A broadly applicable artificial selection system for biomolecule evolution

Lara Sellés Vidal

Submitted for the Degree of Doctor of Philosophy

Supervised by Dr. John T. Heap
Imperial College London
Department of Life Sciences
The Imperial College Centre for Synthetic Biology (IC-CSynB)

October 2019

Abstract

Biocatalysis offers an attractive alternative to traditional chemical catalysis. However, it is often found that an enzyme with the optimal properties for a specific application is not available within the natural repertoire of enzymes. It is then desirable to obtain an improved variant by altering the sequence of a known enzyme, in a process known as protein engineering.

Directed evolution is one of the most powerful tools for protein engineering. In directed evolution, the process of natural evolution is mimicked in the laboratory at a much shorter timescale and selecting for properties that make the enzyme (or any other type of biomolecule) more suitable for an application of human interest. The main bottleneck of directed evolution is the identification of the desired variants amongst a majority of variants without the sought altered or improved property. Selection approaches link the desired activity to an increased survival rate or improved growth. While in principle such methodologies allow for ultra high-throughput analysis of libraries, most selection techniques have a limited scope, and can only be applied to a relatively reduced set of biomolecules or properties.

This thesis presents the most broadly-applicable artificial selection system for the evolution of biomolecules ever reported. The selection platform is based on an engineered *E. coli* strain with impaired regeneration of NAD^+, causing a conditional growth defect during anaerobic fermentation. By directly or indirectly linking the activity of the biomolecules of interest to the oxidation of NADH, cells can be rescued from this growth defect. The efficacy of such selection system has been demonstrated by using it to select alcohol dehydrogenase, imine reductase and nitroreductase variants with altered or enhanced catalytic properties, as well as an isopropanol-producing metabolic pathway with optimised regulatory elements leading to a maximised yield of isopropanol.

These results confirm the wide scope of the developed selection system, which can replace conventional screening currently used in many cases of direct relevance for industrial processes. Increasing the throughput of the variant search process by many orders of magnitude will lead to the discovery of novel biomolecules and accelerate the implementation of biocatalysis.

Declaration

I hereby certify that all the data included in this thesis is my own work, unless stated otherwise.

Lara Sellés Vidal

Acknowledgments

I always thought that saying thank you would be the easiest part to write of my thesis, but it turned out to be quite complicated and very hard to compress it in a few words.

I would like to express my special appreciation and gratitude to my supervisor, Dr. John T. Heap, for giving me the opportunity to work in his lab and guiding me throughout the project. Also, a big thank you to Dr. James Murray for teaching me how a crystal should look like and for fishing the valuable crystals for me. And to BBSRC and Syngenta for funding the project for the past four years.

To all the members of the Imperial College Centre for Synthetic Biology, it has been amazing to work alongside you. You have always managed to make me smile even in my worst days. I would like to give a special thanks to my friends Joyce, Linda, Soo Mei for the constant support and fun in the lab and for the weekend trips around Europe; to Rochelle for always being by my side and feeding me with amazing cakes, to George for being the first test case PhD of our lab and breaking all our lighters, to Chiara for accompanying me late nights and weekends, to Nicolas for being Nicolas. And to Molly, Rhi, Ciarán and Pawel for all the chats and laughs.

Now, I would like to thank in Spanish my family and friends:
En primer lugar, me gustaría agradecer a mis padres por creer en mí y darme la educación que me ha traído hasta aquí. Gracias por todo el apoyo incondicional y consejos que me habéis dado y por haber estado siempre a mi lado a pesar de la distancia. Moltes gràcies Aleix per sempre estar al meu costat encara que durant aquests últims anys haguem estat visquent en països diferents. También me gustaría agradecer a mis abuelos por escucharme siempre que lo he necesitado y por enseñarme muchos de los valores de la vida.

Muchas gracias a todos aquellos que me habéis acompañado durante este largo viaje, ya sea en la lejanía: Silvia (“mi suplemento de queso”), Joris (“café amb nata”), Maria (“pillina”), Bego, Ana, Neus, Isart, Carla, Angel, Alba (“minion”), Miguel (“hay muchas mentes prodigiosas en esta mesa”), Andre y Andrea; como día a día: Silvia (“prou de peix”), Alicia (“crying in the corner”), Albert (“home!”), Joaquín (“by the glory of God”), Rodrigo (“chocolate”), Javi (“la vida golpea”), Ismael (“que viene Cepeda”), and Ivet (“I am not Spanish!”) y que no os habéis cansado de mis infinitos “no tengo tiempo”, “me encantaría, pero no puedo ir” y similares. Por compartir momentos buenos y no tan buenos, risas infinitas, discusiones científicas y no científicas, viajes, coffee breaks…

Por último, pero no menos importante, me gustaría agradecer especialmente a Rafa por todo el apoyo y ayuda constante que he tenido durante estos últimos cuatro años (no sería la científica que soy hoy en día si no fuera por ti). Gracias también por arrancarme esas sonrisas en días largos y duros. Gracias por tu infinita paciencia y por estar siempre a mi lado.

Table of contents

1. Introduction 17

1.1. Background 17

1.2. Rational design 19

1.3. Directed evolution 21

1.4. Generation of a library of variants 23

1.4.1. Random mutagenesis 23

1.4.2. Recombination techniques 30

1.4.3. Rational mutagenesis 33

1.5. Identification of the variants of interest 37

1.5.1. Screening techniques 37

1.5.2. Selection techniques 41

1.6. Aims and objectives 48

2. Materials and Methods 50

2.1. Materials 50

2.2. Computational tools 58

2.3. Microbiological materials and methods 59

2.3.1. Antibiotic selection 59

2.3.2. Bacterial strains and aerobic culture conditions 59

2.3.3. Bacterial strains and anaerobic culture conditions 59

2.3.4. Preparation of electro-competent *E. coli* cells 61

2.3.5. Transformation of *E. coli* by electroporation 62

2.3.6. Construction of *E. coli* mutant strains 62

2.4. Molecular biological materials and methods 63

2.4.1. Oligonucleotide design 63

2.4.2. Storage of synthetic DNA sequences 67

2.4.3. Genomic DNA extraction 67

2.4.4. Amplification of DNA using polymerase chain reaction (PCR) 67

2.4.5. Construction of plasmids for enzyme evolution and characterisation 68

2.4.6. Construction of saturation libraries 70

2.4.7. Golden Gate assembly 71

2.4.8. Extraction and purification of plasmid DNA 71

2.4.9. Quantification of DNA 71

2.4.10. Restriction endonuclease digestion of DNA 71

2.4.11. Agarose gel electrophoresis and extraction of DNA fragments 72

2.4.12. Ligation of DNA 72
2.4.13. Construction of the library of isopropanol production pathways 72
2.4.14. Membrane dialysis of DNA 74
2.4.15. Verification of new genetic constructs 74
2.4.16. Library quality control 74
2.5. Biochemical materials and methods 74
2.5.1. Sodium dodecyl sulfate poly-acrylamide gel electrophoresis 74
2.5.2. Nickel-affinity chromatography purification of histidine-tagged proteins 75
2.5.3. Quantification of protein 75
2.6. Experimental assays 76
2.6.1. Anaerobic growth complementation experiments 76
2.6.2. Toxicity assays 76
2.6.3. Enzymatic assays 76
2.6.4. Crystallization 78
2.6.5. Structure determination 78
2.6.6. Structural modelling 79
2.6.7. Acquisition and processing of 1H -NMR spectra of fermentation broths 79
3. Redox rescue strain: construction and validation 81
3.1. Introduction 81
3.2. Results and discussion 86
3.2.1. Construction and characterisation of the redox rescue *E. coli* strain 86
3.2.2. Validation of the redox rescue system with an endogenous enzyme 87
3.2.3. Validation of the redox rescue system with exogenous reductases 89
3.3. Conclusions 94
3.4. Key outcomes 95
4. Alcohol dehydrogenases: altering cofactor specificity 96
4.1. Introduction 96
4.2. Results and discussion 100
4.2.1. Selection of a novel NAD-dependent CBADH variant 100
4.2.2. Characterization of the selected NAD-dependent CBADH variant 104
4.2.3. Construction of a strict selection system for NAD-dependent activity 106
4.2.4. Selection of NAD-dependent variants with the ALPS strain 112
4.2.5. Characterization of the NAD-dependent TBADH variants 115
4.2.6. Structural insights into cofactor preference reversal 119
4.3. Conclusions 124
4.4. Key outcomes 125

5. Engineering an imine reductase with altered cofactor specificity 126
5.1. Introduction 126
5.2. Results and discussion……128
5.2.1. Selection of a novel NAD-dependent MsIRED variant..... 128
5.2.2. Characterization of the selected NAD-dependent MsIRED variant..... 132
5.2.3. Structural modelling of MsIRED variants 137
5.3. Conclusions……..……139
5.4. Key outcomes 140
6. Alteration of substrate selectivity and improvement of kinetic parameters of a nitroreductase..... 141
6.1. Introduction 141
6.2. Results and discussion……145
6.2.1. Selection of EntNfsB variants with altered substrate selectivity 145
6.2.2. Characterization of the novel EntNfsB variants 151
6.2.3. Structural basis for the change in substrate selectivity of EntNfsB 156
6.3. Conclusions 158
6.4. Key outcomes 158
7. Selection of an entire multi-enzymatic metabolic pathway..... 160
7.1. Introduction 160
7.2. Results and discussion……162
7.2.1. Selection of an optimized isopropanol production pathway……...162
7.2.2. Characterization of selected isopropanol production pathway variants169
7.3. Conclusions 173
7.4. Key outcomes 173
8. Discussion and future perspectives..... 175
9. Appendix 184
9.1. List of appendix Tables and Figures..... 184
9.2. Appendix data……..185
Bibliography……191

List of figures

Figure 1.1. Global enzyme market..... 18
Figure 1.2. Protein fitness landscapes..... 19
Figure 1.3. Random mutagenesis techniques 29
Figure 1.4. Recombination-based mutagenesis techniques..... 32
Figure 1.5. Site-saturation and site-directed mutagenesis 34
Figure 1.6. Screening techniques..... 41
Figure 1.7. Selection techniques 46
Figure 2.1. Schematics of the workflow for the construction of plasmids for anaerobic growth complementation experiments 69
Figure 2.2. Schematics of the two approaches followed for the generation of saturation libraries 70
Figure 3.1. Glucose metabolism in the three main metabolic modes of *E. coli* 83
Figure 3.2. Redox rescue system..... 85
Figure 3.3. Characterization of the genotype of wild-type and AL strain *E. coli* cells..... 86
Figure 3.4. Inability of AL cells to grow under anaerobic fermentation conditions 87
Figure 3.5. Complementation of AL cells with the native *adhE* gene 88
Figure 3.6. Reactions catalysed by the exogenous reductases chosen to attempt anaerobic growth complementation 90
Figure 3.7. Metabolic complementation with exogenous NAD-dependent oxidoreductases 91
Figure 3.8. Toxicity of cyclohexanone, 3-methylcyclohexanone and the corresponding alcohols..... 92
Figure 3.9. Distribution of enzymatic activities……..95
Figure 4.1. Catalytic site of the different types of alcohol dehydrogenases..... 97
Figure 4.2. Enzymatic cofactor regeneration systems…….98
Figure 4.3. Reaction catalysed by CBADH and TBADH……..……..99
Figure 4.4. Characterization of wild-type CBADH 100
Figure 4.5. Cofactor binding pocket of CBADH..... 101
Figure 4.6. Selection of an NAD-dependent CBADH variant 103
Figure 4.7. Metabolic complementation with the selected NAD-dependent CBADH variant ……104
Figure 4.8. Characterization of $CBADH_S$ 107
Figure 4.9. Characterization of the genotype of *E. coli* strains ALS, ALP and ALPS..... 109
Figure 4.10. Inability of ALS, ALP and ALPS cells to grow under anaerobic fermentation conditions..... 109
Figure 4.11. Anaerobic growth complementation with NAD and NADP-dependent

oxidoreductases of AL, ALS, ALP and ALPS cells 111
Figure 4.12. Transhydrogenase-mediated metabolic complementation 111
Figure 4.13. Anaerobic growth complementation of AL and ALPS cells with wild-type TBADH 112
Figure 4.14. Purification and characterization of wild-type TBADH 112
Figure 4.15. Cofactor binding pocket of TBADH 114
Figure 4.16. Mutations found in $TBADH_{S1}$ and $TBADH_{S2}$ 115
Figure 4.17. Characterization of NAD-dependent TBADH variants 117
Figure 4.18. Crystals obtained for $CBADH_S$ and $TBADH_{S1}$ 119
Figure 4.19. Density for the NAD^+ cofactor in the crystallographic maps of $CBADH_S$ and $TBADH_{S1}$ 121
Figure 4.20. Cofactor binding pocket of $CBADH_S$ with NAD^+ bound 122
Figure 4.21. Cofactor binding pocket of $TBADH_{S1}$ with NAD^+ bound 122
Figure 4.22. General workflow for the selection of variants with the redox rescue selection system 124
Figure 5.1. Examples of chiral amines of industrial interest 126
Figure 5.2. Reactions catalysed by imine reductases 127
Figure 5.3. Reaction catalysed by *Streptomyces* GF3587 and *Myxococcus stipitatus* imine reductases 128
Figure 5.4. Cofactor binding pocket of wild-type MsIRED 129
Figure 5.5. Selection of an NAD-dependent MsIRED variant 131
Figure 5.6. Metabolic complementation with the selected NAD-dependent MsIRED variant 132
Figure 5.7. Characterization of wild-type MsIRED, $MsIRED_{C1}$ and $MsIRED_S$ 134
Figure 5.8. Structural basis for the inability of $MsIRED_{C1}$ and $MsIRED_S$ to use NADP 138
Figure 6.1. Gene-directed enzyme prodrug therapy (GDEPT) 142
Figure 6.2. Reactions catalysed by type I and type II nitroreductases 143
Figure 6.3. Overview of the structure of *Enterobacter cloacae* NfsB 145
Figure 6.4. Set of reactions tested with EntNfsB 146
Figure 6.5. Characterization of wild-type EntNfsB 147
Figure 6.6. Toxicity of 4-NBA, 2-NBA and 4-nitrobenzyl alcohol 147
Figure 6.7. Catalytic site of wild-type EntNfsB 148
Figure 6.8. Selection of EntNfsB variants 150
Figure 6.9. Metabolic complementation with the selected EntNfsB variants 152
Figure 6.10. Characterization of $EntNfsB_{S1}$ and $EntNfsB_{S2}$ 153
Figure 6.11. Interactions of wild-type EntNfsB with 4-NBA 156
Figure 6.12. Structural basis for the ability of $EntNfsB_{S1}$ to accept 2-NBA as a substrate 157

Figure 7.1. Engineered isopropanol production pathway 162
Figure 7.2. Metabolic complementation with the isopropanol production pathway 163
Figure 7.3. Balance of NAD^+ regeneration with glucose and gluconate as the carbon sources 165
Figure 7.4. Selection of variants of the isopropanol production pathway..... 166
Figure 7.5. Reactions catalysed by the enzymes encoded in the *sbm* operon 167
Figure 7.6. Comparison of selected and randomly-picked variants of the isopropanol production pathway 170
Figure 8.1. Overview of some of the applications of the redox rescue selection system ... 176
Figure 8.2. Expansion of the application scope of the redox rescue selection system 181
Figure 8.3. Conceptual *in vivo* continuous evolution system..... 183

List of tables

Table 1.1. Comparison of mutation rates for wild-type *E. coli* and two mutator strains 27

Table 1.2. Comaparison of some of the most frequent degeneracy schemes for saturation mutagenesis 36

Table 2.1. Microorganisms used in this project 50

Table 2.2. Plasmids used in this project 52

Table 2.3. Synthetic DNA sequences used in this project 56

Table 2.4. Summary of conditions for anaerobic growth complementation experiments 60

Table 2.5. Oligonucleotides used in this project 63

Table 2.6. PCR conditions for reactions with Phusion and GoTaq polymerases 68

Table 2.7. Summary of saturated positions and employed techniques for the generation of saturation libraries 70

Table 2.8. List of promoters and RBS 73

Table 2.9. List of transcriptional terminators 74

Table 2.10. Recipe for the preparation of poly-acrylamide gels 75

Table 2.11. Conditions for enzymatic activity assays 77

Table 2.12. Characteristic ^{1}H-NMR signals used to quantify each compound of interest 80

Table 3.1. Quantification of metabolites by ^{1}H-NMR of the fermentation broth of wild-type *E. coli* and AL cells complemented with *adhE* 89

Table 3.2. Quantification of metabolites by ^{1}H-NMR of the fermentation broth of AL cells complemented with BDHA, BUDC and TADH .....

Table 4.1. CBADH site-directed and saturation mutagenesis 102

Table 4.2. Quality of the CBADH library 102

Table 4.3. Quantification of metabolites by ^{1}H-NMR of the fermentation broth of AL cells complemented with wild-type CBADH or $CBADH_S$ 106

Table 4.4. TBADH saturation mutagenesis 114

Table 4.5. Quality of the TBADH library 114

Table 4.6. Mutations found in $TBADH_{S1}$ and $TBADH_{S2}$ 115

Table 4.7. Quantification of metabolites by ^{1}H-NMR of the fermentation broth of ALPS cells complemented with $TBADH_{S1}$ and $TBADH_{S2}$ 118

Table 5.1. MsIRED site-directed and saturation mutagenesis 130

Table 5.2. Quality of the MsIRED library 130

Table 5.3. Quantification of metabolites by ^{1}H-NMR of the fermentation broth of AL cells complemented with wild-type MsIRED, $MsIRED_{C1}$ and $MsIRED_S$ 136

Table 6.1. EntNfsB saturation mutagenesis 148

Table 6.2. Quality of the EntNfsB library 149

Table 6.3. Quantification of metabolites by ^{1}H-NMR of the fermentation broth of AL cells complemented with wild-type EntNfsB, EntNfsB$_{S1}$ and EntNfsB$_{S2}$ 155

Table 7.1. Quantification of metabolites by ^{1}H-NMR of the fermentation broth of two spatially separated clones of AL cells transformed with the library of isopropanol production pathways……...168

Table 7.2. Combinations of promoters and RBS found in the selected variants of the isopropanol production pathway..... 169

Table 7.3. Quantification of metabolites by ^{1}H-NMR of the culture broth of wild-type *E. coli* cells transformed with the best selected variant of the isopropanol production pathway and cultured aerobically in rich medium 172

Publications resulting from this work

Sellés Vidal L, James Murray, John T. Heap (in preparation for Nature Biotechnology) "Synthetic defect in universal metabolism enables versatile high-throughput laboratory evolution"

Sellés Vidal L, Kelly CL, Mordaka PM, Heap JT (2018) "Review of NAD(P)H-dependent oxidoreductases: Properties, engineering and application." *Biochimica et Biophysica Acta (BBA)* 1866:327–347. https://doi.org/10.1016/j.bbapap.2017.11.005

Intellectual Property

Sellés Vidal L, Heap JT (in preparation), Novel NADH-dependent acetone reductases

Sellés Vidal L, Heap JT (2019), *Polypeptides,* PCT/GB2019/051727

Note to the reader

Crystal structures were solved in collaboration with Dr. James W. Murray (Imperial College London).

List of abbreviations

2-NBA	2-nitrobenzoic acid
4-NBA	4-nitrobenzoic acid
A	adenine
ACO	aconitase
ADHE	aldehyde-alcohol dehydrogenase
ADP	adenosine diphosphate
ATP	adenosine triphosphate
BDHA	*Bacillus subtilis* 2,3-butanediol dehydrogenase
bp	basepair
BUDC	*Klebsiella pneumoniae* 2,3-butanediol dehydrogenase
C	cytosine
CASP	critical assessment of protein structure prediction
CBADH	*Clostridium beijerinckii* alcohol dehydrogenase
CDM	codon deletion mutagenesis
cDNA	complementary DNA
CDS	coding sequence
CoA	coenzyme A
Cre	causes recombination
CS	citrate synthase
DMSO	dimethyl sulfoxide
DMTr	4,4'-dimethoxytrityl
DNA	deoxyribonucleic acid
dNTPs	deoxynucleotide Triphosphates
E. coli	*Escherichia coli*
EMP	Embden-Meyerhof-Parnas
ENO	enolase
epPCR	error prone PCR
epRCA	error prone RCA
EYFP	enhanced yellow fluorescent protein
FACS	fluorescence-activated cell sorting
FAD	flavin adenine dinucleotide (oxidised)
$FADH_2$	flavin adenine dinucleotide (reduced)
FBA	fructose bisphosphate aldolase
FDA	Food and Drug Administration
FMN	flavin mononucleotide
Fmoc	9-fluorenylmethoxycarbonyl
fmol	femtomole
FRase I	flavin reductase I
FUM	fumarase
G	guanine
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
GDEPT	gene-directed enzyme prodrug therapy
gDNA	genomic DNA
GFP	green fluorescent protein
GFPUV	GFP optimized for excitation by UV light

GOI	gene of interest
GTP	guanosine triphosphate
HK	hexokinase
HPLC	high-performance liquid chromatography
HTSOS	high-throughput screening or selection
IDH	isocitrate dehydrogenase
IDT	Intergrated DNA Technologies
indel	insertions and deletions
IPG	1,2-O-isopropylidene-sn-glycerol
IPTG	isopropyl β-D-1-thiogalactopyranoside
IRED	imine reductase
ITCHY	incremental truncation for the creation of hybrid enzymes
IVC	*in vitro* compartmentalization
kb	kilobase
LB	Luria Bertani
LDHA	lactate dehydrogenase
Loxp	locus of X-over P1
LTR	long terminal repeats
MDH	malate dehydrogenase
MOE-PCR	multiple overlap extension PCR
MORPHING	mutagenic organized recombination process by homologous *in vivo* grouping
MP	metabolic pathway
mRNA	messenger RNA
NAD	nicotinamide adenine dinucleotide (oxidised or reduced)
NAD^+	nicotinamide adenine dinucleotide (oxidised)
NADH	nicotinamide adenine dinucleotide (reduced)
NADP	nicotinamide adenine dinucleotide phosphate (oxidised or reduced)
$NADP^+$	nicotinamide adenine dinucleotide phosphate (oxidised)
NADPH	Nicotinamide adenine dinucleotide phosphate (reduced)
NMR	Nuclear magnetic resonance
OD	optical density
OE-PCR	overlap extension PCR
OGDH	2-oxoglutarate dehydrogenase
ORI	origin of replication
PACE	phage-assisted continuous evolution
PC	pyruvate carboxylase
PCR	polymerase chain reaction
PFK	phosphofructokinase
PFL	pyruvate formate-lyase
PGI	phosphoglucose isomerase
PGK	phosphoglycerate kinase
PGM	phosphoglycerate mutase
POEP	polyacrylamide gel electrophoresis mediated overlap Extension PCR
Pol I	DNA polymerase I
PTA	phosphate acetyltransferase
QUEST	querying for enzymes using the three-hybrid system
RACHITT	random chimeragenesis on transient templates
RAISE	random insertional-deletional strand exchange mutagenesis

RBS	ribosome binding site
RCA	rolling circle amplification
RNA	ribonucleic acid
RNAP	RNA polymerase
rpm	revolutions per minute
rRNA	ribosomal RNA
SCS	succinyl-CoA synthetase
SDH	succinate dehydrogenase
SDM	site-directed mutagenesis
SHIPREC	sequence homology-independent protein recombination
SISD	sequence independent site-directed chimeragenesis
ssDNA	single-stranded DNA
SSM	site-saturation mutagenesis
StEP	staggered expansion process
T	thymine
TADH	*Thermus* sp. ATN-1 alcohol dehydrogenase
Tat	twin-arginine translocation signal peptide
TBADH	*Thermoanaerobacter brockii* alcohol dehydrogenase
TCA	tricarboxylic acid
TdT	terminal deoxynucleotidyl transferase
T_m	melting temperature
TMAO	trimethylamine N-oxide
TNT	2,4,6-trinitrotoluene
TP	terminal protein
TP-DNAP1	TP-DNA polymerase 1
TPIA	triose-phosphate isomerase
TriNEx	trinucleotide exchange
TRINS	tandem repeat insertions
tRNA	transfer RNA
Ty1RT	retrotranscriptase of the Ty1 retrotransposon
UV	ultraviolet
VMA	vacuolar ATPase subunit
w-o	water-in-oil
w-o-w	water-in-oil-in-water
WT	wild type

1. Introduction

1.1. Background

Catalysis is an essential requirement for life. Without catalysis, the complex reactions essential for all life forms would take an extraordinarily long amount of time, some of them ranging in the order of many millions of years (Miller and Wolfenden, 2002). The main biological catalysts are enzymes, biological macromolecules able to catalyse the chemical reactions underlying the biological processes that take place within living organisms.

Enzymes possess several properties that differentiate them from chemical catalysts, such as the high selectivity with which they operate and their regio- and stereospecificity, accepting usually only one or a limited range of isomers of a compound as their substrate and yielding products in an equally selective manner. Furthermore, enzyme-catalysed reactions proceed at greatly enhanced rates, and with the additional benefit of generating significantly fewer undesired byproducts than other catalysts. Finally, since enzymes must be able to catalyse reactions at conditions compatible with life, they can typically perform their activity at relatively mild conditions of, for example, temperature and pressure.

Thanks to such properties, enzymes have attracted a growing interest from an industrial point of view as a more efficient and less costly alternative to certain steps traditionally performed with synthetic chemistry in the manufacturing of many compounds (Johannes, Simurdiak, and Zhao, 2005) (**Figure 1.1**). The range of applications of enzymes is as broad as the types of reactions they catalyse (Chapman *et al.*, 2018). For example, in the pharmaceutical and agrochemical industries, enzymes are often used to catalyse specific modifications affecting only one of the functional groups present on complex substrates as part of a series of sequential steps required to transform a substrate into an active compound (such as medicines, herbicides, pesticides, etc.) where a combination of enzymatically-catalysed reactions and traditional organic synthesis is used (Li *et al.*, 2012; Martinez *et al.*, 2008). Enzymes have also been widely used to act on polymers of biological origin. For instance, proteases and lipases have been long used in laundry detergents to degrade the proteins and fats that form part of stains caused by materials of biological origin, such as milk or blood (Olsen and Falholt, 1998). Numerous enzymes are also used in the food processing industry, both for the bulk generation of a specific product, such as the production of syrups by the amylase-mediated hydrolysis of starch (Flor and Hayashida, 1983), and for the alteration of organoleptic properties of certain foods, such as the debittering of citrus fruit juices by enzymes like naringinase (Narnoliya and Jadaun, 2019). Of particular interest is also the enzymatic generation of biofuels. Several potential strategies are currently in development,

including the use of cellulases and xylanases to degrade lignocellulotic materials for the production of bioethanol (Meites, 1977; Olofsson *et al.*, 2017), and the production of fatty acid alkyl esters through the transesterification of fats with lipases (Tan *et al.*, 2010).

While the naturally available repertoire provides a vast variety of biocatalysts able to catalyse a wide range of reactions, it is often found that an enzyme optimal for specific applications of human interest is not available. Indeed, the same high selectivity that makes enzymes an attractive catalyst, can also be an obstacle by preventing them from efficiently accepting the desired substrate. Furthermore, many natural enzymes suffer from loss of activity when some special reaction conditions are required, such as the presence of organic solvents (Stepankova *et al.*, 2013).

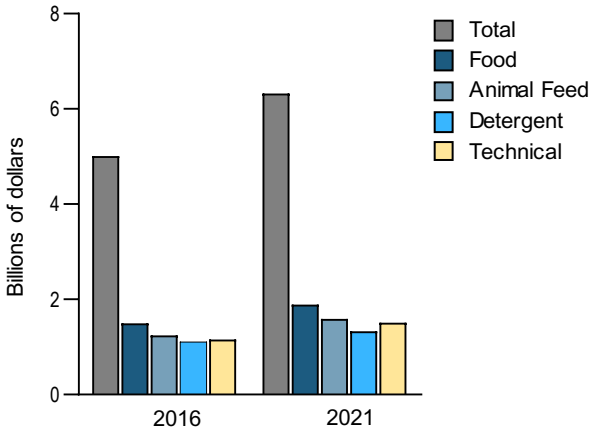


Figure 1.1. Global enzyme market. Enzyme market is estimated to reach a value of over 6.3 billion dollars by 2021, partially due to a shift towards greener catalysts in several industries to avoid fines by environmental agencies. (Figure adapted from Chapman *et al.*, 2018)

Thus, it becomes frequently the case that natural enzymes need to be tailored in order to maximize their performance for each specific application. The process of altering a natural occurring protein, usually with the goal of developing variants with novel or enhanced properties that make them more suitable for specific purposes of human interest, is called protein engineering, and can be described as a transition from a functional protein to another one within the space of all possible protein sequences (Smith, 1970). Even though the number of changes required to transition between the two sequences is usually small, the high-dimensional space contains a massive amount of potential protein sequences. For example, for a short protein of only 100 amino acids, more than 20^{100} possible sequences exist, a number larger than the amount of atoms in the universe (Romero and Arnold, 2009). It is possible to assign a fitness measurement (such as activity towards a given substrate) to each protein in the sequence space. The transition of a parental sequence with a suboptimal fitness

to a variant with increased fitness for a particular application is the goal of protein engineering. Protein engineering can then be treated as an optimization problem, where the landscape of the space to be searched can be taken into account to determine the search strategy (**Figure 1.2**). One of the main problems is the presence of local fitness maxima, which can hinder the achievement of optimal variants unless a temporary decrease in fitness is accepted during the search process. Although the vast size of the sequence space prevents the characterization of the fitness landscape, some general features have been experimentally observed. Firstly, the amount of non-functional variants is immensely larger than that of functional variants, which are present at a very low global density. Additionally, functional variants are not evenly distributed across the sequence space. Instead, they tend to be clustered together, which makes point mutation an efficient evolution mechanism.

Two main approaches can be taken to carry out protein engineering: rational design and directed evolution.

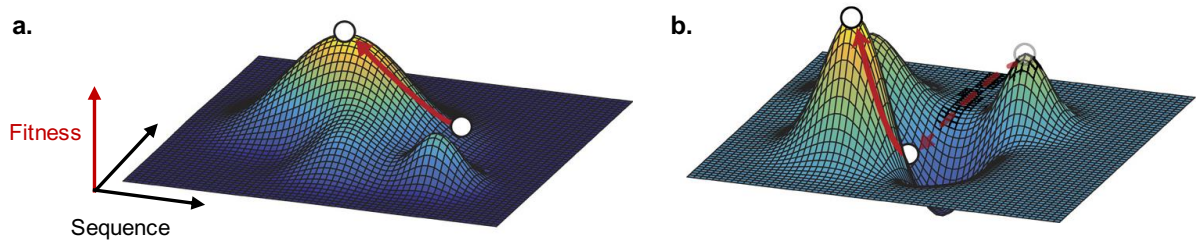


Figure 1.2. Protein fitness landscapes. Each point in the XY plane represents a different sequence, with an associated fitness represented along the Z axis. Several types of fitness landscapes are possible. **a**. Landscape with no local fitness maxima on the shortest path to the global maximum, where every beneficial mutation brings the protein closer to the optimal for the desired application. **b.** Landscape with local fitness maxima. Local maxima pose an obstacle for the consecution of optimal variants, since it can be necessary to go through a series of variants with decreased fitness before reaching the global maximum. (Figure adapted from Steinberg and Ostermeier, 2016)

1.2. Rational design

Rational design involves performing specific point mutations, insertions or deletions in the coding sequence. Mutations must be decided based on structural and functional information about the target biomolecule. After expressing and purifying the generated variant, it must be assessed through functional assays to determine if the expected gain of function was achieved. It is therefore key to have not only structural information about the target protein, but also a good understanding of the structure-function relationship. Several sources of information can be integrated to predict the relationship between sequence, structure and function of a given protein at different levels of details (ranging from whole domains to specific

amino acid residues). Sequence information can be used to search characterized orthologues and paralogues, identify conserved motifs or residues conferring specific functionality by means of multiple sequence alignment or predict the cellular compartment where the protein is expected to perform its function. Structure-based methods for the prediction of protein function include the search of proteins with similar folds, the identification of relatively small local 3D templates associated with certain functions (such as residue arrangements specific to a catalytic site) or the analysis of surface properties, which can be helpful in the prediction of protein-protein interactions and binding of small molecules (Lee *et al.*, 2007). Nevertheless, even with the help of such computational tools to assist rational design, the sequence-structure-function relationship is often difficult to predict accurately, particularly at the single residue level. Additionally, reliable structural information is frequently not available for the protein of interest, and, while progress is being made in methods for protein structure prediction, these still remain rather limited, specially for larger proteins and macromolecular complexes (Schwede, 2013). For these reasons, in many cases rationally designed mutations do not have the expected result or even cause undesired side effects.

Undeniably, rational design has shed a considerable amount of information about the relationship between sequence, structure and function of proteins, and has led to several cases of success at obtaining variant biomolecules with improved properties. A classical example can be found in the early studies done on *Bacillus stearothermophilus* tyrosyl-tRNA synthetase by Greg Winter's laboratory, which employed site-directed mutagenesis to dissect the role of specific residues in the catalytic mechanism, correctly predicting even substitutions that led to increased activity (Wilkinson *et al.*, 1984). Some years later, Wilks *et al.* took advantage of the extensive structural studies of lactate dehydrogenase to alter in several ways its substrate specificity. Initially they changed its substrate specificity to malate with a single point mutation, effectively obtaining a malate dehydrogenase that was even more active than native enzymes (Wilks *et al.*, 1988). Later, they broadened the substrate specificity of the enzyme to larger molecules, such as branched hydroxy-acids and phenylpyruvate (Wilks *et al.*, 1990, 1992).

However, it also became evident that approaches purely based on rational design were limited due to the inability to predict the effect of residue substitutions even when detailed structural information is available. After the substrate-bound structure of trypsin was solved, it was hypothesized that its preference for hydrolyzing peptides at basic residues could be due to the negatively charged side chain of the Asp189 residue found at the substrate binding pocket. William Rutter and collaborators attempted to generate a variant enzyme that would instead cleave at acidic residues by mutating Asp189 to a lysine residue. Nevertheless, no such change in activity was observed (Perona *et al.*, 1995).

Additionally, while it had also been suggested that the absence of an Asp residue at position 189 would cause trypsin to shift its substrate preference towards that of chymotrypsin (Steitz, Henderson, and Blow, 1969), another protease with almost exactly the same active site geometry but cleaving at hydrophobic residues, the Asp189Lys mutant did not exhibit increased chymotrypsin-like activity. It was instead found that a total of 15 residue substitutions were required to change the substrate specificity of trypsin to that of chymotrypsin, most of them in surface loops which did not directly contact the substrate (Steitz, Henderson, and Blow, 1969). Furthermore, Rutter's laboratory did not succeed at modifying the specificity of chymotrypsin to that of trypsin by performing the reverse residue mutations. Instead, in most cases the protease activity was simply reduced, suggesting that chymotrypsin-specific residues had to be mutated to achieve the conversion (Venekei *et al.*, 1996).

It became then obvious that, in order to develop efficient protein variants with novel or altered functionality, it was required to test many different substitutions at any given position, and that, often, several sites had to be mutated to obtain the desired effect. The number of variants to be generated and analyzed became then impractical to tackle by means of traditional rational design strategies, triggering the development of directed evolution.

1.3. Directed evolution

Directed evolution, on the other hand, bypasses the need to determine specific mutations *a priori* by mimicking the process of natural evolution in the laboratory. In nature, mutations which are beneficial for individuals to adapt to the environment are iteratively selected through numerous generations. Directed evolution approaches attempt to reproduce this process with two major differences: the evolutionary process must take place at a much shorter timescale, and it is oriented towards the generation of biomolecules that suit better particular applications of human interest.

Albeit unknowingly, humans have been employing directed evolution for several millennia under the form of selective breeding. By iteratively choosing animal and plant varieties with desired phenotypes and breeding them, an artificial selection pressure is effectively applied where individuals with traits that make them more useful for the breeder are selected.

The first *in vitro* evolution experiments, performed in the laboratory in a much more controlled manner, can be traced back to the 1960s. In a pioneering Darwinian experiment, Sol Spiegelman *et al*. iteratively selected RNA molecules based on their ability to be replicated by Q bacteriophage RNA polymerase (Mills, Peterson, and Spiegelman, 1967). The authors

performed several cycles of short amplification by the RNA polymerase and transfer of the products to a new reaction, eventually obtaining a molecule much more efficiently amplified by the Q bacteriophage RNA polymerase and remarkably smaller. Over the next two decades, such *in vitro* evolution experiments shifted towards more application-driven approaches. This is exemplified by the development of phage display in 1985 by George Smith (Smith, 1985). Smith fused an external sequence to a gene encoding a minor coat protein of a filamentous phage, leading to the assembled viral particles to display the amino acids encoded in the external sequence. A set of phages with different fused sequences could then be subjected to affinity purification with a given binding partner, such as an antibody, to obtain sequences encoding peptides with high affinity towards said partner. The change in the original purpose of the experiments performed by Spiegelman evidences already a shift in the paradigm of evolution experiments: while Spiegelman designed his experiment simply as a means to answer the question of what would happen if an artificial selective pressure was applied *in vitro*, Smith purposely developed his system as a methodology to select biomolecules with specific properties.

The modern concept of directed evolution started to be shaped in the 1990s (Cobb, Chao, and Zhao, 2013). Many successful cases followed each other where a wide variety of novel variants of biomolecules of human interest for specific applications were developed. The range of evolved properties was equally diverse, including greater activity under non-physiological conditions such as the presence of organic solvents or extreme pH, increased thermal stability, altered substrate specificity, improved kinetics, etc.

The advent of directed evolution should, nonetheless, by no means be interpreted as making structural and biochemical information irrelevant for protein engineering. Instead, it provides a way to incorporate such information in a more effective manner, in addition to allowing to tackle protein engineering even when such knowledge is not available.

Directed evolution approaches comprise an ample variety of techniques, but two main steps are always required. The first step consists in generating enough genetic diversity of a given parental sequence to cover the sequence space to be sampled. The resulting set of sequences, or library, often includes a majority of variants that do not display the desired novel or improved property, or simply have become non-functional. Thus, in a second critical step, the variants of interest must be identified and isolated from the library. There are two main type of approaches to do so: screening and selection. Screening approaches consist of evaluating across the library the property sought to be improved or developed, leading to the identification of the best performing variants. On the other hand, selection approaches link an improvement in the evolved property to an increased survival rate. Such approaches discard

automatically all non-functional variants, and thus tend to allow a much higher throughput than screening.

1.4. Generation of a library of variants

Mutations in the genetic material take place spontaneously in living cells as a consequence of errors in DNA replication systems, alterations in DNA caused by reactive oxidative species generated as byproducts of cellular metabolism or the action of external factors present in the environment, such as UV radiation. However, the rate of spontaneous mutations of living cells is usually insufficient for generating the genetic diversity required for obtaining a library large enough to include variants with the desired properties. For example, the mutation rate of wild-type *Escherichia coli* (*E. coli*) is approximately 1×10^{-3} mutations per genome per generation, or 2.2×10^{-10} mutations per base pair per generation (Lee *et al.*, 2012)**.** At such a low mutation rate, over 100,000 generations would be required on average to obtain a single point mutation in a target gene of 1000 base pairs (Greener, Callahan, and Jerpseth 1996).

Therefore, it is necessary to artificially enhance genetic diversification by increasing the mutation rate. A plethora of techniques that allow the generation of large libraries in a reasonable amount of time is available, each one having their own advantages and drawbacks which must be considered before determining the best option for each specific application.

1.4.1. Random mutagenesis

In random mutagenesis approaches, no specific positions of the sequences to be mutated are targeted. Such techniques are particularly useful for directed evolution of proteins for which there is not enough structural information available to determine specific residues that play a determinant role for the property to be evolved. Furthermore, even when structural information is available, the structure-function relationship is not always evident, and mutations expected to have a desirable effect can lead to unpredicted negative side effects. Random mutagenesis techniques are also useful when the property to be evolved cannot be easily attributed to a few specific positions. For instance, enhanced stability under unusual or extreme conditions such as high temperature or presence of large amounts of organic solvents.

For these reasons, the first directed evolution experiments employed random mutagenesis techniques. Some of the earliest mutagenesis techniques were based on the use of chemical or physical mutagens to introduce random mutations in the sequence of interest (Myers, Lerman, and Maniatis, 1985) (**Figure 1.3a**). However, while such mutagenic agents proved to be useful for random deactivation of genes in genome-wide screenings, they were not extensively used for directed evolution due to their strong bias towards certain mutations and

the need to perform further manipulation with the mutagenized sequences to obtain double-stranded DNA fragments.

In one of the landmarks of directed evolution, Chen and Arnold alternated rounds of error prone PCR (epPCR) and screening to obtain a variant with increased activity in the presence of 60% dimethylformamide (Chen and Arnold, 1993) of subtilisin E, a serine protease with multiple industrial applications. The authors used Taq polymerase for the PCR, which has an inherently higher error rate than the high-fidelity polymerases used for standard PCR. Moreover, they further increased the polymerase error rate by modifying the reaction conditions of the PCR. Several PCR conditions can be altered to reduce the fidelity of DNA polymerases, amongst which the addition of Mn^{2+} ions, increasing the concentration of Mg^{2+} and using unequal concentrations of the different deoxyribonucleotides are some of the most frequent. Since its original development in 1992 by Cadwell and Joyce (Cadwell and Joyce, 1992), epPCR has become one of the most widely used techniques for random mutagenesis (**Figure 1.3b**). Other modifications of the original protocol have attempted to further increase the mutation rate by modifying other reaction conditions, such as employing nucleotides modified by chemical treatments (Bratulic and Badran, 2017), incorporating analogues of the four original nucleosides (Bratulic and Badran, 2017) or even using heavy water as the solvent (Minamoto, Wada, and Shimizu, 2012). An inherent problem to the technique is the preference bias towards some mutations. For example, Taq polymerase favours AT to GC transitions over other types of mutations. Some protocols try to work around this limitation by using combinations of several polymerases, which together yield a relatively unbiased mutational spectrum, such as the GeneMorph II random mutagenesis kit (Agilent Technologies, 2015).

Approaches based on epPCR require, nevertheless, further manipulation of the resulting products before they can be transformed into a host strain. These steps often include, at least, treatment with the appropriate restriction enzymes, purification of the correct product and ligation into a vector. Due to the sometimes limited efficiency of the ligation reaction, these additional required steps can lead to a loss of part of the library, effectively limiting its size. This limitation can be overcome by using an amplification method that yields products that can be directly transformed into the host organism instead of PCR, such as rolling circle amplification (RCA) (**Figure 1.3c**). RCA was first devised in 1995 (Fire and Xu, 1995). Briefly, RCA is an isothermal amplification technique where a circular DNA molecule containing the sequence of interest is used as the template. Typically, φ29 DNA polymerase, which possesses a high processivity and strand displacement ability is used together with a universal primer comprising six nucleotides. The amplification reaction results in linear DNA products containing up to several hundred copies in tandem of the sequence of interest. Due to the presence of these tandem repeats, when the products of RCA are directly transformed into

host system, they are recircularized by their recombination systems, resulting in circular replicas of the original plasmid. Fujii *et al.* developed in 2004 a simple protocol for error-prone RCA (epRCA) where the mutagenesis rate is enhanced through the addition of $MnCl_2$ to the reaction and by reducing the concentration of the DNA template (Fujii, Kitaoka, and Hayashi, 2004). The efficiency of the method has been further improved by introducing loxP sites in the template plasmid and treating the obtained tandem repeats with Cre recombinase before transformation (Huovinen *et al.*, 2011).

Both epPCR and epRCA introduce only substitutions in the sequences subjected to mutagenesis. However, insertions and deletions (indels) also play an important role in natural genetic diversification, since they can alter the backbone of the encoded proteins in a way not achievable simply by point mutations (Shortle and Sondek, 1995). Several methodologies for introducing indels in the sequences to be mutated have been developed. Initial efforts focused on the development of technologies which allowed for the introduction of random insertions and deletions, such as the RAISE method (Fujii, Kitaoka, and Hayashi, 2006) (**Figure 1.3d**). In RAISE, fragments of the sequence of interest are randomly extended with terminal deoxynucleotidyl transferase (TdT), an enzyme which plays a key role in the maturation of the hypervariable regions of antibodies. The resulting products are then used for self-primed PCR. By relying on the fact that the random extensions are able to find a complementary region in the original sequence at a random position, fragments of varying length are introduced at random positions across all the sequence.

Insertions by duplication are another of the most frequent naturally occurring indels, with some estimates placing their frequency as high as two thirds of all insertions observed in natural genomes (Kipnis, Dellus-Gur, and Tawfik, 2012). TRINS was developed as a method to mimic such type of insertions *in vitro*, resulting in the formation of tandem repeats of segments of the sequence of interest (Kipnis, Dellus-Gur, and Tawfik, 2012) (**Figure 1.3e**). Firstly, the sequence is digested with DNAse I to generate multiple fragments of random length. Part of the resulting products is then converted to single stranded circular DNA by melting the duplexes and treating with ssDNA ligase. The circular products are then mixed with the remaining linear fragments to perform an assembly PCR. When a linear fragment anneals with a circular one, a polymerization reaction similar to RCA takes place. The DNA polymerase synthesizes a first copy of the circular region, and then proceeds to synthesize additional full or partial copies by displacing the previously synthesized copy. Finally, full-length genes are amplified by means of another PCR where the products of the previous step are used as the template.

Nevertheless, such techniques present the drawback of generating a large proportion of variants with frameshifts, which usually lead to loss of function. Thus, considerable effort has been made towards the development of techniques that allow the generation of more refined indels, many of which are based on modified versions of the mini-Mu transposon, a transposable element which can be easily inserted at random points of a DNA sequence simply by treatment with the MuA transposase (Haapa *et al.*, 1999). Jones modified the mini-Mu transposon to include *MlyI* recognition sites close to both of its termini. *MlyI* is a type IIS restriction enzyme, which cuts at a distance of 5 base pairs from its recognition site. This property can be exploited to generate deletions of a single codon at random positions of a sequence by first inserting the modified mini-Mu transposon and then digesting with *MlyI*. Subsequent ligation then leads to the removal of 3 base pairs due to the specific positioning of the *MlyI* recognition sites in the mini-Mu transposon (Jones, 2005) (**Figure 1.3f**). In 2008, the protocol was adapted to also allow the subsequent insertion of a randomized trinucleotide (Baldwin *et al.*, 2008). The method was termed TriNEx and included an additional step before ligation where a DNA cassette (SubSeq) carrying a degenerated codon at one of its termini is inserted between the two fragments of the gene of interest obtained after *MlyI* treatment. The SubSeq cassette is then removed by another round of *MlyI* digestion, leaving only the degenerated trinucleotide. TriNEx effectively allows random substitutions of single amino acids across an entire sequence while avoiding the drawbacks of PCR-based methods, such as codon bias. The same methodology can be used to insert any sequence, including full protein domains, at random positions simply by using the sequence to be inserted as the SubSeq cassette (Edwards *et al.*, 2008).

Further modifications of the mini-Mu-based deletion technique have expanded its scope to allow the deletion of up to 5 consecutive codons at random positions of the mutated sequence (Liu *et al.*, 2016). In the protocol developed by Liu *et al.*, termed codon deletion mutagenesis (CDM), the sequence of interest is cloned into a plasmid where it becomes fused downstream to a Tat signal peptide and the N-terminal domain of the vacuolar ATPase subunit (VMA) intein. A modified mini-Mu transposon with asymmetric ends and carrying the C-terminal domain of the VMA intein fused to TEM-1 β-lactamase is then transposed into the sequence of interest. Inverse PCR is then performed with combinations of reverse and forward primers that can be adjusted to tune the final number of deleted codons. The resulting products are treated with *BsgI* and Klenow fragment to remove the transposon. Only in the products where the transposon was inserted in the correct reading frame, the Tat signal peptide, both termini of the VMA intein and the TEM-1 β-lactamase will be arranged in a way such that when transformed into the host cells, the intein is excised and a functional TEM-1 β-lactamase with the Tat peptide is produced. Consequently, all products with frameshifts due to transposon

insertion can be automatically discarded by culturing transformed cells in media supplemented with ampicillin. The authors successfully employed CDM to determine critical residues for the fluorescence of GFP, and even identify potential deletions that could enhance fluorescence.

While the above described *in vitro* mutagenesis techniques enable a relatively good control over the mutation rate and spectrum, they also suffer from inherent bottlenecks, such as the limitation on the library size imposed by the transformation efficiency and the requirement, to a varying degree, of manipulating the genetic material. *In vivo* approaches try to bypass these limitations by taking advantage of the cell machinery to directly perform mutagenesis within the host cells. This not only simplifies the process of library generation, but also allows the coupling of mutation and screening or selection cycles, enabling more automated directed evolution workflows.

Mutator strains were the first *in vivo* mutagenesis systems, and took advantage of the increase in mutation rates that occur when DNA repair pathways become inactivated (**Table 1.1**). Greener and Callahan developed the widespread XL1-Red strain (Greener and Callahan, 1994), where three of the primary DNA repair mechanisms were inactivated by mutating several genes including mutS (mismatch repair), mutT (oxo-dGTP hydrolysis) and mutD (3'-5' exonuclease activity of DNA polymerase III) (**Figure 1.3g**). The strain displayed a rate of mutation approximately 5000 times larger than that of the wild-type. While such hypermutator strains provide a simple system for *in vivo* mutagenesis, they also present major drawbacks. These include a relatively low and hard to control mutagenesis rate, biased mutagenesis spectrum and, perhaps most importantly, the indiscriminate mutagenesis of the genome and other sequences out of the gene of interest that occur within mutator strains. This causes a series of undesirable side effects, such as slower growth, reduced transformation efficiency, and potential apparition of false positives.

Table 1.1. Comparison of mutation rates for wild-type *E. coli* and two mutator strains. The frequency of mutations per gene was calculated based on the spontaneous mutation frequency, the size of the *E. coli* genome and the copy number (~100) and size (~4000 bp) of a typical cloning vector. The generations of growth refer to the expected number of cell divisions required to produce one point mutation per kilobase of target DNA. (Adapted from Greener, Callahan, and Jerpseth, 1996)

Host	Spontaneous mutation frequency	Mutations per generation	Generations of growth
Wild-type	10^{-10}	1/4000	120000
mutS	10^{-8}	1/40	1200
XL1-Red	5×10^{-6}	12.5	24

In order to overcome the latter problem, a targeted gene evolution system based on error-prone DNA polymerase I (Pol I) was developed by Camps et al (Camps *et al.*, 2003). In the method, the JS200 strain of *E. coli*, which carries a temperature sensitive allele of Pol I in its genome, is transformed with two plasmids. One of them has a Pol I-independent replication origin, and encodes an error prone version of Pol I. The other one is a high copy number plasmid containing the target sequence to be mutated downstream of a ColE1 replication origin. The ColE1 replication origin causes replication of the plasmid to be initiated by Pol I, before switching to the less error-prone DNA polymerase III. When the error-prone version of Pol I initiates replication of the target plasmid, it introduces an increased number of mutations but only on the first 3000 base pairs downstream of the ColE1 replication origin, thus restricting the increased mutagenesis rate to the sequence of interest.

Based on the same idea of having an alternative error-prone DNA polymerase to replicate the gene of interest, a system for *in vivo* mutagenesis in yeast was recently developed. The system is based on the pGKL1/2 plasmid system of *Kluveromyces lactis* (Ravikumar, Arrieta, and Liu, 2014). Both are linear cytoplasmic plasmids which have a terminal protein (TP) covalently linked to both of their 5' ends. The replication of these plasmids presents the peculiarity of being primed by the covalently bound TP. pGKL2 contains all the components required for the transcription and replication of pGKL1/2, except for TP-DNA polymerase 1 (TP-DNAP1), a fusion protein including TP and a DNA polymerase able to start TP-primed replication of pGKL1, which is encoded in pGKL1. The authors also developed an error-prone version of TP-DNAP1, which they inserted in a circular nuclear yeast plasmid able to replicate autonomously with the replication machinery of the host. When the gene of interest was placed in pGKL1 and host cells were transformed with pGKL1/2 and the circular plasmid carrying error-prone TP-DNAP1, an increased rate of mutagenesis up to 400 times the base mutation rate of the genome was detected exclusively on pGKL1.

Other systems have expanded the idea to use enzymes other than DNA-dependent DNA polymerases as the main *in vivo* source of mutagenesis. This is the case of the Ty1 retrotransposon-based *in vivo* mutagenesis system for yeasts (Crook *et al.*, 2016). Ty1 contains 5' and 3' long terminal repeats (LTR) and encodes a reverse transcriptase (Ty1RT). During its replication cycle, the Ty1 retrotransposon is first transcribed into RNA by the transcription machinery of the cell. The resulting RNA molecule is then read by the translation system as an mRNA and used to synthesize Ty1RT. Ty1RT retrotranscribes the RNA form of Ty1 into its complementary DNA, which is then integrated into a stable genomic locus. The authors took advantage of the retrotransposon replication cycle by introducing the sequence of interest between the Ty1RT coding sequence and the 3' LTR, exposing it to increased mutagenesis rate due to the error-prone nature of Ty1RT. Relatively high transposition rates,

required for obtaining large libraries, were achieved for sequences of up to 5 kilobases, allowing the generation of large enough libraries of single genes and even some multi-enzymatic pathways.

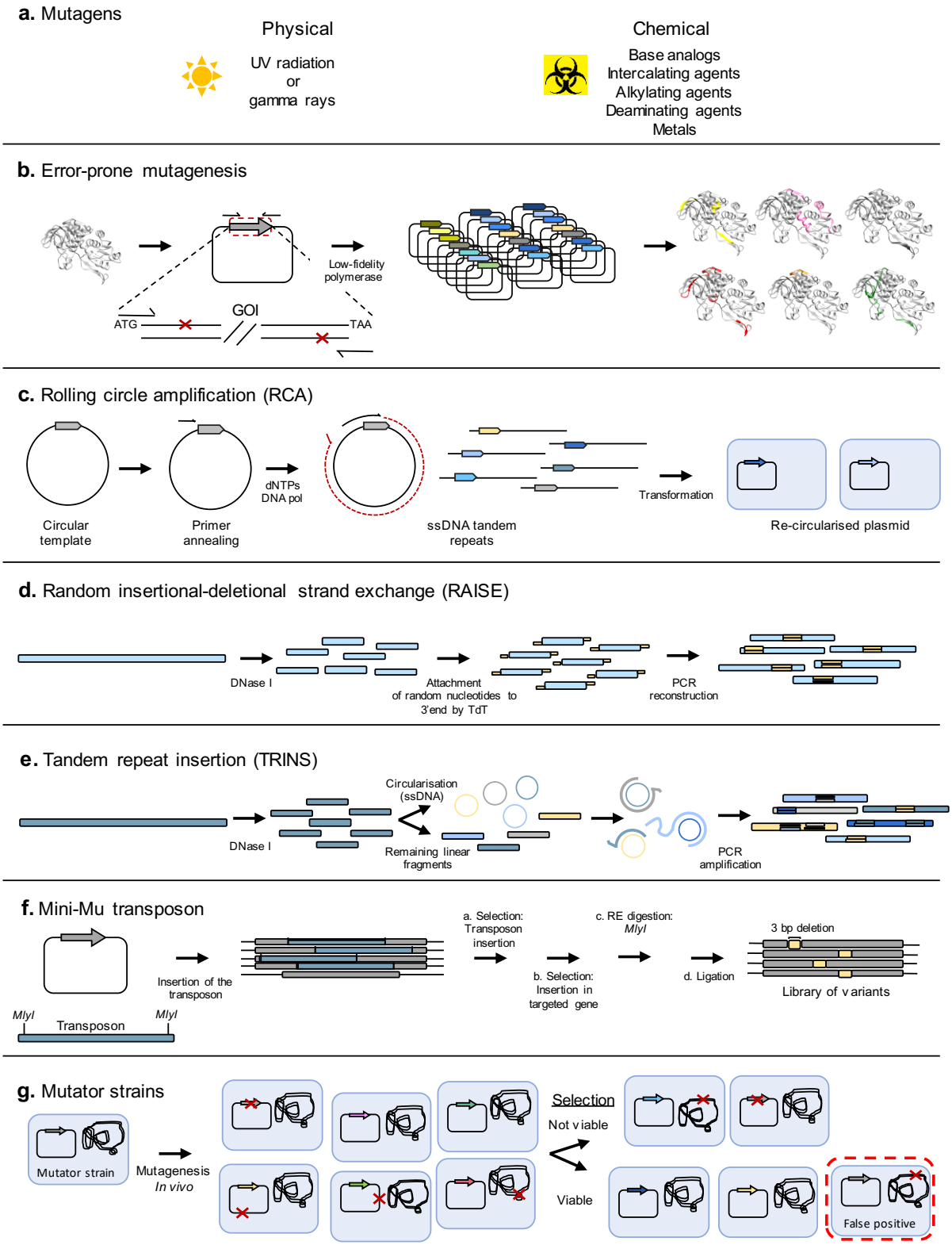


Figure 1.3. Random mutagenesis techniques. Some of the most widely applied random mutagenesis approaches are represented. **a.** Chemical and physical mutagens were the basis of many genome-wide

screening experiments, but have not been extensively used in directed evolution. **b.** In error-prone PCR, random mutations are introduced by a low fidelity polymerase, resulting in linear DNA fragments with point mutations, which generally require further manipulation to be inserted into an appropriate vector. **c.** RCA based methods bypass the need for further manipulation, since the products can be automatically recircularized by the host cells. **d.** RAISE was one of the first methodologies designed to introduce random insertions in the sequence to be mutated. Random small extensions are attached to digested fragments of the parental sequence, and full-length genes are reconstructed through PCR. **e.** TRINS aims to generate repeats of random short fragments of the parental sequence. First, the parental sequence is digested with DNase I. Part of the obtained fragments are circularized and mixed with the remaining linear fragments. An assembly PCR reaction is then performed. When a linear fragment anneals with a circular fragment, a reaction similar to RCA takes place, leading to the replication of multiple copies of the region corresponding to the circularized fragment. **f.** Several mutagenesis techniques based on the mini-Mu transposon have been devised. Such techniques allow the generation of random insertions or even deletions while preserving the appropriate reading frame. **g.** Mutator strains were the first tool enabling *in vivo* random mutagenesis. However, the increased mutagenesis rate applies to the whole genetic material, and not only to the sequence of interest. More sophisticated approaches where only the gene of interest is targeted have been developed.

1.4.2. Recombination techniques

Recombination is one of the most powerful mechanisms that operate in nature to generate the genetic diversification required for adapting more effectively to the environment (Didelot and Maiden, 2010; Stapley *et al.*, 2017). A large variety of recombination-based techniques have been developed, which greatly differ from the previously described random mutagenesis techniques in that they produce libraries of combinatorial variants where segments from several functional biomolecules are combined. In principle, libraries generated through recombination methods possess a higher percentage of functional variants, avoid the introduction of stop codons, and can lead to the discovery of epistatic mutations that cooperate to produce an enhanced effect (Ruff, Dennig, and Schwaneberg, 2013).

The first *in vitro* recombination-based technique for library generation was DNA shuffling, developed in 1994 by Willem Stemmer (Stemmer, 1994). The protocol involves randomly fragmenting a pool of closely related sequences with DNase I treatment, and then reassembling full-length sequences by means of self-priming PCR. Recombination is achieved thanks to annealing of fragments with a high degree of homology derived from different parental sequences (**Figure 1.4a**). The method was originally applied to successfully obtain variants of TEM-1 β lactamase conferring enhanced resistance to cefotaxime by using the variants extracted from several colonies displaying some level of resistance to the antibiotic.

A technically simpler approach where the need to fragment the parental sequences with DNase I is bypassed, termed “staggered expansion process” (StEP), was developed by the group of Francis Arnold in 1998 (Zhao *et al.*, 1998). StEP consists of repeated cycles of denaturation of the templates and very short annealing and extension by DNA polymerase

steps, which are performed until full-length products are obtained. In each cycle, the growing fragment anneals with a different template, causing most of the final full-length sequences to be a combination of several of the templates (**Figure 1.4b**).

However, both DNA shuffling and StEP suffer from a major drawback which limits their scope: they require a high degree of homology to exist between the different parental sequences to be recombined, making them unsuitable for the recombination of distantly related sequences. Thus, at the beginning of the new century, a variety of newly developed methods employed different strategies to reduce the required degree of homology.

One of the first such tools, Random Chimeragenesis on Transient Templates (RACHITT) (Coco *et al.*, 2001), was based on the use of one of a single strand of one of the sequences to be recombined as a scaffolding template. Fragments of the complementary strand of the other sequences are hybridized with the scaffold strand. Then, unhybridized termini are digested, gaps are filled and the remaining nicks are ligated. Finally, scaffolding strands are destroyed and the complementary strands of the recombined sequences are synthesized by means of PCR (**Figure 1.4c**). RACHITT not only resulted in a higher overall number of crossovers, particularly in regions of low sequence identity, but also completely removed the presence of parental sequences in the final library.

Others went even further, seeking to develop techniques that completely removed any requirement for homology between the sequences. The first methodologies allowed only two genes to be recombined, such as Incremental Truncation for the Creation of Hybrid enzymes (ITCHY) (Ostermeier, Shim, and Benkovic, 1999) and Sequence Homology-Independent Protein Recombination (SHIPREC) (Sieber, Martinez, and Arnold, 2001). In the former, 3' and 5' fragments of decreasing length of each gene are obtained by treatment with exonuclease III. Random-length 3' fragments of one of the genes are fused to 5' fragments of the other gene, resulting in a library of hybrids between the two genes (**Figure 1.4d**). An obvious disadvantage of ITCHY is its lack of preservation of gene length, as well as recombination often taking place at sites that are not structurally related. SHIPREC attempted to generate crossover at structurally related sites by selecting fused products that matched the length of either of the parental genes. Nevertheless, both ITCHY and SHIPREC are limited to a single crossover per variant. With said limitation in mind, the developers of ITCHY combined their approach with a subsequent DNA shuffling step where functional variants extracted from a library generated through ITCHY are recombined. The methodology, known as SCRATCHY (Lutz *et al.*, 2001), explored sequence space inaccessible through other recombination techniques, but also displayed some undesired bias in the resulting library. Other methods, such as Sequence Independent Site Directed Chimeragenesis (SISDC) (Hiraga and Arnold,

2003), bypassed the need for sequence homology by pre-establishing the crossover points in the parental sequences. In SISDC, unique cleavable tags are inserted into each crossover point of the parental sequences. The tags can be cleaved by treatment with the appropriate endonuclease, generating site-specific sticky ends that can be annealed between fragments of different parental genes in the correct order to generate chimeric variants.

Recently, the focus of novel recombination methods has shifted towards the *in vivo* generation of libraries, allowing a direct coupling to selection techniques. One of the first such approaches was Mutagenic Organized Recombination Process by Homologous in vivo Grouping (MORPHING), where the authors took advantage of the naturally high frequency of homologous recombination in *Saccharomyces cerevisiae* (Gonzalez-Perez *et al.*, 2014). First, small segments of a parental sequence overlapping by approximately 50 base pairs with adjacent segments are amplified by PCR. Mutagenic PCR is performed for generating genetic variability of certain fragments, whereas standard PCR is chosen for fragments where high fidelity is desired. The amplified segments are then transformed into *Saccharomyces cerevisiae* together with a linearized vector backbone. Finally, the homologous recombination machinery generates a full-length novel variant of the gene integrated in the vector.

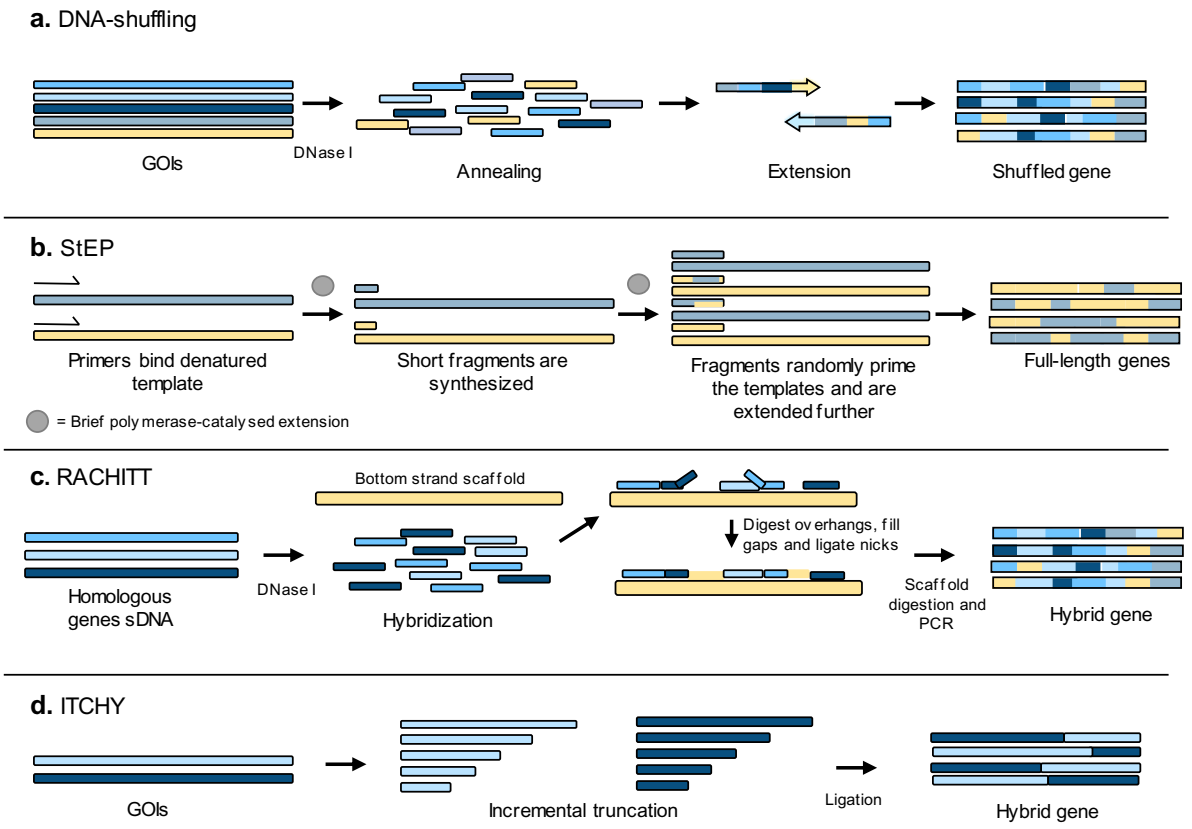


Figure 1.4. Recombination-based mutagenesis techniques. A set of some of the main mutagenesis techniques based on recombination is displayed. **a.** DNA shuffling was the first recombination-based *in vitro* mutagenesis technique to be developed. A set of homologous sequences is treated with DNase I,

and the resulting mix of fragments is used to reassemble full-length sequences through self-priming PCR. **b.** In StEP, a set of homologous sequences is used as templates for a series of cycles of annealing with primers and extension of the primers by a DNA polymerase. In each cycle, the growing primers can anneal with a different template, resulting in chimeric full-length sequences. **c.** RACHITT requires less homology between parental sequences than DNA shuffling and StEP. A set of fragments obtained by treatment with DNase I of the complementary strands of the sequences to be recombined is hybridized to a single-stranded copy of one of the parental sequences. After digesting overhangs, filling the gaps and ligating the nicks, the scaffold strands are digested, and double-stranded chimeric sequences are obtained by means of PCR. **d.** ITCHY decreases even further the sequence homology requirements, but it is limited to a single crossover per variant. Exonuclease III is used to incrementally truncate one of the parental genes from its 3' end, and the other one from its 5' end. Then, random-length fragments of each gene are ligated.

1.4.3. Rational mutagenesis

In contrast to random mutagenesis approaches, rational mutagenesis focuses on mutating only a limited number of positions in the target sequence. The positions to be mutated must be determined by means of prior information about which residues play a critical role in the presence or absence of the property to be evolved. Such information can be provided by several sources such as available structures of biomolecules, multiple sequence alignment, biochemical data and, more recently, computer-based predictions.

The most basic approach is site-directed mutagenesis (SDM), which involves performing a set of specific mutations by means of mutagenic PCR and evaluating the properties of the obtained variant. However, more often than not, rationally designed mutations do not have the expected effect. Furthermore, such strategy can be extremely time-consuming and laborious when the number of mutations to screen is increased to more than a few tens. Therefore, methodologies to quickly generate libraries where all possible mutations at a set of given positions are produced are required.

Site-saturation mutagenesis (SSM) enables fast and efficient generation of libraries where all possible substitutions of the chosen positions are included. The positions to be mutated must be determined based on prior information, such as available structures of the protein of interest or related proteins, multiple sequence alignments or biochemical data providing mechanistic information. Several computational methods that can be used to guide library generation when only limited prior information is available have been developed. Such methods rely on different criteria to predict sets of residues that are likely to play a critical role for the functionality of the protein (Reetz, Carballeira, and Vogel 2006; Nimrod *et al.*, 2005).

Several different strategies have been developed to obtain libraries with saturated positions, but most of them rely on the use of saturated mutagenic primers that mismatch the position to be mutated for PCR, an approach similar to that used for SDM. One of the first such protocols to be developed was overlap extension PCR (OE-PCR), where two sequential steps of

amplification are performed. In the first step, overlapping primers degenerated on the position to be mutated are used, resulting in two overlapping segments of the sequence with the saturated position. The full-length sequence is finally reconstituted in a second amplification step where primers annealing with both ends of the sequence are used (**Figure 1.5**). Further variations of this protocol have been implemented in order to simplify the process. For example, it is possible to reduce the mutagenesis process to a single PCR step by first inserting the target sequence into a circular plasmid and then using a pair of degenerated primers to amplify the whole plasmid. The parental sequences must then be eliminated with *DpnI* treatment, which selectively degrades methylated DNA.

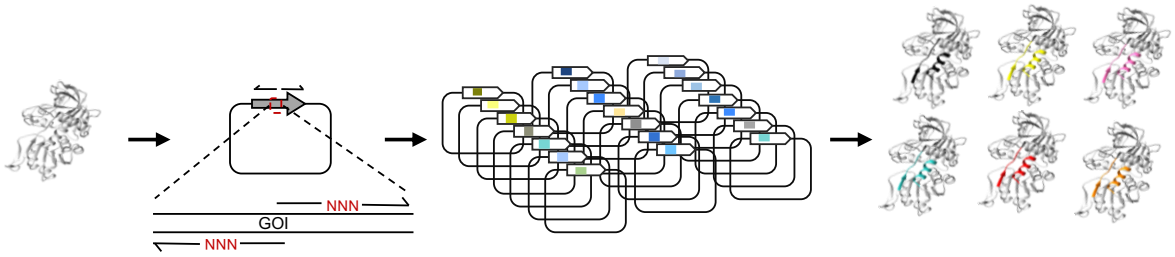


Figure 1.5. Site-saturation and site-directed mutagenesis. Site-saturation mutagenesis allows the introduction of a large range of point-mutations at specific sites, while site-directed mutagenesis introduces a specific set of mutations. In both cases, mutagenic primers, which contain mismatches with the parental sequence in the positions to be mutated, are used for PCR reactions with the parental sequence as the template. The amplification products are the mutated variants. In the case of site-directed mutagenesis, primers carrying specific mutations are used. For site-saturation mutagenesis, degenerated primers containing a range of possible mutations are employed.

While these protocols can be easily used to saturate one or a few very proximal codons (close enough to be included in one or two primers), more specialized strategies are required in order to generate libraries where several distant sites are saturated. One possibility is to use several pairs of mutagenic degenerate primers covering the full sequence, in a manner such that both the reverse primer for a given segment and the forward primer overlap partially and cover the same degenerated codon. The resulting products are separated by agarose (Multiple Overlap Extension PCR, MOE-PCR) or polyacrylamide (Polyacrylamide gel electrophoresis-mediated Overlap Extension PCR, POEP) gels, and full-length sequences are assembled by means of a second PCR with flanking primers (An *et al.* 2005; Peng, Xiong, and Yao, 2006). Another approach combines the use of flanking primers with additional extensions and phosphorylated mutagenic degenerate primers annealing with the different positions to be saturated. After mixing all primers with the target sequence, primers are extended by using T4-DNA polymerase. Treatment with DNA ligase then yields a single-stranded, saturated copy of the

original sequence including the additional extensions on both ends. PCR is then performed with primers annealing with the extensions to obtain the library.

