

NNT : 2017SACLE016

**THÈSE DE DOCTORAT
DE
L'UNIVERSITÉ PARIS-SACLAY
PRÉPARÉE A
L'UNIVERSITÉ D'ÉVRY-VAL-D'ESSONNE**

**ÉCOLE DOCTORALE n° 577
SDSV – Structure et dynamique des systèmes vivants
Specialité de doctorat : Science de la vie et de la santé**

Par

Mme Tiffany Souterre

Optimization of enzymes via directed evolution *in vivo*

Thèse présentée et soutenue au Genoscope, le 28 juin 2017:

Composition du jury :

| | | | |
|------------------------------|------------|-------------------|------------------------|
| Mr Jean-Luc Pernodet | DR | CNRS Paris-Saclay | Président |
| Mr Pierre Alexandre Kaminski | Chercheur | Institut Pasteur | Rapporteur |
| Mr Olivier Tenaillon | DR | Inserm Paris 6 | Rapporteur |
| Mme Magali Remaud-Siméon | Professeur | INSA Toulouse | Examinateuse |
| Mr Marcel Salanoubat | DR | CNRS Paris-Saclay | Directeur de thèse |
| Mme Madeleine Bouzon Bloch | Chercheuse | CEA | Co-encadrante de thèse |

ACKNOWLEDGMENTS

My gratitude to Jean Weissenbach, director of Genoscope, for offering the possibility to pursue my thesis in such a great institute and for giving all the support necessary for this endeavor.

I would like to thank the reporters of my jury, Pierre Alexandre Kaminski and Olivier Tenaillon for accepting to review this manuscript, for their patience and understanding. I also thank Magali Remaud-Siméon and Jean-Luc Pernodet for showing interest in my work and accepting to be part of this jury.

A big thank you to Marcel Salanoubat, my thesis director, for his support throughout those three and a half years. Despite many responsibilities and other Ph.D. students to take care of, he always warmly welcomed me to discuss with a smile and great sense of humor. I strongly appreciated those moments.

I also thank Madeleine Bouzon and Volker Döring for their co-supervision and without whom this project would not have been possible. I am thankful for their disponibility, great scientific discussions and sharing their knowledge about metabolism and directed evolution with me. I am also grateful for the time spent on these last months supporting me for writing my thesis.

I would like to show my gratitude to Alain Perret for stepping in when Ido was putting me through the wringer. His expertise, disponibility, understanding, and optimism were the perfect mix I needed to keep me going. He graciously took the time to explain everything I needed to know about protein purification, mass spectrometry, and biochemistry. I am also grateful for his smile and great laugh that never fails to catch me. I would like to thank Ekaterina Dairy for her help in mass spectrometry, her kindness, and patience.

A great thank you to Gabor Gyapay for his expertise with the Cell Sorter. I greatly appreciated our heated scientific debates that helped me grow scientifically and build confidence. I am grateful for the time he spent helping me with the data, his disponibility and for recommending me interesting books to read.

I am missing words to express how grateful I am to have had the chance to work with Andrew Tolonen. He has never failed to believe in me, always encouraging me to reach for higher goals and cheering me up when I was feeling down. I will miss our long conversations, scientific and personal. I am very thankful for everything he taught me.

I would like to express my gratitude to Karine Labadie for strains sequencing and Stéphane Cruveiller and David Roche for their analysis and help with the MAGE platform. I also thank Aurélie Jouenne, Pierre-Loïc Saaïdi, Carine Vaxelaire and Anne Zaparucha for their help in chemistry and their expertise with the dioxygenases.

A warm thank you to my colleagues who participated in this project, I am indebted to them for their help: Adrien for sharing plasmids and primers, Aline for biochemistry advises, Anne for building strains, for her constant support and great conversations, Christine for protein purification (when it works!) and laughs, Christophe for metabolomes, mass spectrometry and cat conversations ("La paillasse insoumise!"), Emilie for qPCR help and protocols, Isabelle for handling mediums and making great cakes, Ivan for GM3 cultures, for driving me back home when I needed a ride and for being such a good friend, Jean-Louis for... too much to fit on this page but mostly for protocols and mixing Kdo and Ido (yeah, thank you...), Laurence for the MET project, her compassion for me during my thesis and advises on how to take care of my plants, Laurent for GM3 cultures and kindness, Magali for her help with protocols and great restaurants/dinners together, Marielle for protocols, Nadia and Pascal for being amazing Ido partners, Nathalie for GM3 cultures and Cell Sorting, Peggy for protein purification and kindness, Valérie for great help and expertise with protocols and for being so kind to me, Véronique for biochemistry advises and support, Wahiba for optical mapping and to all my colleagues, thank you for making this adventure unforgettable.

A warm thank you to my Ph.D. student fellows: Marion.C, Marion.T, Ombeline, Sophia, Thomas and Tristan for the nice moments spent together, the constant support during rough patches, interesting scientific conversations and for their friendship. Special thanks to Sophia, my Couscous, for her communicative joy, her hilarious stories, for being an amazing friend and for keeping sure I stayed well-fed during those last months writing my thesis. Thank you Thomas, for making sure I had my daily dose of laughter (even unintentionally) after a harassing day of writing. Thank you Tristan for being

there since iGEM, so much happened since then. Thank you for playing such a big part in who I am today.

A big thank you to the LA interns (former and present), Cécile, Emilie, Hela, Parvathi, Sonia and Valentin for the good moments spent together and their encouragements. Special thanks to Cécile Jacry who did a great job on the MET project.

A warm thank you to my dearest friends outside Genoscope: Alice for shining like a sun and advises for my defense, Camille for being the sweetest friend, Cécile for being always there for me, Audam for being my lost little brother and for proofreading my thesis.

I would like to thank my family for their unconditional love and support. Thank you to my grandparents for being so supportive. Mom, thank you for always believing in me, no matter what I do and Dad, thank you for always making me see the big picture and helping me keep going forward. Kevin, thank you for being the best brother, always supportive and confident about what I can achieve, even when I doubt.

Last but not least, the love of my life, Olivier. Thank you for your unconditional love and support, for never failing me and for being such an incredible partner in life. Thank you for sailing beside me through the ups and downs of my research. No doubt that even the hardest day was, after all, not so bad in your arms. Finally Katoun, my rescued kitten found in the Genoscope garbage, thank you for bringing me comfort, love and even trying to write parts of my thesis.

TABLE OF CONTENT

| | |
|--|----|
| ACKNOWLEDGMENTS..... | 3 |
| TABLE OF CONTENT | 6 |
| LIST OF FIGURES | 9 |
| LIST OF TABLES..... | 12 |
| LIST OF APPENDIX..... | 14 |
| LIST OF ABBREVIATIONS | 15 |
| I INTRODUCTION | 18 |
| I.1 General introduction..... | 19 |
| I.2 Tailoring enzymes | 21 |
| I.2.1 <i>De novo</i> design..... | 21 |
| I.2.2 Rational design..... | 22 |
| I.2.3 Directed evolution..... | 24 |
| I.2.4 Directed evolution <i>in vivo</i> | 30 |
| I.3 Directed evolution in the GM3 device..... | 38 |
| I.3.1 Examples of applications | 38 |
| I.3.2 Directed evolution projects | 39 |
| I.4 Iron (II)/ α -keto acid-dependent oxygenases (α -KAOs) | 40 |
| I.4.1 General overview..... | 40 |
| I.4.2 Catalyzed reactions | 41 |
| I.4.3 Catalytic mechanism..... | 43 |
| I.4.4 Candidates for evolution | 44 |
| I.5 Fixation of beneficial mutations in a gene under selection in an evolving cell population..... | 45 |
| I.5.1 Experimental strategy | 45 |

| | | |
|---------|--|----|
| II | MATERIALS & METHODS | 48 |
| II.1 | Molecular biology protocols | 49 |
| II.1.1 | Gene cloning..... | 49 |
| II.1.2 | Directed mutagenesis | 55 |
| II.2 | Molecular genetics protocols..... | 57 |
| II.2.1 | Genetic screen construction..... | 57 |
| II.2.2 | Allelic exchange..... | 60 |
| II.3 | Dioxygenase biochemistry | 64 |
| II.3.1 | Protein production | 64 |
| II.3.2 | Succinic acid assay | 67 |
| II.3.3 | Kinetic parameters determined with mass spectrometry | 68 |
| II.4 | Bioscreen C [©] | 71 |
| II.5 | Metabolomics analysis..... | 72 |
| II.5.1 | Metabolome preparation | 72 |
| II.5.2 | HPLC conditions..... | 72 |
| II.6 | Quantitative PCR and quantitative reverse transcription PCR | 73 |
| II.6.1 | mRNA Stabilization..... | 74 |
| II.6.2 | mRNA Extraction..... | 74 |
| II.6.3 | DNA Degradation..... | 74 |
| II.6.4 | Reverse transcription | 75 |
| II.6.5 | Quantitative PCR (qPCR)..... | 75 |
| II.7 | Cell Sorting and Sequencing | 76 |
| III | RESULTS..... | 78 |
| III.1 | Fixation of beneficial mutations in a gene under selection in an evolving cell population..... | 79 |
| III.1.1 | Genetic screen construction..... | 79 |
| III.1.2 | Variants construction | 80 |

| | | |
|---------|--|-----|
| III.1.3 | Expression formats and growth tests..... | 81 |
| III.1.4 | Reversion dynamics in continuous culture..... | 86 |
| III.1.5 | Intracellular distribution of high copy-number plasmids..... | 94 |
| III.2 | <i>In vivo</i> genetic screen for α -KAOs | 97 |
| III.2.1 | Genetic screen design for α -KAOs..... | 97 |
| III.2.2 | Selection strain construction..... | 101 |
| III.2.3 | Selection strain stability | 101 |
| III.2.4 | Complementation tests of α -KAOs expressed in selection strains..... | 107 |
| III.3 | Biochemical analysis of α -KAOs | 108 |
| III.4 | Evolution <i>in vivo</i> of α -KAOs in a <i>ppc-pck</i> - selection context | 112 |
| III.4.1 | Evolution <i>in vivo</i> of α -KAOs expressed from a plasmidic format..... | 112 |
| III.4.2 | Evolution <i>in vivo</i> of α -KAOs expressed in a chromosomal format | 124 |
| III.4.3 | Evolution <i>in vivo</i> of Kdo2 towards lysine analogs | 131 |
| III.5 | L-isoleucine dioxygenase (Ido) in a <i>ppc-pck</i> - selection context..... | 135 |
| III.5.1 | Genetic screen construction for Ido | 137 |
| III.5.2 | Complementation tests of selection strain expressing Ido | 137 |
| III.5.3 | Evolution <i>in vivo</i> of Ido expressed in a chromosomal format | 139 |
| III.5.4 | Evolution <i>in vivo</i> of Ido towards isoleucine analogs | 147 |
| III.6 | Evolution <i>in vivo</i> of Ido and Kdo2 in a <i>pck+</i> selection context..... | 155 |
| IV | Discussion..... | 195 |
| | Appendix..... | 205 |
| | Bibliography | 216 |

LIST OF FIGURES

| | |
|--|----|
| Figure 1: Process for <i>de novo</i> enzyme design | 22 |
| Figure 2: Schematic overview of directed evolution..... | 25 |
| Figure 3: Diagram of a chemostat | 31 |
| Figure 4: Cultivation fluidics with twin growth vessels in the GM3 cultivation device.... | 34 |
| Figure 5: Pulse-feeding clock of turbidostat (A) and medium SWAP regime (B) of the GM3 automates..... | 37 |
| Figure 6: Hydroxylation reaction catalyzed by α -KAO | 41 |
| Figure 7: The diversity of the reactions catalyzed by α -KAOs | 42 |
| Figure 8: Proposed consensus catalytic mechanism of the α -KAO catalyzed hydroxylation reaction..... | 43 |
| Figure 9: Simplified scheme of the experimental design for studying the fixation of a point mutation in a gene of interest..... | 47 |
| Figure 10: Map of plasmid pSP100..... | 49 |
| Figure 11: Map of plasmid pVDM18..... | 50 |
| Figure 12: Map of plasmid pEVL404..... | 50 |
| Figure 13: Simplified workflow of strain engineering for directed evolution with plasmids | 51 |
| Figure 14: Simplified workflow of directed mutagenesis for directed evolution <i>in vivo</i> with plasmids..... | 55 |
| Figure 15: Design of mutagenesis primers..... | 56 |
| Figure 16: Map of plasmid pCP20 | 59 |
| Figure 17: Construction of single-gene deletion mutants | 60 |
| Figure 18: Simplified workflow of strain engineering for directed evolution with chromosomal insertion | 61 |
| Figure 19: Map of plasmid pEVL411 | 62 |
| Figure 20: Ligation Independent Cloning workflow..... | 64 |
| Figure 21: Bioscreen C [®] Analyzer Components | 71 |
| Figure 22: Reaction catalyzed by the L-homoserine O-succinyltransferase | 79 |
| Figure 23: Growth of descendants of selection strain G3724 ($\Delta metA$) expressing the gene <i>metA</i> WT from different expression formats at 30°C and 37°C | 84 |
| Figure 24: Growth of descendants of selection strain G3724 ($\Delta metA$) expressing the allele <i>metA:E237D</i> from different expression formats at 30°C and 37°C..... | 85 |

| | |
|--|-----|
| Figure 25: Evolution of cultures MET 6, MET 7 and MET 10 in turbidostat regime | 87 |
| Figure 26: Analysis of the evolution of culture MET 10 in turbidostat regime at 30°C.... | 88 |
| Figure 27: Evolution of cultures MET 15, MET 17 and MET 18 in medium SWAP regime | 90 |
| Figure 28: Analysis of the evolution of culture MET 15 in medium SWAP regime at 37°C | 92 |
| Figure 29: Analysis of the evolution of culture MET 18 in medium SWAP regime at 37°C | 93 |
| Figure 30: Sanger sequencing chromatogram of the gene <i>metA</i> at position 711 from culture MET 10 at generation 460..... | 94 |
| Figure 31: Moving average of the percentage of G (corresponding to the revertant <i>metA</i>) at position 711 in the gene <i>metA</i> at generation 186 from culture MET 10 | 96 |
| Figure 32: Succinate selection strain construction <i>ppc</i> - <i>pck</i> - selection context..... | 99 |
| Figure 33: Succinate selection strain construction <i>pck</i> + selection context..... | 100 |
| Figure 34: Evolution of the culture of strain G3513 in medium SWAP regime | 102 |
| Figure 35: Overview of the non-synonym intragenic SNPs along the evolution of the strain G3513 | 105 |
| Figure 36: Comparison of the turnover rate (k_{cat}) between the α -KAOs (Kdo1 to Kdo5) and the average enzyme | 110 |
| Figure 37: Comparison of the enzyme efficiency (k_{cat}/K_M) between the α -KAOs (Kdo1 to Kdo5) and the average enzyme | 110 |
| Figure 38: Enzymes operating within different metabolic groups have significantly different k_{cat} and k_{cat}/K_M values..... | 111 |
| Figure 39: Overview of α -KAOs evolution experiments in plasmidic format..... | 113 |
| Figure 40: Evolution of cultures ODO 8, KDA 10 and KDB 10 in medium SWAP regime | 115 |
| Figure 41: Map of plasmid pGEN1126 | 116 |
| Figure 42: Map of plasmid pGEN1080 | 117 |
| Figure 43: Copy number comparison of gene <i>kdo2</i> between strains G3700, G4096 and G4097 relative to strain G3986..... | 120 |
| Figure 44: Copy number comparison of gene <i>thyA</i> between strains G3700, G4096 and G4097 relative to strain G3986..... | 121 |

| | |
|--|-----|
| Figure 45: mRNA expression comparison of gene <i>kdo2</i> in strain G4096 relative to strain G3986 | 121 |
| Figure 46: ¹³ C enrichment of succinate in metabolome of evolved strain G4097 and control strains..... | 123 |
| Figure 47: Overview of Kdo2 evolution in chromosomal expression format..... | 125 |
| Figure 48: Evolution of cultures KDB 15 and KDB 16 in medium SWAP regime | 126 |
| Figure 49: Scheme of upstream modifications detected in isolates G4275, G4276 and G4292 from culture KDB 15 and KDB 16 | 128 |
| Figure 50: Reaction catalyzed by <i>pdeH</i> | 129 |
| Figure 51: Comparison of mRNA expression level of <i>kdo2</i> in isolate G4275 after evolution relative to the initial strain G4160 | 130 |
| Figure 52: Growth of strain G4275 in the presence of lysine analogs | 132 |
| Figure 53: Evolution of culture KDB 21 and KDB 22 in medium SWAP regime | 134 |
| Figure 54: Reaction catalyzed by Ido | 135 |
| Figure 55: “Shunting” of the TCA cycle in the 2Δ strain expressing Ido | 136 |
| Figure 56: Overview of the evolution of Ido in the chromosomal expression format.... | 140 |
| Figure 57: Evolution of cultures IDO 3 and IDO 6 in medium SWAP regime..... | 141 |
| Figure 58: Map of the <i>ido</i> locus on the chromosome of strain G4110 and strains G4170/G4173 | 143 |
| Figure 59: Comparison of mRNA expression level of <i>ido</i> in the strain G4170 from culture IDO 3 relative to the initial strain G4110 | 144 |
| Figure 60: ¹³ C enrichment of succinate in the metabolome of evolved isolates of cultures IDO 3 and IDO 6 and control strains | 146 |
| Figure 61: Extracted-Ion Chromatogram of 4-OH-isoleucine in the metabolome of strains before and after evolution in cultures IDO 3 and IDO 6 | 147 |
| Figure 62: Evolution of cultures IDO 9, IDO 11 and IDO 13 in medium SWAP regime...150 | |
| Figure 63: Evolution of cultures KDB 26, KDB 27, IDO 17 and IDO 18 in medium SWAP regime | 156 |

LIST OF TABLES

| | |
|--|-----|
| Table 1: Methods for library diversification | 27 |
| Table 2: Candidates enzymes for directed evolution <i>in vivo</i> in GM3 | 45 |
| Table 3: List of vector used for directed evolution <i>in vivo</i> | 49 |
| Table 4: List of primers used for cloning genes..... | 52 |
| Table 5: Table of primers designed for directed mutagenesis of <i>metA</i> | 56 |
| Table 6: Primers used to clone the His-tag..... | 65 |
| Table 7: Extended MRM conditions for the detection of 4-OH-ile and OH-norv (positive mode) | 70 |
| Table 8: Primers used in this study for qRT-PCR..... | 73 |
| Table 9: Mutations introduced in the gene <i>metA</i> | 80 |
| Table 10: Expression formats used in this study | 81 |
| Table 11: Methionine auxotrophy complementation of strains expressing different <i>metA</i> alleles..... | 82 |
| Table 12: Strains used for growth test..... | 83 |
| Table 13: Generation time of descendants of selection strain G3724 (Δ <i>metA</i>) expressing either <i>metA</i> WT or <i>metA:E237D</i> from different expression formats at 30°C..... | 86 |
| Table 14: GM3 cultures launched to study the dynamics of genetic reversion..... | 86 |
| Table 15: Genotype of the different selection strains used for α -KAO evolution..... | 101 |
| Table 16: Growth tests of isolates from evolution of strain G3513 | 103 |
| Table 17: Growth tests for the utilization of citrate as succinate source | 103 |
| Table 18: Succinate complementation tests of α -KAOs expressed from different formats | 108 |
| Table 19: α -KAOs kinetic parameters..... | 109 |
| Table 20: Strains used for α -KAO evolution <i>in vivo</i> in plasmidic format | 113 |
| Table 21: Succinate complementation tests for isolates from culture KDB 10 | 119 |
| Table 22: Succinate complementation tests of isolates from cultures KDB 15 and KDB 16 | 127 |
| Table 23: Succinate complementation test of strain G4434 | 131 |
| Table 24: Lysine analogs tested for evolution..... | 132 |
| Table 25: Succinate complementation tests of strains G4275 and G4276 with lysine analogs..... | 133 |
| Table 26: Genotype of the selection strain G3990 used for Ido evolution..... | 137 |

| | |
|---|-----|
| Table 27: Succinate complementation tests with Ido | 138 |
| Table 28: Succinate complementation tests of isolates from cultures IDO 3 and IDO 6.142 | |
| Table 29: Succinate complementation test of strain G4351..... | 145 |
| Table 30: Genotype of strains used for the evolution of Ido towards isoleucine analogs | 148 |
| Table 31: Isoleucine analogs tested for evolution | 149 |
| Table 32: Succinate complementation for norvaline-positive isolates | 152 |
| Table 33: Succinate complementation tests with norvaline for strain G4525 | 153 |
| Table 34: Ido and Ido:E65K kinetic parameters..... | 154 |
| Table 35: Complementation tests for genes <i>ido</i> and <i>kdo2</i> in <i>pck+</i> selection context.... | 155 |
| Table 36: Kinetic parameters of various dioxygenases | 197 |

LIST OF APPENDIX

| | |
|--|-----|
| Appendix 1: List of all strains cited in this study | 205 |
| Appendix 2: List of all plasmids cited in this study..... | 209 |
| Appendix 3: List of all primers cited in this study | 211 |
| Appendix 4: List of all GM3 cultures cited in this study | 213 |
| Appendix 5: NTA mix composition (1000X)..... | 215 |
| Appendix 6: MS medium composition | 215 |
| Appendix 7: MA medium composition..... | 215 |

LIST OF ABBREVIATIONS

| Abbreviation | Full name |
|-------------------|---|
| 3-OH-arg | 3-hydroxy-L-arginine |
| 3-OH-lys | 3-hydroxy-L-lysine |
| 3-OH-orn | 3-hydroxy-L-ornithine |
| 4-OH-ile or 4-HIL | 4-hydroxy-L-isoleucine |
| 4-OH-lys | 4-hydroxy-L-lysine |
| α -KAO | Iron(II)/ α -ketoacid-dependent Oxygenases |
| α -KG | α -ketoglutarate |
| ACN | Acetonitrile |
| AFN | Amino acids, Fatty acids, and Nucleotides |
| ADO | Assembly of Designed Oligonucleotides |
| ADP | Adenosine-5'-diphosphate |
| Asp | Aspartate |
| ATP | Adenosine-5'-triphosphate |
| BLAST | Basic Local Alignment Search Tool |
| bp | Base pair |
| BRET | Bioluminescence Resonance Energy Transfer |
| CE | Carbon and Energy |
| CE | Collision energy |
| c-di-GMP | Cyclic-diguanosine monophosphate |
| cDNA | Complementary DNA |
| CoA | Coenzyme A |
| CXP | Collision Cell Exit Potential |
| CYP | Cytochrome P450 monooxygenase |
| DAP | Diaminopimelic acid |
| DNA | Deoxyribonucleic acid |
| DP | Declustering Potential |
| DTT | Dithiothreitol |
| dNTP | Deoxynucleotide triphosphate |
| dATP | Deoxyadenosine triphosphate |
| dCTP | Deoxycytidine triphosphate |
| dGTP | Deoxyguanosine triphosphate |
| dTTP | Deoxythymidine triphosphate |
| <i>E. coli</i> | <i>Escherichia coli</i> |
| EDTA | Ethylenediaminetetraacetic acid |
| EIC | Extracted-Ion Chromatogram |
| EMS | Ethyl Methanesulfonate |
| epPCR | Error-Prone PCR |
| epRCA | Error-Prone Rolling Circle Amplification |
| ESI | Electrospray Ionization |
| FACS | Fluorescence-Activated Cell Sorting |

| Abbreviation | Full name |
|-------------------|--|
| FLP | Flippase |
| FPLC | Fast protein liquid chromatography |
| FRET | Fluorescence Resonance Energy Transfer |
| FRT | Flippase Recognition Target |
| Glc | Glucose |
| Glu | Glutamate |
| GM3 | Genemat 3 |
| <i>goi</i> | Gene of interest |
| HEPES | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| Ido | L-isoleucine dioxygenase |
| <i>ido</i> * | gene <i>ido</i> with IS2 upstream |
| Ile | Isoleucine |
| IPTG | Isopropyl β -D-1-thiogalactopyranoside |
| IS | Ion Source |
| IS2 | Insertion Sequence 2 |
| ISOR | Incorporating Synthetic Oligonucleotides via gene Reassembly |
| ITCHY | Incremental Truncation for the Creation of Hybrid enzymes |
| IVC | <i>In vitro</i> Compartmentalization |
| <i>kdo2</i> * | gene <i>kdo2</i> with deletion upstream |
| LA | Laboratory of Applications |
| LB | Lysogeny broth |
| LC/MS | Liquid Chromatography |
| Leu | Leucine |
| L-LDH | L-lactate dehydrogenase |
| LTEE | Long-Term Evolution Experiment |
| Lys | Lysine |
| Met | Methionine |
| MOPS | 3-(N-morpholino)propanesulfonic acid |
| MRM | Multiple Reaction Monitoring |
| mRNA | Messenger RNA |
| MS | Mass Spectrometry |
| N/A | Not Applicable |
| NAD ⁺ | Nicotinamideadenine dinucleotide |
| NADH | Reduced NAD ⁺ |
| NADP ⁺ | Nicotinamide adenine dinucleotide phosphate |
| NADPH | Reduced NADP ⁺ |
| NExT | Nucleotide Exchange and excision Technology |
| Ni-NTA | Nickel-Nitrilotriacetic acid |
| Norl | Norleucine |
| Norv | Norvaline |
| NRR | Non-homologous Random Recombination |
| NTP | Nucleoside triphosphate |

| Abbreviation | Full name |
|---------------------|---|
| OD600 | Optical Density measured at a wavelength of 600 nm |
| OH-norv | Hydroxy-norvaline |
| Orn | Ornithine |
| PACE | Phage-Assisted Continuous Evolution |
| PCR | Polymerase chain reaction |
| PEP | Phosphoenolpyruvate |
| pH | Potential of hydrogen |
| P _i | Inorganic phosphate |
| PK | Pyruvate Kinase |
| qPCR | Quantitative PCR |
| SNP | Single Nucleotide Polymorphism |
| Suc | Succinate |
| REAP | Reconstructed Evolutionary Adaptive Path |
| RACHITT | Random Chimeragenesis on Transient Templates |
| RNA | Ribonucleic acid |
| RT | Reverse Transcription |
| RT-qPCR | Quantitative reverse transcription PCR |
| SCS | Succinyl-CoA synthetase |
| SDS | Sodium Dodecyl Sulfate |
| SDS-PAGE | SDS Polyacrylamide gel Electrophoresis |
| SHIPREC | Sequence Homology-Independent Protein Recombination |
| SISDC | Sequence-Independent Site-Directed Chimeragenesis |
| StEP | Staggered Extension Process |
| TB | Terrific Broth |
| TBE | Tris/Borate/EDTA Buffer |
| TCA | Tricarboxylic acid |
| TE | Tris-EDTA |
| TFA | Trifluoroacetic acid |
| UV | UltraViolet radiation |
| Val | Valine |
| WT | Wild-Type |

I INTRODUCTION

One of the important roles of proteins is to act as enzymes. They have the ability to accelerate biochemical reactions by well over a million fold (Cooper, 2000) drawing much attention for their use in industrial processes. One of the biggest challenges in biotechnology is the optimization of enzymes for such purposes.

The following section reviews the catalytic properties of enzymes and the different methods for their improvement.

I.1 General introduction

Enzymes are powerful catalysts that make life possible by strongly accelerating biochemical reactions that would take years to occur. In any organism, a multitude of chemical reactions continually takes place and practically all of them are catalyzed by enzymes. Frequently, multiple enzymes can ensure a given reaction. Due to this versatility, the list of enzymes used in diverse technical and industrial applications is constantly growing. These applications can be beneficial to society, especially from an environmental point of view. Enzymes have the potential to solve a vast array of technical challenges ranging from the transformation of chemical energy into work (Spudich, 2001) to the ultrasensitive detection of small molecules (Firestein, 2001) and light (Rosenbaum *et al.*, 2009).

Numerous enzymes are used in industrial processes. In 2015, the global enzyme market was valued at approximately \$8.18 billion (Grand View Research, 2016). The market traditionally divides into segments. Technical enzymes, the largest segment with a sales share of 65% of the market, include enzymes used in the detergent, starch, textile, leather, pulp and paper, organic synthesis and biofuel industries. Food enzymes, the second largest segment of the market with a share of 25%, includes enzymes employed in the dairy, brewing, wine and juice, fats and oils, and baking industries. Feed enzymes, comprising enzymes used in animal feeds, contribute to approximately 10% of the market (Grand View Research, 2016).

Over 500 industrial products are currently produced using enzymes and the demand for new catalysts is continuously growing due to their potential to provide sustainable technical solutions (Johannes and Zhao, 2006). Growing population, urbanization and expanding fuel ethanol market are the chief influences driving the upward market demand for industrial enzymes (Kumar *et al.*, 2014).

The catalytic properties of enzymes, besides their power to drastically reduce the activation energy of the reaction, also include regio- and stereoselectivity, enabling the production of chiral compounds which are often difficult to synthesize through classical chemical reactions. Furthermore, they are active under relatively mild conditions of temperature, pressure, pH and they are biodegradable, offering applications within the framework of the green chemistry paradigm (Cherry and Fidantsef, 2003).

The search for novel enzymes in the pool of biodiversity is an excellent starting point for identifying potential candidates for industrial applications (Lee *et al.*, 2010). However, in most cases, natural enzymes are not suitable outside their biological context since they have evolved for fulfilling their role in metabolic pathways to guarantee cell survival. Therefore, they must be tailored to perform in artificial environments for commercial applications (Kuchner and Arnold, 1997). Furthermore, enzymes must compete economically with often entrenched and extremely inexpensive traditional chemical processes although the advantages of replacing chemical treatments with enzymes are often compelling from a societal and environmental viewpoint (Cherry and Fidantsef, 2003). Despite the technological advances of the past 100 years, human-made machines cannot compete with the precision of proteins at the nanoscale (Huang *et al.*, 2016).

Naturally occurring enzymes only represent a tiny fraction of the vast protein sequence space (Huang *et al.*, 2016). Enzymes could be the solution for a number of technical problems impacting society. For this, we need to expand the range of available catalysts. So far, almost all protein engineering has involved the modification of naturally occurring proteins to tune or alter their function (Arnold, 2015; Dougherty and Arnold, 2009; Goldsmith and Tawfik, 2012).

In the search for ever better or even new enzymes, refined for industrial purposes, several methods were developed, summarized in the following section:

- *De novo* design
- Rational redesign
- Directed evolution
- Continuous culture

I.2 Tailoring enzymes

I.2.1 *De novo* design

De novo protein design aims at using the accumulated information about natural proteins to identify patterns and rules underlying the folding of polypeptides into active enzymes and to apply these rules to the design of synthetic proteins. The output of such computational modeling is an amino acid sequence which folds into an optimized protein capable of catalyzing the desired reaction.

The protocol for *de novo* enzyme design (Figure 1) has been first suggested by Sternер and colleagues (Sternér *et al.*, 2008) and more recently optimized by Barrozo and colleagues (Barrozo *et al.*, 2012).

Successful computational enzyme designs have paved the way for encouraging progress in this field such as Kemp-eliminase (Röthlisberger *et al.*, 2008), retro-aldolase (Jiang *et al.*, 2008), Dieals-Alderase (Siegel *et al.*, 2010). However, only a small fraction of computationally designed proteins have detectable activity. Much work needs to be done to understand why current computationally designed enzymes are not better catalysts than natural ones and to pinpoint the steps hampering the design process.

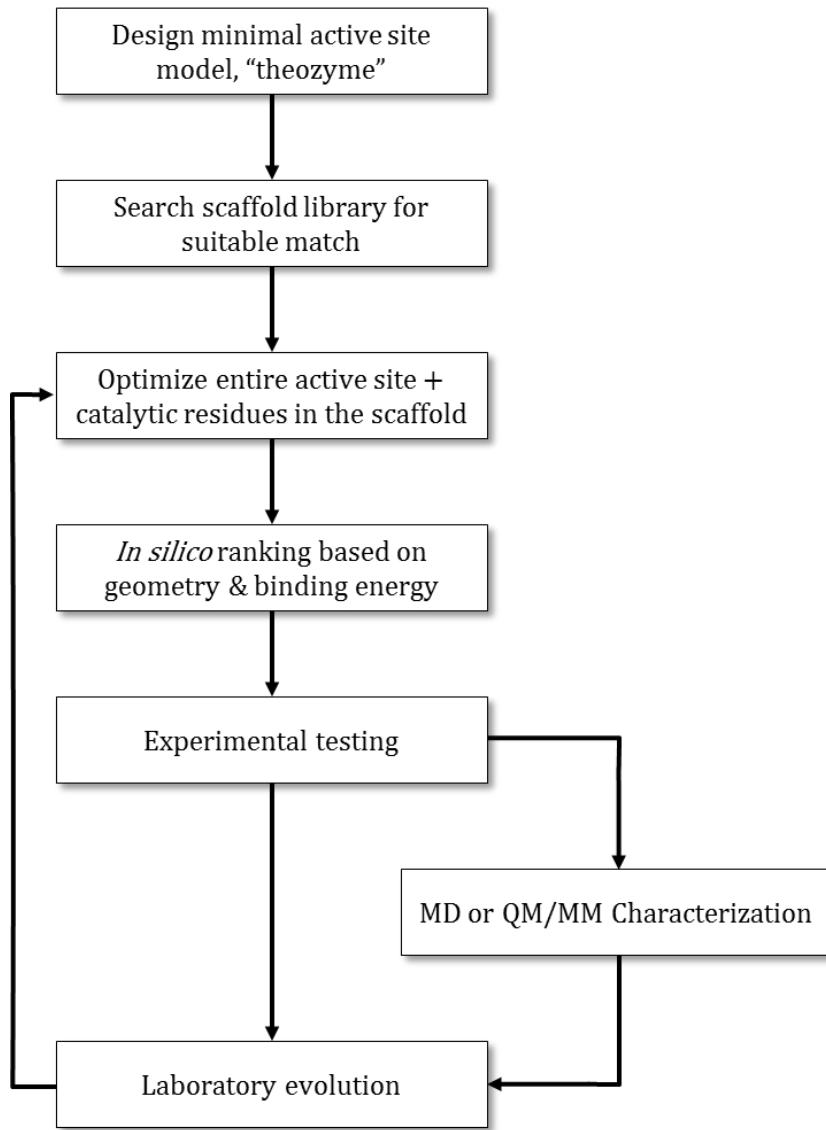


Figure 1: Process for *de novo* enzyme design.

Adapted from Barrozo *et al.*, 2012

I.2.2 Rational design

Since nature offers such a wide choice of versatile and efficient functional proteins, the rational design approach thus starts from existing templates that can be redesigned to fit required properties using computational design algorithms.

Rational design was the earliest approach in protein engineering dating back to the mid-1970s and was used to modify enzyme specificity (Chica *et al.*, 2005). Initially utilized for the study of enzymatic catalysis and binding mechanisms (Gutte, 1975), rational design then became an exploratory tool for improving catalytic activities (Carter *et al.*, 1989).

The variety of computational methods span from quantum mechanical cluster models, QM/MM calculation, MD simulations, to semi-empirical approaches like EVB based on valence bond theory (Kamerlin and Warshel, 2011).

The majority of rational design strategies are employing primitive electrostatic models based on protein crystal structures. Theoretical enzymes are constructed by computing the optimal geometry for transition-state stabilization by functional groups. Since backbone perturbations are often neglected, computational simulations may result in many false-positive protein structures.

Nowadays, reshaping an enzyme specificity is achievable based on an in-depth knowledge of the protein's structure and mechanism (Cedrone *et al.*, 2000). However, not all protein structures and mechanisms have been characterized, limiting rational design to already well-studied enzymes. Despite major scientific advances, the understanding at the molecular level of why one protein performs a certain reaction better than another remains elusive. Small changes in structure or chemical properties can have big, unexpected effects on catalysis (Romero and Arnold, 2009). Indeed, the successful design of one parameter (*e.g.* enhanced stability) is often accompanied by losses elsewhere (*e.g.* catalytic activity) (Arnold, 1998).

I.2.3 Directed evolution

Evolution generated proteins through natural selection despite their complexity. The remarkable sequence variability of proteins allows them to adapt under selection pressure by changing their activity parameters and also their substrate range. Protein engineers exploit this evolvability using directed evolution to generate new proteins (Romero and Arnold, 2009).

Enzyme evolution relies mainly on their promiscuity. Experiments lend credence to the proposal that promiscuity underlies much of enzyme innovation (Copley, 2003). Numerous studies have shown that an enzyme optimized for one reaction can adapt under directed evolution to take on the function of a different member of the superfamily, sometimes with efficiency that rivals nature's solutions (Aharoni *et al.*, 2005; Gerlt and Babbitt, 1998).

Directed evolution consists in mimicking the process of natural selection to obtain proteins with specific/desired traits. Unlike rational design, such an approach does not require prior knowledge of structure or mechanism of a given enzyme, and a vast panoply of methodologies is available for carrying out evolution experiments, both *in vitro* and *in vivo*.

Four fundamental processes need to be maintained to support Darwinian evolution: mutation, translation, selection and replication (Badran and Liu, 2015, Figure 2).

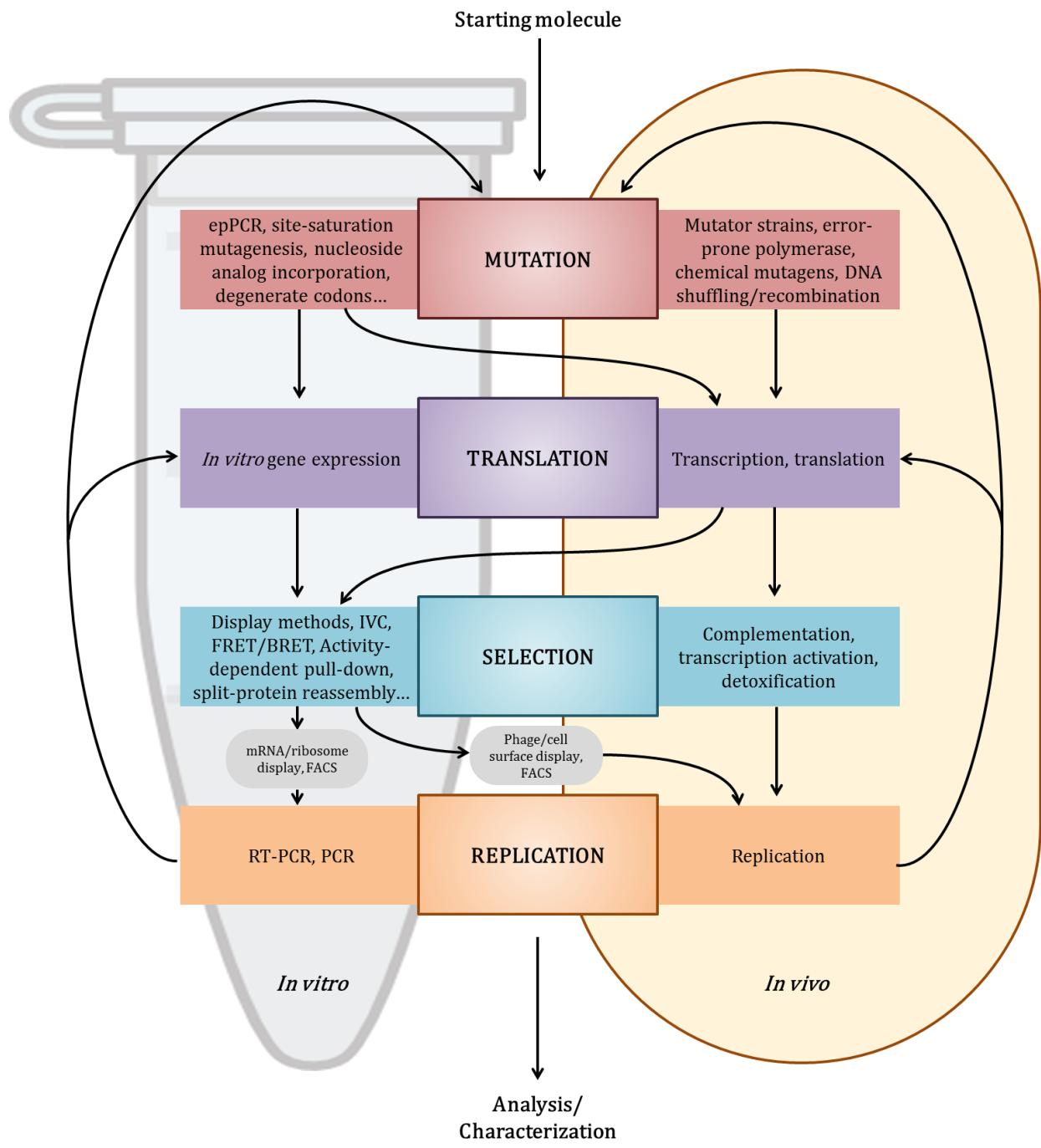


Figure 2: Schematic overview of directed evolution.

Arrows indicate possible routes for connecting individual evolutionary steps.

Abbreviations: DNA, Deoxyribonucleic acid; PCR, Polymerase chain reaction; RT-PCR, Reverse transcription PCR; epPCR, Error-prone PCR; IVC, *In vitro* compartmentalization; FACS Fluorescence-activated cell sorting; FRET/BRET, Fluorescence/Bioluminescence resonance energy transfer; mRNA, messenger Ribonucleic acid.

Adapted from Jäckel *et al.*, 2008

I.2.3.1 Genetic diversity

Creating molecular diversity is the initial step in any directed evolution experiment. The effectiveness of directed protein evolution depends on the quality and the size of the genetic library. The genetic diversity should be large enough to increase the chances of finding a desirable property. At the same time, the growth in library size makes the selection more challenging (Reetz *et al.*, 2008; Shivange *et al.*, 2009). In the literature there is a large spectrum of library generation methods which can be grouped into three main categories:

- Random mutagenesis
- Focused mutagenesis
- Recombination

Random mutagenesis is commonly used when protein structure-function knowledge is lacking (Labrou, 2010). Even when the enzyme structure is available, random generation of diversity is applied to reveal favorable mutations in unexpected positions outside of the active site (Morley and Kazlauskas, 2005). Such mutations influence protein thermostability and tolerance of reaction conditions by inducing conformational changes in the enzyme backbone. In general, random mutagenesis introduces nucleotide variations in the DNA sequence, divided into five categories:

- Transitions, which consists of a substitution of a purine nucleotide with another purine or a pyrimidine nucleotide with another pyrimidine
- Transversions, which consists of a substitution of a purine nucleotide with a pyrimidine or a pyrimidine nucleotide with a purine
- Deletions, where one or more nucleotides are deleted from the sequence
- Insertions, where one or several nucleotides are incorporated into the gene
- Inversions, DNA rearrangements in which a segment of two base pairs (bp) or longer is reversed end to end

Large libraries of variants can be created through different techniques. These include subjecting DNA to UV radiation (Bridges and Woodgate, 1985) or chemical mutagens (Kadonaga and Knowles, 1985; Lai *et al.*, 2004), application of bacterial mutator strains (Greener *et al.*, 1997), error-prone PCR (epPCR) (McCullum *et al.*, 2010) or by error-prone rolling circle amplification (epRCA) (Fujii *et al.*, 2004).

Each method has its pros and cons (Table 1).

Table 1: Methods for library diversification

| Category | Examples and methods | Random or focused? | In vivo or In vitro? | Advantages | Disadvantages |
|------------------------------|--|--------------------|------------------------------------|--|---|
| Chemical mutagenesis | EMS, nitrous acid, UV irradiation, and bisulfite | Random | <i>In vitro</i> and <i>in vivo</i> | Dose-dependent mutation rates | Low mutation rates; uneven mutational spectrum; hazardous chemicals |
| Mutator strains | XL1-red <i>E. coli</i> , mutagenesis plasmid (PACE) and yeast orthogonal replication | Random | <i>In vivo</i> | Easy to use | Low mutation rates; uneven mutational spectrum |
| epPCR | Tag supplemented with Mg ²⁺ .Mn ²⁺ and/or unequal dNTPs; proprietary enzyme mixes (Mutazyme) | Random | <i>In vitro</i> | Permits high mutation rates; easy to use commercial formulations; relatively even mutational spectrum | Random mutagenesis at the nucleotide level but does not evenly sample amino acid codon space |
| Homologous recombination | DNA shuffling, family shuffling, StEP, RACHITT, NExT, heritable recombination, ADO, and synthetic shuffling | NA | <i>In vitro</i> or <i>in vivo</i> | Can identify beneficial combinations of mutations or eliminate passenger mutations; can also shuffle sequences of orthologous proteins to repurpose functional diversity from nature | Rely heavily on sequence homology; evolved clones and natural orthologues can be divergent in nucleotide sequence |
| Non-homologous recombination | ITCHY, SHIPREC, NRR, SISDC and overlap extension PCR | NA | <i>In vitro</i> | Capable of shuffling distantly related sequences, permuting order and combinatorial gene fusions | Highly efficient in niche applications but challenging to implement as a general strategy |

ADO, assembly of designed oligonucleotides; *E. coli*, *Escherichia coli*; EMS, ethyl methanesulfonate; epPCR, error-prone PCR; ISOR, incorporating synthetic oligonucleotides via gene reassembly; ITCHY, incremental truncation for the creation of hybrid enzymes; NA, not applicable; NExT, nucleotide exchange and excision technology; NRR, non-homologous random recombination; PACE, phage-assisted continuous evolution; RACHITT, random chimeragenesis on transient templates; REAP, reconstructed evolutionary adaptive path; SHIPREC, sequence homology-independent protein recombination; SISDC, sequence-independent site-directed chimeragenesis; StEP, staggered extension process.

Adapted from Packer and Liu, 2015

Error-prone PCR (epPCR), which produces random mutations during PCR by reducing DNA polymerase fidelity, is perhaps the most widely used method (Packer and Liu, 2015). Many successful examples of enzyme optimization via mutagenesis by epPCR are cited in the literature such as the successful increase of the enantioselectivity of a lipase (Reetz *et al.*, 1997).

However, since every protein behaves differently, there is no generalized strategy of diversity generation and selection in directed evolution. Many protocols which function for one protein fail when applied to other proteins. In turn, natural evolution remains still unbeaten, turning many protein engineers into evolutionists (Peisajovich and Tawfik, 2007). The advancement in enzyme engineering depends not only on

technological breakthroughs but also on a thorough understanding of how these biomolecules evolved in nature. The reconstruction of mutational trajectories has shown the presence of dead ends in the mechanisms of protein divergence (Weinreich *et al.*, 2005). To explain this observation, it is necessary to examine the effects of protein mutations (Camps *et al.*, 2007). A single mutation could be deleterious, neutral or advantageous for a certain protein property. However, a pleiotropic mutation favoring the gain in one activity might be accompanied by a loss of another property. This effect is also known as mutation trade-off, which usually occurs between:

- New function and stability
- New and existing function
- New function and expression

In reality, the effects of mutations are even more complex, implying the interactions between two or several mutations. The notion of epistasis is introduced when an effect of a given mutation is modulated by other mutations (Poelwijk *et al.*, 2007; Breen *et al.*, 2012). For example, a new-function mutation might rely on another stabilizing, compensatory mutation to be fixed. Likewise, a compensatory mutation can be neutral on its own but beneficial in combination with a destabilizing mutation (Bloom *et al.*, 2006). Under laboratory conditions, the problem of cooperative mutations can be partially overcome by shuffling methods recombining point mutations on the target DNA (Coco *et al.*, 2001).

Another tool for the fine-tuning of protein engineering, which helps to bypass the problem of dead ends in protein evolutionary divergence, is neutral drift selection (Bloom *et al.*, 2007). During first evolution rounds, the selection conditions are chosen such that the accumulation of mutations without altering protein function or stability is favored. The resulting gene pool then contains polymorphic non-adaptive variations enhancing protein evolvability and avoid dead ends. In a second step, a more stringent selection pressure is applied to improve the desired property (Agresti *et al.*, 2010; Aharoni *et al.*, 2005; Bershtein *et al.*, 2008; Gupta and Tawfik, 2008).

I.2.3.2 Selection and screening

The methods for screening and selection are far less developed than the strategies enabling genetic diversification, making selection the limiting step in directed evolution of proteins. While the approaches for diversification are quite general and could be applied to any gene, the selection or screening must be adapted to the enzyme and the desired property that needs to be improved. A suitable selection method for enzyme evolution must fit several criteria:

- Specific to the desired property
- Sensitive to low enzyme activity
- Robust, reproducible, simple in set-up
- Enable a high-throughput approach
- Allow the modulation of selection pressure

The selection scheme must be highly specific to the evolvable property and capable of sensing even small improvements. A high-throughput approach enables the analysis of large mutant libraries, increasing the chances of finding the best variants. Before choosing a selection method, it is necessary to define the objectives and the size of libraries clearly. Through iterative cycles of mutagenesis and amplification of selected library members, beneficial mutations accumulate as in natural Darwinian evolution.

In general, directed evolution methods are sequential and handle each of the four steps discretely (mutation, translation, selection, and replication). Some of these steps can be very time- and labor-intensive.

As mentioned previously, even a protein library containing millions of variants would still only represent a tiny fraction of the sequence space of an average protein. The number of possible permutations for a protein consisting of N residues would be 20^N . For a rather small protein of 100 residues, the number 20^{100} ($= 1.3 \times 10^{130}$) of unique combinations of amino acids already exceeds by far the achievable library size of all known protein library creation methods (Romero and Arnold, 2009), making it infeasible to cover the entire mutational space of a protein.

Biases, with respect to certain types of mutations inherent to the experimental method and the degeneracy of the genetic code, further skew and restrict the sequence

variability of the library (Wong *et al.*, 2006). Rather than addressing these problems through bigger libraries and larger screening scales, some scientists turn to *in vivo* methods of directed evolution.

I.2.4 Directed evolution *in vivo*

Directed evolution *in vivo* relies on spontaneous mutations occurring in a growing cell population. The population, which grows under defined conditions, is constantly diluted. This dilution puts a selective pressure on the cells, leading to an overgrow of the population by cells which carry a mutation enabling them to grow faster under the given conditions. The dilution protocol and culture conditions distinguish the different approaches of directed evolution *in vivo*.

I.2.4.1 Directed evolution in batch culture

The most basic way to conduct *in vivo* directed evolution experiments is the daily dilution of a culture growing in the batch mode. In general, the cells serially go through the distinct phases of growth (lag, exponential, transition, and stationary). A particularly interesting set of evolution experiments in batch culture was performed by Lenski and coworkers which encompassed the fitness analysis of 12 independent *Escherichia coli* (*E. coli*) populations inoculated from the same ancestor. Daily serial transfer propagated these populations for 66 000 generations (from February 1988 to November 2016) in the simple environment of a glucose-supplemented minimal medium in shaking flasks. This experiment, called the long-term evolution experiment (LTEE), has become a model system for studying many fundamental evolutionary questions (Barrick *et al.*, 2009).

Although LTEE have proven to be a very efficient way to mimic evolution in a laboratory (Gerrish and Lenski, 1998; Elena and Lenski, 2003; Barrick *et al.*, 2009; Tenaillon *et al.*, 2016), it also requires the presence of an experimenter for each dilution and absolute sterility to avoid contamination. Furthermore, during batch culture, cells are subjected to alternating periods of growth and stasis upon serial transfer. After a period of exponential growth, a stationary phase is often reached due to exhaustion of one of the nutrients. Alternatively, cells may reach their upper pH limits or become CO₂ limited long before nutrients become the limiting factor (Hinga, 2002). Thus evolutionary

events may arise from selection pressure other than the ones chosen by the experimenter.

I.2.4.2 Directed evolution in continuous culture

Evolution in continuous culture, initially demonstrated as a surrogate for Darwinian evolution by Spiegelman and coworkers (D. R. Mills *et al.*, 1967), is based on growing populations subjected to automated dilution regimes. According to the experimental objectives pursued, two regimes, chemostat and turbidostat, are used.

I.2.4.2.1 Chemostat

The chemostat was independently described by Monod and Novick & Szilard (Monod, 1950; Novick and Szilard, 1950). In a chemostat regime, cells are grown in a vessel with a fixed volume of medium. The medium is continuously diluted by a regular inflow of fresh medium while a corresponding outflow of old medium keeps the whole culture volume constant in the vessel (Figure 3). The growth medium provides all nutrients necessary for growth in excess except one(s), which is limiting.

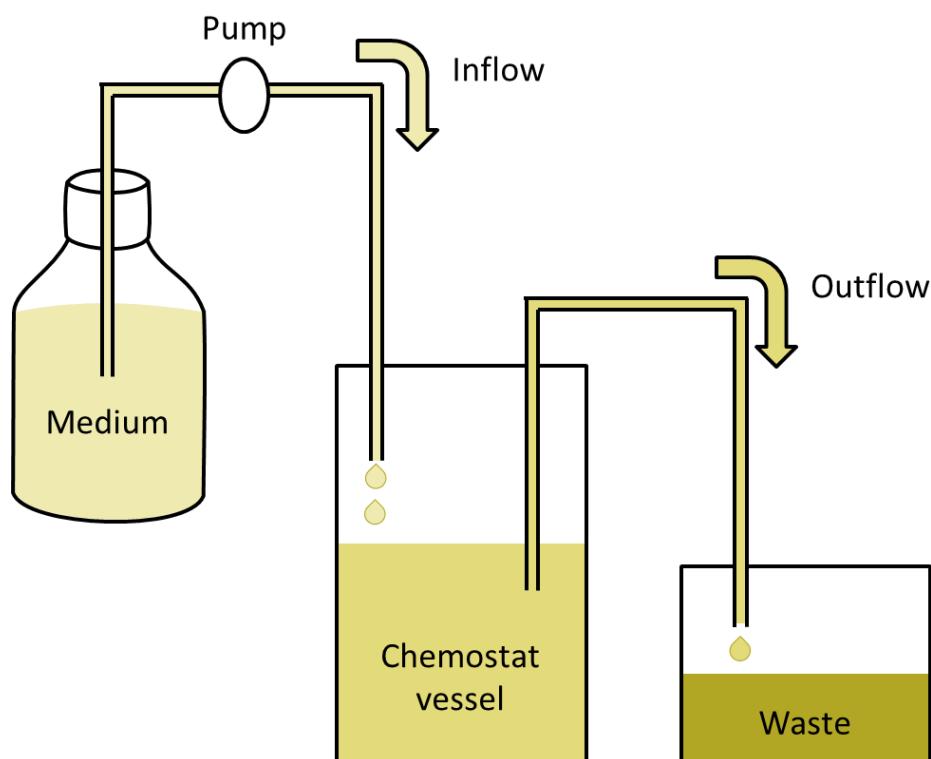


Figure 3: Diagram of a chemostat

Adapted from Ziv et al., 2013

Monod proposed (Monod, 1950) that cell growth occurred analogous to a Michaelis-Menten type relationship depending on external nutrient concentrations (Equation (1)). The rate of change in cell density (x) over time depends on the maximal growth rate of the cells (μ_{max}) and a half-maximal growth rate constant (K_s)

The rate of change in the limiting nutrient concentration (Equation (2)) depends on its concentration in the inflowing media (R), its dilution by outflow (D) and cell consumption of nutrient.

$$(1) \quad \frac{dx}{dt} = \mu_{max} \frac{s}{K_s + s} x - Dx$$

$$(2) \quad \frac{ds}{dt} = DR - Ds - \frac{x}{Y} \mu_{max} \frac{s}{K_s + s}$$

$x:$ Cell density (cells. ml⁻¹)
 $\mu_{max}:$ Maximal growth rate (h⁻¹)
 $s:$ Limiting nutrient concentration (μM)
 $K_s:$ Half max growth rate constante $\mu_{max}/2$ (μM)
 $D:$ Dilution rate (h⁻¹)
 $R:$ Nutrient concentration in media reservoir (μM)
 $Y:$ Yield constant (cells/ml/μM)

The chemostat has reached steady-state when measured variables, such as cell density, remain constant over time. Unlike for the batch culture, a constant nutrient flow and a stable cell density keep the growth environment of the chemostat stable. Mutant cells are selected for a more efficient utilization of the limiting nutrient since they are able to make more biomass and hence will accumulate in the population.

I.2.4.2.2 Turbidostat

In the turbidostat regime, unlike the chemostat, no nutrient is limiting in the growth medium. Dilutions are not occurring at fixed intervals but are triggered by the growth of the culture. The optical density of the culture is constantly monitored and compared with a threshold fixed by the operator. As soon as the cell density exceeds the threshold, the culture receives a dilution. The appearance in the population of a faster-growing mutant will trigger a higher dilution rate which in turn counter selects the slower-growing, non-mutated cells. In this case, the growth rate equals the dilution rate of the culture, the volume being kept constant.

I.2.4.3 An automated continuous culture device: GM3

In continuous culture, a population of bacteria in suspension is constantly diluted with fresh growth medium. The main difficulty in maintaining a culture for a long period of time is the appearance of dilution-resistant variants. These variants escape the selection, normally acting on the organisms in suspension, by forming biofilms (de Crécy-Lagard *et al.*, 2001). The effective elimination of biofilms has proven to be critical for reprogramming and improving the metabolism of microbial populations (Marlière *et al.*, 2011).

To circumvent this difficulty, a continuous culture device, the GM3, was developed (Mutzel and Marliere, 2004, patent No.: US6,686,194 B1). The GM3 automates enable the maintenance of steady-state microbial growth over an extended period of times while preventing biofilm formation (Figure 4).

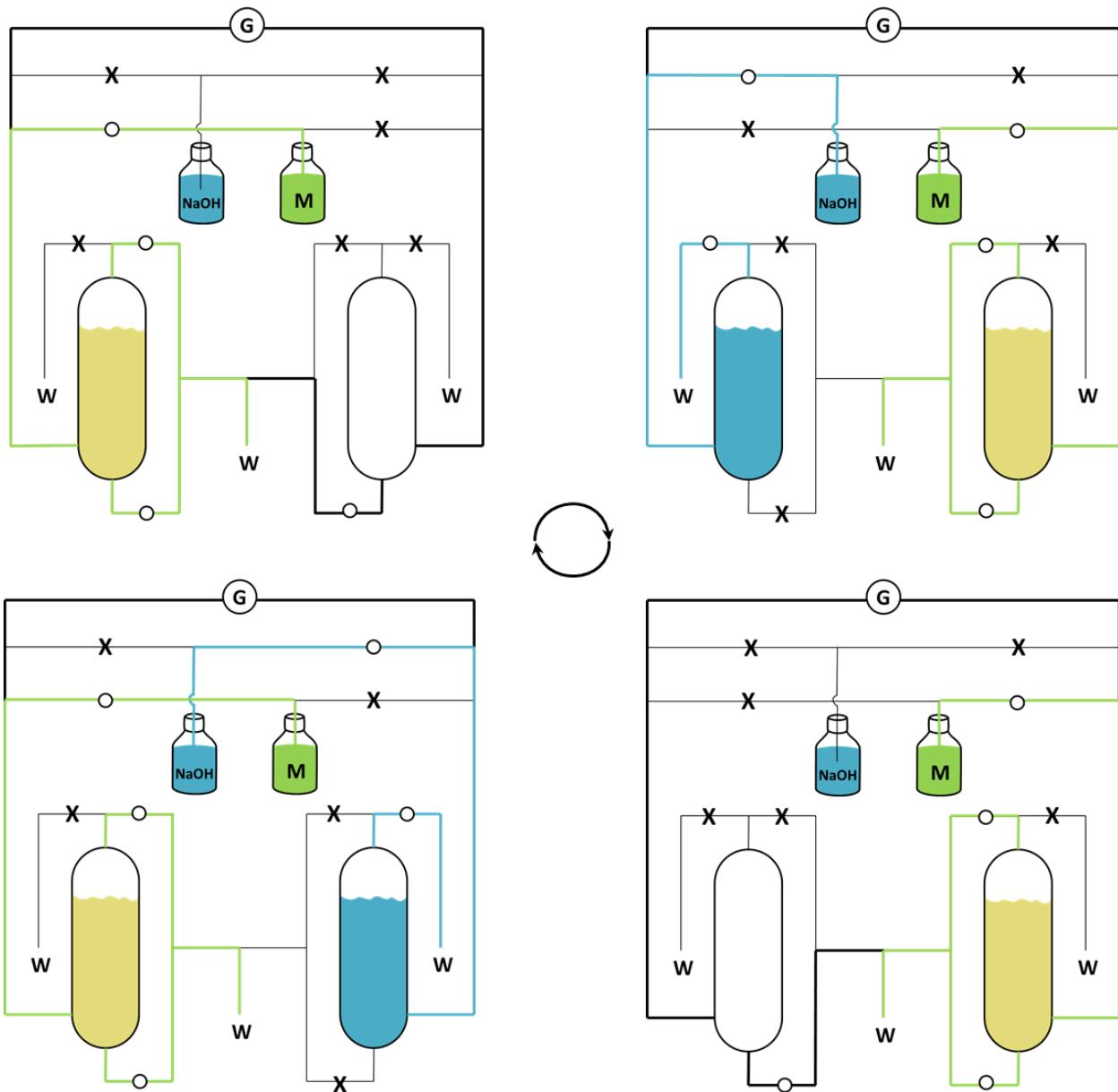


Figure 4: Cultivation fluidics with twin growth vessels in the GM3 cultivation device

Pressurized gas supply (G); medium source (M in green); sterilizing agent (NaOH 5 M in blue); waste (W); Bacterial cultures is shown in beige. Valves are indicated by X and O for the closed and open states, respectively. Flows in the tubing are indicated by bold lines.

Adapted from Marlière *et al.*, 2011

To prevent biofilm formation, two growth vessels are alternatively used and sterilized. Figure 4 is a simplified scheme of key steps in the GM3 cultivation fluidics starting from the top left image and going clockwise. The vessel on the left contains the growing culture. After a defined period, the whole culture is transferred into the second vessel. The first vessel is sterilized with 5 M NaOH washing out any biofilm formed. The culture grows in the second vessel while the sterilizing agent in the first vessel is discarded (W)

and rinsed. The culture is then transferred back into the first vessel. The second vessel is sterilized with 5 M NaOH, rinsed, and the cycle then repeats. Typically, two to four such cleaning cycles are programmed during 24 h, depending on the organism evolving.

The *in vivo* selection of gene variants is based on the spontaneous appearance of point mutations in cells growing in suspension. Thus the setup of the GM3 device needs to be adequate to statistically encompass at least one mutation per base pair in the total chromosome. In *E. coli*, the mutation rate has been determined to be 2.2×10^{-10} per nucleotide per generation (Lee *et al.*, 2012) meaning that statistically each base pair of the chromosome is mutated in a suspension of about 4.5×10^9 cells. Given that a suspension of *E. coli* at OD₆₀₀ = 0.4 (density generally set in the GM3 at steady state) is about $2.8 \times 10^8 \text{ } \mu\text{ml}^{-1}$ (Sezonov *et al.*, 2007), a culture of 16 ml of such a cell suspension should contain at least one mutant for every base pair. Evidently, these statistical values do not take into account the influence of chromosomal topology on mutability (Williams, 2014). The DNA replication process is the principal entry point of genetic mutations, but chemical DNA modifications (nucleobase methylations, etc.) and damage (UV light, etc.) also contribute. A variety of enzymatic machinery exists in the cell capable of correcting all types of mutations. Cells having lost one or more of these error-correcting mechanisms show up to three orders of magnitude higher mutation rates (Marinus, 2010; Yang, 2000).

Different culture regimes can be applied to modulate selective pressure:

- Chemostat regime
- Turbidostat regime
- Medium SWAP regime

I.2.4.3.1 Chemostat and turbidostat regime

As described previously, the chemostat is a regime where the dilution rate is kept constant throughout the evolution experiment, while the turbidostat is characterized by a variable dilution rate triggered by the growth of the culture.

In order to be operational, both regimes require a minimal growth rate under the selection conditions of the culture. Only in growing populations generate the mutations which are the “material” of evolution. Also, from a practical point of view, cultures in the GM3 which are not diluted loose volume through evaporation.

I.2.4.3.2 Medium SWAP regime

The medium SWAP regime has been conceived to enable difficult selections overcoming the limits of the turbidostat and chemostat regimes. The medium SWAP regime consists in diluting the culture alternatively between two different media (Figure 5):

- A relaxing medium composed to sustain stable growth of non-adapted cells
- A stressing medium composed according to the selection of the culture which does not support growth of the non-adapted cells

As for the chemostat regime, pulses of media are sent to the vessel at regular time intervals set by the experimenter. While the frequency of pulses does not depend on cell density, the composition of the medium at a given time point of the experiment does. Depending on the turbidity measured, a pulse of either the relaxing medium (if the density is lower than the fixed threshold) or stressing medium (if the density is higher than the fixed threshold) is sent to the vessel. The culture is thus diluted at a fixed dilution rate. The volume of the medium pulse injected into the culture vessel is set by the experimenter and will impose the generation time. Upon adaptation of the population, the proportion of stressing pulses will rise until reaching 100%. Figure 5 compares the dilution events occurring in a culture growing under a turbidostat and a medium SWAP regime.

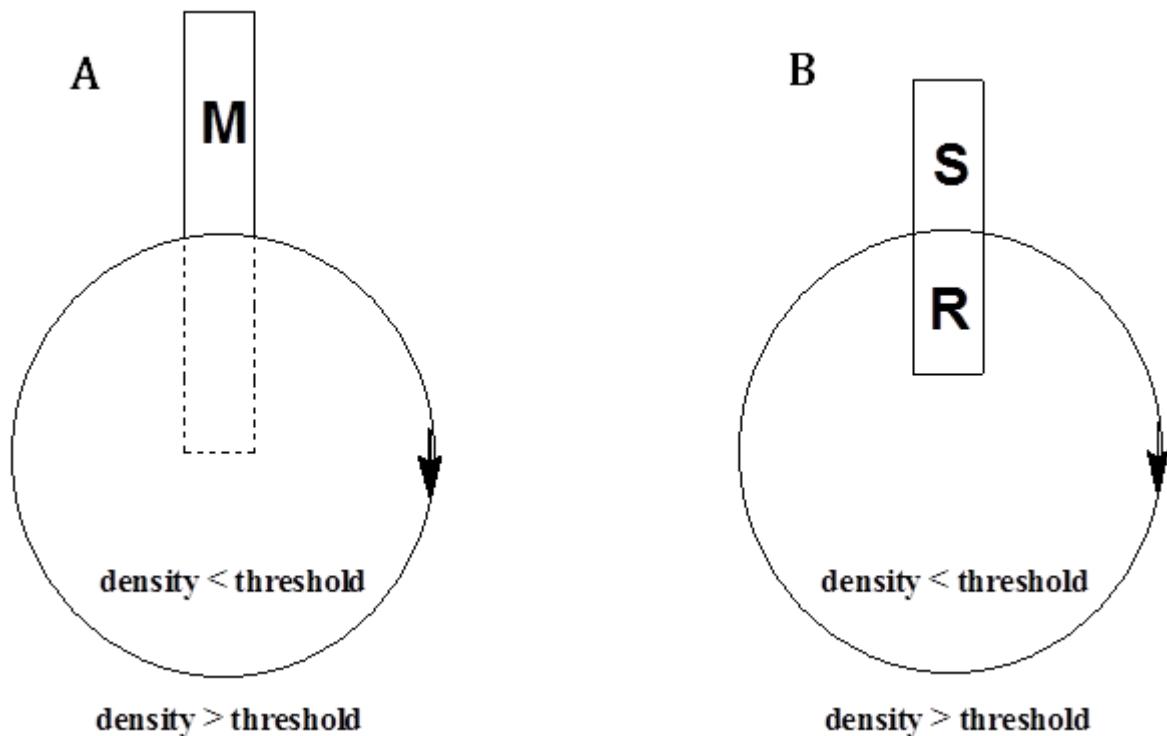


Figure 5: Pulse-feeding clock of turbidostat (A) and medium SWAP regime (B) of the GM3 automates

(A) The circles represent a growth period between two OD measurements, generally 10 minutes. Only one medium (M) dilutes the culture in a turbidostat. Dilution pulses are sent to the vessel only if the cell density is above the fixed threshold. (B) In medium SWAP, relaxing medium (R) or stressing medium (S) dilution pulses are sent to the vessel depending on whether the cell density is lower (medium R) or higher (medium S) than the threshold.

I.3 Directed evolution in the GM3 device

The GM3 device offers the possibility to maintain steady-state microbial growth over extremely long time periods. This enables directed evolution setups to target complex phenotypic traits, as well as pathways or even single enzymes.

I.3.1 Examples of applications

I.3.1.1 Bacterial continuous evolution

In an application of the GM3 device (Marlière *et al.*, 2011), *E. coli* was evolved to substitute the canonical DNA base thymine with an artificial base, the chemical homolog 5-chlorouracil. The initial strain, an *E. coli* thymine auxotroph, was grown in the GM3 device under a medium SWAP regime alternating between:

- A relaxing medium containing 3 µM of thymine
- A stressing medium containing 3 µM of 5-chlorouracil

Every 10 min, a pulse of either relaxing or stressing medium was sent into the vessel depending on the culture turbidity relative to a threshold of $OD_{880} = 1$. After six months of adaptation in the GM3 (around 1000 generations), the culture was able to grow in the stressing medium. The chemical analysis of the chromosomal DNA of isolates showed a 90% substitution of thymine by 5-chlorouracil.

I.3.1.2 Evolution of a specific enzyme

Another successful example of directed evolution with the GM3 device is the evolution of a non-natural chorismate mutase enzyme in *E. coli* (Müller *et al.*, 2013). Chorismate mutase converts chorismate to prephenate, an essential step in the biosynthesis of phenylalanine and tyrosine. The selection strain was an *E. coli* lacking the wild-type (WT) chorismate mutase and expressing a weakly active artificial variant that has been previously designed using a nine amino acid code. The culture was grown in the GM3 device under turbidostat regime applying conditions of increasing stringency:

- First of all, a growth medium supplemented with phenylalanine
- Then, a growth medium lacking both tyrosine and phenylalanine

The culture was propagated for more than 200 days. Chorismate mutase variants with an expanded amino acid set (ten or eleven amino acids) could be isolated. The variants were shown to have improved activity compared to the parent enzyme.

These results show that directed evolution *in vivo* can be an efficient way to explore the vast mutational space of proteins or even entire genomes and to discover optimized and new functions.

I.3.2 Directed evolution projects

The previous examples only show a glimpse of the potential applications for directed evolution in the GM3. The main part of the present work targets, the evolution of specific proteins. In a separated project, a phenotype of a bacterial strain was selected, for which no prior knowledge about the possible genetic targets involved was available. Both projects will be introduced in the following sections.

I.3.2.1 Directed evolution of *Clostridium phytofermentans*

Various evolution projects can be envisioned in the GM3 device including evolving organisms towards the resistance to toxic compounds, the degradation of specific substrates, the production of compounds of interest and so forth.

The directed evolution of an organism in the continuous culture in the GM3 was addressed in collaboration with the Tolonen team (University of Paris-Saclay, Genoscope). The Tolonen team focuses on studying and developing bioengineering tools for a group of plant-fermenting bacteria, the clostridia. Clostridia contribute significantly to the recycling of plant biomass in soil, making them top candidates to transform biomass into fuels and commodities (Lynd *et al.*, 2002).

