminase as main rate-limiting step towards DHB biosynthesis, the strain

CF135 which expresses both Ec-AspC and Ec-Mdh-5Q enzymes was also cultivated under similar conditions (**Figure 3.8b**). Lower DHB production was observed (up to 41.5 mM) with a yield of 0.09 mol mol^{-1} when compared to the strain CF217. Also the extracellular accumulation of the by-products acetate and lactate was kept at low levels.

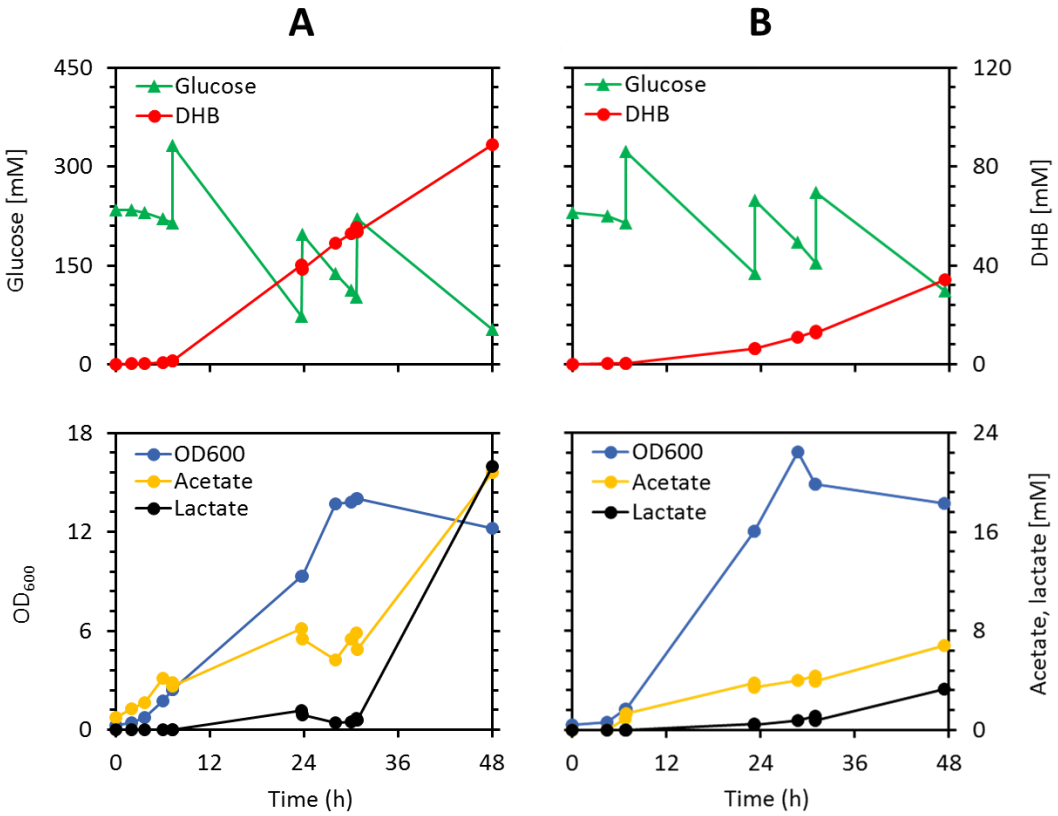

Figure 3.8. Fed-batch cultivation of the DHB-producing strains CF217 **(A)** and CF135 **(B)**. Cells were cultivated on 1.5 L defined mineral medium (pH 7.0) which was supplemented with 1 g L^{-1} threonine and methionine to complement strain auxotrophies. Glucose was manually added to assure non-limiting carbon source concentrations (below 10 g L^{-1}) during cell cultivation. Dissolved oxygen concentration was maintained above 40 % of saturating concentration to assure aerobic conditions. The results are derived from a single experiment.

A comprehensive analysis of the extracellular accumulation of natural amino acids and homoserine DHB precursor was performed at the end of cultivation of each strain to understand in more detail cell metabolism during DHB production. While in both cases homoserine was released at relevant concentrations (> 28 mM), the utilization of Ec-AlaC A142P:Y275D yielded 1.4-fold higher accumulation of the referred DHB precursor (**Figures 3.9a-b**). However, the observed titers of alanine (107.6 mM) suggest a pronounced substrate promiscuity of the enzyme (**Figure 3.9a**). The accumulation of

glutamate (> 17.8 mM) was found to occur upon cultivation of both strains (**Figures 3.9a-b**).

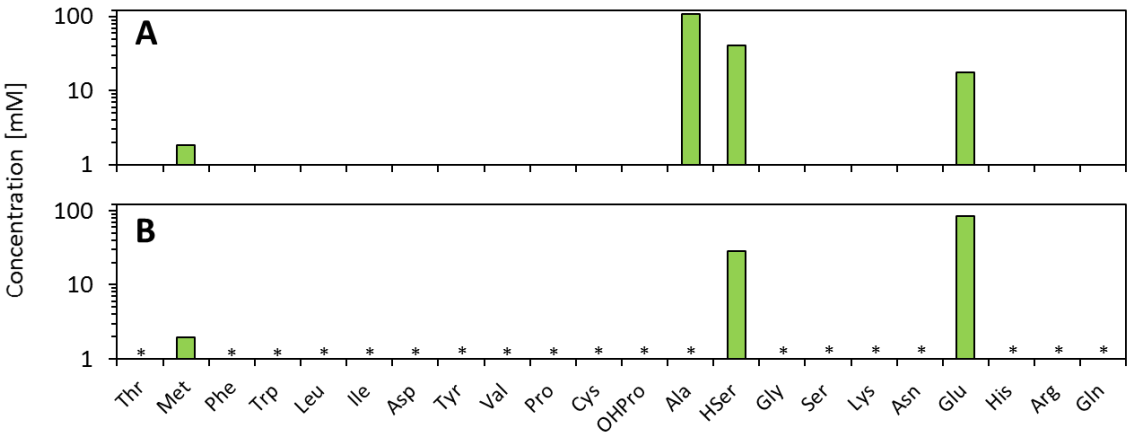

Figure 3.9. Amino acid analysis after 48h of fed-batch cultivation of the DHB-producing strains CF217 **(A)** and CF135 **(B)**. Only those amino acids with concentrations exceeding 1 mM are shown. Cells were cultivated on 1.5 L defined mineral medium (pH 7.0) which was supplemented with 1 g L^{-1} threonine and methionine to complement strain auxotrophies. Glucose was manually added to assure non-limiting carbon source concentrations (below 10 g L^{-1}) during cell cultivation. Dissolved oxygen concentration was maintained above 40 % of saturating concentration to assure aerobic conditions. The results are derived from a single experiment. Legend: OHPro – hydroxyproline, HSer – homoserine, * - not measured.

3.7 Discussion

A new metabolic engineering approach, based on the construction of *de novo* synthetic pathways constituted of novel reactions, has recently emerged enabling the production of non-natural compounds. This new concept however requires the discovery of novel enzyme functions by extensive data mining and/or protein engineering to generate new activities in the absence of naturally available biocatalysts[1,26]. In the development of an artificial route for the production of the non-natural compound DHB from glucose via homoserine, the development of highly-efficient HMS transaminase and OHB reductase enzymes has previously been identified as a major lever to boost the performance of the designed DHB pathway. In this work, the generation of a more efficient OHB reductase is described which was obtained by the step-wise rational engineering of a template enzyme homologue acting on a sterically similar cognate substrate.

Rational protein engineering is a well-established technique for the generation of novel enzyme variants with enhanced activity, stability or altered substrate specificity[1,27,28]. Based on the sequence, structural and functional information about target enzymes, amino acid changes likely to confer the desired enzymatic properties can be predicted *in silico*. The selection of an appropriate template enzyme to be redesigned is crucial to the experimental success of targeted mutagenesis[29]. In this study, Ec-Mdh served as the enzyme to be engineered for OHB reductase activity. Template selection was made on the basis that Ldh enzymes have been previously shown to convert the monocarboxylate OHB to DHB, and that Mdh and Ldh enzymes possess highly similar structures and a shared catalytic mechanism[8,30]. A key exploitable difference between Mdh and Ldh enzymes resides in their strict respective selectivity for di- and monocarboxylic acid substrates[16,18]. However, previous attempts to alter the substrate specificity of Ec-Mdh from the dicarboxylate oxaloacetate to the monocarboxylate pyruvate were met with only partial success[31]. While the engineered Ec-Mdh mutant enzymes reported in the literature exhibited reversed substrate specificity, they were significantly less active than the wild-type Ldh enzyme towards the natural substrate pyruvate, illustrating the stringent substrate specificity of Ec-Mdh[32,33].

Based on comparative structural and sequence analysis combined with molecular modelling, target mutation sites in Ec-Mdh for the introduction of catalytic activity towards the monocarboxylic acid OHB were identified. The predicted amino acid residue exchange sites were located in distinct functional regions: the substrate binding site, the coenzyme nicotinamide ring binding pocket and at a structural contact region on the external surface of the mobile loop covering the active-site during catalysis. Mutations were introduced into wild-type Ec-Mdh in a step-wise manner. Experimental delineation of the observed increases in OHB reductase activity permitted to recombine mutations in different zones of the template protein active-site, leading to the identification of mutant combinations with better optimized kinetic properties. Favourable synergistic effects arising from the combination of correlated mutations have been documented in other studies[34]. In total, 17 Ec-Mdh variants were generated and the most promising amino acid changes bestowing OHB reductase activity were identified as (I12V, R81A, M85E/Q, G179D and/or D86S). Kinetic evaluation of the four best Ec-Mdh mutants revealed an improvement of up to three orders of magnitude in both the OHB reductase activity and

the Michaelis constant for OHB with respect to the wild-type enzyme. However, all four mutants displayed higher catalytic efficiencies for the natural substrate oxaloacetate than for OHB.

The feasibility of the most promising Ec-Mdh mutants for *in vivo* applications was further evaluated by individual expression in the homoserine DHB pathway. All tested enzyme variants provided for approximately two-fold higher DHB production, as compared to the Ll-LdhA Q85C OHB reductase variant. Further improvements were obtained by channeling the carbon flux towards homoserine (Ec-Ppc K620S) and expression of the HMS transaminase variant Ec-AlaC A142P:Y275D in the DHB operon resulting in DHB titers up to 19.8 mM. Fed-batch cultivation under well-controlled conditions of the best DHB-producing strain (CF217) as assessed in cell culture at shake flask level further resulted in extracellular accumulation of DHB to levels up to 89.0 mM (or 10.7 g L^{-1}), corresponding to the highest titer observed to date. The extracellular accumulation of high amounts of the homoserine DHB precursor suggests however that the currently used HMS transaminase is a major rate-limiting step of the pathway. Since the intracellular accumulation of homoserine has previously been shown to inhibit its own biosynthesis[35], only a HMS transaminase with high catalytic efficiency could convert homoserine to OHB without substrate accumulation. The release of such amino acid to the medium may therefore be seen as a strategy of cell metabolism towards the continuous production of DHB. Additionally, the presence of alanine in the culture broth further confirms the substrate promiscuity of the Ec-AlaC A142P:Y275D enzyme, reason for which HMS transaminase may need to be further engineered if industrially relevant DHB titers are desired. The accumulation of glutamate may similarly be a consequence of the 2-ketoglutarate dependent HMS transaminase activity which yields homoserine and glutamate at stoichiometrically equal amounts, or alternatively from the oxidation of the 2-ketoglurate metabolite derived from the Krebs cycle and catalyzed by a glutamate dehydrogenase enzyme. On the other side, other reasons may be behind the presence of lactate and acetate, mainly observed at a late-stage of fermentation in which cells were in a stationary phase. Being both metabolites generated from pyruvate inside cells, the accumulation of lactate was somehow unexpected due to the deletion of (L)-lactate dehydrogenase encoding-gene *ldhA* and since the other two lactate dehydrogenases from *E. coli* (LlDd and Dld) have a marked preference for lactate oxidation. The observed

lactate accumulation may therefore be linked to the observed residual activity of Ec-Mdh-5Q on pyruvate. On the other side, the production of acetate may occur as a metabolic strategy towards the regeneration of intracellular ATP, if it proceeds via acetyl-CoA.

In summary, the work outlined here shows production of DHB via homoserine up to 89.0 mM at a yield of 0.18 mol mol^{-1} glucose. Further enzyme engineering to increase the efficiency of the synthetic pathway is necessary to solve the observed pathway imbalances and remove metabolic inefficiencies. Pathway optimization through combination of modeling technologies with omics data measurements can identify and address additional bottlenecks impeding higher product formation.

3.8 Materials and methods

3.8.1 Chemicals and reagents

All chemicals and solvents were purchased from Sigma-Aldrich unless otherwise stated. Restriction endonucleases and DNA-modifying enzymes were purchased from New England Biolabs and used according to manufacturer's instructions. DNA plasmid isolation was performed using GeneJET Plasmid Miniprep Kit (Thermo Scientific). DNA extraction from agarose gel was carried out using the GeneJET Gel Extraction Kit (Thermo Scientific). DNA sequencing was carried out by Beckman Coulter Genomics (Takeley, United Kingdom) or Eurofins SAS (Ebersberg, Germany).

3.8.2 Protein cloning, mutagenesis, expression and purification

E. coli DH5α (New England Biolabs) was routinely used for construction of plasmids. The *mdh* gene encoding the wild-type Ec-Mdh was amplified by PCR (primers listed in **Table 3.5**) and cloned into the corresponding sites of pET28a (Novagen) using T4 DNA ligase (Biolabs), thereby adding an N-terminal hexa-His tag. Point mutations were introduced on pET28-derived plasmid by inverse PCR using the primer pairs listed in **Table 3.6**. Resulting products were digested by *Dpn*I to remove template DNA and transformed into competent cells. Mutated plasmids were verified by sequencing.

Table 3.5. Primers and restriction enzymes used to clone genes into pET28a expression vector

Gene	Primer sequences (5' – 3')	Restriction enzymes	Resulting vector
Ec-*mdh*	TATAATCATATGAAAGTCGCAGTCCTC TATAATGGATCCTTACTTATTAACGAACTC	NdeI BamHI	pET28-Ec-*mdh*

Table 3.6. Primers used for site-directed mutagenesis in pET28a-Ec-*mdh*

Mutation	Primer sequences (5' – 3')	Restr. site
R81nnk	TTATCTCTGCAGGCGTAGCGNNKAAACCCGGGATGGATCGTTC GAACGATCCATCCCGGGTTTMNNCGCTACGCCTGCAGAGATAA	SmaI
R81A M85E	TTATCTCTGCAGGCGTAGCGGCTAAACCGGGTGAGGATCGTTCCGACCTG CAGGTCGGAACGATCCTCACCCGGTTTAGCCGCTACGCCTGCAGAGATAA	none
R81A M85Q	TTATCTCTGCAGGCGTAGCGGCTAAACCGGGTCAGGATCGTTCCGACCTG CAGGTCGGAACGATCCTGACCCGGTTTAGCCGCTACGCCTGCAGAGATAA	none
I12V	GTCGCAGTCCTCGGCGCCGCTGGCGGTGTCGGCCAGGCGCTTGCAC GTGCAAGCGCCTGGCCGACACCGCCAGCGGCGCCGAGGACTGCGAC	NarI
G179D	CCGGTTATTGGCGGCCACTCTGATGTTACCATTCTGCCGCTGCTG CAGCAGCGGCAGAATGGTAACATCAGAGTGGCCGCCAATAACCGG	EaeI
R81A D86S	GGCGTAGCGGCTAAACCGGGTATGTCTCGTTCCGACCTG CAGGTCGGAACGAGACATACCCGGTTTAGCCGCTACGCC	none
V214T	ATCCAGAACGCGGGTACCGAAGTGACTGAAGCGAAGGCCGGT ACCGGCCTTCGCTTCAGTCACTTCGGTACCCGCGTTCTGGAT	KpnI
T211Y	ACGGATCCAGAACGCCGGCTATGAAGTGGTTGAAGCG CGCTTCAACCACTTCATAGCCGGCGTTCTGGATCCGT	NaeI
R81A M85E D86S	GCGGCCAAACCGGGTGAGTCTCGTTCCGACCTGTTTAACG CGTTAAACAGGTCGGAACGAGACTCACCCGGTTTGGCCGC	HaeIII
R81A M85Q D86S	GCGGCTAAACCGGGCCAGTCTCGTTCCGACCTGTTTAACG CGTTAAACAGGTCGGAACGAGACTGGCCCGGTTTAGCCGC	HaeIII

Enzymes were expressed in *E. coli* BL21(DE3) cells (New England Biolabs) in 200 mL Luria-Bertani (LB) medium supplemented with 50 µg mL^{-1} kanamycin (37 ºC, 200 rpm) that were inoculated from an overnight culture at OD_{600} of 0.05 and grown to OD_{600} of 0.6 before protein expression was induced for 3 h by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) to the culture medium. Cells were harvested by centrifugation (15 min at 4,000 rpm, 4 ºC) and pellets stored at -20 ºC until further analysis. Protein purification starting from frozen cell pellets was performed as described elsewhere[9].

3.8.3 Enzymatic assays

Protein concentrations were determined prior to enzymatic assays by the method of Bradford (Bio-Rad). All enzyme assays were carried out at 37 ºC in 96-well flat-bottomed microtiter plates in a final volume of 250 µl. The reactions were followed by the characteristic absorption of NADH at 340 nm in a microplate reader (Epoch 2, BioTek).

Enzyme activities of wild-type Ec-Mdh and variants were assayed in both senses of the reaction. Assays in the reductive (biosynthetic) direction were carried out by monitoring NADH oxidation during the reduction of 2-keto acids. The assay mixture contained 60 mM Hepes (pH 7), 0.25 mM NADH, 5 mM $MgCl_2$, 50 mM KCl and appropriate amounts of purified enzyme or crude extract. Reactions were started by adding variable concentrations of oxaloacetate, OHB or pyruvate. Since OHB is not commercially available, it was produced *in house* by the action of L-amino acid oxidase in a single-step reaction starting from (L)-homoserine, and quantified by a ketone calibration curve[36]. Assays in the oxidative direction were carried out by following the reduction of NAD^+ during the oxidation of 2-hydroxy acids. The assay mixture contained 179 mM glycine buffer (pH 9), 10 mM NAD^+, 5 mM $MgCl_2$, 50 mM KCl and appropriate amounts of purified enzyme or crude extract. Reactions were started by adding variable concentrations of (L)-malate, DHB or (L)-lactate. (D/L)-DHB consisted of a racemic mixture of stereoisomers kindly provided by Adisseo SAS (France). Values of k_{cat} and K_m were estimated by fitting kinetic data from at least five different substrate concentrations with SigmaPlot v12.0 following non-linear regression of Michaelis-Menten equation, unless uncompetitive substrate inhibition was observed.

3.8.4 Construction of pECO2ppc* (Ec-mdh-X)

The genes coding for the best OHB decarboxylase variants (Ec-mdh-X, in which X determines the mutant) were PCR amplified from pET28A-derived vectors using primer pairs 1707/cf115 (see **Table 3.7**). The primers introduced unique restriction sites flanking the gene of interest, and additionally forward primers inserted a ribosome binding sequence (RBS) immediately upstream coding sequence. PCR products and vector eco2ppc-Ll-ldh were digested with *Not*I and *Bam*HI restriction enzymes and ligated using T4 DNA ligase (New England Biolabs). The resulting constructions were transformed

into NEB 5-α chemically competent *E. coli* cells (NEB) and verified by DNA sequencing to contain the correct insert. Plasmids were then transformed in strain of interest.

Table 3.7. Primers used to construct plasmids expressing the biosynthetic DHB pathway

Primer	Sequence (5' – 3')
1469	CTCTCATGGAAGTTAGGAGTCTGACCCGGGGTTTAACTTTAAGAAGGAG ATATACCATGTTTGAGAACATTACCGCCGCTC
1470	CCTACAAGGATAACTTTTTTACGTTGTTTATCAGCCATGGTATATCTCCTT CTTAAAGTTAAACGCGGCCGCTTACAGCACTGCCACAATCGCTTCGC
1471	GCGGCCGCGTTTAACTTTAAGAAGGAGATATACCATGGCTGATAAACAA CGTAAAAAAGTTATCCTTGTAGG
1472	CGCAATGCGGAATATTGTTCGTTCATGGTACCGAGCTCGAATTCTGTTTC CTGTTCTAGATTAGTTTTTAACTGCAGAAGCAAATTCTTC
1473	TCTAGAACAGGAAACAGAATTCGAGCTCGGTACCATGAACGAACAATAT TCCGCATTGCG
1474	GATGCCTCTAGCACGCGTACCATCCCGGGTTAGCCGGTATTACGCATACC TGCCG
1707(mdh_xbaI_rv)	ATAAT**TCTAGA**TTACTTATTAACGAACTCTTCGCCCAGGGC
cf115 (rbspet_mdh*_NotIF)	ATTAAG**CGGCCGCG**TTTAACTTTAAGAAGGAGATATACCATGAAAGTCG CAGTCCTCGGCGCCGCTGGCGGTGTCGGCCAGGCGCTTGCACTACT

3.8.5 Shake flask cultures for DHB production

All cell cultivation was carried out at 37 ºC on a rotary shaker (Infors HT, France) running at 200 rpm. Pre-cultures were grown overnight in 10 mL of M9 mineral medium supplemented with methionine and threonine (0.2 g L^{-1}) in 50 mL falcon tubes. The biomass needed to start main cultures with a starting OD_{600} of 0.2 was transferred to 250 mL baffled-shake flasks containing 25 mL of M9 mineral medium supplemented with methionine and threonine, and IPTG was added at a concentration of 0.5 mM when OD_{600} reached ~0.6. The antibiotic kanamycin sulphate was added when required at 50 mg L^{-1}. One liter of M9 mineral medium contained: 20 g glucose, 18 g $Na_2HPO_4*12H_2O$, 3 g KH_2PO_4, 0.5 g NaCl, 2 g NH_4Cl, 0.5 g $MgSO_4*7H2O$, 0.015 $CaCl_2*2H_2O$, 1 ml of 0.06 M $FeCl_3$ stock solution prepared in 100 times diluted concentrated HCl, 2 ml of 10 mM thiamine HCl stock solution, 20 g MOPS, and 1 ml of trace element solution (containing per liter: 0.04 g $Na_2EDTA*2H_2O$, 0.18 g $CoCl_2*6H_2O$, $ZnSO_4*7H_2O$, 0.04 g $Na_2MoO_4*2H_2O$, 0.01 g H_3BO_3, 0.12 g $MnSO_4*H_2O$, 0.12 g $CuCl_2*H_2O$). The pH was adjusted to 7, and the medium filter-sterilized.

3.8.6 Fed-batch bioreactor cultures for DHB production

The pre-cultures for the inoculation of the bioreactors were cultivated in 1 L shake flasks containing 150 mL mineral medium (composition as previously indicated) until exponential phase. Cells were then harvested and used to inoculate a 2 L bioreactor (Biostat B Sartorius) that initially contained 1.5 L medium with an OD_{600} of ~0.3. The composition of the fermentation medium was similar to the mineral medium used in the shake flask experiments (see **subsection 3.8.5**), with the exception that it additionally contained 6 g L^{-1} $(NH_4)_2HPO_4$, 0.4 g L^{-1} (NH4)2SO4, 1 g L^{-1} threonine and methionine, 6 g L^{-1} citrate and no MOPS. Initial glucose concentration was 45 g L^{-1}. Glucose concentration was monitored during the fermentation and a concentrated glucose stock solution (500 g L^{-1}) was manually added to assure non-limiting glucose concentrations during the first 45 h of cultivation. The pH of the cultures was kept at 7.0 by the addition of 5 M KOH, and reactors were aerated with air at 0.3–1.5 vvm. Dissolved oxygen tension was maintained above 30 % of the saturating oxygen concentration by adjusting the appropriate agitation speed (300–1500 rpm, Rushton rotor, 28 mm diameter) and aeration rate. Polypropylene glycol (P2000) was used as antifoaming agent during the culture.

3.8.7 Analytical methods

All samples were centrifuged (2 min at 13,000 rpm) and syringe-filtered (0.2 µm), and the resulting supernatant stored at -20 ºC before analysis. A standard calibration curve was obtained by injecting standards and used for all compound analysis.

Extracellular concentrations of glucose, organic acids and DHB were determined on a Dionex Ultimate 3,000 HPLC system (Thermo Scientific, France) equipped with a RI detector (RID-10A, Shimadzu, Japan) and UV/Vis detector (SPD-20A, Shimadzu). The sample injection volume was 20 µL, and the compounds were separated on a Rezex RoA-organic acid H^+ (8%) resin-based column preceded by a SecurityGuard guard cartridge (Phenomenex, USA). The separation was performed at 80 ºC with 0.5 mM H_2SO_4 at 0.5 mL min^{-1} as mobile phase.

The analysis of homoserine and amino acids in culture broth was performed on an Agilent Infinity 1290 LC system coupled to an Agilent 6490 Triple Quadrupole LC/MS equipped

with Agilent Jet Stream and ion funnel technology. The chromatographic separation for amino acids was carried out on an InfinityLab Poroshell 120 HILIC-Z column (Agilent, France), which was maintained at a temperature of 30 °C, and the sample volume injected was 5 µL. Chromatographic separation was performed at a flow rate of 0.6 mL min^{-1} using a gradient with solvent A (10% 200 mM ammonium formate at pH 3 (formic acid) + 90% water) and solvent B (10% 200 mM ammonium formate at pH 3 (formic acid) + 90% acetonitrile). The mass spectrometer was operated in AJS ESI-Positive mode, capillary voltage 4.00 kV, nozzle voltage 1.00 kV, desolvation temperature 400°C, source temperature 250°C, cone gas flow 17 L min^{-1}, desolvation gas flow was 12 L min^{-1}, nebulizer 40 psi, iFunnel parameters (high-pressure and low‑pressure RF) were set to 140 and 60, respectively. Collision energy, iFunnel parameters, and MRM transitions for the 20 amino acids and nozzle voltage were optimized for each amino acid using MassHunter optimizer software. Dwell time was set to 45 ms, and delta EMV(+) to 200 V. Agilent MassHunter quantitative analysis software was used for data analysis.