However, the latter methodology is reliant upon simultaneous annealing of all mutagenic primers to their corresponding complementary sequences. This can sometimes be difficult to achieve when the thermodynamic parameters of each set of primers differ largely, due, for example, to different G-C contents or the presence of secondary structure. OmniChange was developed as an alternative to saturate up to five independent codons in a sequence-independent manner (Dennig *et al.*, 2011). In the OmniChange method, mutagenized fragments containing the degenerate positions are generated as in the MOE-PCR or POEP approaches, with the peculiarity that the twelve nucleotides in the 5' end of primers are bound by phosphorothiodiester bonds instead of the standard phosphorodiester bonds. The phosphorothioate-containing 5' ends are then cleaved by means of iodine-ethanol treatment, generating 5' single-stranded overhangs. Finally, the fragments are assembled by hybridization of the complementary overhangs and transformed into *E. coli*, which repairs the remaining nicks.

In addition to the exact mutagenesis protocol, another key parameter that needs to be decided before undertaking SSM is the degeneracy scheme. It is possible to degenerate the target codons to NNN (where N represents A, G, T or C) to obtain libraries where all possible 64 triplets are present. However, this methodology poses several problems. Firstly, since some amino acids are encoded by more possible codons than others, not all substitutions become equally frequent at the protein level. Furthermore, simply degenerating codons to all other possible triplets introduces premature stop codons, leading to a background of variants with loss of function. Finally, if a large number of codons is targeted, the library size can easily become excessive.

It is possible to reduce the frequency of stop codons by degenerating codons to NNK (where K represents either a G or a T), which reduces the library size by half since only 32 codons are tested at each position. Such degeneration scheme also presents the advantage of only including one possible stop codon instead of three, while still allowing for all 20 amino acids to be encoded in the resulting triplets (Neylon, 2004). More restrictive degeneration strategies can be implemented to completely remove stop codons and balance better the representation of amino acids with different chemical properties, albeit at the cost of not encoding all 20 amino acids (Neylon, 2004) (**Table 1.2**). Tsang *et al.* suggested a potential workaround, consisting of the use of mixtures of primers with restricted degeneration schemes that encode, as an ensemble, all possible substitutions with only one possible codon per amino acid. This allows the generation of "small-intelligent" libraries, where stop codons and rare codons can be

completely removed. DC-Analyzer was developed as a software tool to assist in the generation of such libraries, with the possibility to generate libraries including only polar or hydrophobic residues (Tang *et al.*, 2012). Nevertheless, the approach cannot be easily applied to cases where more than two sites must be randomized.

Table 1.2. Comparison of some of the most frequent degeneracy schemes for saturation mutagenesis. Nucleotide ambiguity is represented by the following code: N – A, G, C or T; K – T or G; S – G or C; D – A, G or T; B – C, G or T; R – A or G.

Degenerated codon	NNN	NNK/NNS	NDT	DBK	NRT
No. of codons	64	32	32	48	8
No. of amino acids	20	20	20	20	8
No. of stop codons	3	1	0	0	0
Encoded amino acids	All	All	RNDCGHILFSYV	ARCGILMFSTWV	RNDCGHSY

Some alternative attempts at improving the standard SSM protocols tackle the generation of optimal degenerated libraries from the point of view of the synthesis of the mutagenic primers. A near optimal solution was provided by the use of mixtures of 20 trinucleotide phosphoramidites, each encoding a different amino acid (Virnekäs *et al.,* 1994; Kayushin, Korosteleva, and Miroshnikov, 2000). Although this procedure allows for the elimination of codon bias and termination codons, the high cost of the technique has prevented its widespread application. Gaytán *et al*. developed a system to produce degenerated mutagenic primers in a cost-effective way where stop and redundant codons are eliminated, named TrimerDimer (Gaytán *et al.*, 2009). The method is based on the use of di and trinucleotides modified to have a protecting group in their 5' hydroxyl group: 9-fluorenylmethoxycarbonyl (Fmoc) in the case of the dinucleotides and 4,4'-dimethoxytrityl (DMTr) for the trinucleotides. Fmoc can be cleaved by treatment with an organic base, while DMTr is sensitive to weak acids. A set of five dinucleotides and five trionucleotides. Initially, the mutagenic primers begin to be synthesized by means of traditional solid phase synthesis with standard nucleotides. Once a position to be saturated is reached, the mixture of protected di- and trinucleotides is added, causing some of the growing oligonucleotides to be extended by three bases (resulting in the position to be degenerated to be mutated to five different codons), and the remaining by two bases. The oligonucleotides are then treated with a weak base to remove the Fmoc protecting groups, leaving only oligonucleotides that had been extended with a dinucleotide unprotected. DMTr-protected cytidine, adenine and guanine are then incorporated, yielding 15 additional random codons in the degenerated position. The synthesis of the primers proceeds then to completion in a traditional way, unless more positions are required to be saturated. The sequences of the dinucleotides and trinucleotides were chosen such that the resulting 20 random codons encode all 20 amino acids and no stop codons. Furthermore, the majority of

the codons are the preferred codons by *E. coli* for the corresponding amino acid. While TrimerDimer provides an attractive way of obtaining well-balanced libraries without variants with premature stop codons, its technical aspects and the required equipment make it not readily available to all standard molecular biology laboratories, limiting its scope.

1.5. Identification of the variants of interest

Thanks to the large variety of mutagenesis techniques developed in the last decades, it is possible to find a suitable approach for generating a library of variants of good quality for almost any biomolecule, or ensemble thereof, to be subjected to directed evolution.

Instead, the main bottleneck has become the identification of the variants of interest exhibiting the desired properties. Typical libraries generated in modern directed evolution experiments can include millions of variants, containing a vast majority of sequences without the desired properties or simply non-functional. Methods allowing for high-throughput, efficient evaluation of large libraries were then developed, resulting in two large groups of approaches: screening and selection.

Screening techniques aim to evaluate every variant for a certain property, such as enzymatic activity. While different strategies have been designed to increase the throughput rate as much as possible, screening techniques tend to be laborious and time-consuming, with a few exceptions such as fluorescence-based techniques. Selection approaches, on the other hand, bypass the need to assess the functionality of each variant, and link the presence of the property to be evolved to a physical separation of the corresponding coding sequence or to increased survival rate. This type of strategies automatically discards all non-functional variants, allowing analysis of larger libraries and at a much higher throughput. While in principle selection strategies appear to offer considerable advantages over screening, the available techniques are relatively specific and cannot be easily applied to a broad range of biomolecules or to select considerably different properties, limiting their widespread application.

1.5.1. Screening techniques

The most basic screening techniques rely on spatially separating each variant and then assessing each one individually with a measurement that can be directly related to the activity of the variant. This is typically achieved by expressing the library of variants in a model organism, such as *E. coli*, and plating in solid media in order to isolate colonies corresponding to clones containing a single variant. In some cases, it is possible to perform directly

colorimetric or fluorescent-based measurements on the colonies obtained in solid-phase culture thanks to automated digital imaging techniques (**Figure 1.6a**). It is also possible to transfer the colonies to multi-well liquid culture devices, and then perform further analysis with the liquid cultures themselves or cell lysates on microtiter plates, which allows tracking hundreds or thousands of assays at the same time and can be handled by robotic systems to increase the throughput (**Figure 1.6b**).

Some biomolecules possess spectral properties that can be directly screened, such as fluorescent proteins. There are indeed multiple cases of successful identification of variants of GFP with more intense fluorescence (Crameri *et al.*, 1996), or altered absorption or emission spectra, where even a variant emitting blue light was identified (Heim, Prasher, and Tsien, 1994).

However, most frequently it is the consumption of a substrate, or the formation of a product, that is measured. When either a substrate or a product has a distinctive absorbance or fluorescence peak at a specific wavelength (usually in the UV-visible spectrum), it is possible to use its signal to evaluate enzymatic activity. Plenty of directed evolution cases where such screening methods were employed are available. Tellier and colleagues obtained a variant of *Thermus thermophilus* β-glycosidase with increased transglycosydase activity and reduced hydrolysis rate by direct imaging of colonies (Koné *et al.*, 2009). When glycosyl-X was added to the solid culture medium but no other substrate was present to perform transglycosylation, colonies with low hydrolytic activity did not gain an intense blue coloration. When the partner substrate was also added, colonies with increased transglycosydase activity became blue, due to the release of the X blue product. By combining both criteria, variants with up to 70-fold increase in the transglycosydase/hydrolysis activity ratio were identified.

If no substrate or product is readily measurable by spectrophotometric techniques, it is possible to use a surrogate substrate with the desired spectral or fluorescent properties to assess the variants. However, after finding potential variants of interest through the use of the surrogate substrate, it is necessary to reassess the candidates again with the original intended substrate to confirm that the optimization observed for the surrogate substrate also applies to the original substrate. Alternatively, the reaction catalysed by the enzyme can be coupled to a secondary reaction which can be monitored spectrophotometrically. Finally, when none of these options is suitable, it is possible to quantify specific compounds through techniques such as nuclear magnetic resonance (NMR), high-performance liquid chromatography (HPLC), gas chromatography and mass spectrometry.

In spite of the simplified and faster handling achieved through robotics and automation, screenings based on spatial separation of each variant remain limited due to the large material

and physical requirements necessary to perform the evaluation of a large number of variants. These limitations can be overcome to some extent by directly analysing the different variants in bulk without previous separation. This became first possible by combining the genotype-phenotype association provided by individual cells with the automated separation of cells achieved through fluorescence-activated cell sorting (FACS) (**Figure 1.6c**).

FACS allows the separation of individual cells by means of flow cytometry based on one or more fluorescent signals, allowing screening of up to 10^8 library variants in around 2 hours (Chen, Dorr, and Liu, 2011), with the condition that the activity of the biomolecule of interest must be linked, in some way, to a change in fluorescence. By using GFP as a reporter gene, the methodology has been exploited in several creative ways. For example, Santoro and Schultz placed two variants of GFP (GFPuv and EYFP) in a reporter plasmid containing two loxP sites and a T7/lac promoter upstream of the EYFP sequence and one of the loxP sites. Cells were transformed with the reporter plasmid and a plasmid carrying a variant of the Cre recombinase. Such an arrangement allows the detection of recombination events by simply measuring the fluorescence pattern: while the unaltered reporter plasmid only allows the expression of EYFP, cells with an active variant of the Cre recombinase would contain a mixture of unaltered and recombined reporter plasmids, leading to equal expression of EYFP and GFPuv. Two types of reporter plasmids were created: the first one carried wild-type loxP sites, whereas the second one included modified loxP sites. By combining a positive screening for variants able to recombine the altered site, and a negative screening for the lack of ability to recombine the wild-type site, they obtained a Cre variant able to operate with similar efficiency as wild-type Cre (Santoro and Schultz, 2002).

It is also possible to use fluorescent substrates that the enzyme of interest converts to a fluorescent product which, unlike the substrate, is unable to leave the cell. After washing off the permeable substrate, cells containing an active variant of the enzyme can be easily identified through the fluorescence of the product. Through the conjugation of standard substrates with fluorogenic molecules, this methodology, known as product entrapment, has been applied to a variety of cases, including glycosyl-transferases, glutathione transferase and β-galactosidase (Yang and Withers, 2009).

FACS also became a suitable method for screening protein-protein interactions thanks to the development of the yeast display technique. In short, a library of variants with one of the binding partners tagged with a known epitope and fused to the yeast adhesion receptor Aga2 is expressed and displayed on the surface of yeast cells. The cells are then exposed to the target protein fused to a second epitope. After washing the unbound protein, cells are treated with fluorescently-labeled antibodies directed against both epitopes, allowing the identification

of cells displaying a variant with the desired binding properties by selecting those exhibiting fluorescent signal corresponding to both antibodies. A modification of yeast display enabled its application to the selection of sortases, a group of enzymes involved in the cleavage of C-terminal peptides which also display weaker transpeptidase activity. Chen and colleagues aimed to obtain a variant of *Staphylococcus aureus* sortase A (SrtA) with increased transpeptidase activity. A library of SrtA variants was fused to Aga2 and displayed in cells where a triglycine acceptor peptide had been fused to Aga1, another yeast surface protein covalently linked to Aga2 by disulfide bonds. The cells were incubated with the peptide substrate for bioconjugation, labeled with biotin. Finally, cells were incubated with a conjugate of streptavidin and a fluorescent molecule, correlating a stronger fluorescent signal with enhanced bioconjugation and allowing the identification of a variant with a 140-fold increase in transpeptidase activity (Chen, Dorr, and Liu, 2011).

There have also been attempts to replace the natural compartmentalization provided by cells with artificial *in vitro* compartmentalization (IVC) (**Figure 1.6d**). IVC offers some advantages over cell compartmentalization, most notably bypassing the need to transform the library, which can limit the effective size of the screened library. Several ways to generate artificial compartmentalization have been proposed, but water-in-oil (w-o) and water-in-oil-in-water (w-o-w) emulsions remain as the most widely used. The application of IVC compartmentalization in directed evolution was first demonstrated by Tawfik and Griffiths, which used it to identify genes encoding the *HaeIII* DNA methyltransferase amongst a 107-fold excess of genes encoding other enzymes (Tawfik and Griffiths, 1998). They generated w-o emulsions where each droplet contained an *in vitro* transcription-translation system, S-adenosyl methionine (the methyl group donor) and, on average, one of the genes present in the library, linked to an additional substrate sequence containing the target site of the *HaeIII* DNA methyltransferase (*HaeIII* R/M). With such a setup, the *HaeIII* R/M site becomes methylated only in droplets containing a gene encoding the methyltransferase, which renders it resistant to cleavage by the *HaeIII* restriction enzyme. After breaking the emulsion, the resulting sequences were amplified through PCR, revealing enrichment factors of up to 5000 times in favour of the methyltransferase.

IVC can also been coupled to FACS, with the requirement that w-o-w must be used in order for the flow cytometer to be able to handle the emulsion. Tawfik and collaborators employed this approach to generate a nanoemulsion containing variants of mammalian paraoxonase 1, which displays low activity towards sarin-like nervous agents. The droplets included fluorogenic surrogates of the nervous agents where a fluorine substituent was replaced with coumarin. By using FACS to measure coumarin fluorescence, a variant with 105 times higher

activity towards cyclosarin was identified, which even demonstrated *in vivo* prophylactic effect (Gupta *et al.*, 2011).

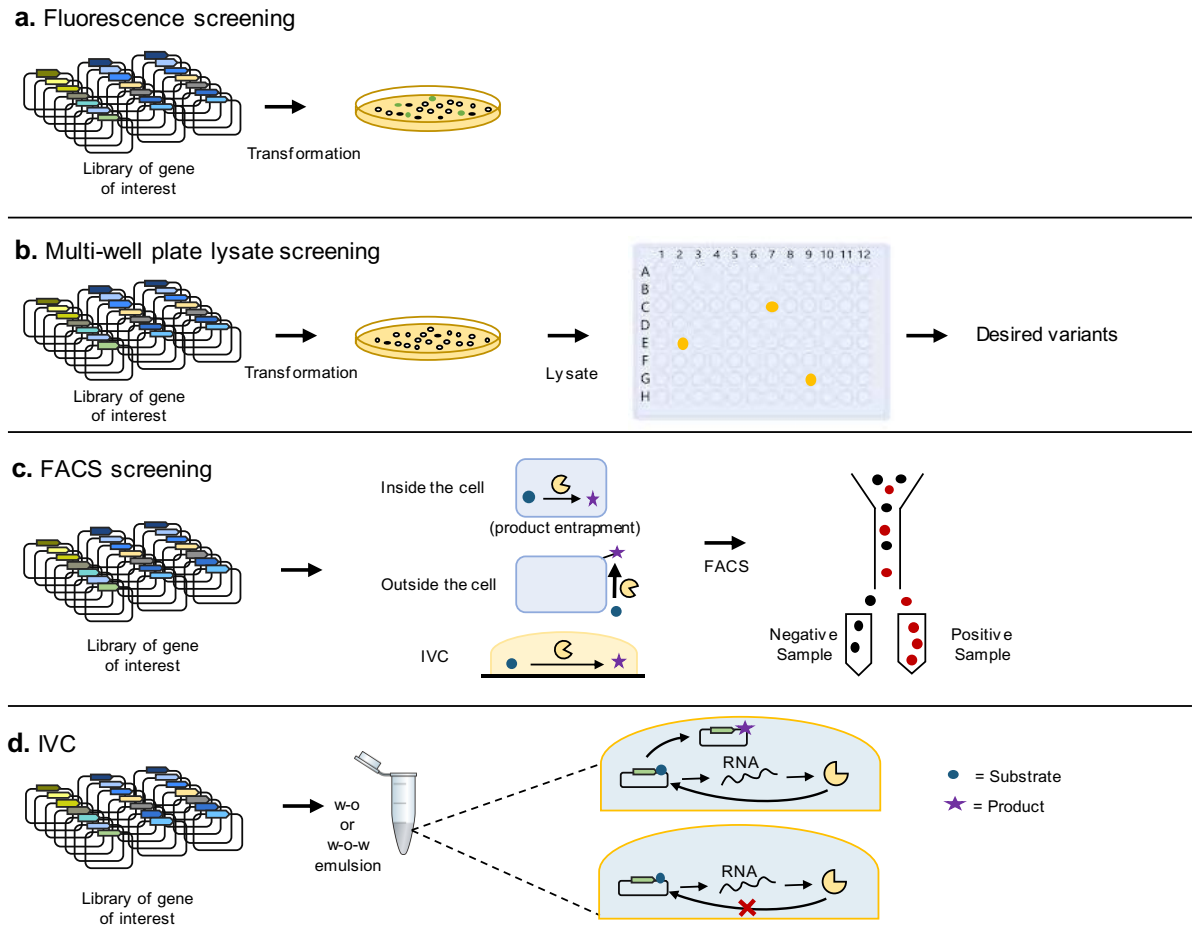


Figure 1.6. Screening techniques. Some of the most frequent screening techniques are depicted. **a.** Variants of proteins that confer fluorescence can be screened by analysing with digital imaging techniques cultures in solid media. **b.** For proteins whose activity can be linked to a colorimetric assay or to the generation of fluorescence, it is possible to automatically transfer individual colonies to liquid cultures by means of automated multi-well liquid culture devices. The liquid cultures or their lysates can then be screened by colorimetric or fluorescent-based assays. **c.** FACS enables the physical separation of individual cells based on their fluorescence properties, allowing for a higher throughput and reduced material and physical requirements. However, it is limited to biomolecules whose activity can be linked to a change in fluorescence. **d.** IVC techniques replace the compartmentalization provided by cells with artificial compartmentalization, most frequently provided by emulsions of water and oil. This allows to bypass the limitation imposed by transformation efficiency, but incompatibilities between the conditions required for transcription and translation and those required for the activity of the biomolecule of interest reduce its scope of application.

1.5.2. Selection techniques

Selection methods allow a higher throughput than screening methods since they are intrinsically designed in a way that automatically discards all undesired variants. They tend to be, however, less generally applicable to a broad range of biomolecules than their screening counterparts, which has led to the development of a large variety of selection schemes for

different directed evolution applications. Two main categories of approaches can be distinguished, depending on whether the activity of a variant of interest leads to a physical segregation of its encoding sequence or to an increased survival rate of the carrying organism.

The first set of techniques, also called display techniques, typically rely on a physical connection between a nucleic acid sequence and the product it encodes, and are generally applied to obtain variants with improved binding to a certain target. Several means of achieving such linking have been devised. The most common method employs the previously described phage display technique to expose variants in the surface of phages which contain the gene encoding the corresponding variant (**Figure 1.7a**). It has indeed become one of the most powerful techniques to select peptides and antibody fragments with specific binding properties, leading to the development of several tens of antibody-based pharmaceuticals which are in late-stage clinical trials or have already been approved (Nixon, Sexton, and Ladner, 2014). The technique has been modified to use other viruses in addition to filamentous phages which allow for better expression and display of certain proteins, such as the T7 phage for cytoplasmic proteins (Binz *et al.*, 2004) or retroviruses and other eukaryotic viruses for eukaryotic proteins (Urban and Merten, 2011).

Another conceptually simpler approach is plasmid display (**Figure 1.7b**), where each encoded variant is fused to a DNA binding protein (such as the p50 subunit of NF-κB) and placed into a vector containing the sequence recognized by the DNA binding protein. After expressing the library of variants in an appropriate organism, the cells are lysed and the resulting variant-plasmid complexes can be subjected to selection by exploiting a functional property of the variant proteins, such as binding affinity for a specific target. For example, Samuelson and colleagues generated a library of the N-glycan binding protein Fbs1 fused to NF-κB p50, and used beads with fetuin (a serum glycoprotein containing complex sialylated N-glycans) to select Fbs1 variants able to bind more efficiently and with less bias than wild-type to a broader range of N-glycans (Chen *et al.*, 2017).

It is also possible to link the protein variants to their corresponding mRNA, instead of the coding DNA. Such techniques bypass the transformation step required in other display methods, which limits the library size. One of the first methodologies to achieve this linking was ribosome display (**Figure 1.7c**), developed in 1997 by Hanes and Plückthun (Hanes and Plückthun, 1997) based on the previously devised polysome display (Mattheakis, Bhatt, and Dower, 1994). In ribosome display, the association between protein variants and the corresponding mRNA is maintained through the interaction of both molecules with the ribosome. The association of full-length proteins with the ribosome is achieved through the removal of the stop codon, which causes the nascent peptide to remain attached to the last

tRNA. Furthermore, after translation is stopped by cooling on ice, the ribosome complexes are further stabilized through the addition of magnesium, making them suitable for functional selection and subsequent identification of the selected coding sequences through reverse transcriptase-PCR. The technique has been mostly applied to the selection of proteins with enhanced binding properties, such as antibodies.

The need of using the ribosome to mediate the interaction between protein and mRNA can be completely bypassed with the mRNA display technique, developed by the laboratory of Jack Szostak (Wilson, Keefe, and Szostak, 2001). Szostak and collaborators generated mRNA from a DNA library with *in vitro* transcription and then ligated them to short DNA oligonucleotides with puromycin attached to their 3' end. Puromycin is an inhibitor of protein translation with a molecular structure that mimics the 3' end of an aminoacyl-tRNA, and that can be incorporated by the ribosome to a nascent peptide. At low concentrations, it is only able to do so during the relatively slow termination step, resulting in puromycin-linked full length proteins (Miyamoto-Sato *et al.*, 2000). This is the case in the setup designed by Szostak and collaborators, which allowed them to covalently link translated full-length peptides to their corresponding mRNA through puromycin. It is then possible to recover the DNA sequence of selected variants by amplification through reverse transcriptase-PCR. mRNA display has been extensively used to obtain small peptides and proteins able to bind a large variety of partners, including RNA (Barrick and Roberts, 2002), small molecules (Keefe and Szostak, 2001) and other proteins (Xu *et al.*, 2002). Seelig and Szostak further tweaked the technique to extend its scope to the selection of bond-forming enzymes, such as ligases. In their modified setup, an additional step of reverse transcription is added after the translation of the protein variants, performed with primers linked to one of the substrates of the enzyme. The other substrate is then conjugated to an anchor group such as biotin and added to the library. Thus, if a library variant is able to efficiently catalyse the bonding of both products, biotin becomes covalently linked to its corresponding cDNA, which can be easily isolated with streptavidin beads. The authors demonstrated the viability of the technique by choosing two RNA molecules as the substrates and using it to select novel RNA ligases evolved from a library generated from the human retinoid-X-receptor, a protein with a zinc-finger domain which does not display any ligase activity (Seelig and Szostak, 2007).

Unlike display techniques, growth complementation techniques are always performed *in vivo*. In growth complementation approaches, the fitness of cells to survive in the experimental conditions becomes linked to the activity desired to be evolved in the biomolecule of interest. Growth complementation techniques offer the advantage of being easily coupled to *in vivo* mutagenesis methods, holding the potential to develop automated continuous evolution techniques. However, so far no widely-applicable growth complementation method has been

developed, with most techniques being restricted to either a narrow range of biomolecules or very specific properties.

Proteins conferring antibiotic resistance are one of the biomolecules most easily evolved by means of growth complementation techniques, given the facility to link their activity to increased survival rate by simply adding the antibiotic towards which resistance is sought to the culture medium (**Figure 1.7d**). Several examples exist where variants of natural antibiotic resistances have been evolved to confer enhanced or novel protection against a particular antibiotic, including proteins such as efflux pumps and β-lactamases (Nyerges *et al.,* 2018; Feiler *et al.*, 2013; Sun *et al.*, 2013; Bokma *et al.*, 2006).

Enzymes involved in essential metabolic reactions can also be easily selected by using auxotrophic mutants, with the only requisite that the enzymatic activity of interest must result in the generation of the component for which the strain is auxotrophic (**Figure 1.7e**). For example, Boersma *et al.* used an aspartate auxotrophic *E. coli* strain to select an enantioselective lipase variant of *Bacillus subtilis* lipase A with high preference towards the (S) isomer of 1,2-O-isopropylidene-sn-glycerol (IPG), a precursor for the synthesis of β-adrenoceptors antagonists (Boersma *et al.*, 2008). In order to do so, cells were cultured in minimal media supplemented with aspartate ester of (S)-(+)-IPG. Upon hydrolysis of the ester, performed more efficiently by lipase variants with increased enantioselectivity for the (S) substrate, aspartate is released and allows auxotrophic cells to grow. Additionally, variants with reduced activity towards (R)-(-)-IPG were achieved through the addition to the culture medium of a phosphonate ester of (R)-(-)-IPG, a suicide inhibitor of lipase A. By means of such dual selection, a variant with similar activity as the wild-type enzyme but an enantiomeric excess of over 70% towards the (S) isomer was achieved.

It is also possible to couple, in an indirect way, the activity of the biomolecule of interest to increased survival rate by linking it to a secondary activity which becomes ultimately responsible for growth recovery, such as an antibiotic resistance (**Figure 1.7f**). A common approach consists of placing the sequence required for growth (such as an antibiotic resistance) under the control of a transcriptional regulator responsive to the substrate or product of an enzyme of interest. For example, in the QUEST method, the selection marker is placed under a synthetic transcriptional activator comprising the AraC DNA-binding domain and a domain able to bind both a molecule inducing dimerization and the substrate of the enzyme of interest, which compete with each other. If both the dimerization-inducing molecule and the substrate are added to the culture medium, only upon conversion of the substrate to a product by the enzyme of interest the transcription factor dimerizes and the selection marker is expressed (Firestine *et al.*, 2000).

There have also been successful attempts at mimicking the growth complementation principle by using *in vitro* compartmentalization. Selection based on *in vitro* compartmentalization is better suited to select enzymes which can directly or indirectly influence the efficiency of the replication or transcription of their coding sequences. For example, Holliger and collaborators compartmentalized DNA polymerases and their coding sequences in a w-o emulsion and provided flanking primers and dNTPs into each aqueous compartment (Ghadessy, Ong, and Holliger, 2001). They then performed PCR cycles, where a higher yield of DNA was obtained in compartments with the most active polymerases, resulting in enrichment of the sequences encoding the fittest variants. By performing several rounds of selection at increasingly higher temperatures or heparin concentrations, they managed to obtain a Taq polymerase variants with enhanced thermal stability and reduced heparin inhibition.

More recently, continuous evolution methods have also been devised, where the required human intervention is minimized. Such techniques couple a mutagenesis system (ideally targeting specifically the gene of interest) with a selection system under continuous dilution. One of the most notorious examples is phage-assisted continuous evolution (PACE), developed by the laboratory of David Liu (Esvelt, Carlson, and Liu, 2011). In the PACE system, the gene of interest replaces gene III of the M13 bacteriophage genome, an essential gene encoding pIII. The modified phages are then used to infect a constant flow of host cells carrying an inducible mutagenesis plasmid (including a defective version of the DNA polymerase proofreading subunit dnaQ296 and an error-prone DNA polymerase) and an accessory plasmid carrying the bacteriophage essential gene III. The system allows for the continuous evolution of the gene of interest as long as its activity can be linked to the production of pIII.

As evidenced by the discussed methodologies, while screening and selection techniques have demonstrated their efficacy in specific cases of directed evolution (such as display techniques for the selection of variants with desired binding properties), none of them allows high-throughput identification of variants of interest of a wide range of biomolecules and properties without requiring complex modifications to the standard protocol. In particular, the available techniques struggle to be routinely applied to obtain variants a large range of enzymes with altered or improved catalytic properties, one of the main targets of directed evolution.

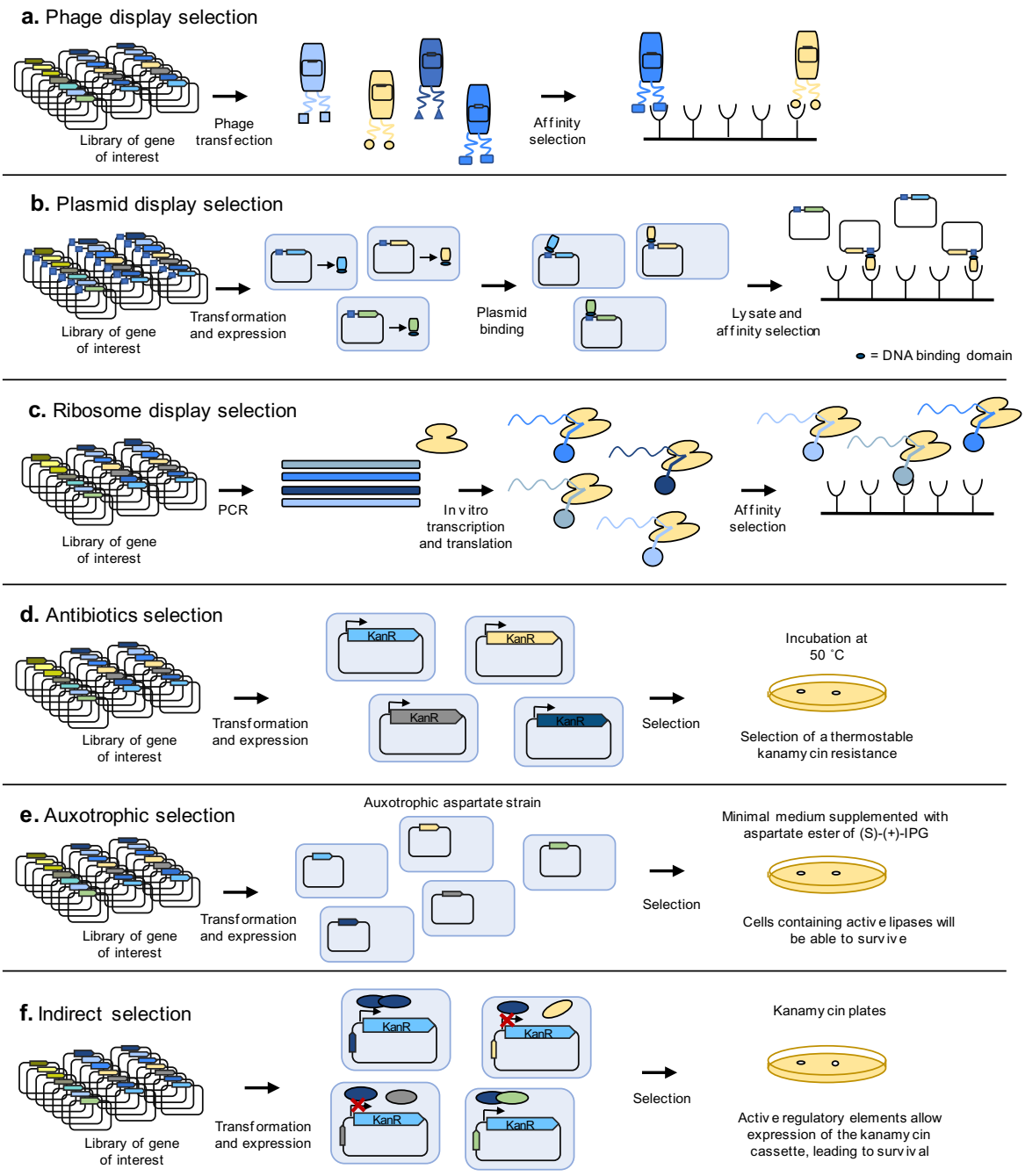


Figure 1.7. Selection techniques. Some of the most frequent selection techniques are represented. **a.** In phage display, protein or peptide variants are exposed on the surface of phages, and selected based on their binding affinity to a target binding partner. **b.** In plasmid display, a DNA-binding protein is fused to each variant. The encoding plasmid contains the target sequence for the DNA-binding protein. After lysing cells, variants can be selected based on their binding affinity to specific target interactors. Variant sequence can then be determined from the associated plasmid thanks to the linkage provided by the DNA-binding protein. **c.** Ribosome display can be used to link protein variants to their corresponding mRNA. Translation is stopped by cooling on ice, and protein-ribosome-mRNA complexes are stabilized through the addition of magnesium, enabling affinity selection and amplification of the sequence of selected variants by treatment with reverse transcriptase, followed by PCR. **d.** Antibiotic resistance selection is one of the most basic types of growth complementation selection techniques. Cells are transformed with a library of variants and grown in selective medium supplemented with a certain antibiotic. Only cells expressing a variant able to confer resistance for the added antibiotic and functional under the selection conditions (such as high temperature) will survive. **e.** In auxotrophy-based selections, cells auxotrophic for a certain metabolite are grown in minimal

medium without said metabolite but with a precursor that can yield the required compound upon transformation by a certain enzymatic activity. Cells are transformed with a library of variants, such that only those carrying a variant able to catalyse the conversion of the precursor will be able to survive. **f.** In indirect growth-complementation based selection techniques, the activity of the gene of interest is not directly responsible for an increased survival rate. Instead, its activity (such as activation of transcription) leads to an increased expression or activity of the biomolecule directly responsible for it, such as an antibiotic resistance.

1.6. Aims and objectives

Directed evolution is one of the most powerful tools for protein engineering. The first step to carry out directed evolution is obtaining a library of variants of the coding sequence of interest. Thanks to the development of molecular biology techniques, a wide variety of mutagenesis methods is available, which facilitate the generation of large, high-quality libraries of almost any biomolecule of interest.

The subsequent identification and isolation of the variant of interest is often the bottleneck of directed evolution projects. While progress has been made in the development of screening and selection methods, major limitations still hinder the application of directed evolution in a general basis. Screening approaches tend to be time-consuming because variants have to be individually evaluated, and most of the selection methods cannot be broadly applied to different types of biomolecules or properties.

The aim of this PhD project was to develop a new selection methodology broadly applicable to the evolution of a wide range of biomolecules and properties, thus overcoming the main bottleneck of directed evolution.

Firstly, in Chapter 3, the metabolism of *E. coli* is introduced, with special emphasis on the requirement of NADH oxidation under anaerobic fermentation conditions. It was hypothesized that an *E. coli* strain with its main fermentative pathways impaired could be used as the basis for a broadly applicable selection system thanks to the metabolic defect imposed by the inability to regenerate oxidised NAD^+. Such a strain (AL strain, Δ*adhE* Δ*ldhA*) was built and the possibility of metabolic complementation was validated with several enzymes.

In Chapter 4 the potential of the AL mutant strain was explored by using it to select variants with altered cofactor specificity of two well-characterised alcohol dehydrogenases for which previous attempts at reversal of cofactor preference had been unsuccessful.

In Chapter 5, the selection system was applied to an imine reductase, an enzyme class of high industrial interest thanks to their ability to produce chiral amines with high enantiospecificity. Since all known naturally-occurring imine reductases exhibit a strong preference for NADP(H), the aim was to obtain an NAD(H)-dependent variant with a considerably simpler workflow than in previous studies.

In Chapter 6, the potential of the selection system to select enzyme variants with modified substrate selectivity was demonstrated, one of the properties most frequently required to be tailored to make an enzyme suitable for specific applications. Nitroreductase was chosen as a test case, given their applications in a large variety of fields, including clinical therapies, bioremediation and industrial processes.

Finally, in Chapter 7, the scope of the selection system was expanded by using it to select variants of an isopropanol-producing metabolic pathway with optimal product yields differing in the regulatory elements controlling the expression of each component gene. Doing so provided proof of the ability of the technique to select complex multi-gene systems, as well as to select a wide range of properties in biomolecules other than enzymes.

Through the range of different test cases, the aim was to demonstrate the broad applicability of the selection system in a series of real cases of practical interest, where, if successful, the evolution of novel improved biomolecules could have a direct impact on the corresponding industrial processes.

2. Materials and Methods

Chapter 2 describes the materials and methods used for the experiments presented in this thesis.

2.1. Materials

Except where otherwise specified, all chemicals, biochemicals and bacterial growth media were supplied by Sigma-Aldrich, Fisher Scientific or VWR. Enzymes and buffers for molecular biology techniques were supplied by New England Biolabs (NEB). The following tables present the microorganisms, plasmids and synthetic DNA sequences used in this project. Synthetic DNA sequences were codon-optimised for expression in *E. coli* with the IDT codon optimisation tool.

Table 2.1. Microorganisms used in this project.

Strain	Genotype	Source
***Escherichia coli* - DH5α**	F^-, φ80dlacZΔM15, Δ(lacZYA-argF) U169, deoR, recA1, endA1, hsdR17(r_K^-, m_K^+), phoA, supE44, λ^-, thi-1, gyrA96, relA1	Invitrogen
***Escherichia coli* - DH10β**	F^- mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 endA1 araD139 Δ(ara-leu)7697 galU galK λ– rpsL(StrR) nupG	Invitrogen
***Escherichia coli* - BW25113**	Wild-type isolate	*E. coli* genetic stock center (CGSC)
***Escherichia coli* - BL21(DE3)**	F^- *ompT* $hsdS_B$ ($r_B^-m_B^-$) *gal dcm rne131* λDE3	Invitrogen
***Escherichia coli* - JW1375**	*Δ*ldhA	Thermo Fisher Scientific (Baba, *et al.*, 2006)
***Escherichia coli* - JW1375**	*Δ*adhE	Thermo Fisher Scientific (Baba, *et al.*, 2006)
***Escherichia coli* - JW1595**	*ΔsthA*	Thermo Fisher Scientific (Baba, *et al.*, 2006)
***Escherichia coli* - JW1595**	*ΔpntA*	Thermo Fisher Scientific (Baba, *et al.*, 2006)
***Escherichia coli* – AL**	*Δ*ldhA, *ΔadhE*	This work
***Escherichia coli* – ALS**	*Δ*ldhA, *ΔadhE, ΔsthA*	This work
***Escherichia coli* – ALP**	*Δ*ldhA, *ΔadhE, ΔpntA*	This work
Escherichia *coli* – ALPS	*Δ*ldhA, *ΔadhE, ΔsthA, ΔpntA*	This work

Table 2.2. Plasmids used in this project.

Plasmid	Description	Antibiotic resistance	Source
pUC19	High copy expression vector (pMB1 ORI) with lacZα – **Negative control** for anaerobic growth complementation experiments	AmpR	Heap Laboratory (Norrander, Kempe, and Messing, 1983)
pCP20	Contains FLP recombinase, temperature sensitive replication origin	AmpR	Heap Laboratory (Cherepanov and Wackernagel, 1995)
pJET1.2	Selection cloning vector	AmpR	Invitrogen
pET28a	Bacterial expression vector with 6xHis-tag	KanR	Invitrogen
pMAK705	Plasmid for knock-out of target genes	CmR	Heap Laboratory (Hamilton, *et al.*, 1989)
pJH28	Plasmid containing *cbadh* gene from *Clostridium beijerinckii* (Uniprot: P25984)	CmR	Heap Laboratory
pStA0	Start-stop Assembly storage vector	AmpR	Heap Laboratory (Taylor *et al*., 2019)
pStA1AB	Start-stop Assembly Level 1 vector (A and B fusion sites)	TetR	Heap Laboratory (Taylor *et al*., 2019)
pStA1BC	Start-stop Assembly Level 1 vector (B and C fusion sites)	TetR	Heap Laboratory (Taylor *et al*., 2019)
pStA1CD	Start-stop Assembly Level 1 vector (C and D fusion sites)	TetR	Heap Laboratory (Taylor *et al*., 2019)
pStA1DE	Start-stop Assembly Level 1 vector (D and E fusion sites)	TetR	Heap Laboratory (Taylor *et al*., 2019)
pStA1EZ	Start-stop Assembly Level 1 vector (E and Z fusion sites)	TetR	Heap Laboratory (Taylor *et al*., 2019)
pGT323	pStA0 containing the BBa_J23100 promoter	AmpR	Heap Laboratory (Taylor *et al*., 2019)
pGT324	pStA0 containing the BBa_J23102 promoter	AmpR	Heap Laboratory (Taylor *et al*., 2019)
pGT326	pStA0 containing the BBa_J23107 promoter	AmpR	Heap Laboratory (Taylor *et al*., 2019)
pGT327	pStA0 containing the BBa_J23116 promoter	AmpR	Heap Laboratory (Taylor *et al*., 2019)
pGT328	pStA0 containing the BBa_J23113 promoter	AmpR	Heap Laboratory (Taylor *et al*., 2019)
pGT336	pStA0 containing the BBa_J23118 promoter	AmpR	Heap Laboratory (Taylor *et al*., 2019)
pGT330	pStA0 containing the RBSc13 RBS	AmpR	Heap Laboratory (Taylor *et al*., 2019)

pGT331	pStA0 containing the RBSc33 RBS	AmpR	Heap Laboratory (Taylor *et al*., 2019)
pGT332	pStA0 containing the RBSc44 RBS	AmpR	Heap Laboratory (Taylor *et al*., 2019)
pGT333	pStA0 containing the RBSc58 RBS	AmpR	Heap Laboratory (Taylor *et al*., 2019)
pGT334	pStA0 containing the RBSc36 RBS	AmpR	Heap Laboratory (Taylor *et al*., 2019)
pGT335	pStA0 containing the RBSc42 RBS	AmpR	Heap Laboratory (Taylor *et al*., 2019)
pGT337	pStA0 containing Terminator 1 (L3S2P55)	AmpR	Heap Laboratory (Taylor *et al*., 2019)
pGT338	pStA0 containing Terminator 2 (L3S2P21)	AmpR	Heap Laboratory (Taylor *et al*., 2019)
pGT339	pStA0 containing Terminator 3 (ECK120033737)	AmpR	Heap Laboratory (Taylor *et al*., 2019)
pGT340	pStA0 containing Terminator 4 (ECK1200196000)	AmpR	Heap Laboratory (Taylor *et al*., 2019)
pStA212	Start-stop Assembly Level 2 vector (fusion sites 1 and 2) – **Negative control** for isopropanol pathway	KanR	Heap Laboratory (Taylor *et al*., 2019)
pLS1	pUC19 containing *adhE* gene from *E. coli* (Uniprot ID: P0A9Q7) – **Positive control** for anaerobic growth complementation experiments	AmpR	This work
pLS2	pUC19 containing *bdhA* gene from *Bacillus subtilis* (Uniprot ID: O34788) - **BDHA**	AmpR	This work
pLS3_1	pJEt1.2 storage plasmid containing *budC* gene from *Klebsiella pneumoniae* (Uniprot ID: Q48436)	AmpR	This work
pLS3	pUC19 containing *budC* gene from *Klebsiella pneumoniae* (Uniprot ID: Q48436) - **BUDC**	AmpR	This work
pLS6	pUC19 containing *cbadh* gene from *Clostridium beijerinckii* (Uniprot: P25984) – **$CBADH_{WT}$**	AmpR	This work
pLS7	pUC19 containing mutated *cbadh* gene from *Clostridium beijerinckii* (S199G, R200G, Y218F) - **$CBADH_{R1}$**	AmpR	This work
pLS8	pUC19 containing mutated *cbadh* gene from *Clostridium beijerinckii* (G198D, S199G, R200G, Y218F) - **$CBADH_{R2}$**	AmpR	This work
pLS10	Library of pUC19 containing mutated *cbadh* gene from *Clostridium* (saturated mutagenesis to four amino acids position: 198,199,200,218) – **$CBADH_{Lib}$**	AmpR	This work

pLS10_3	pUC19 containing selected variant of *cbadh* gene from *Clostridium beijerinckii* (G198D/S199Y/R200R/Y218P) - **CBADH$_S$**	AmpR	This work
pLS12_1	pJEt1.2 storage plasmid containing *tadh* gene from *Thermus* sp. ATN-1 (Uniprot ID: B2ZRE3)	AmpR	This work
pLS12	pUC19 containing *tadh* gene from *Thermus* sp. ATN-1 (Uniprot ID: B2ZRE3) - **TADH**	AmpR	This work
pLS26	pLS6 with a C-terminal 6xHis-tag	AmpR	
pLS39	pMAK705 with 500 bp upstream and 500 bp downstream of *sthA* from *E. coli*	CmR	This work
pLS40	pMAK705 with 500 bp upstream and 500 bp downstream of *pntA* from *E. coli*	CmR	This work
pLS46	pStA0 containing *atoB* gene from *E. coli* (Uniprot ID: P76461)	AmpR	This work
pLS47	pStA0 containing *atoD* gene from *E. coli* (Uniprot ID: P76458)	AmpR	This work
pLS48	pStA0 containing *atoA* gene from *E. coli* (Uniprot ID: P76459)	AmpR	This work
pLS49	pStA0 containing *adc* gene from *Clostridium acetobutylicum* (Uniprot ID: P23670)	AmpR	This work
pLS50	pStA0 containing *cbadh* gene from *Clostridium beijerinckii* (Uniprot ID: P25984)	AmpR	This work
pLS53	pStA1AB containing full transcription unit coding for acetyl-CoA acetyltransferase from *E. coli* (*atoB*) (library of six promoters and six RBS, Terminator 1)	TetR	This work
pLS54	pStA1BC containing full transcription unit coding for acetate CoA-transferase subunit alpha from *E. coli* (*atoD*) (library of six promoters and six RBS, Terminator 2)	TetR	This work
pLS55	pStA1CD containing full transcription unit coding for acetate CoA-transferase subunit beta from *E. coli* (*atoA*) (library of six promoters and six RBS, Terminator 3)	TetR	This work
pLS56	pStA1DE containing full transcription unit coding for acetoacetate decarboxylase from *Clostridium acetobutylicum* (*adc*) (library of six promoters and six RBS, Terminator 4)	TetR	This work
pLS57	pStA1EZ containing full transcription unit coding for NADP-dependent isopropanol dehydrogenase from *Clostridium beijerinckii* (*cadh*) (library of six promoters and six RBS, Terminator 1)	TetR	This work

pLS60	pStA212 combinatorial built of the whole isopropanol pathway library – **MP_{Lib}**	KanR	This work
pLS60_1	pStA212 selected combinatorial built of the whole isopropanol pathway – **MP_1**	KanR	This work
pLS60_2	pStA212 selected combinatorial built of the whole isopropanol pathway – **MP_2**	KanR	This work
pLS63	pMAK705 with 500 bp upstream and 500 bp downstream of *ldhA* from *E. coli*	CmR	This work
pLS67	pJET1.2 storage plasmid containing *tbadh* gene from *Thermoanaerobacter brockii* (Uniprot: P14941)	AmpR	This work
pLS69	pUC19 containing *tbadh* gene from *Thermoanaerobacter brockii* – **$TBADH_{WT}$**	AmpR	This work
pLS73	Library of pUC19 containing mutated *tbadh* gene from *Thermoanaerobacter brockii* (saturated mutagenesis to four amino acids position: 198,199,200,218) – **$TBADH_{Lib}$**	AmpR	This work
pLS73_1	pUC19 containing selected variant of *tbadh* gene from *Thermoanaerobacter brockii* (G198H/S199R/R200A/Y218M/G198'A/R200'H) containing a 153 bp insertion between residues 241 and 242 – **$TBADH_{S2}$**	AmpR	This work
pLS73_2	pUC19 containing selected variant of *tbadh* gene *Thermoanaerobacter brockii (*G198S/S199K/R200P/Y218V) – **$TBADH_{S1}$**	AmpR	This work
pLS90	pET28a containing a gene encoding $TBADH_{S2}$ with N-terminal 6xHis-tag	KanR	This work
pLS91	pET28a containing a gene encoding $TBADH_{S1}$ with N-terminal 6xHis-tag	KanR	This work
pLS97	pET28a containing gene encoding $TBADH_{WT}$ with N-terminal 6xHis-tag	KanR	This work
pLS98	pET28a containing gene encoding selected NADH-dependent variant of CBADH from *Clostridium beijerinckii* (pLS10_3) with N-terminal 6xHis-tag	KanR	This work
pLS129	pJET1.2 containing MYSTI_01767 gene from *Myxococcus stipitatus* (Uniprot: L7U9F5)	AmpR	This work
pLS130	pUC19 containing MYSTI_01767 gene from *Myxococcus stipitatus* – **$MsIRED_{WT}$**	AmpR	This work
pLS131	pUC19 containing mutated MYSTI_01767 gene from *Myxococcus stipitatus* (containing the	AmpR	This work

	following point mutations: N32E, R33Y, T34E, K37R, L67I, L71V) – **MsIRED$_{C1}$**		
pLS133	Library of pUC19 containing mutated MYSTI_01767 gene from *Myxococcus stipitatus* (saturated mutagenesis to four amino acids position: 32, 33, 34, 37) – **MsIRED$_{Lib}$**	AmpR	This work
pLS133_1	pUC19 containing selected variant of MYSTI_01767 gene from *Myxococcus stipitatus* (N32E, R33V, T34R, K37R) – **MsIRED$_{S}$**	AmpR	This work
pLS161	pLS130 with C-terminal 6xHis tag	AmpR	This work
pLS162	pLS131 with C-terminal 6xHis tag	AmpR	This work
pLS164	pLS133_1 with C-terminal 6xHis tag	AmpR	This work
pLS167_1	pJet1.2 containing a gene encoding (R)-imine reductase variant with K40A substitution from *Streptomyces* sp. GF3587 (Uniprot ID: M4ZRJ3)	AmpR	This work
pLS167	pUC19 containing gene encoding (R)-imine reductase variant with K40A substitution from *Streptomyces* sp. GF3587 (Uniprot ID: M4ZRJ3) **Sgf3587IRED$_{C2}$**	AmpR	This work
pLS168_1	pJet1.2 containing *nfsB* gene from *Enterobacter cloacae* (Uniprot: Q01234)	AmpR	This work
pLS168	pUC19 containing *nfsB* gene from *Enterobacter cloacae* (Uniprot: Q01234) – **EntNfsB$_{WT}$**	AmpR	This work
pLS169	Library of pUC19 containing mutated *nfsB* gene from *Enterobacter cloacae* (saturated mutagenesis to four amino acids position: 40, 41, 68, 124) – **EntNfsB$_{Lib}$**	AmpR	This work
pLS169_1	pUC19 containing selected variant of *nfsB* gene from *Enterobacter cloacae* (S40A/T41I/Y68Y/F124A) – **EntNfsB$_{S1}$**	AmpR	This work
pLS169_3	pUC19 containing selected variant of *nfsB* gene from *Enterobacter cloacae* (S40S/T41L/Y68L/F124L) – **EntNfsB$_{S2}$**	AmpR	This work
pLS180	pLS168 with C-terminal 6xHis-tag	AmpR	This work
pLS181	pLS169_1 with C-terminal 6xHis-tag	AmpR	This work
pLS182	pLS169_2 with C-terminal 6xHis-tag	AmpR	This work

Table 2.3. Synthetic DNA sequences used in this project. All synthetic genes were codon optimised for *E. coli* by using the IDT optimisation tool. Restriction sites are highlighted in green and start and stop codons in bold.

Synthetic DNA	Sequence (5'→ 3')	RBS and Restriction sites
Klebsiella pneumoniae ***budC*** **Encoding protein: 2,3-butanediol dehydrogenase**	CCGTTCGCATGCCAATCTTAATCAAATCAGACAGAGAGA GTACAAT**ATG**AAAAAAGTCGCACTTGTTACCGGCGCCG GCCAGGGGATTGGTAAAGCTATCGCCCTTCGTCTGGTG AAGGATGGATTTGCCGTGGCCATTGCCGATTATAACGA CACCACCGCCAAAGCGGTCGCCTCCGAAATCAACCAGG CCGGCGGCCGCGCCATGGCGGTGAAAGTGGATGTCTC CGACCGCGATCAGGTGTTTGCCGCCGTCGAACAGGCG CGCAAAACGCTGGGCGGCTTCGACGTCATCGTCAACAA CGCCGGCGTGGCGCCGTCCACGCCGATCGAGTCCATT ACCCCGGAGATTGTCGATAAAGTCTACAACATCAACGTT AAAGGGGTGATCTGGGGCATTCAGGCGGCGGTCGAGG CCTTTAAGAAAGAGGGTCACGGCGGGAAAATCATCAAC GCCTGTTCCCAGGCCGGCCACGTCGGCAACCCGGAGC TGGCGGTATATAGCTCGAGTAAATTCGCCGTACGCGGC TTAACCCAGACCGCCGCTCGCGACCTCGCGCCGCTGG GCATCACAGTCAACGGCTACTGCCCGGGGATTGTCAAA ACGCCAATGTGGGCCGAAATTGACCGCCAGGTGTCCGA AGCCGCCGGTAAACCGCTGGGTTACGGTACCGCCGAG TTCGCCAAACGCATCACCCTCGGCCGCCTGTCCGAGCC GGAAGATGTCGCCGCCTGCGTCTCCTATCTTGCCAGCC CGGATTCTGATTATATGACCGGTCAGTCATTGCTGATCG ACGGCGGGATGGTGTTTAAC**TAA**GGATCCGCTGAA	*SphI* (5'end) *BamHI* (3'end) Synthetic RBS sequence designed with SALIS calculator: CAATCTTAATCAAATCA GACAGAGAGAGTACAAT
Thermus* sp. ATN-1** ***tadh **Encoding protein: NAD-dependent alcohol dehydrogenase**	CCGTTCGCATGCAGGAGGTACGAACAC**ATG**AAGGGGTT CGCAATGCTGTCTATTGGAAAAGTTGGCTGGATTGAAAA GGAGAAGCCAGCGCCAGGGCCTTTCGACGCAATTGTTC GCCCTTTGGCAGTCGCACCTTGCACGTCTGACATCCAC ACCGTTTTCGAAGGAGCCATTGGTGAACGTCATAACAT GATCTTGGGACACGAAGCGGTAGGTGAGGTTGTAGAG GTCGGTTCTGAAGTTAAGGACTTTAAACCTGGAGACCG CGTGGTGGTGCCCGCGATTACGCCTGACTGGCGTACTT CAGAGGTCCAACGTGGATATCACCAACATAGCGGCGGT ATGCTGGCGGGTTGGAAGTTCTCCAATGTGAAGGACGG TGTTTTCGGAGAATTCTTCCATGTTAATGACGCCGACAT GAATTTGGCGCACCTTCCGAAGGAGATTCCGTTAGAAG CCGCGGTAATGATCCCCGACATGATGACCACCGGCTTT CATGGAGCGGAGCTGGCGGACATCGAGTTGGGCGCTA CCGTGGCTGTACTTGGCATCGGTCCTGTCGGTCTGATG GCGGTGGCAGGGGCCAAGTTGCGTGGTGCAGGACGTA TCATTGCTGTTGGTTCTCGTCCAGTCTGTGTCGACGCTG CCAAGTATTACGGAGCAACCGACATCGTGAACTATAAG GATGGGCCAATTGAGTCACAGATTATGAACCTTACAGAA GGGAAGGGAGTTGATGCAGCTATTATTGCAGGCGGGAA TGCGGATATCATGGCGACAGCCGTCAAGATCGTGAAGC CCGGTGGAACTATTGCTAATGTGAATTACTTTGGTGAGG GAGAAGTTTTGCCGGTGCCTCGCCTGGAATGGGGTTGT GGGATGGCCCACAAAACGATCAAGGGAGGTCTGTGTCC AGGGGGACGTCTGCGCATGGAACGCTTGATTGACCTTG TCTTTTACAAACGTGTGGACCCGAGTAAATTGGTCACAC ACGTATTCCGTGGCTTTGATAACATTGAAAAGGCGTTCA TGTTGATGAAGGATAAACCAAAGGACCTTATCAAACCAG TGGTTATTCTGGCT**TAA**GGATCCTTCAGC	*SphI* (5'end) *BamHI* (3'end) Shine-Dalgarno RBS: AGGAGGTACGAACAC
Thermoanaerobacter brockii ***adh*** **Encoding protein: NADP-dependent isopropanol dehydrogenase**	CCGTTCGCATGCAGGAGGTACGAACAC**ATG**AAGGGGTT CGCAATGCTGTCTATTGGAAAAGTTGGCTGGATTGAAAA GGAGAAGCCAGCGCCAGGGCCTTTCGACGCAATTGTTC GCCCTTTGGCAGTCGCACCTTGCACGTCTGACATCCAC ACCGTTTTCGAAGGAGCCATTGGTGAACGTCATAACAT GATCTTGGGACACGAAGCGGTAGGTGAGGTTGTAGAG GTCGGTTCTGAAGTTAAGGACTTTAAACCTGGAGACCG CGTGGTGGTGCCCGCGATTACGCCTGACTGGCGTACTT CAGAGGTCCAACGTGGATATCACCAACATAGCGGCGGT ATGCTGGCGGGTTGGAAGTTCTCCAATGTGAAGGACGG	*SphI* (5'end) *BamHI* (3'end) Shine-Dalgarno RBS: AGGAGGTACGAACAC

	TGTTTTCGGAGAATTCTTCCATGTTAATGACGCCGACAT GAATTTGGCGCACCTTCCGAAGGAGATTCCGTTAGAAG CCGCGGTAATGATCCCCGACATGATGACCACCGGCTTT CATGGAGCGGAGCTGGCGGACATCGAGTTGGGCGCTA CCGTGGCTGTACTTGGCATCGGTCCTGTCGGTCTGATG GCGGTGGCAGGGGCCAAGTTGCGTGGTGCAGGACGTA TCATTGCTGTTGGTTCTCGTCCAGTCTGTGTCGACGCTG CCAAGTATTACGGAGCAACCGACATCGTGAACTATAAG GATGGGCCAATTGAGTCACAGATTATGAACCTTACAGAA GGGAAGGGAGTTGATGCAGCTATTATTGCAGGCGGGAA TGCGGATATCATGGCGACAGCCGTCAAGATCGTGAAGC CCGGTGGAACTATTGCTAATGTGAATTACTTTGGTGAGG GAGAAGTTTTGCCGGTGCCTCGCCTGGAATGGGGTTGT GGGATGGCCCACAAAACGATCAAGGGAGGTCTGTGTCC AGGGGGACGTCTGCGCATGGAACGCTTGATTGACCTTG TCTTTTACAAACGTGTGGACCCGAGTAAATTGGTCACAC ACGTATTCCGTGGCTTTGATAACATTGAAAAGGCGTTCA TGTTGATGAAGGATAAACCAAAGGACCTTATCAAACCAG TGGTTATTCTGGCT**TAA**GGATCCTTCAGC	
Myxococcus stipitatus **MYSTI_01767** **Encoding protein: NADP-dependent (R)-selective imine reductase**	CCGTTCGCATGCAGGAGGTACGAACAC**ATG**AAACCGAC CCTGACCGTTATTGGCGCTGGCCGTATGGGCTCCGCAC TGATTAAAGCATTCCTGCAATCTGGCTACACGACCACG GTGTGGAACCGTACCAAAGCCAAAAGCGAACCGCTGGC AAAACTGGGCGCACATCTGGCTGATACGGTGCGTGACG CCGTTAAACGCAGCGATATTATCGTGGTTAATGTGCTGG ATTATGACACCTCTGATCAGCTGCTGCGCCAAGACGAA GTGACGCGTGAACTGCGCGGCAAACTGCTGGTTCAGCT GACCAGCGGTTCTCCGGCACTGGCTCGTGAACAGGAAA CGTGGGCGCGCCAACATGGCATTGATTATCTGGACGGT GCGATCATGGCCACCCCGGATTTTATTGGCCAGGCAGA ATGCGCTCTGCTGTACAGTGGTTCCGCGGCCCTGTTCG AAAAACACCGTGCTGTCCTGAATGTGCTGGGCGGTGCC ACCAGCCATGTCGGCGAAGATGTTGGTCATGCCTCAGC ACTGGACAGCGCCCTGCTGTTTCAGATGTGGGGCACCC TGTTCGGTACGCTGCAAGCACTGGCTATTTCTCGCGCA GAAGGCATCCCGCTGGAAAAAACCACGGCGTTTATCAA ACTGACCGAACCGGTCACCCAGGGTGCCGTTGCAGAT GTCCTGACCCGTGTTCAGCAAAATCGCCTGACCGCAGA CGCTCAGACGCTGGCAAGTCTGGAAGCTCATAACGTGG CGTTCCAACACCTGCTGGCCCTGTGTGAAGAACGTAAT ATCCATCGCGGTGTTGCGGATGCCATGTACTCCGTTATT CGTGAAGCGGTCAAAGCCGGCCACGGTAAAGATGACTT TGCAATTCTGACCCGCTTCCTGAAA**TAA**GGATCCTTCAG C	*SphI* (5'end) *BamHI* (3'end) Shine-Dalgarno RBS: AGGAGGTACGAACAC
***Streptomyces sp.* GF3587** **AB747176 locus – K40A mutation** **Encoding protein: NADP-dependent (R)-selective imine reductase with K40A mutation**	CATCTGAAGACAAC**ATG**GGGGACAACCGTACGCCAGTA ACCGTCATTGGTCTGGGTTTAATGGGGCAAGCGCTGGC GGCGGCATTTTTAGAAGCCGGACACACTACAACGGTAT GGAACCGCTCGGCGGGAGCTGCTGAGCAGTTGGTTTC GCAGGGTGCAGTACAAGCAGCAACCCCGGCCGATGCA GTGGCCGCGAGCGAATTAGTAGTGGTCTGCCTGTCAAC ATACGATAACATGCACGATGTAATCGGCTCCCTTGGGG AGAGTCTGCGTGGAAAGGTAATCGTCAATCTGACATCT GGCAGTTCCGATCAAGGGCGCGAAACAGCGGCCTGGG CAGAGAAACAGGGAGTGGAATATCTGGACGGGGCAATT ATGATCACGCCACCTGGGATTGGAACTGAAACGGCGGT CTTATTTTACGCAGGGACTCAAAGTGTATTTGAAAAATAT GAGCCAGCATTGAAGTTGCTGGGCGGGGGCACAACTTA CCTGGGTACAGATCATGGGATGCCGGCATTATATGACG TATCACTTTTGGGGTTGATGTGGGGCACCTTAAATAGCT TCTTGCACGGTGTTGCTGTAGTAGAAACGGCCGGGGTA GGGGCCCAACAATTTTTACCGTGGGCACATATGTGGCT GGAAGCGATTAAAATGTTCACAGCTGATTATGCAGCACA GATTGACGCAGGAGACGGTAAGTTCCCCGCCAACGATG CTACATTGGAGACCCATTTAGCGGCCTTAAAGCACCTTG TTCACGAGTCCGAGGCCTTAGGTATCGACGCAGAGCTG CCTAAGTATTCGGAGGCTCTTATGGAACGTGTTATCAGC CAAGGCCATGCTAAAAACTCCTACGCCGCTGTGCTTAA	*BbsI* (5'end) *BbsI* (3'end)

	AGCGTTTCGCAAGCCGAGCGAA**TAA**GGAGTCTTCAGAG A	
Enterobacter cloacae ***EntNfsB*** **Encoding protein: oxygen-insensitive NAD(P)-dependent nitroreductase**	CATCTGAAGACAAC**ATG**GATATCATTTCTGTCGCCCTGA AACGCCACTCTACCAAGGCGTTCGACGCAAGCAAAAAA CTGACCGCGGAAGAAGCGGAAAAAATCAAAACCCTGCT GCAGTACAGCCCGTCCAGCACCAACTCCCAGCCGTGG CACTTCATTGTAGCCAGCACCGAGGAAGGAAAAGCGCG CGTGGCGAAGTCCGCTGCGGGCACCTATGTGTTCAACG AACGCAAAATGCTGGATGCTTCCCACGTGGTGGTGTTC TGCGCGAAAACCGCGATGGATGACGCCTGGCTGGAGC GCGTCGTGGATCAGGAAGAGGCCGATGGCCGTTTCAA CACGCCGGAAGCCAAAGCCGCAAACCATAAGGGCCGC ACCTACTTCGCCGACATGCACCGCGTGGATCTGAAAGA TGACGACCAGTGGATGGCGAAGCAGGTTTACCTGAACG TCGGCAACTTCCTGCTGGGCGTGGGCGCGATGGGTCT GGACGCGGTACCAATTGAAGGTTTCGACGCCGCTATTC TCGACGAAGAGTTTGGCCTGAAAGAGAAAGGCTTCACC AGCCTGGTGGTGGTACCGGTTGGGCACCACAGCGTGG AAGATTTCAACGCCACGCTGCCGAAATCTCGCCTGCCG CTGAGCACGATTGTGACCGAGTGC**TAA**GGAGTCTTCAG AGA	*BbsI* (5'end) *BbsI* (3'end)

2.2. Computational tools

SnapGene (GSL Biotech, USA) was used to design the construction of oligonucleotides and plasmids, and to visualize sequencing results.

GraphPad Prism (GraphPhad Software, USA) was used to analyse and display growth curves and enzymatic activity assays.

R (R Core Team, 2018) and ggplot2 (Wickham, 2016) were used to obtain and plot a distribution of enzymatic activities across different EC classes. R was also used to perform the t-test for isopropanol titres of selected and randomly-picked variants of the library of metabolic pathways.

CSR-SALAD (Cahn *et al.*, 2017) was used to assist the determination of critical residues for cofactor specificity of CBADH, TBADH and MsIRED.

ProtParam (Gasteiger *et al.*, 2005) was used to calculate molar absorption coefficients for each protein.

The xia2 pipeline (Winter, 2009) and autoPROC (Vonrhein *et al.,* 2011) were used to process X-ray diffraction data.

Phaser (McCoy *et al.*, 2007) was used to solve the structures by molecular replacement.

Coot (Emsley *et al.*, 2010) was used to manually rebuild structural models.

Phenix (Adams *et al.*, 2010) was used to refine the structures.

MolProbity (Chen *et al.*, 2010) was used to validate the structures.

SWISS-MODEL (Waterhouse *et al*., 2018) was used to generate a homology model for wild-type MsIRED.

ChimeraX (Goddard *et al*., 2018) was used to analyse and display structural models, and to generate structural models for MsIRED and EntNfsB variants.

The BLAST web interface (Johnson *et al*., 2008) was used to perform sequence alignments and the search of putative acetoin reductases.

2.3. Microbiological materials and methods

2.3.1. Antibiotic selection

Appropriate antibiotics were used for the selection of cells successfully transformed with plasmid DNA, added as required according to the antibiotic resistance cassette present in the corresponding plasmid. Stock solutions of all antibiotics were prepared in autoclaved deionised water, at concentrations of 100 mg/mL for ampicillin, 25 mg/mL for chloramphenicol, 5 mg/mL for tetracycline and 50 mg/ mL for kanamycin. The antibiotics stock solutions were added in a 1:1000 ratio to the cultures, except for tetracycline, which was added in a 1:500 ratio.

2.3.2. Bacterial strains and aerobic culture conditions

Escherichia coli (*E. coli*) strains DH5α and DH10β (**Table 2.1**) were used throughout the project as a host for cloning and storing plasmids. Unless specified otherwise, aerobic growth was performed at 37 ºC on LB broth with shaking at 230 rpm (orbit of 1.9 cm). Media was supplemented with the appropriate antibiotic (100 µg/mL ampicillin, 50 µg/mL kanamycin, 25 µg/mL chloramphenicol, 10 µg/mL tetracycline). *E. coli* AL, ALS, ALP, ALPS strains (**Table 2.1**) were used throughout the project as a host for selection. *E. coli* BL21(DE3) and BW25113 strains (**Table 2.1**) were used for protein expression.

2.3.3. Bacterial strains and anaerobic culture conditions

E. coli selection strains (AL, ALS, ALP and ALPS) were grown at 37 ºC in an anaerobic cabinet in M9 minimal medium in liquid phase or agar solid phase supplemented with glucose or gluconate, the required antibiotic and an external oxidised substrate if needed for anaerobic growth complementation (**Table 2.4**). Both liquid and solid growth media were made anaerobic by overnight incubation in the anaerobic cabinet.

Table 2.4. Summary of conditions for anaerobic growth complementation experiments.

Mutant strain	Transformed plasmid/library	Antibiotic	Added external substrate	Generated reduced product
AL	pUC19	Ampicillin	As required for the complementation experiment	-
AL	pLS1	Ampicillin	-	-
AL	pLS2	Ampicillin	15 mM acetoin	2,3-butanediol
AL	pLS3	Ampicillin	15 mM acetoin	2,3-butanediol
AL	pLS6	Ampicillin	15 mM acetone	Isopropanol
AL	pLS10	Ampicillin	15 mM acetone	Isopropanol
AL	pLS10_3	Ampicillin	15 mM acetone	Isopropanol
AL	pLS12	Ampicillin	10 mM cyclohexanone or 3-methylcyclohexanone	Cyclohexanol or 3-methylcyclohexanol
AL	pLv2	Kanamycin	-	-
AL	pLS60	Kanamycin	-	Isopropanol
ALPS	pLS69	Ampicillin	15 mM acetone	Isopropanol
ALPS	pLS73	Ampicillin	15 mM acetone	Isopropanol
ALPS	pLS73_1	Ampicillin	15 mM acetone	Isopropanol
ALPS	pLS73_2	Ampicillin	15 mM acetone	Isopropanol
AL	pLS130	Ampicillin	15 mM 2-methyl-1-pyrroline	2-methylpyrrolidine
AL	pLS131	Ampicillin	15 mM 2-methyl-1-pyrroline	2-methylpyrrolidine
AL	pLS133	Ampicillin	15 mM 2-methyl-1-pyrroline	2-methylpyrrolidine
AL	pLS168	Ampicillin	2.5 mM 4-nitrobenzoic acid	4-hydroxylaminobenzoic acid or 4-aminobenzoic acid
AL	pLS168	Ampicillin	8 mM 2-nitrobenzoic acid	2-hydroxylaminobenzoic acid or 2-aminobenzoic acid

AL	pLS168	Ampicillin	2.5 mM 4-nitrobenzyl alcohol	4-hydroxylaminobenzyl alcohol or 4-aminobenzyl alcohol
AL	pLS169	Ampicillin	8 mM 2-nitrobenzoic acid	2-hydroxylaminobenzoic acid or 2-aminobenzoic acid
AL	pLS169_1	Ampicillin	8 mM 2-nitrobenzoic acid	2-hydroxylaminobenzoic acid or 2-aminobenzoic acid
AL	pLS169	Ampicillin	2.5 mM 4-nitrobenzyl alcohol	4-hydroxylaminobenzyl alcohol or 4-aminobenzyl alcohol
AL	pLS169_3	Ampicillin	2.5 mM 4-nitrobenzyl alcohol	4-hydroxylaminobenzyl alcohol or 4-aminobenzyl alcohol

2.3.4. Preparation of electro-competent *E. coli* cells

Fresh overnight cultures of *E. coli* were used to inoculate in a 1:100 ratio 100 mL of LB broth in a 500 mL conical flask. Cells were grown at 37 ºC and 230 rpm shaking (orbit of 1.9 cm) until the exponential growth phase was reached, indicated by an OD_{600} of 0.5-0.8. After the cells reached the required OD_{600}, they were aliquoted in 50 mL tubes and incubated in ice for 30 min. Then, the 50 mL tubes were centrifuged at 4000 g for 20 min. Pellets were kept in ice throughout the whole procedure. They were gently resuspended in 50 mL of autoclaved deionised water at 4 ºC and centrifuged at 4000 g for 20 min. This step was repeated twice to completely remove salt contaminants. Pellets were then gently resuspended in 10 mL of sterile 10 % (v/v) glycerol at 4 ºC and centrifuged at 4000 g for 10 min. Finally, pellets were resuspended with 0. 5 mL of 15 % (v/v) glycerol. The resulting *E. coli* electro-competent cells were stored at -80 ºC in 100 µL aliquots.

2.3.5. Transformation of *E. coli* by electroporation

E. coli electro-competent cells were thawed in ice. 50 ng of pure plasmid DNA or 10 µL of ligation reactions or Golden Gate reactions were added to the cells. After mixing gently, the mixture was transferred to a pre-chilled 1 mm electroporation cuvette. A pulse was applied across the cuvette using a MicroPulser electroporator (Bio-Rad) set to programme 'Ec1' (1.8 kV voltage). Immediately after the pulse, 500 µL of LB broth were added to the cuvette. The content of the cuvette was transferred to a 1.5 mL microtube and incubated for 1 h at 37 ºC and 230 rpm shaking (orbit of 1.9 cm), to allow the antibiotic resistance cassette to be expressed. Cells were centrifuged at 4000 g for 6 min, and 400 µL of supernatant were discarded. Cells were resuspended in the remaining 100 µL supernatant and plated onto appropriate selective media. Alternatively, for cells transformed with libraries, 400 µL of M9 medium with the same carbon source as in the corresponding growth complementation experiments were added and cells were resuspended, resulting in a total volume of 500 µL.