Their primary model, *Clostridium phytofermentans*, is able to ferment plant polysaccharides to ethanol and hydrogen (Warnick *et al.*, 2002). The Tolonen group has shed light on key plant-fermentation mechanisms (Tolonen *et al.*, 2011; Boutard *et al.*, 2014; Tolonen and Haas, 2014; Tolonen *et al.*, 2015; Petit *et al.*, 2015). They also aim at developing strains with novel properties such as resistance to plant-deriving inhibitors.

Plants are primarily composed of lignocellulose, an intricate matrix of polysaccharides and polyphenolic lignin. The plant biomass fermentation by *Clostridium*

phytofermentans releases phenolic and acidic growth inhibitors. A collaboration project between the Laboratory of Applications (LA) and the Tolonen group was undertaken for the directed evolution of *Clostridium phytofermentans* towards increased concentration of plant inhibitors.

I.3.2.2 Directed evolution of Iron (II)/ α -keto acid-dependent oxygenases (α -KAOs)

Iron(II)/ α -ketoacid-dependent oxygenases (α -KAOs) are of great interest because of their capacity to oxidize unactivated C-H bonds. A more detailed presentation of these enzymes will follow in section I.4.

The GM3 device can be used to improve their efficiency and evolve them towards new substrates. In the frame of this study, we address the issue of improving α -KAOs towards natural and new substrates *in vivo*.

I.4 Iron (II)/ α -keto acid-dependent oxygenases (α -KAOs)

I.4.1 General overview

The development of catalytic systems for the stereospecific oxidation of C-H bonds is still a major goal in chemistry. This versatile functionalization of organic molecules is a challenge for chemists but is frequently found in biosynthetic pathways.

Enzymes, such as cytochrome P450 monooxygenases (CYPs), are able to catalyze the insertion of oxygen into unactivated C-H bonds under mild reaction conditions. The remarkable reactivity of CYPs and their substrate promiscuity have sparked interest among researchers (Fasan, 2012) and found applications for the production of a broad range of industrially important substrates. However, the use of these enzymes in industrial processes is hampered since they need the presence of expensive cofactors including NADPH and a redox partner such as cytochrome P450 reductase or a ferredoxin/ferredoxin reductase system (Chefson *et al.*, 2006).

Over the past decade, an increasing number of publications have revealed the diversity of another oxygenase superfamily—the iron(II)/ α -ketoacid-dependent oxygenases (α -KAOs) (Baud *et al.*, 2014). Concomitant to the oxidation of a substrate specific to each dioxygenase by molecular oxygen, most frequently hydroxylation, these enzymes mediate the oxidative decarboxylation of α -ketoglutarate (α -KG) to succinate. Unlike

CYPs, α -KAOs catalyze these reactions without additional cofactors or regenerating enzymatic system.

α -KAOs are involved in various biological processes, for example, DNA and RNA repair (Shen *et al.*, 2014), oxygen sensing (Taabazuing *et al.*, 2014), fatty acid metabolism (Wanders *et al.*, 2011), post-translational modification of collagen (Gorres and Raines, 2010), demethylations related to epigenetic regulation (Johansson *et al.*, 2014), biosynthesis of antibiotics (Hamed *et al.*, 2013) and secondary metabolites (Cheng *et al.*, 2014).

Several bacterial α -KAOs have been reported to hydroxylate the side chain of free amino acids or tethered peptides in non-ribosomal peptide biosynthesis (Hausinger, 2015; Martinez and Hausinger, 2015). Hydroxylated amino acids contain several stereogenic centers and are thus valuable chiral building blocks for fine chemical synthesis. However, except for the production of hydroxyprolines by L-proline hydroxylases, which found applications in pharmaceutical and feed industries, α -KAOs are rarely used in industrial biocatalytic processes (Baud *et al.*, 2014).

I.4.2 Catalyzed reactions

The α -KAOs catalyze the hydroxylation reaction of non-activated Csp3-H. Unlike CYPs, the iron is non haeminic. The reaction catalyzed implicates three different substrates, *i.e.* molecular oxygen, a α -ketodiacid (α -KG in most cases) and a substrate specific for the enzyme. Both oxygen atoms of the O_2 molecules are transferred, one to α -KG which is oxidatively decarboxylated to form succinate upon liberation of carbon dioxide, the other to the specific substrate to form the hydroxyl group (Figure 6).

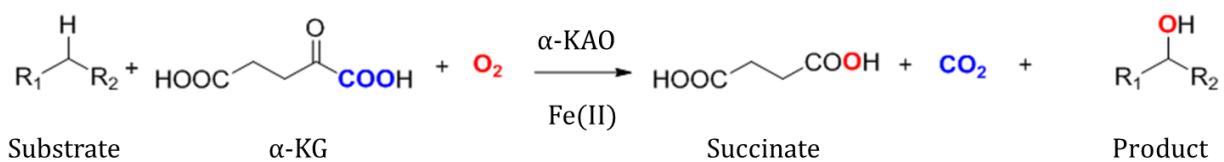


Figure 6: Hydroxylation reaction catalyzed by α -KAO

As shown in Figure 7, α -KAOs do not only catalyze hydroxylation reaction of Csp3-H bonds. They are the largest known subgroup of mononuclear non-haem iron enzymes also catalyzing dealkylation, desaturation, epoxidation, epimerization, halogenation,

cyclization, peroxide formation, and ring expansion/contraction reactions by C-H bond activation (Wu *et al.*, 2016). In these cases, the second oxygen atom is usually not transferred to the second substrate but can be found in a water molecule.

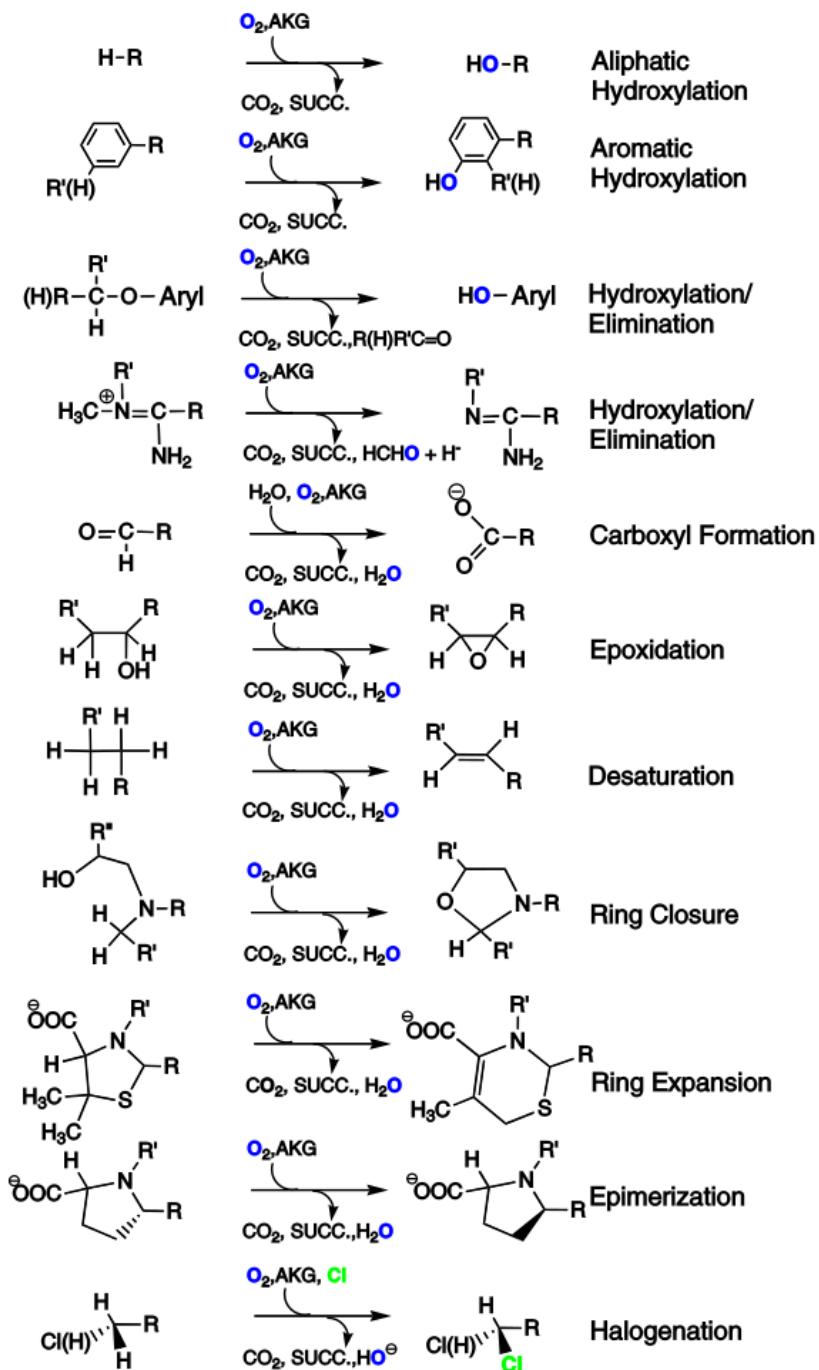


Figure 7: The diversity of the reactions catalyzed by α -KAOs

Reprinted from Purpero and Moran, 2007

I.4.3 Catalytic mechanism

The diverse substrate specificities of the members of the α -KAO enzyme family are reflected by a likewise high diversity in amino acid sequences. However, some structural features are common to these proteins: a double-stranded β -helix fold and a His-Xxx-(Asp/Glu)-(Xxx)n-His consensus sequence which coordinates the mononuclear ferrous iron in the active center (Clifton *et al.*, 2006; Elkins *et al.*, 2002). A general mechanism has been elaborated (Figure 8, Wu *et al.*, 2016), implicating the formation of a Fe^{2+} α -KG bidentate binding complex (II).

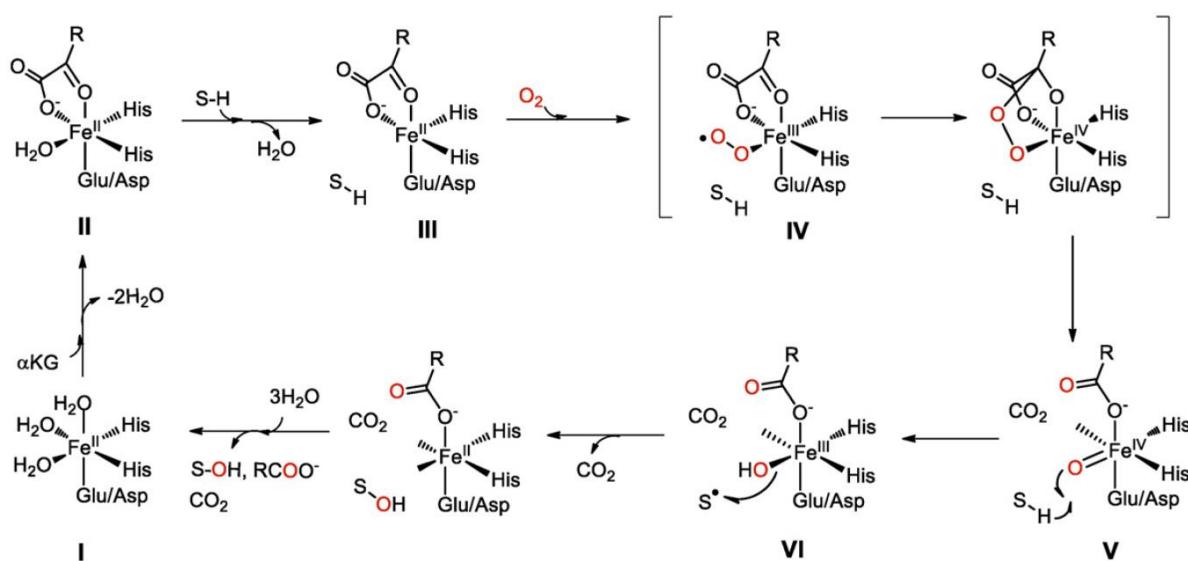


Figure 8: Proposed consensus catalytic mechanism of the α -KAO catalyzed hydroxylation reaction

Reprinted from Wu *et al.*, 2016

The α -KG keto group is positioned opposite to the Glu/Asp, and the α -KG carboxylate is positioned opposite to one of the His residues, thus displacing two metal-bound water molecules (II-III). Then the substrate binds to the sixth coordination site, changing the iron to five-coordinate geometry allowing for O_2 binding/activation (III-IV). The active Fe^{3+} -superoxo intermediate attacks the α -KG group which triggers decomposition of α -KG to CO_2 and succinate leading to the generation of the highly reactive Fe^{4+} species (IV-V).

The crystal structure of some α -KAOs has been solved in the presence of the specific substrate, revealing residues of the active center (Elkins *et al.*, 2002; Helmetag *et al.*, 2009; Qin *et al.*, 2013). One of the most extensively studied α -KAO enzymes, the taurine

dioxygenase TauD of *E. coli*, shows some flexibility in its catalytic site, enabling TauD to be active on a range of sulfonic acid substrates (Eichhorn *et al.*, 1997).

I.4.4 Candidates for evolution

In an attempt to identify new enzymes of the α -KAO family able to broaden the set of substrates hydroxylated, Baud and coworkers screened gene sequences deposited in databases – mainly from prokaryotic genomes - for homologs (Baud *et al.*, 2014). A total of 131 candidate genes were thus identified, amplified via PCR, cloned and expressed in a high throughput format. Enzymatic activities were screened for different substrates using a generic spectrophotometric test. Emphasis was put on candidates showing activity for lysine or ornithine since none of the two amino acids are substrates of a α -KAO described so far. Six new enzymes were found to be active towards L-lysine, L-ornithine, and L-arginine on this screen. Further testing revealed a high stereospecificity and stereoselectivity of these enzymes for the basic amino acids (Baud *et al.*, 2014). However, neither the *in vivo* substrate nor the enzymatic activity parameters have been determined for these catalysts.

Since stereo-controlled hydroxylations of polar amino acids can provide synthons for the production of complex chiral natural or non-natural compounds, these catalysts were chosen for this study (Table 2). The final objective was to ameliorate their activity towards the natural amino acid and to alter their specificity towards non-natural homologs.

In addition to these new α -KAOs, the enzyme L-isoleucine dioxygenase (Ido) was studied. This enzyme was described to hydroxylate stereospecifically isoleucine at carbon 4, and to complement an *E. coli* strain auxotrophic for succinate. It was also shown to exhibit some activity for structural homologs of isoleucine, rendering it suitable as a model enzyme for activity amelioration via directed evolution in continuous culture (Hibi *et al.*, 2011).

Table 2: Candidates enzymes for directed evolution *in vivo* in GM3

| Enzyme | UniProtKB ID | Organism | Substrate | Product |
|--------|--------------|--|---------------------------|---|
| Kdo1 | C7QJ42 | <i>Catenulispora acidiphila</i> (DSM 44928) | L-lysine | (3S)-3-hydroxy-L-lysine |
| Kdo2 | C7PLM6 | <i>Chitinophaga pinensis</i> (DSM 2588) | L-lysine | (4R)-4-hydroxy-L-lysine |
| Kdo3 | A5FF23 | <i>Flavobacterium johnsoniae</i> (DSM 2064) | L-lysine | (4R)-4-hydroxy-L-lysine |
| Kdo4 | G8T8D0 | <i>Niastella koreensis</i> (DSM 17620) | L-lysine | (4R)-4-hydroxy-L-lysine |
| Kdo5 | J3BZS6 | <i>Flavobacterium sp.</i> (strain CF136) | L-lysine | (4R)-4-hydroxy-L-lysine |
| Odo | C7Q942 | <i>Catenulispora acidiphila</i> (DSM 44928) | L-ornithine L-arginine | (3S)-3-hydroxy-L-ornithine (3S)-3-hydroxy-L-arginine |
| Ido | Q81GX0 | <i>Bacillus cereus</i> | L-isoleucine | (4S)-4-hydroxy-L-isoleucine |

I.5 Fixation of beneficial mutations in a gene under selection in an evolving cell population

As detailed previously, the appearance of spontaneous mutations in cells has been the subject of numerous studies which have enabled the determination of mutation frequencies. By contrast, questions concerning the fixation of adaptive mutations in a cell population, influenced by the nature of the selection, but also by the copy number and expression profile of the targeted gene, have not been addressed in detail. Knowledge about the influence of these parameters on the fixation kinetics of mutations in continuous culture regimes may help to optimize the setup of a given genetic screen.

I.5.1 Experimental strategy

In order to study the impact of gene expression formats on the penetration of a beneficial mutation throughout an evolving population, we employed a strategy based on the reversion of a point mutation introduced into a gene of *E. coli* coding for an essential enzymatic activity. The mutation entails a loss of function of the enzyme resulting in an auxotrophic phenotype of the cells harboring the mutated allele. Upon reversion of the point mutation, the auxotrophic growth restriction is abolished. This event can be selected in continuous culture enabling the comparison of the influence of

different growth regimes and expression formats of the mutated gene on the appearance and the fixation of the reversion.

In Figure 9 case A, the essential gene of interest (*goi*) codes for the enzyme that catalyzes the reaction converting the compound α into the compound β in the WT strain. This reaction is essential to the normal growth of the bacteria as it needs compound β to survive.

In case B, the genetic screen is the strain where the essential gene has been deleted (Δgoi) making it unable to produce the essential enzyme. Thus the catalysis of the reaction converting the compound α into the compound β does not occur, and the strain becomes auxotrophic for the compound β .

In case C, variants of the essential gene with no or reduced activity (*goi**) are designed and expressed in the genetic screen whose fitness directly relies on the enzyme activity. Partial or no complementation for compound β depends on the enzyme activity and the expression format.

The case D shows the expected scenario where a point mutation in a targeted gene can be reverted under selection in continuous culture. If a reversion occurs, *i.e.* a mutation that restores the WT activity of the enzyme, the revertant should be selected and fixed in the population over time. The reversion allowing full recovery of prototrophy, the kinetic of fixation of the corresponding point mutation in the population can be monitored following the growth profile of the culture in the GM3.

The following questions will be addressed:

1. What is the influence of the culture regime in the kinetics of evolution?
2. What is the influence of the expression format in evolution?

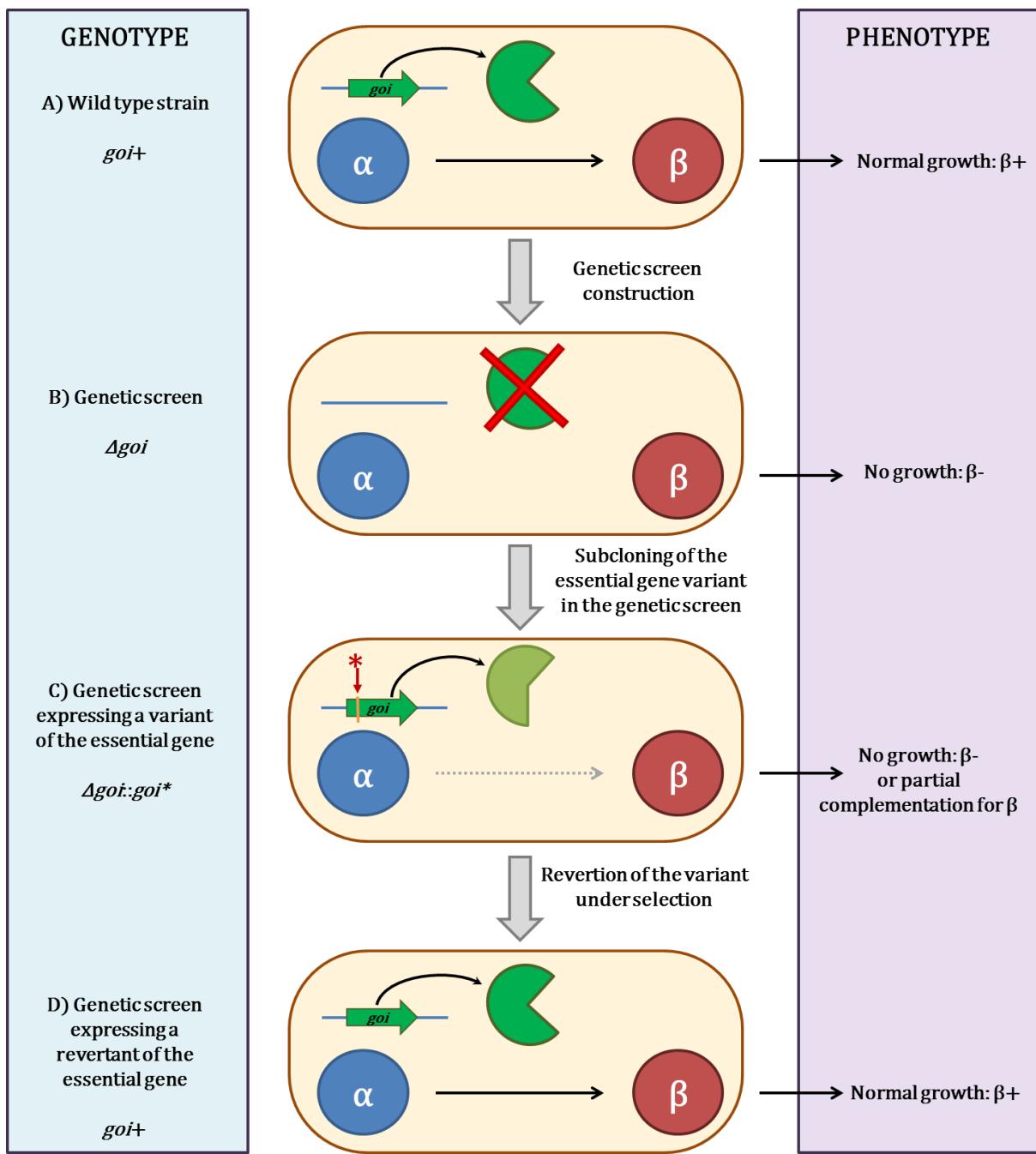


Figure 9: Simplified scheme of the experimental design for studying the fixation of a point mutation in a gene of interest

II MATERIALS & METHODS

II.1 Molecular biology protocols

II.1.1 Gene cloning

Gene cloning consists in introducing the coding sequence of a gene of interest into a vector designed for this purpose, which allows easy selection of the insertion event, possibility of subcloning and selectable transfer in the desired genetic background.

Standard molecular biology protocols were used in order to clone dioxygenases into different vectors (Table 3, Figure 10, Figure 11, and Figure 12).

Table 3: List of vector used for directed evolution *in vivo*

| Name | Selection | ORI | Copy number | Size | Derived from |
|---------|-------------|------|-------------|---------|--------------|
| pSP100 | <i>bla</i> | pMB1 | 20-700 | 2706 bp | pUC18 |
| pVDM18 | <i>cat</i> | P15A | 5-15 | 2317 bp | pSU18 |
| pEVL404 | <i>thyA</i> | pMB1 | 20-700 | 2599 bp | pUC18 |

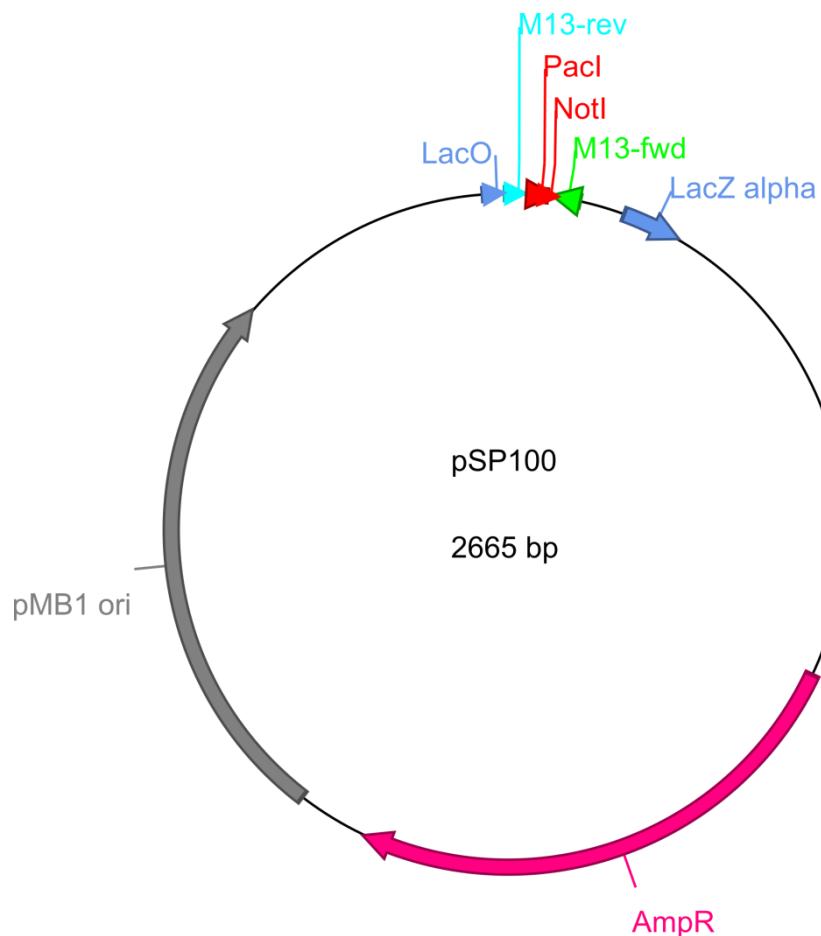


Figure 10: Map of plasmid pSP100

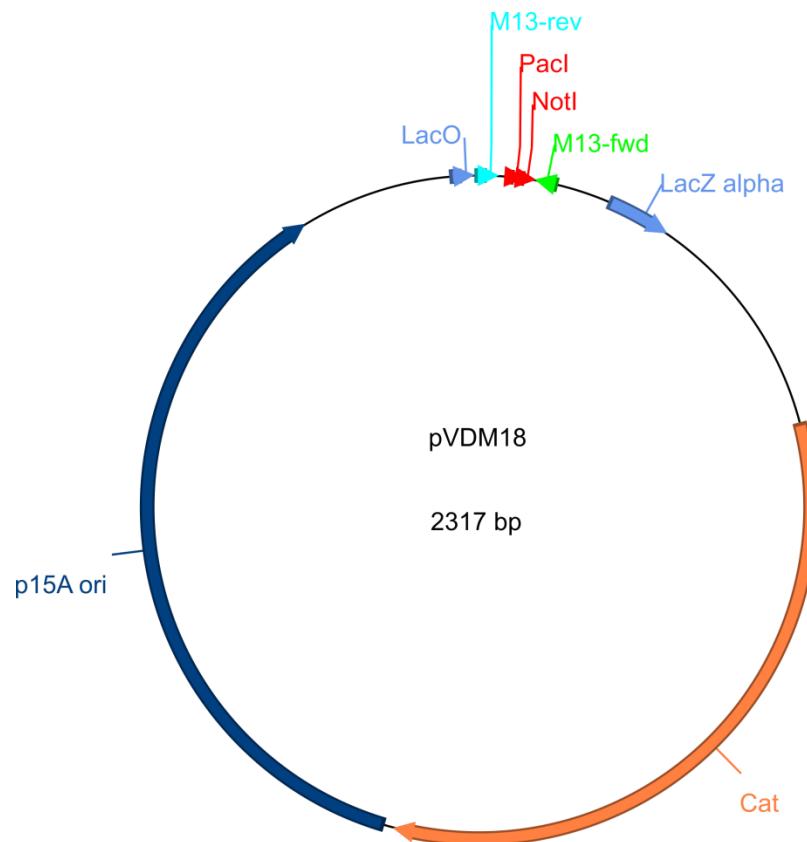


Figure 11: Map of plasmid pVDM18

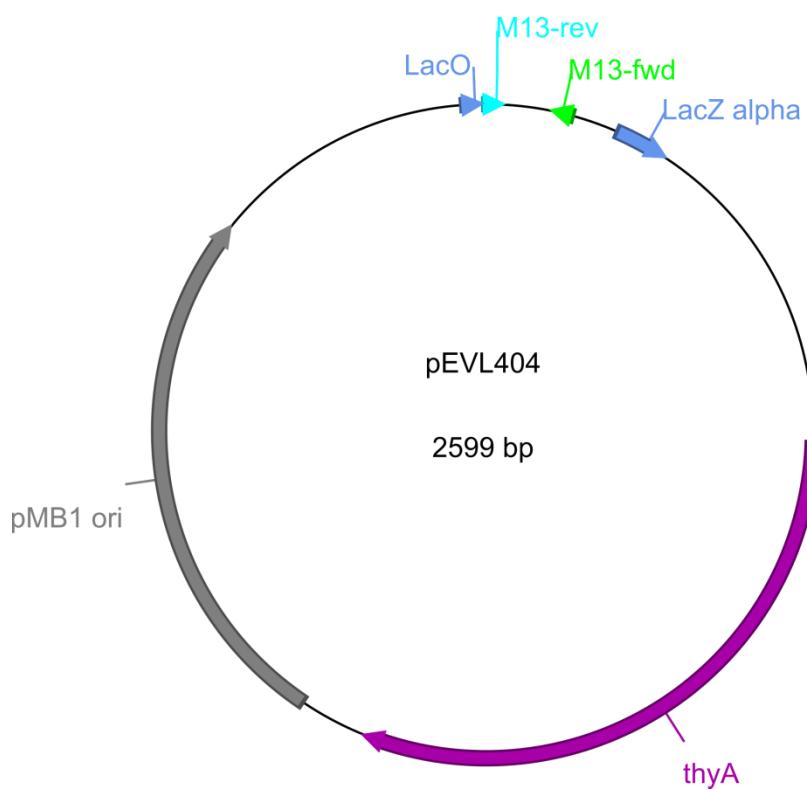


Figure 12: Map of plasmid pEVL404

Figure 13 illustrates a simplified workflow for this procedure. Important steps are described in greater details in the following sections.

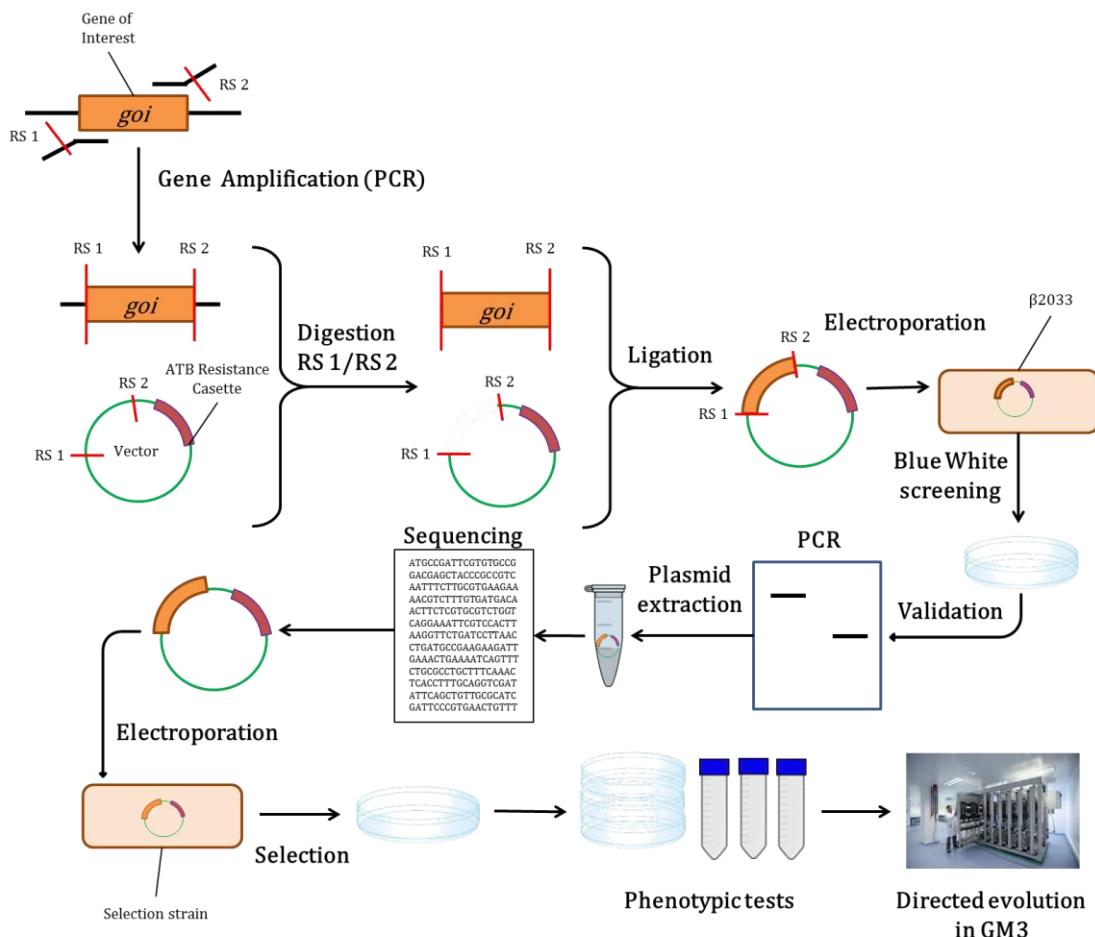


Figure 13: Simplified workflow of strain engineering for directed evolution with plasmids
RS 1: Restriction site 1; RS 2: Restriction site 2; *goi*: gene of interest

II.1.1.1 DNA amplification and purification

Genes of interest were amplified via Polymerase Chain Reaction (PCR). It consists in iterative polymerizations of the gene template with specialized DNA polymerases using specific primers, enabling exponential amplification of the template. For the cloning of the WT genes, specific primers were designed to have restrictions sites flanking the gene sequence (Table 4) and the sequence corresponding to the gene.

PCR reactions were carried out in 96-well microplates in 25, 50 or 100 µl reaction mix with Platinum® Pfx DNA polymerase (ThermoFischer, Ref: 11708021) (for high fidelity reactions) or DreamTaq™ DNA Polymerase (ThermoFischer, Ref: EP0701) (for verification) according to the manufacturer's protocol.

DpnI enzyme (ThermoFischer, Ref: FD1703) which specifically digests methylated DNA was used to eliminate template DNA in order to purify PCR products and avoid background in subsequent reactions.

Table 4: List of primers used for cloning genes

| Gene | Primer | Restriction site | Sequence (5' to 3') | Used for cloning in |
|-------------|---------|------------------|--|--------------------------------|
| | M13 Fwd | | AGGGTTTCCCAGTCACGACGTT | pSP100, |
| | M13 Rev | | GAGCGGATAACAATTACACAGG | pVDM18, pEVL404 |
| <i>metA</i> | 2203 | PacI | CCCTTAATTAATGCCGATTCTGTGCCGG | pSP100, |
| | 2204 | NotI | TATGCGGCCGCTTAATCCAGCGTGGATTAC | pVDM18, pEVL404, pEVL411 |
| <i>ido</i> | 4900 | PacI | CCCTTAATTAATGGAGGTTTTATAATGACGTTTGTT | pSP100, |
| | 4898 | NotI | TATGCGGCCGCTTATTTGTCTCCTTATAAGAAAATGTTAC | pVDM18, pEVL411 |
| <i>odo</i> | 4833 | PacI | CCCTTAATTAATGCACCGCTTGGCCCTGAC | pSP100, |
| | 4834 | NotI | TATGCGGCCGCTTAGTAAATGACCCGGTGTCCG | pVDM18 |
| <i>kdo1</i> | 4823 | PacI | CCCTTAATTAATGAAGAACCTGTCTGCGTATGAAG | pSP100, |
| | 4824 | NotI | TATGCGGCCGCTTAGCTGAACCTCGCAGAGACG | pVDM18 |
| | 4890 | SphI | TATGCATGCTTAGCTGAACCTCGCAGAG | pEVL404 |
| <i>kdo2</i> | 4825 | PacI | CCCTTAATTAATGAGACCCCTAGACGTGACAC | pSP100, |
| | 4826 | NotI | TATGCGGCCGCTTAAAGGTTGCCAGGTGAG | pVDM18, pEVL411 |
| | 5059 | XbaI | TATTCTAGATTAAAGGTTGCCAGGTGAGC | pEVL404 |
| <i>kdo3</i> | 4827 | PacI | CCCTTAATTAATGAAATCACAACTATTAATTG | pSP100, |
| | 4828 | NotI | TATGCGGCCGCTTAAGCCTGATCAAAACTT | pVDM18 |
| | 5060 | XbaI | TATTCTAGATTAAAGCCTGATCAAAACTTTCC | pEVL404 |
| <i>kdo4</i> | 4829 | PacI | CCCTTAATTAATGAAACCACATCGAATCTCG | pSP100, |
| | 4830 | NotI | TATGCGGCCGCTTACTGCTGAGAGTGGAACAGTTACC | pVDM18 |
| | 5061 | XbaI | TATTCTAGATTACTGCTGAGAGTGGAACAGTTAC | pEVL404 |
| <i>kdo5</i> | 4831 | PacI | CCCTTAATTAATGAAATCTCAGTCTATCATGTCTG | pSP100, |
| | 4832 | NotI | TATGCGGCCGCTTAGTCCAGGTGCGAAGATTTACC | pVDM18 |

Restriction sites PacI (**TTAATTA**A), NotI (**GCGGCCG**C), SphI (**GCATG**C), XbaI (**TCTAG**A)

For DNA purification steps, kits from Qiagen were used following manufacturer's instructions:

- QIAQuick PCR Purification kit (Qiagen, Ref: 28014) was used to purify DNA fragments obtained through PCR amplification or plasmid digestion with restriction enzymes;
- QIAprep kit (Qiagen, Ref: 27104) was used to extract plasmids from overnight precultures;
- MinElute kit (Qiagen, Ref: 28004) was used to purify ligation products;
- Gel extraction kit (Qiagen, Ref: 28604) was used to extract DNA fragments of determined length obtained after plasmid digestion with restriction enzymes and separation of the digestion products by electrophoretic migration in 0.8% agarose gels.

Yield and length of PCR products were verified using 0.8% agarose gel electrophoresis in 0.5X TBE in the presence of bromoethidium and visualization under UV.

II.1.1.2 Digestion

The PCR products to be cloned were digested in a final volume of 30 µl along with the backbones by the appropriate enzyme. The digested products were purified with a QIAQuick PCR Purification kit (Qiagen, Ref: 28014) and their concentrations determined by Nanodrop. DNA was stored at -20°C.

II.1.1.3 Ligation

Digested inserts and vectors were mixed in a molar ration of 5:1 and incubated with Quick ligase for 5 minutes at room temperature in a final volume of 10 µl following supplier's instructions (Fermentas). Ligation products were immediately purified (MinElute), and then used for cell transformation or stored at -20°C.

II.1.1.4 Transformation

Electrocompetent cells of strain β 2033 were prepared according to the common protocol. Each plasmid was introduced into competent cells by electroporation. For this, 50 µl of competent cells were mixed on ice with 2 µl of ligation product or 0.5 µl of purified plasmid in an ice-cold 1 mm wide electroporation cuvette. Cells were

electroporated at 1.8 kV, immediately resuspended in 1 ml of LB medium supplemented with nutrients when needed. The transformed strains were incubated at 37°C with agitation for 45 minutes in order to allow the expression of antibiotic resistance gene before applying selection.

II.1.1.5 Blue White screening

Transformants were selected on LB medium-agar plates with antibiotic, the colorless analog of lactose X-galactoside, and IPTG. IPTG induces the transcription of the gene located downward the pLac promoter. X-galactoside is a substrate of β -galactosidase LacZ; its cleavage by LacZ releases 5-bromo-4-chloro-3-hydroxyindole, which spontaneously dimerizes and is rapidly oxidized to form a bright blue insoluble pigment 5,5'-dibromo-4,4'-dichloro-indigo. This reaction allows distinguishing the colonies harboring an empty plasmid (blue colonies) from those harboring a plasmid where the insert of interest has been successfully ligated (white colonies).

II.1.1.6 Validation

The conformity of the plasmids present in the selected white colonies was validated in two steps. Firstly, presence and length of the insert in the extracted plasmid were checked after PCR DNA amplification using primers homologous to vector specific sequences (M13FW and M13RV) and analysis of PCR products in electrophoresis agarose gel with a reference ladder. Secondly, the conformity of the insert DNA sequence was verified by Sanger sequencing.

II.1.2 Directed mutagenesis

Directed mutagenesis protocols are used to introduce targeted point mutations in genes of interest. Figure 14 illustrates a simplified workflow for this procedure. Important steps are described in greater details in the following sections.

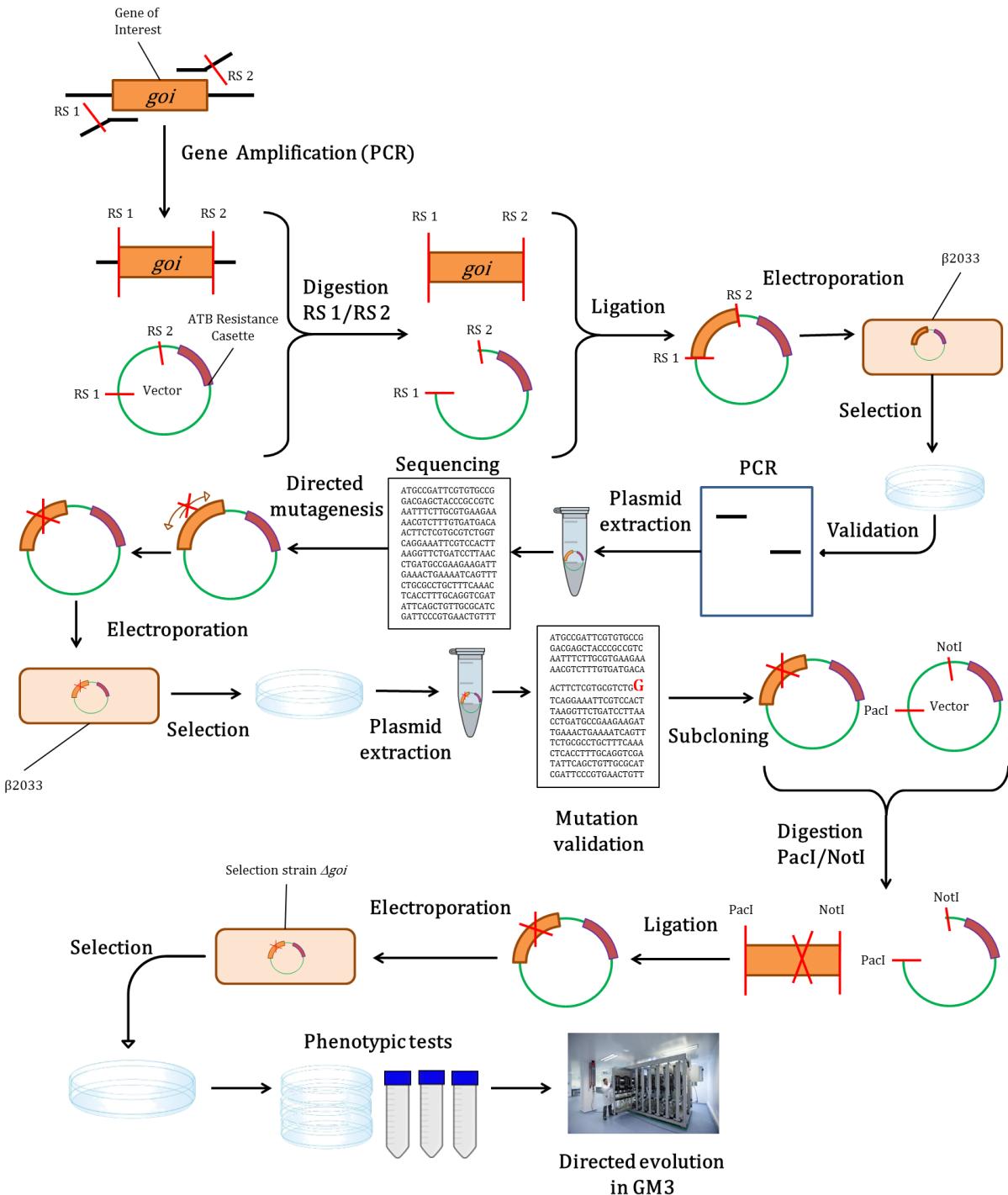


Figure 14: Simplified workflow of directed mutagenesis for directed evolution *in vivo* with plasmids
RS 1: Restriction site 1; RS 2: Restriction site 2; *goi*: gene of interest; Red cross: point mutation

The mutation is introduced using PCR primers homologous to the sequence of the gene locus to be modified but differing by one base in the targeted codon to be modified (Figure 15).

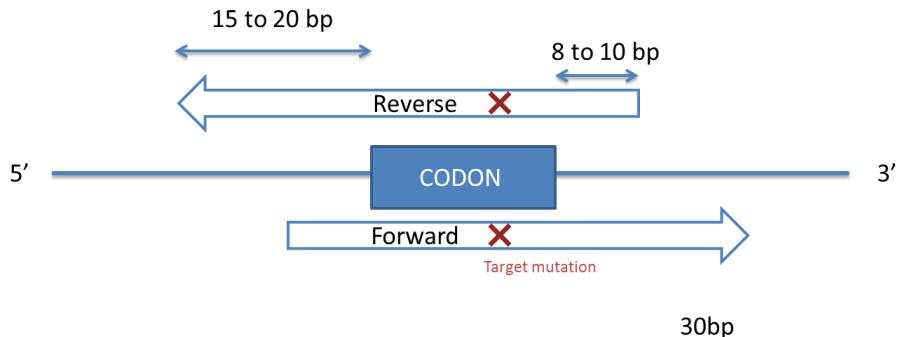


Figure 15: Design of mutagenesis primers

Primers designed and used for directed evolution in this study are listed in Table 5.

Table 5: Table of primers designed for directed mutagenesis of *metA*

| Primers | Gene | Target mutation | Primer sequence (5' -> 3') |
|---------|-------------|-----------------|--|
| 4924 | <i>metA</i> | K46R_fw | CCTGATGCCGA <u>GG</u> AAGATTGAAACTGAAAATCAGTTCTGCCCTG |
| 4925 | <i>metA</i> | K46R_rv | CAGTTTCAATCTT <u>CC</u> CGCATCAGGTTAAGGATCAGAACCTTAAGTGG |
| 4926 | <i>metA</i> | K47R_fw | CCTGATGCCGAAG <u>GG</u> ATTGAAACTGAAAATCAGTTCTGCCCTG |
| 4927 | <i>metA</i> | K47R_rv | CAGTTTCAATCCT <u>CT</u> CGCATCAGGTTAAGGATCAGAACCTTAAGTGG |
| 4928 | <i>metA</i> | K46R K47R_fw | TAACCTGATGCC <u>GG</u> ATTGAAACTGAAAATCAGTTCTGCCCTG |
| 4929 | <i>metA</i> | K46RK47R_rv | TTTCAGTTCAAT <u>CC</u> CGCATCAGGTTAAGGATCAGAACCTTAAG |
| 4951 | <i>metA</i> | C142R_fw | GTTTGTC <u>CG</u> CTGGCGGTACAGGCCGCGCTCAATATCC |
| 4952 | <i>metA</i> | C142R_rv | GTACCGCCC <u>AG</u> CGACAAACAGCGTCGAGGTGACGTG |
| 4953 | <i>metA</i> | C142Y_fw | GTTTGTC <u>TA</u> CTGGCGGTACAGGCCGCGCTCAATATCC |
| 4954 | <i>metA</i> | C142Y_rv | GTACCGCCC <u>CA</u> GTGACAAACAGCGTCGAGGTGACGTG |
| 4955 | <i>metA</i> | H235Y_fw | GACGGGC <u>TA</u> TCCGAATATGATGCCAACGCTGGCGAGGAATTTTC |
| 4956 | <i>metA</i> | H235Y_rv | CATATTGGG <u>AT</u> ACCCCGTACAAAGGCAATGCCCTATCTTACTGGC |
| 4957 | <i>metA</i> | H235R_fw | GACGGGC <u>CG</u> TCCGAATATGATGCCAACGCTGGCGAGGAATTTTC |
| 4958 | <i>metA</i> | H235R_rv | CATATTGGG <u>AC</u> GGCCGTACAAAGGCAATGCCCTATCTTACTGGC |
| 4959 | <i>metA</i> | H235S_fw | GACGGGC <u>AG</u> TCCGAATATGATGCCAACGCTGGCGAGGAATTTTC |
| 4960 | <i>metA</i> | H235S_rv | CATATTGGG <u>AC</u> TGCCGTACAAAGGCAATGCCCTATCTTACTGGC |
| 4961 | <i>metA</i> | E237D_fw | CCATCCC <u>GA</u> TATGATGCCAACGCTGGCGAGGAATTTCCGCG |
| 4952 | <i>metA</i> | E237D_rv | GCGCATCATA <u>AT</u> GGGGATGGCCGTACAAAGGCAATGCCCTATC |
| 4963 | <i>metA</i> | E237A_fw | CCATCCC <u>GA</u> TATGATGCCAACGCTGGCGAGGAATTTCCGCG |
| 4964 | <i>metA</i> | E237A_rv | GCGCATCATA <u>GC</u> GGGGATGGCCGTACAAAGGCAATGCCCTATC |

In order to perform site-directed mutagenesis, the Platinum® Pfx DNA polymerase was used because of its very high levels of fidelity. The PCR mix was performed in 100 µl reaction instead of 50 µl for higher yield. The following PCR protocol was used: melting, 95°C-5'; [melting, 95°C-30"; annealing and amplification, 68°C-4'30"]x20; termination, 68°C-5'. The time of elongation is determined according to the plasmid length and polymerase processivity (approximately 1 minute for 1000 bp for Platinum® Pfx DNA polymerase). Template DNA was digested with DpnI for 2 hours at 37°C. PCR products were verified by electrophoresis on a 0.8% agarose gel and extracted from the gel.

The purified plasmids were electroporated into the β 2033 strain for amplification. Plasmids were extracted with a Qiaprep kit (Qiagen, Ref: 27104) according to manufacturer's protocol. The presence of the mutation was verified by Sanger sequencing. To avoid any other undesired mutation in the backbone, a subcloning step was added where the plasmid containing the mutated gene was digested with PacI and NotI and ligated into the original vector. The selection strain specific to the gene tested was then transformed with the plasmid from the subcloning step. Finally, phenotypic tests were performed in order to determine the growth rate of each mutation and the parameters to set for directed evolution in the GM3.

II.2 Molecular genetics protocols

II.2.1 Genetic screen construction

Various molecular genetics protocols were applied in order to construct the genetic contexts required for directed evolution experiments *in vivo*. For this study, several metabolic screens were used. All genetic constructs derived from the *E. coli* K12 strain MG1655, the selection strains were built by serial substitutions of chromosomal loci with deletion alleles. The various deletion alleles used in this study arose from the *E. coli* K12 single-gene knockout KEIO collection. They were introduced in the recipient bacteria by homologous recombination mediated by transduction with the P1 bacteriophage (Miller, 1972).

The KEIO Collection is a library of 3985 single-gene knockout mutants of *Escherichia coli* K-12 (Baba *et al.*, 2006). Each gene that has been knocked out in this library was replaced by a Kanamycin resistance cassette flanked by FLP recognition target (FRT) sites and 50 bp homologies to adjacent chromosomal sequences. For each deletion, a

phage P1 was prepared of the targeted deletion. The transduction was then performed on the host strain.

II.2.1.1 P1 bacteriophage-mediated allele transduction

Transduction is a process by which DNA is transferred from one bacterium to another by a bacteriophage. P1 bacteriophages were used to transfer one or more selectable alleles in the same recipient strain (Miller, 1972). For this, a stock of P1 phages was prepared by infecting a growing culture of the donor strain with an aliquot of lytic P1 phages. During their lytic infection cycle of *Escherichia coli*, P1 phages can mistakenly encapsidate bacterial DNA fragments in place of normal phage DNA, but remaining infectious. This erroneous phenomenon is rare, but if some selectable marker is present on bacterial DNA, it can be recovered in the convenient recipient strain after genetic transfer by infection and homologous recombination with chromosomal DNA and allelic substitution.

II.2.1.1.1 P1 Bacteriophage preparation

The donor strain was grown in 2 ml LB medium at 37°C, 150 RPM, until OD₆₀₀ = 0.7. To the culture was added CaCl₂ at a final concentration of 2.5 mM and the culture was incubated 30 min at 37°C, 150 RPM. A volume of 100 µl of donor strain culture was added to 100 µl of P1 phage diluted to 10⁻³ and 10⁻⁴ in triplicates. A negative control with 100 µl of water instead of phages was also prepared. All tubes were incubated 20 min at 37°C without shaking. The whole volume of 200 µl is transferred into 4 ml of semi-solid agar (maintained at 60°C) and immediately poured on an LB petri dish and incubated at 37°C overnight. Where phages have formed plaques of lysis, semi-solid agar was harvested, and 800 µl of CHCl₃ was added for each petri dish harvested. Tubes are mixed and centrifuged for 10 min at 10 000 RPM. The supernatant is stored at 4°C with few drops of CHCl₃.

II.2.1.1.2 P1-mediated transduction

The host strain was grown in 5 ml LB medium at 37°C, 150 RPM, until OD₆₀₀ = 0.7. To the culture was added CaCl₂ at a final concentration of 2.5 mM and the culture was incubated 30 min at 37°C, 150 RPM. To 1 ml of the host-strain culture is added 3 µl and 30 µl of P1 phages prepared with the donor strain. A negative control with water instead of P1 phage was also prepared. All tubes were incubated 15 min at 37°C without shaking

then centrifuged for 1 min at 7 000 RPM. The pellet was resuspended in 1 ml of LB citrate 0.1 M, incubated 45 min at 37°C, 150 RPM and centrifuged for 1 min at 7 000 RPM. The pellet was resuspended in 100 µl of LB and spread on a petri dish with the appropriate selection and incubated at 37°C overnight.

II.2.1.2 Cassette excision

To gather multiple gene deletions in a same bacterial construct, the cassette for antibiotic resistance has to be excised to enable its use for further recombinations. In order to perform cassette excision, the pCP20 plasmid was used. The pCP20 contains the yeast FLP recombinase gene, a temperature sensitive replication origin and chloramphenicol and ampicillin resistance genes (Figure 16).

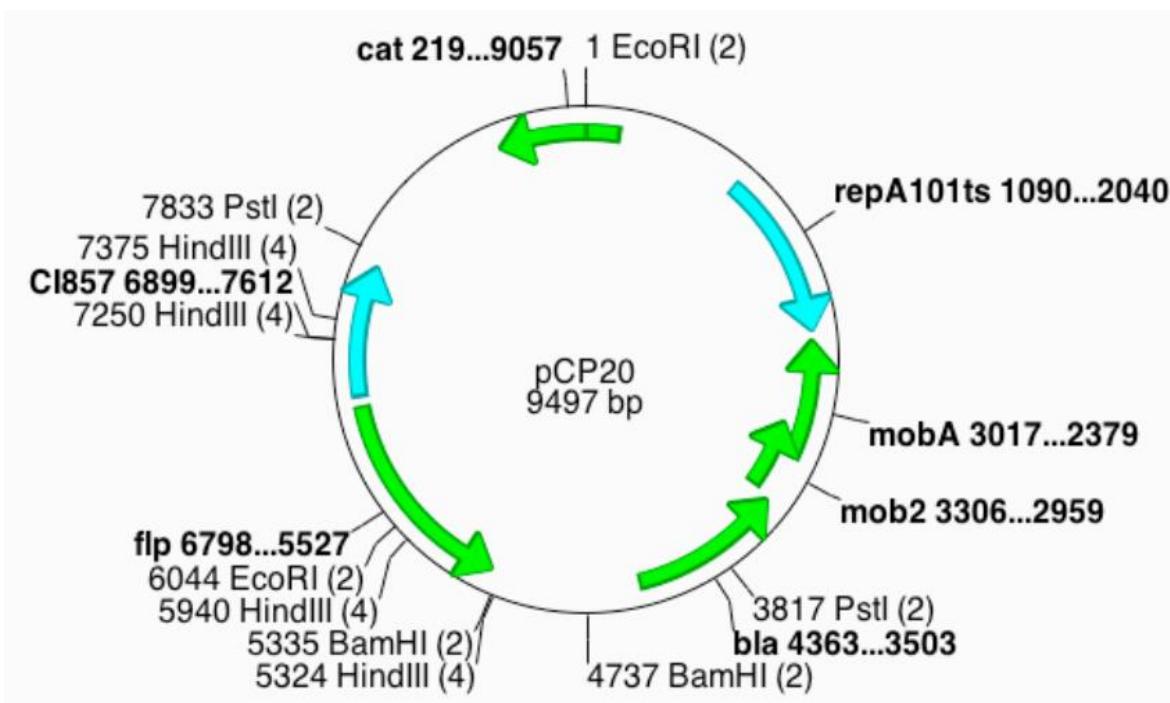


Figure 16: Map of plasmid pCP20

The FLP recombinase recognizes the FRT target sites flanking the resistance cassette (Figure 17). Electrocompetent cells of the strain were prepared and transformed with pCP20 (1.8 kV, cuvette gap 1 mM). Cells were spread on an LB ampicillin petri dish and incubated at 30°C overnight for FLP expression and excision. Selected clones were streaked on an LB petri dish and incubated at 42°C overnight for pCP20 degradation. Clones were tested on LB ampicillin, LB antibiotic, LB at 37°C in order to verify the excision of the targeted cassette and the plasmid degradation.

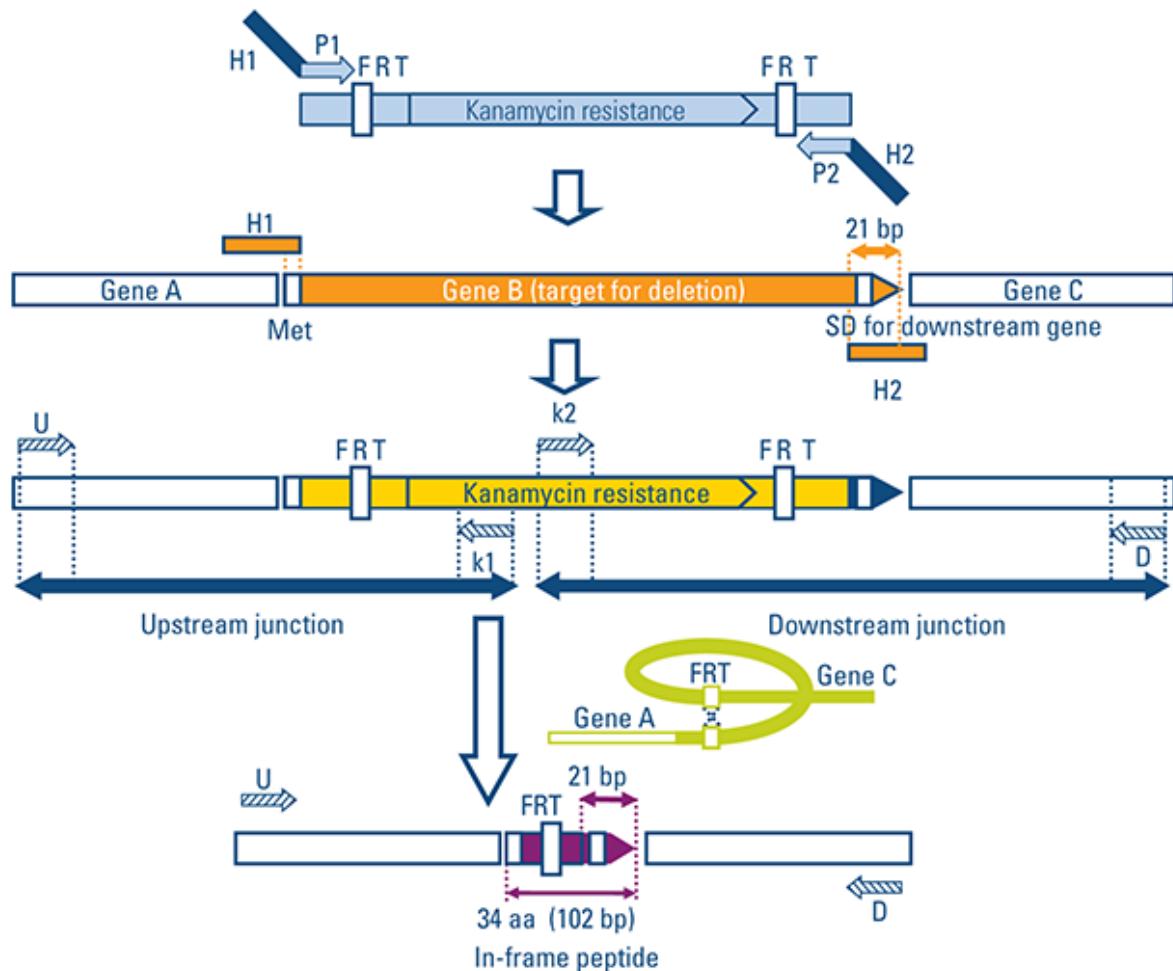


Figure 17: Construction of single-gene deletion mutants

II.2.2 Allelic exchange

Wanner's method (Datsenko and Wanner, 2000) enables efficient allelic substitution on *E. coli* chromosome by homologous recombination using the recombinase of phage lambda. This method has been applied to insert a copy of a gene in the *E.coli* chromosome at the *kdgK* locus (Bouzon, M *et al.*, 2017).

As shown in Figure 18, the simplified workflow for this procedure is the following:

1. Cloning of the gene of interest into the pEVL411 vector (Figure 19)
2. Amplification of the recombinant plasmid in the BW19610 strain
3. Gene recombination in the +2440 strain induced by arabinose
4. Recombinant selection and characterization
5. Transduction of the chromosomal allele in the appropriate genetic screen

Each step is explained in greater details in the following sections.

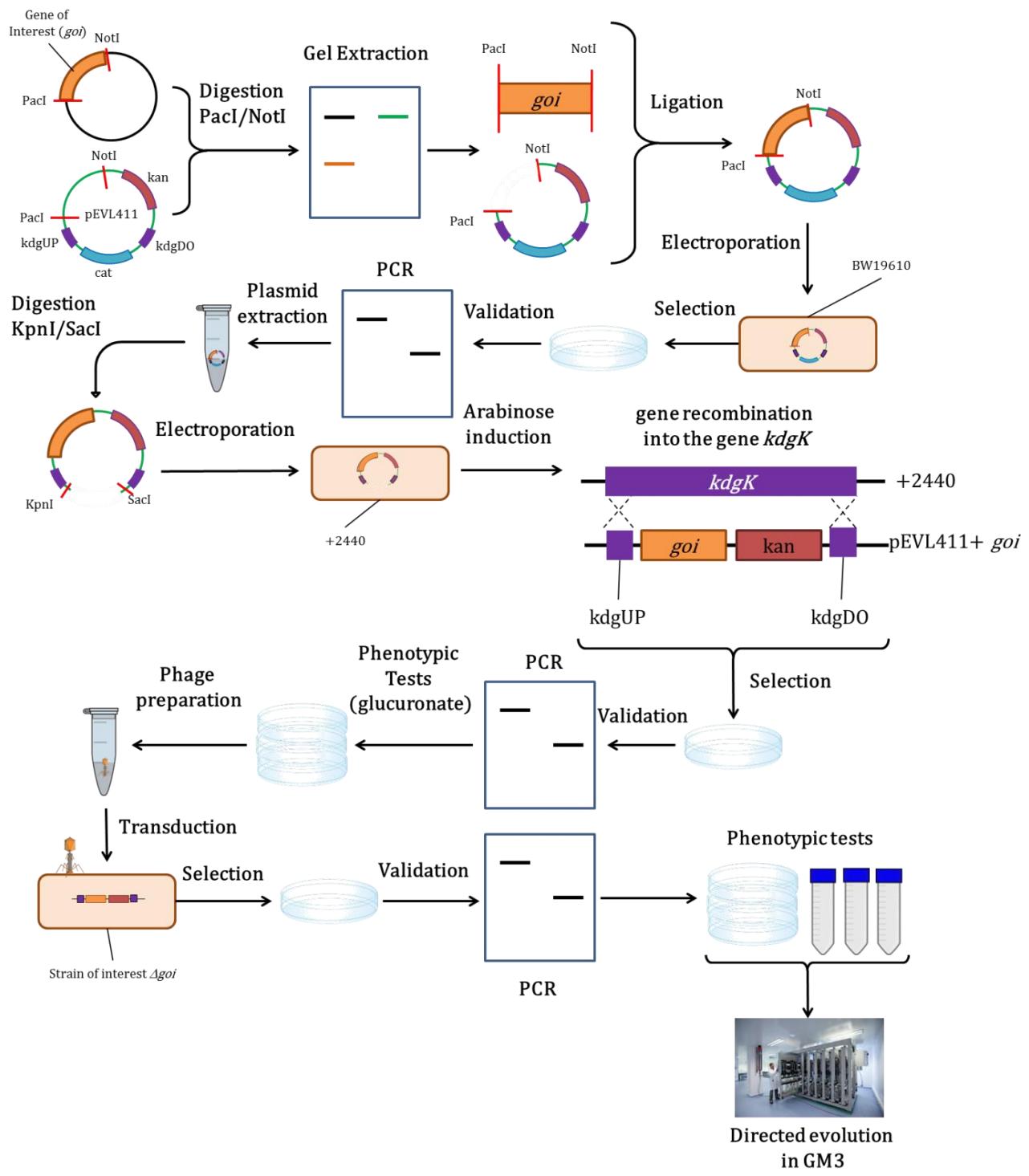


Figure 18: Simplified workflow of strain engineering for directed evolution with chromosomal insertion

II.2.2.1 Gene cloning in pEVL411

To insert a gene in the chromosome of a strain, the plasmid pEVL411 was used. The plasmid contains the sequences homologous upstream (kdgKUP) and downstream (kdgKDO) of the *kdgK* locus, R6K origin, chloramphenicol, and kanamycin resistance cassettes (Figure 19).

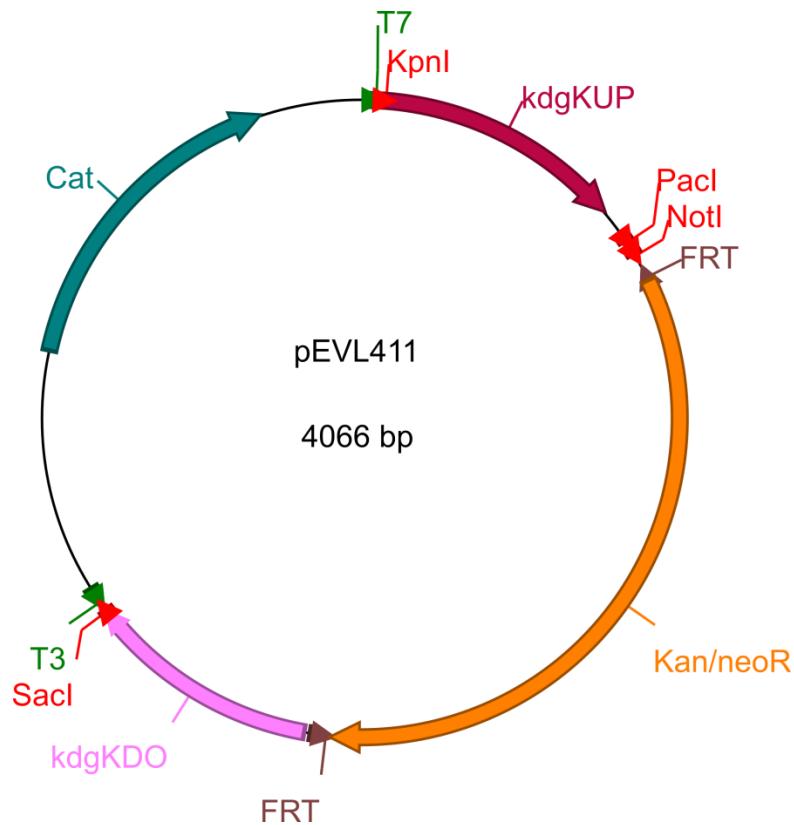


Figure 19: Map of plasmid pEVL411

The gene of interest is cloned in the pEVL411 between the PacI/NotI restriction sites (Table 4) as described in II.1.1 and amplified in the BW19610 strain which allows the propagation of plasmid harboring the R6K origin of replication.

II.2.2.2 Gene recombination

The resulting purified plasmid is linearized by double digestion with SacI/KpnI. A growing culture of +2440 bacteria at 30°C is incubated with L-arabinose for 45 minutes (10 mM). The strain +2440 is an MG1655 containing the plasmid pKD46 λ-red recombinase+. L-arabinose induces the expression of λ-red, the recombinase protein of phage λ, which is under the control of an araBAD promoter. Electrocompetent bacteria are prepared from the +2440 induced bacteria. A volume of 2 µl linearized pEVL411 was used in 50 µl of electrocompetent +2440 for transformation (1.8 kV, cuvette gap 1 mM). Recombinant cells were selected on LB kanamycin at 37°C. The insertion of the gene by recombination event was validated by PCR using specific primer matching outside the construct up and down from *kdgK* locus. It was also tested phenotypically as the *kdgK* gene coding for 2-keto-3-deoxygluconokinase is necessary for *E. coli* to grow with glucuronate as the sole source of carbon. Growth phenotype of recombinant was determined on MS glucuronate (0.2%) plates to assess recombination at the *kdgK* locus. Clones validated by PCR and unable to grow with glucuronate were stored at -80°C.

II.2.2.3 Allele transduction

A preparation of P1 phage as described in II.2.1.1.1 was performed on the +2440 strain with the gene of interest recombined in the *kdgK* locus. The following transduction was performed as described in II.2.1.1.2 on the host strain. Chromosomal insertion was confirmed by PCR verification and glucuronate phenotypic tests as described in II.2.2.2.

II.3 Dioxygenase biochemistry

II.3.1 Protein production

II.3.1.1 Gene cloning with the LIC method

In order to determine the dioxygenases kinetic parameters, each protein must be produced and purified for *in vitro* testing.

Genes of interest were cloned in a pET-22 (+) vector with N-terminal His•Tag® or C-terminal His•Tag® sequence with the Ligation Independent Cloning (LIC) method (Aslanidis and Jong, 1990). This technique uses the 3' → 5' exo activity of T4 DNA Polymerase to create overhangs with complementarity between the vector and the insert. Incorporation of dGTP in the reaction limits the exonuclease processing to the first complementary C residue, and not present in the designed overlap, where the polymerization and exonuclease activities of T4 DNA Polymerase become "balanced". Joined fragments have 4 nicks that are repaired by *E. coli* during transformation. This technique allows efficient creation of scarless recombinant plasmids at many, but not all, positions in a vector (Figure 20).

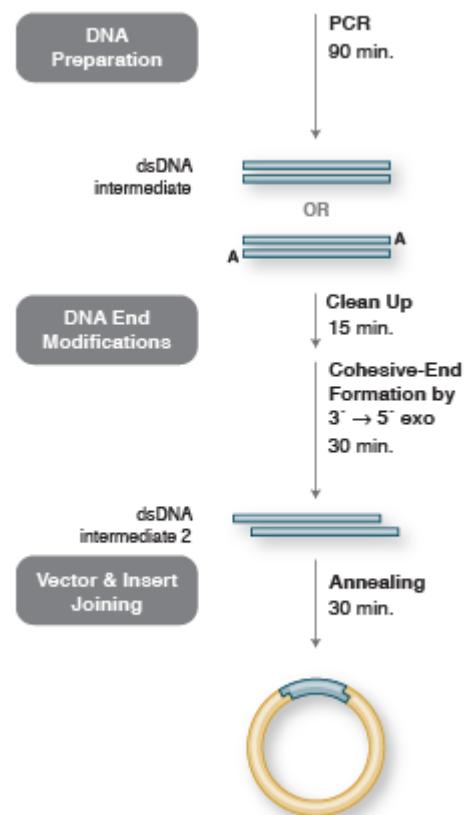


Figure 20: Ligation Independent Cloning workflow

Primers were designed to incorporate the N-terminal or C-terminal His•Tag® (Table 6). Genes of interest were amplified with the appropriate primers and purified using QIAquick PCR purification protocol. The pET22 (+) plasmid was digested with EcoRV and purified. The purified insert and pET22 (+) were treated with 2 units T4 DNA

polymerase in the presence of dGTP 0.5 mM, 100 µg.ml⁻¹ bovine serum albumin in a final volume of 60 µl and incubated 20 min at 37°C then 10 min at 65°C. After purification with QIAquick PCR purification, a volume of 0.5 µl of pET22 (+) vector was added to 5.0 µl of the insert and incubated 30 min at room temperature.