3.8.8 Computational methods

Sequence alignments: Two sets of 1000 amino acid sequences were obtained by pair-wise BLAST searching of non-redundant data banks against Ec-Mdh (UniProtKB P61889) and *Bacillus stearothermophilus* lactate dehydrogenase (UniProtKB code P00344) sequences. The two sequence sets were multiply aligned using MUSCLE[37]. Pfam PF00056 (Ldh_1_N) and PF02866 (Ldh_1_C) domain database alignments[38], respectively corresponding to Ec-Mdh residue ranges Met1 through Gly145 and Thr147 through Val310, were used to explore position-dependent amino acid residue variation in the combined Ldh/Mdh superfamily. Residue frequencies and Shannon information entropy measures of residue variability, corrected for normalised frequencies of residue type occurrence in natural proteins, at selected corresponding positions in Ec-Mdh were calculated from the multiple sequence alignment data. The relative Shannon entropy (H_X) is calculated at each residue alignment position (X) as

$$H_X = \frac{\sum_{i=1}^{20} p_{(i|X)} \ln p_{(i|X)}}{\sum_{i=1}^{20} p_{(i)} \ln p_{(i)}}$$

where $p_{(i/X)}$ is the conditional probability of residue type *(i)* occurrence at the alignment position (X), and p(*i*) is the probability of residue type (*i*) occurrence at any position. To minimize sampling bias, p(*i*) was taken as the globally normalized residue type probability values for all natural proteins tabulated by Ranganathan and colleagues[39]. Values of H_X vary continuously from zero, corresponding to a fully conserved residue position, to unity for a residue position exhibiting no intrinsic residue type preference.

Structure alignments: Crystal structures of L-lactate dehydrogenases from *Bidifobacterium longum* (PDB code 1lth), *Bacillus stearothermophilus* (PDB code 1ldn), *Thermus thermophilus* (PDB code 2v7p), dogfish (PDB code 3ldh), human heart (PDB code 1i0z), and porcine muscle (PDB code 9ldt) were aligned with the binary complex of Ec-Mdh/NAD^+ (PDB code 1ib6). Overlays were obtained by pair-wise structural superposition at all aligned protein chain residue positions with Ec-Mdh subject to a 2Å cut-off separation distance between aligned C^α atom centres.

Molecular modelling: Chains A/B of the X-ray structure of binary complex Ec-Mdh R153C mutant with NAD^+ (PDB code 1ib6) were used as a template dimeric structure for mutant enzyme modelling and OHB ligand docking studies. The mobile active-site loop in this structure is stabilized in the active form by the presence of two SO_4^{2-} ions. The Arg153 side-chain was manually rebuilt in the same fully extended conformation observed in experimental Mdh and Ldh structures. Arg81 was truncated to alanine, and the resulting R81A mutant enzyme complex with NAD^+ was energy minimized with harmonic constraints placed on all heavy atom positions. Minimization was carried out using the ff99SB Amber molecular mechanics force field variant for protein atoms[40], and the GAFF force field[41] for the co-enzyme. The electrostatic model comprised a distance-dependent dielectric constant with $\varepsilon = 4$. Partial atomic charges for NAD^+ were abstracted from the CHARMM27 residue topology file entry[42]. The (A317) sulphate ion stabilizing the mobile loop on the solvent-exposed surface was excised prior to the introduction of mutations in the loop and substrate binding-site. Residue modification was carried out using the interactive Richardson penultimate backbone-dependent side-chain rotamer library[43] search facility and residue mutation functionality in the COOT molecular graphics and modelling package[44]. Mutant structures were re-minimized without constraints on atoms in modified and surrounding residue positions, with harmonic constraints placed on all other heavy atom positions.

Substrate docking: The (A316) sulphate ion and two water molecules (A344, A353) were removed from the active site of energy minimized binary complexes of the mutant enzymes before docking of OHB into the A subunit in the presence of bound NAD^+. Docking was carried out using *in-house* HOMER software and the PANENERGY pair-wise (*non-local*) protein-ligand atomic statistical potential[45] without the application of 1-3 covalent atom connectivity energy scaling. Docked solutions from 200 simulated annealing runs were clustered, and a Boltzmann weighted average interaction energy calculated for each cluster.

3.9 References

(1) Mak, W. S., Tran, S., Marcheschi, R., Bertolani, S., Thompson, J., Baker, D., Liao, J. C., and Siegel, J. B. (2015) Integrative genomic mining for enzyme function to enable engineering of a non-natural biosynthetic pathway. *Nat. Commun. 6*, 10005.

(2) Walther, T., Calvayrac, F., Malbert, Y., Alkim, C., Dressaire, C., Cordier, H., and François, J. M. (2017) Construction of a synthetic metabolic pathway for the production of 2,4-dihydroxybutyric acid from homoserine. *Metab. Eng. 45*, 237–245.

(3) Walther, T., Cordier, H., Dressaire, C., François, J. M., and Huet, R. (2014) Method for the preparation of 2,4-dihydroxybutyrate. Patent WO/2014/009435A1.

(4) Alkim, C., Trichez, D., Cam, Y., Spina, L., François, J. M., and Walther, T. (2016) The synthetic xylulose-1 phosphate pathway increases production of glycolic acid from xylose-rich sugar mixtures. *Biotechnol. Biofuels 9*.

(5) Sun, L., Yang, F., Sun, H., Zhu, T., Li, X., Li, Y., Xu, Z., and Zhang, Y. (2016) Synthetic pathway optimization for improved 1,2,4-butanetriol production. *J. Ind. Microbiol. Biotechnol. 43*, 67–78.

(6) Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., Khandurina, J., Trawick, J. D., Osterhout, R. E., Stephen, R., Estadilla, J., Teisan, S., Schreyer, H. B., Andrae, S., Yang, T. H., Lee, S. Y., Burk, M. J., and Van Dien, S. (2011) Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nat. Chem. Biol. 7*, 445–452.

(7) Walther, T., Dressaire, C., Cordier, H., and François, J. M.. (2013) Method of production of 2,4-dihydroxybutyric acid. Patent WO/2013/160762A3.

(8) Soucaille, P., and Boisart, C. (2010) Method for the preparation of diols. Patent WO/2010/076324A1.

(9) Walther, T., Topham, C. M., Irague, R., Auriol, C., Baylac, A., Cordier, H., Dressaire, C., Lozano-Huguet, L., Tarrat, N., Martineau, N., Stodel, M., Malbert, Y., Maestracci, M., Huet, R., André, I., Remaud-Siméon, M., and François, J. M. (2017) Construction of a synthetic metabolic pathway for biosynthesis of the non-natural methionine precursor 2,4-dihydroxybutyric acid. *Nat. Commun. 8*, 15828.

(10) Explorer, M. (2016, October 14) A modified microorganism for the optimized production of 2,4-dihydroxybutyrate with enhanced 2,4-dihydroxybutyrate efflux.

(11) Hollinshead, W., He, L., and Tang, Y. J. (2014) Biofuel production: an odyssey from metabolic engineering to fermentation scale-up. *Front. Microbiol. 5*, 344.

(12) Nakamura, C. E., and Whited, G. M. (2003) Metabolic engineering for the microbial production of 1,3-propanediol. *Curr. Opin. Biotechnol. 14*, 454–459.

(13) Atsumi, S., Cann, A. F., Connor, M. R., Shen, C. R., Smith, K. M., Brynildsen, M. P., Chou, K. J. Y., Hanai, T., and Liao, J. C. (2008) Metabolic engineering of *Escherichia coli* for 1-butanol production. *Metab. Eng. 10*, 305–311.

(14) Eriksen, D. T., Lian, J., and Zhao, H. (2014) Protein design for pathway engineering. *J. Struct. Biol. 185*, 234–242.

(15) Lane, R. S., and Dekker, E. E. (1969) 2-keto-4-hydroxybutyrate. Synthesis, chemical properties, and as a substrate for lactate dehydrogenase of rabbit muscle. *Biochemistry 8*, 2958–66.

(16) Madern, D. (2002) Molecular evolution within the L-malate and L-lactate dehydrogenase super-family. *J. Mol. Evol. 54*, 825–840.

(17) Minárik, P., Tomášková, N., Kollárová, M., and Antalík, M. (2002) Malate dehydrogenases – structure and function. *Gen. Physiol. Biophys 21*, 257–265.

(18) Chapman, A. D. M., Cortés, A., Dafforn, T. R., Clarke, A. R., and Brady, R. L. (1999) Structural basis of substrate specificity in malate dehydrogenases: crystal structure of a ternary complex of porcine cytoplasmic malate dehydrogenase, α-ketomalonate and tetrahydoNAD. *J. Mol. Biol. 285*, 703–712.

(19) Zaitseva, J., Meneely, K. M., Lamb, A. L., IUCr, E., V. R., and J., B. L. (2009) Structure of *Escherichia coli* malate dehydrogenase at 1.45 Å resolution. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 65*, 866–869.

(20) Bell, J. K., Yennawar, H. P., Wright, S. K., Thompson, J. R., Viola, R. E., and Banaszak, L. J. (2001) Structural analyses of a malate dehydrogenase with a variable active site. *J. Biol. Chem. 276*, 31156–62.

(21) Chothia, C., Lesk, A. M., Tramontano, A., Levitt, M., Smith-Gill, S. J., Air, G., Sheriff, S., Padlan, E. A., Davies, D., Tulip, W. R., Colman, P. M., Spinelli, S.,

Alzari, P. M., and Poljak, R. J. (1989) Conformations of immunoglobulin hypervariable regions. *Nature 342*, 877–883.

(22) Emyanitoff, R. G., and Kelly, P. J. (1982) Kinetic characterization of mitochondrial malate dehydrogenase from *Dictyostelium discoideum*. *J. Gen. Microbiol. 128*, 1767–1771.

(23) Kelly, P. J., Kelleher, J. K., and Wright, B. E. (1979) The tricarboxylic acid cycle in Dictyostelium discoideum. Metabolite concentrations, oxygen uptake and 14C-labelled amino acid patterns. *Biochem. J. 184*, 581–588.

(24) Mokhova, O. N., Kuvaeva, T. M., Golubeva, L. I., Kolokolova, A. V., and Katashkina, J. Y. (2010) A bacterirum belonging to the genus *Pantoea* producing an L-aspartic acid or L-aspartic acid-derived metabolites and a method for producing L-aspartic acid or L-aspartic acid-derived metabolites. Patent WO/2010/038905.

(25) Bouzon, M., Perret, A., Loreau, O., Delmas, V., Perchat, N., Weissenbach, J., Taran, F., and Marlière, P. (2017) A Synthetic Alternative to Canonical One-Carbon Metabolism. *ACS Synth. Biol. 6*, 1520–1533.

(26) Erb, T. J., Jones, P. R., and Bar-Even, A. (2017) Synthetic metabolism: metabolic engineering meets enzyme design. *Curr. Opin. Chem. Biol. 37*, 56–62.

(27) Wijma, H. J., Floor, R. J., and Janssen, D. B. (2013) Structure- and sequence-analysis inspired engineering of proteins for enhanced thermostability. *Curr. Opin. Struct. Biol. 23*, 588–594.

(28) Antikainen, N. M., and Martin, S. F. (2005) Altering protein specificity: techniques and applications. *Bioorg. Med. Chem. 15*, 2701–2716.

(29) Meng, H., Liu, P., Sun, H., Cai, Z., Zhou, J., Lin, J., and Li, Y. (2016) Engineering a D-lactate dehydrogenase that can super-efficiently utilize NADPH and NADH as cofactors. *Sci. Rep. 6*, 24887.

(30) Wu, G., Fiser, A., ter Kuile, B., Sali, A., and Müller, M. (1999) Convergent evolution of *Trichomonas vaginalis* lactate dehydrogenase from malate dehydrogenase. *Proc. Natl. Acad. Sci. 96*, 6285–6290.

(31) Aslan, A. S., Birmingham, W. R., Karagüler, N. G., Turner, N. J., and Binay, B.

(2016) Semi-rational design of *Geobacillus stearothermophilus* L-lactate dehydrogenase to access various chiral α-hydroxy acids. *Appl. Biochem. Biotechnol. 179*, 474–484.

(32) Boernke, W. E., Millard, C. S., Stevens, P. W., Kakar, S. N., Stevens, F. J., and Donnelly, M. I. (1995) Stringency of substrate specificity of *Escherichia coli* malate dehydrogenase. *Arch Biochem Biophys 322*, 43–52.

(33) Yin, Y., and Kirsch, J. F. (2007) Identification of functional paralog shift mutations: conversion of *Escherichia coli* malate dehydrogenase to a lactate dehydrogenase. *Proc. Natl. Acad. Sci. 104*, 17353–17357.

(34) Hoelsch, K., Sührer, I., Heusel, M., and Weuster-Botz, D. (2013) Engineering of formate dehydrogenase: synergistic effect of mutations affecting cofactor specificity and chemical stability. *Appl. Microbiol. Biotechnol. 97*, 2473–2481.

(35) Li, H., Wang, B., Zhu, L., Cheng, S., Li, Y., Zhang, L., Ding, Z. Y., Gu, Z. H., and Shi, G. Y. (2016) Metabolic engineering of *Escherichia coli* W3110 for L-homoserine production. *Process Biochem. 51*, 1973–1983.

(36) Wellner, D., and Lichtenberg, L. A. (1971) Assay of amino acid oxidase. *Methods Enzymol. 17*, 593–596.

(37) Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res. 32*, 1792–1797.

(38) Robert D. Finn, Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Salazar, A. S.-V. G. A., Tate, J., and Bateman, A. (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res. Database 44*, 44:D279-D285.

(39) Halabi, N., Rivoire, O., Leibler, S., and Ranganathan, R. (2009) Protein Sectors: Evolutionary Units of Three-Dimensional Structure. *Cell 138*, 774–786.

(40) Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006) Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins Struct. Funct. Bioinforma. 65*, 712–725.

(41) Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general amber force field. *J. Comput. Chem. 25*,

1157–1174.

(42) Pavelites, J. J., Gao, J., Bash, P. A., and Mackerell, A. D. (1997) A molecular mechanics force field for NAD+, NADH, and the pyrophosphate groups of nucleotides. *J. Comput. Chem. 18*, 221–239.

(43) Lovell, S. C., Word, J. M., Richardson, J. S., and Richardson, D. C. (2000) The penultimate rotamer library. *Proteins 40*, 389–408.

(44) Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr. 66*, 486–501.

(45) Topham, C. M., Barbe, S., and André, I. (2016) An Atomistic Statistically Effective Energy Function for Computational Protein Design. *J. Chem. Theory Comput. 12*, 4146–4168.

Chapter 4. Construction of a synthetic pathway for production of 1,3-propanediol from glucose

4.1 Introduction

A central goal of the bioeconomy consists in reducing our dependence on petroleum by focusing on the development of efficient, sustainable and eco-friendly processes for production of chemicals and fuels[1]. 1,3-Propanediol (PDO) is an important commodity chemical that can serve as a monomer for the synthesis of industrially relevant polymers, including polyesters, polyethers and polyurethanes. Among those, the polyester polytrymethylene terephthalate (PTT) is seen as a major competitor of Nylon in carpet industries, therefore making PDO a molecule of significant industrial interest. While both acrolein hydration, and ethylene oxide hydroformylation followed by reduction of resulting aldehyde enable PDO synthesis, low product yields and harsh reaction conditions impede the successful development of a cost-competitive chemical production process[2,3].

Alternatively, PDO can be microbially produced from glycerol by *Klebsiella* and *Clostridium* species under anaerobic conditions in a two-step pathway that employs vitamin B_{12}-dependent glycerol dehydratase and PDO oxidoreductase enzymatic activities. Albeit expanding the range of substrates would render PDO production more flexible, no natural microorganisms have been found to directly convert sugars to this compound. Previous studies demonstrated the adoption of two-stage and co-fermentation processes to enable a better control of cultivation conditions and a wider flexibility in terms of substrate utilization[4–6]. Another approach is the engineering of microorganisms for the direct production of PDO from sugars via glycerol. Successful examples include engineering *K. pneumoniae* and *Saccharomyces cerevisiae* for enabling PDO biosynthesis from glucose[7]. In addition, Genencor and DuPont have commercialized a process which employs a genetically modified *E. coli* strain able to produce PDO at high titers and yields[8]. In both cases, the engineered PDO synthesis pathways are composed of necessary reaction steps linking the glycolysis-derived metabolite dihydroxyacetone phosphate (DHAP) to PDO via glycerol.

While traditional metabolic engineering resulted in great improvements in PDO biosynthesis, the rapid development of powerful genome mining and protein engineering

tools enables now the creation of *de novo* designed non-natural pathways for the production of economically relevant compounds[9]. In this regard, an artificial pathway enabling PDO to be synthesized from glucose or sucrose was established by extending the naturally-occurring aspartate-homoserine pathway in *E. coli*. In this approach, homoserine is first deaminated by an enzyme with homoserine transaminase activity[10]. The resulting 2-keto-4-hydroxybutyrate (OHB) is then decarboxylated by an enzyme displaying OHB decarboxylase activity, while 3-hydroxypropanal (3-HPA) is finally reduced to PDO by an enzyme with 1,3-propanediol oxidoreductase activity. While the last reaction of this artificial pathway is catalyzed by the broad substrate range NADPH-dependent aldehyde reductase encoded by *yqhD* gene from *E. coli* (Ec-*yqhD*), the two preceding enzymatic reactions are not known to naturally occur in cell metabolism. In demonstration of this pathway, the authors have engineered an *E. coli* strain overexpressing both the native *serC*-encoded transaminase (Ec-SerC) and branched-chain 2-ketoisovalerate decarboxylase from *Lactococcus lactis* (Ll-KivD). Low PDO titers were however obtained (0.3 mM). In another study, Chen and colleagues[11] recently attempted to optimize this pathway by replacing the first catalytic step with an enzyme displaying homoserine dehydrogenase activity obtained by rational engineering of the *E. coli* NADP-dependent glutamate dehydrogenase, but PDO titers were not further improved.

In the present work, the design and construction of an alternative synthetic pathway enabling PDO synthesis from glucose via the TCA cycle intermediate malate through six non-natural enzymatic reaction steps is described. Malate is first converted to the non-natural metabolite (L)-2,4-dihydroxybutyrate (DHB) by employing malate kinase, malate semialdehyde dehydrogenase and malate semialdehyde reductase activities as recently demonstrated by our group[12]. Extending this pathway with DHB dehydrogenase, OHB decarboxylase and PDO oxidoreductase eventually yields PDO. Based on enzyme screening and engineering approaches, the required enzyme activities were demonstrated, improved and found to enable *in vivo* PDO production from DHB. Simultaneous expression of all six enzymatic activities in one *E. coli* strain enabled direct PDO production from glucose, while distributing the malate-to-DHB and DHB-to-PDO pathway individual modules into two *E. coli* strains co-cultivated in mineral medium further improved PDO titers.

4.2 Design of the PDO synthetic pathway

In a previous study, a *de novo* metabolic pathway leading to the production of DHB from glucose via malate was designed and experimentally validated by expressing malate kinase, malate semialdehyde dehydrogenase and malate semialdehyde reductase enzyme activities in *E. coli*[12]. This work demonstrates that the non-natural DHB metabolite may serve as a precursor for the microbial production of PDO. The conversion of DHB to PDO proceeds via three reaction steps (**Figure 4.1**): DHB is first oxidized to OHB; the resulting α-ketoacid is then decarboxylated to yield 3-HPA which is finally reduced into PDO. These reactions are catalyzed by enzymes bearing DHB dehydrogenase, OHB decarboxylase and PDO oxidoreductase reductase activities, respectively. The negative standard Gibbs free energy for the proposed pathway attests its thermodynamic feasibility (**Supplementary information p. 104-105: Note S4.1, Table S4.1**). Stoichiometric analysis of the metabolic network in *E. coli* shows that PDO can be produced from glucose with a theoretical maximum yield of 1.5 mol mol^{-1} (**Supplementary information p. 106: Note S4.2**). The maximum yield is similar to those previously reported on PDO production from glucose via homoserine[11] and glycerol[8]. Implementation of the pathway requires enzymes bearing the three constitutive enzymatic activities. The promiscuous broad range aldehyde reductase YqhD from *E. coli* (Ec-YqhD) was previously shown to catalyze the reduction of 3-HPA to PDO[14], and was therefore also used in this work. In contrast, OHB decarboxylase activity was found for the pyruvate decarboxylase from *Z. mobilis* (Zm-Pdc) but with limited success[11], while DHB dehydrogenase activity has not been reported in literature to date.

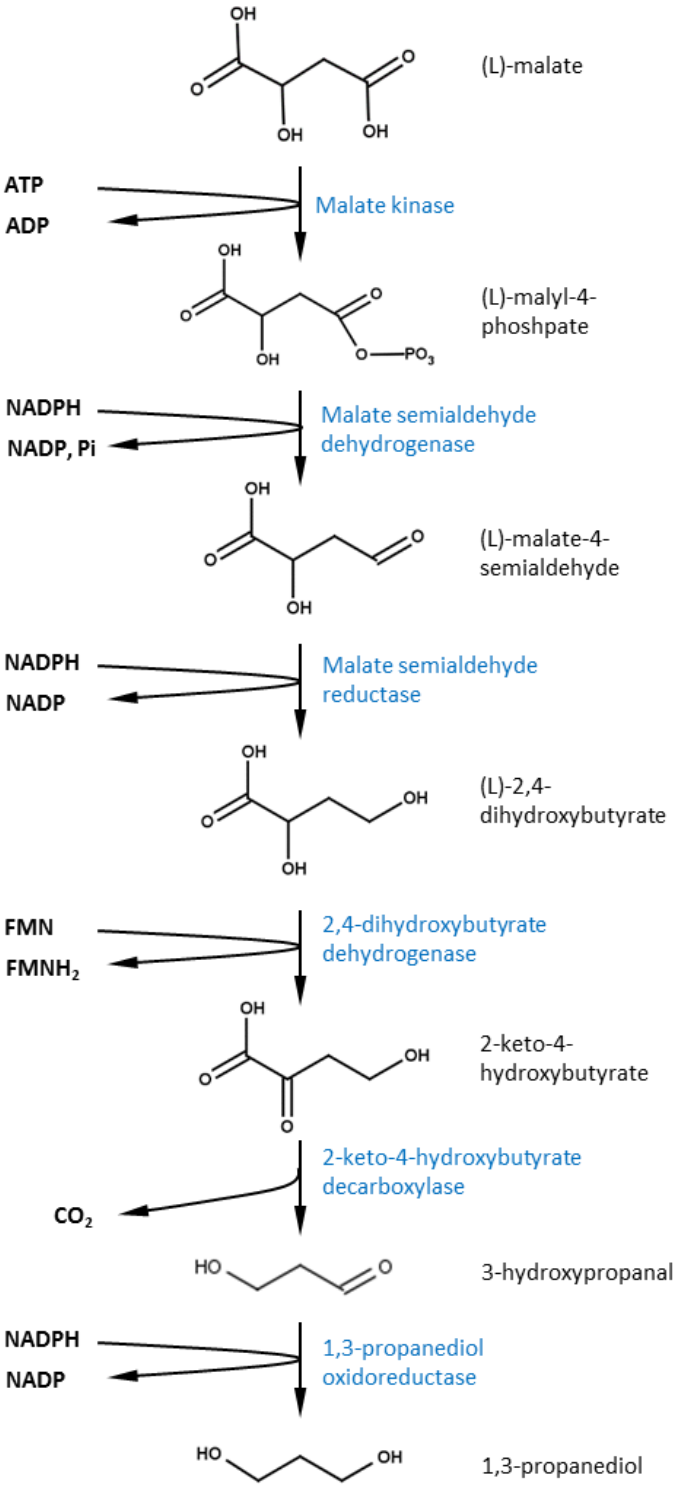

Figure 4.1. The proposed synthetic 1,3-propanediol (PDO) pathway. Based on the group contribution theory, the pathway is thermodynamically favorable and has an overall standard Gibbs free energy of -60.9 kJ mol^{-1} (see **Supplementary information p. 105: Note S4.1**).

4.3 Engineering of DHB dehydrogenase activity

The soluble NAD-dependent (L)-malate dehydrogenase enzyme variant from *E. coli* Ec-Mdh-5Q (I12V:R81Q:M85Q:D86S:G179D) was previously shown to accept DHB as substrate (as discussed in **Chapter 3**). But since the NAD-dependent oxidation of DHB is thermodynamically unfavorable (Δ_rG^0 = 15.7 kJ mol^{-1}) and this enzyme possesses a reduced affinity for (L)-DHB (K_m = 64.1 mM) and a strong preference for OHB reduction (> 100-fold), other candidate enzymes were screened. Crucially, *E. coli* MG1655 was found to possess the capability to assimilate the (L)-form of DHB, but not its D-stereoisomer (data not shown). The identification of those enzymes from *E. coli* endogenous metabolism with DHB dehydrogenase activity was therefore attempted. In particular, the utilization of 2-hydroxyacid dehydrogenase enzymes acting on substrates structurally similar to the target (L)-form of DHB molecule was envisaged. In this regard, *E. coli* displays of a set of three lactate dehydrogenase enzymes able to catalyze the interconversion between lactate and pyruvate (encoded by *lldD*, *ldhA* and *dld* genes), but of which only the membrane-associated (L)-lactate dehydrogenase (Ec-LldD) possesses activity on (L)-stereoisomers[15–17]. Additionaly, Ec-LldD relies on a FMN-dependent co-factor system[16] which would render (L)-DHB oxidation thermodynamically favorable (Δ_rG^0 = -23.1 kJ mol^{-1}; **Supplementary information p. 105: Note S4.1**) as opposed to NAD-dependent enzymes. For these reasons, the Ec-LldD enzyme was further characterized towards DHB dehydrogenase activity. To this end, the wild-type Ec-LldD enzyme was produced from a pET28-derived vector expressed in *E. coli* BL21(DE3) cells and the corresponding activity determined in crude extract. As indicated in **Table 4.1**, the Ec-LldD enzyme showed measurable activity on (L)-DHB (V_{max} = 0.6 U mg^{-1}) which was however 4-fold lower than that on (L)-lactate (V_{max} = 2.1 U mg^{-1}). In addition, a specificity (expressed as V_{max}/K_m) of 3-orders of magnitude higher for (L)-lactate than for (L)-DHB was observed, reason for which Ec-LldD was rationally engineered for increased DHB dehydrogenase activities as described next.