For cells transformed with libraries, the efficiency of transformation was measured by plating serial dilutions of transformed cells in aerobic LB plates with the corresponding antibiotic and counting the colonies in plates where the spacing between individual colonies allowed them to be distinguished. The number of cells successfully transformed was calculated with the following formula:

$$N_{transformed} = N_{colonies} \cdot D \cdot \frac{V_{transformation}}{V_{dilutions}}$$

Where $N_{transformed}$ is the number of cells successfully transformed, $N_{colonies}$ is the counted number of colonies, D is the dilution factor by which cells were diluted before being plated, $V_{transformation}$ is the transformation volume and $V_{dilutions}$ is the volume of transformed cells used to prepare the serial dilutions.

2.3.6. Construction of *E. coli* mutant strains

E. coli strain JW1228 from the Keio collection (Baba *et al.*, 2006), characterized for its *ΔadhE* genotype, was used as the starting point for constructing all mutant strains. The protocol described in Hamilton *et al.* (Hamilton *et al.*, 1989) was followed to successively construct the AL (Δ*adhE* Δ*ldhA*), ALP (Δ*adhE* Δ*ldhA* Δ*pntB*), ALS (Δ*adhE* Δ*ldhA* Δ*sthA*) and ALPS (Δ*adhE* Δ*ldhA* Δ*pntB* Δ*sthA*) mutant strains. Selection of cells with the pMAK705-derived plasmid successfully integrated in the chromosome was done in medium supplemented with chloramphenicol at 44 ºC, which inhibits the replication of unintegrated pMAK705-derived plasmids due to their temperature-sensitive origin of replication. pMAK705-derived plasmids were constructed by PCR amplifying 500 bp upstream and 500 bp downstream of the gene to be knocked out from *E. coli* gDNA with *BamHI* and *HindIII* overhangs. The pMAK705 plasmid

was then treated with *BamHI* and *HindIII* and ligated with the PCR fragment with *BamHI* and *HindIII* overhangs corresponding to the target gene.

First, a variant of pMAK705 designed to target the *ldhA* locus was used to transform JW1228 cells, resulting in the AL strain. The AL strain was then subjected to the same protocol with pMAK705 variants targeting the *pntA* and *sthA* loci, yielding strains ALP and ALS respectively. Finally, ALS was treated with the pMAK705 variant targeting *pntA* to obtain the quadruple mutant strain ALPS.

2.4. Molecular biological materials and methods

2.4.1. Oligonucleotide design

Oligonucleotides were designed with SnapGene. All oligonucleotides used in this project were synthesised by Intergrated DNA Technologies (IDT) (**Table 2.5**). The preparation of the oligonucleotides was done following the manufacturer's instructions, resulting in oligonucleotide solutions at a final concentration of 10 µM in all cases.

Table 2.5. Oligonucleotides used in this project.

Plasmid	Gene	Oligo ID	Description	Sequence (5' → 3')
pUC19	-	oligoLS315	Fw - *BbsI*	TCTCTGAAGACCCTAAGGATCCCCGGGTACC
		oligoLS314	Rv - *BbsI*	TCTCTGAAGACTCCATGTGTTCGTACCTCCTGCATG
pLS1	*adhE*	oligoLS19	Fw - *SphI*	CCGTTCGCATGCAGGAGGTACGAACACATGGCTGTTACTAATGT
		oligoLS20	Rv - *BamHI*	GCTGAAGGATCCTTAAGCGGATTTTTTCG
-	*adhE*	oligoLS1	Fw - internal	CCTGTGGTGTTCTGTCTG
		oligoLS2	Rv - internal	TAGATTTCGGAATACCCA
		oligoLS3	Fw - external	GGCGAAAAGCGATGCTG
		oligoLS4	Rv - external	CGGTGGGAAGGTGTTCTGC
-	*ldhA*	oligoLS5	Fw - internal	GCCGCCCGGTGCTGGAAG
		oligoLS6	Rv - internal	GGCGACGGAATACGTCAT
		oligoLS7	Fw - external	GAAGGTTGCGCCTACACT
		oligoLS8	Rv - external	CACCAAAGCTGATTTCTG
pLS2	*bdhA*	oligoLS23	Fw - *SphI*	CCGTTCGCATGCAGGAGGTACGAACACATGAAGGCAGCAAGATG
		oligoLS24	Rv - *BamHI*	GCTGAAGGATCCTTAGTTAGGTCTAACAAGGATTTTGACT
pLS6	*cbadh*	oligoLS87	Fw - *SphI*	GTTCGCATGCATTCGGATCTATACAGATAAGGAGAAAGAGATGAAAGGCTTTGCCATGCTGGG
		oligoLS88	Rv - *BamHI*	CTTCCATGGATCCTCACTATTAGAGGATAACTACGGCC

pLS7	*cbadh*$_{R1}$	oligoLS107	Rv	CTTGGCGGCCTCAACGCAAATAGGCCCACCGC CGACACCAATAATCCGACCTGC
		oligoLS109	Fw	TTCTACGGCGCGACCGACATTCTGAATTTCAAA AATGGCCATATTGTGGAC
pLS8	*cbadh*$_{R2}$	oligoLS108	Rv	CTTGGCGGCCTCAACGCAAATAGGCCCACCGT CGACACCAATAATCCGACCTGC
		oligoLS109	Fw	TTCTACGGCGCGACCGACATTCTGAATTTCAAA AATGGCCATATTGTGGAC
pLS10	*cbadh*$_{Lib}$	oligoLS112	Rv	CTTGGCGGCCTCAACGCAAATAGGNNNNNNNN NGACACCAATAATCCGACCTGC
		oligoLS113	Fw	TTCTACGGCGCGACCGACATTCTGAATNNNAAA AATGGCCATATTGTGGAC
pLS39	500 bp upstream and downstream of *sthA*	oligoLS208	Fw - *BamHI*	TTCAGCGGATCCAATGTATCTGCATGAAGCACA GACCCACCAGTTACTGG
		oligoLS212	Rv	AACAGGTAAGCCCTACCATGTAAAACTTTATCG AAATGGCCATCCATTCTTGCGCGG
		oligoLS213	Fw	GCCATTTCGATAAAGTTTTACATGGTAGGGCTT ACCTGTTCTTATACATAAAAGCAACAGAATGG
		oligoLS209	Rv - *HindIII*	TTCAGCAAGCTTCATTAAACCGCTCTCATCAAC CATGGTCAGACCCAGTTCG
-	*sthA*	oligoLS194	Fw - internal	GATGGAACAAAATTTTCAGCGTGCC
		oligoLS193	Rv - internal	ATAGTAATAGGTTCCGGCCC
		oligoLS195	Fw - external	CAGGCAATGGGTTTCTGTTTTG
		oligoLS196	Rv - external	CGAACTGGGTCTGACCATGGTTGATGAGAGCG GTTTAATG
pLS40	500 bp upstream and downstream of *pntA*	oligoLS216	Fw - *BamHI*	TTCAGCGGATCCGAAACGACCAGAGCCGCCAG GTTCA
		oligoLS217	Rv	CCGATGGAAGGGAATATCATGTAAGGGGTAAC ATATGTCTGGAGGATTAGTTACAGCTGCATACA TTGTTGCCGC
		oligoLS219	Fw	CCAGACATATGTTACCCCTTACATGATATTCCC TTCCATCGGTTTTATTGATG
		oligoLS218	Rv - *HindIII*	TTCAGCAAGCTTCAGGAGGGTGTTCTTAAGCTT CATAAAAATAATCCTTCGCCTTGCGCAAA
-	*pntA*	oligoLS223	Fw - internal	GTGCTCCGACAACAATAATCC
		oligoLS224	Rv - internal	TGATGGTGATTGGTGCGGGTG
		oligoLS216	Fw - external	GAAACGACCAGAGCCGCCAGGTTCA
		oligoLS221	Rv - external	TTTGCGCAAGGCGAAGGATTATTTTTATGAAGC
pLS46	*atoB*	oligoLS230	Fw - *BsaI*	AAGGGGTTGGTCTCATGTGCTCTTCGATGAAAA ATTGTGTCATCGTCAGTGCGGTACG
		oligoLS231	Rv - *BsaI*	AAGGGGTTGGTCTCTGGTCTTACGCTCTTCATT AATTCAACCGTTCAATCACCATCGCAATTCCC

pLS47	*atoD*	oligoLS234	Fw - *BsaI*	AAGGGGTTGGTCTCATGTGGCTCTTCGATGAAA ACAAAATTGATGACATTACAAGACG
		oligoLS243	Rv - *BsaI*	AAGGGGTTGGTCTCTGGTCTTACGCTCTTCATT ATTTGCTCTCCTGTGAAACGATGATGTG
pLS48	*atoA*	oligoLS235	Fw - *BsaI*	AAGGGGTTGGTCTCTGGTCTTACGCTCTTCATT ATAAATCACCCCGTTGCGTATTC
		oligoLS242	Rv - *BsaI*	AAGGGGTTGGTCTCATGTGGCTCTTCGATGGA TGCGAAACAACGTATTGCGC
pLS49	*adc*	oligoLS228	Fw - *BsaI*	AAGGGGTTGGTCTCATGTGGCTCTTCGATGTTA AAGGATGAAGTAATTAAACAAATTAGCACG
		oligoLS229	Rv - *BsaI*	AAGGGGTTGGTCTCTGGTCTTACGCTCTTCATT ACTTAAGATAATCATATATAACTTCAGCTCTAGG C
pLS50	*cbadh*	oligoLS232	Fw - *BsaI*	AAGGGGTTGGTCTCATGTGGCTCTTCGATGAAA GGCTTTGCCATGCTGGGTATTAAC
		oligoLS233	Rv - *BsaI*	AAGGGGTTGGTCTCTGGTCTTACGCTCTTCATT AGAGGATAACTACGGCCTTAATGAGATCTTTAG G
pLS63	500 bp upstream and downstream of *ldhA*	oligoLS244	Fw - *BamHI*	TTCAGCGGATCCTGTCTGTTTTGCGGTCGCCA G
		oligoLS247	Rv	CACTGGAGAAAGTCTTATGTAATCTTGCCGCTC CCCTGCATTCCAG
		oligoLS246	Fw	CAGGGGAGCGGCAAGATTACATAAGACTTTCT CCAGTGATGTTGAATC
		oligoLS245	Rv - *HindIII*	TTCAGCAAGCTTCAAGCAGAATCAAGTTCTACC GTGC
pLS73	$tbadh_{Lib}$	oligoLS258	Rv	AGCGTCGACACAGACTGGNNNNNNNNNAACAG CAATGATACGTCCTGC
		oligoLS259	Fw	GCCAAGTATTACGGAGCAACCGACATCGTGAA CNNNAAGGATGGGCC
pLS90	$tbadh_{S2}$	oligoLS288	Fw - *NdeI*	GCAGCCATATGATGAAGGGGTTCGCAATGCTG TCTATTGG
		oligoLS289	Rv - *BlpI*	TTATTGCTCAGCTTAAGCCAGAATAACCACTGG TTTGATAAGGTCCTTTGG
pLS91	$tbadh_{S1}$	oligoLS288	Fw - *NdeI*	GCAGCCATATGATGAAGGGGTTCGCAATGCTG TCTATTGG
		oligoLS289	Rv - *BlpI*	TTATTGCTCAGCTTAAGCCAGAATAACCACTGG TTTGATAAGGTCCTTTGG
pLS97	$tbadh_{WT}$	oligoLS288	Fw - *NdeI*	GCAGCCATATGATGAAGGGGTTCGCAATGCTG TCTATTGG
		oligoLS289	Rv - *BlpI*	TTATTGCTCAGCTTAAGCCAGAATAACCACTGG TTTGATAAGGTCCTTTGG
pLS98	$cbadh_{S}$	oligoLS294	Fw - *NdeI*	GCAGCCATATGATGAAAGGCTTTGCCATGCTG GGTATTAACAAATTAGG
		oligoLS295	Rv - *BlpI*	TTATTGCTCAGCTTAGAGGATAACTACGGCCTT AATGAGATCTTTAGGTTTATCTTTCATGAG
pLS131	$MsIRED_{C1}$	oligoLS344	Rv	ACGATAATATCGCTGCGTTTAAC
		oligoLS345	Fw	CTGGCAAAACTGGGCGCACATC

		oligoLS342	Rv	/5'Phos/CGGTTCGCTACGGGCTTTTTCATATTCCCACACCGTGGTCG
		oligoLS343	Fw	/5'Phos/GGTTAATGTGATTGATTATGACGTTTCTGATCAGCTGCTG
pLS133	*MsIRED*$_{Lib}$	oligoLS337	Fw	GCTGAGAAGACCGACCACGGTGTGGNNNNNNNNNAAAGCCNNNAGCGAACCGCTGGCAAAACTG
		oligoLS338	Rv	GCTGAGAAGACCGTGGTCGTGTAGCCAGATTGCAGGAATGCTTTAATCAGTGCGGAGCCCATACGGCC
pLS161	*MsIRED*$_{WT}$, C-terminal 6xHis-tag	oligoLS358	Rv - *BbsI* 6xHis-tag	TCTCTGAAGACTCCTTAGTGGTGGTGGTGGTGGTGTTTCAGGAAGCGGGTCAGAATTGCAAAG
		oligoLS359	Fw - *BbsI*	TCTCTGAAGACAACATGAAACCGACCCTGACCGTTATTGGC
pLS162	*MsIRED*$_{C1}$, C-terminal 6xHis-tag	oligoLS358	Rv - *BbsI* 6xHis-tag	TCTCTGAAGACTCCTTAGTGGTGGTGGTGGTGGTGTTTCAGGAAGCGGGTCAGAATTGCAAAG
		oligoLS359	Fw- *BbsI*	TCTCTGAAGACAACATGAAACCGACCCTGACCGTTATTGGC
pLS164	*MsIRED*$_{S}$, C-terminal 6xHis-tag	oligoLS358	Rv - *BbsI* 6xHis-tag	TCTCTGAAGACTCCTTAGTGGTGGTGGTGGTGGTGTTTCAGGAAGCGGGTCAGAATTGCAAAG
		oligoLS359	Fw - *BbsI*	TCTCTGAAGACAACATGAAACCGACCCTGACCGTTATTGGC
pLS169	*EntNfsB*$_{Lib}$	oligoLS363	Rv - *BbsI*	TCTCTGAAGACTCGGTGCTGGCTACAATGAAGTGCCACGGCTGGGAGTTNNNNNNGGACGGGCTGTACTGC
		oligoLS366	Fw - *BbsI*	CTCTGAAGACCAGTGGATGGCGAAGCAGGTTTACCTGAACGTCGG
		oligoLS364	Fw - *BbsI*	CTCTGAAGACAGCACCGAGGAAGGAAAAGCGCGCGTGGCGAAGTCCGCTGCGGGCACCNNNGTGTTCAACGAACG
		oligoLS365	Rv - *BbsI*	TCTCTGAAGACATCCACTGGTCGTCATCTTTCAGATCCACGCGGTGCATGTCGGCNNNGTAGGTGCGGCC
pLS180	*EntNfsB*$_{WT}$, C-terminal 6xHis-tag	oligoLS385	Rv - *BbsI* 6xHis-tag	TCTCTGAAGACTCCTTAGTGGTGGTGGTGGTGGTGGCACTCGGTCACAATCGTGCTCAGC
		oligoLS386	Fw - *BbsI*	TCTCTGAAGACAACATGGATATCATTTCTGTCGCCCTG
pLS181	*EntNfsB*$_{S1}$, C-terminal 6xHis-tag	oligoLS385	Rv - *BbsI* 6xHis-tag	TCTCTGAAGACTCCTTAGTGGTGGTGGTGGTGGTGGCACTCGGTCACAATCGTGCTCAGC
		oligoLS386	Fw - *BbsI*	TCTCTGAAGACAACATGGATATCATTTCTGTCGCCCTG
pLS182	*EntNfsB*$_{S2}$, C-terminal 6xHis-tag	oligoLS385	Rv - *BbsI* 6xHis-tag	TCTCTGAAGACTCCTTAGTGGTGGTGGTGGTGGTGGCACTCGGTCACAATCGTGCTCAGC
		oligoLS386	Fw - *BbsI*	TCTCTGAAGACAACATGGATATCATTTCTGTCGCCCTG
-	-	M13	Rv	CAGGAAACAGCTATGACC
-	-	T7	Fw	TAATACGACTCACTATAGGG

-	-	oligoLS275	Fw	CATCCTATGGAACTGCCTCG
-	-	oligoLS276	Fw	GAAAGTGAATGTCAACGGCG
-	-	oligoLS277	Fw	CATCGTTGCGACACACTTGGC
-	-	oligoLS278	Fw	CTGCACCATGCCACTCACTG
-	-	oligoLS279	Fw	CCGTACATGAAGCTTGGACAGG

2.4.2. Storage of synthetic DNA sequences

Genes obtained as gBlocks from IDT were stored into pJET1.2 by mixing the synthetic DNA fragments and the commercial linearized vector in the appropriate ratio and ligating both components with the supplied ligase and buffers according to the manufacturer's protocol.

2.4.3. Genomic DNA extraction

In order to extract genomic DNA from *E. coli* and *Bacillus subtilis*, 5 mL cultures were grown overnight in aerobic conditions. Cells were harvested by centrifuging 500 µL of the overnight cultures at 4000 g for 10 min at 4 ºC. The supernatant was discarded, and the pellet was resuspended with 100 µL of sterile deionised water and boiled at 100 ºC for 10 min. The suspension was centrifuged at 15000 g for 5 min. The resulting supernatant was transferred to a fresh tube for immediate usage or frozen at -20 ºC for long-term storage.

2.4.4. Amplification of DNA using polymerase chain reaction (PCR)

DNA fragments of interest for cloning and plasmid construction were amplified by PCR using the high-fidelity Phusion DNA polymerase (New England Biolabs) with the corresponding buffer supplied by the manufacturer. The conditions of the PCR were set according to the manufacturer's instructions (**Table 2.6**).

For diagnostic or colony PCR (used to confirm the results of plasmid construction), GoTaq DNA polymerase (Invitrogen) was used following the manufacturer's instructions.

Oligonucleotides used for PCR are summarized in **Table 2.5**.

PCR products were purified with Qiagen QIAprep Spin Miniprep kit, following the manufacturer's instructions.

Table 2.6. PCR conditions for reactions with Phusion and GoTaq polymerases.

Component	Phusion Reaction mixture	GoTaq reaction mixture
10x Buffer (HF) or 5x Gotaq Buffer	10 µL	10 µL
5M Betaine	5 µL	-
Mg Cl_2	-	6 µL
Forward primer (10µM)	2.5 µL	1.5 µL
Reverse primer (10µM)	2.5 µL	1.5 µL
dNTPs (10µM)	1 µL	2 µL
DNA polymerase	0.5 µL	0.5 µL
ddH_2O	27.5 µL	27.5 µL
Total	50 µL	50 µL

Step	Temperature		Time		Number of cycles
	Phusion	GoTaq	Phusion	Gotaq	
Initial denaturation	98 °C	95 °C	30 s	2 min	1
Denaturation	98 °C	95 °C	10 - 30 s	0.5 - 1 min	30
Annealing	50 - 72 °C	42 - 65 °C	30 s	0.5 - 1 min	-
Extension	72 °C		30 s/kb	1 min/kb	1
Final extension	72 °C		5 min		1

2.4.5. Construction of plasmids for enzyme evolution and characterization

Plasmids for anaerobic growth complementation experiments with wild-type enzymes and the variant of Sgf3587IRED with a K40A mutation were constructed by cloning the gene of interest into the pUC19 vector backbone. The gene of interest was obtained from synthetic DNA gBlocks stored in pJET1.2 for the genes listed in **Table 2.3**, from PCR amplification from the pJH28 plasmid (previously constructed in the Heap Laboratory) for CBADH or from PCR amplification from genomic DNA in all other cases.

For all genes of interest except those encoding Sgf3587IRED with the K40A mutation and EntNfsB, the pUC19 vector and the corresponding pJET1.2 or purified PCR fragment were treated with *BamHI* and *SphI*. The appropriate fragments were purified, mixed and ligated with T4 DNA ligase (**Figure 2.1**).

For the construction of plasmids with the genes encoding Sgf3587IRED with the K40A mutation and EntNfsB, the pUC19 vector backbone was amplified with primers introducing overhangs with *BbsI* restriction sites and start and stop codons in the fusion sites created upon treatment with *BbsI*. The amplified PCR product was treated with *DpnI* and purified. The resulting linearized pUC19 vector with *BbsI* restriction sites was then mixed with the

corresponding pJET1.2 storage plasmid (carrying either Sgf3587IRED with the K40A mutation or EntNfsB) to perform a Golden Gate assembly reaction (**Figure 2.1**).

The construction of plasmids carrying rationally predicted variants of CBADH and previously described variant of MsIRED was performed by inverse PCR following a similar procedure as described for the generation of libraries by inverse PCR **(Figure 2.2)**, but with oligonucleotides introducing the appropriate mutations instead of degenerate primers.

In order to construct plasmids carrying specific mutated variants with a 6xHis-Tag, the mutated gene was first amplified by PCR. For constructing plasmids based on the pET28 vector, primers were designed to introduce overhangs with *NdeI* and *BlpI* restriction sites. The pET28 vector was treated with *NdeI* and *BlpI* and ligated with the PCR fragment of the corresponding mutated gene by treatment with T4 DNA ligase, resulting in a plasmid containing the gene with an N-terminal 6xHis-tag. For plasmids based on the pUC19 vector, primers were designed to introduce a C-terminal 6xHis-tag and overhangs with *BbsI* restriction sites. The resulting PCR fragment was purified and used to perform a Golden Gate assembly reaction with linearized pUC19 vector with *BbsI* restriction sites obtained as previously described.

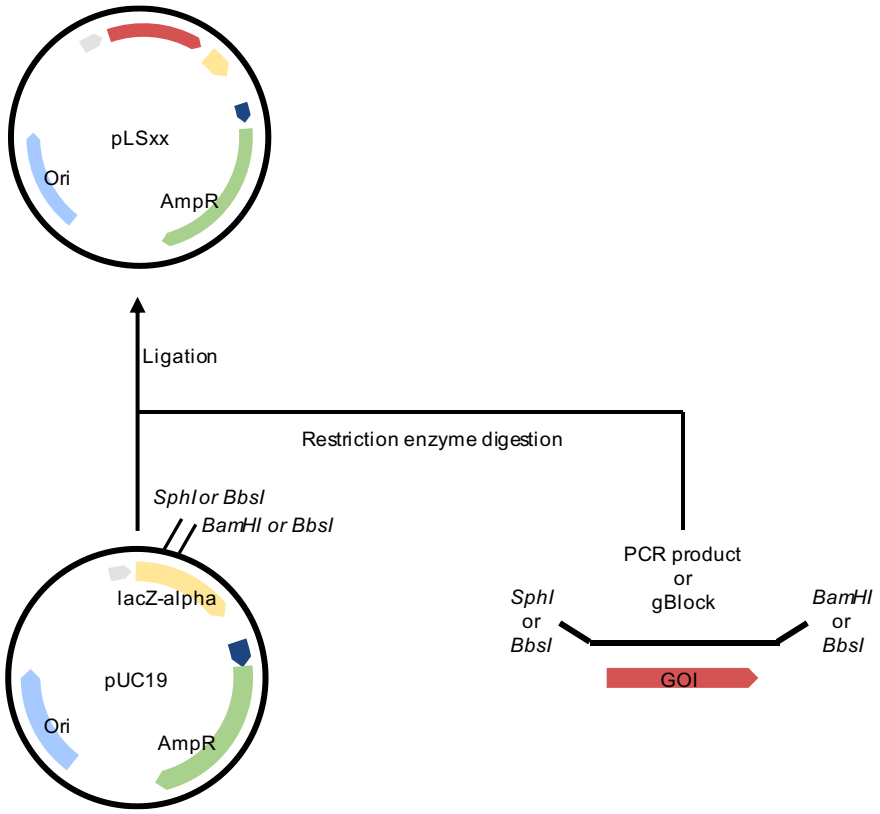


Figure 2.1. Schematics of the workflow for the construction of plasmids for anaerobic growth complementation experiments. The pUC19 vector and the gene of interest (obtained either by PCR or as a gBlock) were treated with the appropriate restriction enzymes (*SphI* and *BamHI*, or just *BbsI*). The resulting digestion products were ligated with T4 DNA ligase.

2.4.6. Construction of saturation libraries

Libraries of variants were constructed by saturation mutagenesis of the desired positions with either inverse PCR or Golden Gate assembly (**Table 2.7**).

Table 2.7. Summary of saturated positions and employed techniques for the generation of saturation libraries.

Enzyme	Organism	Saturated positions	Technique	Oligonucleotides
CBADH	*Clostridium beijerinckii*	198/199/200/218	Inverse PCR	LS112/LS113
TBADH	*Thermoanaerobacter brockii*	198/199/200/218	Inverse PCR	LS258/LS259
MsIRED	*Myxococcus stipitatus*	32/33/34/37	Golden Gate	LS337/LS338
EntNfsB	*Enterobacter cloacae*	40/41/68/124	Golden Gate	LS363/LS364/LS365/LS366

For inverse PCR reactions, two oligonucleotides were designed with the degenerated nucleotides for each reaction, and one of the oligonucleotides contained a phosphorylated 5' end to allow the subsequent T4 ligation. (**Figure 2.2**). The reactions were carried out with Phusion High-Fidelity DNA polymerase (New England Biolabs) using the wild type genes integrated in the pUC19 vector as the template. The reactions were set up following the protocol provided by the manufacturer in a total volume of 50 µL.

For Golden Gate assembly reactions (Püllmann *et al*., 2019), first the pUC19 vector backbone was amplified by PCR with oligonucleotides with overhangs matching *BbsI* restriction sites. The wild-type genes were amplified with one or more pairs of oligonucleotides containing the saturated positions and overhangs matching *BbsI* restriction sites. The resulting PCR amplified fragments contained the saturated positions and covered the entire sequence of the gene. The amplified vector backbone and PCR fragments were used to carry out a Golden Gate assembly reaction in the presence of *DpnI*.

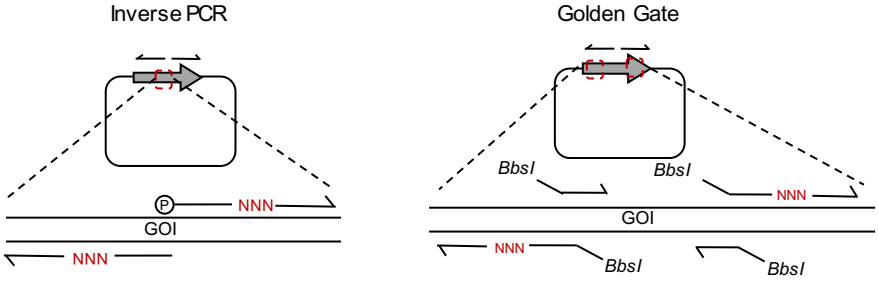


Figure 2.2. Schematics of the two approaches followed for the generation of saturation libraries. In libraries generated by inverse PCR, primers with opposite directionalities targeting the gene of

interest and and carrying degenerated codons in the positions to be saturated were employed to perform a PCR reaction. One of the primers contained a phosphorylated 5' end to allow ligation of the PCR amplification products with T4 DNA ligase. For libraries generated with Golden Gate assembly, one or more pairs of oligonucleotides with overhangs with *BbsI* restriction sites and containing the degenerated codons were used to amplify wild-type genes. The pUC19 vector was also amplified by PCR with oligonucleotides with *BbsI* restriction sites. The resulting PCR amplification products were combined to carry out a Golden Gate assembly reaction, in the presence of *DpnI* to cleave parental DNA.

2.4.7. Golden Gate assembly

Golden Gate reactions were performed with DNA solutions of both the backbone and the fragment to be cloned at an equimolar concentration of 50 fmol/µL. The reaction mixes were prepared as follows: 0.2 µL of backbone (linearized pUC19 with *BbsI* restriction sites obtained as previously described), 1 µL of the corresponding DNA fragment, 2 µL T4 DNA ligase buffer, 1 µL T4 DNA Ligase, 1 µL restriction enzyme *BbsI*-HF, and 14.8 µL H_2O, resulting in a total reaction volume of 20 µL. The reaction mixture was then incubated in a thermocycler with 30 cycles alternating 5-min steps at 37 ˚C with 5-min steps at 16 ºC, before performing a single final deactivation step at 65 ºC for 20 min.

2.4.8. Extraction and purification of plasmid DNA

In order to purify plasmid DNA, *E. coli* DH5α or DH10β cells were transformed with the desired plasmid and plated into agar-LB plates supplemented with the appropriate antibiotic. A single colony was inoculated in 10 mL of LB broth and cultured overnight aerobically. Cells were pelleted by centrifuging at 4000 g for 10 min and plasmid DNA was extracted with the Qiagen QIAprep Spin Miniprep kit (based on the alkaline lysis method) following manufacturer's instructions.

2.4.9. Quantification of DNA

DNA resulting from purification of plasmid DNA, PCR or gel extraction was quantified spectrophotometrically by measuring the absorbance at 260 nm with a Nanodrop 1000 spectrophotometer (Ausubel *et al.*, 2001). The purity of DNA was confirmed by A_{260}/A_{280} and A_{260}/A_{230} absorbance ratios ranging between 1.8 and 1.9.

2.4.10. Restriction endonuclease digestion of DNA

DNA was digested by following the manufacturer's instructions for each specific restriction enzyme and using the supplied optimal buffers. Reactions were allowed to proceed for 2 h at 37 ºC, except for digestions of parental plasmid DNA with *DpnI*, which were allowed to proceed for 4 h at 37 ºC.

2.4.11. Agarose gel electrophoresis and extraction of DNA fragments

Agarose gels were used for diagnosing PCR amplification and digestion with restriction enzymes digestion. DNA fragments were separated by electrophoresis at 100-115 V for 40-60 min in 1% agarose gels. The resulting gels were stained with SYBR safe (Invitrogen) and visualised under blue light. DNA fragments of interest were extracted and purified with the Qiagen QIAquick Gel Extraction kit, following the manufacturer's instructions.

2.4.12. Ligation of DNA

DNA fragments were ligated by using T4 DNA ligase (New England Biolabs) following manufacturer's instructions. The ligation reaction was carried out for 2 h at room temperature, using an insert:vector molar ratio of 3:1. Reactions were done with 50 ng of vector, and the following formula was used in order to calculate the mass of insert DNA required to achieve a 3:1 molar ratio of insert to vector:

$$\mathrm{m_{insert}} = \frac{3 \cdot \mathrm{bp_{insert}} \cdot \mathrm{m_{vector}}}{\mathrm{bp_{vector}}}$$

Where m_{insert} is the mass of insert DNA in ng, bp_{insert} is the length of the insert fragment in base pairs, m_{vector} is the mass of vector DNA in ng (which was always set to 50 ng) and bp_{vector} is the length of the vector in base pairs

2.4.13. Construction of the library of isopropanol production pathways

The library of isopropanol production pathways was based on a previously described pathway (Hanai *et al.*, 2007), which contained *atoB*, *atoD*, *atoA*, *adc* and *cbadh*. The *atoA*, *atoD* and *atoB* genes were obtained by PCR amplification from *E. coli* gDNA. The *adc* gene was obtained by PCR amplification from *Clostridium acetobutylicum* gDNA previously extracted in the Heap Laboratory. The *cbadh* gene was obtained by PCR amplification from the pLS6 plasmid. In all cases, primers designed to introduce *BsaI* overhangs were used.

Individual parts, transcription units, and full metabolic pathways were constructed by following the Start-Stop assembly hierarchical workflow (Taylor *et al.*, 2018). Firstly, functional genetic parts were cloned in the Level 0 storage plasmid (pStA0) by mixing the PCR amplified fragments with *BsaI*-digested vector and ligating with T4 ligase. Level 0 storage plasmids for promoters, RBS and terminators were already available in the Heap Laboratory. Level 0 mixes of six different promoters and six different RBS were prepared (**Table 2.8**). Level 1 plasmids (containing whole expression units including a promoter, a RBS, a functional gene and a transcriptional terminator) were obtained by Start-Stop assembly reactions (**Table 2.2**). For each functional gene, the reaction mixture had a total volume of 20 µL and included 20 fmol

of the destination vector (pStA1AB, pStA1BC, pStA1CD, pStA1DE or pStA1EZ), 40 fmol of the mixture of Level 0 plasmids carrying promoters, 40 fmol of the mixture of Level 0 plasmids carrying RBS, 40 fmol of Level 0 plasmid carrying the corresponding gene, 40 fmol of Level 0 plasmid carrying a terminator out of four different possibilities (**Table 2.9**)**,** T4 DNA ligase buffer for promoters, 400 units of T4 DNA ligase and 10 units of the *BsaI* restriction enzyme. The reaction mixture was incubated with a thermocycler for 30 cycles of 37 ˚C for 5 min and 16 ˚C for 5 min, before a single final step of deactivation at 65 ˚C for 20 min. The resulting products were used to transform *E. coli* DH5α cells, and plasmid DNA was isolated. The Level 2 library (containing variants of the full metabolic pathway differing in the combination of promoters and RBS) was generated with a similar Start-Stop assembly reactions, but the *SapI* restriction enzyme was used instead of *BsaI*.

Table 2.8. List of promoters and RBS. Six promoters previously characterised and available in the Anderson collection (Registry of Standard Biological Parts [http://parts.igem.org]; Galdzicki *et al*., 2011) were chosen, with expression strengths equally spaced over a wide range. Six RBS previously characterised in the Heap Laboratory (Taylor *et al*., 2018) were selected based on the same criteria than for the promoters. Both promoters and RBS are presented in descending order of strength.

Promoters	DNA Sequence (5' →3')
J23100	TTGACGGCTAGCTCAGTCCTAGGTACAGTGCTAGCCCATGAAGAGCGTAAGACCTCTAGGGCGGCG
J23102	TTGACAGCTAGCTCAGTCCTAGGTACTGTGCTAGCCCATGAAGAGCGTAAGACCTCTAGGGCGGCG
J23118	TTGACGGCTAGCTCAGTCCTAGGTATTGTGCTAGCCCATGAAGAGCGTAAGACCTCTAGGGCGGCG
J23107	TTTACGGCTAGCTCAGCCCTAGGTATTATGCTAGCCCATGAAGAGCGTAAGACCTCTAGGGCGGCG
J23116	TTGACAGCTAGCTCAGTCCTAGGGACTATGCTAGCCCATGAAGAGCGTAAGACCTCTAGGGCGGCG
J23113	CTGATGGCTAGCTCAGTCCTAGGGATTATGCTAGCCCATGAAGAGCGTAAGACCTCTAGGGCGGCG

RBS	DNA Sequence (5' →3')
RBSc44	TCGCCAATCGGATTGGATCCAAAGGAGGTTATACCGATGTGAAGAGCGTAAG
RBSc33	AAAACACTAGACTGGAAAGGAGGTAGAGAATATGTGAAGAGCGTAAGACCTCTAGGGCGGCG
RBSc13	CTACGTTTTTTAGAAAAAGGAGGTATGCGAGATGTGAAGAGCGTAAGACCTCTAGGGCGGCG
RBSc58	CTCTTCGCCACATGATCGAATGATTAAAGGAGGTTGGAGGTATGTGAAGAGCGTAAGACCTC
RBSc42	TTCGCCAACAGGATACATCTGTAAAGGAGGTAACGATGATGTGAAGAGC
RBSc36	TCTTCGCCAAGCTCCTTAGCTCCTAAAGGAGGTAGTACATATGTGAAGAG

Table 2.9. List of transcriptional terminators.

Terminators	Genes	DNA Sequence (5' →3')
Terminator 1	*atoB, cbadh*	CTCGGTACCAAAGACGAACAATAAGACGCTGAAAAGCGTCTTTTTTCGT TTTGGTCC
Terminator 2	*atoD*	CTCGGTACCAAATTCCAGAAAAGAGGCCTCCCGAAAGGGGGGCCTTTTT TCGTTTTGGTCC
Terminator 3	*atoA*	GGAAACACAGAAAAAAGCCCGCACCTGACAGTGCGGGCTTTTTTTTTCG ACCAAAGG
Terminator 4	*adc*	TTCAGCCAAAAAACTTAAGACCGCCGGTCTTGTCCACTACCTTGCAGTA ATGCGGTGGACAGGATCGGCGGTTTTCTTTTCTCTTCTCAA

2.4.14. Membrane dialysis of DNA

The products of ligation, Golden Gate or Start-Stop assembly reactions were dialysed prior to transformation of electro-competent cells. The DNA to be dialyzed was applied onto Millipore 0.025 µm nitrocellulose filters, which were then placed over deionised water for 2 h.

2.4.15. Verification of new genetic constructs

All new genetic constructs were verified by running an agarose gel of the products of either colony PCR or restriction enzyme digestion, followed by Sanger sequencing of the full fragment inserted into the vector by Source Bioscience (Cambridge, UK). Successfully generated constructs were retransformed into *E. coli* DH5α cells and stored at -80 °C with 15% (v/v) glycerol.

2.4.16. Library quality control

To verify diversity in the libraries, 5 to 10 colonies were randomly picked and their plasmid DNA was sequenced.

2.5. Biochemical materials and methods

2.5.1. Sodium dodecyl sulfate poly-acrylamide gel electrophoresis (SDS-PAGE)

Cell lysates and purified proteins were analysed by SDS-PAGE. Protein samples were diluted with Laemmli buffer, containing β-mercaptoethanol (to reduce disulphide bonds) and boiled at 100 °C for 10 min to denature proteins. 20 µL of denatured samples were loaded onto poly-acrylamide gels (**Table 2.9**) and separated by electrophoresis at 80 V for 15 min and then at 120 V for 45 min in 1x SDS running buffer (25 mM Tris, 192 mM glycine, 0.1% SDS at pH 8.3) (**Table 2.10**). After electrophoresis, the gel was rinsed three times with 50 mL of deionised water, stained with InstantBlue (Expedeon) in a shaken tray overnight and de-stained with deionised water in a shaken tray for 24 h. Gels were visualized with a GelDoc XR+ instrument.

Table 2.10. Recipe for the preparation of poly-acrylamide gels.

Component	Resolving gel (10%)	Stacking gel (6%)
1M Tris HCl	2.5 mL (pH 8.8)	0.65 mL (pH 6.8)
Water	4 mL	2.5 mL
30% Acrylamide/Bis acrylamide solution	3.3 mL	0.8 mL
20% SDS	50 µL	20 µL
10% APS	100 µL	40 µL
TEMED	10 µL	4 µL

2.5.2. Nickel-affinity chromatography purification of histidine-tagged proteins

BL21(DE3) *E. coli* cells were transformed with either pET28 or pUC19 expression plasmids encoding histidine-tagged proteins and plated in agar-LB plates supplemented with antibiotics as required. Individual colonies were inoculated in 10 mL of LB with the required antibiotic and cultured overnight at 37 ºC and 230 rpm (orbit of 1.9 cm). Cells were then inoculated at a 1:1000 ratio in LB with the required antibiotic and incubated at 37 ºC and 230 rpm (orbit of 1.9 cm). When an OD_{600} of 0.6 was reached, cells were induced with 1 mM IPTG and incubated overnight at 18 ºC and 230 rpm (orbit of 1.9 cm). Cells were harvested by centrifugation at 4000 g for 30 min at 4 ºC and lysed by sonication in purification buffer (50 mM Tris-HCl at pH 7.4, 500 mM NaCl, 10 % (w/v) glycerol and 1 mM TCEP) supplemented with Base Muncher DNAse, lysozyme and one cOmplete Protease Inhibitor Tablet per liter of culture. The lysate was centrifuged at 21000 g for 1 h at 4 ºC to pellet cell debris. The supernatant was filtered through a 0.21 µm filter and loaded onto a 5 mL HisTrap column (GE HealthCare) with an ÄKTA pure chromatography system. The column was extensively washed with purification buffer and eluted with a gradient of 25–500 mM imidazole. Peak fractions were tested for enzymatic activity, and purity was assessed by means of SDS-page. Pure proteins were concentrated to 40-80 µM, flash frozen with liquid nitrogen and stored at -80 ºC.

2.5.3. Quantification of protein

The concentration of purified protein was determined by measuring the absorbance at 280 nm with a Nanodrop 1000 spectrophotometer. Concentrations were determined by applying the Beer-Lambert law, with molar absorption coefficients calculated for each protein from their sequence with the ProtParam tool (Gasteiger *et al.,* 2005). Protein purity was assessed based on the A_{280}/A_{260} ratio, and additionally the A_{280}/A_{454} ratio (lower than 5) for EntNfsB and its variants.

2.6. Experimental assays

2.6.1. Anaerobic growth complementation experiments

AL, ALS, ALP or ALPS cells were transformed with pure plasmid containing wild-type sequences or variants or libraries of enzymes or pathways to be tested for metabolic complementation under anaerobic fermentation conditions in a Whitley A35 anaerobic workstation with anaerobic growth gas mix (10% CO_2, 10% H_2 and 80% N_2). Transformed cells were transferred into 15 mL of M9 medium and left aerobically for 6 h to recover. Cells were then inoculated at a 1:100 dilution ratio into anaerobic M9 media with 18 mM glucose or gluconate, 100 µg/mL ampicillin or 50 µg/mL kanamycin antibiotics, 1 mM IPTG and an external oxidised substrate if required. Culture media was made anaerobic by placing it in the anaerobic workstation for 24 h before inoculation. Metabolic complementation was monitored by measuring the OD_{600}. In all growth complementation experiments, AL cells transformed with *adhE* were used as the positive control and AL cells transformed with empty pUC19 vector were used as the negative control.

For metabolic complementation experiments performed with libraries, after plateau phase of growth was reached, the selected variants were isolated by culturing the grown cells in LB-agar plates with the corresponding antibiotics and sequencing plasmid DNA from single colonies.

Growth curves were generated by fitting a Gompertz growth equation to the OD_{600} values measured at different time points.

2.6.2. Toxicity assays

The toxicity of cyclic ketones (cyclohexanone and 3-methylcyclohexanone), cyclic alcohols (cyclohexanol and 3-methylcyclohexanol) and nitroaromatic compounds (4-NBA, 2-NBA and 4-nitrobenzyl alcohol) was analysed by growing AL cells aerobically in minimal medium supplemented with 10 mM of each cyclic ketone or cyclic alcohol, or 2.5 mM of each nitroaromatic compound. Growth was monitored by measuring the OD_{600}, and growth curves were generated by fitting a Gompertz growth equation to the OD_{600} values measured at different time points.

2.6.3. Enzymatic assays

Enzymatic assays at different substrate concentrations were performed at 37 ºC with purified protein by monitoring NAD(P)H absorbance at 340 nm or 370 nm with an Eppendorf BioSpectrometer kinetic instrument. Assays were performed anaerobically for EntNfsB and its variants and aerobically for all other enzymes. All reactions were performed in a total volume

of 200 µL for aerobic reactions and 400 µL for anaerobic reactions, with protein concentration ranging between 26 nM and 1.3 µM and saturating concentrations of cofactor (**Table 2.11**). Oxidation reactions were performed in assay buffer A (50 mM Tris-HCl at pH 8.8 and 50 mM NaCl) and reduction reactions were carried out in assay buffer B (50m M Tris-HCl at pH 7.4 and 50 mM NaCl). Reaction rates were determined by using the initial linear rates. These were calculated from the change in absorbance by using the known molar absorption coefficients of NADH and NADPH at 340 nm of $\varepsilon_{340} = 6.22 \times 10^3\ M^{-1}\ cm^{-1}$ and of NADH at 370 nm of $\varepsilon_{370} = 2.66 \times 10^3\ M^{-1}\ cm^{-1}$ (Pitsawong *et al.*, 2014). Triplicate measurements were acquired for all initial reaction rates. For all reactions except those catalysed by EntNfsB and its variants, equimolar amounts of cofactor and substrate are consumed as the reaction progresses, and thus the consumption of cofactor was directly converted to consumption of substrate. For reactions catalysed by EntNfsB and its variants, the substrates were assumed to be reduced only to the corresponding hydroxylamines, due to previous reports indicating that the wild-type enzyme does not catalyse the reduction to amines *in vitro*. Additionally, the reduction from the nitroso intermediate to the corresponding hydroxylamine was assumed to happen immediately. As a consequence, and considering that both the reduction to a nitroso compound and to a hydroxylamine require 2 NADH, calculations were performed on the basis of 4 mol of cofactor being consumed for each consumed mol of substrate. Kinetic parameters were obtained by fitting a Michaelis-Menten model with GraphPad Prism, including substrate inhibition in the model when a reduction of initial rates was observed at higher substrate concentrations. k_{cat} values were derived from the fitted V_{max} values (in µM product/s) by dividing them by the concentration of enzyme used for the corresponding assay in µM units.

Table 2.11. Conditions for enzymatic activity assays.

Enzyme	Substrate	Cofactor	Enzyme concentration	Absorbance (nm)
$CBADH_{WT}$	Isopropanol	1 mM $NADP^+$	110 nM	340
$CBADH_{WT}$	Isopropanol	10 mM NAD^+	110 nM	340
$CBADH_S$	Isopropanol	1 mM $NADP^+$	110 nM	340
$CBADH_S$	Isopropanol	10 mM NAD^+	110 nM	340
$TBADH_{WT}$	Isopropanol	1.2 mM $NADP^+$	110.8 nM	340
$TBADH_{WT}$	Isopropanol	10 mM NAD^+	110.8 nM	340
$TBADH_{S1}$	Isopropanol	1.2 mM $NADP^+$	110 nM	340
$TBADH_{S1}$	Isopropanol	10 mM NAD^+	110 nM	340
$TBADH_{S2}$	Isopropanol	2.35 mM $NADP^+$	51.55 nM	340
$TBADH_{S2}$	Isopropanol	2.35 mM NAD^+	51.55 nM	340
$MsIRED_{WT}$	2-methyl-1-pyrroline	0.25 mM NADPH	1.2 µM	340
$MsIRED_{WT}$	2-methyl-1-pyrroline	0.25 mM NADH	1.2 µM	340

$MsIRED_{C1}$	2-methyl-1-pyrroline	0.25 mM NADH	1.26 µM	340
$MsIRED_{S}$	2-methyl-1-pyrroline	0.25 mM NADH	1.25 µM	340
$EntNfsB_{WT}$	4-nitrobenzoic acid	0.3 mM NADH	60.8 nM	340
$EntNfsB_{WT}$	2-nitrobenzoic acid	0.3 mM NADH	60.8 nM	370
$EntNfsB_{WT}$	4-nitrobenzyl alcohol	0.3 mM NADH	26.3 nM	370
$EntNfsB_{S1}$	4-nitrobenzoic acid	0.3 mM NADH	45 nM	370
$EntNfsB_{S1}$	2-nitrobenzoic acid	0.3 mM NADH	45 nM	370
$EntNfsB_{S2}$	4-nitrobenzyl alcohol	0.3 mM NADH	26.3 nM	370

2.6.4. Crystallization

Purified $CBADH_S$, $TBADH_{S1}$ and $TBADH_{S2}$ were concentrated to ~10 mg/mL for crystallization with sitting drop vapor diffusion at 17 °C using commercial sparse matrix screens Wizard 1, 2, 3 and 4 (Rigaku Reagents) with a Mosquito robot (TTPLabTech). Crystallization hits were detected with a light-UV microscope and further optimized in manually set up hanging drop vapor diffusion experiments. For $CBADH_S$, several concentrations of tribasic sodium citrate were tested (1000, 900, 800, 700, 600 and 500 mM) in the presence of 100 mM imidazole. For $TBADH_{S1}$, several concentrations of PEG 3K (24, 22, 20, 18 and 16% w/v) were tested in 100 mM sodium citrate at pH 5.5. Final $CBADH_S$ crystals came from 900 mM sodium citrate, 100 mM imidazole at pH 8 mixed with an equal volume of protein. $CBADH_S$ crystals were soaked in the mother liquor with 1 mM NAD^+, then cryoprotected for several minutes in mother liquor with 30% volume PEG 400 added and flash-cooled in liquid nitrogen. Final $TBADH_{S1}$ crystals came from 20% (w/v) PEG 3K and 100 mM sodium citrate at pH 5.5 mixed with an equal volume of protein. $TBADH_{S1}$ crystals were soaked in the mother liquor with 1 mM NAD^+, then cryoprotected for several minutes in mother liquor with 30% volume PEG 400 and flash-cooled in liquid nitrogen.

2.6.5. Structure determination

X-ray diffraction data were collected at Diamond Light Source synchrotron (beamlines I03 and I04), and processed with the xia2 (Winter, 2009) pipeline using AutoPROC (Vonrhein *et al.,* 2011). Structures were solved by molecular replacement in Phaser (McCoy *et al.*, 2007) using PDB 1KEV for CBADH and PDB 1YKF for TBADH. The models were rebuilt in Coot (Emsley *et al.,* 2010) with cycles of refinement in phenix.refine (Adams *et al.*, 2010), and validated with MolProbity (Chen *et al.*, 2010). Data collection and refinement information are presented in Appendix Table A2.

2.6.6. Structural modelling

A homology model of wild-type MsIRED was generated with SWISS-MODEL using the structure of *Streptomyces kanamyceticus* imine reductase (PDB code 3ZHB).

Structural models of MsIRED and EntNfsB variants were generated by mutating the corresponding residues in ChimeraX. Models of $EntNfsB_{s1}$ bound to 2-NBA was generated by docking in 2-NBA aligning its nitro group and aromatic ring to those of 4-NBA in the wild-type.

2.6.7. Acquisition and processing of ^{1}H-NMR spectra of fermentation broths

Grown anaerobic cultures were centrifuged for 15 min at 8000 g to pellet the cells. 0.5 mL of the supernatant were extracted and mixed with 0.5 mL of NMR buffer (75 mM Na_2HPO_4 buffer at pH 7.4, 4.6 mM 3-(trimethylsilyl)-[2,2,3,3-^{2}H4]-propionate (TSP) and 20% (v/v) D_2O). The resulting solution was transferred into 3 mm NMR tubes. 1D ^{1}H-NMR spectra were acquired on a Bruker Avance III 400 MHz spectrometer operating at 293 K using the zg30 pulse sequence. A total of 16 scans for each sample were collected into 32,768 data points with a spectral window of 20 ppm. Additionally, for each metabolite of interest, spectra of M9 medium supplemented with 10 mM of the metabolite were acquired in order to determine characteristic peaks of each metabolite (**Table 2.12**). The data were automatically processed with the MestReNova suite. After performing phase and baseline correction, spectra were calibrated to the TSP signal at δ 0 ppm. Metabolite assignment was confirmed with spiking experiments. The concentration of each metabolite of interest was calculated by integrating the characteristic signal of each compound (**Table 2.12**) and comparing it to the area of the signal corresponding to TSP, taking into account the different number of protons corresponding to each signal with the following formula:

$$c_{met} = 2 \cdot c_{ref} \cdot \frac{A_{met}}{A_{ref}} \cdot \frac{NH_{ref}}{NH_{met}}$$

Where c_{met} is the concentration of the metabolite of interest, c_{ref} is the concentration of the reference compound (TSP) present at a known concentration in the sample being measured (2.3 mM in all cases), A_{met} is the area of the characteristic signal of the metabolite of interest, A_{ref} is the area of the signal of the reference compound (set to be at 0 ppm), NH_{ref} is the number of protons corresponding to the signal of the reference compound (in the case of TSP, 9) and NH_{met} is the number of protons corresponding to the characteristic signal of the metabolite of interest. NH_{met} for each signal was determined by comparing experimental spectra of each metabolite to assigned spectra available in the HMDB database (Wishart *et al.,* 2018) or previous studies (Samsonowicz *et al*., 2007, 2005; Shi *et al*., 2012). The result was multiplied by a factor of 2 in order to calculate the concentration of each metabolite in the fermentation broth to account for the dilution introduced during the preparation of samples for NMR.

Table 2.12. Characteristic ^{1}H-NMR signals used to quantify each compound of interest. The multiplicity of each signal (s – singlet, d – doublet, t – triplet, q – quartet, m – multiplet) and the number of contributing protons is showed between brackets.

Metabolite	δ for characteristic signal of the metabolite (ppm)
Ethanol	1.19 (t,3)
Lactate	1.32 (d,3)
Succinate	2.41 (s,4)
Acetate	1.92 (s,3)
Formate	8.46 (s,1)
Acetoin	1.38 (d, 3)
2,3-butanediol	1.15 (d,6)
Cyclohexanone	2.40 (t,4)
Cyclohexanol	1.24 (m, 5)
3-methylcyclohexanone	1.01 (d,3)
3-methylcyclohexanol	0.88 (d,3)
Acetone	2.24 (s,6)
Isopropanol	1.18 (d,6)
Propionate	2.19 (q, 2)
2-methyl-1-pyrroline	2.42 (s,3)
2-methylpyrrolidine	1.38 (d,3)
4-nitrobenzoic acid	8.01 (d,2)
4-hydroxylaminobenzoic acid or 4-aminobenzoic acid	7.73 (d,2)
2-nitrobenzoic acid	8.11 (d,1)
2-hydroxylaminobenzoic acid or 2-aminobenzoic acid	7.30 (t,1)
4-nitrobenzyl alcohol	8.27 (d,1)
4-hydroxylaminobenzyl alcohol or 4-aminobenzyl alcohol	7.23 (d,2)

3. Redox rescue strain: construction and validation

3.1. Introduction

E. coli is a facultative anaerobic microorganism, able to use a wide range of organic molecules to sustain heterotrophic growth. Sugars, and particularly glucose, are often used as both the carbon and energy source when culturing *E. coli*. Glucose is oxidised to pyruvate with NAD^+ as the electron acceptor, mainly through the Embden-Meyerhof-Parnas (EMP) glycolytic pathway (Romano and Conway, 1996), which results in the net production of 2 pyruvate, 2 ATP and 2 NADH molecules per glucose (**Figure 3.1**).

Aerobic respiration, where oxygen is used as the terminal electron acceptor, is the preferred metabolic pathway since it maximizes energy yield. This is evidenced by the fact that oxygen represses the terminal reductases required for anaerobic respiration with alternative electron acceptors and key components of fermentation pathways (Unden and Trageser, 1991; Shalel-Levanon, San, and Bennett, 2005). In the presence of oxygen, the pyruvate dehydrogenase complex converts pyruvate to acetyl-CoA and CO_2. Acetyl-CoA is then oxidised to generate three NADH, one $FADH_2$ and one GTP via the tricarboxylic acid (TCA) cycle (**Figure 3.1**). A series of dehydrogenases and terminal oxidases, comprising the respiratory chain, employ reduced NADH and $FADH_2$ to generate a proton gradient across the periplasmic membrane, which serves as the driving force for ATP synthesis by oxidative phosphorylation (Green and Guest, 1998).

If oxygen is not available, other molecules can be used as alternative terminal acceptors of the respiratory chain to perform anaerobic respiration. Terminal reductases, instead of oxidases, are then expressed to transfer electrons to the final acceptors (Green and Guest, 1998). Additionally, pyruvate is metabolized by pyruvate formate-lyase instead of the pyruvate dehydrogenase complex, yielding formate and acetyl-CoA. Formate can be used as a substrate for formate dehydrogenase, directly contributing to the generation of the proton gradient (Sawers, 1994). The TCA cycle is also altered by inhibiting the conversion of 2-oxoglutarate into succinate. The cycle is then split into an oxidative and a reductive branch, which yield respectively 2-oxoglutarate (for biosynthetic purposes) and succinate (**Figure 3.1**) (Green and Guest, 1998). *E. coli* is able to use both inorganic and organic electron acceptors for anaerobic respiration, including nitrate, nitrite, fumarate, DMSO and TMAO (McCrindle *et al.,* 2005). There is a hierarchy of preference for the different terminal acceptors for anaerobic respiration, based on the attainable energetic yield, which is directly related to the redox potential of the electron acceptor. In accordance with this criterion, nitrate is the preferred terminal acceptor in the absence of oxygen, and represses the terminal reductases of

fumarate and DMSO, which have a lower redox potential than nitrate (Unden and Bongaerts, 1997; Green and Guest, 1998).

When neither oxygen nor alternative terminal electron acceptors are available, respiratory chains and oxidative phosphorylation cannot be used anymore as a source of ATP. Under such circumstances, metabolism swaps to an anaerobic fermentation mode, where internal organic metabolites take the role of electron acceptors, and undergo reduction reactions where the redox equivalents generated from the partial oxidation of glucose are used (Clark, 1989). Several types of fermentations have been found across different organisms, but in all of them the net oxidation state of fermentation products is equal to the net oxidation state of the substrates used as energy sources. *E. coli* performs mixed-acid fermentation, where a variety of products, including several organic acids, are generated (Clark, 1989). During anaerobic fermentation, the single source of ATP becomes substrate-level phosphorylation, such as the ones catalysed by phosphoglycerate kinase and pyruvate kinase during glycolysis or the transfer of a phosphate group from acetyl-phosphate to ADP catalysed by acetate kinase (Clark, 1989). In order to maintain a constant glycolytic flux to supply ATP, the generated NADH must be reoxidised to provide the NAD^+ required for glucose oxidation. The oxidation of NADH during fermentation is coupled to the reduction of pyruvate. This can be achieved through two main pathways. Firstly, pyruvate can be cleaved by pyruvate formate-liase to yield formate and acetyl-CoA. Acetyl-CoA can then be reduced to ethanol. Reduction of acetyl-CoA to ethanol is carried out by the binfunctional aldehyde-alcohol dehydrogenase (*adhE*), and allows the reoxidation of 2 NADH molecules (Clark, 1989). Alternatively, acetyl-CoA can be converted to acetyl phosphate by phosphate acetyltransferase, and then into acetate by acetate kinase (producing 1 ATP molecule) (Sawers & Clark, 2004). On the other hand, on the late exponential and stationary phases of growth, pyruvate can be reduced to lactate by lactate dehydrogenase (*ldhA*) (Tarmy & Kaplan, 1968a; Tarmy & Kaplan, 1968b), oxidising 1 NADH molecule in the process (**Figure 3.1**). Since anaerobic fermentation results only in partial oxidation of the substrate (such as glucose), the ATP yield is considerably reduced compared to other metabolic modes, particularly aerobic respiration (**Figure 3.1**).

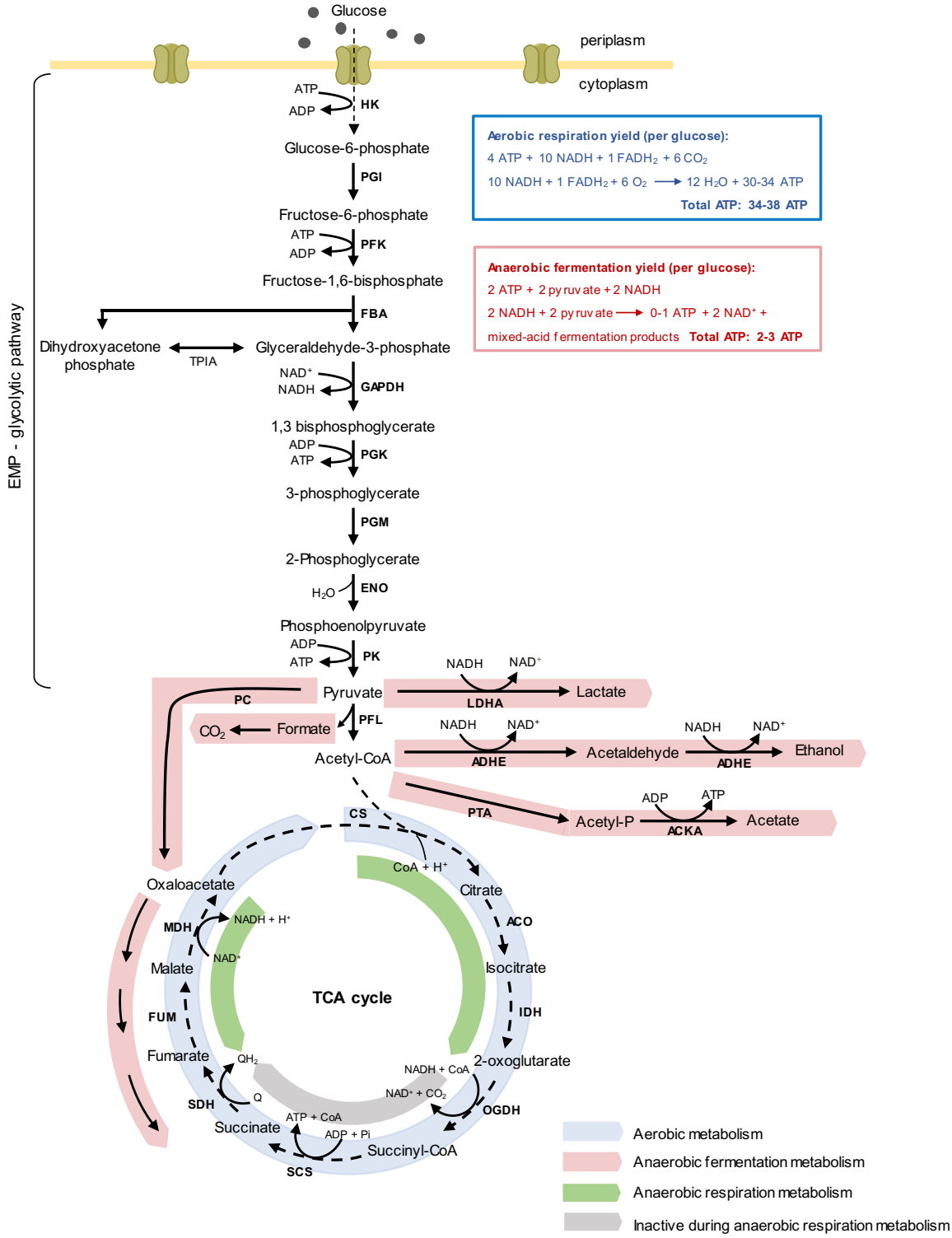


Figure 3.1. Glucose metabolism in the three main metabolic modes of *E. coli*. Glucose is first metabolized through the EMP glycolytic pathway, resulting in the net generation of 2 ATP, 2 NADH and 2 pyruvate molecules per glucose. During aerobic respiration (light blue), pyruvate is converted into acetyl-CoA, which then enters the TCA cycle to generate more NADH, $FADH_2$ and ATP. NADH and $FADH_2$ can then be oxidised through the respiratory chain using O_2 as the final electron acceptor to generate the proton gradient across the membrane required for ATP synthesis through oxidative phosphorylation. In anaerobic respiration metabolism (green), a terminal acceptor different than O_2 is used for the respiratory chain, such as nitrate, fumarate or DMSO. The TCA cycle is not fully functional, and is instead split into two branches. The NADH and ATP yields depend on the growth conditions and terminal electron acceptors available, but are always lower than during aerobic respiration. Finally, in

anaerobic respiration (red) neither oxygen nor alternative terminal electron acceptors are available, and respiratory chains are not active. ATP is generated by substrate-level phosphorylation only, and the NADH produced during glycolysis is used to reduce internal metabolites to provide the constant supply of NAD^+ required to sustain glycolysis. Some additional ATP can be generated by acetate kinase.

In the present work, the requirement of NAD^+ reoxidation to sustain growth under anaerobic fermentation conditions was exploited to develop a high-throughput, broadly applicable selection system based on growth complementation. As the basis of the selection system, I constructed an *E. coli* strain unable to sustain anaerobic growth through mixed-acid fermentation due to the accumulation of reduced NADH. I hypothesized that this could be achieved by knocking out the genes coding two of the main enzymes involved in mixed-aced fermentation, alcohol dehydrogenase (*adhE*) and lactate dehydrogenase (*ldhA*), which would be sufficient to make *E. coli* unable to regenerate oxidised NAD^+ at a high enough rate to sustain growth under anaerobic fermentation conditions. While it is widely accepted that NAD^+ regeneration is the main reason why anaerobic fermentation supports anaerobic growth, there is no direct evidence that restoration of NADH recycling alone can reestablish growth under anaerobic fermentation conditions of strains with impaired anaerobic fermentation. Thus, I also aimed to test if such metabolic impairment could be complemented by transforming cells with a plasmid leading to the expression of a reductase activity able to transfer electrons from NADH to an electron acceptor, which could be provided into the culture media if required (**Figure 3.2**). Such recovery of anaerobic growth by regeneration of oxidised NAD^+ was termed “redox rescue”.

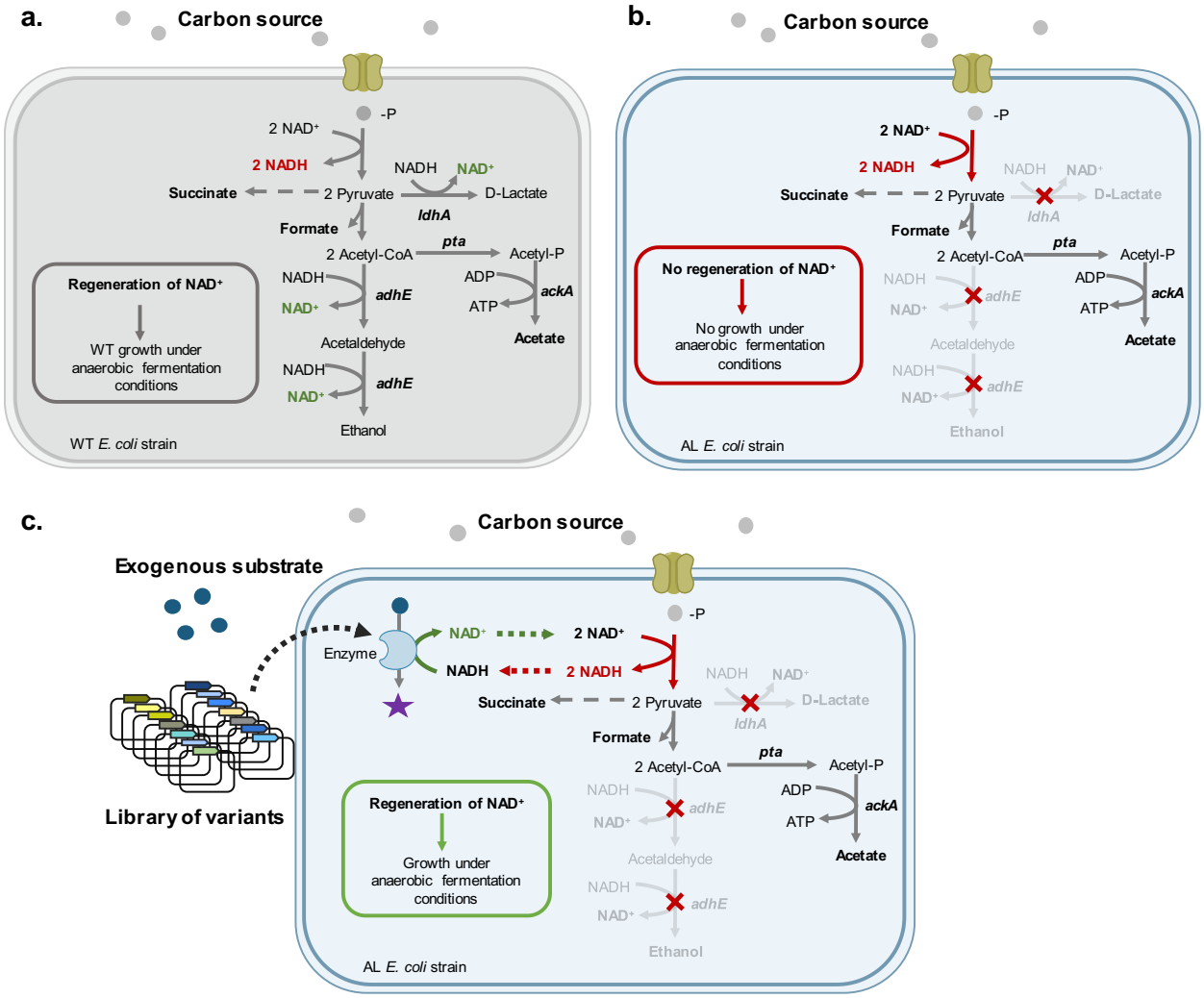


Figure 3.2. Redox rescue system. **a.** Wild-type *E. coli* can grow anaerobically in the absence of alternative terminal electron acceptors by performing anaerobic fermentation. In anaerobic fermentation, the NADH generated during glycolysis must be oxidised to provide a constant supply of NAD^+ and allow anaerobic glycolysis to continue providing ATP through substrate-level phosphorylation. This is achieved by using NADH to reduce internal metabolites. NAD^+ regeneration is mostly due to the reduction of acetyl-CoA to ethanol by alcohol dehydrogenase (*adhE*) and of pyruvate to lactate by lactate dehydrogenase (*ldhA*). Other fermentation products are acetate, formate and succinate. **b.** As the basis of the selection system, I generated a mutant *E. coli* strain (the AL strain) with impaired anaerobic fermentation by knocking out the *adhE* and *ldhA* genes. Due to its inability to regenerate NAD^+, the AL strain is unable to grow under anaerobic fermentation conditions. **c.** I hypothesized growth under anaerobic fermentation conditions of the AL strain could be restored by transforming cells with an exogenous NADH-dependent oxidoreductase and supplementing, if required, the culture medium with the oxidised substrate for the exogenous enzyme. The oxidoreductase would then employ NADH to reduce the substrate, producing also NAD^+ and therefore allowing anaerobic growth. If cells were transformed with a library of variants of an oxidoreductase, only those receiving a functional variant would be able to grow under anaerobic fermentation conditions, enabling the use of the AL strain as a selection platform.

3.2. Results and discussion

3.2.1. Construction and characterization of the redox rescue *E. coli* strain

Previous studies have reported that the major *E. coli* fermentation products are ethanol, lactate, acetate and formate (Clark, 1989). Amongst these, only the production of ethanol and lactate is directly coupled to the oxidation of NADH. Thus, it was hypothesized that mutating *adhE* and *ldhA* would result in impaired anaerobic fermentation. The *E. coli* AL double mutant strain (Δ*adhE,* Δ*ldhA*) was derived from the JW1228 strain (Δ*adhE*) of the Keio collection (Baba *et al*., 2006). First, the kanamycin resistance cassette present in the *adhE* locus of the JW1228 strain was removed through FLP-FRT recombination with pCP20 (Cherepanov and Wackernagel, 1995). Then, the *ldhA* gene was knocked out following the protocol developed by Hamilton *et al*. based on homologous recombination (Hamilton *et al*., 1989).

In order to confirm the successful knock out of both *adhE* and *ldhA*, genomic DNA was extracted and PCR with external and internal primers targeting both genes were performed. External primers were designed to bind flanking sequences upstream and downstream of each gene, whereas internal primers bind to regions contained within the coding sequences. Electrophoresis in agarose gel was employed to determine the size of the amplified products for both the AL strain and the wild-type BW25113 strain (**Figure 3.3**). No amplification products were detected when internal primers were used for either gene. Additionally, the products obtained with external primers for genomic DNA of the AL strain were shorter than for the BW25113 strain, confirming the absence of *adhE* and *ldhA* in the genome of the AL strain.

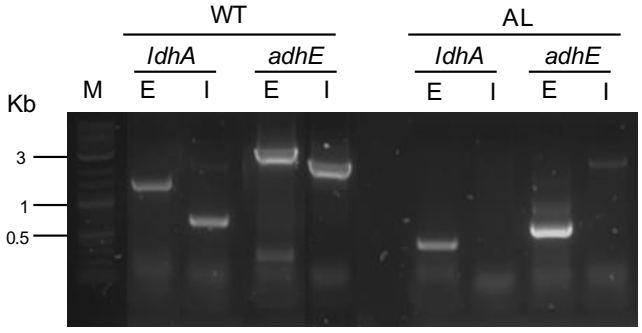


Figure 3.3. Characterization of the genotype of wild-type and AL strain *E. coli* cells. The size of PCR products obtained with primers targeting flanking (E) or internal (I) regions of the *adhE* and *ldhA* genes was determined for both wild-type and AL strain *E. coli* cells. The products obtained for wild-type cells were as expected (for *adhE*, 3167 base pairs with E primers and 2190 base pairs with I primers; for *ldhA*, 1404 base pairs with E primers and 681 base pairs with I primers). On the other hand, for AL cells no bands where observed with I primers, and smaller products were obtained with E primers (414 base pairs for *ldhA* and 491 base pairs for *adhE*) due to the deletion of *ldhA* and *adhE*. M denotes the DNA size marker.

AL cells were able to grow under aerobic conditions as efficiently as wild-type cells confirming that *adhE* and *ldhA* are only required under anaerobic fermentation conditions. However, they were unable to grow under anaerobic conditions in M9 minimal medium with glucose, indicating that anaerobic fermentation was impaired (**Figure 3.4**).