Table 6: Primers used to clone the His-tag

| Gene | Primer sequence (5'->3') | His tag position |
|-------------|--|------------------|
| <i>ido</i> | <u>AAAGAAGGAGATAGGATCATG</u> CATCATCACCATCACCAT GAGGTTTTATAATGACGTT TGTTCTTAG <u>GTGTAATGGATAGT</u> GATCTTATTTGTCTCCTTATAAGAAAATGTTAC | N-terminal |
| <i>ido</i> | <u>AAAGAAGGAGATAGGATCATG</u> GAGGTTTTATAATGACGTTGTTCT GTGTAATGGATAGT <u>GATCTTA</u> ATGGTGATGGTATGATG TTTGTCCTTATAAGAAA ATGTTACTAATAAAATATC | C-terminal |
| <i>odo</i> | <u>AAAGAAGGAGATAGGATCATG</u> CATCATCACCATCACCAT CACCGCTTGGCCCTG GTGTAATGGATAGT <u>GATCTT</u> AGTAAATGACCCGGTC | N-terminal |
| <i>kdo1</i> | <u>AAAGAAGGAGATAGGATCATG</u> CATCATCACCATCACCAT AAGAACCTGTCTGCGTATG GTGTAATGGATAGT <u>GATCTT</u> AGCTGAACCTCGCAGAGAC | N-terminal |
| <i>kdo2</i> | <u>AAAGAAGGAGATAGGATCATG</u> CATCATCACCATCACCAT AGACCCTTAGACGTGACACC GTGTAATGGATAGT <u>GATCTTAAAGGTTGCCAGGTGAG</u> | N-terminal |
| <i>kdo3</i> | <u>AAAGAAGGAGATAGGATCATG</u> CATCATCACCATCACCAT AAATCACAAATCATTAAATT GTGTAATGGATAGT <u>GATCTTAAAGCCTGATCAAAACTTTCC</u> | N-terminal |
| <i>kdo4</i> | <u>AAAGAAGGAGATAGGATCATG</u> CATCATCACCATCACCAT GAAACCACATCATCGAACATC GTGTAATGGATAGT <u>GATCTTACTGCTGAGAGTGGAACAG</u> | N-terminal |
| <i>kdo5</i> | <u>AAAGAAGGAGATAGGATCATG</u> CATCATCACCATCACCAT AAATCTCAGTCTATCATGTCT GTTG GTGTAATGGATAGT <u>GATCTTACTGCTCAGGTCGAAGATTTAC</u> | N-terminal |

pET22 analogous sequence are underlined and His tags are in **bold blue**.

II.3.1.2 Chemical transformation

A volume of 5 µl of plasmids previously prepared was added to 20 µl of chemically competent BL21 DE3 codon+ cells and incubated for 30 min on ice. Cells were then incubated for 1 min at 42°C for heat shock and put back on the ice for 2 min. Cells were then resuspended in 100 µl of LB medium and incubated for 1 h at 37°C. Culture was spread on LB amp 0.1 mg.ml⁻¹ and incubated at 37°C overnight. A volume of 10 ml LB amp was inoculated with single colonies and incubated at 37°C overnight.

II.3.1.3 Protein overexpression

Flasks of 500 ml of TB amp 0.1 mg.ml⁻¹ were inoculated to obtain a starting OD600 of 0.1 and incubated at 37°C. When OD600 reached 2.0, IPTG was added to the culture at a final concentration of 1 mM and incubated at 20°C overnight.

The protein overproduction was verified by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). A 10 µl sample of each culture was mixed with 4 µl of loading buffer, 1 µl of SDS and heated 10 min at 90°C. The samples were then loaded on a 12% acrylamide gel in Nu PAGE MOPS Buffer and migrated for 1 h at 200 V. Gels were stained overnight in Coomassie blue (Simply Blue Safe Stain, Novex).

The rest of the protein was centrifuged 30 min at 4°C, 10 000 g. Supernatant was discarded and pellets stored at -80°C.

II.3.1.4 Cell lysis

II.3.1.4.1 For KDO1 to KDO5

Pellets were resuspended in 60 ml lysis buffer (50 mM Phosphate (Na/K) pH 8.0, 500 mM NaCl, 10 mM imidazole, 15% glycerol, 1 mM DTT, 1 mM Pefabloc®) and incubated with the lysis buffer 30 min at room temperature with agitation. Samples were treated by sonication cycles (1 cycle = 3 x 30 sec ultrasounds at 70%) and centrifuged 30 min at 4°C, 10 000 g. Lysate concentrations were estimated with a Bradford protein assay, and concentrations were adjusted to 3 mg.ml⁻¹.

II.3.1.4.2 For IDO

Pellets were resuspended in 32 ml lysis buffer (50 mM Phosphate (Na/K) pH 8.0, 500 mM NaCl, 30 mM imidazole, 30% glycerol, 1 X BugBuster (Millipore, Ref: 70921), 1 mM DTT, 1 mM Pefabloc®) and incubated with the lysis buffer 30 min at room temperature with agitation. Samples were centrifuged 30 min at 4°C, 20 000 g. Lysate concentrations were estimated with a Bradford protein assay, and concentrations were adjusted to 3 mg.ml⁻¹. Proteins were filtered on a 0.22 µm membrane (Millipore, Ref: 1609203).

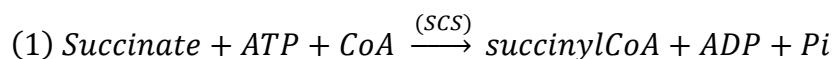
II.3.1.5 Protein purification

Fast protein liquid chromatography (FPLC) was performed with a Ni-NTA agarose nickel-charged affinity resin specifically used for purification of proteins containing a polyhistidine sequence. Ni-NTA column was equilibrated with 10 ml lysis buffer without DTT and then 10 ml lysis buffer with 1 mM DTT. The volume of lysate loading was 0.5

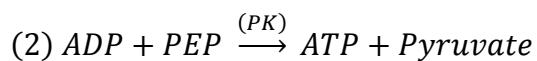
$\text{ml} \cdot \text{min}^{-1}$, for cleaning was $5 \text{ ml} \cdot \text{min}^{-1}$ and for elution was $2 \text{ ml} \cdot \text{min}^{-1}$ with a gradient from 0% to 100% of elution buffer (50 mM Phosphate (Na/K) pH 8.0, 500 mM NaCl, 250 mM imidazole, 15% glycerol, 1 mM DTT). Pools of purified proteins were desalted on a Hi Prep 26/10 Desalting column (Ref: 17508701). Columns were equilibrated with 50 mM Tris pH 8.0, 50 mM NaCl, Glycerol 15%. Fractions of interest were pooled, and concentrations were estimated with a Bradford protein assay and samples were then loaded on a 12% acrylamide gel in Nu-PAGE MOPS Buffer and migrated for 1 h at 200 V for purification verification.

II.3.2 Succinic acid assay

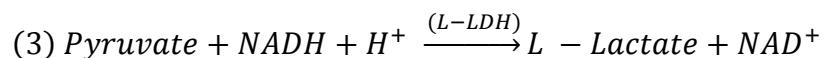
The succinic acid assay performed with the Megazyme kit (Ref: K-SUCC) function as follows. In the presence of adenosine-5'-triphosphate (ATP), succinic acid (succinate) is converted to succinyl-CoA by the enzyme succinyl-CoA synthetase (SCS), with the concurrent formation of adenosine-5'- diphosphate (ADP) and inorganic phosphate (P_i) (1).



In the presence of pyruvate kinase (PK), ADP reacts with phosphoenolpyruvate (PEP) to form pyruvate and ATP (2).



The pyruvate produced is reduced to L-lactate by L-lactate dehydrogenase (L-LDH) in the presence of reduced nicotinamideadenine dinucleotide (NADH), with the production of NAD^+ (3).



The amount of NAD^+ formed in the above-coupled reactions is stoichiometric with the amount of succinic acid in the sample. The NADH consumption is measured by the decrease in absorbance at 340 nm.

For the α -KG activity assay, the enzymatic mix contained 50 mM HEPES Buffer (pH 7.5), NADH (kit), ATP/PEP/CoA (kit), 1 mM enzyme's specific substrate, 0.5 mM FeSO₄, PK/L-LDH (kit), SCS (kit) and successive dilution of α -KG starting from 2.5 mM to 0 mM. The reaction was started by adding the enzyme to a final concentration of 0.005 $\mu\text{g}\cdot\mu\text{l}^{-1}$ and the decrease in absorbance at 340 nm followed at room temperature in real time.

For the specific substrate assay, the enzymatic mix contained 50 mM HEPES Buffer (pH 7.5), NADH (kit), ATP/PEP/CoA (kit), 0.5 mM α -KG, 0.5 mM FeSO₄, PK/L-LDH (kit), SCS (kit) and successive dilution of specific substrate starting from 2.5 mM to 0 mM. The reaction was started by adding the enzyme to a final concentration of 0.005 $\mu\text{g}\cdot\mu\text{l}^{-1}$ and the decrease in absorbance at 340 nm followed at room temperature in real time.

Kinetic parameters were determined by plotting the initial velocities at different substrate concentrations. Determinations of V_m , K_M and k_{cat} were calculated with the SigmaPlot software by iterative curve fitting.

II.3.3 Kinetic parameters determined with mass spectrometry

II.3.3.1 Enzymatic assay

The Megazyme kit previously mentioned was not adequate for the characterization of Ido due to pH incompatibility. Thus the kinetic parameters were determined by mass spectrometry. Reactions were performed with various concentrations of substrate (10, 25, 50, 100, 175, 250, 375, 500 and 1000 μM) at different time points (0, 1, 2 and 5 min).

For the α -KG activity assay, the enzymatic mix contained 50 mM MES Buffer (pH 6), 0.5 mM ascorbate, 0.5 mM FeSO₄, 1 mM isoleucine, and various concentrations (10, 25, 50, 100, 175, 250, 375, 500 and 1000 μM) of α -KG in a final volume of 100 μl .

For the isoleucine activity assay, the enzymatic mix contained 50 mM MES Buffer (pH 6), 0.5 mM ascorbate, 0.5 mM FeSO₄, 0.5 mM α -KG, and various concentrations (10, 25, 50, 100, 175, 250, 375, 500 and 1000 μM) of isoleucine in a final volume of 100 μl .

Enzymatic reactions were triggered by the addition of Ido at a final concentration of 0.04 $\mu\text{g}\cdot\mu\text{l}^{-1}$ at room temperature. The reactions were stopped by adding 1 μl of pure

trifluoroacetic acid (Sigma, Ref: T6508) after 0, 1, 2 and 5 min of incubation. The pH was adjusted with the addition of 1.3 µl of K₂CO₃ (5 M) and mixes were filtered on a 0.22 µm membrane (Milex, Ref: SLGV004SL). Samples were diluted 1:4 in mobile phase (80% acetonitrile and 20% ammonium formate).

II.3.3.2 LC/MS Method for 4-hydroxy-isoleucine and hydroxy-norvaline

The detection of 4-hydroxy-isoleucine (4-OH-ile) and hydroxy-norvaline (OH-norv) was performed using Thermo Scientific™ Dionex™ UltiMate™ 3000 Thermostatted Column Compartment Rapid Separation (TCC-3000RS) LC-MS system (Thermo Fisher Scientific Courtaboeuf, France) coupled to QTRAP 5500 hybrid mass spectrometer from ABSciex (Toronto, Canada) equipped with electrospray ionization (ESI) source.

II.3.3.2.1 HPLC conditions

HPLC separations were performed on a Sequant ZICpHILIC columns 5 µm, 2.1x50 mm (Merck, Darmstadt, Germany) thermostated at 40°C. The flow rate was 200 µl·min⁻¹. Phase A was aqueous solution of 100 mM of ammonium formate at pH 3 and phase B was acetonitrile. The following gradient was used: 0.5 min isocratic step at 80% of solvent B, linear gradient from 80% to 40% of solvent B for 6 min followed by isocratic elution at the achieved composition (40% of solvent B) for 1.5 min, return to the initial composition (80% of solvent B) in 1.5 min, and finally, reconditioning under these conditions for 4.3 min.

II.3.3.2.2 Mass spectrometry

The QTRAP mass spectrometer was operated in positive ionization mode. Mass spectrometry analyses on a QTRAP hybrid instrument were performed using multiple reaction monitoring (MRM) mode with the following in source parameters: ion source (IS) 5.5 kV, curtain gas (CUR) 20 a.u., temperature (TEM) 500°C, gas 1 (GS1) 45 a.u., gas 2 (GS2) 60 a.u., CAD – medium.

Extended method parameters and transitions for 4-OH-Ile and OH-norv are specified in Table 7.

Table 7: Extended MRM conditions for the detection of 4-OH-Ile and OH-norv (positive mode)

| | Q1 m/z | Q3 m/z | DP (Volts) | CE (Volts) | CXP (Volts) |
|----------|--------|--------|------------|------------|-------------|
| 4-OH-Ile | 148 | 74 | 41 | 19 | 14 |
| | 148 | 84 | 41 | 19 | 10 |
| | 148 | 102 | 41 | 13 | 6 |
| OH-Norv | 134 | 70 | 41 | 19 | 10 |
| | 134 | 74 | 41 | 19 | 10 |
| | 134 | 88 | 41 | 13 | 6 |

(DP: Declustering Potential, CE: collision energy, CXP: Collision Cell Exit Potential)

The data analysis was performed using Analyst software (ABSciex).

The calibration curve for 4-OH-Ile quantitation was performed with various dilutions of a commercial preparation of 4-OH-Ile (Sigma, Ref: 50118). However, OH-norv is not commercially available. The calibration curve for OH-norv was conducted using an enzymatic production in the presence of isoleucine and α-KG. OH-norv concentration was then estimated quantifying succinate using the Megazyme kit as described in section II.3.2.

II.4 Bioscreen C[®]

To analyze bacterial growth, a Bioscreen C[®] apparatus was used. It consists of a thermostatic incubator chamber, a growth monitoring device (OD reader) and an automated growth curve analyzer (Figure 21).

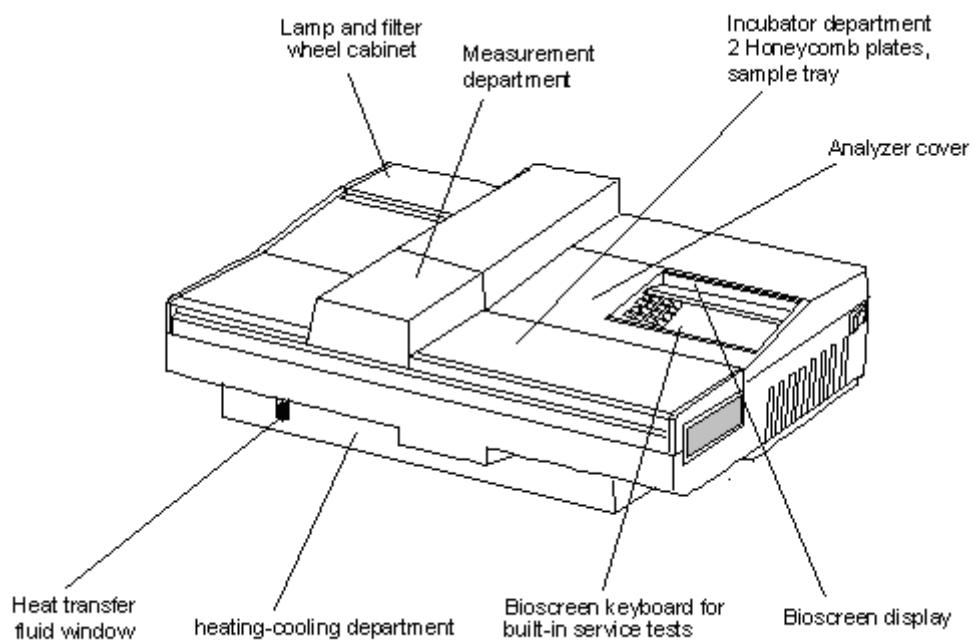


Figure 21: Bioscreen C[®] Analyzer Components

The strains to be tested and controls were grown overnight at 37°C, 150 RPM. The cultures were adjusted to OD₆₀₀ = 1.0 in a neutral medium (MS or MA) and then diluted at 1:200 in the test media. Final volumes of 200 µl of various cell suspensions were distributed in the honeycomb plate in triplicates.

Settings in the EZ Experiment software during the run were:

- Temperature: 37°C
- Shaking: continuous, high, normal
- Stop shaking: 5 seconds before measurement
- Measurement interval: 15 min
- Time of the experiment: 48 h
- Filter: 600 nm

II.5 Metabolomics analysis

II.5.1 Metabolome preparation

The media to be tested were inoculated with bacteria from over-night cultures at a starting OD600 of 0.2 and incubated at 37°C, 150 RPM, until OD600 = 0.7. The cultures were centrifuged 10 min at 4°C, 4000 RPM and the cell pellets were resuspended in 200 µl of extract solution (H₂O + methanol + chloroform 1 : 3 : 1). Bacteria were lysed by 3 freeze/thaw cycles by freezing in -80°C ethanol + dry ice bath for 10 min and then vortexed at room temperature for 30 sec. Samples were incubated 1 h at 4°C, 150 RPM then centrifuged 3 min at 12000 rpm - 4°C. The supernatants were transferred into new tubes. Solvent was evaporated using a Speed Vacuum overnight. Lyophilizates were resuspended in 200 µl of mobile phase (80% acetonitrile and 20% methanol).

The detection of (non-labeled and labeled by ¹³C) succinate in metabolomes was carried out using Thermo Scientific™ Dionex™ UltiMate™ 3000 Thermostatted Column Compartment Rapid Separation (TCC-3000RS) LC-MS system (Thermo Fisher Scientific Courtaboeuf, France) coupled to ultra-high resolution Orbitrap Elite hybrid mass spectrometer (Thermo Fisher Scientific Courtaboeuf, France) equipped with electrospray ionization (ESI) source.

II.5.2 HPLC conditions

HPLC separation was performed on a Sequant ZICpHILIC column 5 µm, 4.6 × 150 mm (Merck, Darmstadt, Germany) at 40° C with a mobile phase gradient flow rate of 0.5 ml/min. Aqueous solution of 10mM (NH₄)₂CO₃ (at a pH of 9.9 adjusted by NH₄OH) was used as phase A and acetonitrile was used as solvent B. Elution was conducted using following gradient conditions: 2 min isocratic step at 80% of phase B; 20 min linear gradient from 80 to 40% of phase B; keeping constant this ration for 8 min, return to 80% of phase B in 15 min and a reconditioning step under these conditions of 15 min.

II.5.2.1.1 Mass spectrometry

The Orbitrap Elite mass spectrometer was operated in the ESI negative ion mode. For LC-MS and MS/MS (using collision induces dissociation, CID) experiments the resolving power was set at 60000 m/Δm (full width at half maximum, FWHM at m/z=400). Ion

spray (IS) was set at -4.0 kV and capillary temperature at 275°C. Sheath gas and auxiliary gas flow rates were set at 60 and 50 arbitrary units (a. u.), respectively. Mass spectra were acquired over an *m/z* range from *m/z* 50 up to *m/z* 1000.

Raw data were analyzed using the Qual-browser module of Xcalibur version 2.2 (Thermo Fisher Scientific, Courtaboeuf, France).

The percentage of ¹³C incorporation in succinate (as metabolite) was calculated using the ratios of intensities of monoisotopic and first isotopic (involving one ¹³C atom) peaks from mass spectra and CID spectra of standard (Sigma, Ref: 398055) and succinate from metabolome.

II.6 Quantitative PCR and quantitative reverse transcription PCR

Couples of primers were designed to best fit the qPCR standard criteria:

- Target amplified region of exactly 100 bp
- GC content around 50-60%
- Primer length of 20 bp
- Tm difference less than 1°C
- Max 3' self-complementary of 1
- Max poly-x of 3

A couple of primers for each gene of interest was designed and a couple of primers for a conserved region of the 16S as a reference for normalization (Table 8).

Table 8: Primers used in this study for qRT-PCR

| Primer number | Gene | Function | Sequence |
|---------------|-------------|---------------------------|------------------------|
| 5355 | <i>kdo2</i> | Forward primer for 100 bp | GCAGGAAACCGACCATGAA |
| 5356 | <i>kdo2</i> | Reverse primer for 100 bp | TCGCTGCGCGAAATACTC |
| 5357 | 16s | Forward primer for 100 bp | CCCTTACGACCAGGGCTACA |
| 5358 | 16s | Reverse primer for 100 bp | ATCCGGACTACGACCGCACTT |
| 5376 | <i>ido</i> | Forward primer for 100 bp | ACTATGACGATGGCGGGAAA |
| 5377 | <i>ido</i> | Reverse primer for 100 bp | AATTCACTATCGAAACGCACGA |
| 5425 | <i>thyA</i> | Forward primer for 100 bp | GCGCCGCATTATTGTTC |
| 5426 | <i>thyA</i> | Reverse primer for 100 bp | TTTGCCGTCTGCCACATAG |

II.6.1 mRNA Stabilization

RNA was extracted from a log phase culture. The media to be tested were inoculated at a starting OD600 of 0.2 and incubated at 37°C, 150RPM, until OD600 = 0.7. For each sample, 2 volumes of RNA Protect Bacteria Reagent (Qiagen, Ref: 76506) were added to 1 volume of bacterial culture, vortexed and incubated 5 min at room temperature. Samples were centrifuged 10 min at 4°C, 7 000 RPM. The supernatant is discarded, and samples can be stored at -70°C for up to 4 weeks.

II.6.2 mRNA Extraction

The RNA extraction was performed using the RNeasy® Protect Bacteria Mini Kit (Qiagen, Ref: 74524) with the QIAcube. The QIAcube is a fully automated robot for purification of DNA, RNA or proteins. The full program name installed on the QIAcube to execute this protocol was "*Purification of total RNA from up to 5×10⁸ bacteria with enzymatic lysis*". TE buffer RNase-Free pH8.0 (Sigma-Aldrich, Ref: AM9849) + 1 mg.ml⁻¹ lysozyme (Sigma-Aldrich, Ref: L4919) and RLT buffer (RNeasy® Protect Bacteria Mini Kit) + 0.04 M DTT (ThermoFisher, Ref: P2325) were always prepared the day of the extraction. Reagents were positioned as described in the Qiabube protocol sheet.

II.6.3 DNA Degradation

RNA concentration was measured with Nanodrop and adjusted to 100 ng.µl⁻¹ with 5 µl 10 X DNase buffer (ThermoFisher, Ref: EN0525), 5 µl DNase I (ThermoFisher, Ref: EN0525) and RNase free water up to a final volume of 50 µl and incubated 30 min at 37°C. DNase was then degraded by a thermic treatment (10 min at 65°C) in the presence of EDTA (ThermoFisher, Ref: EN0525) at a final concentration of 5 mM for the stabilization of RNAs.

II.6.4 Reverse transcription

The High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Ref: 4368814) was used to perform reverse transcription.

Master Mix:

- $V_{RTBuffer} = 2\mu\text{l}$
- $V_{dNTP} = 0.8\mu\text{l}$
- $V_{RTRandomPrimers} = 2\mu\text{l}$
- $V_{Nucleasefreewater} = 4.2\mu\text{l}$
- $V_{MultiScribeRT} = 1\mu\text{l}$
- $V_{RNA} = 10\mu\text{l}$

Thermal cycler program:

$$\left\{ \begin{array}{l} \text{Step 1: } 25^\circ\text{C for 10min} \\ \text{Step 2: } 37^\circ\text{C for 120min} \\ \text{Step 3: } 85^\circ\text{C for 5min} \\ \text{Step 4: } 4^\circ\text{C for } \infty \end{array} \right.$$

II.6.5 Quantitative PCR (qPCR)

For qPCR, the KAPA SYBR® FAST qPCR Kits (Kapa Biosystems, Ref: KK4620) was used. cDNA concentration was measured with Nanodrop and adjusted to $10 \text{ ng}\cdot\mu\text{l}^{-1}$. Each sample was measured in triplicates where $2 \mu\text{l}$ of cDNA sample was added to $18 \mu\text{l}$ of Master Mix for each well in the plate. In the case of a qPCR from bacterial suspension, $5.3 \mu\text{l}$ of culture adjusted at $\text{OD}_{600} = 0.1$ was used per reaction. The volume was calculated to approximately reach 20 ng of DNA per reaction.

The plate was put in a Quantitative PCR thermocycler Stratagene MX 3000 and the software used to follow the measurements in real time was MXPro (Agilent). For the analysis of relative gene expression, the $2^{-\Delta\Delta CT}$ method was used(Livak and Schmittgen, 2001).

Master Mix:

- $V_{Kapa} = 10\mu\text{l}$
- $V_{FwdPrimer} = 0.4\mu\text{l}$
- $V_{RevPrimer} = 0.4\mu\text{l}$
- $V_{water} = 7.2\mu\text{l}$

Thermal cycler program:

- i. Step Enzyme activation: $95^\circ\text{C for 3 min}$
- ii. $30 \text{ cycles} = \left\{ \begin{array}{l} \text{Step Denature: } 95^\circ\text{C for 1 - 3 sec} \\ \text{Step Anneal/Extend : Primer TM for 30 sec} \end{array} \right.$
- iii. $\text{Step Dissociation} = \left\{ \begin{array}{l} 95^\circ\text{C for 1 min} \\ \text{Primer Tm for 30 sec} \\ 95^\circ\text{C for 30 sec} \end{array} \right.$

II.7 Cell Sorting and Sequencing

Cultured bacteria were centrifuged (10000 rpm; 10 min) and then the pellet was resuspended in PBS in order to obtain a cell count of 106 cells.ml⁻¹. For the cell sorting a MoFlo Astrios (Beckman Coulter) machine was used equipped with a 70 nm nozzle. The cells were collected in a 384 well flat bottom micro plates where in each well 2 µl of PBS were deposited in order to avoid the adherence of the cells to the bottom of the wells. For cell detection, we used the 488 nm laser and the plots of forward scatter (FSC) and side scatter (SSC) were used to identify single cells (doublet discrimination gating was used to exclude the doublets from sorting).

For the single cell genome amplification, the Repli-g Single Cell Kit (Qiagen) was used following the proposed protocol of Qiagen. This kit uses isothermal genome amplification termed Multiple Displacement Amplification (MDA), which involves the binding of random hexamers to denatured DNA, followed by strand displacement synthesis at constant temperature using the φ29 polymerase.

For the bacterial lyses and the DNA denaturation, alkaline lysis was used at 65°C for 10 min, and the solution was neutralized. For amplification the master mix and the enzyme of the kit were added and then incubated overnight at 30°C. The amplification reaction was stopped at 65°C for 3 minutes. Generally at the end of the amplification procedure the yield was 400-450 ng/µl in 25 µl of total volume.

Amplified genomic DNA was sonicated using the E210 Covaris instrument (Covaris, Woburn, MA). The fragments were end-repaired, then 3'adenylated and Illumina compatible adaptors (Bioo Scientific, Austin, TX) were added by using NEBNext Sample Reagent Set (New England Biolabs, Ipswich, MA). Ligation products were purified by AMPure XP (Beckman Coulter Genomics, Danvers, MA). All these steps were performed using a FXP robot (Beckman Coulter Genomics, Danvers, MA). DNA fragments were then PCR-amplified using Illumina adapter-specific primers and one of the two enzymes: Platinum Pfx DNA polymerase (Invitrogen, Carlsbad, CA) or Kapa Hifi DNA polymerase (Kapa Biosystems, Wilmington, MA). Libraries were purified with AMPure XP beads by using a FXP robot (Beckman Coulter Genomics, Danvers, MA).

They were quantified by qPCR using the KAPA Library Quantification Kit for Illumina Libraries (Kapa Biosystems, Wilmington, MA). The library profiles were assessed using the DNA High Sensitivity LabChip kit on a Bioanalyzer (Agilent Technologies, Santa Clara, CA). Libraries were sequenced on HiSeq 2000 instrument (Illumina, San Diego, CA) using 100 base-length read v3 chemistry in a paired-end mode.

III RESULTS

III.1 Fixation of beneficial mutations in a gene under selection in an evolving cell population

A central question for an optimized experimental setup aiming at a targeted *in vivo* evolution of an enzymatic activity concerns the expression format of the gene specifying the enzyme to be evolved. The influence of different expression formats on the appearance and the fixation of adaptive mutations in cell populations were studied. A strategy based on the reversion of a point mutation introduced into a gene of *E. coli* coding for an essential enzymatic activity was used. For this purpose, a genetic selection screen had to be constructed.

III.1.1 Genetic screen construction

Genetic selection screens are strains lacking essential genes for growth and are auxotrophic for one or several compounds needed for proliferation. We choose to mutate the gene *metA* of *E. coli*, coding for L-homoserine O-succinyltransferase. This enzyme catalyzes the synthesis of O-succinyl-homoserine from L-homoserine and succinyl-CoA (Figure 22), an essential step of the methionine biosynthetic pathway. A number of catalytic residues have been identified, and the impact of mutations at these positions on the MetA activity was studied (Ziegler *et al.*, 2007; Zubieta *et al.*, 2008).

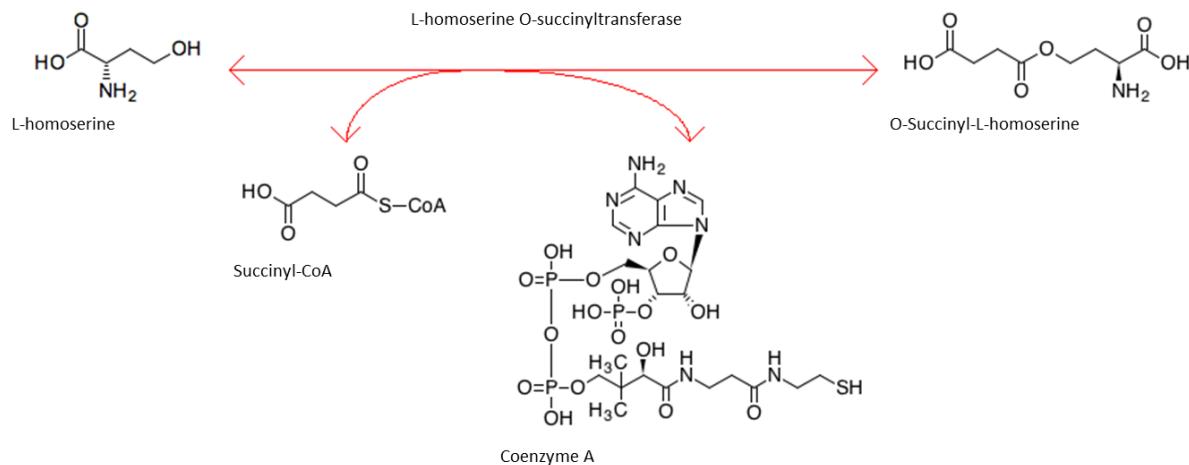


Figure 22: Reaction catalyzed by the L-homoserine O-succinyltransferase

The selection strain G3724 (*ΔmetA::kan+*) was constructed by replacing the *metA* gene by a kanamycin resistance cassette. The strain construction involved the preparation of a stock of P1 phages from the strain JW3973 contained in the Keio collection of *E. coli*

single-gene knockout mutants (Baba *et al.*, 2006; Yamamoto *et al.*, 2009). The P1 preparation was used to transduce the *E. coli* K12 strain MG1655 according to the protocol described in section II.2.1. Phenotypic tests confirmed methionine auxotrophy for strain G3724.

III.1.2 Variants construction

Point mutations were introduced into the gene *metA* changing catalytically important residues. These residues have previously been identified through *in vitro* activity tests (Ziegler *et al.*, 2007; Zubieta *et al.*, 2008). Since the *in vivo* enzyme activity of a mutated variant cannot easily be extrapolated from *in vitro* activity data, a large set of point mutations was constructed and tested for *in vivo* activity. According to the published results, the mutations can be separated into two groups:

- Mutations leading to enzyme inactivation (labeled “Essential”)
- Mutations leading to a reduced enzyme activity (labeled “Reduced activity”)

The list of selected variants is given in Table 9.

Table 9: Mutations introduced in the gene *metA*

| Amino acid | Mutation | Codon | Group | Mutation type |
|-----------------|----------|--------------------------------|------------------|--|
| C142 | C142R | UGC to CGC | Essential | Transition C to U |
| | C142Y | UGC to UAC | | Transition A to G |
| H235 | H235Y | CAU to UAU | Essential | Transition U to C |
| | H235R | CAU to CGU | | Transition G to A |
| | H235S | CAU to AGU | | Transversion A to C Transition G to A |
| E237 | E237D | GAA to GAU | Reduced activity | Transversion U to A |
| | E237A | GAA to GCA | | Transversion C to A |
| K46 | K46R | AAG to AGG | Reduced activity | Transition G to A |
| K47 | K47R | AAG to AGG | Reduced activity | Transition G to A |
| K46, K47 | K46RK47R | AAGAAG to AGGAGG | Reduced activity | Transition G to A |

Mutations are in **bold red**.

III.1.3 Expression formats and growth tests

Three replicons were chosen as expression formats for the gene *metA* WT and the mutated alleles listed in Table 9. The copy numbers of the replicons in the cell are 1, 5-15 and 50-70, respectively (Table 10).

Table 10: Expression formats used in this study

| Expression format | Derivation | Selection cassette | ORI | Copy number |
|-----------------------|------------|--------------------|------|-------------|
| Chromosomal insertion | N/A | N/A | N/A | 1 |
| pVDM18 | pSU18 | <i>cat</i> | P15A | 5-15 |
| pSP100 | pUC18 | <i>bla</i> | pMB1 | 50-70 |

The plasmid constructions were realized following classical ligation protocols of PCR amplified genes. For the one copy constructions, genes were inserted into the *E. coli* chromosome via allelic exchange at the *kdgK* locus coding for 2-keto-3-deoxygluconokinase, an enzyme implicated in the galacturonic acid degradation (Bouzon, M *et al.*, 2017). The inserted fragment comprised a T7 derived promoter enabling constitutive expression of the inserted gene. Details of the allelic exchange protocol and the tests which were conducted to verify the insertions are given in section II.2.2.

As shown in Table 11, the selection strains expressing the gene *metA* WT and the mutated alleles were tested for complementation at 37° as well as at 30°C. It is commonly seen that residual enzymatic activity can sustain cellular growth at conditions of attenuated proliferation rates. This was observed for the mutations at residue E237, *i.e.* the alleles E237A and E237D, for which complementation was only obtained at 30°C. This finding is in accordance with *in vitro* activity data: the E237D mutation was shown to reduce the activity of the enzyme by a factor of twelve (Zubieta *et al.*, 2008).

Table 11: Methionine auxotrophy complementation of strains expressing different *metA* alleles

| Allele | Expression format | Plasmid | Strain | Growth | |
|--------|-------------------|----------|--------|--------|------|
| | | | | 30°C | 37°C |
| WT | pSP100 | pGEN118 | G4008 | +++ | +++ |
| | pVDM18 | pGEN1084 | G4009 | +++ | +++ |
| | Chromosomal | none | G3997 | +++ | +++ |
| K46R | pSP100 | pGEN992 | | N/A | N/A |
| | pVDM18 | pGEN1041 | G3841 | + | + |
| | Chromosomal | none | | N/A | N/A |
| K47R | pSP100 | pGEN993 | | N/A | N/A |
| | pVDM18 | pGEN1042 | G3842 | N/A | N/A |
| | Chromosomal | none | | N/A | N/A |
| C142Y | pSP100 | pGEN1001 | | N/A | N/A |
| | pVDM18 | pGEN1044 | G3844 | - | - |
| | Chromosomal | none | | N/A | N/A |
| C142R | pSP100 | pGEN1002 | G3748 | - | - |
| | pVDM18 | pGEN1045 | G3845 | N/A | N/A |
| | Chromosomal | none | | N/A | N/A |
| H235S | pSP100 | pGEN1005 | G3749 | - | - |
| | pVDM18 | pGEN1048 | G3848 | - | - |
| | Chromosomal | none | | N/A | N/A |
| E237D | pSP100 | pGEN1086 | G4011 | + | - |
| | pVDM18 | pGEN1085 | G4010 | + | - |
| | Chromosomal | none | G3880 | + | - |
| E237A | pSP100 | pGEN1007 | | N/A | N/A |
| | pVDM18 | pGEN1050 | G3850 | + | - |
| | Chromosomal | none | | N/A | N/A |

All transformed strains showed in this table are descendant of strain G3724 ($\Delta metA$). Complementation tests on solid medium were performed at 37°C for 48 h. The bacterial growth observed is indicated with signs for normal (+++), medium (++) residual (+), no complementation (-) and data not available (N/A).

The conditional phenotype of the E237 alleles was particularly suitable for our reversion strategy. It enabled the maintenance of the culture in a turbidostat regime at 30°C for both plasmidic and chromosomal expression. Since the mutated gene was indispensable for growth, no loss of plasmid had to be anticipated.

The reversion event and its spreading throughout the culture, which should result in a decrease in the generation time, can be monitored in the GM3. The number of generations necessary for the restoration of WT growth of the cell population due to reversion will – after verification of the reversion event by sequencing – be compared between the different expression setups.

The impact of the mutation E237D on the growth kinetics at 30°C and 37°C was further investigated and compared to the corresponding WT constructs. The strains used are listed in Table 12.

Table 12: Strains used for growth test

| Strain | Genotype | Expression format |
|--------|---|--------------------|
| G4011 | $\Delta metA$, pGEN1086 (pSP100:: <i>metA:E237D+</i>) | High copy number |
| G4008 | $\Delta metA$, pGEN118 (pSP100:: <i>metA+</i>) | High copy number |
| G4010 | $\Delta metA$, pGEN1085 (pVDM18:: <i>metA:E237D+</i>) | Low copy number |
| G4009 | $\Delta metA$, pGEN1084 (pVDM18:: <i>metA+</i>) | Low copy number |
| G3880 | $\Delta metA$, $\Delta kdgK::metA:E237D+$ | Single copy number |
| G3997 | $\Delta metA$, $\Delta kdgK::metA+$ | Single copy number |

Strains G4008, G4009, and G3997 (Table 12), which harbor the gene *metA* WT, were grown in mineral MS glc (0.2%) medium at 30°C and 37°C. When comparing expression formats at 37°C (Figure 23, triangles), there is a significant difference between the growth of strain G3997 (chromosomal format) and strains G4008 and G4009 (plasmid formats). As previously described in the literature (Bentley *et al.*, 1990; Bhattacharya and Dubey, 1995; Glick, 1995), extensive gene expression due to multicopy plasmids causes a significant reduction in the growth rate of cells, mainly due to metabolic burden. Reasons for this metabolic burden are various. The introduction of foreign DNA containing coding sequences in a host organism drains energy to maintain and express the recombinant protein. Furthermore, the metabolic load has been described to increase with plasmid copy number (Birnbaum and Bailey, 1991), promoter and ribosome binding site strength (Popov *et al.*, 2011).

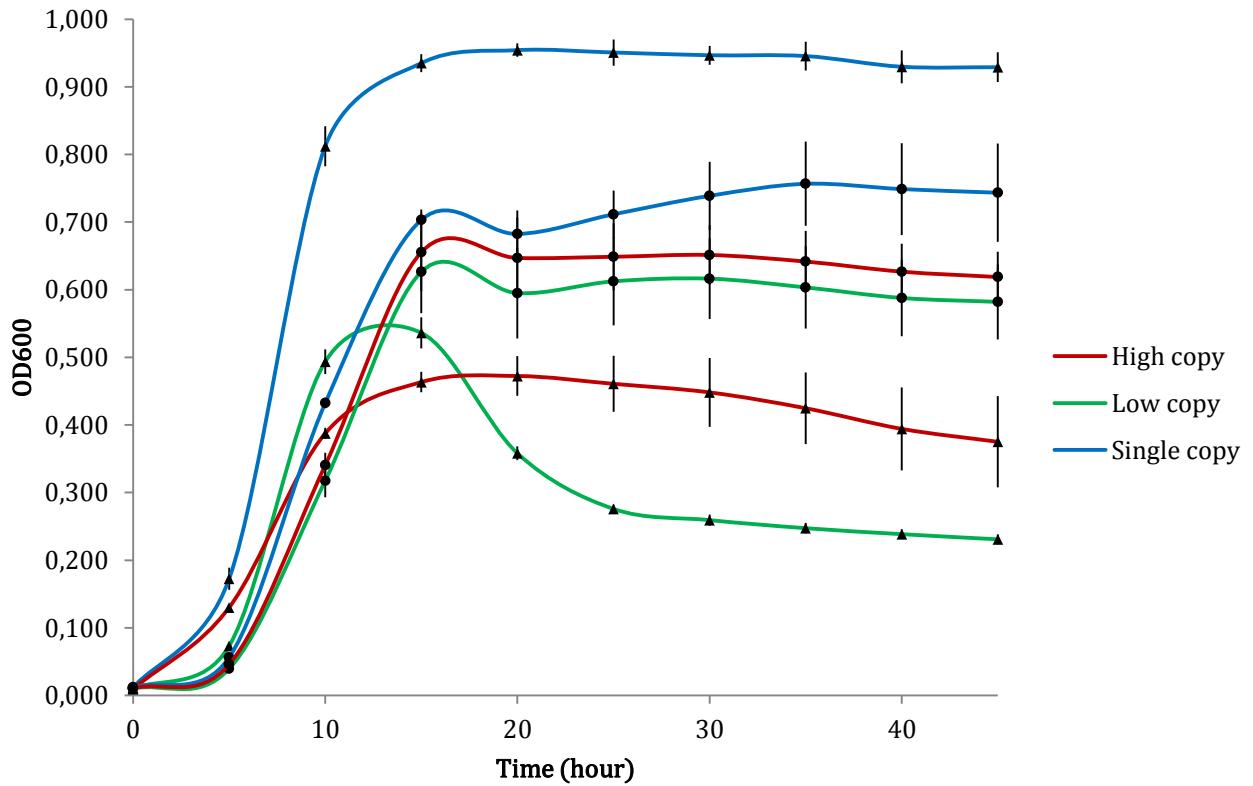


Figure 23: Growth of descendants of selection strain G3724 (*ΔmetA*) expressing the gene *metA* WT from different expression formats at 30°C and 37°C

Shown is the OD600 of cultures grown in MS glc (0.2%) at 30°C (circles) and 37°C (triangles) over time. Bars show means of triplicate measurements \pm sd.

Cultures of strain G4008 (*ΔmetA*, pGEN118 (pSP100::*metA*+)): red line.

Cultures of strain G4009 (*ΔmetA*, pGEN1084 (pVDM18::*metA*+)): green line.

Cultures of strain G3997 (*ΔmetA*, *ΔkdgK*::*metA*+): blue line.

By contrast at 30°C (Figure 23, circles), growth rates are very similar independently of the expression format. At lower temperature, *E. coli* metabolism is slowed down which probably leads to an attenuation of the plasmid metabolic load observed at 37°C.

As expected, when conducting the same experiment with strains expressing the allele *metA*:E237D (strains G4011, G4010, and G3880 in Table 12), the influence of the temperature on cellular growth was found to be even more pronounced. None of the strains grew at 37°C within 45h (Figure 24, triangles), whereas the three strains grew at 30°C (Figure 24, circles) with growth rates depending on their *metA*:E237D expression format.

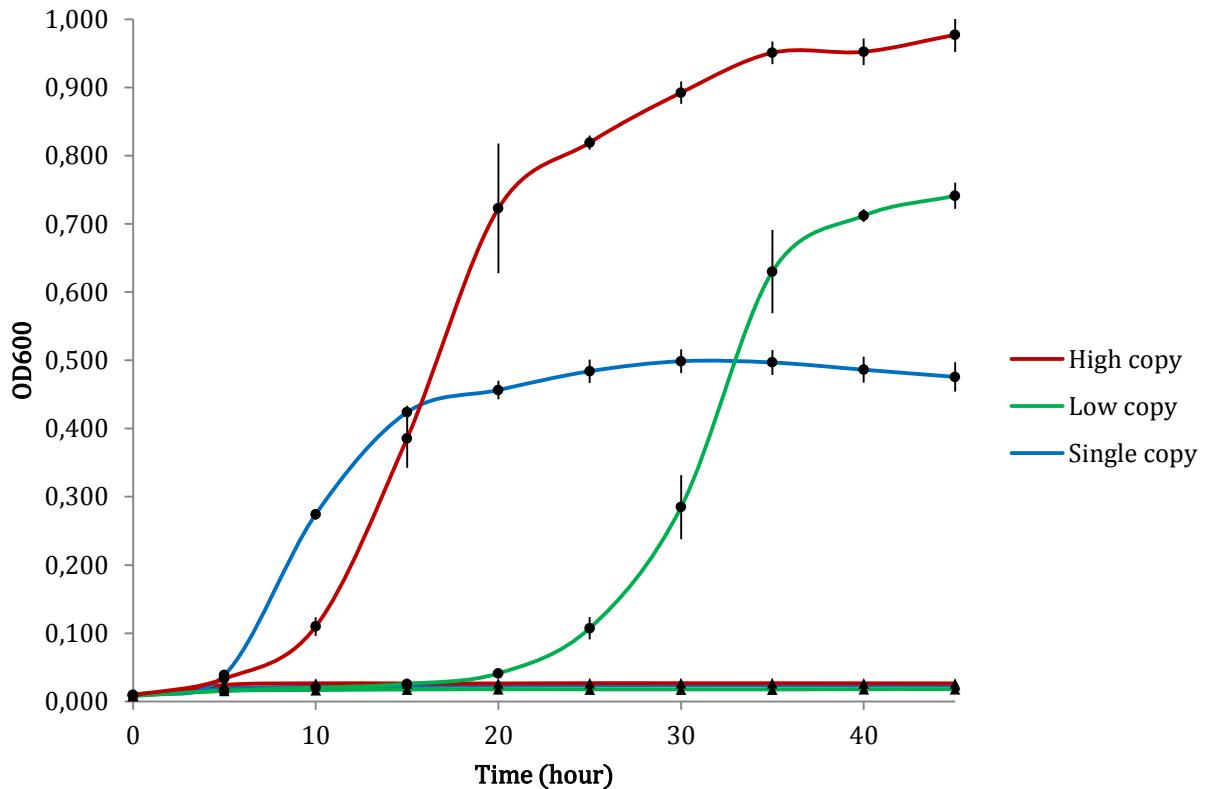


Figure 24: Growth of descendants of selection strain G3724 ($\Delta metA$) expressing the allele *metA:E237D* from different expression formats at 30°C and 37°C

Shown is the OD600 of cultures grown in MS glc (0.2%) at 30°C (circles) and 37°C (triangles) over time. Bars show means of triplicate measurements \pm sd.

Cultures of G4011 ($\Delta metA$, pGEN1086 (pSP100::*metA:E237D*+)): **red** line.

Cultures of G4010 ($\Delta metA$, pGEN1085 (pVDM18::*metA:E237D*+)): **green** line.

Cultures of G3880 ($\Delta metA$, $\Delta kdgK$::*metA:E237D*+): **blue** line.

These observations confirm that the temperature can be used to modulate the selective pressure. It also indicates that the turbidostat is applicable at 30°C as the culture regime for reversion studies of the *metA:E237D* mutation, but not at 37°C.

Unlike the results obtained for strains G4008, G4009 and G3997 at 30°C (Figure 23, circles), where growth rates were similar for gene *metA* WT expressed from different formats, growth rates for strains G4011, G4010 and G3880 were expression format dependent at 30°C (Figure 24, circles). The best growth for strains expressing the variant *metA:E237D* variant was obtained for the high-copy number expression format (Figure 24, circles, red line), showing that the accumulation of mutated enzymes exhibiting low activity can outweigh the growth burden caused by the high number of plasmid replicons present in the cell.

When comparing the growth at 30°C (Table 13) of the selection strain expressing the gene *metA* WT with the corresponding strain expressing the allele E237D, we observe an increase in the generation time of the latter of around 1 hour, except for the chromosomal expression format where generation times are comparable. Differences in growth rate between WT and mutant are a prerequisite for selection; if these differences are too small, a turbidostat regime might not be appropriate for the fixation of a given mutation.

Table 13: Generation time of descendants of selection strain G3724 ($\Delta metA$) expressing either *metA* WT or *metA:E237D* from different expression formats at 30°C

| Strain | Genotype | Expression format | Generation time at 30°C |
|--------|---|--------------------|-------------------------|
| G4011 | $\Delta metA$, pGEN1086 (pSP100:: <i>metA:E237D</i> +) | High copy number | 2h55 |
| G4008 | $\Delta metA$, pGEN118 (pSP100:: <i>metA</i> +) | High copy number | 2h06 |
| G4010 | $\Delta metA$, pGEN1085 (pVDM18:: <i>metA:E237D</i> +) | Low copy number | 3h11 |
| G4009 | $\Delta metA$, pGEN1084 (pVDM18:: <i>metA</i> +) | Low copy number | 2h00 |
| G3880 | $\Delta metA$, $\Delta kdgK::metA:E237D$ + | Single copy number | 2h15 |
| G3997 | $\Delta metA$, $\Delta kdgK::metA$ + | Single copy number | 2h02 |

III.1.4 Reversion dynamics in continuous culture

Each culture in the GM3 device was given a 3-letter code corresponding to the project name (in the present case MET) followed by an iterative number increasing each time a new culture is launched. All cultures mentioned in the following section are listed in Table 14.

Table 14: GM3 cultures launched to study the dynamics of genetic reversion

| GM3 culture name | Strain | Genotype | Regime |
|------------------|--------|---|-------------|
| MET 6 | G3880 | $\Delta metA$, $\Delta kdgK::metA:E237D$ + | Turbidostat |
| MET 7 | G4010 | $\Delta metA$, pGEN1085 (pVDM18:: <i>metA:E237D</i> +) | Turbidostat |
| MET 10 | G4011 | $\Delta metA$, pGEN1086 (pSP100:: <i>metA:E237D</i> +) | Turbidostat |
| MET 15 | G3880 | $\Delta metA$, $\Delta kdgK::metA:E237D$ + | Medium SWAP |
| MET 17 | G4010 | $\Delta metA$, pGEN1085 (pVDM18:: <i>metA:E237D</i> +) | Medium SWAP |
| MET 18 | G4011 | $\Delta metA$, pGEN1086 (pSP100:: <i>metA:E237D</i> +) | Medium SWAP |

III.1.4.1 Turbidostat regime

Strains G4011, G4010 and G3880 (respectively culture MET 10, MET 7 and Met 6 in Table 14) were inoculated into the culture vessel of GM3 devices and cultured in mineral salt glucose medium (MS glc 0.2%) at 30°C following a turbidostat regime. The change in generation time throughout the experiment was recorded (Figure 25).

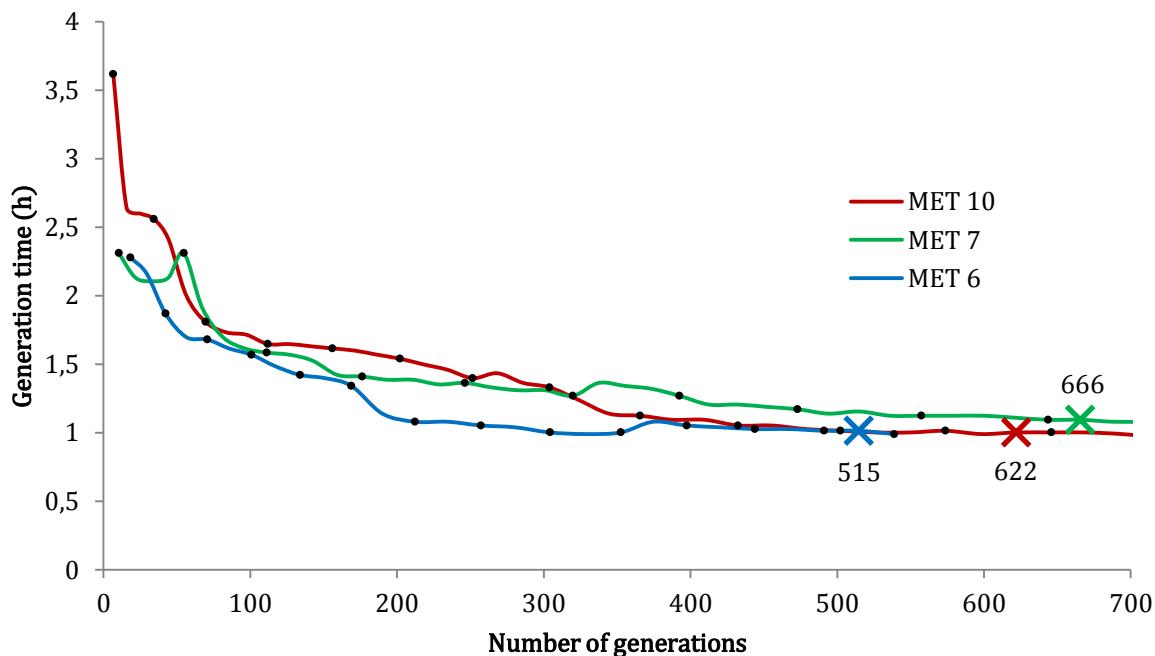


Figure 25: Evolution of cultures MET 6, MET 7 and MET 10 in turbidostat regime

Shown is the generation time of cultures in turbidostat regime grown in MS glc (0.2%) at 30°C with respect to the number of generations. Crosses indicate sampling points, the corresponding number of generations is indicated.

Culture MET 10 of strain G4011 ($\Delta metA$, pGEN1086 (pSP100::*metA*:E237D+)): red line.

Culture MET 7 of strain G4010 ($\Delta metA$, pGEN1085 (pVDM18::*metA*:E237D+)): green line.

Culture MET 6 of strain G3880 ($\Delta metA$, $\Delta kdgK$::*metA*:E237D+): blue line.

For all three cultures, the generation time diminished, without significant differences between the evolution experiments. Samples of the cultures were collected at 515 (culture MET 6), 622 (culture MET 10) and 666 (culture MET 7) generations, corresponding to a time point where the generation time seemed to have reached a plateau of around 1 h. The same plateau of around 1 h was observed for selection strains expressing the gene *metA* WT on the same expression formats grown in the same conditions (data not shown).

Isolates from samples were obtained on solid mineral salt glucose medium. The gene *metA* was sequenced for 3 isolates of each culture. In the case of culture MET 10, codon 237 GAT was mutated to GAG coding for glutamic acid in all three isolates. This mutation

is equivalent to a reversion event since the aspartic acid 237 changed to glutamic acid, which is the WT residue at position 237 of MetA. This observation confirms that the revertants appeared during the evolution experiment and not from a contamination of the original strain, which carries the codon 237 GAA. Despite the similar evolution of their generation time, no reversion events were detected for the isolates obtained from cultures MET 6 and MET 7.

To assess the fixation kinetics of the reversion of genes expressed on a high copy number plasmid in turbidostat, culture samples along the evolution timeline of culture MET 10 were analyzed. For each sample, 100 colonies were isolated on solid mineral glucose medium at 30°C (permissive growth conditions) and patched on solid mineral glucose medium grown at 37°C (nonpermissive growth conditions). Figure 26 shows the proportion of clones which grew at 37°C for each time point (blue line), together with the evolution of the generation time of the culture (red line).

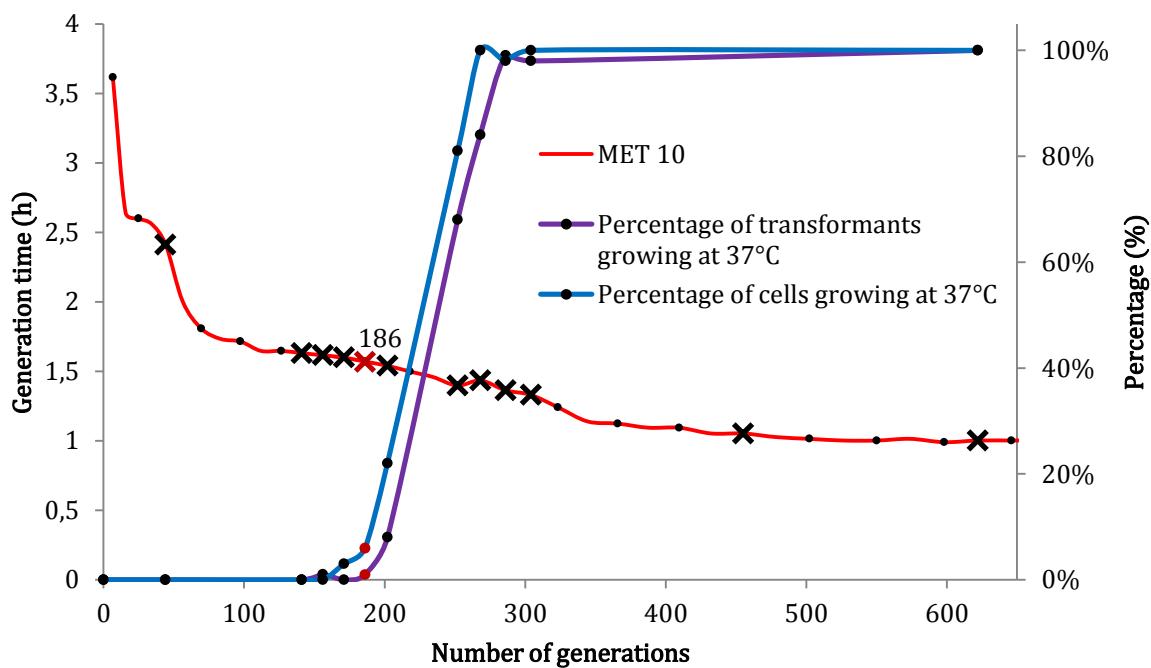


Figure 26: Analysis of the evolution of culture MET 10 in turbidostat regime at 30°C

Shown is the generation time of culture MET 10 of strain G4011 (*ΔmetA*, pGEN1086 (pSP100::*metA:E237D+*)) in turbidostat regime grown in MS glc (0.2%) at 30°C (red line, left axis). Crosses indicate sampling points. Red symbols show data corresponding to the sample at generation 186. Percentage of cells growing at 37°C (blue line, right axis) and percentage of transformants growing at 37°C (purple line, right axis) are also shown.

In addition, plasmid DNA was purified from each MET 10 sample analyzed and transformed into the selection strain G3724. For each transformation, 100 colonies were

tested for growth on solid mineral glucose medium at 37°C and the percentage of colonies which grew at that condition were recorded (Figure 26, purple line). Sequencing of a fraction of these colonies growing at 37°C (up to 10 for each time point) revealed 100% reversion of the *metA* gene (codon 237 GAT to GAG).

The following conclusions can be drawn from these data:

- Growth of cells at 37°C was related to the presence of a reversion of the gene *metA* carried on the plasmid.
- Despite the high copy number of plasmids in a cell, it only took around 100 generations from the appearance of the first cells exhibiting the mutated phenotype (at generation 170) to the complete fixation of the reverted phenotype (at generation 286).
- The lower number of positive transformants (Figure 26, purple line) with respect to the positive cells present in each sample (Figure 26, blue line) can be attributed to the fact that the cells in the evolving population growing at 37°C contain a mixture of plasmids, carrying either the original *metA* allele 237Asp or the reverted allele 237Glu. The question of plasmid populations within a cell will be dealt with in section III.1.5.

III.1.4.2 Medium SWAP regime

Strains G4011, G4010 and G3880 (respectively culture MET 18, MET 17 and Met 15 in Table 14) were inoculated into the culture vessel of GM3 devices in a medium SWAP regime (Figure 27).

Conditions of the medium SWAP regime for cultures MET 15 (strain G3880), MET 17 (strain G4010) and MET 18 (strain G4011) were as follows:

- Relaxing medium: MS glc (0.2%) met (0.05mM)
- Stressing medium: MS glc (0.2%)
- Generation time 2 h at 37°C

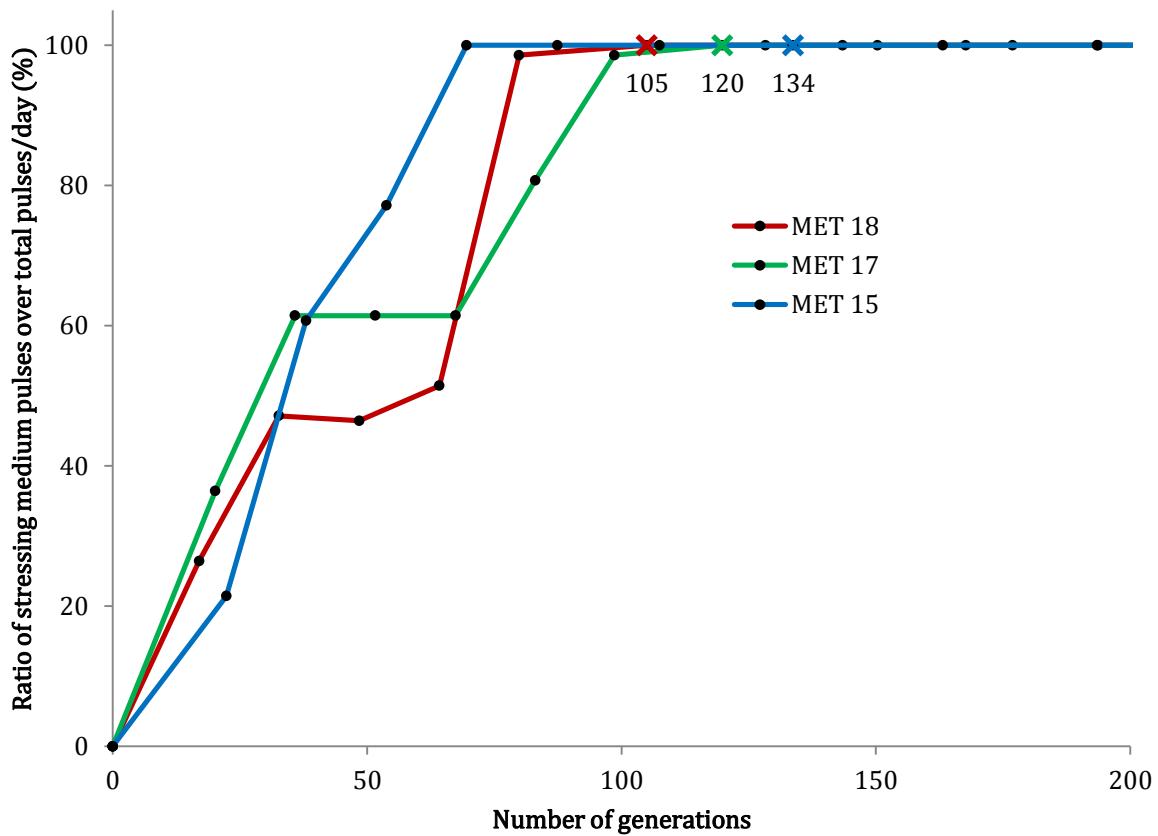


Figure 27: Evolution of cultures MET 15, MET 17 and MET 18 in medium SWAP regime

Shown is the ratio of stressing medium dilution pulses per day in medium SWAP regime. The relaxing medium was MS glc (0.2%) met (0.05 mM) and the stressing medium was MS glc (0.2%). The generation time was fixed at 2 h, and growth temperature at 37°C. Crosses indicate sampling points, the corresponding number of generations is indicated.

Culture **MET 18** of strain G4011 ($\Delta metA$, pGEN1086 (pSP100::metA:E237D+)): **red** line.

Culture **MET 17** of strain G4010 ($\Delta metA$, pGEN1085 (pVDM18::metA:E237D+)): **green** line.

Culture **MET 15** of strain G3880 ($\Delta metA$, $\Delta kdgK$::metA:E237D+): **blue** line.

Cultures MET 15, MET 18 and MET 17 were diluted with 100% of stressing medium at growth points corresponding to 70, 80 and 99 generations respectively, signifying the growth of the bacterial population without methionine. After reaching this point, all cultures were switched to a turbidostat regime in MS glc (0.2%) at 37°C. Generation times were then stable at around 50 min for all cultures (data not shown). Samples were collected at generation 105 (culture MET 18), 120 (culture MET 17) and 134 (culture MET 15), colonies were isolated and sequenced as described in the previous section. Unlike for the turbidostat adaptation, codon 237 was reverted to GAG for all *metA* genes sequenced, regardless of the expression format. This suggests that the medium SWAP regime exerted a higher selection pressure in this particular setup, certainly due to the fact that the cultures were operated at 37°C. As seen previously, the temperature

modulates the growth rate. At 37°C, the optimum growth temperature for *E. coli*, the metabolism works at its fastest speed. When one enzyme is deficient, *i.e.* the homoserine O-succinyltransferase MetA, it creates a metabolic bottleneck and displaces fluxes. At 30°C, the metabolism is slowed down, diminishing the stress around the metabolic bottleneck. In consequence, the selective pressure in the turbidostat adaptation might be attenuated. As in the turbidostat adaptation, the reversion event selected at codon 237, in medium SWAP regime, is a T to G transversion in the three evolution experiments. This transversion type seems favored over the T to A transversion, in line with the rates of spontaneous mutations determined in *E. coli* as published by Lee *et al.*, 2012.

To assess the fixation kinetics of the reversion in the medium SWAP regime, culture samples along the evolution timeline of cultures MET 15 and MET 18 were analyzed. As described previously for the analysis of culture MET 10 (Figure 26), 100 colonies were isolated for each sample at 30°C and patched on solid mineral glucose medium grown at 37°C. The proportion of clones which grew at 37°C for each time point, together with the evolution of the culture is shown in Figure 28 for culture MET 15 and Figure 29 for culture MET 18.

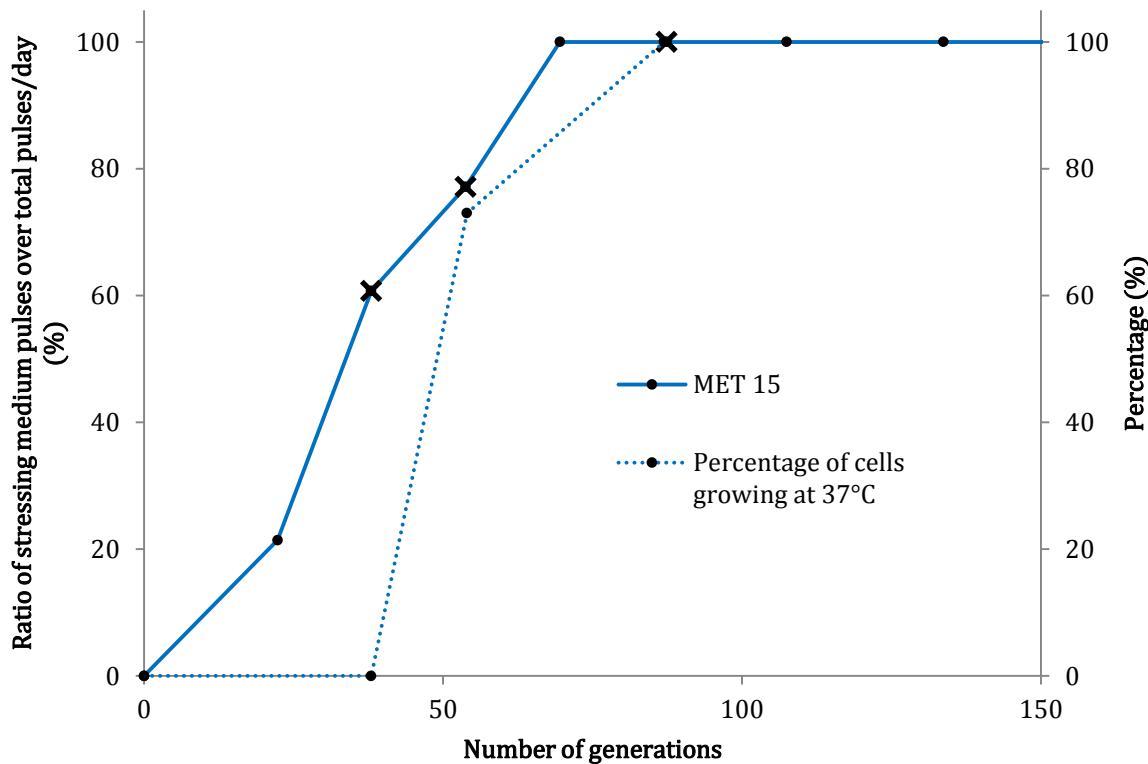


Figure 28: Analysis of the evolution of culture MET 15 in medium SWAP regime at 37°C

Shown is the ratio of stressing medium dilution pulses per day (left axis) in medium SWAP regime. The relaxing medium was MS glc (0.2%) met (0.05 mM) and the stressing medium was MS glc (0.2%). The generation time was fixed at 2 h, and growth temperature at 37°C. Crosses indicate sampling points. Culture MET 15 of strain G3880 (*ΔmetA, ΔkdgK::metA:E237D+*): blue line.

Percentage of cells growing at 37°C for samples from culture MET 15 (dotted blue line) is shown in right axis.