Table 4.1. Kinetic parameters of candidate DHB dehydrogenases on (L)-DHB and corresponding natural substrates. Data are presented as the mean (±S.D.) of at least two replicates.

Enzyme	(L)-lactate			(L)-2,4-dihydroxybutyrate			Specificity[b]
	V_{max} [U mg^{-1}][a]	K_m [10^{-3} M]	V_{max}/K_m (U mg^{-1} M^{-1})	V_{max} [U mg^{-1}][a]	K_m [10^{-3} M]	V_{max}/K_m [U mg^{-1} M^{-1}]	
Ec-LlDd	2.1 (±0.2)	0.04 (±0.01)	48,509 (±14,491)	0.6 (±0.1)	10.6 (±0.7)	58 (±10)	834
Ec-LlDd V108C	3.0 (±0.5)	0.2 (±0.0)	16,946 (±3,205)	0.5 (±0.1)	6.6 (±0.8)	82 (±24)	207

[a] Activities were estimated in U mg^{-1} cell crude extract.

[b] Specificity corresponds to the ratio of (V_{max}/K_m) between natural substrate and (L)-DHB.

Amino acid residues responsible for substrate specificity in Ec-LldD were identified by visual inspection of the crystal structure of the homologous protein from *Saccharomyces cerevisiae* (Sc-Cyb2) bound to the reaction product pyruvate and the co-factor FMN[18]. Key residues in the active site of Sc-Cyb2 are shown in **Figure 4.2**. The negative charge of the α-carboxylate group of pyruvate is neutralized by electrostatic interaction with Arg376, while residues Leu199, Leu230, Leu286, Tyr254 and Phe325 delimit the active site pocket. The catalytic residue His373 deprotonates the α-hydroxyl group of (L)-lactate, but the mechanism of substrate oxidation is still not clear. The pivotal role of those residues in substrate binding was previously discussed by Mowat and colleagues[19]. In the same study, amino acid substitutions at residue Leu230 (corresponding to Val108 in Ec-LldD), which is in contact with the methyl group of pyruvate, were demonstrated to confer activity on larger substrates (e.g. mandelate). The amino acid valine in position 108 was therefore replaced by smaller amino acid residues such as alanine, serine, cysteine and glycine, and the activities of resulting mutants were analyzed (data not shown). Only the substitution V108C resulted however in an improved catalytic efficiency for DHB (V_{max}/K_m = 82 U mg^{-1} M^{-1}) when compared to the wild-type enzyme (V_{max}/K_m = 58 U mg^{-1} M^{-1}) (**Table 4.1**). Therefore, the Ec-LldD V108C mutant enzyme was chosen as the DHB dehydrogenase enzyme to be employed in the synthetic PDO pathway.

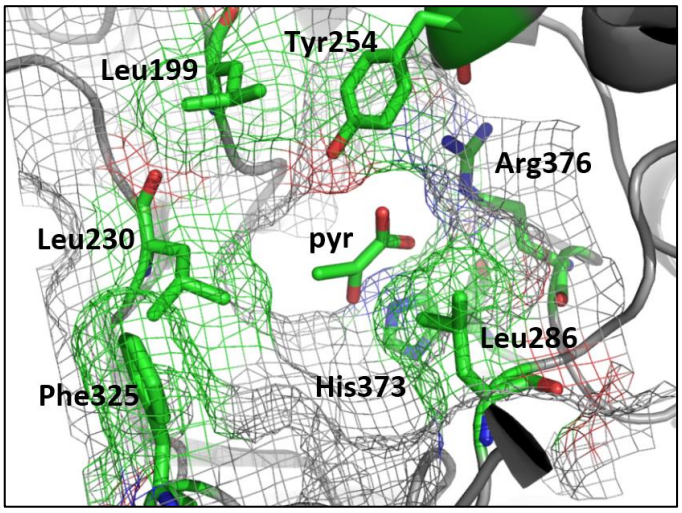

Figure 4.2. Active site region in the X-ray crystal structure of *S. cerevisiae* flavocytochrome B2 (Sc-Cyb2) (PDB code 1fcb). The residues in contact with reaction product pyruvate (pyr) are highlighted.

4.4 Engineering of OHB decarboxylase activity

The enzymes Zm-Pdc and the branched-chain ketoacid decarboxylase from *L. lactis* (Ll-KdcA) were previously demonstrated to possess activity on a wide range of 2-ketoacids, yielding corresponding aldehydes and carbon dioxide as reaction products[11]. Due to the structural similarity between OHB and 2-ketoacids, activities of both enzymes on their natural substrates and OHB were examined. While both enzymes displayed OHB decarboxylase activities, a strong preference for the natural substrate of up to 3-orders of magnitude was observed (**Table 4.2**). For this reason, a rational engineering strategy was employed aiming at increasing enzyme activities towards the OHB substrate through site-directed mutagenesis of both biocatalysts.

Manual docking of OHB into the active site pocket of Zm-Pdc revealed the substrate binding site to be too small for its accommodation (**Figure 4.3A**). In particular, the steric clash of Trp392 with the OHB substrate can be avoided by replacement with glutamine (**Figure 4.3B**), which may then be able to hydrogen bond with the 4-OH hydroxyl group, or leucine. Replacement of tryptophan by glutamine at residue 392 in Zm-Pdc resulted in 1.4-fold increased catalytic efficiency on OHB and a more than 100-fold improved specificity towards the non-natural substrate (**Table 4.2**). However, the measured enzyme activities for the variant W392Q still remained at low levels (V_{max} = 0.02 U mg^{-1}).

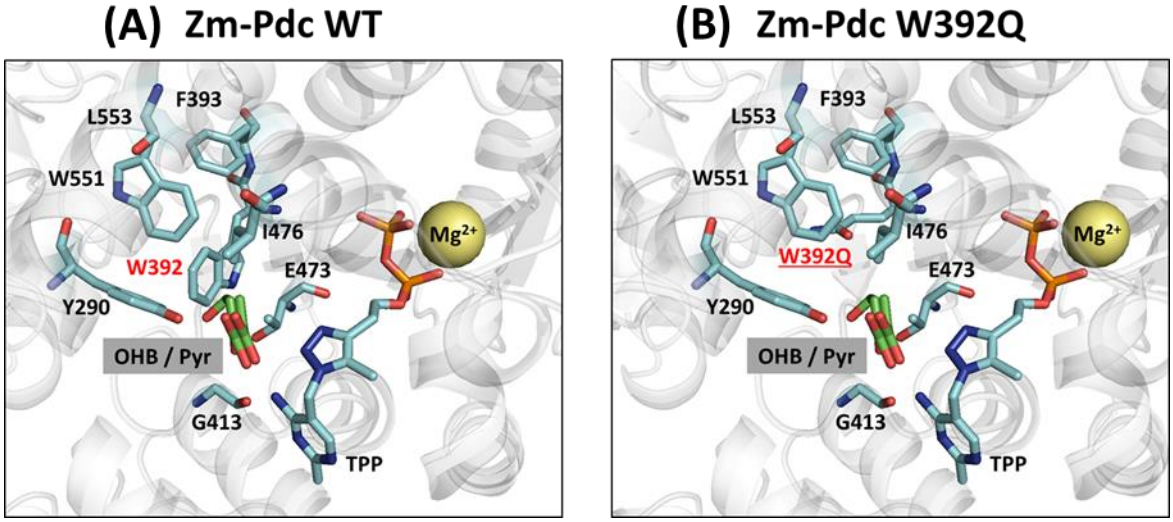

Figure 4.3. Active site region in the crystal structure of *Z. mobilis* pyruvate decarboxylase (Zm-Pdc) wild-type **(A)** and W392Q **(B)**. The synthetic 2-keto-4 hydroxybutyrate (OHB) substrate was manually docked into the X-ray structure of the enzyme with bound pyruvate (Pyr) and a TPP analogue complexed with a Mg^{2+} ion (PDB code 2wva). The individual mutation at residue 392 was introduced to Zm-Pdc manually in Pymol and resulting change in pocket size is shown.

Table 4.2. Kinetic parameters of candidate OHB decarboxylases on OHB and corresponding natural substrates. Data are presented as the mean (±S.D.) of at least two replicates.

Enzyme	Natural substrate[a]			2-keto-4-hydroxybutyrate				Specificity[c]
	V_{max} [U mg^{-1}]	K_m [10^{-3} M]	V_{max}/K_m [U mg^{-1} M^{-1}][b]	V_{max} [U mg^{-1}]	K_m [10^{-3} M]	K_i [10^{-3} M]	V_{max}/K_m [U mg^{-1} M^{-1}][b]	
Zm-Pdc	49.56 (±1.20)	1.7 (±0.2)	29,416 (±4,625)	0.01 (±0.0003)	0.4 (±0.2)	-	34 (±13)	864
Zm-Pdc W392Q	1.74 (±0.32)	5.2 (±0.3)	335 (±83)	0.02 (±0.004)	0.5 (±0.2)	5.0 (±0.4)	46 (±23)	7
Ll-KdcA	2.66 (±0.17)	2.7 (±0.2)	416 (±43)	0.02 (±0.001)	2.5 (±1.6)	-	9 (±5)	48
Ll-KdcA V461I	2.89 (±0.52)	5.1 (±1.1)	590 (±233)	0.04 (±0.0001)	1.3 (±0.5)	-	31 (±12)	19

[a] Natural substrates were used as follows: pyruvate for Zm-Pdc, 3-methyl-2-ketobutyric acid for Ll-KdcA.
[b] Activities were estimated in U mg^{-1} purified protein.
[c] Specificity corresponds to the ratio of (V_{max}/K_m) between natural substrate and OHB.

Contrarily to Zm-Pdc, the substrate binding site in Ll-KdcA can accept a wide range and more voluminous molecules including branched chain substrates. Visual inspection of the X-ray crystal structure of an enzyme-bound inhibitory TPP analogue reaction intermediate[20] manually docked with pyruvate and OHB substrates, enabled the identification of target residue positions for mutation (**Figure 4.4A**). In particular, Val461 lies at the entrance of the S-pocket to which it can control access. Replacement by a bulkier isoleucine residue, as observed in Zm-Pdc (Ile472), is expected to further hinder binding in the S-pocket, and to improve interactions with the OHB in the main substrate channel (**Figure 4.4B**). The active site of Ll-KdcA is lined by several other key residues previously identified to play a critical role in controlling S-pocket sized and acceptance of bulkier substrates, including Phe381, Gly402, Met538, and Phe542[21]. Substitution of Val461 by isoleucine in Ll-KdcA provided a 2-fold increase in maximum enzyme activity on OHB while decreasing affinity towards this non-natural substrate by approximately two-fold (**Table 4.2**). Also the replacement of Gly402 with a serine residue in Ll-KdcA was attempted but with no noticeable improvements (data not shown).

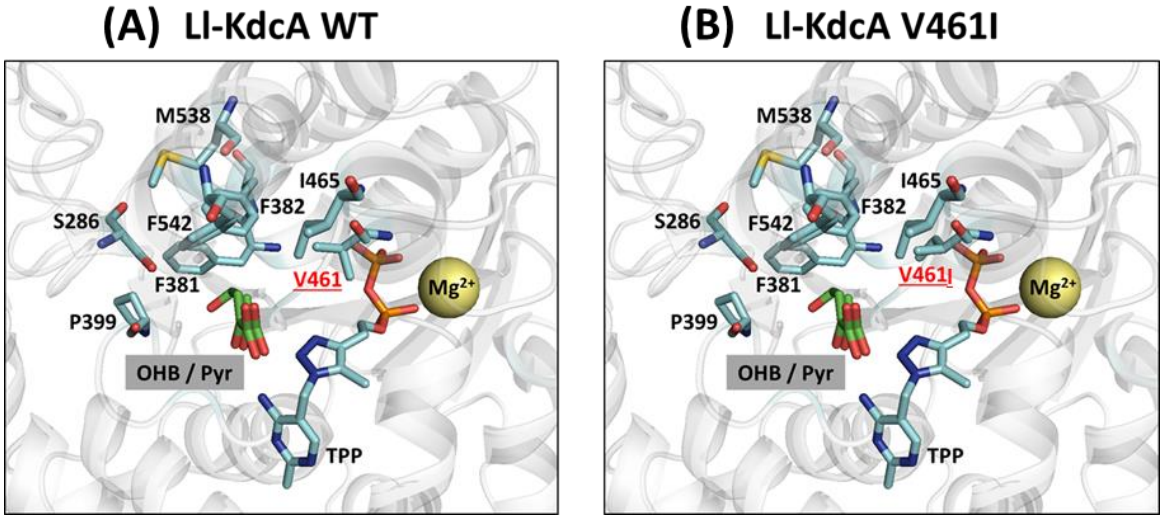

Figure 4.4. Active site region in the crystal structure of *L. lactis* branched chain keto-acid decarboxylase (Ll-KdcA) wild-type **(A)** and V461I **(B)**. The X-ray structure bound with a TPP analogue complexed with a Mg^{2+} ion (PDB code 2vbg) was aligned with Zm-Pdc structure (PDB code 2wva) crystalized with pyruvate bound. Upon protein structure overlay, the Zm-Pdc structure was hidden, leaving only the pyruvate, TPP analogue and Mg^{2+} and the Ll-KdcA active site shown. The pyruvate molecule served as an indicator of the putative binding position and orientation of the 2-ketoacid substrates within Ll-KdcA, after which the synthetic 2-keto-4 hydroxybutyrate (OHB) substrate was manually docked. The individual mutation at residue 461 was introduced to Ll-KdcA manually in Pymol and resulting change in pocket size is shown.

4.5 PDO production from DHB

To evaluate the possibility of converting DHB to PDO, an operon composed of genes encoding for the DHB dehydrogenase Ec-LldD V108C variant, OHB decarboxylase variant from either Zm-Pdc or Ll-KdcA and the aldehyde reductase Ec-YqhD was assembled. The operons were cloned into the medium-copy pACT3 vector[22] and resulting plasmids were transformed into the wild-type *E. coli* MG1655 strain. PDO production from a racemic mixture of 40 mM (D/L)-DHB after 47h of incubation was obtained for all constructs except for the strain Pen946 which harbored an empty plasmid (**Table 4.3**). These results indicate that the proposed reaction sequence to transform DHB into PDO is feasible. The strains which expressed the OHB decarboxylase variants Ll-KdcA V461I (strain Pen913) and Zm-Pdc W392Q (strain Pen966) showed the highest PDO titers (3.75 and 2.82 mM, respectively). Thus, increased OHB decarboxylase activity of the mutants translated into increased PDO production. But since DHB consumption by cells was not further improved, DHB oxidation and/or DHB uptake were hypothesized as limiting factor(s) towards PDO biosynthesis. To evaluate this possibility, the PDO-producing strain Pen913 was incubated with various DHB concentrations (10-100 mM) and PDO titers evaluated (**Figure 4.5**).

Table 4.3. Production of PDO from (D/L)-DHB by *E. coli* MG1655 wild-type host strain harboring DHB downstream pathway. Cells were cultivated in 250 mL non-baffled shake flasks on mineral medium containing 20 g L^{-1} glucose. At OD_{600} ~0.6, IPTG (0.5 mM) and a racemic mixture of (D/L)-DHB (40 mM) were added to the medium. After 40 h incubation time, PDO and (D/L)-titers were measured. Data are presented as the mean (±SD) of at least two replicates. All plasmids are derived from the pACT3 medium-copy number plasmid[22] which express, by this order, the genes encoding the enzymes: DHB dehydrogenase, Ec-L1dD V108C; OHB decarboxylase; PDO oxidoreductase, Ec-YqhD. As a control, an empty pACT3 plasmid was used.

Strain	Genes expressed	(D/L)-DHB consumed [mM]	PDO produced [mM]	PDO yield [mol mol^{-1} (D/L)-DHB]
Pen946	pACT3	4.5 (±0.5)	0.0 (±0.0)	0.00 (±0.00)
Pen911	pACT3-llDd$_{V108C}$-kdcA-yqhD	9.1 (±1.5)	1.3 (±0.0)	0.14 (±0.03)
Pen913	pACT3-llDd$_{V108C}$-kdcA$_{V461I}$-yqhD	10.2 (±1.4)	3.8 (±0.2)	0.37 (±0.08)
Pen965	pACT3-llDd$_{V108C}$-pdc-yqhD	7.0 (±0.9)	0.5 (±0.0)	0.07 (±0.00)
Pen966	pACT3-llDd$_{V108C}$-pdc$_{W392Q}$-yqhD	8.7 (±1.4)	2.8 (±0.5)	0.32 (±0.00)

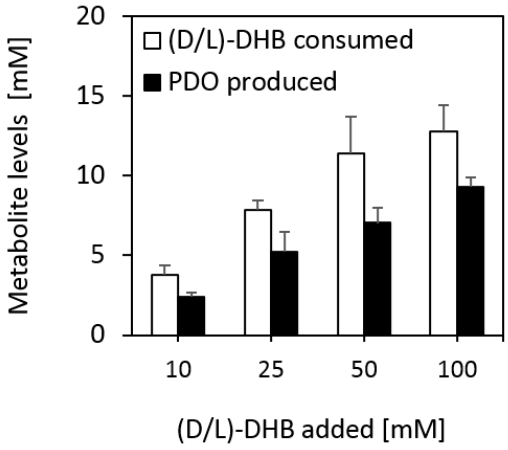

Figure 4.5. Production of PDO from various concentrations of (D/L)-DHB by Pen913 strain (*E. coli* MG1655 pACT3-llDd$_{V108C}$-kdcA$_{V461I}$-yqhD). Cells were cultivated in 250 mL non-baffled shake flasks on mineral medium containing 20 g L^{-1} glucose. At OD$_{600}$ ~0.6, IPTG (1 mM) and a racemic mixture of (D/L)-DHB were added to the medium. After 47h incubation time, product and substrate titers were measured. Data are presented as the mean (±S.D.) of at least two replicates.

Indeed, cells were able to assimilate more substrate at increasingly higher DHB concentrations. Crucially, supplementation with 50 mM and 100 mM (D/L)-DHB yielded however a marginal 12% increase in DHB consumption by cells (11.4 and 12.8 mM consumed, respectively). This occurrence suggests that DHB uptake systems may be DHB unspecific and/or saturated. Dischert and colleagues[23] recently proposed various permeases from *E. coli* as putative DHB transporters, including glycolate permease (GlcA) and α-ketoglutarate permease (KgtP). In line with this information, DHB consumption from strains expressing the synthetic pathway in parallel with the aforementioned DHB uptake systems was therefore investigated. Interestingly, DHB consumption was substantially increased within 48h only in strains that overexpressed GlcA (**Figure 4.6**), supporting the fact that DHB uptake was rate-limiting for PDO production. The lower PDO yields observed (*glcA*^{proD}: 0.44 mol mol^{-1}) in comparison to the wild-type host strain (0.62 mol mol^{-1}) may however indicate downstream catalytic inefficiencies and/or DHB conversion into other products.

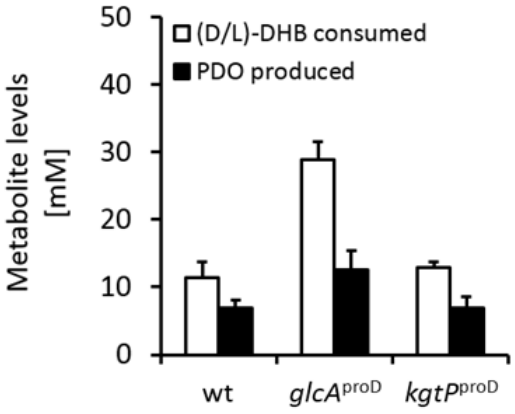

Figure 4.6. Production of PDO from (D/L)-DHB by various *E. coli* MG1655 host strains (as indicated in X-axis) harboring the pACT3-llDd$_{V108C}$-kdcA$_{V461I}$-yqhD plasmid. For each condition PDO yields were as follows, wt: 0.62 ±0.23 mol mol^{-1}, *glcA*^{proD}: 0.44 ±0.14 mol mol^{-1}, and *kgtP*^{proD}: 0.54 ±0.17 mol mol^{-1}. Cells were cultivated in 250 mL non-baffled shake flasks on mineral medium containing 20 g L^{-1} glucose. At OD$_{600}$ ~0.6, IPTG (1 mM) and a racemic mixture of 50 mM (D/L)-DHB were added to the medium. After 47h incubation time, product and substrate titers were measured. Data are presented as the mean (±S.D.) of at least two replicates.

4.6 Synthesis of PDO from glucose

To demonstrate PDO production directly from glucose, the genes coding for the complete six-reaction step pathway were cloned into two different plasmids. The previously constructed medium-copy pDHBop(ppc*) vector[12] provided all enzymatic activities required to transform malate into DHB by driving the expression of malate kinase Ec-LysC V115A:E119S:E250K:E434V, malate semialdehyde dehydrogenase Bs-Asd E218Q, and malate semialdehyde reductase Ms-Ssr H39R:N43H enzymes. Overexpression of the malate-insensitive phosphoenolpyruvate carboxylase variant Ec-Ppc K620S from the pDHBop(ppc*) plasmid was previously shown to greatly increase DHB production[12]. The three genes responsible for the conversion of DHB to PDO were cloned into a high-copy vector yielding pEXT20-lldD$_{V108C}$-pdc$_{W392Q}$-yqhD or pEXT20-lldD$_{V108C}$-kdcA$_{V461I}$-yqhD. The pDHBop(ppc*) and pEXT20-derived vectors were transformed into *E. coli* strains which were cultivated for 24h on mineral medium containing 20 g L^{-1} glucose as the carbon source. Resulting strains CF317 and CF318 which expressed all pathway genes produced small amounts of PDO (up to 0.1 mM), whereas the control strain CF316 which only harbored pDHBop(ppc*) and an empty control pEXT20 plasmid was unable to produce PDO (**Figure 4.7**). While it is still unclear why DHB excretion was 5-fold higher in CF318 than

CF317, these results suggested that the very low PDO production from glucose can be associated to a weaker catalytic capacity of the downstream DHB-to-PDO pathway as compared to the DHB-yielding pathway.

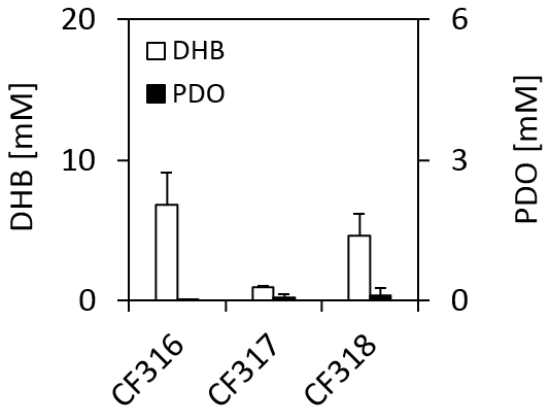

Figure 4.7. PDO production from glucose after 24h of cultivation of an *E. coli* mono-culture system. Cells were cultivated in 250 mL baffled shake flasks on mineral medium containing 20 g L^{-1} glucose. At OD_{600} ~0.6, IPTG (1 mM) was added to the medium. Data are presented as the mean (±S.D.) of at least two replicates.

4.7 Employing a two-strain co-cultivation mode to increase PDO production

To get some insights about which of the two synthetic pathways is limiting in the PDO production from glucose, a co-cultivation strategy was considered in which an *E. coli* strain expresses the synthetic DHB-yielding pathway while another one bears the PDO-producing pathway. Adopting such strategy would additionaly reduce the metabolic burden imposed by the expression of the whole synthetic pathway in a single production strain. This idea is further supported by the previous demonstration of bacterial co-cultures as a mean towards increased production of small molecules, including muconic acid[24], 3-amino-benzoic acid[25] and *n*-butanol[26]. The design of the co-culture strategy is shown in **Figure 4.8.**

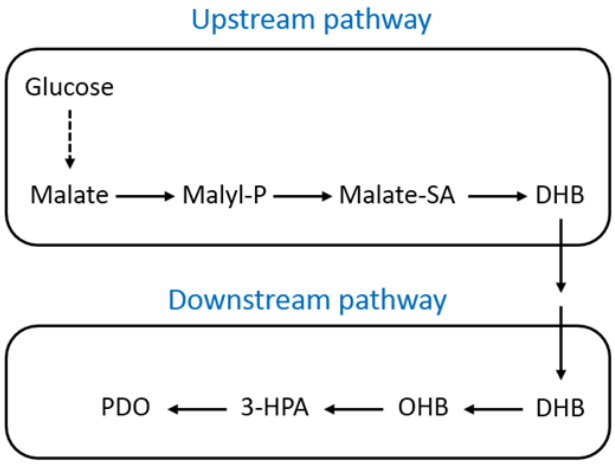

Figure 4.8. Scheme of the co-culture design, in which a first cell expresses the upstream pathway enabling DHB synthesis from glucose, while a second cell incorporates extracellularly accumulated DHB and converts it to PDO due to expression of the downstream pathway.

The upstream DHB-yielding pathway was expressed from the CF285 strain whilst the downstream part of the pathway which converted DHB into PDO was expressed in the strains Pen913 or Pen966 (which differed in the expression of the OHB decarboxylase mutants, Ll-KdcA V461I and Zm-Pdc W392Q, respectively). To evaluate the performance of the two-strain co-cultivation strategy, the DHB- and PDO-producing strains were cultivated alone or simultaneously (in the latter case, at 1:1 inoculation ratio) for 24h on glucose-containing mineral medium (**Figure 4.9**).