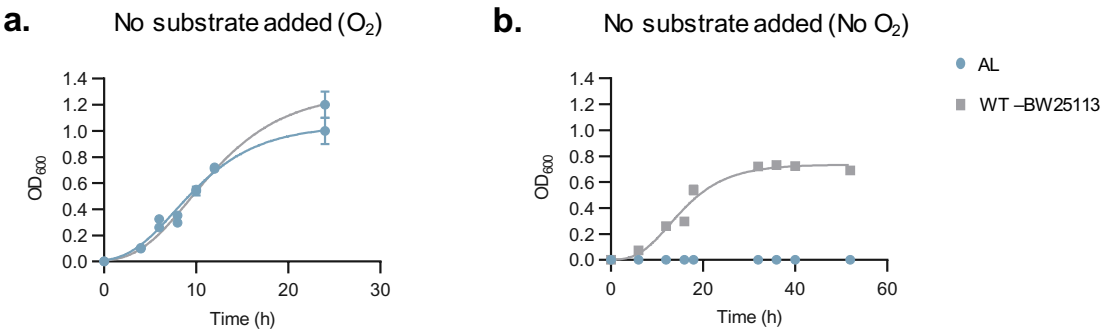


Figure 3.4. Inability of AL cells to grow under anaerobic fermentation conditions. Both wild-type and AL cells were able to grow in minimal medium in the presence of oxygen. However, AL cells were unable to grow anaerobically in minimal medium, due to impaired anaerobic fermentation. Error bars show the standard deviation for each time point.

3.2.2. Validation of the redox rescue system with an endogenous enzyme

I first confirmed that *E. coli* AL cells could be complemented by transforming them with a plasmid containing the native *E. coli adhE* gene in a pUC19 vector backbone and growing them anaerobically. AL cells transformed with *adhE* and cultured anaerobically were able to grow as efficiently as wild-type cells, reaching a maximum optical density at 600 nm (OD_{600}) of 0.8 after 23 h and indicating that *adhE* alone is able to regenerate NAD^+ at a high enough rate to sustain growth through anaerobic fermentation (**Figure 3.5**). This is in accordance with the fact that ethanol production from acetyl-CoA requires the oxidation of two NADH molecules, resulting in the highest yield of oxidised NAD^+ per glucose molecule amongst the different fermentative products.

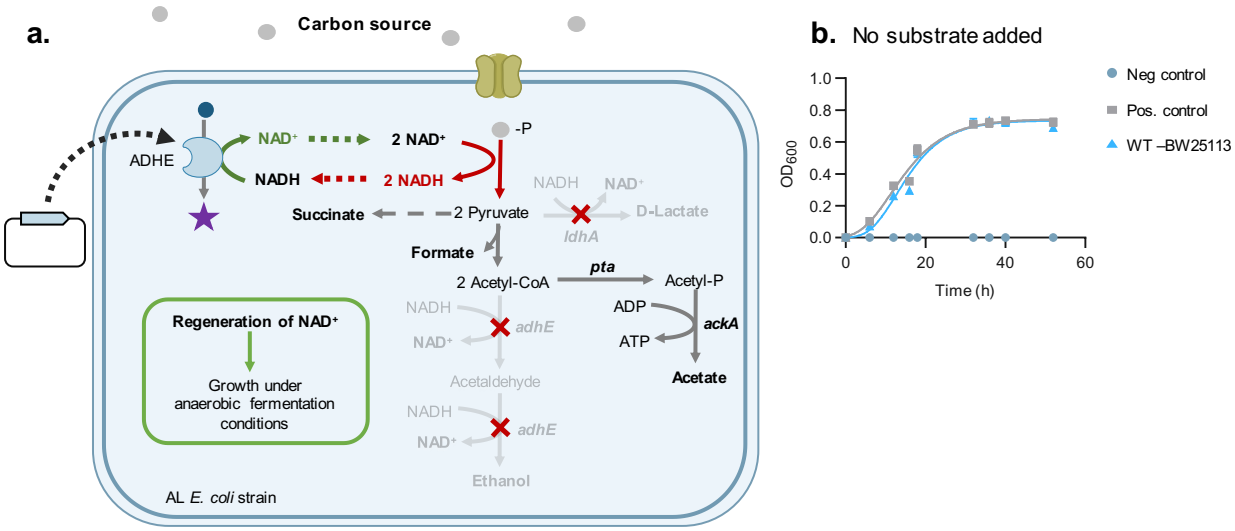


Figure 3.5. Complementation of AL cells with the native *adhE* gene. a. Schematics of complementation with *adhE*. Upon transformation with *adhE*, AL cells recover the ability to reduce acetyl-CoA to ethanol, which allows them to generate NAD^+ and grow anaerobically. **b**. AL cells transformed with *adhE* (Pos. control) were able to grow under anaerobic fermentation conditions as efficiently as wild-type cells, whereas those transformed with the empty pUC19 vector (Neg. control) were unable to grow. In the subsequent growth complementation experiments, AL cells transformed with *adhE* were used as the positive control, while AL cells transformed with empty pUC19 were used as the negative control. Error bars show the standard deviation for each time point.

The concentration of typical *E. coli* fermentation products in the fermentation broth of both wild-type and AL cells transformed with *adhE* was determined in order to confirm that *adhE* activity was responsible for the restoration of anaerobic growth. Due to its ability to determine the concentration of several metabolites in a single experiment easily and accurately, proton nuclear magnetic resonance (^{1}H-NMR) was the chosen technique to analyse the fermentation broths. The concentration of the different metabolites was determined by comparing the area of a characteristic peak of each of them to the area of the signal corresponding to TSP, the reference compound added before the acquisition of ^{1}H-NMR spectra. The metabolic profile of wild-type cells confirmed the presence of ethanol, lactate, succinate, formate and acetate as the main fermentation products, in accordance with previous characterization of *E. coli* mixed-acid fermentation. A similar profile was observed for AL cells transformed with *adhE*, with the exception that no lactate was detected due to the lack of a functional *ldhA* gene (**Table 3.1**).

Table 3.1. Quantification of metabolites by ^{1}H-NMR of the fermentation broth of wild-type *E. coli* and AL cells complemented with *adhE*. All typical mixed-acid fermentation products were detected in the fermentation broth of wild-type *E. coli* cells. No lactate was detected in the fermentation broth of AL cells transformed with *adhE*, but the presence of all other products was confirmed.

Strain	Plasmid	Encoded enzyme	[Ethanol] (mM)	[Lactate] (mM)	[Succinate] (mM)	[Acetate] (mM)	[Formate] (mM)
BW25113- WT	-	-	14.08	2.35	2.7	16.7	34.3
AL	pLS1	ADHE	13.8	0	5.2	12.1	20.7

Due to its ability to restore anaerobic growth without the need of the addition of an exogenous substrate to the culture medium, *adhE* was used as the positive control in all growth complementation experiments.

3.2.3. Validation of the redox rescue system with exogenous reductases

Next, I attempted to restore anaerobic growth of AL cells by means of exogenous reductases whose substrates are not commonly produced by *E. coli*. Metabolic complementation was tested with a set of three different NADH-dependent reductases: acetoin reductases from *Bacillus subtilis* (*bdhA,* Uniprot accession number: O34788) and *Klebsiella pneuomoniae* (*budC*, Uniprot accession number: Q48436), and alcohol dehydrogenase from *Thermus* sp. ATN1 (TADH, Uniprot accession number: B2ZRE3). The enzymes were chosen on the basis that their substrates were not expected to be generated or metabolized by *E. coli* native enzymes under the conditions of the growth complementation experiments.

Both acetoin reductases catalyse the reduction of acetoin to 2,3-butanediol coupled to the oxidation of NADH. This is a key reaction for organisms performing 2,3-butanediol fermentation, which include a number of bacteria from the *Klebsiella*, *Bacillus*, *Serratia* and *Pseudomonas* genera, but not *E. coli* (Garg and Jain, 1995). TADH, on the other hand, is a thermostable NADH-dependent alcohol dehydrogenase able to reduce a range of aldehydes and cyclic ketones to the corresponding alcohols (Höllrigl *et al*., 2008) (**Figure 3.6**). Two related substrates previously described to be accepted by TADH were chosen to attempt metabolic complementation: cyclohexanone, and 3-methylcyclohexanone.

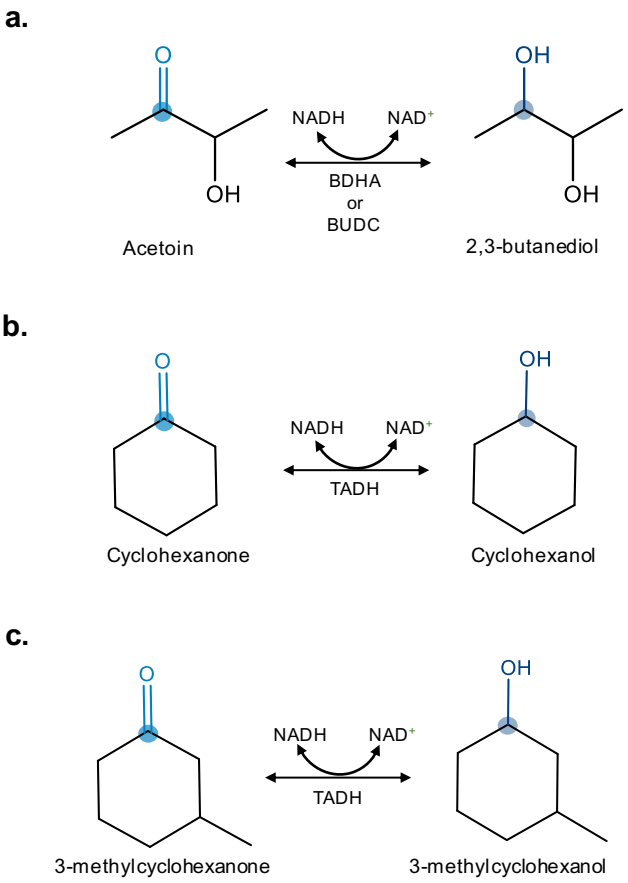


Figure 3.6. Reactions catalysed by the exogenous reductases chosen to attempt anaerobic growth complementation. a. Both BDHA and BUDC are acetoin reductases able to catalyse the reduction of acetoin to 2,3-butanediol. **b, c.** Some of the reactions catalysed by TADH. TADH is able to catalyse the reduction of large range of ketones and oxidation of the corresponding alcohols, using NAD as the cofactor. The reactions for the substrates used for anaerobic growth complementation are displayed.

AL cells transformed with any of the exogenous reductases were able to grow anaerobically in media supplemented with the corresponding oxidised substrate (acetoin for *bdhA* and *budC*, and cyclohexanone or 3-methylcyclohexanone for TADH**, Figure 3.7**). Cells transformed with either acetoin reductase reached an OD_{600} higher than 0.6 during the stationary growth phase, similar to control cells transformed with a functional copy of *adhE* to complement the Δ*adhE* mutation. In the case of cells transformed with TADH, the OD_{600} at the stationary phase remained below 0.2, likely due to the toxic effects of cyclohexanone, 3-methylcyclohexanone and the corresponding alcohols (Lee *et al*., 2015). The toxicity of all four compounds was confirmed by culturing wild-type *E. coli* cells aerobically in M9 medium supplemented with glucose and each of the compounds (**Figure 3.8**).

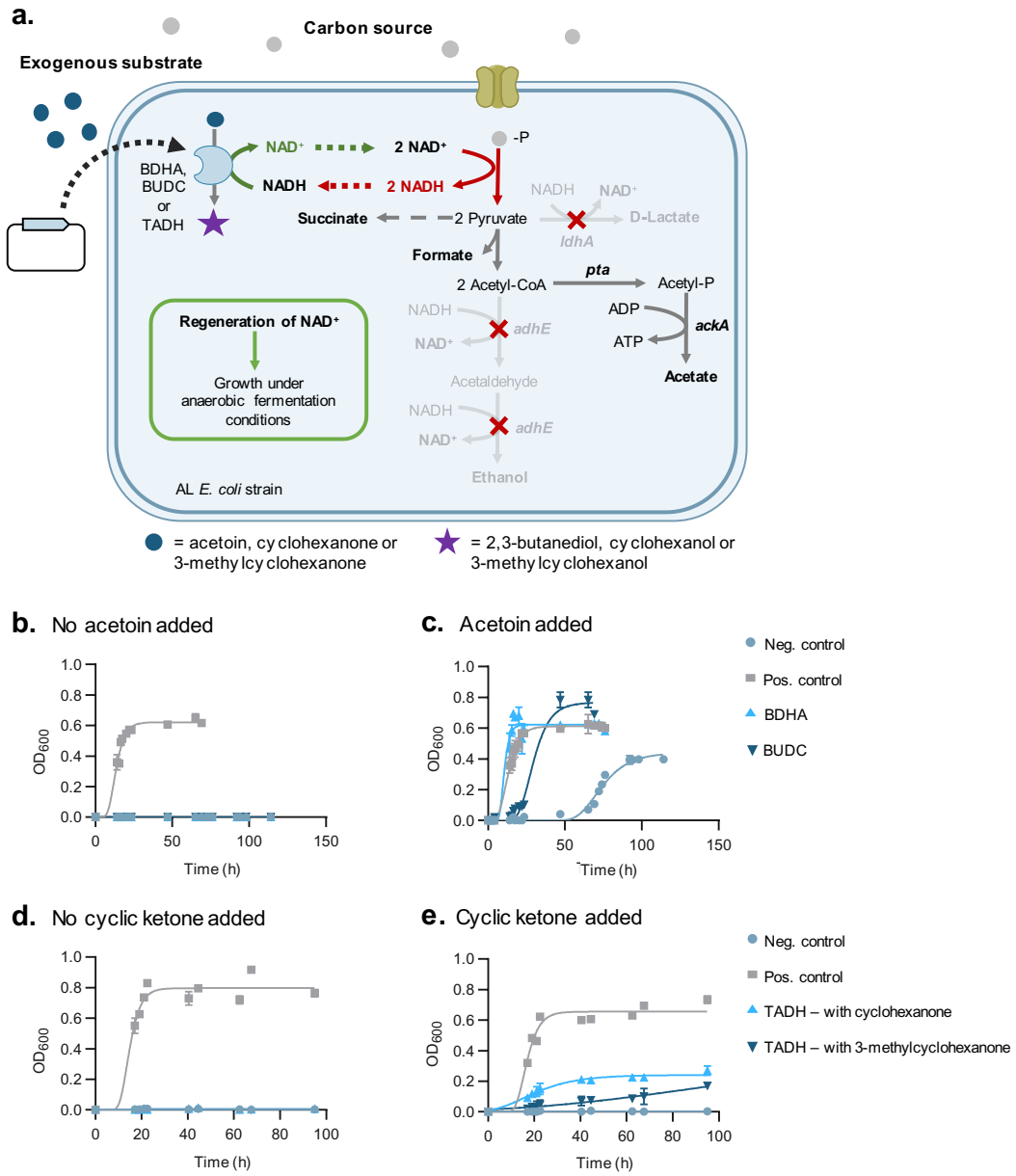


Figure 3.7. Metabolic complementation with exogenous NAD-dependent oxidoreductases. a. When AL cells are transformed with an exogenous NAD-dependent oxidoreductase whose oxidised substrate is not natively produced by *E. coli* cells, the substrate must be added to the medium. Transformed cells can then reduce it thanks to the activity of the exogenous oxidoreductase, restoring the ability to grow under anaerobic fermentation conditions. Three NAD-dependent oxidoreductases were tested: two acetoin reductases (BDHA and BUDC), and one alcohol dehydrogenase (TADH). The oxidised substrates was supplied in the culture medium (acetoin for BDHA and BUDC, or cyclohexanone or 3-methylcyclohexanone for TADH). **b, c.** AL cells transformed with *bdhA* or *budC* were able to grow anaerobically in medium supplemented with acetoin, but not in medium without acetoin. Interestingly, the negative control (AL cells transformed with an empty pUC19 vector) were also able to eventually grow in medium with acetoin, probably due to the activity of an unidentified endogenous acetoin reductase. **d, e.** AL cells transformed with *tadh* were able to grow in medium supplemented with cyclohexanone or 3-methylcyclohexanone, but not in medium without any of these substrates. The maximum OD_{600} reached by AL cells expressing TADH and grown in the presence of cyclohexanone or 3-methylcyclohexanone remained below 0.2, due to the toxicity of the substrates and the corresponding alcohols. Error bars show the standard deviation for each time point.

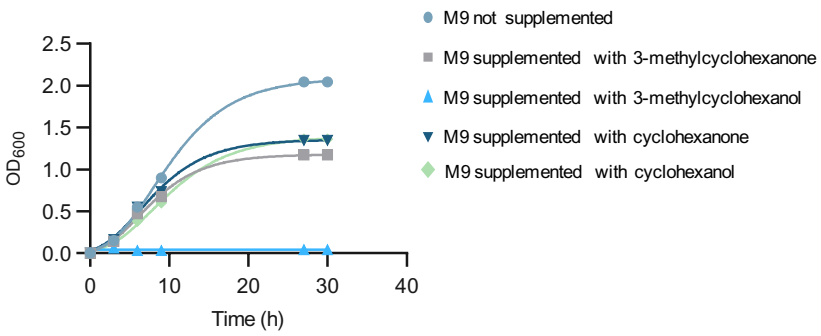


Figure 3.8. Toxicity of cyclohexanone, 3-methylcyclohexanone and the corresponding alcohols. The aerobic growth of wild-type *E. coli* cells was negatively affected by the presence in the culture medium of 10 mM cyclohexanone, 3-methylcyclohexanone, cyclohexanol or 3-methylcyclohexanol. Particularly, 3-methylcyclohexanol reduced very severely cellular growth. Error bars show the standard deviation for each time point.

Interestingly, anaerobic growth was also eventually observed when AL cells transformed with a control empty plasmid were grown in medium supplemented with acetoin, and ^{1}H-NMR spectra revealed the presence of 2,3-butanediol in the fermentation broth (**Table 3.2**). However, the lag phase lasted over 50 h, a considerably larger period than the 10 and 20 h observed for BDHA and BUDC, respectively. This result suggests that one or more endogenous enzymes are able to support anaerobic growth of AL cells in the presence of acetoin. Most likely, this is due to the activity of an endogenous acetoin reductase not expressed under standard culture conditions. In order for cells to grow, changes in the gene expression pattern must take place, which lead to an increased lag phase when compared to cells where the required reductase can be readily expressed from a high-copy plasmid. While *E. coli* has not been shown to be a native 2,3-butanediol producer, several sequences in its genome, such as the *yohF* and *idnO* loci (Reed *et al.*, 2003), have been annotated based on sequence similarity with other dehydrogenases as putative acetoin reductases or diacetyl reductases, some of which have been described to also reduce acetoin to 2,3-butanediol (Yang *et al*., 2014; Zhao *et al.,* 2015).

^{1}H-NMR spectra confirmed the oxidised substrate was consumed by AL cells when transformed with the appropriate reductase. Accordingly, the reduced substrate was produced in all cases, confirming that growth recovery was indeed coupled to regeneration of oxidised NAD^{+} by the activity of the exogenous enzyme. Acetate and formate were also detected in the fermentation broth of AL cells complemented with any of the exogenous reductases, while only small amounts of succinate were produced (**Table 3.2**).

Table 3.2. Quantification of metabolites by ^{1}H-NMR of the fermentation broth of AL cells complemented with BDHA, BUDC and TADH. The expected reduced product was detected in all cases, as well as all mixed-acid fermentation products except ethanol and lactate.

Plasmid	Encoded enzyme	Exogenous substrate	Resulting product	Substrate δ * (ppm)	Product δ * (ppm)	[Substrate] (mM)	[Product] (mM)	[Ethanol] (mM)	[Lactate] (mM)	[Succinate] (mM)	[Acetate] (mM)	[Formate] (mM)
pUC19	-	Acetoin	2,3-butanediol	1.38 (d, 3)	1.15 (d,6)	3.2	5.6	0	0	1.1	6.6	4.1
pLS2	BDHA	Acetoin	2,3-butanediol	1.38 (d, 3)	1.15 (d,6)	0	8.4	0	0	0.5	6.7	3.7
pLS3	BUDC	Acetoin	2,3-butanediol	1.38 (d, 3)	1.15 (d,6)	0.2	8.3	0	0	0.7	7.9	5
pLS12	TADH	Cyclohexanone	cyclohexanol	2.40 (t, 4)	1.24 (m,5)	0	10.4	0	0	0.8	10.1	8.3
pLS12	TADH	3-methylcyclohexanone	3-cyclohexanol	1.01 (d,3)	0.88 (d,3)	4.4	4	0	0	0.4	3.7	2.9

* δ of characteristic signal of substrate or product (ppm). The multiplicity of each signal (s – singlet, d – doublet, t – triplet, q – quartet, m – multiplet) and the number of contributing protons is shown between brackets.

3.3. Conclusions

An *E. coli* strain with impaired fermentation due to an abolishment of NAD^+ regeneration has been built. Such metabolic impairment was achieved by knocking out *adhE* and *ldhA* in the AL strain, which are responsible for most of the NAD^+ regeneration under anaerobic fermentation conditions. Since such genes are not essential for growth in other conditions, the only observed phenotype of AL cells was the inability to sustain anaerobic growth through fermentation.

It has been demonstrated that the activity of a reductase able to oxidise NADH can restore anaerobic fermentative growth, allowing a sustained ATP supply through substrate-level phosphorylation. The simplest metabolic complementation case consists of expressing a reductase whose substrate is typically produced by *E. coli*, such as the *adhE*. It has also been demonstrated that, even if an appropriate reductase substrate is not endogenously produced in the cells, metabolic complementation can be achieved by adding it to the culture medium.

Such a system provides a powerful platform for the high-throughput selection of variants of a large variety of biomolecules. The most immediate application is the selection of novel NADH-dependent oxidoreductase variants, which can be achieved simply by transforming AL cells with a library of the oxidoreductase of interest and growing them anaerobically in media supplemented with the substrate towards which activity is desired. NADH-dependent oxidoreductases are one of the largest enzyme groups, comprising over 50% of oxidoreductases and 15% of all enzymatic activities described in BRENDA (Sellés Vidal *et al*., 2018) (**Figure 3.9**). However, the scope of the selection system can be easily expanded well beyond these limits by indirectly linking the activity of the biomolecule of interest to an increased NADH-dependent reductase activity.

In the next chapters, the versatility and robustness of the selection system are demonstrated by applying it to select novel variants of a range of different biomolecules of special industrial interest with a variety of new or improved properties.

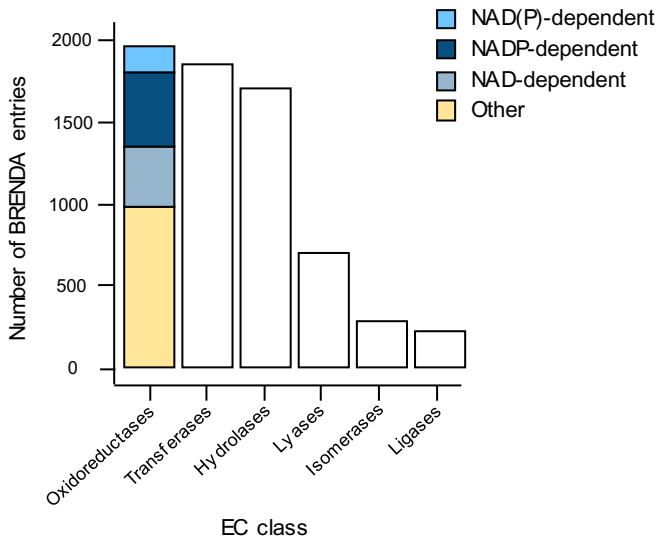


Figure 3.9. Distribution of enzymatic activities. Based on BRENDA entries, oxidoreductases constitute roughly 30% of all the BRENDA enzymatic activities, among which around 50% use NAD and/or NADP as a cofactor. The distribution of the enzymatic activities was obtained from BRENDA by means of manual queries and plotted with R and the ggplot2 package. (Figure adapted from Sellés Vidal *et al.*, 2018)

3.4. Key outcomes

- An *E. coli* strain (AL strain) with impaired anaerobic fermentation has been engineered by knocking out the *adhE* and *ldhA* genes.
- Anaerobic growth can be restored by transforming AL cells with an NAD-dependent oxidoreductase and providing the oxidised substrate in the culture medium.
- The AL strain is unable to grow anaerobically in minimal medium due to its inability to regenerate oxidised NAD^+.

4. Alcohol dehydrogenases: altering cofactor specificity

4.1. Introduction

Alcohol dehydrogenases (EC 1.1.1) constitute one of the most diverse groups of NAD(P)-dependent oxidoreductases. They are able to catalyse the oxidation of a wide range of alcohols, and the reduction of the corresponding aldehydes and ketones, coupled to the transfer of a hydride group to or from the C4 atom of the nicotinamide ring of an NAD(P) cofactor. Three main types of alcohol dehydrogenases have been found, which do not share sequence homology and differ in their three-dimensional structure and catalytic mechanism. Nevertheless, most of them contain at least two domains: a cofactor-binding domain (most commonly comprising a set of two Rossmann folds), and a more variable substrate-binding domain, which provides the residues directly involved in catalysis and determines substrate specificity (Kavanagh *et al.*, 2008; Kutzenko, Lamzin, and Popov, 1998).

Type I alcohol dehydrogenases, which include the classical horse liver alcohol dehydrogenase (Pettersson, 1987), belong to the superfamily of medium-chain dehydrogenases/reductases (MDR). They are dimeric or tetrameric enzymes which typically require one or more zinc ions as cofactors. Most frequently, one of the Zn^{2+} ions plays a pivotal role in the catalytic mechanism. In the absence of the substrate, the catalytic Zn^{2+} is coordinated by cysteine and histidine residues, as well as a water molecule. When the substrate binds to the enzyme, it displaces the water molecule and coordinates the Zn^{2+} ion. The substrate then transfers a proton to the solvent (in the case of the alcohol oxidation reaction), generating an alkoxide intermediate which is stabilized by the Zn^{2+} ion. Finally, the alkoxide ion transfers a hydride group to $NAD(P)^+$, collapsing to an aldehyde or ketone (Baker *et al*., 2009) (**Figure 4.1**).

Type II alcohol dehydrogenases were first identified in *Drosophila melanogaster* (Sofer and Ursprung, 1968), and have been found to form part of the short-chain dehydrogenases/reductases (SDR) superfamily. They do not require zinc as a cofactor, and instead rely on a highly conserved tyrosine residue as the central acid-base catalyst. The pKa of the tyrosine hydroxyl is lowered through the interactions with an adjacent lysine residue and the nicotinamide ring of NAD(P), which facilitate its role as a proton/acceptor donor. When a proton is abstracted from the alcohol substrate, a hydride is transferred to the oxidised $NAD(P)^+$ cofactor (Kavanagh *et al*., 2008) (**Figure 4.1**).

Finally, type III alcohol dehydrogenases comprise iron-dependent enzymes, and have been less extensively studied than their type I and type II counterparts (Gaona-López, Julián-Sánchez, and Riveros-Rosas, 2016). They contain an Fe^{2+} ion coordinated generally by one aspartate and three histidine residues, which lowers the pKa of the hydroxyl group of the

alcohol substrate and thus facilitates the abstraction of a proton from the substrate and the transfer of a hydride to $NAD(P)^+$ (Moon *et al.*, 2011) (**Figure 4.1**). A characteristic feature of type III alcohol dehydrogenases is their sensitivity to oxygen: they become quickly inactivated under aerobic conditions, possibly due to the oxidation of Fe^{2+} and its replacement by other metals, such as Zn^{2+} (Ying *et al.*, 2009, 2007).

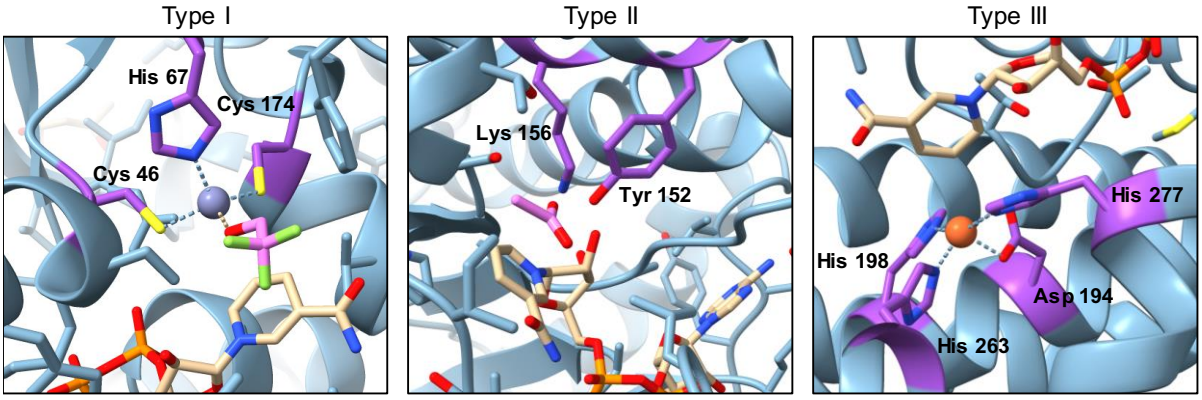


Figure 4.1. Catalytic site of the different types of alcohol dehydrogenases. Representative examples of the structure of the catalytic site of each of the three types of alcohol dehydrogenases are shown. In type I dehydrogenases, a Zn^{2+} ion is coordinated by cysteine and histidine residues, as well as a water molecule which is displaced by the substrate upon its binding. Horse liver alcohol dehydrogenase bound to NAD^+ and trifluoroethanol is displayed (PDB code 4NFS). In type II dehydrogenases, there is no Zn^{2+} ion at the active site. Instead, a tyrosine residue acts as the central acid-base catalyst. *Drosophila melanogaster* alcohol dehydrogenase bound to NADH and acetate is displayed (PDB code 1MG5). Type III dehydrogenases contain a catalytic Fe^{2+} ion, generally coordinated by aspartate and histidine residues. *Zymomonas mobilis* ZM4 alcohol dehydrogenase bound to NAD^+ is displayed (PDB code 3OX4).

Thanks to their ability to perform the asymmetric synthesis of chiral alcohols with high stereospecificity under mild conditions (Hummel, 1997), alcohol dehydrogenases have been extensively used in industrial applications where enantiomerically pure compounds are desired (Patel, 2013). Alcohol dehydrogenases can be used to yield chiral alcohols via the asymmetric reduction of an aldehyde or ketone, but also through less direct techniques such as dynamic kinetic resolution of a racemic mixture or the asymmetrization of meso-diols (Hall and Bommarius, 2011). This is of particular interest in the pharmaceutical industry, where complex and highly specific transformations are required to obtain a pure active compound. Indeed, alcohol dehydrogenase-catalysed steps have been incorporated into the synthesis of a wide variety of pharmaceuticals, including antidepressants, antineoplasics, antiobiotics and antihypertensive and antithrombotic drugs, among many others. Additionally, alcohol dehydrogenases have also proven useful for the bulk synthesis of enantiomerically pure chiral

compounds in other industries, such as hydroxyl esters, lactones, phenylethanol derivatives and specific isomers of 2,3-butanediol (Zheng *et al*., 2017).

Alcohol dehydrogenases have been redesigned through protein engineering to make them more suitable for applications where naturally available enzymes present some limitations, achieving, for example, enhanced kinetic properties (Hummel, 1997), improved activity towards non-natural substrates (Patel, 2013), the ability to efficiently perform the desired catalysis under the non-physiological conditions often required for industrial processes (Wehrmann and Klebensberger, 2018) or altered cofactor preference (Solanki, Abdallah, and Banta, 2017). Altering the cofactor preference of NAD(P)-dependent alcohol dehydrogenases, and in general of NAD(P)-dependent enzymes, is of particular interest due to the higher cost of NADP cofactors (bulk prices per mole of 5,000 $ for $NADP^+$ and 215,000 $ for NADPH, compared to 710 $ for NAD^+ and 3,000 $ for NADH), as well as their lower stability compared to NAD cofactors (Green Catalysis: Biocatalysis, 2013). Cofactor regeneration systems have been developed. Several strategies are available, with enzymatic methods perhaps being one of the most advantageous types (Wu *et al.,* 2013). In enzymatic cofactor regeneration technologies, the oxidation or reduction of a secondary sacrificial substrate by the same enzyme catalysing the main reaction or by a different enzyme is used to regenerate the cofactor in the redox state required for the main reaction (**Figure 4.2**). Furthermore, while efficient cofactor regeneration systems are available for NADH, such as the formate/formate dehydrogenase system, NADPH regeneration systems remain more limited and costly due to the lower efficiency of the available enzymes able to use cheap sacrificial substrates which do not lead to the generation of undesired byproducts (Seelbach *et al*., 1996; Woodyer, Zhao, and Van Der Donk, 2005; Wang *et al.,* 2017; Green Catalysis: Biocatalysis, 2013). There is thus often a great interest in altering the cofactor preference of NADP-dependent enzymes to make them able to catalyse NAD-dependent reactions.

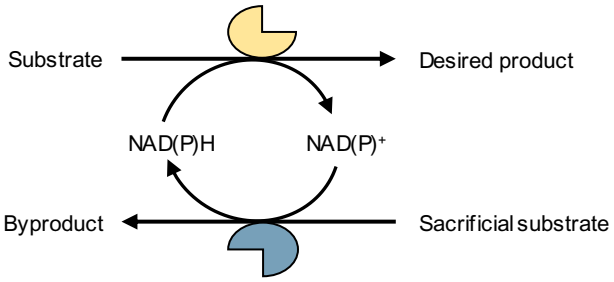


Figure 4.2. Enzymatic cofactor regeneration systems. The main reaction (catalysed by the yellow enzyme) leads to the production of oxidised cofactor. Then, the reduced cofactor is regenerated by means of a secondary reaction (blue enzyme) where a sacrificial substrate is oxidised. The secondary reaction can be catalysed by the same enzyme as the main reaction, or by a different one. Similar systems are available for the regeneration of oxidised cofactors, where the sacrificial substrate is reduced during the secondary reaction.

Cofactor preference reversal was first demonstrated by Scrutton *et al.*, who managed to engineer an NAD-dependent variant of the NADP-dependent *E. coli* glutathione reductase through rational mutagenesis, guided by available structures of related enzymes and sequence alignment information (Scrutton *et al*., 1990). Over a decade later, Rosell *et al.* managed to completely reverse the cofactor preference of *Rana perezi* ADH8, an alcohol dehydrogenase thought to be involved in the reduction of retinal to retinol *in vivo* (Rosell *et al.,* 2003), from NADP to NAD. By mutating three consecutive residues which formed part of the binding pocket of the 2' phosphate group of NADP, they obtained a mutant enzyme with a catalytic efficiency (k_{cat}/K_m) with NAD similar to that of the wild-type enzyme with NADP (Rosell *et al*., 2003).

Nevertheless, such rational approaches have also led to the generation of mutant proteins that not only did not display the expected change in cofactor preference, but also exhibited undesired negative side effects, such as loss of activity or decreased protein solubility. Korkhin *et al.* performed extensive structural studies of *Clostridium beijerinckii* alcohol dehydrogenase (CBADH) as well as *Thermoanaerobacter brockii* alcohol dehydrogenase (TBADH), a thermostable enzyme closely related to CBADH. Both are type I alcohol dehydrogenases with a very strict preference for NADP able to catalyse the reduction of acetone to isopropanol (**Figure 4.3**). After solving the crystal structure of apo and NADP-bound forms for both alcohol dehydrogenases, they predicted a set of point mutations that could potentially reverse the cofactor specificity from NADP to NAD (Korkhin *et al*., 1998). Nevertheless, when such substitutions were tested in *Clostridium autoethanogenum* alcohol dehydrogenase (CAADH), which shares 86% sequence identity with CBADH, they resulted in an insoluble protein variant (Maddock, Patrick, and Gerth, 2015).

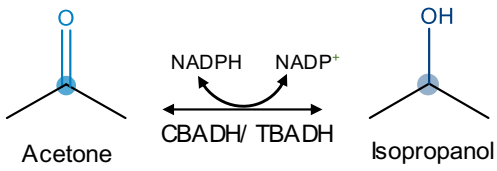


Figure 4.3. Reaction catalysed by CBADH and TBADH. Both enzymes are able to catalyse the reduction of acetone to isopropanol coupled to the oxidation of NADPH to $NADP^+$, as well as the inverse reaction.

More recently, the CSR-SALAD software was developed to guide library generation when attempting to reverse cofactor preference (Cahn *et al*., 2017). CSR-SALAD integrates structural information to predict critical residues for determining cofactor specificity, and suggests a set of potential point mutations to obtain the desired cofactor specificity reversal. However, even when cofactor specificity reversal is achieved, it frequently leads to a

remarkable loss of activity, requiring additional rounds of mutagenesis and screening to identify the optimal variant with the desired cofactor specificity (Chánique and Parra, 2018).

As the first proof-of-concept for the selection system, the selection system was used to obtain an efficient NAD-dependent variant of CBADH. Surprisingly, metabolic complementation was also found to be possible with NADP-dependent enzymes, due to the activity of endogenous transhydrogenases. Therefore, an alternative selection strain was developed, which allows one to select strictly NAD-dependent enzymes. Its efficacy was demonstrated by using it to select NAD-dependent variants of TBADH.

4.2. Results and discussion

4.2.1. Selection of a novel NAD-dependent CBADH variant

Before beginning to generate variants of CBADH, the reported properties of the native enzyme were first verified, particularly in order to confirm the preference of CBADH for NADP and its inability to catalyse the redox reaction with NAD. Wild-type CBADH tagged with a hexa-histidine affinity tag was expressed in *E. coli* and purified by means of affinity chromatography with a nickel column (**Figure 4.4**). Pure wild-type CBADH was used to perform enzymatic assays to test the ability of the enzyme to oxidise isopropanol with NAD^+ and $NADP^+$. CBADH only catalysed the reaction with $NADP^+$ (**Figure 4.4**), and no activity with NAD^+ was observed (**Figure A.1a**).

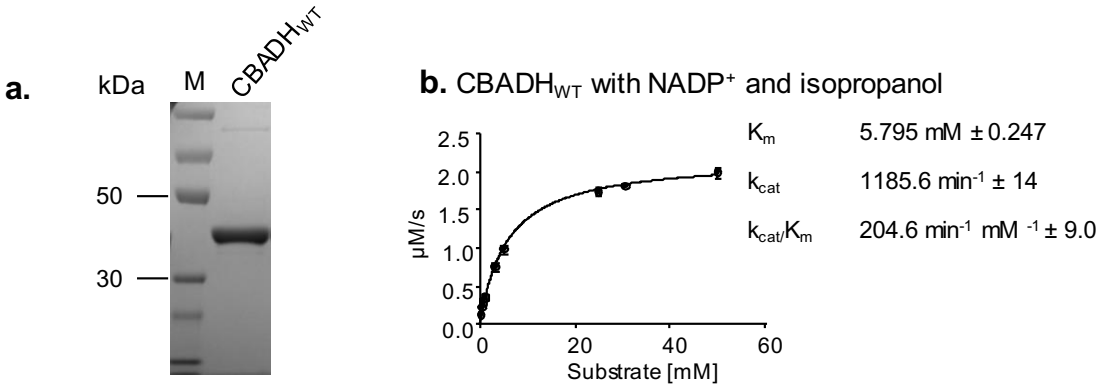


Figure 4.4. Characterization of wild-type CBADH. a. SDS-PAGE confirming the purity of wild-type CBADH purified by nickel-affinity chromatography, expected to have a molecular weight of 38.5 kDa including the hexa-histidine tag. **b.** Fitting of a Michaelis-Menten model to the initial reaction rates obtained with $NADP^+$ and different isopropanol concentrations. No activity with NAD^+ was detected. Error bars show the standard deviation for each concentration.

A library of CBADH variants was generated by site-saturation mutagenesis with the goal of obtaining an NAD-dependent variant. Based on the residues predicted by Korkhin *et al*. to play a key role in determining cofactor specificity, a set of four residues were randomised: Gly198, Ser199, Arg200 and Tyr218 (**Table 4.1**). The pivotal role of such residues was further suggested by the CSR-SALAD software, which ranked them as the top residues to be mutated in order to reverse cofactor preference from NADP to NAD. Inspection of the NADP-bound crystal structure of CBADH (PDB code 1KEV) confirmed the location of all four residues in the proximity of the 2' phosphate group of NADP (**Figure 4.5**). The library was generated by inverse PCR using degenerate primers designed to saturate with an NNN scheme the codons corresponding to the targeted residues, resulting in a library size of over 16 million variants at the nucleotide level, or 160,000 variants at the protein level (without considering variants with premature stop codons). Because of the ability of the selection system to automatically discard all non-functional variants, and the possibility to sample a number of variants in the range of the theoretical library size thanks to the transformation efficiency of *E. coli*, the NNN degeneration scheme was chosen instead of other more restrictive schemes, such as NNK. Plasmid DNA of five randomly-chosen colonies of AL cells transformed with the library was sequenced to confirm the quality of the library. All five colonies were found to carry different CBADH variants (**Table 4.2**).

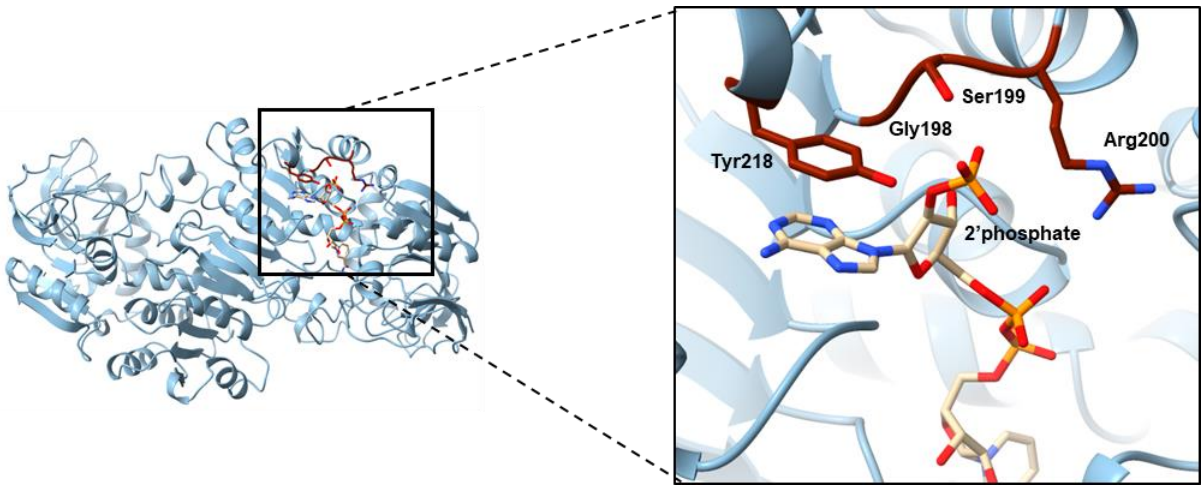


Figure 4.5. Cofactor binding pocket of CBADH. Ser199, Arg200 and Tyr218 establish interactions with the 2' phosphate of NADP. The Tyr218 residue also establishes a stacked-ring interaction with the adenine moiety of NADP. PDB code 1KEV.

Anaerobic growth recovery of AL cells was attempted by transforming them with the library of CBADH variants and supplementing the culture medium with 15 mM acetone. Transformation efficiency was estimated by aerobically plating serial dilutions of transformed cells in aerobic LB medium and counting the number of colonies, resulting in approximately three million cells

successfully transformed per transformation. Therefore, anaerobic growth experiments were set up for six independent transformations of the library, resulting in 18 million variants being sampled. Additionally, metabolic complementation with wild-type CBADH was also tested, as well as with the two sets of point mutations that Korkhin *et al*. originally proposed to switch cofactor preference of CBADH to NAD: Ser199Gly, Arg200Gly and Tyr218Phe ($CBADH_{R1}$); and Gly198Asp, Ser199Gly, Arg200Gly and Tyr218Phe ($CBADH_{R2}$) (**Table 4.1**).

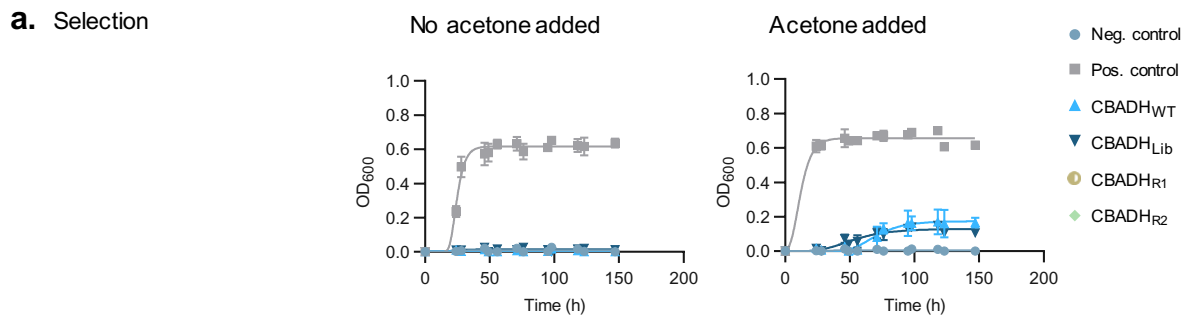
Table 4.1. CBADH site-directed and saturation mutagenesis. Rational design based on predictions by Korkhin *et al.* and site saturation mutagenesis approaches were taken to test the potential of the selection system.

Amino acid position	198	199	200	218
$CBADH_{WT}$	GGC	AGT	CGG	TAC
$CBADH_{R1}$	GGC	GGT	GGG	TTC
$CBADH_{R2}$	GAC	GGT	GGG	TTC
$CBADH_{Lib}$	NNN	NNN	NNN	NNN

Table 4.2. Quality of the CBADH library. The variants contained by five random clones were sequenced, revealing that all variants were different between them and from the wild-type.

Amino acid position	198	199	200	218
$CBADH_{WT}$	GGC	AGT	CGG	TAC
$CBADH_{1}$	TCC	CGG	ACC	TTG
$CBADH_{2}$	TTA	TTC	TAA	Deletion of 210bp
$CBADH_{3}$	TTC	CGT	ATA	CGG
$CBADH_{4}$	TCA	GCT	GTA	TAG
$CBADH_{5}$	AAT	GTG	ACA	GGG

Anaerobic growth of AL cells transformed with the library was observed in medium supplemented with acetone for three independent transformations, after a lag phase of 46 h on average (**Figure 4.6**). Cells grown in each independent transformation of the library were sub-cultured into anaerobic medium with acetone, and allowed to grow anaerobically once again. Growth was observed after 20 h. Then, plasmid DNA was extracted from each culture and sequenced, revealing the presence of the same variant at the protein level in all of them with three residue substitutions: Gly198Asp, Ser199Tyr and Tyr218Pro ($CBADH_{S}$) (**Figure 4.6**). This indicates that a strong selection had taken place, confirming the effectivity of the selective pressure applied with the selection system.



c.

Amino acid position	198	199	200	218
$CBADH_{WT}$	Gly (GGC)	Ser (AGT)	Arg (CGG)	Tyr (TAC)
$CBADH_{R1}$	Gly (GGC)	**Gly (GGT)**	**Gly (GGG)**	**Phe (TTC)**
$CBADH_{R2}$	**Asp (GAC)**	**Gly (GGT)**	**Gly (GGG)**	**Phe (TTC)**
$CBADH_{S}$	**Asp (GAC)**	**Tyr (TAT)**	Arg (AGA)	**Pro (CCG)**

Figure 4.6. Selection of an NAD-dependent CBADH variant. a. AL cells transformed with the library of CBADH variants were able to grow anaerobically in medium supplemented with acetone, but not in medium without acetone. AL cells transformed with wild-type CBADH were also able to grow in medium supplemented with acetone, due to the activity of endogenous transhydrogenases. However, AL cells transformed with the rationally desgined variants of CBADH were unable to grow anaerobically even in the presence of acetone, indicating that they had no activity with NAD and that the mutations had also negatively affected the activity with NADP. **b.** Sub-cultures into the same conditions of the cells transformed with the library and grown in the presence of acetone were performed to further enrich the selected variant before isolating plasmid DNA to sequence it. A similar behavior as for the first culture was observed, with the difference that the sub-culture of cells transformed with the library grew faster thanks to the presence of a larger initial population of cells with the selected variant. **c.** Amino acid residues present at the saturated positions in wild-type CBADH and rationally designed and selected variants. The corresponding DNA codons are shown in brackets. Error bars show the standard deviation for each time point.

Interestingly, while no metabolic complementation was observed for AL cells transformed with the rationally-designed variants $CBADH_{R1}$ or $CBADH_{R2}$, wild-type CBADH allowed the recovery of anaerobic growth in the presence of acetone after a lag phase of 60 h (**Figure**

4.6). Since CBADH did not display any activity with NAD, it was hypothesized that the support of growth by wild-type CBADH could be due to the activity of transhydrogenases, which are able to transfer a hydride group from NADH to $NADP^+$, generating NAD^+ and NADPH, and vice versa. Relevant transhydrogenases are introduced in the next section. On the other hand, the inability of $CBADH_{R1}$ and $CBADH_{R2}$ to restore anaerobic growth of AL cells indicates a loss of activity even with NADP, or another factor such loss of solubility causing these variants to aggregate. The possibility of transhydrogenase-mediated growth complementation was further investigated by generating transhydrogenase knock outs, as described in section **4.2.3**.

4.2.2. Characterization of the selected NAD-dependent CBADH variant

AL cells transformed with a plasmid containing the selected $CBADH_S$ variant were able to grow under anaerobic fermentation conditions when supplemented with acetone. A lag phase of only 8 h was observed. This is a similar lag phase to that observed for cells transformed with *adhE* as the positive control, and a markedly shorter period than when cells where transformed with the library of variants, where a much smaller initial population of cells contain optimized variants (**Figure 4.7**). ^{1}H-NMR spectra of the fermentation broth confirmed the consumption of acetone and the presence of isopropanol, as expected due to the activity of CBADH. Succinate, formate and acetate were also observed (**Table 4.3**).

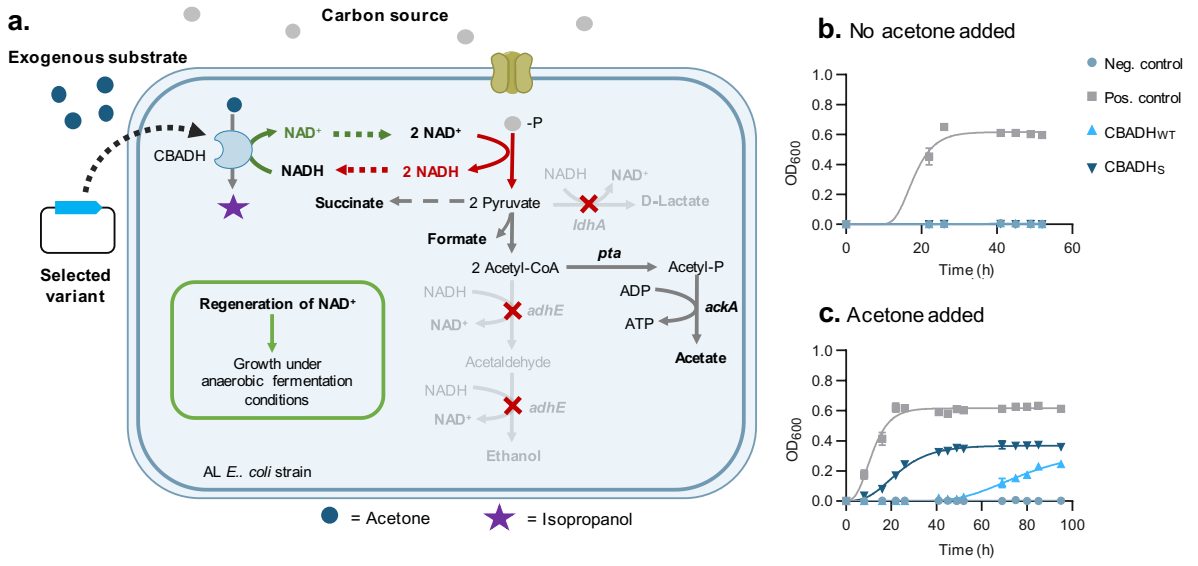


Figure 4.7. Metabolic complementation with the selected NAD-dependent CBADH variant. a. AL cells transformed with the NAD-dependent variant of CBADH would be able to grow anaerobically in medium supplemented with acetone thanks to the regeneration of NAD^+ provided by the CBADH variant upon reduction of acetone to isopropanol. **b, c**. AL cells transformed with $CBADH_S$ were able to grow anaerobically in medium supplemented with acetone, but not in medium without acetone. Cells transformed with wild-type CBADH were also able to grow anaerobically in medium with acetone, although with a lag phase much longer than cells transformed with $CBADH_S$. Error bars show the standard deviation for each time point.

$CBADH_S$ was expressed in *E. coli* and purified in a similar manner to wild-type CBADH, using a hexa-histidine affinity tag, and enzymatic assays were performed with pure protein (**Figure 4.8**). In contrast to wild-type CBADH, $CBADH_S$ was able to catalyse the oxidation of isopropanol with NAD^+, but no activity with $NADP^+$ was detected (**Figure A.1b**), confirming that cofactor preference reversal was successfully achieved. Kinetic parameters for the oxidation of isopropanol were determined, revealing a decrease of around 3.5 times in k_{cat} and a 3-fold increase in K_m when compared to the NADP-dependent reaction catalysed by wild-type CBADH (**Figure 4.8**). The approximately 10-fold reduced catalytic efficiency is in accordance to previous reports showing that mutating residues directly involved in the cofactor-binding pocket often lead to at least partial loss of activity. Nevertheless, the decrease in activity in $CBADH_S$ was considerably less severe than most cases where the cofactor preference of an alcohol dehydrogenase is swapped from NADP to NAD (Chánique and Parra, 2018), thanks to the simultaneous selection of optimal activity which takes place with the selection system.

Table 4.3. Quantification of metabolites by ^{1}H-NMR of the fermentation broth of AL cells complemented with wild-type CBADH or $CBADH_S$. Isopropanol was detected in both cases.

Strain	Plasmid	Encoded enzyme	Exogenous substrate	Resulting product	Substrate δ * (ppm)	Product δ * (ppm)	[Substrate] (mM)	[Product] (mM)	[Ethanol] (mM)	[Lactate] (mM)	[Succinate] (mM)	[Acetate] (mM)	[Formate] (mM)
AL	pLS6	$CBADH_{WT}$	Acetone	Isopropanol	2.24 (s, 6)	1.18 (d,6)	0.1	10.8	0	0	1.1	8.1	5.8
AL	pLS10s	$CBADH_S$	Acetone	Isopropanol	2.24 (s, 6)	1.18 (d,6)	0.5	12.7	0	0	1.1	13	9.9

* δ of characteristic signal of substrate or product (ppm). The multiplicity of each signal (s – singlet, d – doublet, t – triplet, q – quartet, m – multiplet) and the number of contributing protons is shown between brackets.

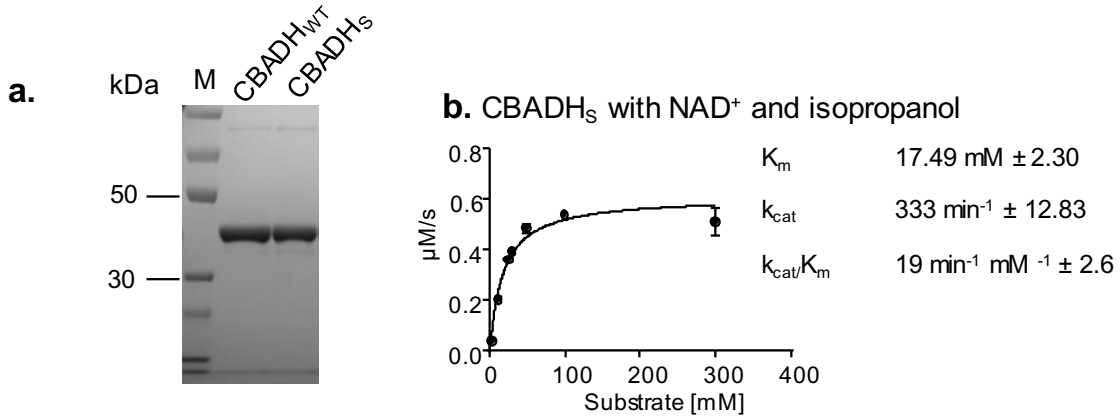


Figure 4.8. Characterization of CBADH$_S$. a. SDS-PAGE confirming the purity of CBADH$_S$ purified by nickel-affinity chromatography, expected to have a molecular weight of 38.5 kDa including the hexa-histidine tag. **b.** Fitting of a Michaelis-Menten model to the initial reaction rates obtained with NAD$^+$ and different isopropanol concentrations. No activity with NADP$^+$ was detected. Error bars show the standard deviation for each concentration.

Interestingly, the Gly198Asp mutation observed in CBADH$_S$ is among those suggested by Korkhin *et al*. The glycine residue is part of a motif in the Rossmann fold involved in binding the additional phosphate of NADP and comprising a glycine residue followed by two more hydrophilic residues, which has been found in alcohol dehydrogenases from distantly related organisms, including CBADH, TBADH and others (Rosell *et al*., 2003). Such a glycine residue is thought to play a major role in determining the cofactor specificity of alcohol dehydrogenases. Indeed, a Gly223Asp (the equivalent position to CBADH Gly198) substitution was one of the mutations intentionally introduced by Rosell *et al.* to switch the cofactor specificity of NADP-preferring *Rana perezi* ADH8. Nevertheless, other mutations (Thr224Ile and His225Asn) of *Rana perezi* ADH8 were indispensable for obtaining the desired effect on the cofactor preference, since a single mutant containing only the Gly223Asp mutation still used NADP preferentially over NAD, displaying only a modest effect on the catalytic efficiency with both NAD and NADP (Rosell *et al*., 2003). Furthermore, Maddock *et al.* also found that a Gly198Asp substitution in CAADH, in conjunction with other point mutations (Ser199Val and Pro201Glu), resulted in a variant still highly active with NADP and with barely any activity with NAD, requiring an additional Tyr218Ala substitution to abolish NADP-dependent activity and gain an NAD-dependent acetone reductase activity high enough to characterize kinetic parameters (Maddock *et al*., 2015).

4.2.3. Construction of a strict selection system for NAD-dependent activity

In order to test the hypothesis that wild-type CBADH was able to restore growth under anaerobic fermentation conditions thanks to the activity of transhydrogenases (by regenerating oxidised NAD^+ from NADH and $NADP^+$), I decided to test if metabolic complementation could also take place in an *E. coli* strain where, in addition to *adhE* and *ldhA*, transhydrogenase genes were also knocked out.

Two types of transhydrogenase have been identified: energy-linked and non-energy linked transhydrogenases. Energy-linked transhydrogenases are integral membrane proteins that couple the transhydrogenation reaction to translocation of protons across the membrane where they are located. Therefore, in order to perform their reaction, they use the proton motive force, spending energy at the expense of ATP generation. They are found in the mitochondrial inner membrane in eukaryotic cells, as well as in the cell membrane of some heterotrophic and photosynthetic bacteria. On the other hand, non-energy linked transhydrogenases are soluble flavoproteins which require FAD as a cofactor. This type of transhydrogenases has only been identified in certain heterotrophic bacteria. Most frequently, only one of the two types is present in the genome of any given organism, with the exception of the *Enterobacterioceae* family, which includes *E. coli*, whose genomes encode one transhydrogenase of each type (Sauer *et al*., 2004). The physiological role of both types of transhydrogenase remains unclear. It is thought that, in general, energy-linked transhydrogenases yield NADPH and NAD^+, while non energy-linked ones generate mostly $NADP^+$ and NADH. This is in accordance with the typically observed high NAD^+:NADH ratio but low $NADP^+$:NADPH ratio within the cell (Bennett *et al*., 2009), which would make the generation of NADH and $NADP^+$ the thermodynamically favoured direction without energy input. Nevertheless, the preferred directionality of each transhydrogenase type is controversial, and possibly varies between different organisms and growth conditions (Sellés Vidal *et al*., 2018). The reactions catalysed by transhydrogenases can be summarized by the following equations:

Energy linked transhydrogenases: $\mathrm{NADH} + \mathrm{NADP^+} + \mathrm{H^+_{out}} \rightleftarrows \mathrm{NAD^+} + \mathrm{NADPH} + \mathrm{H^+_{in}}$

Non-energy linked transhydrogenases: $\mathrm{NADH} + \mathrm{NADP^+} \rightleftarrows \mathrm{NAD^+} + \mathrm{NADPH}$

In *E. coli*, the *sthA* gene (also named *udhA*) encodes a non-energy linked transhydrogenase, while the *pntAB* genes encode the two subunits of an energy-linked transhydrogenase. Due to the ongoing controversy about the directionality of the reaction for each enzyme, individual knock outs for each of the transhydrogenases were constructed, as well as a double knock out. The AL strain was used as the parental strain, resulting in three new mutants: ALS (Δ*adhE*

Δ*ldhA* Δ*sthA*), ALP (Δ*adhE* Δ*ldhA* Δ*pntA*) and ALPS (Δ*adhE* Δ*ldhA* Δ*pntA* Δ*sdhA*) (**Figure 4.9**).

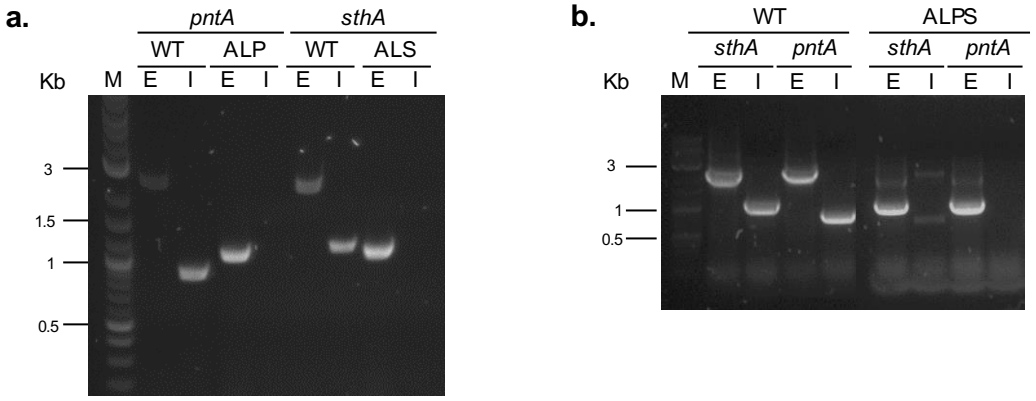


Figure 4.9. Characterization of the genotype of *E. coli* strains ALS, ALP and ALPS. a. Characterization of the ALS and ALP strains. The size of PCR products obtained with primers targeting flanking (E) or internal (I) regions of the *sthA* or *pntA* genes was determined for ALS or ALP cells respectively, as well as wild-type *E. coli* cells. The products obtained for wild-type cells were as expected (for *sthA*, 2503 base pairs with E primers and 1102 base pairs with I primers; for *pntA*, 2609 base pairs with E primers and 903 base pairs with I primers). On the other hand, for ALS and ALP cells no bands where observed with I primers, and smaller products were obtained with E primers (1102 base pairs for *sthA* in ALS cells, and 1069 base pairs for *pntA* in ALP cells) due to the deletion of *sthA* or *pntA*. M denotes the DNA size marker. **b**. ALPS cells combined the genotypes observed for ALS and ALP cells, with the sizes of products with primers targeting *sthA* or *pntA* matching those obtained with the ALS or ALP strains respectively, due to the deletion of both *sthA* and *pntA*.

First the viability of the triple and quadruple mutant strains was tested aerobically and anaerobically. As expected, the ALS, ALP and ALPS strains were unable to grow under anaerobic fermentation conditions, similar to the AL strain. All three strains grew aerobically in minimal medium as efficiently as wild-type and *E. coli* AL strain cells (**Figure 4.10**), suggesting that transhydrogenases do not play an essential role under normal aerobic respiration culture conditions with glucose as the main energy and carbon source.

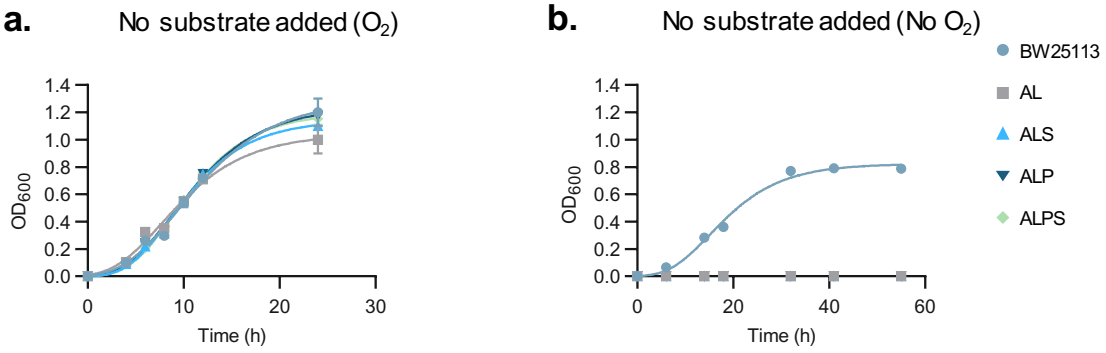


Figure 4.10. Inability of ALS, ALP and ALPS cells to grow under anaerobic fermentation conditions. Similarly to AL cells, the ALS, ALP and ALPS strains were able to grow in minimal medium in the presence of oxygen as efficiently as wild-type cells. However, they were unable to grow anaerobically in minimal medium. Error bars show the standard deviation for each time point.

I then tested if anaerobic growth complementation in the ALS, ALP and ALPS strains was possible with both wild-type CBADH and the previously-selected NAD-dependent variant $CBADH_S$. Either enzyme was able to restore anaerobic growth in ALS and ALP (**Figure 4.11)**, indicating that both *sthA* and *pntAB* are able to catalyse the transhydrogenation reaction in the direction of NADPH and NAD^+ production. In the case of ALS cells, a longer lag phase than for the ALP strain was observed (70 h and 35 h, respectively), suggesting that a larger pool of oxidised $NADP^+$ must be built up in order for *pntAB* to yield NAD^+.

However, in the case of ALPS, no anaerobic growth was observed when transformed with NADP-dependent wild-type CBADH, and only when transformed with NAD-dependent $CBADH_S$ was metabolic complementation achieved (**Figure 4.11**). These results confirm that transhydrogenase activities are responsible for the anaerobic growth observed in the AL strain when transformed with an NADP-dependent enzyme. Furthermore, the ALPS strain provides a valuable tool for a more stringent selection of NAD-dependent enzymes, since NADP-dependent activities cannot support its growth under anaerobic fermentation conditions (**Figure 4.12**). The redox rescue selection system can then be operated in two modes: a permissive system with the AL strain allowing growth recovery to take place by the regeneration of either NAD^+ or $NADP^+$, which could be desirable when there is no preference over the cofactor used by the evolved enzymes (for example, when purely aiming to alter substrate selectivity) since it could potentially allow for a larger range of mutations to be accessible, or a stricter system with the ALPS strain where only regeneration of NAD^+ supports anaerobic growth.

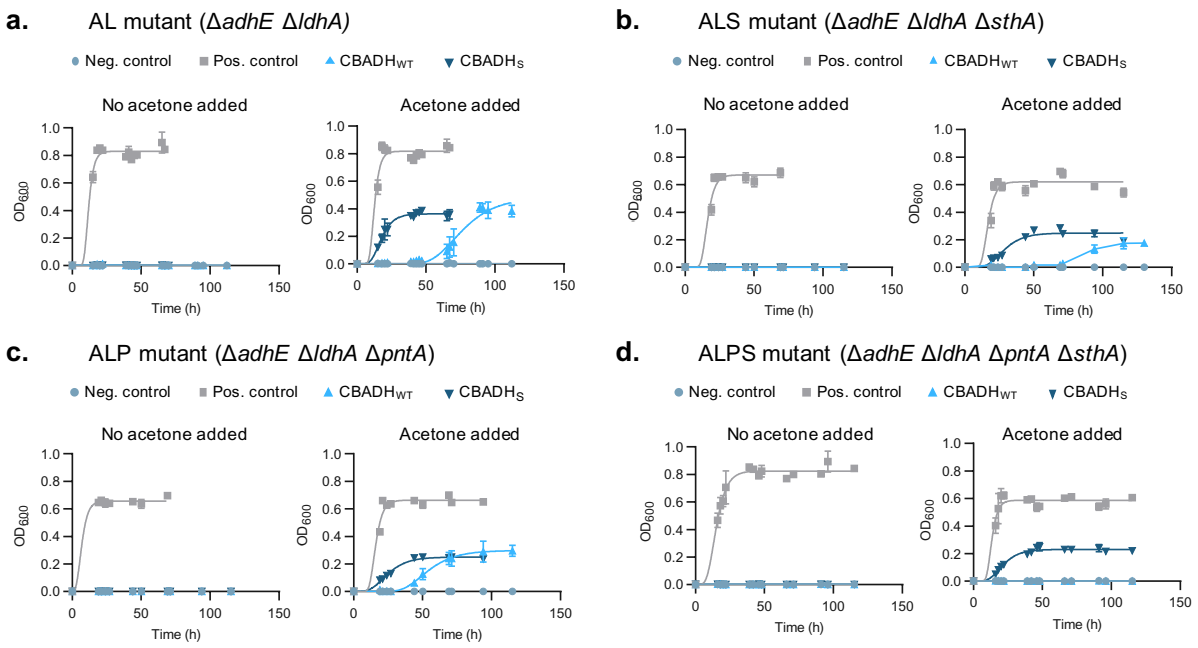


Figure 4.11. Anaerobic growth complementation with NAD and NADP-dependent oxidoreductases of AL, ALS, ALP and ALPS cells. For each *E. coli* mutant strain, anaerobic growth with (right panels) and without (left panels) acetone supplemented to the culture media was followed. **a.** AL mutant (Δ*adhE* Δ*ldhA*). **b.** ALS mutant (Δ*adhE* Δ*ldhA* Δ*sthA*). **c.** ALP mutant (Δ*adhE* Δ*ldhA* Δ*pntA*). **d.** ALPS mutant (Δ*adhE* Δ*ldhA* Δ*pntA* Δ*sthA*). Anaerobic growth of cells with at least one active transhydrogenase was recovered upon transformation of either an NAD or an NADP-dependent oxidoreductase. However, in the case of ALPS cells, where both transhydrogenases were knocked-out, only the NAD-dependent enzyme restored anaerobic growth, indicating that metabolic complementation by NADP-dependent enzymes is mediated by transhydrogenases. Error bars show the standard deviation for each time point.

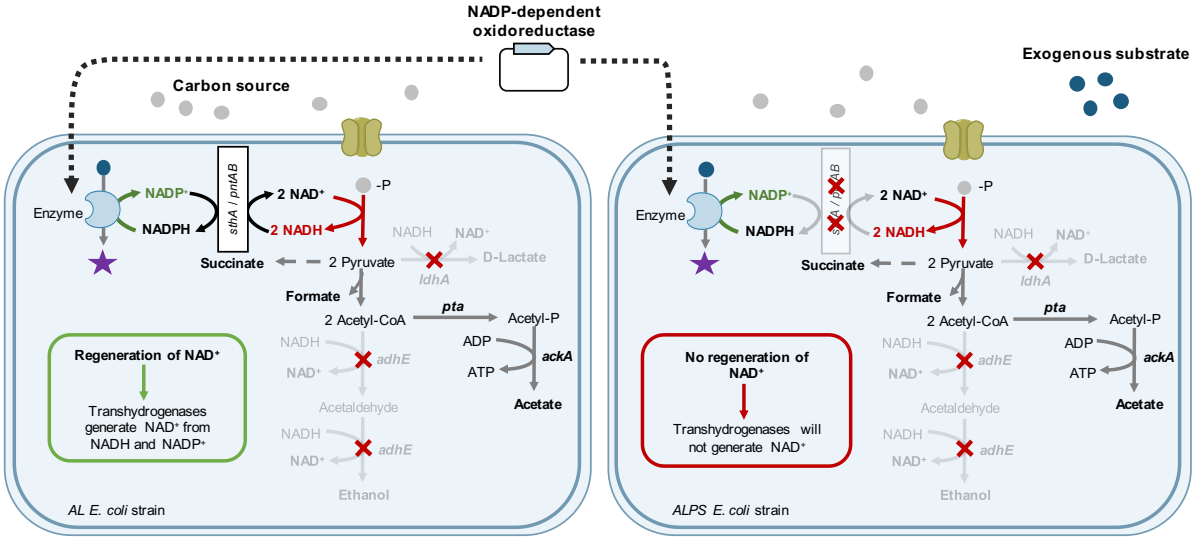


Figure 4.12. Transhydrogenase-mediated metabolic complementation. When ALPS cells (right) are transformed with an NADP-dependent oxidoreductase and its oxidised substrate is added to the culture medium, $NADP^+$ is produced, but no NAD^+ is regenerated, and therefore cells are unable to grow anaerobically. In the case of AL cells (left), when they are transformed with an NADP-dependent oxidoreductase, transhydrogenases use the $NADP^+$ produced by the exogenous oxidoreductase to regenerate NAD^+, which allows them to grow under anaerobic fermentation conditions.