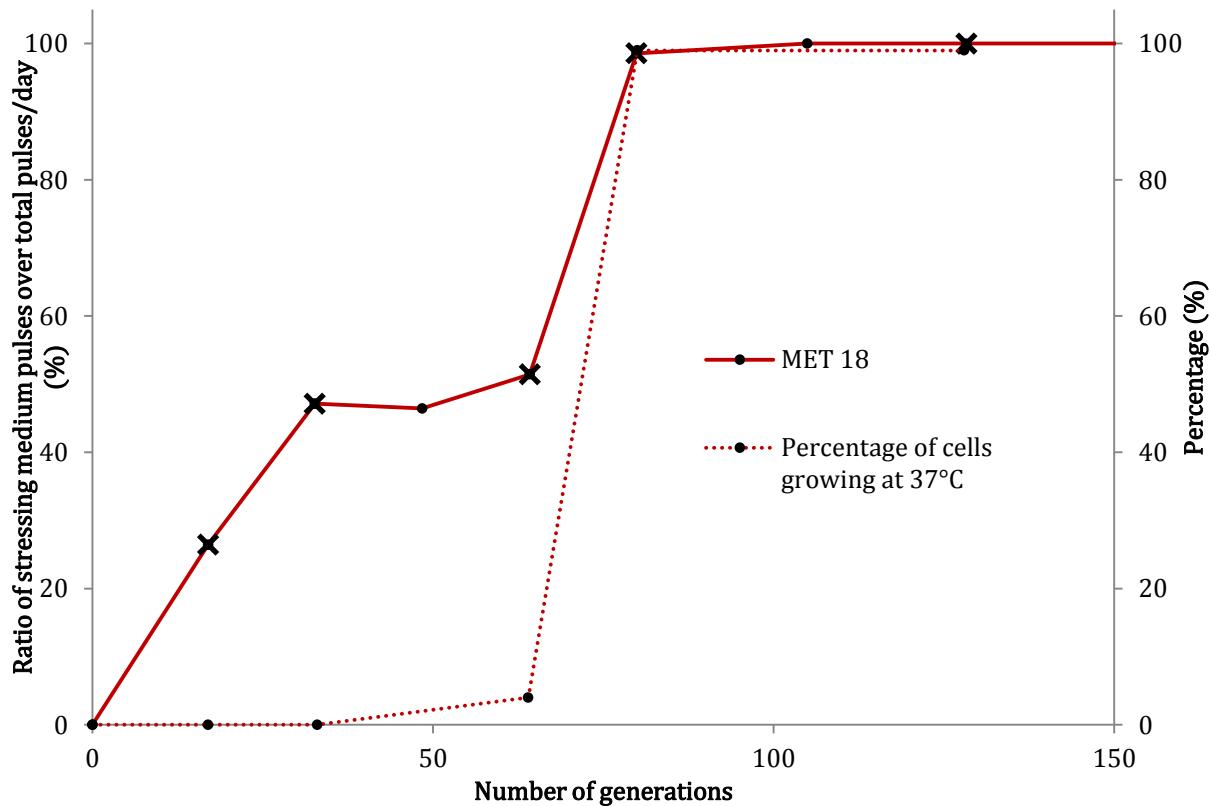


Figure 29: Analysis of the evolution of culture MET 18 in medium SWAP regime at 37°C

Shown is the ratio of stressing medium dilution pulses per day (left axis) in medium SWAP regime. The relaxing medium was MS glc (0.2%) met (0.05 mM) and the stressing medium was MS glc (0.2%). The generation time was fixed at 2 h, and growth temperature at 37°C. Crosses indicate sampling points. Culture MET 18 of strain G4011 ($\Delta metA$, pSP100:: $metA:E237D+$): red line.

Percentage of cells growing at 37°C for samples from culture MET 18 (dotted red line) is shown in right axis.

The following conclusions can be drawn from these data:

- The data corroborate the previous results obtained for the turbidostat (section III.1.4.1) showing that the growth of cells is related to the reversion of the gene *metA* carried on the plasmid.
- Independently of the expression format, it took less than 50 generations from the appearance of the first cells exhibiting the mutated phenotype to the complete fixation of the reverted phenotype for both cultures MET 15 and MET 18. This is about twice as fast as in turbidostat.

III.1.5 Intracellular distribution of high copy-number plasmids

The samples withdrawn from culture MET 10 (Figure 26) contained an increasing number of cells showing growth at 37°C. When the plasmids extracted from a culture sample of an intermediary time point (generation 460) were sequenced, the presence of both *metA* alleles – bearing either codon 237GAT or 237GAG – could be demonstrated (Figure 30).

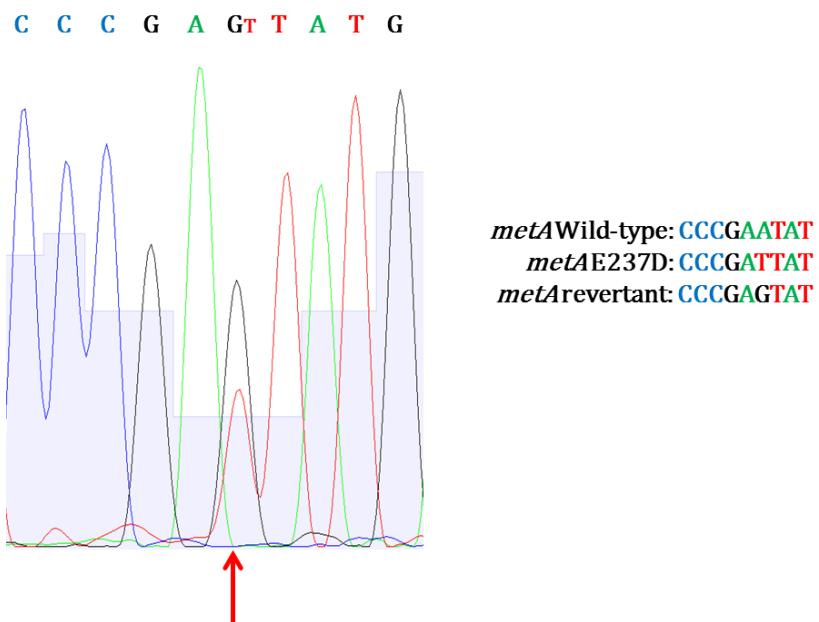


Figure 30: Sanger sequencing chromatogram of the gene *metA* at position 711 from culture MET 10 at generation 460

Color code in chromatogram: **C** base is blue; **G** base is black; **A** base is green, and **T** base is red. The red arrow shows the point mutation.

In the light of the experimental results obtained, we can ask the following questions:

1. What is the dynamic of intracellular reversion accumulation throughout the selection?
2. What is the minimal percentage of reverted *metA* copies in a cell required for growth at 37°C?
3. Do all cells of a culture adapted to growth at 37°C contain the reverted allele at 100%?

To address these questions, experimental methods enabling the collection of statistical data are necessary. A cell sorting/single cell sequencing approach was performed to obtain a first response. The sample at generation 186 from culture Met 10 (Figure 26) was analyzed with this method. Cells were separated in a 384-well microtiter plate. Single cells were deposited in each well with a negligible error rate. Cell membranes were disrupted and the whole genome sequenced. For each well, sequencing reads were aligned against the reference genome for *E. coli* and the plasmid bearing the gene *metA*. Each read mapping on the *metA* sequence was analyzed at the position 711 where the point mutation is located. Data were cleaned from base calling errors and only cells for which the coverage at position 711 was superior to 100 were considered. The ratio of reads at the position 711 of the gene *metA* containing the base G (corresponding to the revertants) per cell was determined. The ratio was also calculated for reads at the position 711 containing the other bases (A and C) but systematically amounted to less than 2% (data not shown). In order to smooth out uncertainties due to amplification biases, a moving average was calculated and the results are plotted in Figure 31.

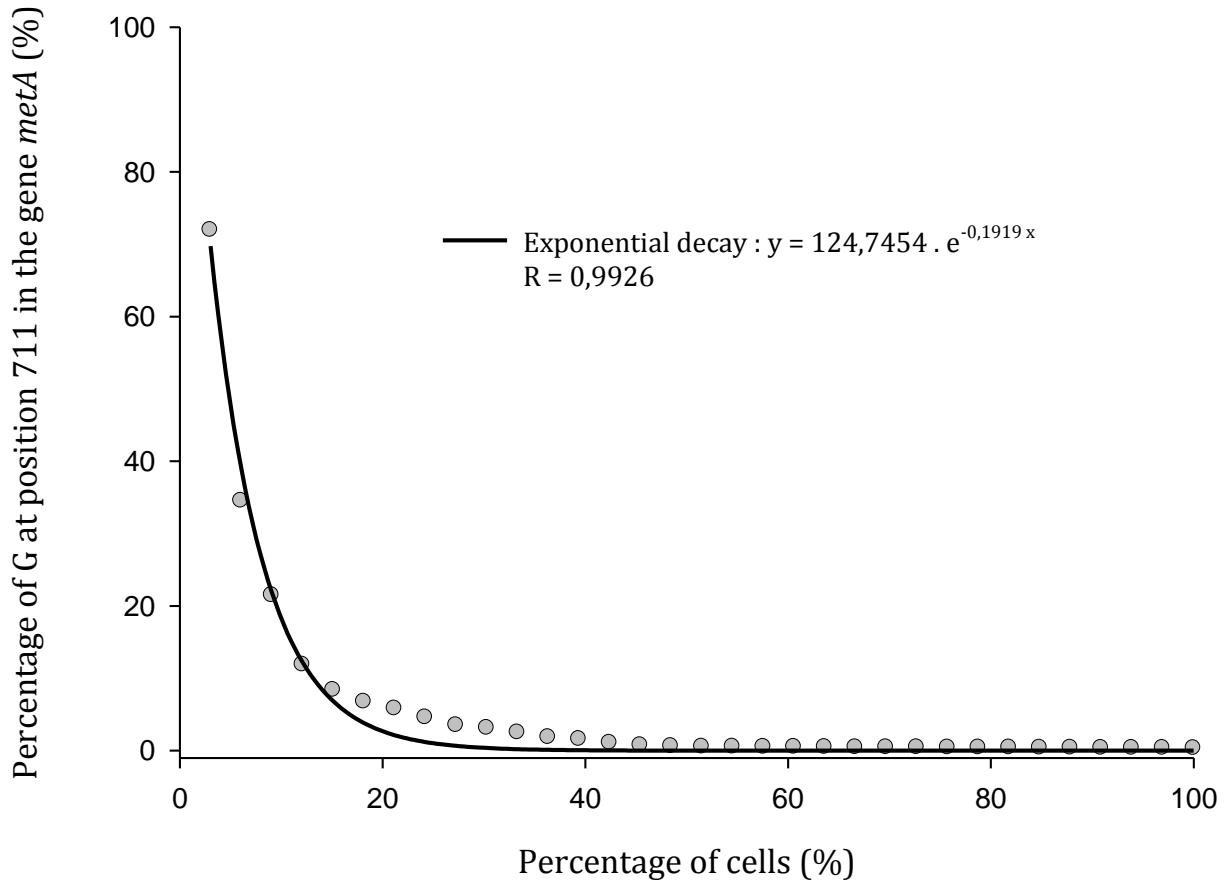


Figure 31: Moving average of the percentage of G (corresponding to the revertant *metA*) at position 711 in the gene *metA* at generation 186 from culture MET 10

Shown is the moving average (gray dots) of the percentage of reads (Coverage threshold ≥ 100) containing the base G at position 711 in the gene *metA* per cell. At position 711, the base in the gene *metA* WT is an A, in the variant *metA:E237D* it is a T and in the revertant gene *metA* from culture MET 10 it is a G. The black line is the iterative curve fitting following an exponential decay with $\lambda = 0.1919 \pm 0.0082$ and $a = 124.7454 \pm 5.1866$

The distribution of the moving average of the percentage of G at position 711 of the gene *metA* follows an exponential decay.

The sample at generation 186 from culture MET 10 was previously phenotypically analyzed (Figure 26, blue line). At this time point, the percentage of cells from the culture growing at 37°C was 6% meaning that 6% of the cells expressed enough reverted MetA to allow growth at 37°C.

When applying this data to the exponential decay (Figure 31), we find $y = 124.7454 \times e^{-0.1919 \times 6} = 39.44\% \pm 3.58\%$. This suggests that approximately 39.44% of reverted *metA* copies per cell is required for growth at 37°C. The sequencing of the additional samples from culture MET 10 is currently in progress to confirm these observations.

III.2 *In vivo* genetic screen for α -KAOs

The main objective of this thesis was to evolve α -KG dependent dioxygenases, through directed evolution *in vivo*, towards better activity on their natural substrate or new activity on substrate analogs. The α -KAOs studied in this work have previously been identified in an activity screen for functioning with basic amino acids (Baud *et al.*, 2014). In a first step, a generic genetic screen for the activity of α -KAOs was designed and implemented into *E. coli*.

III.2.1 Genetic screen design for α -KAOs

A selection strain is a strain whose growth strictly depends on the production of an indispensable metabolite by the enzymatic activity or metabolic pathway under selection. In our study, the fitness of the selection strain should depend on the reaction catalyzed by a α -KAO (Figure 6).

Succinate is a central intermediate in the carbon metabolism. Different metabolic pathways lead to the production of succinate. Interruption of these pathways creates a dependence of succinate for the growth of the selection strain. Any α -KAO activity can thus be put under selective pressure in the cells, which in turn exhibit a proliferation rate directly correlated to the activity of the α -KAO expressed.

Succinate is a product common to every α -KAO-catalyzed reaction, arising through the decarboxylation of the enzyme co-substrate α -ketoglutarate (α -KG). We chose to design a genetic context deficient for the production of succinate. Succinate is a pivotal intermediate of the carbon central metabolism and the energetic metabolism of the cell. Succinate being a common α -KAO product, such a genetic screen applies, in theory, for all enzymes in the dioxygenase family providing that the specific substrate of the enzyme under study can enter the cell and that there is no toxic effect.

Succinate can be formed through different pathways in *E. coli* metabolism:

- When *E. coli* is grown in aerobiosis with glycolytic carbon sources, it derives from α-KG in two steps, oxidative decarboxylation of α-KG and ATP-forming cleavage of succinyl-CoA,
- When *E. coli* is grown in aerobiosis with acetate as a carbon source, the glyoxylate shunt is activated, and succinate is formed by the cleavage of isocitrate by the enzyme isocitrate lyase,
- Succinate constitutes an end-product of the arginine and putrescine degradation pathways (γ -aminobutyrate shunt),
- Succinate can be obtained as well by the reductive functioning of the TCA from oxaloacetate or malate.

Selection schemes can be designed where the entire carbon for biomass formation – with the exception of the C5 amino acids, produced from α-KG – would come from the α-KAO activity. This screen would thus be of maximal stringency. Selections with lowered stringency are also possible, relying on the addition of a second carbon source in the growth medium. Two different generic selection contexts were constructed for this study, deviating in the stringency of the selection:

- A *ppc-pck*-selection context
- A *pck+* selection context

Given the fact that α-KAOs all have a specific second substrate hydroxylated during the reaction, this compound needs to be present in the cell in sufficient concentrations. In general, it is added to the growth medium and thus has to enter the cells.

III.2.1.1 The *ppc-pck*-selection context

The less stringent selection strain was designed to install a succinate auxotrophy by blocking the different pathways listed above (Figure 32).

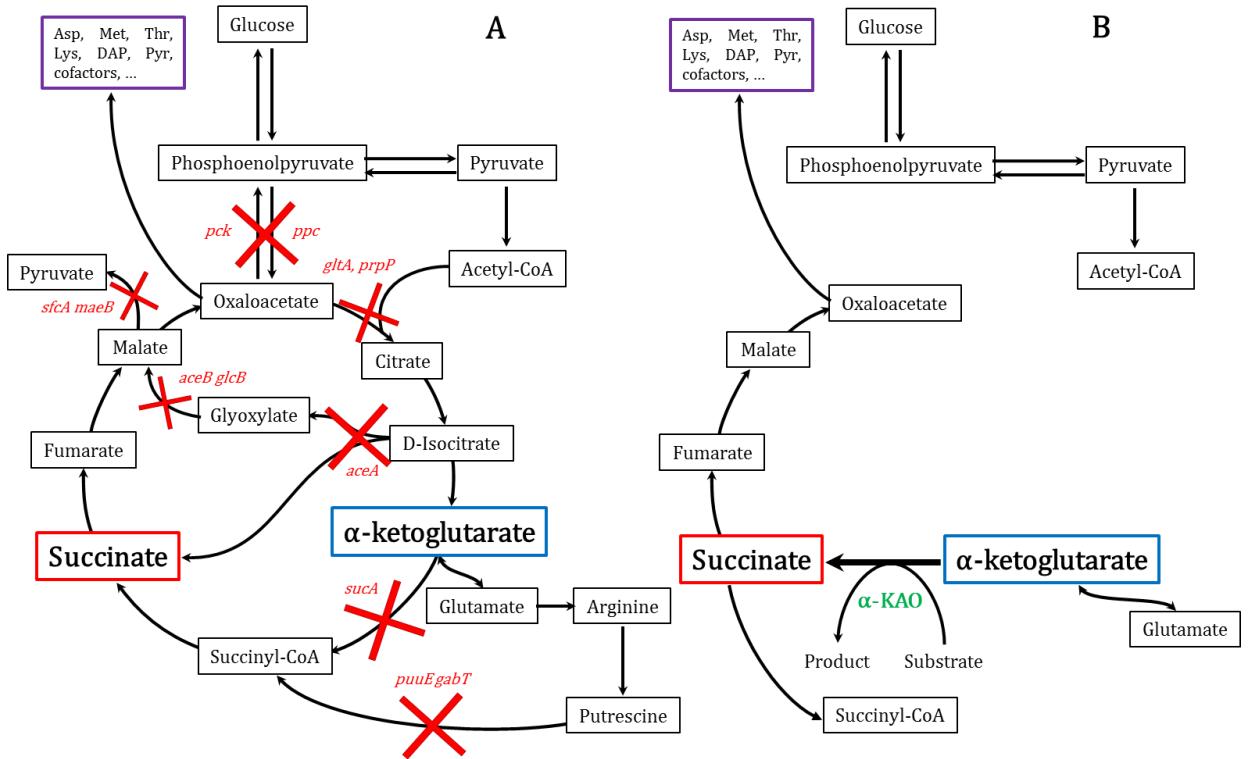


Figure 32: Succinate selection strain construction *ppc-pck*-selection context

A: Simplified scheme showing metabolic pathways involved in the production of succinate; deleted genes are indicated (red). B: Simplified scheme of central carbon metabolism in the *ppc-pck*-selection context after deletion of the genes involved in the production of succinate.

The TCA oxidative branch leading to succinyl-CoA was blocked by deleting the gene encoding 2-oxoglutarate decarboxylase (*sucA*). The glyoxylate shunt allowing the conversion of D-isocitrate to succinate and glyoxylate was blocked with the deletion of the isocitrate lyase (*aceA*). The degradation pathways of arginine and putrescine were blocked with the deletion of the two genes coding for 4-aminobutyrate aminotransferase (*gabT*, *puuE*). The synthesis of succinate through the TCA reductive branch was impeded by deleting the genes encoding the malate synthases (*maeB*, *sfcA*), the phosphoenolpyruvate carboxylase (*ppc*) and the phosphoenolpyruvate carboxykinase (*pck*). The genes of citrate synthase (*gltA*) and methylcitrate synthase (*prpP*) were deleted as well. In this genetic context, the carbon metabolism is decoupled as followed:

- Sugars such as D-glucose (C6), pentoses (C5) or pyruvate are provided to feed the glycolytic pathway and the pentose phosphate pathway.
- Glutamate is provided as source of the amino acids C5 (glutamate, glutamine, arginine, proline) and of the α-KAO co-substrate α-KG.
- Succinate, the product of the α-KAO reaction by the oxidative decarboxylation of α-KG, is the precursor of the amino acids C4 and the pyrimidines.

III.2.1.2 The *pck+* selection context

The more stringent genetic context was designed to select for a succinate production from α -KAO-catalyzed α -KG decarboxylation able to sustain all the carbon requirements of the cell (Figure 33).

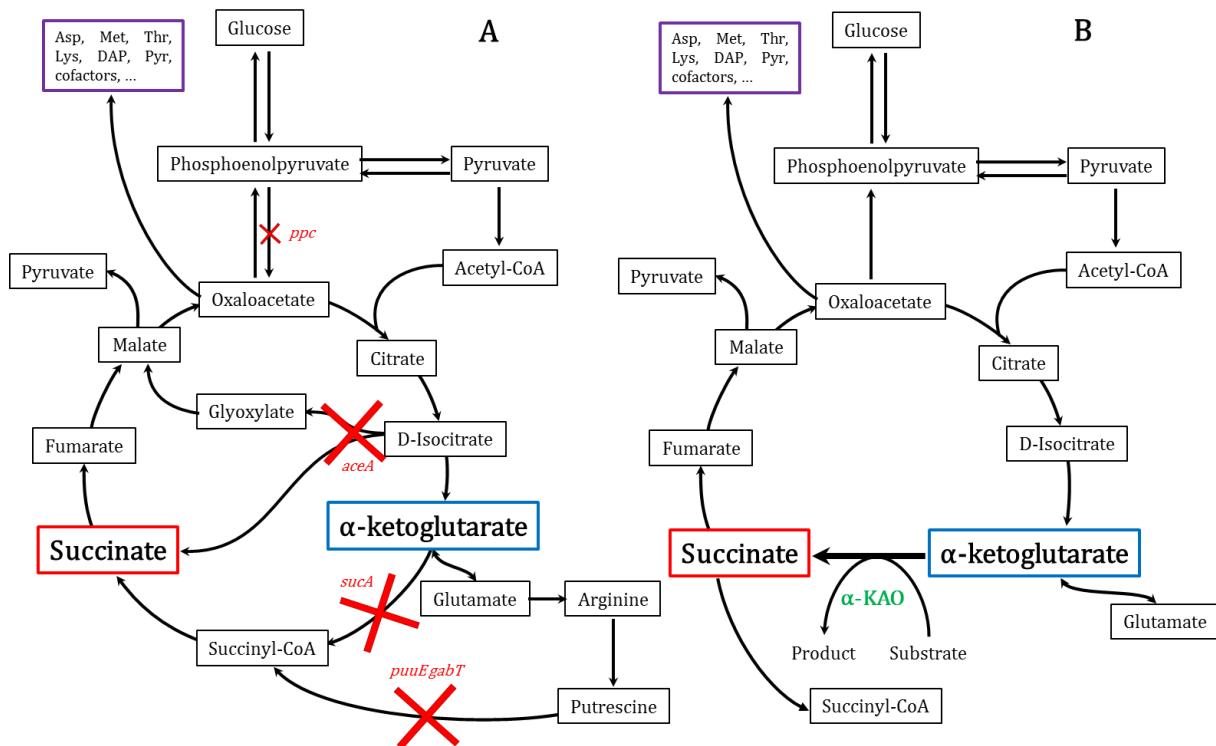


Figure 33: Succinate selection strain construction *pck+* selection context

A: Simplified scheme with showing metabolic pathways involved in the production of succinate; deleted genes are indicated (red). B: Simplified scheme of central carbon metabolism in the *pck+* selection context after deletion of the genes involved in the production of succinate.

In this genetic context, the carbon metabolism is decoupled as followed:

- Glutamate is provided as the sole source of carbon. It is the precursor of the amino acids C5 (glutamate, glutamine, arginine, proline) and of α -KG as α -KAO co-substrate.
- Succinate, the product of the α -KAO reaction, feeds the whole metabolism of carbon, apart from the amino acids C5 via pyruvate formation from PEP, gluconeogenesis and the biosynthesis pathways of the amino acids C4 and of the pyrimidines.

III.2.2 Selection strain construction

Selection strains constructed for this study derived from the *E. coli*K12 strain MG1655.

For each selection strain, successive deletions of chromosomal loci were performed by serial transductions as described in section II.2.1. The analysis of the PCR products obtained using specific primers, and the chromosomal DNA of the various genetic constructs as template confirmed the genotype. The resulting strains are listed in Table 15. It has to be noted that strain G3700 is a derivative of G3513 just lacking a kanamycin resistance cassette.

Table 15: Genotype of the different selection strains used for α -KAO evolution

| Strain | Genotype | Phenotype |
|--------|---|--|
| G3513 | C> $\Delta ArgF-lac$ $\Delta gltA$ $\Delta icd::aad+$ Δpck Δppc $\Delta aceBAK$ $\Delta glcDEFGB$ $\Delta sfcA$ $\Delta maeB$ $\Delta puuE$ $\Delta gabT$ $\Delta sucA::kan+$ | Glutamate and succinate auxotroph |
| G3700 | C> $\Delta ArgF-lac$ $\Delta gltA$ $\Delta icd::aad+$ Δpck Δppc $\Delta aceBAK$ $\Delta glcDEFGB$ $\Delta sfcA$ $\Delta maeB$ $\Delta puuE$ $\Delta gabT$ $\Delta sucA$ | Glutamate and succinate auxotroph |
| G4343 | C> $\Delta aceAK$ $\Delta sucA$ Δppc $\Delta puuE$ $\Delta gabT$ | Succinate auxotroph with glucose as carbon source/Succinate source of carbon |

III.2.3 Selection strain stability

An important feature of a selection strain is its genetic stability. Due to enzyme promiscuity, cryptic pathways might exist which upon mutations can evolve and circumvent genetic selections. For strain G3513, which needs a sugar complement like glucose in its growth medium, this possibility had to be envisioned.

The stability of selection strain G3513 was tested in continuous culture under a medium SWAP regime (Figure 34). Conditions of the medium SWAP regime for the culture of G3513 were as follows:

- Relaxing medium: MS glc (0.2%) glu (2 mM) suc (100 μ M)
- Stressing medium: MS glc (0.2%) glu (2 mM)
- Generation time 2 h at 37°C

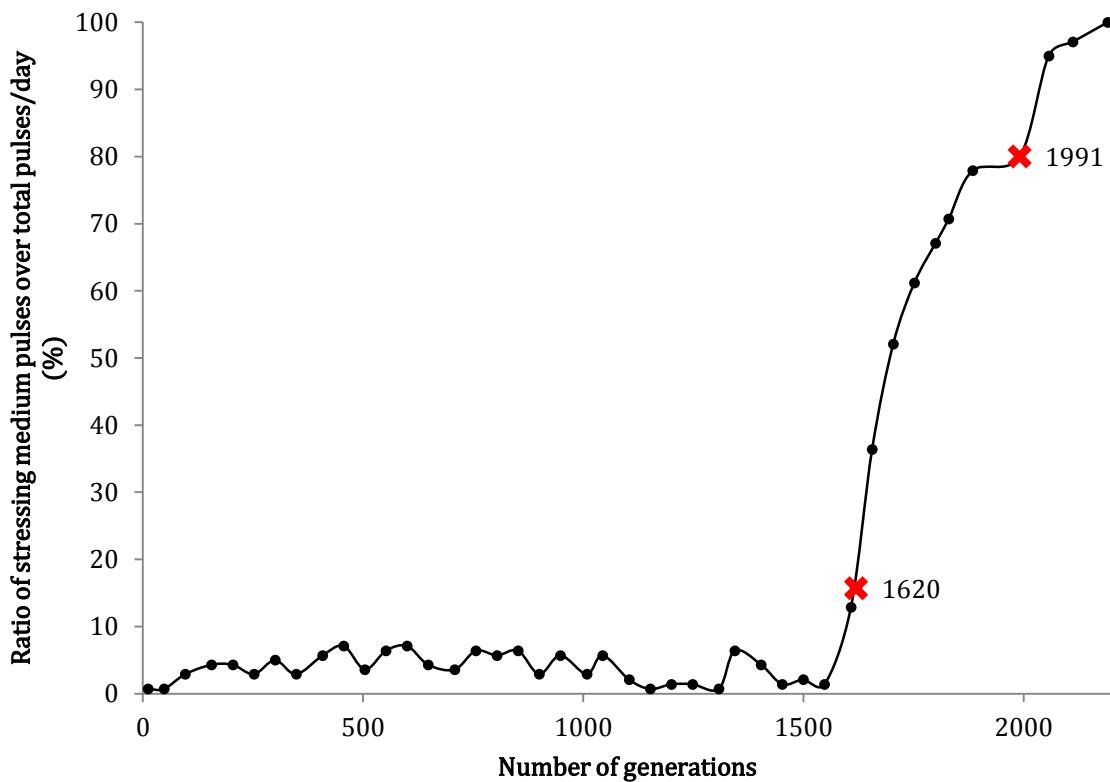


Figure 34: Evolution of the culture of strain G3513 in medium SWAP regime

Shown is the ratio of stressing medium dilution pulses per day in medium SWAP regime. The relaxing medium was MS glc (0.2%) glu (2 mM) suc (100 μ M) and the stressing medium was MS glc (0.2%) glu (2 mM). The generation time was fixed at 2 h, and growth temperature at 37°C. Crosses indicate sampling points with the corresponding number of generations.

The low concentration of succinate in the relaxing medium (100 μ M) strongly limited the daily number of stressing medium pulses at the onset of the culture. The daily ratio of stressing medium pulses remained stable for about 1500 generations (50 days) before detecting a sudden increase in the number of dilution pulses of stressing medium. It took 2200 generations for the culture (91 days) to reach 100% of stressing medium.

Samples were collected before the increase in the ratio of stressing medium (generation 1620) and around a ratio of 80% stressing medium (generation 1991). From each sample, two clones were isolated, G3886 and G3887 at generation 1620 and G3685 and G3686 at generation 1991.

Phenotypic tests (Table 16) showed that the isolates collected at generation 1620 started to have residual succinate complementation. A clear shift in retrieving the succinate prototrophy happened for both isolates from generation 1991.

Table 16: Growth tests of isolates from evolution of strain G3513

| Strain | Genotype | Growth in MS glc (0.2%) glu (5 mM) | |
|--------------------|--|------------------------------------|--------------|
| | | Ø | suc (0.1 mM) |
| G3513 | <i>C>ΔArgF-lac ΔgltA Δicd::aad+ Δpck Δppc</i> <i>ΔaceBAK ΔglcDEFGB Δsfca ΔmaeB ΔpuuE ΔgabT</i> <i>ΔsucA::kan+</i> | - | +++ |
| G3886 and G3887 | Isolates sampled from culture of G3513 at generation 1620 | + | +++ |
| G3685 and G3686 | Isolates sampled from culture of G3513 at generation 1991 | +++ | +++ |

Complementation tests on solid medium were performed at 37°C for 48 h. The bacterial growth observed is indicated with signs for normal (+++), medium (++) residual (+) and no complementation (-).

These data suggest that some mutations have been fixed in the population allowing a cryptic pathway to generate succinate when bacteria are grown in the stressing medium. The carbon sources available in this medium are D-glucose, glutamate, and citrate (present at a final concentration of 4 mM for chelating metal ions, Appendix 6). We tested whether the growth in the stressing medium of the evolved isolates depended on the presence of citrate. For this, phenotypic tests were performed replacing MS medium by MA medium which lacks citrate (Appendix 7, Table 17).

Table 17: Growth tests for the utilization of citrate as succinate source

| Strain | Genotype | Growth in MA glc (0.2%) glu (5 mM) | |
|--------------------|--|------------------------------------|------------|
| | | Ø | cit (4 mM) |
| G3513 | <i>C>ΔArgF-lac ΔgltA Δicd::aad+ Δpck Δppc</i> <i>ΔaceBAK ΔglcDEFGB Δsfca ΔmaeB ΔpuuE ΔgabT</i> <i>ΔsucA::kan+</i> | - | - |
| G3886 and G3887 | Isolates sampled from culture of G3513 at generation 1620 | - | - |
| G3685 and G3686 | Isolates sampled from culture of G3513 at generation 1991 | - | +++ |

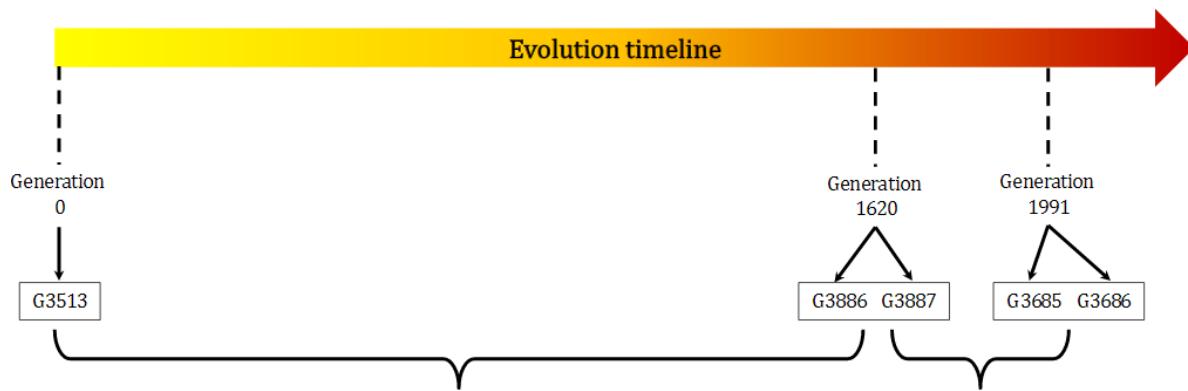
Complementation tests on solid medium were performed at 37°C for 48 h. The bacterial growth observed is indicated with signs for normal (+++), medium (++) residual (+) and no complementation (-).

These results show that the isolates from generation 1991 acquired the ability to aerobically use citrate.

The absence of growth with citrate as a carbon source in aerobic conditions is a characteristic trait of the *E. coli* species. However, the isolation of *E. coli* spontaneous mutants able to utilize citrate aerobically as carbon and energy source has been described by various authors, and the genetic basis of this gain of function explored. Aerobic citrate utilization was selected in the Long Term Evolution Experiment (LTEE) of Lenski after more than 30 000 generations (Blount *et al.*, 2012) but it can appear much more rapidly in directed selections (Hall, 1982; Hofwegen *et al.*, 2016). The acquisition of a stable citrate utilization phenotype requires the expression of two transporters, the citrate transporter, encoded by the gene *citT*, whose expression is down-regulated in aerobic conditions and the *dctA*-encoded dicarboxylic acid transporter, known to be involved in the transport of succinate, fumarate, malate, orotate and citrate (Hofwegen *et al.*, 2016; Quandt *et al.*, 2014). Upregulation of these two genes can arise following various genetic events such as gene duplication, insertion of an IS which places the gene under a strong promoter, or various chromosomal rearrangements leading to promoter capture.

These previous experiments were all carried out with *E. coli* strains expressing an unmodified central carbon metabolism, where once citrate is transported into the cell, it is metabolized through the oxidative branch of TCA. However, in our experimental setup, the strain G3513, as well as the citrate-utilizing isolates G3685 and G3686, are devoid of functional TCA (Δicd) and glyoxylate shunt ($\Delta aceAK$). This raises the question of how citrate is metabolized in these evolved isolates.

The genomes of strain G3513 and its derivatives G3886, G3887, G3685, and G3686 were sequenced (Myseq, Illumina); the reads were analyzed with PALOMA (Polymorphism Analyses in Light Of Massive DNA sequencing) in the Microscope platform (Score ≥ 0.5 , HQ Reads ≥ 5). In total, 24 non-synonymous genic SNPs, 5 intergenic SNPs and 23 small in/dels common to both isolates from each time-point were detected. The evolution timeline can be split in two, the period before the increase of stressing medium in the culture (before 1620 generations), and the period after. An overview of the non-synonymous intragenic SNPs is presented in Figure 35.



| Nº | Gene | Description | Amino acid change |
|----|------|---|-------------------|
| 1 | allA | ureidoglycolate hydrolase | V93A |
| 2 | asnA | asparagine synthetase A | G143R |
| 3 | atpA | F1 sector of membrane-bound ATP synthase, alpha subunit | A202T |
| 4 | crl | DNA-binding transcriptional regulator | E127K |
| 5 | dctA | C4-dicarboxylic acid, orotate and citrate transporter | E65G |
| 6 | envZ | sensory histidine kinase in two-component regulatory system with OmpR | A193V |
| 7 | gadX | DNA-binding transcriptional dual regulator | E129G |
| 8 | mazG | nucleoside triphosphate pyrophosphohydrolase | D196G |
| 9 | metB | cystathionine gamma-synthase, PLP-dependent | G103A |
| 10 | mutS | methyl-directed mismatch repair protein | A789V |
| 11 | relA | (p)ppGpp synthetase I/GTP pyrophosphokinase | L198R |
| 12 | sbcD | exonuclease, dsDNA, ATP-dependent | L375P |
| 13 | topB | DNA topoisomerase III | T598A |
| 14 | torZ | trimethylamine N-oxide reductase system III, catalytic subunit | T380A |
| 15 | ybaE | conserved hypothetical protein | G542S |
| 16 | ybbS | putative DNA-binding transcriptional activator of the allD operon | T52A |
| 17 | ybgP | putative exported chaperone, PapD family | H122R |
| 18 | ydgI | putative arginine/ornithine antiporter transporter | R64H |
| 19 | yfbN | hypothetical protein | C231R |
| 20 | yhgN | putative antibiotic transporter | V53A |

| Nº | Gene | Description | Amino acid change |
|----|------|--|-------------------|
| 21 | fruR | DNA-binding transcriptional dual regulator | T61A |
| 22 | ompC | outer membrane porin protein C | V10A |
| 23 | yhbC | conserved hypothetical protein | A144V |
| 24 | yihP | putative transporter | N114S |

Figure 35: Overview of the non-synonym intragenic SNPs along the evolution of the strain G3513

Mutations listed from PALOMA (Score ≥ 0.5 , HQ Reads ≥ 5). Only non-synonymous intragenic SNPs common to both isolates from the same time-point are showed. **Green:** Mutations found at generation 1620 but not at generation 1991 compared to the initial strain G3513. **Yellow:** Mutations found at both time-points (1620 and 1991 generations) compared to the initial strain G3513. **Orange:** Mutations found only at generation 1991 compared to generation 1620.

At generation 1620, 20 non-synonymous intragenic SNPs are common to both isolates G3886 and G3887 (Figure 35, mutations in green and yellow). Among those mutations, we notice a mutation in *mutS* (Figure 35, mutation n°10). The protein MutS is one of the components of the MutHLS complex which has a role in the methyl-directed mismatch repair pathway. A mutation in *mutS* could result in a mutator phenotype for which the spontaneous mutation frequency has increased compared to the initial strain. It has been shown that mutator genes can accelerate adaptation of populations growing in a new environment (Taddei *et al.*, 1997).

Also, among the 20 mutations detected at generation 1620, 3 mutations are located in genes related to transport mechanisms: *ydgI*, *yhgN*, and the aforementioned *dctA*, which encodes the dicarboxylic acid transporter. Moreover, two intergenic SNPs are detected 21 and 25 bp upstream of the ATG START codon of *dctA*, which might modulate the expression level of DctA.

These combined upstream and intragenic mutations affecting the *dctA* locus were not sufficient on their own to allow the isolates G3886 and G3887 to metabolize citrate. This corroborates the observations made in the phylogenetic analysis of the Cit⁺ trait in the LTEE (Quandt *et al.*, 2014).

Four additional mutations were detected in the genomes of the isolates G3685 and G3686 sampled at generation 1991 (Figure 35, mutations in orange) as well as a two-fold amplification of the read coverage profile of the *rhsB-rhsA* region (140 kbp long) where *dctA* is situated, compared to the coverage of the rest of the chromosome (x100). Interestingly, this amplification suggests a genetic event constituting an independent duplicate described in Van Hofwegen *et al.*, 2016. As citrate utilization as a carbon and energy source requires the activation of both CitT and DctA transporters, the *citT* locus of our isolates was PCR amplified using specific pairs of primers and the DNA fragments obtained were sequenced. Sequence analysis excluded the presence of an IS upstream of the *citT* gene or local rearrangements which would have modified the promoting region of this gene.

No evident clue was found which could indicate how citrate might be converted into succinate. However, some activities specified by the *cit* operon (*citCDEXG*), notably the citrate lyase, are likely to be involved in this metabolic pathway, despite the fact that this enzyme is described to be active under anaerobic conditions (Lütgens and Gottschalk, 1980). Further analyses are required to understand the mechanisms of citrate utilization but would be beyond the scope of the present work.

Overall, we can conclude from this experiment that:

- The genetic screen was stable for about 1500 generations (50 days), which allows us to consider this strain for the selection of α -KAOs variants (within the timeframe of 1500 generations).
- An unknown cryptic pathway converting citrate to succinate was activated. To avoid the selection of the mutations leading to its activation, MA mineral medium which lacks citrate was used in subsequent long-term evolution experiments (Appendix 7).¹

III.2.4 Complementation tests of α -KAOs expressed in selection strains

The α -KAOs active with basic amino acids (Table 2) were cloned in different expression formats (Table 10). Constructs were then introduced in the selection strain G3513 (Table 15), and succinate complementation tests were performed. As can be seen in Table 18, no succinate complementation was detected for any of the constructs tested.

The following hypothesis can be put forward to explain this lack of complementation:

- The activity of the enzyme is too low to produce enough succinate required for triggering growth of the selection strain
- The enzyme is poorly expressed
- The enzyme is inactive in the heterologous host *E. coli* because of misfolding
- The hydroxylated amino acid resulting from the enzymatic reaction is toxic

¹ Some evolution experiments were started prior to the aforementioned results and were thus conducted in MS mineral medium instead of MA mineral medium.

Table 18: Succinate complementation tests of α -KAOs expressed from different formats

| Enzyme | Expression format | Selection | | MS glc (0.2%) glu (5mM) | | | |
|--------|-------------------|-----------|--------|-------------------------|------------|------------|------------|
| | | strain | Strain | \emptyset | suc (5 mM) | orn (5 mM) | lys (5 mM) |
| Odo | pSP100 | G3513 | G3542 | - | ++ | - | |
| | pVDM18 | G3513 | G3543 | - | +++ | - | |
| Kdo1 | pSP100 | G3513 | G3544 | - | ++ | | - |
| | pVDM18 | G3513 | G3545 | - | +++ | | - |
| Kdo2 | pSP100 | G3513 | G3982 | - | ++ | | - |
| | pVDM18 | G3513 | G3547 | - | +++ | | - |
| | Chromosomal | G3700 | G4160 | - | +++ | | - |
| Kdo3 | pSP100 | G3513 | G3548 | - | ++ | | - |
| | pVDM18 | G3513 | G3549 | - | +++ | | - |
| Kdo4 | pSP100 | G3513 | G4087 | - | ++ | | - |
| | pVDM18 | G3513 | G3984 | - | +++ | | - |
| Kdo5 | pSP100 | G3513 | G4089 | - | ++ | | - |

Complementation tests on solid medium were performed at 37°C for 48 h. The bacterial growth observed is indicated with signs for normal (+++), medium (++) residual (+) and no complementation (-). Refer to Table 15 or Appendix 1 for a precise genotype of selection strains.

III.3 Biochemical analysis of α -KAOs

The α -KAOs were expressed in the selection strain, but no succinate complementation was detected. In order to determine if the activity of the enzyme was too low to produce enough succinate required for detectable growth of the selection strain, the kinetic parameters of the α -KAOs were determined.

α -KAOs were expressed with a N-terminal His-tag from vector pET22 (Table 6) and purified for *in vitro* analysis. Purification and enzymatic assays were performed as described in section II.3. Kinetic parameters were determined for both the α -KG and the amino acid substrate (Table 19).

Table 19: α -KAOs kinetic parameters

| Enzyme | Substrate | K_M (μM) | | | k_{cat} (s^{-1}) | | | k_{cat}/K_M ($\text{s}^{-1} \cdot \text{M}^{-1}$) | | |
|--------|--------------|-------------------------|-------|-------|--------------------------------------|-------|-------|--|----------|--------|
| Kdo1 | Lysine | 29.97 | \pm | 5.69 | 0.64 | \pm | 0.04 | 2.12 | \times | 10^4 |
| | α -KG | 11.44 | \pm | 3.06 | 0.80 | \pm | 0.05 | 6.98 | \times | 10^4 |
| Kdo2 | Lysine | 96.24 | \pm | 19.13 | 1.79 | \pm | 0.10 | 1.86 | \times | 10^4 |
| | α -KG | 31.09 | \pm | 4.18 | 2.20 | \pm | 0.07 | 7.07 | \times | 10^4 |
| Kdo3 | Lysine | 73.64 | \pm | 6.48 | 2.14 | \pm | 0.05 | 2.90 | \times | 10^4 |
| | α -KG | 21.46 | \pm | 2.61 | 2.20 | \pm | 0.05 | 1.02 | \times | 10^5 |
| Kdo4 | Lysine | 48.26 | \pm | 3.13 | 1.81 | \pm | 0.003 | 3.75 | \times | 10^4 |
| | α -KG | 18.98 | \pm | 2.43 | 1.80 | \pm | 0.005 | 9.49 | \times | 10^4 |
| Kdo5 | Lysine | 129.22 | \pm | 25.77 | 4.34 | \pm | 0.26 | 3.36 | \times | 10^4 |
| | α -KG | 27.92 | \pm | 1.97 | 3.20 | \pm | 0.005 | 1.15 | \times | 10^5 |

The α -KAOs tested exhibit similar turnover rates (k_{cat}) and enzymatic efficiencies (k_{cat}/K_M). The enzymatic efficiency is 3 to 5 times higher for α -KG than for lysine, the deviation being mainly due to differences in K_M values. Statistics of known enzymatic activity parameters concerning all types of enzymes have been exploited to define an “average enzyme” (Arren Bar-Even *et al.*, 2011). According to this study, the average enzyme exhibits a $k_{\text{cat}} \sim 10 \text{ s}^{-1}$ and a $k_{\text{cat}}/K_M \sim 10^5 \text{ s}^{-1} \cdot \text{M}^{-1}$. The parameters determined for the α -KAOs (Table 19) are compared to the average enzyme in Figure 36 and Figure 37.

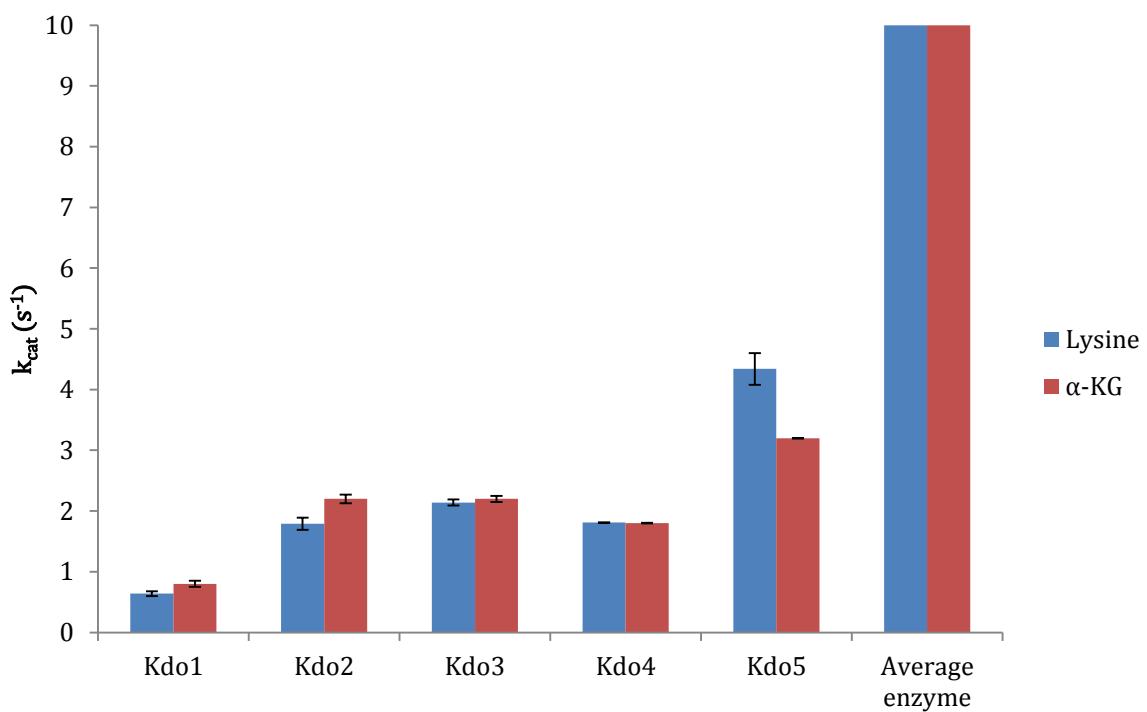


Figure 36: Comparison of the turnover rate (k_{cat}) between the α -KAOs (Kdo1 to Kdo5) and the average enzyme

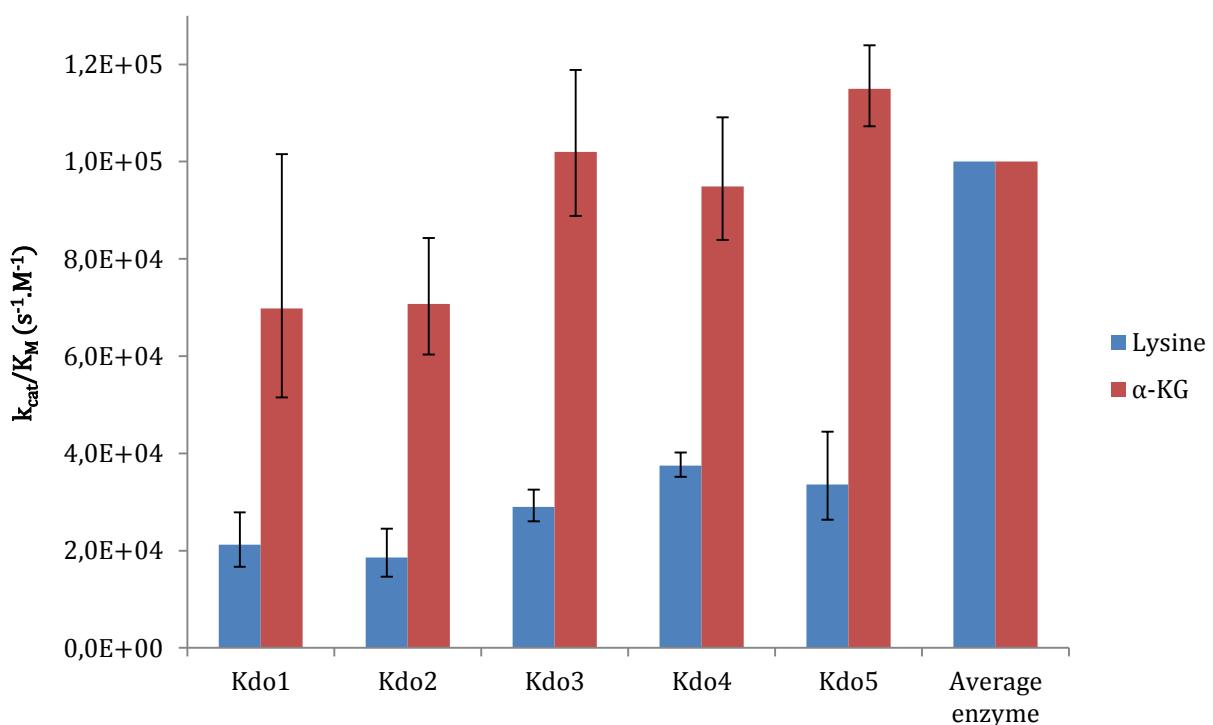
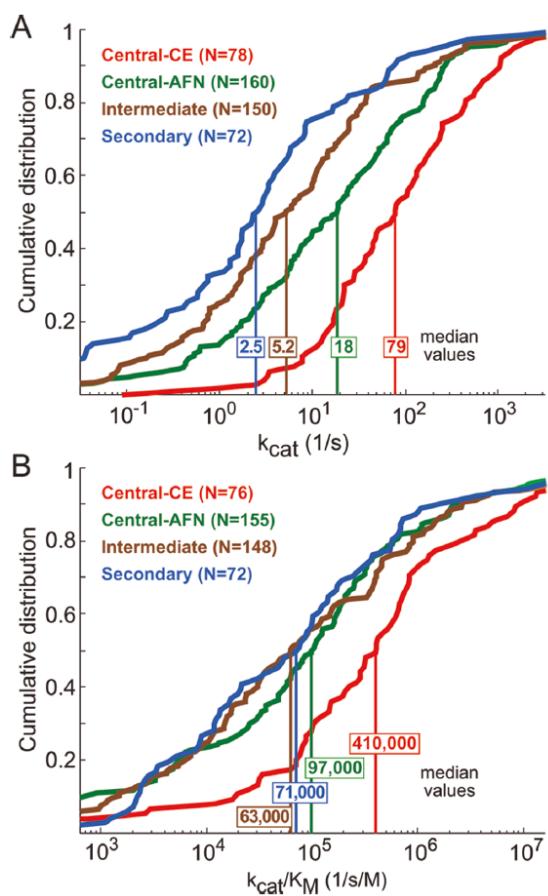


Figure 37: Comparison of the enzyme efficiency (k_{cat}/K_M) between the α -KAOs (Kdo1 to Kdo5) and the average enzyme

The α -KAOs have a turnover rate 2 to 15 times lower than the average enzyme. The enzymatic efficiency for lysine is about 2 to 5 times lower than the average enzyme and close to the average enzyme for α -KG. According to Bar-Even and colleagues, the median turnover rate of reactions associated with secondary metabolism is around 2.5 s^{-1} (Figure 38, (A)) and the median enzymatic efficiency is around $7.1 \times 10^4 \text{ s}^{-1} \cdot \text{M}^{-1}$ (Figure 38, (B)).



(A) Distribution of k_{cat} values for enzyme-substrate pairs belonging to different metabolic contexts. All distributions are significantly different with a p value of <0.0005 (rank-sum test), except for intermediate versus secondary metabolisms ($p < 0.05$). (B) Distribution of k_{cat}/K_M values for enzyme-substrate pairs belonging to different metabolic contexts. Central-CE (carbohydrate and energy) metabolism has significantly higher k_{cat}/K_M values than all other metabolic groups [$p < 0.0005$ (rank-sum test)]. Abbreviations: CE, carbon and energy; AFN, amino acids, fatty acids, and nucleotides. Numbers in parentheses represent the numbers of enzyme-substrate pairs included in each set.

Reprinted from Arren Bar-Even *et al.*, 2011

Figure 38: Enzymes operating within different metabolic groups have significantly different k_{cat} and k_{cat}/K_M values.

These data suggest that the α -KAOs tested in this study belong to the group of enzymes catalyzing reactions of the secondary metabolism. In the secondary metabolism, selection pressure might be lower on the kinetic parameters as low fluxes can be maintained with relatively weak enzymes compared to those operating in the central metabolism (Arren Bar-Even *et al.*, 2011). However, in the present genetic screen, the α -KAOs have to produce enough succinate to fulfill the needs for the C4 amino acids and

the pyrimidines by bridging the TCA cycle. These observations could explain why no succinate complementation is observed with the α -KAOs tested (Table 18).

This opens the possibility of improving the activity of the α -KAOs for a higher succinate production and in parallel a higher hydroxylation rate of lysine as second substrate using the low-stringency genetic screen constructed (described in section III.2.1.1).

III.4 Evolution *in vivo* of α -KAOs in a *ppc-pck*-selection context

The kinetic parameters of the α -KAOs determined in the previous section showed that they might belong to the group of enzymes involved in the secondary metabolism. These observations could explain why no succinate complementation was observed (Table 18). The following section summarizes the evolution experiments of α -KAOs in continuous culture.

III.4.1 Evolution *in vivo* of α -KAOs expressed from a plasmidic format

The α -KAOs expressed on a high copy number plasmid did not complement the succinate requiring selection strain G3513 (Table 18). Succinate had to be added to the medium to obtain growth of the strains. Under these conditions, the possibility of plasmid loss had to be considered since the presence of the gene is not necessary for growth. To stabilize the plasmids in the evolving cell population, antibiotic resistance genes are not suitable since resistance mutations are likely to appear. Therefore a plasmid, the pEVL404, was constructed containing the essential gene *thyA* of *E. coli* coding for thymidylate synthase (Munch, 1983) in place of the antibiotic resistance gene. Selection strains carrying a *thyA* deletion in their chromosome depend on the presence of thymidine in the medium for growth. In theory, the *thyA*+ plasmid substitutes for this growth factor requirement and cannot be lost by cells growing in the absence of exogenous thymidine.

The genes coding for the α -KAOs, Odo, Kdo1 and Kdo2 (Table 2) were cloned in the plasmid pEVL404 (Figure 12) and the resulting constructs transformed into the strain G3752, a thymidine auxotroph derivative of the selection strain G3700 (Table 20).

Table 20: Strains used for α -KAO evolution *in vivo* in plasmidic format

| Strain | Genotype | GM3 culture name |
|--------|--|------------------|
| G3700 | $\Delta ArgF-lac$ $\Delta gltA$ $\Delta icd::aad+$ Δpck Δppc $\Delta aceBAK$ $\Delta glcDEFGB$ $\Delta sfCA$ $\Delta maeB$ $\Delta pruuE$ $\Delta gabT$ $\Delta sucA$ | |
| G3752 | G3700 $\Delta thyA::erm+$ | |
| G3935 | G3752 pGEN1066 (pEVL404:: <i>odo</i> +) | ODO |
| G3951 | G3752 pGEN1070 (pEVL404:: <i>kdo1</i> +) | KDA |
| G3986 | G3752 pGEN1080 (pEVL404:: <i>kdo2</i> +) | KDB |

In Figure 39 are depicted the experimental steps involved in the evolution of α -KAOs in plasmidic format.

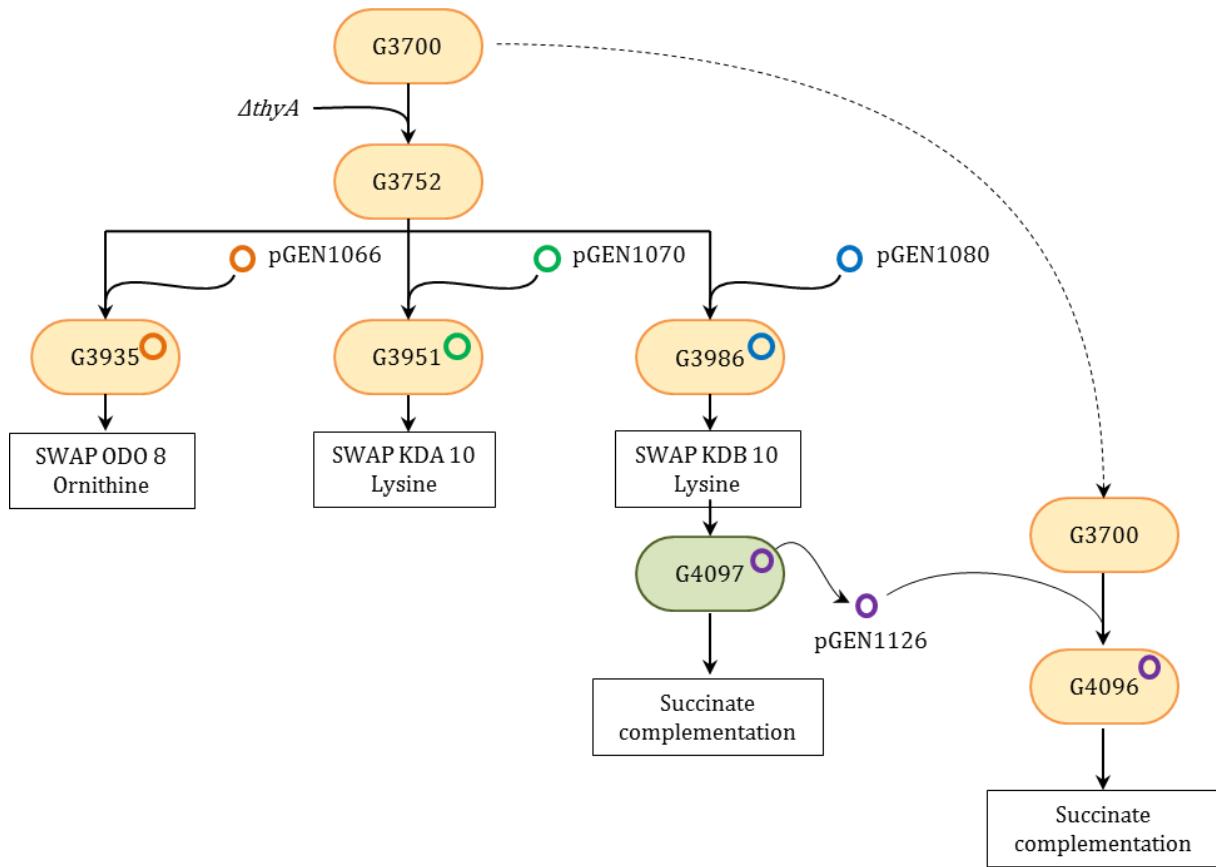


Figure 39: Overview of α -KAOs evolution experiments in plasmidic format

Strains are represented by an oval shape. Green ovals represent strains that have evolved in the GM3. Plasmids are represented by a circle: pGEN1066 (orange), pGEN1070 (green), pGEN1080 (blue), and pGEN1126 (purple). For every medium SWAP regime, generation time was fixed at 2 h, and growth temperature at 37°C.

III.4.1.1 Evolution *in vivo* of Odo, Kdo1 and Kdo2 in medium SWAP regime

Strains G3935, G3951 and G3986 (Table 20) were inoculated in GM3 devices and cultivated under medium SWAP regime. Each culture was given a 3-letter code corresponding to the project name followed by an iterative number increasing each time a new culture is launched (Table 20).

Conditions of the medium SWAP regime for the culture ODO 8 (strain G3935 grown with ornithine, Figure 40, orange line) were as follows:

- Relaxing medium: MS glc (0.2%) glu (5 mM) orn (5 mM) suc (0.8 mM)
- Stressing medium: MS glc (0.2%) glu (5 mM) orn (5 mM)
- Generation time 2 h at 37°C

Conditions of the medium SWAP regime for the culture of KDA 10 (strain G3951 grown with lysine, Figure 40, green line) were as follows:

- Relaxing medium: MS glc (0.2%) glu (5 mM) lys (5 mM) suc (0.8 mM)
- Stressing medium: MS glc (0.2%) glu (5 mM) lys (5 mM)
- Generation time 2 h at 37°C

Conditions of the medium SWAP regime for the culture KDB 10 (strain G3986 grown with lysine, Figure 40, blue line) were as follows:

- Relaxing medium: MS glc (0.2%) glu (5 mM) lys (5 mM) suc (0.8 mM)
- Stressing medium: MS glc (0.2%) glu (5 mM) lys (5 mM)
- Generation time 2 h at 37°C

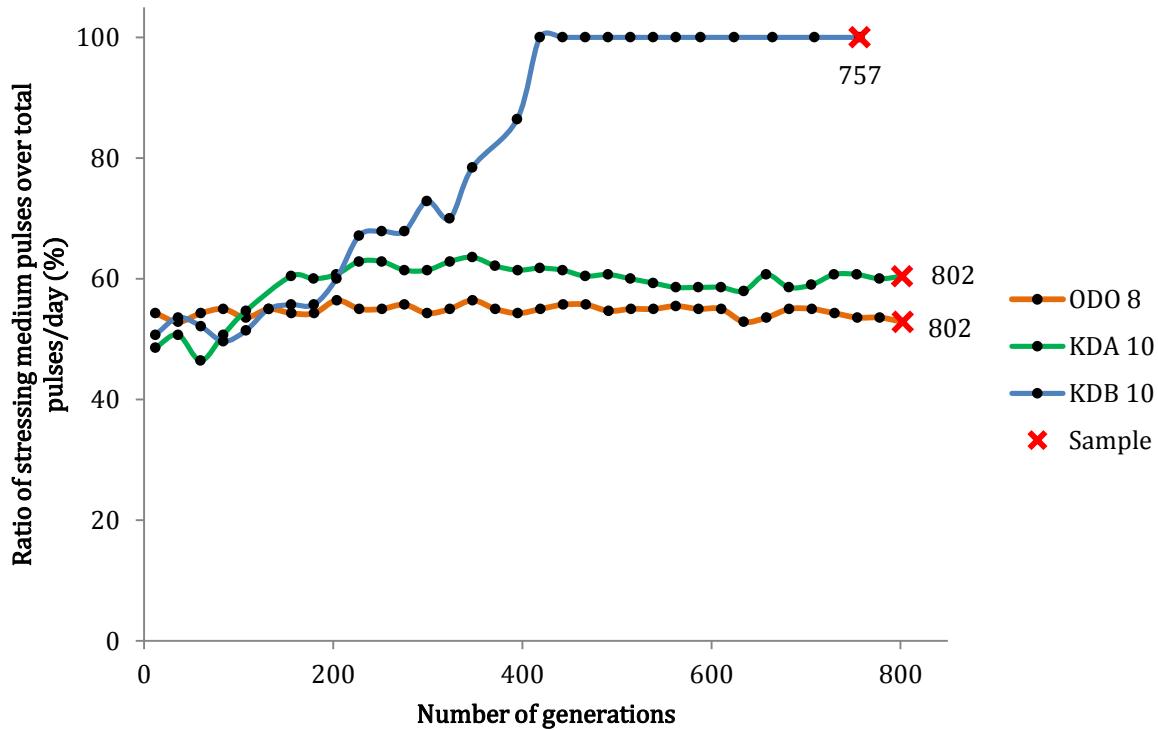


Figure 40: Evolution of cultures ODO 8, KDA 10 and KDB 10 in medium SWAP regime

Shown is the ratio of stressing medium dilution pulses per day. The generation time was fixed at 2 h, and growth temperature at 37°C. Crosses indicate sampling points with the corresponding number of generations.

Culture ODO 8 of strain G3935 (G3752 pGEN1066 (pEVL404::*odo*+)): orange line. Relaxing medium was MS glc (0.2%) glu (5 mM) orn (5 mM) suc (0.8 mM) and stressing medium was MS glc (0.2%) glu (5 mM) orn (5 mM).

Culture KDA 10 of strain G3951 (G3752 pGEN1070 (pEVL404::*kdo*1+)): green line. Relaxing medium was MS glc (0.2%) glu (5 mM) lys (5 mM) suc (0.8 mM) and stressing medium was MS glc (0.2%) glu (5 mM) lys (5 mM).

Culture KDB 10 of strain G3986 (G3752 pGEN1080 (pEVL404::*kdo*2+)): blue line. Relaxing medium was MS glc (0.2%) glu (5 mM) lys (5 mM) suc (0.8 mM) and stressing medium was MS glc (0.2%) glu (5 mM) lys (5 mM).

The ratio of stressing medium pulses over total pulses per day was stable at around 60% for more than 1000 generations in the case of cultures ODO 8 and KDA 10. By then, the cultures were stopped. In the case of culture KDB 10, the ratio of stressing medium pulses over total pulses per day increased steadily and reached 100% of stressing medium pulses after 450 generations.

III.4.1.2 Sequencing analysis of isolates from culture KDB 10

The gene *odo* and *kdo2* from samples of cultures ODO 8 and KDA 10 were sequenced; no mutations were detected in the genes.

Isolates were obtained from a sample of the culture KDB 10. Plasmids were extracted from one of the clones (strain G4097, Figure 39), and designated pGEN1126 (Figure 41). No mutation was found in the gene *kdo2*. Sequencing of the whole plasmid revealed a deletion of about 1500 bp when compared with the initial plasmid pGEN1080 (Figure 42). The excised part contained the gene *thyA*. Besides this excision, no other mutations were detected in the plasmid pGEN1126.

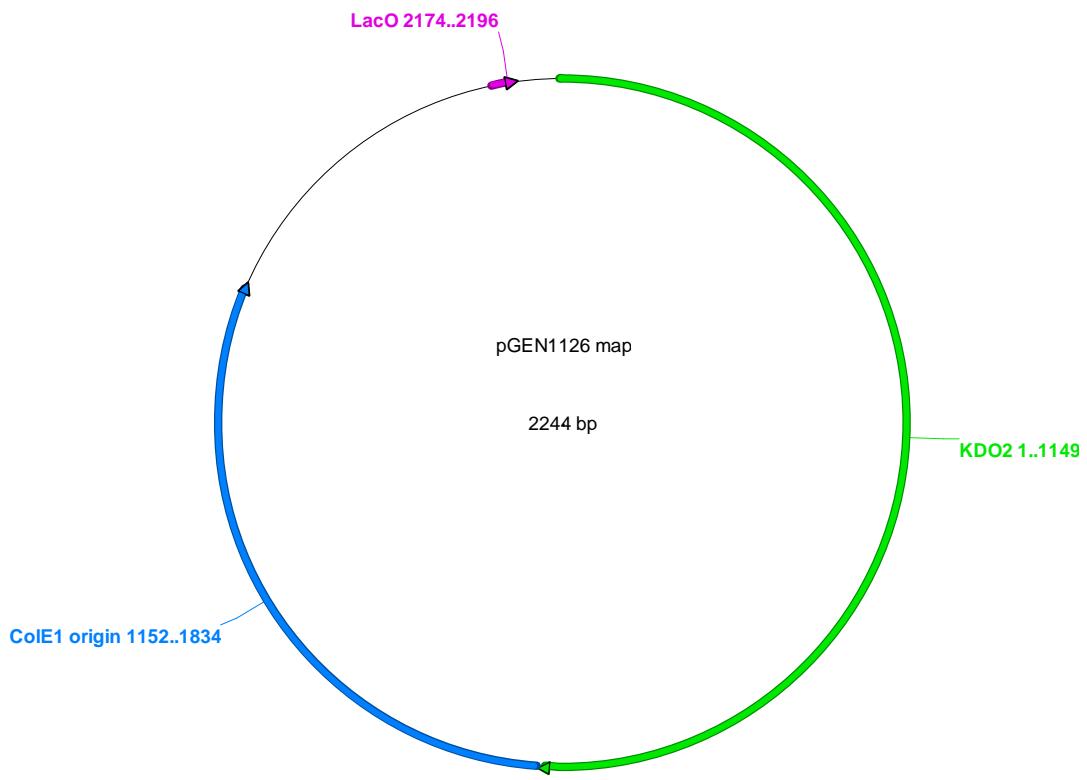


Figure 41: Map of plasmid pGEN1126

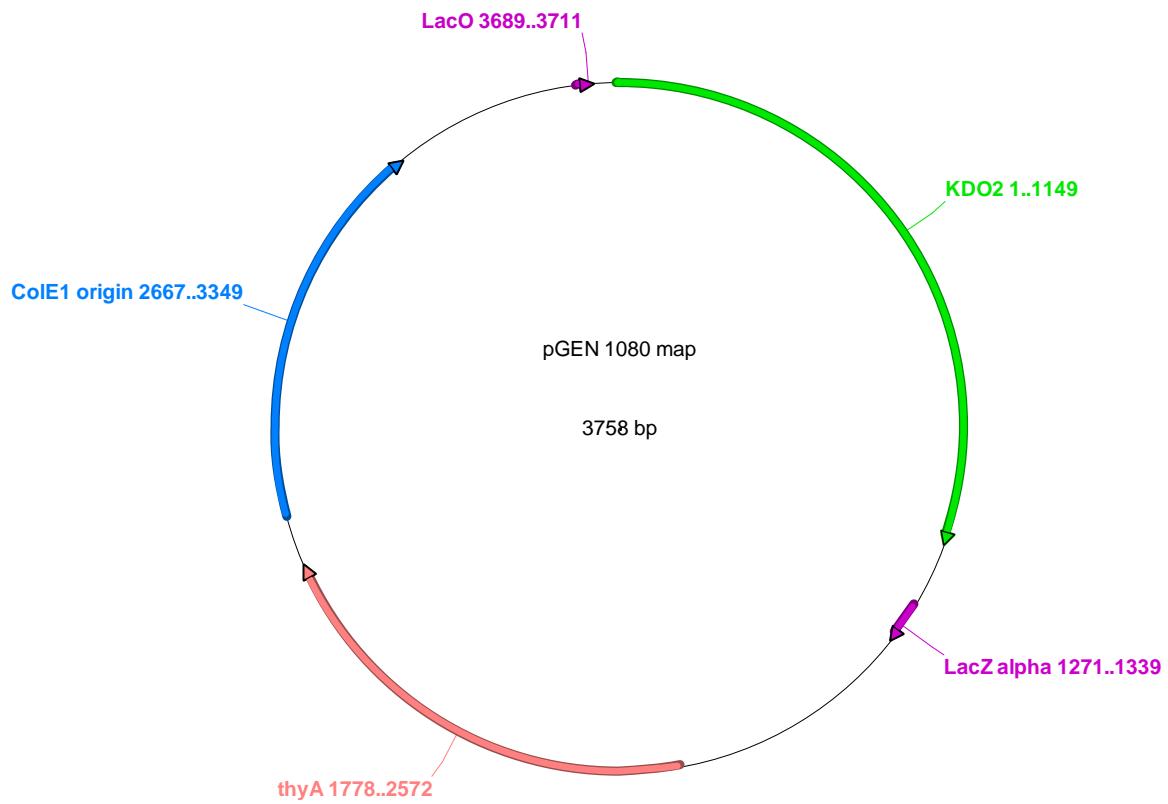


Figure 42: Map of plasmid pGEN1080

Imbalanced dNTP levels, which can be provoked by the overexpression of genes coding for enzymes involved in the biosynthesis of DNA precursor molecules, have been shown to be mutagenic (Kumar *et al.*, 2011; Kunz, 1988). Thus, it is probable that the expression of *thyA* on the plasmid might have limited its copy number. This limitation could have been released through excision of the gene *thyA* from the plasmid and its integration in the genome of the host strain. This might, in turn, have enabled the relaxation of the control of the copy number of the plasmid pGEN1126 up to a point where the number of *kdo2* gene copies would be sufficient to assure succinate complementation in the selection strain. This hypothesis was tested in two ways:

- Transformation of plasmid pGEN1126 into the selection strain G3700 and complementation test
- Determination of the copy number of the plasmid pGEN1126, the gene *thyA* and the gene *kdo2* in the strain G4097 (Figure 39)

III.4.1.3 Phenotypic analysis of isolates from culture KDB 10

Previous results showed that an isolate (strain G4097, Figure 39) from culture KDB 10 capable of succinate complementation contained plasmids with a 1500 bp excision (plasmid pGEN1126, Figure 41) compared to the initial plasmid (plasmid pGEN1080, Figure 42). The excised sequence contained the gene *thyA*.