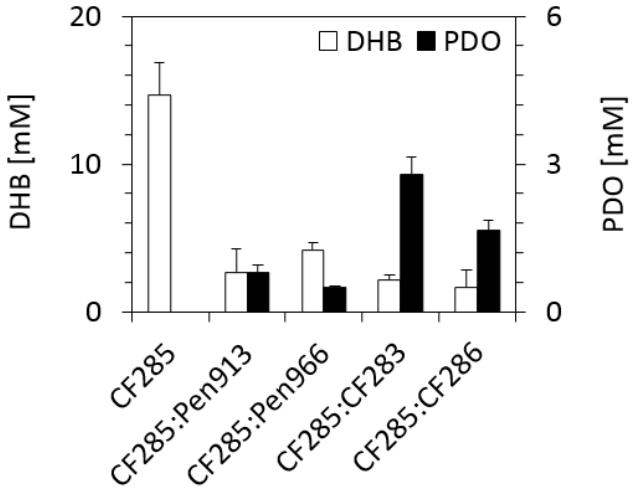

Figure 4.9. PDO production from glucose after 24h of cultivation of *E. coli-E. coli* co-culture systems at an inoculation ratio equal to 1:1 (in which 1 corresponds to an OD_{600} of 0.2). Improved PDO titers were achieved by varying PDO producing strains (Pen913, Pen966, CF283, CF286) while keeping unchanged the DHB producing strain CF285. Cells were cultivated in 250 mL baffled shake flasks on mineral medium containing 20 g L^{-1} glucose. After 3h of cell cultivation, IPTG (1 mM) was added to the medium. Data are presented as the mean (±S.D.) of at least two replicates.

In a control experiment, strain CF285 alone produced 14.7 mM DHB but no PDO. The co-cultivation of CF285:Pen913 resulted in two-fold higher PDO production than of CF285:Pen966 (0.8 mM vs 0.4 mM), indicating that the catalytic efficiency of chosen OHB decarboxylase is critical in the DHB-PDO pathway. In both cases, DHB was extracellularly accumulated, which accordingly was the highest in the co-culture producing PDO at lower titers. This observation is supported by the idea that the PDO-producing pathway bearing the DHB dehydrogenase variant Ec-LlDd V108C and either one of the OHB decarboxylase mutants (Ll-KdcA V461I and Zm-Pdc W392Q) is catalytically less efficient that the DHB-yielding pathway. As GlcA overexpression was found to enhance DHB uptake, this co-cultivation strategy was repeated using a PDO-producing strain that overexpressed *glcA* (CF283: MG1655 *glcA*^{proD} pACT3-llDd$_{V108C}$-kdcA$_{V461I}$-yqhD or CF286: MG1655 *glcA*^{proD} pACT3-llDd$_{V108C}$-pdc$_{W392Q}$-yqhD). This resulted in increased PDO production up to 2.8 mM with CF285 and CF283 strain co-cultivation, whereas only 1.5 mM was obtained with co-culture of CF285 with CF286, confirming the previous results suggesting that the OHB decarboxylase mutant Ll-KdcA V461I is more efficient than the Zm-Pdc W382Q (see **Table 4.3**).

In an attempt to fine-tune the best-performing CF285:CF2833 co-cultivation system, the inoculation ratios were varied (**Figure 4.10**). Specifically, the starting cell concentration (expressed as OD_{600}) of PDO-producing strain CF283 was step-wise decreased while keeping constant the initial cell concentration of CF285 to assure DHB production at levels not inferior to those observed above. At least 2-fold higher accumulation of DHB was observed when less CF283 was added to the co-culture system, but improvements in PDO production were only marginal. In the overall, co-cultivation of CF285 and CF283 strains at an inoculation ratio of 1:0.5 resulted in the highest production of PDO (3.4 mM) from glucose.

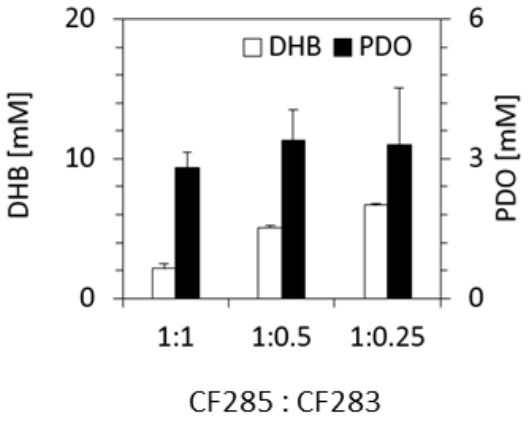

Figure 4.10. PDO production from glucose after 24h of cultivation of DHB-producing strain CF285 and PDO-producing strain CF283 at various inoculation ratios (in which 1 corresponds to an OD_{600} of 0.2). Cells were cultivated in 250 mL baffled shake flasks on mineral medium containing 20 g L^{-1} glucose. After 3h of cell cultivation, IPTG (1 mM) was added to the medium. Data are presented as the mean (±S.D.) of at least two replicates.

4.8 Discussion

In this work, a synthetic pathway enabling PDO biosynthesis from glucose under aerobic conditions via the Krebs cycle intermediate (L)-malate is presented. The proposed route does neither involve supplementation of the expensive vitamin-B12 nor the use of glycerol as a precursor unlike naturally occurring PDO pathways. Instead, it is based on the extension of a previously published (L)-DHB metabolic pathway with three additional reaction steps. But as an energy and NADPH-intensive pathway which requires one mol of ATP and three mol of NAD(P)H per each mol of PDO generated from malate, co-factor supply is likely essential for achieving relevant PDO titers. In this scope, our group[27] has previously engineered *E. coli* malate metabolism via Krebs cycle and glyoxylate shunt, which provides increased ATP and NADH amounts at the expense of a decreased maximum malate yield (1.33 mol mol^{-1} glucose). Implementation of the complete PDO pathway in a malate-overproducing strain may therefore translate into increased PDO titers.

One of the major challenges found during the construction of non-natural pathways consists in finding desired enzyme activities at rates compatible with in *vivo* applications. The proposed route uses a combination of natural and engineered enzymes acting on non-natural substrates. While the three first catalytic steps linking (L)-malate to DHB were demonstrated elsewhere[12], the discovery of the three remaining steps that would render PDO production

possible was targeted in this work. Substrate promiscuity is a feature displayed by around one-tenth of naturally occurring enzymes and particularly useful towards the discovery of new enzyme functions[28]. Based on the selection of template enzymes performing similar reactions and acting on sterically cognate substrates, candidate (L)-DHB dehydrogenase and OHB decarboxylase enzymes were first screened. The chosen enzymes exhibited however weak activities on non-natural substrates, reason for which a protein rational engineering was considered towards optimization of desired catalytic efficiencies. But the relatively limited success in increased specificities towards these non-natural substrates obtained with engineered Ec-LlDd V108C and OHB decarboxylase variants (Ll-KdcA V461I and Zm-Pdc W392Q) demand for alternative strategies. In this regard, the recently developed transcriptional-based metabolite sensor for aldehyde detection (see **Chapter 5**), including the resulting 3-HPA product of OHB decarboxylation, may be used as a tool for *in vivo* screening of OHB decarboxylase and/or other rate-limiting upstream enzyme(s) mutant libraries generated by (semi-)random mutagenesis.

For a more comprehensive understanding of the pathway, the feasibility of the proposed DHB-PDO pathway was first evaluated, but the uncomplete substrate consumption together with low PDO yields revealed an important limitation. This bioconversion experiments also indicated that the DHB uptake is a bottleneck, which was in part relieved by overexpression of the glycolate encoded transporter GlcA. However, even in a strain in which uptake of DHB has been improved, PDO yields were not increased which further suggests OHB decarboxylases to limit the rate at which DHB is converted to PDO.

When the direct production of PDO from glucose was evaluated by expressing the six-step synthetic pathway in an *E coli* strain, a very weak production titer of 0.1 mM PDO was obtained from 110 mM glucose. In the same time, metabolic burden imposed by the expression of this whole synthetic pathway is likely to occur due to the presence of a medium- and high-copy plasmids in cells. Nevertheless, this experiment revealed that the 3-step synthetic pathway that converts DHB to PDO was likely less efficient than the three-step synthetic pathway yielding DHB from malate since a significant amount of DHB accumulated in the growth medium. As a mean to overcome the hypothetical metabolic burden arising from expression of this six-step synthetic pathway, a co-cultivation strategy was attempted in which the whole pathway was split into an *E. coli* strain expressing the DHB-yielding and another *E. coli* strain that expressed the PDO-producing pathway. This

strategy resulted in increased product formation (up to 3.4 mM PDO), thereby showing that the performance of a synthetic pathway can, in principle, be improved by splitting into functional modules that are expressed in different strains. In addition, a lower accumulation of DHB in the medium was observed when employing this co-cultivation approach as compared to the expression of the whole synthetic pathway in a single *E. coli* cell, which can be ascribed in part to a better functioning of the PDO-producing pathway when it is expressed alone. For industrial applications this is however not a preferable solution, in particular, for production of bulk chemical which requires extremely high carbon efficiency that cannot be achieved by co-culturing two strains. Overcoming the identified limiting steps of the DHB-PDO synthetic pathway, additional strain engineering towards increased availability of DHB to the PDO-producing pathway, as well as resolving the metabolic burden caused by expression of the six-step synthetic pathway are essential to achieve higher production of PDO from glucose in a single strain.

4.9 Materials and methods

4.9.1 Chemicals and reagents

All chemicals and solvents were purchased from Sigma-Aldrich unless otherwise stated. Restriction endonucleases and DNA-modifying enzymes were purchased from New England Biolabs and used according to manufacturer's instructions. DNA plasmid isolation was performed using GeneJET Plasmid Miniprep Kit (Thermo Scientific). DNA extraction from agarose gel was carried out using the GeneJET Gel Extraction Kit (Thermo Scientific). DNA sequencing was carried out by Beckman Coulter Genomics (Takeley, United Kingdom) or Eurofins SAS (Ebersberg, Germany). The racemic mixture of sodium (D/L)-DHB (purity, 70%) was chemically synthetized by Adisseo SA (France).

4.9.2 Protein cloning and mutagenesis

E. coli DH5α (New England Biolabs) was routinely used for construction of plasmids. Wild-type Ec-*lldD* and Zm-*pdc* genes were amplified from genomic DNA (extracted from *E. coli* MG1655 and *Zymomonas mobilis* ATCC® 31821, respectively) by PCR. The used primers are listed in **Table 4.4**. The gene Ll-*kdcA* from *Lactococcus lactis* B1157-NIZO was codon-

optimized for expression in *E. coli* and synthesized by Eurofins. The resulting DNA fragments were digested with suitable restriction enzymes (see **Table 4.4**) and cloned into the corresponding sites of pET28a vector system (Novagen) using T4 DNA ligase (Biolabs), thereby adding an N-terminal hexa-His tag. Point mutations were introduced on pET28-derived plasmid by inverse PCR using the primers listed in **Table 4.5**. Resulting products were digested by DpnI to remove template DNA and transformed into competent cells. The introduction of desired mutation was verified by sequencing.

Table 4.4. Primers and restriction enzymes used to clone genes into pET28a expression vector

Gene*	Primer sequences (5' – 3')	Restriction enzymes	Resulting vector
Ec-*lldD*	CATATGATTATTTCCGCAGCCAGC AGATCTCTATGCCGCATTCCCTTTC	NdeI BgI2	pET28-Ec-lldD
Zm-*pdc*	CATATGAGTTATACTGTCGGTACC GGATCCCTAGAGGAGCTTGTTAAC	NdeI BamHI	pET28-Zm-pdc
Ll-*kdcA*	provided in vector by Eurofins	NheI EcoRI	pET28-Ll-kdcA

* Ec (from *Escherichia coli)*, Zm (from *Zymomonas mobilis)*, Ll (from *Lactococcus lactis)*

Table 4.5. Primers used for site-directed mutagenesis

Matrix	Mutated position	Primer sequences (5' – 3')	Restr. site
pET28-Ec-lldD	V108C	TTCCGTTTACTCTGTCGACGTGTTCCGTTTGCCCGA TCGGGCAAACGGAACCCGTCGACAGAGTAAACGGAA	HincII
pET28-Zm-Pdc	W392Q	GTTATTGCTGAAACCGGTGACTCTCAGTTCAATGCGCA GCGCATGAAGC GCTTCATGCGCTGCGCATTGAACTGAGAGTCACCGGT TTCAGCAATAAC	FspI
pET28-Ll-kdcA	V461I	TTTGCTTTATCATTAATAATGACGGCTACACAATCGAG CGCGAAATTCA TGAATTTCGCGCTCGATTGTGTAGCCGTCATTATTAAT GATAAAGCAAA	AseI

4.9.3 Protein expression and purification

Enzymes were expressed in *E. coli* BL21 (DE3) cells (New England Biolabs) in 200 mL Luria-Bertani (LB) medium supplemented with 50 µg/mL kanamycin (37 ºC, 200 rpm).

Expression cultures were inoculated from an overnight culture at OD_{600} of 0.05 and grown to OD_{600} of 0.6 before protein expression was induced by addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) to the culture medium. Cells were harvested after 3 h of incubation by centrifugation (15 min at 4,000 rpm, 4 ºC) and pellets stored at -20 ºC until further analysis.

Frozen cell pellets were resuspended in 1 mL of lysis buffer (50 mM Hepes, 300 mM NaCl, pH 7.5) and disrupted by four successive rounds of sonication (sonication interval: 30 s, power output: 30%, sonicator: Bioblock Scientific, VibraCell™ 72437). The resulting cell crude extract was directly used for measurement of FMN-dependent 2-hydroxyacid dehydrogenase activities, while 2-ketoacid decarboxylase enzymes were purified as described elsewhere[12].

4.9.4 Enzymatic assays

Protein concentrations were determined prior to enzymatic assays by the method of Bradford[29]. All enzyme assays were performed in a microplate reader (Epoch 2, BioTek) at 37ºC in 96-well flat-bottomed microtiter plates in a final volume of 250 µL.

FMN-dependent 2-hydroxyacid dehydrogenase activity: Activity of wild-type membrane-associated (L)-lactate dehydrogenase and mutant variant V108C assayed in the oxidative direction by monitoring reduction of 2,6-dichloroindophenol (DCIP) at 655 nm ($\varepsilon = 5.9\ mM^{-1}\ cm^{-1}$) during oxidation of 2-hydroxy acids. The assay mixture contained 60 mM Hepes (pH 7), 50 mM KCl, 0.06 mM DCIP and appropriate amounts of crude protein extract. Reactions were started by adding appropriate concentrations of (L)-lactate or a racemic mixture of (D/L)-DHB (Adisseo SAS, France).

2-Ketoacid decarboxylase activity: Activity of keto-acid decarboxylase wild-type and mutant variants was assayed by coupling the decarboxylation reactions to the NAD(P)H-dependent reduction of the produced aldehydes. The decarboxylation of natural substrates (pyruvate, 3-methyl-2-oxobutyric acid) was coupled to the NADH-dependent reduction of acetaldehyde catalyzed by yeast alcohol dehydrogenase (Sigma, A7011). OHB decarboxylase activity was assayed by coupling decarboxylase activity to the NADPH-dependent reduction of the released 3-hydroxypropanal by purified aldehyde reductase Ec-

YqhD. The assay mixture contained 60 mM Hepes (pH 7), 50 mM KCl, 5 mM $MgCl_2$, 0.25 mM NAD(P)H, 0.5 mM thiamine pyrophosphate, 100 µg mL^{-1} auxiliary enzyme and appropriate amounts of purified enzyme. Reactions were started by adding appropriate concentrations of pyruvate, 3-methyl-2-oxobutyric acid or OHB. The latter was synthesized in-house as previously described[30].

4.9.5 Construction of plasmids and strains for biosynthesis of PDO

All plasmids and strains constructed and used in this study for PDO production are listed in **Table 4.6**.

Plasmid construction: The gene Ec-*yqhD* was amplified by PCR from genomic DNA with primer pairs pen290/pen291, whilst the remaining genes were amplified from pET28-28 derived vectors using primer pairs listed in **Table 4.7**. DNA fragments were purified and assembled by homologous recombination with BamHI/SalI digested pEXT20 or pACT3 vector using the NEBuilder® HiFi DNA Assembly kit (New England Biolabs). The resulting plasmids were transformed into DH5α competent E. coli cells and assembled operons verified by DNA sequencing.

Strain construction: E. coli K-12 substr. MG1655 (ATCC 47076) was used as the parental strain for all constructions in this study. Expression of *glcA* and *ktgP* genes was rendered constitutive by replacing the native chromosomal 5'-UTR of each gene by the synthetic constitutive and insulated promoter proD[31]. The proD sequence was preceded by a kan resistance cassette which was amplified by PCR adding 50 bp flanking sequences that were homologous to the target locus. The resulting DNA fragment was used to replace the natural gene promoter by homologous recombination[32]. Primers used are listed in **Table 4.7**. Positive clones were selected on LB agar plates containing kanamycin (50 µg mL^{-1}) and verified by PCR analysis. The kan cassette was removed from the genome by expressing FLP recombinase from the pCP20 plasmid[33] and correct excision of the cassette was verified by PCR using locus specific primers (**Table 4.7**). Plasmids were transformed into the target *E. coli* strains using standard protocols[34].

Table 4.6. Strains used in this study.

Strain	Genotype	Source
NEB5-α	*E. coli fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80Δ (lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	NEB
BL21(DE3)	*E. coli fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS*	NEB
MG1655	*E. coli* $F^-\lambda^-$ ilvG-rfb-50 rph-1	ATCC 47076
CF220	MG1655 $glcA^{proD}$	This work
CF221	MG1655 $kgtP^{proD}$	This work
Pen946	MG1655 / pACT3	This work
Pen911	MG1655 / pACT3-$lldD_{V108C}$-kdcA-yqhD	This work
Pen913	MG1655 / pACT3-$lldD_{V108C}$-$kdcA_{V461I}$-yqhD	This work
Pen965	MG1655 / pACT3-$lldD_{V108C}$-pdc-yqhD	This work
Pen966	MG1655 / pACT3-$lldD_{V108C}$-pdc_{W392Q}-yqhD	This work
CF285	MG1655 / $pDHB_{op}$ (ppc*)	Walther et al[12]
CF316	CF285 / pEXT20	This work
CF317	CF285 / pEXT20-$lldD_{V108C}$-pdc_{W392Q}-yqhD	This work
CF318	CF285 / pEXT20-$lldD_{V108C}$-$kdcA_{V461I}$-yqhD	This work
CF283	CF220 / pACT3-$lldD_{V108C}$-$kdcA_{V461I}$-yqhD	This work
CF284	CF221 / pACT3-$lldD_{V108C}$-$kdcA_{V461I}$-yqhD	This work
CF286	CF220 / pACT3-$lldD_{V108C}$-pdc_{W392Q}-yqhD	This work

Table 4.7. Primers used in strain construction.

Primer	DNA sequences (5' – 3')
Plasmid construction	
pen286 (fw-IldD)	CGGTACCCGGGGATCCTGCCCTAGCCTATTCGATTAAGGAGGTCAAATATGATTATTTCCGCAGCCAG
	TTGGTGTTATTTTTGACTATGCCGCATTCCCTTTCGC
pen287 (rv-IldD-kdcA)	TCAAAAATAACACCAAATCAATAAGGAGGAACGTTAtgtataccgttggggattatc
pen288 (fw-kdcA)	TCCTTAGTGGATATCGttatttgttctgttcagcaaac
pen289 (rv-kdcA)	CGATATCCACTAAGGAGGTAACTAATGAACAACTTTAATCTGCA
pen290 (fw-yqhD)	ATGCCTGCAGGTCGACTTAGCGGGCGGCTTCGTATA
pen291 (rv-yqhD)	ATATGTTTTATCTGATCTATGCCGCATTCCCTTTCGC
pen292 (rv-lldD-pdc)	ATCAGATAAAACATATTTTAGAGGAGGTAGTAAATGAGTTATACTGTCGGTACC
pen293 (fw-pdc)	TCCTTAGTGGATATCGTTATTTGTTCTGTTCAGCAAAC
pen294 (rv-pdc)	
Strain construction	
cf245 (kgtP-prod_fw)	ATGGTTAAGGTTGCATAATGATATGCAACAAATGTATAATATTTCGTGTAGGCTGGAGCTGCTTC
cf246 (kgtP-prod_rv)	GTATCACTACTTGTCAGTTTGCTGTCTGCCGTTACAGTACTTTCAGCcatATAATACCTCCTAAAGTTAAACAAA
	ATTATTTGTAG
cf253 (glca-prod_fw)	CCGAACCGTTATTACACGCCTGGCGTTTACGCGAAAAAGAAAGTCATTAAGTGTAGGCTGGAGCTGCTTC
	GCCCCAGTCCTCCCATCGGCATATACATTTGGGTCCAGGTAACcatATAATACCTCCTAAAGTTAAACAAAATTA
cf254 (glca-prod_rv)	TTTGTAG
Verification primers	
cf247 (kgtP_ver_fw)	ATACGTGTCC TCCTTACCAG
cf248 (kgtP_ver_rv)	AGTAGAGTGAACAGAACGAGTA
cf255 (glcA_ver_fw)	AATTTCGCTA ACTCGTG
cf256 (glcA_ver_rv)	ATGTCCTTTCAGACGTAATA

4.9.6 Synthesis of PDO from DHB

All cell cultivation was carried out at 37 ºC on a rotary shaker (Infors HT, France) running at 200 rpm. Pre-cultures were grown in 5 mL of LB in 50 mL falcon tubes. After ~10h, 500 µL were used to inoculate a second pre-culture (10 mL of 90 % v/v M9 mineral medium supplemented with 20 g L^{-1} glucose and 10 % v/v LB in 50 mL falcon tubes) that was cultivated overnight. The biomass needed to start main cultures with a starting OD_{600} of 0.2 was transferred to 250 mL baffled shake flasks containing 25 mL of 90 % v/v M9 mineral medium supplemented with 20 g L^{-1} glucose and 10 % v/v LB, and IPTG and (D/L)-DHB were added (at amounts indicated in each associated figure or table) when OD_{600} reached ~0.6. The antibiotic chloramphenicol was added when required at 25 mg L^{-1}.

4.9.7 Synthesis of PDO from glucose

Pre-cultures were grown in 5 mL of LB in 50 mL falcon tubes. After ~10h, 500 µL were used to inoculate a second pre-culture (10 mL of M9 mineral medium supplemented with 20 g L^{-1} glucose in 50 mL falcon tubes) that was cultivated overnight. The biomass needed to start main cultures with a starting OD600 of 0.2 was transferred to 250 mL baffled shake flasks containing 25 mL of M9 mineral medium supplemented with 20 g L^{-1} glucose. The antibiotics ampicillin, kanamycin sulphate and chloramphenicol were added when required, respectively, at 100, 50 and 25 mg L^{-1}. IPTG (1 mM) was added after 3 h of cell cultivation. In co-cultivation experiments, the different strains were inoculated at a ratio described in the text.

4.9.8 Analytical methods

Concentrations of glucose and (D/L)-DHB were determined on a Dionex Ultimate 3,000 HPLC system (Thermo Scientific, France) equipped with a RI detector (RID-10A, Shimadzu, Japan) and UV/Vis detector (SPD-20A, Shimadzu). The sample injection volume was 20 µL, and the compounds were separated on a Rezex RoA-organic acid H^+ (8%) resin-based column preceded by a SecurityGuard guard cartridge (Phenomenex, USA). The separation was performed at 80 ºC with 0.5 mM H_2SO_4 at 0.5 mL min^{-1} as mobile phase.

All samples were centrifuged (2 min at 13,000 rpm) and syringe-filtered (0.2 μm), and the resulting supernatant stored at -20 ºC before analysis. A standard calibration curve was obtained by injecting standards and used for all compound analysis.

4.10 Supplementary information

Table S4.1. Gibbs free energy of formation of PDO pathway intermediates.

Compound	**Contributing groups**	**$\Delta_f G^{0'}$ [kJ mol^{-1}]**
malate^{2-}	2x –COO^{1-} 1x -OH 1x -CH$_2$- 1x –CH<	-840.98
malyl-P^{2-}	1x –COO^{1-} 1x -OH 1x -CH$_2$- 1x –CH< 1x -CO-OPO$_3$H$^-$	-1740.1
malate-SA$^-$	1x -COO^{1-} 1x –OH 1x -COH 1x -CH$_2$- 1x –CH<	-620.5
DHB$^-$	1x -COO^{1-} 2x –OH 2x -CH$_2$- 1x –CH<	-660.2
OHB$^-$	1x -COO^{1-} 1x –OH 2x -CH$_2$- 1x >CO	-626.6
3-HPA	1x –OH 2x -CH$_2$- 1x >CO	-287.3
PDO	2x –OH 3x -CH$_2$-	-326.9
ATP3- ADP2- NADPH NADP+ HPO42- H+ NADH NAD+ FMN FMNH2 CO2		-2819.4 -1949.1 -3082.9 -3104.9 -1096.1 -39.9 -2193.8 -2215.8 -1352.4 -1449.0 -386.0

Note S4.1. Standard Gibbs free energy of the DHB pathway

The stoichiometry of the reactions malate kinase (MK), malate semialdehyde dehydrogenase (MSD) and malate semialdehyde reductase (MSR), 2,4-dihydroxybutyrate dehydrogenase (DD), OHB decarboxylase (OD) and aldehyde reductase (AR) is as follows (Equations S4.1-S4.6):

MK: $malate^{2-} + ATP^{3-} + H^{+} = malyl\text{-}P^{2-} + ADP^{2-}$ (S4.1)

MSD: $malyl\text{-}P^{2-} + NADPH = malate\text{-}SA^{-} + NADP^{+} + HPO_4^{2-}$ (S4.2)

MSR: $malate\text{-}SA^{-} + NADPH + H^{+} = DHB^{-} + NADP^{+}$ (S4.3)

DD: $DHB^{-} + FMN + H^{+} = OHB^{-} + FMNH_2$ (S4.4)

OD: $OHB^{-} + H^{+} = 3\text{-}HPA + CO_2$ (S4.5)

AR: $3\text{-}HPA + NADPH + H^{+} = PDO + NADP^{+}$ (S4.6)

The standard Gibbs free energy for the formation of the DHB pathway intermediates ($\Delta_f G^{0'}$) was calculated based on the group contribution theory[35] using the dataset published by Jankowski[36] (Supplementary Table 4). The standard Gibbs free energy of a reaction ($\Delta_r G^{0'}$) can be estimated according to Equation S8 from the standard Gibbs free energy of the formation of the participating compounds ($\Delta_f G^{0'}$), with ν_i being the stoichiometric coefficients.

$$\Delta_r G^{0\prime} = \sum_i \nu_i \cdot \Delta_f G_i^{0\prime} \quad \text{(S4.8)}$$

Accordingly, the standard Gibbs free energy of the reactions S1 – S6 are MK: 11 kJ mol^{-1}, MSD: 1.5 kJ mol^{-1}, MSR: -21.8 kJ mol^{-1}, DD: -23.1 kJ mol^{-1}, OD: -6.8 kJ mol^{-1}; AR: -21.7 kJ mol^{-1}. The upstream DHB pathway has a standard Gibbs free energy of -9.3 kJ mol^{-1}, while downstream PDO pathway displays a standard Gibbs free energy of -51.6 kJ mol^{-1}.