4.2.4. Selection of NAD-dependent variants with the ALPS strain

In order to demonstrate the viability of the ALPS strain to act as a selection platform in a similar fashion to the AL strain, it was employed to select NAD-dependent variants of TBADH, an NADP-dependent alcohol dehydrogenase from *Thermoanaerobacter brockii* closely related to CBADH. Similarly to CBADH, wild-type TBADH was able to support the anaerobic growth of AL cells in medium supplemented with acetone, but not of ALPS cells (**Figure 4.13**). Enzymatic assays with purified protein revealed that, as expected, it was only able to accept $NADP^+$ as the cofactor (**Figures 4.14 and A.1c**).

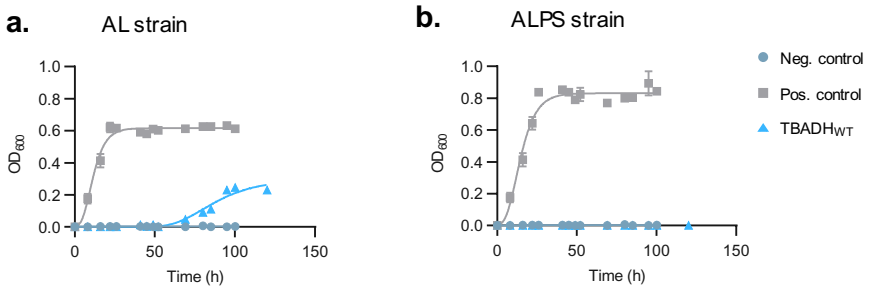


Figure 4.13. Anaerobic growth complementation of AL and ALPS cells with wild-type TBADH. Transformation with wild-type TBADH was able to support the anaerobic growth of AL cells grown anaerobically in medium with acetone, but not of ALPS cells. This is in accordance with the high selectivity of TBADH for NADP, which makes transhydrogenases necessary to support anaerobic growth. Error bars show the standard deviation for each time point.

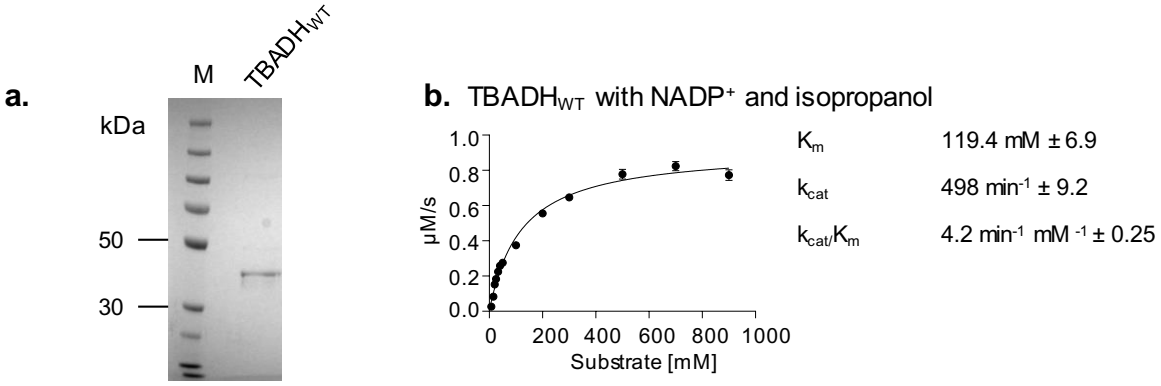


Figure 4.14. Purification and characterization of wild-type TBADH. a. SDS-PAGE confirming the purity of wild-type TBADH purified by nickel-affinity chromatography, expected to have a molecular weight of 38.4 kDa including the hexa-histidine tag. **b.** Fitting of a Michaelis-Menten model to the initial reaction rates obtained with $NADP^+$ and different isopropanol concentrations. No activity with NAD^+ was detected. Error bars show the standard deviation for each concentration.

CBADH and TBADH share a sequence identity of 76%. Most critically, Gly198, Ser199, Arg200 and Tyr218 are conserved in TBADH, and form part of the binding pocket of the 2' phosphate of NADP as shown by the available crystal structure of TBADH bound to $NADP^+$

(PDB code 1YKF) (**Figure 4.15**). Thus, a TBADH library was constructed by means of saturation mutagenesis of said residues (**Table 4.4**). The library was generated by means of inverse PCR with primers designed to degenerate the codons corresponding to the targeted residues with an NNN degeneration scheme, resulting in a library size of over 16 million variants at the nucleotide level and 160,000 variants at the protein level (not including variants where a premature stop codon was introduced). A transformation efficiency of approximately 3 million successfully transformed cells was observed. Six independent transformations of ALPS cells with the library were performed, and cells were inoculated into the selective condition: anaerobic medium with glucose and acetone. A dilution of one of the transformations was grown under aerobic non-selective conditions, and plasmid DNA of five random colonies were sequenced to confirm the quality of the library (**Table 4.5)**. When growth was observed, after 28 h on average, cells were sub-cultured again into minimal medium with glucose and acetone, and grown anaerobically. Plasmid DNA of all three cultures was isolated and sequenced, revealing the presence two different variants ($TBADH_{S1}$ and $TBADH_{S2}$). $TBADH_{S1}$ contained four residue substitutions at the protein level: Gly198Ser, Ser199Lys, Arg200Pro and Tyr218Val. Surprisingly, $TBADH_{S2}$ contained a duplication of part of the Rossmann fold (residues 191 to 241), including the 2' phosphate binding pocket, and all the residues targeted for mutagenesis, inserted between residues 241 and 242 (**Figure 4.16**)**.** Substitutions in the targeted residues were also observed, both in the positions of the original sequence and in the corresponding positions in the insertion. Overall, the observed substitutions were Gly198His, Ser199Arg, Arg200Ala, Tyr218Met, Gly198'Ala and Arg200'Lys, where 198' and 200' denote positions 249 and 251 of $TBADH_{S2}$, which correspond to the positions in the insertion equivalent to the original 198 and 200 residues (**Table 4.6**). Since the duplication was not intentionally introduced during the generation of the library of variants, it can be concluded that it was most likely randomly generated within cells after being transformed with the library, and thus happened at an extremely low frequency, highlighting the potential of the selection system to enrich optimal variants exhibiting the desired properties even when these are present in a very low proportion in the initial library.

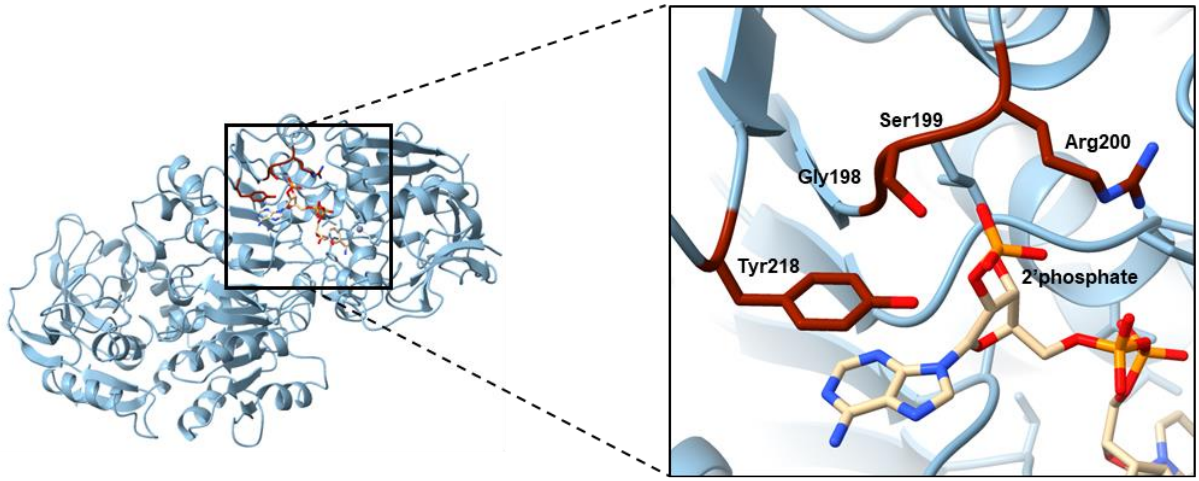


Figure 4.15. Cofactor binding pocket of TBADH. The cofactor binding pocket of TBADH is very similar to that of CBADH, with residues 198, 199, 200 and 218 conserved in both proteins. PDB code 1YKF.

Table 4.4. TBADH saturation mutagenesis.

Amino acid position	198	199	200	218
$TBADH_{WT}$	GGT	TCT	CGT	TAT
$TBADH_{Lib}$	NNN	NNN	NNN	NNN

Table 4.5. Quality of the TBADH library. The variants contained by five random clones were sequenced, revealing that all variants were different between them and from the wild-type.

Amino acid position	198	199	200	218
$TBADH_{WT}$	GGT	TCT	CGT	TAT
$TBADH_{1}$	TGA	CGC	CGA	GTT
$TBADH_{2}$	AAT	CAA	AAG	AGG
$TBADH_{3}$	AGC	ACC	CGA	CGA
$TBADH_{4}$	AAC	TGA	ACT	GTA
$TBADH_{5}$	CCC	CAG	AGG	AGA

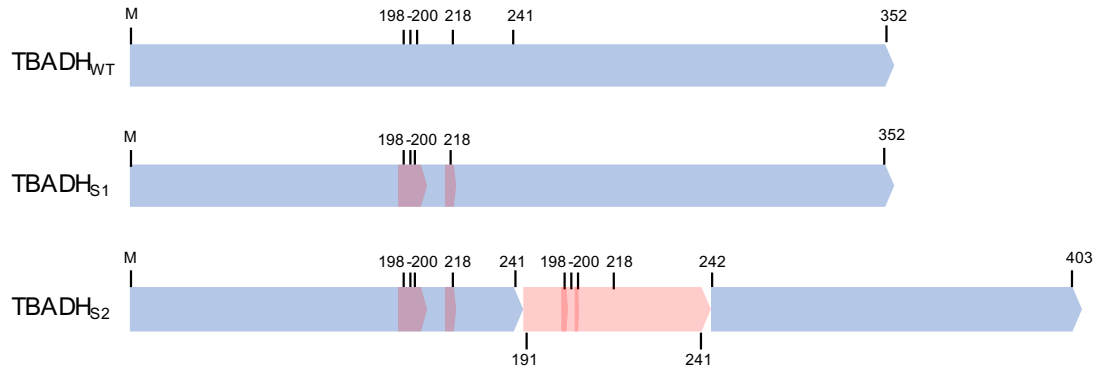


Figure 4.16. Mutations found in TBADH$_{S1}$ and TBADH$_{S2}$. Point mutations are displayed in dark red. TBADH$_{S1}$ contained substitutions in all saturated positions (198, 199, 200 and 218). TBADH$_{S2}$ also contained substitutions in all saturated positions. Additionally, it contained a duplication (displayed in pink) of residues 191 to 241 inserted between positions 241 and 242 of the original sequence. The insertion comprised the saturated positions, some of which were also substituted compared to the original sequence. In particular, residues 198' and 200' (corresponding to positions 249 and 251 in the sequence of TBADH$_{S2}$) were substituted.

Table 4.6. Mutations found in TBADH$_{S1}$ and TBADH$_{S2}$. 198', 199', 200' and 218' denote the positions equivalent to residues 198, 199, 200 and 218 of the original sequence in the duplication present in TBADH$_{S2}$, which correspond to positions 249, 250, 251 and 269 in the sequence of TBADH$_{S2}$. The corresponding DNA codons are shown in brackets.

Amino acid position	198	199	200	218	198'	199'	200'	218'
TBADH$_{WT}$	Gly (GGT)	Ser (TCT)	Arg (CGT)	Tyr (TAT)	-	-	-	-
TBADH$_{S1}$	**Ser (TCA)**	**Lys (AAA)**	**Pro (CCG)**	**Val (GTA)**	-	-	-	-
TBADH$_{S2}$	**His (CAC)**	**Arg (CGG)**	**Ala (GCC)**	**Met (ATG)**	**Ala (GCA)**	**Ser (TCA)**	**Lys (AAA)**	**Tyr (TAC)**

4.2.5. Characterization of the selected NAD-dependent TBADH variants

TBADH$_{S1}$ and TBADH$_{S2}$ were able to support anaerobic growth of ALPS cells in the presence of acetone, with lag phases of 10 and 13 h on average respectively (**Figure 4.17**). ^{1}H-NMR spectra of the fermentation broth of ALPS cells transformed with either variant confirmed the production of isopropanol (**Table 4.7**). After successfully expressing in *E. coli* and purifying by means of nickel affinity chromatography both TBADH variants tagged with a hexa-histidine tag (**Figure 4.17**), their kinetic parameters for the oxidation of isopropanol with NAD and NADP were determined and compared to those of wild-type TBADH. TBADH$_{S1}$ did not display any activity with NADP^{+} (**Figure A.1d**). However, when NAD^{+} was provided as the cofactor, isopropanol oxidation was readily detected, certifying the reversal in cofactor preference (**Figure 4.17**). While the k_{cat} of TBADH$_{S1}$ for the oxidation of isopropanol with NAD^{+} was 4.5 times lower than the k_{cat} of wild-type TBADH for the same reaction with NADP^{+}, a 32-fold

decrease in the K_m for isopropanol was observed when compared to the K_m for isopropanol of wild-type TBADH in the presence of $NADP^+$. The combined effects of the changes in both kinetic parameters resulted in a catalytic efficiency for the oxidation of isopropanol 7.1 times higher than that of wild-type TBADH with $NADP^+$. The highest relative catalytic efficiency previously reported for an alcohol dehydrogenase variant with reversed cofactor preference was 5.89, achieved by Cahn *et al*. for *Saccharomyces cerevisiae* cinnamyl alcohol dehydrogenase after two rounds of mutagenesis and screening (Cahn *et al*., 2017). Therefore, $TBADH_{S1}$ displays, to my knowledge, the highest relative catalytic efficiency obtained in any case of reversal of cofactor specificity for an alcohol dehydrogenase (either from NADP to NAD or vice versa), and the best for reversal of preference from NADP to NAD for any enzyme, as supported by the data reviewed by Chánique and Parra (Chánique and Parra, 2018).

In contrast to $TBADH_{S1}$, $TBADH_{S2}$ was able to oxidise isopropanol both with $NADP^+$ and NAD^+ (**Figure 4.17**). While a decrease in k_{cat} similar to that observed in $TBADH_{S1}$ also took place in $TBADH_{S2}$ for $NADP^+$ and NAD^+, the K_m for isopropanol improved in the presence of both cofactors, albeit to a lesser extent than in $TBADH_{S1}$. The effect on K_m was more marked in the presence of NAD^+, resulting in a relative catalytic efficiency of 1.2 when compared to the wild-type enzyme with $NADP^+$.

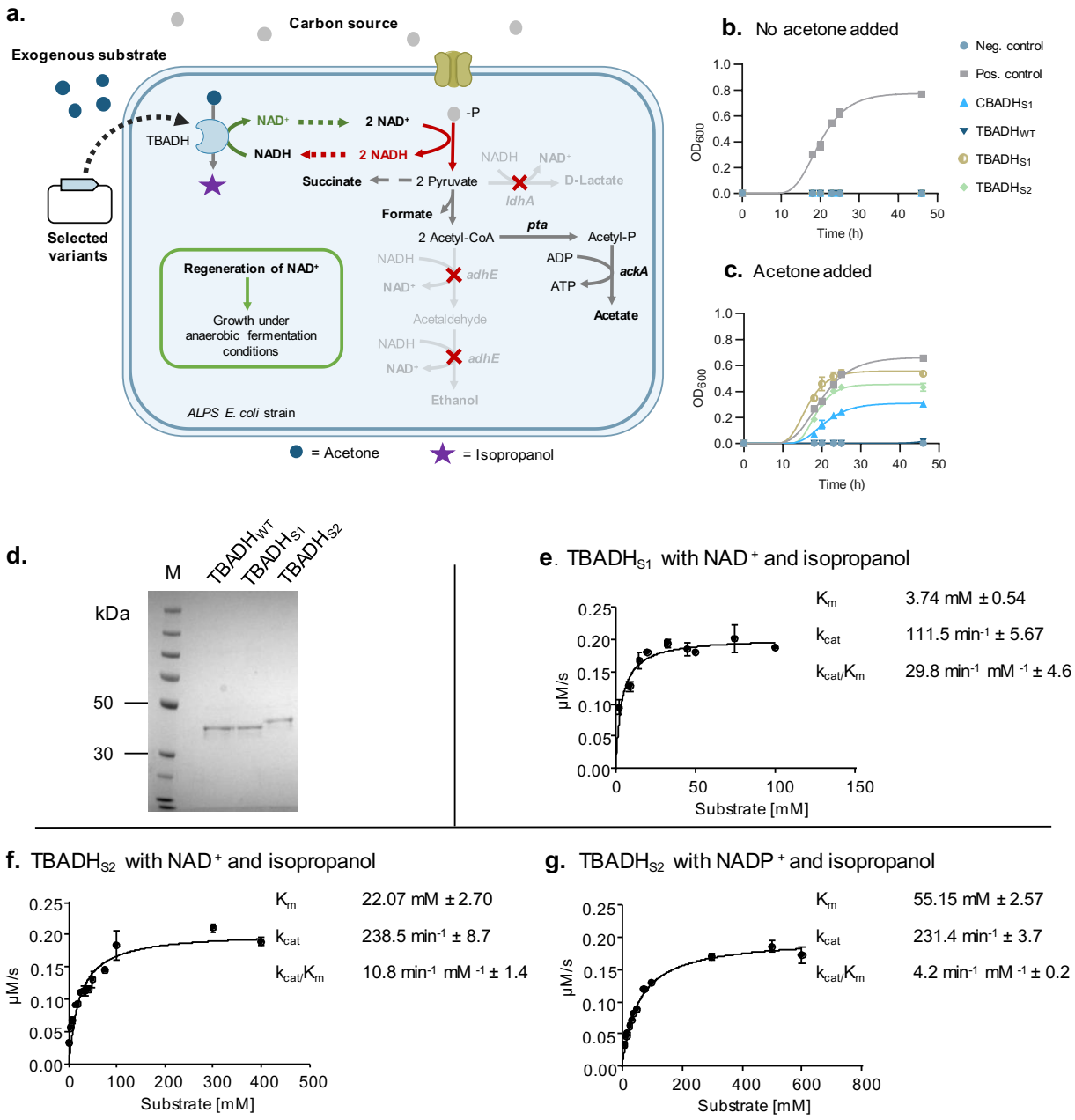


Figure 4.17. Characterization of NAD-dependent TBADH variants. a. Metabolic complementation of ALPS cells with TBADH variants. Since ALPS cells cannot perform transhydrogenase-mediated complementation of anaerobic growth, the transformed TBADH variants must be NAD-dependent. ALPS cells transformed with such variants are able to grow in medium with acetone thanks to the NAD^+ regenerated by the TBADH variants during the reduction of acetone. **b, c.** ALPS cells transformed with either $TBADH_{S1}$ or $TBADH_{S2}$ were able to grow anaerobically in medium supplemented with acetone, but not in medium without acetone. Cells transformed with wild-type TBADH were unable to grow anaerobically in medium with acetone. **d**. SDS-PAGE confirming the purity of $TBADH_{S1}$ and $TBADH_{S2}$ purified by nickel-affinity chromatography, expected to have a molecular weight of 38.4 kDa and 43.7 kDa respectively, including the hexa-histidine tags. **e, f, g.** Fitting of a Michaelis-Menten model to the initial reaction rates obtained with NAD^+ and different isopropanol concentrations for $TBADH_{S1}$ and $TBADH_{S2}$, and also with $NADP^+$ and different isopropanol concentrations for $TBADH_{S2}$. No activity with $NADP^+$ was detected for $TBADH_{S1}$. Error bars show the standard deviation for each time point or concentration.

Table 4.7. Quantification of metabolites by ^{1}H-NMR of the fermentation broth of ALPS cells complemented with $TBADH_{S1}$ and $TBADH_{S2}$. Isopropanol was detected in both cases.

Strain	Plasmid	Encoded enzyme	Exogenous substrate	Resulting product	Substrate δ * (ppm)	Product δ * (ppm)	[Substrate] (mM)	[Product] (mM)	[Ethanol] (mM)	[Lactate] (mM)	[Succinate] (mM)	[Acetate] (mM)	[Formate] (mM)
ALPS	pLS69	$TBADH_{WT}$	Acetone	Isopropanol	2.24 (d, 6)	1.18 (d,6)	8.5	0	0	0	0	0	0
ALPS	pLS73s1	$TBADH_{s1}$	Acetone	Isopropanol	2.24 (d, 6)	1.18 (d,6)	0.1	14.9	0	0	1.9	16	10.8
ALPS	pLS73s2	$TBADH_{s2}$	Acetone	Isopropanol	2.24 (d, 6)	1.18 (d,6)	0.1	15.2	0	0	1.7	17.4	12.8

* δ of characteristic signal of substrate or product (ppm). The multiplicity of each signal (s – singlet, d – doublet, t – triplet, q – quartet, m – multiplet) and the number of contributing protons is shown between brackets.

4.2.6. Structural insights into cofactor preference reversal

In order to investigate the mechanism through which the mutations found in $CBADH_S$, $TBADH_{S1}$ and $TBADH_{S2}$ led to a reversal of the cofactor preference from NADP to NAD (accompanied, in the case of $CBADH_S$ and $TBADH_{S1}$, by a complete loss of activity with NADP), we attempted to crystallize all three variants by means of the sitting drop method with a range of commercially available matrix screens.

Positive hits were obtained under several conditions for CBADHs and $TBADH_{S1}$ after 2-3 days. Optimization of the crystallization conditions was performed for the best hits of each variant. The resulting crystals (**Figure 4.18**) were soaked in NAD^+ before cooling them as preparation for the acquisition of diffraction patterns. The crystals diffracted to a maximum resolution of 2.2 Å for $CBADH_S$ and 2.85 Å for $TBADH_{S1}$.

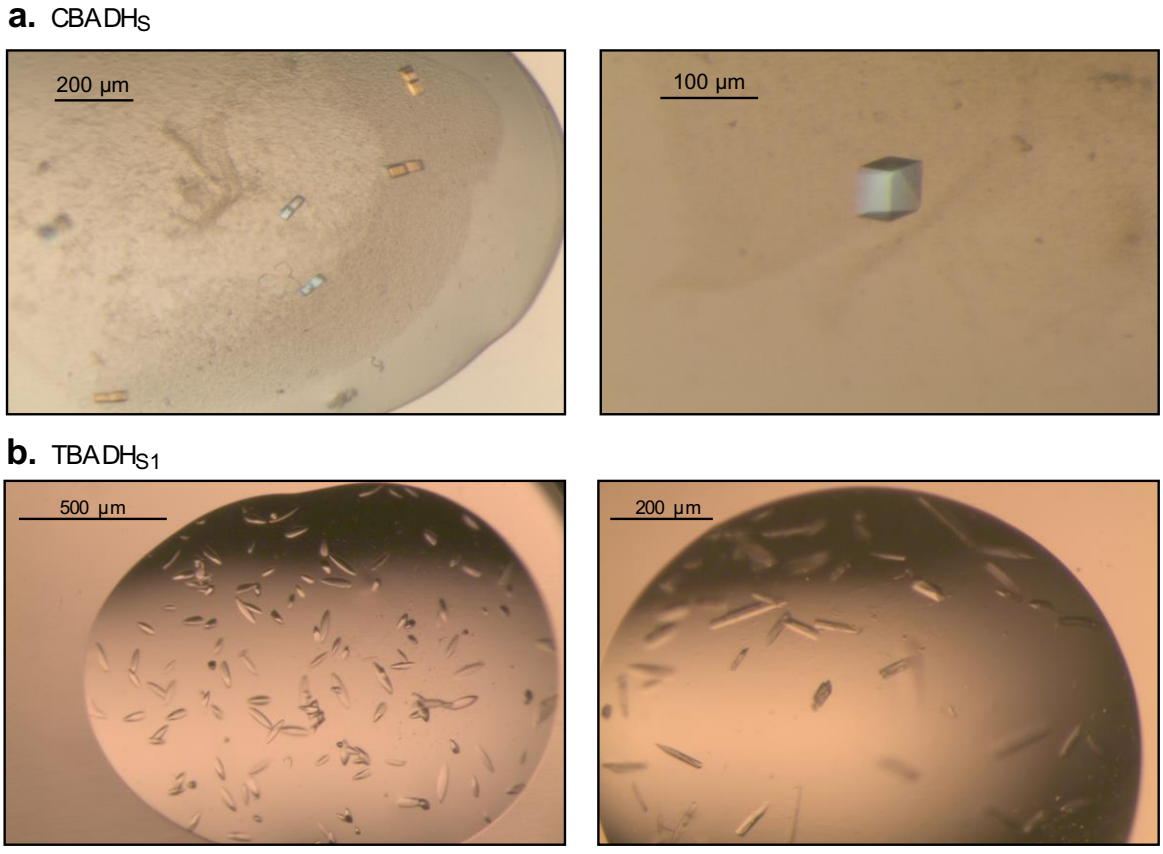


Figure 4.18. Crystals obtained for $CBADH_S$ and $TBADH_{S1}$. Images of crystals obtained after optimization of crystallization conditions are shown. Crystals were obtained by mixing equal volumes of concentrated protein with 900 mM sodium citrate, 100 mM imidazole pH 8 (for $CBADH_S$) or with 20% (w/v) PEG 3K and 100 mM sodium citrate at pH 5.5 (for $TBADH_{S1}$).

The obtained density clearly showed bound NAD^+ for $CBADH_S$, while only partial occupancy of the cofactor-binding pocket was observed for $TBADH_{S1}$ (**Figure 4.19**). The structures

revealed that the size of the cofactor-binding pocket is reduced due to the substitution of residues 198 and 199 by others with bulkier side chains (Gly198Asp and Ser199Tyr for $CBADH_S$, and Gly198Ser and Ser199Lys for $TBADH_{S1}$), which sterically impede the positioning of the 2' phosphate of NADP (**Figures 4.20 and 4.21**). In the case of $CBADH_S$, such effect is further enhanced by the replacement with an aspartate residue, which would also impede NADP binding through electrostatic repulsion between the carboxylate group of its side chain and the 2' phosphate of NADP (**Figure 4.20**). Interestingly, while the Arg200 residue was conserved in $CBADH_S$, it was replaced by a proline in $TBADH_{S1}$, a residue that can have a negative impact on protein stability when incorporated into the polypeptide chain because of its contrained φ dihedral angle and its lack of an amide hydrogen (Bajaj *et al.*, 2007). The introduction of a proline residue as one of the mutations for cofactor preference reversal would not have been an intuitive choice, highlighting the potential of the selection system for the identification of variants with non-obvious changes. It is possible that such substitution allows for a more flexible positioning of the NAD cofactor, compensating potential hindrances in NAD binding introduced by the rest of the substitutions. It is unclear why such substitution was not selected in CBADH. Given the presence of a proline residue in position 201 in both enzymes, at the beginning of an α-helix, it is possible that, in the context of CBADH, the Arg200Pro substitution leads to a destabilization of said α-helix due to the presence of two consecutive prolines. This could result in a partial loss of structure of the cofactor-binding pocket, or the exposing of hydrophobic areas that could increase the tendency to aggregate. In contrast, TBADH might be able to stand such a substitution due to the higher overall stability of its structure, as evidenced by its higher T_m (Bogin *et al*., 1998).

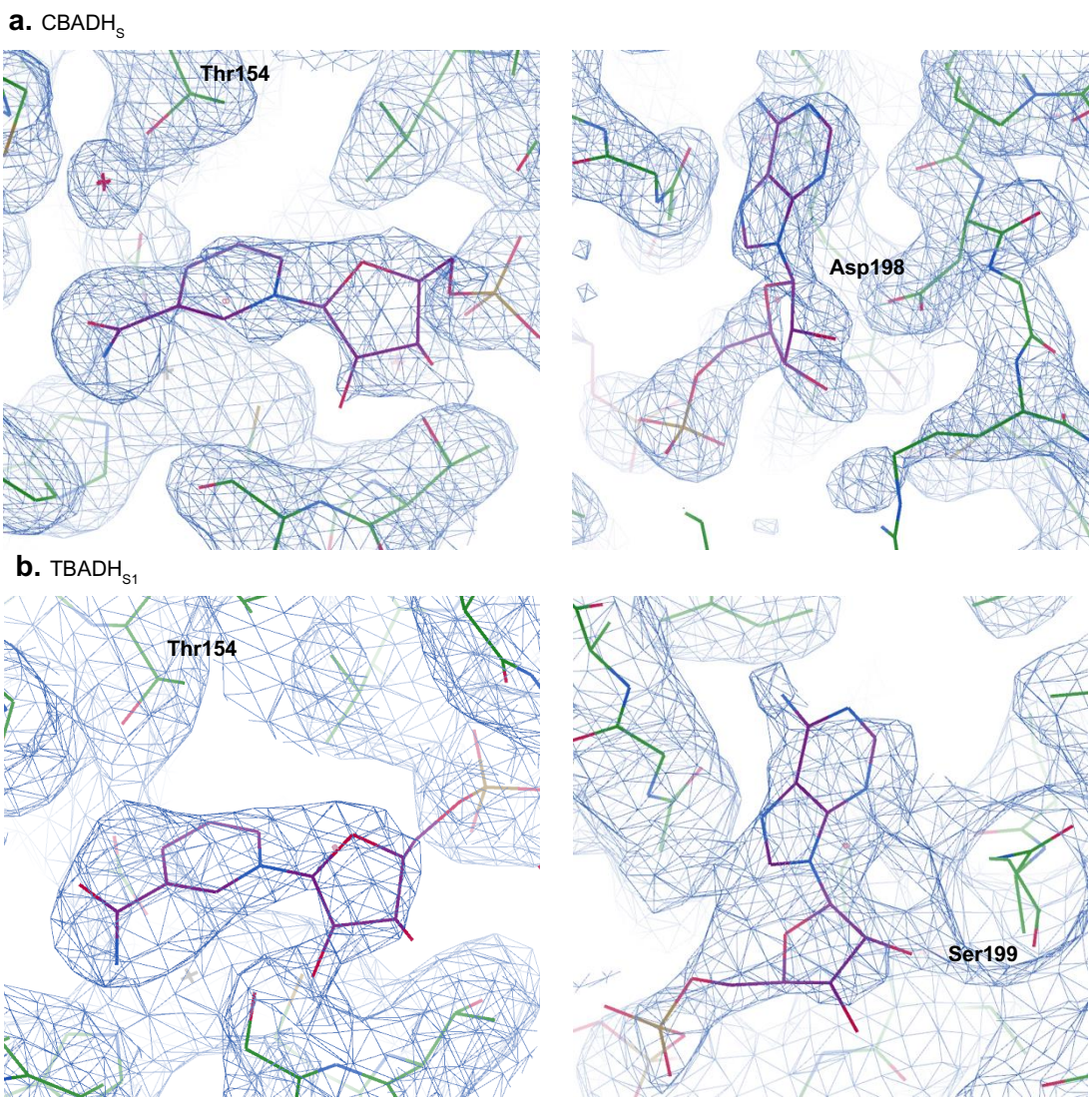


Figure 4.19. Density for the NAD$^+$ cofactor in the crystallographic maps of CBADH$_S$ and TBADH$_{S1}$. Left panels display the density for the nicotinamide part of the cofactor, while right panels show the density for the adenine part. The density for the cofactor was stronger in CBADH$_S$ (**a**) than in TBADH$_{S1}$ (**b**), due to partial occupancy of the cofactor in the latter. $2F_o$-F_c maps are shown at sigma level of 1 for CBADH$_S$ and 0.8 for TBADH$_{S1}$.

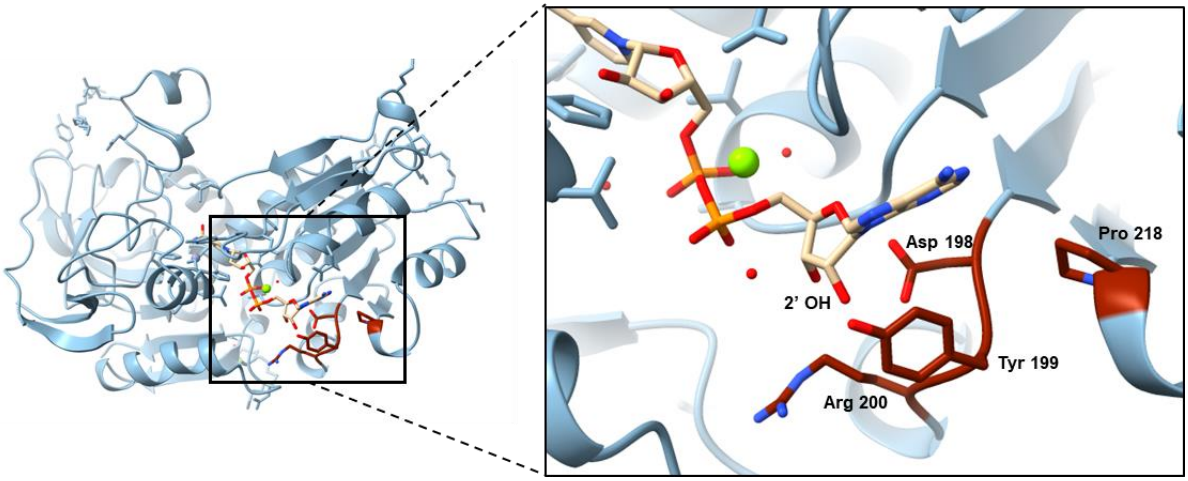


Figure 4.20. Cofactor binding pocket of CBADH$_S$ with NAD$^+$ bound. The substitutions identified in CBADH$_S$ allow for the binding of NAD, but would prevent the binding of NADP by steric impediments and electrostatic repulsion with the side chains of Asp198 and Tyr199. Additionally, the stacked-ring interaction of the adenine moiety with Tyr218 observed in wild-type CBADH cannot be established anymore due to the Tyr218Pro substitution, possibly enabling a more flexible binding of the cofactor.

Another common feature of both CBADH$_S$ and TBADH$_{S1}$ is the substitution of Tyr218. The side chain of Tyr218 is known to undergo a 120º rotation in the wild-type enzymes to allow stacking to the adenine moiety of NADP and the formation of a hydrogen bond with the 2' phosphate through its hydroxyl group (Korkhin *et al.*, 1998). As shown by the CBADH$_S$ and TBADH$_{S1}$ structures, the side chains of the substituted residues are not proximal enough to the 2' phosphate to interact with it, and would only be able to form, if any, hydrophobic interactions with the adenine moiety (**Figures 4.20 and 4.21**).

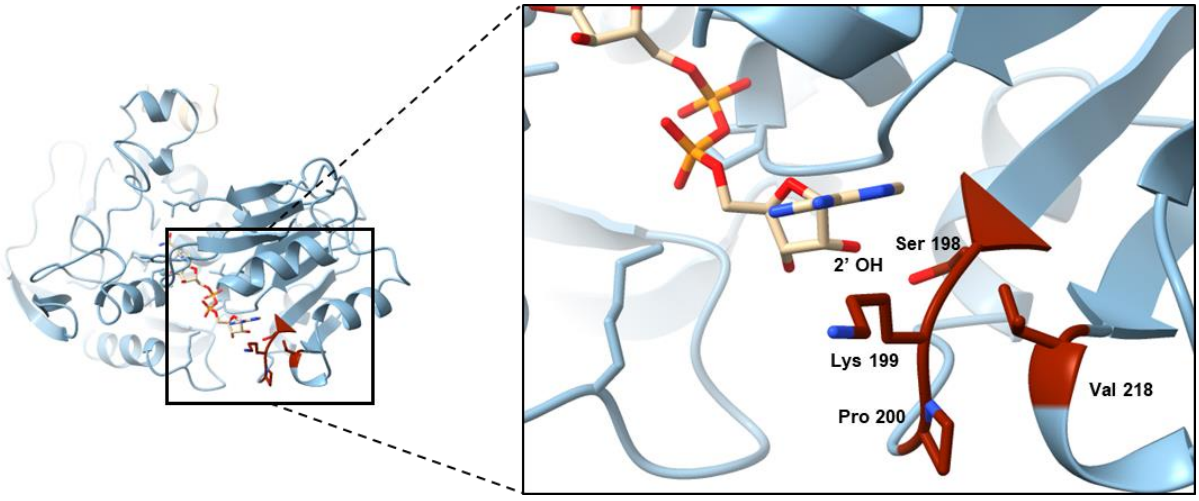


Figure 4.21. Cofactor binding pocket of TBADH$_{S1}$ with NAD$^+$ bound. While TBADH$_{S1}$ can accommodate NAD in its cofactor binding pocket, the binding of NADP would be prevented by steric impediments caused by the side chains of Ser198 and Lys199. As in CBADH$_S$, the substitution of Tyr218 prevents the formation of a stacked-ring interaction with the adenine ring of the cofactor, possibly enabling a more flexible binding of NAD.

The lack of crystals for $TBADH_{S2}$ prevented us from evaluating the impact of the partial duplication of the Rossmann fold in the overall structure and in the cofactor-binding pocket to determine the reason of the observed switch of cofactor preference and why activity with NADP was not lost. Nevertheless, it is possible that the presence of the additional residues, protruding directly from the binding pocket of the 2' phosphate, makes the cofactor-binding site more labile, allowing for a more flexible placing of the cofactor. This would result in the enzyme acquiring the ability to accept NAD, while still being able to accommodate NADP.

4.3. Conclusions

NAD-dependent variants of NADP-dependent oxidoreductases with optimal activity (within the explored sequence space) towards a substrate of choice have been successfully identified with the novel selection system. Such variants are of great industrial interest, due to the reduced cost of NAD cofactors over NADP and the higher efficiency of *in vitro* regeneration systems for NAD. Previous attempts at cofactor specificity reversal have frequently resulted in a severe loss of activity, proving to be particularly challenging when attempting to convert an NADP-dependent alcohol dehydrogenase into an NAD-dependent enzyme (Chánique and Parra, 2018). With a single round of selection, the cofactor preference of CBADH and TBADH was reversed, resulting in the variant of an NAD(P)-dependent dehydrogenase with altered cofactor preference with the second-best relative catalytic efficiency ever reported, and best for the NADP-to-NAD direction of reversal. Furthermore, the ability of the system to enrich unexpected variants not explicitly generated with the chosen library design (an indicator of the strength of the system to perform deep and exhaustive searches in the sequence space) has been demonstrated.

These results provide the first demonstration of the effectiveness of the selection system, which provides a powerful selection tool without the requirement for a technically demanding workflow (**Figure 4.22**). The selection system can also be readily applied to the improvement of kinetic properties of NAD-dependent enzymes, since an enhanced activity towards the substrate provided in the culture medium would result in a selective advantage under anaerobic conditions due to the increased rate of NAD^+ regeneration. Two variants of the system have been developed: one allowing for the selection of biomolecules able to regenerate either NAD^+ or $NADP^+$ (the AL strain), and another one strictly requiring regeneration of NAD^+ by the selected biomolecules, although $NADP^+$ may also be regenerated (the ALPS strain). The AL strain can therefore also be applied to the selection of improved NADP-dependent oxidoreductases, provided that the library of variants is carefully tailored to maintain the structure of the binding pocket of the 2' phosphate of NADP. Conversely, if activity with NAD is the only sought property, the usage of the ALPS strain might be preferable, since residual NADP activity would not contribute at all towards the restoration of anaerobic growth, and therefore it would not be selected.

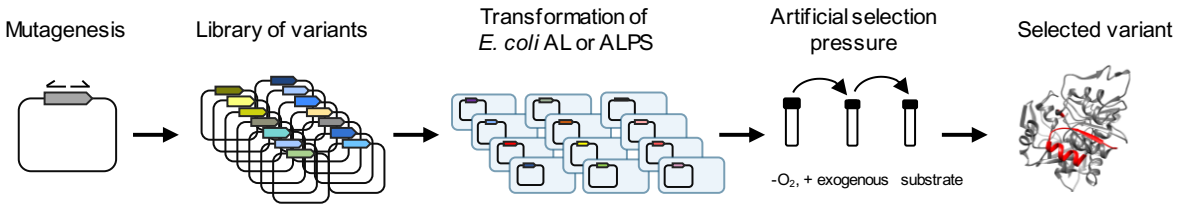


Figure 4.22. General workflow for the selection of variants with the redox rescue selection system. After obtaining a library of variants by mutagenesis of a gene of interest, AL or ALPS cells are transformed with the library of variants. Transformed cells are then grown anaerobically in minimal medium supplemented with a carbon source and an oxidised substrate that is expected to be reduced by a variant of the library. Sub-cultures can be performed to further enrich the selected variant, which can then by identified by isolating and sequencing plasmid DNA.

4.4. Key outcomes

- A variant of CBADH (a strictly NADP-dependent secondary alcohol dehydrogenase) with reversed cofactor specificity has been selected with the AL strain.
- The AL strain allows selection of NAD or NADP-dependent oxidoreductases thanks to the activity of transhydrogenases.
- Transhydrogenases are not essential for aerobic growth of *E. coli* in minimal medium with glucose, and both of them can catalyse the transhydrogenation reaction in the direction of generation of NAD^+ and NADPH.
- Another mutant *E. coli* strain (the ALPS strain) enabling a selection strictly requiring oxidoreductases able to regenerate NAD^+ (although $NADP^+$ might also be regenerated) has been engineered by knocking out transhydrogenase genes.
- One of the variants selected with the ALPS strain, $TBADH_{S1}$, displayed the highest relative catalytic efficiency for any case of reversal of cofactor preference from NADP to NAD.
- Another variant selected with the ALPS strain, $TBADH_{S2}$, contained an unexpected sequence duplication which had not been deliberately introduced during the generation of the library, and was able to use both NAD and NADP as cofactors.

5. Engineering an imine reductase with altered cofactor specificity

5.1. Introduction

Chiral amines are characterized by the presence of a stereogenic carbon adjacent to the nitrogen atom of an amine group. They are fundamental molecules in the pharmaceutical and agrochemical industries, where respectively 40% and 20% of the active ingredients are estimated to contain a chiral amine moiety (Schrittwieser *et al.*, 2015). Some examples are presented in **Figure 5.1**. Additionally, chiral amines are also widely used as resolving agents to solve racemic mixtures of organic acids through diastereomeric salt crystallization, where the addition of the chiral amine adds a new chiral centre that converts the original enantiomers to be separated into diastereomers, which have different physical properties and can be separated through crystallization.

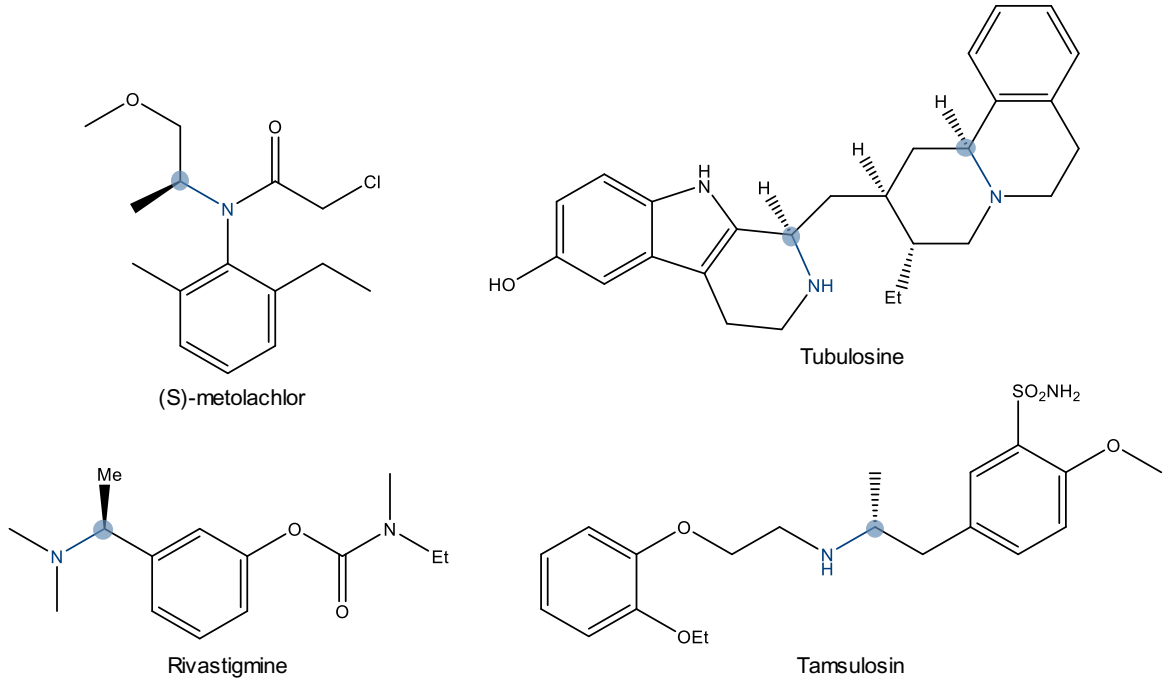


Figure 5.1. Examples of chiral amines of industrial interest. Asymmetric carbon atoms bound to amine groups are highlighted in blue. (S)-metolachlor is a herbicide, tubulosine is an anti-tumour, rivastigmine is an acetylcholinesterase inhibitor (used to treat Alzheimer's disease and Parkinson's disease) and tamsulosin is used to treat benign prostatic hyperplasia and facilitate the passage of kidney stones.

Thus, chiral amines have become an attractive target for asymmetric synthesis, specially since the creation by the FDA in 1997 of a fast-track registration procedure allowing companies to benefit from shortened registration times for drugs provided that an enantiomerically pure product, and not a racemic mixture, is marketed (Ghislieri and Turner, 2014). One of the

devised synthetic strategies for the synthesis of chiral amines is the asymmetric hydrogenation of the corresponding pro chiral imines, which are relatively readily available. Such a reaction can be performed with traditional chemical catalysts, but require first an activation of the imine with groups such as phenyl, benzyl, phosphinoyl or sulfonyl. Then, the imine group must be reduced, after which the activating group must be cleaved. The difficulty of this process is increased in the case of acyclic imines, which require the use of transition metals to undergo asymmetric hydrogenation. Most typically, this can be achieved with the use of iridium-based catalysts, which are expensive and limit the industrial applicability of this approach (Nugent and El-Shazly, 2010).

An alternative approach for the asymmetric reduction of imines is provided by the use of imine reductases (IREDs), enzymes able to catalyse the one-step reduction of imines coupled to the oxidation of NADPH. Additionally, IREDs are also known to be able to catalyse the reductive amination of ketones, where a carbonyl group is converted into an amine group via an imine intermediate which is reduced with NADPH (Bornadel *et al.*, 2019) (**Figure 5.2**). For this reason, they are also sometimes referred to as reductive aminases. Reductive amination is another widely used approach for the synthesis of chiral amines, which can be preferable when the necessary pro chiral imine cannot be easily achieved. The ability to catalyse both asymmetric imine reduction, as well as reductive aminations, enables imine reductases to produce not only primary amines, like most other biocatalytic approaches for the synthesis of chiral amines, but also secondary (Aleku *et al.*, 2017) and tertiary (Matzel *et al.*, 2017) amines. The catalytic mechanism of IREDs is still poorly understood, although several have been described to probably proceed through an acid-base catalytic mechanism where a residue with a protic side chain (aspartate, lysine or tyrosine) is proximal to the area of interaction between the substrate and the cofactor, and acts as a proton donor to the nascent amine (Rodríguez-Mata *et al.*, 2013; Man *et al.*, 2015).

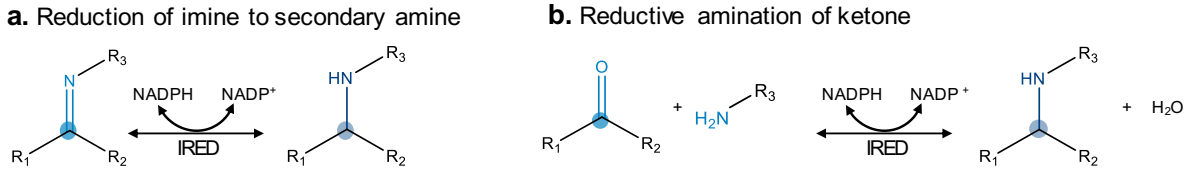


Figure 5.2. Reactions catalysed by imine reductases. a. The primary reaction catalysed by imine reductases is the reduction of imines to the corresponding amines coupled to the oxidation of NADPH, and the inverse reaction. **b**. They have also been found to catalyse the reductive amination of ketones, where a keto group is replaced by an amine at the expense of NADPH oxidation.

Since a naturally occurring IRED with optimal properties for a specific purpose is often not available, there is a great interest in engineering IREDs to make them more suitable for industrial applications. Several properties have been targeted, including substrate specificity and cofactor preference. While no natural NADH-dependent imine reductases are known, in a previous study Gand *et al*. found that the Lys40Ala substitution modestly increased the ability of *Streptomyces* GF3587 (R)-selective imine reductase (Sgf3587IRED, Uniprot accession number: M4ZRJ3) to reduce 2-methyl-1-pyrroline to 2-methylpyrrolidine (**Figure 5.3**) with NAD^+ instead of $NADP^+$ (Gand *et al.*, 2016). More recently, Borlinghaus and Nestl engineered *Myxococcus stipitatus* (R)-selective imine reductase (MsIRED, Uniprot accession number: L7U9F5), which also catalyses the reduction of 2-methyl-1-pyrroline (**Figure 5.3**) (Borlinghaus and Nestl, 2018). By means of sequential rounds of mutagenesis and well-plate screening, they managed to obtain variants able to accept NAD as the cofactor, with the best variant containing six residue substitutions.

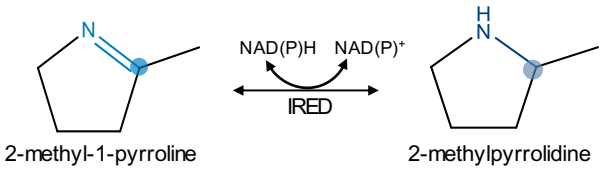


Figure 5.3. Reaction catalysed by *Streptomyces* GF3587 and *Myxococcus stipitatus* imine reductases. Both enzymes are able to use NADPH to reduce 2-methyl-1-pyrroline to the R isomer of 2-methylpyrrolidine. The variant of *Streptomyces* GF3587 imine reductase designed by Gand *et al.* had a modest activity with NADH.

In order to demonstrate the ability of the redox rescue selection system to be applied to oxidoreductases other than alcohol dehydrogenases, I decided to attempt to obtain an optimized NAD-dependent MsIRED variant. Furthermore, by directly comparing my results to those obtained in an established case of successful identification by screening, I aimed to demonstrate the advantages provided by the redox rescue selection system over said approaches.

5.2. Results and discussion

5.2.1. Selection of a novel NAD-dependent MsIRED variant

Due to the lack of a crystal structure of MsIRED, a homology model of MsIRED was constructred based on the structure of the related R-specific IRED from *Streptomyces kanamyceticus* (PDB code 3ZHB). The homology model was generated with SWISS-MODEL, and the NADPH cofactor was placed in the same position as in the template structure. The

structure prediction was then combined with the CSR-SALAD software to guide the generation of a library of MsIRED variants with the goal of obtaining an NAD-dependent variant.

CSR-SALAD predicted residues Asn32, Arg33, Thr34 and Lys37 to be critical for the determination of cofactor specificity. Inspection of the homology model revealed that all four residues constituted part of the binding pocket for the 2' phosphate of NADPH, establishing charged interactions or hydrogen bonds with it through their sidechains (**Figure 5.4**). Thus, a library of variants was generated by saturation mutagenesis of all four residues (**Table 5.1**). These residues were among those that were mutated by Borlinghaus and Nestl to achieve their best performing NAD-dependent MsIRED variants (Borlinghaus and Nestl, 2018). Furthermore, the Lys37 residue is equivalent to the Lys40 residue which was mutated by Gand *et al*. to increase the NAD-dependent activity of Sgf3587IRED (Gand *et al*., 2016). The library was generated by Golden Gate assembly with primers designed to degenerate the targeted positions with an NNN scheme. The resulting theoretical maximum library size was over 16 million variants at the nucleotide level, or 160,000 different variants at the protein level (without including variants with premature stop codons). Sequencing of random library members confirmed diversity of variants, with only one variant matching the parental sequence (**Table 5.2**). Additionally, plasmids with the best MsIRED variant identified by Borlinghaus and Nestl (**Table 5.1**) and the Lys40Ala variant of Sgf3587IRED described by Gand *et al*. were also constructed. These protein variants served as controls, and are therefore referred to as $MsIRED_{C1}$ and $Sgf3587IRED_{C2}$ respectively.

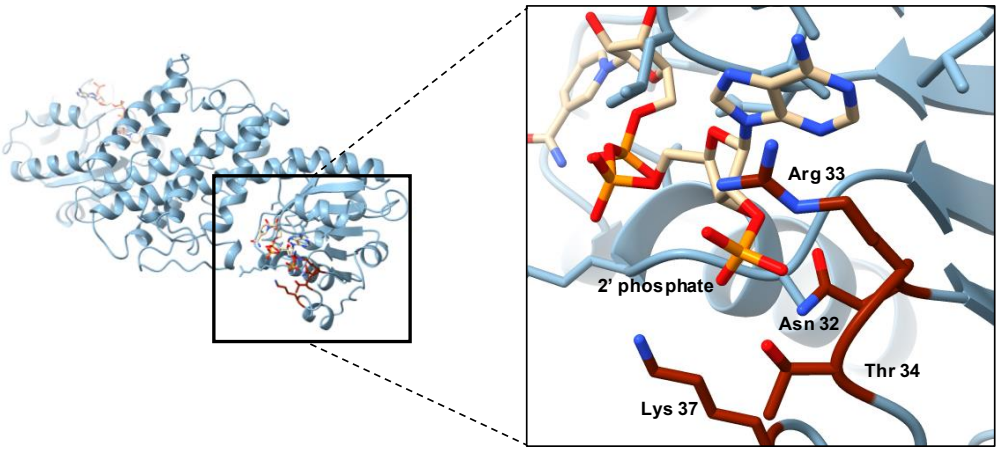


Figure 5.4. Cofactor binding pocket of wild-type MsIRED. A homology model was constructed based on the crystal structure of the related R-specific IRED from *Streptomyces kanamyceticus* (PDB code 3ZHB). The side chains of Asn32, Arg33, Thr34 and Lys37 establish interactions with the 2' phosphate group of NADP. The side chain of Arg33 also establishes a cation-π interaction with the adenine ring.

Table 5.1. MsIRED site-directed and saturation mutagenesis.

Amino acid position	32	33	34	37	67	71
$MsIRED_{WT}$	AAC	CGT	ACC	AAA	ATC	ACC
$MsIRED_{C1}$	GAA	TAT	GAA	CGT	ATT	GTT
$MsIRED_{Lib}$	NNN	NNN	NNN	NNN	ATC	ACC

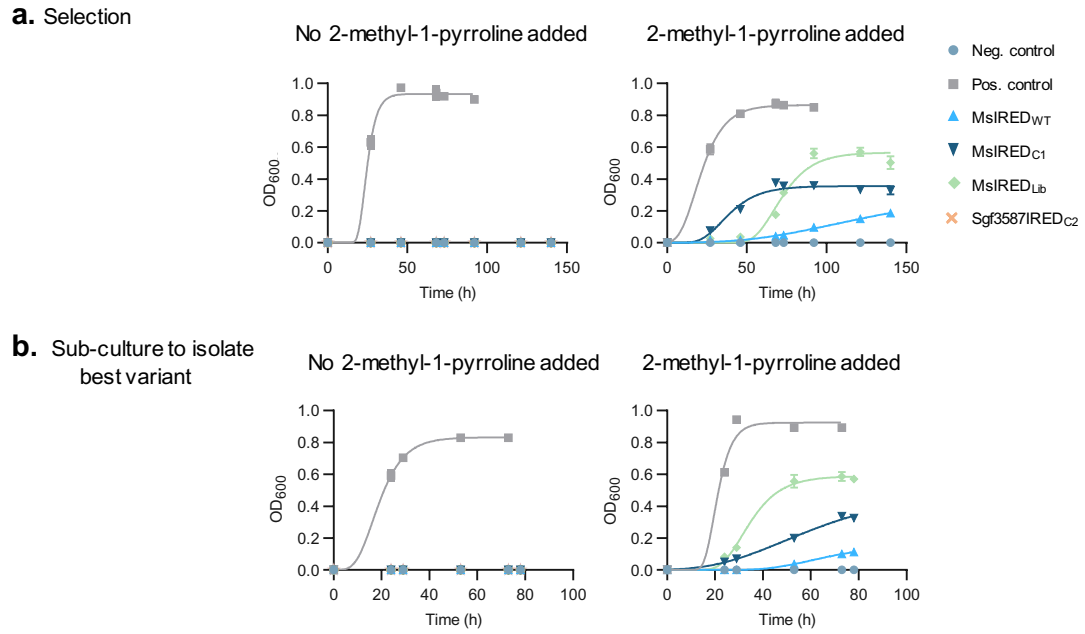
Table 5.2. Quality of the MsIRED library. The variants contained by five random clones were sequenced, with only one variant matching the parental sequence.

Amino acid position	32	33	34	37
$MsIRED_{WT}$	AAC	CGT	ACC	AAA
$MsIRED_{1}$	GGT	GGT	GAT	GGG
$MsIRED_{2}$	AAC	CGT	ACC	AAA
$MsIRED_{3}$	GTG	TCT	GCT	GGG
$MsIRED_{4}$	CTT	CGT	ANT	TTT
$MsIRED_{5}$	CAT	GAT	GCG	CAG

AL cells were transformed with the library of MsIRED variants, as well as with $MsIRED_{C1}$ and $Sgf3587IRED_{C2}$. The transformation efficiency for AL cells transformed with the library was estimated to be approximately three million cells per transformation. Therefore, six independent transformations with the library were performed for growth complementation experiments. Metabolic complementation experiments were carried out by culturing transformed cells anaerobically in M9 medium containing glucose and supplemented with 2-methyl-1-pyrroline.

$Sgf3587IRED_{C2}$ was not able to support anaerobic growth, indicating that the gain in NAD-dependent activity was not sufficient to regenerate NAD^+ efficiently, and the NADP-dependent activity was not large enough to support transhydrogenase-mediated growth recovery. On the other hand, both NADP-dependent wild-type MsIRED and NAD-dependent $MsIRED_{C1}$ were able to restore anaerobic growth in M9 medium containing glucose and supplemented with 2-methyl-1-pyrroline after 68 h and 25 h, reaching a maximum OD_{600} of 0.2 and 0.36 respectively (**Figure 5.5**). Cells complemented with the library were able to grow after 46 h, and reached a maximum OD_{600} of 0.6. When cells which grew in the initial library selection culture were sub-cultured in the same selective conditions (medium containing glucose and 2-methyl-1-

pyrroline), they were able to reach the same maximum OD_{600} with a shorter lag phase of 22 h (**Figure 5.5**). The higher OD_{600} reached by cells complemented with the library can be attributed to a more efficient NAD regeneration allowing a higher glycolytic flux, which indicates a higher NADH-dependent 2-methyl-1-pyrroline reduction activity. The sequence of the MsIRED variants contained in cultures grown with each of the three independent transformations of the library was determined, revealing the selection of the same variant in all cases containing Asn32Glu, Arg33Val, Thr34Arg and Lys37Arg substitutions ($MsIRED_S$) (**Figure 5.5**).



c.

Amino acid position	32	33	34	37	67	71	40*
$MsIRED_{WT}$	Asn (AAC)	Arg (CGT)	Thr (ACC)	Lys (AAA)	Leu (CTG)	Thr (ACC)	
$MsIRED_{C1}$	Glu (GAA)	Tyr (TAT)	Glu (GAA)	Arg (CGT)	Ile (ATT)	Val (GTT)	
$Sgf3587IRED_{C2}$							Ala (GCT)
$MsIRED_S$	**Glu (GAG)**	**Val (GTG)**	**Arg (CGG)**	**Arg (CGG)**	Leu (CTG)	Thr (ACC)	

* Only applies to $Sgf3587IRED_{C2}$. Position 40 of Sgf3587IRED is equivalent to position 37 of MsIRED

Figure 5.5. Selection of an NAD-dependent MsIRED variant. a. AL cells transformed with the library of MsIRED variants were able to grow anaerobically in medium supplemented with 2-methyl-1-pyrroline, but not in medium without it. AL cells transformed with $MsIRED_{C1}$ or wild-type MsIRED were also able to grow in medium supplemented with 2-methyl-1-pyrroline, thanks the NAD-dependent activity of the variant in the first case and the activity of endogenous transhydrogenases in the case of wild-type MsIRED. However, AL cells transformed with $Sgf3587IRED_{C2}$ were unable to grow anaerobically even when the medium was supplemented with 2-methyl-1-pyrroline, indicating that its NAD-dependent

activity was too low to support growth and that the NADP-dependent activity had also been reduced by the introduction of the Lys40Ala mutation. **b.** Sub-cultures into the same conditions of the cells transformed with the library and grown in the presence of 2-methyl-1-pyrroline were performed to further enrich the selected variant before isolating plasmid DNA to sequence it. A similar behavior as for the first culture was observed, with the difference that the sub-culture of cells transformed with the library grew faster thanks to the presence of a larger initial population of cells with the selected variant. **c.** Amino acid residues present at the saturated positions in wild-type MsIRED, MsIRED$_{C1}$ and the selected MsIRED$_S$ variant. The substitution present in Sgf3587IRED$_{C2}$ is also shown. The corresponding DNA codons are shown in brackets. Error bars show the standard deviation for each time point.

5.2.2. Characterization of the selected NAD-dependent MsIRED variant

Plasmid DNA encoding MsIRED$_S$ was isolated and used to perform growth complementation experiments. AL cells transformed with the MsIRED$_S$ variant were able to grow under anaerobic fermentation conditions in M9 medium supplemented with glucose and 2-methyl-1-pyrroline, exhibiting a lag phase of 23 h (**Figure 5.6**). ^{1}H-NMR spectra of the fermentation broth confirmed the presence of 2-methylpyrrolidine, resulting from the reduction of 2-methyl-1-pyrroline (**Table 5.3**).

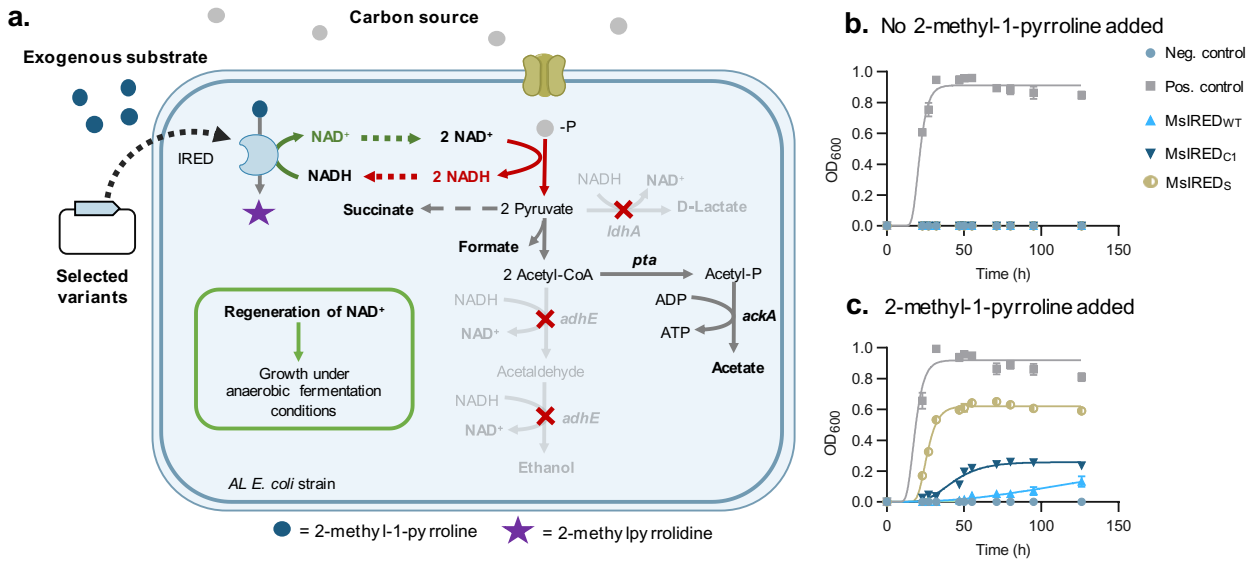


Figure 5.6. Metabolic complementation with the selected NAD-dependent MsIRED variant. a. AL cells transformed with the NAD-dependent variant of MsIRED would be able to grow anaerobically in medium supplemented with 2-methyl-1-pyrroline thanks to the regeneration of NAD$^+$ provided by the MsIRED variant upon reduction of 2-methyl-1-pyrroline to 2-methylpyrrolidine. **b, c**. AL cells transformed with MsIRED$_S$ were able to grow anaerobically in medium supplemented with 2-methyl-1-pyrroline, but not in medium without it. As previously observed, cells transformed with MsIRED$_{C1}$ or wild-type MsIRED were also able to grow anaerobically in medium with 2-methyl-1-pyrroline, although less efficiently than cells transformed with MsIRED$_S$ in both cases. Error bars show the standard deviation for each time point.

Wild-type MsIRED, MsIRED$_{C1}$ and MsIRED$_S$ were expressed in *E. coli* and purified using a hexa-histidine affinity tag in order to characterize the novel MsIRED$_S$ variant and compare it

to wild-type MsIRED and the best previously found NAD-dependent variant ($MsIRED_{C1}$). Successful expression and purification was achieved for the three enzymes (**Figure 5.7**), and purified protein was used for assaying 2-methyl-1-pyrroline reductase activity with NADH and NADPH.

Wild-type MsIRED was able to reduce 2-methyl-1-pyrroline only with NADPH (**Figure A.2a**). On the other hand, $MsIRED_S$ and $MsIRED_{C1}$ were able to use NADH as the cofactor, but no activity with NADPH was detected (**Figures A.2b and A.2c**). In all cases, a decrease in the activity was observed at the highest concentrations of 2-methyl-1-pyrroline, due to substrate inhibition (**Figure 5.7**). Thus, a modified version of the Michaelis-Menten equation accounting for substrate inhibition was fitted to the data:

$$V = \frac{V_{max} \cdot [S]}{K_m + [S] \cdot \frac{1 + [S]}{K_i}}$$

Where V is the initial reaction rate, [S] is the substrate concentration, V_{max} is a constant representing the maximum reaction rate that would be achieved in the absence of substrate inhibition, K_m is a constant representing the concentration at which half of the maximum reaction rate would be achieved in the absence of substrate inhibition and K_i is the dissociation constant for substrate binding, assuming that two substrate molecules can bind the enzyme and that the substrate can bind to the non-catalytic, inhibitory site only after another substrate molecule has bound to the catalytic site.

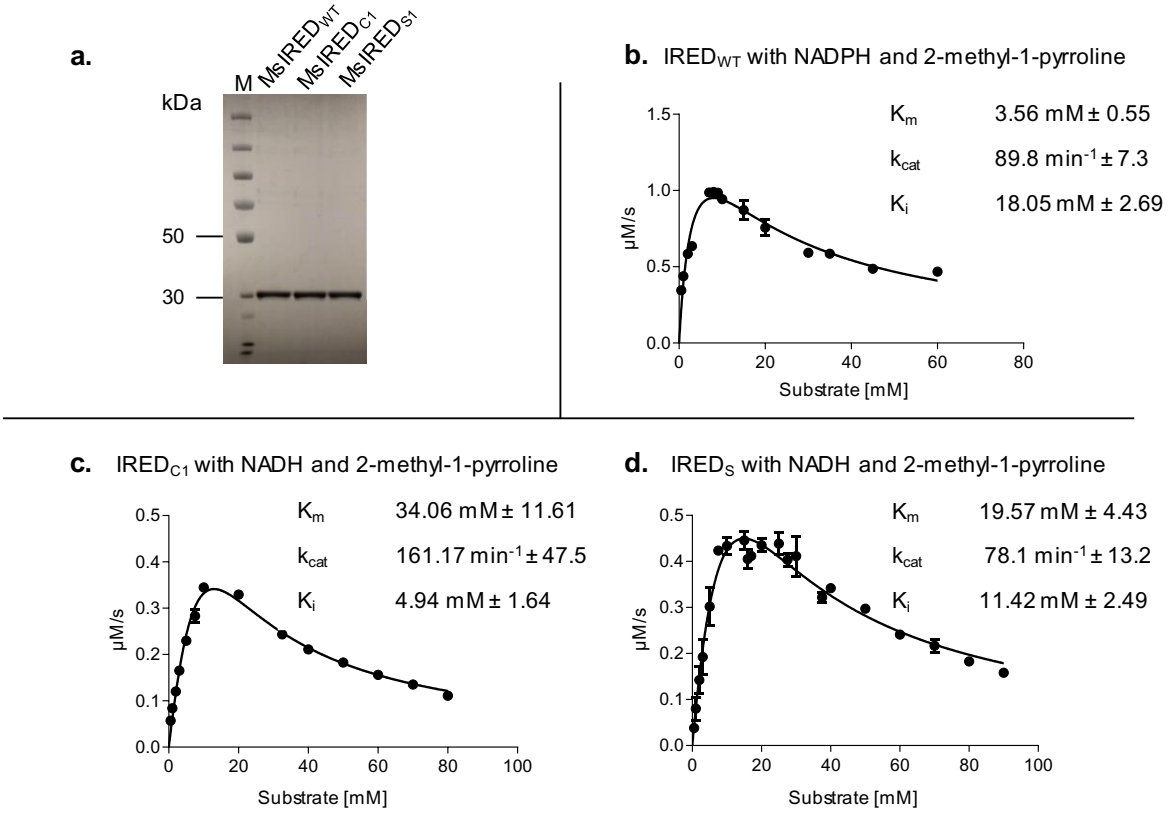


Figure 5.7. Characterization of wild-type MsIRED, MsIRED$_{C1}$ and MsIRED$_S$. a. SDS-PAGE confirming the purity of wild-type MsIRED, MsIRED$_{C1}$ and MsIRED$_S$ purified by nickel-affinity chromatography, all expected to have a molecular weight of 32.3 kDa including the hexa-histidine tags. **b, c, d.** Fitting of a Michaelis-Menten model modified to account for substrate inhibition to the initial reaction rates obtained with NADPH and different 2-methyl-1-pyrroline concentrations for wild-type MsIRED, and with NADH and different 2-methyl-1-pyrroline concentrations for MsIRED$_{C1}$ and MsIRED$_S$. No activity with NADH was detected for wild-type MsIRED, and no activity with NADPH was detected for MsIRED$_{C1}$ and MsIRED$_S$. Error bars show the standard deviation for each concentration.

Wild-type MsIRED catalysed a more efficient reaction with NADPH than either MsIRED$_{C1}$ or MsIRED$_S$ with NADH, displaying the best kinetic parameters for the reduction of 2-methyl-1-pyrroline and the smallest substrate inhibition. Nevertheless, MsIRED$_S$ performed better with NADH than MsIRED$_{C1}$. This is demonstrated by the greater initial rate observed at all tested substrate concentrations, in spite of a lower k_{cat} determined from the V_{max} values obtained after fitting the modified Michaelis-Menten equation. In such a model, V_{max} represents the hypothetical maximum initial reaction rate that could be achieved under saturating substrate concentrations if the enzyme did not display substrate inhibition, and therefore the derived k_{cat} values cannot be used to assess the performance of each enzyme, since at high substrate concentrations the inhibitory effects of the substrate severely reduce the reaction rate. Instead, it is necessary to consider also the strength of the inhibition exercised by the substrate in each variant, reflected in the value of K_i. Indeed, the K_i for MsIRED$_S$ was more than twice as large

as that of MsIRED$_{C1}$, indicating that substrate inhibition is stronger in MsIRED$_{C1}$. Therefore, MsIRED$_{S}$ represents the best NAD-dependent IRED reported to my knowledge.