Is succinate complementation phenotype related to the evolved plasmid pGEN1126?

To test the impact of the modified plasmid pGEN1126 on the phenotype, it was transformed back into the initial genetic screen (Figure 39). This allows focusing on the impact of the plasmid pGEN1126 on succinate complementation while excluding the influence of any other potential mutations in the genome that might have occurred during the adaptation in the GM3 device.

Since the plasmid pGEN1126 was lacking the gene *thyA*, the plasmid had to be transformed back into the selection strain G3700 (*thyA+*) instead of the strain G3752 (*ΔthyA*). Transformants had to be selected on their capacity to grow without succinate. Based on the hypothesis that the modification in the plasmid pGEN1126 is involved in the succinate complementation phenotype, it was theoretically possible to select transformants on plates containing MS mineral glc (0.2%) glu (5 mM) lys (5 mM) medium.

It was noted that transformation efficiency was unusually low for competent bacteria transformed by electroporation, suggesting that the transformants isolated might have been selected for other traits such as increased plasmid copy levels or different mRNA expression levels.

One resulting transformant, strain G4096 (G3700 pGEN1126, Figure 39) was tested for succinate complementation (Table 21).

Table 21: Succinate complementation tests for isolates from culture KDB 10

| Strain | Genotype | MS glc (0.2%) glu (5 mM) | | |
|--------|---|--------------------------|------------|--------------|
| | | Ø | lys (5 mM) | suc (0.8 mM) |
| G3700 | C> Δ <i>ArgF-lac</i> Δ <i>gltA</i> Δ <i>icd::aad+</i> Δ <i>pck</i> Δ <i>ppc</i> Δ <i>aceBAK</i> Δ <i>glcDEFGB</i> Δ <i>sfca</i> Δ <i>maeB</i> Δ <i>puuE</i> Δ <i>gabT</i> Δ <i>sucA</i> | - | - | +++ |
| G3752 | G3700 Δ <i>thyA::erm+</i> | - | - | - |
| G3986 | G3752 pGEN1080 (Figure 42) | - | - | ++ |
| G4097 | Isolate sampled from culture KDB 10 | - | +++ | +++ |
| G4096 | G3700 pGEN1126 (Figure 41) | - | +++ | +++ |

Complementation tests on solid medium were performed at 37°C for 48 h. The bacterial growth observed is indicated with signs for normal (+++), medium (++) residual (+) and no complementation (-).

The strains G4097 and G4096 both showed succinate complementation depending on the presence of lysine, strongly suggesting that an active Kdo2 enzyme was responsible for this phenotype. Given the absence of mutations within the coding region of the gene, an increase in the copy number of the plasmid and therefore an increase in the levels of *kdo2* expression is the most likely explanation for this result.

III.4.1.4 Plasmid and mRNA copy number analysis in isolates from culture KDB 10

In the hypothesis that the copy number of plasmid pGEN1126 – and therefore the number of *kdo2* gene copies in the cells - might have been impacted in strains G4097 and G4096, qPCR assays were performed on culture samples (as described in section II.6.5) to determine the *kdo2* gene copy number in the cells.

The results of the qPCR experiments are shown in Figure 43. Strain G4097 (isolate after evolution) shows a three-fold increase in the copy number of *kdo2* compared to the initial strain G3986. This observation is in accordance with the hypothesis that the level of expression of *kdo2* was increased in the evolved strain G4097.

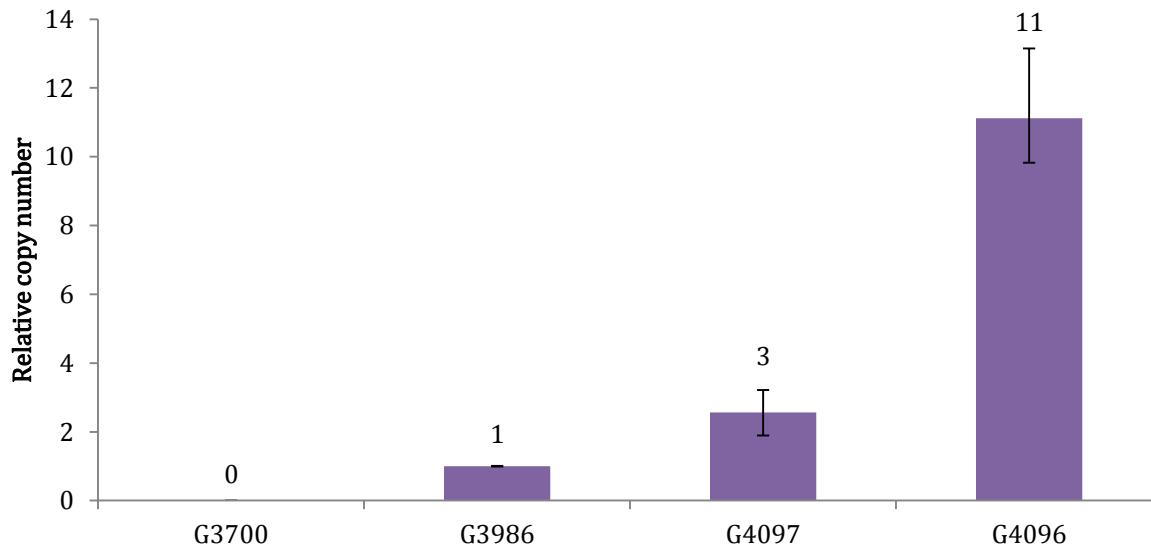


Figure 43: Copy number comparison of gene *kdo2* between strains G3700, G4096 and G4097 relative to strain G3986

Copy number of gene *kdo2* in strains G3700 (selection strain), G4096 (G3700 pGEN1126) and G4097 (isolate from culture KDB 10) relative to the initial strain G3986 (G3752 pGEN1080) before evolution in the GM3. Copy number was measured by qPCR and quantified as $2^{-\Delta\Delta Ct}$ normalized to 16S copy number; bars show means of triplicate measurements \pm sd.

The strain G4096 shows an 11-fold increase in the copy number of the gene *kdo2* compared to strain G3986. This is also almost four times more than the copy number of the gene *kdo2* in strain G4097. As mentioned previously, low transformation efficiency was observed when strain G3700 was transformed with plasmid pGEN1126 and selected for *kdo2* activity. It is likely that the transformants obtained exhibited a genetic predisposition for plasmid copy number upregulation.

To further investigate the adaptation process observed for culture KDO 10, *thyA* gene copies were quantified for the evolved and retransformed strains. Figure 44, shows that the strain G3986 contains around 138 times more copies of the gene *thyA* than the strain G3700 which has one copy in the chromosome. In the strain G4097, an isolate from culture KDB 10, the copy number of the gene *thyA* has considerably decreased compared to strain G3986 and is only two-fold higher than the strain G3700. This result corroborates the hypothesis that the gene *thyA* has been excised from the plasmid pGEN1080 and integrated into the genome of the evolved strain G4097.

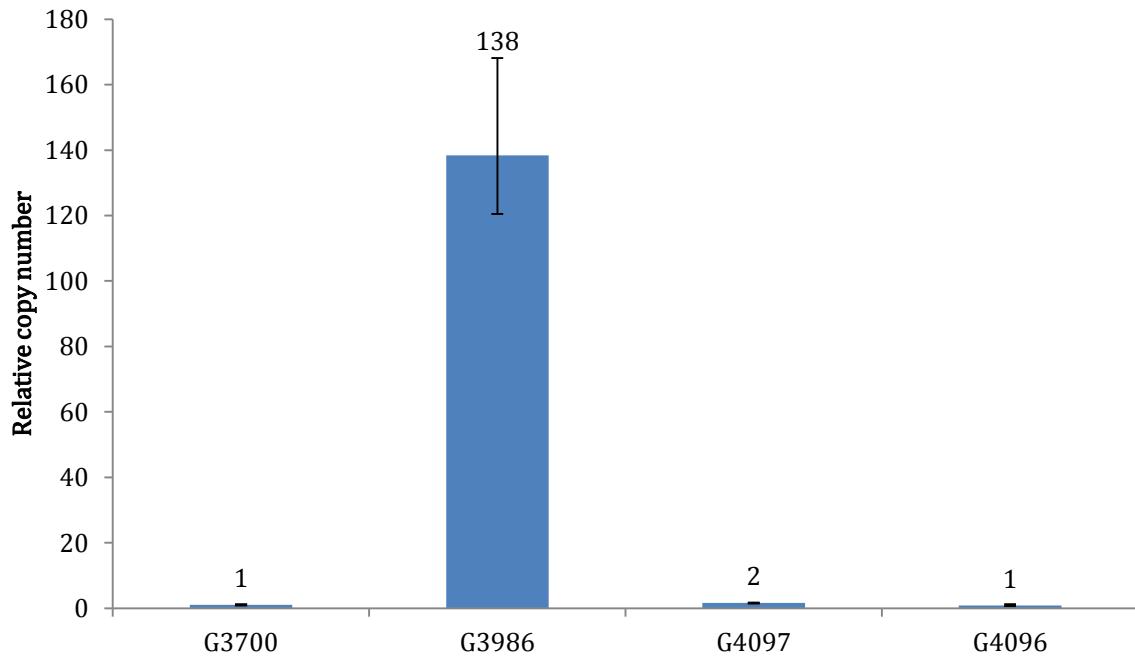


Figure 44: Copy number comparison of gene *thyA* between strains G3700, G4096 and G4097 relative to strain G3986

Copy number of gene *thyA* in strains G3700 (selection strain), G4096 (G3700 pGEN1126) and G4097 (isolate from culture KDB 10) relative to the strain G3986 (G3752 pGEN1080) before evolution in the GM3. Copy number was measured by qPCR and quantified as $2^{-\Delta\Delta Ct}$ normalized to 16S copy number; bars show means of triplicate measurements \pm sd.

To determine the actual level of expression of gene *kdo2*, RT-qPCR assays were also performed (Figure 45).

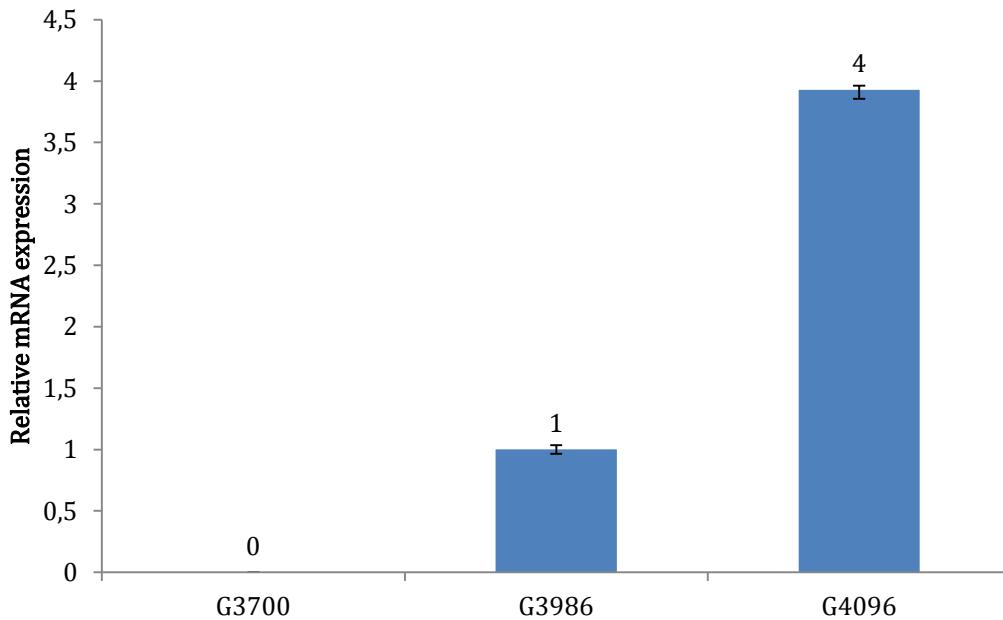


Figure 45: mRNA expression comparison of gene *kdo2* in strain G4096 relative to strain G3986

mRNA expression of gene *kdo2* in the strain G4096 (3700 pGEN1126) relative to strain G3986 (G3752 pGEN1080) before evolution in the GM3. Expression was measured by RT-qPCR and quantified as $2^{-\Delta\Delta Ct}$ normalized to 16S mRNA expression; bars show means of triplicate measurements \pm sd.

The RT-qPCR results show that the mRNA expression of gene *kdo2* in strain G4096 is around four times higher than in the initial strain G3986. This observation is consistent with the results of the phenotypic and qPCR analysis obtained for this strain.

We can conclude that:

- Succinate complementation results from an increase in the level of expression of the gene coding for Kdo2.
- The increase of gene *kdo2* expression is mediated by the increase of plasmid copy number.
- Overexpression of gene *thyA* was counter-selected, probably due to DNA precursor pool imbalances, leading to a *thyA* free plasmid.
- The expression of plasmid pEVL404 is unstable in a thymidine auxotroph and not suitable for directed evolution in continuous culture.

III.4.1.5 Isotopic labeling analysis of isolates from culture KDB 10

To confirm the *in vivo* activity of Kdo2 other than by growth phenotype of selection strains and the actual formation of succinate by α -KG decarboxylation, an isotopic labeling analysis was performed. The GM3-evolved strain G4097, as well as control strains, were grown in the presence of [1,2- $^{13}\text{C}_2$]-labeled glutamate. Glutamate being the direct precursor for α -KG, the succinate formed by the reaction through Kdo2 activity should be ^{13}C -labeled at one carbon atom.

Bacteria of the studied strains were grown in the two following media:

- Medium I: MS glc (0.2%) lys (5mM) [1,2- $^{13}\text{C}_2$]-glu (5 mM) suc (0.8 mM)
- Medium II: MS glc (0.2%) lys (5mM) [1,2- $^{13}\text{C}_2$]-glu (5 mM)

Metabolomes of cultures in the presence of [1,2- $^{13}\text{C}_2$]-labeled glutamate were analyzed by liquid chromatography–mass spectrometry (LC-MS). Results are shown in Figure 46.

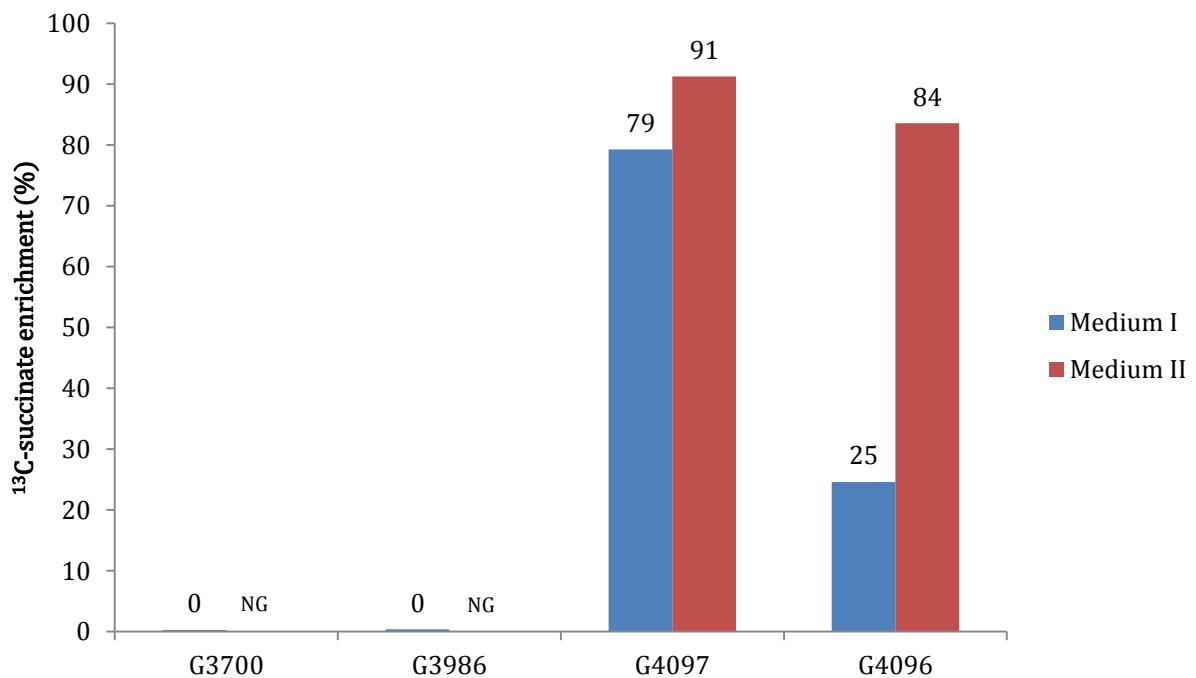


Figure 46: ^{13}C enrichment of succinate in metabolome of evolved strain G4097 and control strains

^{13}C -succinate enrichment in metabolome of strains G3700 (selection strain), G3986 (G3752 pGEN1080), G4096 (G3700 pGEN1126) and G4097 (isolate from culture KDB 10). Medium I (blue): MS glc (0.2%) lys (5mM) ^{13}C -glu (5 mM) suc (0.8 mM) and medium II (red): MS glc (0.2%) lys (5mM) ^{13}C -glu (5 mM). NG: No growth

The isotopic analysis of succinate shows significant ^{13}C -succinate enrichment in the metabolomes of the evolved strain G4097 and the transformed strain G4096 grown in the presence of succinate and labeled glutamate.

The dilution of ^{13}C -succinate by standard succinate is greater in the metabolome of the strain G4096. No enrichment is detected in the strain G3700 which lacks gene *kdo2*. This strongly suggests that Kdo2 is responsible for succinate production from α -KG. When strains G4097 and G4096 were grown in the absence of standard succinate the percentage of ^{13}C -succinate in succinate peak separated by liquid chromatography is even higher, although a maximal ratio would be expected.

No enrichment is detected in the metabolome of the strain G3986, although the latter harbors multiple copies of gene *kdo2* as indicated by qPCR quantification of the copy number of gene *thyA* (Figure 44). These results combined with the quantification of the *kdo2* transcript mRNA (Figure 45) suggests that the gene *kdo2* is weakly transcribed or repressed in strain G3986.

III.4.2 Evolution *in vivo* of α-KAOs expressed in a chromosomal format

We previously concluded from experimental observations that the plasmidic expression format pEVL404 was not suitable for the directed evolution of gene *kdo2*. To avoid difficulties linked to plasmid loss, the gene *kdo2*, which was found to complement the succinate dependence of the selection strain G3700 upon gene overexpression (section III.4.1.4), was inserted in the *kdgK* locus of the selection strain resulting in the strain G4160 (Figure 47).

This chromosomal locus has previously been used for the constitutive expression of heterologous genes in *E. coli*. In accordance with the results from plasmidic expression formats, succinate complementation tests were negative for the chromosomal format (Table 18).

The rational and final goals of evolving strain G4160 in continuous culture was the same as for the plasmid-bearing strains:

1. Evolution towards succinate complementation with lysine by selecting either higher expression levels of the gene *kdo2* or a mutated variant of the gene *kdo2* with improved kinetic parameters.
2. Evolution towards lysine analogs to broaden the scope of application of Kdo2 to new substrates.

Given the absence of complementation, the medium SWAP was the continuous culture regime of choice to evolve the strain G4160. A general overview of the experimental steps involved in the evolution of the α-KAO Kdo2 in chromosomal format is depicted in Figure 47.

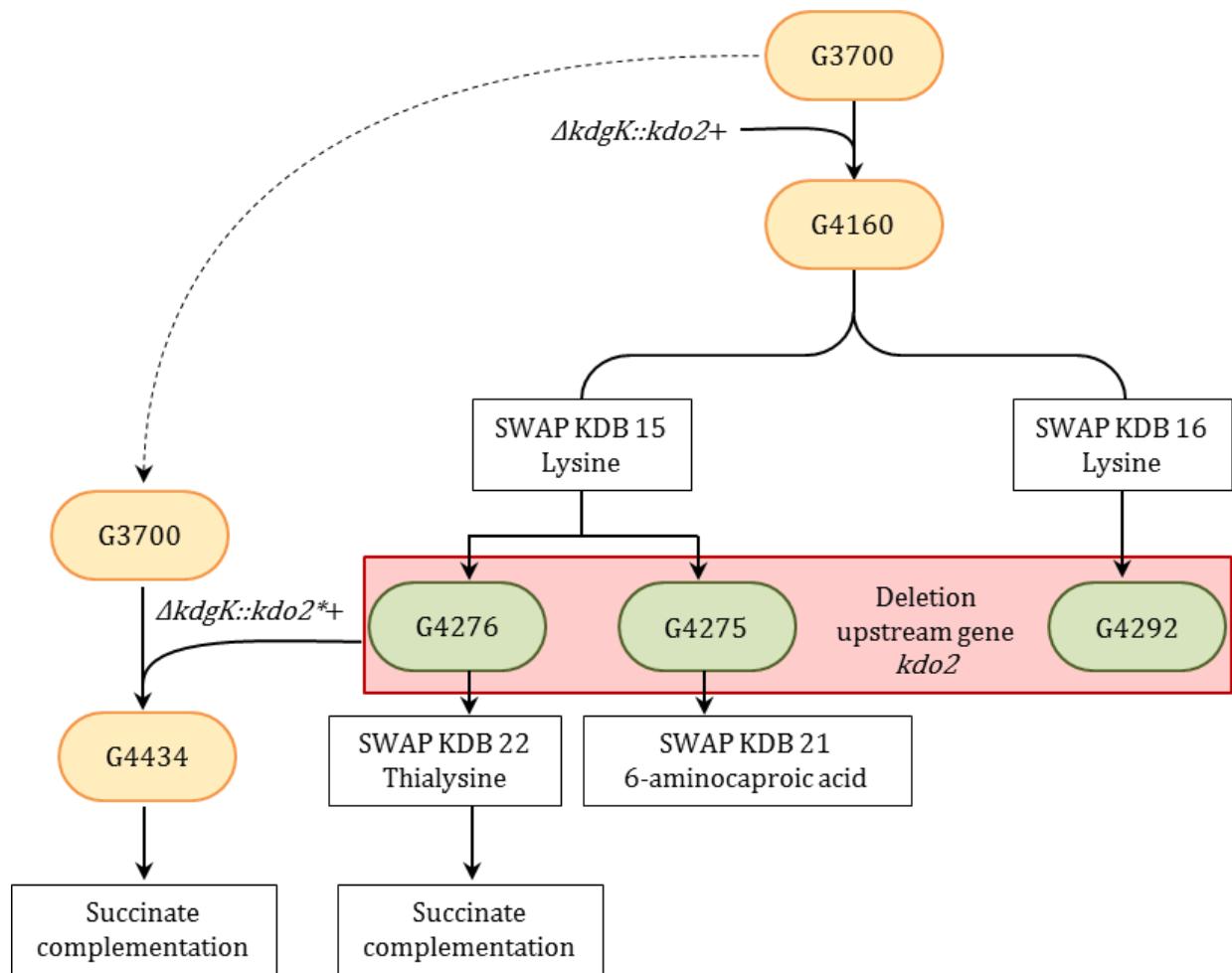


Figure 47: Overview of Kdo2 evolution in chromosomal expression format

Strains are represented by an oval shape. Green ovals represent strains that have evolved in the GM3. Strains embedded in a red box contain a deletion upstream of the gene *kdo2*. The star symbol (*) indicates that the gene *kdo2* has a deletion upstream of its coding sequence (*kdo2**). For all medium SWAP regimes, generation time was fixed at 2 h, and growth temperature at 37°C.

III.4.2.1 Evolution of Kdo2 in medium SWAP regime

Since the strain G4160 showed no succinate complementation in the presence of lysine (Table 18), a medium SWAP regime was imposed (Figure 48) with a relaxing medium containing succinate.

Conditions of the medium SWAP regime for the cultures KDB 15 and KDB 16 (strain G4160 grown with lysine) were as follows:

- Relaxing medium: MS glc (0.2%) glu (5 mM) lys (5 mM) suc (0.8 mM)
- Stressing medium: MS glc (0.2%) glu (5 mM) lys (5 mM)
- Generation time 2 h at 37°C

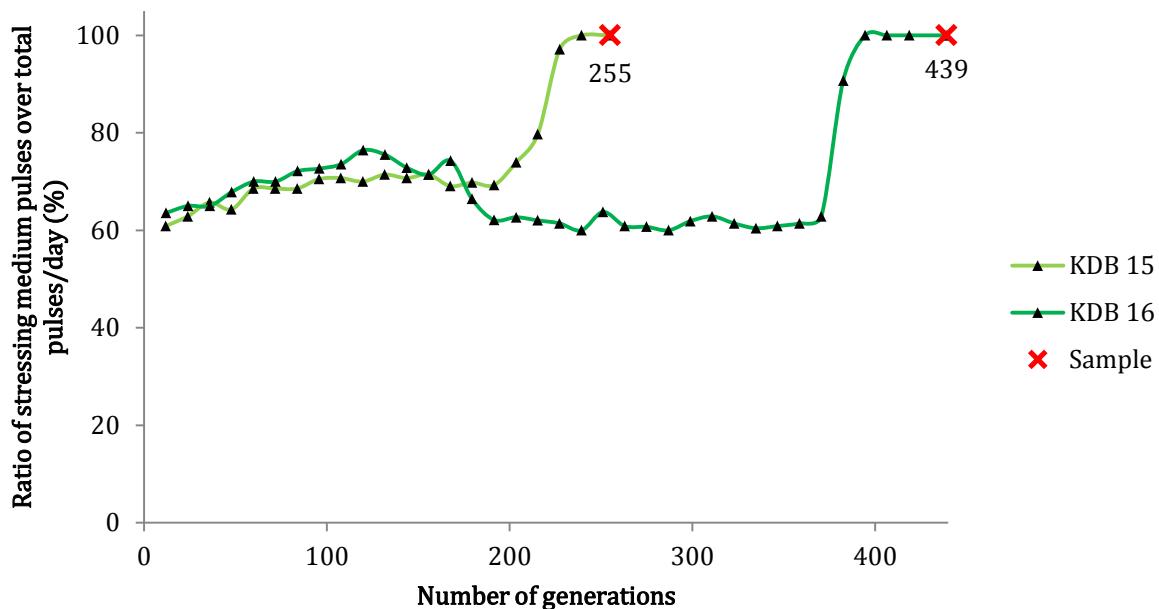


Figure 48: Evolution of cultures KDB 15 and KDB 16 in medium SWAP regime

Shown is the ratio of stressing medium dilution pulses per day in medium SWAP regime. The relaxing medium was MS glc (0.2%) glu (5 mM) lys (5 mM) suc (0.8 mM) and the stressing medium was MS glc (0.2%) glu (5 mM) lys (5 mM). The generation time was fixed at 2 h, and growth temperature at 37°C. Crosses indicate sampling points with the corresponding number of generations.

Culture of strain G4160 (G3700 *ΔkdgK::kdo2+*) is in light green for culture KDB 15 and dark green for KDB 16

It took around 240 generations to reach 100% of stressing medium pulses for culture KDB 15 and around 400 generations for culture KDB 16. Culture KDB 15 was sampled at generation 255 and culture KDB 16 at generation 439.

III.4.2.2 Phenotypic analysis of isolates from culture KDB 15 and culture KDB 16

Isolates obtained from cultures KDB 15 and KDB 16 at the indicated time points were tested for succinate complementation in the presence of lysine. Results are shown in Table 22.

Table 22: Succinate complementation tests of isolates from cultures KDB 15 and KDB 16

| Strain | Genotype | Growth in MS glc (0.2%) glu (5 mM) | | |
|--------|--|------------------------------------|------------|-------------------------|
| | | Ø | lys (5 mM) | Lys (5 mM) suc (0.8 mM) |
| G3700 | C> <i>ΔArgF-lac ΔgltA Δicd::aad+ Δpck Δppc</i> <i>ΔaceBAK ΔglcDEFGB ΔsfCA ΔmaeB ΔpuuE</i> <i>ΔgabT ΔsucA</i> | - | - | +++ |
| G4160 | G3700 <i>ΔkdgK::kdo2+</i> | - | - | +++ |
| G4275 | Isolate sampled from culture KDB 15 | - | +++ | +++ |
| G4276 | Isolate sampled from culture KDB 15 | - | +++ | +++ |
| G4292 | Isolate sampled from culture KDB 16 | - | +++ | +++ |

Complementation tests on solid medium were performed at 37°C for 48 h. The bacterial growth observed is indicated with signs for normal (+++), medium (++) residual (+) and no complementation (-).

In the absence of succinate, the growth of the isolates was dependent on the presence of lysine in the medium.

III.4.2.3 Sequence analysis of isolates from culture KDB 15 and culture KDB 16

Sequencing of the gene *kdo2* revealed a WT coding sequence. However, a modification upstream of the gene *kdo2* was detected for all three isolates G4275, G4276 and G4292 (Figure 49).

Taking the ATG START codon of gene *kdo2* as a reference, a deletion of 14 bp was detected from position -322 to -336 in isolates G4275 and G4276, from culture KDB 15. In isolate G4292 from culture KDB 16, a deletion of 3 bp was detected, overlapping the region deleted in strains G4275 and G4276 from position -333 to -336. An overlapping deletion found in independent isolates from separate cultures strongly suggests that this modification is involved in an evolutionary adaptation.

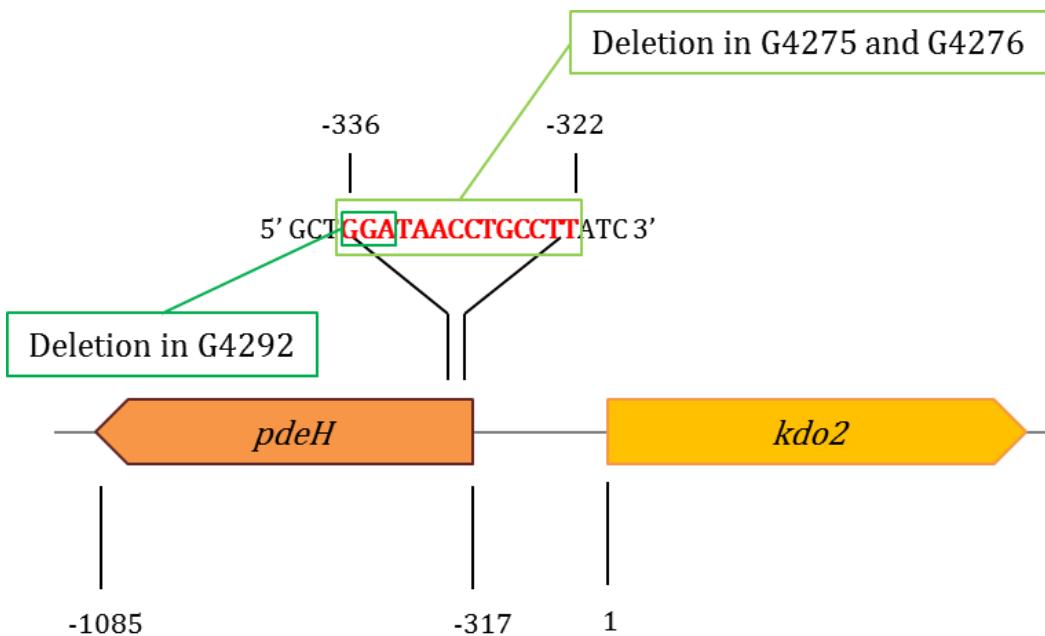


Figure 49: Scheme of upstream modifications detected in isolates G4275, G4276 and G4292 from culture KDB 15 and KDB 16

Isolates G4275 and G4276 from culture KDB 15 with 14 bp deletion at position -322 to -336 and isolate G4292 from culture KDB 16 with 3 bp deletion at position -333 to -336.

In Figure 49, we can see that the deletions are located at the beginning of the gene *pdeH* (GenBank: [AO071725](#)).

What could be the consequences of a 3 or 14 bp deletion in the gene *pdeH*?

In the case of the 3 bp deletion, we observe the substitution of 2 codons (ATC-CAG) by one codon (AAG) that might result in an altered activity of PdeH protein. The 14 bp deletion causes a frameshift at the beginning of the *pdeH* sequence suggesting a complete loss of function of PdeH. PdeH is a c-di-GMP phosphodiesterase that mediates the (3'-5')-cyclic-diguanosine monophosphate (c-di-GMP) degradation through the reaction in Figure 50. The loss of function of PdeH might thus result in increased cellular levels of c-di-GMP.

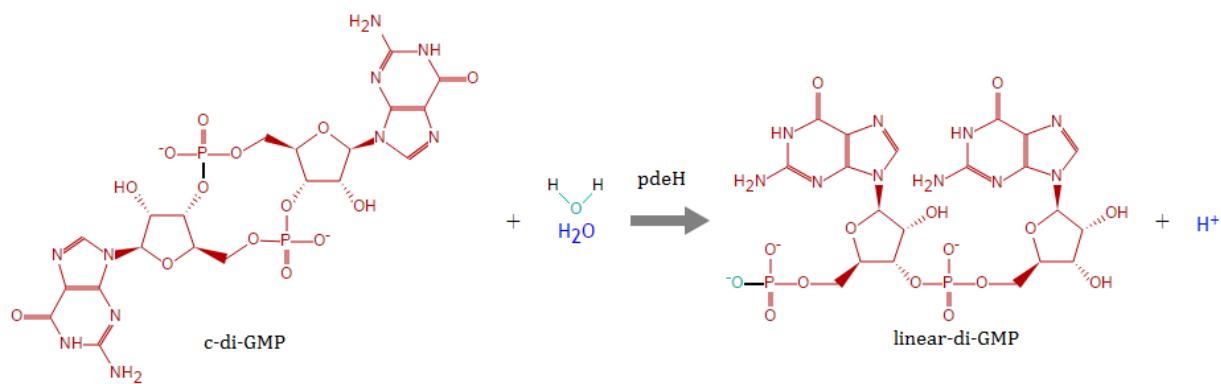


Figure 50: Reaction catalyzed by *pdeH*

C-di-GMP is a bacterial second messenger involved in the regulation of a number of complex physiological processes including cell-cell signaling, biofilm formation, motility, differentiation, virulence, etc (Olsén *et al.*, 1989; Zhao *et al.*, 2007; Pesavento *et al.*, 2008). It has also been shown to control gene expression (transcription and translation), activity, stability, localization and interaction of specific proteins (Jenal *et al.*, 2017). Cellular levels of c-di-GMP are regulated in response to internal and environmental cues.

Overall, the fact that c-di-GMP is involved in various important biological processes suggests that the deletion in the gene *pdeH*, probably resulting in PdeH function loss and thus increased levels of c-di-GMP, might have an impact (directly or indirectly) on the expression of gene *kdo2*. Alternatively, the deletions upstream of the gene *kdo2* might have a cis-acting effect through altering expression levels of gene *kdo2*.

In order to test this hypothesis, the mRNA expression levels of *kdo2* were determined for the initial strain G4160, and the evolved isolate G4275 by RT-qPCR and compared.

III.4.2.4 mRNA expression of Kdo2 in isolate from culture KDB 15

In the hypothesis that the aforementioned genetic event, *i.e.* short deletions might have impacted the expression of *kdo2* in the evolved strains, RT-qPCR assays were performed (Figure 51).

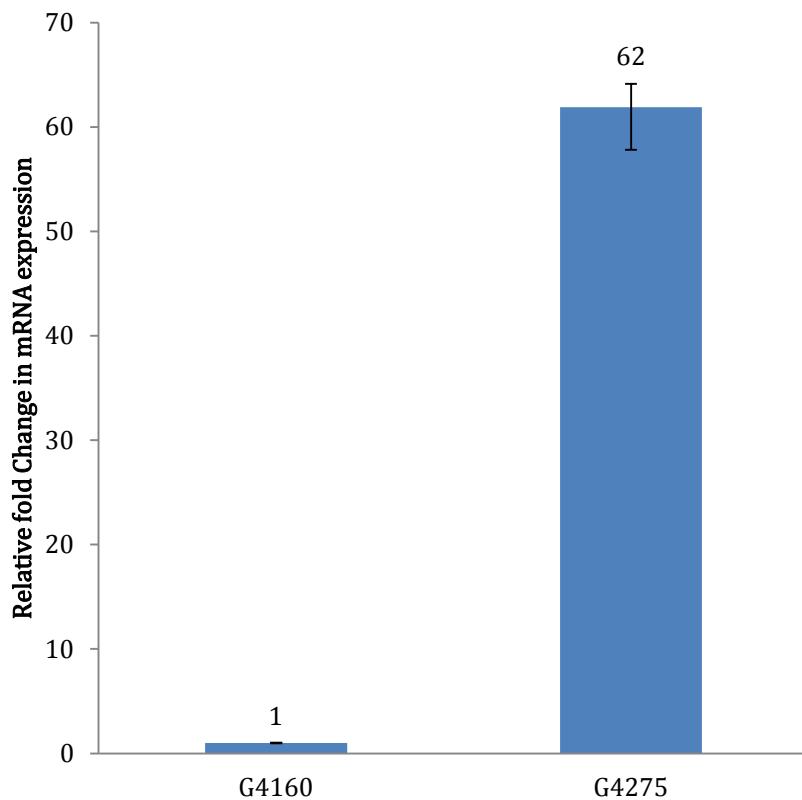


Figure 51: Comparison of mRNA expression level of *kdo2* in isolate G4275 after evolution relative to the initial strain G4160

Strain G4160 (G3700 $\Delta kdgK::kdo2+$) and strain G4275 (isolate from culture KDB 15). Expression was measured by RT-qPCR and quantified as $2^{-\Delta\Delta Ct}$ normalized to 16S mRNA expression; bars show means of triplicate measurements \pm sd.

The results of the RT-qPCR experiments show that the mRNA expression of Kdo2 in the evolved isolate G4275 is 62 times higher than in the initial strain G4160. This observation confirms that the expression of Kdo2 has been upregulated. As seen previously with the plasmidic format (III.4.1.4), succinate complementation in the presence of lysine could result from an increase in the level of expression of *kdo2*. However, other mutations elsewhere in the genome can also have an impact. To determine the actual influence of the short deletion upstream of the gene *kdo2*, the deletion was introduced in the initial strain G3700 (Figure 47) via transduction.

III.4.2.5 Transduction of gene *kdo2* with the upstream deletion in the initial strains

The gene *kdo2* along with the upstream 14 bp deletion (noted *kdo2**) was transduced back in the initial strain G3700, resulting in the strain G4434 (G3700 *ΔkdgK::kdo2**+) (Figure 47). Succinate complementation tests were performed to determine if the upstream deletion was sufficient to reestablish succinate prototrophy (Table 23).

Table 23: Succinate complementation test of strain G4434

| Strain | Genotype | Growth in MS glc (0.2%) glu (5 mM) | | |
|--------|--|------------------------------------|------------|-------------------------|
| | | Ø | lys (5 mM) | lys (5 mM) suc (0.8 mM) |
| G3700 | C> <i>ΔArgF-lac ΔgltA Δicd::aad+ Δpck Δppc</i> <i>ΔaceBAK ΔglcDEFGB ΔsfCA ΔmaeB ΔpuuE</i> <i>ΔgabT ΔsucA</i> | - | - | +++ |
| G4434 | G3700 <i>ΔkdgK::kdo2*</i> + | - | ++ | +++ |

Complementation tests on solid medium were performed at 37°C for 48 h. The bacterial growth observed is indicated with signs for normal (+++), medium (++) residual (+) and no complementation (-).

The strain G4434 shows succinate prototrophy in the presence of lysine. This observation confirms that the locus harboring the 14 bp deletion upstream of the gene *kdo2* and the gene *kdo2* are necessary and sufficient for triggering the succinate complementation phenotype.

The present evolution experiments led to the selection of mutants capable of succinate complementation. These isolates have evolved towards lysine, the natural substrate of Kdo2, for which an increase in the expression level of the gene *kdo2* was sufficient to enable succinate complementation. These isolates can be used as starting points for evolution towards new substrates.

III.4.3 Evolution *in vivo* of Kdo2 towards lysine analogs

α-KAOs have been reported to be very stereospecific and stereoselective (Baud *et al.*, 2014). Evolution of α-KAOs towards new substrates could provide new synthons of interest for the chemical and pharmaceutical industry.

This section will focus on the evolution of Kdo2 towards two lysine analogs, thialysine and 6-aminocaproic acid (Table 24). An *in vitro* substrate promiscuity survey revealed that Kdo2 has some activity towards thialysine (Baud *et al.*, 2014).

Table 24: Lysine analogs tested for evolution

| Name | Chemical formula | Structural formula |
|---------------------|--|--------------------|
| L-Lysine | C ₆ H ₁₄ N ₂ O ₂ | |
| 6-aminocaproic acid | C ₆ H ₁₃ NO ₂ | |
| Thialysine | C ₅ H ₁₂ N ₂ O ₂ S | |

III.4.3.1 Phenotypic analysis with lysine analogs

Toxicity tests were performed on strain G4275 in order to calibrate the culture conditions in the GM3 (Figure 52).

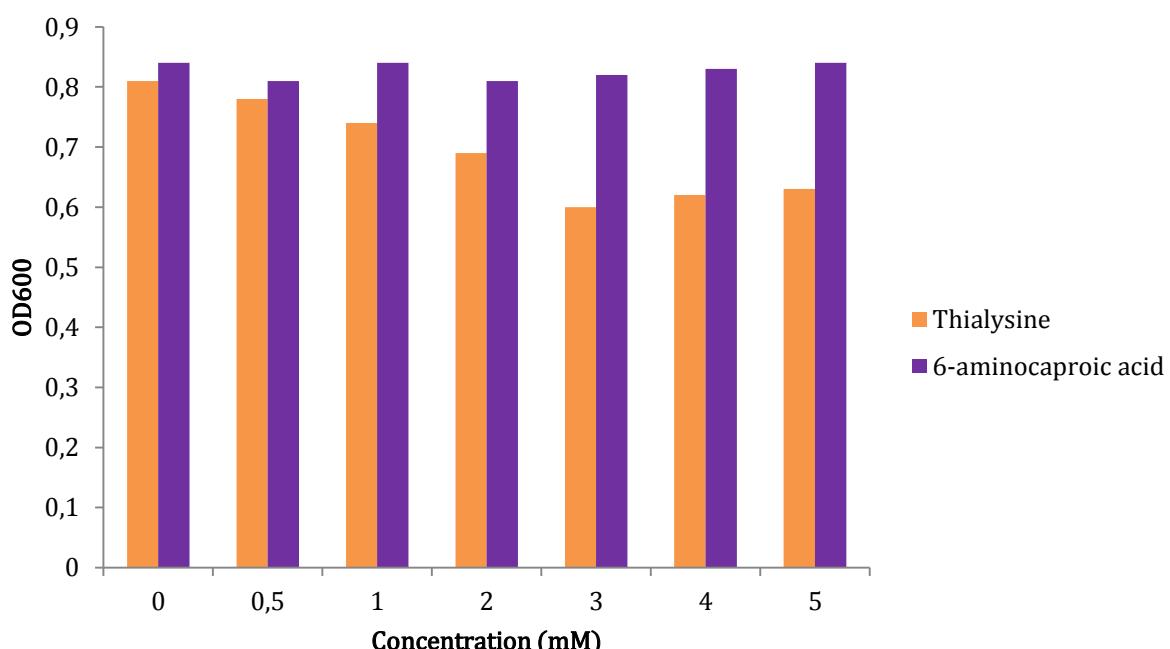


Figure 52: Growth of strain G4275 in the presence of lysine analogs

Medium: MA glc (0.2%) glu (5 mM) suc (5 mM) thialysine (orange) or 6-aminocaproic acid (purple), at 37°C for 24 h with agitation.

The lysine analog 6-aminocaproic acid did not show toxicity at the concentrations tested while thialysine slightly diminished growth above 3 mM. For both analogs, a concentration of 1.5 mM was fixed for evolution experiments. At that concentration, no succinate complementation with these compounds was observed (Table 25).

Table 25: Succinate complementation tests of strains G4275 and G4276 with lysine analogs

| Strain | Genotype | Growth in MA glc (0.2%) glu (5 mM) | |
|--------|--|------------------------------------|-------------------------------|
| | | Thialysine (1.5 mM) | 6-amino Caproic acid (1.5 mM) |
| G3700 | C> ΔArgF-lac ΔgltA Δicd::aad+ Δpck Δppc ΔaceBAK ΔglcDEFGB Δsfca ΔmaeB ΔpuuE ΔgabT ΔsucA | - | - |
| G4160 | G3700 ΔkdgK::kdo2+ | - | - |
| G4275 | Isolates sampled from culture KDB 15 | - | - |
| G4276 | Isolates sampled from culture KDB 15 | - | - |

Complementation tests on solid medium were performed at 37°C for 48 h. The bacterial growth observed is indicated with signs for normal (+++), medium (++) residual (+) and no complementation (-).

III.4.3.2 Evolution *in vivo* of Kdo2 towards lysine analogs in medium SWAP regime

Since neither the isolate G4275 nor G4276 showed succinate complementation in the presence of thialysine or 6-aminocaproic acid (Table 25), a medium SWAP regime was imposed for evolution towards these compounds (Figure 53).

Conditions of the medium SWAP regime for culture KDB 21 (strain G4275 grown with 6-aminocaproic acid, Figure 53, purple line) were as follows:

- Relaxing medium: MS glc (0.2%) glu (5 mM) 6-aminocaproic acid (1.5 mM) suc (0.8 mM)
- Stressing medium: MS glc (0.2%) glu (5 mM) 6-aminocaproic acid (1.5 mM)
- Generation time 2 h at 37°C

Conditions of the medium SWAP regime for culture KDB 22 (strain G4276 grown with thialysine, Figure 53, orange line) were as follows:

- Relaxing medium: MS glc (0.2%) glu (5 mM) thialysine (1.5 mM) suc (0.8 mM)
- Stressing medium: MS glc (0.2%) glu (5 mM) thialysine (1.5 mM)
- Generation time 2 h at 37°C

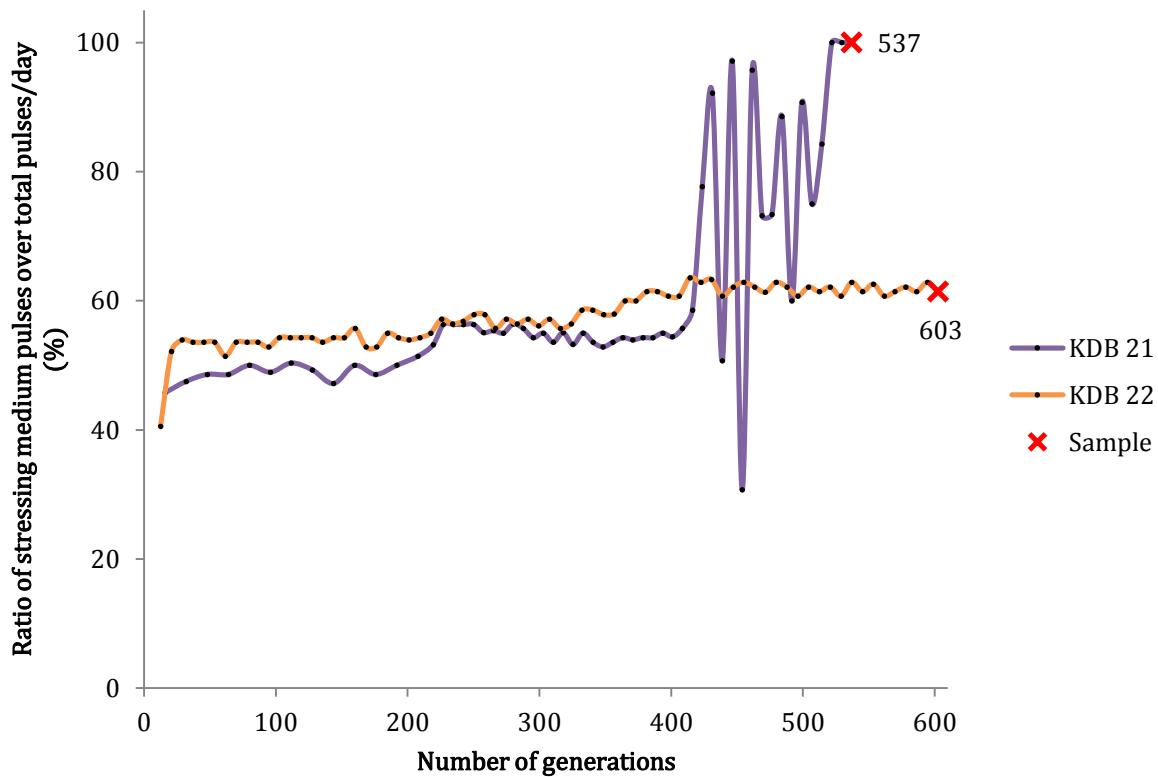


Figure 53: Evolution of culture KDB 21 and KDB 22 in medium SWAP regime

Shown is the ratio of stressing medium dilution pulses per day in medium SWAP regime. The generation time was fixed at 2 h, and growth temperature at 37°C. Crosses indicate sampling points with the corresponding number of generations.

Culture KDB 21 of strain G4275: **purple** line. Relaxing medium was MA glc (0.2%) glu (5 mM) 6-aminocaproic acid (1.5 mM) suc (0.8 mM) and stressing medium was MA glc (0.2%) glu (5 mM) 6-aminocaproic acid (1.5 mM).

Culture KDB 22 of strain G4276: **orange** line. Relaxing medium was MA glc (0.2%) glu (5 mM) thialysine (1.5 mM) suc (0.8 mM) and stressing medium was MA glc (0.2%) glu (5 mM) thialysine (1.5 mM).

For about 600 generations, the culture KDB 22 (strain G4276 with thialysine) remained stable with only a slight increase of the ratio of stressing medium pulses called in the growth vessel. Sequencing analysis of isolates obtained from samples after 600 generations did not show mutations in the gene *kdo2*. The growth yield of these isolates grown in test tubes with limiting succinate did not augment upon addition of thialysine in the medium, demonstrating that the stressing medium ratio increase was not due to a positive thialysine effect. By contrast, lysine could no longer complement the succinate auxotrophy of the isolates (data not shown). This observation suggests a loss of function of the gene *kdo2* in the evolved descendants of strain G4276 as a response to the toxicity of thialysine. An alternative explanation would be the loss of uptake capacity of the cells for lysine and its homologs.

The culture KDB 21 (strain G4275 with 6-aminocaproic acid) showed an unusual evolution pattern. A rapid increase at generation 400 in the amount of stressing medium pulses was followed by an oscillating behavior spanning for 100 generations before reaching 100% of stressing medium in the vessel around generations 520. A sample of the culture was streaked on solid medium composed of the ingredients of the stressing medium, but no clone could be isolated and stored. Sequencing analysis of culture samples did not show any mutation in the gene *kdo2*.

The evolution experiments of Kdo2 towards lysine analogs did not yield to variants which was the goal of these experiments. We decided not to pursue our efforts trying other lysine analogs, but to study the evolving capacity of another member of the α -KAO enzyme family.

III.5 L-isoleucine dioxygenase (Ido) in a *ppc-pck*-selection context

The L-isoleucine dioxygenase (Ido) belongs to the family of the α -KAOs and catalyzes the following reaction (Figure 54).

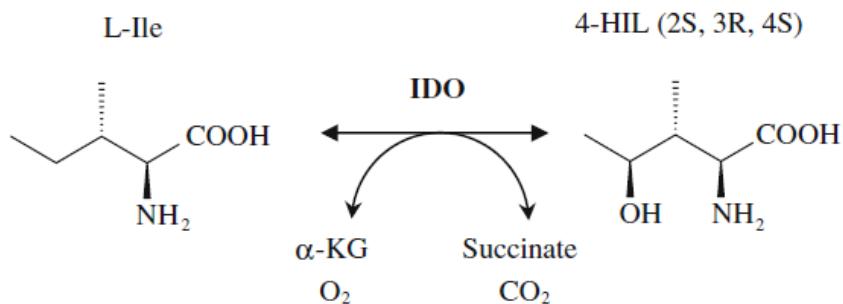


Figure 54: Reaction catalyzed by Ido

Reprinted from (Smirnov *et al.*, 2010)

Smirnov *et al.*, reported the expression of Ido from *Bacillus thuringiensis* in an *E. coli* strain named 2 Δ which lacked the activities of α -KG dehydrogenase (Δ sucAB) and isocitrate lyase (Δ aceAK) (Smirnov *et al.*, 2010). The 2 Δ strain also lacked the succinyl-CoA necessary for lysine, diaminopimelate and methionine synthesis. The expression of Ido on a high copy number plasmid restored the growth in mineral medium (M9-salts glucose (137 mM) glycerol (136 mM)) supplemented with L-isoleucine (137 mM) (Figure 55).

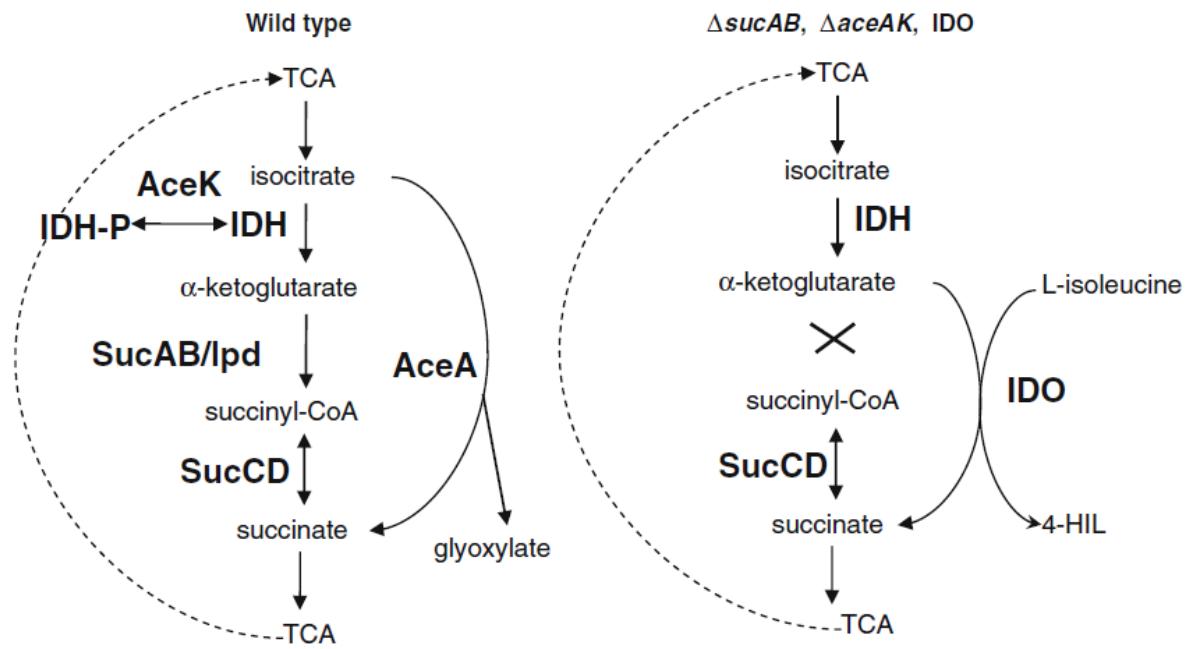


Figure 55: "Shunting" of the TCA cycle in the 2Δ strain expressing Ido

Reprinted from (Smirnov *et al.*, 2010)

This experiment showed that the need for succinate of the 2Δ strain could be complemented by the activity of a α -KAO expressed on a high copy number plasmid. However, the 2Δ strain is not suitable for directed evolution *in vivo* in continuous culture as this genetic screen is not tight. As it appears in Smirnov *et al.* 2010, the 2Δ bacteria were able to proliferate in M9 glucose medium without isoleucine after prolonged time of incubation.

We expressed Ido in the genetic screen designed in this study to test whether succinate complementation was possible through the activity of a α -KAO. To be suitable for evolving Ido in continuous culture, the screen had to be modified.

III.5.1 Genetic screen construction for Ido

For the expression of Ido in our genetic screen, the additional deletion of the ilv operon was preferable to avoid interference in the biosynthesis of branched-chain amino acids by the addition of exogenous isoleucine in the medium (Szentirmai and Horváth, 1976). Synthesis of branched-chain amino acids in *E. coli* K12 is controlled by end-product inhibition:

- Isoleucine inhibits threonine deaminase
- Valine inhibits acetohydroxy acid synthetase
- Leucine inhibits α -isopropylmalate synthetase

Thus the ilv operon was deleted from the initial selection strain G3513 and all branched-chain amino acid added to the medium (Table 26). The genotype of the resulting strain G3990 was confirmed by PCR, and phenotypic tests confirmed succinate, isoleucine, leucine and valine auxotrophies.

Table 26: Genotype of the selection strain G3990 used for Ido evolution

| Strain | Genotype | Phenotype |
|--------|--|--|
| G3513 | C> $\Delta ArgF-lac$ $\Delta gltA$ $\Delta icd::aad+$ Δpck Δppc $\Delta aceBAK$ $\Delta glcDEFGB$ $\Delta sfcA$ $\Delta maeB$ $\Delta puuE$ $\Delta gabT$ $\Delta sucA::kan+$ | Succinate auxotroph |
| G3990 | C> $\Delta ArgF-lac$ $\Delta gltA$ $\Delta icd::aad+$ Δpck Δppc $\Delta aceBAK$ $\Delta glcDEFGB$ $\Delta sfcA$ $\Delta maeB$ $\Delta puuE$ $\Delta gabT$ $\Delta sucA$ $\Delta ilv::Cm+$ | Succinate, isoleucine, leucine and valine auxotroph |

III.5.2 Complementation tests of selection strain expressing Ido

The gene *ido* (GenBank ref: [HM358019](#)) used by Smirnov and colleagues (Smirnov *et al.*, 2010), was cloned from the *Bacillus thuringiensis* strain 2-e-2 (acc. no. AKU 0251) from the culture collection of the Faculty of Agriculture, Kyoto University (Kodera *et al.*, 2009).

Using the Blast research algorithm, we found a gene from *Bacillus cereus* ATCC 14579 (GenBank ref: [AE016877](#)) showing 98% sequence identity with *ido* from *Bacillus thuringiensis* strain 2-e-2. The DNA of this strain was already available in the Genoscope genome library (DSMZ ref: [DSM31T](#)).

The gene *ido* from *Bacillus cereus* was PCR amplified (primers Table 4), cloned into the plasmid pSP100 (Figure 10) and transformed into the selection strain G3990 (Table 26). The gene *ido* was also inserted into the *kdgK* locus of the selection strain G3990 to obtain a one-copy expression format. Succinate complementation tests were performed with the resulting strains G4136 and G4110 (Table 27).

Table 27: Succinate complementation tests with Ido

| Strain | Genotype | MS glc (0.2%) glu (5 mM) leu val (0.3 mM) | | |
|--------|---|---|------------|-------------------------|
| | | Ø | ile (5 mM) | ile (5 mM) suc (0.8 mM) |
| G3990 | C>Δ <i>ArgF-lac</i> Δ <i>gltA</i> Δ <i>icd::aad+</i> Δ <i>pck</i> Δ <i>ppc</i> Δ <i>aceBAK</i> Δ <i>glcDEFGB</i> Δ <i>sfcA</i> Δ <i>maeB</i> Δ <i>puuE</i> Δ <i>gabT</i> Δ <i>sucA</i> Δ <i>ilv::Cm+</i> | - | - | +++ |
| G4136 | G3990 pGEN997 (pSP100:: <i>ido</i> +) - | - | +++ | ++ |
| G4110 | G3990 Δ <i>kdgK::ido</i> + | - | - | +++ |

Complementation tests on solid medium were performed at 37°C for 48 h. The bacterial growth observed is indicated with signs for normal (+++), medium (++) residual (+) and no complementation (-).

Succinate complementation in the presence of isoleucine in the growth medium was observed when Ido was expressed on a high copy number format, demonstrating that the α-KAO activity of Ido was able to supply sufficient succinate for the growth of the selection strain in this context. Expression from a single chromosomal copy was insufficient for complementation. The result obtained with strain G4136 is in contrast to those obtained with the lysine-depending α-KAOs studied so far (Table 18), where complementation with a pSP100-derived plasmid required secondary events in the cells leading to an increase in plasmid copy number well above the usually observed levels.

III.5.3 Evolution *in vivo* of Ido expressed in a chromosomal format

Since the strain G4136 (G3990 pGEN997 (pSP100::*ido*)) already showed normal growth through succinate complementation with isoleucine (Table 27), evolution experiments towards a more efficient isoleucine hydroxylation in the GM3 using this expression format were of little interest. By contrast, strain G4110 (G3900 $\Delta kdgK::ido$), expressing Ido in a chromosomal format, showed no succinate complementation in the presence of isoleucine (Table 27). This strain was thus suitable for directed evolution *in vivo* towards a better isoleucine hydroxylation.

A general overview of the experimental steps for the evolution of Ido in the chromosomal format is depicted in Figure 56.

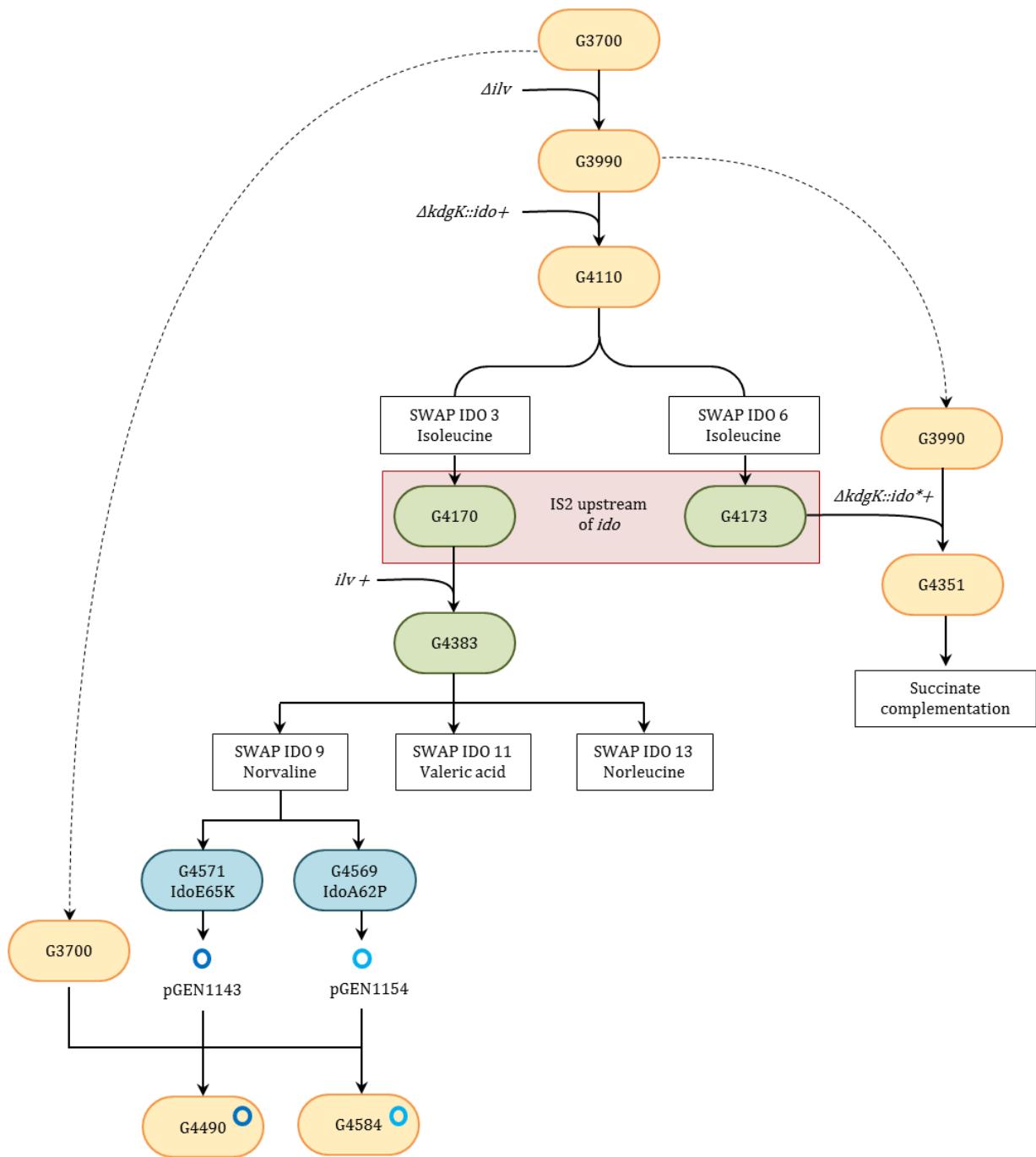


Figure 56: Overview of the evolution of Ido in the chromosomal expression format

Strains are represented by an oval shape. Green ovals represent strains that have evolved in the GM3. Blue ovals represent strains that have evolved in the GM3 twice on two different media. Strains containing an IS2 upstream of the gene *ido* are embedded in a red box. The star symbol (*) indicates the presence of the IS2 element (*ido**). Plasmids are represented by a circle: pGEN1143 (pSP100::*ido*:*E65K*+) (dark blue), pGEN1154 (pSP100::*ido*:*A62P*+) (light blue). For all medium SWAP regimes, generation time was fixed at 2 h, and growth temperature at 37°C.

III.5.3.1 Evolution *in vivo* of Ido in medium SWAP regime

The strain G4110 was inoculated in the GM3 device, and a medium SWAP regime experiment started to adapt the cells to grow on isoleucine without succinate (Figure 57). To lower the stringency of the selection, two of the essential metabolites derived from succinate, L-methionine (met) and diaminopimelic acid (DAP), were added to the media together with the aliphatic amino acids valine (val) and leucine (leu).

Conditions of the medium SWAP regime for the cultures IDO 3 and IDO 6 (strain G4110 grown with isoleucine) were as follows:

- Relaxing medium: MS glc (0.2%) glu (5 mM) val (0.3 mM) leu (0.3 mM) met (0.1 mM) DAP (0.3 mM) ile (3 mM) suc (0.8 mM)
- Stressing medium: MS glc (0.2%) glu (5 mM) val (0.3 mM) leu (0.3 mM) met (0.1 mM) DAP (0.3 mM) ile (3 mM)
- Generation time 2h at 37°C

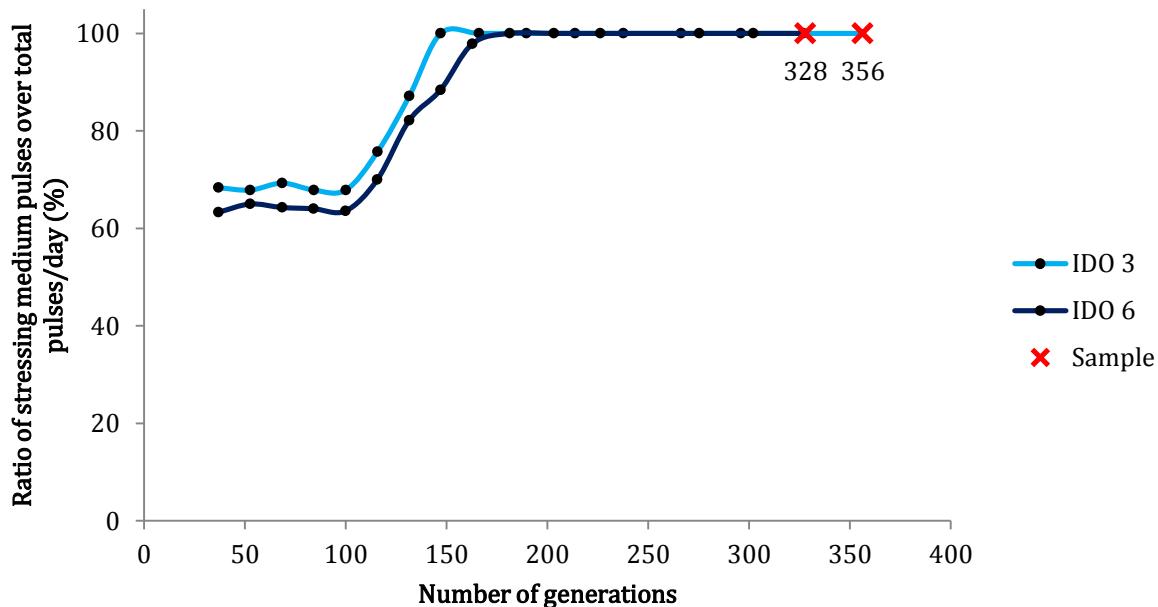


Figure 57: Evolution of cultures IDO 3 and IDO 6 in medium SWAP regime

Shown is the ratio of stressing medium dilution pulses per day in medium SWAP regime. The relaxing medium was MS glc (0.2%) glu (5 mM) val (0.3 mM) leu (0.3 mM) met (0.1 mM) DAP (0.3 mM) ile (3 mM) suc (0.8 mM) and the stressing medium was MS glc (0.2%) glu (5 mM) val (0.3 mM) leu (0.3 mM) met (0.1 mM) DAP (0.3 mM) ile (3 mM). The generation time was fixed at 2 h, and growth temperature at 37°C. Crosses indicate sampling points with the corresponding number of generations.

Culture of strain G4110 (G3990 *ΔkdgK::ido+*) is in light blue for culture IDO 3 and dark blue for culture IDO 6.

For both cultures IDO 3 and IDO 6, the evolution profiles were similar. The ratio of 100% dilution pulses per day of stressing medium was reached around 150 generations in continuous culture. Culture IDO 3 was sampled at generation 356 and culture IDO 6 at generation 328. Isolates were obtained on solid medium.

III.5.3.2 Phenotypic analysis of isolates from cultures IDO 3 and IDO 6

Isolates from cultures IDO 3 (G4170) and IDO 6 (G4173) were tested for succinate complementation in the presence of isoleucine. Results are shown in Table 28.