The complete PDO pathway has a standard Gibbs free energy of -60.9 kJ mol^{-1}.

Note S4.2. Calculation of the theoretical 2,4-dihydroxybutyric acid yield

A previously published stoichiometric model of the central carbon metabolism in *E. coli*[37] was extended by 6 reactions steps enabling PDO production from malate. The theoretical yield was calculated based on elementary mode analysis using the CellNetAnalyzer software package[38]. The flux map showing one of the predicted carbon flux distributions that provide maximum yield was determined (data not shown). In the absence of cell growth, the maximum PDO yield is 1.5 mol per mol glucose.

4.11 References

(1) Mak, W. S., Tran, S., Marcheschi, R., Bertolani, S., Thompson, J., Baker, D., Liao, J. C., and Siegel, J. B. (2015) Integrative genomic mining for enzyme function to enable engineering of a non-natural biosynthetic pathway. *Nat. Commun. 6*, 10005.

(2) Saxena, R. K., Anand, P., Saran, S., and Isar, J. (2009) Microbial production of 1,3-propanediol: recent developments and emerging opportunities. *Biotechnol. Adv. 27*, 895–913.

(3) Kraus, G. A. (2008) Synthetic Methods for the Preparation of 1,3-Propanediol. *CLEAN - Soil, Air, Water* 36. 648–651.

(4) Xin, B., Wang, Y., Tao, F., Li, L., Ma, C., and Xu, P. (2016) Co-utilization of glycerol and lignocellulosic hydrolysates enhances anaerobic 1,3-propanediol production by *Clostridium diolis*. *Sci. Rep. 6*.

(5) Biebl, H., and Marten, S. (1995) Fermentation of glycerol to 1,3-propanediol: use of cosubstrates. *Appl. Microbiol. Biotechnol. 44*, 15–19.

(6) Hartlep, M., Hussmann, W., Prayitno, N. Meynial-Salles, I., and Zeng, A.-P. (2002) Study of two-stage processes for the microbial production of 1,3-propanediol from glucose. *Appl. Microbiol. Biotechnol. 60*, 60–66.

(7) Nakamura, C. E., Gatenby, A. A., Hsu, A. K., La Reau, R. D., Haynie, S. L., Diaz-Torres, M., Trimbur, D. E., Whited, G. M., Nagarajan, V., Payne, M. S., Picataggio, S. K., Nair, R. V. (1997) Method for the production of 1,3-propanediol by recombinant microorganisms. Patent US/6013494A.

(8) Nakamura, C. E., and Whited, G. M. (2003) Metabolic engineering for the microbial production of 1,3-propanediol. *Curr. Opin. Biotechnol. 14*, 454–459.

(9) Mori, Y., and Shirai, T. (2018) Designing artificial metabolic pathways, construction of target enzymes, and analysis of their function. *Curr. Opin. Biotechnol. 54*, 41–44.

(10) Soucaille, P., and Boisart, C. (2014) Method for the preparation of 1,3-propanediol from sucrose. Patent WO/2012/004247A1.

(11) Chen, Z., Geng, F., and Zeng, A.-P. (2015) Protein design and engineering of a de novo pathway for microbial production of 1,3-propanediol from glucose. *Biotechnol. J. 10*, 284–289.

(12) Walther, T., Topham, C. M., Irague, R., Auriol, C., Baylac, A., Cordier, H., Dressaire, C., Lozano-Huguet, L., Tarrat, N., Martineau, N., Stodel, M., Malbert, Y., Maestracci, M., Huet, R., André, I., Remaud-Siméon, M., and François, J. M. (2017) Construction of a synthetic metabolic pathway for biosynthesis of the non-natural methionine precursor 2,4-dihydroxybutyric acid. *Nat. Commun. 8*, 15828.

(13) Deck, P., Exner, K. M., and Buschhaus, B. (2009) Method for the production of D,L-2-hydroxy-4-alkylthio butyric Acid. Patent WO/2008/022953A1.

(14) Jarboe, L. R. (2011) YqhD: a broad-substrate range aldehyde reductase with various applications in production of biorenewable fuels and chemicals. *Appl. Microbiol. Biotechnol. 89*, 249–257.

(15) Clark, D. P., Nikolova, S., and Jiang, G. R. (2001) Regulation of the *ldhA* gene, encoding the fermentative lactate dehydrogenase of *Escherichia coli*. *Microbiology*. *147*, 2437-2446.

(16) Futai, M., and Kimura, H. (1977) Inducible membrane-bound L-lactate dehydrogenase from *Escherichia coli*. Purification and properties. *J. Biol. Chem. 252*, 5820–7.

(17) Futai, M. (1973) Membrane D-lactate dehydrogenase from *Escherichia coli*. Purification and properties. *Biochemistry 12*, 2468–2474.

(18) Tegoni, M., and Cambillau, C. (2008) The 2.6-Å refined structure of the *Escherichia coli* recombinant *Saccharomyces cerevisiae* flavocytochrome b2-sulfite complex. *Protein Sci. 3*, 303–313.

(19) Mowat, C. G., Wehenkel, A., Green, A. J., Walkinshaw, M. D., Reid, G. A., and Chapman, S. K. (2004) Altered substrate specificity in flavocytochrome b_2: structural insights into the mechanism of L-lactate dehydrogenation. *Biochemistry 43*, 9519–9526.

(20) Berthold, C. L., Gocke, D., Wood, M. D., Leeper, F. J., Pohl, M., and Schneider, G. (2007) Structure of the branched-chain keto acid decarboxylase (KdcA) from *Lactococcus lactis* provides insights into the structural basis for the chemoselective and enantioselective carboligation reaction. *Acta Crystallogr. Sect. D Biol. Crystallogr. 63*, 1217–1224.

(21) Chen, G. S., Siao, S. W., and Shen, C. R. (2017) Saturated mutagenesis of

ketoisovalerate decarboxylase V461 enabled specific synthesis of 1-pentanol via the ketoacid elongation cycle. *Sci. Rep. 7*.

(22) Dykxhoorn, D. M., St. Pierre, R., and Linn, T. (1996) A set of compatible tac promoter expression vectors. *Gene 177*, 133–136.

(23) Dischert, W., Dumon-Seignovert, L., Vasseur, P., Bestel-Corre, G., and Soucaille, P. (2016) A modified microorganism for the optimized production of 2,4-dihydroxybutyrate with enhanced 2,4-dihydroxybutyrate efflux *1*. Patent WO/2016/162442A1.

(24) Zhang, H., Pereira, B., Li, Z., and Stephanopoulos, G. (2015) Engineering *Escherichia coli* coculture systems for the production of biochemical products. *Proc. Natl. Acad. Sci. 112*, 8266–8271.

(25) Zhang, H., and Stephanopoulos, G. (2016) Co-culture engineering for microbial biosynthesis of 3-amino-benzoic acid in *Escherichia coli*. *Biotechnol. J. 11*, 981–987.

(26) Saini, M., Hong Chen, M., Chiang, C. J., and Chao, Y. P. (2015) Potential production platform of *n*-butanol in *Escherichia coli*. *Metab. Eng. 27*, 76–82.

(27) Trichez, D., Auriol, C., Baylac, A., Irague, R., Dressaire, C., Carnicer-Heras, M., Heux, S., François, J. M., and Walther, T. (2018) Engineering of *Escherichia coli* for Krebs cycle-dependent production of malic acid. *Microb. Cell Fact. 17*, 113.

(28) Erb, T. J., Jones, P. R., and Bar-Even, A. (2017) Synthetic metabolism: metabolic engineering meets enzyme design. *Curr. Opin. Chem. Biol. 37*, 56–62.

(29) Bradford, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem. 72*, 248–254.

(30) Walther, T., Calvayrac, F., Malbert, Y., Alkim, C., Dressaire, C., Cordier, H., and François, J. M. (2018) Construction of a synthetic metabolic pathway for the production of 2,4-dihydroxybutyric acid from homoserine. *Metab. Eng. 45*, 237–245.

(31) Davis, J. H., Rubin, A. J., and Sauer, R. T. (2011) Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res. 39*, 1131.

(32) Datsenko, K. A., and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. 97*, 6640–

6645.

(33) Cherepanov, P. P., and Wackernagel, W. (1995) Gene disruption in *Escherichia coli:* TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene 158*, 9–14.

(34) Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular cloning: a laboratory manual. *Mol. cloning a Lab. manual.*

(35) Mavrovouniotis, M. L. (1991) Estimation of standard Gibbs energy changes of biotransformations. *J. Biol. Chem. 266*, 14440–5.

(36) Jankowski, M. D., Henry, C. S., Broadbelt, L. J., and Hatzimanikatis, V. (2008) Group contribution method for thermodynamic analysis of complex metabolic networks. *Biophys. J. 95*, 1487–1499.

(37) Stelling, J., Klamt, S., Bettenbrock, K., Schuster, S., and Gilles, E. D. (2002) Metabolic network structure determines key aspects of functionality and regulation. *Nature 420*, 190–193.

(38) Klamt, S., Saez-Rodriguez, J., and Gilles, E. D. (2007) Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Syst. Biol. 1*, 2.

Chapter 5. Development of a metabolite sensor for high-throughput detection of aldehydes in *Escherichia coli* [1]

5.1 Introduction

Aldehydes are a class of chemicals with a wide range of applications, namely in the synthesis of rubbers, plastics and formulation of flavors and fragrances. Formaldehyde, butanal and isobutyraldehyde are among the aldehydes produced in quantities greater than 1 Mt per year, while the aromatic aldehydes vanillin and benzaldehyde are the two most widely used flavoring agents in food products[1,2]. Producing these compounds from inexpensive sugar feedstocks using microbial species is therefore an alternative to the currently established chemical synthesis and plant-extraction processes[3]. Beyond their use as end-products, many aldehydes are intermediates of metabolic pathways resulting in multiple and diversified products (e.g. fatty acids, alcohols, alkanes, carboxylic acids), including those developed by our group leading to the production of glycolic acid (via glycolaldehyde), 1,3-propanediol (via 3-hydroxypropanal) and 2,4-dihydroxybutyrate (via malate semialdehyde)[4–7]. Optimizing those enzymatic activities that produce aldehydes is an essential task during the improvement of the above mentioned synthetic metabolic pathways. However, the rational or evolutionary engineering of such new enzymatic activities can be subject to various bottlenecks.

Recent advances in synthetic biology together with decreasing DNA synthesis costs allow today for rapid design and assembly of DNA sequences that may encode individual enzymes or even entire metabolic pathways[8,9]. While an increasingly large sequence space becomes accessible at lower costs and higher rates, the available analytical systems to screen this vast number of sequences for positive variants often do not provide the necessary throughput. Small molecules such as aldehydes can be detected and quantified through the use of conventional chromatography (GC-MS, HPLC-MS)[10] and colorimetric techniques (Schiff's test). However, the analytical throughput of these techniques is far from being exploitable for laboratory evolution campaigns that are based on random mutagenesis. The development

[1] A modified version of the published manuscript Frazão CJR, Maton V, François JM, Walther T. (2018) Development of a metabolite sensor for high-throughput detection of aldehydes in Escherichia coli. *Front. Bioeng. Biotechnol.* 6, 118.

of high-throughput screening and selection methods is therefore of crucial importance for successful directed evolution of strains and enzymes aiming at aldehyde overproduction[11].

A variety of devices sensitive to the accumulation of intra- or extracellular small molecules, ions or changes in physical parameters have been evolved using microbial cells[12]. Among those, ligand-responsive transcription factors (TFs) play a major role in cell physiological adaptation. They are DNA-binding proteins that regulate gene expression by physically interacting with specific target molecules[13,14]. Therefore, they are interesting devices for a broad range of applications, including the construction of whole-cell biosensors or metabolite sensors for detection of, respectively, extracellular or intracellular target molecules[15–17]. Metabolite sensors are usually modular. While a sensing module contains a transcriptional regulator which is activated in the presence of a target ligand, a reporter module consisting of a corresponding cognate promoter which drives transcription of a reporter gene (e.g. *lacZ*, *gfp* and mutant variants) enables the output of measurable signals[18]. The utilization of metabolite sensors has recently gained particular interest for evolution of bacterial strains[19–21] and enzymes[22,23].

When metabolite sensors that detect the intracellular production of target molecules are combined with FACS systems, ultra-high throughput analyses and sorting of individual cells become feasible at rates higher than 10^7 cells screened per hour[24]. While these numbers make metabolite sensors a highly attractive tool for strain and enzyme engineering, the use of these sensors and respective implementation in a screening protocol is still far from being an "off-the-shelf technology" thereby requiring significant research efforts for both, the optimization of the sensor and the screening protocols. In this work, a metabolite sensor for detection of various aldehydes in *E. coli* was developed which employs the aldehyde-responsive transcription factor YqhC to drive the expression of the yellow fluorescent reporter protein SYFP2. Based on 5'-UTR engineering of the sensor and reporter modules, the gain of the fluorescence signal in response to the model compound glycolaldehyde was strongly increased. The best sensor variant detected various extracellularly added aldehydes at concentrations in the range of 1-10 mM. In addition, intracellular production of aldehydes via two in-house synthetic pathways was reliably detected. This result showed that the metabolite sensor can be applied in screening systems that rely on the detection of intracellular production of a target aldehyde in live cells.

5.2 YqhC can be employed as an aldehyde sensor

In a previous study, YqhD was identified by our research group as the major glycolaldehyde reductase in *E. coli*[4], while others demonstrated this enzyme to be active on a broad range of short-chain aldehydes (e.g. butyraldehyde, 3-hydroypropanal, acrolein)[25]. In a genomic context, YqhD is expressed from the *yqhD-dkgA* operon that is known to be induced by the divergently transcribed transcriptional regulator YqhC (**Figure 5.1**)[26,27]. Genome-wide transcriptome studies from our group, furthermore, revealed *yqhD* and *dkgA* genes to be strongly up-regulated (by 26- and 10-fold, respectively) when wild-type *E. coli* cells were exposed to 10 mM glycolaldehyde[4]. On the basis of these observations, the transcription factor YqhC may in principle be used to engineer an *in vivo* aldehyde-sensor system in *E. coli*.

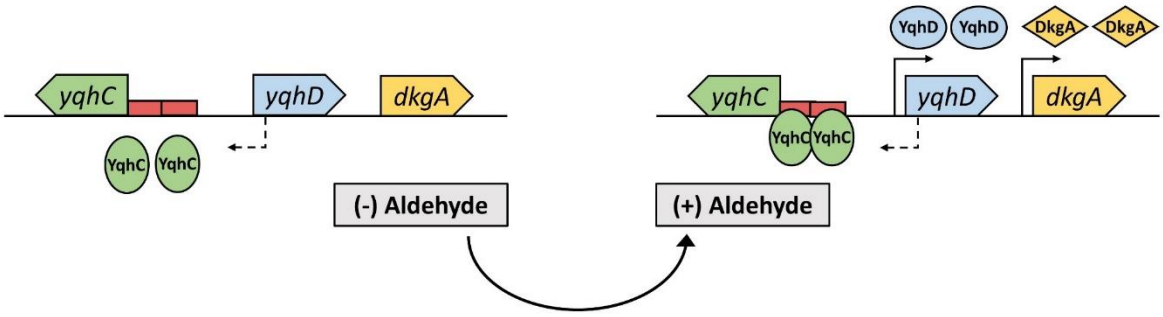

Figure 5.1. The *yqhC/D-dkgA* operon in *E. coli* MG1655[28]. When cells are exposed to aldehydes, the constitutively expressed transcription factor YqhC binds to the promoter region of *yqhD* that contains a SoxS-like binding sequence as well as a 24-bp palindrome (red rectangles), enhancing/activating transcription of *yqhD* and *dkgA* genes, resulting in the increased expression of the NADPH-dependent aldehyde reductases YqhD and DkgA. Solid and dashed arrows represent the transcription start site of divergently transcribed genes.

To investigate this possibility, a bi-modular system was constructed in which the sensing module (named pSENS-13) drove constitutive expression of the regulatory protein YqhC from a low-copy number vector. In the reporter module (named pREP-14), expression of the super yellow fluorescent protein SYFP2[29] was placed under transcriptional control of the *yqhD* promoter region, that included 150-nt of the *yqhC* coding sequence and the adjacent 108-nt *yqhC/D* intergenic region (until 28-nt downstream the transcription start site where the putative ribosome binding site pre-sequence upstream *yqhD* gene starts) (**Figure 5.2a**). To maximize protein expression, a 35-nt strong ribosome binding site (RBS_01) was designed using the RBS calculator tool[30] and placed in front of the *syfp2* reporter gene (**Table 5.1** shows DNA sequence). Reporter and sensing modules were then co-transformed

into the parent strain CF30 (MG1655 Δ*sad* Δ*yqhD*), that was chosen as the host for evaluating biosensor-strains. Deletion of the *yqhD* and *sad* genes was expected to minimize intracellular degradation of short-chain aldehydes and semialdehydes, respectively. All resulting biosensor-strains were cultivated in M9 mineral medium and candidate aldehydes were added to the exponentially growing cells when OD_{600} reached ~0.6. It is of note that the utilization of a minimal medium resulted in a strong reduction of cell background fluorescence, as opposed to utilization of Luria Broth (LB) and 2x Yeast tryptone (YT) rich media (data not shown). Since glycolaldehyde was previously found as a potent inducer of the YqhC-dependent transcriptional response[4], the behaviour of the constructed sensor was evaluated in response to this compound at a non-lethal concentration of 5 mM. After a 4h incubation period, fluorescence intensity was measured at the single-cell level by flow cytometry and found to be increased by 6.8-fold when compared to cells which were cultivated in the absence of this aldehyde (**Figure 5.2b**). In a control experiment, the host strain CF30 was co-transformed with the pZS23 empty plasmid and the reporter module pREP-14. The resulting strain displayed no fluorescence increase upon aldehyde exposure, thereby confirming the feasibility of the developed metabolite sensor.

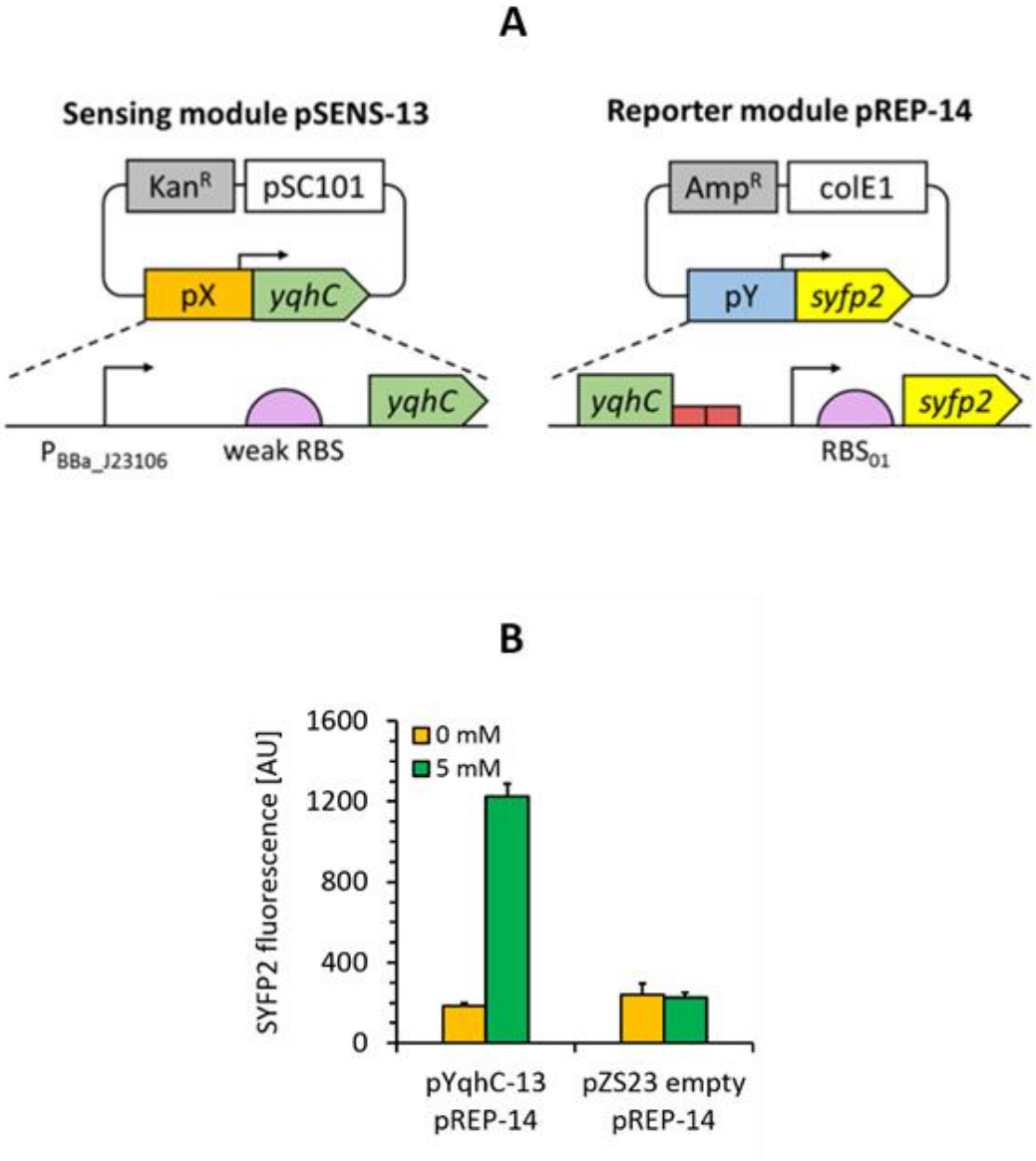

Figure 5.2. Design of the YqhC-based aldehyde sensor. **(A)** The sensing module pSENS-13 consists of a low-copy plasmid in which *yqhC* is under control of a medium-strength constitutive promoter (P_{BBa_J23106}) and a weak ribosome binding site. The reporter module pREP-14 was built by fusing the *syfp2* reporter gene (preceded by a strong RBS) to the YqhC cognate promoter in a high-copy vector. The 5'-UTR regions containing regulatory elements responsible by transcription of *yqhC* and reporter genes were named pX and pY, respectively. In each module, the antibiotic resistance marker and origin of replication is shown (grey and white boxes, respectively). **(B)** Fluorescence variation upon aldehyde exposure of engineered *E. coli* strain co-transformed with pSENS-13 and pREP-14. In a control experiment, the host strain was transformed with pZS23 and pREP-14. All strains are derived from the host strain CF30 (*E. coli* MG1655 Δ*sad* Δ*yqhD*). Cells were cultivated in M9 mineral medium containing 20 g L^{-1} glucose and incubated for 4 h with 5 mM glycolaldehyde when OD_{600} reached ~0.6. SYFP2 fluorescence was calculated using cytometry data based on geometric mean. The reported values represent the mean ± S.D. ($n \geq 2$).

Table 5.1. Constitutive promoters and RBS used for construction of sensing modules in the metabolite sensor.

Reference	Sequence (5' – 3')	Strength
Promoters		
BBa_J23106 [a]	TTTACGGCTAGCTCAGTCCTAGGTATAGTGCTAGC	1,185 a.u. [a]
BBa_J23114 [a]	TTTATGGCTAGCTCAGTCCTAGGTACAATGCTAGC	256 a.u. [a]
BBa_J23113 [a]	CTGATGGCTAGCTCAGTCCTAGGGATTATGCTAGC	21 a.u. [a]
proD, insulated	CACAGCTAACACCACGTCGTCCCTATCTGCTGCCCTAGGTCTATGAGTG GTTGCTGGATAACTTTACGGGCATGCATAAGGCTCGTATAATATATTCA GGGAGACCACAACGGTTTCCCTCTACAAATAATTTTGTTTAACTTT	5,191. a.u. [b]
RBS [c]		
weak	GGTCCACCGCTTACCCCCCCAAGGGACGAATAAA	10,000 T.I.R.
very weak	GTCTTAACAAAGGAAAAAATTTACT	1,000 T.I.R.
medium	AAATTTACTTATAAAGGAGGAGATAG	100,000 T.I.R.
strong	TCGGAAGAAGAATCGAGGAGGAGGTATCA	1,000,000 T.I.R.

[a] Nomenclature of promoters and relative promoter strengths as in Registry of Standard Biological Parts (http://partsregistry.org/Main_Page).

[b] Strength calculated based on information available at Registry of Standard Biological Parts (http://parts.igem.org/Part:BBa_J23114) and study from Davis and co-workers[31].

[c] RBS with various strengths were designed using online RBS calculator tool (https://salislab.net/software/forward).