Taken together, these results highlight the effectiveness of the redox rescue selection system to select variant enzymes where multiple kinetic parameters are enhanced simultaneously, resulting in novel enzymes with optimal activity towards the desired substrate. Furthermore, several rounds of mutagenesis and screening were required to identify the previously-reported variant MsIRED$_{C1}$ (Borlinghaus and Nestl, 2018). In the first three rounds, cofactor preference reversal was sought by targeting residues Asn32, Arg33, Thr34 and Lys37, while the final rounds aimed to recover the loss of activity due to the first round of mutagenesis by saturating residues Leu67, Thr71 and Leu75. With the new approach, both cofactor preference reversal and activity optimization were performed simultaneously in a single round of selection, yielding a variant which performed better than MsIRED$_{C1}$ despite the smaller number of residues for which mutations were explored.

Table 5.3. Quantification of metabolites by ^{1}H-NMR of the fermentation broth of AL cells complemented with wild-type MsIRED, MsIRED$_{C1}$ and MsIRED$_{S}$. 2-methylpyrrolidine was detected in all cases.

Strain	Plasmid	Encoded enzyme	Exogenous substrate	Resulting product	Substrate δ * (ppm)	Product δ * (ppm)	[Substrate] (mM)	[Product] (mM)	[Ethanol] (mM)	[Lactate] (mM)	[Succinate] (mM)	[Acetate] (mM)	[Formate] (mM)
AL	pLS130	MsIRED$_{WT}$	2-methyl-1-pyrroline	2-methylpyrrolidine	2.42 (s,3)	1.38 (d,3)	5.5	6.6	0	0	1.2	4.3	2.1
AL	pLS131	MsIRED$_{C1}$	2-methyl-1-pyrroline	2-methylpyrrolidine	2.42 (s,3)	1.38 (d,3)	3.45	9.2	0	0	1.7	10.7	7.9
AL	pLS133s	MsIRED$_{S}$	2-methyl-1-pyrroline	2-methylpyrrolidine	2.42 (s,3)	1.38 (d,3)	0.8	13.4	0	0	2	15.9	12.8

* δ of characteristic signal of substrate or product (ppm). The multiplicity of each signal (s – singlet, d – doublet, t – triplet, q – quartet, m – multiplet) and the number of contributing protons is shown between brackets.

5.2.3. Structural modelling of MsIRED variants

Structural models of $MsIRED_S$ and $MsIRED_{C1}$ with NAD and NADP were constructed by performing the appropriate residue mutations and placing the different cofactors with ChimeraX to gain insights into the structural basis of the reversal of cofactor preference in MsIRED.

Both variants contained the Asn32Glu substitution. The side chain of the resulting Glu32 was modelled to be in close proximity of the position of the 2' phosphate if an NADP cofactor was bound (**Figure 5.8**). Such positioning would lead to an electrostatic repulsion between the negatively charged phosphate side chain of Glu32 and the 2' phosphate of the cofactor, preventing binding of NADP. The Lys37Arg mutation was also common to $MsIRED_{C1}$ and $MsIRED_S$. I hypothesize that the long and positively charged side chain is able to interact with one of the phosphates of the phosphodiester bond of NAD, allowing for a stabilizing effect of NAD binding.

However, $MsIRED_{C1}$ and $MsIRED_S$ differed in the substitutions found in positions 33 and 34. In $MsIRED_{C1}$, Arg33 was substituted by a tyrosine residue, which is potentially able to form a tilted stacked-ring interaction with the adenine moiety of the cofactor (**Figure 5.8**). On the other hand, a valine residue was found in the same position in $MsIRED_S$. Due to its smaller side chain, it is possible that the Arg33Val substitution offers a more flexible cofactor binding site (**Figure 5.8**), which could compensate for the distortion of the cofactor binding site induced by the rest of the substitutions and lead to an increased activity with NAD when compared to $MsIRED_{C1}$.

Regarding position 34, $MsIRED_{C1}$ contained a Thr34Glu substitution, which contributes to the reversal of cofactor preference by causing additional electrostatic repulsion with the 2' phosphate (**Figure 5.8**). Interestingly, $MsIRED_S$ contained a Thr34Arg substitution. It is possible that, in spite of having a positively charged residue at position 34, the Arg residue is unable to contribute to an increased affinity towards NADP. Instead, due to the large size of its side chain, it would lead to a clash with the 2' phosphate of NADP upon its binding in a regular manner (**Figure 5.8**). Thus, the Thr34Arg substitutions prevents the binding of NADP through steric impediments.

Overall, the switched preference of cofactor of MsIRED seems to be guided by a combination of both alteration of the electrostatic properties of the cofactor binding pocket around the 2' phosphate of NADP and the introduction of bulky side chains which sterically prevent correct binding of NADP to the enzyme. Such a combination of mutations is not obvious, and it is unlikely it could have been identified purely by rational design. Furthermore, subtle changes in the cofactor binding pocket can lead to different activity and kinetic parameters in the final

variants, highlighting the advantages of the presented selection method over screening methods. In such methods, a deep search of the mutagenesis space cannot be easily performed, making it difficult to find the optimal combinations of substitutions, which often have cooperative effects towards a maximized activity.

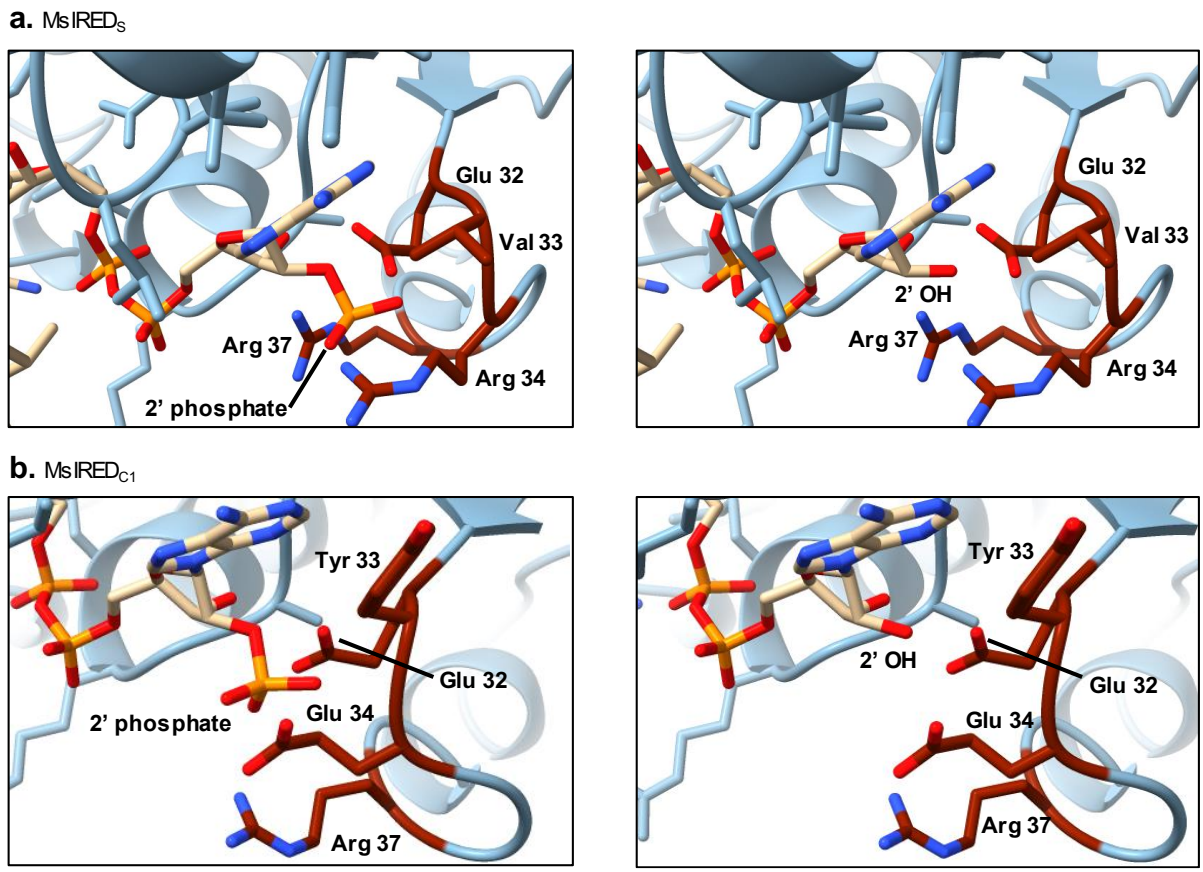


Figure 5.8. Structural basis for the inability of MsIRED$_{C1}$ and MsIRED$_S$ to use NADP. Left panels display models with NADP$^+$ bound, while right panels display models with NAD$^+$ bound. Negatively charged glutamate residues are present in both variants next to where the 2' phosphate of NADP would be placed if an NADP cofactor was bound, which prevent its binding by steric impediments and electrostatic repulsions. In MsIRED$_S$, an arginine residue is present close to where the 2' phosphate of NADP would be located. However, in spite of its positive charge, its bulky side chain would clash with the 2' phosphate, preventing the binding of NADP by steric impediments. The smaller side chain of Val33 found in MsIRED$_S$ compared to the Tyr33 present in MsIRED$_{C1}$ could possibly enable a more flexible cofactor binding.

5.3. Conclusions

The applicability of the redox rescue selection system to enzymes able to act on functional groups other than alcohols has been demonstrated by using it to select the best NAD-dependent IRED reported so far. Furthermore, the selection system has proven to be suitable to not only select an NAD-dependent variant of a given enzyme, but, at the same time, to optimize its kinetic parameters. One of the particular features observed in $MsIRED_S$, the MsIRED variant selected with the redox rescue selection system, is a partial relief of the substrate inhibition compared to the best NAD-dependent MsIRED previously reported, $MsIRED_{C1}$ (Borlinghaus and Nestl, 2018). Around 20% of all known enzymes are estimated to display at least one form of substrate inhibition (Reed, Lieb, and Nijhout, 2010). While such a feature offers an additional level of regulation *in vivo*, it can reduce the effectiveness of the enzyme in applications of human interest, and particularly in industrial processes, where high concentrations of substrate are typically used. There is thus a great interest in reducing the extent to which enzymes of interest are affected by substrate inhibition.

Additionally, the ability of the selection system to optimize multiple properties of an enzyme simultaneously enables it to bypass the need for further rounds of screening and mutation, often required when attempting to reverse cofactor specificity due to the practical impossibility of performing a deep, thorough search of the sequence space with conventional methods dependent upon screening. Thus, variants with relatively high NAD-dependent activity can be achieved in a single round of mutagenesis and selection. Alternatively, such variants can provide a better starting point for performing further mutagenesis and selection rounds aimed at improving other properties of the enzyme.

The redox rescue selection system provides a potential platform for the easy identification of novel IREDs with altered substrate specificity able to synthesize specific chiral amines, simply by providing the corresponding imine in the culture medium of AL cells transformed with either a library of naturally occurring IREDs or a library of mutant variants of a specific IRED. The selection system could potentially be used to improve the reductive aminase activity of IREDs too, by providing a ketone and an amine group donor in the culture medium instead of an imine. In combination with its capacity to provide NAD-dependent enzymes with enhanced kinetic parameters, the system holds the potential to facilitate the replacement of traditional approaches for the synthesis of chiral amines with cheaper and more efficient IRED-based biocatalysts.

5.4. Key outcomes

- An NAD-dependent IRED, MsIRED$_S$, has been selected from a library of *Myxococcus stipitatus* IRED variants by using the AL strain.
- MsIRED$_S$ is the best NAD-dependent IRED reported so far.
- MsIRED$_S$ displayed a relief in substrate inhibition compared to MsIRED$_{C1}$, the best NAD-dependent IRED variant previously reported.
- The relief of substrate inhibition observed in MsIRED$_S$ highlights the potential of the selection system for optimizing multiple kinetic parameters simultaneously.
- Selection with the AL strain allowed for a more efficient identification of the optimal variant than previous approaches, while requiring a less complex workflow.

6. Alteration of substrate selectivity and improvement of kinetic parameters of a nitroreductase

6.1. Introduction

Thanks to their ability to reduce nitroaromatic and nitroheterocyclic compounds, nitroreductases have attracted an increasingly large interest for their potential applications in bioremediation of soils contaminated with nitroaromatic pollutants (Spain, 1995; Ramos *et al.* 2005), the industrial scale synthesis of compounds of human interest (Calle *et al.,* 2019) and even in chemotherapy as a means to activate prodrugs targeting specifically malignant cells (Searle *et al.,* 2004; Gungor *et al.,* 2018; Güngör *et al.*, 2019). One of the most frequent types of pollutants amenable to bioremediation with nitroreductases are explosive nitroaromatic compounds, such as 2,4,6-trinitrotoluene (TNT), which have been shown to be mutagenic and carcinogenic compounds in several species, including humans (Honeycutt, Jarvis, and McFarland, 1996; Nyanhongo *et al.,* 2005). Furthermore, such compounds can easily leak through contaminated soils to groundwater reservoirs, leading to serious health and environmental concerns. The use of plants to decontaminate polluted sites, known as phytoremediation, offers a potentially cheap and easily applicable solution with a low environmental impact. Nevertheless, most plants lack the enzymes required for the metabolization of nitroaromatic contaminants. Thus, several groups have developed transgenic plants expressing exogenous nitroreductases which are expected to provide an efficient means for the cleanup of contaminated sites, such as an *Arabidopsis thaliana* carrying the nitroreductase A gene from *Enterobacter cloacae* constructed by You *et al*. (You *et al.*, 2015).

The use of nitroreductases as substitutes of traditional catalysts for the hydrogenation of nitro groups is particularly attractive thanks to their ability to specifically reduce nitro groups in the context of organic molecules including a variety of different functional groups, which is hard to achieve through traditional organic synthesis, often leading to the generation of undesired side products (Hoogenraad *et al.*, 2004).

Regarding the clinical relevance of nitroreductases, it is generally recognized that they offer a vast potential for the treatment of cancer in combination with nitroaromatic prodrugs, particularly through the use of gene-directed enzyme prodrug therapy (GDEPT). In GDEPT, a nitroaromatic prodrug is administered, which must be converted to the active drug by an exogenous enzyme (**Figure 6.1**). A gene encoding said enzyme is delivered specifically to tumour cells, typically by using a tumour-tropic bacterial or viral vector, where it is expressed, allowing for a localized conversion of the prodrug to its active form. Nitroaromatic prodrugs

are activated through the reduction of nitro groups by nitroreductases, which converts them into DNA-damaging metabolites. This confers them the ability of killing also quiescent tumour cells, unlike other drugs, such as nucleotide analogues, which only target actively dividing cells (Williams *et al.*, 2015). Additionally, activated nitroaromatic drugs can often diffuse easily between cells, providing a "bystander effect" essential for the success of GDEPT therapies. The most widely studied nitroaromatic prodrug for GDEPT is CB1954, discovered in the late 1960s (Khan and Ross, 1969). CB1954 must be activated by converting it to hydroxylamine derivatives, which undergo further spontaneous reactions to become potent DNA alkylating and crosslinking agents. The *E. coli* NfsB nitroreductase was found to reduce CB1954 to both its 2-hydroxylamine and 4-hydroxylamine derivatives, and has been explored as an enzyme candidate for GDEPT in combination with CB1954. However, the low tolerance of humans to the prodrug due to the activity of endogenous liver nitroreductases, together with a low efficiency of CB1954 reduction at low concentrations by *E. coli* NfsB (Kestell *et al.*, 2000), led to a severe limitation of the success of the CB1954-NfSB pair in clinical trials (Patel *et al.*, 2009). There is thus a great interest in developing novel nitroreductase variants displaying improved kinetic parameters for the reduction of CB1954, as well as other alternative nitroaromatic prodrugs which could offer some advantages over CB1954, such as reduced toxicity in humans (Heap *et al.*, 2014; Williams *et al.*, 2015).

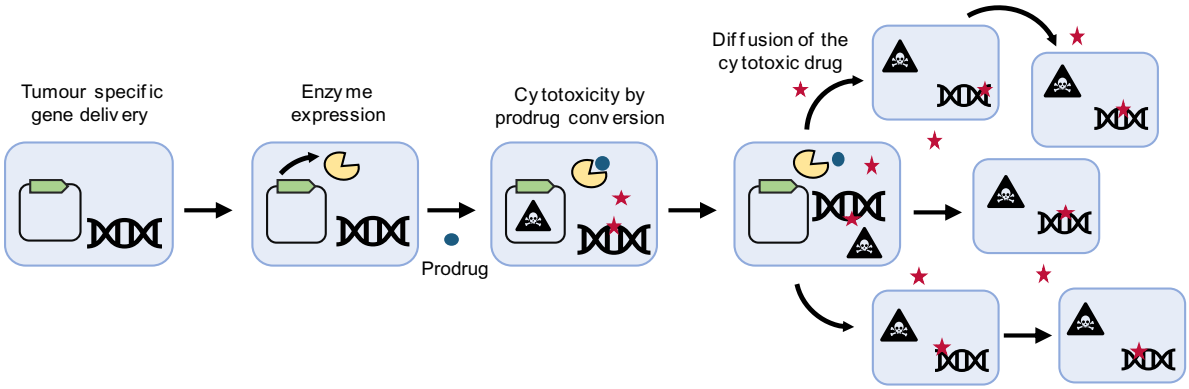


Figure 6.1. Gene-directed enzyme prodrug therapy (GDEPT). In GDEPT, a gene encoding an enzyme is delivered specifically to tumour cells by means of a tumour tropic vector, such as viruses able to replicate in tumour cells but not in normal cells, or obligate anerobic bacteria which can only survive in the hypoxic environment generated in tumours. The gene is then expressed. Higher level of selectivity can be achieved by means of the regulatory elements controlling the expression of the gene, such as tissue-specific promoters. At the same time, a prodrug (ideally non-toxic) is systemically administered. In tumour cells, the prodrug is converted to a cytotoxic drug by the activity of the exogenous enzyme, leading to their death. The activated form of the drug should be able to diffuse to nearby cells (bystander effect) in order to maximize its therapeutic effect. (Figure adapted from Williams *et al.*, 2015)

Nitroreductases are flavoproteins which transfer electrons from NAD(P)H to a wide range of nitroaromatic compounds via a flavin prosthetic group, which can be either flavin

mononucleotide (FMN) or flavin adenine dinucleotide (FAD). Two types of nitroreductases are known, depending on the mechanism employed for the reduction of nitro groups. Type I nitroreductases, or oxygen-insensitive nitroreductases, use FMN as their prosthetic group and catalyse up to three sequential transfers of electron pairs from NAD(P)H to the nitroaromatic compound, generating first a nitroso derivative, then a hydroxylamine, and finally an amine (**Figure 6.2**). The nitroso intermediate is usually not detected, since it is very quickly reduced to a hydroxylamine. Not all type I nitroreductases display the ability to further reduce the hydroxylamine to amine. While it is unclear why some nitroreductases are able to perform this third reduction step, it is thought to depend on factors such as the redox potential of the FMN prosthetic group (influenced by the protein environment) and the geometry of the active site (Miller *et al.*, 2018).

On the other hand, type II nitroreductases can bind either FMN or FAD as a prosthetic group and perform sequential transfers of a single electron to the nitroaromatic substrate. The first electron transfer reduces the nitro group to a nitro anion radical, which reacts readily with molecular oxygen to yield the original nitroaromatic compound and superoxide anion, leading to a futile cycle and preventing the formation of reduced products (**Figure 6.2**). For this reason, type II nitroreductases are also known as oxygen-sensitive nitroreductases (Roldán *et al.*, 2008).

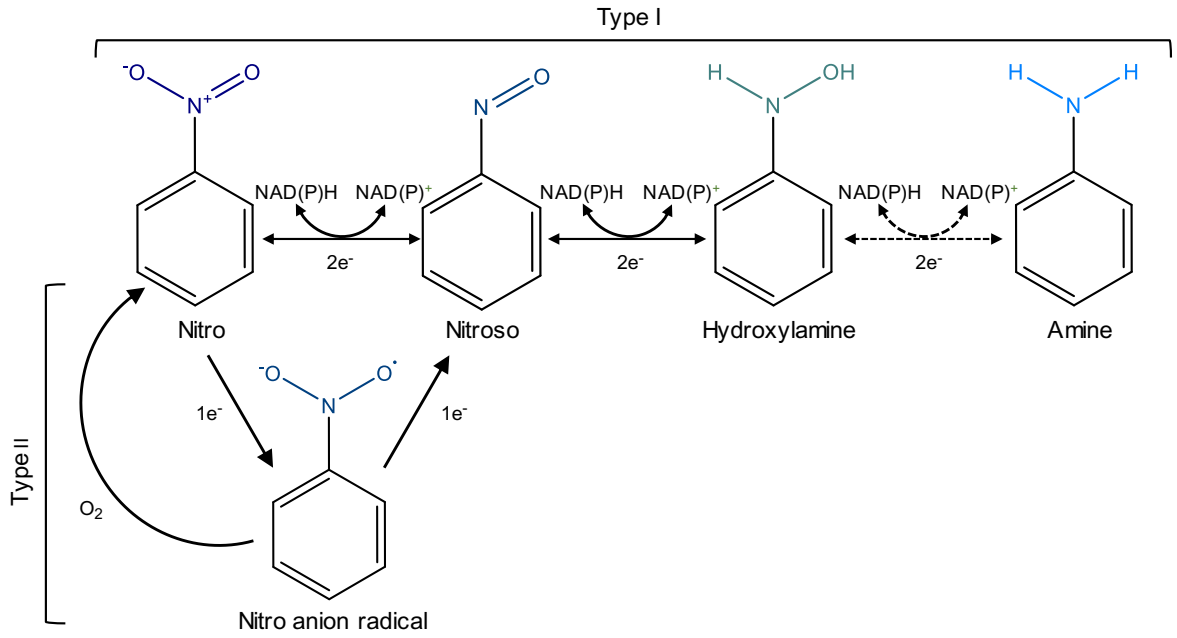


Figure 6.2. Reactions catalysed by type I and type II nitroreductases. Type I nitroreductases are able to catalyse sequential reductions of a nitroaromatic to a nitroso derivative and then to a hydroxylamine, transferring the two electrons required for each reduction from NAD(P)H. Some type I nitroreductases have also been found to further reduce the hydroxylamine to an aromatic amine, but this is not always the case. Type II nitroreductases are able to reduce nitroaromatic compounds to a nitroso derivative by means of two sequential one-electron transfers, generating an intermediate with a nitro anion radical in the process. The nitro anion radical intermediate is readily oxidised to the original

nitroaromatic by O_2, explaining the oxygen-sensitive nature of type II nitroreductases. Nitrobenzene is shown as the initial substrate for simplicity, but nitroreductases are known to act on a large range of nitroaromatic compounds.

Type I nitroreductases, which are found mostly in bacteria, have been more extensively studied biochemically and structurally than their type II counterparts. In type I nitroreductases, the reaction proceeds through a ping-pong kinetic mechanism. First, NAD(P)H binds to the enzyme and transfers two electrons to the FMN prosthetic group. Upon release of the oxidised $NAD(P)^+$ cofactor, the oxidised substrate binds to the enzyme, and two electrons are transferred from the reduced form of FMN. Due to the thermodynamics of the FMN group in the environment of type I nitroreductases, its semiquinone radical form is strongly destabilized, which prevents such nitroreductases from performing single electron transfers from reduced FMN to the substrate (Koder *et al.*, 2002).

One of the best characterized type I nitroreductases is NfsB from *Enterobacter cloacae* (referred to here as EntNfsB), which is known to reduce a broad range of nitroaromatic compounds to the corresponding hydroxylamines, including the explosive TNT, as well as several others such as 2,4-dinitrotoluene and 4-nitrobenzoic acid (4-NBA). The structures of the enzyme bound to a nicotinamide cofactor analogue and 4-NBA was solved by Pitsawong *et al*. (PDB codes 5J8D and 5J8G) (Pitsawong *et al.*, 2017), providing detailed information about the placement of the substrate in the active site. The enzyme acts as a homodimer, where the FMN and substrate binding pocket is located at the interface between two monomers (**Figure 6.3**). In this chapter, the aim was to select novel variants of EntNfsB with altered substrate selectivity. I aimed to obtain variants able to catalyse the reduction of 2-nitrobenzoic acid (2-NBA), which were identified as not being accepted by the native enzyme. Additionally, I also sought to select a variant able to act more efficiently on 4-nitrobenzyl alcohol, whose reduction is catalysed by the wild-type enzyme less efficiently than for 4-nitrobenzoic acid due to the different *para* substituent, in accordance with the previously described effect of substituents on the reduction rate of nitroaromatics depending on their electron-donating or electron-accepting nature (Miller *et al.*, 2018).

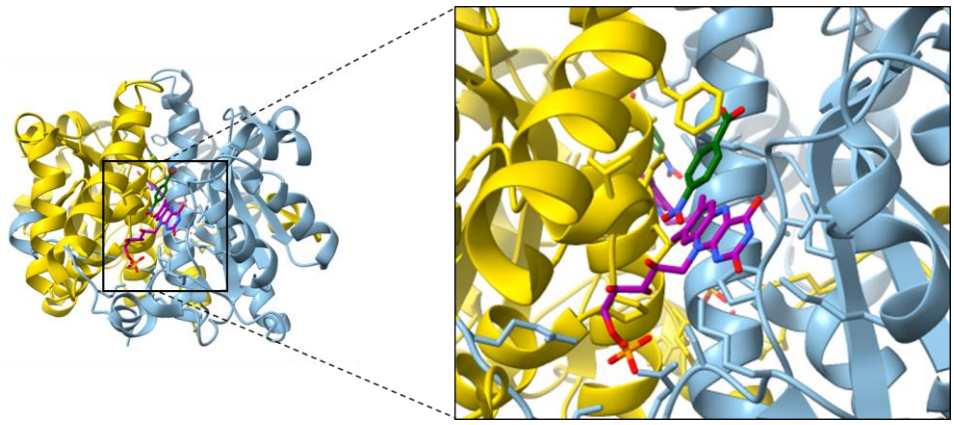


Figure 6.3. Overview of the structure of *Enterobacter cloacae* NfsB. The structure of EntNfsB bound to 4-NBA (PDB code 5J8G) shows that the enzyme acts as a homodimer where the FMN (purple) and substrate (dark green) binding pocket is located at the interface between the two monomers.

6.2. Results and discussion

6.2.1. Selection of EntNfsB variants with altered substrate selectivity

In order to determine the target substrates towards which to seek improved activity, the activity of native EntNfsB with several nitroaromatic substrates was characterized (**Figure 6.4**). The wild-type enzyme was successfully expressed and purified in *E. coli*, and the activity of the enzyme was tested with 4-NBA, 2-NBA (a positional isomer of 4-NBA) and 4-nitrobenzyl alcohol (**Figure 6.5**), which differs from 4-NBA in the *para* substituent. The presence of an electron-withdrawing *para* substituent in derivatives of nitrobenzene has been shown to facilitate their reduction by nitroreductases (McCormick *et al*., 1976). Thus, a decrease in the efficiency of catalysis with 4-nitrobenzyl alcohol compared to 4-NBA was expected.

Enzymatic activity assays revealed that EntNfsB was able to reduce 4-NBA and 4-nitrobenzyl alcohol, but not 2-NBA (**Figures 6.5 and A.3**). Furthermore, the ability to accept 4-nitrobenzyl alcohol as a substrate was much lower than for 4-NBA, as evidenced by the fact that saturation was not reached even with the highest substrate concentrations tested, which prevented the calculation of kinetic parameters (**Figure 6.5**). The difference in the efficiency of catalysis for the reduction of 4-NBA and 4-nitrobenzyl alcohol is in accordance with the presence of a *para* substituent with less electron-withdrawing power (Miller *et al.*, 2018; Hansch, Leo, and Taft, 1991; Girault and Dana, 1977). Then, the ability of wild-type EntNfsB to support anaerobic growth was tested by transforming AL cells with it and growing them in anaerobic medium supplemented with 4-NBA. Growth was observed after 133 hours, reaching a maximum OD_{600} of 0.09. I hypothesized that the lower maximum OD_{600} and considerably longer lag phase observed in metabolic complementation with EntNfsB could be possibly due to the toxicity of 4-NBA, which is known to inhibit partially *E. coli* growth (Davis, 1951). The inhibitory effect of

4-NBA was confirmed with aerobic growth curves of wild-type *E. coli* cells in the presence of 4-NBA. 2-NBA and 4-nitrobenzyl alcohol were also found to have a detrimental effect on *E. coli* growth, although to a lesser extent than 4-NBA (**Figure 6.6**).

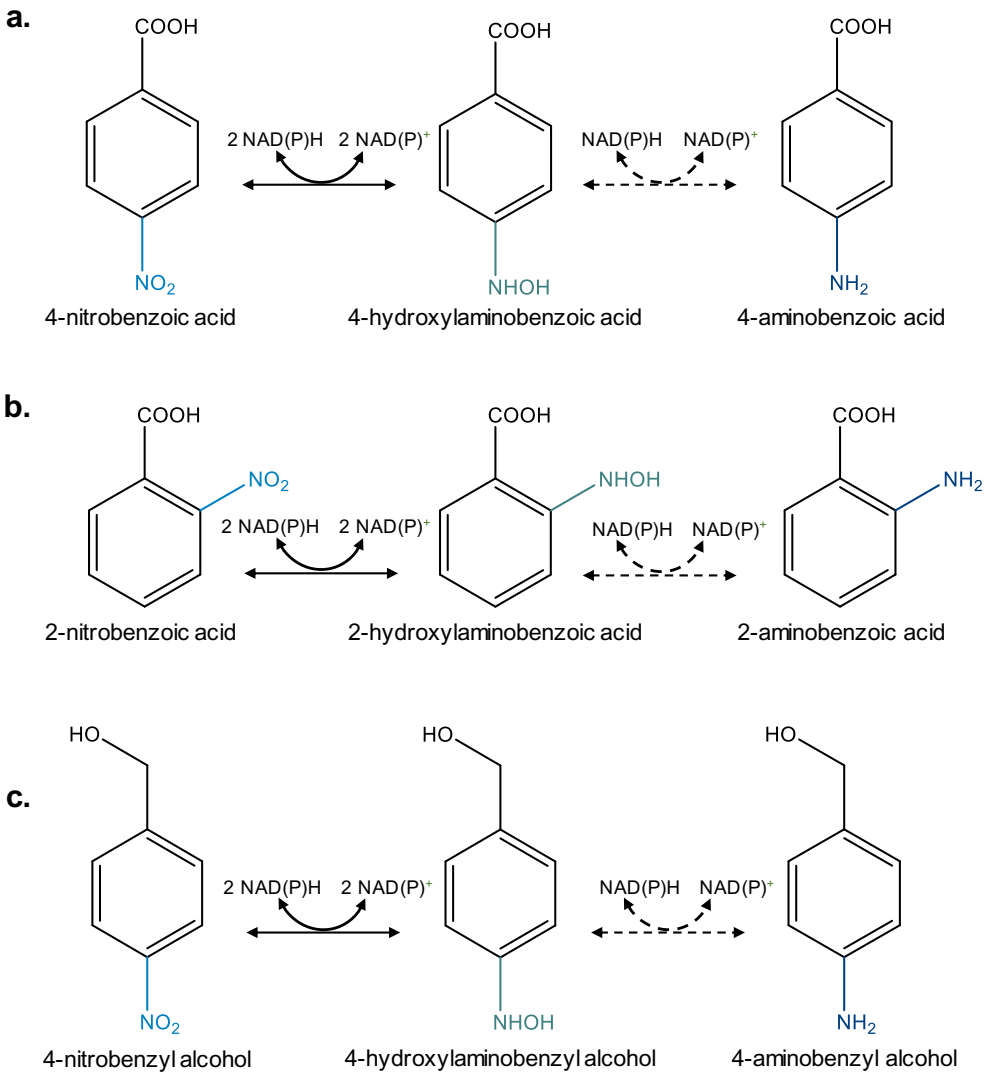


Figure 6.4. Set of reactions tested with EntNfsB. EntNfsB is known to reduce 4-NBA (**a**) to the corresponding hydroxylamine. It has been described as unable to further reduce aromatic hydroxylamines to the corresponding aromatic amines (dashed lines). 2-NBA (**b**) and 4-nitrobenzyl alcohol (**c**) were also tested as substrates.

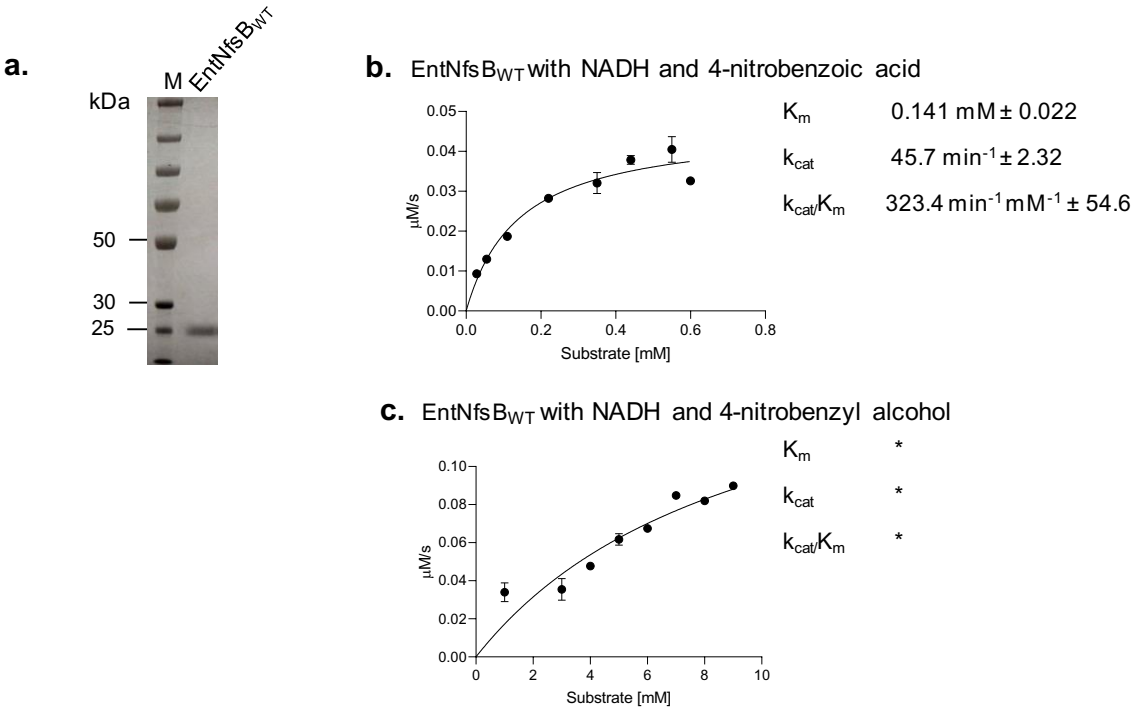


Figure 6.5. Characterization of wild-type EntNfsB. a. SDS-PAGE confirming the purity of wild-type EntNfsB purified by nickel-affinity chromatography, expected to have a molecular weight of 24.8 kDa including the hexa-histidine tag. **b.** Fitting of a Michaelis-Menten model to the initial reaction rates obtained with NADH and different 4-NBA concentrations for wild-type EntNfsB. The kinetic parameters were calculated assuming that 4-NBA was reduced to 4-hydroxylaminebenzoic acid only, and that the reduction of the nitroso derivative happened immediately after the first reduction step. **c.** Plot of initial reaction rates obtained with NADH and different 4-nitrobenzyl alcohol concentrations. Wild-type EntNfsB was able to catalyse its reduction, but with a lower efficiency than for 4-NBA, which prevented us from determining the kinetic parameters for the reaction given the limited solubility of 4-nitrobenzyl alcohol. No activity with 2-NBA was detected. Error bars show the standard deviation for each concentration.

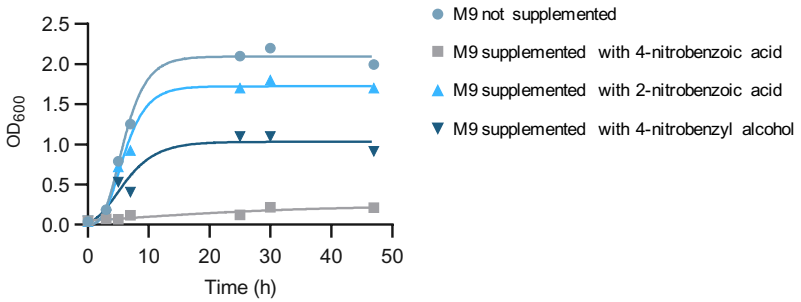


Figure 6.6. Toxicity of 4-NBA, 2-NBA and 4-nitrobenzyl alcohol. The aerobic growth of wild-type *E. coli* cells was negatively affected by the presence in the culture medium of 2.5 mM 4-NBA, 2-NBA or 4-nitrobenzyl alcohol. 4-NBA exhibited the highest toxicity. Error bars show the standard deviation for each time point.

In order to determine which residues could play a key role in determining the substrate specificity of EntNfsB, the available 4-NBA-bound crystal structure (Pitsawong *et al.*, 2017) (PDB code 5J8G) was inspected. Residues Ser40, Thr41, Tyr68 and Phe124 formed part of the substrate binding pocket while at the same time did not establish direct interactions with the essential FMN group (**Figure 6.7**), thus minimizing the possibility of detrimental effects on activity upon mutation. Therefore, a library of EntNfsB variants was generated by Golden Gate assembly with primers designed to perform saturation mutagenesis of these four residues with an NNN degeneration scheme (**Table 6.1**), resulting in a maximum theoretical library size of over 16 million variants at the nucleotide level, or 160,000 variants at the protein level (without considering variants with a premature stop codon). Library quality was confirmed by sequencing of five random variants, which were found to be all different (**Table 6.2**).

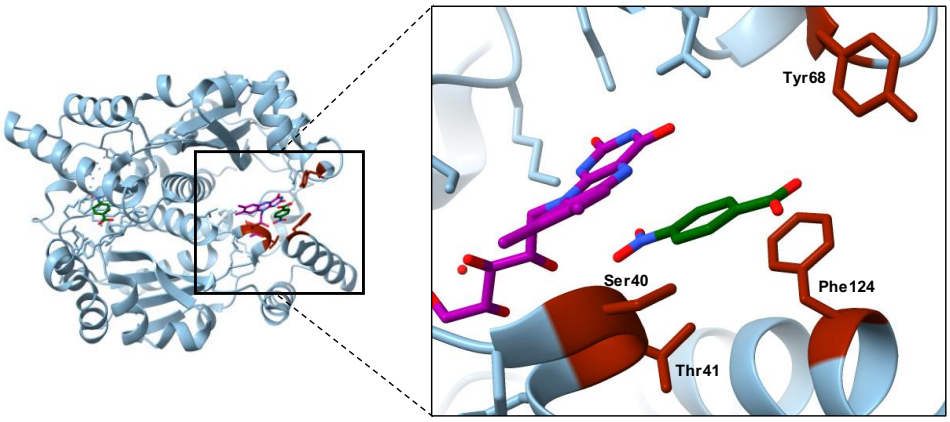


Figure 6.7. Catalytic site of wild-type EntNfsB. As observed in the structure of EntNfsB bound to 4-NBA (PDB code 5J8G), residues Ser40, and Thr41, Phe124 and Tyr68 were in close proximity of the substrate (dark green), but did not establish any interaction with the FMN prosthetic group (purple), minimizing the potential detrimental effects caused by their mutation. Therefore, they were chosen as targets for saturation mutagenesis.

Table 6.1. EntNfsB saturation mutagenesis.

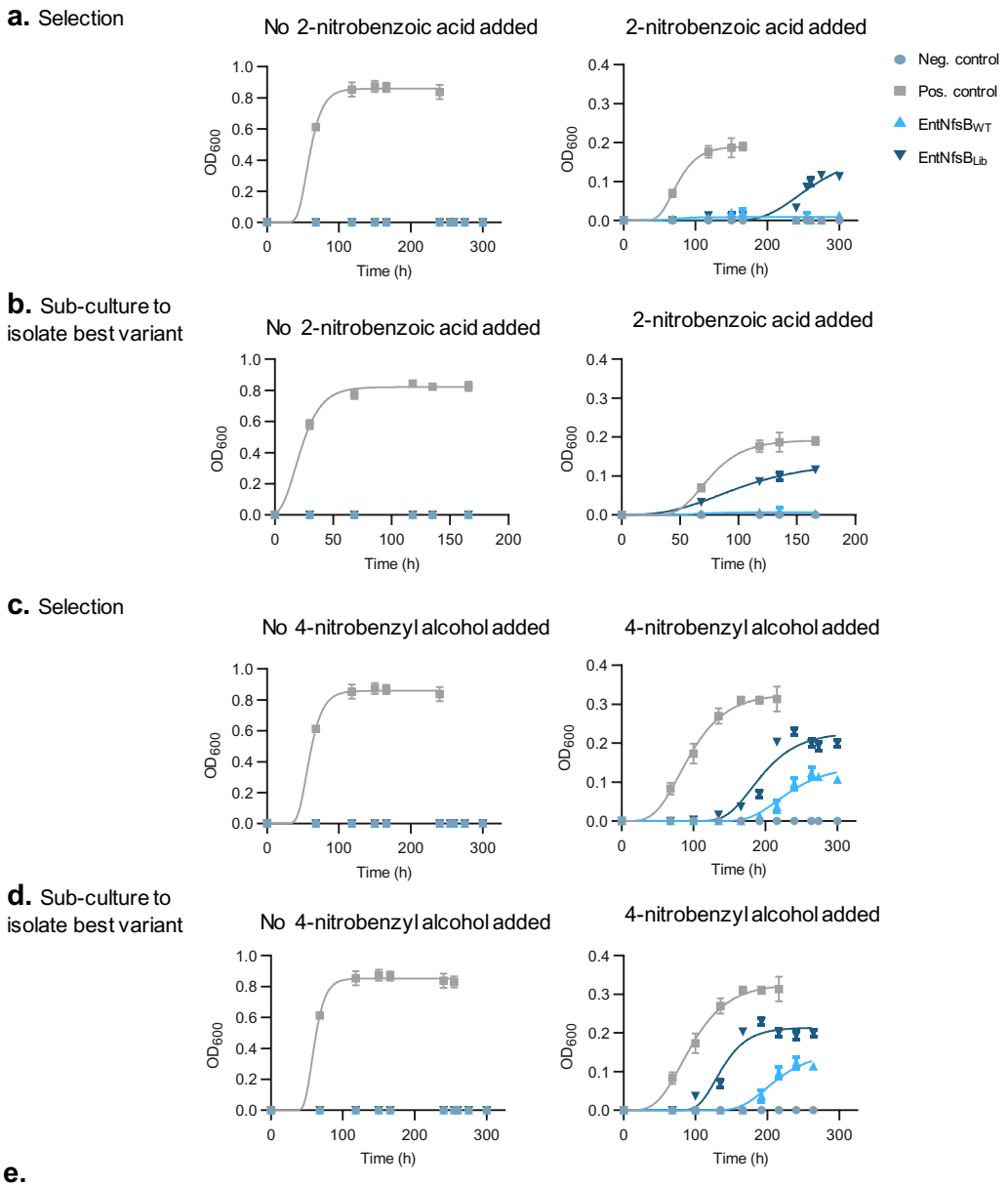
Amino acid position	40	41	68	124
$EntNfsB_{WT}$	AGC	ACC	TAT	TTC
$EntNfsB_{Lib}$	NNN	NNN	NNN	NNN

Table 6.2. Quality of the EntNfsB library. The variants contained by five random clones were sequenced, revealing that all variants were different between them and from the wild-type.

Amino acid position	40	41	68	124
$EntNfsB_{WT}$	AGC	ACC	TAT	TTC
$EntNfsB_{1}$	TTA	TTA	GTG	CAA
$EntNfsB_{2}$	AAC	GTA	TAG	CCC
$EntNfsB_{3}$	TCA	GTA	AAT	GTA
$EntNfsB_{4}$	TAC	ATA	CCT	ACC
$EntNfsB_{5}$	GTA	TTA	TGT	TGC

AL cells were transformed with the library of EntNfsB variants or wild-type EntNfsB. The transformation efficiency for AL cells transformed with the library was estimated to be approximately three million cells per transformation. Therefore, six independent transformations with the library were performed for growth complementation experiments. Metabolic complementation experiments were carried out by culturing transformed cells anaerobically in M9 medium supplemented with glucose and either 2-NBA or 4-nitrobenzyl alcohol. Cells transformed with wild-type EntNfsB were able to grow in medium supplemented with 4-nitrobenzyl alcohol (observable after ten days and reaching a maximum OD_{600} of 0.13), but not with 2-NBA, in accordance with the ability to reduce 4-nitrobenzyl alcohol but not 2-NBA previously observed in enzymatic assays. However, cells transformed with the library of variants were able to grow both in medium with 2-NBA and 4-nitrobenzyl alcohol, after ten and seven days and to maximum OD_{600} values of 0.12 and 0.23 respectively (**Figure 6.8**). Grown cultures with cells transformed with the library were used to inoculate fresh anaerobic medium with the appropriate substrate in order to further enrich the selected clones, making them easier to isolate. Anaerobic growth was then observed after three days for cells grown in the presence of 2-NBA, and after four days with 4-nitrobenzyl alcohol.

Plasmid DNA was isolated from the second round of anaerobic cultures of AL cells transformed with the library. Sequencing revealed that for cultures both with 2-NBA and 4-nitrobenzyl alcohol a single variant had been selected. Variant $EntNfsB_{S1}$, selected in the presence of 2-NBA, contained the Ser40Ala, Thr41Ile and Phe124Ala substitutions, while in $EntNfsB_{S2}$, selected in the presence of 4-nitrobenzyl alcohol, the Thr41Leu, Tyr68Leu and Phe124Leu substitutions were found.



e.

Amino acid position	40	41	68	124
EntNfsB$_{WT}$	Ser (AGC)	Thr (ACC)	Tyr (TAT)	Phe (TTC)
EntNfsB$_{S1}$	**Ala (GCA)**	**Ile (ATA)**	Tyr (TAT)	**Ala (GCA)**
EntNfsB$_{S2}$	Ser (TCA)	**Leu (CTA)**	**Leu (CTT)**	**Leu (CTC)**

Figure 6.8. Selection of EntNfsB variants. a, c. AL cells transformed with the library of EntNfsB variants were able to grow anaerobically in medium supplemented with 2-NBA (**a**) or 4-nitrobenzyl alcohol (**c**), but not in medium without any of them. AL cells transformed with wild-type EntNfsB were also able to grow in medium supplemented with 4-nitrobenzyl alcohol, but not in medium with 2-NBA, in accordance with the previous observation that wild-type EntNfsB can reduce 4-nitrobenzyl alcohol but not 2-NBA. **b, d.** Sub-cultures into the same conditions of the cells transformed with the library and grown in the presence of 2-NBA (**b**) or 4-nitrobenzyl alcohol (**d**) were performed to further enrich the selected variants before isolating plasmid DNA to sequence it. A similar behavior as for the first cultures

was observed, with the difference that the sub-cultures of cells transformed with the library grew faster thanks to the presence of larger initial populations of cells with the selected variants. **e.** Amino acid residues present at the saturated positions in wild-type EntNfsB and the selected $EntNfsB_{S1}$ and $EntNfsB_{S2}$ variants. The corresponding DNA codons are shown in brackets. Error bars show the standard deviation for each time point.

6.2.2. Characterization of the novel EntNfsB variants

AL cells were transformed with $EntNfsB_{S1}$ or $EntNfsB_{S2}$ and grown anaerobically in M9 medium supplemented with glucose and 2-NBA or 4-nitrobenzyl alcohol respectively (**Figure 6.9**). The lag phase lasted 118 h for $EntNfsB_{S1}$ with 2-NBA and 115 h for $EntNfsB_{S2}$ with 4-nitrobenzyl alcohol. ^{1}H-NMR spectra of the fermentation broth of both cultures, as well as of AL cells transformed with wild-type EntNfsB and grown anaerobically in M9 medium supplemented with glucose and 4-NBA, were acquired in order to determine product formation. Due to the unavailability of 2-hydroxylaminobenzoic acid, 4-hydroxylaminobenzoic acid and 4-hydroxylaminobenzyl alcohol in most commercial suppliers, the acquired spectra were compared with reference spectra of 2-aminobenzoic acid, 4-aminobenzoic acid and 4-aminobenzyl alcohol. Surprisingly, the spectra revealed the presence of the aromatic amine corresponding to the supplied nitroaromatic compound in all cultures (**Table 6.3**), in spite of previous reports indicating the inability of EntNfsB to reduce 4-NBA to 4-aminobenzoic acid (Pitsawong, Hoben, and Miller, 2014; Miller *et al.,* 2018). Nevertheless, said reports only assayed the *in vitro* activity, over periods of up to 24 h. It is possible the highly reducing conditions found in the cellular context of the AL strain, combined with longer periods of incubation under anaerobic conditions, allows for the formation of the amine products at detectable levels. However, due to the lack of spectral data of the hydroxylamines corresponding to 2-NBA, 4-NBA and 4-nitrobenzyl alcohol, it cannot be ruled out that the chemical shifts of their aromatic protons overlap with the chemical shifts of the aromatic amines. Further analysis where the spectra of the fermentation broths are compared to the spectra of the aromatic hydroxylamines will be required in order to unambiguously determine the formed product.

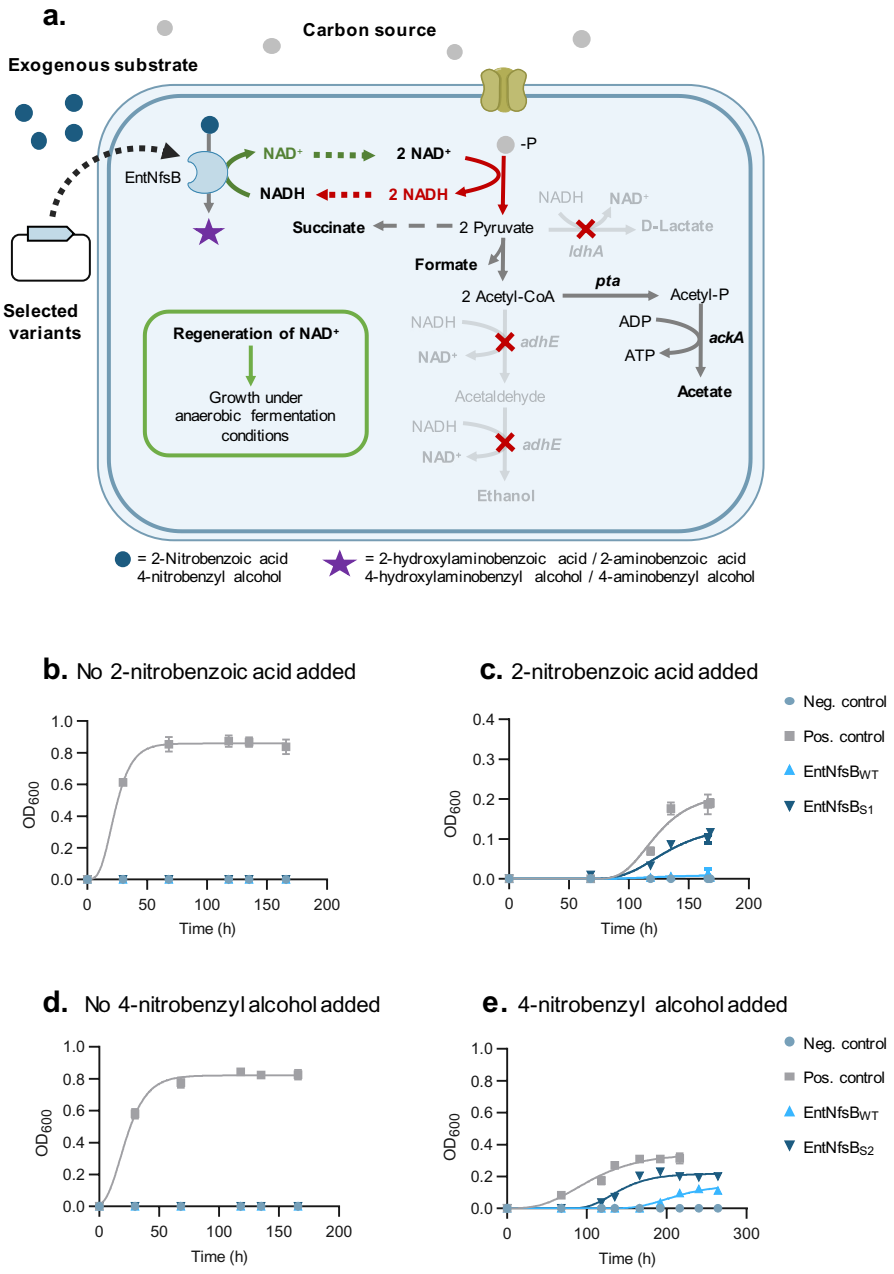


Figure 6.9. Metabolic complementation with the selected EntNfsB variants. a. AL cells transformed with $EntNfsB_{S1}$ or $EntNfsB_{S2}$ would be able to grow anaerobically in medium supplemented with 2-NBA or 4-nitrobenzyl alcohol respectively, thanks to the regeneration of NAD^+ provided by the EnfNfsB variants upon reduction of 2-NBA or 4-nitrobenzyl alcohol. **b, c.** AL cells transformed with $EntNfsB_{S1}$ were able to grow anaerobically in medium supplemented with 2-NBA, but not in medium without it. **d, e.** AL cells transformed with $EntNfsB_{S2}$ were able to grow anaerobically in medium supplemented with 4-nitrobenzyl alcohol, but not in medium without it. Error bars show the standard deviation for each time point.

After purifying both variants, enzymatic assays were performed with each of them. Both variants retained the ability of reducing 4-NBA. $EntNfsB_{S1}$ was also able to reduce 2-NBA, in accordance with its ability to restore anaerobic growth of AL cells in the presence of 2-NBA,

displaying a K_m for 2-NBA of 0.808 mM and a k_{cat} of 28.4 min^{-1}, with the assumptions that the substrate is only reduced to 2-hydroxylaminobenzoic acid and that the two sequential two-electron transfers happen immediately one after the other one (**Figure 6.10**). Interestingly, the K_m for 4-NBA of EntNfsB$_{S1}$ (0.051 mM) was not worse than that of wild-type EntNfsB, suggesting that the introduced substitutions do not have a negative effect on the binding of 4-NBA. EntNfsB$_{S2}$ was able to reduce 4-nitrobenzyl alcohol more efficiently than wild-type EntNfsB, with a K_m for the substrate of 1.11 mM and a k_{cat} of 205.0 min^{-1} under the same assumptions than for EntNfsB$_{S1}$ (**Figure 6.10**).

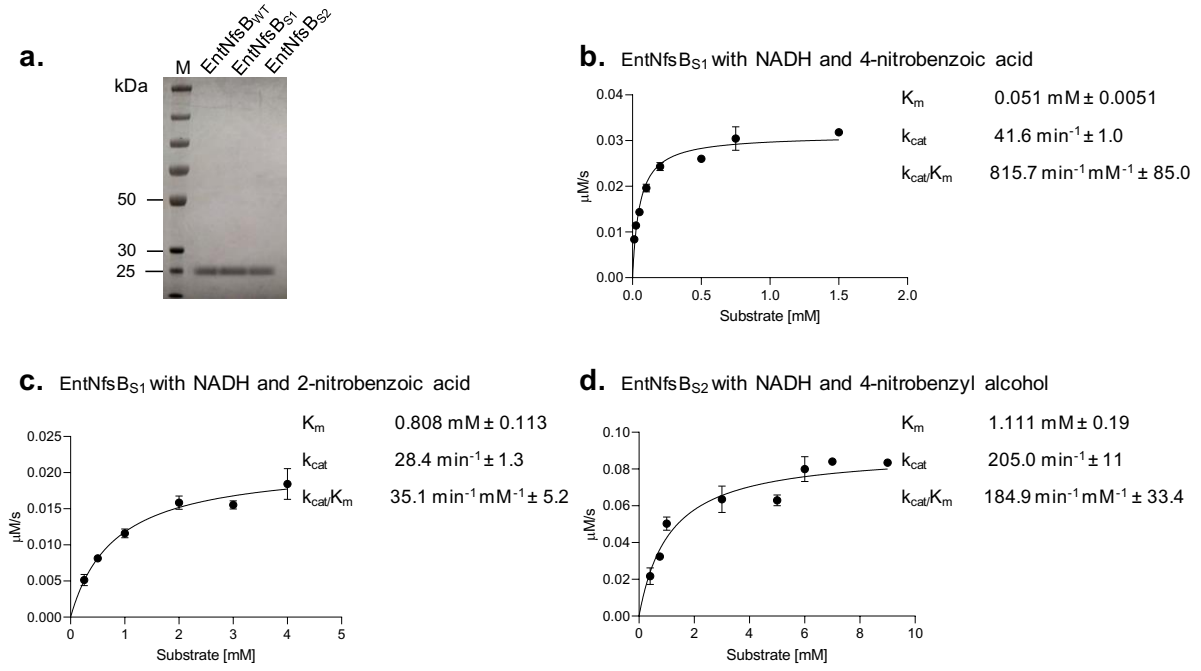


Figure 6.10. Characterization of EntNfsB$_{S1}$ and EntNfsB$_{S2}$. a. SDS-PAGE confirming the purity of EntNfsB$_{S1}$ and EntNfsB$_{S2}$ purified by nickel-affinity chromatography, both expected to have a molecular weight of 24.8 kDa including the hexa-histidine tags. **b, c.** Fitting of a Michaelis-Menten model to the initial reaction rates obtained with NADH and different concentrations of 4-NBA and 2-NBA for EntNfsB$_{S1}$. **d.** Fitting of a Michaelis-Menten model to the initial reaction rates obtained with NADH and different 4-nitrobenzyl alcohol concentrations for EntNfsB$_{S2}$. The kinetic parameters were calculated assuming that the substrates were reduced to the corresponding hydroxylamines, and that the reduction of the nitroso derivatives happened immediately after the first reduction step. Error bars show the standard deviation for each concentration.

Additionally, both EntNfsB$_{S1}$ and EntNfsB$_{S2}$ also displayed FMN reductase activity. This is in accordance with a previous report by Zenno *et al*. where they observed that the mutation of Phe124 of *E. coli* NfsB (which shares 88% sequence identity with EntNfsB) to any other residue is sufficient to convert NfsB to a flavin reductase, in most cases without a considerable loss of nitroreductase activity (Zenno *et al*., 1996). Interestingly, a protein family has been identified including nitroreductases such as EntNfsB and its homologues in other organisms,

as well as flavin reductases like flavin reductase I (FRase I) from *Vibrio fischeri*. This suggests that the genes encoding NfsB-like nitroreductases and FRase I share a common ancestor, which later diverged to give raise to enzymes specializing either in nitroreductase or flavin reductase activity (Zenno *et al*., 1996).

Table 6.3. Quantification of metabolites by ^{1}H-NMR of the fermentation broth of AL cells complemented with wild-type EntNfsB, EntNfsB$_{S1}$ and EntNfsB$_{S2}$. The spectra suggested the presence of the corresponding aromatic amines, although it could not be definitely confirmed if the generated products were the aromatic amines or hydroxylamines corresponding to the nitroaromatic substrates.

Strain	Plasmid	Encoded enzyme	Exogenous substrate	Resulting product	Substrate δ * (ppm)	Product δ * (ppm)	[Substrate] (mM)	[Product] (mM)	[Ethanol] (mM)	[Lactate] (mM)	[Succinate] (mM)	[Acetate] (mM)	[Formate] (mM)
AL	pLS168	EntNfsB$_{WT}$	4-nitrobenzoic acid	4-hydroxylaminobenzoic or 4-aminobenzoic acid	8.01 (d, 2)	7.73 (d, 2)	0	0.4	0	0	0.2	2.8	2.1
AL	pLS168	EntNfsB$_{WT}$	2-nitrobenzoic acid	2-hydroxylaminobenzoic or 2-aminobenzoic acid	8.11 (d, 1)	7.30 (t, 1)	7.9	0	0	0	0	0	0
AL	pLS168	EntNfsB$_{WT}$	4-nitrobenzyl alcohol	4-hydroxylaminobenzyl alcohol or 4-aminobenzyl alcohol	8.27 (d, 2)	7.23 (d, 2)	0.2	0.5	0	0	0.3	4.1	3.3
AL	pLS169s1	EntNfsB$_{s1}$	2-nitrobenzoic acid	2-hydroxylaminobenzoic or 2-aminobenzoic acid	8.11 (d, 1)	7.30 (t, 1)	2.5	4.1	0	0	1.1	11	8.3
AL	pLS169s2	EntNfsB$_{s2}$	4-nitrobenzyl alcohol	4-hydroxylaminobenzyl alcohol or 4-aminobenzyl alcohol	8.27 (d, 2)	7.23 (d, 2)	0	1	0	0	0.4	6.1	5.4

* δ of characteristic signal of substrate or product (ppm). The multiplicity of each signal (s – singlet, d – doublet, t – triplet, q – quartet, m – multiplet) and the number of contributing protons is shown between brackets.

6.2.3. Structural basis for the change in substrate selectivity of EntNfsB

The previously solved structure of 4-NBA-bound wild-type EntNfsB revealed the binding mode of 4-NBA in the active site (Pitsawong *et al*., 2017), PDB code 5J8G). The nitro group of 4-NBA is stacked on top of the isoalloxazine ring of FMN, allowing an efficient direct hydride transfer from the reduced flavin N5 atom. The nitro group interacts with the backbone of Thr41, the benzene ring establishes a tilted stacked-ring interaction with the side chain of Phe124 and the carboxyl group forms a hydrogen bond with a water molecule which also interacts with the side chains of Glu165 and His128 (**Figure 6.11**).

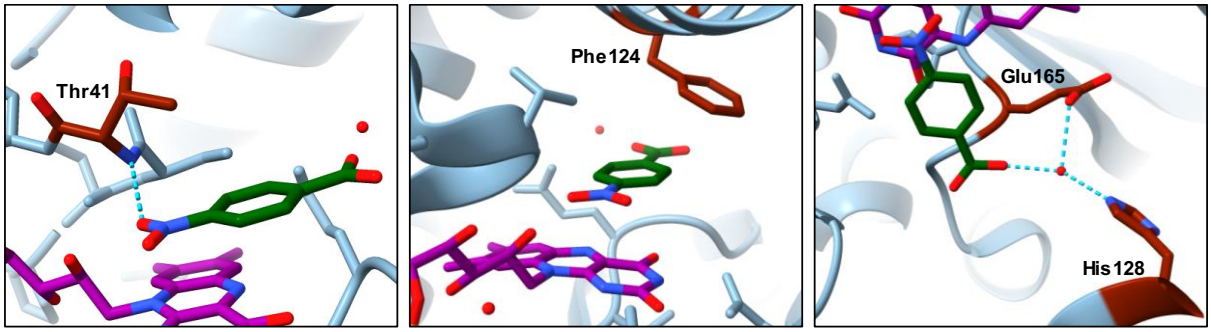


Figure 6.11. Interactions of wild-type EntNfsB with 4-NBA. As observed in the crystal structure (PDB code 5J8G), wild-type EntNfsB establishes a hydrogen bond with the nitro group of 4-NBA (dark green) through the backbone of Thr41 (left panel) and a tilted stacked-ring interaction with its aromatic ring through the side chain of Phe124 (middle panel). Additionally, a crystallographic water mediates the interaction between the carboxyl group of 4-NBA and the side chains of His128 and Glu165 by means of hydrogen bonds (right panel).

Substituting 4-NBA by 2-NBA in the wild-type structure with a positioning of the nitro group and aromatic ring similar to that of 4-NBA led to clashes between the carboxylate group of 2-NBA and the side chain and backbone atoms of Ser40 and Thr41 in one of the possible binding modes (**Figure 6.12a**). In the other possible binding mode, a clash with the side chain of Phe128 was detected (**Figure 6.12b**). I hypothesize that the Ser40Ala mutation in EntNfsB$_{S1}$ possibly allows for an alternative conformation of the loop where it is located thanks to its smaller side chain, which would alleviate the steric impediment for the binding of 2-NBA caused by the protein backbone of said loop in the first hypothesized binding mode. Additionally, the Phe124Ala substitution would eliminate the clash with the carboxyl group of 2-NBA in the other potential binding modes due to the introduction of a much less bulkier alanine side chain (**Figure 6.12c**). The introduction of this mutation also eliminates the tilted stacked-ring interaction established with the aromatic ring of the substrate in the wild-type enzyme, possibly allowing for more flexibility in the binding of the substrate. Nevertheless, other alternative placements of the substrate cannot be ruled out, given the known promiscuity

with which nitroreductases bind their substrates (Johansson *et al.*, 2003; Race *et al.*, 2005; Pitsawong *et al.*, 2017).

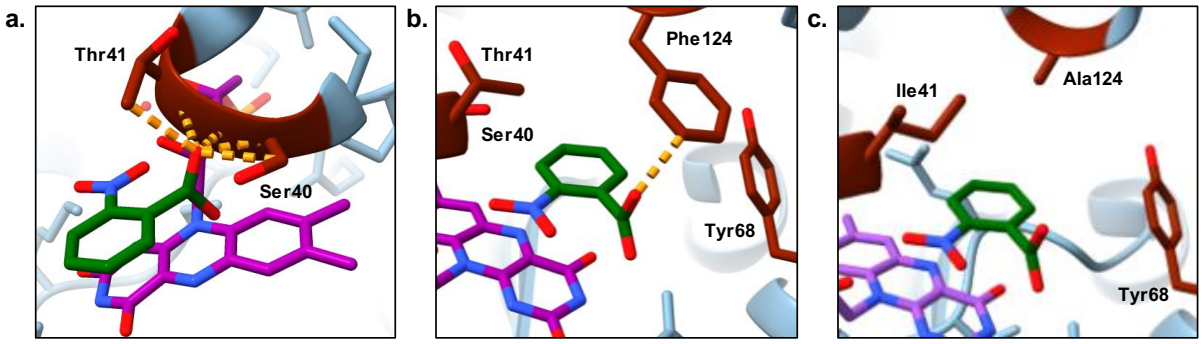


Figure 6.12. Structural basis for the ability of EntNfsB$_{S1}$ to accept 2-NBA as a substrate. Structural models were generated by replacing 4-NBA with 2-NBA in the crystal structure of wild-type EntNfsB (PDB code 5J8G), keeping the same position for the nitro group and the aromatic ring. **a.** In wild-type EntNfsB, one of the possible placements of 2-NBA (dark green) would lead to clashes (oranges) between the carboxyl group of the substrate (in *ortho* position with respect to the nitro group) and the backbone of Ser40 and Thr41. It is possible that the substitution of Ser40 by an alanine residue in EntNfsB$_{S1}$ allows for a different conformation of the loop containing residues 40 and 41 which would not clash with 2-NBA, thanks to the smaller side-chain of alanine. **b.** In the other possible placement of 2-NBA, the carboxyl group would clash with the side chain of Phe124 in wild-type EntNfsB. **c**. However, in EntNfsB$_{S1}$, such a clash would not exist thanks to the Phe124Ala substitution.

While it is harder to rationalize the increase in activity towards 4-nitrobenzyl alcohol observed in EntNfsB$_{S2}$, it is worth noting that the Ser40 residue was not substituted and Phe124 was substituted by a residue containing also a bulky side chain, in accordance with the lack of a need to accommodate a substituent in the *ortho* position with respect to the nitro group. Nevertheless, as in the EntNfsB$_{S1}$ variant, the Phe124 residue was also substituted by a non-aromatic residue. Interestingly, in a previous study certain mutations of Phe124 in *E. coli* NfsB resulted in improved activity towards the CB1954 prodrug, suggesting a role for Phe124 in substrate specificity in addition to determining the absence of flavin reductase activity (Grove *et al.*, 2003).

6.3. Conclusions

The selection of EntNfsB$_{S1}$ and EntNfsB$_{S2}$ demonstrates the potential of the redox rescue selection system for the development of variant enzymes with novel substrate selectivity, as well as variants with improved activities towards substrates that wild-type enzymes are already able to accept. The selection system enabled the identification of the first nitroreductase to my knowledge able to reduce both 2-NBA and 4-NBA, as well as a variant with improved activity towards 4-nitrobenzyl alcohol. Both mononitrobenzoates (which include 2-NBA and 4-NBA) and nitrobenzyl alcohols are toxic compounds employed during the manufacturing of explosives and other products frequently found in industrial wastes and groundwaters near ammunition facilities (Grummt *et al*., 2006). The selected EntNfsB variants, or others that could be selected with the same selection system, could prove to be useful tools for the engineering of microorganisms able to efficiently degrade such environmental pollutants.

Nitroreductases also offer a great potential for both bioremediation and cancer therapy. However, their actual applicability has been hindered by suboptimal activity towards the desired substrates and the lack of a fast and simple methodology to tailor existing enzymes to match the requirements for each specific application. For example, previous studies aiming to obtain variants of enzymes of the NfsB family with improved activity towards CB1954 have often required the screening of hundreds or thousands of variants, resulting in a laborious and time-consuming process. Furthermore, in many cases only one position of the protein sequence is targeted in each library, preventing the finding of substitutions with a cooperative effect towards the improved activity with the substrate of interest (Grove *et al*., 2003; Swe *et al.*, 2012; Jaberipour *et al.*, 2010).

The presented workflow to obtain variants of EntNfsB exhibiting the ability to reduce 2-NBA or improved activity towards 4-nitrobenzyl alcohol could be easily replicated to obtain variants of clinical significance simply by replacing the exogenous compound added to the culture medium to restore anaerobic growth. Such an approach creates the possibility of performing ultra-high throughput searches of variants of a given nitroreductase against a panel of prodrugs with hardly any need of human intervention, potentially allowing for a large expansion of the scope of GDEPT-based cancer therapies.

6.4. Key outcomes

- A variant of *Enterobacter cloacae* NfsB nitroreductase able to reduce 2-NBA, a substrate that wild-type EntNfsB is unable to reduce, has been selected,

demonstrating the potential of the redox rescue selection system for altering substrate selectivity.

- A variant of *Enterobacter cloacae* NfsB nitroreductase able to reduce 4-nitrobenzyl alcohol considerably more efficiently than wild-type EntNfsB has been selected, demonstrating the potential of the redox rescue selection system for improving the efficiency of catalysis with substrates already accepted by a given enzyme.
- $EntNfsB_{S1}$ is the first known nitroreductase able to reduce both 2-NBA and 4-NBA.

7. Selection of an entire multi-enzymatic metabolic pathway

7.1. Introduction

While originally devised as a means to obtain novel variants of specific biomolecules, directed evolution has also been extended to the evolution of more complex systems, such as entire metabolic pathways where a set of enzymes work together towards the generation of a certain product. Nevertheless, only a reduced amount of successful cases of directed evolution of metabolic pathways have been reported. This is due to the additional challenges posed by the evolution of metabolic pathways able to perform optimally in a cellular context. Even if the individual enzymes comprising a metabolic pathway are optimized individually (through protein sequence modifications) using directed evolution approaches, their combination does not necessarily lead to a maximized yield of the final desired product (Chatterjee and Yuan, 2006). This can be due to a number of reasons, which can cause unexpected effects when the engineered pathway is placed within the larger context of global cellular metabolism. For instance, the presence of an engineered, non-native metabolic pathway, which often is not subjected to the endogenous regulatory mechanisms, can lead to a disruption of the equilibrium in the intracellular pools of metabolites, causing a deleterious effect in key cellular processes with the consequent reduced flux across the pathway of interest (Lee *et al*., 2012). For example, some *E. coli* strains where an isopentenyl pyrophosphate biosynthetic pathway was engineered in order to serve as the basis for the synthesis of isoprenoids were found to have their growth severely inhibited due to the accumulation of 3-hydroxy-3-methylglutaryl-CoA, which inhibits the synthesis of fatty acids (Kizer *et al*., 2008).

Metabolic engineering approaches that allow to model metabolic fluxes across large metabolic networks are available, such as flux balance analysis. However, these approaches suffer from limitations due to inaccuracies in the employed metabolic model. These can arise from the lack of consideration of compartmentalization, as well as the existence of underground metabolism, a set of uncharacterized side reactions performed by promiscuous enzymes and non-enzymatic reactions which are not accounted for in the models (Noda-Garcia *et al*., 2018). Additionally, even if a steady-state assumption is made (each internal metabolite is consumed at the same rate than it is produced), the solution space satisfying this condition is often too large to obtain meaningful insight or make predictions about the system of interest. Therefore, further constraints must be made, typically relying on experimental findings or manual curation of literature. If such additional constraints are not accurate, they can lead to the prediction of physiologically unlikely metabolic fluxes and incorrect conclusions about the system (Orth, Thiele, and Palsson, 2010; Smith *et al*., 2017).