Table 28: Succinate complementation tests of isolates from cultures IDO 3 and IDO 6

| Strain | Genotype | MS glc (0.2%) glu (5 mM) leu val (0.3 mM) | | |
|--------|--|---|--------------|---------------------------|
| | | Ø | ile (1.5 mM) | ile (1.5 mM) suc (0.8 mM) |
| G3990 | C> <i>ΔArgF-lac ΔgltA Δicd::aad+</i> <i>Δpck</i> <i>Δppc ΔaceBAK ΔglcDEFGB ΔsfCA ΔmaeB</i> <i>ΔpuuE ΔgabT ΔsucA Δilv</i> | - | - | +++ |
| G4110 | G3990 <i>ΔkdgK::ido+</i> | - | - | +++ |
| G4170 | Isolate sampled from culture IDO 3 | - | +++ | +++ |
| G4173 | Isolate sampled from culture IDO 6 | - | +++ | +++ |

Complementation tests on solid medium were performed at 37°C for 48 h. The bacterial growth observed is indicated with signs for normal (+++), medium (++) residual (+) and no complementation (-).

Both isolates G4170 and G4173 showed succinate complementation depending on the presence of isoleucine.

III.5.3.3 Sequence analysis of isolates from cultures IDO 3 and IDO 6

Isolates G4170 and G4173 were sequenced at the *ido* gene region and no mutation inside the gene was detected. However, an insertion of approximately 1.3 kb was noticed upstream of the gene *ido*. A blast analysis on the 1.5 kb insertion matched the sequence of an IS2 element (Insertion Sequence 2) with 100% identity.

An IS is a short DNA sequence that behaves like a transposable element. IS2 has been described as a turn-off and turn-on element of gene activity depending on its orientation with respect to adjacent genes (Saedler and Reif, 1974). In orientation I (polar), IS2 has been described to serve as a very efficient turn-off signal, and in orientation II it is a

turn-on signal. From the sequencing data of isolates G4170 and G4173, it could be concluded that the IS2 inserted directly upstream of *ido* is in orientation II, probably leading to a higher expression of the gene *ido*. Figure 58 shows a simplified scheme of the gene *ido* and its surrounding genomic context as it can be deduced for the initial strain G4110 and the evolved isolates G4170 and G4173.

Similarly to the cultures KDB 15 and KDB 16 (section III.4.2.3), this genetic event occurred in isolates obtained from independent cultures suggesting that they had a direct impact on the acquisition of the new phenotype. It is also noteworthy that the two cultures evolved in a parallel manner suggesting the insertion event to have taken place at the same moment in the growing cell population.

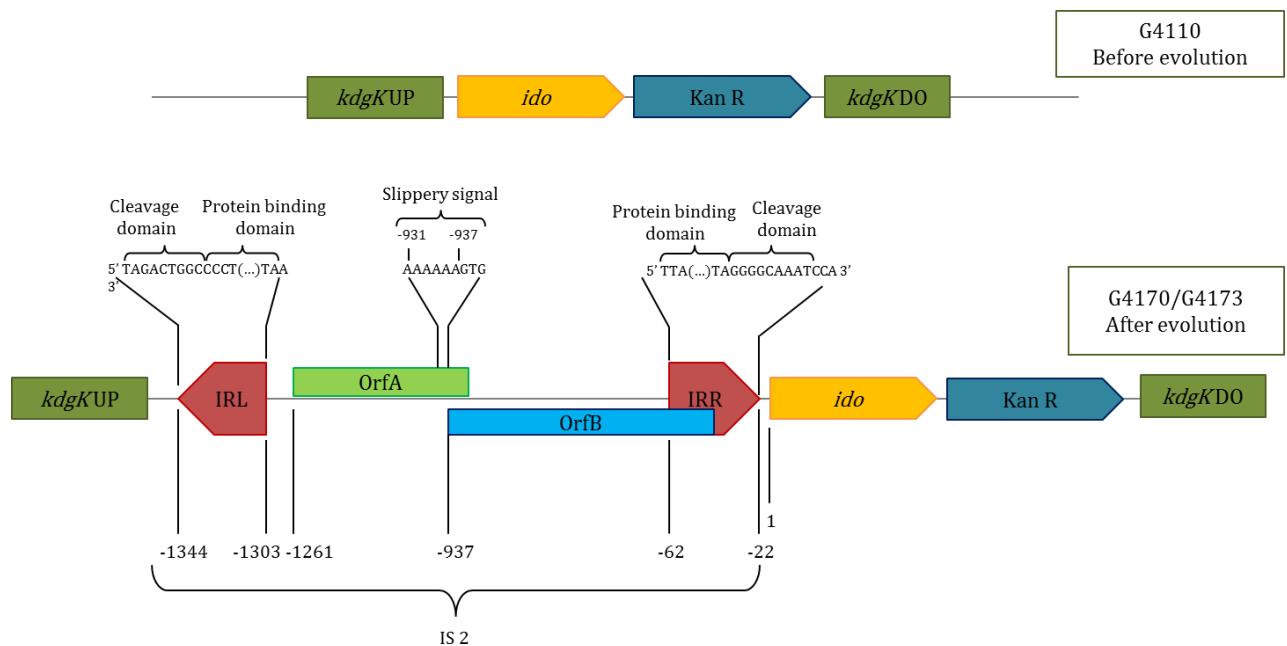


Figure 58: Map of the *ido* locus on the chromosome of strain G4110 and strains G4170/G4173

III.5.3.4 mRNA expression of Ido in isolate from culture IDO 3

In the hypothesis that the aforementioned genetic event, *i.e.* IS2 insertion upstream of gene *ido*, impacted the expression of the gene in the evolved strains, RT-qPCR assays were performed to determine the *ido*-mRNA content in the cells (Figure 59).

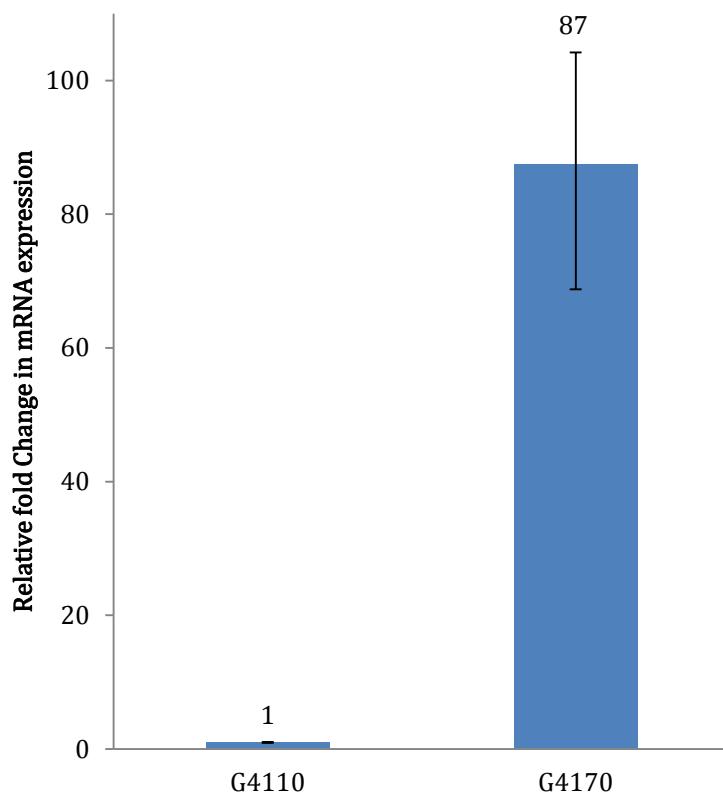


Figure 59: Comparison of mRNA expression level of *ido* in the strain G4170 from culture IDO 3 relative to the initial strain G4110

Expression was measured by RT-qPCR and quantified as $2^{-\Delta\Delta Ct}$ normalized to 16S mRNA expression; bars show means of triplicate measurements \pm sd.

The results of the RT-qPCR experiment show that the mRNA expression of *ido* in the evolved strain G4170 is 87 times higher than in the initial strain G4110.

Similarly to what was observed during the evolution of *kdo2* towards lysine (III.4.2.4), we observe that the expression of *ido* has been upregulated. In order to determine if the insertion of the IS2 upstream of *ido* is sufficient for succinate complementation, the *ido* chromosomal region containing the IS2 element was transferred in the selection strain G3990.

III.5.3.5 Transduction in the initial strains of the modified *ido* gene region with the upstream IS2

A P1 transduction was performed to transfer the chromosomal region containing the gene *ido* along with the upstream IS2 (Noted *ido** from the evolved strain G4170 into the selection strain G3700. The resulting strain G4351 (G3700 *ΔkdgK::ido**+) was then tested for succinate complementation to determine if the IS2 upstream of the gene *ido* was sufficient to establish succinate prototrophy in the presence of isoleucine (Table 29).

Table 29: Succinate complementation test of strain G4351

| Strain | Genotype | Growth in MS glc (0.2%) glu (5 mM) | | |
|--------|--|------------------------------------|--------------|---------------------------|
| | | Ø | ile (1.5 mM) | ile (1.5 mM) suc (0.8 mM) |
| G3700 | C> <i>ΔArgF-lac ΔgltA Δicd::aad+ Δpck</i> <i>Δppc ΔaceBAK ΔglcDEFGB Δsfca ΔmaeB</i> <i>ΔpuuE ΔgabT ΔsucA</i> | - | - | +++ |
| G4351 | G3700 <i>ΔkdgK::ido*</i> + | - | +++ | +++ |

Complementation tests on solid medium were performed at 37°C for 48 h. The bacterial growth observed is indicated with signs for normal (+++), medium (++), residual (+) and no complementation (-).

As indicated in Table 29, the strain G4351 shows succinate complementation depending on the presence of isoleucine. This observation confirms that the IS2 upstream of the gene *ido* has a direct influence on the succinate complementation phenotype.

The present evolution experiments led to the selection of mutants capable of efficient succinate complementation due to an increase in the expression level of the gene *ido*. These isolates have evolved with the natural substrate of Ido, the isoleucine.

III.5.3.6 Isotopic analysis of isolates from cultures IDO 3 and IDO 6

An isotopic labeling analysis was performed to confirm the activity of Ido *in vivo* as it was performed for Kdo2 (Section III.4.1.5). Bacteria of the strains studied were grown in the following media:

- Medium I: MS glc (0.2%) [1,2-¹³C₂]-glu (5 mM) val (0.3 mM) leu (0.3 mM) met (0.1 mM) DAP (0.3 mM) ile (3 mM) suc (0.8 mM)
- Medium II: MS glc (0.2%) [1,2-¹³C₂]-glu (5 mM) val (0.3 mM) leu (0.3 mM) met (0.1 mM) DAP (0.3 mM) ile (3 mM)

Metabolomes of cultures in the presence of [1,2-¹³C₂]-labeled glutamate were analyzed by Liquid chromatography–mass spectrometry (LC-MS). Results are shown in Figure 60.

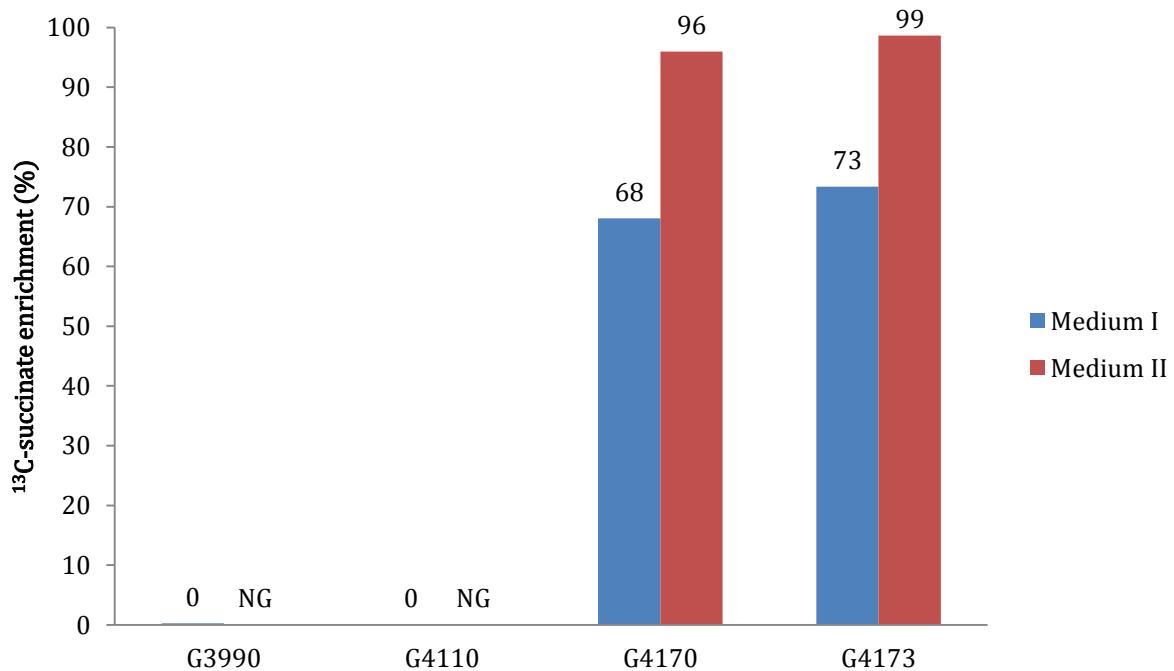


Figure 60: ¹³C enrichment of succinate in the metabolome of evolved isolates of cultures IDO 3 and IDO 6 and control strains

¹³C-succinate enrichment in metabolome of strain G3990 (selection strain), G4110 (G3990 *ΔkdgK::ido+*), strain G4170 (isolate from culture IDO 3) and strain G4173 (isolate from culture IDO 6). Medium I (blue): MS glc (0.2%) ¹³C-glu (5 mM) val (0.3 mM) leu (0.3 mM) met (0.1 mM) DAP (0.3 mM) ile (3 mM) suc (0.8 mM) and medium II (red): MS glc (0.2%) ¹³C-glu (5 mM). val (0.3 mM) leu (0.3 mM) met (0.1 mM) DAP (0.3 mM) ile (3 mM). NG: No growth

No ¹³C-succinate has been detected in the metabolome of the selection strain G3990 which does not express Ido. In the case of the evolved strains G4170 and G4173, we observe similar significant ¹³C-succinate enrichments. In the absence of succinate supplementation during growth, the succinate in the metabolome of strain G4170 is 96% ¹³C-enriched and 99% for strain G4173. After growth in the presence of standard succinate, the whole succinate content of the metabolome of strain G4170 showed a similar decrease of the ¹³C-labeled form of succinate.

In the case of the non-evolved strain G4110, no ¹³C-succinate enrichment has been detected, although the strain harbors one copy of gene *ido* at the *kdgK* locus. The expression of Ido in strain G4110 might be so low that the ¹³C-succinate enrichment originating from α-KG conversion cannot be distinguished from natural occurrence of ¹³C-in standard succinate by the detection method. This is at odds with the results

obtained with the gene *metA*, whose expression from the very same genetic configuration fully complemented the methionine needs of a *ΔmetA* strain.

In addition to these results, the specific product of Ido, 4-OH-isoleucine, has been identified in the metabolome of strains G4170 and G4173 (Figure 61). The identity of the product was confirmed with a sample of commercial 4-OH-isoleucine (Sigma Aldrich, Ref: 50118) on the basis of its retention time and the exact mass (*m/z* = 148.09682 in the positive ionization mode) (data not shown).

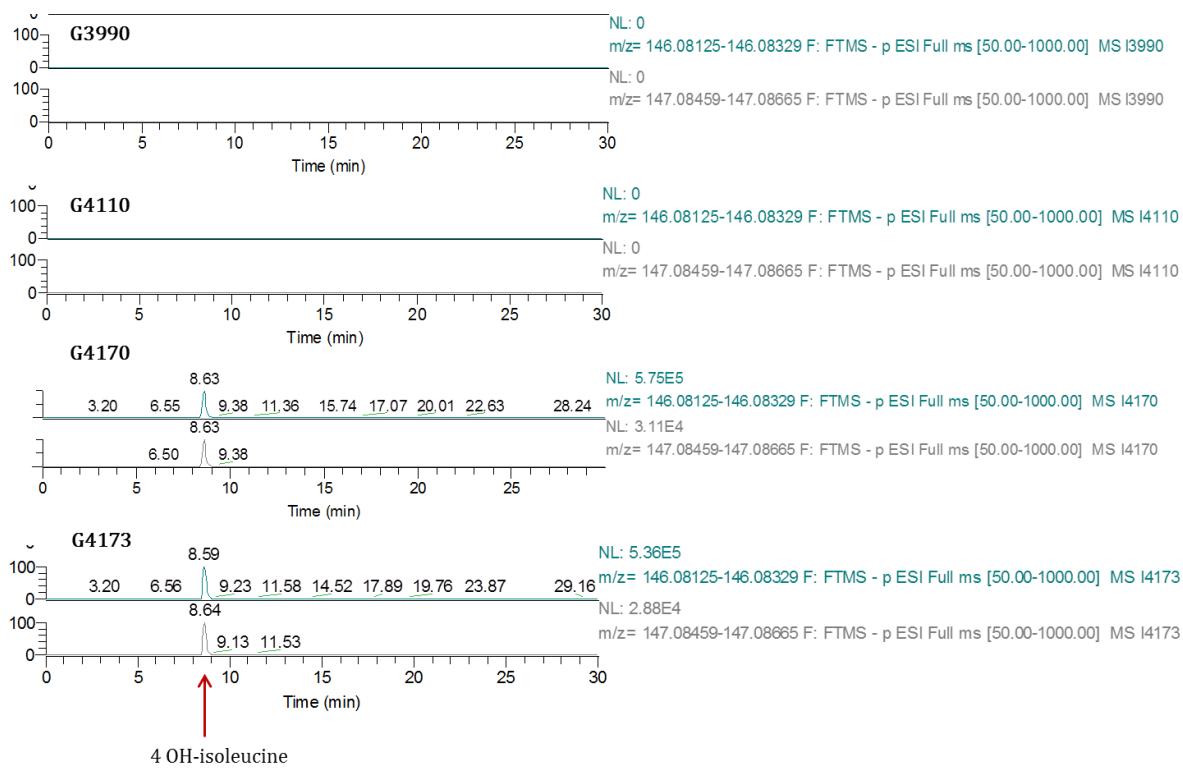


Figure 61: Extracted-Ion Chromatogram of 4-OH-isoleucine in the metabolome of strains before and after evolution in cultures IDO 3 and IDO 6

Extracted-Ion Chromatogram in positive and negative mode in the metabolome of strain G3990 (selection strain), strain G4110 (G3990 *ΔkdgK::ido+*), strain G4170 (isolate from culture IDO 3) and strain G4173 (isolate from culture IDO 6).

These observations confirm the activity of Ido *in vivo*.

III.5.4 Evolution *in vivo* of Ido towards isoleucine analogs

Similarly to the gene *kdo2*, we attempted to evolve the gene *ido* towards new hydroxylation substrates. The strain G4170 was used as a starting point. However, this strain was deleted for the *i/v* locus requiring the addition of the aliphatic amino acids ile, leu and val in the growth medium. The presence of these amino acids can have

uncontrolled influences on the evolution experiments with analogs, notably concerning the cellular uptake of the compounds. Also, isoleucine as natural substrate of Ido is in competition with any analog for which enzymatic adaptation is sought.

We therefore restored the *ilv* operon for further evolution experiments towards isoleucine analogs in strain G4170 (Figure 56). Strain G4170 was transduced with a P1 phage prepared on the *E. coli* K12 strain MG1655, and the *ilv* operon was restored resulting in the strain G4383 (Table 30).

Table 30: Genotype of strains used for the evolution of Ido towards isoleucine analogs

| Strain | Genotype |
|--------|--|
| G3990 | C> <i>ΔArgF-lac ΔgltA Δicd::aad+ Δpck Δppc ΔaceBAK ΔglcDEFGB ΔsfcA ΔmaeB ΔpuuE ΔgabT ΔsucA Δilv::Cm+</i> |
| G4110 | G3990 <i>ΔkdgK::ido+</i> |
| G4170 | Isolate sampled from culture IDO 3 (<i>ΔkdgK::ido*</i>) |
| G4383 | G4170 <i>ilv+</i> |

As analogs, the amino acids norvaline and norleucine, possessing a simplified side chain with respect to isoleucine were chosen. Preliminary results have been published showing some activity on these compounds in an *in vitro* test (Hibi *et al.*, 2011). In addition valeric acid, lacking the alpha amino group, was chosen (Table 31).

Table 31: Isoleucine analogs tested for evolution

| Name | Chemical formula | Structural formula |
|--------------|--|--------------------|
| Isoleucine | C ₆ H ₁₃ NO ₂ | |
| Valeric acid | C ₅ H ₁₀ O ₂ | |
| Norvaline | C ₅ H ₁₁ NO ₂ | |
| Norleucine | C ₆ H ₁₃ NO ₂ | |

III.5.4.1 Evolution *in vivo* of Ido towards isoleucine analogs in medium SWAP regime

The strain G4383 was inoculated in the GM3 in medium SWAP regime for the evolution of Ido towards isoleucine analogs (Figure 62).

Conditions of the medium SWAP regime for the culture IDO 11 (strain G4383 grown with valeric acid, Figure 62, orange line) were as follows:

- Relaxing medium: MA glc (0.2%) glu (5 mM) valeric acid (1.5 mM) suc (0.8 mM)
- Stressing medium: MA glc (0.2%) glu (5 mM) valeric acid (1.5 mM)
- Generation time 2 h at 37°C

Conditions of the medium SWAP regime for the culture IDO 13 (strain G4383 grown with norleucine, Figure 62, green line) were as follows:

- Relaxing medium: MA glc (0.2%) glu (5 mM) norleucine (1.5 mM) suc (0.8 mM)
- Stressing medium: MA glc (0.2%) glu (5 mM) norleucine (1.5 mM)
- Generation time 2 h at 37°C

Conditions of the medium SWAP regime for the culture IDO 9 (strain G4383 with norvaline, Figure 62, blue line) were as follows:

- Relaxing medium: MA glc (0.2%) glu (5 mM) norvaline (1.5 mM) suc (0.8 mM)
- Stressing medium: MA glc (0.2%) glu (5 mM) norvaline (1.5 mM)
- Generation time 2 h at 37°C

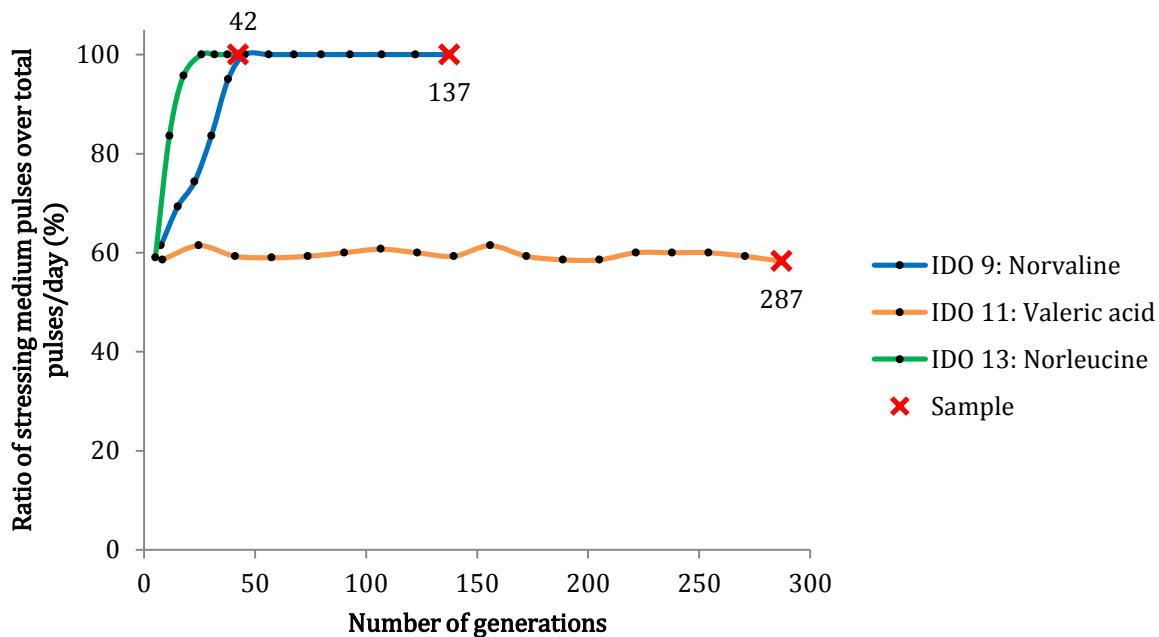


Figure 62: Evolution of cultures IDO 9, IDO 11 and IDO 13 in medium SWAP regime.

Shown is the ratio of stressing medium dilution pulses per day in medium SWAP regime. The generation time was fixed at 2 h, and growth temperature at 37°C. Crosses indicate sampling points, the corresponding number of generations is indicated.

Culture IDO 9 of strain G4383: **blue** line. Relaxing medium was MA glc (0.2%) glu (5 mM) norvaline (1.5 mM) suc (0.8 mM) and stressing medium was MA glc (0.2%) glu (5 mM) norvaline (1.5 mM)

Culture IDO 11 of strain G4383: **orange** line. Relaxing medium was MA glc (0.2%) glu (5 mM) valeric acid (1.5 mM) suc (0.8 mM) and stressing medium was MA glc (0.2%) glu (5 mM) valeric acid (1.5 mM)

Culture IDO 13 of strain G4383: **green** line. Relaxing medium was MA glc (0.2%) glu (5 mM) norleucine (1.5 mM) suc (0.8 mM) and stressing medium was MA glc (0.2%) glu (5 mM) norvaline (1.5 mM)

The culture IDO 11 growing in the presence of valeric acid remained stable for 300 generations. As expected, sequencing of isolates obtained at generation 287 showed no mutation in the gene suggesting that Ido has no promiscuous activity towards valeric acid.

Both cultures IDO 9 and IDO 13 rapidly reached 100% of stressing medium. Samples at generation 42 for culture IDO 13, and at generation 186 for culture IDO 9, were withdrawn. Isolates were obtained on plates of the corresponding stressing medium and the *ido* gene locus sequenced.

No variants of the gene *ido* were identified for isolates from culture IDO 13. Complementation tests with a sample obtained from this culture showed growth to depend on the presence of either norleucine or isoleucine. Different processes might have taken place during the medium SWAP adaptation. As seen previously, an event leading to upregulation of the *ido* expression could explain this result. Alternatively, under the hypothesis that the norleucine hydroxylation product exerts some toxicity, an adaptive mutation could have occurred abolishing this inhibitory effect.

In the culture IDO 9, variants of Ido were identified. A sequencing analysis of 53 isolated clones yielded a 15% abundance of the allele *ido:E65K* (resulting from a GA to AG transition) and 7.5% of the allele *ido:A62P* (resulting from a GC to CG transversion). These observations demonstrate that different genetic variants were selected in the conditions applied (medium swap regime followed by a one-week period of turbidostat regime in the stressing medim).

III.5.4.2 Phenotypic analysis of isolates from culture IDO 9

Isolates G4571 ($\Delta kdgK::ido:E65K$) and G4569 ($\Delta kdgK::ido:A62P$) from culture IDO 9 were tested for succinate complementation in the presence of isoleucine and norvaline. The *ido* gene variants encoding for Ido:E65K and Ido:A62P were subcloned in the pSP100 vector, yielding pGEN1143 and pGEN1154 respectively. The plasmids were transformed into the genetic screen, strain G3700 (Table 32).

Table 32: Succinate complementation for norvaline-positive isolates

| Strain | Genotype | MA glc (0.2%) glu (5 mM) | | |
|--------|--|--------------------------|--------------|---------------|
| | | Ø | ile (1.5 mM) | norv (1.5 mM) |
| G3700 | C> <i>ΔArgF-lac ΔgltA Δicd::aad+ Δpck Δppc ΔaceBAK</i> <i>ΔglcDEFGB ΔsfcA ΔmaeB ΔpuuE ΔgabT ΔsucA</i> | - | - | - |
| G4383 | C> <i>ΔArgF-lac ΔgltA Δicd::aad+ Δpck Δppc ΔaceBAK</i> <i>ΔglcDEFGB ΔsfcA ΔmaeB ΔpuuE ΔgabT ΔsucA</i> <i>ΔkdgK::ido*(IS2 upstream)</i> | - | +++ | - |
| G4571 | Isolate from culture IDO 9 with <i>ΔkdgK::ido:E65K+</i> | - | +++ | ++ |
| G4569 | Isolate from culture IDO 9 with <i>ΔkdgK::ido:A62P+</i> | - | +++ | ++ |
| G4495 | G3700 pGEN1138 (pSP100:: <i>ido</i> +) | - | +++ | - |
| G4490 | G3700 pGEN1143 (pSP100:: <i>ido:E65K</i> +) | - | +++ | - |
| G4584 | G3700 pGEN1154 (pSP100:: <i>ido:A62P</i> +) | - | +++ | - |

Complementation tests on solid medium were performed at 37°C for 48 h. The bacterial growth observed is indicated with signs for normal (+++), medium (++) residual (+) and no complementation (-).

As depicted in Table 32, the evolved strains G4571 and G4569 show succinate complementation in the presence of both isoleucine and norvaline. The transformants G4490 and G4584, strains which have not evolved in continuous culture, did not show succinate complementation towards norvaline after incubation of 48 h. However, after a prolonged incubation of 72 h, some clones appeared in the case of strain G4490. One of these clones (strain G4525) was isolated. The genotype of strain G4525 was confirmed by PCR excluding the possibility of a contamination. Additional succinate complementation tests were performed on this strain to confirm the norvaline-positive phenotype (Table 33).

Table 33: Succinate complementation tests with norvaline for strain G4525

| Strain | Genotype | Growth in MA glc (0.2%) glu (5 mM) | | |
|--------|---|------------------------------------|--------------|---------------|
| | | Ø | ile (1.5 mM) | norv (1.5 mM) |
| G3700 | C> <i>ΔArgF-lac ΔgltA Δicd::aad+</i> <i>Δpck Δppc ΔaceBAK</i> <i>ΔglcDEFGB ΔsfcA ΔmaeB ΔpuuE ΔgabT ΔsucA</i> | - | - | - |
| G4490 | G3700 pGEN1143 (pSP100:: <i>ido:E65K+</i>) | - | +++ | - |
| G4525 | Isolate from strain G4490 adapted on norvaline | - | +++ | ++ |

Complementation tests on solid medium were performed at 37°C for 48 h. The bacterial growth observed is indicated with signs for normal (+++), medium (++) residual (+) and no complementation (-).

These observations suggest that:

- Ido has been shown to have some activity towards norvaline. This promiscuous activity might be enhanced in the Ido:E65K variant.
- Alternatively, the activity of the Ido:E65K variant might be enhanced for the substrate α-KG. Under conditions of norvaline saturation, this would also lead to an augmentation of the succinate production of the mutated Ido.
- The overexpression of Ido:E65K on a high copy plasmid allows succinate complementation within 48 h in the presence of isoleucine and a weak complementation after 72 h in the presence of norvaline. This suggests that the activity of Ido:E65K towards norvaline is lower than its activity towards isoleucine.
- The strain G4525 appeared as a single colony when the strain G4490 was kept on norvaline plates for a prolonged time (72 h). This points to a secondary adaptation event necessary for succinate complementation of the non-evolved genetic background of strain G3700 in the presence of norvaline.

In order to further study the impact of the mutations E65K and A62P on Ido, biochemical analysis is necessary.

III.5.4.3 Biochemical analysis of Ido and variants

The enzyme Ido was expressed with a N-terminal His-tag (Table 6). However, the formation of inclusion bodies constrained us to change the His-tag position to the C-termini instead of the N-termini. Also, the working buffer of the activity assays for succinate detection used for the determination of the kinetic parameters for the other α -KAOs (II.3.2) was in the alkaline range (pH 8.4), whereas the optimum pH for Ido activity is slightly acidic (pH 6.0). To cope with this difficulty, several methods were tested including decoupling the reaction using different buffers for each step, but none of them were adequate for the determination of the kinetic parameters of Ido.

We finally opted for the determination of the kinetic parameters via mass spectrometry as described in II.3.3. Kinetic parameters were determined for Ido and Ido:E65K for the α -KG, isoleucine and norvaline (Table 34).

Table 34: Ido and Ido:E65K kinetic parameters

| Enzyme | Substrate | Kinetic constants | | |
|----------|---------------------------|-------------------------|--|--|
| | | K_M (μM) | k_{cat} (min^{-1}) | k_{cat}/K_M ($\text{s}^{-1} \cdot \text{M}^{-1}$) |
| Ido (WT) | α -KG ^a | 29.8 ± 3.4 | | $1,7 \cdot 10^3$ |
| | Isoleucine ^b | 96.1 ± 9.1 | 3.0 ± 0.1 | $5,2 \cdot 10^2$ |
| | Norvaline ^b | 108.0 ± 11.1 | 2.5 ± 0.1 | $3,9 \cdot 10^2$ |
| Ido:E65K | α -KG ^a | 75.9 ± 3.4 | | $1,1 \cdot 10^3$ |
| | Isoleucine ^b | 65.9 ± 10.4 | 5.1 ± 0.4 | $1,3 \cdot 10^3$ |
| | Norvaline ^b | 64.8 ± 9.4 | 6.0 ± 0.3 | $1,5 \cdot 10^3$ |

^a Concentration d'isoleucine : 500 μM .

^b Concentration de 2-oxoglutarate: 500 μM .

The catalytic efficiency of Ido:E65K for isoleucine shows an increase by a factor of 2.5 as compared to Ido WT and 3.8 for norvaline.

III.6 Evolution *in vivo* of Ido and Kdo2 in a *pck+* selection context

In the previous evolution experiments towards natural substrates, *i.e.* L-lysine and L-isoleucine, isolates obtained after evolution in continuous culture exhibited a higher expression level of α -KAOs. These modifications fixed during the adaptation were sufficient for succinate complementation without mutations in the gene encoding for the α -KAOs kept under selection.

In the hypothesis that a higher selection pressure might enable the selection of variants of α -KAOs, a tighter genetic screen was tested.

As introduced in detail in section III.2.1.2, the *pck+* selection context is a tighter genetic screen for succinate complementation. In this genetic context, glutamate is provided as sole carbon source for the cells. Glutamate is converted into α -KG which is then converted by the α -KAOs in succinate, which provides all carbon necessary for biomass production.

The selection strain G4343 was constructed by deleting the genes *aceAK*, *sucA*, *ppc*, *puuE*, and *gabT*. The genes *kdo2* and *ido* with their upstream modification, *i.e.* the 14 bp deletion and the IS2 respectively, were introduced in the *kdgK* locus of strain G4343. The resulting strains were tested for succinate complementation (Table 35) and no complementation was observed.

Table 35: Complementation tests for genes *ido* and *kdo2* in *pck+* selection context

| Strain | Genotype | Growth in MA glu (5 mM) | | | |
|--------|--|-------------------------|------------|------------|------------|
| | | \emptyset | lys (5 mM) | ile (5 mM) | suc (5 mM) |
| G4343 | C> Δ <i>aceAK</i> Δ <i>sucA</i> Δ <i>ppc</i> Δ <i>puuE</i> Δ <i>gabT</i> | - | - | - | +++ |
| G4497 | G4343 Δ <i>kdgK::kdo2*</i> + <i>kan+</i> | - | - | N/A | +++ |
| G4499 | G4343 Δ <i>kdgK::ido*</i> + <i>kan+</i> | - | N/A | - | +++ |

Complementation tests on solid medium were performed at 37°C for 48 h. The bacterial growth observed is indicated with signs for normal (+++), medium (++) residual (+) and no complementation (-).

The strains G4497 and G4499 were inoculated in the GM3 in medium SWAP regime (Figure 63).

Conditions of the medium SWAP regime for the cultures KDB 26 and KDB 27 (strain G4497 grown with lysine) were as follows:

- Relaxing medium: MA glu (10 mM) lys (10 mM) suc (8 mM)
- Stressing medium: MA glu (10 mM) lys (10 mM)
- Generation time 2 h at 37°C

Conditions of the medium SWAP regime for the cultures IDO17 and IDO18 (strain G4499 grown with isoleucine) were as follows:

- Relaxing medium: MA glu (10 mM) ile (10 mM) suc (8 mM)
- Stressing medium: MA glu (10 mM) ile (10 mM)
- Generation time 2 h at 37°C

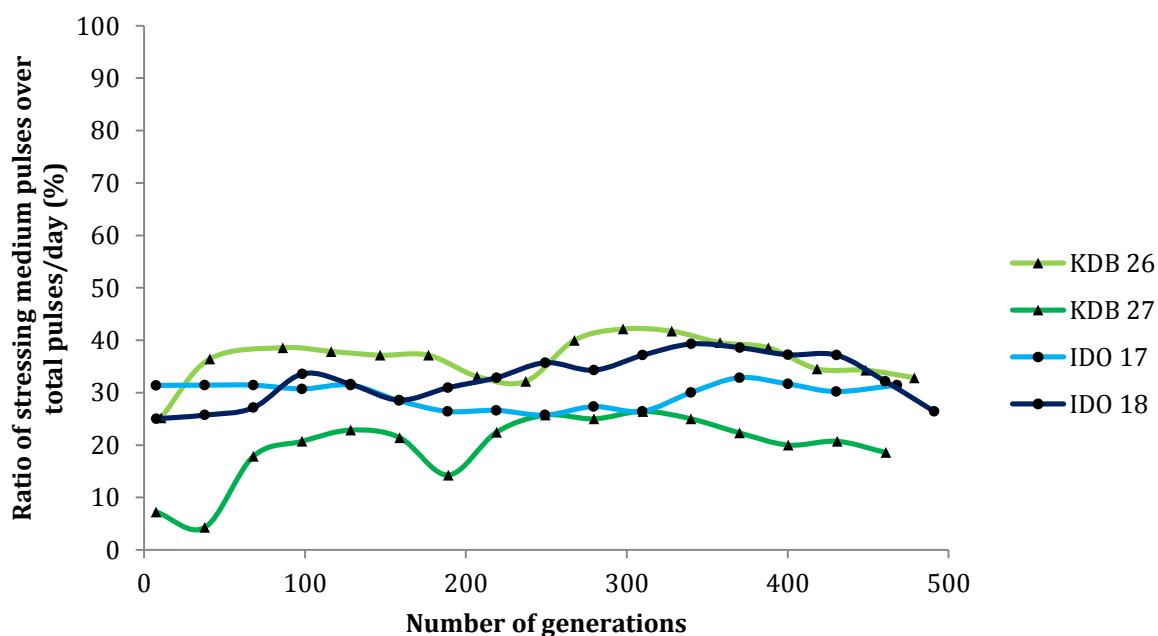


Figure 63: Evolution of cultures KDB 26, KDB 27, IDO 17 and IDO 18 in medium SWAP regime

Shown is the ratio of stressing medium dilution pulses per day in medium SWAP regime. The generation time was fixed at 2 h, and growth temperature at 37°C.

Cultures KDB 26 and KDB 27 of strain G4497 (triangles) are in light green and dark green respectively. Relaxing medium was MA glu (10 mM) lys (10 mM) suc (8 mM) and stressing medium was MA glu (10 mM) lys (10 mM).

Cultures IDO 17 and IDO 18 of strain G4499 (circles) are in light blue and dark blue respectively. Relaxing medium was MA glu (10 mM) ile (10 mM) suc (8 mM) and stressing medium was MA glu (10 mM) ile (10 mM).

After about 500 generations, none of the cultures has shown a significant increase in the amount of stressing medium pulses called in the vessels. The cultures are currently still running.

III.6.1.1 Directed evolution of *Clostridium phytofermentans*

The GM3 device has shown to be a very efficient tool for the directed evolution of *E. coli* (Marlière *et al.*, 2011). Most directed evolution studies have used *E. coli* albeit other organisms are of great interest such as plant-fermenting bacteria. *Clostridium phytofermentans* is a great candidate for directed evolution because of its capacity to ferment plant biomass that could provide a renewable alternative to fossil fuels. However, the fermentation process of plant biomass, *i.e.* the degradation of lignocellulose, also releases phenolics and acidic inhibitors.

In collaboration with the Tolonen's team (University of Paris Saclay, Genoscope); the directed evolution of *Clostridium phytofermentans* towards increased concentration of plant inhibitors was undertaken.

A survey of a panel of biomass inhibitors including phenolics, furans, and aliphatic acids was performed on *Clostridium phytofermentans* growth. Among these compounds, ferulic acid, one of the most abundant phenolic inhibitors, was chosen for directed evolution. The strain was cultured in the GM3 device with increased concentrations of ferulic acid. Ferulate resistant isolates were sequenced and transcription patterns analyzed.

The results are the subject of a publication:

"*Evolution of a biomass-fermenting bacterium to resist lignin phenolics*", Tristan Cerisy, Tiffany Souterre, Ismael Torres-Romero, Magali Boutard, Ivan Dubois, Julien Patrouix, Karine Labadie, Wahiba Berrabah, Marcel Salanoubat, Volker Döring, Andrew Tolonen. Appl. Environ. Microbiol. doi:10.1128/AEM.00289-17.

Evolution of a biomass-fermenting bacterium to resist
lignin phenolics

Tristan Cerisy^{1-4*}, Tiffany Souterre^{1-4*}, Ismael Torres-Romero¹⁻⁴, Magali Boutard^{1,2}, Ivan Dubois^{1,2}, Julien Patrouix^{1,2}, Karine Labadie^{1,2}, Wahiba Berrabah^{1,2}, Marcel Salanoubat¹⁻⁴, Volker Doring¹⁻⁴, Andrew Tolonen^{1-4†}

¹CEA, Genoscope, Évry France

²CNRS-UMR8030, Évry France

³Université Paris-Saclay, Évry France

⁴Université d'Évry, Évry France

*co-first authors

†correspondence: atolonen@genoscope.cns.fr

39

Abstract

40

Increasing the resistance of plant-fermenting bacteria to lignocellulosic inhibitors is useful to understand microbial adaptation and to develop candidate strains for consolidated bioprocessing. Here we study and improve inhibitor resistance in *Clostridium phytofermentans* (also called *Lachnoclostridium phytofermentans*), a model anaerobe that ferments lignocellulosic biomass. We survey the resistance of this bacterium to a panel of biomass inhibitors, and then evolve strains that grow in increasing concentrations of the lignin phenolic, ferulic acid, by automated, long-term growth selection in an anaerobic GM3 automat. Ultimately, strains resist multiple inhibitors and grow robustly at the solubility limit of ferulate while retaining the ability to ferment cellulose. We analyze genome-wide transcription patterns during ferulate stress and genomic variants that arose along the ferulate growth selection, revealing how cells adapt to inhibitors by changes in gene dosage and regulation, membrane fatty acid structure, and the surface layer. Collectively, this study demonstrates an automated framework for evolution of anaerobes and gives insight into the genetic mechanisms by which bacteria survive exposure to chemical inhibitors.

56

57

58

Importance

59

60

Fermentation of plant biomass is a key part of carbon cycling in diverse ecosystems. Further, industrial biomass fermentation could provide a renewable alternative to fossil fuels. Plants are primarily composed of lignocellulose, a matrix of polysaccharides and polyphenolic lignin. Thus, when microorganisms degrade lignocellulose to access sugars, they also release phenolic and acidic inhibitors. Here, we study how the plant-fermenting bacterium *Clostridium phytofermentans* resists plant inhibitors using the lignin phenolic, ferulic acid. We examine how the cell responds to abrupt ferulate stress by measuring changes in gene expression. We evolve increasingly resistant strains by automated, long-term cultivation at progressively higher ferulate concentrations and sequence their genomes to identify mutations associated with acquired ferulate resistance. Our study develops an inhibitor-resistant bacterium that ferments cellulose and provides insights into genomic evolution to resist chemical inhibitors.

73

74

75

Introduction

76

Fermentation of lignocellulosic biomass by bacteria like *Clostridium phytofermentans* is central to the functioning of soil, aquatic, and intestinal microbiomes. In addition, industrial fermentation of lignocellulosic biomass into fuels and chemicals could contribute significantly to global energy needs without impacting food production

Accepted Manuscript Posted Online
Applied and Environmental Microbiology
AEM
80 (1). Plant biomass is primarily composed of a macromolecular network of
81 polysaccharides linked with lignin, a polymer of phenylpropanoid subunits with aromatic
82 rings of varying degrees of methoxylation (2). Thus, when microorganisms hydrolyze
83 lignocellulose to access sugars, they also liberate three main types of inhibitors:
84 aliphatic acids, furans, and solubilized phenolics. The relative amounts of inhibitors
85 depend on the species and condition of the plant matter (3), but hydrolysates generally
86 contain inhibitors at concentrations that impede growth of microorganisms (4) by
87 damaging the cell membrane, metabolic enzymes, and nucleic acids (5). The most
88 abundant aliphatic acids are generally acetate, particularly in acetylxyan-rich
89 hardwoods (6), and formate from furan breakdown. The main furans are furfural and
90 hydroxymethylfurfural (5-HMF) that are formed by the dehydration of pentose and
91 hexose sugars, respectively. The most potent inhibitors released during biomass
92 hydrolysis are generally phenolics released from lignin (7).

93 The resistance of model, sugar-fermenting bacteria such as *E. coli* to biomass
94 inhibitors has been well-studied for aliphatic acids (8), (9), furans (7), and phenolics
95 (10). However, much less is known about resistance in bacteria like *C. phytofermentans*
96 that hydrolyze and ferment lignocellulose, even though plant inhibitors are important to
97 the ecology of these species. Moreover, developing inhibitor-resistant microorganisms
98 that directly metabolize biomass is needed for consolidated bioprocessing in a single
99 reactor, which is generally regarded as the most economical configuration for microbial
100 transformation of biomass into value-added chemicals (11).

101 Here we study and increase resistance to plant-derived inhibitors in *C.*
102 *phytofermentans*, an anaerobic bacterium in Clostridium cluster XIVa that expresses

Accepted Manuscript Posted Online
AEM
Applied and Environmental Microbiology
103 dozens of carbohydrate-active enzymes to degrade lignocellulosic biomass into
104 hexoses and pentoses, which it then ferments to ethanol, H₂, and acetate (12). We
105 initially define the effects of a panel of biomass inhibitors including phenolics, furans,
106 and aliphatic acids on *C. phytofermentans* growth. Among these compounds, we focus
107 on ferulic acid, a guaiacyl lignin precursor that is one of the most abundant phenolic
108 inhibitors in woods, grasses, and agriculturally important crops (13). We examine the
109 transcriptional response to ferulate stress by quantifying genome-wide mRNA
110 expression changes. We apply long term, anaerobic growth selection in a GM3 device
111 (14) to isolate a series of increasingly ferulate-resistant strains. We examined the
112 phenotypes of clones from along the selection and sequenced their genomes to identify
113 positively-selected genomic point mutations, small insertions and deletions (indels), and
114 large structural rearrangements. Finally, we discuss how these results improve our
115 understanding of the genetic basis of how bacteria evolve to resist chemical inhibitors.

116

Methods

118

119 Cell cultivation: *C. phytofermentans* ISDg (ATCC 700394) was cultured
120 anaerobically in GS2 medium (15). Growth of batch cultures containing inhibitors (Table
121 S1) was measured in 100-well microtiter plates (Bioscreen 9502550) containing 400 µL
122 GS2 medium with 3 g L⁻¹ glucose supplemented with a given inhibitor neutralized to pH
123 7. Wells were inoculated with 1:10 volume cells grown to log phase in minus-inhibitor
124 medium. The plates were sealed in the anaerobic chamber (2% H₂, 98% N₂) by press-
125 fitting adhesive sheets (Qiagen 1018104) (16) and incubated at 37°C in a Thermo

126 Scientific Bioscreen C. The cell densities (OD_{600}) were measured every 15 minutes with
127 30 seconds shaking before each reading. Cellulose cultures were inoculated into GS2
128 containing 10 g L^{-1} cellulose (0.5×5 cm strips of Whatman filter paper 1001-090, >98%
129 cellulose content). Cellulose degradation was measured as the dry mass of cellulose
130 remaining in culture by collecting the remaining cellulose on 11 μ filters by vacuum
131 filtration and drying it overnight at 65°C (17).

132 Ferulate-resistant *C. phytofermentans* clones were selected using a GM3
133 automat (14), a dual chamber continuous culture device that maintained anaerobic
134 conditions by flushing cultures with 100% N2 gas. A 50 ml culture was maintained at
135 30°C with optical density readings every 30 seconds, and was transferred between
136 growth chambers every 12 h to clean the empty chamber with 5N sodium hydroxide.
137 Cells were acclimated to increased ferulate using medium-swap mode, a modified
138 chemostat (6h generation time) with dilutions every 30 minutes of stressing medium
139 (high ferulate) if the cell density exceeded the density threshold (measured as OD_{880} 30,
140 which is equivalent to OD_{600} 0.4) and relaxing medium (low ferulate) otherwise. Once
141 cell densities stabilized at a constant cell density in the stressing medium for 24 h, the
142 GM3 was run as a turbidostat using the stressing medium until the culture reattained a
143 3.75 h generation time, similar to the WT strain in minus-ferulate medium. In turbidostat
144 mode, 20% of the culture volume was replaced with fresh medium each time the cell
145 density reached OD_{880} 30. Initially, the stressing medium contained 1 g L^{-1} ferulate, the
146 highest concentration at which a WT culture could be established in the GM3, and the
147 relaxing medium lacked ferulate. The medium-swap/turbidostat approach was iterated
148 by incrementing the stress medium by 0.5 g L^{-1} ferulate and replacing the relaxing

149 medium with the previous stressing medium. Samples from the GM3 culture were
150 plated to isolate colonies called CFY1, CFY2 and CFY3 clones at the end of the
151 turbidostat selections in 1, 2, and 3 g L⁻¹ ferulate, respectively.

152

153 RNA-seq: Log phase cultures (OD₆₀₀ 0.8) of WT *C. phytofermentans* ISDg were
154 diluted with 1 volume medium either lacking ferulate (-ferulate) or containing 4 g L⁻¹
155 ferulate (+ferulate, 2 g L⁻¹ final concentration). Samples for RNA and cell densities were
156 taken from duplicate cultures for each treatment immediately before ferulate addition
157 and 0.5 and 4 h afterwards. Total RNA was extracted using TRI reagent (Sigma 93289)
158 and 20 µg RNA was treated with 4U Turbo DNase (Ambion AM2238) for 30 min at
159 37°C. RNA was purified by Zymoclean (Zymo Research R1015) to capture RNA>200
160 bp. Five µg total RNA was depleted of rRNA by Ribo-Zero (Illumina MRZMB126),
161 yielding 200-400 ng RNA, and purified by Zymo Concentrator-5 (total capture) into 10 µl
162 water. cDNA libraries were prepared from 100 ng RNA using the Truseq Stranded
163 mRNA kit (Illumina 15031047) and sequenced on an Illumina HiSeq2000 sequencer
164 with paired-end 150 bp reads. Reads were aligned to the *C. phytofermentans* ISDg
165 genome (NCBI NC_010001.1) using Bowtie 2 (18). Gene expression was calculated as
166 reads per kb of gene per million reads (RPKM) using the easyRNASeq Bioconductor
167 package (19). Differential expression was defined as a greater than 4-fold change in
168 expression and a DESeq (20) p-value<0.01 after Bonferroni correction for multiple
169 testing of the 3,902 genes in the genome.

170

171 Genome sequencing: Genomes were sequenced for clones isolated from the
172 GM3 samples: CFY1 (2 clones CFY1A-B), CFY2 (2 clones CFY2C-D), and CFY3 (4
173 clones CFY3E-H). Genomic DNA (15-20 µg) was extracted from 3 ml cultures using the
174 Sigma GenElute Bacterial Genomic DNA kit (NA2110). DNA (100-250 ng) was
175 fragmented by Covaris E220 (Covaris, Inc., Woburn, MA, USA) to a 600 bp mean
176 fragment size. The DNA was end-repaired, 3' A-tailed, and ligated to Illumina
177 compatible adapters using the NEBNext DNA Sample Prep Master Mix Set 1 (New
178 England Biolabs E6040). Ligation products were purified with 1 volume Solid Phase
179 Reversible Immobilisation (SPRI) beads (Beckman Coulter A63880) and amplified by 12
180 cycles PCR using Kapa Hifi HotStart NGS library Amplification kit (Kapa Biosystems
181 KK2611) with P5/P7 primers. PCR products were purified (0.8 volume SPRI beads), run
182 on a 2% agarose gel, and DNA (700-800 bp) was excised and purified using the
183 Nucleospin Extract II DNA purification kit (Macherey-Nagel 740609). cDNA libraries
184 were sequenced using 300 bp paired-end reads on an Illumina MiSeq instrument.
185 Reads were quality filtered by Picard (<https://github.com/broadinstitute/picard>) and
186 aligned to the *C. phytofermentans* ISDg reference genome (NCBI NC_010001.1) using
187 Bowtie 2 (18). Sequence variants (SNPs, indels) in the CFY strains relative to the
188 reference genome were identified using the GATK (21) as described previously (22).
189 Structural variations were detected using the Breseq split-read analysis tool (23).
190 Insertion sequences (IS) were identified using ISfinder (24).

191

192 Optical genome mapping: High molecular-weight DNA of strain CFY2C was
193 extracted in agar plugs, which were solubilized with 0.4 U of GELase (Epicentre

194 G09200) and dialyzed for 45 minutes. DNA was treated using IrysPrep® Reagent Kit
195 (BioNano Genomics) to prepare NLRS (Nicked, Labeled, Repaired and Stained) DNA.
196 Briefly, 300 ng of DNA was nicked with 10 U Nt.BspQI (NEB R0644S) for 2 h at 37 °C.
197 Nicked DNA was incubated for 1 h at 72 °C with fluorescently-labelled Alexa546-dUTP
198 and Taq Polymerase (NEB M0273). Nicks were ligated using Taq ligase (NEB M0208)
199 with dNTPs. DNA was counterstained with YOYO-1 (Life Technologies). NLRS DNA
200 was loaded into IrysChips® (BioNano Genomics) and data were collected on the Irys®
201 instrument (BioNano Genomics) until reaching ≥1000-fold coverage of molecules
202 ≥100 kb.

203 CFY2C DNA molecules were filtered using BioNano IrysView software (version
204 2.5.1) retaining molecules ≥ 100 kb with at least 6 label sites, yielding 32,359 molecules
205 with a N50 of 172 kb. The NCBI assembly (NC_010001) was *in silico* digested with
206 BspQI (5'-GCTCTTC-3') and used to align and assemble CFY2C molecules using the
207 BioNano assembly pipeline (Pipeline version 4618, RefAligner and Assembler version
208 4704) with the parameters used for small genomes. Molecules ≥400 kb were aligned
209 against the NCBI assembly with a tandem duplication of bp 2,689,393-3,023,191 joined
210 by an ISL3-2 element in order to identify molecules spanning the duplicated zone.

211
212 Quantitative PCR: We measured mRNA expression by quantitative reverse
213 transcription PCR (qRT-PCR) as described previously (25). Briefly, RNA was extracted
214 from log phase WT and CFY3E cultures as for RNA-seq. RNA was reverse transcribed
215 (Applied Biosystems 4368814) and mRNA expression was quantified by qPCR (KAPA
216 KK4621) with primers in Table S2. Expression values are means of triplicate

217 measurements of duplicate cultures calculated as $2^{-\text{Ct}}$ (26), normalized to 16S rRNA
218 levels and multiplied by a scaling factor of 10^6 . To calculate the relative abundance of
219 DNA variants in the GM3 cultures, genomic DNA was extracted as for genome
220 sequencing from samples directly taken from the GM3 at 8 time points. The abundance
221 of a DNA variant was measured by qPCR (KAPA KK4621) relative to 16S (primers in
222 Table S2). The abundance of the DNA variant in the mixed population was calculated
223 relative to the $2^{-\text{Ct}}$ of a CFY clone that bears the variant in 100% of cells.

224

225 Mass spectrometry and chromatography: Ferulate concentrations were
226 compared in WT and CFY3E cultures after 5 days growth in GS2 medium containing 6
227 g L⁻¹ ferulate by LC/ESI-MS and MS/MS using a Dionex TCC-3000RS chromatographic
228 system (Thermo Fisher Scientific) coupled to an Orbitrap Elite mass spectrometer
229 (Thermo Electron Corporation) equipped with a HESI source. HPLC separation was
230 performed on a 5 µm, 4.6 × 150 mm Sequant ZICpHILIC column (Merck) at 40° C with a
231 flow rate of 0.5 ml min⁻¹ and a mobile phase of 10 mM (NH₄)₂CO₃ pH 9.9 (phase A) and
232 acetonitrile (phase B). Elution was conducted using the following gradient conditions: 2
233 min at 80% phase B, 20 min linear gradient from 80 to 40% of phase B, 8 min at 40%
234 phase B, 5 min increase to 80% phase B, and 15 min of 80% phase B. The mass
235 spectrometer was operated in ESI negative ion mode using a -4.5 kV ion spray, a 275°C
236 capillary temperature, and a mass resolution of 60,000. Sheath gas, auxiliary gas, and
237 sweep gas flow rates were set to 60, 10 and 2 arbitrary units, respectively. Mass
238 spectra were analyzed using Xcalibur version 2.2 (Thermo Fisher Scientific).

239 Cellular fatty acids were analyzed in ± ferulate WT and CFY3E cultures and -
240 ferulate CFY1B, CFY2C, CFY2D, and CFY3F cultures. Late log phase cells were
241 collected by centrifugation from cultures grown in medium either with 2 g L⁻¹ ferulate or
242 lacking ferulate. Fatty acid methyl esters (FAME) were obtained from 100 mg cells by
243 saponification, methylation, and extraction (27) and were identified using the DSMZ
244 Identification Service (Braunschweig, Germany). Briefly, FAME mixtures were
245 separated using the Sherlock Microbial Identification System (MIS) (Microbial ID, USA):
246 an Agilent model 6890N gas chromatograph with a 5% phenyl-methyl silicone capillary
247 column (0.2 mm x 25 m), a flame ionization detector, and an automatic sampler (Agilent
248 model 7683A). Peaks were integrated and fatty acid names and percentages calculated
249 using Sherlock MIS Standard Software (Microbial ID, USA). Plasmalogens were
250 quantified as dimethyl acetyl fatty acids. Polar lipids were extracted from 100 mg cells
251 using a chloroform:methanol:0.3% aqueous NaCl mixture 1:2:0.8 (v/v/v) by stirring
252 overnight. Cells were centrifuged and the polar lipids were recovered in the chloroform
253 phase by adjusting the chloroform:methanol:0.3% aqueous NaCl mixture to 1:1:0.9
254 (v/v/v). Polar lipids were resolved by 2D silica gel thin layer chromatography: dimension
255 1 was chloroform:methanol:water (65:25:4 v/v/v), dimension 2 was
256 chloroform:methanol:acetic acid:water (80:12:15:4 v/v/v/v). Total lipids were detected
257 using molybdatophosphoric acid and specific functional groups identified using spray
258 reagents specific for defined functional groups (28).

259

260 **Results**

261 Native *C. phytofermentans* inhibitor resistance: We measured growth of *C.*
262 *phytofermentans* in various concentrations of 12 lignocellulosic inhibitors (Table S1) to
263 gain a general understanding of the relative effects of aliphatic acids, furans, and
264 phenolics (Fig 1, S1). Both aliphatic acids reduce growth; acetate (Fig 1A) was less
265 toxic than formate (Fig 1B) on a mass per volume basis, but both acids had similar
266 effects in terms of molarity (Fig S1). At low furan concentrations, we observed normal
267 growth rates after an extended lag phase (Fig 1 C-D), similar to other bacteria that
268 reduce and detoxify furans (29), (30). Growth lags are proposed to be due to alcohol
269 dehydrogenase (ADH) reducing furan, causing NADH depletion and acetaldehyde
270 accumulation (31). Supportingly, the ADH protein Cphy1179 shares 28% amino acid
271 identity with a furfural-reducing, Zn-dependent ADH (32). However, if *C.*
272 *phytofermentans* detoxifies furans, this mechanism is abruptly overwhelmed at
273 concentrations above 2 g L⁻¹ 5-HMF and 1 g L⁻¹ furfural.

274 We examined the toxicities of two types of phenolic acids: hydrocinnamic acids
275 (p-coumarate and ferulate) and hydroxybenzoic acids (vanillate and 4-hydroxybenzoic
276 acid). We found that hydrocinnamic acids (Fig 1E-F) are more toxic than
277 hydroxybenzoic acids (Fig 1G,H), supporting the propionic group on the benzene ring in
278 hydrocinnamic acids enhances toxicity, likely by affecting how the molecules partition
279 into the membrane. Moreover, we found that phenolic acids are typically less toxic than
280 the corresponding aldehydes (Fig I-K) and catechol (Fig 1L). For example, vanillate (Fig
281 1G) is much less toxic than vanillin (Fig 1I) and 4-hydroxybenzoic acid (Fig 1H) is
282 similarly less toxic than benzaldehyde (Fig 1J). The enhanced toxicity of aldehydes is

283 likely due to their reactivity, resulting in formation of adducts with nucleophilic sites on
284 DNA, proteins, and other macromolecules (33).

285

286 Genome-wide mRNA expression during ferulate stress: We quantified genome-
287 wide mRNA expression changes at two timepoints ($t=0.5$ h, $t=4$ h) following
288 supplementation of mid-log cultures with 2 g L⁻¹ ferulate, which reduced growth (Fig 2A)
289 similar to the initial growth screen (Fig 1F). Three to five million read pairs were aligned
290 to the genome for each culture (Table S3A) to calculate gene expression levels (Table
291 S3B). The number of differentially-expressed genes (Table S3C-E) increased from 0
292 genes before ferulate addition (Fig 2B) to 78 genes after 30 minutes (Fig 2C), then
293 declined to 47 genes after 4 hours (Fig 2D). The most abundant functional categories of
294 differentially-expressed genes at $t=0.5$ h relate to repression of energy production,
295 coenzyme metabolism, and lipids (Fig 2E). The coenzyme-associated genes enable
296 siroheme biosynthesis, which is repressed in clostridia in response to redox stress (34).
297 Lipid genes include the *fab* gene cluster (*cphy0516-23*) for fatty acid biosynthesis, which
298 was strongly repressed at $t=0.5$ h (Fig 2E) and recovered by $t=4$ h. While cultures
299 continued active growth after the sampling, many of the differences between ± ferulate
300 cultures at $t=4$ h indicate the minus-ferulate treatment had depleted nutrients in the
301 medium, triggering expression of genes to assimilate alternative carbohydrates (Fig 2F).

302 Gene expression at $t=0.5$ h shows abrupt ferulate stress induces expression of
303 genes encoding the efflux pump *cphy1055-6*, which is similar to *E. coli* *mdtAB*
304 conferring resistance to organic solvents (35). Many of the genes up-regulated at $t=0.5$
305 h are co-located in two genomic regions. The first region encodes *tad* (tight adherence)

306 *cphy0029-40* genes for Flp-type type IV pili assembly. Type IV pili are widespread in
307 clostridia (36) for adhesion to solid substrates to form protective biofilms (37), reflecting
308 how ferulate represses motility genes in *C. beijerinckii* (38). The other cluster *cphy1838-*
309 *45* includes genes for the flavin mononucleotide (FMN) binding proteins WrbA (39) and
310 two NADPH:FMN reductases. NADPH:FMN reductase inactivation confers ferulate
311 resistance in *C. beijerinckii* by an unknown mechanism (40). While this appears in
312 opposition to our data that NADPH:FMN reductases are up-regulated by ferulate, both
313 results support the importance of FMN-mediated oxidoreduction in ferulate resistance.
314 This island also includes genes encoding an acetyltransferase and Cphy1845 that
315 shares 41% amino acid identity and the metal coordination with *E. coli* YhhW, which
316 cleaves the plant phenolic quercetin (41) *C. phytofermentans* may thus up-regulate
317 genes to transform or detoxify plant phenolics, similar to some ruminal clostridia (42).
318

319 Selection and physiology of ferulate-resistant strains: We selected *C.*
320 *phytofermentans* strains with increased ferulate resistance by cultivation in a GM3
321 automat, a dual chamber, continuous-culture device that automates delivery of fresh
322 medium and transfers the evolving cell suspension between twin growth chambers to
323 prevent biofilm formation. During acclimation to increased ferulate in medium-swap
324 mode, cell densities oscillated for 2-5 days because high densities triggered pulses of
325 stressing medium (high ferulate) that reduced culture density, which in turn resulted in
326 delivery of relaxing medium (low ferulate) that enabled recovery (Fig 3A). As such, the
327 ferulate-based selection in medium-swap mode is modulated by the ratio of relaxing and
328 stressing medium. Once cell densities stabilized in the stressing medium, growth rate at

329 the higher ferulate concentration was improved in turbidostat mode (Fig 3B). We
330 initiated the growth selection with stressing medium containing 1 g L⁻¹ ferulate, the
331 highest concentration at which we could establish a stable WT culture in the GM3. After
332 93 days (~500 generations) of continuous, log phase growth selection with
333 incrementally higher ferulate, the culture grew with the same 3.75 h generation time in 3
334 g L⁻¹ ferulate medium as WT in the absence of ferulate (Fig 3C). Clones isolated along
335 the growth selection are progressively more ferulate resistant in batch culture (Fig 3D-
336 G); while no growth was observed above 2 g L⁻¹ ferulate in the WT strain (Fig 3D),
337 CFY3 clones grow robustly at the ferulate solubility limit (6 g L⁻¹) (Fig 3G). We assessed
338 the ferulate resistance of 2 clones from each of the CFY1 (CFY1A-B) and CFY2
339 (CFY2C-D) time points and 4 clones from the CFY3 time point (CFY3E-H). The
340 duplicate CFY1 and CFY2 clones showed similar ferulate resistance, but CFY3H is
341 much less ferulate-resistant than the 3 other clones (Fig S2), showing cells in the GM3
342 culture are heterogeneous with respect to ferulate resistance.

343 We examined whether selection for ferulate resistance in glucose medium
344 resulted in physiological changes impacting cellulose fermentation and resistance to
345 other inhibitors. CFY3 strains degrade cellulose similar to WT (Fig 4A) and show
346 accelerated cellulose degradation in medium supplemented with ferulate (Fig 4B),
347 supporting the evolved strains are potentially improved candidates for fermentation of
348 lignocellulose. Moreover, the evolved resistance mechanisms extend to other biomass
349 inhibitors as CFY3 strains are also more resistant to vanillate and acetate (Fig S3),
350 albeit with considerable variability between strains. We also used mass spectrometry to
351 investigate if ferulate was consumed or transformed in WT and CFY3 cultures, revealing

352 the ferulate concentration was unaltered with no products corresponding to reduced,
353 demethoxylated, or decarboxylated ferulate (Fig S4). Thus, even though *C.*
354 *phytofermentans* up-regulates potential phenol-degrading enzymes in response to
355 ferulate, the cell adapted to ferulate by reinforcing the cell or excluding this molecule,
356 rather than detoxifying it.

357 As toxicity of aromatic molecules is often associated with disruption of the cell
358 membrane, we profiled fatty acids (FA) to determine if ferulate resistance is associated
359 with altered membrane phospholipids (Table S4). We found that when WT was exposed
360 to ferulate, the plasmalogen (vinyl ether phospholipid) content in the membranes
361 increased 18-fold. Moreover, CFY strains retained elevated plasmalogens even in the
362 absence of ferulate (Fig 4C). In particular, the CFY1B plasmalogen content in minus-
363 ferulate medium was 185-fold higher than WT. Related clostridia similarly increase
364 plasmalogens in response to aliphatic alcohol stress (43) (44), likely to fine tune
365 membrane fluidity and protect from redox-mediated damage (45). The distribution of FA
366 chain lengths in WT cells (Fig 4D) is similar to other clostridia, but with fewer
367 unsaturated FA and more cyclopropanes (46), both of which reduce membrane fluidity
368 to protect from solvent stress (47). While the addition of ferulate had little immediate
369 effect on the FA chains of WT cells (Fig S5A-B), the CFY strains showed altered FA
370 relative to WT in the absence of ferulate (Fig 4D). The CFY1B FA profile was the most
371 perturbed with increased hydroxylated C16 and unsaturated fatty acids, largely C18:1,
372 which is associated with increased ethanol tolerance in *E. coli* (48). CFY3F shifted to
373 branched FA (especially C15) and longer chain lengths (C18, C20), which increases
374 membrane rigidity (10) to potentially combat the membrane fluidizing effects of ferulate.

375 *C. phytofermentans* fatty acids are decorated with a diversity of phospho-, glyco-, and
376 amino- head groups (Fig S5C). While we did not detect changes in these head groups
377 in the WT response to ferulate or in the CFY strains, we consider it likely that they
378 participate in the response to solvents, similar to some other bacteria (10).

379

380 Genomes of ferulate-resistant isolates: We sequenced the genomes of eight
381 CFY1-3 clones giving between 106 and 705-fold coverage (Table S5A) to identify DNA
382 variants relative to wild-type (Table S5B). Seven single-nucleotide variations (SNV) and
383 short insertions/deletions (indels) are present in all the CFY genomes (Fig 5A), which
384 likely fixed in the population during an early selective sweep. These variants caused
385 non-synonymous changes in 5 proteins including a homolog of Cap5F (Cphy3503), a
386 protein for biosynthesis of capsular polysaccharides (49) that is associated with biofilm
387 formation (50) and stress resistance (51). Strains subsequently accrued strain-specific
388 mutations consistent with the population exploring alternative mutational pathways to
389 improve ferulate resistance, particularly by modifying sensor kinases that could
390 transduce signals associated with ferulate stress, fatty acid biosynthesis, and the
391 surface layer (S-layer) (Fig 5A). For example, the CFY1 and CFY2 strains incurred
392 coding variants in 3 genes putatively encoding fatty acid biosynthesis proteins:
393 Cphy3113 for anaerobic synthesis of unbranched fatty acids (52), the fatty acid
394 dehydratase FabZ (Cphy0520), and the reductase FabV (Cphy1286) for the final step in
395 fatty acid elongation. The genomes of CFY3E-G (high resistance) and CFY3H (low
396 resistance) differ by variants in Cphy3510, the most highly expressed protein in the
397 proteome that is proposed to form the S-layer (53). The S-layer is a protein lattice that

398 provides mechanical stabilization, sites for extracellular protein attachment, and a
399 selective barrier for molecules (54).

400 Intergenic changes that arose in the CFY genomes affect the expression levels
401 of adjacent genes. For example, a 15 bp sequence between the first two genes of the
402 ABC glucose transporter operon (*cphy2241-3*) was duplicated in the CFY3 strains (Fig
403 S6A-B). The repeated sequence forms an inverted repeat (IR) similar to Repeated
404 Extragenic Palindrome (REP) sequences, a widespread mechanism in bacteria to tune
405 gene expression by modulating the stability of different mRNA segments within an
406 operon (55). Duplication of this putative REP increases the mRNA secondary structure
407 of the *cphy2243-2242* intergenic region (Fig S6B), supporting it functions similar to REP
408 that increase expression by forming stable stem-loop structures that protect mRNA from
409 ribonucleases (56) (57). Supportingly, we found that mRNA expression of the two genes
410 downstream of the insertion were elevated (Fig 5B), which could have increased fitness
411 because the GM3 growth selections were done in glucose medium. The mRNA
412 expression of genes in two co-located operons with upstream point mutations were up-
413 regulated in the CFY strains (Fig 5C). The A-to-G transition upstream of *cphy1464*
414 created a TG di-nucleotide 2 bp upstream of the Pribnow hexamer (Fig S6A) that
415 enhances transcription in other bacteria (58) (59) and is present in the consensus -10
416 promoter sequence in *C. phytofermentans* (60). We propose the up-regulated operons
417 *cphy1459-61* and *cphy1464-5* either enable increased production of malonyl-CoA for
418 fatty acid biosynthesis or neutralize intracellular pH in response to ferulic acid stress
419 through production of ammonium and lactate (7) and bicarbonate buffering (Fig S6C).