5.3 Engineering biosensor modules for improved aldehyde detection

The effect of varying *yqhC* expression levels on the strength of the fluorescence signal by engineering promoter and RBS sequences in the sensing module was next evaluated (**Table 5.2**). The impact of three alternative constitutive promoters with different characteristics (strength, insulation) immediately upstream of the weak RBS that controlled protein expression from the *yqhC* coding sequence was first tested (see DNA sequences in **Table 5.1**). Whilst keeping reporter module pREP-14 unaltered in the host strain, co-transformation with the engineered pSENS-X plasmid variants (in which X = 16-18) resulted in distinct behaviors regarding fluorescence induction ratios upon glycolaldehyde exposure (**Table 5.2**). The expression of *yqhC* gene under control of a weak constitutive promoter (P_{BBa_J23114}) resulted in the highest fluorescence induction ratios (18.9-fold, $p < 0.001$), and the corresponding plasmid pSENS-16 was therefore used as the backbone for further modifications. The implementation of ribosome binding sites of variable strengths (pSENS-X, in which X = 19-21) resulted in further improved fluorescence induction ratios of up to 32.6-fold when the pSENS-20 plasmid (medium-strength RBS, $p < 0.01$) was used.

Table 5.2. Fluorescence induction upon aldehyde exposure of engineered *E. coli* strains co-transformed with various sensing modules and pREP-14 as reporter module. All strains are derived from the host strain CF30 (*E. coli* MG1655 Δ*sad* Δ*yqhD*). Cells were cultivated in M9 mineral medium containing 20 g L^{-1} glucose and incubated for 4 h with 5 mM glycolaldehyde when OD_{600} reached ~0.6. SYFP2 fluorescence was calculated using cytometry data based on geometric mean. The reported values represent the mean ± S.D. ($n \geq 2$).

Sensing module			Reporter module	Fluorescence induction[b]
Name	**Promoter strenght**	**RBS strength[a]**		
pSENS-13	medium (P_{BBa_J23106})	weak	pREP-14	6.8 (±1.0)
Promoter engineering				
pSENS-16	weak (P_{BBa_J23114})	weak	pREP-14	18.9 (±1.1)***
pSENS-17	very weak (P_{BBa_J23113})	weak	pREP-14	4.6 (±0.5)
pSENS-18	strong (proD, insulated)	weak	pREP-14	13.0 (±1.9)
RBS engineering				
pSENS-19	weak (P_{BBa_J23114})	very weak	pREP-14	28.2 (±0.4)
pSENS-20	weak (P_{BBa_J23114})	medium	pREP-14	32.6 (±1.6)**
pSENS-21	weak (P_{BBa_J23114})	strong	pREP-14	24.6 (±6.9)

[a] Strengths of RBS driving yqhC expression were determined using the RBS calculator tool and vary across 3-orders of magnitude.

[b] Ratio between single-fluorescence values of aldehyde-induced and uninduced cells.

*** $p < 0.001$, pSENS16 significant when compared to sensing modules with weak RBS strength (pSENS13,17,18).

** $p < 0.01$, pSENS20 significant when compared to sensing modules with weak promoter strength (pSENS16,19,21).

Having optimized the sensing module, the aldehyde sensor was further improved by engineering the promoter region of the reporter module pREP-14 (**Figure 5.3**). To this end, the plasmid pREP-15 was first created by extending the promoter region to include the full *yqhC/D* intergenic region plus the adjacent 150-nt downstream coding region of *yqhD* followed by a stop codon (**Figure 5.3a**). The additional introduction of this nucleotidic region was made in an attempt of including possibly missing uncharacterized motifs present in the genome of *E. coli*. However, significantly lower fluorescence induction ratios (10.9-fold) were observed upon aldehyde exposure when compared to pREP-14 (32.6-fold, $p < 0.001$) (**Figure 5.3c**). For this reason, the optimization of the promoter region in the pREP-14 module was rather attempted to further enhance SYFP2 expression. In the genome of *E. coli,* the *yqhD* gene is preceded by the -35 and -10 elements TTGAGA and CACAAT, respectively, to which RNA polymerase and the sigma factor σ^{70} bind to initiate transcription (with C as the initiation element)[26]. This configuration was maintained in pREP-14 to control expression of SYFP2. But since the sequence of the -10 promoter core element deviates from the consensus sequence for σ^{70}-dependent promoters (-35: TTGACA; -10: TATAAT; initiation element: A)[32], the -10 core and downstream distal elements in pREP-14 were replaced by those found in the well-characterized synthetic IPTG-inducible and σ^{70}-dependent promoter $P_{LlacO-1}$ (**Figure 5.3b**). When the resulting plasmid pREP-22 was used to detect glycolaldehyde, a 70-fold fluorescence induction ratio was observed (**Figure 5.3c**), which corresponded to a 2-fold improvement when compared to the utilization of plasmid pREP-14 as reporter module ($p < 0.001$). Therefore, the plasmids pSENS-20 and pREP-22 were used as sensing and reporter modules for subsequent experiments.

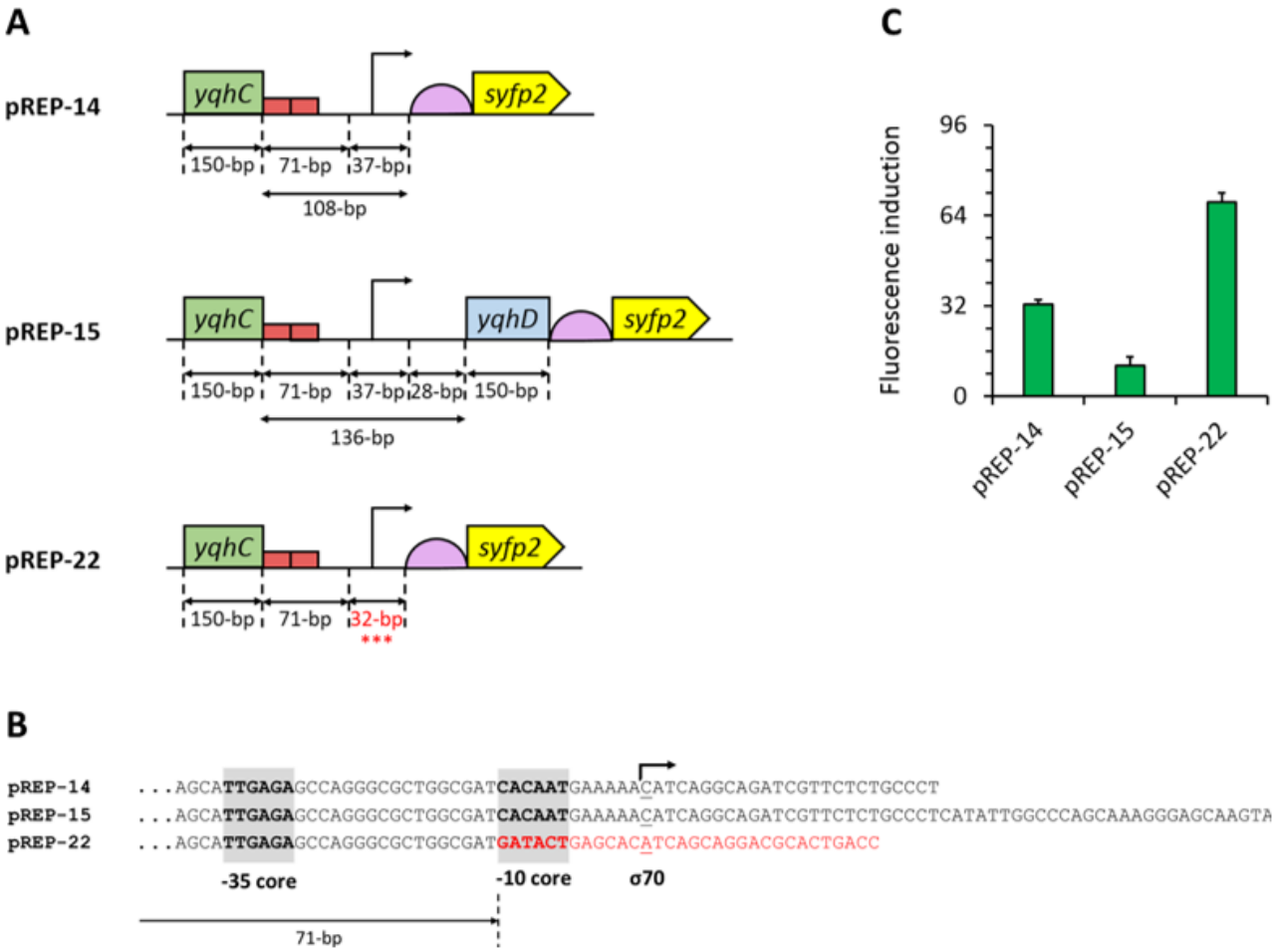

Figure 5.3. Engineering of the 5'-UTR of the reporter module. **(A)** Schematic map of characteristic sequence elements in the sensor module. **(B)** Comparison of promoter sequences in the pREP-X reporter modules. **(C)** Fluorescence induction upon aldehyde exposure of engineered *E. coli* strains co-transformed with pSENS-13 and various reporter module. All strains are derived from the host strain CF30 (*E. coli* MG1655 Δ*sad* Δ*yqhD*). Cells were cultivated in M9 mineral medium containing 20 g L^{-1} glucose and incubated for 4 h with 5 mM glycolaldehyde when OD_{600} reached ~0.6. SYFP2 fluorescence was calculated using cytometry data based on geometric mean. The reported values represent the mean ± S.D. ($n \geq 2$).

5.4 The YqhC-based aldehyde sensor detects various aldehydes

The best metabolite sensor was next characterized in response to the presence of various aldehydes at a concentration of 5 mM (**Figure 5.4**). After incubating exponentially-growing cells with candidate aldehydes, a clear correlation between the fluorescence induction and the chain length or the chemical structure of the aldehyde could not be established, although the presence of a benzyl group in aldehyde molecules (vanillin and phenylacetaldehyde) resulted in comparatively high fluorescence. Additionally, the presence of the short-chain aldehydes acrolein and succinic semialdehyde led to inductions up to 20-fold when compared to non-induced cells, whereas the biosensor was almost insensitive to butanal, furfural and hexanal.

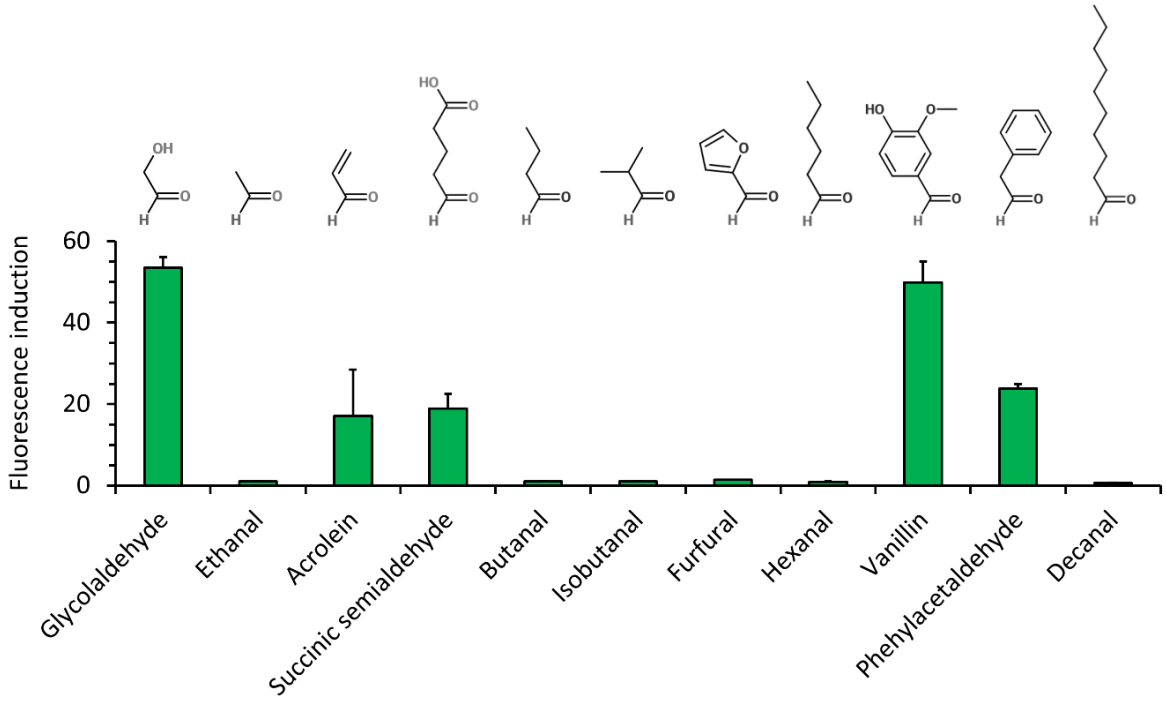

Figure 5.4. Aldehyde detection spectrum (at a concentration of 5 mM) of *E. coli* host strain CF30 (MG1655 Δ*sad* Δ*yqhD*) harboring pSENS-20 and pREP-22 as sensing and reporter modules, respectively. Cells were cultivated in M9 mineral medium containing 20 g L^{-1} glucose and incubated for 12 h with aldehyde inducer when OD_{600} reached ~0.6. SYFP2 fluorescence was calculated using cytometry data based on geometric mean. The reported values represent the mean ± S.D. ($n \geq 2$).

The dose-response curves for the fluorescence-inducing compounds are depicted in **Figure 5.5**. With the exception of vanillin, they show that extracellular aldehyde concentrations above 1 mM were necessary to trigger a significant increase of the fluorescence signal.

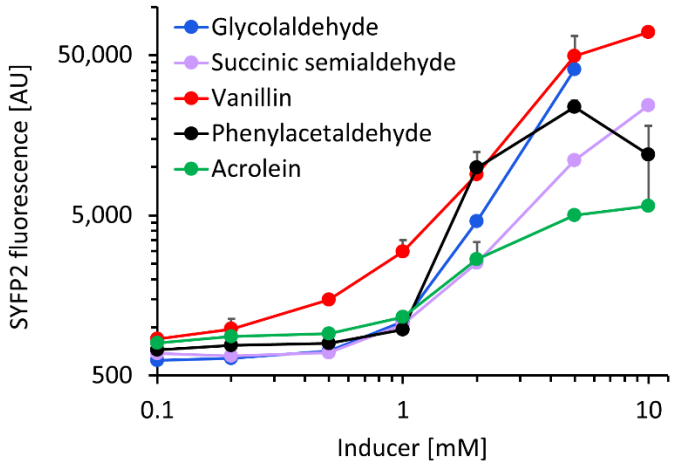

Figure 5.5. Dose-response curve of *E. coli* host strain CF30 (MG1655 Δ*sad* Δ*yqhD*) harboring pSENS-20 and pREP-22 as sensing and reporter modules, respectively, when exposed to various aldehydes in the concentration range of 0.1-10 mM. Cells were cultivated in M9 mineral medium containing 20 g L^{-1} glucose and incubated for 12 h with aldehyde inducer when OD_{600} reached ~0.6. SYFP2 fluorescence was calculated using cytometry data based on geometric mean. The reported values represent the mean ± S.D. ($n \geq 2$).

5.5 The aldehyde sensor can detect *in vivo* aldehyde production

To explore the applicability of the biosensor as a tool for strain and/or enzyme evolution, intracellular aldehyde production during induction of two in-house synthetic metabolic pathways was monitored.

5.5.1 Xylulose-1-phosphate synthetic pathway

In the xylulose-1-phosphate (X1P) pathway (**Figure 5.6a**), xylose is first converted into (D)-xylulose by xylose isomerase (XylA, *E. coli*)[4]. Xylulose is then phosphorylated by an enzyme with xylulose-1-kinase activity (KhkC, *Homo sapiens*) before the resulting xylulose-1-phosphate is cleaved into glycolaldehyde and dihydroxyacetone phosphate (DHAP) by a xylulose-1-phosphate aldolase (AldoB, *H. sapiens*). While glycolaldehyde can be further metabolized into either ethylene glycol or glycolic acid, DHAP is metabolized through the Embden-Meyerhof-Parnas pathway thereby enabling cell growth. It was previously shown that deletion of the xylulose-5-kinase encoding gene *xylB* was necessary to deviate the xylose-derived carbon flux into the synthetic pathway[4]. Therefore, we evaluated whether the glycolaldehyde production via the synthetic pathway could be monitored by our metabolite sensor in an *E. coli* Δ*xylB* Δ*yqhD* mutant strain that expresses the aldehyde sensor modules (pSENS-20, pREP-22) and the synthetic pathway (pZA33-

khkC-aldoB). Deletion of *yqhD* served to minimize reduction of glycolaldehyde to ethylene glycol and under these conditions, the production of glycolic acid is very low[4]. The resulting strain was cultivated in mineral medium containing either 20 g L^{-1} glucose or a mixture of 1 g L^{-1} glucose and 10 g L^{-1} xylose as carbon sources. In both cases, expression of the synthetic pathway was induced by the addition of 1 mM IPTG at the start of the culture. Rapid growth was observed during the cultivation on glucose, resulting in a depletion of carbon source in the first 24 h of the cultivation (**Figure 5.6b**). Cells growing on glucose exhibited no increase in fluorescence or extracellular accumulation of glycolaldehyde. A different behavior was observed when cells were cultivated on the glucose/xylose mixture. As expected, glucose was first consumed resulting in a doubling of the cell density after the first 3h of the cultivation. During the diauxic shift which lasted approximately 3h, cell density increased only very marginally before the growth rate increased again concomitant with utilization of xylose as carbon source. The fluorescence signal further mirrors this behavior in that there was no observable increase of fluorescence during growth on glucose and during the diauxic shift. Only upon induction of the X1P pathway by xylose, an increase in single-cell fluorescence by more than 40-fold was measured, which nicely correlated with the accumulation of glycolaldehyde in the medium.

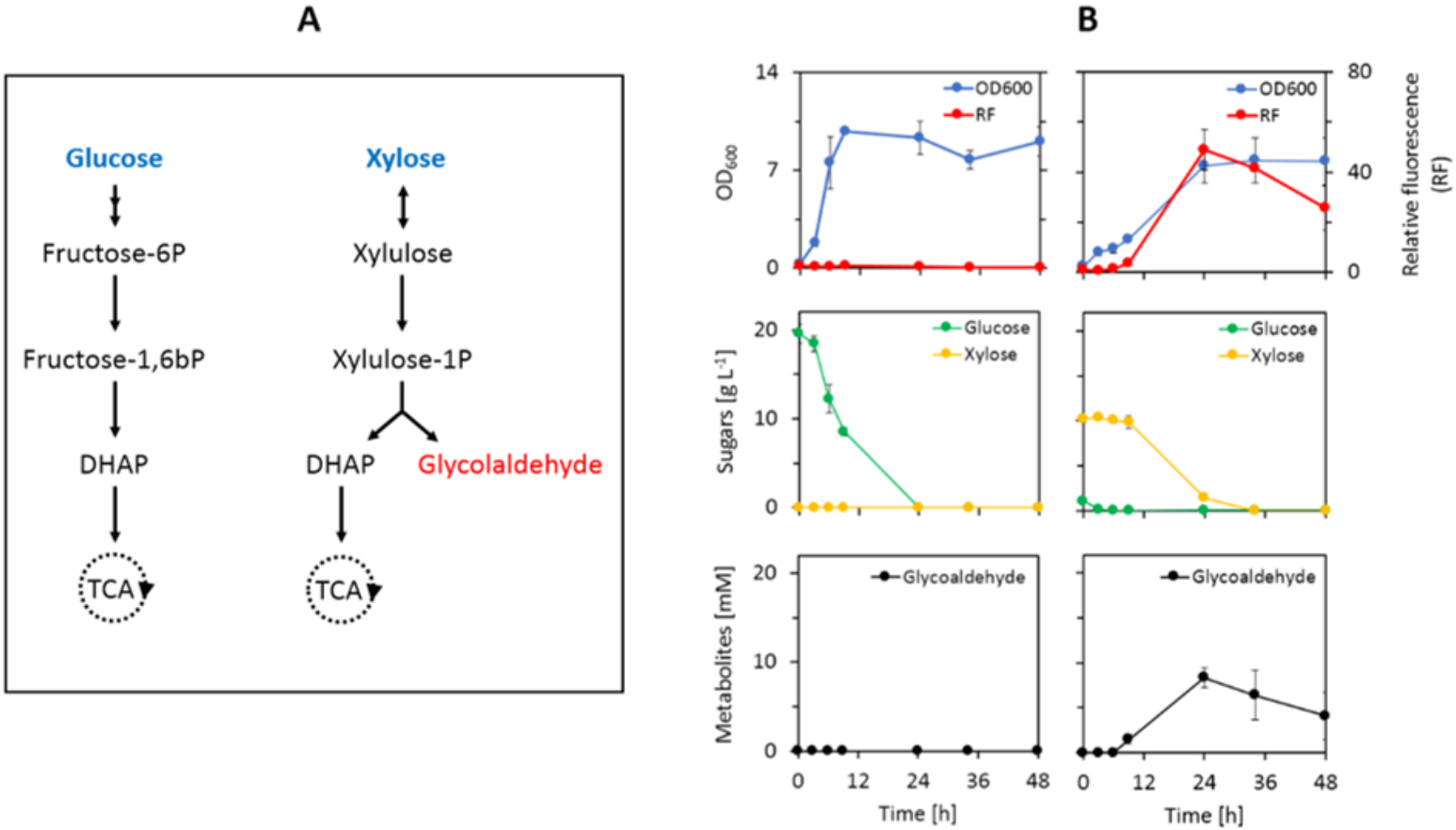

Figure 5.6. Monitoring of *in vivo* production of glycolaldehyde by an artificial metabolic pathway using the aldehyde metabolite sensor. **(A)** Natural glucose assimilation pathway (left) and synthetic xylulose-1-phosphate pathway (right). **(B)** Fermentation profile of E. coli host strain CF272 (MG1655 Δ*xylB* Δ*yqhD*) co-transformed with pYqhC-20, pREP-22 and pZA33-khkC-aldoB plasmids in the presence of glucose (left panel) or a mixture of glucose / xylose (right panel). Relative fluorescence was calculated from the ratio of SYFP2 fluorescence at a given time point and the SYFP2 fluorescence at the beginning of the cultivation. Fluorescence values were calculated using cytometry data based on geometric mean. The reported values represent the mean ± S.D. (n ≥ 2). Glycolaldehyde was determined in the extracellular medium by HPLC coupled to RI detection system. As this compound elutes almost at the same retention time as glycolic acid, and that since this later can be detected by UV, a simple visualization of the UV chromatograms confirmed that there was no glycolic acid produced under this condition.

5.5.2 1,3-Propanediol synthetic pathway from 2,4-dihydroxybutyrate

In **Chapter 4**, a *de novo* six-step pathway enabling biosynthesis of 1,3-propanediol (PDO) from malate was conceived and experimentally validated. During the study, metabolic inefficiencies linked, among other factors, to the low catalytic efficiency of 2-keto-4-hydroxybutyrate (OHB) decarboxylases were observed, reason for which improving such enzymes by directed evolution campaigns may ultimately render the designed pathway more efficient. Therefore, this subsection aims at detecting the 3-hydroxypropanal (3-HPA) product released by the enzyme since previous incubation of the metabolite sensor with exogenously added 3-hydroxypropanal yielded no fluorescence increase (data not shown) presumably due to the high-instability of the commercial version of the aldehyde.

In line with this information, an alternative strategy for demonstration of intracellular accumulation of the 3-HPA compound was employed and it is based on the expression of a fraction of the PDO pathway that links DHB to 3-HPA (**Figure 5.7a**). In this sub-pathway, DHB is first converted into OHB by a (L)-DHB dehydrogenase. OHB is then decarboxylated by an enzyme with OHB decarboxylase activity. Since expression of the aldehyde reductase Ec-YqhD was previously shown to catalyze the conversion of the released 3-HPA molecule to PDO, deletion of *yqhD* and *sad* genes was expected to avoid aldehyde degradation. The production of 3-HPA via the synthetic pathway was first monitored by the metabolite sensor in an *E. coli* Δ*sad* Δ*yqhD* $lldD^{proD}$ $glcA^{proD}$ $lacI^{Q}$::Sp host strain (named CF267) that harbored the aldehyde sensor modules (pSENS-20, pREP-22) and a medium-copy pZA33 plasmid expressing various OHB decarboxylase variants. The overexpression of DHB dehydrogenase and DHB importer encoding genes *lldD* and *glcA* permitted increased flux towards OHB synthesis from DHB, while the constitutive expression of $lacI^{Q}$ ensured a strong repression of target genes expressed from the pZA33 plasmid in the absence of IPTG inducer molecule.

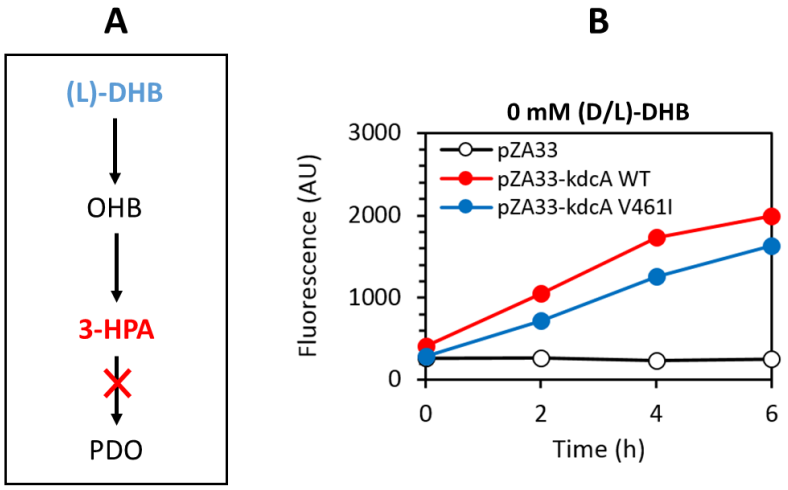

Figure 5.7. The monitoring process of *in vivo* production of 3-hydropropanal (3-HPA) from a racemic mixture of (D/L)-2,4-dihydroxybutyrate (L-DHB) **(A)** is hampered by the unstable levels of background fluorescence in the absence of DHB precursor molecule from the *E. coli* host strain CF267 (MG1655 Δ*sad* Δ*yqhD lldD*^{proD} *glcA*^{proD} *lacI*^{Q}::Sp) co-transformed with pYqhC-20, pREP-22 and pZA33-derived plasmids **(B)**. Cells were cultivated in M9 mineral medium containing 20 g L^{-1} glucose and IPTG (1 mM) was added when OD_{600} reached ~0.6 (t = 0 h), after which fluorescence was monitored. SYFP2 fluorescence was calculated using cytometry data based on geometric mean. The reported values represent the mean ± S.D. ($n \geq 2$).