One of the key parameters that must be fine-tuned when engineering a metabolic pathway is the amount of each enzyme that is produced. Achieving the optimal level of expression is not only crucial to avoid the previously mentioned imbalance in metabolite pools, but also to maintain an efficient and stoichiometric metabolic flux across the newly introduced pathway. Such an approach was applied by Ajikumar *et al*., who split a pathway for the production of taxadiene (an intermediate for the production of the anticancer drug taxol) into two modules and screened for different combinations of promoters and replication origins to find the optimal expression level and copy number of each module that allowed to maximize the taxadiene yield (Ajikumar *et al*., 2010). Similar modular optimization approaches have been since then applied to a variety of metabolic pathways (Jeschek, Gerngross, and Panke, 2017).

Nevertheless, adjusting such regulatory parameters tends to be a laborious process, often requiring a thorough analysis of the levels of the involved metabolites and quantification of the expression of each component of the pathway. The difficulty to rationally determine the best regulatory elements for a given pathway is further increased due to the frequent lack of information about the *in vivo* activity of the involved enzymes, as well as the possibility of unexpected responses from the host organism upon the expression of the engineered pathway (Nielsen and Keasling, 2016). Therefore, a broadly applicable system enabling the selection of metabolic pathways without the need of extensive and time-consuming screening would be a valuable tool of general interest for the field of metabolic engineering.

In this chapter, the aim was to demonstrate the applicability of the redox rescue selection system to entire metabolic pathways. As a test case, I sought to obtain optimized metabolic pathways for the production of isopropanol. Isopropanol has several applications of human interest, such as its uses as a disinfectant, solvent of non-polar compounds, rubbing alcohol and for cleaning, in addition to its laboratory applications. It can also be used to esterify fat and oil to produce biodiesel with reduced tendency to solidify at low temperatures (Lee *et al*., 1995). While isopropanol is natively produced by several species of *Clostridium*, such as *Clostridium beijerinckii*, due to the difficulty of genetic manipulation of these strains Hanai *et al*. decided to change the host to *E. coli*, which in spite of not being a native isopropanol producer, is a much better studied and easier to engineer organism (Jeschek, Gerngross, and Panke, 2017). They successfully engineered an efficient synthetic isopropanol production pathway in *E. coli*, which included five genes from three different organisms: *Clostridium acetobutyilicum thl* (acetyl-CoA acetyltransferase; or thiolase) or *E. coli atoB* (acetyl-CoA acetyltransferase), *E. coli atoA* and *atoD* (acetoacetyl-CoA transferase), *Clostridium acetobutyilicum adc* (acetoacetate decarboxylase) and *Costridium beijerinckii adh* (CBADH) (**Figure 7.1**) (Hanai *et al*., 2007).

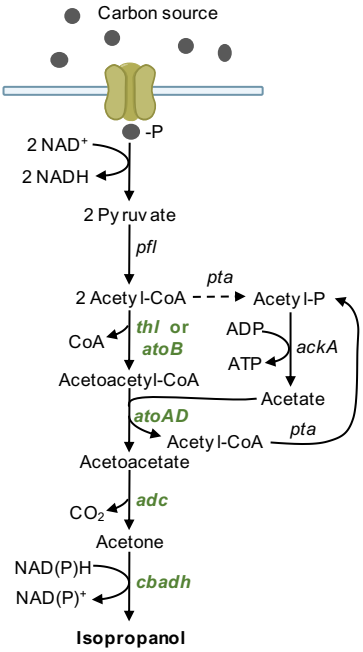


Figure 7.1. Engineered isopropanol production pathway. Hanai *et al*. engineered a synthetic isopropanol production pathway in *E. coli*. An acetyl-CoA acetyltransferase (encoded by *thl* or *atoB*) catalyses the condensation of two acetyl-CoA molecules into one acetoacetyl-CoA, releasing also one free CoA molecule. Then, an acetoacetyl-CoA transferase (comprised by two subunits encoded by *atoA* and *atoD*) transfers the CoA group from acetoacetyl-CoA to acetate, yielding acetoacetate and acetyl-CoA. Acetoacetate is decarboxylated by an acetoacetate decarboxylase (encoded by *adc*), resulting in acetone and CO_2. Finally, acetone is reduced to isopropanol by CBADH. The acetate molecule consumed during the reaction catalysed by the acetoacetyl-CoA transferase can be regenerated from the resulting acetyl-CoA by the actions of phosphate acetyltransferase and acetate kinase (encoded by *pta* and *ackA* respectively).

The pathway developed by Hanai *et al*. was used as the basis for the construction of a library of variants differing in the regulatory elements controlling the transcription and expression of each of the genes, and demonstrated the ability of the redox rescue selection system to select variants yielding the highest titres of isopropanol.

7.2. Results and discussion

7.2.1 Selection of an optimized isopropanol production pathway

In the pathway designed by Hanai *et al*., only a single combination of regulatory elements for the components of the pathway was tested. The *thl* or *atoB*, *atoAD* and *adc* genes were cloned into a plasmid with a ColE1 replication origin and placed under the control of a single $P_L lacO_1$ promoter. The CBADH gene was cloned into a separate plasmid with a p15A replication origin, also under the control of a $P_L lacO_1$ promoter. A ribosome binding site (RBS) derived from the pET-31b(+) expression vector was used for all genes except *thl* or *atoB*, which were cloned

directly after the RBS found in the parental pZE21-MCS1 vector (Hanai *et al*., 2007). It is difficult to predict how sensitive this system is to the expression levels of the enzymes, and impossible to know how suitable this single design is, but it is very unlikely to be optimal. I hypothesized that expressing such pathway in AL cells would allow the recovery of growth under anaerobic fermentation conditions thanks to the regeneration of oxidised $NADP^+$ by CBADH. The $NADP^+$ would in turn be used by transhydrogenases to yield NAD^+, enabling metabolic complementation. Furthermore, since the initial substrate for the pathway is the endogenously produced acetyl-CoA, the addition of an external oxidised substrate to act as the electron acceptor would not be required (**Figure 7.2**). By using the redox rescue selection system to couple isopropanol production to growth through the associated NAD^+ generation, I sought to find the best expression profile leading to an optimized trade-off between high flux, avoiding accumulation of intermediate metabolites and avoiding toxicity or excessive metabolic burden due to the overexpression of the enzymes of the exogenous metabolic pathway.

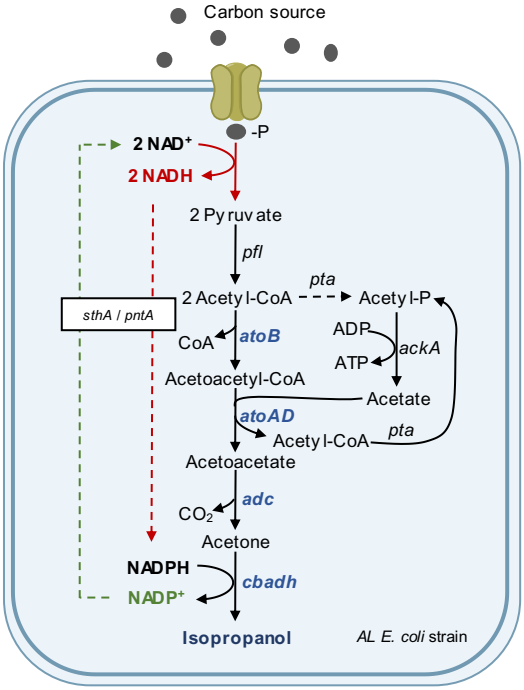


Figure 7.2. Metabolic complementation with the isopropanol production pathway. AL cells containing the isopropanol production pathway designed by Hanai *et al*. would be able to grow under anaerobic fermentation conditions thanks to the $NADP^+$ produced by CBADH, which would be used by transhydrogenases to regenerate NAD^+. Additionally, since the rest of the enzymes of the synthetic pathway can generate acetone from acetyl-CoA, produced endogenously from the pyruvate resulting from glycolysis, no exogenous substrate would be required to be added to the culture medium.

In order to have a wide variety of expression levels for each protein, two elements controlling the final level of expression of each gene were targeted: the promoter (controlling transcription) and the RBS (controlling the efficiency of the translation of mRNA to protein). A

set of six promoters and six RBS previously characterized spanning a wide and evenly spaced range of transcription and translation strengths (Taylor *et al*., 2019) were chosen to be tested for each of the genes, and a library was generated by means of the Start-Stop assembly methodology (Taylor *et al*., 2019). Instead of the *Clostridium acetobutylicum thl* gene, the *E. coli atoB* gene was used, which was also tested by Hanai *et al*. and yielded isopropanol levels very similar to those obtained with *thl*. Considering all possible 36 combinations of promoters and RBS for each of the five genes, a total of $36^5 \approx 60 \times 10^6$ different potential variants existed. The transformation efficiency of AL cells with the metabolic pathway library (MP_{Lib}) was estimated to be roughly two million successfully transformed cells per transformation. After performing 15 independent transformations of AL cells with the MP_{Lib} (resulting in 30 million variants being sampled, and therefore covering approximately half of the library), they were inoculated in anaerobic M9 medium supplemented with gluconate, with the expectation that in cells transformed with the most efficient variants anaerobic growth would be restored. Gluconate was chosen as the carbon source instead of glucose due to a better stoichiometry for the regeneration of oxidised NAD^+ (**Figure 7.3**). As previously explained, when *E. coli* metabolizes glucose through the EMP route 2 NADH molecules are generated for each glucose molecule. Nevertheless, even if all the acetyl-CoA derived from glucose entered the exogenous isopropanol production pathway without any flux towards other metabolic routes, only one molecule of NADH would be reoxidised to NAD^+ for each glucose molecule (**Figure 7.3**). Gluconate, on the other hand, is metabolized mostly through the Entner-Doudoroff pathway (Eisenberg and Dobrogosz, 1967), resulting in the generation of a single NADH molecule per gluconate molecule, in accordance with the more oxidised state of gluconate compared to glucose. The isopropanol pathway would thus allow for a theoretical stoichiometric reoxidation of NAD^+ (**Figure 7.3**). It should be noticed that even though the essential acetoacetyl-CoA transferase step stoichiometrically consumes acetate and produces acetyl-CoA (**Figures 7.1 and 7.2**), this is effectively catalytic, and no net consumption or production of these species is required, because acetyl-CoA can be converted back to acetate via acetyl-phosphate.

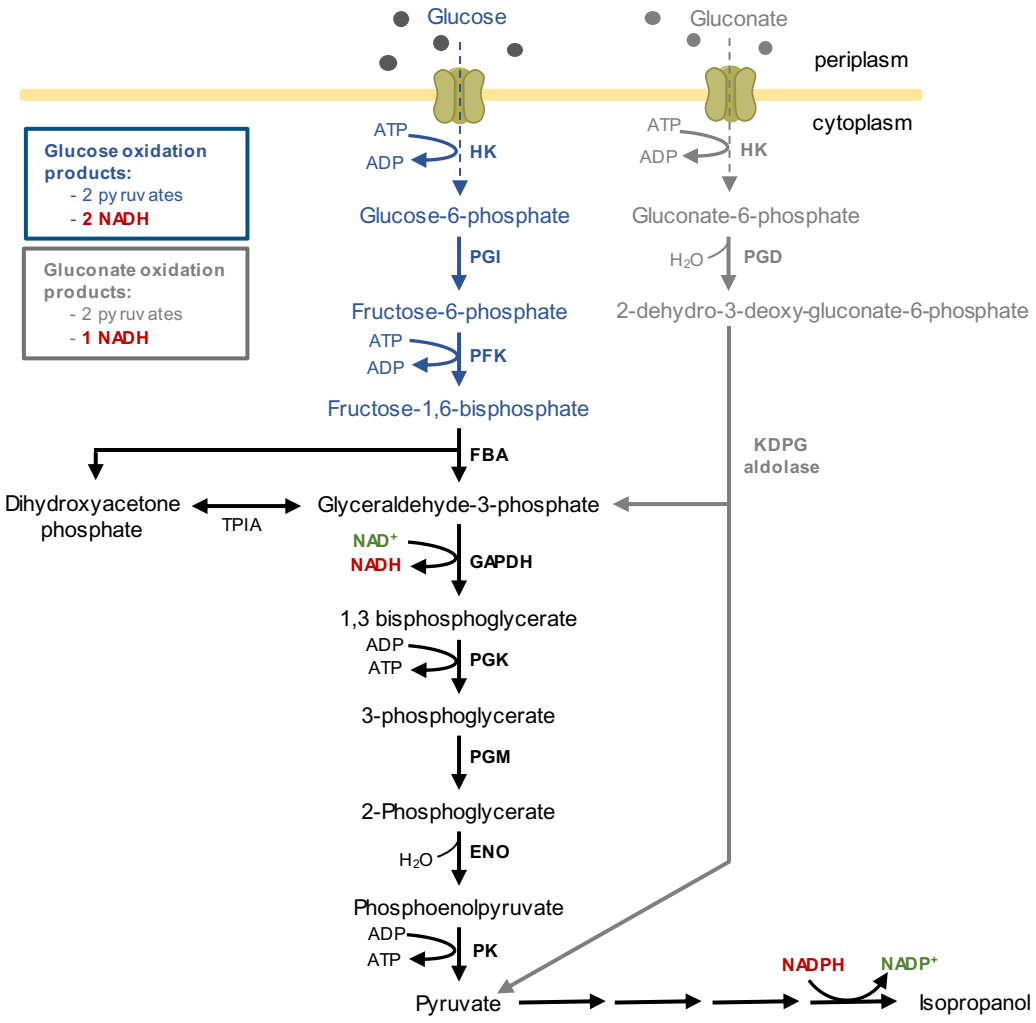


Figure 7.3. Balance of NAD⁺ regeneration with glucose and gluconate as the carbon sources. During oxidation of glucose to pyruvate, 2 NADH molecules per glucose are produced. However, during the oxidation of gluconate to pyruvate, only 1 NADH molecule per gluconate is produced, in accordance with the more oxidised state of gluconate compared to glucose. Since the isopropanol production pathway is able to regenerate 1 $NADP^+$ molecule per produced isopropanol molecule (used to regenerate 1 NAD^+ by transhydrogenases), only with gluconate as the carbon source a stoichiometric regeneration of NAD^+ would be achieved.

Nevertheless, in spite of the use of gluconate as the carbon source, only a very limited anaerobic growth was observed for AL cells transformed with the MP_{Lib}, reaching a maximum OD_{600} of 0.05, after which cells died. It was hypothesized that this could be an artifact arising from the fact that the library selection culture is actually a co-culture of functionally different variants able to communicate metabolically thanks to the permeability of cellular membranes to acetone. Specifically, clones with incomplete variants of the pathway able to reduce acetone, but lacking one or more of the enzymes required to produce acetone, would be unable to grow autonomously. However, they could uptake and reduce acetone diffused from cells containing a pathway with all required enzymes for the production of acetone. This would

lead to a depletion of the acetone pool, preventing further growth of cells with a fully-functional isopropanol production pathway.

I aimed to overcome this issue by spatially isolating clones and reducing the diffusion rate of acetone. In order to do so, an agar-based solid culture medium was used. Indeed, when AL cells transformed with the MP_{Lib} were plated in anaerobic M9-agar with gluconate, and incubated anaerobically, colonies were observed after 65 h (**Figure 7.4**). A set of 10 individual colonies were then picked and inoculated in anaerobic liquid M9 with gluconate. After eight days, growth was observed in two of the cultures, with the rest of cultures growing after a period of 10 to 15 days.

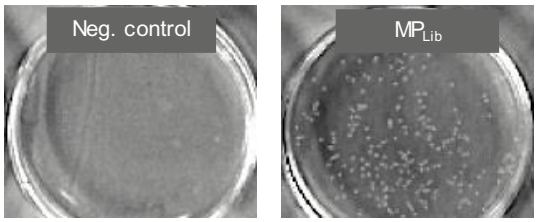


Figure 7.4. Selection of variants of the isopropanol production pathway. Cells transformed with the library of variants of the isopropanol production pathway were able to grow when plated in anaerobic M9-agar with gluconate as the carbon source and incubated anaerobically (right panel). No growth was observed for cells transformed with an empty pLv2 plasmid without any functional genes, used as the negative control (left panel).

^{1}H-NMR spectra of the fermentation broth of the cultures of the two spatially isolated clones that grew the fastest in liquid medium were acquired, confirming the presence of isopropanol (**Table 7.1**). Surprisingly, in one of the cultures only small amounts of isopropanol were detected, but large amounts of propionate were present (**Table 7.1**), a metabolite not natively produced by *E. coli* as a glucose fermentation product (Clark, 1989). Nevertheless, a four-gene native operon encoding a potential cobalamin-dependent propionate production pathway has been identified in the genome of *E. coli*. The operon, named *sbm* (sleeping beauty mutase) operon and initially thought to be silent *in vivo*, contains the following genes: *sbm* (encoding a methylmalonyl-CoA mutase), *ygfD* (a GTPase of unknown function), *ygfG* (a methylmalonyl-CoA decarboxylase) and *ygfH* (a propionyl-CoA:succinate CoA transferase) (**Figure 7.5**). While originally proposed as a potential pathway for the production of propionate from glucose in *E. coli*, the ability of the enzymes encoded in the *sbm* operon to produce propionate on their own in minimal medium with sugars as the only carbon source has been later questioned due to the inability of the methylmalonyl-CoA mutase encoded by *sbm* to produce the (S) isomer of methylmalonyl-CoA (Gonzalez-Garcia *et al.*, 2017). Instead, it has been observed to produce only the (R) isomer from its activity on succinyl-CoA, which is not

accepted by the methylmalonyl-CoA decarboxylase encoded in *ygfG* as the substrate (Dayem *et al*., 2002).

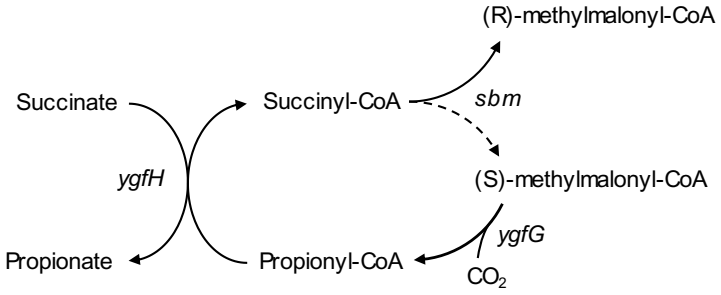


Figure 7.5. Reactions catalysed by the enzymes encoded in the *sbm* operon. *ygfH* encodes a propionyl-CoA:succinate CoA transferase, which can transfer CoA from propionyl-CoA to succinate to yield propionate and succinyl-CoA. *sbm* encodes a methylmalonyl-CoA mutase, able to catalyse the conversion of succinyl-CoA into methylmalonyl-CoA. In previous studies, it was found to produce only (R)-methylmalonyl-CoA, and not the S isomer. *ygfG* encodes a methylmalonyl-CoA decarboxylase able to decarboxylate (S)-methylmalonyl-CoA, producing propionyl-CoA and CO_2. The set of enzymes would be able to produce propionate in cells cultured in minimal medium if the methylmalonyl-CoA mutase encoded by *sbm* was able to produce (S)-methylmalonyl-CoA (dashed line).

The generation of propionic acid as a glucose fermentation product would thus require the activity of a methylmalonyl-CoA epimerase able to convert the (R) isomer of methylmalonyl-CoA to the (S) isomer. While such an enzyme has been found in native producers of propionate like *Propionibacterium acidipropionici* (Guan *et al*., 2014), it has not been identified in the *sbm* operon or elsewhere in the genome of *E. coli,* requiring the expression of a heterologous methylmalonyl-CoA epimerase to achieve the production of propionate in minimal media. It has been observed, however, that *E. coli* is able to produce propionate in complex media, potentially due to the ability to obtain (S)-methylmalonyl-CoA through the degradation of amino acids (Haller *et al*., 2000).

The observation of propionate in the fermentation broth of AL cells grown in M9 minimal medium supplemented with gluconate and carrying variants of the isopropanol production pathway is thus intriguing. It is possible that the native *sbm* methylmalonyl-CoA mutase is able to produce small amounts of (S)-methylmalonyl-CoA under the conditions of high succinate production due to the blockage of traditional alcoholic and lactic fermentations. Alternatively, an unidentified, more complex pathway requiring additional steps could be functioning to generate the required (S)-methylmalonyl-CoA. Finally, it cannot be ruled out that further adaptations have taken place, such as mutations in the *sbm* gene leading to a gain of the ability to produce (S)-methylmalonyl-CoA.

Table 7.1. Quantification of metabolites by ^{1}H-NMR of the fermentation broth of two spatially separated clones of AL cells transformed with the library of isopropanol production pathways. Spectra were acquired for the two clones spatially isolated in solid medium that grew the fastest when transferred to liquid medium.

Plasmid	Encoded enzymes	Resulting product	Product δ * (ppm)	[Acetone] (mM)	[Isopropanol] (mM)	[Ethanol] (mM)	[Lactate] (mM)	[Succinate] (mM)	[Acetate] (mM)	[Formate] (mM)	[Propionate] (mM)
pLv2	-	Isopropanol	1.18 (d, 6)	0	0	0	0	0	0	0	0
$pLS60_{Clone1}$	AtoB/AtoA/AtoD/ $ADC/CBADH_{WT}$	Isopropanol	1.18 (d, 6)	0	3.45	0	0	13.2	13.2	5.0	0
$pLS60_{Clone2}$	AtoB/AtoA/AtoD/ $ADC/CBADH_{WT}$	Isopropanol	1.18 (d, 6)	0	0.1	0	0	0.5	1.9	0.8	9.1

* δ of characteristic signal of product (ppm). The multiplicity of each signal (s – singlet, d – doublet, t – triplet, q – quartet, m – multiplet) and the number of contributing protons is shown between brackets.

7.2.2. Characterization of selected isopropanol production pathway variants

In order to determine the effectiveness of the redox rescue selection system in selecting optimal isopropanol production pathways, the sequence of the variants contained by the two spatially isolated clones that grew the fastest when transferred to anaerobic liquid cultures, the variants present in eight additional colonies selected in anaerobic M9-gluconate-agar plates as well as ten random variants of the library was determined.

The two clones that grew the fastest in anaerobic liquid cultures were found to contain the same variant of the isopropanol production pathway, Metabolic Pathway Selected 1 (MP_{S1}) (**Table 7.2**). Furthermore, the rest of variants selected in plates contained either the same MP_{S1} variant or a different Metabolic Pathway Selected 2 (MP_{S2}) variant, with no other combination of promoters and RBS detected (**Table 7.2**). Interestingly, a feature shared both by MP_{S1} and MP_{S2} is the presence of different sets of regulatory elements for *atoA* and *atoD*, which encode respectively the β and α subunits of the acetoacetyl-CoA transferase complex (Sramek and Frerman, 1975). Given the presence of two copies of each subunit in the functional complex (Sramek and Frerman, 1975), in principle it might seem logical to have the same set of regulatory elements for both genes in order to obtain an optimal stoichiometry. Nevertheless, it is possible that the number of copies of each subunit in the cell is influenced by other factors not accounted for simply by the promoter and RBS, such as the context dependence of RBS, different codon usage, folding efficiency or propensity to degradation. Alternatively, since both genes are also endogenously present and expressed in the *E. coli* genome, the combination of copies expressed from both the native genes and the transformed plasmid, and not simply the amount of each gene expressed from the plasmid, might lead to an optimal stoichiometry. It is also worth to note that the set of regulatory elements present in the selected pathways do not correspond to elements associated to the highest transcription and translation levels. Very likely, the selected combinations instead lead to finely tuned expression levels that allow for an optimized flux across the metabolic pathway, while at the same time minimizing the disruption of essential pools of native metabolites.

Table 7.2. Combinations of promoters and RBS found in the selected variants of the isopropanol production pathway. The sequence of all promoters and RBS can be found in Table 2.8 of Chapter 2.

Gene	*atoB*		*atoD*		*atoA*		*adc*		*cbadh*	
	P	RBS	P	RBS	P	RBS	P	RBS	P	RBS
MP- S1	J23116	33	J23113	42	J23102	36	J23102	42	J23113	58
MP- S2	J23116	13	J23113	13	J23116	36	J23100	58	J23113	42

On the other hand, the combination of regulatory elements present in random variants of the library were not found to follow a clear pattern, with no single combination of parts occurring more than once in the chosen set of 10 random variants. Furthermore, among the randomly-picked variants, the presence of defective variants lacking one or more of the genes was confirmed (**Figure 7.6**), revealing that the selective pressure had acted at two different levels. Firstly, defective variants with one or more absent or inactive genes had been eliminated. Secondly, a strong preference for a reduced number of promoters and RBS was observed in the selected variants, indicating that specific combinations leading to levels of expression for each enzyme that maximize the production of isopropanol had been selected.

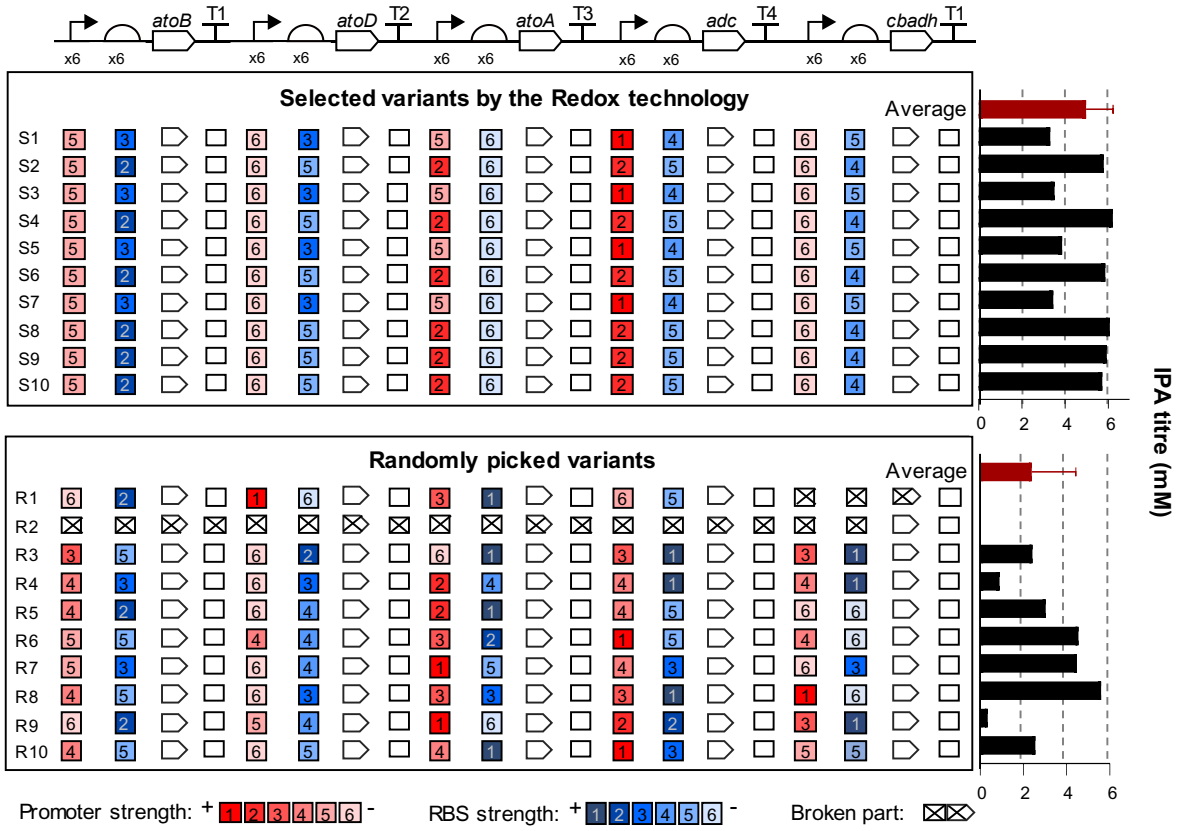


Figure 7.6. Comparison of selected and randomly-picked variants of the isopropanol production pathway. Only two different combinations of promoter and RBS were identified in selected variants, while randomly-picked variants contained much more variability and no repeated combination was found among the ten randomly-picked variants that were analysed. Additionally, some of the randomly-picked variants contained variants lacking one or all of the genes encoding the enzymes of the pathway, while all selected variants contained fully functional pathways. Finally, selected variants led to significantly higher titres of isopropanol than randomly-picked variants.

The isopropanol titre obtained by all of the selected and random variants when wild-type *E. coli* cells were transformed with them and cultured aerobically in LB medium was quantified by acquiring ^{1}H-NMR spectra of the culture broth 17 h after inoculation. Aerobic conditions

were used for the comparison of isopropanol titres because although the selected variants would be expected to grow anaerobically, the randomly-picked variants might not. An average of 4.97 mM isopropanol concentration was obtained with selected variants, while random variants produce on average 2.39 mM, or 2.99 mM without considering variants with incomplete pathways, resulting either way in a significantly higher production of isopropanol by selected variants as demonstrated by the results of a t-test (when incomplete variants were included, $t(9) = 3.39$ and $p = 0.0016$; when excluding them, $t(9) = 2.72$ and $p = 0.0076$) (**Figure 7.6**).

Finally, the performance of the best selected variant (MP_{S1}) was compared with that of the original pathway developed by Hanai *et al*. under the same culture conditions. When providing glucose as the main carbon source, the maximum theoretically achievable yield is one isopropanol molecule per glucose, if all the glycolytic flux was completely coupled to the isopropanol production pathway. Hanai *et al*. employed rich culture medium supplemented with 111 mM glucose (SD8 medium) under aerobic conditions to maximize the isopropanol titre. They obtained a maximum titre of 44.8 mM isopropanol with a consumption of 103 mM of glucose, resulting in a yield of 43.5%. The isopropanol titre achieved after wild-type *E. coli* cells were cultured aerobically for 20 h in the same medium was measured, obtaining an average titre of 61.9 mM isopropanol (**Table 7.3**). Assuming all glucose was consumed, the corresponding yield is 56%, the highest isopropanol yield ever reported for any organism, including native producers.

Table 7.3. Quantification of metabolites by ^{1}H-NMR of the culture broth of wild-type *E. coli* cells transformed with the best selected variant of the isopropanol production pathway and cultured aerobically in rich medium. Three independent transformations were cultured in SD8 medium and analysed after 20 h of aerobic growth.

Plasmid	Encoded enzymes	Resulting product	Product δ * (ppm)	[Acetone] (mM)	[Isopropanol] (mM)	[Ethanol] (mM)	[Lactate] (mM)	[Succinate] (mM)	[Acetate] (mM)	[Formate] (mM)
pLv2	-	Isopropanol	1.18 (d, 6)	0	0	0	0	0	0	0
MP_{S1}	AtoB/AtoA/AtoD/ $ADC/CBADH_{WT}$	Isopropanol	1.18 (d, 6)	7.1	67.0	0	0	2.9	4.4	5.8
MP_{S1}	AtoB/AtoA/AtoD/ $ADC/CBADH_{WT}$	Isopropanol	1.18 (d, 6)	13.1	58.4	0	0	0.5	1.4	2.1
MP_{S1}	AtoB/AtoA/AtoD/ $ADC/CBADH_{WT}$	Isopropanol	1.18 (d, 6)	14.6	60.4	0	0	2.6	2.9	1.7

* δ of characteristic signal of product (ppm). The multiplicity of each signal (s – singlet, d – doublet, t – triplet, q – quartet, m – multiplet) and the number of contributing protons is shown between brackets.

7.3. Conclusions

The construction and optimization of heterologous metabolic pathways allowing for a maximized yield of a desired product is a non-trivial task involving different levels of design which pertain not only to the constituent genes themselves, but also to their arrangement and the regulatory elements controlling their expression. Furthermore, the heterologous pathway should not excessively disrupt the native essential metabolic processes and pools of endogenous metabolites. Due to the difficulty in predicting the global effects of the introduction of a heterologous metabolic pathway, finding the optimal setup for a given pathway becomes a challenging task, often requiring intensive and time-consuming manual screening of multiple variants because of the lack of selection methods generally applicable to metabolic pathways.

As a test case, the redox rescue selection system was applied to obtain an improved variant of a previously described isopropanol production pathway with optimized combinations of regulatory elements. A variant enabling the generation of the most efficient isopropanol-producing organism reported so far was selected, demonstrating the large margin of optimization available in metabolic pathways for the production of specific compounds by simply altering the arrangement and regulatory elements of well described functional components.

The implication for such a result is two-fold. Firstly, it proves the suitability of the redox rescue selection system to more complex systems containing several genes, such as metabolic pathways. Additionally, it demonstrates its ability to select improved variants not only of enzymes or other proteins, but in general of any biomolecule or genetic element exhibiting any property that can be indirectly linked to the generation of NAD^+, such as regulatory sequences. These features make the selection system an attractive novel tool suitable for the selection of a wide variety of metabolic pathways, and particularly for those including a direct step where NAD(P)H is consumed. In such pathways, large amounts of redox cofactor to maximize the flux across the pathway must be provided, while at the same time maintaining the redox and energetic cellular balances, making the redox rescue selection system an ideal choice for their optimization.

7.4 Key outcomes

- An optimized isopropanol production pathway has been selected from a library of variants differing in the regulatory elements controlling the expression of the genes composing a previously described pathway.

- The best selected variant enabled the production of the highest isopropanol yield ever reported for any organism.
- The redox rescue selection system is suitable for the selection of genetic regulatory elements, as well as complex systems comprising several biomolecules.

8. Discussion and future perspectives

Biocatalysis is now established as a solid component of the toolkit available for industrial manufacturing processes, offering in many cases a more efficient, sustainable and cheaper alternative to traditional chemical catalysis (Itoh and Hanefeld, 2017). Nevertheless, it is often found that the set of available biocatalysts does not include an enzyme with the optimal properties for a specific application (Osbon and Kumar, 2019). New, better suited enzymes can then be discovered through a search of the repertoire of naturally-occurring biomolecules. Alternatively, variants with altered or improved properties can be developed from already known enzymes through protein engineering.

The properties of an enzyme depend largely on its three-dimensional structure, which is ultimately determined by its sequence of amino acids with a contribution from the environment in which folding occurs (Macdonald and Johnson, 2001). However, despite recent improvements, the methods for the prediction of protein structure are still unable to consistently provide correct approximations of the tertiary structure of a given protein from its primary structure. This is particularly true for the modelling of side chains, absolutely required for predicting the structure of the active site and making predictions about the role of specific residues. Even though algorithms for the prediction of protein structure through template-based modelling are becoming increasingly effective at predicting the general conformation of the backbone when only the structure of a distantly related homologue is available, the prediction of more detailed features is still heavily dependent on the availability of a closely related homologue, and even then it is still far from perfect as evidenced by the results of the latest iteration of the Critical Assessment of protein Structure Prediction (CASP), CASP13 (Croll *et al.,* 2019). The situation only gets worse when not even a remotely related structure that can be used as a template is not available. Nevertheless, a surprising improvement was provided by the recently developed AlphaFold algorithm, based on deep learning, which achieved an accuracy of almost 60% on average for the prediction of gross protein topology. However, its accuracy for the prediction of finer details of protein structure remained around 40% (AlQuraishi, 2019). Furthermore, the structure-function relationship remains poorly understood in many cases, making it challenging to rationally develop a novel protein variant with the desired properties even when structural information is available or can be obtained through protein structure prediction methods. Most methods for the prediction of function from structure rely on the comparison of structural features between the target protein and other proteins of known function. Improvements have been observed when similarity between local structural motifs are considered instead of global similarities in the overall folds, but even with such an approach only around 50% of the annotated molecular functions were correctly

predicted. The binding to specific molecules was found particularly hard to be correctly predicted and associated to a set of structural motifs (Hvidsten *et al.*, 2009).

Directed evolution offers an attractive alternative to rational design, since, instead of determining a reduced set of mutations potentially having the desired effect, a large range of mutations is instead tested, mimicking the process of natural evolution in the laboratory at a much faster pace (Arnold, 2018). Methods based on artificial selection pressures where the target property to be evolved in a given biomolecule is coupled to cell growth offer one of the most promising approaches, thanks to their ability to perform ultra-high throughput searches of the sequence space in an automated manner and without requiring extensive human intervention. However, the applicability of the methods developed so far has been limited to a rather narrow range of biomolecules or properties (Xiao, Bao, and Zhao, 2015).

In the present work, developed a novel selection method based on the regeneration of the universal metabolite NAD^+ has been developed, providing the broadest applicable artificial selection method ever reported (**Figure 8.1**). The usefulness of the redox rescue selection system has been demonstrated by applying it to a series of real cases of industrial interest, which are also each representative examples of broad classes of challenges to which this approach could be successfully applied.

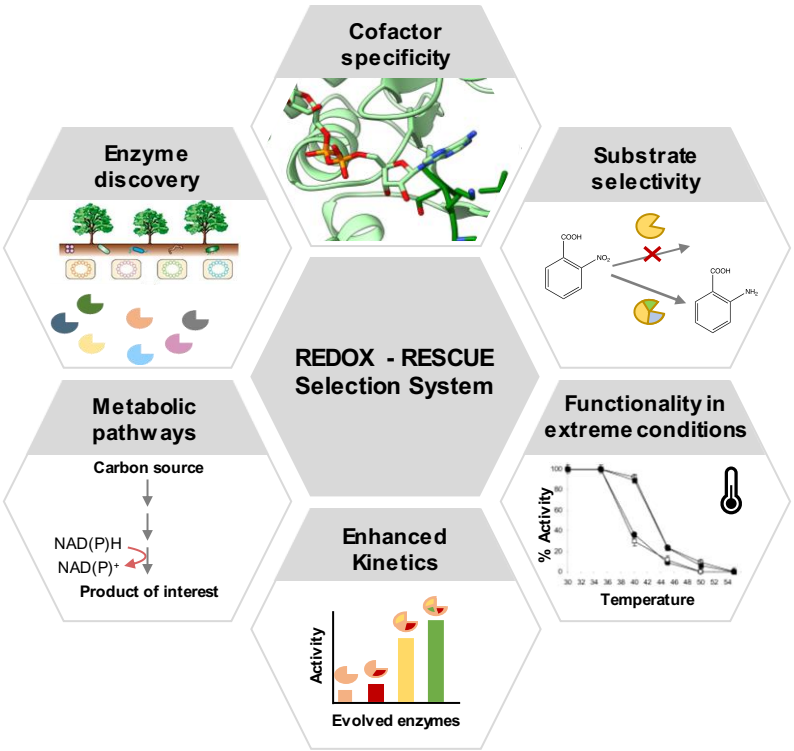


Figure 8.1. Overview of some of the applications of the redox rescue selection system. The redox rescue system allows for the selection of a wide range of biomolecules and properties of interest. The system could also be potentially applied to the discovery of novel natural enzymes from libraries constructed from metagenomic samples instead of by mutagenesis of a single parental sequence.

As the first test case, the ability of the system to select NAD-dependent variants of CBADH and TBADH, secondary alcohol dehydrogenases exhibiting a very strong preference for NADP, was demonstrated. While previous attempts at reversing the cofactor specificity of these enzymes failed (Maddock *et al.*, 2015), the redox rescue selection system enabled the fast identification of such variants by simply transforming the selector strains with a library of variants and culturing them under anaerobic fermentation conditions in medium supplemented with acetone, one of the substrates of the alcohol dehydrogenases. The efficacy of the system was further evidenced by the fact that one of the selected variants displayed the highest relative catalytic efficiency to date in a case of reversal of cofactor preference from NADP to NAD, where usually large loses of activity are observed, and the second highest for any case of reversal of cofactor preference (Chánique and Parra, 2018). Additionally, a variant with an unexpected insertion has been identified. Such insertion was not purposely introduced during the mutagenesis procedure, and thus must have occurred at an extremely low rate, proving the potential of the system to select highly underrepresented variants. Surprisingly, such a variant was able to use both NAD and NADP efficiently, providing a potentially useful tool for whole-cell biocatalysis given its ability to benefit from both the NAD and NADP redox equivalent pools.

I then aimed to demonstrate how the redox rescue selection system can outperform other screening-based approaches by using it to select an NAD-dependent variant of *Myxococcus stipitatus* IRED. IREDs catalyse the asymmetric reduction of imines and the reductive amination of ketones, yielding in both cases chiral amines, one of the most important building blocks in the pharmaceutical and agricultural industries (Schrittwieser, Velikogne, and Kroutil, 2015). Nevertheless, all known natural IREDs are NADP-dependent, thus granting a special interest in developing NAD-dependent variants (Gand *et al.*, 2016). Within a single round of mutagenesis and selection, such a variant of *Myxococcus stipitatus* IRED was obtained. The selected variant was able to perform better than the best variant identified in a previous study, where up to three rounds of mutagenesis of a larger set of residues and manual screening were required (Borlinghaus and Nestl, 2018). Additionally, a relief in substrate inhibition was observed, evidencing the ability of the redox rescue selection system to optimize multiple kinetic parameters simultaneously. This could be of particular interest for the development of novel biomolecules as part of whole-cell biocatalysts, where multiple types of variations in the properties of an enzyme can make it more suitable for performing optimally in the cellular context, including reduced inhibition by the substrate itself but also by other metabolites present in the cells. It is almost certain that even better variants exist, but can only be accessed by exploring a larger sequence space. While such a deep exploration of mutagenic space cannot be performed in a reasonable amount of time with traditional screening-based

methods, the redox rescue selection system is certainly a suitable tool for the task. Even in cases where the number of variants to be sampled is much larger than can be reasonably achieved due to the transformation efficiency of *E. coli*, a smart library approach could be taken to alleviate this problem, such as the generation of multiple libraries, each one targeting subsets of the residues of interest. Such a “libraries of libraries” approach would scale up better than simply creating a library targeting all the residues of interest simultaneously, which might require an unreasonably number of independent transformations to sample a large enough number of variants.

As the next proof-of-concept, I sought to obtain variant enzymes with altered substrate selectivity, a critical property that is often required to be engineered when developing novel biocatalysts for industrial applications. Because of their relevance for a wide variety of applications ranging from novel cancer therapies to bioremediation, in addition to their potential for chemical synthesis (Bai *et al.*, 2015), nitroreductases were chosen as a test case. Novel variants of *Enterobacter cloacae* NfsB nitroreductase were developed using 2-NBA and 4-nitrobenzyl alcohol as the target substrates. The wild-type enzyme does not display any activity towards 2-NBA, while it is able to reduce 4-nitrobenzyl alcohol. In both cases, the selected variants displayed a gain of activity or improved activity towards the chosen substrate, proving the applicability of the redox rescue selection system in two different situations. Firstly, it can be applied to obtain variant enzymes able to act on novel substrates of synthetic origin which are not accepted by native enzymes. This is of high relevance for the development of nitroreductase catalysts for the efficient production of specific aylhydroxylamines (and possibly aromatic amines also), important intermediates in the synthesis of pharmaceuticals and other synthetical processes (Smith *et al.*, 1999; Bai *et al.*, 2015; Ho and Lau, 2000), particularly for cases where several nitro groups are present in the precursor molecule. By applying the selection system, it is possible to obtain nitroreductase variants able to act on nitro groups of the target molecule that would not be reduced by the naturally available enzymes. Additionally, if a native enzyme displays poor activity towards the substrate of interest, the system can be used to identify a variant with improved activity. This application mode could be of special interest for the field of nitroreductase-mediated cancer therapies, where often a promising prodrug candidate is available but natural nitroreductases are not able to convert it to the active drug efficiently enough. For example, the PR104 prodrug, belonging to the group of dinitrobenzamide mustards, has been shown to have several advantages over CB1954, such as a better bystander effect and a higher potency. However, its clinical progression has been limited by the fact that it is unexpectedly reduced under aerobic conditions by human aldo-ketoreductase 1C3, which is expressed in myeloid progenitor cells. The dose limitation derived from the resulting myelotoxicity means a highly

active nitroreductase, able to catalyse its reduction even at very low concentration, is required, which is not currently available (Williams *et al.*, 2015).

Finally, the scope of the redox rescue selection system was expanded by using it to select optimal multienzymatic isopropanol production pathways differing in the combination of regulatory elements controlling the expression of each of its genes. By succeeding in selecting optimal combinations of promoters and RBS within the chosen arrangement, I confirmed the applicability of the selection system not only to complex systems comprising several biomolecules where only one of them is able to regenerate $NAD(P)^+$, but also to biomolecules different than enzymes and properties not necessarily related to catalysis, such as promotion of transcription or translation. Previous selection methods have been limited to the identification of enhanced variants of single biomolecules, and the exploration of different arrangements or regulatory elements of the functional genes comprising a pathway has been performed mostly through screening techniques, which require the application of strategies to reduce the number of variants and limit the sequence space that can be sampled (Jeschek, Gerngross, and Panke, 2017). The redox rescue selection system provides a useful platform for the identification of optimal pathways without the need to make any *a priori* assumption about the suitability of different regulatory elements or arrangements which could compromise the ability to find the best performing variants.

It is easy to devise many other cases where the redox rescue selection system could be applied by linking in an indirect way the activity of the biomolecule of interest to the regeneration of NAD^+. The most straightforward approach would involve the use a secondary enzymatic step able to regenerate oxidised $NAD(P)^+$ by reducing the product resulting from the reaction catalysed by the enzyme being selected. Through such a selection mode, already demonstrated by the selection of a variant of an isopropanol production pathway, the scope of enzymes amenable to evolution by means of the redox rescue selection method is further expanded beyond the already wide group of NAD(P)-dependent oxidoreductases (**Figure 8.2**).

By using different arrangements, biomolecules other than enzymes acting on small molecules can be selected. For example, to select a variant of an RNA-polymerase able to initiate transcription more efficiently from a specific promoter, it would suffice to place an NAD(P)-dependent oxidoreductase under the control of such a promoter and transform cells with a library of the RNA-polymerase (**Figure 8.2**). A similar strategy could be applied to select other biomolecules able to promote in any way the expression of a gene encoding an NAD(P)-dependent oxidoreductase, such as DNA recombinases, as well as the corresponding genetic regulatory elements. Furthermore, such a setup could also be applied to select biomolecules

exhibiting specific binding properties by combining it with a two-hybrid-like system, with the only difference that the reporter gene would be replaced by a gene encoding an NAD(P)-dependent oxidoreductase. Finally, the system is also suitable for the selection of transporters, provided that the imported substrate can be coupled to the regeneration of oxidised $NAD(P)^+$ either directly by means of an NAD(P)-dependent oxidoreductase acting on the supplied substrate or indirectly by a set of several enzymes (**Figure 8.2**). An additional requisite would be that the cellular membrane is, at least partially, impermeable to the substrate. Alternatively, it might also be possible to use the selection system to select transporters able to export the reduced products generated by the activity of the reductase directly responsible for $NAD(P)^+$ regeneration, since its efficient removal from the interior of cells would shift the equilibrium of the reaction catalysed by the reductase towards the generation of more oxidised cofactor, thus allowing for better growth recovery (**Figure 8.2**).

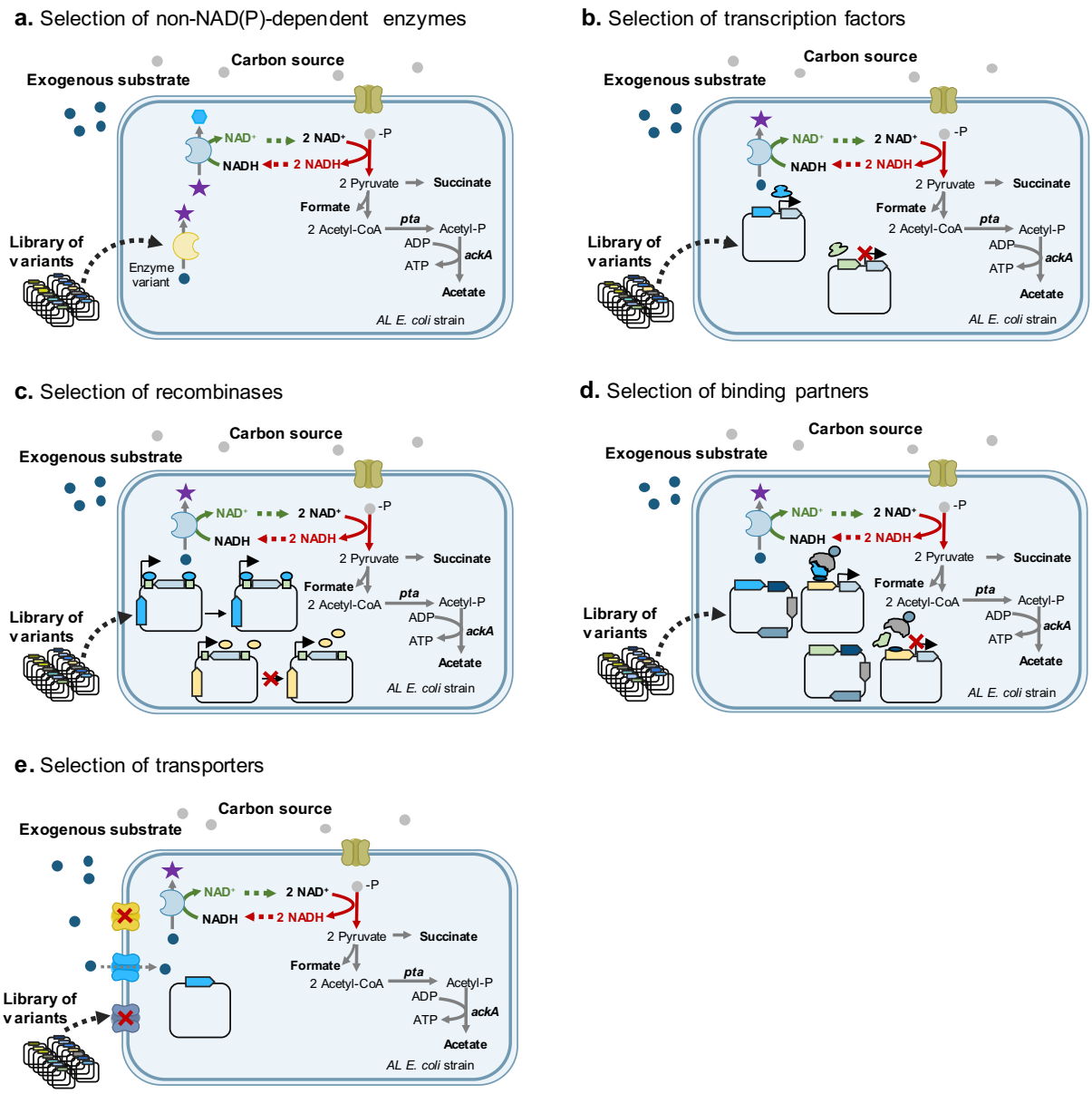


Figure 8.2. Expansion of the application scope of the redox rescue selection system. a. Enzymes other than NAD(P)-dependent oxidoreductases could be selected by coupling their activity to the regeneration of $NAD(P)^+$ by means of one or more intermediary reactions. **b.** Transcription factors, such as RNA polymerases, could be selected by placing an NAD-dependent oxidoreductase under the control of regulatory elements activated by the transcription factor. A similar approach can be used to select for variants of the regulatory elements themselves, instead of the transcription factor. **c.** Recombinases could be selected by placing an NAD-dependent oxidoreductase flanked by inverted repeated recombination sequences in a wrong arrangement that would not allow for the correct transcription and expression of the NAD-dependent oxidoreductase unless the recombinase was able to invert the gene by means of recombination. Other arrangements could be used for other types of enzymes acting on DNA. **d.** Biomolecules with specific binding properties could be selected by using a two-hybrid system where only upon interaction of the biomolecule of interest with a target interactor the transcription of an NAD-dependent oxidoreductase would be activated. **e.** Transporters could also be selected, provided that the cellular membrane is, at least partially, impermeable to the compound of interest. Cells would also be required to be transformed with an enzyme (or a set of them) able to act on the substrate to produce $NAD(P)^+$. Alternatively, it would also be possible to select transporters able to export the product generated together with $NAD(P)^+$, since its removal from the cell would shift the equilibrium for the reaction catalysed by the oxidoreductase directly responsible for $NAD(P)^+$ regeneration towards the production of oxidised cofactor.

The universality of the metabolic defect lying at the core of the redox rescue selection system provides even more opportunities to expand its scope. Indeed, a large variety of facultative anaerobe microorganisms are able to grow anaerobically by performing different fermentation patterns, including the mixed-acid fermentation exhibited by *E. coli* and others forming a variety of end products with the goal of maintaining a constant supply of ATP by means of substrate-level phosphorylation (Müller, 2001). Of special interest are fermentative extremophiles, such as thermophiles (Krahe *et al.*, 1996). Given the ability of the latter to grow at higher temperatures than mesophilic organisms such as *E. coli*, a strain functionally analogous to AL and ALPS *E. coli* strains (although not necessarily with the same mutations) but generated from a thermophilic organism would provide a useful platform for the selection of thermostable biomolecules by following a similar methodology to the one presently described, with the only difference that cultures for metabolic complementation should be performed at high temperatures. The generation of such a strain is being currently pursued by using *Geobacillus thermoglucosidasius* as the parental organism, which is able to perform, as *E. coli*, a mixed-acid fermentation under anaerobic conditions, relying mostly on the production of ethanol and lactate as fermentation products to regenerate oxidised NAD^+ under anaerobic fermentation conditions (Cripps *et al.*, 2009; Zhou, Wu, and Rao, 2016).

Another attractive feature of growth complementation-based selection methods is the potential opportunity of coupling them to an *in vivo* mutagenesis method to provide a completely automated continuous evolution system. A successful example was provided with the development of the PACE methodology, although its scope was limited to the selection of proteins whose activity could be linked to the expression of the pIII gene, such as polymerases, recombinases and proteins able to interact with a target binding partner (Esvelt *et al*., 2011). To that extent, *E. coli* offers the advantages of being a key model organism for which a large variety of genetic manipulation tools have been developed. The DNA Pol I-based *in vivo* mutagenesis method conceived by Camps *et al*. (described in Chapter 1) provides a technically simple approach targeting specifically the gene of interest (Camps *et al.*, 2003). With the aim to couple it to the redox rescue selection system, a novel *E. coli* strain including the mutations found in both the AL strain and the JS200 mutator strain built by Camps *et al.* is being created. Such a strain holds the potential of becoming a completely automated continuous evolution system with a broad application scope, both in terms of evolved biomolecules and targeted properties (**Figure 8.3**).

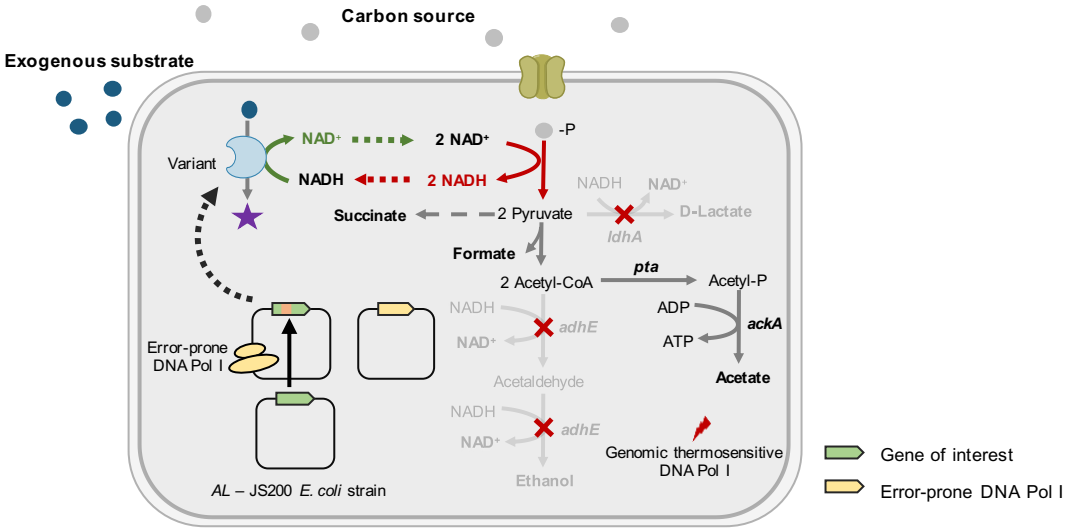


Figure 8.3. Conceptual *in vivo* continuous evolution system. It might be possible to obtain a system for the *in vivo* continuous evolution of biomolecules by combining the redox rescue selection system with the *in vivo* mutagenesis system designed by Camps *et al*. The gene of interest would be placed immediately downstream of a ColE1 replication origin. Cells would also be transformed with a plasmid containing an error-prone version of DNA polymerase I, which replicates the first kilobases of DNA downstream of ColE1 replication origins, and would therefore introduce mutations specifically in the gene of interest. The mutagenesis efficiency could be enhanced by constructing an *E. coli* strain similar to the AL strain but using the JS200 strain as the parental strain, which has a temperature sensitive allele of DNA polymerase I in its genome. Cycles of culture under aerobic conditions at a temperature high enough to inactivate the genomic, non error-prone copy of DNA polymerase I and culture under anaerobic fermentation conditions to select the best variants could then be performed, eliminating the need to generate a library of variants before the selection step.

A novel, highly versatile selection system has been developed. The versatility of the selection system can be observed at three different levels. Firstly, it can be applied to a wide range of biomolecules. Second, the repertoire of properties that can be selected with it is equally broad. Finally, it is expected to be easily ported to a variety of host organisms, enhancing even further the possibilities of the system. Thanks to these properties, the scope of the selection system is beyond that of previous selection technologies, bypassing many of the frequent limitations encountered when undertaking directed evolution of a biomolecule or a set of them. The efficacy of the selection system in a varied range of cases has been demonstrated, providing the foundational basis for its application. These results allow to anticipate that it will be adopted by a number of research groups and industrial companies in the short to medium term, revealing novel, non-obvious implementation modes.

9. Appendix

9.1 List of appendix Tables and Figures

Table A1. Protein sequence of selected CBADH, TBADH, MsIRED and EntNfsB variants 185

Table A2. Data collection parameters and refinement statistics for the crystal structures of $CBADH_S$ and $TBADH_{S1}$ 186

Figure A1. Activity assays showing lack of activity with NAD^+ or $NADP^+$ for wild-type CBADH, $CBADH_S$, wild-type TBADH and $TBADH_{S1}$ 187

Figure A2. Activity assays showing lack of activity with NADH or NADPH for wild-type MsIRED, $MsIRED_{C1}$ and $MsIRED_S$ 187

Figure A3. Activity assay showing lack of activity with 2-NBA for wild-type EntNfsB . 187

Figure A4. Characteristic ^{1}H-NMR signal used to quantify each of the analysed metabolites..... 188

9.2 Appendix data

Table A1. Protein sequence of selected CBADH, TBADH, MsIRED and EntNfsB variants.

Protein	Sequence
$CBADH_{S1}$	MKGFAMLGINKLGWIEKERPVAGSYDAIVRPLAVSPCTSDIHTVFEGALGDRKNMILGHEAV GEVVEVGSEVKDFKPGDRVIVPCTTPDWRSLEVQAGFQQHSNGMLAGWKFSNFKDGVFG EYFHVNDADMNLAILPKDMPLENAVMITDMMTTGFHGAELADIQMGSSVVVIGIGAVGLMGI AGAKLRGAGRIIGVDYRPICVEAAKFYGATDILNPKNGHIVDQVMKLTNGKGVDRVIMAGG GSETLSQAVSMVKPGGIISNINYHGSGDALLIPRVEWGCGMAHKTIKGGLCPGGRLRAEML RDMVVYNRVDLSKLVTHVYHGFDHIEEALLLMKDKPKDLIKAVVIL
$TBADH_{S1}$	MKGFAMLSIGKVGWIEKEKPAPGPFDAIVRPLAVAPCTSDIHTVFEGAIGERHNMILGHEAV GEVVEVGSEVKDFKPGDRVVVPAITPDWRTSEVQRGYHQHSGGMLAGWKFSNVKDGVF GEFFHVNDADMNLAHLPKEIPLEAAVMIPDMMTTGFHGAELADIELGATVAVLGIGPVGLMA VAGAKLRGAGRIIAVSKPPVCVDAAKYYGATDIVNVKDGPIESQIMNLTEGKGVDAAIIAGGN ADIMATAVKIVKPGGTIANVNYFGEGEVLPVPRLEWGCGMAHKTIKGGLCPGGRLRMERLI DLVFYKRVDPSKLVTHVFRGFDNIEKAFMLMKDKPKDLIKPVVILA
$TBADH_{S2}$	MKGFAMLSIGKVGWIEKEKPAPGPFDAIVRPLAVAPCTSDIHTVFEGAIGERHNMILGHEAV GEVVEVGSEVKDFKPGDRVVVPAITPDWRTSEVQRGYHQHSGGMLAGWKFSNVKDGVF GEFFHVNDADMNLAHLPKEIPLEAAVMIPDMMTTGFHGAELADIELGATVAVLGIGPVGLMA VAGAKLRGAGRIIAVHRAPVCVDAAKYYGATDIVNMKDGPIESQIMNLTEGKGVDAAIIAGRII AVASKPVCVDAAKYYGATDIVNYKDGPIESQIMNLTEGKGVDAAIIAGGNADIMATAVKIVKP GGTIANVNYFGEGEVLPVPRLEWGCGMAHKTIKGGLCPGGRLRMERLIDLVFYKRVDPSK LVTHVFRGFDNIEKAFMLMKDKPKDLIKPVVILA
$MsIRED_S$	MKPTLTVIGAGRMGSALIKAFLQSGYTTTVWEVRKARSEPLAKLGAHLADTVRDAVKRSDII VVNVLDYDTSDQLLRQDEVTRELRGKLLVQLTSGSPALAREQETWARQHGIDYLDGAIMAT PDFIGQAECALLYSGSAALFEKHRAVLNVLGGATSHVGEDVGHASALDSALLFQMWGTLF GTLQALAISRAEGIPLEKTTAFIKLTEPVTQGAVADVLTRVQQNRLTADAQTLASLEAHNVAF QHLLALCEERNIHRGVADAMYSVIREAVKAGHGKDDFAILTRFLK
$EntNfsB_{S1}$	MDIISVALKRHSTKAFDASKKLTAEEAEKIKTLLQYSPSAINSQPWHFIVASTEEGKARVAKS AAGTYVFNERKMLDASHVVVFCAKTAMDDAWLERVVDQEEADGRFNTPEAKAANHKGRT YAADMHRVDLKDDDQWMAKQVYLNVGNFLLGVGAMGLDAVPIEGFDAAILDEEFGLKEKG FTSLVVVPVGHHSVEDFNATLPKSRLPLSTIVTEC
$EntNfsB_{S2}$	MDIISVALKRHSTKAFDASKKLTAEEAEKIKTLLQYSPSSLNSQPWHFIVASTEEGKARVAKS AAGTLVFNERKMLDASHVVVFCAKTAMDDAWLERVVDQEEADGRFNTPEAKAANHKGRT YLADMHRVDLKDDDQWMAKQVYLNVGNFLLGVGAMGLDAVPIEGFDAAILDEEFGLKEKG FTSLVVVPVGHHSVEDFNATLPKSRLPLSTIVTEC

Table A2. Data collection parameters and refinement statistics for the crystal structures of $CBADH_S$ and $TBADH_{S1}$.

Beamline	$CBADH_s$ variant Diamond I03	$TBADH_{S1}$ variant Diamond I04
Wavelength	0.9762	0.9795
Resolution range	59.36 - 2.199 (2.278 - 2.199)	71.68 - 2.85 (2.952 - 2.85)
Space group	P 1 21 1	P 21 21 21
Unit cell	75.8115 99.5695 114.089 90 102.768 90	79.1234 123.946 169.24 90 90 90
Total reflections	265743 (16967)	258210 (26258)
Unique reflections	82558 (7335)	39642 (3923)
Multiplicity	3.2 (2.3)	6.5 (6.7)
Completeness (%)	98.45 (87.66)	99.87 (99.62)
Mean I/sigma(I)	7.57 (0.71)	6.92 (1.11)
Wilson B-factor	34.89	69.95
R-merge	0.09525 (1.064)	0.1918 (1.671)
R-meas	0.114 (1.345)	0.2086 (1.811)
R-pim	0.06202 (0.8095)	0.08117 (0.6942)
CC1/2	0.996 (0.345)	0.991 (0.554)
CC*	0.999 (0.716)	0.998 (0.844)
Reflections used in refinement	82766 (7315)	39597 (3913)
Reflections used for R-free	4037 (357)	2015 (182)
R-work	0.1639 (0.2548)	0.1746 (0.3201)
R-free	0.2068 (0.2927)	0.2390 (0.3647)
CC(work)	0.970 (0.619)	0.955 (0.750)
CC(free)	0.960 (0.587)	0.940 (0.651)
Number of non-hydrogen atoms	11354	10622
macromolecules	10666	10618
ligands	332	4
solvent	356	0
Protein residues	1417	1416
RMS(bonds)	0.007	0.009
RMS(angles)	0.93	1.06
Ramachandran favored (%)	95.87	95.10
Ramachandran allowed (%)	4.13	4.69
Ramachandran outliers (%)	0.00	0.21
Rotamer outliers (%)	0.97	0.27
Clashscore	5.11	4.48
Average B-factor	45.12	77.08
macromolecules	44.13	77.04
ligands	78.13	174.49
solvent	43.81	-
Number of TLS groups	18	15

Statistics for the highest-resolution shell are shown in parentheses.

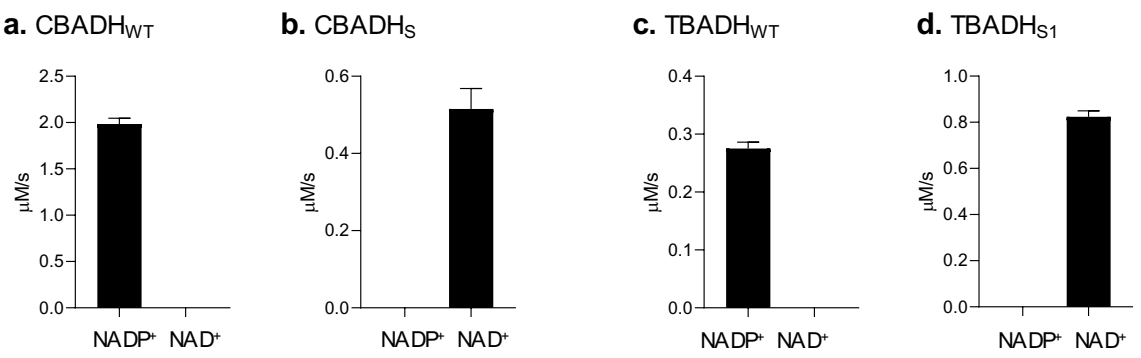


Figure A1. Activity assays showing lack of activity with NAD^+ or $NADP^+$ for wild-type CBADH, $CBADH_S$, wild-type TBADH and $TBADH_{S1}$. In each case, reactions were performed under the conditions described in Table 2.11 of Chapter 2 with the concentrations of isopropanol yielding the highest activity during enzymatic activity assays for the determination of kinetic parameters with the cofactor with which activity was detected. Wild-type CBADH and TBADH were able to catalyse the oxidation of isopropanol only with $NADP^+$, while $CBADH_S$ and $TBADH_{S1}$ could catalyse the reaction only with NAD^+. Error bars show the standard deviation.

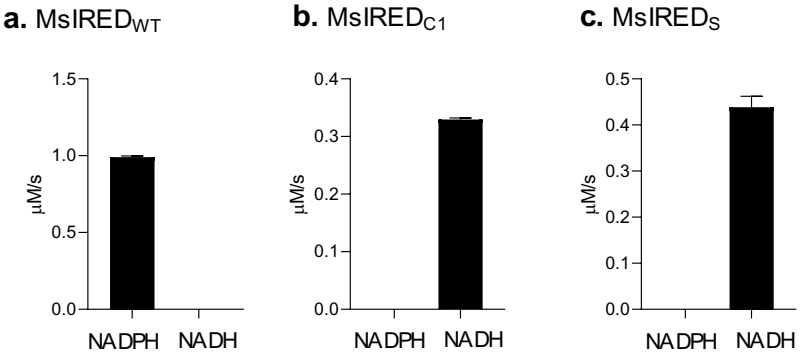


Figure A2. Activity assays showing lack of activity with NADH or NADPH for wild-type MsIRED, $MsIRED_{C1}$ and $MsIRED_S$. In each case, reactions were performed under the conditions described in Table 2.11 of Chapter 2 with the concentrations of 2-methyl-1-pyrroline yielding the highest activity during enzymatic activity assays for the determination of kinetic parameters with the cofactor with which activity was detected. Wild-type MsIRED was able to catalyse the reduction of 2-methyl-1-pyrroline only with NADPH, while $MsIRED_{C1}$ and $MsIRED_S$ could catalyse the reaction only with NADH. Error bars show the standard deviation.

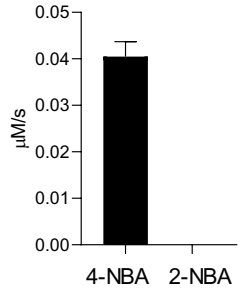
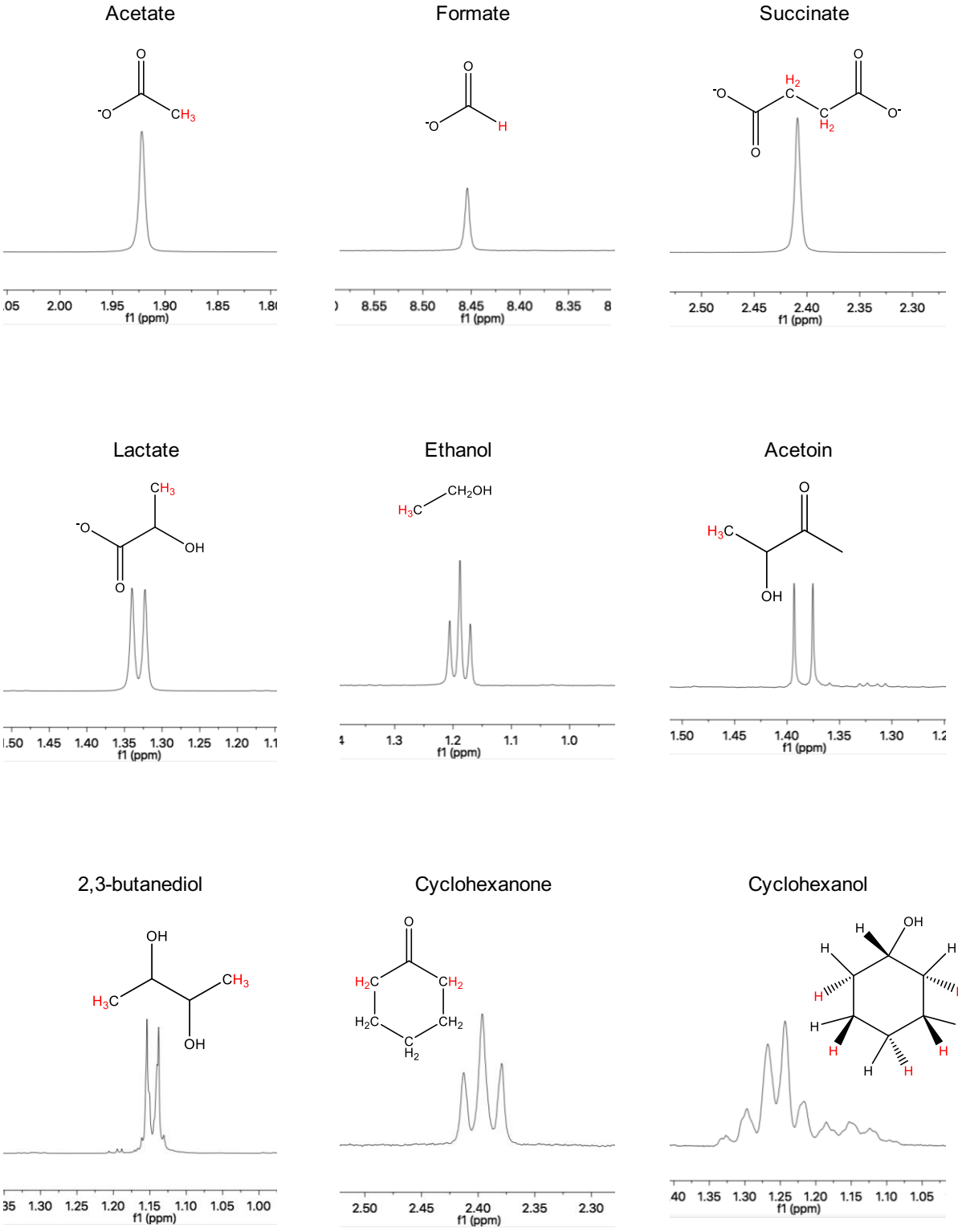
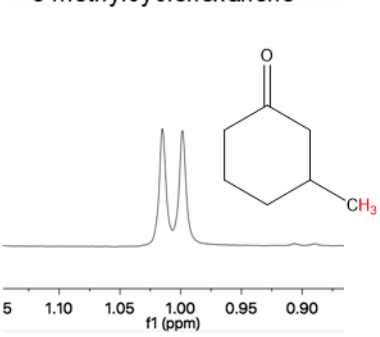


Figure A3. Activity assay showing lack of activity with 2-NBA for wild-type EntNfsB. The reaction was performed under the conditions described in Table 2.11 of Chapter 2 with the concentrations of 4-NBA or 2-NBA yielding the highest activity during enzymatic activity assays for the determination of kinetic parameters of wild-type EntNfsB or $EntNfsB_{S1}$, respectively. Error bars show the standard deviation.