420 Adaptive function can be imparted by structural changes to the genome resulting
421 from recombination and transposition of insertion sequences (IS elements). C.
422 *phytofermentans* encodes 31 IS elements (Table S6) including 12 ISL3 comprised of 2
423 isoforms: 8 ISL3-1 elements and 4 ISL3-2 elements. IS elements inactivate genes
424 through their transposition and act as substrates for homologous recombination. In
425 addition, the IR of all 12 ISL3 contain a 5'-TTGACA-3' sequence matching an outward-
426 facing, consensus -35 box from this organism (60) (Table S6), suggesting ISL3 could
427 activate expression of adjacent genes (61). An ISL3-1 was precisely deleted in all CFY
428 genomes as evidenced by reduced read coverage (Fig 6A) as well as BioNano optical
429 mapping and Sanger sequencing (Fig S7A-B), showing ISL3 are active in C.
430 *phytofermentans*. Further, the CFY2C,D strains share a 333 kb duplication from
431 *cphy2178* to *cphy2461* (276 genes) (Fig 6B), which we showed by PCR exists as a
432 tandem duplication joined by a novel ISL3-2 insertion (Fig 6C, Fig S7C). We did not
433 observe any extrachromosomal DNA by pulsed-field gel in the CFY2C strain, supporting
434 the *cphy2178*-ISL3-*cphy2461* fragment did not excise as a circular molecule. Further,
435 BioNano sequencing of DNA molecules greater than 400 kb spanning the junctions of
436 the duplicated region localize the rearrangement as a genomic, tandem duplication (Fig
437 6C). We quantified the relative abundance of cells bearing this duplication by
438 quantitative PCR of the *cphy2178*-ISL3-*cphy2461* fragment (Fig 6D). The duplication
439 arose between day 13 and 20 and overtook the population to comprise 68% cells by
440 day 40, supporting it was the subject of positive selection. Subsequently, this variant
441 declined in the population, representing 1% cells at day 63, as it was gradually replaced
442 by mutants with higher fitness; it was not present in any of the CFY3 genomes.

443

Discussion

445

446 When plant-fermenting bacteria like *C. phytofermentans* degrade lignocellulosic
447 biomass to access sugars, they also release various biomass-derived inhibitors
448 including ferulic acid. The abilities of these bacteria to survive exposures to these
449 inhibitors thus influences both their ecology and industrial potential. Fatty-acid (FA)
450 biosynthesis genes such as *cphy0520* (*fabZ*) and *cphy3113* were associated with both
451 the transcriptional and evolutionary response to ferulate. In particular, the FA profiles
452 were massively perturbed in CFY1B, which has a strain-specific variant in *FabZ*, a
453 dehydratase for unsaturated fatty acid synthesis. *FabZ* inactivation in CFY1B is
454 consistent with the accumulation of hydroxylated C16 and decrease of C17
455 cyclopropanes, which are synthesized from C16:1. Our results also show important
456 differences in the cellular processes implicated in the mRNA response to short-term
457 stress and the DNA changes enabling long-term resistance. Faced with an abrupt
458 increase in ferulate, the cell slows growth and up-regulates transcription of genes for
459 efflux pumps, biofilm formation, and flavoproteins including two NADPH:FMN
460 reductases that are associated with ferulate stress in *C. beijerinckii* (40). Over longer
461 time periods, natural selection of strains with robust growth in the presence of ferulate
462 resulted in DNA changes associated with metabolism, gene regulation, and the cell
463 surface (S-layer).

464

465 In Gram-negative bacteria, the outer membrane protects against influx of toxic
compounds. Gram-positives lack this outer membrane and instead have a thick

Accepted Manuscript Posted Online
AEM
Applied and Environmental Microbiology
466 peptidoglycan wall that cannot exclude solvents. Consequently, Gram-positive bacteria
467 are generally more sensitive to hydrophobic solvents (10). The S-layer is a lattice often
468 composed of a single protein that covers many Gram-positive bacteria. It functions as a
469 permeability barrier (54), potentially excluding ferulate, and stabilizes the cell through
470 non-covalently linkage to cell wall polysaccharides using threonine residues (62). A
471 three residue (NTT) insertion near the C-terminus of the S-layer protein Cphy3510 was
472 the sole mutation exclusively present in the three highly resistant CFY3 strains (CFY3E-
473 G); the CFY3 strain with lower resistance (CFY3H) had a N573D variant in Cphy3510.
474 S-layer proteins differ greatly among bacteria and neither of the Cphy3510 variants are
475 in known domains, but our results support that modification of the S-layer could be a
476 effective strategy to improve inhibitor resistance.

477 In addition to minor genomic changes (SNV and indels) that alter gene
478 expression or protein activity, both CFY2 genomes contain a tandem duplication of a
479 333 kb region joined by a novel ISL3-2 insertion, supporting this large genome
480 rearrangement was positively-selected during ferulate selection. Tandem duplications of
481 large chromosomal regions have been detected in other bacteria (55) and can improve
482 fitness by increasing gene dosage. A tandem duplication of regions joined by an IS
483 element in *E. coli* was proposed to have arisen following insertion of IS elements into
484 each copy of the duplication; when the IS recombined with each other, the intervening
485 region was deleted to leave a single, central IS element (63). The duplication observed
486 in CFY2 strains arose early in the experiment, perhaps because stress induced IS
487 element activity. Strains containing the duplication rapidly took over the population,
488 supporting its enhanced fitness, then gradually declined to represent 1 in 10^5 cells at the

489 end of the experiment (Fig 6D), likely because this strain was outcompeted by others
490 with higher fitness.

491 Our approach uses continuous, directed evolution as a framework for 'real time'
492 study of natural selection by analyzing the succession of microbial strains with
493 progressively higher fitness. Genome analysis of these strains using both high-
494 coverage, short reads and long-range, optical mapping reveals both the small and large
495 genomic changes that underlie a complex phenotype. When coupled with RNA-seq to
496 study the transcriptional response to abrupt change, this approach gives a portrait of
497 how the cell adapts to a given perturbation on different time scales. These results can
498 be applied to prioritize genes to engineer bacterial stress resistance. For example,
499 abrupt ferulate stress could be mitigated by overexpressing efflux pumps and
500 flavoproteins, whereas long-term ferulate resistance could be improved by altering the
501 primary surface layer protein and membrane biosynthesis (*fab* genes) to favor longer
502 fatty acids.

503

References

- 504 1. Metzger JO, Hüttermann A. 2009. Sustainable global energy supply based on
505 lignocellulosic biomass from afforestation of degraded areas. *Naturwissenschaften*
506 **96**:279–288. <http://dx.doi.org/10.1007/s00114-008-0479-4>
- 507 2. Klinke HB, Thomsen AB, Ahring BK. 2004. Inhibition of ethanol-producing yeast
508 and bacteria by degradation products produced during pre-treatment of biomass. *Appl
509 Microbiol Biotechnol* **66**:10–26. <http://dx.doi.org/10.1007/s00253-004-1642-2>
- 510 3. Galbe M, Zacchi G. 2007. Pretreatment of lignocellulosic materials for efficient
511 bioethanol production. *Adv Biochem Eng Biotechnol* **108**:41–65.
512 http://dx.doi.org/10.1007/10_2007_070

- 517
- 518 4. **Zaldivar J, Martinez A, Ingram LO.** 1999. Effect of selected aldehydes on the
519 growth and fermentation of ethanologenic *Escherichia coli*. Biotechnol Bioeng **65**:24–33.
520 [http://dx.doi.org/10.1002/\(SICI\)1097-0290\(19991005\)65:1<24::AID-BIT4>3.0.CO;2-2](http://dx.doi.org/10.1002/(SICI)1097-0290(19991005)65:1<24::AID-BIT4>3.0.CO;2-2)
521
- 522 5. **Mills TY, Sandoval NR, Gill RT.** 2009. Cellulosic hydrolysate toxicity and tolerance
523 mechanisms in *Escherichia coli*. Biotechnol Biofuels **2**:26.
524 <http://dx.doi.org/10.1186/1754-6834-2-26>
525
- 526 6. **Jönsson LJ, Alriksson B, Nilvebrant N-O.** 2013. Bioconversion of lignocellulose:
527 inhibitors and detoxification. Biotechnol Biofuels **6**:16.
528 <http://dx.doi.org/10.1186/1754-6834-6-16>
529
- 530 7. **Ibraheem O, Ndimba BK.** 2013. Molecular adaptation mechanisms employed by
531 ethanologenic bacteria in response to lignocellulose-derived inhibitory compounds. Int J
532 Biol Sci **9**:598–612. <http://dx.doi.org/10.7150/ijbs.6091>
533
- 534 8. **Axe DD, Bailey JE.** 1995. Transport of lactate and acetate through the energized
535 cytoplasmic membrane of *Escherichia coli*. Biotechnol Bioeng **47**:8–19.
536 <http://dx.doi.org/10.1002/bit.260470103>
537
- 538 9. **Roe AJ, McLaggan D, Davidson I, O'Byrne C, Booth IR.** 1998. Perturbation of
539 anion balance during inhibition of growth of *Escherichia coli* by weak acids. J Bacteriol
540 **180**:767–772.
541
- 542 10. **Weber FJ, de Bont JA.** 1996. Adaptation mechanisms of microorganisms to the
543 toxic effects of organic solvents on membranes. Biochim Biophys Acta **1286**:225–245.
544 [http://dx.doi.org/10.1016/S0304-4157\(96\)00010-X](http://dx.doi.org/10.1016/S0304-4157(96)00010-X)
545
- 546 11. **Lynd LR, van Zyl WH, McBride JE, Laser M.** 2005. Consolidated bioprocessing of
547 cellulosic biomass: an update. Curr Opin Biotechnol **16**:577–583.
548 <http://dx.doi.org/10.1016/j.copbio.2005.08.009>
549

- 550 12. **Warnick TA, Methé BA, Leschine SB.** 2002. *Clostridium phytofermentans* sp.
 551 nov., a cellulolytic mesophile from forest soil. *Int J Syst Evol Microbiol* **52**:1155–1160.
 552 <http://dx.doi.org/10.1099/00207713-52-4-1155>
 553
- 554 13. **Skerker JM, Leon D, Price MN, Mar JS, Tarjan DR, Wetmore KM, Deutschbauer
 555 AM, Baumohl JK, Bauer S, Ibáñez AB, Mitchell VD, Wu CH, Hu P, Hazen T, Arkin
 556 AP.** 2013. Dissecting a complex chemical stress: chemogenomic profiling of plant
 557 hydrolysates. *Mol Syst Biol* **9**:674. <http://dx.doi.org/10.1038/msb.2013.30>
 558
- 559 14. **Mutzel R, Marliere P.** June 2000. Method and Device for Selecting Accelerated
 560 Proliferation of Living Cells in Suspension.WO/2000/034433.
 561
- 562 15. **Cavedon K, Leschine SB, Canale-Parola E.** 1990. Cellulase system of a free-
 563 living, mesophilic clostridium (strain C7). *J Bacteriol* **172**:4222–4230.
 564 <http://dx.doi.org/10.1128/jb.172.8.4222-4230.1990>
 565
- 566 16. **Boutard M, Cerisy T, Nogue P-Y, Alberti A, Weissenbach J, Salanoubat M,
 567 Tolonen AC.** 2014. Functional diversity of carbohydrate-active enzymes enabling a
 568 bacterium to ferment plant biomass. *PLoS Genet* **10**:e1004773.
 569 <http://dx.doi.org/10.1371/journal.pgen.1004773>
 570
- 571 17. **Tolonen AC, Chilaka AC, Church GM.** 2009. Targeted gene inactivation in
 572 *Clostridium phytofermentans* shows that cellulose degradation requires the family 9
 573 hydrolase Cphy3367. *Mol Microbiol* **74**:1300–1313.
 574 <http://dx.doi.org/10.1111/j.1365-2958.2009.06890.x>
 575
- 576 18. **Langmead B, Salzberg SL.** 2012. Fast gapped-read alignment with Bowtie 2. *Nat
 577 Methods* **9**:357–359. <http://dx.doi.org/10.1038/nmeth.1923>
 578
- 579 19. **Delhomme N, Padoleau I, Furlong EE, Steinmetz LM.** 2012. easyRNASeq: a
 580 bioconductor package for processing RNA-Seq data. *Bioinforma Oxf Engl* **28**:2532–
 581 2533. <http://dx.doi.org/10.1093/bioinformatics/bts477>
 582
- 583 20. **Anders S, Huber W.** 2010. Differential expression analysis for sequence count
 584 data. *Genome Biol* **11**:R106. <http://dx.doi.org/10.1186/gb-2010-11-10-r106>

- 585
- 586 21. **Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-**
 587 **Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV,**
 588 **Altshuler D, Gabriel S, DePristo MA.** 2013. From FastQ data to high confidence
 589 variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc*
 590 *Bioinforma* **43**:11.10.1-33. <http://dx.doi.org/10.1002/0471250953.bi1110s43>
- 591
- 592 22. **Tolonen AC, Zuroff TR, Ramya M, Boutard M, Cerisy T, Curtis WR.** 2015.
 593 Physiology, genomics, and pathway engineering of an ethanol-tolerant strain of
 594 *Clostridium phytofermentans*. *Appl Environ Microbiol* **81**:5440–5448.
 595 <http://dx.doi.org/10.1128/AEM.00619-15>
- 596
- 597 23. **Barrick JE, Colburn G, Deatherage DE, Traverse CC, Strand MD, Borges JJ,**
 598 **Knoester DB, Reba A, Meyer AG.** 2014. Identifying structural variation in haploid
 599 microbial genomes from short-read resequencing data using breseq. *BMC Genomics*
 600 **15**:1039. <http://dx.doi.org/10.1186/1471-2164-15-1039>
- 601
- 602 24. **Siguier P, Gourbeyre E, Chandler M.** 2014. Bacterial insertion sequences: their
 603 genomic impact and diversity. *FEMS Microbiol Rev* **38**:865–891.
 604 <http://dx.doi.org/10.1111/1574-6976.12067>
- 605
- 606 25. **Tolonen AC, Cerisy T, El-Sayed H, Boutard M, Salanoubat M, Church GM.**
 607 2015. Fungal lysis by a soil bacterium fermenting cellulose. *Environ Microbiol* **17**:2618–
 608 2627. <http://dx.doi.org/10.1111/1462-2920.12495>
- 609
- 610 26. **Schmittgen TD, Livak KJ.** 2008. Analyzing real-time PCR data by the comparative
 611 C(T) method. *Nat Protoc* **3**:1101–1108. <http://dx.doi.org/10.1038/nprot.2008.73>
- 612
- 613 27. **Miller LT.** 1982. Single derivatization method for routine analysis of bacterial whole-
 614 cell fatty acid methyl esters, including hydroxy acids. *J Clin Microbiol* **16**:584–586.
- 615
- 616 28. **Daniels L, Hanson RS, Phillips JA.** 2007. Chemical Analysis, p 462-503. In
 617 Methods for General and Molecular Microbiology, Third Edition.
 618 <http://dx.doi.org/10.1128/9781555817497.ch18>
- 619

- 620 29. **Ezeji T, Qureshi N, Blaschek HP.** 2007. Butanol production from agricultural
621 residues: Impact of degradation products on *Clostridium beijerinckii* growth and butanol
622 fermentation. *Biotechnol Bioeng* **97**:1460–1469. <http://dx.doi.org/10.1002/bit.21373>
623
- 624 30. **Miller EN, Jarboe LR, Yomano LP, York SW, Shanmugam KT, Ingram LO.** 2009.
625 Silencing of NADPH-dependent oxidoreductase genes (*yqhD* and *dkgA*) in furfural-
626 resistant ethanologenic *Escherichia coli*. *Appl Environ Microbiol* **75**:4315–4323.
627 <http://dx.doi.org/10.1128/AEM.00567-09>
628
- 629 31. **Palmqvist E, Almeida JS, Hahn-Hägerdal B.** 1999. Influence of furfural on
630 anaerobic glycolytic kinetics of *Saccharomyces cerevisiae* in batch culture. *Biotechnol
631 Bioeng* **62**:447–454.
632 [http://dx.doi.org/10.1002/\(SICI\)1097-0290\(19990220\)62:4<447::AID-BIT7>3.0.CO;2-0](http://dx.doi.org/10.1002/(SICI)1097-0290(19990220)62:4<447::AID-BIT7>3.0.CO;2-0)
633
- 634 32. **Kang C, Hayes R, Sanchez EJ, Webb BN, Li Q, Hooper T, Nissen MS, Xun L.**
635 2012. Furfural reduction mechanism of a zinc-dependent alcohol dehydrogenase from
636 *Cupriavidus necator* JMP134. *Mol Microbiol* **83**:85–95.
637 <http://dx.doi.org/10.1111/j.1365-2958.2011.07914.x>
638
- 639 33. **LoPachin RM, Gavin T.** 2014. Molecular mechanisms of aldehyde toxicity: a
640 chemical perspective. *Chem Res Toxicol* **27**:1081–1091.
641 <http://dx.doi.org/10.1021/tx5001046>
642
- 643 34. **Sander K, Wilson CM, Rodriguez M, Klingeman DM, Rydzak T, Davison BH,
644 Brown SD.** 2015. *Clostridium thermocellum* DSM 1313 transcriptional responses to
645 redox perturbation. *Biotechnol Biofuels* **8**:211.
646 <http://dx.doi.org/10.1186/s13068-015-0394-9>
647
- 648 35. **Foo JL, Jensen HM, Dahl RH, George K, Keasling JD, Lee TS, Leong S,
649 Mukhopadhyay A.** 2014. Improving microbial biogasoline production in *Escherichia coli*
650 using tolerance engineering. *mBio* **5**:e01932. <http://dx.doi.org/10.1128/mBio.01932-14>
651
- 652 36. **Melville S, Craig L.** 2013. Type IV pili in Gram-positive bacteria. *Microbiol Mol Biol
653 Rev MMBR* **77**:323–341. <http://dx.doi.org/10.1128/MMBR.00063-12>
654

- 655 37. **Tomich M, Fine DH, Figurski DH.** 2006. The TadV protein of *Actinobacillus*
 656 *actinomycetemcomitans* is a novel aspartic acid prepilin peptidase required for
 657 maturation of the Flp1 pilin and TadE and TadF pseudopilins. *J Bacteriol* **188**:6899–
 658 6914. <http://dx.doi.org/10.1128/JB.00690-06>
- 659
- 660 38. **Lee S, Lee JH, Mitchell RJ.** 2015. Analysis of *Clostridium beijerinckii* NCIMB
 661 8052's transcriptional response to ferulic acid and its application to enhance the strain
 662 tolerance. *Biotechnol Biofuels* **8**:68. <http://dx.doi.org/10.1186/s13068-015-0252-9>
- 663
- 664 39. **Grandori R, Khalifah P, Boice JA, Fairman R, Giovanielli K, Carey J.** 1998.
 665 Biochemical characterization of WrbA, founding member of a new family of multimeric
 666 flavodoxin-like proteins. *J Biol Chem* **273**:20960–20966.
 667 <http://dx.doi.org/10.1074/jbc.273.33.20960>
- 668
- 669 40. **Liu J, Guo T, Shen X, Xu J, Wang J, Wang Y, Liu D, Niu H, Liang L, Ying H.** 2016. Engineering *Clostridium beijerinckii* with the Cbei_4693 gene knockout for
 670 enhanced ferulic acid tolerance. *J Biotechnol* **229**:53–57.
 671 <http://dx.doi.org/10.1016/j.jbiotec.2016.04.052>
- 672
- 673
- 674 41. **Adams M, Jia Z.** 2005. Structural and biochemical analysis reveal pirins to possess
 675 quercetinase activity. *J Biol Chem* **280**:28675–28682.
 676 <http://dx.doi.org/10.1074/jbc.M501034200>
- 677
- 678 42. **Schoefer L, Mohan R, Schwierz A, Braune A, Blaut M.** 2003. Anaerobic
 679 degradation of flavonoids by *Clostridium orbiscindens*. *Appl Environ Microbiol* **69**:5849–
 680 5854. <http://dx.doi.org/10.1128/AEM.69.10.5849-5854.2003>
- 681
- 682 43. **Kolek J, Patáková P, Melzoch K, Sigler K, Řezanka T.** 2015. Changes in
 683 membrane plasmalogens of *Clostridium pasteurianum* during butanol fermentation as
 684 determined by lipidomic analysis. *PLoS One* **10**:e0122058.
 685 <http://dx.doi.org/10.1371/journal.pone.0122058>
- 686
- 687 44. **Timmons MD, Knutson BL, Nokes SE, Strobel HJ, Lynn BC.** 2009. Analysis of
 688 composition and structure of *Clostridium thermocellum* membranes from wild-type and
 689 ethanol-adapted strains. *Appl Microbiol Biotechnol* **82**:929–939.
 690 <http://dx.doi.org/10.1007/s00253-009-1891-1>

- 691
- 692 45. **Nagan N, Zoeller RA.** 2001. Plasmalogens: biosynthesis and functions. *Prog Lipid Res* **40**:199–229. [http://dx.doi.org/10.1016/S0163-7827\(01\)00003-0](http://dx.doi.org/10.1016/S0163-7827(01)00003-0)
- 693
- 694
- 695 46. **Johnston NC, Goldfine H.** 1983. Lipid composition in the classification of the
696 butyric acid-producing clostridia. *J Gen Microbiol* **129**:1075–1081.
697 <http://dx.doi.org/10.1099/00221287-129-4-1075>
- 698
- 699 47. **Zhao Y, Hindorff LA, Chuang A, Monroe-Augustus M, Lyristis M, Harrison ML,**
700 **Rudolph FB, Bennett GN.** 2003. Expression of a cloned cyclopropane fatty acid
701 synthase gene reduces solvent formation in *Clostridium acetobutylicum* ATCC 824. *Appl Environ Microbiol* **69**:2831–2841. <http://dx.doi.org/10.1128/AEM.69.5.2831-2841.2003>
- 702
- 703
- 704 48. **Ingram LO, Vreeland NS, Eaton LC.** 1980. Alcohol tolerance in *Escherichia coli*.
705 *Pharmacol Biochem Behav* **13 Suppl 1**:191–195.
706 [http://dx.doi.org/10.1016/S0091-3057\(80\)80030-X](http://dx.doi.org/10.1016/S0091-3057(80)80030-X)
- 707
- 708 49. **Kneidinger B, O'Riordan K, Li J, Brisson J-R, Lee JC, Lam JS.** 2003. Three
709 highly conserved proteins catalyze the conversion of UDP-N-acetyl-D-glucosamine to
710 precursors for the biosynthesis of O antigen in *Pseudomonas aeruginosa* O11 and
711 capsule in *Staphylococcus aureus* type 5. Implications for the UDP-N-acetyl-L-
712 fucosamine biosynthetic pathway. *J Biol Chem* **278**:3615–3627.
713 <http://dx.doi.org/10.1074/jbc.M203867200>
- 714
- 715 50. **Branda SS, Vik S, Friedman L, Kolter R.** 2005. Biofilms: the matrix revisited.
716 *Trends Microbiol* **13**:20–26. <http://dx.doi.org/10.1016/j.tim.2004.11.006>
- 717
- 718 51. **Ophir T, Gutnick DL.** 1994. A role for exopolysaccharides in the protection of
719 microorganisms from desiccation. *Appl Environ Microbiol* **60**:740–745.
- 720
- 721 52. **Isabella VM, Clark VL.** 2011. Identification of a conserved protein involved in
722 anaerobic unsaturated fatty acid synthesis in *Neisseria gonorrhoeae*: implications for
723 facultative and obligate anaerobes that lack FabA. *Mol Microbiol* **82**:489–501.
724 <http://dx.doi.org/10.1111/j.1365-2958.2011.07826.x>
- 725

- 726 53. Tolonen AC, Haas W, Chilaka AC, Aach J, Gygi SP, Church GM. 2011.
 727 Proteome-wide systems analysis of a cellulosic biofuel-producing microbe. *Mol Syst Biol*
 728 **7**:461. <http://dx.doi.org/10.1038/msb.2010.116>
- 729
- 730 54. Sára M, Sleytr UB. 1987. Molecular sieving through S layers of *Bacillus*
 731 *stearothermophilus* strains. *J Bacteriol* **169**:4092–4098.
 732 <http://dx.doi.org/10.1128/jb.169.9.4092-4098.1987>
- 733
- 734 55. Darmon E, Leach DRF. 2014. Bacterial genome instability. *Microbiol Mol Biol Rev*
 735 **MMBR** **78**:1–39. <http://dx.doi.org/10.1128/MMBR.00035-13>
- 736
- 737 56. Newbury SF, Smith NH, Robinson EC, Hiles ID, Higgins CF. 1987. Stabilization
 738 of translationally active mRNA by prokaryotic REP sequences. *Cell* **48**:297–310.
 739 [http://dx.doi.org/10.1016/0092-8674\(87\)90433-8](http://dx.doi.org/10.1016/0092-8674(87)90433-8)
- 740
- 741 57. Xu C, Huang R, Teng L, Jing X, Hu J, Cui G, Wang Y, Cui Q, Xu J. 2015.
 742 Cellulosome stoichiometry in *Clostridium cellulolyticum* is regulated by selective RNA
 743 processing and stabilization. *Nat Commun* **6**:6900.
 744 <http://dx.doi.org/10.1038/ncomms7900>
- 745
- 746 58. Graves MC, Rabinowitz JC. 1986. In vivo and in vitro transcription of the
 747 *Clostridium pasteurianum* ferredoxin gene. Evidence for “extended” promoter elements
 748 in Gram-positive organisms. *J Biol Chem* **261**:11409–11415.
- 749
- 750 59. Helmann JD. 1995. Compilation and analysis of *Bacillus subtilis* sigma A-
 751 dependent promoter sequences: evidence for extended contact between RNA
 752 polymerase and upstream promoter DNA. *Nucleic Acids Res* **23**:2351–2360.
 753 <http://dx.doi.org/10.1093/nar/23.13.2351>
- 754
- 755 60. Boutard M, Ettwiller L, Cerisy T, Alberti A, Labadie K, Salanoubat M,
 756 Schildkraut I, Tolonen AC. 2016. Global repositioning of transcription start sites in a
 757 plant-fermenting bacterium. *Nat Commun* **7**:13783.
 758 <http://dx.doi.org/10.1038/ncomms13783>
- 759
- 760 61. Zafarullah M, Charlier D, Glansdorff N. 1981. Insertion of IS3 can “turn-on” a
 761 silent gene in *Escherichia coli*. *J Bacteriol* **146**:415–417.

- 762
 763 62. **Fagan RP, Fairweather NF.** 2014. Biogenesis and functions of bacterial S-layers.
 764 *Nat Rev Microbiol* **12**:211–222.
 765 <http://dx.doi.org/10.1038/nrmicro3213>
 766
- 767 63. **Raeside C, Gaffé J, Deatherage DE, Tenaillon O, Briska AM, Ptashkin RN,**
 768 **Cruveiller S, Médigue C, Lenski RE, Barrick JE, Schneider D.** 2014. Large
 769 chromosomal rearrangements during a long-term evolution experiment with *Escherichia*
 770 *coli*. *mBio* **5**:e01377-01314. <http://dx.doi.org/10.1128/mBio.01377-14>
 771
- 772 64. **Tatusov RL, Galperin MY, Natale DA, Koonin EV.** 2000. The COG database: a
 773 tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res*
 774 **28**:33–36. <https://doi.org/10.1093/nar/28.1.33>
 775
 776

777 Acknowledgements

778
 779 This work was funded by the Genoscope-CEA, a CNRS Chaire d'Excellence (to
 780 ACT), and the Agence Nationale de la Recherche Grant ANR-16-CE05-0020 (to ACT).
 781 We thank P Marlière for the GM3, E Darii for mass spectrometry, and the Genoscope-
 782 CEA sequencing platform for DNA and RNA sequencing.
 783
 784

785 Competing Interests

786 The authors declare no competing financial interests.
 787
 788

789 Data Availability

790 The authors confirm that all data underlying the findings are fully available without
 791 restriction. Sequencing files in FASTQ format are available in the European Nucleotide
 792 Archive under study accession ERP018602 (RNA-seq) and ERP018603 (whole genome
 793 sequencing).
 794
 795

796 Figures

797
 798 **Fig 1** *C. phytofermentans* growth GS2 glucose medium containing different
 799 concentrations of **A** acetate, **B** formate, **C** 5-hydroxymethylfurfural, **D** furfural, **E**
 800 coumarate, **F** ferulate, **G** vanillate, **H** 4-hydroxybenzoic acid, **I** vanillin, **J** benzaldehyde,
 801 **K** syringaldehyde, **L** catechol. Colors show inhibitor concentrations (g L⁻¹): 0 (gray), 1
 802
 803

804 (purple), 2 (blue), 3 (dark green), 5 (light green), 10 (yellow), 20 (orange), and 30 (red).
 805 Data shows mean cell density (OD_{600}) of 4 cultures \pm standard deviation (sd).

806
 807 **Fig 2** *C. phytofermentans* growth and gene expression during ferulate stress. **A** Growth
 808 in log phase cultures in medium either lacking ferulate (-ferulate, circles) or containing 2
 809 g L⁻¹ ferulate (+ferulate, triangles). Points are mean cell density (OD_{600}) of duplicate
 810 cultures \pm sd with red points showing times sampled for RNA-seq: t=0 h (immediately
 811 before dilution), t=0.5 h, and t=4 h. mRNA expression from cultures at **B** t=0h, **C** t=0.5 h,
 812 **D** t=4 h. Differentially-expressed genes in **B-D** are red triangles; unchanged genes are
 813 gray circles. Five most abundant COG functional categories (64) of differentially-
 814 expressed genes at **E** t=0.5 h and **F** t=4 h. Positive y-axis is up-regulated genes and
 815 negative y-axis is repressed genes.

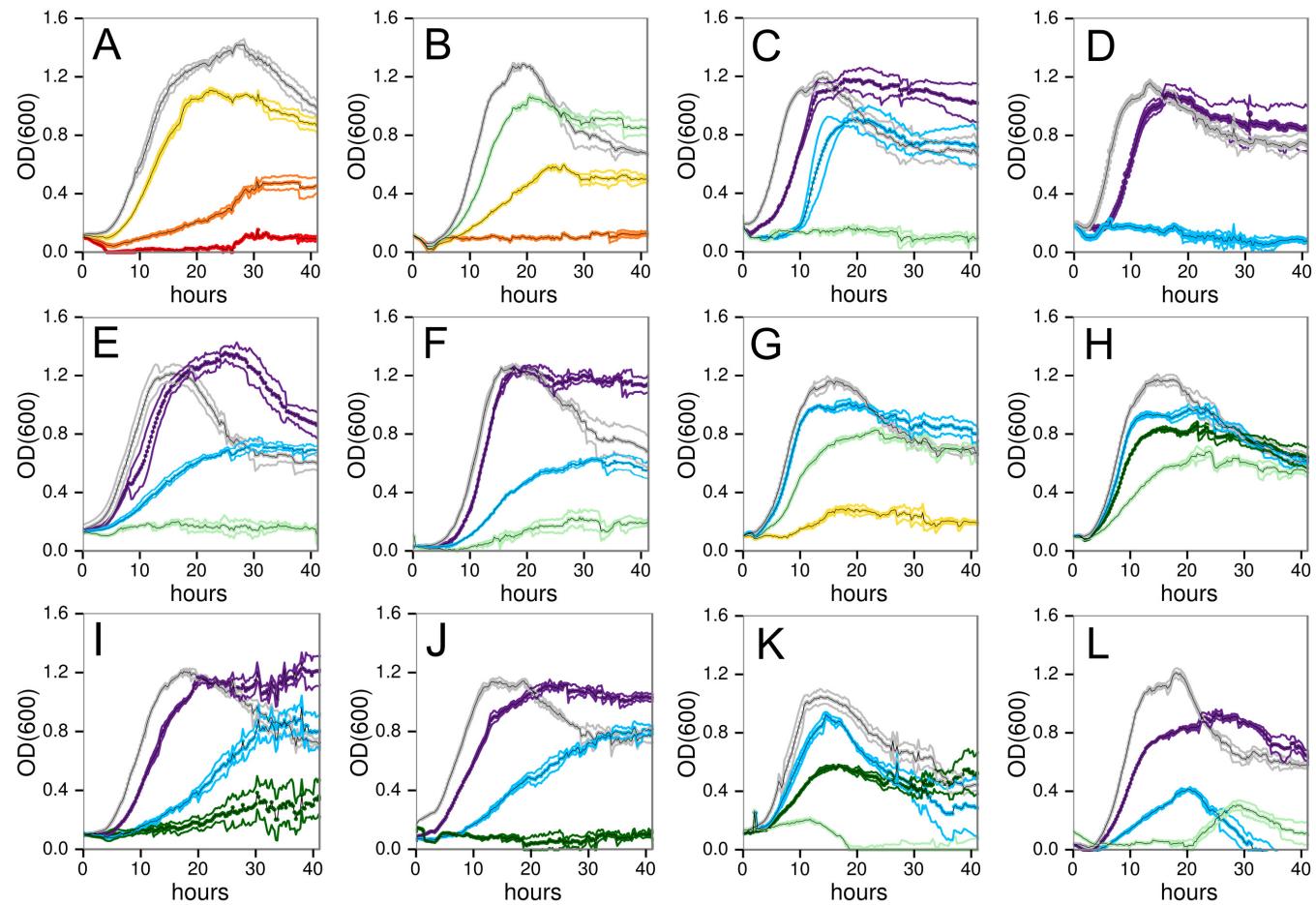
816
 817 **Fig 3** Growth improvement of *C. phytofermentans* GM3 strains in ferulate medium. **A**
 818 Cells were acclimated to increased ferulate using medium swap mode, a chemostat
 819 with dilutions of stressing medium if density exceeds a threshold and otherwise with
 820 relaxing medium. **B** Growth rate was improved using turbidostat mode in which the
 821 culture was diluted each time it reached the threshold. In **A** and **B**, dashed red lines
 822 show cell density threshold (OD_{880} 30) and red arrows are when the growth chamber
 823 was sterilized. **C** *C. phytofermentans* growth rate over 93 day GM3 experiment in
 824 medium with increasing ferulate concentrations (shown above plot). Shaded areas are
 825 periods of medium-swap with fixed 6h generation time. Black line shows average daily
 826 generation time (h) during turbidostat growth selection. Red points are sample times for
 827 physiology and genome sequences (CFY1, CFY2, CFY3). Batch culture growth (OD_{600})
 828 of **D** wild-type and clones **E** CFY1A, **F** CFY2C, and **G** CFY3E in GS2 glucose medium
 829 containing either 0 (gray), 1 (purple), 2 (blue), 4 (yellow), or 6 (red) g L⁻¹ ferulate. Data
 830 shows mean cell density (OD_{600}) of 4 cultures \pm sd.

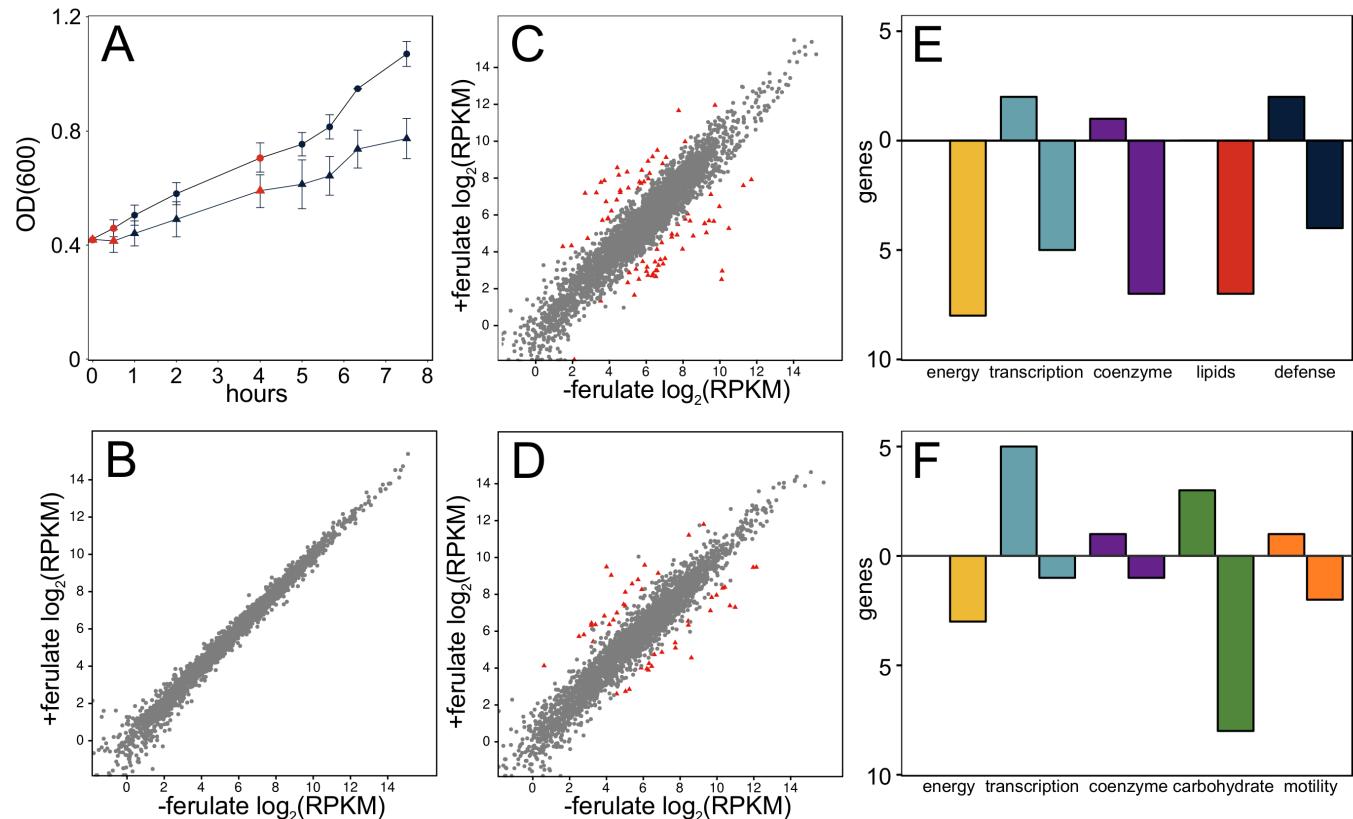
831
 832 **Fig 4** Growth physiology and membrane fatty acid composition of *C. phytofermentans*
 833 WT and GM3 strains. Rate of cellulose degradation by CFY3E (blue triangles) and WT
 834 (red circles) in **A** medium lacking ferulate and **B** medium supplemented with 2 g L⁻¹
 835 ferulate. **C** Plasmalogen content expressed as percent of total fatty acids of WT and
 836 GM3 strains grown in presence (+F) or absence (-F) of ferulate. **D** Cellular fatty acid
 837 profiles of log phase WT, CFY1B, CFY2C, and CFY3F cultures in -ferulate medium.
 838 Fatty acids classified by acyl chain length (C12-20) and whether acyl chains were
 839 saturated (dark blue), unsaturated (light blue), hydroxylated (green), cyclopropane (red),
 840 branched (yellow). Data in **A-C** show mean of duplicate cultures \pm sd.

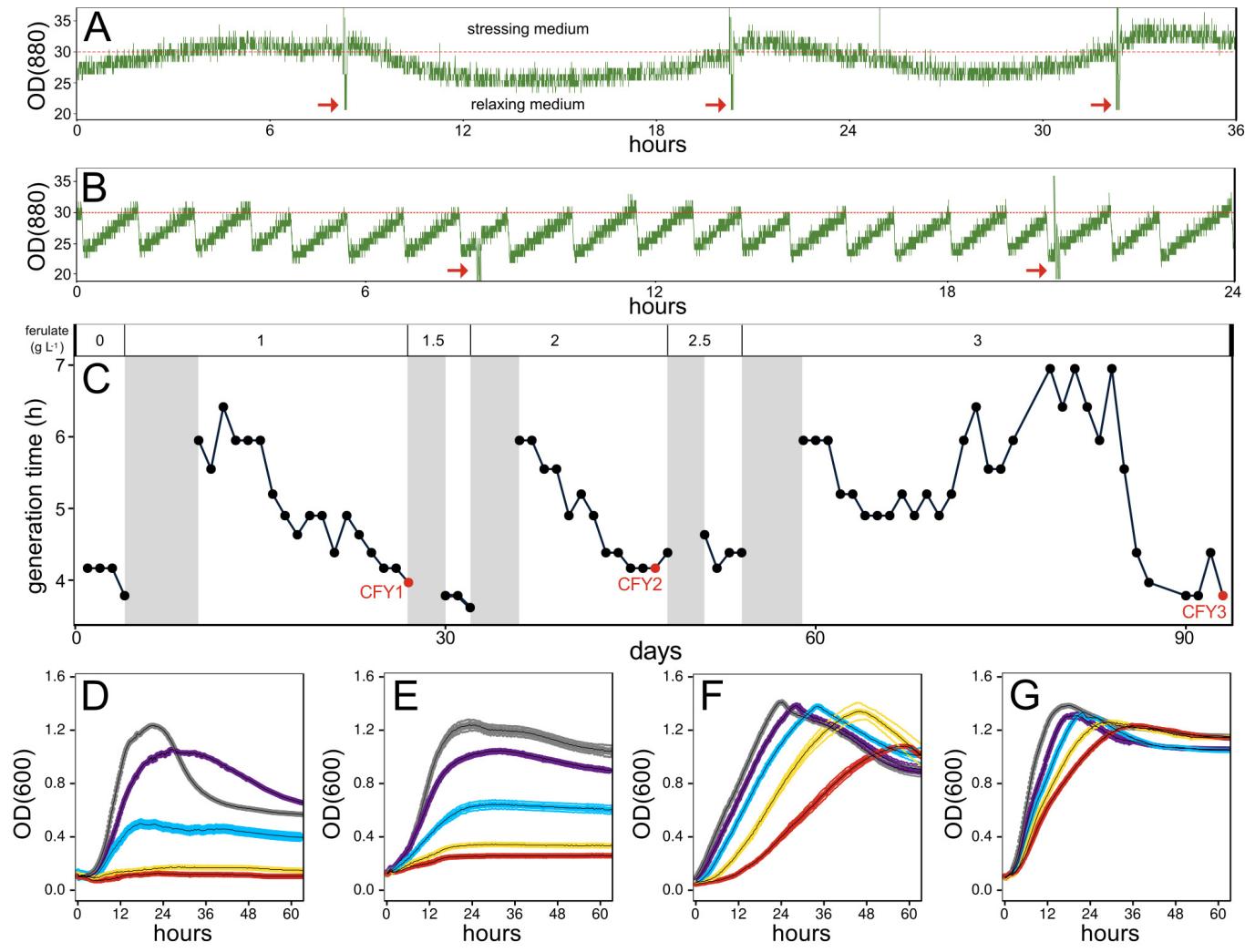
841
 842 **Fig 5** Small-scale genome differences in *C. phytofermentans* GM3 strains. **A**
 843 Accumulation of single nucleotide variants and small indels in the genomes of clones
 844 isolated from the CFY1-3 time points. mRNA expression of the **B** *cphy2241-3* operon
 845 and the **C** *cphy1459-cphy1465* genes in WT (blue) and CFY3E (red) strains. Genes are
 846 shown above plots with asterisks denoting positions of DNA changes. Expression was
 847 measured by qRT-PCR and quantified as $2^{-\Delta Ct}$ normalized to 16S rRNA expression; bars
 848 show means of triplicate measurements \pm sd.

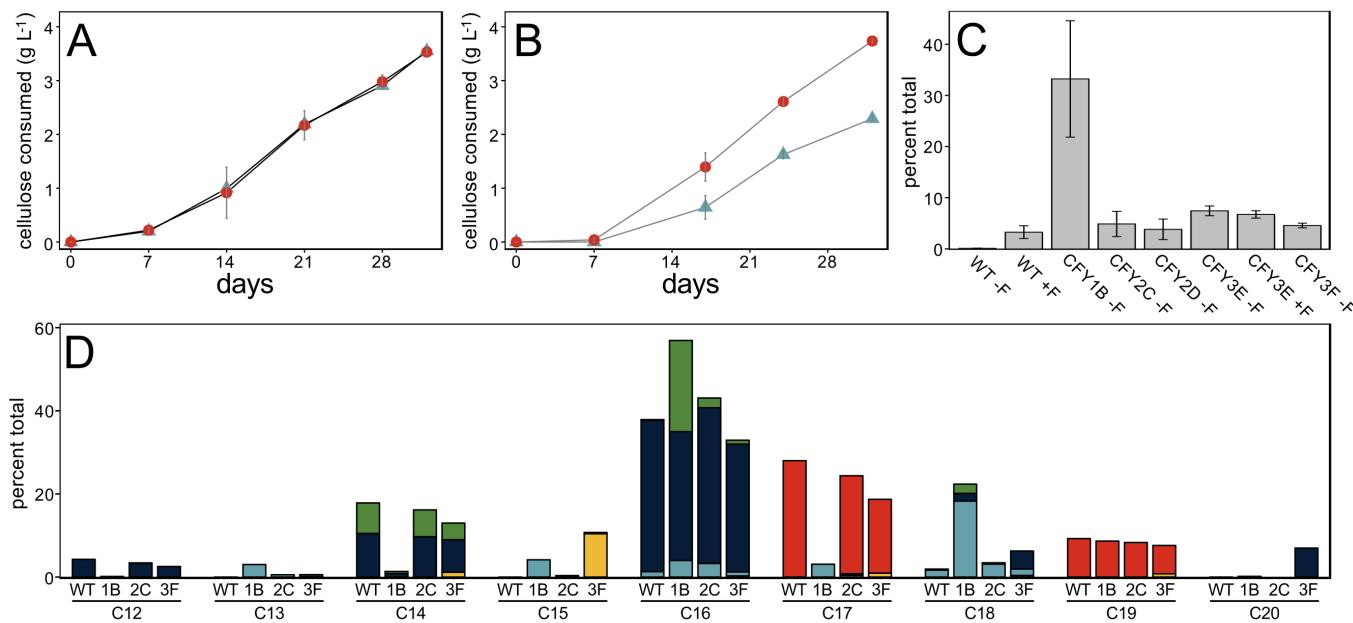
849

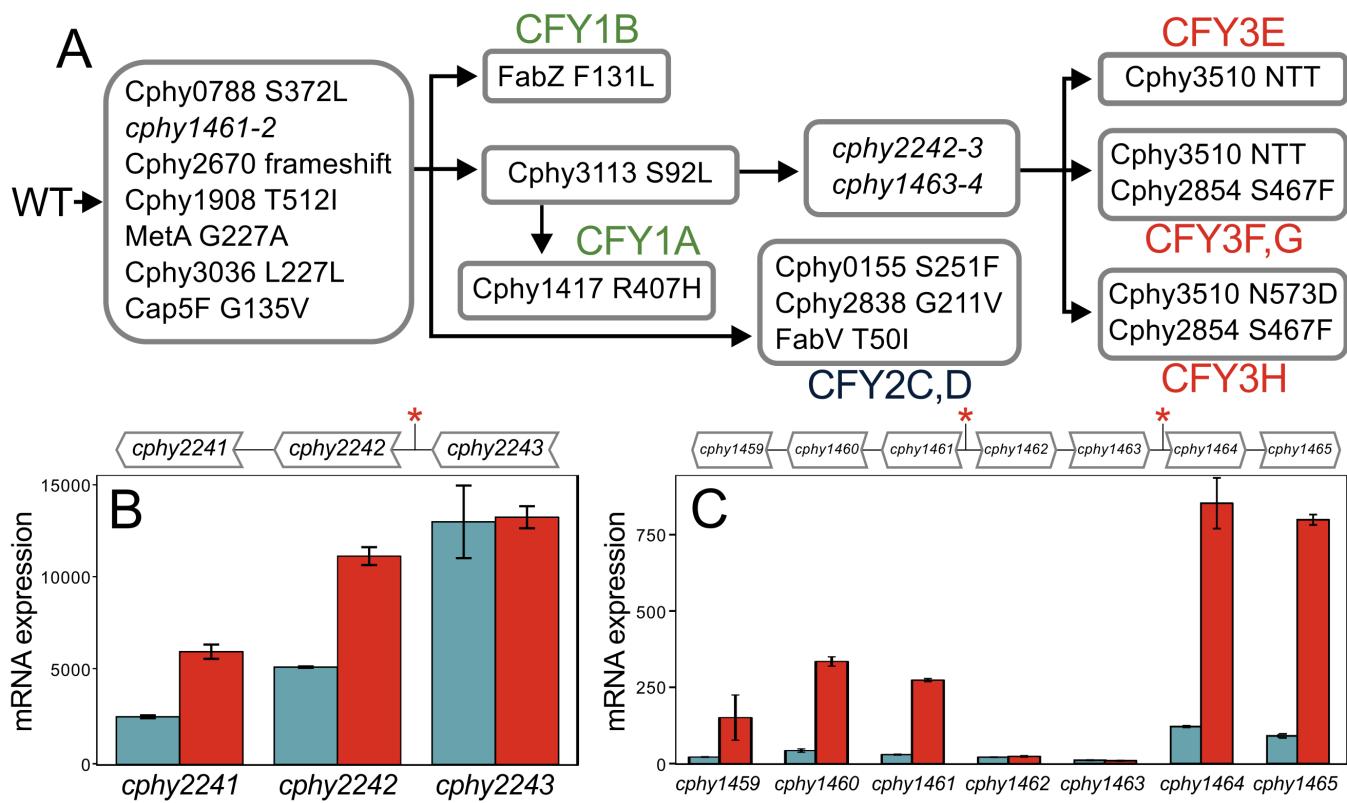
850 **Fig 6** ISL3-associated genome changes in CFY strains. **A** Read coverage showing
851 deletion of the ISL3-1 element including the *cphy1314* transposase gene (CFY1A
852 shown). **B** Read coverage in CFY2C and CFY2D (shown) reveals the duplication of a
853 333 kb region from *cphy2178-cphy2461* (genome position 2,689,393-3,023,191). **C** The
854 *cphy2178-cphy2461* region is a tandem duplication joined by an ISL3-2 element.
855 Positions of single molecule restriction fragments >400 kb shown by BioNano optical
856 mapping to span the duplicated region are shown. **D** Relative abundance of the
857 *cphy2178-ISL3-cphy2461* junction in the GM3 culture from day 6 to 83. Genomic
858 locations of the PCR primers are shown in **C**. Abundance was measured by qPCR,
859 calculated as $2^{-\Delta Ct}$ of the junction fragment in the GM3 culture relative to the purified
860 CFY2C, which has the junction present in 100% cells.
861

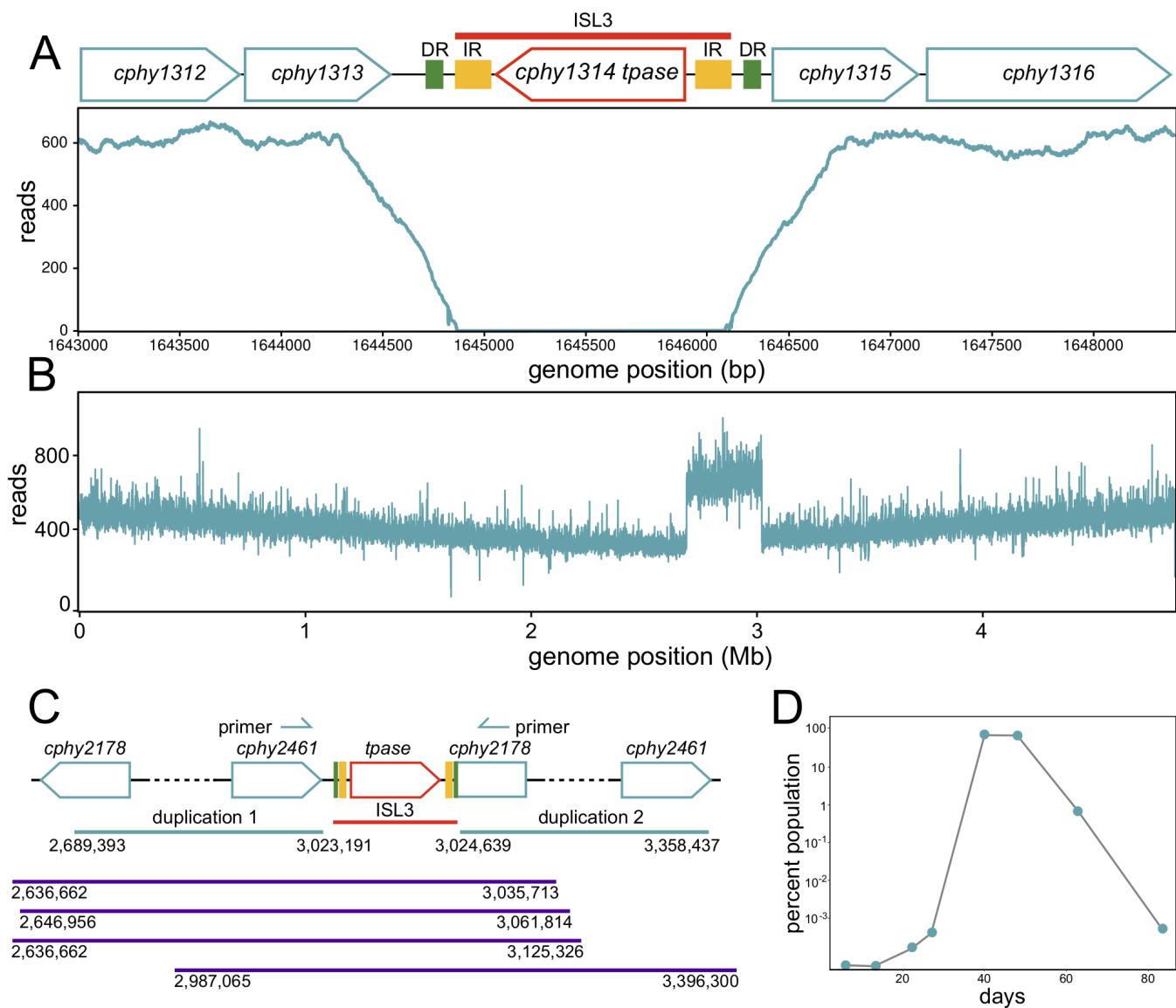












IV DISCUSSION

Improving enzymes catalytic capacities is a persisting challenge in biotechnology. Directed evolution *in vivo* in continuous culture is an alternative to the numerous *in vitro* methods of enzyme mutation and activity screening.

The main objective of this study was to determine the experimental parameters enabling the selection of enhanced enzymes variants. The α -KAO enzyme family, catalyzing reactions difficult to realize with the methods of classical organic chemistry, was chosen as a target. These enzymes hydroxylate non-activated C-H bonds in a stereospecific manner, giving rise to synthons of value for chemical synthesis.

The reaction catalyzed by the α -KAOs produces succinate via decarboxylation of α -KG. A generic screen based on succinate auxotrophy was implemented in *E. coli*. Succinate is a precursor for essential C4-compounds of the metabolism. The expression of a member of the α -KAOs, isoleucine α -KG dioxygenase (Ido) from *Bacillus cereus*, led to complementation of the succinate auxotrophy depending on the presence of isoleucine and α -KG in the growth medium, thus validating the selection screen.

The screen was then used to evolve α -KAOs which had been identified in a sequence-based screen and shown to hydroxylate basic amino acids, *i.e.* L-lysine and L-ornithine *in vitro* (Baud *et al.*, 2014).

The activity of these enzymes (Odo and Kdo1 to Kdo5) expressed from a high copy number plasmid was found insufficient for complementation of the succinate auxotrophy of the selection strain.

The determination of enzymatic activity parameters revealed K_M and turnover numbers typical for enzymes involved in pathways of secondary metabolism. Table 36 compares the activity constants determined for Kdo2 with constants published for a number of α -KAOs with different substrate specificities. The affinity and activity parameters of Kdo2 are within the range of these enzymes, which are involved in the secondary metabolism of the host cells. Surprisingly, the catalytic efficiency measured in this study for Ido is lower than the ones for Kdos despite the *in vivo* observations that the selection strain showed succinate complementation when expressing Ido on a high copy number plasmid (Table 27) and not when expressing Kdos (Table 18). This suggests that the catalytic efficiency of Kdos is not the limiting factor explaining the lack of complementation.

Table 36: Kinetic parameters of various dioxygenases

| Enzyme | Organism | Substrates | Kinetic parameters | | | Source |
|-----------------------------|----------------------------------|------------------------------------|------------------------|--|---|---------------------------------|
| | | | K _M (mM) | k _{cat} (s ⁻¹) | k _{cat} /K _M (s ⁻¹ .M ⁻¹) | |
| L-lysine hydroxylase (Kdo2) | <i>Chitinophaga pinensis</i> | L-lysine | 0.096 | 1.79 | 1.86 x 10 ⁴ | This study |
| L-isoleucine hydroxylase | <i>Bacillus cereus</i> | L-isoleucine | 0.096 | 0.05 | 5.2 x 10 ² | This study |
| L-asparagine hydroxylase | <i>Streptomyces coelicolor</i> | L-asparagine | 0.48 | 4.98 | 1.04 x 10 ⁴ | (Stricker <i>et al.</i> , 2007) |
| L-arginine Hydroxylase | <i>Streptomyces vinaceus</i> | L-arginine | 3.4 | 43.5 | 1.28 x 10 ⁴ | (Helmetag <i>et al.</i> , 2009) |
| L-proline 3-hydroxylase | <i>Streptomyces sp.</i> | L-proline | 0.4 | 3.2 | 0.8 x 10 ⁴ | (Mori <i>et al.</i> , 1997) |
| L-proline 4-cis-hydroxylase | <i>Mesorhizobium loti</i> | L-proline | 0.54 | 25 | 4.6 x 10 ⁴ | (Hara and Kino, 2009) |
| taurine dioxygenase | <i>Escherichia coli</i> | taurine | 0.052 | 3 | 5.8 x 10 ⁴ | (Luo <i>et al.</i> , 2006) |
| clavaminate synthase | <i>Streptomyces clavuligerus</i> | proclavaminate | 0.19 | 0.76 | 0.40 x 10 ⁴ | (Salowe <i>et al.</i> , 1990) |
| γ-butyrobetaine dioxygenase | <i>Homo sapiens</i> | 4-(trimethylammonio) butanoic acid | 0.16 | 5.3 | 3.3 x 10 ⁴ | (Rydzik <i>et al.</i> , 2012) |
| flavonol synthase | <i>Arabidopsis thaliana</i> | dihydroquercetin | 0.059 | 0.056 | 1.7 x 10 ⁴ | (Chua <i>et al.</i> , 2008) |

In the absence of growth, a medium SWAP regime establishing a gradient of succinate, which decreases as a response to accelerated growth of the selection strain, is adequate to select enzyme variants with higher activity. A high copy number plasmid enabling stable propagation under the non-selective culture conditions, which prevails during growth in the relaxing medium containing succinate, was used for α-KAO expression. Conservation of the plasmid is ensured through co-expression of an essential gene of the DNA precursor synthesis, thymidylate synthase. The chromosomal copy of the gene *thyA* has been deleted in the selection strain for this purpose.

Selection strains expressing gene *kdo2* adapted to the stressing medium in the SWAP regime and acquired prototrophy for succinate (strain G4097). Sequencing of the plasmid showed no mutation in the *kdo2* coding sequence but revealed the absence of the *thyA* gene (plasmid pGEN1126). The determination of *kdo2* gene copy number in the evolved cells showed an increase with respect to the initial strain G3986. The concomitant increase in *kdo2* transcript level strongly suggests that the copy number of the plasmid rose to enable a higher expression of the gene. For this augmentation, the transfer of the gene *thyA* from the plasmid to the chromosome seemed to be a prerequisite to avoid toxic overexpression of this gene.

The truncated plasmid pGEN1126 has been introduced into the selection strain G3700 selecting for growth on lysine without succinate (Table 21, strain G4096). The *kdo2* gene copy number was 11 times higher in this strain than in strain G3986 harboring the initial *thyA+* plasmid construct (plasmid pGEN1080). Given a plasmid copy number of about 100 in strain G3986, as calculated from the *thyA* gene number, strain G4096 contains about 1000 copies of the truncated plasmid. Obviously, complementation required a large amount of *kdo2* gene templates, pointing to some inefficiency in the expression mechanisms of this gene. Also, misfolding or low stability of the protein might play a role.

The strain obtained grew well on lysine without succinate and therefore was no longer a starting point for the selection of enhanced kdo-variants. The studies on the appearance and fixation of mutations cancelling a loss of function phenotype (section III.1) have shown that the chromosomal format of gene expression is suitable for selecting beneficial point mutations in genes specifying enzymes under selection. Albeit a multicopy plasmid offers a higher probability to generate beneficial variants, the metabolic burden inherently linked to the copy number of a plasmid can be a hurdle for the directed evolution. In addition, the process of fixation of a beneficial mutation in the evolving population has to be considered. As was shown for the *metA* reversion experiment, a threshold ratio of mutated over unchanged plasmid copies needs to be surpassed for the manifestation of the phenotype (Section III.1.5, Figure 31). This can be an obstacle for the fixation of mutations with only a small beneficial increment on growth. Also, the chromosomal format has the advantage of stability since the gene copy cannot be lost.

Gene *kdo2* was inserted into the chromosome of the selection strain and evolved in a medium SWAP regime as was done for the plasmid context. Two parallel cultures, KDB 15 and KDB 16, were performed. Both evolved towards succinate prototrophy. While the time needed for adaptation was different for the two cultures (255 and 439 generations), a sharp increase in the number of dilution pulses with stressing medium leading to prototrophy was observed for both cultures (Figure 48), suggesting a single or few mutational event to be responsible for the evolution. Sequencing of the *kdo2* insertion locus of culture isolates showed an unchanged coding sequence. However, short deletions upstream of the gene were found, in the N-terminal coding region of gene *pdeH*. This gene codes for an enzyme involved in the inactivation of the second messenger c-di-GMP which has pleiotropic effects on cell proliferation. Transcript copy number of gene *kdo2* was greatly enhanced (62 times) as compared with the initial strain G4160. Whether this effect is due to an altered phosphodiesterase activity of the PdeH enzyme variants or to a direct effect on RNA polymerase template recognition has not been studied. Again, a mutational event effecting gene expression and not enzymatic activity parameters was selected.

The study was extended for the enzyme Ido hydroxylating isoleucine. This enzyme, when expressed from a high copy number plasmid, complemented the succinate auxotrophy of the selection strain, but failed to do so when expressed from one chromosomal copy. A medium SWAP evolution of the corresponding strain resulted in succinate prototrophic cells depending on isoleucine for growth. Sequencing of the *ido* chromosomal locus of evolved isolates revealed the presence of an IS2 element known to enhance transcription of downstream genes, while the coding region was found unchanged. The *ido* transcript copy number was strongly increased (87 times) as compared to the initial strain G4110, suggesting that, as was seen for the KDB cultures, evolution "acted" on the regulation of gene expression rather than enzymatic activity.

A more stringent selection context for succinate complementation was also tested. Glutamate, precursor of α -KG, was provided as sole carbon source and the succinate produced by the α -KAO provides all carbon necessary for growth. Both genes *kdo2* and *ido* were inserted into the chromosome of the selection strain and evolved in a medium SWAP regime. After 500 generations, the daily ratio of stressing medium was still stable

(about 40%) suggesting that no beneficial mutations, neither in the coding region nor in chromosomal loci regulating expression of the genes *kdo2* or *ido*, have appeared.

When comparing those observations with the results of the study on the fixation of a beneficial mutation (section III.1), it is noteworthy that it took less than 50 generations in medium SWAP regime for the reversion of gene *metA* to be fixed in the population regardless of the expression format. This suggests that the number of protein residues in α -KAOs which, upon a change to another amino acid, would improve enzyme activity is limited, and the beneficial change might require more than one base pair to mutate.

The selection strains growing with glutamate and either lysine or isoleucine as second substrate for the respective α -KAO expressed from the chromosome were evolved to grow on analogs of the amino acid substrate. The experiments were conducted using the medium SWAP regime of continuous culture. To enable selection of enzyme variants, the relaxing medium contained succinate but not the natural substrate. In general, a shift in substrate specificity of an enzyme provoked by mutations increasing the affinity or the turnover number for a substrate analog decreases the kinetic parameters for the natural substrate. Such mutations would not be fixed in a medium SWAP culture setup where the relaxing growth conditions would depend on wild-type activity.

For Kdo2, continuous cultures were conducted for two analogs of L-lysine, thialysine (S-2-aminoethylcysteine) and 6-aminocaproic acid. Kdo2 was previously found to have some activity on thialysine (Baud *et al.*, 2014). However, this compound is toxic for *E. coli* due to its incorporation into proteins where it can replace lysine (Girolamo *et al.*, 1982). The concentration of thialysine in the culture media of the medium SWAP was fixed at 1.5 mM. Thialysine exhibits only a moderate growth inhibition at this concentration. The cultures evolved only very slightly, gaining 10% of stressing pulses per day after 600 generations. The *kdo2* gene sequence of isolates of the cultures obtained at this time point was unchanged. The cells did not grow in the presence of lysine as was the case for the initial strain. The adaptation was, therefore, most probably due to a loss of function, either of Kdo2 or of the transport systems which channels lysine/thialysine into the cells.

An unusual evolution pattern was observed for the cultures conducted for 6-aminocaproic acid adaptation. The stressing/relaxing dilution ratio showed strong oscillations and finally reached 100% stressing pulses. The cell growth in suspension could not be reproduced on solid medium; no isolates could therefore be obtained. The coding sequence of the gene *kdo2* was found unchanged when a DNA preparation of a culture sample was sequenced. A deeper analysis of the cell population, for instance through sequencing of the DNA of single cells, would be necessary to better understand the evolution of these cultures.

The lysine dependent α-KAOs, which were evolution targets in this study, have not been characterized in detail. By contrast, the isoleucine α-ketoglutarate dioxygenase from *Bacillus thuringiensis* has been studied more extensively; it was shown that isoleucine analogs like norvaline function as substrates for Ido (Hibi *et al.*, 2011). Evolution in medium SWAP continuous culture regime for the growth of the selection strain overexpressing the gene *ido* from the chromosome (strain G4383) with norvaline yielded suspensions stably proliferating without succinate. Isolates harbored *ido* gene variants with a changed coding sequence, affecting either residue 62 (alanine to proline) or 65 (glutamate to lysine). In addition, some succinate prototrophic isolates contained an unchanged *ido* gene. Obviously, the medium SWAP regime permitted cells with different adaptive genotypes to prevail in the population. Similar observations have been made for other medium SWAP evolution experiments (Marlière *et al.*, 2011). This continuous culture regime resembles a chemostat in that the doubling time of the culture is fixed, lowering the selective pressure which – in a turbidostat – would lead to a homogenous population of the fastest growing cells.

Expression of the E65K *ido* from a high copy number plasmid in the selection strain complemented succinate auxotrophy in the presence of either isoleucine or norvaline. As judged from growth profiles on solid medium, complementation with norvaline, but not with isoleucine, seemed to require additional adaptations of the cells.

Ido WT allows succinate complementation in the presence of isoleucine but not in the presence of norvaline in the strain G4383, suggesting that the activity of Ido towards norvaline is insufficient to allow cellular growth.

The kinetic characterization of Ido E65K variant showed that the k_{cat} for norvaline was increased by a factor of 2.4 as compared to the wild-type enzyme. These results can be linked to those published by Hibi and colleagues who reported for Ido WT that the formation reaction rate of 4-OH-ile was 2.6 times higher than the one measured for OH-norv. In those conditions, an increase in the catalytic activity of Ido by a factor of 2.4 in the presence of norvaline should allow the production of enough succinate to ensure cellular growth. This is precisely the effect of the mutation on the catalytic constant k_{cat} for norvaline. The catalytic efficiency (k_{cat}/k_M) of Ido:E65K is 4 times higher for norvaline than the one of Ido WT. However, it is noticeable that the catalytic constant of Ido WT found for isoleucine and norvaline are similar, diverging from the published data by Hibi and colleagues. It is important to stress that the quantitation of OH-norv could not be performed with a reference standard as described in section II.3.3.2.2. This could have led to overestimate the k_{cat} .

The conservation of the amino acid residues at each position of Ido was evaluated using a set of 1859 non-redundant sequences of α -KAOs collected with BLAST CD-search and aligned with Cobalt alignment analyzer. The amino acids A62 and E65 are weakly conserved (7.2% and 4.3%, respectively). Interestingly, some neighboring positions are highly conserved (89.9% for L58, 81.1% for D61, 93% for R67 and 88.4% for R69) and proline at position 62 is significantly more frequent (26%). The mutated amino acids seem to belong to a rather conserved region of the proteins. However, in the absence of any 3D structure for Ido, the selected mutations cannot easily be linked to possible functional alterations.

A BLAST alignment of the sequences of Ido and of the L-proline cis-4-hydroxylase of *Mesorhizobium loti* (Pdo) shows weak homology between the two proteins (17% identities, 29% positives over total length) but local higher similarities in our region of interest, as shown in the following alignment:

| | | | |
|-----|----|-------------------------------|-----|
| Ido | 57 | NLEDDAYGECRARSYSRYIKYVDSPDYI | 84 |
| | | N D + + + Y+R I+ ++ I | |
| Pdo | 77 | NELLDNHFDRDSIRYARIIIRISENACII | 104 |

The structural analysis of L-proline cis-4-hydroxylase in the presence of the natural substrate L-proline points to the interaction of Pdo R93 with the carboxylic acid function of L-proline (Koketsu *et al.*, 2015). One can hypothesize that the corresponding R73 of Ido might interact with Ido substrates and that these interactions might be modulated by the selected mutations A62P and E65K.

In addition, the directed evolution of an organism in continuous culture in the GM3 was studied. The plant fermenting strain, *Clostridium phytofermentans*, was adapted in medium SWAP regime to increased concentration of ferulic acid, a phenolic inhibitor. Ferulate resistant isolates exhibited transcriptional adaptations similar to what was obtained for the evolution of Kdo2 and Ido. Intergenic changes arose affecting the expression levels of adjacent genes in response to ferulic acid stress.

Also, IS elements related to duplication events were found to arise early in the evolution then gradually declined at the end of the experiment. These results show important differences in the cellular processes implicated in the mRNA response to short-term stress and the DNA changes enabling long-term adaptation. Strains with early transcriptional adaptation rapidly take over the population, allowing them to subsequently adapt with long-term DNA changes.

The following lessons can be drawn from the experimental work performed for studying the conditions of the *in vivo* mutagenesis of proteins:

- In the case that heterologous genes have to be evolved, mutation processes modifying expression are favored. In general, heterologous genes are less well expressed than native genes which co-evolved with the strain. Optimizations can involve codon usage, but also protein folding and stability.
- The chromosome is a valid expression context for the target gene, avoiding loss of independent replicons and assuring rapid fixation of adaptive mutations in the population.
- The medium SWAP regime of continuous culture enables the selection of mutations conveying only small increments in enzyme activity.