The resulting strains were cultivated in mineral medium containing 20 g L^{-1} glucose, and IPTG was added at 1 mM at OD_{600} ~0.6 (t = 0 h) to express the OHB decarboxylase variants Ll-KdcA wt or V461I. In a control experiment in which the DHB precursor was not added to the cultivation medium (**Figure 5.7b**), the expression of both decarboxylases resulted however in a fluorescence increase over time thereby suggesting that both enzyme variants were able to convert other intracellularly available 2-ketoacids into corresponding aldehydes to which the metabolite sensor is sensitive. This was further confirmed by the expression of the empty plasmid pZA33, which resulted in nearly no fluorescence variation over time in the absence of DHB. Crucially, the Ll-KdcA wild-type enzyme has been previously shown to possess activity on a broad range of 2-ketoacids involved in the synthesis of branched chain amino acids (BCAA)[33]. The BCAA metabolic pathways leading to the synthesis of valine, leucine and isoleucine are represented in **Figure 5.8**.

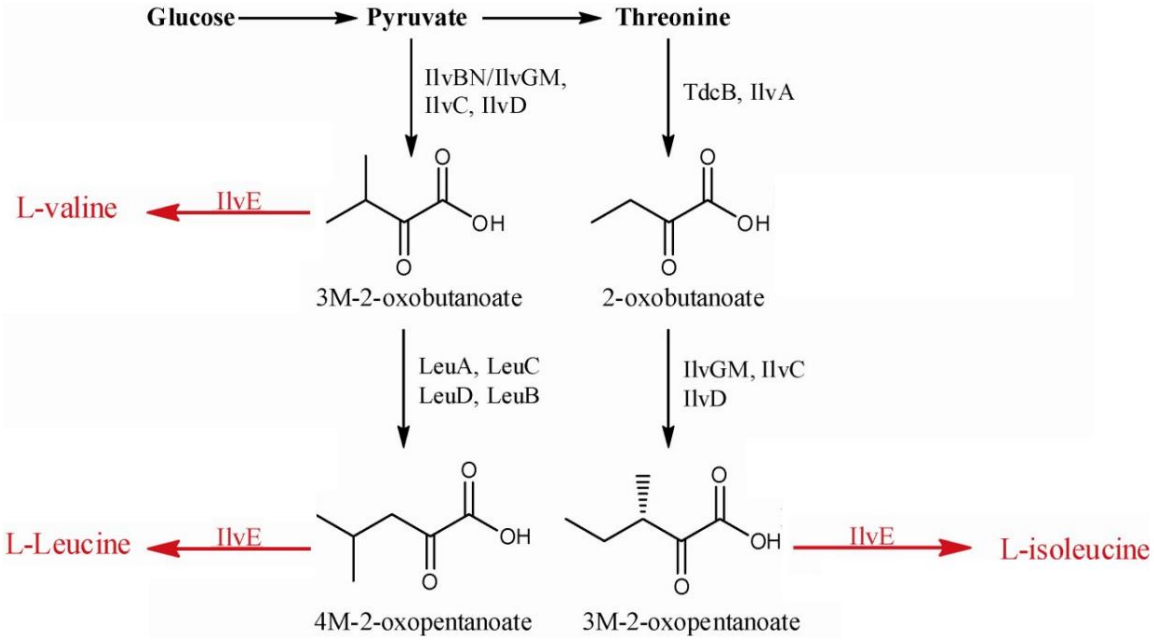

Figure 5.8. Branched-chain amino acid pathway in *E. coli* involves the synthesis of intermediate 2-ketoacids. Adapted from Dietrich, 2013[34].

In particular, the enzymes encoded from the *ilvEDA* operon are determinant for the anabolism/catabolism of valine, leucine and isoleucine. Whilst the addition of extracellular excess amounts of those amino acids has previously been reported to result in the transcriptional repression of *ilvEDA* operon and subsequent abolishment of BCAA synthesis[35], no perceptible improvements in cell fluorescence background were observed when supplementing the cultivation medium with BCAAs at a concentration of 0.5 mM (data not shown). For this reason, step-wise gene deletion of *ilvEDA* operon from the host strain was envisaged. While *ilvE* gene deletion yielded no further improvements, the additional genetic deletion of *ilvD* and *ilvA* resulted in a low fluctuation in fluorescence levels up to 4 h post-induction in the absence of DHB (**Figure 5.9**). Therefore, the host strain CF267 Δ*ilvEDA* was chosen to continue the studies aiming at demonstrating 3-HPA production from extracellularly added DHB.

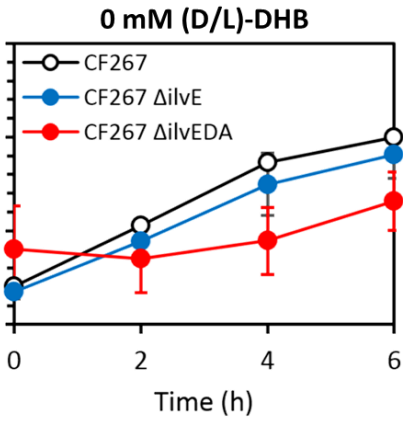

Figure 5.9. Engineering the host strain CF267 harboring pYqhC-20, pREP-22 and pZA33-kdcA plasmids towards decreased levels of intracellular branched-chain 2-keto acids yielded a strong reduction in background fluorescence in the absence of DHB. Cells were cultivated in M9 mineral medium containing 20 g L^{-1} glucose and IPTG (1 mM) was added when OD_{600} reached ~0.6 (t = 0 h), after which fluorescence was measured. SYFP2 fluorescence was calculated using cytometry data based on geometric mean. The reported values represent the mean ± S.D. ($n \geq 2$).

Aiming at demonstrating intracellular production of 3-HPA, the strain CF267 Δ*ilvEDA* harboring the metabolite sensor and pZA33-kdcA was cultivated in mineral medium and IPTG (1 mM) and a racemic mixture of (D/L)-DHB (20 mM) were added during cell exponential phase (OD ~0.6). Increased fluorescence levels were observed after 4h post-induction reaching its maximum at 6h post-induction time (**Figure 5.10**). Crucially, expression of the catalytically more efficient OHB decarboxylase Ll-KdcA V461I as assessed by *in vitro* studies (see **Chapter 4**) yielded higher fluorescence values when compared to the utilization of its wild-type counterpart (**Figure 5.10**). On the other side, transformation of the host strain with the empty pZA33-plasmid resulted in a stable but low fluorescence background. These results suggest therefore that under optimal conditions the metabolite sensor is able to detect 3-HPA released from the synthetic pathway and that in principle *E. coli* does not possess any enzymes with sufficient OHB decarboxylase activity from its endogenous metabolism. The additional discrimination between enzyme variants possessing variable OHB decarboxylase activities further suggests the applicability of the metabolite sensor towards directed evolution campaigns.

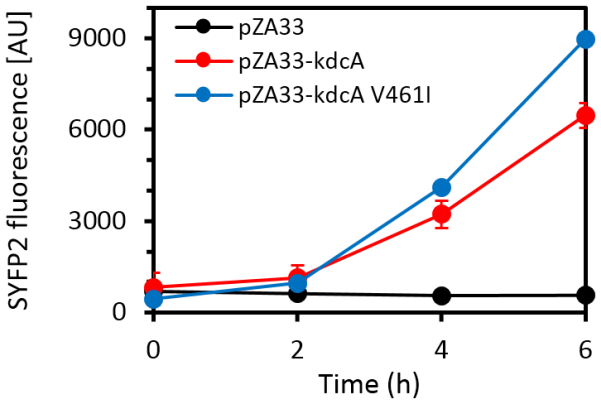

Figure 5.10. Monitoring of *in vivo* production of 3-hydropropanal (3-HPA) by an artificial metabolic pathway using the aldehyde metabolite sensor. The *E. coli* host strain CF267 Δilv*EDA* (MG1655 Δ*sad* Δ*yqhD* $lldD^{proD}$ $glcA^{proD}$ $lacI^{Q}$::Sp Δilv*EDA*) was co-transformed with pYqhC-20, pREP-22 and various pZA33-derived plasmids. Cells were cultivated in M9 mineral medium containing 20 g L^{-1} glucose and IPTG (1 mM) and 20 mM (D/L)-DHB was added when OD_{600} reached ~0.6, after which fluorescence was measured. SYFP2 fluorescence was calculated using cytometry data based on geometric mean. The reported values represent the mean ± S.D. ($n \geq 2$).

5.6 Discussion

The success in improving individual enzymes and/or entire metabolic pathways by directed evolution is often hindered by the lack of high-throughput screening/selection methods enabling detection of target phenotypes[11]. Specific detection of small molecules is often required when large mutant libraries are screened for over-producing strains or improved enzyme activities. In the present study, the design and implementation of a metabolite sensor system for the detection of various aldehydes in *E. coli* was demonstrated. While other aldehyde-sensing systems have previously been developed on the basis of invasive and indirect sampling methods[36], the herein developed system allows for direct aldehyde detection and may additionally enable FACS-based high-throughput screening and selection for aldehyde-producing enzymes or strains.

In this work, YqhC served as the transcriptional regulator to detect the presence of an aldehyde and to induce expression of a fluorescent marker protein. Turner and colleagues previously constructed a YqhC-based metabolite sensor in *E. coli*, in which the firefly luciferase-encoding gene was put under transcriptional control of putative *yqhD* promoter region in a low-copy number plasmid[27]. Albeit the genomically expressed YqhC levels ensured a light response upon aldehyde exposure, the low measurable output signals observed from cell lysates were not compatible with high-throughput screening purposes, in which cell sorting is desired. In those cases, engineering poor ligand-transcription factor binding through protein or 5'-UTR engineering has previously been reported to dramatically

change the response profiles of metabolite sensors[37,38]. To this end, an alternative bi-modular aldehyde sensor based on YqhC was constructed. A sensing module drives constitutive expression of *yqhC* gene, while a reporter module drives expression of *syfp2* reporter gene under transcriptional control of *yqhD* promoter region. One of the major challenges when developing metabolite sensors is to optimize their sensitivity and dynamic range. Therefore, SYFP2 was selected as a reporter protein since it allows for optimized folding, maturation and superior brightness when compared to sfGFP and eYFP autofluorescent protein variants[29]. Fine-tuning of the expression levels of YqhC by engineering the promoter and RBS region of the reporter module further resulted in a 10-fold improvement over the initial sensor system. These results suggest the previously reported concept of two-module metabolite sensors[16,39] to be particularly useful when it comes to optimization tasks, since it enables to independently fine-tune each of its parts.

The specificity and dynamic range responses of the constructed biosensor were evaluated with respect to a spectrum of aldehydes, including short, medium carbon chain and aromatic compounds. At variance to Turner and colleagues[27] who reported a YqhC-dependent transcriptional activation of the firefly-luciferase encoding reporter gene to ethanal, propanal, butanal, methylglyoxal and lignocellulose inhibitors (furfural, cinnamaldehyde), the biosensor was almost insensitive to some of these compounds, but responded to others such as glycolaldehyde, vanillin and phenylacetaldehyde. The discrepancy between the results herein achieved and those from Turner and colleagues can be explained in part by the difference in the genomic design of the constructed sensor. In this specific case, a bi-modular sensor was employed in which the *yqhC* gene is used out of its natural genomic context and expressed constitutively from a low-copy plasmid, while the reporter module bears the *syfp2* gene preceded by a strong RBS and flanked upstream by a promoter region that included 150-nt of *yqhC* coding sequence plus 108-nt of adjacent *yqhC/D* intergenic region. On the other side, a different reporter gene (leading to the expression of the firefly luciferase) immediately preceded by the *yqhD* promoter region was constructed by the Ingram group to investigate the response to aldehydes in a *E. coli* strain expressing YqhC from its genomic locus[27]. In addition, and whatsoever the difference between the two sensors, the structural diversity of the aldehydes that exhibit transcriptional induction raised questions about the mechanism by which YqhC may interact with these compounds. While the apparent lack of specificity can be regarded as a disadvantage, it can instead be

considered as an appropriate feature for the *in vivo* screening of synthetic pathways in which a single aldehyde is produced as an intermediate or end-product, and whose accumulation can be correlated to the catalytic efficiency of the upstream enzymatic reaction or can indicate a bottleneck in the downstream reaction in that pathway. This is nicely exemplified by the demonstration of glycolaldehyde detection produced by the synthetic X1P pathway and 3-HPA production via the artificial PDO pathway.

Another key feature determining biosensor performance is its dynamic range of detection. We found that the sensitivity of our biosensor to aldehyde was in the range of 1-10 mM. However, in spite of this elevated minimum concentration, the application of the sensor should be possible when aiming at detecting aldehyde over-producing strains or enzymes. In addition, this relatively poor sensitivity could be an advantage as it may avoid perturbation resulting from potential endogenous aldehydes, whose concentrations are actually relatively low due to the presence of many aldehyde reductases in *E. coli*[2]. Also, fluorescence saturation from the metabolite sensor at higher concentrations may constitute an advantage in metabolic engineering projects, in which high product titers are desired. Indeed, as a first proof-of-principle application, the feasibility of the system to monitor intracellular aldehyde production from a strain harboring the xylulose-1-phosphate pathway that can generate glycolaldehyde from xylose at a theoretical yield of 1 mol mol^{-1}. A remarkable increase of the fluorescence signal dependent on xylose consumption was observed. In a second proof-of-concept, the intracellular production of 3-HPA was monitored from a strain harboring a partial PDO synthetic pathway. The expression of OHB decarboxylases with distinct catalytic efficiencies matched well with the observed differences in fluorescence levels, thereby confirming the potential applicability of the developed high-throughput biosensor system towards metabolic and/or enzyme engineering projects upon minor optimization adjustments in each case scenario.

5.7 Materials and methods

5.7.1 Chemicals and reagents

All chemicals and solvents were purchased from Sigma-Aldrich unless otherwise stated. Restriction endonucleases and DNA-modifying enzymes were purchased from New England Biolabs and used according to instructions of the manufacturer. DNA plasmid isolation was performed using GeneJET Plasmid Miniprep Kit (Thermo Scientific). DNA extraction from agarose gel was carried out using the GeneJET Gel Extraction Kit (Thermo Scientific). DNA sequencing was carried out by Eurofins SAS (Ebersberg, Germany).

5.7.1 Plasmid construction

All plasmids and primers used in this study are listed in **Tables 5.3** and **5.4**, respectively.

*Construction of sensing modules***:** The upstream regions (including variable strength constitutive promoters and ribosome binding sites) were introduced into the forward primer that together with the reverse primer CF154 served to amplify the wild-type *yqhC* gene from genomic DNA of *E. coli* K-12 *substr.* MG1655 (ATCC 47076). Additionally, the primers used allowed the insertion of unique restriction sites upstream and downstream of the amplified fragments. Resulting PCR products and the low-copy vector backbone pZS23 (Expressys, Germany) were digested with *Xho*I and *Bam*HI restriction enzymes, gel purified and ligated with T4 DNA ligase. The resulting plasmids were transformed into DH5α competent *E. coli* cells (New England Biolabs) and inserts verified by DNA sequencing.

*Construction of reporter modules***:** The YqhD promoter region was PCR amplified from genomic DNA with primers pairs CF155/CF156, while the *syfp2* gene (synthesized by Eurofins) was amplified with primers CF157/CF158. DNA fragments were gel purified and assembled by homologous recombination with the *Xho*I/*Bam*HI-digested pZE13 vector using the NEBuilder® HiFi DNA Assembly kit (New England Biolabs). The resulting plasmid was named pREP-14. For construction of pREP-15 and pREP-22 vectors, promoter regions of *yqhD* gene were PCR-amplified from *E. coli* genomic DNA using respectively the primer pairs CF159/160 and CF155/324. The obtained PCR products and pREP-14 plasmid were digested (*Xho*I, *Hind*III), gel-purified and complementary-ends ligated. After sequencing analysis, the resulting plasmids were named pREP-15 and pREP-22, respectively.

*Construction of pZA33-khkC-aldoB***:** The khkC-aldoB operon was amplified by PCR from pEXT20-khkC-aldoB [4] using the primer pairs Pen268/Pen269, while the medium-copy vector pZA33 (Expressys, Germany) was PCR-linearized using primer pairs Pen321/Pen322. Resulting PCR products were gel purified and assembled by homologous recombination using the NEBuilder® HiFi DNA Assembly kit (New England Biolabs). The resulting plasmid was transformed into DH5α competent *E. coli* cells and the assembled operon verified by DNA sequencing.

Construction of pZA33-kdcA variants: The *kdcA* gene variants were amplified from pET-28 derived vectors using the primer pairs CF183/CF192. Resulting PCR products and the medium-copy vector pZA33 were digested with *Eco*RI and *Bam*HI restriction enzymes, gel purified and ligated with T4 DNA ligase. The resulting plasmids were transformed into DH5α competent *E. coli* cells (New England Biolabs) and inserts verified by DNA sequencing.

Table 5.3. Strains and plasmids used in this study.

Name	Description	Source
Strains	**Genotype**	
DH5α	*E. coli fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80Δ (lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	NEB
MG1655	F^- λ^- ilvG- rfb-50 rph-1	ATCC 47076
CF30	MG1655 Δ*sad* Δ*yqhD*	This work
Pen155	MG1655 Δ*xylB*	Cam et al., 2016[4]
CF272	Pen155 Δ*yqhD*	This work
CF267	CF30 $glcA^{proD}$ $lldD^{proD}$ $lacI^{Q}$::Sp	This work
	CF267 Δ*ilvE*	This work
	CF267 Δ*ilvEDA*	This work
Plasmids	**Relevant characteristics**	
pCP20	Amp^R, temperature-sensitive replicon, expressing FLP recombinase	Cherepanov et al., 1995[40]
pZE13	Amp^R; colE1 ori; promoter P_{A1}lacO1	Expressys
pZS23	Kan^R; pSC101 ori; promoter P_{A1}lacO1	Expressys
pZA33	Chm^R; p15A ori; promoter P_{A1}lacO1	Expressys
pEXT20-khkC-aldoB	AmpR; colE1 ori; promoter pTAC : *khkC* : *aldoB*	Cam et al., 2016[4]
pSENS-13	pZS23 derivative; promoter BBa_J23106 : RBS_{weak} : *yqhC*	This work
pSENS-16	pZS23 derivative; promoter BBa_J23114 : RBS_{weak} : *yqhC*	This work
pSENS-17	pZS23 derivative; promoter BBa_J23113 : RBS_{weak} : *yqhC*	This work
pSENS-18	pZS23 derivative; promoter proD : RBS_{weak} : *yqhC*	This work
pSENS-19	pZS23 derivative; promoter BBa_J23113 : $RBS_{very\ weak}$: *yqhC*	This work
pSENS-20	pZS23 derivative; promoter BBa_J23113 : RBS_{medium} : *yqhC*	This work
pSENS-21	pZS23 derivative; promoter BBa_J23113 : RBS_{strong} : *yqhC*	This work
pREP-14	pZE13 derivative; promoter P_{yqhD}: RBS_{01} : *syfp2*	This work
pREP-15	pZE13 derivative; promoter P_{yqhD} extended : RBS_{01} : *syfp2*	This work
pREP-22	pZE13 derivative; promoter P_{yqhD} hybrid : RBS_{01} : *syfp2*	This work
pZA33-khkC-aldoB	pZA33 derivative; promoter P_{A1}lacO1 : *khkC* : *aldoB*	This work
pZA33-kdcA	pZA33 derivative; promoter P_{A1}lacO1 : *Ll-kdcA*	This work
pZA33-$kdcA_{V461}$	pZA33 derivative; promoter P_{A1}lacO1 : *Ll-kdcA*$_{V461I}$	This work

Table 5.4. Primers used in this study for plasmid construction and strain validation.

Primer	Sequence (5' – 3')
Construction of sensing modules (pSENS)	
CF149	tatataGtcgagtttacggctagctcagtcctaggtatagtgctagcGGTCCACCGCTTACCCCCCCA AGGGACGAATAAA**atg**CTACAAAATTGCGCACA
CF154	tgcttaggatcCttaATTCCCCTGCATCG
CF161	taagcactcgagtttatggctagctcagtcctaggtacaatgctagcGGTCCACCGCTTACCCCCCCA AGGGACGAATAAA**atg**CTACAAAATTGCGCACA
CF162	taagcactcgagctgatggctagctcagtcctagggattatgctagcGGTCCACCGCTTACCCCCCCA AGGGACGAATAAA**atg**CTACAAAATTGCGCACA
CF163	taagcactcgagCACAGCTAACACCACGTC
CF164	TGTGCGCAATTTTGTAGcatTTTATTCGTCCCTTGGGGGGGTAAGCGGTGGACCAAA GTTAAACAAAATTATTTGTAGAGG
CF257	taagcactcgagtttatggctagctcagtcctaggtacaatgctagc**GTCTTAACAAAGGAAAAAAT TTACT**atgCTACAAAATTGCGCACA
CF258	taagcactcgagtttatggctagctcagtcctaggtacaatgctagc**AAATTTACTTATAAAGGAGG AGATAG**atgCTACAAAATTGCGCACA
CF259	taagcactcgagtttatggctagctcagtcctaggtacaatgctagc**TCGGAAGAAGAATCGAGGA GGAGGTATCA**atgCTACAAAATTGCGCACA
Construction of reporter modules (pREP)	
CF155	gaggccctttcgtcttcacctcgagttaCACATCGGGCAACAGTCC
CF156	gtatttaagttggaaagcttAGGGCAGAGAACGATCTG
CF157	tctctgccctaagctttccaacttaaatacaaggaaaataaggaggtcaacATGGTTAGCAAGGGCGA AG
CF158	gtacgcgtaccatgggatccTTATTATTTATACAGCTCATCCATACCC
CF159	TAAGCActcgagttaCACATCGGGCAACAGTC
CF160	tgcttaAAGCTTTTAAACTTGATCGAGAACGCC
CF324	tgcttaAAGCTTGGTCAGTGCGTCCTGCTGATGTGCTCAGTATCATCGCCAGCGCCCTG
Construction of xylulose-1-phosphate pathway	
Pen268	cggctgctaacaaagcccg
Pen269	gaattctgtgtgaaattgttatccgc
Pen321	**tttcacacagaattcGTTTAACTTTAA*GAAGGAGATATACC*ATG**GAAGAGAAGCAGATC CTGTGC
Pen322	**ctttgttagcagccgggatcctca**TTAATACGTGTAACAGGCCGTAAACAGA
Construction of pZA33-kdcA variants	
CF183	cggataacaatttcacacagaattcatcgacgctaaaacaaaaatataaggaggaacataATGTATACC GTTGGGGATTATC
CF192	cagtggtggtggtggtggtacgcgtaccatgggatccTTATTTGTTCTGTTCAGCAAAC
Strain construction	
glcA-proD_fw (CF253)	CCGAACCGTTATTACACGCCTGGCGTTTACGCGAAAAAGAAAGTCATTAAGTGTAG GCTGGAGCTGCTTC
glcA-proD_rv (CF254)	GCCCCAGTCCTCCCATCGGCATATACATTTGGGTCCAGGTAACcatATAATACCTCCT AAAGTTAAACAAAATTATTTGTAG
lldP-prod_fw (CF249)	CAATTCTCTGATGAGGATTGCCCTTTTCTTTACCAGACATCTCCCCCCACGTGTAGGC TGGAGCTGCTTC
lldP-prod_rv (CF250)	CCAGATATTCCCGGCGGGATCGTAGTTTTGTTGCCAGAGATTcatATAATACCTCCTA AAGTTAAACAAAATTATTTGTAG

Table 5.4. (continued)

Primer	Sequence (5' – 3')
Strain verification	
Δsad-ver-fw	CTGCCAGCTTCGGCAA
Δsad-ver-rv	GGGTAAAGTCGCGGATTAT
ΔyqhD-ver-fw (Pen15)	CAAGCGGCAAATCTCTTCAC
ΔyqhD-ver-rv (CF346)	TGGATTAGCCATACGTTCCT
glcA_ver_fw (CF255)	AATTTCGCTAACTCGTG
glcA_ver_rv (CF256)	ATGTCCTTTCAGACGTAATA
lldP_ver_fw (CF251)	ATTCTGCACATTCCTATAGG
lldP_ver_rv (CF252)	CATTTTATAGAACAGCAAAG
LacIQ_fw (CL20)	CCAATCAGCAACGACTGTTT
SpecR_rv (CL21)	CGTACATTTGTACGGCTCC

5.7.2 Strain construction and growth conditions

E. coli K-12 substr. MG1655 (ATCC 47076) was used as the parental strain for all constructions in this study. Single gene deletion (*yqhD, sad, ilvE)* and constitutive expression of $lacI^Q$ was achieved using the phage transduction method adapted from Miller[41]. Expression of *glcA* and *lldD* genes was rendered constitutive by replacing the native chromosomal 5'-UTR of each gene by the synthetic constitutive and insulated promoter proD. The proD sequence was preceded by a kan resistance cassette which was amplified by PCR adding 50-bp flanking sequences that were homologous to the target locus. The resulting DNA fragment was used to replace the natural gene promoter by homologous recombination[31]. Primers used are listed in **Table 5.4**. Positive clones were selected on LB agar plates containing either kanamycin (50 µg mL^{-1}) or spectinomycin (50 µg mL^{-1}) and verified by PCR analysis. The kan cassette was removed from the genome by expressing FLP recombinase from the pCP20 plasmid[40] and correct excision of the cassette was verified by PCR using locus specific primers (**Table 5.4**). Plasmids were transformed into the target *E. coli* strains using standard protocols[42].