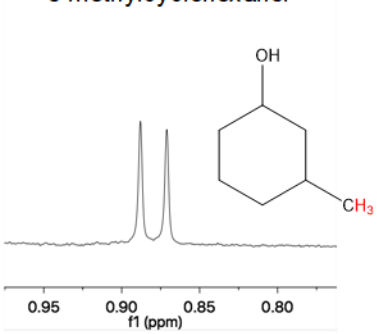
Figure A4. Characteristic ^{1}H-NMR signal used to quantify each of the analysed metabolites.



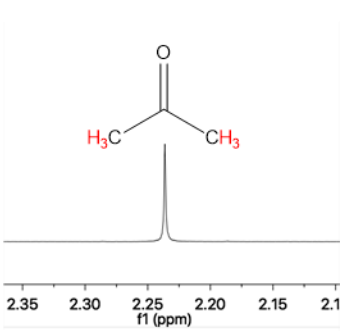
3-methylcyclohexanone



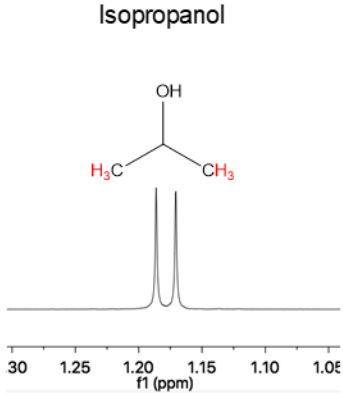
3-methylcyclohexanol



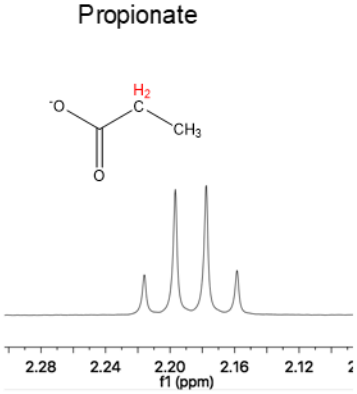
Acetone



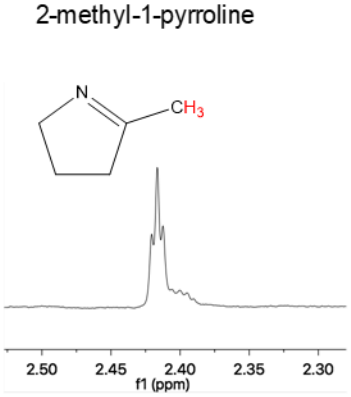
Isopropanol



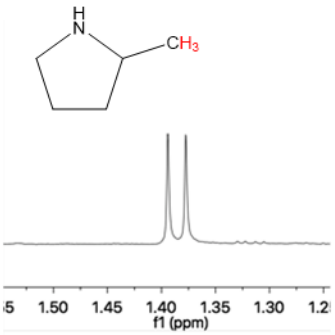
Propionate



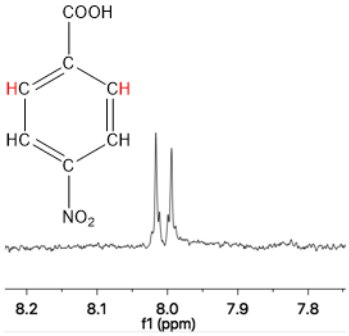
2-methyl-1-pyrroline



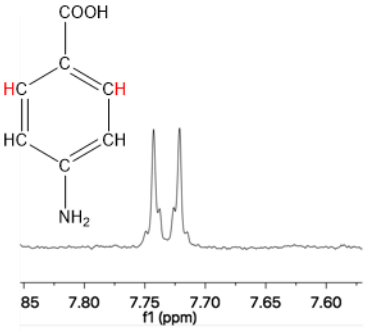
2-methylpyrrolidine



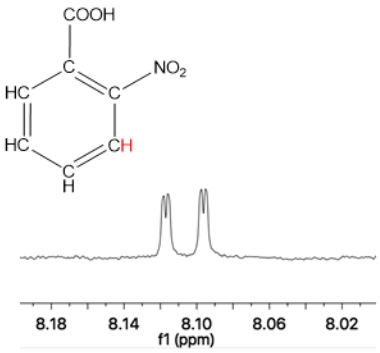
4-nitrobenzoic acid



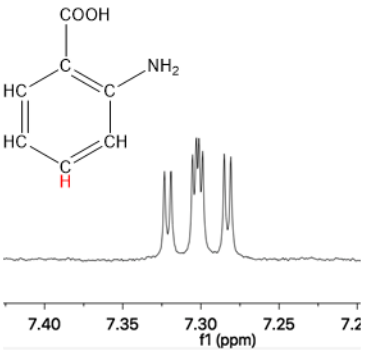
4-aminobenzoic acid



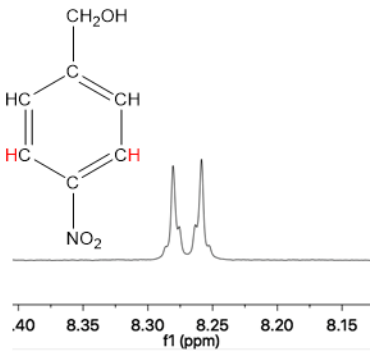
2-nitrobenzoic acid



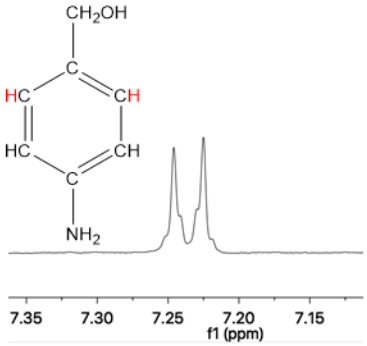
2-aminobenzoic acid



4-nitrobenzyl alcohol



4-aminobenzyl alcohol



Bibliography

Adams, Paul D., Pavel V. Afonine, Gábor Bunkóczi, Vincent B. Chen, Ian W. Davis, Nathaniel Echols, Jeffrey J. Headd, et al. 2010. “PHENIX: A Comprehensive Python-Based System for Macromolecular Structure Solution.” *Acta Crystallographica*. *Section D, Biological Crystallography* 66 (Pt 2): 213–21.

Agilent Technologies. 2015. “GeneMorph II Random Mutagenesis Kit”. Version C.0. Santa Clara, CA, Agilent Technologies.

Ajikumar, Parayil Kumaran, Wen-Hai Xiao, Keith E. J. Tyo, Yong Wang, Fritz Simeon, Effendi Leonard, Oliver Mucha, Too Heng Phon, Blaine Pfeifer, and Gregory Stephanopoulos. 2010. “Isoprenoid Pathway Optimization for Taxol Precursor Overproduction in Escherichia Coli.” *Science* 330 (6000): 70–74.

Aleku, Godwin A., Scott P. France, Henry Man, Juan Mangas-Sanchez, Sarah L. Montgomery, Mahima Sharma, Friedemann Leipold, Shahed Hussain, Gideon Grogan, and Nicholas J. Turner. 2017. “A Reductive Aminase from Aspergillus Oryzae.” *Nature Chemistry* 9 (10): 961–69.

AlQuraishi, Mohammed. 2019. “AlphaFold at CASP13.” *Bioinformatics*, May.

An, Yingfeng, Jianfei Ji, Wenfang Wu, Anguo Lv, Ribo Huang, and Yutuo Wei. 2005. “A Rapid and Efficient Method for Multiple-Site Mutagenesis with a Modified Overlap Extension PCR.” *Applied Microbiology and Biotechnology* 68 (6): 774–78.

Arnold, Frances H. 2018. “Directed Evolution: Bringing New Chemistry to Life.” *Angewandte Chemie* 57 (16): 4143–48.

Ausubel, Frederick M., Roger Brent, Robert E. Kingston, David D. Moore, J. G. Seidman, John A. Smith, and Kevin Struhl, eds. 2001. “Quantitation of DNA and RNA with Absorption and Fluorescence Spectroscopy.” *In Current Protocols in Molecular Biology*, 100:188. Hoboken, NJ, USA: John Wiley & Sons, Inc.

Baba, Tomoya, Takeshi Ara, Miki Hasegawa, Yuki Takai, Yoshiko Okumura, Miki Baba, Kirill A. Datsenko, Masaru Tomita, Barry L. Wanner, and Hirotada Mori. 2006. “Construction of Escherichia Coli K-12 In-frame, Single-gene Knockout Mutants: The Keio Collection.” *Molecular Systems Biology* 2 (1).

Bai, Jing, Yong Zhou, Qi Chen, Qing Yang, and Jun Yang. 2015. “Altering the Regioselectivity of a Nitroreductase in the Synthesis of Arylhydroxylamines by Structure-Based Engineering.” *Chembiochem: A European Journal of Chemical Biology* 16 (8): 1219–25.

Bajaj, Kanika, M. S. Madhusudhan, Bharat V. Adkar, Purbani Chakrabarti, C. Ramakrishnan, Andrej Sali, and Raghavan Varadarajan. 2007. “Stereochemical Criteria for Prediction of the Effects of Proline Mutations on Protein Stability.” *PLoS Computational Biology* 3 (12): e241.

Baker, Patrick J., K. Linda Britton, Martin Fisher, Julia Esclapez, Carmen Pire, Maria Jose Bonete, Juan Ferrer, and David W. Rice. 2009. “Active Site Dynamics in the Zinc-Dependent Medium Chain Alcohol Dehydrogenase Superfamily.” *Proceedings of the National Academy of Sciences of the United States of America* 106 (3): 779–84.

Baldwin, Amy J., Kathy Busse, Alan M. Simm, and D. Dafydd Jones. 2008. “Expanded Molecular Diversity Generation during Directed Evolution by Trinucleotide Exchange (TriNEx).” *Nucleic Acids Research* 36 (13): e77.

Barrick, Jeffrey E., and Richard W. Roberts. 2002. “Sequence Analysis of an Artificial Family of RNA-Binding Peptides.” *Protein Science: A Publication of the Protein Society* 11 (11): 2688–96.

Bennett, Bryson D., Elizabeth H. Kimball, Melissa Gao, Robin Osterhout, Stephen J. Van Dien, and Joshua D. Rabinowitz. 2009. “Absolute Metabolite Concentrations and Implied Enzyme Active Site Occupancy in Escherichia Coli.” *Nature Chemical Biology* 5 (8): 593–99.

Binz, H. Kaspar, Patrick Amstutz, Andreas Kohl, Michael T. Stumpp, Christophe Briand, Patrik Forrer, Markus G. Grütter, and Andreas Plückthun. 2004. “High-Affinity Binders Selected from Designed Ankyrin Repeat Protein Libraries.” *Nature Biotechnology* 22 (5): 575–82.

Boersma, Ykelien L., Melloney J. Dröge, Almer M. van der Sloot, Tjaard Pijning, Robbert H. Cool, Bauke W. Dijkstra, and Wim J. Quax. 2008. “A Novel Genetic Selection System for Improved Enantioselectivity of Bacillus Subtilis Lipase A.” *Chembiochem: A European Journal of Chemical Biology* 9 (7): 1110–15.

Bogin, Oren, Moshe Peretz, and Yigal Burstein. 1998. “Probing Structural Elements of Thermal Stability in Bacterial Oligomeric Alcohol Dehydrogenases. I. Construction and Characterization of Chimeras Consisting of Secondary ADHs fromThermoanaerobacter Brockii andClostridium Beijerinckii.” *Letters in Peptide Science*: LIPS 5 (5): 399–408.

Bokma, Evert, Eva Koronakis, Sune Lobedanz, Colin Hughes, and Vassilis Koronakis. 2006. “Directed Evolution of a Bacterial Efflux Pump: Adaptation of the E. Coli TolC Exit Duct to the Pseudomonas MexAB Translocase.” FEBS Letters 580 (22): 5339–43.

Borlinghaus, Niels, and Bettina M. Nestl. 2018. “Switching the Cofactor Specificity of an Imine Reductase.” *ChemCatChem* 10 (1): 183–87.

Bornadel, Amin, Serena Bisagni, Ahir Pushpanath, Sarah L. Montgomery, Nicholas J. Turner, and Beatriz Dominguez. 2019. “Technical Considerations for Scale-Up of Imine-Reductase-Catalyzed Reductive Amination: A Case Study.” *Organic Process Research & Development* 23 (6): 1262–68.

Bratulic, Sinisa, and Ahmed H. Badran. 2017. “Modern Methods for Laboratory Diversification of Biomolecules.” *Current Opinion in Chemical Biology* 41 (December): 50–60.

Cadwell, R. C., and G. F. Joyce. 1992. “Randomization of Genes by PCR Mutagenesis.” PCR Methods and Applications 2 (1): 28–33.

Cahn, Jackson K. B., Caroline A. Werlang, Armin Baumschlager, Sabine Brinkmann-Chen, Stephen L. Mayo, and Frances H. Arnold. 2017. “A General Tool for Engineering the NAD/NADP Cofactor Preference of Oxidoreductases.” *ACS Synthetic Biology* 6 (2): 326–33.

Calle, Maria Elena de la, Gema Cabrera, Domingo Cantero, Antonio Valle, and Jorge Bolivar. 2019. “A Genetically Engineered Escherichia Coli Strain Overexpressing the Nitroreductase NfsB Is Capable of Producing the Herbicide D-DIBOA with 100% Molar Yield.” *Microbial Cell Factories* 18 (1): 86.

Camps, Manel, Jussi Naukkarinen, Ben P. Johnson, and Lawrence A. Loeb. 2003. “Targeted Gene Evolution in Escherichia Coli Using a Highly Error-Prone DNA Polymerase I.” *Proceedings of the National Academy of Sciences of the United States of America* 100 (17): 9727–32.

Chánique, Andrea M., and Loreto P. Parra. 2018. “Protein Engineering for Nicotinamide Coenzyme Specificity in Oxidoreductases: Attempts and Challenges.” *Frontiers in Microbiology* 9 (February): 194.

Chapman, Jordan, Ahmed E. Ismail, and Cerasela Zoica Dinu. 2018. “Industrial Applications of Enzymes: Recent Advances, Techniques, and Outlooks.” *Catalysts* 8 (6): 238.

Chatterjee, Ranjini, and Ling Yuan. 2006. “Directed Evolution of Metabolic Pathways.” *Trends in Biotechnology* 24 (1): 28–38.

Chen, Irwin, Brent M. Dorr, and David R. Liu. 2011. “A General Strategy for the Evolution of Bond-Forming Enzymes Using Yeast Display.” *Proceedings of the National Academy of Sciences of the United States of America* 108 (28): 11399–404.

Chen, K., and F. H. Arnold. 1993. “Tuning the Activity of an Enzyme for Unusual Environments: Sequential Random Mutagenesis of Subtilisin E for Catalysis in Dimethylformamide.” *Proceedings of the National Academy of Sciences of the United States of America* 90 (12): 5618–22.

Chen, Minyong, Xiaofeng Shi, Rebecca M. Duke, Cristian I. Ruse, Nan Dai, Christopher H. Taron, and James C. Samuelson. 2017. “An Engineered High Affinity Fbs1 Carbohydrate Binding Protein for Selective Capture of N-Glycans and N-Glycopeptides.” *Nature Communications* 8 (May): 15487.

Chen, Vincent B., W. Bryan Arendall 3rd, Jeffrey J. Headd, Daniel A. Keedy, Robert M. Immormino, Gary J. Kapral, Laura W. Murray, Jane S. Richardson, and David C. Richardson. 2010. “MolProbity: All-Atom Structure Validation for Macromolecular

Crystallography.” *Acta Crystallographica. Section D, Biological Crystallography* 66 (Pt 1): 12–21.
Cherepanov, P. P., and W. Wackernagel. 1995. “Gene Disruption in Escherichia Coli: TcR and KmR Cassettes with the Option of Flp-Catalyzed Excision of the Antibiotic-Resistance Determinant.” *Gene* 158 (1): 9–14.
Clark, D. P. 1989. “The Fermentation Pathways of Escherichia Coli.” *FEMS Microbiology Reviews* 5 (3): 223–34.
Cobb, Ryan E., Ran Chao, and Huimin Zhao. 2013. “Directed Evolution: Past, Present and Future.” *AIChE Journal. American Institute of Chemical Engineers* 59 (5): 1432–40.
Coco, W. M., W. E. Levinson, M. J. Crist, H. J. Hektor, A. Darzins, P. T. Pienkos, C. H. Squires, and D. J. Monticello. 2001. “DNA Shuffling Method for Generating Highly Recombined Genes and Evolved Enzymes.” *Nature Biotechnology* 19 (4): 354–59.
Crabtree, Robert H. 2013 “Green Catalysis: Biocatalysis”. John Wiley & Sons, Germany.
Crameri, A., E. A. Whitehorn, E. Tate, and W. P. Stemmer. 1996. “Improved Green Fluorescent Protein by Molecular Evolution Using DNA Shuffling.” *Nature Biotechnology* 14 (3): 315–19.
Cripps, R. E., K. Eley, D. J. Leak, B. Rudd, M. Taylor, M. Todd, S. Boakes, S. Martin, and T. Atkinson. 2009. “Metabolic Engineering of Geobacillus Thermoglucosidasius for High Yield Ethanol Production.” *Metabolic Engineering* 11 (6): 398–408.
Croll, Tristan I., Massimo D. Sammito, Andriy Kryshtafovych, and Randy J. Read. 2019. “Evaluation of Template-Based Modeling in CASP13.” *Proteins*, August.
Crook, Nathan, Joseph Abatemarco, Jie Sun, James M. Wagner, Alexander Schmitz, and Hal S. Alper. 2016. “In Vivo Continuous Evolution of Genes and Pathways in Yeast.” *Nature Communications* 7 (October): 13051.
Davis, B. D. 1951. “Inhibition of Escherichia Coli by P-Aminobenzoic Acid and Its Reversal by P-Hydroxybenzoic Acid.” *The Journal of Experimental Medicine* 94 (3): 243–54.
Dayem, Linda C., John R. Carney, Daniel V. Santi, Blaine A. Pfeifer, Chaitan Khosla, and James T. Kealey. 2002. “Metabolic Engineering of a Methylmalonyl-CoA Mutase-Epimerase Pathway for Complex Polyketide Biosynthesis in Escherichia Coli.” *Biochemistry* 41 (16): 5193–5201.
Dennig, Alexander, Amol V. Shivange, Jan Marienhagen, and Ulrich Schwaneberg. 2011. “OmniChange: The Sequence Independent Method for Simultaneous Site-Saturation of Five Codons.” *PloS One* 6 (10): e26222.
Didelot, Xavier, and Martin C. J. Maiden. 2010. “Impact of Recombination on Bacterial Evolution.” *Trends in Microbiology* 18 (7): 315–22.
Edwards, Wayne R., Kathy Busse, Rudolf K. Allemann, and D. Dafydd Jones. 2008. “Linking the Functions of Unrelated Proteins Using a Novel Directed Evolution Domain Insertion Method.” *Nucleic Acids Research* 36 (13): e78.
Eisenberg, R. C., and W. J. Dobrogosz. 1967. “Gluconate Metabolism in Escherichia Coli.” *Journal of Bacteriology* 93 (3): 941–49.
Emsley, Paul, and Kevin Cowtan. 2004. “Coot: Model-Building Tools for Molecular Graphics.” Acta Crystallographica. Section D, *Biological Crystallography* 60 (Pt 12 Pt 1): 2126–32.
Emsley, P., B. Lohkamp, W. G. Scott, and K. Cowtan. 2010. “Features and Development of Coot.” Acta Crystallographica. Section D, *Biological Crystallography* 66 (Pt 4): 486–501.
Esvelt, Kevin M., Jacob C. Carlson, and David R. Liu. 2011. “A System for the Continuous Directed Evolution of Biomolecules.” *Nature* 472 (7344): 499–503.
Feiler, Christian, Adam C. Fisher, Jason T. Boock, Matthew J. Marrichi, Lori Wright, Philipp A. M. Schmidpeter, Wulf Blankenfeldt, Martin Pavelka, and Matthew P. DeLisa. 2013. “Directed Evolution of Mycobacterium Tuberculosis β-Lactamase Reveals Gatekeeper Residue That Regulates Antibiotic Resistance and Catalytic Efficiency.” *PloS One* 8 (9): e73123.
Fire, A., and S. Q. Xu. 1995. “Rolling Replication of Short DNA Circles.” *Proceedings of the National Academy of Sciences of the United States of America* 92 (10): 4641–45.

Firestine, S. M., F. Salinas, A. E. Nixon, S. J. Baker, and S. J. Benkovic. 2000. “Using an AraC-Based Three-Hybrid System to Detect Biocatalysts in Vivo.” *Nature Biotechnology* 18 (5): 544–47.

Flor, P. Q., and S. Hayashida. 1983. “Continuous Production of High-Glucose Syrup by Chitin-Immobilized Amylase.” *Biotechnology and Bioengineering* 25 (8): 1973–80.

Fujii, Ry, Motomitsu Kitaoka, and Kiyoshi Hayashi. 2004. “One-Step Random Mutagenesis by Error-Prone Rolling Circle Amplification.” *Nucleic Acids Research* 32 (19): e145.

Fujii, Ry, Motomitsu Kitaoka, and Kiyoshi Hayashi. 2006. “RAISE: A Simple and Novel Method of Generating Random Insertion and Deletion Mutations.” *Nucleic Acids Research* 34 (4): e30.

Galdzicki, Michal, Cesar Rodriguez, Deepak Chandran, Herbert M. Sauro, and John H. Gennari. 2011. “Standard Biological Parts Knowledgebase.” *PloS One* 6 (2): e17005.

Gand, Martin, Christian Thöle, Hubertus Müller, Henrike Brundiek, Ghader Bashiri, and Matthias Höhne. 2016. “A NADH-Accepting Imine Reductase Variant: Immobilization and Cofactor Regeneration by Oxidative Deamination.” *Journal of Biotechnology* 230 (July): 11–18.

Gaona-López, Carlos, Adriana Julián-Sánchez, and Héctor Riveros-Rosas. 2016. “Diversity and Evolutionary Analysis of Iron-Containing (Type-III) Alcohol Dehydrogenases in Eukaryotes.” *PloS One* 11 (11): e0166851.

Garg, S. K., and A. Jain. 1995. “Fermentative Production of 2,3-Butanediol: A Review.” *Bioresource Technology* 51 (2): 103–9.

Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., Bairoch A. 2005. “Protein Identification and Analysis Tools on the ExPASy Server”; (In) John M. Walker (ed): *The Proteomics Protocols Handbook*, Humana Press .pp. 571-607

Gaytán, Paul, Casandra Contreras-Zambrano, Mónica Ortiz-Alvarado, Alfredo Morales-Pablos, and Jorge Yáñez. 2009. “TrimerDimer: An Oligonucleotide-Based Saturation Mutagenesis Approach That Removes Redundant and Stop Codons.” *Nucleic Acids Research* 37 (18): e125.

Ghadessy, F. J., J. L. Ong, and P. Holliger. 2001. “Directed Evolution of Polymerase Function by Compartmentalized Self-Replication.” *Proceedings of the National Academy of Sciences of the United States of America* 98 (8): 4552–57.

Ghislieri, Diego, and Nicholas J. Turner. 2014. “Biocatalytic Approaches to the Synthesis of Enantiomerically Pure Chiral Amines.” *Topics in Catalysis* 57 (5): 284–300.

Girault, Jean-Pierre, and Gilbert Dana. 1977. “A Re-Evaluation of the Hammett σp Values for the Hydroxymethyl and Formyl Groups.” J*ournal of the Chemical Society, Perkin Transactions* 2, no. 7 (January): 993–993.

Goddard, Thomas D., Conrad C. Huang, Elaine C. Meng, Eric F. Pettersen, Gregory S. Couch, John H. Morris, and Thomas E. Ferrin. 2018. “UCSF ChimeraX: Meeting Modern Challenges in Visualization and Analysis*.” Protein Science: A Publication of the Protein Society* 27 (1): 14–25.

Gonzalez-Garcia, Ricardo Axayacatl, Tim McCubbin, Annalena Wille, Manuel Plan, Lars Keld Nielsen, and Esteban Marcellin. 2017. “Awakening Sleeping Beauty: Production of Propionic Acid in Escherichia Coli through the Sbm Operon Requires the Activity of a Methylmalonyl-CoA Epimerase.” *Microbial Cell Factories* 16 (1): 121.

Gonzalez-Perez, David, Patricia Molina-Espeja, Eva Garcia-Ruiz, and Miguel Alcalde. 2014. “Mutagenic Organized Recombination Process by Homologous IN Vivo Grouping (MORPHING) for Directed Enzyme Evolution.” *PloS One* 9 (3): e90919.

Green Catalysis: Biocatalysis. 2013. John Wiley & Sons.

Green, Jeffrey, and John R. Guest. 1998. “The Citric Acid Cycle and Oxygen-Regulated Gene Expression in Escherichia Coli.” *In Molecular Microbiology*, 17–39. Springer Berlin Heidelberg.

Greener, Alan, and M. Callahan. 1994. “XL1-Red: A Highly Efficient Random Mutagenesis Strain.” *Strategies* 7: 32–34.

Greener, Alan, Marie Callahan, and Bruce Jerpseth. 1996. “An Efficient Random Mutagenesis Technique Using an E. Coli Mutator Strain.” *In In Vitro Mutagenesis Protocols*, edited by Michael K. Trower, 375–85. Totowa, NJ: Humana Press.

Grove, Jane I., Andrew L. Lovering, Christopher Guise, Paul R. Race, Christopher J. Wrighton, Scott A. White, Eva I. Hyde, and Peter F. Searle. 2003. “Generation of Escherichia Coli Nitroreductase Mutants Conferring Improved Cell Sensitization to the Prodrug CB1954.” *Cancer Research* 63 (17): 5532–37.

Grummt, Tamara, Heinz-Günter Wunderlich, Asima Chakraborty, Michael Kundi, Bernhard Majer, Franziska Ferk, Armen K. Nersesyan, Wolfram Parzefall, and Siegfried Knasmüller. 2006. “Genotoxicity of Nitrosulfonic Acids, Nitrobenzoic Acids, and Nitrobenzylalcohols, Pollutants Commonly Found in Ground Water near Ammunition Facilities.” *Environmental and Molecular Mutagenesis* 47 (2): 95–106.

Guan, Ningzi, Hyun-Dong Shin, Rachel R. Chen, Jianghua Li, Long Liu, Guocheng Du, and Jian Chen. 2014. “Understanding of How Propionibacterium Acidipropionici Respond to Propionic Acid Stress at the Level of Proteomics.” *Scientific Reports* 4 (November): 6951.

Güngör, Tuğba, Ferah Cömert Önder, Esra Tokay, Ünzile Güven Gülhan, Nelin Hacıoğlu, Tuğba Taşkın Tok, Ayhan Çelik, Feray Köçkar, and Mehmet Ay. 2019. “Prodrugs for nitroreductase based cancer therapy- 2: Novel amide/Ntr Combinations Targeting PC3 Cancer Cells.” *European Journal of Medicinal Chemistry* 171 (June): 383–400.

Gungor, Tugba, Gulden Yetis, Ferah C. Onder, Esra Tokay, Tugba T. Tok, Ayhan Celik, Mehmet Ay, and Feray Kockar. 2018. “Prodrugs for Nitroreductase Based Cancer Therapy- 1: Metabolite Profile, Cell Cytotoxicity and Molecular Modeling Interactions of Nitro Benzamides with Ssap-NtrB.” *Medicinal Chemistry* 14 (5): 495–507.

Gupta, Rinkoo D., Moshe Goldsmith, Yacov Ashani, Yair Simo, Gavriel Mullokandov, Hagit Bar, Moshe Ben-David, *et al.* 2011. “Directed Evolution of Hydrolases for Prevention of G-Type Nerve Agent Intoxication.” *Nature Chemical Biology* 7 (2): 120–25.

Haapa, S., S. Suomalainen, S. Eerikäinen, M. Airaksinen, L. Paulin, and H. Savilahti. 1999. “An Efficient DNA Sequencing Strategy Based on the Bacteriophage Mu in Vitro DNA Transposition Reaction.” *Genome Research* 9 (3): 308–15.

Hall, Mélanie, and Andreas S. Bommarius. 2011. “Enantioenriched Compounds via Enzyme-Catalyzed Redox Reactions.” Chemical Reviews 111 (7): 4088–4110.

Haller, T., T. Buckel, J. Rétey, and J. A. Gerlt. 2000. “Discovering New Enzymes and Metabolic Pathways: Conversion of Succinate to Propionate by Escherichia Coli.” *Biochemistry* 39 (16): 4622–29.

Hamilton, C. M., M. Aldea, B. K. Washburn, P. Babitzke, and S. R. Kushner. 1989. “New Method for Generating Deletions and Gene Replacements in Escherichia Coli.” *Journal of Bacteriology* 171 (9): 4617–22.

Hanai, T., S. Atsumi, and J. C. Liao. 2007. “Engineered Synthetic Pathway for Isopropanol Production in Escherichia Coli.” *Applied and Environmental Microbiology* 73 (24): 7814–18.

Hanes, J., and A. Plückthun. 1997. “In Vitro Selection and Evolution of Functional Proteins by Using Ribosome Display.” *Proceedings of the National Academy of Sciences of the United States of America* 94 (10): 4937–42.

Hansch, Corwin, A. Leo, and R. W. Taft. 1991. “A Survey of Hammett Substituent Constants and Resonance and Field Parameters.” *Chemical Reviews* 91 (2): 165–95.

Heap, John T., Jan Theys, Muhammad Ehsaan, Aleksandra M. Kubiak, Ludwig Dubois, Kim Paesmans, Lieve Van Mellaert, et al. 2014. “Spores of Clostridium Engineered for Clinical Efficacy and Safety Cause Regression and Cure of Tumors in Vivo.” *Oncotarget* 5 (7): 1761–69.

Heim, R., D. C. Prasher, and R. Y. Tsien. 1994. “Wavelength Mutations and Posttranslational Autoxidation of Green Fluorescent Protein.” *Proceedings of the National Academy of Sciences of the United States of America* 91 (26): 12501–4.

Hiraga, Kaori, and Frances H. Arnold. 2003. “General Method for Sequence-Independent Site-Directed Chimeragenesis.” *Journal of Molecular Biology* 330 (2): 287–96.

Ho, Chi-Ming, and Tai-Chu Lau. 2000. “Copper-Catalyzed Amination of Alkenes and Ketones by Phenylhydroxylamine.” *New Journal of Chemistry = Nouveau Journal de Chimie* 24 (11): 859–63.

Höllrigl, Volker, Frank Hollmann, Andreas C. Kleeb, Katja Buehler, and Andreas Schmid. 2008. “TADH, the Thermostable Alcohol Dehydrogenase from Thermus Sp. ATN1: A Versatile New Biocatalyst for Organic Synthesis.” *Applied Microbiology and Biotechnology* 81 (2): 263–73.

Honeycutt, M. E., A. S. Jarvis, and V. A. McFarland. 1996. “Cytotoxicity and Mutagenicity of 2,4,6-Trinitrotoluene and Its Metabolites.” *Ecotoxicology and Environmental Safety* 35 (3): 282–87.

Hoogenraad, Marcel, Johannes B. van der Linden, Alan A. Smith, Bob Hughes, Andrew M. Derrick, Laurence J. Harris, Paul D. Higginson, and Alan J. Pettman. 2004. “Accelerated Process Development of Pharmaceuticals: Selective Catalytic Hydrogenations of Nitro Compounds Containing Other Functionalities.” *Organic Process Research & Development* 8 (3): 469–76.

Hummel, W. 1997. “New Alcohol Dehydrogenases for the Synthesis of Chiral Compounds.” *Advances in Biochemical Engineering/biotechnology* 58: 145–84.

Huovinen, Tuomas, Marja Julin, Hanna Sanmark, and Urpo Lamminmäki. 2011. “Enhanced Error-Prone RCA Mutagenesis by Concatemer Resolution.” *Plasmid* 66 (1): 47–51.

Hvidsten, Torgeir R., Astrid Laegreid, Andriy Kryshtafovych, Gunnar Andersson, Krzysztof Fidelis, and Jan Komorowski. 2009. “A Comprehensive Analysis of the Structure-Function Relationship in Proteins Based on Local Structure Similarity.” *PloS One* 4 (7): e6266.

Itoh, Toshiyuki, and Ulf Hanefeld. 2017. “Enzyme Catalysis in Organic Synthesis.” *Green Chemistry: An International Journal and Green Chemistry Resource:* GC 19 (2): 331–32.

Jaberipour, Mansooreh, Simon O. Vass, Christopher P. Guise, Jane I. Grove, Richard J. Knox, Longqin Hu, Eva I. Hyde, and Peter F. Searle. 2010. “Testing Double Mutants of the Enzyme Nitroreductase for Enhanced Cell Sensitisation to Prodrugs: Effects of Combining Beneficial Single Mutations.” *Biochemical Pharmacology* 79 (2): 102–11.

Jeschek, Markus, Daniel Gerngross, and Sven Panke. 2017. “Combinatorial Pathway Optimization for Streamlined Metabolic Engineering.” *Current Opinion in Biotechnology* 47 (October): 142–51.

Johannes, Tyler, Michael R. Simurdiak, and Huimin Zhao. 2005. “Biocatalysis.” *In Encyclopedia of Chemical Processing* (Online), 101–10. CRC Press.

Johansson, Eric, Gary N. Parkinson, William A. Denny, and Stephen Neidle. 2003. “Studies on the Nitroreductase Prodrug-Activating System. Crystal Structures of Complexes with the Inhibitor Dicoumarol and Dinitrobenzamide Prodrugs and of the Enzyme Active Form.” *Journal of Medicinal Chemistry* 46 (19): 4009–20.

John Wiley & Sons, Ltd, ed. 2001. “Bacterial Fermentation.” In Encyclopedia of Life Sciences, 10:286. Chichester, UK: *John Wiley & Sons*, Ltd.

Johnson, Mark, Irena Zaretskaya, Yan Raytselis, Yuri Merezhuk, Scott McGinnis, and Thomas L. Madden. 2008. “NCBI BLAST: A Better Web Interface.” *Nucleic Acids Research* 36 (Web Server issue): W5–9.

Jones, D. Dafydd. 2005. “Triplet Nucleotide Removal at Random Positions in a Target Gene: The Tolerance of TEM-1 Beta-Lactamase to an Amino Acid Deletion.” *Nucleic Acids Research* 33 (9): e80.

Kavanagh, K. L., H. Jörnvall, B. Persson, and U. Oppermann. 2008. “Medium- and Short-Chain Dehydrogenase/reductase Gene and Protein Families: The SDR Superfamily: Functional and Structural Diversity within a Family of Metabolic and Regulatory Enzymes.” *Cellular and Molecular Life Sciences*: CMLS 65 (24): 3895–3906.

Kayushin, A., M. Korosteleva, and A. Miroshnikov. 2000. “Large-Scale Solid-Phase Preparation of 3’-Unprotected Trinucleotide Phosphotriesters--Precursors for Synthesis of Trinucleotide Phosphoramidites.” *Nucleosides, Nucleotides & Nucleic Acids* 19 (10-12): 1967–76.

Keefe, A. D., and J. W. Szostak. 2001. “Functional Proteins from a Random-Sequence Library.” *Nature* 410 (6829): 715–18.
Kestell, Philip, Frederik B. Pruijn, Bronwyn G. Siim, Brian D. Palmer, and William R. Wilson. 2000. “Pharmacokinetics and Metabolism of the Nitrogen Mustard Bioreductive Drug 5-[N,N-bis(2-Chloroethyl)amino]-2,4-Dinitrobenzamide (SN 23862) and the Corresponding Aziridine (CB 1954) in KHT Tumour-Bearing Mice.” *Cancer Chemotherapy and Pharmacology*.
Khan, A. H., and W. C. J. Ross. 1969. “Tumour-Growth Inhibitory Nitrophenylaziridines and Related Compounds: Structure-Activity Relationships. II.” *Chemico-Biological Interactions*.
Kipnis, Yakov, Eynat Dellus-Gur, and Dan S. Tawfik. 2012. “TRINS: A Method for Gene Modification by Randomized Tandem Repeat Insertions.” *Protein Engineering, Design & Selection: PEDS* 25 (9): 437–44.
Kizer, Lance, Douglas J. Pitera, Brian F. Pfleger, and Jay D. Keasling. 2008. “Application of Functional Genomics to Pathway Optimization for Increased Isoprenoid Production.” *Applied and Environmental Microbiology* 74 (10): 3229–41.
Koder, Ronald L., Chad A. Haynes, Michael E. Rodgers, David W. Rodgers, and Anne-Frances Miller. 2002. “Flavin Thermodynamics Explain the Oxygen Insensitivity of Enteric Nitroreductases.” *Biochemistry* 41 (48): 14197–205.
Koné, Fankroma M. T., Mickaël Le Béchec, Jean-Pierre Sine, Michel Dion, and Charles Tellier. 2009. “Digital Screening Methodology for the Directed Evolution of Transglycosidases.” *Protein Engineering, Design & Selection: PEDS* 22 (1): 37–44.
Korkhin, Y., A. J. Kalb(Gilboa), M. Peretz, O. Bogin, Y. Burstein, and F. Frolow. 1998. “NADP-Dependent Bacterial Alcohol Dehydrogenases: Crystal Structure, Cofactor-Binding and Cofactor Specificity of the ADHs of Clostridium Beijerinckii and Thermoanaerobacter Brockii.” *Journal of Molecular Biology* 278 (5): 967–81.
Krahe, M., G. Antranikian, and H. Märkl. 1996. “Fermentation of Extremophilic Microorganisms.” *FEMS Microbiology Reviews* 18 (2-3): 271–85.
Kutzenko, A. S., V. S. Lamzin, and V. O. Popov. 1998. “Conserved Supersecondary Structural Motif in NAD-Dependent Dehydrogenases.” *FEBS Letters* 423 (1): 105–9.
Lee, David, Oliver Redfern, and Christine Orengo. 2007. “Predicting Protein Function from Sequence and Structure.” *Nature Reviews. Molecular Cell Biology* 8 (12): 995–1005.
Lee, Heewook, Ellen Popodi, Haixu Tang, and Patricia L. Foster. 2012. “Rate and Molecular Spectrum of Spontaneous Mutations in the Bacterium Escherichia Coli as Determined by Whole-Genome Sequencing.” *Proceedings of the National Academy of Sciences of the United States of America* 109 (41): E2774–83.
Lee, Inmok, Lawrence A. Johnson, and Earl G. Hammond. 1995. “Use of Branched-Chain Esters to Reduce the Crystallization Temperature of Biodiesel.” *Journal of the American Oil Chemists’ Society* 72 (10): 1155–60.
Lee, Jeong Wook, Dokyun Na, Jong Myoung Park, Joungmin Lee, Sol Choi, and Sang Yup Lee. 2012. “Systems Metabolic Engineering of Microorganisms for Natural and Non-Natural Chemicals.” *Nature Chemical Biology* 8 (6): 536–46.
Lee, Na-Rae, Ji-Yeong Yun, Sun-Mee Lee, and Jin-Byung Park. 2015. “Cyclohexanone-Induced Stress Metabolism of Escherichia Coli and Corynebacterium Glutamicum.” *Biotechnology and Bioprocess Engineering: BBE* 20 (6): 1088–98.
Li, Tao, Jack Liang, Alexandre Ambrogelly, Tim Brennan, Guy Gloor, Gjalt Huisman, James Lalonde, *et al.* 2012. “Efficient, Chemoenzymatic Process for Manufacture of the Boceprevir Bicyclic [3.1.0] proline Intermediate Based on Amine Oxidase-Catalyzed Desymmetrization.” *Journal of the American Chemical Society* 134 (14): 6467–72.
Liu, Shu-Su, Xuan Wei, Qun Ji, Xiu Xin, Biao Jiang, and Jia Liu. 2016. “A Facile and Efficient Transposon Mutagenesis Method for Generation of Multi-Codon Deletions in Protein Sequences.” *Journal of Biotechnology* 227 (June): 27–34.
Lutz, S., M. Ostermeier, G. L. Moore, C. D. Maranas, and S. J. Benkovic. 2001. “Creating Multiple-Crossover DNA Libraries Independent of Sequence Identity.” *Proceedings of the National Academy of Sciences of the United States of America* 98 (20): 11248–53.

Macdonald, J. R., and W. C. Johnson Jr. 2001. “Environmental Features Are Important in Determining Protein Secondary Structure.” *Protein Science: A Publication of the Protein Society* 10 (6): 1172–77.

Maddock, Danielle J., Wayne M. Patrick, and Monica L. Gerth. 2015. “Substitutions at the Cofactor Phosphate-Binding Site of a Clostridial Alcohol Dehydrogenase Lead to Unexpected Changes in Substrate Specificity.” *Protein Engineering, Design & Selection: PEDS* 28 (8): 251–58.

Man, Henry, Elizabeth Wells, Shahed Hussain, Friedemann Leipold, Sam Hart, Johan P. Turkenburg, Nicholas J. Turner, and Gideon Grogan. 2015. “Structure, Activity and Stereoselectivity of NADPH-Dependent Oxidoreductases Catalysing the S-Selective Reduction of the Imine Substrate 2-Methylpyrroline.” Chembiochem: *A European Journal of Chemical Biology* 16 (7): 1052–59.

Martinez, Carlos A., Shanghui Hu, Yves Dumond, Junhua Tao, Patrick Kelleher, and Liam Tully. 2008. “Development of a Chemoenzymatic Manufacturing Process for Pregabalin.” *Organic Process Research & Development* 12 (3): 392–98.

Mattheakis, L. C., R. R. Bhatt, and W. J. Dower. 1994. “An in Vitro Polysome Display System for Identifying Ligands from Very Large Peptide Libraries.” *Proceedings of the National Academy of Sciences of the United States of America* 91 (19): 9022–26.

Matzel, Philipp, Lukas Krautschick, and Matthias Höhne. 2017. “Photometric Characterization of the Reductive Amination Scope of the Imine Reductases from Streptomyces Tsukubaensis and Streptomyces Ipomoeae.” *Chembiochem: A European Journal of Chemical Biology* 18 (20): 2022–27.

McCormick, N. G., F. E. Feeherry, and H. S. Levinson. 1976. “Microbial Transformation of 2,4,6-Trinitrotoluene and Other Nitroaromatic Compounds.” *Applied and Environmental Microbiology* 31 (6): 949–58.

McCoy, Airlie J., Ralf W. Grosse-Kunstleve, Paul D. Adams, Martyn D. Winn, Laurent C. Storoni, and Randy J. Read. 2007. “Phasercrystallographic Software.” *Journal of Applied Crystallography*.

McCrindle, Sharon L., Ulrike Kappler, and Alastair G. McEwan. 2005. “Microbial Dimethylsulfoxide and Trimethylamine-N-Oxide Respiration.” *In Advances in Microbial Physiology*, edited by Robert K. Poole, 50:147–201e. Academic Press.

Meites, J. 1977. “Evaluation of Research on Control of Prolactin Secretion.” *Advances in Experimental Medicine and Biology* 80: 135–52.

Miller, Anne-Frances, Jonathan T. Park, Kyle L. Ferguson, Warintra Pitsawong, and Andreas S. Bommarius. 2018. “Informing Efforts to Develop Nitroreductase for Amine Production.” *Molecules* 23 (2).

Miller, Brian G., and Richard Wolfenden. 2002. “Catalytic Proficiency: The Unusual Case of OMP Decarboxylase.” *Annual Review of Biochemistry* 71: 847–85.

Mills, D. R., R. L. Peterson, and S. Spiegelman. 1967. “An Extracellular Darwinian Experiment with a Self-Duplicating Nucleic Acid Molecule.” *Proceedings of the National Academy of Sciences of the United States of America* 58 (1): 217–24.

Minamoto, Toshifumi, Eitaro Wada, and Isamu Shimizu. 2012. “A New Method for Random Mutagenesis by Error-Prone Polymerase Chain Reaction Using Heavy Water.” *Journal of Biotechnology* 157 (1): 71–74.

Miyamoto-Sato, E., N. Nemoto, K. Kobayashi, and H. Yanagawa. 2000. “Specific Bonding of Puromycin to Full-Length Protein at the C-Terminus.” *Nucleic Acids Research* 28 (5): 1176–82.

Moon, Ji-Hyun, Hyun-Ju Lee, Suk-Youl Park, Jung-Mi Song, Mi-Young Park, Hye-Mi Park, Jiali Sun, Jeong-Hoh Park, Bo Yeon Kim, and Jeong-Sun Kim. 2011. “Structures of Iron-Dependent Alcohol Dehydrogenase 2 from Zymomonas Mobilis ZM4 with and without NAD+ Cofactor.” *Journal of Molecular Biology* 407 (3): 413–24.

Myers, R. M., L. S. Lerman, and T. Maniatis. 1985. “A General Method for Saturation Mutagenesis of Cloned DNA Fragments.” *Science* 229 (4710): 242–47.

Narnoliya, Lokesh Kumar, and Jyoti Singh Jadaun. 2019. “Biotechnological Avenues for Fruit Juices Debittering.” *In Green Bio-Processes: Enzymes in Industrial Food Processing*,

edited by Binod Parameswaran, Sunita Varjani, and Sindhu Raveendran, 119–49. Singapore: Springer Singapore.
Neylon, Cameron. 2004. “Chemical and Biochemical Strategies for the Randomization of Protein Encoding DNA Sequences: Library Construction Methods for Directed Evolution.” *Nucleic Acids Research* 32 (4): 1448–59.
Nielsen, Jens, and Jay D. Keasling. 2016. “Engineering Cellular Metabolism.” *Cell* 164 (6): 1185–97.
Nimrod, G., F. Glaser, D. Steinberg, N. Ben-Tal, and T. Pupko. 2005. “In Silico Identification of Functional Regions in Proteins.” *Bioinformatics*.
Nixon, Andrew E., Daniel J. Sexton, and Robert C. Ladner. 2014. “Drugs Derived from Phage Display: From Candidate Identification to Clinical Practice.” *mAbs* 6 (1): 73–85.
Noda-Garcia, Lianet, Wolfram Liebermeister, and Dan S. Tawfik. 2018. “Metabolite-Enzyme Coevolution: From Single Enzymes to Metabolic Pathways and Networks.*” Annual Review of Biochemistry* 87 (June): 187–216.
Norrander, J., T. Kempe, and J. Messing. 1983. “Construction of Improved M13 Vectors Using Oligodeoxynucleotide-Directed Mutagenesis.” *Gene* 26 (1): 101–6.
Nugent, Thomas C., and Mohamed El-Shazly. 2010. “Chiral Amine Synthesis - Recent Developments and Trends for Enamide Reduction, Reductive Amination, and Imine Reduction.” *Advanced Synthesis & Catalysis*.
Nyanhongo, Gibson S., Marc Schroeder, Walter Steiner, and Georg M. Gübitz. 2005. “Biodegradation of 2,4,6-Trinitrotoluene (TNT): An Enzymatic Perspective.” *Biocatalysis and Biotransformation* 23 (2): 53–69.
Nyerges, Ákos, Bálint Csörgő, Gábor Draskovits, Bálint Kintses, Petra Szili, Györgyi Ferenc, Tamás Révész, et al. 2018. “Directed Evolution of Multiple Genomic Loci Allows the Prediction of Antibiotic Resistance.” *Proceedings of the National Academy of Sciences of the United States of America* 115 (25): E5726–35.
Olofsson, Johanna, Zsolt Barta, Pål Börjesson, and Ola Wallberg. 2017. “Integrating Enzyme Fermentation in Lignocellulosic Ethanol Production: Life-Cycle Assessment and Techno-Economic Analysis.” *Biotechnology for Biofuels* 10 (February): 51
Olsen, Hans Sejr, and Per Falholt. 1998. “The Role of Enzymes in Modern Detergency.” *Journal of Surfactants and Detergents, Surfactant Science Series*, 1 (4): 555–67.
Orth, Jeffrey D., Ines Thiele, and Bernhard Palsson. 2010. “What Is Flux Balance Analysis?” Nature Biotechnology 28 (3): 245–48.
Osbon, Yauheniya, and Manish Kumar. 2019. “Biocatalysis and Strategies for Enzyme Improvement.” *In Biophysical Chemistry*. IntechOpen.
Ostermeier, M., J. H. Shim, and S. J. Benkovic. 1999. “A Combinatorial Approach to Hybrid Enzymes Independent of DNA Homology.” *Nature Biotechnology* 17 (12): 1205–9.
Patel, Prashant, J. Graham Young, Vivien Mautner, Daniel Ashdown, Sarah Bonney, Robert G. Pineda, Stuart I. Collins, *et al.* 2009. “A Phase I/II Clinical Trial in Localized Prostate Cancer of an Adenovirus Expressing Nitroreductase with CB1984.” *Molecular Therapy: The Journal of the American Society of Gene Therapy* 17 (7): 1292–99.
Patel, Ramesh N. 2013. “Biocatalytic Synthesis of Chiral Alcohols and Amino Acids for Development of Pharmaceuticals.” *Biomolecules* 3 (4): 741–77.
Peng, Ri-He, Ai-Sheng Xiong, and Quan-Hong Yao. 2006. “A Direct and Efficient PAGE-Mediated Overlap Extension PCR Method for Gene Multiple-Site Mutagenesis.” *Applied Microbiology and Biotechnology* 73 (1): 234–40.
Perona, J. J., L. Hedstrom, W. J. Rutter, and R. J. Fletterick. 1995. “Structural Origins of Substrate Discrimination in Trypsin and Chymotrypsin.” *Biochemistry* 34 (5): 1489–99.
Pettersson, G. 1987. “Liver Alcohol Dehydrogenase.” *CRC Critical Reviews in Biochemistry* 21 (4): 349–89.
Pitsawong, Warintra, Chad A. Haynes, Ronald L. Koder Jr, David W. Rodgers, and Anne-Frances Miller. 2017. “Mechanism-Informed Refinement Reveals Altered Substrate-Binding Mode for Catalytically Competent Nitroreductase.” *Structure* 25 (7): 978–87.e4.

Pitsawong, Warintra, John P. Hoben, and Anne-Frances Miller. 2014. “Understanding the Broad Substrate Repertoire of Nitroreductase Based on Its Kinetic Mechanism.” *The Journal of Biological Chemistry* 289 (22): 15203–14.

Püllmann, Pascal, Chris Ulpinnis, Sylvestre Marillonnet, Ramona Gruetzner, Steffen Neumann, and Martin J. Weissenborn. 2019. “Golden Mutagenesis: An Efficient Multi-Site-Saturation Mutagenesis Approach by Golden Gate Cloning with Automated Primer Design.” *Scientific Reports* 9 (1): 10932.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Race, Paul R., Andrew L. Lovering, Richard M. Green, Abdelmijd Ossor, Scott A. White, Peter F. Searle, Christopher J. Wrighton, and Eva I. Hyde. 2005. “Structural and Mechanistic Studies of Escherichia Coli Nitroreductase with the Antibiotic Nitrofurazone. Reversed Binding Orientations in Different Redox States of the Enzyme.” *The Journal of Biological Chemistry* 280 (14): 13256–64.

Ramos, Juan L., M. Mar González-Pérez, Antonio Caballero, and Pieter van Dillewijn. 2005. “Bioremediation of Polynitrated Aromatic Compounds: Plants and Microbes Put up a Fight.” *Current Opinion in Biotechnology* 16 (3): 275–81.

Ravikumar, Arjun, Adrian Arrieta, and Chang C. Liu. 2014. “An Orthogonal DNA Replication System in Yeast.” *Nature Chemical Biology* 10 (3): 175–77.

Reed, Jennifer L., Thuy D. Vo, Christophe H. Schilling, and Bernhard O. Palsson. 2003. “An Expanded Genome-Scale Model of Escherichia Coli K-12 (iJR904 GSM/GPR).” *Genome Biology* 4 (9): R54.

Reed, Michael C., Anna Lieb, and H. Frederik Nijhout. 2010. “The Biological Significance of Substrate Inhibition: A Mechanism with Diverse Functions.” *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology* 32 (5): 422–29.

Reetz, Manfred T., José D. Carballeira, and Andreas Vogel. 2006. “Iterative Saturation Mutagenesis on the Basis of B Factors as a Strategy for Increasing Protein Thermostability.” *Angewandte Chemie* 45 (46): 7745–51.

Rodríguez-Mata, María, Annika Frank, Elizabeth Wells, Friedemann Leipold, Nicholas J. Turner, Sam Hart, Johan P. Turkenburg, and Gideon Grogan. 2013. “Structure and Activity of NADPH-Dependent Reductase Q1EQE0 from Streptomyces Kanamyceticus, Which Catalyses the R-Selective Reduction of an Imine Substrate.” *Chembiochem: A European Journal of Chemical Biology* 14 (11): 1372–79.

Roldán, María Dolores, Eva Pérez-Reinado, Francisco Castillo, and Conrado Moreno-Vivián. 2008. “Reduction of Polynitroaromatic Compounds: The Bacterial Nitroreductases.” *FEMS Microbiology Reviews* 32 (3): 474–500.

Romano, A. H., and T. Conway. 1996. “Evolution of Carbohydrate Metabolic Pathways.” *Research in Microbiology* 147 (6-7): 448–55.

Romero, Philip A., and Frances H. Arnold. 2009. “Exploring Protein Fitness Landscapes by Directed Evolution.” *Nature Reviews. Molecular Cell Biology* 10 (12): 866–76.

Rosell, Albert, Eva Valencia, Wendy F. Ochoa, Ignacio Fita, Xavier Parés, and Jaume Farrés. 2003. “Complete Reversal of Coenzyme Specificity by Concerted Mutation of Three Consecutive Residues in Alcohol Dehydrogenase.” *The Journal of Biological Chemistry* 278 (42): 40573–80.

Rosell, Albert, Eva Valencia, Xavier Parés, Ignacio Fita, Jaume Farrés, and Wendy F. Ochoa. 2003. “Crystal Structure of the Vertebrate NADP(H)-Dependent Alcohol Dehydrogenase (ADH8).” *Journal of Molecular Biology* 330 (1): 75–85.

Ruff, Anna J., Alexander Dennig, and Ulrich Schwaneberg. 2013. “To Get What We Aim for--Progress in Diversity Generation Methods.” *The FEBS Journal* 280 (13): 2961–78.

Samsonowicz, M., T. Hrynaszkiewicz, R. Świsłocka, E. Regulska, and W. Lewandowski. 2005. “Experimental and Theoretical IR, Raman, NMR Spectra of 2-, 3- and 4-Aminobenzoic Acids.” *Journal of Molecular Structure* 744-747 (June): 345–52.

Samsonowicz, M., R. Świsłocka, E. Regulska, and W. Lewandowski. 2007. “Experimental and Theoretical IR, Raman, NMR Spectra of 2-, 3-, and 4-Nitrobenzoic Acids.” *International Journal of Quantum Chemistry* 107 (2): 480–94.
Santoro, Stephen W., and Peter G. Schultz. 2002. “Directed Evolution of the Site Specificity of Cre Recombinase.” *Proceedings of the National Academy of Sciences of the United States of America* 99 (7): 4185–90.
Sauer, Uwe, Fabrizio Canonaco, Sylvia Heri, Annik Perrenoud, and Eliane Fischer. 2004. “The Soluble and Membrane-Bound Transhydrogenases UdhA and PntAB Have Divergent Functions in NADPH Metabolism of Escherichia Coli.” *The Journal of Biological Chemistry* 279 (8): 6613–19.
Sawers, G. 1994. “The Hydrogenases and Formate Dehydrogenases of Escherichia Coli.” *Antonie van Leeuwenhoek* 66 (1-3): 57–88.
Schrittwieser, Joerg H., Stefan Velikogne, and Wolfgang Kroutil. 2015. “Biocatalytic Imine Reduction and Reductive Amination of Ketones.” *Advanced Synthesis & Catalysis* 357 (8): 1655–85.
Schwede, Torsten. 2013. “Protein Modeling: What Happened to the ‘Protein Structure Gap” *Structure* 21 (9): 1531–40.
Scrutton, N. S., A. Berry, and R. N. Perham. 1990. “Redesign of the Coenzyme Specificity of a Dehydrogenase by Protein Engineering.” *Nature* 343 (6253): 38–43.
Searle, Peter F., Ming-Jen Chen, Longqin Hu, Paul R. Race, Andrew L. Lovering, Jane I. Grove, Chris Guise, et al. 2004. “Nitroreductase: A Prodrug-Activating Enzyme for Cancer Gene Therapy.” *Clinical and Experimental Pharmacology & Physiology* 31 (11): 811–16.
Seelbach, Karsten, Bettina Riebel, Werner Hummel, Maria-Regina Kula, Vladimir I. Tishkov, Alexey M. Egorov, Christian Wandrey, and Udo Kragl. 1996. “A Novel, Efficient Regenerating Method of NADPH Using a New Formate Dehydrogenase.” *Tetrahedron Letters* 37 (9): 1377–80.
Seelig, Burckhard, and Jack W. Szostak. 2007. “Selection and Evolution of Enzymes from a Partially Randomized Non-Catalytic Scaffold.” *Nature* 448 (7155): 828–31.
Sellés Vidal, Lara, Ciarán L. Kelly, Paweł M. Mordaka, and John T. Heap. 2018. “Review of NAD(P)H-Dependent Oxidoreductases: Properties, Engineering and Application.” *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*.
Shalel-Levanon, Sagit, Ka-Yiu San, and George N. Bennett. 2005. “Effect of ArcA and FNR on the Expression of Genes Related to the Oxygen Regulation and the Glycolysis Pathway in Escherichia Coli under Microaerobic Growth Conditions.” *Biotechnology and Bioengineering* 92 (2): 147–59.
Shi, Lei, Yingying Liu, Qingfeng Liu, Bin Wei, and Guisheng Zhang. 2012. “Selective Reduction of Aldehydes and Ketones to Alcohols with Ammonia Borane in Neat Water.” *Green Chemistry: An International Journal and Green Chemistry Resource: GC* 14 (5): 1372–75.
Shortle, D., and J. Sondek. 1995. “The Emerging Role of Insertions and Deletions in Protein Engineering.” *Current Opinion in Biotechnology* 6 (4): 387–93.
Sieber, V., C. A. Martinez, and F. H. Arnold. 2001. “Libraries of Hybrid Proteins from Distantly Related Sequences.” *Nature Biotechnology* 19 (5): 456–60.
Smith, Anthony C., Filmon Eyassu, Jean-Pierre Mazat, and Alan J. Robinson. 2017. “MitoCore: A Curated Constraint-Based Model for Simulating Human Central Metabolism.” *BMC Systems Biology* 11 (1): 114.
Smith, D. G., A. D. Gribble, D. Haigh, R. J. Ife, P. Lavery, P. Skett, B. P. Slingsby, R. Stacey, R. W. Ward, and A. West. 1999. “The Inhibition of Human Cytomegalovirus (hCMV) Protease by Hydroxylamine Derivatives.” *Bioorganic & Medicinal Chemistry Letters* 9 (21): 3137–42.
Smith, G. P. 1985. “Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface.” *Science* 228 (4705): 1315–17.
Smith, J. M. 1970. “Natural Selection and the Concept of a Protein Space.” *Nature* 225 (5232): 563–64.

Sofer, W., and H. Ursprung. 1968. “Drosophila Alcohol Dehydrogenase. Purification and Partial Characterization.” *The Journal of Biological Chemistry* 243 (11): 3110–15.
Solanki, Kusum, Walaa Abdallah, and Scott Banta. 2017. “Engineering the Cofactor Specificity of an Alcohol Dehydrogenase via Single Mutations or Insertions Distal to the 2’-Phosphate Group of NADP(H).” *Protein Engineering, Design & Selection: PEDS* 30 (5): 373–80.
Spain, J. C. 1995. “Biodegradation of Nitroaromatic Compounds.” *Annual Review of Microbiology* 49: 523–55.
Sramek, S. J., and F. E. Frerman. 1975. “Purification and Properties of Escherichia Coli Coenzyme A-Transferase.” *Archives of Biochemistry and Biophysics* 171 (1): 14–26.
Stapley, Jessica, Philine G. D. Feulner, Susan E. Johnston, Anna W. Santure, and Carole M. Smadja. 2017. “Recombination: The Good, the Bad and the Variable.” *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 372 (1736).
Steinberg, Barrett, and Marc Ostermeier. 2016. “Environmental Changes Bridge Evolutionary Valleys.” *Science Advances* 2 (1): e1500921.
Steitz, T. A., R. Henderson, and D. M. Blow. 1969. “Structure of Crystalline Alpha-Chymotrypsin. 3. Crystallographic Studies of Substrates and Inhibitors Bound to the Active Site of Alpha-Chymotrypsin.” *Journal of Molecular Biology* 46 (2): 337–48.
Stemmer, W. P. 1994. “Rapid Evolution of a Protein in Vitro by DNA Shuffling.” *Nature* 370 (6488): 389–91.
Stepankova, Veronika, Sarka Bidmanova, Tana Koudelakova, Zbynek Prokop, Radka Chaloupkova, and Jiri Damborsky. 2013. “Strategies for Stabilization of Enzymes in Organic Solvents.” *ACS Catalysis* 3 (12): 2823–36.
Sun, Song, Wei Zhang, Bengt Mannervik, and Dan I. Andersson. 2013. “Evolution of Broad Spectrum β-Lactam Resistance in an Engineered Metallo-β-Lactamase.” *The Journal of Biological Chemistry* 288 (4): 2314–24.
Swe, P. M., J. N. Copp, L. K. Green, C. P. Guise, A. M. Mowday, J. B. Smaill, A. V. Patterson, and D. F. Ackerley. 2012. “Targeted Mutagenesis of the Vibrio Fischeri Flavin Reductase FRase I to Improve Activation of the Anticancer Prodrug CB1954.” *Biochemical Pharmacology* 84 (6): 775–83.
Tan, Tianwei, Jike Lu, Kaili Nie, Li Deng, and Fang Wang. 2010. “Biodiesel Production with Immobilized Lipase: A Review.” *Biotechnology Advances* 28 (5): 628–34.
Tang, Lixia, Hui Gao, Xuechen Zhu, Xiong Wang, Ming Zhou, and Rongxiang Jiang. 2012. “Construction of ‘Small-Intelligent’ Focused Mutagenesis Libraries Using Well-Designed Combinatorial Degenerate Primers.” *BioTechniques* 52 (3): 149–58.
Tawfik, D. S., and A. D. Griffiths. 1998. “Man-Made Cell-like Compartments for Molecular Evolution.” *Nature Biotechnology* 16 (7): 652–56.
Taylor, George M., Paweł M. Mordaka, and John T. Heap. 2019. “Start-Stop Assembly: A Functionally Scarless DNA Assembly System Optimized for Metabolic Engineering.” *Nucleic Acids Research* 47 (3): e17.
Unden, G., and J. Bongaerts. 1997. “Alternative Respiratory Pathways of Escherichia Coli: Energetics and Transcriptional Regulation in Response to Electron Acceptors.” *Biochimica et Biophysica Acta* 1320 (3): 217–34.
Unden, G., and M. Trageser. 1991. “Oxygen Regulated Gene Expression in Escherichia Coli: Control of Anaerobic Respiration by the FNR Protein.” *Antonie van Leeuwenhoek* 59 (2): 65–76.
Urban, Johannes H., and Christoph A. Merten. 2011. “Retroviral Display in Gene Therapy, Protein Engineering, and Vaccine Development.” *ACS Chemical Biology* 6 (1): 61–74.
Venekei, I., L. Szilágyi, L. Gráf, and W. J. Rutter. 1996. “Attempts to Convert Chymotrypsin to Trypsin.” *FEBS Letters* 383 (1-2): 143–47.
Virnekäs, B., L. Ge, A. Plückthun, K. C. Schneider, G. Wellnhofer, and S. E. Moroney. 1994. “Trinucleotide Phosphoramidites: Ideal Reagents for the Synthesis of Mixed Oligonucleotides for Random Mutagenesis.” *Nucleic Acids Research* 22 (25): 5600–5607.

Vonrhein, C., and G. Bricogne. 2008. “AutoPROC—a Framework for Automated Data Processing.” *Acta Crystallographica. Section A, Crystal Physics, Diffraction, Theoretical and General Crystallography* 64: C78.

Vonrhein, Clemens, Claus Flensburg, Peter Keller, Andrew Sharff, Oliver Smart, Wlodek Paciorek, Thomas Womack, and Gérard Bricogne. 2011. “Data Processing and Analysis with the autoPROC Toolbox.” *Acta Crystallographica*. Section D, Biological

Waterhouse, Andrew, Martino Bertoni, Stefan Bienert, Gabriel Studer, Gerardo Tauriello, Rafal Gumienny, Florian T. Heer, et al. 2018. “SWISS-MODEL: Homology Modelling of Protein Structures and Complexes.” *Nucleic Acids Research* 46 (W1): W296–303.

Wehrmann, Matthias, and Janosch Klebensberger. 2018. “Engineering Thermal Stability and Solvent Tolerance of the Soluble Quinoprotein PedE from Pseudomonas Putida KT2440 with a Heterologous Whole-Cell Screening Approach.” *Microbial Biotechnology* 11 (2): 399–408.

Wickham, H., ggplot2: Elegant Graphics for Data Analysis.Springer-Verlag New York, 2016.

Wilkinson, A. J., A. R. Fersht, D. M. Blow, P. Carter, and G. Winter. 1984. “A Large Increase in Enzyme-Substrate Affinity by Protein Engineering.” *Nature* 307 (5947): 187–88.

Wilks, H. M., D. J. Halsall, T. Atkinson, W. N. Chia, A. R. Clarke, and J. J. Holbrook. 1990. “Designs for a Broad Substrate Specificity Keto Acid Dehydrogenase.” *Biochemistry* 29 (37): 8587–91.

Wilks, H. M., K. W. Hart, R. Feeney, C. R. Dunn, H. Muirhead, W. N. Chia, D. A. Barstow, T. Atkinson, A. R. Clarke, and J. J. Holbrook. 1988. “A Specific, Highly Active Malate Dehydrogenase by Redesign of a Lactate Dehydrogenase Framework.” *Science* 242 (4885): 1541–44.

Wilks, H. M., K. M. Moreton, D. J. Halsall, K. W. Hart, R. D. Sessions, A. R. Clarke, and J. J. Holbrook. 1992. “Design of a Specific Phenyllactate Dehydrogenase by Peptide Loop Exchange on the Bacillus Stearothermophilus Lactate Dehydrogenase Framework.” *Biochemistry* 31 (34): 7802–6.

Williams, Elsie M., Rory F. Little, Alexandra M. Mowday, Michelle H. Rich, Jasmine V. E. Chan-Hyams, Janine N. Copp, Jeff B. Smaill, Adam V. Patterson, and David F. Ackerley. 2015. “Nitroreductase Gene-Directed Enzyme Prodrug Therapy: Insights and Advances toward Clinical Utility.” *Biochemical Journal* 471 (2): 131–53.

Wilson, D. S., A. D. Keefe, and J. W. Szostak. 2001. “The Use of mRNA Display to Select High-Affinity Protein-Binding Peptides.” *Proceedings of the National Academy of Sciences of the United States of America* 98 (7): 3750–55.

Winter, G. 2009. “xia2: An Expert System for Macromolecular Crystallography Data Reduction.” *Journal of Applied Crystallography* 43 (1): 186–90.

Wishart, David S., Yannick Djoumbou Feunang, Ana Marcu, An Chi Guo, Kevin Liang, Rosa Vázquez-Fresno, Tanvir Sajed, et al. 2018. “HMDB 4.0: The Human Metabolome Database for 2018.” *Nucleic Acids Research* 46 (D1): D608–17.

Woodyer, Ryan, Huimin Zhao, and Wilfred A. Van Der Donk. 2005. “Mechanistic Investigation of a Highly Active Phosphite Dehydrogenase Mutant and Its Application for NADPH Regeneration: Investigation of an NADPH Regeneration Catalyst.” *The FEBS Journal* 272 (15): 3816–27.

Wu, Hong, Chunyong Tian, Xiaokai Song, Chuang Liu, Dong Yang, and Zhongyi Jiang. 2013. “Methods for the Regeneration of Nicotinamide Coenzymes.” *Green Chemistry: An International Journal and Green Chemistry Resource*: GC 15 (7): 1773–89.

Xiao, Han, Zehua Bao, and Huimin Zhao. 2015. “High Throughput Screening and Selection Methods for Directed Enzyme Evolution.” *Industrial & Engineering Chemistry Research* 54 (16): 4011–20.

Xu, Lihui, Patti Aha, Ke Gu, Robert G. Kuimelis, Markus Kurz, Terence Lam, Ai Ching Lim, *et al.* 2002. “Directed Evolution of High-Affinity Antibody Mimics Using mRNA Display.” *Chemistry & Biology* 9 (8): 933–42.

Yang, Guangyu, and Stephen G. Withers. 2009. “Ultrahigh-Throughput FACS-Based Screening for Directed Enzyme Evolution.” *Chembiochem: A European Journal of Chemical Biology* 10 (17): 2704–15.

Yang, Taek Ho, Chelladurai Rathnasingh, Hee Jong Lee, and Doyoung Seung. 2014. “Identification of Acetoin Reductases Involved in 2,3-Butanediol Pathway in Klebsiella Oxytoca.” *Journal of Biotechnology* 172 (February): 59–66.

Ying, Xiangxian, Amy M. Grunden, Lin Nie, Michael W. W. Adams, and Kesen Ma. 2009. “Molecular Characterization of the Recombinant Iron-Containing Alcohol Dehydrogenase from the Hyperthermophilic Archaeon, Thermococcus Strain ES1.” *Extremophiles: Life under Extreme Conditions* 13 (2): 299–311.

Ying, Xiangxian, Ying Wang, Hamid R. Badiei, Vassili Karanassios, and Kesen Ma. 2007. “Purification and Characterization of an Iron-Containing Alcohol Dehydrogenase in Extremely Thermophilic Bacterium Thermotoga Hypogea.” *Archives of Microbiology* 187 (6): 499–510.

You, Shuang-Hong, Bo Zhu, Hong-Juan Han, Bo Wang, Ri-He Peng, and Quan-Hong Yao. 2015. “Phytoremediation of 2,4,6-Trinitrotoluene by Arabidopsis Plants Expressing a NAD(P)H-Flavin Nitroreductase from Enterobacter Cloacae.” *Plant Biotechnology Reports* 9 (6): 417–30.

Zenno, S., H. Koike, M. Tanokura, and K. Saigo. 1996. “Conversion of NfsB, a Minor Escherichia Coli Nitroreductase, to a Flavin Reductase Similar in Biochemical Properties to FRase I, the Major Flavin Reductase in Vibrio Fischeri, by a Single Amino Acid Substitution.” *Journal of Bacteriology* 178 (15): 4731–33.

Zhao, H., L. Giver, Z. Shao, J. A. Affholter, and F. H. Arnold. 1998. “Molecular Evolution by Staggered Extension Process (StEP) in Vitro Recombination.” *Nature Biotechnology* 16 (3): 258–61.

Zhao, X., X. Zhang, Z. Rao, T. Bao, X. Li, M. Xu, T. Yang, and S. Yang. 2015. “Identification and Characterization of a Novel 2,3-Butanediol Dehydrogenase/acetoin Reductase from Corynebacterium Crenatum SYPA5-5.” *Letters in Applied Microbiology* 61 (6): 573–79.

Zheng, Yu-Guo, Huan-Huan Yin, Dao-Fu Yu, Xiang Chen, Xiao-Ling Tang, Xiao-Jian Zhang, Ya-Ping Xue, Ya-Jun Wang, and Zhi-Qiang Liu. 2017. “Recent Advances in Biotechnological Applications of Alcohol Dehydrogenases.” *Applied Microbiology and Biotechnology* 101 (3): 987–1001.

Zhou, Jiewen, Kang Wu, and Christopher V. Rao. 2016. “Evolutionary Engineering of Geobacillus Thermoglucosidasius for Improved Ethanol Production.” *Biotechnology and Bioengineering* 113 (10): 2156–67.