APPENDIX

Appendix 1: List of all strains cited in this study

| Strain | Genotype | Derived from |
|--------|--|--------------|
| G3513 | C> $\Delta ArgF-lac \Delta gltA \Delta icd::aad+$ $\Delta pck \Delta ppc \Delta aceBAK \Delta glcDEFGB \Delta sfcA \Delta maeB$ $\Delta puuE \Delta gabT \Delta sucA::kan+$ | MG1655 |
| G3542 | C> $\Delta ArgF-lac \Delta gltA \Delta icd::aad+$ $\Delta pck \Delta ppc \Delta aceBAK \Delta glcDEFGB \Delta sfcA \Delta maeB$ $\Delta puuE \Delta gabT \Delta sucA::kan+$ P> pGEN966 (pSP100::odo+) | G3513 |
| G3543 | C> $\Delta ArgF-lac \Delta gltA \Delta icd::aad+$ $\Delta pck \Delta ppc \Delta aceBAK \Delta glcDEFGB \Delta sfcA \Delta maeB$ $\Delta puuE \Delta gabT \Delta sucA::kan+$ P> pGEN967 (pVDM18::odo+) | G3513 |
| G3544 | C> $\Delta ArgF-lac \Delta gltA \Delta icd::aad+$ $\Delta pck \Delta ppc \Delta aceBAK \Delta glcDEFGB \Delta sfcA \Delta maeB$ $\Delta puuE \Delta gabT \Delta sucA::kan+$ P> pGEN968 (pSP100::kdo1+) | G3513 |
| G3545 | C> $\Delta ArgF-lac \Delta gltA \Delta icd::aad+$ $\Delta pck \Delta ppc \Delta aceBAK \Delta glcDEFGB \Delta sfcA \Delta maeB$ $\Delta puuE \Delta gabT \Delta sucA::kan+$ P> pGEN969 (pVDM18::kdo1+) | G3513 |
| G3547 | C> $\Delta ArgF-lac \Delta gltA \Delta icd::aad+$ $\Delta pck \Delta ppc \Delta aceBAK \Delta glcDEFGB \Delta sfcA \Delta maeB$ $\Delta puuE \Delta gabT \Delta sucA::kan+$ P> pGEN971 (pVDM18::kdo2+) | G3513 |
| G3548 | C> $\Delta ArgF-lac \Delta gltA \Delta icd::aad+$ $\Delta pck \Delta ppc \Delta aceBAK \Delta glcDEFGB \Delta sfcA \Delta maeB$ $\Delta puuE \Delta gabT \Delta sucA::kan+$ P> pGEN972 (pSP100::kdo3+) | G3513 |
| G3549 | C> $\Delta ArgF-lac \Delta gltA \Delta icd::aad+$ $\Delta pck \Delta ppc \Delta aceBAK \Delta glcDEFGB \Delta sfcA \Delta maeB$ $\Delta puuE \Delta gabT \Delta sucA::kan+$ P> pGEN973 (pVDM18::kdo3+) | G3513 |
| G3685 | Isolate sampled from culture of G3513 at generation 1991 | G3513 |
| G3686 | Isolate sampled from culture of G3513 at generation 1991 | G3513 |
| G3700 | C> $\Delta ArgF-lac \Delta gltA \Delta icd::aad+$ $\Delta pck \Delta ppc \Delta aceBAK \Delta glcDEFGB \Delta sfcA \Delta maeB$ $\Delta puuE \Delta gabT \Delta sucA$ | G3513 |
| G3724 | C> $\Delta metA$ | MG1655 |
| G3748 | C> $\Delta metA$ P> pGEN1002 (pSP100::metA:C142R+) | G3724 |
| G3749 | C> $\Delta metA$ P>pGEN1005 (pSP100::metAH235S+) | G3724 |
| G3752 | C> $\Delta ArgF-lac \Delta gltA \Delta icd::aad+$ $\Delta pck \Delta ppc \Delta aceBAK \Delta glcDEFGB \Delta sfcA \Delta maeB$ | G3700 |

| Strain | Genotype | Derived from |
|--------|--|--------------|
| G3841 | $\Delta puuE \Delta gabT \Delta sucA \Delta thyA::erm+$ C> $\Delta metA$ P> pGEN1041 (pVDM18::metA:K46R+) | G3724 |
| G3842 | C> $\Delta metA$ P> pGEN1042 (pVDM18::metA:K47R+) | G3724 |
| G3844 | C> $\Delta metA$ P> pGEN1044 (pVDM18::metA:C142Y+) | G3724 |
| G3845 | C> $\Delta metA$ P> pGEN1045 (pVDM18::metA:C142R+) | G3724 |
| G3848 | C> $\Delta metA$ P>pGEN1048 (pVDM18::metA:H235S+) | G3724 |
| G3850 | C> $\Delta metA$ P>pGEN1050 (pVDM18::metA:E237A+) | G3724 |
| G3880 | C> $\Delta metA, \Delta kdgK::metA:E237D+$ | G3724 |
| G3886 | Isolate sampled from culture of G3513 at generation 1620 | G3513 |
| G3887 | Isolate sampled from culture of G3513 at generation 1620 | G3513 |
| G3935 | C> $\Delta ArgF-lac \Delta gltA \Delta icd::aad+$ $\Delta pck \Delta ppc \Delta aceBAK \Delta glcDEFGB \Delta sfcA \Delta maeB$ $\Delta puuE \Delta gabT \Delta sucA \Delta thyA::erm+$ P> pGEN1066 (pEVL404::odo+) | G3752 |
| G3951 | C> $\Delta ArgF-lac \Delta gltA \Delta icd::aad+$ $\Delta pck \Delta ppc \Delta aceBAK \Delta glcDEFGB \Delta sfcA \Delta maeB$ $\Delta puuE \Delta gabT \Delta sucA \Delta thyA::erm+$ P> pGEN1070 (pEVL404::kdo1+) | G3752 |
| G3982 | C> $\Delta ArgF-lac \Delta gltA \Delta icd::aad+$ $\Delta pck \Delta ppc \Delta aceBAK \Delta glcDEFGB \Delta sfcA \Delta maeB$ $\Delta puuE \Delta gabT \Delta sucA::kan+$ P> pGEN1074 (pSP100::kdo2+) | G3513 |
| G3984 | C> $\Delta ArgF-lac \Delta gltA \Delta icd::aad+$ $\Delta pck \Delta ppc \Delta aceBAK \Delta glcDEFGB \Delta sfcA \Delta maeB$ $\Delta puuE \Delta gabT \Delta sucA::kan+$ P> pGEN1076 (pVDM18::kdo4+) | G3513 |
| G3986 | C> $\Delta ArgF-lac \Delta gltA \Delta icd::aad+$ $\Delta pck \Delta ppc \Delta aceBAK \Delta glcDEFGB \Delta sfcA \Delta maeB$ $\Delta puuE \Delta gabT \Delta sucA \Delta thyA::erm+$ P> pGEN1080 (pEVL404::kdo2+) | G3752 |
| G3990 | C> $\Delta ArgF-lac \Delta gltA \Delta icd::aad+$ $\Delta pck \Delta ppc \Delta aceBAK \Delta glcDEFGB \Delta sfcA \Delta maeB$ $\Delta puuE \Delta gabT \Delta sucA \Delta ilv::Cm+$ | G3700 |
| G3997 | C> $\Delta metA, \Delta kdgK::metA+$ | G3724 |
| G4008 | C> $\Delta metA$ P> pGEN118 (pSP100::metA+) | G3724 |

| Strain | Genotype | Derived from |
|--------|---|--------------|
| G4009 | C> $\Delta metA$ P> pGEN1084 (pVDM18::metA+) | G3724 |
| G4010 | C> $\Delta metA$ P> pGEN1085 (pVDM18::metA:E237D+) | G3724 |
| G4011 | C> $\Delta metA$ P> pGEN1086 (pSP100::metA:E237D+) | G3724 |
| G4087 | C> $\Delta ArgF-lac \Delta gltA \Delta icd::aad+$ $\Delta pck \Delta ppc \Delta aceBAK \Delta glcDEFGB \Delta sfcA \Delta maeB$ $\Delta puuE \Delta gabT \Delta sucA::kan+$ P> pGEN1101 (pSP100::kdo4+) | G3513 |
| G4089 | C> $\Delta ArgF-lac \Delta gltA \Delta icd::aad+$ $\Delta pck \Delta ppc \Delta aceBAK \Delta glcDEFGB \Delta sfcA \Delta maeB$ $\Delta puuE \Delta gabT \Delta sucA::kan+$ P> pGEN1102 (pSP100::kdo5+) | G3513 |
| G4096 | C> $\Delta ArgF-lac \Delta gltA \Delta icd::aad+$ $\Delta pck \Delta ppc \Delta aceBAK \Delta glcDEFGB \Delta sfcA \Delta maeB$ $\Delta puuE \Delta gabT \Delta sucA$ P> pGEN1126 (pEVL404*::kdo2+) | G3700 |
| G4097 | Isolate from culture KDB 10 | G3986 |
| G4110 | C> $\Delta ArgF-lac \Delta gltA \Delta icd::aad+$ $\Delta pck \Delta ppc \Delta aceBAK \Delta glcDEFGB \Delta sfcA \Delta maeB$ $\Delta puuE \Delta gabT \Delta sucA \Delta ilv::Cm+$ $\Delta kdgK::ido+$ | G3990 |
| G4136 | C> $\Delta ArgF-lac \Delta gltA \Delta icd::aad+$ $\Delta pck \Delta ppc \Delta aceBAK \Delta glcDEFGB \Delta sfcA \Delta maeB$ $\Delta puuE \Delta gabT \Delta sucA \Delta ilv::Cm+$ P> pGEN997 (pSP100::ido+) | G3990 |
| G4160 | C> $\Delta ArgF-lac \Delta gltA \Delta icd::aad+$ $\Delta pck \Delta ppc \Delta aceBAK \Delta glcDEFGB \Delta sfcA \Delta maeB$ $\Delta puuE \Delta gabT \Delta sucA \Delta kdgK::kdo2+$ | G3700 |
| G4170 | Isolate from culture IDO 3 ($\Delta kdgK::ido^*$) | G4110 |
| G4173 | Isolate from culture IDO 6 ($\Delta kdgK::ido^*$) | G4110 |
| G4275 | Isolate from culture KDB 15 ($\Delta kdgK::kdo2^*$) | G4160 |
| G4276 | Isolate from culture KDB 15 ($\Delta kdgK::kdo2^*$) | G4160 |
| G4292 | Isolate from culture KDB 16 ($\Delta kdgK::kdo2^*$) | G4160 |
| G4343 | C> $\Delta aceAK \Delta sucA \Delta ppc \Delta puuE \Delta gabT$ | MG1655 |
| G4351 | C> $\Delta ArgF-lac \Delta gltA \Delta icd::aad+$ $\Delta pck \Delta ppc \Delta aceBAK \Delta glcDEFGB \Delta sfcA \Delta maeB$ $\Delta puuE \Delta gabT \Delta sucA \Delta kdgK::ido^*$ | G3700 |
| G4383 | Restitution of the <i>ilv</i> operon in strain G4170 | G4170 |
| G4434 | C> $\Delta ArgF-lac \Delta gltA \Delta icd::aad+$ $\Delta pck \Delta ppc \Delta aceBAK \Delta glcDEFGB \Delta sfcA \Delta maeB$ $\Delta puuE \Delta gabT \Delta sucA \Delta kdgK::kdo2^*$ | G3700 |
| G4490 | C> $\Delta ArgF-lac \Delta gltA \Delta icd::aad+$ $\Delta pck \Delta ppc \Delta aceBAK \Delta glcDEFGB \Delta sfcA \Delta maeB$ $\Delta puuE \Delta gabT \Delta sucA$ P> pGEN1143 (pSP100::ido:E65K+) | G3700 |

| Strain | Genotype | Derived from |
|--------|--|--------------|
| G4495 | C> $\Delta ArgF-lac$ $\Delta gltA$ $\Delta icd::aad+$ Δpck Δppc $\Delta aceBAK$ $\Delta glcDEFGB$ $\Delta sfcA$ $\Delta maeB$ $\Delta puuE$ $\Delta gabT$ $\Delta sucA$ P> pGEN1138 (pSP100:: <i>ido</i> +) | G3700 |
| G4497 | C> $\Delta aceAK$ $\Delta sucA$ Δppc $\Delta puuE$ $\Delta gabT$ $\Delta kdgK::kdo2^*$ + <i>kan</i> + | G4343 |
| G4499 | C> $\Delta aceAK$ $\Delta sucA$ Δppc $\Delta puuE$ $\Delta gabT$ $\Delta kdgK::ido^*$ + <i>kan</i> + | G4343 |
| G4569 | Isolate from culture IDO 9 ($\Delta kdgK::ido:A62P+$) | G4383 |
| G4571 | Isolate from culture IDO 9 ($\Delta kdgK::ido:E65K+$) | G4383 |
| G4584 | C> $\Delta ArgF-lac$ $\Delta gltA$ $\Delta icd::aad+$ Δpck Δppc $\Delta aceBAK$ $\Delta glcDEFGB$ $\Delta sfcA$ $\Delta maeB$ $\Delta puuE$ $\Delta gabT$ $\Delta sucA$ P> pGEN1154 (pSP100:: <i>ido:A62P+</i>) | G3700 |

Appendix 2: List of all plasmids cited in this study

| Plasmid | Cloning | Insert | Backbone | Selection cassette |
|----------|-------------|--|----------|-------------------------|
| pCP20 | N/A | N/A | | <i>bla+</i> <i>cat+</i> |
| pET22 | N/A | N/A | | <i>bla+</i> |
| pSP100 | N/A | N/A | pUC18 | <i>bla+</i> |
| pVDM18 | N/A | N/A | pSU18 | <i>cat+</i> |
| pEVL404 | N/A | N/A | pUC18 | <i>thyA+</i> |
| pGEN118 | PacI/NotI | <i>metA</i> (<i>Escherichia coli</i>) | pSP100 | <i>bla+</i> |
| pGEN966 | PacI/NotI | <i>odo</i> (<i>Catenulisporea acidiphila</i>) | pSP100 | <i>bla+</i> |
| pGEN967 | PacI/NotI | <i>odo</i> (<i>Catenulisporea acidiphila</i>) | pVDM18 | <i>cat+</i> |
| pGEN968 | PacI/NotI | <i>kdo1</i> (<i>Catenulisporea acidiphila</i>) | pSP100 | <i>bla+</i> |
| pGEN969 | PacI/NotI | <i>kdo1</i> (<i>Catenulisporea acidiphila</i>) | pVDM18 | <i>cat+</i> |
| pGEN971 | PacI/NotI | <i>kdo2</i> (<i>Chitinophaga pinensis</i>) | pVDM18 | <i>cat+</i> |
| pGEN972 | PacI/NotI | <i>kdo3</i> (<i>Flavobacterium johnsoniae</i>) | pSP100 | <i>bla+</i> |
| pGEN973 | PacI/NotI | <i>kdo3</i> (<i>Flavobacterium johnsoniae</i>) | pVDM18 | <i>cat+</i> |
| pGEN992 | PacI/NotI | <i>metA:K46R</i> | pSP100 | <i>bla+</i> |
| pGEN993 | PacI/NotI | <i>metA:K47R</i> | pSP100 | <i>bla+</i> |
| pGEN997 | PacI/NotI | <i>ido</i> (<i>Bacillus cereus</i>) | pSP100 | <i>bla+</i> |
| pGEN1001 | PacI/NotI | <i>metA:C142Y</i> | pSP100 | <i>bla+</i> |
| pGEN1002 | PacI/NotI | <i>metA:C142R</i> | pSP100 | <i>bla+</i> |
| pGEN1005 | PacI/NotI | <i>metA:H235S</i> | pSP100 | <i>bla+</i> |
| pGEN1007 | PacI/NotI | <i>metA:E237A</i> | pSP100 | <i>bla+</i> |
| pGEN1041 | PacI/NotI | <i>metA:K46R</i> | pVDM18 | <i>cat+</i> |
| pGEN1042 | PacI/NotI | <i>metA:K47R</i> | pVDM18 | <i>cat+</i> |
| pGEN1044 | PacI/NotI | <i>metA:C142Y</i> | pVDM18 | <i>cat+</i> |
| pGEN1045 | PacI/NotI | <i>metA:C142R</i> | pVDM18 | <i>cat+</i> |
| pGEN1048 | PacI/NotI | <i>metA:H235S</i> | pVDM18 | <i>cat+</i> |
| pGEN1050 | PacI/NotI | <i>metA:E237A</i> | pVDM18 | <i>cat+</i> |
| pGEN1066 | PacI/NotI | <i>odo</i> (<i>Catenulisporea acidiphila</i>) | pEVL404 | <i>thyA+</i> |
| pGEN1070 | PacI/NotI | <i>kdo1</i> (<i>Catenulisporea acidiphila</i>) | pEVL404 | <i>thyA+</i> |
| pGEN1074 | PacI/NotI | <i>kdo2</i> (<i>Chitinophaga pinensis</i>) | pSP100 | <i>bla+</i> |
| pGEN1076 | PacI/NotI | <i>kdo4</i> (<i>Niastella koreensis</i>) | pVDM18 | <i>cat+</i> |
| pGEN1080 | EcoRII/XbaI | <i>kdo2</i> (<i>Chitinophaga pinensis</i>) | pEVL404 | <i>thyA+</i> |
| pGEN1084 | PacI/NotI | <i>metA</i> (<i>Escherichia coli</i>) | pVDM18 | <i>cat+</i> |
| pGEN1085 | PacI/NotI | <i>metA:E237D</i> | pVDM18 | <i>cat+</i> |
| pGEN1086 | PacI/NotI | <i>metA:E237D</i> | pSP100 | <i>bla+</i> |
| pGEN1101 | PacI/NotI | <i>kdo4</i> (<i>Niastella koreensis</i>) | pSP100 | <i>bla+</i> |
| pGEN1102 | PacI/NotI | <i>kdo5</i> (<i>Flavobacterium sp</i>) | pSP100 | <i>bla+</i> |
| pGEN1126 | PacI/NotI | <i>kdo2</i> (<i>Chitinophaga pinensis</i>) | pEVL404* | |

| Plasmid | Cloning | Insert | Backbone | Selection cassette |
|----------------|----------------|---------------------------------------|-----------------|---------------------------|
| pGEN1138 | PacI/NotI | <i>ido</i> (<i>Bacillus cereus</i>) | pSP100 | <i>bla</i> + |
| pGEN1143 | PacI/NotI | <i>ido:E65K</i> | pSP100 | <i>bla</i> + |
| pGEN1154 | PacI/NotI | <i>ido:A62P</i> | pSP100 | <i>bla</i> + |

Appendix 3: List of all primers cited in this study

| Primer | Gene | Use | Sequence (5' to 3') |
|---------|-------------|---------|--|
| M13 Fwd | | Cloning | AGGGTTTCCCAGTCACGACGTT |
| M13 Rev | | Cloning | GAGCGATAACAATTACACAGG |
| | <i>odo</i> | His-tag | <u>AAAGAAGGAGATAGGATCATG</u> CATCATCACCATCACCAT CACCGCTTGGCC CTG |
| | <i>odo</i> | His-tag | <u>GTGTAATGGATAAGTGATCTTAGTAAATGACCCGGTC</u> |
| | <i>kdo1</i> | His-tag | <u>AAAGAAGGAGATAGGATCATG</u> CATCATCACCATCACCAT AAGAACCTGTCT GCGTATG |
| | <i>kdo1</i> | His-tag | <u>GTGTAATGGATAAGTGATCTTAGCTGAACCTCGCAGAGAC</u> |
| | <i>kdo2</i> | His-tag | <u>AAAGAAGGAGATAGGATCATG</u> CATCATCACCATCACCAT AGACCCTTAGAC GTGACACC |
| | <i>kdo2</i> | His-tag | <u>GTGTAATGGATAAGTGATCTAAAGGTTGCCAGGTGAG</u> |
| | <i>kdo3</i> | His-tag | <u>AAAGAAGGAGATAGGATCATG</u> CATCATCACCATCACCAT AAATCACAATCA TTAATT |
| | <i>kdo3</i> | His-tag | <u>GTGTAATGGATAAGTGATCTTAAGCCTGATCAAAAACTTTCC</u> |
| | <i>kdo4</i> | His-tag | <u>AAAGAAGGAGATAGGATCATG</u> CATCATCACCATCACCAT GAAACCATCATC GAATC |
| | <i>kdo4</i> | His-tag | <u>GTGTAATGGATAAGTGATCTTACTGCTGAGAGTGGAACAG</u> |
| | <i>kdo5</i> | His-tag | <u>AAAGAAGGAGATAGGATCATG</u> CATCATCACCATCACCAT AAATCTCAGTCT ATCATGTCTGTTG |
| | <i>kdo5</i> | His-tag | <u>GTGTAATGGATAAGTGATCTTAGTCCAGGTCGAAGATTTCAC</u> |
| 2203 | <i>metA</i> | Cloning | CCCTTAATTAA <u>TGCCGATTCTGTGTGCCGG</u> |
| 2204 | <i>metA</i> | Cloning | TAT <u>GCGGCCGC</u> TTAATCCAGCGTTGGATTTC |
| 4823 | <i>kdo1</i> | Cloning | CCCTTAATTAA <u>TGAAGAACCTGTCTCGTATGAAG</u> |
| 4824 | <i>kdo1</i> | Cloning | TAT <u>GCGGCCGC</u> TTAGCTGAACCTCGCAGAGACG |
| 4825 | <i>kdo2</i> | Cloning | CCCTTAATTAA <u>TGAGACCCTTAGACGTGACAC</u> |
| 4826 | <i>kdo2</i> | Cloning | TAT <u>GCGGCCGC</u> TTAAAGGTTGCCAGGTGAG |
| 4827 | <i>kdo3</i> | Cloning | CCCTTAATTAA <u>TGAAATCACAATCATTAAATTG</u> |
| 4828 | <i>kdo3</i> | Cloning | TAT <u>GCGGCCGC</u> TTAACGCCTGATCAAAACTT |
| 4829 | <i>kdo4</i> | Cloning | CCCTTAATTAA <u>TGGAAACCATCATCGAATCTCG</u> |
| 4830 | <i>kdo4</i> | Cloning | TAT <u>GCGGCCGC</u> TTACTGCTGAGAGTGGAACAGTTTAC |
| 4831 | <i>kdo5</i> | Cloning | CCCTTAATTAA <u>TGAAATCTCAGTCTATCATGTCTG</u> |
| 4832 | <i>kdo5</i> | Cloning | TAT <u>GCGGCCGC</u> TTAGTCCAGGTCGAAGATTTCAC |
| 4833 | <i>odo</i> | Cloning | CCCTTAATTAA <u>TGCACCGCTGGCCCTGAC</u> |
| 4834 | <i>odo</i> | Cloning | TAT <u>GCGGCCGC</u> TTAGTAAATGACCCGGTCGTCCG |
| 4890 | <i>kdo1</i> | Cloning | TAT <u>GCATGC</u> TTAGCTGAACCTCGCAGAG |
| 4898 | <i>ido</i> | Cloning | TAT <u>GCGGCCGC</u> TTATTTGTCTCCTTATAAGAAAATGTTAC |
| 4900 | <i>ido</i> | Cloning | CCCTTAATTAA <u>TGGAGGTTTTATAATGACGTTGTT</u> |
| 4924 | <i>metA</i> | Dir mut | CCTGATGCC <u>AG</u> GAAGATTGAAACTGAAATCAGTTCTGCGCCTG |
| 4925 | <i>metA</i> | Dir mut | CAGTTCAATCTT <u>CC</u> CGGCATCAGGTTAAGGATCAGAACCTTAAGTGG |
| 4926 | <i>metA</i> | Dir mut | CCTGATGCC <u>AG</u> GAAG <u>GG</u> ATTGAAACTGAAATCAGTTCTGCGCCTG |
| 4927 | <i>metA</i> | Dir mut | CAGTTCAATCCT <u>TT</u> CGGCATCAGGTTAAGGATCAGAACCTTAAGTGG |
| 4928 | <i>metA</i> | Dir mut | TAACCTGATGCC <u>AG</u> <u>GG</u> AGGATTGAAACTGAAATCAGTTCTGCGCCTG |
| 4929 | <i>metA</i> | Dir mut | TTTCAGTTCAAT <u>CC</u> <u>CT</u> CGGCATCAGGTTAAGGATCAGAACCTTAAG |
| 4951 | <i>metA</i> | Dir mut | GTTTGTC <u>CG</u> <u>CT</u> GGCGGTACAGGCCGCGCTCAATATCC |
| 4952 | <i>metA</i> | Dir mut | GTACCGCC <u>AG</u> <u>CC</u> GACAAACAGCGTCGAGGTGACGTG |
| 4953 | <i>metA</i> | Dir mut | GTTTGTC <u>ACT</u> GGCGGTACAGGCCGCGCTCAATATCC |
| 4954 | <i>metA</i> | Dir mut | GTACCGCC <u>AG</u> <u>TA</u> GACAAACAGCGTCGAGGTGACGTG |
| 4955 | <i>metA</i> | Dir mut | GACGGG <u>T</u> <u>AT</u> CCCGAATATGATGCGCAAACGCTGGCGCAGGAATTTC |
| 4956 | <i>metA</i> | Dir mut | CATATTGG <u>GA</u> <u>TA</u> GGCCGTACAAAGGCAATGCGCTTATTTACTGGC |

| Primer | Gene | Use | Sequence (5' to 3') |
|--------|-------------|---------|---|
| 4957 | <i>metA</i> | Dir mut | GACGGGCC <u>GT</u> CCCGAATATGATGCGCAAACGCTGGCGCAGGAATTTC |
| 4958 | <i>metA</i> | Dir mut | CATATTGGG <u>AC</u> GGCCCGTCACAAAGGCAATGCCCTATCTTACTGGC |
| 4959 | <i>metA</i> | Dir mut | GACGGGC <u>AGT</u> CCCGAATATGATGCGCAAACGCTGGCGCAGGAATTTC |
| 4960 | <i>metA</i> | Dir mut | CATATTGGG <u>AT</u> GCCCGTCACAAAGGCAATGCCCTATCTTACTGGC |
| 4961 | <i>metA</i> | Dir mut | CCATCCC <u>GAT</u> TATGATGCGCAAACGCTGGCGCAGGAATTTC |
| 4952 | <i>metA</i> | Dir mut | GCGCATCATA <u>ATC</u> GGGATGCCCGTCACAAAGGCAATGCCTTATC |
| 4963 | <i>metA</i> | Dir mut | CCATCCC <u>GAT</u> ATGATGCGCAAACGCTGGCGCAGGAATTTC |
| 4964 | <i>metA</i> | Dir mut | GCGCATCATA <u>G</u> GGGATGCCCGTCACAAAGGCAATGCCTTATC |
| 5059 | <i>kdo2</i> | Cloning | TAT <u>TCTAGA</u> TTAAAGGTTGCCAGGTGAGC |
| 5060 | <i>kdo3</i> | Cloning | TAT <u>TCTAGA</u> TTAACGCTGATCAAAACTTTCC |
| 5061 | <i>kdo4</i> | Cloning | TAT <u>TCTAGA</u> TTACTGCTGAGAGTGGAACAGTTAC |
| 5089 | <i>ido</i> | His-tag | <u>AAAGAAGGAGATAGGATCATG</u> CATCATCACCATCACCAT GAGGTTTTATA ATGACGTTGTTCTTAG |
| 5091 | <i>ido</i> | His-tag | <u>GTGTAATGGATAAGT</u> GATCTTATTTGTCCTTATAAGAAAATGTTAC |
| 5140 | <i>ido</i> | His-tag | <u>AAAGAAGGAGATAGGAT</u> CATGGAGGTTTTATAATGACGTTGTTCT |
| 5092 | <i>ido</i> | His-tag | <u>GTGTAATGGATAAGT</u> GATCTTA ATGGTATGGTATGATG TTTGTCCTT ATAAGAAAATGTTACTAATAAAATATC |
| 5355 | <i>kdo2</i> | qPCR | GCAGGAAACCGACCAGTAA |
| 5356 | <i>kdo2</i> | qPCR | TCGCTGCGGAAATACTC |
| 5357 | 16s | qPCR | CCCTTACGACCAGGGCTACA |
| 5358 | 16s | qPCR | ATCCGGACTACGACGCAC |
| 5376 | <i>ido</i> | qPCR | ACTATGACGATGGGGAAA |
| 5377 | <i>ido</i> | qPCR | AATTCAAGTATCGAACGACGA |
| 5425 | <i>thyA</i> | qPCR | GCGCCGCATTATTGTTTC |
| 5426 | <i>thyA</i> | qPCR | TTTGGCGTCTGCCACATAG |

Restriction sites PstI (**TTAATTAA**), NotI (**GC**GGCCGC), SphI (**GCATGC**), XbaI (**TCTAGA**), pET22 analogous sequence are underlined and His tags are in **bold blue**. Primers for directed mutagenesis (Dir mut) are underlined at the mutated codon

Appendix 4: List of all GM3 cultures cited in this study

| GM3 culture | Strain name | Regime | Generation time, Temperature | Relaxing medium | Stressing medium |
|-------------|-------------|----------------|------------------------------|---|--|
| G3513 | G3513 | Medium SWAP | 2 h, 37°C | MS glc (0.2%) glu (2 mM) suc (100 μM) | MS glc (0.2%) glu (2 mM) |
| IDO 3 | G4110 | Medium SWAP | 2 h, 37°C | MS glc (0.2%) glu (5 mM) val (0.3 mM) leu (0.3 mM) met (0.1 mM) DAP (0.3 mM) ile (3 mM) suc (0.8 mM) | MS glc (0.2%) glu (5 mM) val (0.3 mM) leu (0.3 mM) met (0.1 mM) DAP (0.3 mM) ile (3 mM) |
| IDO 6 | G4110 | Medium SWAP | 2 h, 37°C | MS glc (0.2%) glu (5 mM) val (0.3 mM) leu (0.3 mM) met (0.1 mM) DAP (0.3 mM) ile (3 mM) suc (0.8 mM) | MS glc (0.2%) glu (5 mM) val (0.3 mM) leu (0.3 mM) met (0.1 mM) DAP (0.3 mM) ile (3 mM) |
| IDO 9 | G4383 | Medium SWAP | 2 h, 37°C | MA glc (0.2%) glu (5 mM) norvaline (1.5 mM) suc (0.8 mM) | MA glc (0.2%) glu (5 mM) norvaline (1.5 mM) |
| IDO 11 | G4383 | Medium SWAP | 2 h, 37°C | MA glc (0.2%) glu (5 mM) valeric acid (1.5 mM) suc (0.8 mM) | MA glc (0.2%) glu (5 mM) valeric acid (1.5 mM) |
| IDO 13 | G4383 | Medium SWAP | 2 h, 37°C | MA glc (0.2%) glu (5 mM) norleucine (1.5 mM) suc (0.8 mM) | MA glc (0.2%) glu (5 mM) norleucine (1.5 mM) |
| IDO 17 | G4499 | Medium SWAP | 2 h, 37°C | MA glu (10 mM) ile (10 mM) suc (8 mM) | MA glu (10 mM) ile (10 mM) |
| IDO 18 | G4499 | Medium SWAP | 2 h, 37°C | MA glu (10 mM) ile (10 mM) suc (8 mM) | MA glu (10 mM) ile (10 mM) |
| KDA 10 | G3951 | Medium SWAP | 2 h, 37°C | MS glc (0.2%) glu (5 mM) lys (5 mM) suc (0.8 mM) | MS glc (0.2%) glu (5 mM) lys (5 mM) |
| KDB 10 | G3986 | Medium SWAP | 2 h, 37°C | MS glc (0.2%) glu (5 mM) lys (5 mM) suc (0.8 mM) | MS glc (0.2%) glu (5 mM) lys (5 mM) |
| KDB 15 | G4160 | Medium SWAP | 2 h, 37°C | MS glc (0.2%) glu (5 mM) lys (5 mM) suc (0.8 mM) | MS glc (0.2%) glu (5 mM) lys (5 mM) |
| KDB 16 | G4160 | Medium SWAP | 2 h, 37°C | MS glc (0.2%) glu (5 mM) lys (5 mM) suc (0.8 mM) | MS glc (0.2%) glu (5 mM) lys (5 mM) |

| GM3 | Strain | Regime | Generation time, | Relaxing medium | Stressing medium |
|---------|--------|----------------|------------------|---|--|
| culture | | | Temperature | | |
| | | | | | |
| KDB 21 | G4275 | Medium SWAP | 2 h, 37°C | MS glc (0.2%) glu (5 mM) 6-aminocaproic acid (1.5 mM) suc (0.8 mM) | MS glc (0.2%) glu (5 mM) 6-aminocaproic acid (1.5 mM) |
| KDB 22 | G4276 | Medium SWAP | 2 h, 37°C | MS glc (0.2%) glu (5 mM) thialysine (1.5 mM) suc (0.8 mM) | MS glc (0.2%) glu (5 mM) thialysine (1.5 mM) |
| KDB 26 | G4497 | Medium SWAP | 2 h, 37°C | MA glu (10 mM) lys (10 mM) suc (8 mM) | MA glu (10 mM) lys (10 mM) |
| KDB 27 | G4497 | Medium SWAP | 2 h, 37°C | MA glu (10 mM) lys (10 mM) suc (8 mM) | MA glu (10 mM) lys (10 mM) |
| MET 6 | G3880 | Turbidostat | 30°C | MS glc (0.2%) | |
| MET 7 | G4010 | Turbidostat | 30°C | MS glc (0.2%) | |
| MET 10 | G4011 | Turbidostat | 30°C | MS glc (0.2%) | |
| MET 15 | G3880 | Medium SWAP | 2 h, 37°C | MS glc (0.2%) met (0.05mM) | MS glc (0.2%) |
| MET 17 | G4010 | Medium SWAP | 2 h, 37°C | MS glc (0.2%) met (0.05mM) | MS glc (0.2%) |
| MET 18 | G4011 | Medium SWAP | 2 h, 37°C | MS glc (0.2%) met (0.05mM) | MS glc (0.2%) |
| ODO 8 | G3935 | Medium SWAP | 2 h, 37°C | MS glc (0.2%) glu (5 mM) orn (5 mM) suc (0.8 mM) | MS glc (0.2%) glu (5 mM) orn (5 mM) |

Appendix 5: NTA mix composition (1000X)

| Reagent | Chemical formula | Final concentration (mM) |
|------------------------|---|--------------------------|
| Nitrilotriacetic acid | C ₆ H ₉ NO ₆ | 10 |
| Calcium chloride | CaCl ₂ , 2 H ₂ O | 3 |
| Iron(III) chloride | FeCl ₃ , 6 H ₂ O | 3 |
| Manganese(II) chloride | MnCl ₂ , 4 H ₂ O | 1 |
| Zinc chloride | ZnCl ₂ | 1 |
| Boric acid | BH ₃ O ₃ | 0.1 |
| Chromium(III) chloride | CrCl ₃ , 6 H ₂ O | 0.1 |
| Cobalt(II) chloride | CoCl ₂ , 6 H ₂ O | 0.1 |
| Copper(II) chloride | CuCl ₂ , 2 H ₂ O | 0.1 |
| Nickel(II) chloride | NiCl ₂ , 6 H ₂ O | 0.1 |
| Sodium molybdate | Na ₂ MoO ₄ , 2 H ₂ O | 0.1 |
| Sodium selenite | Na ₂ O ₃ Se, 5 H ₂ O | 0.1 |

Appendix 6: MS medium composition

| Reagent | Chemical formula | Final concentration (mM) |
|-----------------------|---|--------------------------|
| Citric acid | C ₆ H ₈ O ₇ , H ₂ O | 4 |
| Magnesium sulfate | MgSO ₄ , 7 H ₂ O | 1 |
| Ammonium chloride | NH ₄ Cl | 20 |
| Dipotassium phosphate | K ₂ HPO ₄ , 3 H ₂ O | 50 |
| NTA Mix | See Appendix 5 | 1 X |

Appendix 7: MA medium composition

| Reagent | Chemical formula | Final concentration (mM) |
|-------------------------|--|--------------------------|
| Nitrilotriacetic acid | C ₆ H ₉ NO ₆ | 0.2 |
| Magnesium sulfate | MgSO ₄ , 7 H ₂ O | 1 |
| Disodium phosphate | Na ₂ HPO ₄ , 12 H ₂ O | 400 |
| Ammonium chloride | NH ₄ Cl | 95 |
| Monopotassium phosphate | KH ₂ PO ₄ | 125 |
| NTA Mix | See Appendix 5 | 1 X |

BIBLIOGRAPHY

- Agresti, J.J., Antipov, E., Abate, A.R., Ahn, K., Rowat, A.C., Baret, J.-C., Marquez, M., Klibanov, A.M., Griffiths, A.D., and Weitz, D.A. (2010). Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proc. Natl. Acad. Sci.* *107*, 4004–4009.
- Aharoni, A., Gaidukov, L., Khersonsky, O., Gould, S.M., Roodveldt, C., and Tawfik, D.S. (2005). The “evolvability” of promiscuous protein functions. *Nat. Genet.* *37*, 73–76.
- Arnold, F.H. (1998). When blind is better: Protein design by evolution. *Nat. Biotechnol.* *16*, 617–618.
- Arnold, F.H. (2015). The nature of chemical innovation: new enzymes by evolution. *Q. Rev. Biophys.* *48*, 404–410.
- Arren Bar-Even, Elad Noor, Yonatan Savir, Wolfram Liebermeister, Dan Davidi, Dan S. Tawfik, and Ron Milo (2011). The moderately efficient enzyme: Evolutionary and Physicochemical trends shaping enzyme parameters. *Biochemistry (Mosc.)* *50*, 4402–4410.
- Aslanidis, C., and Jong, P.J. de (1990). Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res.* *18*, 6069–6074.
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* *2*, 2006.0008.
- Badran, A.H., and Liu, D.R. (2015). In vivo continuous directed evolution. *Curr. Opin. Chem. Biol.* *24*, 1–10.
- Barrick, J.E., Yu, D.S., Yoon, S.H., Jeong, H., Oh, T.K., Schneider, D., Lenski, R.E., and Kim, J.F. (2009). Genome evolution and adaptation in a long-term experiment with Escherichia coli. *Nature* *461*, 1243–1247.
- Barrozo, A., Borstnar, R., Marloie, G., and Kamerlin, S.C.L. (2012). Computational Protein Engineering: Bridging the Gap between Rational Design and Laboratory Evolution. *Int. J. Mol. Sci.* *13*, 12428–12460.
- Baud, D., Saaidi, P.-L., Monfleur, A., Harari, M., Cuccaro, J., Fossey, A., Besnard, M., Debard, A., Mariage, A., Pellouin, V., Petit, J.-L., Salanoubat, M., Weissenbach, J., de Berardinis, V., and Zaparucha, A. (2014). Synthesis of Mono- and Dihydroxylated Amino Acids with New α -Ketoglutarate-Dependent Dioxygenases: Biocatalytic Oxidation of C-H Bonds. *ChemCatChem* *6*, 3012–3017.
- Bentley, W.E., Mirjalili, N., Andersen, D.C., Davis, R.H., and Kompala, D.S. (1990). Plasmid-encoded protein: the principal factor in the “metabolic burden” associated with recombinant bacteria. *Biotechnol. Bioeng.* *35*, 668–681.
- Bershtein, S., Goldin, K., and Tawfik, D.S. (2008). Intense Neutral Drifts Yield Robust and Evolvable Consensus Proteins. *J. Mol. Biol.* *379*, 1029–1044.

Bhattacharya, S.K., and Dubey, A.K. (1995). Metabolic burden as reflected by maintenance coefficient of recombinant *Escherichia coli* overexpressing target gene. *Biotechnol. Lett.* *17*, 1155–1160.

Birnbaum, S., and Bailey, J.E. (1991). Plasmid presence changes the relative levels of many host cell proteins and ribosome components in recombinant *Escherichia coli*. *Biotechnol. Bioeng.* *37*, 736–745.

Bloom, J.D., Labthavikul, S.T., Otey, C.R., and Arnold, F.H. (2006). Protein stability promotes evolvability. *Proc. Natl. Acad. Sci.* *103*, 5869–5874.

Bloom, J.D., Romero, P.A., Lu, Z., and Arnold, F.H. (2007). Neutral genetic drift can alter promiscuous protein functions, potentially aiding functional evolution. *Biol. Direct* *2*, 17.

Blount, Z.D., Barrick, J.E., Davidson, C.J., and Lenski, R.E. (2012). Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* *489*, 513–518.

Boutard, M., Cerisy, T., Nogue, P.-Y., Alberti, A., Weissenbach, J., Salanoubat, M., and Tolonen, A.C. (2014). Functional Diversity of Carbohydrate-Active Enzymes Enabling a Bacterium to Ferment Plant Biomass. *PLOS Genet.* *10*, e1004773.

Bouzon, M., Perret, A., Loreau, O., Delmas, V., Perchat, N., Weissenbach, J., Taran, F., and Marlière, P. (2017). A synthetic alternative to one-carbon metabolism. *ACS Synth. Biol.*

Breen, M.S., Kemeny, C., Vlasov, P.K., Notredame, C., and Kondrashov, F.A. (2012). Epistasis as the primary factor in molecular evolution. *Nature* *490*, 535–538.

Bridges, B.A., and Woodgate, R. (1985). Mutagenic repair in *Escherichia coli*: products of the recA gene and of the umuD and umuC genes act at different steps in UV-induced mutagenesis. *Proc. Natl. Acad. Sci. U. S. A.* *82*, 4193–4197.

Camps, M., Herman, A., Loh, E., and Loeb, L.A. (2007). Genetic constraints on protein evolution. *Crit. Rev. Biochem. Mol. Biol.* *42*, 313–326.

Carter, P., Nilsson, B., Burnier, J.P., Burdick, D., and Wells, J.A. (1989). Engineering subtilisin BPN' for site-specific proteolysis. *Proteins Struct. Funct. Bioinforma.* *6*, 240–248.

Cedrone, F., Ménez, A., and Quéméneur, E. (2000). Tailoring new enzyme functions by rational redesign. *Curr. Opin. Struct. Biol.* *10*, 405–410.

Chefson, A., Zhao, J., and Auclair, K. (2006). Replacement of Natural Cofactors by Selected Hydrogen Peroxide Donors or Organic Peroxides Results in Improved Activity for CYP3A4 and CYP2D6. *ChemBioChem* *7*, 916–919.

Cheng, A.-X., Han, X.-J., Wu, Y.-F., and Lou, H.-X. (2014). The Function and Catalysis of 2-Oxoglutarate-Dependent Oxygenases Involved in Plant Flavonoid Biosynthesis. *Int. J. Mol. Sci.* *15*, 1080–1095.

Cherry, J.R., and Fidantsef, A.L. (2003). Directed evolution of industrial enzymes: an update. *Curr. Opin. Biotechnol.* *14*, 438–443.

Chica, R.A., Doucet, N., and Pelletier, J.N. (2005). Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Curr. Opin. Biotechnol.* *16*, 378–384.

Chua, C.S., Biermann, D., Goo, K.S., and Sim, T.-S. (2008). Elucidation of active site residues of *Arabidopsis thaliana* flavonol synthase provides a molecular platform for engineering flavonols. *Phytochemistry* *69*, 66–75.

Clifton, I.J., McDonough, M.A., Ehrismann, D., Kershaw, N.J., Granatino, N., and Schofield, C.J. (2006). Structural studies on 2-oxoglutarate oxygenases and related double-stranded beta-helix fold proteins. *J. Inorg. Biochem.* *100*, 644–669.

Coco, W.M., Levinson, W.E., Crist, M.J., Hektor, H.J., Darzins, A., Pienkos, P.T., Squires, C.H., and Monticello, D.J. (2001). DNA shuffling method for generating highly recombined genes and evolved enzymes. *Nat. Biotechnol.* *19*, 354–359.

Cooper, G.M. (2000). The Central Role of Enzymes as Biological Catalysts.

Copley, S.D. (2003). Enzymes with extra talents: moonlighting functions and catalytic promiscuity. *Curr. Opin. Chem. Biol.* *7*, 265–272.

de Crécy-Lagard, V.A., Bellalou, J., Mutzel, R., and Marlière, P. (2001). Long term adaptation of a microbial population to a permanent metabolic constraint: overcoming thymineless death by experimental evolution of *Escherichia coli*. *BMC Biotechnol.* *1*, 10.

D. R. Mills, R. I. Peterson, and S. Spiegelman (1967). An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule.pdf. *PNAS* *58*.

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci.* *97*, 6640–6645.

Dougherty, M.J., and Arnold, F.H. (2009). Directed evolution: new parts and optimized function. *Curr. Opin. Biotechnol.* *20*, 486–491.

Eichhorn, E., van der Ploeg, J.R., Kertesz, M.A., and Leisinger, T. (1997). Characterization of alpha-ketoglutarate-dependent taurine dioxygenase from *Escherichia coli*. *J. Biol. Chem.* *272*, 23031–23036.

Elena, S.F., and Lenski, R.E. (2003). Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat. Rev. Genet.* *4*, 457–469.

Elkins, J.M., Ryle, M.J., Clifton, I.J., Dunning Hotopp, J.C., Lloyd, J.S., Burzlaff, N.I., Baldwin, J.E., Hausinger, R.P., and Roach, P.L. (2002). X-ray crystal structure of *Escherichia coli* taurine/alpha-ketoglutarate dioxygenase complexed to ferrous iron and substrates. *Biochemistry (Mosc.)* *41*, 5185–5192.

Fasan, R. (2012). Tuning P450 Enzymes as Oxidation Catalysts. *ACS Catal.* *2*, 647–666.

Firestein, S. (2001). How the olfactory system makes sense of scents. *Nature* *413*, 211–218.

- Fujii, R., Kitaoka, M., and Hayashi, K. (2004). One-step random mutagenesis by error-prone rolling circle amplification. *Nucleic Acids Res.* *32*, e145.
- Gerlt, J.A., and Babbitt, P.C. (1998). Mechanistically diverse enzyme superfamilies: the importance of chemistry in the evolution of catalysis. *Curr. Opin. Chem. Biol.* *2*, 607–612.
- Gerrish, P.J., and Lenski, R.E. (1998). The fate of competing beneficial mutations in an asexual population. *Genetica* *102–103*, 127.
- Girolamo, M.D., Busiello, V., Cini, C., Foppoli, C., and Marco, C.D. (1982). Thialysine utilization by *E. coli* and its effects on cell growth. *Mol. Cell. Biochem.* *46–107*, 43–48.
- Glick, B.R. (1995). Metabolic load and heterologous gene expression. *Biotechnol. Adv.* *13*, 247–261.
- Goldsmith, M., and Tawfik, D.S. (2012). Directed enzyme evolution: beyond the low-hanging fruit. *Curr. Opin. Struct. Biol.* *22*, 406–412.
- Gorres, K.L., and Raines, R.T. (2010). Prolyl 4-hydroxylase. *Crit. Rev. Biochem. Mol. Biol.* *45*, 106–124.
- Grand View Research (2016). Enzymes Market By Type And Segment Forecasts To 2024.
- Greener, A., Callahan, M., and Jerpseth, B. (1997). An efficient random mutagenesis technique using an *E. coli* mutator strain. *Mol. Biotechnol.* *7*, 189–195.
- Gupta, R.D., and Tawfik, D.S. (2008). Directed enzyme evolution via small and effective neutral drift libraries. *Nat. Methods* *5*, 939–942.
- Gutte, B. (1975). A synthetic 70-amino acid residue analog of ribonuclease S-protein with enzymic activity. *J. Biol. Chem.* *250*, 889–904.
- Hall, B.G. (1982). Chromosomal mutation for citrate utilization by *Escherichia coli* K-12. *J. Bacteriol.* *151*, 269–273.
- Hamed, R.B., Gomez-Castellanos, J.R., Henry, L., Ducho, C., McDonough, M.A., and Schofield, C.J. (2013). The enzymes of β-lactam biosynthesis. *Nat. Prod. Rep.* *30*, 21–107.
- Hara, R., and Kino, K. (2009). Characterization of novel 2-oxoglutarate dependent dioxygenases converting l-proline to cis-4-hydroxy-l-proline. *Biochem. Biophys. Res. Commun.* *379*, 882–886.
- Hausinger, R. (2015). Biochemical Diversity of 2-Oxoglutarate-Dependent Oxygenases. 1–58.
- Helmetag, V., Samel, S.A., Thomas, M.G., Marahiel, M.A., and Essen, L.-O. (2009). Structural basis for the erythro-stereospecificity of the L-arginine oxygenase VioC in viomycin biosynthesis. *FEBS J.* *276*, 3669–3682.
- Hibi, M., Kawashima, T., Kodera, T., Smirnov, S.V., Sokolov, P.M., Sugiyama, M., Shimizu, S., Yokozeki, K., and Ogawa, J. (2011). Characterization of *Bacillus thuringiensis* l-

Isoleucine Dioxygenase for Production of Useful Amino Acids. *Appl. Environ. Microbiol.* **77**, 6926–6930.

Hinga, K.R. (2002). Effects of pH on coastal marine phytoplankton. *Mar. Ecol. Prog. Ser.* **238**, 281–300.

Hofwegen, D.J.V., Hovde, C.J., and Minnich, S.A. (2016). Rapid Evolution of Citrate Utilization by *Escherichia coli* by Direct Selection Requires *citT* and *dctA*. *J. Bacteriol.* **198**, 1022–1034.

Huang, P.-S., Boyken, S.E., and Baker, D. (2016). The coming of age of de novo protein design. *Nature* **537**, 320–327.

Jäckel, C., Kast, P., and Hilvert, D. (2008). Protein Design by Directed Evolution. *Annu. Rev. Biophys.* **37**, 153–173.

Jenal, U., Reinders, A., and Lori, C. (2017). Cyclic di-GMP: second messenger extraordinaire. *Nat. Rev. Microbiol.* **15**, 271–284.

Jiang, L., Althoff, E.A., Clemente, F.R., Doyle, L., Röthlisberger, D., Zanghellini, A., Gallaher, J.L., Betker, J.L., Tanaka, F., Barbas, C.F., Hilvert, D., Houk, K.N., Stoddard, B.L., and Baker, D. (2008). De Novo Computational Design of Retro-Aldol Enzymes. *Science* **319**, 1387–1391.

Johannes, T.W., and Zhao, H. (2006). Directed evolution of enzymes and biosynthetic pathways. *Curr. Opin. Microbiol.* **9**, 261–267.

Johansson, C., Tumber, A., Che, K., Cain, P., Nowak, R., Gileadi, C., and Oppermann, U. (2014). The roles of Jumonji-type oxygenases in human disease. *Epigenomics* **6**, 89–120.

Kadonaga, J.T., and Knowles, J.R. (1985). A simple and efficient method for chemical mutagenesis of DNA. *Nucleic Acids Res.* **13**, 1733–1745.

Kamerlin, S.C.L., and Warshel, A. (2011). The empirical valence bond model: theory and applications. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **1**, 30–45.

Kodera, T., Smirnov, S.V., Samsonova, N.N., Kozlov, Y.I., Koyama, R., Hibi, M., Ogawa, J., Yokozeki, K., and Shimizu, S. (2009). A novel l-isoleucine hydroxylating enzyme, l-isoleucine dioxygenase from *Bacillus thuringiensis*, produces (2S,3R,4S)-4-hydroxyisoleucine. *Biochem. Biophys. Res. Commun.* **390**, 506–510.

Koketsu, K., Shomura, Y., Moriwaki, K., Hayashi, M., Mitsuhashi, S., Hara, R., Kino, K., and Higuchi, Y. (2015). Refined Regio- and Stereoselective Hydroxylation of l-Pipecolic Acid by Protein Engineering of l-Proline cis-4-Hydroxylase Based on the X-ray Crystal Structure. *ACS Synth. Biol.* **4**, 383–392.

Kuchner, O., and Arnold, F.H. (1997). Directed evolution of enzyme catalysts. *Trends Biotechnol.* **15**, 523–530.

Kumar, D., Abdulovic, A.L., Viberg, J., Nilsson, A.K., Kunkel, T.A., and Chabes, A. (2011). Mechanisms of mutagenesis in vivo due to imbalanced dNTP pools. *Nucleic Acids Res.* *39*, 1360–1371.

Kumar, V., Sangwan, P., Singh, D., and Gill, P.K. (2014). Global scenario of industrial enzyme market.

Kunz, B.A. (1988). Mutagenesis and deoxyribonucleotide pool imbalance. *Mutat. Res. Mol. Mech. Mutagen.* *200*, 133–147.

Labrou, N.E. (2010). Random mutagenesis methods for in vitro directed enzyme evolution. *Curr. Protein Pept. Sci.* *11*, 91–100.

Lai, Y.-P., Huang, J., Wang, L.-F., Li, J., and Wu, Z.-R. (2004). A new approach to random mutagenesis in vitro. *Biotechnol. Bioeng.* *86*, 622–627.

Lee, H., Popodi, E., Tang, H., and Foster, P.L. (2012). Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc. Natl. Acad. Sci. U. S. A.* *109*, E2774–E2783.

Lee, H.S., Kwon, K.K., Kang, S.G., Cha, S.-S., Kim, S.-J., and Lee, J.-H. (2010). Approaches for novel enzyme discovery from marine environments. *Curr. Opin. Biotechnol.* *21*, 353–357.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. *Methods* *25*, 402–408.

Luo, L., Pappalardi, M.B., Tummino, P.J., Copeland, R.A., Fraser, M.E., Grzyska, P.K., and Hausinger, R.P. (2006). An assay for Fe(II)/2-oxoglutarate-dependent dioxygenases by enzyme-coupled detection of succinate formation. *Anal. Biochem.* *353*, 69–74.

Lütgens, M., and Gottschalk, G. (1980). Why a Co-substrate is Required for Anaerobic Growth of *Escherichia coli* on Citrate. *Microbiology* *119*, 63–70.

Lynd, L.R., Weimer, P.J., van Zyl, W.H., and Pretorius, I.S. (2002). Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol. Mol. Biol. Rev. MMBR* *66*, 506–577, table of contents.

Marinus, M.G. (2010). DNA methylation and mutator genes in *Escherichia coli* K-12. *Mutat. Res. Mutat. Res.* *705*, 71–76.

Marlière, P., Patrouix, J., Döring, V., Herdewijn, P., Tricot, S., Cruveiller, S., Bouzon, M., and Mutzel, R. (2011). Chemical Evolution of a Bacterium's Genome. *Angew. Chem. Int. Ed.* *50*, 7109–7114.

Martinez, S., and Hausinger, R.P. (2015). Catalytic Mechanisms of Fe(II)- and 2-Oxoglutarate-dependent Oxygenases. *J. Biol. Chem.* *290*, 20702–20711.

McCullum, E.O., Williams, B.A.R., Zhang, J., and Chaput, J.C. (2010). Random mutagenesis by error-prone PCR. *Methods Mol. Biol. Clifton NJ* *634*, 103–109.

- Miller, J.H. (1972). Experiments in molecular genetics (Cold Spring Harbor Laboratory).
- Monod, J. (1950). La technique de culture continue, theorie et applications. Ann. Inst. Pasteur 79, 390–410.
- Mori, H., Shibasaki, T., Yano, K., and Ozaki, A. (1997). Purification and cloning of a proline 3-hydroxylase, a novel enzyme which hydroxylates free L-proline to cis-3-hydroxy-L-proline. J. Bacteriol. 179, 5677–5683.
- Morley, K.L., and Kazlauskas, R.J. (2005). Improving enzyme properties: when are closer mutations better? Trends Biotechnol. 23, 231–237.
- Müller, M.M., Allison, J.R., Hongdilokkul, N., Gaillon, L., Kast, P., Gunsteren, W.F. van, Marlière, P., and Hilvert, D. (2013). Directed Evolution of a Model Primordial Enzyme Provides Insights into the Development of the Genetic Code. PLOS Genet. 9, e1003187.
- Munch, P. (1983). Transport of nucleic acid precursors. Metab. Nucleotides 259–305.
- Mutzel, R., and Marliere, P. (2004). Method and device for selecting accelerated proliferation of living cells in suspension.
- Novick, A., and Szilard, L. (1950). Description of the Chemostat. Science 112, 715–716.
- Olsén, A., Jonsson, A., and Normark, S. (1989). Fibronectin binding mediated by a novel class of surface organelles on Escherichia coli. Nature 338, 652–655.
- Packer, M.S., and Liu, D.R. (2015). Methods for the directed evolution of proteins. Nat. Rev. Genet. 16, 379–394.
- Peisajovich, S.G., and Tawfik, D.S. (2007). Protein engineers turned evolutionists. Nat. Methods 4, 991–994.
- Pesavento, C., Becker, G., Sommerfeldt, N., Possling, A., Tschenkow, N., Mehlis, A., and Hengge, R. (2008). Inverse regulatory coordination of motility and curli-mediated adhesion in Escherichia coli. Genes Dev. 22, 2434–2446.
- Petit, E., Coppi, M.V., Hayes, J.C., Tolonen, A.C., Warnick, T., Latouf, W.G., Amisano, D., Biddle, A., Mukherjee, S., Ivanova, N., Lykidis, A., Land, M., Hauser, L., Kyripides, N., Henrissat, B., Lau, J., Schnell, D.J., Church, G.M., Leschine, S.B., and Blanchard, J.L. (2015). Genome and Transcriptome of Clostridium phytofermentans, Catalyst for the Direct Conversion of Plant Feedstocks to Fuels. PLOS ONE 10, e0118285.
- Poelwijk, F.J., Kiviet, D.J., Weinreich, D.M., and Tans, S.J. (2007). Empirical fitness landscapes reveal accessible evolutionary paths. Nature 445, 383–386.
- Popov, M., Petrov, S., Nacheva, G., Ivanov, I., and Reichl, U. (2011). Effects of a recombinant gene expression on ColE1-like plasmid segregation in Escherichia coli. BMC Biotechnol. 11, 18.
- Purpero, V., and Moran, G.R. (2007). The diverse and pervasive chemistries of the α -keto acid dependent enzymes. JBIC J. Biol. Inorg. Chem. 12, 587–601.

Qin, H.-M., Miyakawa, T., Jia, M.Z., Nakamura, A., Ohtsuka, J., Xue, Y.-L., Kawashima, T., Kasahara, T., Hibi, M., Ogawa, J., and Tanokura, M. (2013). Crystal Structure of a Novel N-Substituted L-Amino Acid Dioxygenase from *Burkholderia ambifaria* AMMD. *PLOS ONE* *8*, e63996.

Quandt, E.M., Deatherage, D.E., Ellington, A.D., Georgiou, G., and Barrick, J.E. (2014). Recursive genomewide recombination and sequencing reveals a key refinement step in the evolution of a metabolic innovation in *Escherichia coli*. *Proc. Natl. Acad. Sci.* *111*, 2217–2222.

Reetz, M.T., Zonta, A., Schimossek, K., Jaeger, K.-E., and Liebeton, K. (1997). Creation of Enantioselective Biocatalysts for Organic Chemistry by In Vitro Evolution. *Angew. Chem. Int. Ed. Engl.* *36*, 2830–2832.

Reetz, M.T., Kahakeaw, D., and Lohmer, R. (2008). Addressing the Numbers Problem in Directed Evolution. *ChemBioChem* *9*, 1797–1804.

Romero, P.A., and Arnold, F.H. (2009). Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* *10*, 866–876.

Rosenbaum, D.M., Rasmussen, S.G.F., and Kobilka, B.K. (2009). The structure and function of G-protein-coupled receptors. *Nature* *459*, 356–363.

Röthlisberger, D., Khersonsky, O., Wollacott, A.M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J.L., Althoff, E.A., Zanghellini, A., Dym, O., Albeck, S., Houk, K.N., Tawfik, D.S., and Baker, D. (2008). Kemp elimination catalysts by computational enzyme design. *Nature* *453*, 190–195.

Rydzik, A.M., Leung, I.K.H., Kochan, G.T., Thalhammer, A., Oppermann, U., Claridge, T.D.W., and Schofield, C.J. (2012). Development and Application of a Fluoride-Detection-Based Fluorescence Assay for γ -Butyrobetaine Hydroxylase. *ChemBioChem* *13*, 1559–1563.

Saedler, H., and Reif, H.J. (1974). IS2, A Genetic Element for Turn-off and Turn-on of Gene Activity in *E. coli*. *Molec. Genet.* *265*–289.

Salowe, S.P., Marsh, E.N., and Townsend, C.A. (1990). Purification and characterization of clavaminate synthase from *Streptomyces clavuligerus*: an unusual oxidative enzyme in natural product biosynthesis. *Biochemistry (Mosc.)* *29*, 6499–6508.

Sezonov, G., Joseleau-Petit, D., and D'Ari, R. (2007). *Escherichia coli* Physiology in Luria-Bertani Broth. *J. Bacteriol.* *189*, 8746–8749.

Shen, L., Song, C.-X., He, C., and Zhang, Y. (2014). Mechanism and function of oxidative reversal of DNA and RNA methylation. *Annu. Rev. Biochem.* *83*, 585–614.

Shivange, A.V., Marienhagen, J., Mundhada, H., Schenk, A., and Schwaneberg, U. (2009). Advances in generating functional diversity for directed protein evolution. *Curr. Opin. Chem. Biol.* *13*, 19–25.

- Siegel, J.B., Zanghellini, A., Lovick, H.M., Kiss, G., Lambert, A.R., St.Clair, J.L., Gallaher, J.L., Hilvert, D., Gelb, M.H., Stoddard, B.L., Houk, K.N., Michael, F.E., and Baker, D. (2010). Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction. *Science* *329*, 309–313.
- Smirnov, S.V., Kodera, T., Samsonova, N.N., Kotlyarova, V.A., Rushkevich, N.Y., Kivero, A.D., Sokolov, P.M., Hibi, M., Ogawa, J., and Shimizu, S. (2010). Metabolic engineering of *Escherichia coli* to produce (2S, 3R, 4S)-4-hydroxyisoleucine. *Appl. Microbiol. Biotechnol.* *88*, 719–726.
- Spudich, J.A. (2001). The myosin swinging cross-bridge model. *Nat. Rev. Mol. Cell Biol.* *2*, 387–392.
- Sternер, R., Merkl, R., and Raushel, F.M. (2008). Computational Design of Enzymes. *Chem. Biol.* *15*, 421–423.
- Stricker, M., Kopp, F., Mahlert, C., Essen, L.-O., and Marahiel, M.A. (2007). Mechanistic and Structural Basis of Stereospecific C β -Hydroxylation in Calcium-Dependent Antibiotic, a Daptomycin-Type Lipopeptide. *ACS Chem. Biol.* *2*, 187–196.
- Szentirmai, A., and Horváth, I. (1976). Regulation of branched-chain amino acid biosynthesis. *Acta Microbiol. Acad. Sci. Hung.* *23*, 137–149.
- Taabazuing, C.Y., Hangasky, J.A., and Knapp, M.J. (2014). Oxygen sensing strategies in mammals and bacteria. *J. Inorg. Biochem.* *133*, 63–72.
- Taddei, F., Radman, M., Maynard-Smith, J., Toupance, B., and al, et (1997). Role of mutator alleles in adaptive evolution. *Nature* *387*, 700–702.
- Tenaillon, O., Barrick, J.E., Ribeck, N., Deatherage, D.E., Blanchard, J.L., Dasgupta, A., Wu, G.C., Wielgoss, S., Cruveiller, S., Médigue, C., Schneider, D., and Lenski, R.E. (2016). Tempo and mode of genome evolution in a 50,000-generation experiment. *Nature* *536*, 165–170.
- Tolonen, A.C., and Haas, W. (2014). Quantitative Proteomics Using Reductive Dimethylation for Stable Isotope Labeling. *JoVE J. Vis. Exp.* e51416–e51416.
- Tolonen, A.C., Haas, W., Chilaka, A.C., Aach, J., Gygi, S.P., and Church, G.M. (2011). Proteome-wide systems analysis of a cellulosic biofuel-producing microbe. *Mol. Syst. Biol.* *7*, 461.
- Tolonen, A.C., Zuroff, T.R., Mohandass, R., Boutard, M., Cerisy, T., and Curtis, W.R. (2015). Physiology, genomics, and pathway engineering of an ethanol-tolerant strain of *Clostridium phytofermentans*. *Appl. Environ. Microbiol.* *AEM.00619-15*.
- Wanders, R.J.A., Komen, J., and Ferdinandusse, S. (2011). Phytanic acid metabolism in health and disease. *Biochim. Biophys. Acta* *1811*, 498–507.
- Warnick, T.A., Methé, B.A., and Leschine, S.B. (2002). *Clostridium phytofermentans* sp. nov., a cellulolytic mesophile from forest soil. *Int. J. Syst. Evol. Microbiol.* *52*, 1155–1160.

Weinreich, D.M., Watson, R.A., and Chao, L. (2005). Perspective: Sign epistasis and genetic constraint on evolutionary trajectories. *Evol. Int. J. Org. Evol.* *59*, 1165–1174.

Williams, A.B. (2014). Spontaneous mutation rates come into focus in *Escherichia coli*. *DNA Repair* *24*, 73–79.

Wong, T.S., Zhurina, D., and Schwaneberg, U. (2006). The Diversity Challenge in Directed Protein Evolution. *Comb. Chem. High Throughput Screen.* *9*, 271–288.

Wu, L.-F., Meng, S., and Tang, G.-L. (2016). Ferrous iron and α -ketoglutarate-dependent dioxygenases in the biosynthesis of microbial natural products. *Biochim. Biophys. Acta* *1864*, 453–470.

Yamamoto, N., Nakahigashi, K., Nakamichi, T., Yoshino, M., Takai, Y., Touda, Y., Furubayashi, A., Kinjyo, S., Dose, H., Hasegawa, M., Datsenko, K.A., Nakayashiki, T., Tomita, M., Wanner, B.L., and Mori, H. (2009). Update on the Keio collection of *Escherichia coli* single-gene deletion mutants. *Mol. Syst. Biol.* *5*, 335.

Yang, W. (2000). Structure and function of mismatch repair proteins. *Mutat. Res. Repair* *460*, 245–256.

Zhao, K., Liu, M., and Burgess, R.R. (2007). Adaptation in bacterial flagellar and motility systems: from regulon members to “foraging”-like behavior in *E. coli*. *Nucleic Acids Res.* *35*, 4441–4452.

Ziegler, K., Noble, S.M., Mutumanje, E., Bishop, B., Huddler, D.P., and Born, T.L. (2007). Identification of Catalytic Cysteine, Histidine, and Lysine Residues in *Escherichia coli* Homoserine Transsuccinylase. *Biochemistry (Mosc.)* *46*, 2674–2683.

Ziv, N., Brandt, N.J., and Gresham, D. (2013). The Use of Chemostats in Microbial Systems Biology. *J. Vis. Exp. JoVE*.

Zubieta, C., Arkus, K.A.J., Cahoon, R.E., and Jez, J.M. (2008). A Single Amino Acid Change Is Responsible for Evolution of Acyltransferase Specificity in Bacterial Methionine Biosynthesis. *J. Biol. Chem.* *283*, 7561–7567.

Titre : Optimisation de catalyseurs enzymatique par évolution dirigée *in vivo*

Mots clés : Evolution dirigée, Dioxygénase, GM3, α -KAO

RÉSUMÉ: De nombreuses méthodes *in vitro* d'évolution d'enzymes existent mais elles sont souvent coûteuses en temps et en main d'œuvre. Une approche alternative, l'évolution dirigée *in vivo*, permet la sélection de variantes d'enzymes améliorées ou modifiées à moindre coût et peut être appliquée à des enzymes de structure inconnue. Des dispositifs automatisés de culture continue (GM3), ont été développés et construits au Genoscope. Le GM3 a été conçu pour maintenir indéfiniment des cultures bactériennes en suspension tout en les soumettant à des régimes de culture sélectifs de type turbidostat et chémostat. Le principal objectif du présent travail était l'optimisation par évolution dirigée *in vivo* d'enzymes de la famille des dioxygénases à α -cétoglutarate et fer-dépendantes (α -KAO). Les α -KAOs catalysent l'hydroxylation d'un atome de carbone non substitué et sont donc d'intéressants outils pour la fonctionnalisation de molécules en synthèse organique. La dynamique de fixation d'une mutation bénéfique dans une population en croissance en fonction du régime de culture et du format d'expression a été étudiée. Des souches *d'Escherichia coli* ont été génétiquement modifiées pour que leur croissance dépende strictement d'une activité α -KAO. Des populations de ces bactéries exprimant une α -KAO ont été soumises à des protocoles d'évolution dirigée *in vivo* pour sélectionner: dans un premier temps, des variants d'expression optimisée sur le substrat naturel de l'enzyme, dans un second temps, des variants utilisant des analogues de substrat non naturels. Deux variants de la L-isoleucine dioxygénase (Ido:E65K, Ido:A62P) ont été obtenus. L'efficacité catalytique de Ido:E65K vers la norvaline a été augmentée d'un facteur quatre par rapport à celle de Ido sauvage. En collaboration avec l'équipe du Dr. Tolonen, la culture continue a été appliquée à l'évolution de la bactérie cellulolytique *Clostridium phytofermentans*. Cette bactérie anaérobie stricte qui dégrade la lignocellulose en éthanol et H₂ présente un intérêt applicatif pour la valorisation de la biomasse. La dégradation de la lignocellulose libère des composés phénoliques, en particulier l'acide férulique. *C. phytofermentans* a été adapté à résister à des concentrations croissantes de ce composé inhibiteur. Des isolats résistant au ferulate ont été analysés et leur génomes séquencés (Cerisy *et al.*, 2017).

Title : Optimization of enzymes via Directed Evolution *in vivo*

Keywords : Directed evolution *in vivo*, Dioxygenase, GM3, α -KAO

ABSTRACT: Numerous methods of directed evolution *in vitro* have been developed. These methods are often expensive and time-consuming. An alternative approach, directed evolution *in vivo*, allows the selection of protein variants at lower cost and does not require prior knowledge of the protein's structure and mechanism. An automated continuous culture device, the GM3, has been developed and built at the Genoscope. The GM3 technology enables the maintenance of steady-state microbial growth over an extended period of time.

The main objective of the present study was the optimization by directed evolution *in vivo* of enzymes of the family of the Iron(II)/ α -ketoacid-dependent oxygenases (α -KAOs). The α -KAOs catalyze the hydroxylation of non-activated C-H bonds and thus are interesting tools for the functionalization of organic molecules in synthetic chemistry.

The dynamics of fixation of beneficial mutations in a growing cell population depending on the expression format and the culture regime has been explored.

Strains of *Escherichia coli* have been genetically modified such that their growth

strictly depended on a α -KAO activity. Populations of these bacteria expressing a α -KAO were subjected to protocols of directed evolution in continuous culture in order to select, in a first round, variants showing optimized expression on natural substrates of the enzyme, and in a second round, variants with activity towards non-natural analogs of the substrate. Two variants of L-isoleucine dioxygenase (Ido:E65K, Ido:A62P) were obtained. The catalytic efficiency of Ido:E65K was found to be four times higher for norvaline as compared to Ido WT.

In collaboration with the group of Dr. Tolonen, directed evolution in continuous culture was applied to the cellulolytic bacteria *Clostridium phytofermentans*. This anaerobic bacterium degrades lignocellulose into ethanol and H₂ and can thus be applied for biomass valorization. The degradation of lignocellulose liberates phenolic compounds like ferulic acid known to be toxic to the bacteria. *C. phytofermentans* was adapted in the GM3 to resist to growing concentrations of this inhibitor. Resistant bacteria were analyzed and their genomes sequenced (Cerisy et al., 2017).