The cultures of *E. coli* strains were carried out at 37 ºC on a rotary shaker running at 200 rpm in a M9 mineral medium which, unless otherwise stated, contained per liter: 20 g glucose, 18 g $Na_2HPO_4*12H_2O$, 3 g KH_2PO_4, 0.5 g NaCl, 2 g NH_4Cl, 0.5 g $MgSO_4*7H_2O$, 0.015 $CaCl_2*2H_2O$, 1 ml of 0.06 M $FeCl_3$ stock solution prepared in 100 times diluted concentrated HCl, 2 ml of 10 mM thiamine HCl stock solution, 20 g MOPS, and 1 ml of trace element solution (containing per liter: 0.04 g Na2EDTA*2H2O, 0.18 g $CoCl_2*6H_2O$, $ZnSO_4*7H_2O$, 0.04 g $Na_2MoO_4*2H_2O$, 0.01 g H_3BO_3, 0.12 g $MnSO_4$*H2O, 0.12 g $CuCl_2*H_2O$). The pH was adjusted to 7 and the medium was filter-sterilized. The antibiotics ampicillin and kanamycin sulfate were added when required at concentrations of 100 mg L^{-1} and 50 mg L^{-1}, respectively.

For assaying *in vivo* production of glycolaldehyde by the *E. coli* strain CF272 harboring the xylulose-1-phosphate pathway and metabolite sensor, the experiments were carried out as follows. Pre-cultures were grown in 10 mL M9 mineral medium in the presence of the antibiotics ampicillin, kanamycin sulphate and chloramphenicol at 100, 50 and 35 mg mL^{-1}. After an overnight incubation, cells were spun down by centrifugation (4000 rpm, 10 min at 4 ºC) and washed with sterile water. They were resuspended at an initial OD_{600} of 0.5 in 25 mL of fresh M9 mineral medium containing 1 mM IPTG and appropriate antibiotics and

in the presence of glucose or a mixture of glucose / xylose. Cell growth was followed by monitoring OD_{600}, and samples were regularly withdrawn for flow cytometry analyses and sugars consumption.

For assaying *in vivo* production of 3-HPA by various *E. coli* host strains harboring the metabolite sensor and pZA33-derived plasmids, the experiments were carried out as follows. Pre-cultures were grown in 5 mL M9 mineral medium in the presence of the antibiotics ampicillin, kanamycin sulphate and chloramphenicol at 100, 50 and 35 mg mL^{-1}. After an overnight incubation, cells were added at an initial OD_{600} of 0.2 in 10 mL of fresh M9 mineral medium containing 20 g L^{-1} glucose and appropriate antibiotics. When OD_{600} reached ~0.6, IPTG (1 mM) and a racemic mixture of (D/L)-DHB (20 mM) were added to the cultivation medium. Samples were regularly withdrawn for flow cytometry analyses.

5.7.3 Microtiter plate screening system

Screening of biosensor-strains: Pre-cultures were grown overnight in 5 mL of M9 mineral medium (37 ºC, 200 rpm). They were used to inoculate 10 mL of M9 mineral medium supplemented with the appropriate antibiotics starting at an initial OD_{600} of 0.2 in 50 mL falcon tube flasks placed on a rotary shaker set at 200 rpm and at 37°C. When OD_{600} reached ~0.6, 200 µL of cell culture were transferred in a 96-well plate and supplemented with glycolaldehyde at a final concentration of 5 mM. During the induction phase, microplates were incubated at 37 ºC with an orbital frequency at 807 rpm (Epoch 2, BioTek). After 4 h of incubation, single-cell fluorescence was measured by flow cytometry.

*Aldehyde dose-response curves***:** Pre-cultures were grown overnight in 5 mL of M9 mineral medium (37 ºC, 200 rpm). The biomass needed to start main cultures with a starting OD_{600} of 0.2 was transferred to 250 non-baffled shake flasks containing 25 mL of M9 mineral medium with appropriate antibiotics (37 ºC, 200 rpm). When OD_{600} reached ~0.6, 200 µL of cell culture were inoculated in a 96-well plate and supplemented with aldehydes at the desired concentrations. During the induction phase, microplates were incubated at 37 ºC with an orbital frequency at 807 rpm (Epoch 2, BioTek). After 12 h of incubation, single-cell fluorescence was measured by flow cytometry.

5.7.4 Flow cytometry

Flow cytometry measurements were performed with an Attune™ Acoustic Flow Cytometer (Life Technologies) with 488 nm excitation. Forward-scatter characteristics (FSC) and side-scatter characteristics (SSC) were detected as small-angle and large-angle scatters of the 488 nm laser, respectively. SYFP2 fluorescence was detected using a 530/30 nm (channel BL1) band-pass filter set. Data were analyzed using the Attune™ software (Life Technologies). A total of 100,000 events was recorded per sample, and electronic gating was applied on the densest subset of cells on the basis of forward- versus side-scatter height. The same gate was used to estimate geometric mean levels of SYFP2 fluorescence.

5.7.5 Analytical methods

The concentrations of glucose and xylose were determined on a Dionex Ultimate 3,000 HPLC system (Thermo Scientific, France) equipped with a RI detector (RID-10A, Shimadzu, Japan). The sample injection volume was 20 µL, and the compounds were separated in an Aminex HPX-87H column protected by a Micro-Guard Cation H+ pre-column (BioRad, USA). The separation was performed at 35 ºC with 1.25 mM H_2SO_4 at 0.5 mL min^{-1} as mobile phase. All samples were centrifuged (2 min at 13,000 rpm) and syringe-filtered (0.2 µm), and the resulting supernatant kept at -20 ºC until analysis.

5.7.6 Statistical methods

All statistical analyses were conducted in Microsoft Excel® using the Analysis ToolPak package. A two-tailed unpaired *t* test was to use to compare fluorescence induction levels, in which an alpha level of $p < 0.05$ was set for significance.

5.8 References

(1) Kunjapur, A. M., Tarasova, Y., and Prather, K. L. J. (2014) Synthesis and accumulation of aromatic aldehydes in an engineered strain of *Escherichia coli*. *J. Am. Chem. Soc. 136*, 11644–11654.

(2) Rodriguez, G. M., and Atsumi, S. (2014) Toward aldehyde and alkane production by removing aldehyde reductase activity in *Escherichia coli*. *Metab. Eng. 25*, 227–37.

(3) Kunjapur, A. M., and Prather, K. L. J. (2015) Microbial engineering for aldehyde synthesis. *Appl. Environ. Microbiol. 81*, 1892–901.

(4) Cam, Y., Alkim, C., Trichez, D., Trebosc, V., Vax, A., Bartolo, F., Besse, P., François, J. M., and Walther, T. (2016) Engineering of a synthetic metabolic pathway for the assimilation of (D)-xylose into value-added chemicals. *ACS Synth. Biol. 5*, 607–618.

(5) Akhtar, M. K., Turner, N. J., and Jones, P. R. (2013) Carboxylic acid reductase is a versatile enzyme for the conversion of fatty acids into fuels and chemical commodities. *Proc. Natl. Acad. Sci. 110*, 87–92.

(6) Walther, T., Topham, C. M., Irague, R., Auriol, C., Baylac, A., Cordier, H., Dressaire, C., Lozano-Huguet, L., Tarrat, N., Martineau, N., Stodel, M., Malbert, Y., Maestracci, M., Huet, R., André, I., Remaud-Siméon, M., and François, J. M. (2017) Construction of a synthetic metabolic pathway for biosynthesis of the non-natural methionine precursor 2,4-dihydroxybutyric acid. *Nat. Commun. 8*, 15828.

(7) Walther, T., and Francois, J. M. (2014) A microorganism modified for production of 1,3-propanediol. Patent WO/2014/009432A3.

(8) Erb, T. J., Jones, P. R., and Bar-Even, A. (2017) Synthetic metabolism: metabolic engineering meets enzyme design. *Curr. Opin. Chem. Biol. 37*, 56–62.

(9) Mori, Y., and Shirai, T. (2018) Designing artificial metabolic pathways, construction of target enzymes, and analysis of their function. *Curr. Opin. Biotechnol. 54*, 41–44.

(10) Berdyshev, E. V. (2011) Mass spectrometry of fatty aldehydes. *Biochim. Biophys. Acta 1811*, 680–693.

(11) Cheng, F., Tang, X.-L., and Kardashliev, T. (2018) Transcription factor-based biosensors in high-throughput screening: advances and applications. *Biotechnol. J. 13*.

(12) Mahr, R., and Frunzke, J. (2016) Transcription factor-based biosensors in

biotechnology: current state and future prospects. *Appl. Microbiol. Biotechnol. 100*, 79-90.

(13) Browning, D. F., and Busby, S. J. W. (2004) The regulation of bacterial transcription initiation. *Nat. Rev. Microbiol. 2*, 57–65.

(14) Dietrich, J. A., McKee, A. E., and Keasling, J. D. (2010) High-throughput metabolic engineering: advances in small-molecule screening and selection. *Annu. Rev. Biochem. 79*, 563–590.

(15) Siedler, S., Schendzielorz, G., Binder, S., Eggeling, L., Bringer, S., and Bott, M. (2014) SoxR as a single-cell biosensor for NADPH-consuming enzymes in *Escherichia coli*. *ACS Synth. Biol. 3*, 41–47.

(16) Li, H., Liang, C., Chen, W., Jin, J.-M., Tang, S.-Y., and Tao, Y. (2017) Monitoring in vivo metabolic flux with a designed whole-cell metabolite biosensor of shikimic acid. *Biosens. Bioelectron. 98*, 457–465.

(17) Kasey, C. M., Zerrad, M., Li, Y., Cropp, T. A., and Williams, G. J. (2018) Development of transcription factor-based designer macrolide biosensors for metabolic engineering and synthetic biology. *ACS Synth. Biol. 7*, 227–239.

(18) Chong, H., and Ching, C. B. (2016) Development of colorimetric-based whole-cell biosensor for organophosphorus compounds by engineering transcription regulator DmpR. *ACS Synth. Biol. 5*, 1290–1298.

(19) Morgan, S.-A., Nadler, D. C., Yokoo, R., and Savage, D. F. (2016) Biofuel metabolic engineering with biosensors. *Curr. Opin. Chem. Biol. 35*, 150–158.

(20) Rogers, J. K., Taylor, N. D., and Church, G. M. (2016) Biosensor-based engineering of biosynthetic pathways. *Curr. Opin. Biotechnol. 42*, 84–91.

(21) Liu, Y., Liu, Y., and Wang, M. (2017) Design, Optimization and application of small molecule biosensor in metabolic engineering. *Front. Microbiol. 8*, 2012.

(22) Cheng, F., Kardashliev, T., Pitzler, C., Shehzad, A., Lue, H., Bernhagen, J., Zhu, L., and Schwaneberg, U. (2015) A competitive flow cytometry screening system for directed evolution of therapeutic enzyme. *ACS Synth. Biol. 4*, 768–775.

(23) Kwon, K. K., Lee, D.-H., Kim, S. J., Choi, S.-L., Rha, E., Yeom, S.-J., Subhadra, B., Lee, J., Jeong, K. J., and Lee, S.-G. (2018) Evolution of enzymes with new specificity

by high-throughput screening using DmpR-based genetic circuits and multiple flow cytometry rounds. *Sci. Rep. 8*, 2659.

(24) Schallmey, M., Frunzke, J., and Eggeling, L. (2014) Looking for the pick of the bunch: high-throughput screening of producing microorganisms with biosensors. *Curr. Opin. Biotechnol. 26*, 148–154.

(25) Jarboe, L. R. (2011) YqhD: a broad-substrate range aldehyde reductase with various applications in production of biorenewable fuels and chemicals. *Appl. Microbiol. Biotechnol. 89*, 249–257.

(26) Lee, C., Kim, I., Lee, J., Lee, K.-L., Min, B., and Park, C. (2010) Transcriptional activation of the aldehyde reductase YqhD by YqhC and its implication in glyoxal metabolism of *Escherichia coli* K-12. *J. Bacteriol. 192*, 4205–14.

(27) Turner, P. C., Miller, E. N., Jarboe, L. R., Baggett, C. L., Shanmugam, K. T., and Ingram, L. O. (2011) YqhC regulates transcription of the adjacent *Escherichia coli* genes *yqhD* and *dkgA* that are involved in furfural tolerance. *J. Ind. Microbiol. Biotechnol. 38*, 431–439.

(28) Keseler, I. M., Mackie, A., Santos-Zavaleta, A., Billington, R., Bonavides-Martínez, C., Caspi, R., Fulcher, C., Gama-Castro, S., Kothari, A., Krummenacker, M., Latendresse, M., Muñiz-Rascado, L., Ong, Q., Paley, S., Peralta-Gil, M., Subhraveti, P., Velázquez-Ramírez, D. A., Weaver, D., Collado-Vides, J., Paulsen, I., and Karp, P. D. (2017) The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Res. 45*, D543–D550.

(29) Kremers, G.-J., Goedhart, J., van Munster, E. B., and Gadella, T. W. J. (2006) Cyan and yellow super fluorescent proteins with improved brightness, protein folding, and FRET Förster radius. *Biochemistry 45*, 6570–6580.

(30) Salis, H. M., Mirsky, E. A., and Voigt, C. A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol. 27*, 946–950.

(31) Davis, J. H., Rubin, A. J., and Sauer, R. T. (2011) Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res. 39*, 1131–1141.

(32) Hook-Barnard, I., Johnson, X. B., and Hinton, D. M. (2006) *Escherichia coli* RNA polymerase recognition of a sigma70-dependent promoter requiring a -35 DNA

element and an extended -10 TGn motif. *J. Bacteriol. 188*, 8352–9.

(33) Smit, B. A., van Hylckama Vlieg, J. E. T., Engels, W. J. M., Meijer, L., Wouters, J. T. M., and Smit, G. (2005) Identification, cloning, and characterization of a *Lactococcus lactis* branched-chain alpha-keto acid decarboxylase involved in flavor formation. *Appl. Environ. Microbiol. 71*, 303–11.

(34) Dietrich, J. A., Shis, D. L., Alikhani, A., and Keasling, J. D. (2013) Transcription factor-based screens and synthetic selections for microbial small-molecule biosynthesis. *ACS Synth. Biol. 2*, 47–58.

(35) Park, J. H., and Lee, S. Y. (2010) Fermentative production of branched chain amino acids: a focus on metabolic engineering. *Appl. Microbiol. Biotechnol. 85*, 491–506.

(36) Pariente, F., Lorenzo, E., Tobalina, F., and Abruna, H. D. (1995) Aldehyde biosensor based on the determination of NADH enzymically generated by aldehyde dehydrogenase. *Anal. Chem. 67*, 3936–3944.

(37) Kim, M. N., Park, H. H., Lim, W. K., and Shin, H. J. (2005) Construction and comparison of *Escherichia coli* whole-cell biosensors capable of detecting aromatic compounds. *J. Microbiol. Methods 60*, 235–245.

(38) Becskei, A., Kaufmann, B. B., and Van Oudenaarden, A. (2005) Contributions of low molecule number and chromosomal positioning to stochastic gene expression. *Nat. Genet.* 37, 937–944.

(39) Xue, H., Shi, H., Yu, Z., He, S., Liu, S., Hou, Y., Pan, X., Wang, H., Zheng, P., Cui, C., Viets, H., Liang, J., Zhang, Y., Chen, S., Zhang, H. M., and Ouyang, Q. (2014) Design, construction, and characterization of a set of biosensors for aromatic compounds. *ACS Synth. Biol. 3*, 1011–1014.

(40) Cherepanov, P. P., and Wackernagel, W. (1995) Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene 158*, 9–14.

(41) Miller, J. H. (1992) A short course in bacterial genetics : a laboratory manual and handbook for *Escherichia coli* and related bacteria. Cold Spring Harbor Laboratory Press.

(42) Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press.

Chapter 6. Conclusions and future directions

As discussed throughout this work, a central goal of this thesis has been the construction of synthetic metabolic pathways for the synthesis of two industrially relevant commodity chemicals: 1,3-propanediol and (L)-2,4-dihydroxybutyrate. In this context, rational enzyme engineering was applied towards the generation of enzyme activities necessary for pathway operation, while the construction of an aldehyde-metabolite sensor may contribute for further improvements of the established non-natural routes. This thesis is likely to mirror the challenges in creation of metabolic pathways for which required biocatalysts are not known to exist in nature.

6.1 A set of pathways with DHB as key intermediate molecule

While (L)-2,4-dihydroxybutyrate is a non-natural metabolite, recent efforts have proved synthetic pathway engineering to enable its biosynthesis in *E. coli*. In particular, this work outlined enzyme engineering as a mean towards optimization of a metabolic route yielding DHB as final product via homoserine (see **Chapter 3**). An alternative three-reaction step synthetic route aiming at DHB biosynthesis from the Krebs cycle intermediate malate was previously developed by our group. Extension of the malate DHB pathway with three additional steps further allowed conversion of DHB into PDO (see **Chapter 4**). This work contributed therefore to the optimization of metabolic pathways in which DHB can be seen as a key player molecule (**Figure 6.1**).

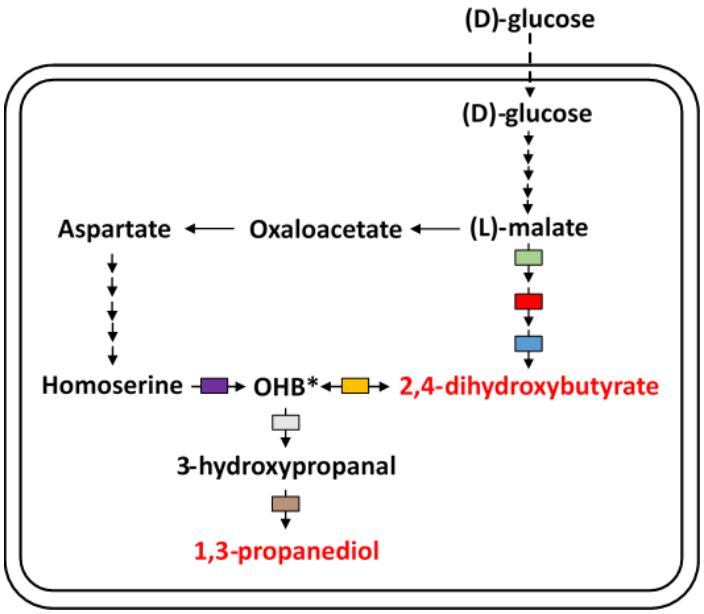

Figure 6.1. Synthetic metabolic pathways with 2,4-dihydroxybutyrate (DHB) as key (intermediate) molecule. Each rectangle represents an engineered enzyme. Legend: OHB, 2-keto-4-hydroxybutyrate.

6.2 Engineering DHB pathway via homoserine

Construction of a catalytically efficient functional OHB reductase which could be employed as part of the DHB pathway via homoserine proved to be a difficult and labor-intensive task. As detailed in **Chapter 3**, sequence comparative analysis and computationally-aided protein design permitted the identification of amino acid substitutions likely to introduce OHB reductase activity into the template enzyme. Characterized by possessing a very strong affinity towards the reduction of oxaloacetate, substitution of Arg81 by an alanine residue in Ec-Mdh yielded an enzyme variant with 287-fold improved activity on OHB substrate. This substitution in particular has previously been reported as essential towards a shift in Ec-Mdh activity in dicarboxylic acids. But no saturation was observed under the tested range of OHB concentrations. Only when amino acid substitutions were simultaneously introduced into three different regions of the protein, highly active variants with low K_m values on OHB were observed. However, the increased OHB reductase activity as observed in most-promising variants was accompanied by an increase in activity on the natural oxaloacetate substrate, yielding variants with catalytic efficiencies comparable between both substrates.

To demonstrate the applicability of the best mutant enzymes for *in vivo* applications, each of those was individually expressed as part of the DHB pathway via homoserine in *E. coli*. Shake-flask experiments resulted in increased DHB titers and yields of up to two-fold when compared with the previously reported OHB reductase variant Ll-LdhA Q85C. The simultaneous expression of the best-performing OHB reductase Ec-Mdh-5Q variant (I12V:R81A:M85Q:G179D:D86S) and the Ec-AlaC A142P:Y275D homoserine transaminase variant in the homoserine-overproducing ECO4 parent strain, together with carbon flux redirection allowed further improvements in the designed synthetic pathway. Cell cultivation under well-controlled conditions in fed-batch mode further resulted in the production of up to 10.7 g L^{-1} DHB. Whilst it corresponds to the highest titer reported to date, the concomitant accumulation of high amounts of homoserine precursor and alanine anticipate HMS transaminase as a major rate-limiting step of the pathway. As a part of a competitive project with an industrial partner, several strategies are in current development by our team in an attempt to overcome this bottleneck, which could not be solved in this thesis.

6.3 Construction of a PDO pathway via DHB

Product derivatization (e.g. DHB) into other compounds of industrial interest (e.g. PDO, 1,2,4-butanetriol) is notably important in the context of biorefinery, thereby allowing for a flexible commercialization of a range of related but diversified products according to market needs.

In this scope, a synthetic metabolic pathway enabling direct biosynthesis of 1,3-propanediol (PDO) from glucose via the Krebs cycle intermediate malate was conceived. The proposed route extends the previously published pathway for the synthesis of DHB with three additional reaction steps catalyzed by DHB dehydrogenase, OHB decarboxylase and PDO oxidoreductase. Screening and structure-guided protein engineering provided a (L)-DHB dehydrogenase from the membrane-associated (L)-lactate dehydrogenase from *E. coli* and OHB decarboxylase variants derived from either the branched-chain ketoacid decarboxylase encoded by *kdcA* from *L. lactis* or pyruvate decarboxylase from *Z. mobilis*, whose simultaneous overexpression with the *ydhD*-encoded aldehyde reductase enabled PDO biosynthesis from DHB. Simultaneous expression of the six enzymatic activities in a single

E. coli strain resulted however in a low production of 0.1 mM PDO from 110 mM glucose. A metabolic burden and the occurrence of bottlenecks in the pathway may be behind the observed weak titers. However, a co-cultivation strategy in which an *E. coli* strain expressing the DHB-yielding pathway and another strain bearing the plasmid with the DHB-to-PDO pathway resulted in increased PDO titers up to 3.4 mM. While co-cultivation of two *E. coli* strains proved effective in increasing product formation, this strategy is less desired for industrial applications.

6.4 A metabolite sensor for high-throughput aldehyde detection

In this work, a fluorescence-based metabolite sensor enabling *in vivo* detection of various aldehydes of biotechnological interest in *E. coli* was developed. By assuming that a native transcription factor promoter pair which is capable to detect any small molecules can be identified in Nature, the *E. coli* YqhC-P_{yqhD} system was selected in an attempt to demonstrate its potential application as high-throughput screening and selection devices. YqhC is a transcriptional regulator that is known to be involved in the upregulation of the *yqhD-dgkA* operon in the presence of aldehydes. By taking advantage of this property, a bi-modular biosensor was constructed, in which a sensing module constitutively expresses *yqhC* while a reporter module drives the expression of the *syfp2* reporter gene that is put under control of the *yqhD* promoter.

The sensitivity of the sensor has been optimized by engineering the 5'-untranslated region of both the sensing and reporter modules, resulting in a 70-fold gain of fluorescence in response to the model compound glycolaldehyde at 5 mM. The optimized sensor further responded to other aldehydes when supplemented to the cultivation medium at concentrations of 1-10 mM.

Additional studies further demonstrated that this metabolite sensor was functional *in vivo* upon expression of two in house synthetic pathways. In particular, it responded to the presence of glycolaldehyde that is specifically produced upon induction of a synthetic xylulose-1-phosphate pathway expressed in *E. coli*, as well as to 3-HPA produced from a fraction of the PDO pathway via DHB. This bi-modular sensor has been shown to constitute a potential tool for FACS-based ultra-high throughput screening of aldehyde (over-) producing enzymes.

6.5 Future work

As previously outlined, this thesis contributed to the establishment of a set of synthetic pathways in which DHB can be seen as a central metabolite. In particular, production of DHB was accomplished from homoserine or (L)-malate, while extension of the malate-to-DHB pathway allowed for PDO biosynthesis. But this work further opens the possibility to produce PDO from homoserine if the herein identified HMS transaminase, OHB decarboxylase and PDO oxidoreductase enzymes are expressed in a homoserine-overproducing strain (**Figure 6.1**). Interestingly, PDO synthesis from homoserine has the advantage of only requiring three additional reaction steps, as opposed to its formation from malate via DHB. But in both cases further enzyme engineering is likely critical so that proposed routes can operate at efficient rates.

In particular, the designed pathways as assessed in **Chapters 3 and 4** were shown to possess several botlenecks including rate-limiting steps (e.g. HMS transaminase, OHB decarboxylase), which resulted in the accumulation of intermediate molecules. In line with this information, the metabolite sensor as developed in **Chapter 5** can be applied in combination with FACS technology in high-throughput screening studies aiming at the selection of randomly-mutated enzyme variants with higher catalytic efficiencies, as far as 3-HPA can be generated as end-product from a single enzyme or combination of multiple enzymes (as shown in **Figure 6.1**) with minor adjustments of the metabolite sensor. The metabolite sensor has, additionaly, the advantage of being well-suited for alternative metabolic engineering projects, in which randomly mutated strains expressing pathways yielding 3-HPA can be detected and selected.

While identification of highly-efficient enzymes is crucial in the construction of metabolic pathways, it only comprises one step in the long process of metabolic engineering. A more comprehensive analysis of cell metabolism can identify further bottlenecks, which can be addressed by host strain engineering. Also, process engineering and optimization of conditions of cell cultivation in well-controlled bioreactors will allow to reach higher productivities and titers.

"Valeu a pena? Tudo vale a pena
Se a alma não é pequena.
Quem quer passar além do Bojador
Tem que passar além da dor.
Deus ao mar o perigo e o abismo deu,
Mas nele é que espelhou o céu."
—Fernando Pessoa

The End